

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Mouloud Mammeri de Tizi-Ouzou

Faculté des Sciences Economiques, Commerciales et des

Sciences de Gestion

Département des Sciences Financières et Comptabilité



Mémoire

En vue de l'obtention du diplôme de Master en
sciences financières et comptabilité

Option : Finance d'entreprise

Sujet

La finance d'entreprise à l'heure de la *data science*

**Cas : Gestion de portefeuille avec R et
Python Bourse d'Alger**

Réalisé par

IRBOUH Mélina

HARIRECHE Celina

Devant le jury composé de

Présidente : Mme MEKACHER Amal ,

Maitre conférence B, UMMTO

Rapporteur : Mr GHEDDACHE Lyes ,

Professeur , UMMTO

Examineur : Mr YAMOUN Karim ,

Maître assistant , UMMTO

Promotion : Juin 2023

Remerciements

Avant toute choses nous remercions Dieu le tout puissant de nous avoir donné le courage, la volonté et la patience tout au long de notre cursus.

Nous tenons à exprimer notre sincère gratitude envers notre promoteur M. Gheddache Lyes, pour son accompagnement tout au long de notre projet. Votre soutien, votre disponibilité et votre gentillesse ont été d'une grande valeur pour nous. Nous vous adressons nos plus vifs remerciements pour vos remarques constructives qui ont contribué à l'amélioration de notre travail.

De plus, nous souhaitons également exprimer notre profonde reconnaissance envers le personnel de la bourse pour leur accompagnement chaleureux et leur précieuse assistance. Particulièrement Mme Talhi qui nous a accompagnées avec persévérance tout au long de ce projet. Sa gentillesse, sa patience et son soutien inébranlable ont été d'une valeur inestimable pour nous. Nous sommes extrêmement reconnaissantes envers elle pour son dévouement et sa contribution à notre réussite.

Nous tenons aussi à remercier les membres de jury pour avoir accepté d'examiner et évaluer ce mémoire de fin d'études.

Enfin, nos profonds remerciements vont également à toutes les personnes qui nous ont aidées et soutenues de près ou de loin pour la réalisation de ce travail.

Dédicaces

Je dédie ce modeste mémoire réalisé grâce au bon Dieu :

A ma chère maman

Je suis incapable de trouver les mots justes pour exprimer pleinement l'affection et l'amour que je ressens envers toi. Puissent tes soutiens et tes sacrifices être récompensés par ce travail. Que Dieu veille sur toi, t'accorde santé et bonheur, et que tu demeures toujours l'éclat de ma vie

A mon cher père

Il est impossible de trouver les mots justes pour exprimer pleinement la gratitude et la reconnaissance que j'ai envers toi. Tu as toujours été à mes côtés, me soutenant et m'épaulant sans relâche. Ce mémoire est dédié à toi, en témoignage de ton dévouement et de mon amour profond.

Je tiens aussi à remercier mes frères Rayan et Danny ainsi que ma petite sœur Kamilia puisse dieu vous accorder santé, bonheur et longue vie et faire en sorte que jamais je ne vous déçoive

Je tiens aussi à remercier ma tante Ouerdia pour ses engagements et pour toujours avoir été là pour moi.

Et enfin à toutes personnes qui m'aiment.



Dédicace

En guise de reconnaissance, je dédie ce travail

À mes très chers parents :

Je vous offre ce mémoire, à vous qui avez été mes sources constantes de soutien, d'amour et d'inspiration. Votre présence inconditionnelle tout au long de ma vie et de mes études a été la clé de ma réussite. Vos encouragements constants, votre confiance en moi et votre croyance en mes capacités ont été des piliers essentiels qui m'ont permis de franchir chaque étape avec détermination. Que le Tout-Puissant vous accorde santé, bonheur et une longue vie, afin que je puisse vous rendre à mon tour tout ce que vous avez fait pour moi.

A mes sœurs : Lilia, Amel et Melina & mon frère Amine :

Je ne saurais envisager ma vie sans votre présence. Je tiens à vous exprimer ma profonde gratitude pour votre encouragement tout au long de mon parcours. Que ce travail soit l'expression de mon estime pour vous et que Dieu vous accorde, bonheur, réussite dans votre vie.

A mon cher confident : Boualem .K

Pour l'amour qui nous unisse, je ne saurais exprimer ma profonde reconnaissance pour le soutien continu dont tu as toujours fait preuve. Tu as toujours été là pour m'encourager et me pousser à donner le meilleur de moi-même.

A mes chères amies Horia et Sabrina :

Je vous dédie ce travail en reconnaissance de votre soutien et encouragements, que vous m'offrez quotidiennement.

HARIRECHE Celina

- ACM** : Association for Computer Machinery
- ALL** : Alliance Assurance
- API** : Application Programming Interface (interface de programmation)
- AR** : modèles autorégressifs
- AUR** : Aurassi
- BADR** : Banque D'agriculture et De Développement Rural
- BDL** : Banque De Développement Local
- BEA** : Banque Extérieure d'Algérie
- BI** : Business Intelligence
- BIO** : BIOPHARM
- BNA** : Banque Nationale d'Algérie
- BOC** : Bulletin Officiel de la Cote
- CAPM** : *Capital Asset Pricing Model* (Modèle d'évaluation des actifs financiers)
- CARN** : *The Comprehensive R Archive Network*
- CNEP** : Caisse Nationale d'épargne et de Prévoyance-Banque
- Cov** : Covariance
- CPA** : Crédit Populaire d'Algérie
- CSV** : *comma-separated values* (valeurs séparé par une virgule)
- CVAR** : *Conditional Value at Risk* (Valeur conditionnelle à risque)
- DBSCAN** : *Density-Based Spatial Clustering of Applications with Noise*
- DM** : *Digital Marketing*
- Doc** : Documents
- DS** : *Data Science*
- ES** : *Expected Shortfall*
- FCP** : Fonds Communs de Placement
- FFT** : *Fast Fourier Transform* (La Transformation de Fourier Rapide)
- GFS** : *Google File System*
- HDFS** : *Hadoop Distributed File System*
- HFT** : High-frequency trading (Transactions à haute fréquence)
- IA** : Intelligence Artificielle
- IBM** : International Business Machines
- IDE** : Integrated development environment (Environnement de développement intégré)
- IOB** : Intermédiaires en Opération de Bourse

IoT : *Internet of things*

KPI : Indicateurs clés de performance

MA : Les modèles à moyenne mobile

ML : *Machine Learning* (Apprentissage automatique)

NDFS : *Nutch distributed File system*

OAT : Obligations Assimilables Au Trésor

OLTP : online transactional processing (traitement des transactions en ligne)

OPCVM : Organismes de Placements Collectifs en Valeurs Mobilières

PME : Petit et Moyenne Entreprises

RDBMS : *Relational Data Base Management System* (système de gestion de base de données relationnelle)

RGPD : Règlement Général sur la Protection des Données

Rm : rentabilité d'un marché

Rp : Rentabilité d'un portefeuille

SAI : SAIDAL

SAP : Systems, Applications & Products in data processing

SAS : *Statistical Anaslysis System*

SGVB : La Société de Gestion de la Bourse des Valeurs Mobilières

SICAV : Sociétés d'Investissement à Capital Variable

SQL : Structured Query Language

SQOOP : SQL to HADOOP

StdDev : *Standard Deviation*

SVM : Support Vector Machine

TTR: *Technical Trading Rules*(règles techniques de négociation)

VaR : Value at Risk (Valeur à risque)

Var : Variance

XTS : *eXtensible Time Series* (séries temporelles extensibles)

Yarn : *(Yet Another Resource Negot*

Listes des tableaux et figures

Figure 1 les 5 V de la Big Data.....	14
Figure 2 Données structuré Vs données non-structuré.....	17
Figure 3 Business Intelligence VS Data Science	22
Figure 4 Processus d'un traitement MapReduce	33
Figure 5 L'écosystème Hadoop	35
Figure 6 Compilation/Interprétation Python.....	42
Figure 7 les tops 10 des langages de programmation selon TIOBE index	43
Figure 8 TIOBE Index Classement des langages sur la base de divers moteurs de recherche	56
Figure 9 installation du logiciel R	56
Figure 10 Environnement R.....	57
Figure 11 Cours des actions AAPL.....	62
Figure 12 Graphe des cours d'action AAPL.....	65
Figure 13 Le système Algérien mis en place pour l'organisation du marché boursier	90
Figure 14 l'acquisition et la vente de titres.....	93
Figure 15 Déroulement de la séance de cotation à la Bourse d'Alger	95
Figure 16 Définitions des variables et bibliothèques.....	96
Figure 17 Lecture des fichiers et du portefeuille	97
Figure 18 Calcul du rendement du portefeuille.....	98
Figure 19 Rendements journaliers des actions.....	99
Figure 20 Espérance de rendement du portefeuille	100
Figure 21 Rendement et risque annuel du portefeuille.....	101
Figure 22 rendement mensuelle et annuelle du portefeuille	102
Figure 23 Graphique des rendements du portefeuille.....	102
Figure 24 Calcul des ratios	104
Figure 25 Covariance des rendements de portefeuille.....	105
Figure 26 Sharpe ratio	106
Figure 27 Calcule du CAPM Beta.....	107
Figure 28 Importation des fichier et assignation des variables	111
Figure 29 Importation des prix du portefeuille Python	112
Figure 30 les rendements journaliers des actions Python.....	113
Figure 31 Rendements journaliers du portefeuille Python	113
Figure 32 Espérance de rendement des actifs	114

Listes des tableaux et figures

Figure 33 Calcul des ratios Python.....	115
Figure 34 Covariance des actifs	115
Figure 35 Fonction sharpe ratio Python	116
Figure 36 Calcule du rendement espéré avec la méthode CAPM.....	117

Listes des tableaux

Tableau 1 Classification des technologie de la Data Science	22
Tableau 2 : Comparaison entre Python et R.....	74

Sommaire

Remerciements	I
Liste des abréviations	IV
Liste des tableaux et figures	VII
Sommaire	VIII
Introduction générale	1

Chapitre I : Le Big Data et la *Data Science* : Un domaine en plein essor

Introduction	05
Section 1 : Qu'est ce que le <i>Big Data et la Data science</i>	06
Section 2 : <i>Business intelligence VS Data science</i>	18
Section 3 : Les grands défis en matière de <i>Big Data</i> dans la finance.....	21
Conclusion	22

Chapitre II : Les technologies de la *data science* applique à la finance

Introduction	26
Section 1 : Apache Hadoop.....	27
Section 2 Le logiciel « Python ».....	37
Section 3 Le logiciel« R »	51
Conclusion	68

Chapitre III : Les Uses Cases de la *Data science* en finance

Introduction	70
Section 1 : Applications de la <i>data science</i> pour l'optimisation des processus et la croissance des entreprises	70
Section 2 : la <i>Data science</i> pour les entreprise moderne et Startups.....	78
Conclusion	81

Chapitre IV : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

Introduction.....	83
-------------------	----

Section 01 : Présentation du marché financier algérien SGBV	84
Section 02 : Gestion de portefeuille avec R studio.....	93
Section 03 :Gestion de portefeuille avec Spyder (Python)	107
Conclusion	114
Conclusion générale.....	116

Bibliographie

Annexes

Table des matières

Introduction générale

Au cours des dernières décennies, les progrès rapides des technologies de l'information et des communications ont donné lieu à une explosion de la quantité d'informations disponibles, marquant ainsi l'avènement de l'ère des mégadonnées. Chaque jour, des milliards d'octets de données sont générés à partir de diverses sources telles que les capteurs, les appareils mobiles, les transactions en ligne et les réseaux sociaux. Cet énorme effort a conduit à la naissance d'un nouveau domaine interdisciplinaire appelé "science des données", dont les principes et les techniques visent à l'extraction automatique d'informations et de connaissances potentiellement utiles à partir des données.

Le secteur financier mondial, qui est l'une des industries les plus lucratives au monde, a été l'un des premiers à identifier le potentiel de profit des données massives (big data). Cela n'a rien de surprenant : les investisseurs ont toujours cherché à prédire les évolutions du marché afin de réaliser les meilleurs investissements et d'acquérir un avantage concurrentiel.

Cependant, la quantité de données collectées par le secteur financier pose un problème . L'un des plus grands problèmes est de savoir comment utiliser au mieux toutes ces données non structurées (c'est-à-dire des données complètement désorganisées et dépourvues de modèle cohérent) pour en tirer profit .

Face à cette réalité, une question essentielle se pose : comment ces nouvelles technologies vont-elles concrètement impacter le monde de la finance ? Il est clair que les technologies Big Data et la Data Science auront un impact significatif, mais il est crucial de comprendre en quoi elles façonneront concrètement le futur de la finance moderne. Ainsi, cette problématique devient la ligne directrice de notre travail : **Comment les technologies Big Data et la Data Science peuvent-elles façonner le future de la finance moderne notamment dans le contexte du marché financier algérien ?**

Dans le but de traiter notre problématique, il est utile de répondre aux questions suivantes :

- 1- En quoi consiste le Big data et la Data science ?
- 2- Quels sont les outils et méthodes de la data sciences utilisées dans la finance moderne ?
- 3- Comment la science des données peut-elle être appliquée aux différents domaines de la finance ?
- 4- Comment les logiciels R et Python peuvent-ils aider à analyser un portefeuille d'actions cotées à la Bourse algérienne de manière approfondie ?

Les hypothèses

Afin de répondre aux questions posées, nous nous sommes appuyées sur les hypothèses suivantes

- 1- La data science représente une avancée technologique majeure pour la finance moderne, offrant un potentiel considérable pour l'avenir du secteur.
- 2- L'exploitation de la science des données et des algorithmes de la big data permet une évaluation plus précise des portefeuilles d'actifs, offrant ainsi des améliorations significatives en termes d'efficacité et de pertinence dans ce domaine.

Choix et intérêt du sujet :

Nous avons choisi ce thème pour diverses raisons :

- Sa relation avec le domaine de notre spécialité (finance d'entreprise) ;
- Un Sujet qui n'a jamais été traité auparavant au sein de notre université
- Combiner nos connaissances théoriques déjà acquises aux nouvelles technologies modernes
- Cette discipline offre un avenir prometteur et intéressant en raison de l'émergence du domaine à l'échelle internationale
- Comprendre l'importance de la gestion de portefeuille pour garantir le profit des investisseurs

Méthodologie de recherche :

Afin de mener à bien notre recherche et répondre à ses préoccupations, nous avons opté pour la méthode d'étude de cas que nous trouvons plus adéquate pour ce type de sujet.

Dans un premier lieu, nous avons privilégié la recherche bibliographique basée sur la consultation des ouvrages, documents, rapports, Sites internet, mémoires relatifs à notre thématique.

Une méthode analytique qui en premier lieu consiste à acquérir les bases et les fondements de l'utilisation des logiciels R et Python, ensuite la création d'un algorithme pour l'optimisation du portefeuille

Une méthode comparative qui consiste à comparer les résultats obtenus par les deux algorithmes des deux différents logiciels

Structure de la recherche :

Le travail est structuré en quatre parties, chacune ayant son propre axe de développement. Chapitre 1 : Nous débutons par une vue générale du sujet, où nous avons définissons les termes "Big Data" et "Data Science", ainsi que le contexte dans lequel ces technologies émergent. Il est essentiel de garder à l'esprit certains éléments lors de notre exploration de ces nouveaux domaines. Cette première partie vise également à fournir aux lecteurs des clés de compréhension pour mieux appréhender les départements de Data Science, qui émergent de plus en plus dans de nombreuses entreprises, y compris dans les services financiers. Il est crucial de définir le contexte et les rôles de ces équipes, car elles sont au cœur de cette évolution.

Chapitre 2 : Nous abordons un aspect plus technique de notre travail en nous concentrant sur la définition des technologies utilisées dans notre développement informatique. Nous justifions leur apparition et expliquons pourquoi elles sont les mieux adaptées pour exploiter le Big Data.

Ensuite, le travail se tourne vers le deuxième volet : la finance. Nous examinons l'environnement technologique actuel de la finance et explorons pourquoi le Big Data s'impose dans ce domaine. Nous décrivons également les types de données exploitables dans les services financiers et les analyses à la lumière des concepts présentés dans notre première partie. Une sous-partie importante est consacrée à une dizaine de cas d'utilisation de la Data Science à la finance. Ces exemples apportent des éléments de réponse essentiels à notre problématique. Quelles sont les opportunités offertes par le Big Data dans le domaine de la finance ? Bien que les cas d'utilisation soient présentés de manière concise, nous avons tenté d'englober les principes les plus importants tels que le trading, le marketing, la gestion de portefeuille, la gestion des risques et la détection de fraudes au sein des entreprises.

Le quatrième et dernier chapitre se concentre sur une analyse du risque et du rendement d'un portefeuille d'investissement, ainsi que sur sa position sur le marché des titres à revenu fixe, formulée à l'aide des langages de programmations R et Python. Nous déterminons ainsi la position de l'investisseur sur le marché.

Enfin, le travail s'achève par une conclusion générale qui va permettre d'affirmer ou d'infirmer les hypothèses de départ.

Chapitre 1
Le Big Data et la Data Science
Un domaine en plein essor

Chapitre 1 : Le Big Data et la Data Science : Un domaine en plein essor

Introduction

Avec l'évolution constante des systèmes informatiques, les entreprises et les industries sont confrontées à une quantité de données sans précédent. Les prévisions de croissance des volumes de données dépassent désormais les limites des technologies traditionnelles. Ce flux continu de données massives, connu sous le nom de big data, est devenu une ressource essentielle pour les organisations cherchant à améliorer leur efficacité, leur prise de décisions et leur avantages concurrentiel.

Parmi les domaines qui ont embrassé cette révolution des données, la finance se distingue en tant que secteur où le big data et la data science ont connu une adoption rapide et un impact significatif.

Dans ce chapitre, nous allons explorer les généralités du big data et de la data science, ainsi que leur impact progressif dans la finance. Le Big data fait référence à l'ensemble des données volumineuses, variées et générées à grande vitesse, qui dépassent la capacité des outils traditionnels de gestion et d'analyse. Ces données peuvent provenir de sources telles que les médias sociaux, les transactions financières, les capteurs, les dispositifs connectés, et bien d'autres encore.

Le chapitre est divisé en 3 sections :

Section 1 : Qu'est ce que le *Big Data et la Data science*

Section 2 : *Business intelligence VS Data science*

Section 3 : Les grands défis en matière de *Big Data* dans la finance

Chapitre 1 : Le Big Data et la Data Science : Un domaine en plein essor

Section 1 : Qu'est ce que le *Big Data* et la *Data science* ?

1.1. Historique et origine de l'évolution du *Big Data*

Le *Big Data* est un terme vaste et flou pour une grande partie des professionnels et particuliers. Derrière ce concept se cache un bon nombre d'idées : l'explosion des données (volume, variété et vélocité), la potentialité de valeur que l'on peut en tirer, sans oublier bien sûr les technologies pour les traiter. Pour comprendre ces concepts et imaginer leurs développements futurs, il est indispensable d'en connaître l'origine et l'évolution. Revenons brièvement sur la petite histoire – à forts rebondissements – du *Big Data* .

Le concept de "*Big Data* " est relativement récent et est apparu pour la première fois dans les années 1990. Cependant, la notion de stockage et de traitement de grandes quantités de données remonte beaucoup plus loin. Des papyrus utilisés par les scribes de l'Égypte ancienne il y a plus de 3000 ans, en passant par les tablettes d'argile mésopotamiennes, jusqu'aux bases de données et logiciels informatiques que l'on connaît aujourd'hui : la donnée financière a toujours su épouser, à chaque époque, le format qui lui allait le mieux.

Les premières traces de l'humanité stockant et analysant des données datent de 18 000 ans avant notre ère. L'os d'Ishango a été découvert en 1960 dans ce qui est maintenant connu comme l'Ouganda et est considéré comme l'une des plus anciennes preuves de stockage de données préhistoriques. Les membres des tribus du paléolithique marquaient des encoches sur des bâtons ou des os pour suivre les activités commerciales ou les approvisionnements. Ils comparaient les bâtons et les encoches pour effectuer des calculs rudimentaires, ce qui leur permettait de prédire, par exemple, la durée de vie de leurs réserves de nourriture.

Vers 300 av. J.-C. Les anciens Égyptiens ont déjà essayé de capturer toutes les "données" existantes dans la bibliothèque d'Alexandrie. De plus, l'Empire romain avait l'habitude d'analyser attentivement les statistiques de leur armée pour déterminer la distribution optimale pour leurs troupes.¹

Le développement de la machine de lecture de cartes perforées par IBM pour l'Administration de la sécurité sociale dans les années 1930 est un exemple significatif de l'utilisation de la technologie de traitement des données au 20^{ème} siècle. La loi sur la sécurité

¹ GEORGE Firican "The history of *Big Data* ", light on Data ,2022

Chapitre 1 : Le Big Data et la Data Science : Un domaine en plein essor

sociale de 1935 a établi un programme pour fournir des prestations de retraite aux travailleurs américains, ce qui a nécessité que le gouvernement suive les contributions et l'éligibilité aux avantages à grande échelle. La machine de lecture de cartes perforées d'IBM a permis au gouvernement de traiter et d'analyser de grandes quantités de données rapidement et avec précision, ce qui était essentiel pour l'administration du programme de sécurité sociale. Ce projet est un exemple précoce de la façon dont les technologies de données massives peuvent être utilisées pour résoudre des problèmes complexes dans les gouvernements et les entreprises².

Un autre exemple illustrant l'utilisation de la *Big Data* remonte à l'an 1943 lorsque les anglais ont développé la première machine de traitement de l'information pour déchiffrer les codes nazis pendant la Seconde Guerre mondiale . Cet appareil, baptisé Colossus, recherchait des motifs dans les messages interceptés à une vitesse de 5 000 caractères par seconde, réduisant ainsi le temps nécessaire à cette tâche de plusieurs semaines à quelques heures seulement.

Au tournant du siècle, les géants de la technologie comme Google, Yahoo! et Amazon ont commencé à collecter et à stocker des quantités massives de données sur les comportements des utilisateurs et les activités en ligne. Ces données ont été utilisées pour améliorer les résultats de recherche, recommander des produits et services, et cibler la publicité en ligne.

En 2005, Roger Mougala a donné le nom de *Big Data* à un vaste ensemble de données qui, à l'époque, étaient presque impossibles à gérer et à traiter à l'aide des outils de veille stratégique traditionnels disponibles ce qui a fait de lui le « père fondateur » de la *Big Data* . La même année, Hadoop, qui pouvait traiter les *Big Data* , a été créé. Hadoop était basé sur un cadre logiciel en libre accès appelé Nutch et a été fusionné avec MapReduce de Google.

Les *Big Data* marquent le début d'une nouvelle révolution. À l'avenir, et plus tôt que nous ne le pensons, de nombreux aspects de notre monde seront augmentés ou remplacés par des systèmes informatiques qui sont aujourd'hui l'apanage du jugement humain. Non seulement la conduite ou le matchmaking, mais même des tâches plus complexes. Après tout, Amazon peut recommander le livre idéal, Google peut classer les sites web les plus pertinents, Facebook connaît nos préférences et LinkedIn devine qui nous connaissons. Les mêmes technologies

² The Age of *Big Data* " James Manyika, Michael Chui et Brad Brown

Chapitre 1 : Le Big Data et la Data Science : Un domaine en plein essor

seront appliquées au diagnostic des maladies, à la recommandation de traitements, peut-être même à l'identification de "criminels" avant qu'ils ne commettent réellement un crime. Tout comme Internet a radicalement changé le monde en ajoutant la communication aux ordinateurs, les *Big Data* changeront également des aspects fondamentaux de la vie en lui donnant une dimension quantitative qu'elle n'avait jamais eue auparavant.

Les *Big Data* seront aussi une source de nouvelle valeur économique et d'innovation. L'émergence de celle-ci a créé de nouvelles opportunités pour les entreprises d'extraire de la valeur et de générer de l'innovation à partir de la quantité immense de données générées chaque jour. Cette science a le potentiel de fournir des informations qui étaient auparavant impossibles à obtenir, permettant aux entreprises de prendre des décisions plus éclairées, d'optimiser leurs opérations et de créer de nouveaux produits et services

1.2. Définitions des termes et contexte

La première mention du terme "*Big Data*" est apparue dans la bibliothèque de l'Association for Computer Machinery (ACM) il y a plus de deux décennies. Michael Cox et David Ellsworth ont écrit : "La visualisation pose un défi intéressant pour les systèmes informatiques : les ensembles de données sont généralement très volumineux, ce qui sollicite les capacités de la mémoire principale, du disque local et même du disque distant. Nous appelons cela le problème des "*Big Data*". Lorsque les ensembles de données ne rentrent pas dans la mémoire principale (en mémoire centrale) ou lorsqu'ils ne rentrent même pas sur le disque local, la solution la plus courante est d'acquérir plus de ressources." En d'autres termes, à l'époque, la définition de *Big Data* était essentiellement "des données qui ne pouvaient plus tenir sur le matériel disponible".

En avançant rapidement de seulement plus de deux décennies - après l'explosion d'Internet, des smartphones, de l'Internet des objets et du *cloud computing* la définition de *Big Data* s'est étendue bien au-delà des limites d'un "disque local". Selon divers ressources on peut le définir comme :

"*Big Data* est un terme utilisé pour décrire des ensembles de données volumineux, complexes et variables qui nécessitent des outils avancés de traitement, de stockage et d'analyse pour en extraire des connaissances et des informations utiles"³ (McKinsey & Company).

³ Mc Kinsey Global Institute "*Big Data* : The next frontier for innovation, competition and productivity" –

Chapitre 1 : Le Big Data et la Data Science : Un domaine en plein essor

"Le terme *Big Data* fait référence à des données volumineuses, complexes et diverses qui ne peuvent pas être facilement traitées à l'aide d'outils de gestion de base de données traditionnels⁴" (Forbes).

"Le *Big Data* sont des informations très volumineuses, très rapides et/ou très variées qui exigent des formes rentables et innovantes de traitement de l'information permettant d'améliorer la compréhension, la prise de décision et l'automatisation des processus.⁵" (Gartner).

"Les *Big Data* sont des données complexes et variées générées à partir de sources multiples, qui ne peuvent pas être traitées efficacement à l'aide d'outils traditionnels de gestion de base de données⁶" (IBM).

"*Big Data* est un terme qui décrit les grandes quantités de données - structurées et non structurées qui inondent une entreprise au quotidien" (SAS).

"Le terme "*Big Data* " désigne des ensembles de données volumineux et complexes que les outils traditionnels de traitement des données ne sont pas en mesure de traiter. " (TechTarget)

" Le *Big Data* désigne la quantité massive de données structurées et non structurées générées par les technologies numériques, telles que les médias sociaux, les smartphones, les dispositifs de l'internet des objets (IoT) et les capteurs... ⁷" (Oracle).

"Le terme "*Big Data* " est utilisé pour décrire des ensembles de données si volumineux et complexes qu'il devient difficile de les traiter à l'aide des outils traditionnels de gestion de bases de données. " (Microsoft).

Dans l'ensemble, la *Big Data* fait référence à des ensembles de données volumineux, variés et complexes, qui nécessitent des outils et des techniques avancés pour être collectés, stockés, gérés et analysés efficacement. Les données peuvent provenir de diverses sources, y compris des appareils IoT, des réseaux sociaux, des capteurs, des appareils mobiles, des bases de données, etc. La valeur de la *Big Data* réside dans sa capacité à fournir des informations précieuses et à aider les entreprises et particuliers à prendre des décisions plus éclairées.

⁴ <https://www.forbes.com/sites/lisaarthur/2013/08/15/what-is-big-data/>

⁵ <https://www.gartner.com/en/information-technology/glossary/big-data>

⁶ IT and data management research -Making sense of *Big Data* : A day in the life of an enterprise architect

⁷ <https://www.oracle.com/fr/big-data/what-is-big-data/>

1.3. Les 5 « V » de la Big Data

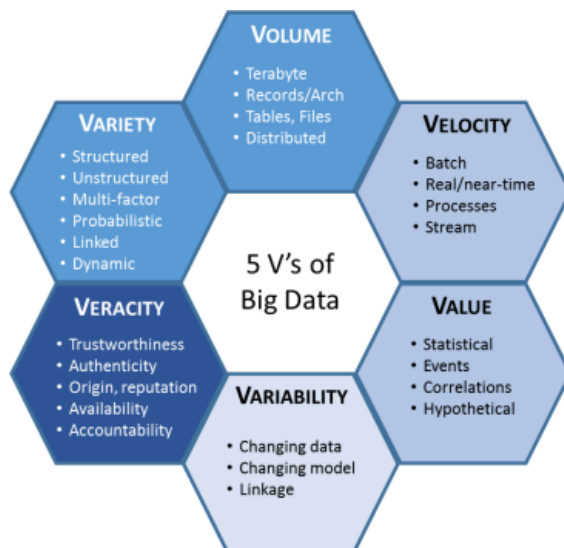
Le mythe des "3V" qui est l'origine des "5V" a été popularisé par Doug Laney , actuellement VP Analyst chez Gartner, En 2001 il publie un article décrivant le changement de paradigme qui s'opèrerait dans le « Data Management » des nouveaux E-Business. En deux pages, Doug Laney décompose la révolution des données en 3 V et a proposé que les données peuvent être caractérisées par trois dimensions : volume, variété et vélocité. Selon cette théorie, la valeur des données augmente avec leur volume (la quantité de données collectées), leur variété (la diversité des types de données collectées) et leur vélocité (la rapidité avec laquelle les données sont générées et doivent être traitées)⁸ .

Cependant, cette théorie a été critiquée pour être incomplète et ne pas prendre en compte d'autres dimensions importantes des données, telles que la qualité, la pertinence, la sécurité et la confidentialité. En outre, les avancées technologiques ont conduit à la création de nouveaux types de données, tels que les données géospatiales, les données de réseaux sociaux, les données IoT et les données de capteurs, qui ne sont pas couverts par les "3V" de Laney.

Par conséquent, bien que le concept des "3V" puisse être utile pour comprendre certaines caractéristiques des données, il est important de ne pas s'y limiter et de reconnaître que la valeur des données dépend de plusieurs facteurs qui vont au-delà de ces trois dimensions. C'est ainsi que la théorie des 5 V fut son apparition : le volume, la variété, la vélocité, la véracité et la valeur.

⁸ Regards croisés sur l'économie -Les 5V du *Big Data* Thomas Bourany 2018/2 (n° 23), pages 27 à 31

Figure 1 les 5 V de la Big Data



Source : *Exploring Techniques of Improving Security and Privacy in Big Data*, Bledi Shaqiri ,2017

D'après la figure ci-dessus, nous pouvons indiquer ce qui suit :

1.3.1 Le volume

C'est le premier V de la *Big Data*. Il se réfère à la quantité de données qui sont collectées et stockées. Avec l'explosion des technologies connectées, des réseaux sociaux et des appareils intelligents, la quantité de données générées par les entreprises et les particuliers a augmenté de manière exponentielle ces dernières années. La gestion de ces quantités massives de données est donc devenue un véritable défi pour les organisations, car il faut des outils et des infrastructures de stockage spécifiques pour gérer ces données. L'augmentation du volume de données est également liée à la nécessité de traiter des données en temps réel, ce qui nous amène à la deuxième V, la vélocité.

1.3.2. La vélocité

Elle se réfère à la vitesse à laquelle les données sont générées et traitées. Dans certains secteurs, comme les services financiers et les entreprises de commerce électronique, les données doivent être traitées en temps réel pour prendre des décisions critiques. Les entreprises doivent donc disposer de solutions qui leur permettent de traiter les données rapidement, avec une grande

Chapitre 1 : Le Big Data et la Data Science : Un domaine en plein essor

précision. Les technologies telles que le *cloud computing*, l'informatique en temps réel et les réseaux de diffusion de contenu sont utilisées pour répondre à ces besoins.

1.3.3. La variété

Elle fait référence aux différents types de données qui sont générées par les entreprises. Les données peuvent être structurées ou non structurées, provenir de différentes sources et avoir des formats différents. Par exemple, les entreprises peuvent collecter des données à partir de sources telles que les médias sociaux, les objets connectés, les capteurs, les images et les vidéos. Elles doivent donc être en mesure de traiter et d'analyser ces données variées pour en tirer des informations utiles.

1.3.4 La véracité.

Elle se réfère à la qualité des données collectées. Les données peuvent être erronées ou incomplètes, ce qui peut conduire à des décisions inexactes. Les entreprises doivent donc disposer d'outils pour nettoyer et préparer les données avant de les utiliser. Cela implique souvent de vérifier la source des données et de les comparer à d'autres sources pour s'assurer de leur exactitude.

1.3.5 La valeur

Elle fait référence à l'importance et à la pertinence des informations que les entreprises peuvent tirer de leurs données. Les données ont peu d'intérêt si elles ne sont pas exploitées pour prendre des décisions importantes. Les entreprises doivent donc être en mesure d'extraire des informations utiles de leurs données et de les utiliser pour prendre des décisions éclairées.

En conclusion, les 5V de la *Big Data* sont un concept important pour comprendre les défis et les opportunités offertes et jouent aussi un rôle crucial dans la manière dont les entreprises gèrent et analysent les données massives pour en tirer des insights significatifs et améliorer leurs performances.

1.4. Les Types de Données

Une structure de données englobe un ensemble de données accompagnées de leurs relations, ainsi que des fonctions ou des opérations permettant de les manipuler. Elle est utilisée

Chapitre 1 : Le Big Data et la Data Science : Un domaine en plein essor

pour organiser et stocker les données dans un ordinateur, facilitant leur accès et leur modification de manière efficace.

En ce qui concerne les Big Data, les données collectées, stockées et traitées proviennent de divers domaines et sources hétérogènes. Cela génère une grande quantité de données de types différents, à la fois structurées, non structurées et semi-structurées.⁹

1.4.1. Les données structurées

Les données structurées sont des informations présentant un format et une taille précise, faciles à stocker et à analyser, et hautement organisées. Elles sont organisées de manière identifiable pour pouvoir répondre aux demandes d'information dans un but d'organisation.

Les bases de données relationnelles, telles que le langage de requête structuré (SQL), représentent un exemple typique de données structurées. Elles contiennent des nombres, des dates, des groupes de mots et des chaînes de texte organisés. Grâce à la structure claire de la base de données, il est possible d'effectuer des recherches directes et simples en fonction du type de données dans le contenu réel en utilisant des algorithmes de recherche.¹⁰

1.4.2. Les données non structurées

Les données non structurées sont des informations qui se présentent sous différentes formes et qui ne correspondent pas aux modèles de données conventionnels. Par conséquent, elles ne sont généralement pas appropriées pour une base de données relationnelle traditionnelle. Le traitement et l'analyse de ces données non structurées sont donc souvent très difficiles et prennent beaucoup de temps.

Selon Feldman et Sanger, les données non structurées ne suivent pas de format spécifique. Elles incluent généralement des images, du texte, des e-mails et d'autres types de données qui ne sont pas intégrés à la base de données.

⁹Hurwitz, Judith S., et al. « Big data for dummies ». John Wiley & Sons, 2013.

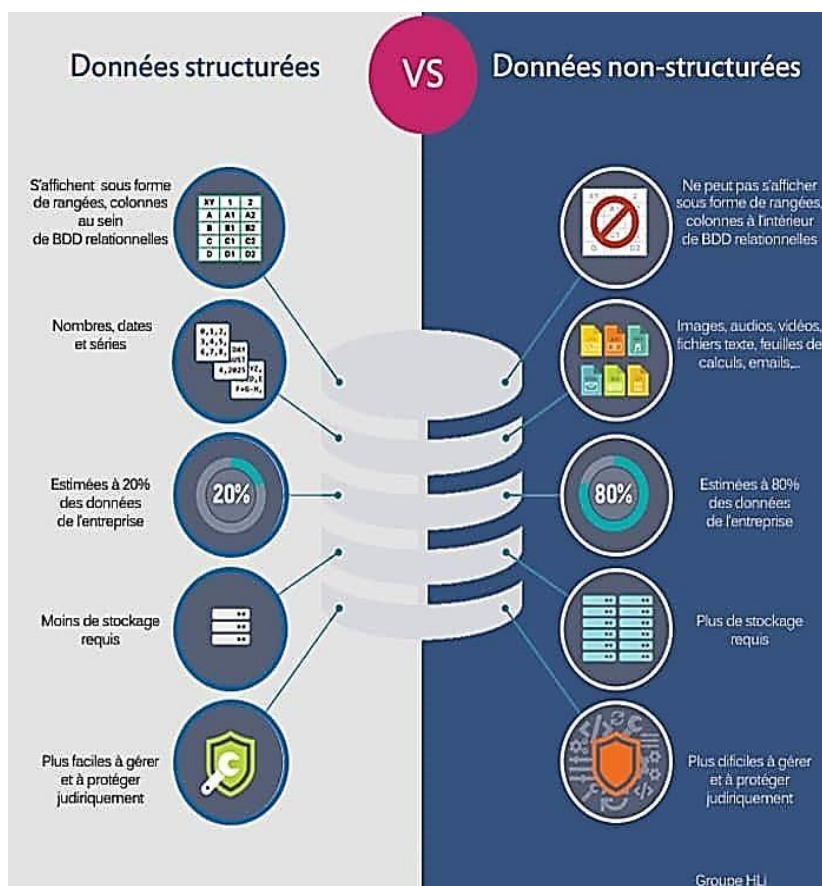
¹⁰Eberendu, Adanma Cecilia. "Unstructured Data: an overview of the data of Big Data." International Journal of Computer Trends and Technology 3.1 (2016): 46-50.

1.4.3. Les données semi-structurées :

Les données semi-structurées sont des informations qui sont irrégulières et qui peuvent être incomplètes. Leur structure peut changer rapidement ou de manière imprévisible, mais elles ne sont pas conformes à un schéma fixe ou explicite.

Selon Hanig, Schierle et Trabold, le modèle de données semi-structurées permet de regrouper en un ensemble des informations provenant de plusieurs sources avec des propriétés liées mais différentes. Par exemple, il peut s'agir de fichiers de courrier électronique, de fichiers XML ou de fichiers Doc.¹¹

Figure 2 Données structurées Vs données non-structurées



Source : <https://www.groupe-hli.com/processus-metiers-donnees-non-structurees/>

¹¹ Feldman, Ronen, and James Sanger. The text mining handbook: advanced approaches in analyzing unstructured data. Cambridge university press, 2007.

1.5. Du Big Data a la *Data science*

Les défis de stockage, d'organisation et de compréhension d'une quantité aussi importante d'informations issu du *Big Data* ont conduit au développement de nouvelles technologies dans différents domaines de la statistique, de l'apprentissage automatique et de l'extraction de données, en interaction également avec des domaines de l'ingénierie et de l'intelligence artificielle (IA), entre autres. Cet effort considérable a conduit à la naissance du nouveau domaine interdisciplinaire appelé "*Data Science*", dont les principes et les techniques visent à l'extraction automatique d'informations et de connaissances potentiellement utiles à partir des données. Bien que les technologies de *la Data science* aient été appliquées avec succès dans de nombreux domaines différents (par exemple, la santé , la maintenance prédictive et la gestion de la chaîne d'approvisionnement , entre autres), leurs potentiels ont été plus explorés dans les domaines de l'économie et de la finance.

Selon Plusieurs Data Scientiste et spécialiste du domaine ,cette science peut être définie comme :

- "La *Data science* est l'ensemble des outils, des techniques et des méthodes qui permettent d'extraire des connaissances à partir de données massives et complexes.¹²"
Yann LeCun, Rachel Schutt et Catherine Cramer
- "La *Data science* est un domaine multidisciplinaire qui utilise des techniques statistiques, informatiques et mathématiques pour extraire des connaissances et des insights à partir de données structurées et non structurées.¹³" Jake VanderPlas
- "La *Data science* est une discipline qui consiste à collecter, stocker, analyser, interpréter et communiquer des données en utilisant des techniques de traitement de données, des algorithmes et des modèles statistiques et mathématiques.¹⁴" Nadeem Iftikhar

¹² Yann LeCun, Rachel Schutt et Catherine Cramer *Data scienc* pour l'entreprise

¹³ VanderPlas Jake VanderPlas *Python Data scienc Handbook*, O`Reilly 2016

¹⁴ Nadeem Iftikhar ,*Data scienc Essentials*, Canterbury Christ Church University 2021

Chapitre 1 : Le Big Data et la Data Science : Un domaine en plein essor

- "La *Data science* est un domaine émergent qui combine des compétences en informatique, en mathématiques, en statistiques et en analyse de données pour résoudre des problèmes complexes à grande échelle¹⁵." Cathy O'Neil et Rachel Schutt

En résumé, la *Data science* est une discipline qui utilise des outils connus sous le nom de « *Machine Learning* » et des techniques de traitement de données pour extraire des connaissances et des *insights* à partir de données massives et complexes. Elle implique une combinaison de compétences en informatique, en mathématiques, en statistiques et en analyse de données pour résoudre des problèmes à grande échelle dans différents domaines d'application.

1.6. Les grandes tendances Big Data

Les technologies informatiques font indéniablement partie des compétences des Personnes employées dans la *Data Science*. Avec plus ou moins d'expertise, chacun à son niveau dans une équipe analytique doit connaître certains outils afin de tirer parti des données à sa disposition. Les données dont nous parlons sont générées et stockées par des processus informatiques, il est donc intuitif que leurs utilisations se fassent à travers le même biais. Le tableau ci-dessous divise les compétences technologiques en 3 grandes divisions.

Tableau 1 Classification des technologie de la *Data Science*¹⁶

<i>Classification</i>	<i>Exemples</i>
Les applications	Tableau, PowerBI, Excel, SAS...
Les langages de programmation	R, Python, SQL, Java, NoSQL..
Les infrastructures	Hadoop, Cassandra, Pig, Hive...

Cette classification est étroitement liée avec les différents rôles décrits précédemment

¹⁵ Cathy O'Neil et Rachel Schutt Doing Data Science, O'Reilly 2013

¹⁶ Zinovy Radovitsky , Vishwanath Hegde , Anuja Acharya, Uma Uma- Skills Requirements of Business Data Analytics and Data Science Jobs: A Comparative Analysis , Journal of Supply Chain and Operations Management, Volume 16, Number 1, March 2018

Chapitre 1 : Le Big Data et la Data Science : Un domaine en plein essor

dans le milieu de l'analyse de données. Les applications sont les principaux outils des *Data/Business Analyst* alors les langages de programmation sont ceux des data scientists. Les différentes infrastructures sont par contre le domaine d'expertise des Data Engineers et Architects. Bien évidemment ces classes ne sont pas hermétiques pour chacun des rôles. Chacun peut choisir ces armes parmi ces différents domaines pour amener son équipe à trouver les meilleures solutions

Chapitre 1 : Le Big Data et la Data Science : Un domaine en plein essor

Section 2 : Business intelligence VS Data science

De nombreuses entreprises utilisent des données pour automatiser leurs processus et optimiser leur performance, en améliorant la gestion des stocks, en analysant les comportements d'achat ou encore en détectant des écarts financiers. Toutefois, le véritable enjeu consiste à collecter les données de manière adéquate et à les fournir à la bonne personne, sous la bonne forme et au bon moment. C'est là que la *Business intelligence* (BI) et le *Big Data* interviennent, en aidant l'entreprise à prendre des décisions éclairées et à orienter son développement. Bien que ces deux notions semblent tous deux travailler avec les données pour résoudre des problèmes et prendre des décisions, quelle est la véritable différence entre les deux ? Revenons aux bases en examinant les similitudes et les différences de chacun en ce qui concerne leurs fonctions principales, leurs livrables et leur rôle global en ce qui concerne la prise de décision axée sur les données.

2.1 Définition de la Business Intelligence

La Business Intelligence est une sous-catégorie de la *Data science* qui se concentre sur la fourniture d'informations pour soutenir la prise de décision en entreprise. Elle implique la collecte, l'analyse et la visualisation de données pour aider les entreprises à prendre des décisions stratégiques. La Business Intelligence implique généralement de travailler avec des données structurées, telles que les données provenant des systèmes de vente, de marketing et financiers. L'objectif de la BI est de créer, à partir des données de l'entreprise mais externe à celle-ci, l'information et le savoir aidant les membres de l'entreprise, des cadres dirigeants aux opérationnels dans leur pilotage¹⁷.

D'autre part, la *Data science* est un domaine multidisciplinaire qui consiste à utiliser des techniques statistiques, informatiques et d'apprentissage automatique pour extraire des informations et des connaissances à partir des données. Elle implique généralement de travailler avec des ensembles de données volumineux et complexes, souvent avec des données non structurées ou semi-structurées. Les data scientist utilisent leurs compétences pour construire

¹⁷ Cigref, Business Intelligence : place de la BI et des projets décisionnelle _CIGREF_2009

Chapitre 1 : Le Big Data et la Data Science : Un domaine en plein essor

des modèles prédictifs, identifier des schémas et développer des solutions basées sur les données pour résoudre des problèmes du monde réel.

2.2 Différence entre le *Business Intelligence* et *Data science*

La *Business intelligence* et la *Data science* diffèrent de plusieurs manières, allant du type de données sur lequel elles travaillent aux livrables et approches de projet. la figure ci-dessous illustre une distinction visuelle entre les attributs les plus courants des deux.¹⁸

Figure 3 Business Intelligence VS Data Science

	BUSINESS INTELLIGENCE	DATA SCIENCE
Perspective	Looking Toward Past & Present	Looking Toward the Future
Data Types	Structured	Structured & Unstructured
Deliverables	Dashboards, Reports, Ad-Hoc Requests	Statistical & Predictive Models, Hypothesis Testing
Process	Descriptive & Static	Exploratory
Business Value	Drives Decisions	Strategic Planning

Source : <https://www.qualitatformacio.com/data-analytics-business-intelligence>

2.2.1 Perspective

La *Business intelligence* se concentre sur le présent tandis que la *Data science* regarde vers l'avenir et prédit ce qui pourrait arriver ensuite. La BI travaille avec des données historiques afin de déterminer une action réactive, tandis que la *Data science* crée des modèles prédictifs qui reconnaissent les opportunités futures.

¹⁸ Pierre Delort « Le Big Data », Edition Que sais-je ? 2018

Chapitre 1 : Le Big Data et la Data Science : Un domaine en plein essor

2.2.2 Types de données :

La *Business intelligence* travaille avec des données structurées qui sont généralement stockées dans des entrepôts de données ou des silos de données. De même, la *Data science* travaille également avec des données structurées, mais elle est principalement chargée de traiter des données non structurées et semi-structurées, ce qui entraîne une plus grande période consacrée au nettoyage et à l'amélioration de la qualité des données.

2.2.3 Livrables

Les rapports sont le cœur de métier en matière d'*Business intelligence*. Les autres livrables de la *Business intelligence* incluent la création de tableaux de bord et l'exécution de demandes AD HOC. Les livrables de la *Data science* ont le même objectif final en tête, mais se concentrent davantage sur des projets à long terme et tournés vers l'avenir. Les projets incluront la construction de modèles en production plutôt que de travailler à partir d'outils de visualisation d'entreprise. Ces projets mettent également l'accent sur la prédiction des résultats futurs, par opposition à la concentration de l'BI sur l'état actuel d'une organisation.

2.2.4 Processus

La distinction entre les processus de chacun revient à la perspective du temps, de la même manière qu'elle influence la nature des livrables. L'*Business intelligence* tourne autour de l'analyse descriptive, qui est la première étape de l'analyse et pose les bases de ce qui s'est déjà produit. C'est là que les utilisateurs métier non techniques peuvent comprendre et interpréter les données grâce à des visualisations.¹⁹ Par exemple, les responsables d'entreprise peuvent déterminer combien d'articles X ont été vendus en juillet à partir de courriels promotionnels par rapport au trafic direct sur le site Web. Cela conduit ensuite à des fouilles et analyses supplémentaires sur les raisons pour lesquelles certains canaux ont mieux fonctionné que d'autres. En poursuivant avec l'exemple précédent, la science des données adopterait une approche exploratoire. Cela signifie enquêter sur les données à travers ses attributs, tester des hypothèses et explorer les tendances communes plutôt que de répondre d'abord aux questions

¹⁹Roger Magoulas *Data science and Business Intelligence: A Practitioner's Guide to Building Intelligent Data Systems*

Chapitre 1 : Le Big Data et la Data Science : Un domaine en plein essor

métier sur les performances. Les scientifiques des données commencent souvent par une question ou un problème complexe, mais cela évolue généralement au cours de l'exploration²⁰.

En résumé, bien que la *Data science* et la *Business Intelligence* impliquent toutes deux de travailler avec des données, la *Data science* a tendance à se concentrer sur le développement de nouveaux modèles et algorithmes pour extraire des informations, étudier et prédire le futur d'une entreprise, tandis que la *Business intelligence* se concentre davantage sur l'utilisation d'outils et de techniques existants (passé ou présent) pour générer des informations qui soutiennent la prise de décision en entreprise.

Section 3 : Les grands défis en matière de *Big Data* dans la finance

Les technologies permettant de relever ces défis sont d'ores et déjà disponibles. Cependant, les entreprises doivent comprendre comment gérer le *Big Data*, aligner leur organisation sur les nouvelles initiatives technologiques et surmonter la résistance interne générale. Les enjeux spécifiques du *Big Data* dans le domaine de la finance sont donc un peu plus complexes que ceux des autres secteurs, pour de nombreuses raisons. Sécurité des données : les données financières sont souvent très sensibles, et leur sécurité est cruciale pour éviter les violations de confidentialité et les cyberattaques.²¹

3.1 Défis en termes de données

En termes de données, la *data science* est confrontée à plusieurs défis qui peuvent affecter la qualité, la disponibilité et l'exploitation des données. Voici quelques-uns de ces défis

3.1.1 Qualité des données

La qualité des données est un défi majeur en finance. Les données peuvent être incomplètes, inexacts ou mal structurées, ce qui peut entraîner des erreurs dans l'analyse. Aussi avec la prolifération des attaques de piratage et des menaces avancées et persistantes,

²⁰ Oreilly Media *Big Data Now: Current Perspectives from O'Reilly Radar*" et "*Data science and Business Intelligence: A Practitioner's Guide to Building Intelligent Data Systems*".

²¹ Utiliser le Big Data dans le secteur de la finance – www.talend.com

Chapitre 1 : Le Big Data et la Data Science : Un domaine en plein essor

les mesures de gouvernance des données sont devenues essentielles pour atténuer les risques associés au secteur des services financiers.

3.1.2 Intégration des données :

Celle-ci est un défi majeur pour le secteur financier car les données financières sont souvent dispersées dans de nombreux systèmes différents. Cette fragmentation rend difficile la création d'une base de données cohérente qui intègre toutes les sources de données nécessaires. Les entreprises doivent relever ce défi pour pouvoir exploiter efficacement les données financières et obtenir des informations précieuses pour la prise de décision .

3.1.3 Analyse des données

L'analyse des données massives est complexe et exige des compétences techniques et analytiques avancées. Les données financières sont souvent volumineuses, complexes et hétérogènes, ce qui rend leur analyse difficile. Les entreprises doivent disposer de personnel qualifié, tel que des *data scientists*, pour extraire des informations significatives des données financières. Des outils d'analyse de données avancés, tels que l'apprentissage automatique et l'analyse prédictive, peuvent être utilisés pour aider à traiter les données financières massives et obtenir des informations précieuses pour la prise de décision. Les entreprises doivent investir dans la formation et le développement des compétences de leurs employés pour garantir que leur personnel dispose des compétences nécessaires pour analyser efficacement les données financières massives.

3.2 Autres défis

Dans le domaine de la data science, outre les défis techniques et analytiques, il existe une multitude d'autres défis qui peuvent influencer la réussite et l'application des méthodes de traitement des données. On peut citer :

3.2.1 Réglementation

Les réglementations financières sont strictes et doivent être respectées lors du traitement de données financières. Elles sont soumises à des lois strictes concernant la protection des données financières, telles que la loi sur la protection des données et le règlement général sur la protection des données (RGPD) en Europe.

Chapitre 1 : Le Big Data et la Data Science : Un domaine en plein essor

Les entreprises doivent s'assurer que les données financières sont traitées conformément aux normes réglementaires pour garantir la confidentialité, la sécurité et la protection des données personnelles des clients. Les entreprises doivent également être en mesure de prouver la conformité à ces réglementations en cas d'audit ou de demande de régulateurs. Les entreprises peuvent faire appel à des experts en conformité réglementaire pour les aider à s'assurer que leurs processus et leurs pratiques respectent les réglementations en vigueur.

3.2.2 Gestion de la capacité :

Le traitement de grandes quantités de données peut nécessiter une infrastructure informatique très performante et coûteuse. Les entreprises doivent être en mesure de gérer efficacement leur capacité informatique pour répondre aux besoins de traitement de données en temps réel, tout en évitant les temps d'arrêt et les temps de latence qui peuvent avoir des conséquences financières graves. Elles peuvent investir dans des solutions de stockage et de traitement de données haute performance, telles que des serveurs dédiés ou des solutions de *cloud computing*, pour gérer efficacement leur capacité informatique mais qui coutent cher .

3.2.3 Éthique

L'utilisation de données massives peut soulever des questions éthiques et morales sur la protection de la vie privée et la transparence dans la prise de décision. Celle-ci doivent respecter les normes de protection de la vie privée des consommateurs et des utilisateurs. L'utilisation de données personnelles pour la prise de décisions doit être transparente et les consommateurs doivent être informés de l'utilisation qui en est faite. Elles doivent également prendre des mesures pour garantir la sécurité des données et prévenir les violations de données, qui peuvent avoir des conséquences négatives sur les consommateurs et les entreprises elles-mêmes.

Ces défis doivent être abordés de manière proactive pour permettre aux entreprises financières de bénéficier pleinement des avantages des *Big Data* tout en minimisant les risques associés.

Chapitre 1 : Le Big Data et la Data Science : Un domaine en plein essor

Conclusion

A travers notre recherche approfondie, nous avons pu conclure que le Big data et la *data science* jouent un rôle essentiel dans la transformation du paysage financier. Avec l'augmentation exponentielle du volume, et la variété et de la vélocités des données, les institutions financières sont confrontées à des opportunités et à des défis sans précédent.

Ces données proviennent de sources diverses telles que les transactions financières, les données de marché, les médias sociaux et les indicateurs économiques. L'exploitation efficace de ces données massives offre des avantages substantiels, tels que la compréhension approfondie des comportements du marché, l'identification des tendances émergentes et l'évaluation précise des risques.

En conclusion, l'analyse du big data nécessite la collecte de données provenant de sources variées, qui sont ensuite traitées à l'aide d'algorithmes et de structures adaptées afin de générer des informations pertinentes. Les solutions de Big Data reposent sur les technologies de la *Data science* pour comprendre et analyser ces vastes ensembles de données, permettant ainsi d'obtenir des résultats fiables et significatifs.

Il est impératif pour les acteurs financiers de maîtriser les compétences et les techniques de la *Data science* afin de tirer pleinement parti de ces opportunités et de faire face aux défis inhérents au Big Data. Dans les chapitres à venir, nous explorerons en détail les technologies spécifiques de la *Data science* appliquées à la finance, afin de fournir une vision claire et approfondie de leur utilisation et de leur impacte dans ce domaine en constante évolution.

Chapitre 02
Les technologies de la *data science*
applique à la finance

Chapitre 02 : Les technologies de la data science applique à la finance

Avec la croissance exponentielle du volume d'informations disponibles, le Big Data émerge comme une alternative aux solutions classiques de gestion et d'analyse de données, telles que les serveurs SQL et les plateformes de Business Intelligence. Les entreprises pionnières du secteur du web, telles que Yahoo, Google et Facebook, ont été parmi les premières à adopter ces technologies en raison de leurs besoins précurseurs en matière de gestion de quantités massives de données.

Dans cette partie, nous découvrirons comment ces données massives peuvent être collectées, stockées et analysées pour en extraire des informations précieuses. Nous mettrons également en évidence les défis et les opportunités qu'elles présentent pour les institutions financières.

Ce chapitre est structuré en trois sections :

- Apache Hadoop
- Le logiciel « Python »
- Le logiciel « R »

Section 01 : APACHE HADOOP

Pendant de nombreuses décennies, les entreprises ont principalement stocké leurs données sur des bases de données relationnelles (RDBMS) pour les stocker et effectuer des requêtes.

Cependant, ce types de base de données n'est pas adapté pour stocker des données non structurées et ne peut pas gérer les vastes volumes du big data.

En conséquences, la digitalisation et l'apparition de nombreuses technologies comme l'IoT (*Internet of things*) et les Smartphones ont provoqué une explosion du volume de données brutes. Face à cette révolution, de nouvelles technologies sont nécessaires pour le stockage et le traitement de données. C'est là qu'intervient le *framework* logiciel Hadoop, conçu pour répondre à ces nouveaux défis.

1.1.Définition de Hadoop

Hadoop est un framework logiciel qui repose sur Java, il est conçu spécialement pour stocker et traiter de vastes volumes de données (Big data). Il s'agit d'un projet open source soutenu par la fondation *Apache Software Foundation* basé sur le principe MapReduce et Google File System.

Les systèmes de données Hadoop ne sont pas limités en termes de capacité, ce qui signifie qu'il donne la possibilité d'ajouter facilement de hardware et de clusters pour supporter une charge plus lourd sans avoir besoin d'une reconfiguration ou de licences logicielles coûteuses.

1.2 Historique de Hadoop

Tout a commencé peu de temps après la naissance de la *world wide web*. L'avènement d'internet dans les années 90 a révolutionné le monde et lancé une nouvelle ère technologique. Cette période a également vu l'émergence des moteurs de recherche conçus pour distribuer et échanger des informations entre les internautes. Au début, les résultats étaient envoyés par des humains, mais avec l'augmentation du nombre de pages, l'automatisation est devenue indispensable pour gérer le flux croissant d'informations.

Hadoop a vu le jour grâce au projet Apache Lucene, un moteur de recherche libre lancé en 2002. Les développeurs de Lucene ont rencontré des difficultés liées à la volumétrie des données pendant le développement du projet, mais Doug Cutting a réussi à résoudre ce problème en adaptant une version *open source* des outils propriétaires développés par Google. Cette adaptation a finalement donné naissance au projet Hadoop.

Les dates marquantes de l'histoire du projet Hadoop sont : ²²

- En 2003, Sanjay Ghemawat, Howard Gobioff, et Shun-Tak Leung, ingénieurs chez Google, a publié un article décrivant son système de gestion de fichiers distribués appelé *Google File System* (GFS), qui était utilisé pour stocker les fichiers générés dans le cadre de l'exploitation web et des processus d'indexation. En 2004, Apache a implémenté une version open source du système de fichiers GFS sous le nom de *Nutch distributed File system* (NDFS).
- En 2004, Google a publié l'article qui a fait connaître MapReduce au monde entier.
- En 2005, les développeurs de Nutch ont mis en place une implémentation fonctionnelle de MapReduce dans leur système, et au cours de l'année, tous les principaux algorithmes de Nutch ont été adaptés pour fonctionner avec MapReduce et NDFS.
- En février 2006, ils ont quitté Nutch pour former un sous-projet indépendant de Lucene appelé Hadoop (dont le nom est inspiré par Doug Cutting du doudou de son fils, un éléphant en peluche jaune, pour le logo ainsi que pour le nom de ce nouveau framework Java, qui est devenu un projet indépendant d'Apache. Peu de temps après, Doug Cutting a rejoint Yahoo! qui a mis à sa disposition une équipe dédiée et les ressources nécessaires pour développer Hadoop en un système capable de fonctionner à grande échelle sur le web).

²² White, T. (2015) "HADOOP", The Definitive Guide, Edition Morgan Kaufmann, page 12

Chapitre 02 : Les technologies de la data science appliquée à la finance

- En 2008, Doug Cutting exploite Hadoop pour améliorer l'indexation de leur moteur de recherche et devient le système le plus rapide à trier un terabyte de données sur un cluster de 900 nœuds en seulement 209 secondes.

De nos jours, Hadoop est utilisé comme une plateforme polyvalente pour stocker et analyser de grandes quantités de données (big data). Il existe également d'autres distributions Hadoop proposées par différents fournisseurs tels qu'IBM, Microsoft, Oracle, Cloudera, etc.

1.3 Evolution de HADOOP

Hadoop à un système de fichiers distribué qui favorise un taux élevé de transfert de données entre les nœuds et permet un fonctionnement ininterrompu du système en cas de défaillances d'un entre eux. Cette approche diminue le risque de panne majeure, même lorsqu'un un nombre important de nœuds deviennent inopérants

Hadoop se base sur deux concepts fondamentaux, sont le système de fichiers distribué HDFS et l'algorithme de traitement de données distribué MapReduce. HDFS permet le stockage de données volumineuses et leur répartition sur un cluster de serveurs, tandis que MapReduce permet de traiter ces données en parallèle à travers les nœuds du cluster en utilisant une approche de traitement distribué.

1.3.1 Le système de fichier distribué d'Hadoop HDFS

Hadoop exploite son propre système de fichiers virtuels, connu sous le nom de HDFS (*Hadoop Distributed File System*). Celui-ci est distribué, extensible et portable, et s'inspire du *Google File System* (GFS). Le HDFS a été conçu pour stocker de très grandes quantités de données sur un grand nombre de machines équipées de disques durs standardisés. Il permet l'abstraction de l'architecture physique de stockage, ce qui permet de manipuler un système de fichiers distribués comme s'il s'agissait d'un seul et unique disque dur.²³

Cependant, HDFS se distingue d'un système de fichiers traditionnel pour les raisons principales suivantes :²⁴

²³ <https://blent.ai/blog/a/hdfs-hadoop> consulté le 13/04/2023

²⁴ <https://mbaron.developpez.com/tutoriels/bigdata/hadoop/introduction-hdfs-mapreduce> consulté le 15/04/2023

Chapitre 02 : Les technologies de la data science appliquée à la finance

- HDFS est indépendant du noyau du système d'exploitation, ce qui lui confère une grande portabilité et lui permet d'être déployé sur divers systèmes d'exploitation. Néanmoins, il nécessite l'utilisation d'une application externe pour monter une unité de disque HDFS
- HDFS est un système de fichiers distribué qui permet de stocker de très grandes quantités de données sur un grand nombre de machines standardisées. Contrairement aux systèmes de fichiers classiques, HDFS dispose d'un service central appelé Namenode qui se charge de gérer les métadonnées.
- HDFS dispose également d'un système de réplication des blocs configurable (pendant la phase d'écriture En cas d'indisponibilité d'un bloc sur un nœud, des copies de ce bloc sont disponibles sur d'autres nœuds).

Le HDFS se compose des nœuds suivants :

A. Le nœud principal (NameNode)

Est un service central (généralement appelé aussi maître) qui s'occupe de gérer l'état du système de fichiers.

Le namenode se caractérise par les tâches suivantes :

- Responsable de la distribution et de la réplication des blocs
- Serveur d'informations du Hdfs pour le client Hdfs
- Stocke et gère les métadonnées
- Comporte la liste des blocs pour chaque fichier (dans le cas de lecture)
- Contient la liste des DataNodes pour chaque bloc (dans le cas de l'écriture)
- Tenir les attributs des fichiers (ex : nom, date de création, facteur de réplication)
- Logs toute métadonnée et toute transaction sur un support persistant
- Créations/suppressions

B. Le nœud de données ou DataNode

Les Datanodes sont sous les ordres du Namenode et sont surnommés les Workers.

Il se caractérise par les tâches suivantes :

- Stocke des blocs de données dans le système de fichier local.
- Maintenir des métadonnées sur les blocs possédés.
- Namenode reçoit un « Heartbeat » d'un Datanode qui indique le bon fonctionnement de ce dernier...

C. Secondary NameNode

Le NameNode dans l'architecture Hadoop est un point unique de défaillance. Si ce service est arrêté, il n'y a pas moyen de pouvoir extraire les blocs d'un fichier donné. Pour résoudre ces problèmes, un NameNode secondaire appelé Secondary NameNode a été mis en place dans l'architecture Hadoop. Son rôle consiste à :

- Télécharge régulièrement les logs sur le Namenode.
- crée une nouvelle image en fusionnant les logs avec l'image HDFS.
- renvoie la nouvelle image au Namenode.

1.3.2 MapReduce :

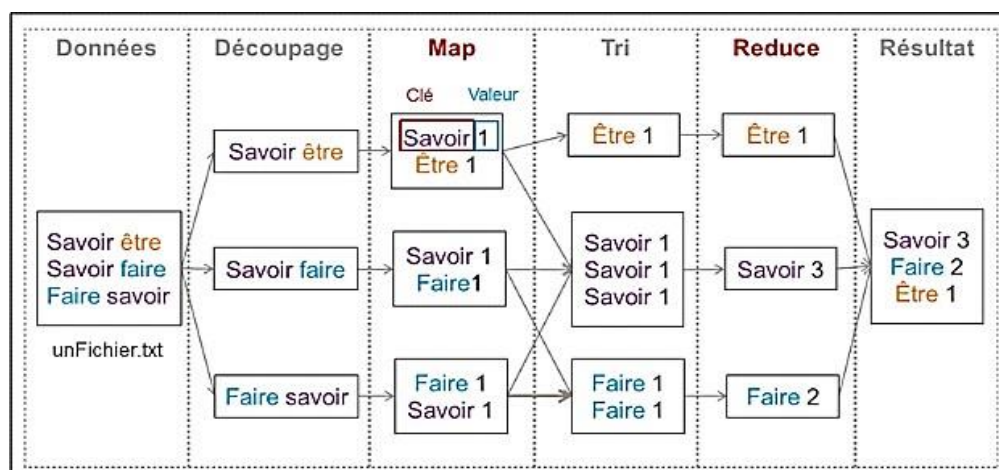
MapReduce est un modèle de programmation parallèle proposé par Google, utilisé principalement pour le traitement distribué des données au sein de clusters de nœuds. Son concept est simple il s'agit de déviser une tâche qui manipule un gros volume en plusieurs tâches traitant chacune un sous ensemble. Ainsi ce modèle peut se résumer en deux fonctions Map () et Reduce ()

Map : séparer les tâches d'un algorithme à travers l'intégralité du cluster et ainsi permettre à chaque nœud de travailler sur ces propres données.

Reduce : chaque nœud renvoie son résultat au master node. Ce dernier se chargera d'associer l'ensemble des résultats pour offrir aux clients un seul et même résultat.

Le schéma ci-dessous présente le processus d'un traitement MapReduce

Figure n°04: Processus d'un traitement MapReduce



Source : https://www.memoireonline.com/05/14/8890/m_Big-data-rapport-de-stage.html

Le MapReduce présenté sur le schéma permet de trouver le nombre d'occurrence des mots d'un fichier nommé ici « unFichier.txt »

Durant la phase de « Découpage », les lignes du fichier sont découpées en blocs.

Puis lors de la phase « Map », des clés sont créées avec une valeur associée. Dans cet exemple une clé est un mot et la valeur est 1 pour signifier que le mot est présent une fois.

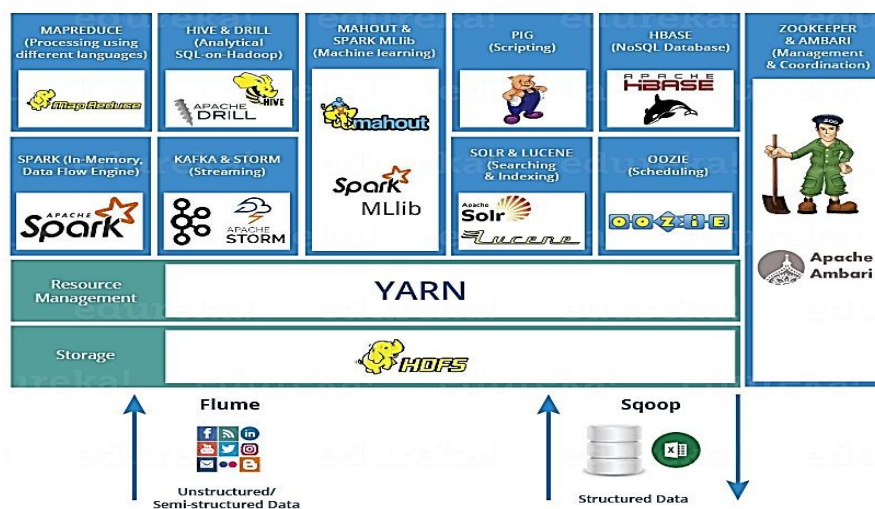
Lors du « Tri », toutes les clés identiques sont regroupées, ici ce sont tous les mots identiques. Ensuite lors de la phase « Reduce » un traitement est réalisé sur toutes les valeurs d'une même clé. Dans cet exemple on additionne les valeurs ce qui permet d'obtenir le nombre d'occurrence des mots.

1.4 L'écosystème Hadoop :

De nos jours l'écosystème s'est développé pour offrir aux professionnels un ensemble cohérent de solutions. Alors que HDFS et MapReduce restent fondamentaux pour l'architecture Hadoop, il existe désormais de nouveaux outils spécifiques qui améliorent et étendent les capacités de gestion des données de Hadoop, dans les principaux sont :

- **Yarn**(*Yet Another Resource Negotiator*) : Introduit dans la deuxième version de Hadoop, il s'agit d'un composant qui sépare le traitement des données MapReduce et la ressource management, c'est-à-dire la gestion effectuée par le JobTracker et le NameNode; (Plusieurs tâches peuvent être gérées plus efficacement dans le cluster).
- **SQOOP** : est un connecteur qui associe SQL et Hadoop, et qui facilite l'intégration de données provenant de système de gestion de base de données relationnelles (RDBMS) classiques vers HDFS.
- **Flume** : Flume permet la collecte et l'agrégation des fichiers logs, destinés à être stockés et traités par Hadoop. Il s'interface directement avec HDFS au moyen d'une API native.
- **Hive** : c'est créé une base de données relationnelle dans le système de fichiers HDFS. Le projet permet aux développeurs d'écrire des requêtes, dans un langage proche de SQL appelé HiveQL, qui sont ensuite traduites comme des programmes MapReduce sur le cluster. L'avantage est de pouvoir fournir un langage que les développeurs connaissent pour l'écriture des programmes MapReduce.
- **Spark** : c'est un moteur de calcul distribué qui effectue des traitements en mémoire sur un cluster. Il fonctionne en mode interactif. C'est-à-dire, il charge les données en mémoire avant de les traiter. Spark est donc particulièrement adapté au traitement de machine Learning, car les opérations en mémoire sont beaucoup plus rapides que les opérations sur disque

Figure n°05: L'écosystème Hadoop



Source : <https://www.edureka.co/blog/hadoop-ecosystem>

1.5 Hadoop dans la banque et la finance :

Après la crise de 2008, les contraintes réglementaires ont exigé l'enregistrement et la déclaration d'un plus grand nombre de données que jamais, et, comme les besoins en fonds propres et en réserves de liquidité ont également augmenté, la nécessité de savoir exactement combien de fonds propres doit être réservé, en fonction des expositions actuelles, est critique. Afin de suivre les tendances du marché et de gérer les services de manière plus efficace, Apache Hadoop est en train de gagner du terrain.

1.5.1 Analyse des données

L'analyse des données tel que les portefeuilles, valorisation des actifs, liquidités, capitaux propres, le commerce et les prêts est un élément essentiel dans le secteur financier et bancaire, et Hadoop peut être utilisé pour effectuer cette tâche d'une manière efficace et fournit également des informations précises pour aider à prendre la bonne décision.

Les avantages de l'analyse des données :

- Séparation des données et des calculs pour économiser la bande passante du réseau et accélérer les calculs

Chapitre 02 : Les technologies de la data science appliquée à la finance

- Le big data financier nécessite d'effectuer beaucoup de calculs en quelques secondes. Les solutions traditionnelles importent des données sur le serveur, ce qui provoque des goulots d'étranglement inutiles sur le réseau et interfère avec d'autres processus.
- Hadoop dans la finance peut être exploité pour effectuer des calculs localement et utiliser le réseau uniquement pour transmettre les résultats, ce qui permet d'économiser un espace réseau précieux pour des processus plus importants.

1.5.2 La détection de la fraude

Hadoop répond efficacement aux défis les plus courants du secteur, en analysant les points de vente, les autorisations et les transactions, et d'autres points, pour que les banques puissent identifier et atténuer la fraude.

L'augmentation de la dépendance à l'internet et au commerce électronique a entraîné une hausse significative des fraudes. Avec Hadoop, les industries peuvent utiliser l'apprentissage automatique et l'analyse prédictive pour créer des outils de regroupement qui identifient les tendances et modèles de fraudes. Par exemple, Hadoop peut alerter le secteur bancaire en cas de comportements inhabituels liés à une carte de crédit volée, permettant ainsi de bloquer temporairement la carte en attendant de vérifier l'identité du titulaire du compte.

Des algorithmes tels que K-means et SVM sont utilisés pour la détection des fraudes, en sélectionnant les caractéristiques et en classifiant les données dans les catégories frauduleuses ou non. Le processus de détection des fraudes comprend l'obtention d'échantillons de données, l'entraînement du modèle avec différents algorithmes d'apprentissage automatique, et enfin le test et le déploiement du modèle.

1.5.3 Gestion des risques

Les banques, les compagnies d'assurance, les sociétés financières et les sociétés de titres doivent évaluer les risques avec précision et Hadoop leur permet de le faire en identifiant les risques associés à l'exposition au crédit. Les solutions Hadoop offrent une vision complète et précise du risque et de son impact, permettant aux banques et aux sociétés financières d'analyser le comportement des clients, de suivre les transactions entrantes par rapport aux individus et de prendre les mesures appropriées si l'activité semble suspecte.

1.5.4 Stockage et sécurité des données :

La protection, le stockage facile et l'accès aux données financières sont les besoins optimaux des banques et des sociétés financières. Alors que *Hadoop Distributed File System* (HDFS) fournit un stockage de données évolutif et fiable conçu pour couvrir de grands clusters de serveurs de base,

Chapitre 02 : Les technologies de la data science appliquée à la finance

MapReduce traite chaque nœud en parallèle, transférant uniquement le code de package pour ce nœud. Cela signifie que les informations sont stockées dans plusieurs clusters, mais avec une sécurité supplémentaire pour fournir une option de stockage de données meilleure et plus sûre. Ainsi que Hadoop propose également des mécanismes d'authentification et d'autorisation pour contrôler l'accès aux données stocker.

1.5.5 Segmentation de la clientèle

Le secteur bancaire et financier doit analyser les données non structurées concernant ses clients qui se trouvent dans les profils des médias sociaux, les courriels, les appels, les registres de plaintes et les forums de discussion. Hadoop représente une opportunité incroyable pour les entreprises de connaître leurs clients et d'analyser leur sentiment et leur expérience client - ainsi que dans la segmentation des clients en fonction de leurs comportements communs.

L'une des techniques de regroupement les plus répandues est celle des K-moyennes, qui est largement utilisée pour regrouper des points de données similaires. Il s'agit d'un algorithme d'apprentissage non supervisé, ce qui signifie que les données sur lesquelles il est appliqué n'ont pas d'étiquettes et ne possèdent pas de correspondance entrée-sortie. La segmentation de la clientèle aide les institutions financières de diverses manières, notamment

- Identification des clients en fonction de leur rentabilité.
- Segmentation des clients en fonction de leur utilisation des services bancaires.
- Renforcer les relations avec leurs clients.
- Fournir des programmes et des services appropriés qui attirent des clients spécifiques.
- Analyser les segments de clientèle pour mettre en œuvre et améliorer les services.

Au fil des années, le secteur bancaire a dû faire face à des changements importants pour assurer la sécurité des dépôts dans des coffres forts solides. Afin d'éviter les crises et améliorer l'efficacité et la détection des fraudes, il est essentiel de disposer rapidement d'un système de sécurité avancé. C'est pourquoi les domaines du marketing et de la finance se tournent vers Apache Hadoop pour bénéficier d'un cryptage puissant et d'une analyse approfondie avec une précision statistique de 100 %.

De plus, Étant une technologie open-source, Hadoop est accessible aux petites entreprises financières et aux startups, car elles n'ont pas à payer de coûts de licence élevés.

Section 02 : Le logiciel Python

De nos jours, Python est sans aucun doute l'une des principales plateformes technologiques stratégiques dans l'industrie financière. Grace à son puissant écosystème de packages d'analyse de données, de visualisation et de machine Learning

En 2014, de nombreuses institutions financières importantes - telles que Bank of America Merrill Lynch avec son projet Quartz ou JP Morgan Chase avec le projet Athena - utilisent stratégiquement Python aux côtés d'autres technologies établies pour construire, améliorer et maintenir certains de leurs systèmes informatiques principaux.

Python est utilisé par les programmeurs débutants ainsi que par les développeurs experts hautement qualifiés, dans les écoles, les universités, les entreprises Web, les grandes entreprises et les institutions financières, ainsi que dans tous les domaines scientifiques.

2.1 Définition de Python

C'est un langage de haut niveau, interprété et multiplateforme, qui est utilisé pour une variété d'applications, notamment en informatique scientifique, en traitement des données, en intelligence artificielle et en finance. Le Python se distingue des autres langages de programmation par sa syntaxe simple et lisible, qui facilite la compréhension et la maintenance du code.

Python prend en charge les modules et les packages, ce qui encourage la modularité du programme et la réutilisation du code. En outre, l'interpréteur Python et la vaste bibliothèque standard sont disponibles gratuitement pour toutes les principales plates-formes.

Cela décrit assez bien pourquoi Python est devenu l'un des principaux langages de programmation d'aujourd'hui.

2.2 Historique de Python

Bien que Python puisse encore sembler nouveau pour certains, il existe depuis un certain temps. En fait, les efforts de développement ont commencé dans les années 1980 par Guido van Rossum des Pays-Bas. Il est toujours actif dans le développement de Python et a été nommé Dictateur Bienveillant à vie par la communauté Python (cf. http://en.wikipedia.org/wiki/History_of_Python). Les étapes suivantes peuvent être considérées comme des repères dans le développement de Python :

- Python 0.9.0 publié en 1991 (première version)
- Python 1.0 publié en 1994

- Python 2.0 publié en 2000
- Python 2.6 publié en 2008
- Python 2.7 publié en 2010
- Python 3.0 publié en 2008
- Python 3.3 publié en 2010
- Python 3.4 publié en 2014

Il est remarquable, et parfois déroutant pour les nouveaux venus en Python, qu'il existe deux versions majeures disponibles, toujours en développement et, plus important encore, en utilisation parallèle depuis 2008.

2.3 Caractéristique du Python

Entre autres, Python est caractérisé par les fonctionnalités suivantes :

2.3.1 Open source

Python et la majorité des bibliothèques et outils de soutien disponibles sont open source et généralement livrés avec des licences assez flexibles et ouvertes.

2.3.2 Interprété

L'implémentation de référence de Python, CPython, est un interpréteur du langage qui traduit le code Python en bytecode exécutable pendant l'exécution

2.3.3 Multiparadigme

Python prend en charge différents paradigmes de programmation et d'implémentation, tels que l'orientation objet, impérative, fonctionnelle ou procédurale.

2.3.4 Multi-usages

Python peut être utilisé pour le développement de code interactif rapide ainsi que pour la construction de grandes applications. Il peut être utilisé pour les opérations de système de bas niveau ainsi que pour les tâches d'analyse de haut niveau.

2.3.5 Multiplateforme

Python est disponible pour les systèmes d'exploitation les plus importants, tels que Windows, Linux et Mac OS. Il est utilisé pour construire des applications de bureau ainsi que des applications Web.

Chapitre 02 : Les technologies de la data science applique à la finance

Il peut être utilisé sur les plus grands clusters et les serveurs les plus puissants ainsi que sur des petits appareils tels que le Raspberry Pi (cf. <http://www.raspberrypi.org>).

2.3.6 Typage dynamique

Les types en Python sont généralement déduits pendant l'exécution et non déclarés statiquement comme dans la plupart des langages compilés.

2.3.7 Sensible à l'indentation

Contrairement à la majorité des autres langages de programmation, Python utilise l'indentation pour marquer les blocs de code au lieu des parenthèses, des crochets ou des points-virgules.

2.3.8 Collecte des déchets

Python dispose d'une collecte automatisée des déchets, évitant ainsi la nécessité pour le programmeur de gérer la mémoire.

Python est utilisé dans de nombreux projets, de types différents. Il est notamment utilisé comme langage de script pour des applications web. Il est également très présent dans des applications de calculs scientifiques, et est intégré dans de nombreux logiciels de modélisation. Enfin, il est aussi utilisé dans le développement de jeux vidéo ou dans l'écriture de tâches dans le domaine de l'intelligence artificielle.

Figure 6 Compilation/Interprétation Python





Source : Apprendre à programmer avec Python Swinnen, Gérard O'Reilly 2005

Python permet ainsi d'écrire rapidement, facilement, et de tester des algorithmes « *on-the-spot* » indépendamment de leur complexité. Ceci en fait un outil parfait pour l'exploration et l'analyse de données. En outre, depuis l'apparition de notebook tels que Jupyter, le partage de script entre les équipes et la visualisation de données ont été grandement facilités et de fait la collaboration également.

Chapitre 02 : Les technologies de la data science applique à la finance

Désormais Python peut être déployé à moindre coût pour effectuer des calculs scientifiques, du machine-learning, de la manipulation de données ou même de la programmation web. Cet aspect très versatile la rendu populaire à travers des petites, mais également de grandes entreprises comme Google, YouTube ou encore Facebook (Python, 2018). Le Tiobe Index le place on 1^{er} des langages les plus couramment utilisés et le « Developer Survey » de StackOverflow place Python en première position des langages les plus souhaités presque 15 % des voix comme on peut le constater dans la figure ci-dessus :

Figure 7 les tops 10 des langages de programmation selon TIOBE index

Mar 2023	Mar 2022	Change	Programming Language	Ratings	Change
1	1		 Python	14.83%	+0.57%
2	2		 C	14.73%	+1.67%
3	3		 Java	13.56%	+2.37%
4	4		 C++	13.29%	+4.64%
5	5		 C#	7.17%	+1.25%
6	6		 Visual Basic	4.75%	-1.01%
7	7		 JavaScript	2.17%	+0.09%
8	10	▲	 SQL	1.95%	+0.11%
9	8	▼	 PHP	1.61%	-0.30%
10	13	▲	 Go	1.24%	+0.26%
11	9	▼	 Assembly language	1.11%	-0.79%
12	15	▲	 MATLAB	1.08%	+0.28%
13	12	▼	 Delphi/Object Pascal	1.06%	-0.06%
14	23	▲	 Scratch	1.00%	+0.47%
15	17	▲	 Classic Visual Basic	0.98%	+0.38%
16	11	▼	 R	0.93%	-0.44%
17	30	▲	 Fortran	0.79%	+0.40%

Source : <https://www.blogdumoderateur.com/classement-langages-informatiques-tiobe-mars-2023/>

2.4 Les concepts de bases de python :

2.4.1 Variable :

Les variables en Python sont des éléments fondamentaux de la programmation. Elles permettent de stocker des valeurs en mémoire pour les réutiliser plus tard dans le programme.

En finance, les variables peuvent être utilisées pour stocker des données financières telles que des prix d'actions, des taux d'intérêt, des dates de transactions, des quantités échangées, etc.

Chapitre 02 : Les technologies de la data science applique à la finance

Elles permettent aussi aux analystes financiers et aux professionnels de la finance de manipuler ces données de manière efficace pour résoudre des problèmes.

Par exemple, une entreprise peut stocker les prix de ses actions dans une variable et utiliser des fonctions de calcul financières pour calculer des métriques financières telles que le rendement quotidien, le rendement annuel, la volatilité, etc

Il est important de noter que les variables en Python sont dynamiques, ce qui signifie qu'elles peuvent changer de type et de valeur au cours de l'exécution du programme. Par exemple, une variable qui stocke un prix d'action peut être convertie en une variable qui stocke le rendement de l'action en effectuant un calcul sur cette variable.

2.4.2 Types de données :

En Python, il existe plusieurs types de données de base, tels que :

- Les booléens (bool) : ce sont des valeurs binaires True ou False, souvent utilisées pour les opérations logiques.
- Les nombres entiers (int) : ce sont des nombres entiers, positifs ou négatifs, sans décimale.
- Les nombres à virgule flottante (float) : ce sont des nombres à virgule, positifs ou négatifs.

Par exemple, les nombres sont couramment utilisés pour représenter les données financières telles que les prix des actions, les taux d'intérêt, les devises, les montants de transactions, Les opérations arithmétiques sont effectuées sur ces nombres pour des calculs financiers tels que le calcul du rendement d'un investissement, le calcul du taux de croissance d'une entreprise.

- Les chaînes de caractères (str) : ce sont des séquences de caractères, utilisées pour stocker du texte. Les chaînes de caractères sont utilisées pour stocker des données financières sous forme de texte, telles que des noms de sociétés, des descriptions de transactions, des codes de symboles boursiers, etc. De plus, les opérations de chaîne telles que la concaténation, la recherche, le remplacement, la modification de la casse peuvent être utilisées pour manipuler ces données.
- Les listes et les tuples sont des types de données couramment utilisés pour stocker des données financières sous forme de séquences. Les listes sont modifiables tandis que les tuples sont constantes. Ces types de données sont utilisés pour stocker les prix historiques des actions, les volumes de transactions, les ratios financiers, etc... Les opérations de liste et de tuple telles que l'indexation, la tranche, la concaténation, la copie peuvent être utilisées pour manipuler ces données.

Chapitre 02 : Les technologies de la data science applique à la finance

Les dictionnaires servent à stocker des données sous forme de paires clé-valeur, telles que les informations sur les clients, les données de portefeuille et de transaction. Les opérations courantes sur les dictionnaires, telles que l'ajout, la modification, la suppression et la recherche, permettent de manipuler efficacement ces données.

2.4.3 Structures de contrôle de flux :

Les structures de contrôle de flux de Python, telles que les instructions conditionnelles (if/elif/else) et les boucles (for/while), ont leur utilité dans la résolution de problèmes en finance. Pour l'analyse des méga-données financières, l'utilisation d'une boucle peut être pertinente pour parcourir chaque élément de l'ensemble et effectuer des calculs spécifiques.

Les instructions conditionnelles peuvent être utilisées pour prendre des décisions en fonction de certaines conditions. Ainsi que, si le prix d'une action atteint un certain seuil, une décision d'achat ou de vente peut être prise en fonction de la situation.

De plus, les structures de contrôle de flux peuvent être utilisées pour implémenter des stratégies d'investissement, une stratégie qui consiste à acheter une action si elle a baissé de plus de 10% au cours des 3 derniers jours peut être implémentée à l'aide d'une instruction conditionnelle.

Enfin, les structures de contrôle de flux peuvent être utilisées pour automatiser certaines tâches financières, telles que la génération de rapports financiers ou la gestion de portefeuille.

2.4.4 Les fonctions :

Les fonctions en Python sont couramment utilisées en finance pour encapsuler des blocs de code qui effectuent des tâches spécifiques. Cela permet de réduire la duplication de code et de rendre le code plus facile à maintenir et à modifier. Les fonctions peuvent également être utilisées pour automatiser des tâches répétitives et pour créer des modules réutilisables.

Par exemple, une fonction peut être créée pour calculer le rendement annuel d'un investissement en utilisant les données de prix historiques. Cette fonction peut être appelée plusieurs fois pour différents actifs et permettre une analyse comparative plus facile. Les fonctions peuvent également être utilisées pour créer des modèles financiers personnalisés en combinant différentes fonctions pour effectuer des tâches spécifiques.

Elles sont aussi utilisées pour effectuer des calculs de la volatilité d'un portefeuille, ou la création d'un système de trading algorithmique et exécuter des tâches de traitement de données comme la lecture et l'écriture des fichiers de données, la transformation de données et la fusion de données à partir de plusieurs sources. En outre on les utilise aussi pour automatiser des tâches de reporting

Chapitre 02 : Les technologies de la data science applique à la finance

telles que la génération de rapports de performances de portefeuille, l'analyse de risque et la production de graphiques.

2.4.5 Les modules :

Les modules en Python sont des fichiers contenant du code qui peuvent être importés dans d'autres programmes Python en utilisant le mot-clé « import », pour fournir des fonctionnalités supplémentaires. Ils peuvent être utilisés pour résoudre des problèmes financiers en fournissant des fonctions, des classes et des méthodes spécialisées pour des tâches spécifiques.

De nombreux autres modules spécialisés en finance comme "pandas" et "numpy" utilisés pour la simulation de Monte Carlo afin d'évaluer les options, le calcul de la valeur à risque (VaR) ou encore la récupération de données financières en temps réel à partir de sources en ligne

2.5 Fonctionnement et environnement technologiques (Écosystème de Python)

Le Python est un langage interprété, ce qui signifie que le code source est exécuté directement sans avoir besoin d'être compilé au préalable. Il est principalement utilisé pour la programmation orientée objective, la manipulation et l'analyse de données.

Pour travailler avec Python, il est courant d'utiliser un environnement de développement intégré (IDE) tel que Spyder ou Jupyter Notebook. Ces IDE offrent une interface utilisateur graphique pour écrire, exécuter et décrire du code Python, ainsi que des fonctionnalités pour gérer des projets Python, installer des bibliothèques externes et collaborer avec d'autres développeurs.

Le fonctionnement de Python repose sur l'utilisation de modules et de bibliothèques. Les modules sont des fichiers contenant du code Python qui peuvent être importés dans d'autres programmes. Les bibliothèques sont des ensembles de modules qui sont conçus pour effectuer des tâches spécifiques, telles que NumPy, Pandas et Matplotlib, qui sont largement utilisées en finance pour l'analyse de données, la visualisation et la modélisation. Les bibliothèques de machine learning telles que Scikit-learn et TensorFlow peuvent également être utilisées pour la classification et la prévision de séries chronologiques.

Chapitre 02 : Les technologies de la data science appliquée à la finance

2.6 Les différentes bibliothèques et frameworks disponibles pour la finance

Si le Python s'est érigé comme le meilleur langage de programmation pour le Big Data, c'est grâce à ses différents packages et bibliothèques de science des données. Voici les plus populaires.

2.6.1 NumPy :

NumPy est une bibliothèque Python qui fournit des structures de données et des outils pour le calcul numérique en Python. Elle est utilisée pour manipuler des tableaux multidimensionnels (vecteurs, matrices, tenseurs) et pour effectuer des opérations mathématiques et statistiques sur ces tableaux de manière efficace.

Cette bibliothèque est un outil incontournable pour les professionnels de la finance qui cherchent à effectuer des calculs numériques sur des données financières de manière rapide et efficace.

2.6.2 Pandas

Pandas fait partie des bibliothèques de la *Data science* les plus populaires. Elle a été développée par des data scientists habitués à R et Python. Désormais, il est utilisé par un grand nombre de scientifiques et d'analystes.

La bibliothèque Pandas permet de créer des structures de données appelées *DataFrames* qui permettent de stocker des données tabulaires et de les manipuler facilement. Les *DataFrames* sont très utiles en finance pour stocker des données de marché, des informations sur les entreprises et des données économiques, et pour les analyser.

Les fonctionnalités de Pandas comprennent la lecture de données à partir de diverses sources de données, y compris des fichiers CSV et Excel, des bases de données SQL et des sites web, la fusion et le regroupement de données, la manipulation de données manquantes et la suppression de données dupliquées.

En finance, Pandas est souvent utilisé pour effectuer des analyses de rendement, des analyses de risque, des simulations de portefeuille et pour créer des graphiques et des visualisations pour communiquer les résultats des analyses.

2.6.3 Scipy

Scipy est une bibliothèque pour les calculs techniques et scientifiques. Elle regroupe des modules pour les tâches de la *Data science* et d'ingénierie.

La bibliothèque SciPy étend les fonctionnalités de base de NumPy et offre une vaste gamme de fonctions utiles pour le traitement de données financières. En utilisant les fonctions de SciPy, les

Chapitre 02 : Les technologies de la data science applique à la finance

analystes financiers peuvent implémenter des méthodes sophistiquées pour traiter des données complexes et obtenir des résultats plus précis.

2.6.4 Scikit-learn

Scikit-learn est très utile pour les algorithmes de classification, de régression ou de clustering. Elle fournit des outils simples et efficaces pour l'analyse de données et la modélisation prédictive, y compris dans le domaine de la finance :

- Prédire les prix des actions ou des produits financiers en utilisant des modèles de régression.
- Détecter les fraudes financières en utilisant des modèles de classification et d'analyse des anomalies.
- Analyser le sentiment des investisseurs sur les réseaux sociaux et les forums financiers en utilisant des modèles de classification de texte.
- *Clustering* : Scikit-learn peut être utilisé pour regrouper les investisseurs en fonction de leurs préférences d'investissement ou de leur comportement en utilisant des modèles de clustering.

2.6.5 Matplotlib :

Une fois que vous avez collecté les données financières dont vous avez besoin, il est probable que vous souhaitiez les visualiser pour mieux comprendre les tendances et les relations. Dans ce cas, le module Matplotlib peut être très utile. Il s'agit d'une bibliothèque de visualisation qui permet de tracer presque tous les types de graphiques et de diagrammes en 2 ou 3 dimensions. Cependant, il faut noter que l'API de Matplotlib peut être un peu difficile à maîtriser. Elle a été conçue pour imiter la fonctionnalité de Matlab, ce qui peut être curieux pour les programmeurs Python natifs.

2.6.6 QuantLib

Les professionnels de l'analyse quantitative en finance qui se concentrent sur la structuration et la valorisation ont la chance de pouvoir utiliser une version Python encapsulée de la bibliothèque QuantLib en C++, largement utilisée pour évaluer et calculer les risques des produits dérivés financiers. Grâce à sa base en C++, cette version est très rapide, mais il y a une courbe d'apprentissage pour l'utiliser car QuantLib n'est pas une bibliothèque native de Python et il n'y a pas de documentation spécifique à Python. Il est important de noter qu'une bibliothèque de prix dérivés entièrement écrite en Python serait trop lente pour faire des calculs de prix en temps réel sur un bureau de trading animé.

2.6.7 Pyfolio

Pyfolio est une bibliothèque Python open source pour l'analyse de portefeuilles et l'évaluation de la performance en finance. Elle est utilisée pour analyser les stratégies de trading, les portefeuilles, les fonds et les stratégies d'investissement.

Elle permet d'analyser la performance de portefeuilles en utilisant des statistiques de performance telles que le ratio de Sharpe, le Beta de marché et d'autres mesures de risque et de performance. Elle fournit également des graphiques interactifs pour visualiser les résultats de l'analyse.

Pyfolio peut être utilisé en conjonction avec d'autres bibliothèques Python pour l'analyse financière et la modélisation, telles que Pandas, NumPy et Matplotlib. Elle est utilisée par les investisseurs, les gestionnaires de portefeuilles, les analystes financiers, les universitaires et les étudiants pour l'analyse quantitative de portefeuilles.

2.6.8 TensorFlow

Développé par Google Brain, TensorFlow est une bibliothèque de Machine Learning. Ses graphiques de data flow et son architecture flexible permettent d'effectuer des opérations et des calculs de données.

En finance, TensorFlow est utilisé pour la modélisation prédictive, l'analyse de données en temps réel, la détection de fraudes et la reconnaissance de texte pour l'analyse de documents. Elle est aussi utilisée pour prédire les tendances du marché financier, identifier les risques et les opportunités d'investissement.

2.7 Les avantages de l'écosystème Python pour la finance :

- **Une communauté active et en constante évolution**

L'écosystème Python pour la finance est très actif et en constante évolution. De nouvelles bibliothèques et *framework* sont régulièrement développés pour répondre aux besoins des utilisateurs, ce qui permet aux financiers de bénéficier des dernières avancées technologiques.

- **Une grande flexibilité Python**

Cette flexibilité permet aux financiers de créer des analyses personnalisées et des modèles adaptés à leurs besoins spécifiques. En utilisant des bibliothèques Python spécialisées en finance,

Chapitre 02 : Les technologies de la data science applique à la finance

les analystes financiers peuvent facilement effectuer des analyses de portefeuille, des prévisions de prix.

- **Une grande rapidité de développement**

Grâce à la rapidité d'analyser les données, les décisions financières doivent être prises rapidement en fonction des fluctuations du marché.

- **Un coût d'utilisation très faible**

Sa programmation open source permet aux professionnels du secteur financier de réduire leurs coûts d'utilisation de la technologie tout en profitant des avantages d'un écosystème dynamique et florissant.

- **La compatibilité avec les données financières**

Les données financières sont généralement stockées sous forme de tableaux, ce qui convient parfaitement à la manipulation de données en Python. En utilisant des bibliothèques comme Pandas, les analystes financiers peuvent facilement importer, manipuler et analyser les données financières.

2.8 Python pour la Finance

Maintenant que python est devenu un choix naturel pour les professionnels de la finance qui cherche à automatiser les tâches et analyser des données complexes des données financières. Dans cette partie, nous allons explorer les différents cas d'utilisation de Python en finance, ainsi que quelques exemples de projets Python réussis dans le domaine

2.8.1 Développement de modèles prédictifs

Les modèles prédictifs sont une autre utilisation courante de Python, car ils sont utilisés pour prédire des événements futurs tels que le prix d'un actif financier, le risque associé à un investissement ou la performance d'un portefeuille. Python est souvent utilisé pour développer ces modèles en raison de sa facilité d'utilisation, de sa flexibilité et de sa grande variété de bibliothèques de modélisation prédictive.

Scikit-learn est l'une des bibliothèques les plus populaires pour la modélisation prédictive en Python. Elle offre une grande variété de modèles prédictifs, y compris la régression linéaire, la classification. Par exemple, une entreprise d'investissement pourrait utiliser Scikit-learn pour développer un modèle de prévision de prix des actions en utilisant des données historiques de prix d'actions. Ce modèle pourrait être utilisé pour aider l'entreprise à prendre des décisions éclairées en matière d'investissement en prédisant les tendances futures du marché.

Chapitre 02 : Les technologies de la data science applique à la finance

En fin de compte, l'utilisation de Python pour développer des modèles prédictifs permet aux *Data Scientists* d'obtenir des informations précieuses sur les tendances du marché et d'optimiser leurs stratégies d'investissement en conséquence.

2.8.2 Trading algorithmique

Le trading algorithmique est une méthode de trading qui utilise des algorithmes informatiques pour prendre des décisions d'achat et de vente de manière autonome en fonction de données de marché en temps réel. Cela peut inclure l'analyse de données historiques, la surveillance des actualités et des événements économiques.

Python est largement utilisé dans le trading algorithmique en raison de sa facilité d'utilisation, de sa souplesse et de la richesse de ses bibliothèques d'analyse de données. Ces algorithmes peuvent ainsi prendre des décisions de manière autonome en fonction des données du marché, ce qui permet aux traders de gagner du temps et d'optimiser leur stratégie de trading. Un exemple de projet réussi est l'utilisation de Python par JP Morgan pour développer un algorithme de trading destiné aux matières premières. Cet algorithme se base sur l'analyse de données historiques afin de prédire les variations des prix des matières premières et de prendre des décisions de trading en se basant sur ces prévisions

Cependant, il peut également présenter des risques, notamment en cas de mauvaise programmation ou de défaillance des systèmes informatiques. Il est donc important que les traders utilisant des algorithmes de trading soient conscients des risques potentiels et prennent des mesures appropriées pour les atténuer.

2.8.3 Gestion de portefeuille

La gestion de portefeuille est une pratique courante dans le domaine de la finance, qui consiste à évaluer la performance d'un portefeuille d'investissement en utilisant des données historiques.

Tout d'abord, il est possible de collecter des données historiques sur les actifs du portefeuille en utilisant des sources de données telles que Yahoo Finance ou Google Finance. Ensuite, ces données peuvent être manipulées et nettoyées en utilisant Pandas pour éliminer les valeurs manquantes, les données incorrectes ou les données redondantes.

Une fois les données nettoyées, il est possible d'effectuer une analyse de la performance du portefeuille en utilisant des techniques statistiques telles que le calcul de la volatilité, du rendement, du ratio de Sharpe et d'autres mesures clés de performance financière.

Chapitre 02 : Les technologies de la data science applique à la finance

Enfin, les résultats de l'analyse peuvent être présentés sous forme de graphiques et de tableaux interactifs à l'aide de Matplotlib, permettant ainsi une visualisation complète de la performance du portefeuille.

Un exemple concret est BlackRock, l'un des plus importants gestionnaires d'actifs au monde, qui utilise Python pour analyser les risques de ses investissements. BlackRock se sert de bibliothèques comme Pandas, NumPy et Scikit-learn pour évaluer les performances de ses portefeuilles, évaluer les risques associés et prédire les tendances du marché. Grâce à Python, BlackRock est en mesure d'effectuer des analyses approfondies, d'optimiser ses investissements et de prendre des décisions éclairées basées sur des données précises.

2.8.4 Logiciels bancaires

Python est largement utilisé dans le secteur bancaire pour améliorer la structure des systèmes de paiement, des applications bancaires en ligne et même des distributeurs automatiques de billets. Les institutions financières tirent parti de la simplicité et de la polyvalence de Python pour développer ces solutions.

Par exemple, Goldman Sachs a utilisé Python pour concevoir un modèle de prévision de défaut de paiement pour les prêts hypothécaires. Ce modèle utilise des techniques de machine learning pour prédire la probabilité de défaut de paiement en prenant en compte divers facteurs tels que l'historique de crédit de l'emprunteur, le type de prêt et le taux d'intérêt. En utilisant Python, Goldman Sachs a pu développer un modèle puissant et précis pour évaluer les risques liés aux prêts hypothécaires.

2.9 Les obstacles à l'utilisation de Python en finance

Python est devenu un langage de programmation très populaire dans le domaine de la finance, en raison de sa flexibilité, sa simplicité et sa puissance. Toutefois, malgré ces avantages, il y a des obstacles à l'utilisation de Python dans ce secteur.

L'un des principaux obstacles est la technologie et le talent. Les professionnels de la finance sont souvent confrontés à des technologies obsolètes, des systèmes complexes et des processus bureaucratiques. Il peut donc être difficile pour eux d'adopter de nouvelles technologies telles que Python, même si cela peut améliorer l'efficacité et la rentabilité de leur entreprise. En outre, il peut être difficile de trouver des talents qualifiés qui maîtrisent à la fois la finance et la programmation Python.

Chapitre 02 : Les technologies de la data science applique à la finance

Un autre obstacle à l'adoption de Python dans les grandes entreprises financières est la culture d'entreprise. Les entreprises financières sont souvent très réglementées et ont des processus stricts en place pour la gestion des risques et de la conformité. Les équipes informatiques sont souvent séparées des équipes de la finance et ont leur propre ensemble de processus et de normes. Il peut donc être difficile d'introduire Python dans un environnement où les équipes ont des cultures différentes.

En outre, les grandes entreprises financières ont souvent investi massivement dans des technologies propriétaires et des systèmes personnalisés qui ne sont pas compatibles avec Python. Les entreprises peuvent être réticentes à abandonner ces investissements et à passer à une technologie plus ouverte comme Python.

Enfin, il y a des préoccupations en matière de sécurité et de confidentialité qui peuvent entraver l'utilisation de Python dans la finance. Les entreprises financières sont soumises à des règles strictes en matière de protection des données et de la vie privée, et elles peuvent être réticentes à utiliser des langages de programmation qui ne sont pas considérés comme sûrs.

Section 03 : Le logiciel « R »

Le système R connaît depuis plus d'une décennie une progression remarquable dans ses fonctionnalités, dans la variété de ses domaines d'application ou, plus simplement, dans le nombre de ses utilisateurs. Cette section propose ainsi une présentation générale du logiciel R, son fonctionnement ainsi que son utilisation en finance

3.1 Initiation au « R » :

3.1.1 Origines du logiciel

À l'origine fut le S, un langage pour « programmer avec des données » développé chez Bell Laboratories à partir du milieu des années 1970 par une équipe de chercheurs menée par John M. Chambers. Au fil du temps, le S a connu quatre principales versions communément identifiées par la couleur du livre dans lequel elles étaient présentées : version « originale » (Brown Book ; Becker et Chambers, 1984), version 2 (Blue Book ; Becker et collab., 1988), version 3 (White Book ; Chambers et Hastie, 1992) et version 4 (Green Book ; Chambers, 1998) ; voir aussi Chambers (2000) et Becker (1994) pour plus de détails.²⁵

Dès la fin des années 1980 et pendant près de vingt ans, le S a principalement été popularisé par une mise en œuvre commerciale nommée S-PLUS. En 2008, Lucent Technologies a vendu le langage S à Insightful Corporation, ce qui a effectivement stoppé le développement du langage par ses auteurs originaux. Aujourd'hui, le S est commercialisé de manière relativement confidentielle sous le nom Spotfire S+ par TIBCO Software.

Ce qui a fortement contribué à la perte d'influence de S-PLUS, c'est une nouvelle mise en œuvre du langage développée au milieu des années 1990. Inspirés à la fois par le S et par Scheme (un dérivé du Lisp), Ross Ihaka et Robert Gentleman proposent un langage pour l'analyse de données et les graphiques qu'ils nomment R (Ihaka et Gentleman, 1996). À la suggestion de Martin Maechler de l'ETH de Zurich, les auteurs décident d'intégrer leur nouveau langage au projet GNU 1, faisant de R un logiciel libre.

Ainsi disponible gratuitement et ouvert aux contributions de tous, R gagne rapidement en popularité là même où S-PLUS avait acquis ses lettres de noblesse, soit dans les milieux académiques. De simple dérivé « not unlike S », R devient un concurrent sérieux à S-PLUS, puis le surpasse lorsque les efforts de développement se rangent massivement derrière le projet libre.

²⁵ Vincent Goulet - Introduction à la programmation en R- Cinquième édition 2016

Chapitre 02 : Les technologies de la data science applique à la finance

D'ailleurs John Chambers place aujourd'hui ses efforts de réflexion et de développement dans le projet R (Chambers, 2008).

3.1.2 Description sommaire de R

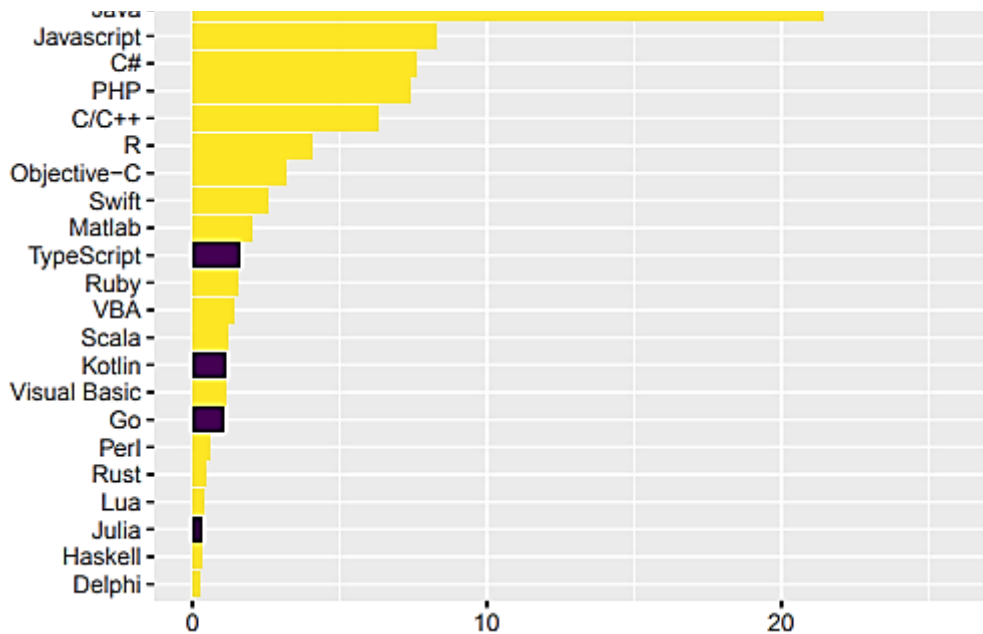
Tel que mentionné précédemment, le R est un langage principalement inspiré du S et de Scheme (Abelson et collab, 1996). Il est à la fois un langage informatique et un environnement de travail : les commandes sont exécutées grâce à des instructions codées dans un langage relativement simple, les résultats sont affichés sous forme de texte et les graphiques sont visualisés directement dans une fenêtre qui leur est propre. C'est un clone du logiciel S-plus qui est fondé sur le langage de programmation orienté objet S, développé par AT&T Bell Laboratoires en 1988. Ce logiciel sert à manipuler des données, à tracer des graphiques et à faire des analyses statistiques sur ces données²⁶.

Le R est un langage particulièrement puissant pour les applications mathématiques et statistiques (et donc actuarielles) puisque précisément développé dans ce but. Parmi ses caractéristiques particulièrement intéressantes, on note :

- L'accessibilité de l'outil en termes de coûts: tous les langages de programmation ne sont pas gratuits comme R
- L'accessibilité du langage en termes de syntaxe: R est très compréhensible (surtout pour quelqu'un qui se retrouve un peu avec la langue anglaise).
- Des procédures efficaces de traitement des données et des capacités de stockage de ces données ;
- Une suite d'opérateurs pour des calculs sur des tableaux et en particulier sur des matrices
- Une vaste et cohérente collection de procédures statistiques pour l'analyse de données et des capacités graphiques évoluées.
- Un langage de programmation simple et efficace intégrant les conditions, les boucles, la récursivité, et des possibilités d'entrée-sortie.
- La dynamique de développement du langage: le langage étant un investissement en soit, il est important de miser sur ceux qui présentent un avenir. Et ceux-ci sont ceux qui mutent avec la technologie et les besoins des utilisateurs. A ce niveau encore, R présente des arguments. Il dispose du réseau CRAN alimenté par des milliers de contributeurs, divers aussi bien de par leur position dans le monde que de par leur discipline..

²⁶ Le logiciel R Maîtriser le langage -Collection Statistique et probabilités appliquées- Pierre Lafaye de Micheaux Rémy Drouilhet Benoît Lique

Figure 8 TIOBE Index Classement des langages sur la base de divers moteurs de recherche 2023



Source: <http://pypl.github.io/PYPL.html>

Ce qui apparait dans cette figure, c'est que R parvient à se tailler une place parmi les langages les plus populaires au monde. Et cela, malgré le fait que c'est un langage spécialisé. Si sur les dix dernières années, le langage s'est enrichi avec la diversification de ses contributeurs, il reste à la base un langage élaboré par des statisticiens pour des statisticiens. De ce fait, il est excellent pour l'analyse de données.

3.2 Installation de R et RStudio

R est un logiciel de statistique distribué librement par le CRAN (Comprehensive R Archive Network) à l'adresse suivante <http://cran.r-project.org>. L'installation de R varie selon le système d'exploitation (Windows, Mac OS X ou Linux) mais les fonctionnalités sont exactement les mêmes et les programmes sont portables d'un système à l'autre. L'installation de R est très simple, il suffit de suivre les instructions.

L'utilisation du logiciel R est facilitée par l'utilisation d'un environnement de développement intégré (IDE) : **Le RStudio**. Comme explique dans la figure ci-dessus

Figure 9 installation du logiciel R

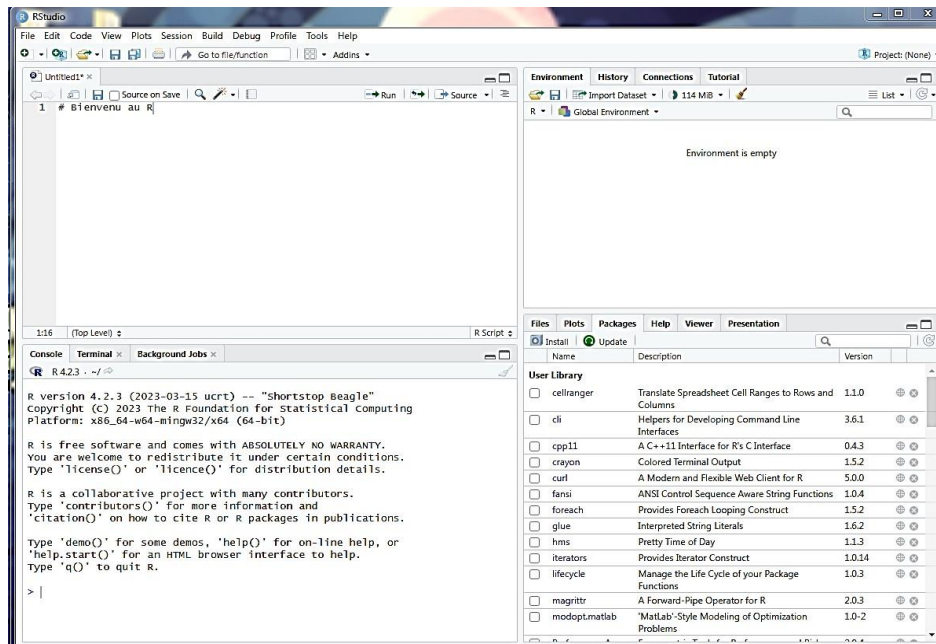
The screenshot shows the Posit website's download page for RStudio Desktop. At the top, there is a dark teal banner with the text "Grow your data science skills at posit::conf(2023) | September 17th-20th in Chicago" and a "LEARN MORE" button. Below the banner is the Posit logo and navigation links: PRODUCTS, SOLUTIONS, LEARN & SUPPORT, EXPLORE MORE, and PRICING. The main heading is "RStudio Desktop" under the sub-heading "DOWNLOAD". A descriptive paragraph states: "Used by millions of people weekly, the RStudio integrated development environment (IDE) is a set of tools built to help you be more productive with R and Python." The page is divided into two columns. The left column is titled "1: Install R" and contains the text "RStudio requires R 3.3.0+. Choose a version of R that matches your computer's operating system." and a "DOWNLOAD AND INSTALL R" button. The right column is titled "2: Install RStudio" and contains a "DOWNLOAD RSTUDIO DESKTOP FOR WINDOWS" button. Below this button, technical details are listed: "Size: 208.08 MB | SHA-256: 885432DB | Version: 2023.03.0+386 | Released: 2023-03-16".

Source : <https://posit.co/download/rstudio-desktop/>

Le RStudio (<https://www.rstudio.com/products/RStudio/>) est un environnement de développement intégré (IDE) qui propose des outils et facilite l'écriture de scripts et l'usage de R au quotidien. C'est une interface bien supérieure à celles fournies par défaut lorsqu'on installe R sous Windows ou sous Mac. Son installation suppose d'avoir préalablement installé R mais se fait sans difficulté (<https://www.rstudio.com>).

Une fois installer, L'environnement RStudio se présente sous la forme d'une fenêtre globale (voir Fig. ci dessous) scindée en 4 sous-fenêtres distinctes :

Figure 10 Environnement R



Source : Logiciel R

- La fenêtre de scripts (en haut à gauche),
- La console (en bas à gauche),
- La fenêtre d'environnement et d'historique (en haut à droite),
- La fenêtre des fichiers, graphes, packages et d'aide (en bas à droite)

Il existe plusieurs autres interfaces utilisateur pour R qui permettent aux utilisateurs de travailler avec R de manière plus conviviale hors le R studio .parmi on peut citer

- **Jupyter Notebook** : c'est une interface utilisateur populaire pour R et d'autres langages de programmation. Elle offre une expérience interactive pour l'écriture de code, l'affichage des résultats et la création de visualisations.
- **Tinn-R** : c'est une interface utilisateur pour R qui offre un éditeur de texte avancé, des fonctionnalités de coloration syntaxique et une intégration avec les commandes R.
- **RKward** : c'est une interface utilisateur pour R qui offre un éditeur de texte avancé, des fonctionnalités de débogage, des fonctionnalités de visualisation et une intégration avec les commandes R.

Chapitre 02 : Les technologies de la data science applique à la finance

Il existe également d'autres interfaces utilisateur pour R, telles que R Commander, Rattle et Deducer. Les utilisateurs peuvent choisir l'interface utilisateur qui convient le mieux à leurs besoins en fonction de leur niveau d'expérience en programmation, de leurs préférences personnelles et de la complexité de leur projet.

3.2.1 Mécanismes de fonctionnement du « R » :

Le langage de programmation R est un langage open source très populaire utilisé pour l'analyse de données, la visualisation et la modélisation statistique. Les mécanismes de fonctionnement de R reposent sur un ensemble de types de données et d'opérations de base qui permettent de manipuler et d'analyser les données.

Tout d'abord, R est un langage interprété, pas compilé, ce qui signifie que toutes les commandes tapées au clavier sont directement exécutées sans nécessiter la construction d'un programme complet comme dans la plupart des langages informatiques (C, Fortran, Pascal, etc.). Deuxièmement, la syntaxe de R est très simple et intuitive. Par exemple, une régression linéaire peut être effectuée avec la commande `lm(y ~ x)` qui signifie "ajuster un modèle linéaire avec y comme réponse et x comme prédicteur".

En R, pour être exécutée, une fonction doit toujours être écrite avec des parenthèses, même s'il n'y a rien à l'intérieur (par exemple, `ls()`). Si l'on tape simplement le nom d'une fonction sans parenthèses, R affichera le contenu de la fonction.

Lorsque R est en cours d'exécution, les variables, les données, les fonctions, les résultats, etc., sont stockés dans la mémoire active de l'ordinateur sous forme d'objets qui ont un nom. L'utilisateur peut effectuer des actions sur ces objets avec des opérateurs (arithmétiques, logiques, de comparaison, etc.) et des fonctions (qui sont elles-mêmes des objets).

A. Types de données

R dispose d'un ensemble de types de données, notamment :

- Les types de données numériques, tels que les nombres entiers (`integer`) et les nombres à virgule flottante (`numeric`).
- Les vecteurs ; est une collection de nombres ou de chaînes de caractères (textes)
- Les types de données logiques (`logical`), qui prennent les valeurs `TRUE` ou `FALSE`.
- Les types de données de caractères (`character`), qui représentent des chaînes de caractères.

Chapitre 02 : Les technologies de la data science applique à la finance

- Les types de données factoriels (factor), qui sont utilisés pour représenter des variables catégorielles.
- Les types de données de dates et heures (date/time), qui sont utilisés pour représenter des valeurs temporelles.

B. Opérations de base

Les opérations de base en R sont similaires à celles d'autres langages de programmation, telles que l'affectation de variables, les opérations mathématiques et les fonctions.

B.1) Affectation de variables

L'affectation de variables en R se fait en utilisant le symbole `<-` ou `=`. Par exemple :

```
x<-5
```

on assigne la valeur 5 à la variable x.

B.2) Opérations mathématiques

R prend en charge les opérations mathématiques de base, telles que l'addition, la soustraction, la multiplication et la division. Par exemple :

```
x<-5  
y<- 10  
z<- x + y
```

On assigne la somme de x et y à la variable z.

B.3) Fonctions

R dispose d'un grand nombre de fonctions intégrées pour effectuer des opérations telles que le calcul de la moyenne, la somme, le maximum et le minimum. Par exemple :

```
x <- c(1, 2, 3, 4, 5)  
  
mean(x)
```

`mean(x)` calcule la moyenne des valeurs dans le vecteur x.

C. Fonctions et packages

Un package est une bibliothèque de programmes externes , c'est-à-dire un ensemble de programmes R qui complète les fonctions de base et permet d'augmenter les fonctionnalités de R. Certains sont considérés comme indispensables et sont installés par défaut. Avant d'être utilisé, un package doit être installé (s'il ne l'est pas déjà) grâce à la commande (install.packages), puis obligatoirement chargé grâce à la commande(library)²⁷ .

```
install.packages(quantmod) #charger le package QUANTMOD avec ses dépendances
```

```
library(quantmod) # charger un package existant.
```

C.1) Les packages les plus utilisées en Finance

Les packages R utilisés en finance sont des bibliothèques de fonctions et de méthodes spécialement conçues pour les analyses financières. Ces packages offrent des outils pour l'analyse de données financières, la modélisation de risques, la création de graphiques et de visualisations, la gestion de portefeuille, l'optimisation de portefeuille, etc. les plus connus sont :

- **Quantmod**

Quantmod est un package très puissant conçu pour les traders quantitatifs afin d'explorer et de construire des modèles de trading quantitatifs. Il est principalement utilisé pour télécharger des données boursières pertinentes pour la gestion de portefeuille, bien qu'il dispose d'autres fonctionnalités pour les techniques avancées. Il est livré avec les packages xts et zoo, qui sont tous deux d'excellents programmes pour manipuler les données de séries temporelles. Les principales fonctionnalités de quantmod sont :

- Récupération de données financières à partir de diverses sources telles que Yahoo Finance, Google Finance.
- Visualisation de données financières en utilisant des graphiques interactifs.
- Calcul de mesures financières telles que les rendements, les ratios financiers et les indicateurs techniques.
- Modélisation de séries temporelles financières à l'aide de modèles VAR.

²⁷ *Statistiques avec R* de Cornillon et al

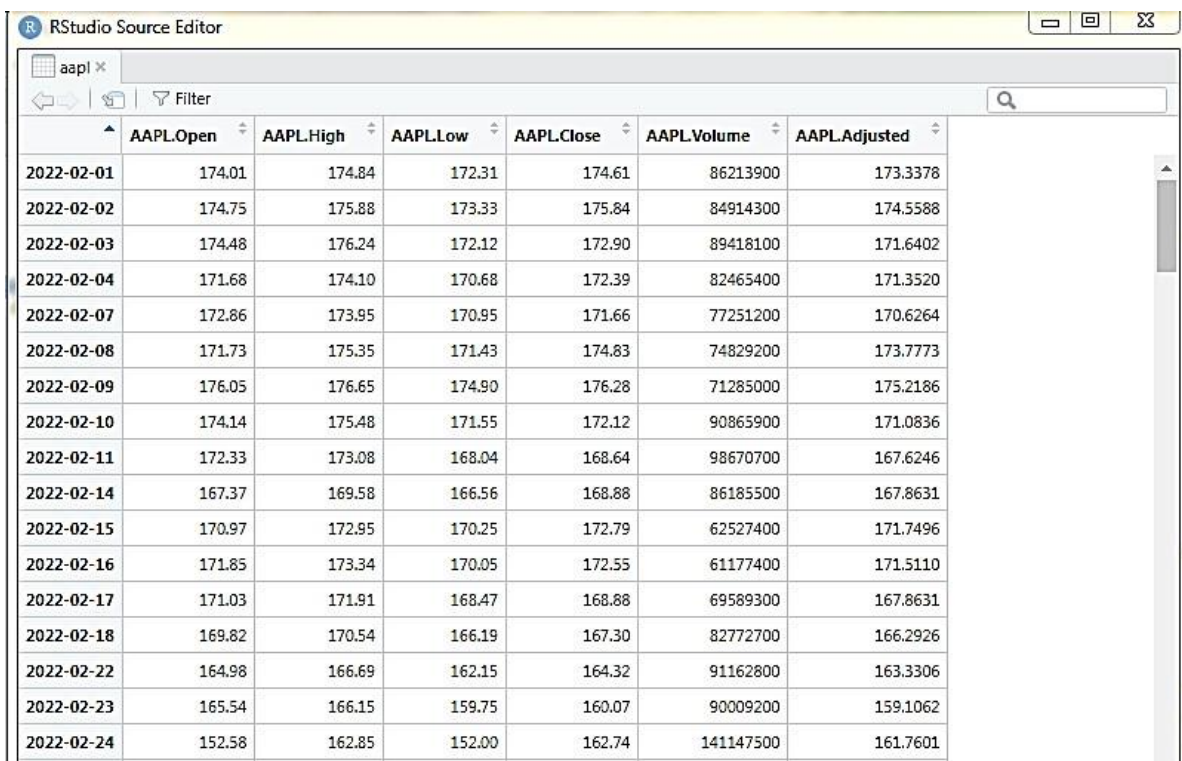
Chapitre 02 : Les technologies de la data science applique à la finance

Quantmod peut être aussi utilisé pour de nombreuses autres tâches en finance, telles que, la prédiction de prix, la modélisation de risque et l'analyse de tendances. Dans l'exemple de code ci-dessous, nous vous montrons comment charger les données historiques de prix pour AAPL à partir de YahooFinance (veuillez noter que ces prix sont en USD) :

```
install.packages("quantmod")
library(quantmod)
dt<- "2022-2-1"
aapl<- getSymbols.yahoo("AAPL", from=01, auto.assign= F)
```

Une fois le code exécuté le package fournira des données du cours des actions d'APPLE depuis le 01/02/2022 comme ceci :

Figure 11 Cours des actions AAPL



	AAPL.Open	AAPL.High	AAPL.Low	AAPL.Close	AAPL.Volume	AAPL.Adjusted
2022-02-01	174.01	174.84	172.31	174.61	86213900	173.3378
2022-02-02	174.75	175.88	173.33	175.84	84914300	174.5588
2022-02-03	174.48	176.24	172.12	172.90	89418100	171.6402
2022-02-04	171.68	174.10	170.68	172.39	82465400	171.3520
2022-02-07	172.86	173.95	170.95	171.66	77251200	170.6264
2022-02-08	171.73	175.35	171.43	174.83	74829200	173.7773
2022-02-09	176.05	176.65	174.90	176.28	71285000	175.2186
2022-02-10	174.14	175.48	171.55	172.12	90865900	171.0836
2022-02-11	172.33	173.08	168.04	168.64	98670700	167.6246
2022-02-14	167.37	169.58	166.56	168.88	86185500	167.8631
2022-02-15	170.97	172.95	170.25	172.79	62527400	171.7496
2022-02-16	171.85	173.34	170.05	172.55	61177400	171.5110
2022-02-17	171.03	171.91	168.47	168.88	69589300	167.8631
2022-02-18	169.82	170.54	166.19	167.30	82772700	166.2926
2022-02-22	164.98	166.69	162.15	164.32	91162800	163.3306
2022-02-23	165.54	166.15	159.75	160.07	90009200	159.1062
2022-02-24	152.58	162.85	152.00	162.74	141147500	161.7601

Source : Logiciel R

Chapitre 02 : Les technologies de la data science applique à la finance

- Xts et Dplyr

Le package xts est le package idéal pour manipuler des données de séries temporelles. Il étend le populaire package zoo. Comme les données financières prennent souvent la forme de données de séries temporelles, nous pouvons nous attendre à ce que les objets xts soient très utiles, car il permet d'effectuer de nombreuses opérations différentes sur ces matrices, telles que l'extraction de segments de données spécifiques dans le temps. Par exemple, si vous souhaitez prévoir les prix, mais que vous ne voulez pas inclure les heures d'ouverture et de fermeture volatiles du marché, vous pouvez choisir d'exclure ces deux intervalles de temps de votre modèle.

Le package dplyr quant à lui peut être utilisé pour la manipulation de données. Il peut filtrer, trier, résumer, sélectionner et muter vos données. Dans l'analyse financière, cela peut être utile lorsqu'on cherche :

- Des instruments financiers qui sont liés les uns aux autres,
- À obtenir certaines mesures telles que l'écart-type, la moyenne, la plage, etc.
- À agréger des informations de prix provenant de différentes bourses.

- Readr

Le package readr est un package R utilisé pour lire des données rectangulaires (par exemple, des fichiers .csv, excel) dans R de manière rapide et efficace. Il fournit des fonctions de lecture de données plus rapides que les fonctions de base de R, tout en permettant une analyse et une manipulation des données plus faciles en raison de la structure en tibble qu'il fournit.

Les fonctions de readr permettent de spécifier le type de données de chaque colonne de manière explicite ou implicite, ce qui facilite la conversion de données brutes en données prêtes à être utilisées pour une analyse. De plus, il fournit des fonctions pour gérer les valeurs manquantes, ainsi que pour les dates et les heures.

- TTR

Le package TTR (Technical Trading Rules) est une bibliothèque de fonctions R pour l'analyse technique des marchés financiers. Il permet de calculer des mesures de volatilité, telles que la covariance et la volatilité historique.

Ce dernier est largement utilisé dans l'industrie financière pour l'analyse technique des marchés et pour la création de modèles de trading. Les fonctions du package TTR permettent d'appliquer des indicateurs techniques aux données de marché, de créer des signaux d'achat et de vente en fonction de ces indicateurs, et de tester des stratégies de trading sur des données historiques

Chapitre 02 : Les technologies de la data science applique à la finance

- **PortfolioAnalytics**

PortfolioAnalytics est un package spécialement conçu pour optimiser les portefeuilles en fonction de critères spécifiques, par exemple l'optimisation du portefeuille moyenne-variance de Markowitz. En utilisant le package PortfolioAnalytics, on peut obtenir des solutions et des visualisations pour des problèmes de portefeuille avec des objectifs et des contraintes complexes. Il prend en charge différents types d'objectifs tels que le rendement, le risque, le budget de risque et la concentration de poids. Il dispose de solides fonctions pour les problèmes d'optimisation. Il peut également être utilisé pour créer des graphiques tels que les budgets de risque et la frontière efficiente.

Le package PortfolioAnalytics est également accompagné du package PerformanceAnalytics, qui permet de calculer les rendements des portefeuilles. Il est utilisé pour l'analyse de la performance et du risque des instruments financiers ou des portefeuilles. Il comprend plusieurs fonctions permettant de calculer diverses mesures telles que la valeur conditionnelle à risque (CVAR), l'écart type, la perte attendue (Expected Tail Loss) et (Expected Shortfall).

- **FPortfolio**

FPortfolio est l'un des packages les plus puissants en finance. Il permet de créer et de gérer des portefeuilles d'investissement et fournit des outils pour construire des portefeuilles diversifiés et optimisés en fonction des objectifs de l'investisseur, tels que la maximisation des rendements, la minimisation des risques ou la gestion des actifs.

FPortfolio prend en compte les contraintes de l'investisseur, telles que les limites de pondération, les contraintes de couverture et les contraintes de liquidité. Il permet également d'analyser les performances du portefeuille et de mesurer son risque à l'aide de mesures telles que la volatilité, le ratio de Sharpe.

- **Tidyquant**

Tidyquant est le pont entre le tidyverse et zoo, xts, quantmod et TTR. Il permet de récupérer et de nettoyer des données de marché à partir de diverses sources, telles que Yahoo Finance, Google Finance et FRED (Federal Reserve Economic Data) et d'appliquer des fonctions d'analyse financière courantes, telles que l'analyse des rendements, des risques et des corrélations. Il permet également de créer des graphiques et des visualisations pour aider à la prise de décisions de trading.

Chapitre 02 : Les technologies de la data science applique à la finance

Tidyquant est conçu pour fonctionner en harmonie avec d'autres packages populaires de la communauté R, tels que dplyr, tidyr et ggplot2. Il utilise la syntaxe de programmation «tidyverse» pour rendre l'analyse de données financières plus simple et plus intuitive pour les utilisateurs de R.

- **FinancialInstrument**

Le package *FinancialInstrument* permet de définir et de stocker des métadonnées pour les contrats négociables (appelés instruments, par exemple les actions, les contrats à terme, les options, etc.) Il peut être utilisé pour créer n'importe quelle classe d'actifs et de produits dérivés, dans plusieurs devises.

FinancialInstrument faisait à l'origine partie d'un paquetage complémentaire, blotter, qui fournit des fonctionnalités de comptabilité de portefeuille. Aujourd'hui, il fournit des fonctionnalités pour récupérer les données de marché pour les instruments financiers, ainsi que pour calculer des statistiques de performance et de risque pour les portefeuilles.

- **Ggplot2**

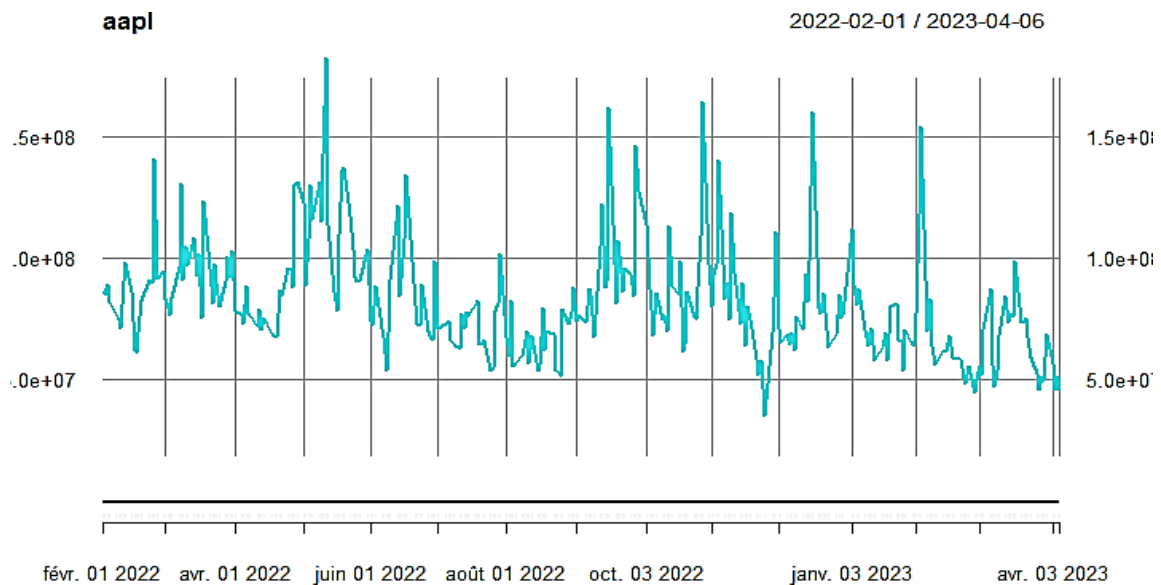
Ggplot2 est un autre membre de tidyverse, il permet de créer une grande variété de graphiques, tels que des graphiques en barres, des graphiques à nuage de points, des graphiques en boîte et des graphiques en courbes. Il offre également des fonctionnalités pour personnaliser les graphiques, comme la mise en forme des axes, l'ajout de titres et de légendes, ainsi que la personnalisation des couleurs et des styles.

C'est un package qui est très flexible et peut être utilisé pour visualiser des données financières, représenter les données historiques de prix ou de rendement, visualiser la corrélation entre deux variables financières, telles que le rendement d'un actif par rapport à un indice de marché.etc . Il est également intégré à d'autres packages populaires de la communauté R, tels que dplyr et tidyr. Par exemple, on peut l'utiliser pour visualiser les données d'APPLE précédemment collecté

```
install.packages("ggplot2")
library(ggplot2)
aapl<- getSymbols.yahoo("AAPL", from=dt, auto.assign= F)
plot(aapl)
```

le Graphe s'affiche donc ainsi :

Figure 12 Graphe des cours d'action AAPL



Source : Sortie R

Le graphe ci-dessous est une représentation graphique du cours de l'action AAPL de la figure 9

3.3 Utilisations du langage « R » dans la science de gestion

Le langage R est de plus en plus utilisé dans les sciences de gestion pour analyser les données de marché, prédire les ventes, segmenter la clientèle et réaliser des analyses de portefeuilles. Ces applications sont importantes pour les entreprises, car elles permettent d'optimiser leur performance et de prendre des décisions plus éclairées. Dans cette sous-section, nous explorerons chacune de ces applications en détail.

3.3.1 Utilisation de R dans l'analyse de données de marché

R est couramment utilisé pour l'analyse de données de marché, notamment pour l'analyse de séries chronologiques, la modélisation de la volatilité.

Les techniques d'analyse de données de marché comprennent l'analyse fondamentale, l'analyse technique et l'analyse quantitative. L'analyse fondamentale consiste à examiner les états financiers et les indicateurs clés de performance d'une entreprise pour évaluer sa santé financière. L'analyse technique utilise des graphiques pour identifier les tendances et les modèles de prix. L'analyse quantitative utilise des techniques mathématiques et statistiques pour évaluer les données de marché.

Chapitre 02 : Les technologies de la data science applique à la finance

et les tendances. Il est possible de combiner ces techniques d'analyse de marché avec R pour produire des résultats plus précis et plus complets.

Par exemple, il est possible d'utiliser des packages R tels que "dplyr", "tidyr" et "data.table" pour filtrer, regrouper, fusionner et transformer les données financières, facilitant ainsi leur préparation pour l'analyse ensuite utiliser les packages "quantmod", "TTR" (Technical Trading Rules), "PerformanceAnalytics" et "fPortfolio" pour calculer des indicateurs financiers tels que les moyennes mobiles, les rendements, les volatilités, les ratios financiers, etc. Elles permettent également de réaliser des analyses graphiques, tels que les graphiques de chandeliers japonais, les graphiques en barres, les graphiques en courbes, etc., pour visualiser les tendances et les motifs des données financières ou pour identifier des modèles de prix et des tendances de marché, puis d'utiliser des packages d'analyse statistique tels que "stats" pour ajuster des modèles linéaires et non linéaires pour prédire les tendances futures.

En ce qui concerne les états financiers, R offre des packages tels que "xbrl", "financials" et "XBRL", qui permettent de lire, analyser et manipuler les données financières provenant de rapports financiers publics, tels que les rapports annuels et les rapports trimestriels. Ces packages permettent de calculer des ratios financiers, d'effectuer des analyses comparatives, de créer des tableaux de bord financiers, etc.

3.3.2 la Gestion des risques avec « R »

R peut être utilisé pour construire et mettre en œuvre des modèles statistiques sophistiqués permettant de quantifier et de gérer les différents types de risques auxquels les institutions financières sont exposées tels que les modèles de valeur en risque (VaR)²⁸, les modèles de risque de valeur conditionnelle (CVaR) et les simulations de Monte Carlo. Ces modèles sont largement utilisés en finance pour quantifier et gérer divers types de risques, tels que le risque de marché, le risque de crédit et le risque opérationnel.²⁹

En ce qui concerne les modèles de valeur en risque (VaR), R offre des bibliothèques telles que "quantmod", "PerformanceAnalytics" et "rugarch", qui permettent de calculer la VaR pour différents types de portefeuilles d'actifs, tels que les actions, les obligations, les options, etc. Ces bibliothèques offrent également des fonctionnalités pour effectuer des analyses de sensibilité, des

²⁸ La Value-At-Risk représente la perte potentielle maximale d'un investisseur sur la valeur d'un actif ou d'un portefeuille d'actifs financiers qui ne devrait être atteinte qu'avec une probabilité donnée sur un horizon donné.

²⁹ Risk Management and Simulation in the Insurance Industry" (Gestion des risques et simulation dans l'industrie de l'assurance) de Thomas Zimmermann et Alexander J. McNeil

Chapitre 02 : Les technologies de la data science applique à la finance

simulations historiques et des approches basées sur la simulation de Monte Carlo pour estimer la VaR.

Les modèles de risque de valeur conditionnelle (CVaR) sont également largement utilisés en gestion des risques, et R offre des packages tels que "Rsolnp" et "RiskPortfolios" qui permettent de calculer le CVaR pour différents portefeuilles d'actifs. Ces packages offrent des fonctionnalités pour optimiser les portefeuilles en tenant compte du CVaR comme mesure de risque, et pour générer des rapports détaillés sur les risques associés aux portefeuilles.

3.3.3 Optimisation de portefeuille

L'optimisation de portefeuille est le processus de sélection du meilleur portefeuille (répartition des actifs), parmi l'ensemble de tous les portefeuilles considérés, en fonction d'un objectif. L'objectif maximise généralement des facteurs tels que le rendement attendu et minimise les coûts tels que le risque financier.

L'une des bibliothèques d'optimisation les plus couramment utilisées dans R est "quadprog", qui permet d'optimiser les portefeuilles en utilisant la méthode de programmation quadratique. Cette méthode permet de résoudre des problèmes d'optimisation sous contraintes, ce qui est souvent nécessaire dans le contexte de l'optimisation de portefeuille. Avec "quadprog", on peut spécifier divers objectifs d'optimisation, tels que maximiser le rendement attendu, minimiser le risque ou atteindre un certain niveau de rendement tout en respectant des contraintes spécifiques, comme la limitation de l'exposition à certains actifs ou la diversification du portefeuille.

En utilisant R, il est également possible d'effectuer l'analyse de performance des portefeuilles. Cela peut inclure des calculs de mesures de performance courantes, telles que le rendement, la volatilité, le ratio de Sharpe, le ratio de Sortino, etc. R permet également de réaliser des tests rétrospectifs, qui consistent à évaluer la performance d'un portefeuille sur une période passée en utilisant les données historiques. Ces tests permettent de mesurer la performance d'un portefeuille par rapport à des critères spécifiques et d'évaluer son adéquation par rapport à un objectif d'investissement donné.

3.3.4 Analyse de segmentation de la clientèle avec « R » (clustering)

L'analyse de segmentation de la clientèle plus connu sous *Le clustering*, est une technique courante qui joue un rôle important dans le *machine learning* moderne, il consiste à diviser les données en groupes. Dans le domaine du marketing, nous utilisons *Le clustering* pour explorer et

Chapitre 02 : Les technologies de la data science applique à la finance

sélectionner les clients qui sont des acheteurs potentiels du produit. Cela permet de différencier les clients les plus sympathiques de ceux qui ont le moins tendance à acheter le produit. Une fois les groupes constitués, les entreprises peuvent suivre leurs clients et prendre les décisions nécessaires pour les conserver dans ce groupe. En d`autre terme, on diviser une population de clients en groupes homogènes pour mieux comprendre leurs besoins et leurs comportements .

Dans la programmation R, le *clustering* est utile car Il permet d'identifier les tendances du marché et, par conséquent, les acheteurs potentiels. Le fait d'identifier les centres d'intérêt des clients à l'aide du *clustering* et de leur présenter le même produit que celui qui les intéresse peut augmenter les chances qu'ils l'achètent. La bibliothèque *cluster* fournit une gamme de méthodes de clustering pour l'analyse de segmentation de la clientèle, y compris les méthodes hiérarchiques et non hiérarchiques. La bibliothèque *factoextra* fournit des fonctions pour visualiser les résultats de clustering, y compris des graphiques en nuage de points et des graphiques en nuage de silhouette ainsi que la bibliothèque *scikit-learn* qui offre également plusieurs algorithmes de clustering populaires tels que K-means, ainsi que des outils pour évaluer les performances des modèles de clustering.

3.3.5 L`analyse des séries temporelles avec « R »

L'analyse des séries temporelles est une technique d'analyse de données qui est largement utilisée dans les entreprises pour comprendre, modéliser et prévoir les données temporelles, telles que les ventes, les revenus, les demandes de produits, les données de trafic, les données de production, etc. Cette approche permet d'extraire des informations significatives à partir de données séquentielles et de prendre des décisions éclairées pour améliorer les performances commerciales.

R possède une gamme complète de fonctionnalités avancées pour l'analyse des séries temporelles, qui sont largement utilisées dans le domaine de la finance pour anticiper les mouvements des prix des actions, les variations des taux d'intérêt, les fluctuations des taux de change et d'autres variables financières clés. Les capacités de R dans ce domaine comprennent des outils sophistiqués de visualisation, de modélisation, de prévision et d'évaluation des performances des modèles.

En ce qui concerne la modélisation, R propose une vaste gamme de packages pour construire et mettre en œuvre divers modèles financiers, tels que les modèles d'évaluation des options, les modèles d'obligations et les modèles de risque de crédit. Ces modèles sont utilisés en finance pour fixer les prix des instruments financiers dérivés, évaluer les titres à revenu fixe et évaluer le risque de crédit dans les prêts et les portefeuilles de crédit.

Chapitre 02 : Les technologies de la data science applique à la finance

3.4 Python ou R : quel est le meilleur pour la *data science* en finance

Le choix entre R et Python pour les tâches de finance dépend de divers facteurs tels que les besoins spécifiques de l'entreprise, les compétences de l'équipe, les préférences personnelles .

Tableau 2 : comparaison entre Python et R

Caractéristiques	R	Python
Domaines d'application	Statistiques, analyse de données, data science, recherche opérationnelle	Développement web, science des données, apprentissage automatique, automatisation de tâches
Syntaxe	Syntaxe associée aux statistiques et mathématiques, fonctionnelle et procédurale	Syntaxe simple et lisible, orientée objet et fonctionnelle
Communauté	Grande communauté, avec de nombreux packages et ressources dédiées à la statistique et à l'analyse de données	Communauté dynamique et vaste, avec de nombreux forfaits dans divers domaines
Les performances	Performances optimisées pour les calculs statistiques et les analyses de données	Performances élevées pour le traitement de données massives et l'apprentissage automatique
Visualisation	Bonnes fonctionnalités de visualisation avec des packages tels que ggplot2	Bibliothèques de visualisation puissantes comme Matplotlib
Simplicité d'apprentissage	Courbe d'apprentissage relativement raide pour les débutants	Syntaxe simple et intuitive, facilitant l'apprentissage pour les débutants
Disponibilité des forfaits	Vaste collection de packages dédiés à la statistique et à l'analyse de données	Large écosystème de packages pour divers domaines, y compris l'apprentissage automatique
Système d'intégration	Peut être intégré avec d'autres langages et outils, mais principalement orienté vers l'analyse de données	Peut être intégré dans des flux de travail système plus complexes grâce à des bibliothèques spécifiques
Popularité	Populaire dans les domaines universitaires et de recherche	Populaire dans l'industrie et la recherche, avec une adoption croissante
Documentation	Documentation complète et précise disponible, avec de nombreux exemples	Documentation étendue et communautaire, avec une variété de tutoriels et de guides

En conclusion... Pour la pure data science, R a toujours un léger avantage sur Python, bien que l'écart se soit considérablement réduit. Néanmoins, les applications plus larges de Python en font le meilleur choix.

CONCLUSION

En conclusion de ce chapitre consacré aux technologies de la Data Science appliquées à la finance, nous pouvons affirmer que les langages de programmation Hadoop , Python et R jouent un rôle fondamental dans la mise en œuvre de ces avancées. Ces langages sont largement utilisés par les professionnels de la finance pour analyser, modéliser et visualiser les données, grâce à leur richesse en bibliothèques et packages dédiés à la Data Science.

Ces deux langages offrent une communauté dynamique et des ressources abondantes. Leur adoption croissante témoigne de leur pertinence et de leur efficacité pour répondre aux défis spécifiques auxquels sont confrontées les institutions financières.

Le chapitre suivant approfondira ces aspects en explorant les cas d'utilisation spécifiques de Python et R dans des contextes financiers concrets. Nous explorerons les exemples concrets de leur application, tels que la prévision des prix des actifs financiers, la gestion des portefeuilles d'investissement, l'analyse des risques et la détection des fraudes.

CHAPITRE 3
Les uses cases de la data science en
finance

Chapitre 3 : Les *Uses Cases* de la *Data science* en finance

Dans le précédent chapitre, nous avons exploré les outils de la *Data science* utilisés en finance, tels que R et Hadoop. Dans ce chapitre, nous allons approfondir notre compréhension en examinant les cas d'utilisation améliorés de cette technologie dans le domaine financier. Nous mettrons en évidence comment ces outils et techniques peuvent être appliqués de manière spécifique pour résoudre des problèmes complexes et améliorer les processus financiers

La première section sera dédiée à l'application de la *data science* pour l'optimisation des processus et la croissance des entreprises ou nous parlerons de son application en marketing, gestion de performance, trading et détection de fraude > tandis que la section 2 sera dédiée au rôle de la *Data science* pour les entreprises modernes et Startups

Section 1 : Applications de la *data science* pour l'optimisation des processus et la croissance des entreprises

La *data science* a révolutionné la façon dont les entreprises abordent leurs activités, en leur offrant des opportunités sans précédent pour optimiser leurs processus, améliorer leurs performances et stimuler leur croissance. Dans cette section, nous explorerons les applications de la *data science* dans différents domaines clés tels que le marketing, la détection de fraude, la gestion de performance et le trading.

1.1 Applications de la *data science* pour le marketing

Depuis le début du 21^{ème} siècle, tant le Digital Marketing (DM) que les *Data science s* (DS) ont remarquablement évolué en termes d'usage et de rentabilité (Tla DSGo & Verissimo, 2014). Cela a conduit à l'émergence d'un écosystème numérique, qui connecte les utilisateurs 24h/24 et 7j/7 et qui a façonné les nouvelles habitudes et comportements des utilisateurs (Mayer Schönberger & Cukier, 2013).

La DS a un vaste potentiel en marketing. Elle aide à multiplier les informations et les sources de données, à améliorer les capacités de gestion des données du logiciel et à concevoir des algorithmes complexes et avancés. La DS change la façon dont les marques et les utilisateurs interagissent les uns avec les autres. Son application dépend fortement de la nature du site Web et du type d'entreprise. Les spécialistes du marketing peuvent désormais se concentrer davantage sur le client et répondre à ses besoins en temps réel. En utilisant La DS ils peuvent

Chapitre 3 : Les *Uses Cases* de la *Data science* en finance

déterminer rapidement quel contenu cibler pour les clients, quel canal utiliser et à quel moment, grâce aux données collectées et générées par ses algorithmes. Les utilisateurs se sentent à l'aise et sont plus enclins à acheter ce qui est proposé lorsque la DS est utilisée pour personnaliser leurs expériences.

Les outils de la DS peuvent également être utilisés pour analyser les performances des campagnes d'un concurrent et identifier les attentes de ses clients. Cela permet aux spécialistes du marketing d'ajuster leurs propres stratégies en conséquence et de rester compétitifs sur le marché.

La DS s'intéresse principalement à la fidélisation des utilisateurs et à la conversion des prospects dans le marketing numérique. Elle peut guider un utilisateur dans la direction qui correspond aux objectifs de l'entreprise en utilisant des chats bots, un marketing par e-mail intelligent, une conception Web interactive et d'autres services de marketing numérique. Plusieurs facteurs déterminent l'impact de celle-ci sur le marketing numérique.

Grâce à des logiciels spécialisés en DS, une analyse approfondie des entreprises et de leurs clients peut être réalisée en ligne. L'incorporation de la DS et du ML dans la stratégie marketing permet aux entreprises d'exploiter plus efficacement les données disponibles et de cibler les clients potentiels avec des publicités attrayantes au moment le plus opportun.

La DS permet aussi d'éviter des dépenses excessives en publicité numérique et garantit une utilisation efficace des ressources financières. La capacité des ordinateurs à reconnaître des objets, des scènes et des activités dans les images est appelée vision par ordinateur. La vision par ordinateur trouve des applications dans l'analyse d'imagerie, la reconnaissance faciale, la sécurité publique et la surveillance. Lorsque la vision par ordinateur est combinée à des algorithmes appropriés, les robots peuvent prédire les événements futurs et agir en conséquence en fonction des changements potentiels.

L'intégration de la DS dans le marketing a également simplifié la création de profils clients et la compréhension du parcours client. Les marques sont ainsi en mesure de fournir rapidement et facilement un contenu personnalisé et précieux à différents profils de clients, à chaque étape du processus marketing et sur tous les canaux. En se basant sur des données historiques, les applications de la DS dans le marketing numérique peuvent déterminer quel contenu est le plus susceptible de ramener les clients sur un site. De plus, la *data science* permet d'identifier les

Chapitre 3 : Les *Uses Cases* de la *Data science* en finance

clients les plus enclins à se désabonner d'un service spécifique et d'analyser les caractéristiques communes parmi les désabonnés. Ces analyses permettent aux spécialistes du marketing de planifier leurs futures campagnes et de mettre en place des stratégies visant à fidéliser leur clientèle.

En somme, La *data science* joue un rôle considérable dans le marketing digital en permettant d'exploiter pleinement le potentiel des données pour prendre des décisions stratégiques et améliorer l'expérience des clients. Grâce à l'utilisation de techniques avancées de *data science*, les plateformes de marketing numérique telles que Facebook et Instagram peuvent évaluer minutieusement les informations des utilisateurs et les utiliser pour orienter les offres publicitaires de manière personnalisée, répondant ainsi aux envies et aux besoins spécifiques des clients.

1.2 Applications de la *data science* pour la gestion de performance

L'importance de la gestion de la performance était déjà reconnue par les chercheurs dans les années 1920 (De Nisi & Pritchard, 2006 ; Justin & Joy, 2021). Depuis lors, les approches de gestion ont été sujettes à de fréquents changements et se sont développées de manière excessive au fil du temps. Partant du concept de « management par exception », où les supérieurs hiérarchiques n'interviendraient qu'en cas de conflit (Komaki, 1986), le management de la performance décrit aujourd'hui une pratique répétée à forte interaction humaine et technologique

L'activité de gestion de la performance est définie comme le « processus continu d'identification, de mesure et de développement de la performance des individus et des équipes et d'alignement de la performance sur les objectifs stratégiques de l'organisation » (Aguinis, 2013, p. 2). Cela indique que le processus de gestion de la performance implique différentes activités visant à fournir une rétroaction fiable à la main-d'œuvre pour ajuster les compétences individuelles à la stratégie globale de l'organisation

En général, le processus de gestion des performances est un processus continu et ses activités sont organisées en une séquence circulaire. Il commence par la définition des exigences de performance clés, où les attentes en matière de performance sont discutées et les objectifs sont planifiés (Armstrong, 2000). Ensuite, les employés s'engagent dans leur travail en mettant en

Chapitre 3 : Les *Uses Cases* de la *Data science* en finance

œuvrent le plan défini, tout en étant surveillés par leurs superviseurs pour évaluer leurs progrès et fournir des retours d'information (Aguinis, 2013).

Cependant, l'introduction des technologies du *Big Data*, a révolutionné la gestion de la performance. Ces avancées technologiques ont permis de collecter, analyser et interpréter de vastes quantités de données provenant de différentes sources, offrant ainsi aux gestionnaires des informations plus précises et approfondies sur les performances individuelles et organisationnelles.

Grâce à l'utilisation de l'analyse de données, les entreprises peuvent maintenant recueillir des données en temps réel sur les performances des employés, en utilisant des indicateurs clés de performance spécifiques et des métriques de performance. Cela permet aux gestionnaires d'avoir une vision plus claire et objective de la performance de l'entité, en identifiant les domaines de forces et de faiblesses, et en prenant des décisions basées sur des données factuelles.

De plus, l'utilisation de la *data science* dans la gestion de la performance permet également d'identifier des modèles et des tendances qui peuvent influencer la performance, en identifiant les facteurs clés qui contribuent aux performances élevées et en mettant en évidence les opportunités d'amélioration, les dirigeants sont en mesure de mettre en place des stratégies de développement des compétences ciblées et des programmes de formation adaptés pour optimiser les performances individuelles et collectives.

Pour mesurer et gérer leur performance, la plupart des organisations utilisent actuellement des systèmes de gestion de la performance qui intègrent des méthodes, des techniques, des outils de support et des composants de Business Intelligence de gestion. Les principales fonctions de ces systèmes sont les suivantes :

- Collecter et stocker régulièrement différentes mesures de l'entreprise (indicateurs de l'état actuel de la performance de l'entreprise).
- Recueillir et stocker des repères, des cibles (valeurs seuils) et interprétations des résultats de comparaison entre les indicateurs de performance actuels et les valeurs projetées.

Chapitre 3 : Les *Uses Cases* de la *Data science* en finance

- Faciliter les roll ups³⁰ et les drill downs³¹ des indicateurs analysés ainsi que les critères d'agrégation hiérarchiques (mesures de performance structurées).

En résumé, dans le contexte du *Big Data*, le rôle du jugement, de l'intuition et de l'expertise humaine est réévalué au profit de l'analyse quantitative basée sur les données. Les analyses sont plus basées sur les données plutôt que sur des opinions et les décisions reposent plus sur les résultats issus des données, plutôt que sur des opinions subjectives. Il est tout de même crucial pour l'analyste d'avoir une compréhension globale de l'organisation afin de communiquer efficacement avec la direction. Cela signifie qu'il doit être en mesure de traduire les résultats de l'analyse en termes compréhensibles pour les décideurs, en fournissant des informations pertinentes et exploitables.

1.3 . Applications de la *data science* dans la détection de fraude

Le développement de technologies de plus en plus avancées a certainement un impact positif sur différents domaines, tels que l'industrie, les affaires et la finance. Tout d'abord, la technologie peut automatiser les processus commerciaux et encourager les gens à trouver des innovations liées à leur activité. Deuxièmement, la comptabilité informatisée peut fournir des prédictions ou des projections concernant les perspectives d'avenir de l'entreprise au profit des parties prenantes. Troisièmement, avec l'existence d'un système informatisé, il n'est plus nécessaire de traiter manuellement la préparation des rapports financiers (Naqvi, 2020).

L'impact de la fraude a affecté les performances dans divers domaines et professions, dont l'audit. La fraude entraîne l'émergence de nouveaux domaines qui requièrent l'attention de l'auditeur. En 2020, chaque personne générera en moyenne 1,7 Mo de données par seconde, ce qui augmentera de manière exponentielle chaque année. Les données doivent être analysées pour s'assurer que le contrôle de la fraude est effectué de manière approfondie. Si les données ne sont pas contrôlées, les développements informatiques peuvent devenir une nouvelle porte ouverte à la fraude, car les fraudeurs y trouvent de nombreuses cachettes. Ces progrès technologiques entraînent également de nouveaux risques, tels que les virus qui attaquent le système, la cybercriminalité et les activités illégales sur les réseaux informatiques et sur l'internet. En outre, le risque auquel il faut prêter attention est la fraude numérique, qui peut

³⁰ Les rolls up désigne le processus de visualisation des données à un niveau de détail accru

³¹ les drill downs désigne le processus de visualisation des données à un niveau de détail décroissant.

Chapitre 3 : Les *Uses Cases* de la *Data science* en finance

être interprétée comme l'utilisation de la technologie basée sur le web pour commettre des fraudes et abuser des actifs afin d'obtenir des avantages financiers.

C'est dans ce contexte qu'est apparue la nécessité d'utiliser les technologies émergentes pour soutenir les performances des auditeurs. L'étendue de la population de données a eu pour conséquence que la méthode d'échantillonnage utilisée dans les audits conventionnels n'est plus efficace. Sans un examen approfondi de chaque donnée, de nombreux domaines critiques risquent de passer inaperçus. En outre, l'utilisation de techniques d'audit conventionnelles nécessite plus de temps, de ressources et de coûts pour contrôler un champ d'application plus étendu. Ainsi, les techniques d'audit conventionnelles augmentent le risque d'échec de la détection de la fraude et réduisent l'efficacité et l'efficience de l'audit. Le secteur de l'audit tarde encore à adopter les technologies émergentes pour soutenir les activités d'audit (Satyawan et al., 2021).

Aujourd'hui, la grande majorité des institutions financières se sont dotées, de processus éprouvés et de professionnels qualifiés en matière de prévention de la fraude, comme des experts en *data science*, lesquels utilisent des algorithmes pour analyser les tendances parmi les incidents passés afin de prédire les prochains. Ces institutions tirent profit de leur exposition à la fraude en tirant des leçons de celle-ci tandis que de nombreuses autres organisations utilisent des méthodes de détection qui s'appuient uniquement sur leur expérience passée et peuvent devenir obsolètes.

La *data science*, et particulièrement le *machine learning*, sont déjà largement utilisés pour des problématiques liées à la fraude. En effet, grâce à ces algorithmes, il est possible de détecter les anomalies d'après des commandes frauduleuses réelles afin d'aider l'organisation à créer des règles plus sophistiquées. Selon les dernières recherches il a été constaté que les clients qui utilisaient un sous-ensemble de fournisseurs de services gratuits de messagerie électronique présentaient un risque de fraude plus élevé que ceux qui commandaient à partir d'adresses de courriel d'entreprise. L'introduction de ces algorithmes a permis à l'organisation de mettre en place un programme de prévention de la fraude plus efficace, réduisant à la fois les erreurs de détection et les omissions.

Pour établir un algorithme efficace de détection de la fraude, les *data scientists* ont généralement besoin de beaucoup de données, ce qui est difficile à faire avec peu d'opérations frauduleuses relevées. L'une des solutions consiste à utiliser un auto-encodeur, c'est-à-dire un

Chapitre 3 : Les *Uses Cases* de la *Data science* en finance

réseau neuronal qui peut générer des milliers d'exemples fictifs basés sur quelques cas réels. Les réseaux neuronaux ³²sont un concept au cœur de l'intelligence artificielle; ils simulent le fonctionnement du cerveau et sont entre autres capables de générer de nombreux exemples semblables à partir d'un petit nombre d'entrées. Ainsi, un grand nombre de cas de fraude artificiels fondés sur quelques cas réels peuvent servir à entraîner les algorithmes de prévention de la fraude.

Pour conclure on peut dire que *la data science* être utile pour détecter les fraudes financières, même si elle présente certaines limites et représente un grand potentiel d'avenir .

1.4 Applications de la *data science* pour le trading algorithmique

Prédire le marché boursier est l'objectif des investisseurs depuis qu'il existe. Chaque jour, des milliards de dollars sont échangés en bourse, et derrière chaque dollar se cache un investisseur qui espère en tirer profit d'une manière ou d'une autre. Si un investisseur est capable de prédire avec précision les mouvements du marché, il s'agit d'une promesse alléchante de richesse et d'influence. Il n'est donc pas étonnant que le marché boursier et les défis qui lui sont associés soient présents dans l'imaginaire du public à chaque fois qu'il se comporte mal.

Depuis les années 1990 et l'introduction des méthodes informatiques dans la finance, de nombreuses recherches se sont concentrées sur l'application de l'intelligence artificielle aux investissements financiers sur le marché boursier. Les principaux avantages de l'utilisation d'approches informatiques pour automatiser le processus d'investissement financier comprennent l'élimination de l'irrationalité momentanée" ou des décisions prises sur la base des émotions, la capacité de reconnaître et d'explorer des modèles qui sont examinés par les humains, et la consommation immédiate d'informations en temps réel. Bien qu'un ordinateur effectue la grande majorité des transactions de fonds spéculatifs de manière automatisée, 90 % de ces opérations sont encore réalisées par une procédure codée en dur. Ainsi, l'application toujours croissante de la *data science* a encore un grand potentiel de développement.

Le trading algorithmique est la pratique consistant à acheter ou à négocier des titres selon un ensemble de règles prescrites et testées sur des données passées ou historiques. Ces

³² Un réseau neuronal est une méthode d'intelligence artificielle qui apprend à des ordinateurs à traiter les données d'une manière inspirée par le cerveau humain. C'est un type de processus de machine learning appelé deep learning, qui exploite des nœuds, ou neurones, interconnectés dans une structure à plusieurs couches similaire au cerveau humain. Il crée un système adaptatif utilisé par les ordinateurs pour apprendre de leurs erreurs et s'améliorer en continu.

Chapitre 3 : Les *Uses Cases* de la *Data science* en finance

règles sont basées sur des graphiques, des indicateurs, des analyses techniques ou des données essentielles sur les actions

L'avantage des systèmes de trading basée sur la *data science* par rapport aux traders humains est qu'elle rend la négociation plus systématique. Ainsi, ils peuvent analyser un vaste ensemble de données provenant de différentes sources en une fraction de seconde et effectuer de véritables transactions à haute fréquence (HFT) qui peuvent tirer profit des anomalies du marché et des différences de prix

Un système de trading automatisé est un type de système de prise de décision basé sur l'information qui permet aux traders d'établir des règles spécifiques pour l'entrée et la sortie des transactions et qui est exécuté automatiquement par l'intermédiaire d'un ordinateur. Dans ce contexte, La *data science* est appliqué pour modifier le mode de fonctionnement des systèmes de négociation classiques. Celle-ci permet de prendre des mesures dans le but de maximiser les récompenses et d'obtenir les meilleurs résultats. . Par conséquent, au lieu de prendre des décisions basées sur des prévisions de prix explicitement, les modèles sont entraînés à exécuter directement des positions de trading.

Les algorithmes de trading fonctionnent en suivant des règles prédéfinies et en analysant les données du marché pour identifier des opportunités de trading. Ces règles peuvent être basées sur une variété de facteurs tels que les graphiques, les indicateurs techniques, les données fondamentales ou même l'apprentissage automatique.

Lorsque l'algorithme détecte une configuration ou un signal de trading conforme à ses règles, il peut prendre automatiquement des décisions d'achat ou de vente de titres, et ce, sans intervention humaine. Ils peuvent également être configurés pour exécuter des transactions à haute fréquence, tirant profit des variations rapides des prix sur les marchés.

En conclusion, l'application de la *data science* dans le trading algorithmique offre de nombreuses opportunités et avantages significatifs. Les avancées technologiques et les énormes quantités de données disponibles ont permis aux traders d'exploiter des modèles et des algorithmes sophistiqués pour prendre des décisions éclairées et améliorer les performances de trading.

Section 2 : la *Data science* pour les entreprises moderne et Startups

Les start-up, les entreprises modernes innovantes en forte croissance, attirent une attention médiatique et politique considérable. Elles sont souvent perçues comme des entités "cool" et caractérisées par une structure organisationnelle moins hiérarchisée. Les start-up se présentent comme un modèle économique, organisationnel et idéologique alternatif par rapport aux entreprises traditionnelles. Ces dernières bénéficient pleinement des avantages de la *data science*, ce que nous allons explorer dans cette section.

2.1 Les « *Machines Learning* » au profit de l'innovation et des Startup

Le succès des jeunes entreprises (appelées startups) joue un rôle crucial dans notre économie car ces entreprises agissent souvent en tant que créateurs nets d'emplois nouveaux et poussent, grâce à leurs innovations produits et processus, la frontière technologique de la société.

Les histoires à succès des entrepreneurs qui ont remodelé des industries entières sont très marquantes, mais d'un point de vue probabiliste, il est estimé que seulement 10% des startups restent en activité à long terme. Les algorithmes d'apprentissage automatique, entraînés sur une grande quantité de données, sont généralement adaptés pour prédire le succès des startups, en particulier parce que les facteurs de réussite sont souvent inconnus et que leurs interactions sont complexes.

De même que pour la prédiction du succès au niveau de l'entreprise, les algorithmes d'apprentissage automatique peuvent être utilisés pour prédire le succès pour des projets individuels. Les données non structurées, telles que des plans d'affaires, peuvent être combinées avec des données structurées pour mieux prédire les chances de succès. Pour distinguer avec succès comment classer les réussites des échecs, les algorithmes sont généralement alimentés avec des entrées spécifiques à l'entreprise, au fondateur et à l'investisseur, qui peuvent aller d'une poignée d'attributs à quelques centaines.

La *Data science* peut offrir également de nombreux avantages aux startups en leur permettant de :

- Mieux comprendre leurs clients : Les startups peuvent utiliser des techniques d'analyse de données pour mieux comprendre les besoins et les comportements de leurs clients.

Chapitre 3 : Les *Uses Cases* de la *Data science* en finance

Cela peut leur permettre de développer des produits et des services plus adaptés aux attentes de leur marché cible.

- Optimiser leur stratégie marketing : En analysant les données de leurs campagnes publicitaires et de leur présence en ligne, les startups peuvent améliorer leur stratégie marketing en identifiant les canaux les plus efficaces pour toucher leur public et en adaptant leur message en fonction de l'audience visée.
- Prendre des décisions plus éclairées : La *Data science* peut aider les startups à prendre des décisions plus éclairées en leur fournissant des informations précises et factuelles sur leur entreprise, leur marché et leur environnement concurrentiel.
- Réduire les coûts et améliorer l'efficacité : En utilisant des techniques d'analyse de données, les startups peuvent identifier des opportunités d'optimisation des coûts et d'amélioration de l'efficacité dans leurs processus internes, ce qui peut leur permettre de gagner du temps et de l'argent.
- Innover et se différencier : En utilisant la *Data science* pour identifier des tendances émergentes dans leur marché, les startups peuvent innover et se différencier de leurs concurrents en proposant des produits et des services nouveaux et innovants.

2.2 Les *Data scientists* des « *assets* » pour les entreprises moderne :

Compte tenu des nombreuses possibilités offertes par la *Data science* pour le monde de la finance, les *Data Scientists* sont très recherchés dans ce secteur. Ces experts occupent désormais une place cruciale.

« *Data Scientist* » a été nommée en 2017 le métier en plus forte croissance par LinkedIn, et en 2018, Glassdoor a classé *data scientist* comme le meilleur métier aux États-Unis. De plus, une récente étude de *PriceWaterhouseCoopers* déclare que « *The best jobs right now in America include titles like data scientist, data engineer, and business analyst* »³³.

Selon le U.S. Bureau of Labor Statistics la demande pour ce métier a augmenté de 260% depuis 2012 et approximativement 11.5 million de poste d'emploi sera ouverte d'ici 2026 avec un salaire annuel de \$84,000 .

Ceci montre à quel point ces scientifiques jouent un rôle majeur dans les entreprises du 21^{ème} siècle . En effet , ces experts possèdent des compétences techniques avancées ainsi qu'une

³³ *Data science: One of the fastest growing occupations* - California University of Pennsylvania

Chapitre 3 : Les *Uses Cases* de la *Data science* en finance

connaissance approfondie de l'industrie. Ils sont en mesure d'utiliser des techniques de *Data science* pour, par exemple, détecter les fraudes ou créer des expériences personnalisées afin d'améliorer l'expérience client ou prédire les performances de l'entreprise en une année voir plus.

2.3 Comment le *Big Data* aide-t-il les entreprises numériques ?

L'emploi du *Big Data* est en plein essor grâce justement au large éventail de sources et de volumes de données disponibles en ligne. Réseaux sociaux, achats en ligne, films en streaming, des millions de dispositifs connectés... Tout cela fournit des informations précieuses qui permettent aux entreprises numériques de mieux connaître leur public, d'adapter leurs offres aux désirs et besoins des consommateurs et de se rapprocher d'eux du mieux possible pour le bénéfice de tous.

En analysant les données sur les habitudes de navigation et les préférences des utilisateurs, les entreprises numériques peuvent personnaliser l'expérience utilisateur pour répondre aux besoins spécifiques de chaque individu. Elle leur permet aussi de collecter des données sur les performances de leur site web ou de leur application, ainsi que sur les tendances du marché et les comportements des concurrents. Ces informations peuvent être utilisées pour prendre des décisions éclairées sur les investissements futurs et les changements à apporter à leur produit ou service.

En utilisant aussi les données collectées à partir de l'utilisation des produits et services, les entreprises numériques peuvent identifier les problèmes et les points d'amélioration, et ainsi proposer des améliorations pour améliorer la qualité de leur produit ou service et optimiser leurs processus internes en identifiant les domaines à faible efficacité et en proposant des améliorations pour rationaliser et automatiser les processus.

Des plateformes aussi diverses que **Spotify** et **Amazon** utilisent aujourd'hui le *Big Data* pour adapter leurs offres, tant en termes de contenu que de produit, à leurs utilisateurs. Connaître leurs goûts, leurs tendances de consommation, savoir quel produit ils ont acheté précédemment et à quelle fréquence ils changent de service, par exemple, sont les éléments qui leur permettent de se rapprocher de leurs clients de manière plus naturelle, assurant ainsi leur développement communautaire.

Chapitre 3 : Les *Uses Cases* de la *Data science* en finance

Conclusion

La data science est en train de devenir une composante essentielle de toutes les entreprises à l'échelle mondiale, et cette tendance devrait se maintenir à long terme. Les récents développements en matière d'automatisation basée sur l'intelligence artificielle ont entraîné d'importants changements dans le domaine de la *data science*. Cela se traduit par de nouvelles idées, de nouveaux intérêts et des investissements réorientés vers l'adoption de la *data science* par les entreprises. Cette technologie est suffisamment sophistiquée pour reconnaître les visages et les objets, ce qui a d'énormes implications pour diverses applications commerciales notamment le marketing, la détection de fraude le trading.....

Enfin, la *data science* joue également un rôle important pour les entreprises modernes et les startups. Elle leur donne un avantage concurrentiel en leur permettant d'exploiter efficacement les données pour prendre des décisions éclairées et plus rapides. La *data science* aide les startups à identifier les tendances du marché, à comprendre les besoins des clients et à ajuster rapidement leurs produits et services en conséquence

Chapitre 4
La gestion d'un portefeuille avec les
langages R et Python cas : Bourse
d'Alger

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

INTRODUCTION

Après avoir obtenu une vue d'ensemble de l'application de la *data science* en finance, nous nous concentrerons sur son application spécifique à la gestion de portefeuille, qui est le sujet principal de notre projet d'étude. Nous examinerons comment la *data science* peut être appliquée à un portefeuille composé d'actions d'entreprises cotées au sein de la SGBV d'Alger pour l'améliorer.

Dans ce chapitre, nous allons aborder différents aspects liés à la gestion de portefeuille dans le contexte du marché financier algérien, représenté par la SGBV (Société de la Gestion de la Bourse des Valeurs). Nous commencerons par présenter une vue d'ensemble du marché financier algérien, en mettant en avant son évolution et sa structure actuelle.

Pendant notre stage, nous avons acquis une compréhension approfondie du fonctionnement de la bourse et des mécanismes qui déterminent les cours des actions présentes dans notre jeu de données. Nous examinerons ces aspects dans le cadre de notre analyse, afin de mieux comprendre les facteurs qui influencent les performances des actions dans notre portefeuille. Dans le cadre de notre étude, notre portefeuille est constitué d'actions d'entreprises cotées en bourse (SGVB) telles que Alliance Assurance (ALL), Groupe SAIDAL (SAI), EGH EL AURASSI (AUR) et BIOPHARM (BIO). Nous avons investi des sommes spécifiques dans chacune de ces actions. Afin d'analyser ce portefeuille, nous ferons appel à des logiciels de programmation offrant des fonctionnalités avancées en termes de traitement des données et d'analyse financière.

Au cours de notre étude, nous explorerons différentes applications de la *data science* dans la gestion de portefeuille, telles que la prévision des rendements, l'analyse des risques, le calcul du bêta, le modèle d'évaluation des actifs financiers (CAPM), et bien d'autres. En utilisant des techniques de modélisation statistique et d'apprentissage automatique, nous chercherons à améliorer notre compréhension des performances de ces actions individuelles et de l'ensemble du portefeuille et savoir **Comment l'utilisation de la *data science* avec les langages de programmation R et Python peut-elle être bénéfique pour les acteurs du marché algérien dans l'analyse approfondie des risques liés à un portefeuille d'actions ?**

En combinant l'expertise financière traditionnelle avec les outils puissants de la *data science*, nous espérons obtenir des informations précieuses qui nous aideront à prendre des décisions d'investissement éclairées et à maximiser les rendements tout en gérant les risques.

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

Grâce à notre analyse approfondie et à l'utilisation des logiciels R et Python, nous visons à fournir des analyses concrets pour une future amélioration de notre portefeuille et pour la prise de décisions basées sur des données solides."

Section 01 : Présentation du marché financier algérien SGBV

Dans cette section, nous aborderons le processus de création de la bourse d'Alger, ainsi que l'évolution de sa structure au fil du temps.

1.1. Présentation de la SGBV :

La Société de Gestion de la Bourse des Valeurs , sous la forme abrégé SGBV, sise au 27 Bd Colonel Amirouche Alger, est une société par actions au capital social de 485 200 000, 00 DA, créée par le décret législatif n°93-10 du 23 mai 1993 et constituée le 25 mai 1997 ; elle constitue un cadre organisé et réglementé au service des IOB en tant que professionnels pour leur permettre d'exercer leurs missions conformément aux lois et règlements en vigueur.

Depuis sa création, elle s'est attelée à mettre en place les dispositifs opérationnels et techniques nécessaires aux transactions sur les valeurs mobilières admises en Bourse.

1.2. Fonctionnement de la bourse

La cote officielle de la Bourse des valeurs mobilières comporte un marché de titres de capital et un marché de titres de créance.

1.2.1. Le marché des titres de capital

Il est composé d'un:

A) Marché principal

Destiné aux grandes entreprises. Actuellement, quatre (04) sociétés sont cotées sur le marché principal à savoir :

- *Groupe SAIDAL* : activant dans le secteur pharmaceutique ;
- *EGH EL AURASSI* : activant dans le secteur du tourisme ;
- *ALLIANCE ASSURANCES* : activant dans le secteur des assurances ;
- *BIOPHARM* : activant dans le secteur pharmaceutique.

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

B) Marché PME

Réservé aux Petites et Moyennes Entreprises : créé en 2012 par le règlement COSOB n°12-01 du 12 janvier 2012 modifiant et complétant le règlement n°97-03 du 18 novembre 1997 relatif au règlement général de la bourse des valeurs mobilières. (Paru au JORA n°41 du 15 juillet 2012).

Ce marché peut offrir aux PME en démarrage une alternative pour accéder aux capitaux. C'est une excellente occasion de croissance pour les PME, tout en procurant aux investisseurs un marché bien réglementé pour leurs placements.

Une seule société activant dans le domaine du tourisme est cotée sur ce marché, il s'agit bel et bien de la société AOM Invest.

1.2.2 Le marché des titres de créance , composé d'un :

- A) Marché des obligations : titres de créance émis par les sociétés par actions.
- B) Marché bloc OAT : réservé aux Obligations Assimilables émises par le Trésor public Algérien. Il a été créé en 2008 et compte actuellement plus de 25 lignes cotées englobant près de 400 milliard de Dinar algérien .
- C) Les Obligations Assimilables du Trésor présentant des maturités de 7, 10 et 15 ans sont négociées par l'entremise des Intermédiaires en Opérations de Bourse et les compagnies d'assurances ayant le statut de Spécialistes en Valeurs du Trésor à concurrence de cinq séances par semaine.

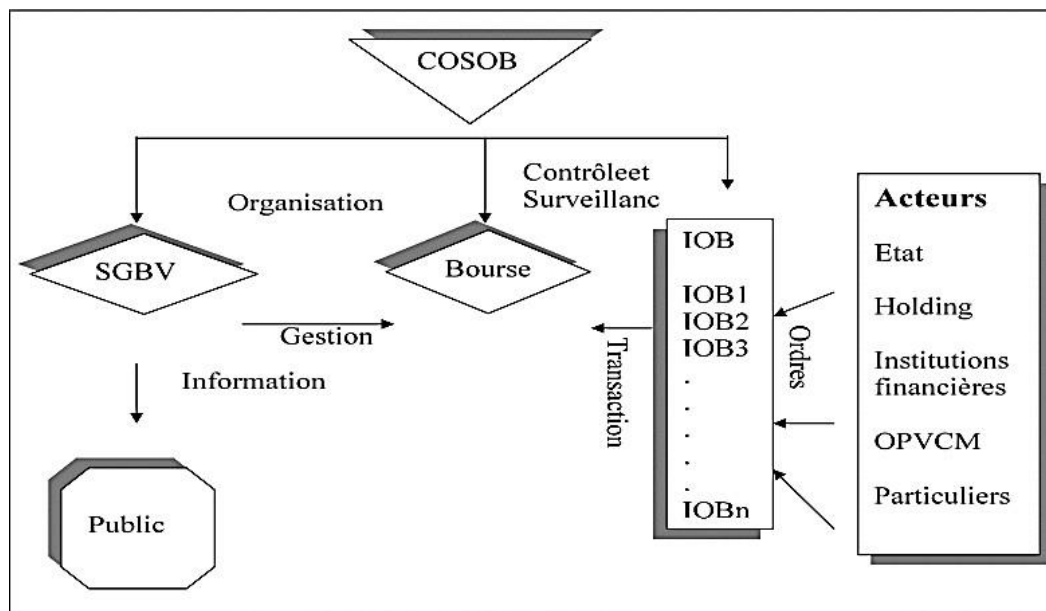
1.3. Structure et organisation de la Bourse d'Alger :

Dans le but de concrétiser la création d'un marché des capitaux en Algérie, la Bourse des Valeurs Mobilières d'Alger a été établie en janvier 1998 grâce à un partenariat entre la Commission d'Organisation et de Surveillance des Opérations de Bourse (COSOB) d'Algérie et une agence d'exécution québécoise (Canada), Coopers & Lybrand.

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

L'organisation de la bourse des valeurs mobilières d'Alger et la mission de ses acteurs peut être représentée par le schéma suivant :

Figure 13 Le système Algérien mis en place pour l'organisation du marché boursier



Source : SONATRACH la revue N 8, Octobre 1997, P18

Le schéma ci-dessus, représente les différents acteurs intervenants sur le marché boursier, qui sont :

- La société de gestion des valeurs mobilières (SGBV).
- La commission d'organisation et de surveillance des opérations de bourse (COSOB).
- Les intermédiaires en opération de bourse (IOB).
- Les autres acteurs qui sont : Etat, holding, entreprises, institutions financières, OPCVM et particuliers.

1.3.1 La commission d'organisation & de surveillance des opérations de bourse (COSOB)

La Commission d'Organisation et de Surveillance des Opérations de Bourse (COSOB) ont pour mission d'organiser et de surveiller le marché des valeurs mobilières en veillant notamment :

- à la protection des investisseurs en valeurs mobilières ;
- au bon fonctionnement et à la transparence du marché des valeurs mobilières.

Pour assurer ces missions, elle est dotée des pouvoirs :

- réglementaire ;

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

- surveillance et de Contrôle ;
- disciplinaire et Arbitral.

composition :

- d'un Président nommé par un décret présidentiel pour un mandat de 04 ans ;
- six membres nommés par arrêté du Ministre des Finances pour une durée de 04 ans.

1.3.2 Les intermédiaires en opération de bourse (IOB)

L'activité d'Intermédiaire en Opération de Bourse est réglementée, surveillée et contrôlée par la Commission d'organisation et de Surveillance du marché, elle est soumise à l'agrément de celle-ci.

L'exercice de l'une ou de plusieurs activités (citées en dessous) est soumis à l'agrément de la COSOB.

L'activité d'I.O. B peut être exercée par des sociétés commerciales ayant pour objet principal l'intermédiation en opérations de bourse, ainsi que les banques et les établissements financiers, ayant reçu l'agrément de la commission.(voir Règlement COSOB n° 15-01 du 25 Joumada Rthania 1436 correspondant au 25 avril 2015 relatif aux conditions d'agrément, aux obligations et au contrôle des intermédiaires en opérations de Bourse).

Actuellement la place financière d'Alger compte onze IOB représentant les Banques publiques suivantes : BADR, BDL, BEA, BNA, CNEP-Banque, CPA et quatre banques privées BNP Paribas El Djazaïr, Société Générale Algérie, Banque Al Baraka d'Algérie, Al Salam Bank Algeria et une société privée Tell Markets.

Missions:

L'intermédiaire en opérations de bourse désigne tout intermédiaire agréé qui intervient en bourse pour négocier pour le compte de ses clients, ou pour son propre compte des valeurs mobilières et autres produits financiers cotés et droits s'y rapportant.

L'intermédiaire en opérations de bourse peut exercer, également, une ou plusieurs des activités suivantes :

- Le conseil en placement de valeurs mobilières ;
- La gestion individuelle de portefeuille en vertu d'un contrat écrit ;
- La gestion de portefeuille d'organismes de placement collectif en valeurs mobilières ;

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

- Le placement de valeurs mobilières et de produits financiers ;
- La garantie de bonne fin et la prise ferme d'émission de titres ;
- La négociation pour le propre compte de l'intermédiaire ;
- La conservation et l'administration de valeurs mobilières ;
- Le conseil aux entreprises en matière de structure de capital, de fusion et de rachat d'entreprises.

1.3.3 Les organismes de placement collectifs en valeurs mobilières OPCVM

Les OPCVM sont des sociétés ou fonds, qui ont pour objet de constituer et gérer un portefeuille de valeurs mobilières et autres produits financiers pour le compte de tiers.³⁴

Ils sont habilités à collecter de l'épargne, aussi modeste soit elle, en vue d'être investie sur le marché financier selon une politique de placement bien définie.

On distingue deux grandes familles d'OPCVM : les FCP (fonds communs de placement) et les SICAV (sociétés d'investissement à capital variable). Ces deux entités se distinguent non pas par le mode de leur financement, mais par leur nature juridique : la SICAV est une société par action tandis que le FCP est un fonds de porteurs de capitaux

1.3.4 Algérie clearing

Algérie clearing est un organisme veillé à une plus grande sécurité dans le déroulement des opérations sur titres, une plus grande célérité dans le dénouement des transactions ainsi qu'une gestion professionnelle des titres tout en appliquant une mise à niveau par rapport aux standards internationaux

1.4 Le processus de cotations et négociation des titres

La Société de la Gestion de la Bourse de Valeurs (SGBV) établit le programme des séances de cotation pour l'année en cours, ainsi que les horaires d'ouverture et de clôture de ces séances. Les titres peuvent être cotés en continu ou en fixing, en fonction de leur liquidité. Après la réception de l'ordre par l'Intermédiaire en Opérations de Bourse (IOB), celui-ci est transmis à la bourse pour être exécuté.

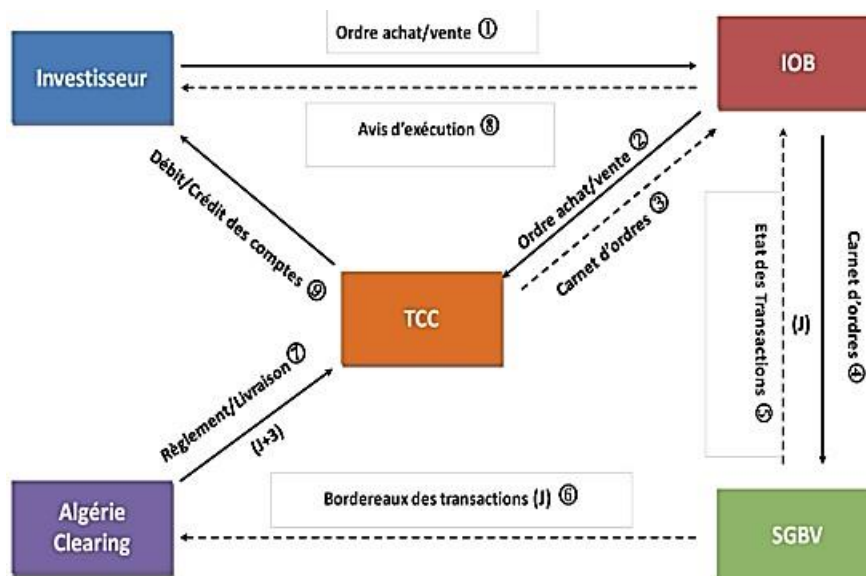
³⁴ <https://www.cosob.org/opcvm/>

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

Les séances de négociations de bourse s'effectuent au niveau du parquet de la bourse les dimanche , mardis et les jeudis de 9h30 à 11h00.

L'acquisition et / ou la vente de titres via la bourse peut être schématisée comme suit :

Figure 14 l'acquisition et la vente de titres



Source : <https://www.algerieclearing.dz/>

Les négociations se font automatiquement par le biais d'un système de cotation électronique avec la méthode du fixing. :

La cotation au fixing consiste en l'application d'un seul cours coté à l'ensemble des transactions pour chaque titre lors d'une séance de négociation. Ce cours permet de :

- Maximiser les volumes des titres échangés ;
- Réduire le déséquilibre du coté acheteur ou vendeur ;
- Minimiser la volatilité du cours par rapport au cours de clôture de la dernière séance de bourse.

La fin de la séance, sont réalisées les opérations de blocs qui représentent les ordres d'achat et de vente portant sur un grand nombre d'actions et/ou d'obligations. Ces transactions doivent respecter les cours cotés sur le marché.

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

Le processus de négociation se déroule selon les étapes suivantes ³⁵

1.4.1 Ouverture de la séance :

Les IOB se présentent à la bourse pour la séance de négociation pour transmettre leurs ordres d'achat et de vente transcrits sur des registres d'ordres normalisés dans une option manuelle, ou pour les saisir dans une option automatisée. Après vérification de la conformité des ordres par rapport aux règles définies par la SGBV (écarts par rapport aux cours de référence³⁶, pas de cotation, taille de lots...), les ordres sont acceptés

1.4.2 Détermination du cours :

La cotation s'effectue au fixing et, selon le cas, la cote est calculée avec les ordres clients et non clients (livre combiné).

1.4.3 Allocation des ordres :

Après avoir déterminé le volume des titres à allouer, les IOB sont servis selon un cours unique. Si on est en situation de déséquilibre, un ou deux ratios d'allocation partielle sont calculés. Les IOB reçoivent des états récapitulant leurs ordres et les allocations auxquels seront annexés des billets de transactions (un billet par transaction) qui serviront comme support à la compensation.

1.4.4 Effectuer les transactions en bloc :

Les transactions en bloc sont des transactions dont le volume des titres de capital est supérieur à 5000 titres ou dont la valeur des titres de créance est supérieure à 3 000 000 DA, ces transactions subissent un traitement particulier. Elles ont lieu à la fin de la séance au prix du fixing arrêté plus ou moins 1%

1.4.5 Clôture de la séance de négociation :

Cette étape est caractérisée par :

- l'annonce des résultats par la SGBV ;
- la remise des registres d'ordres et des billets de transactions aux IOB ;
- la diffusion des supports d'information : BOC (bulletin officiel de la cote), rapport de

³⁵ Repris dans REMIDI Djoumana « le marché boursier algérien, contraintes et perspectives », mémoire de magister

en sciences économique option Monnaie Finance et Banque, Tizi Ouzou, 2013, p.100

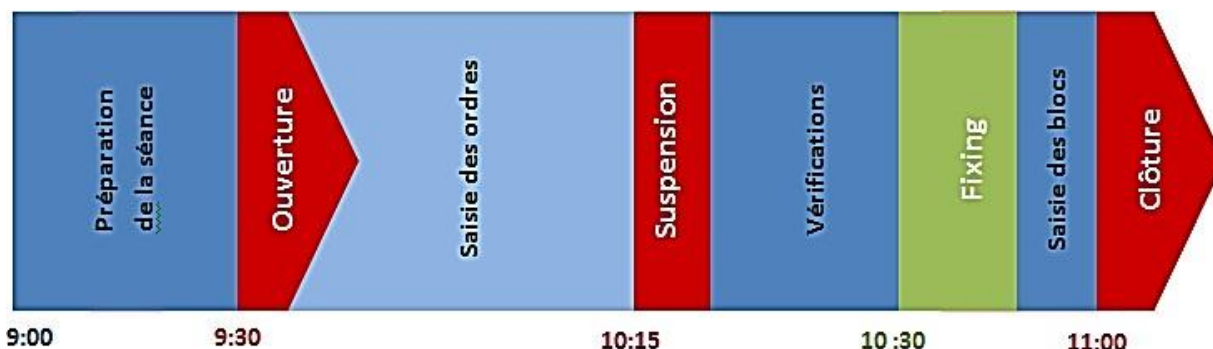
³⁶ Cours de référence : c'est le dernier cours lors d'une séance de bourse

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

transactions, sommaire de transaction...

La séance de cotation peut être synthétisée comme suit :

Figure 15 Déroulement de la séance de cotation à la Bourse d'Alger



Source : www.sgbv.dz

Au sein de la bourse, et surtout lors de la cotation, un logiciel d'interface inspiré par Fox Pro ou SQL Server a été développé pour la gestion des cotations et des calculs comme nous l'avons constaté. Ces logiciels permettent de collecter et de traiter les données en temps réel lors des séances de cotation.

Le logiciel d'interface joue un rôle clé dans le processus de cotation en fournissant une plateforme permettant de recueillir les informations relatives aux ordres d'achats et de ventes, prix de clôture des actions, aux volumes échangés et aux variations de valeurs. Il peut également effectuer des calculs et des analyses sur ces données en utilisant des formules et des algorithmes spécifiques. Le logiciel d'interface est généralement connecté aux systèmes de négociation et de gestion de la bourse, lui permettant ainsi d'accéder aux données en temps réel et de les enregistrer de manière organisée.

Une fois la séance de cotation clôturée, toutes les données pertinentes de cette séance sont enregistrées. Cela inclut les prix de clôture des actions, les volumes échangés et les variations de valeurs. Ces données sont essentielles pour analyser les performances du marché et pour la gestion de portefeuille.

Ces données enregistrées, comprenant également les données historiques des séances précédentes, seront utilisées dans notre étude sur la gestion de portefeuille dans la section suivante. Elles nous permettront d'analyser les tendances, les fluctuations et les performances des différentes actions sur une base annuelle. En examinant ces données, nous pourrions évaluer

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

la rentabilité des investissements, identifier les opportunités et les risques, et prendre des décisions éclairées pour la gestion du portefeuille.

Dans la section précédente, nous avons mis en évidence le fonctionnement de la bourse. Dans la section suivante, nous aborderons la gestion de portefeuille et présenterons des données de cotations améliorées. Cette section mettra l'accent sur la manière dont nous gérons nos investissements et les informations clés utilisées pour prendre des décisions éclairées.

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

Section 2: Gestion de portefeuille avec R studio

Pour notre recherche, nous allons utiliser des données Excel contenant les cours des actions des entreprises cotées en bourse pour la période du 8 février 2022 au 7 février 2023 (voir annexe 1). Nous effectuerons plusieurs analyses sur ces données, notamment le calcul du rendement du portefeuille, du risque et du ratio de Sharpe et ceux en utilisant les logiciel R et Python .

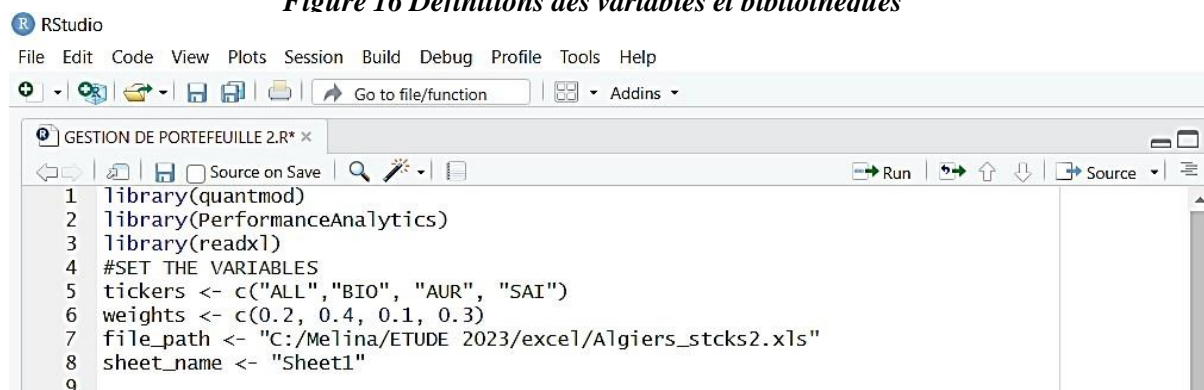
Cette section sera dédiée en premier lieu a la gestion du portefeuille avec le R studio et ces bibliothèques quantmod et Performance Analytics

2.1 Préparation des données

Avant de commencer l'analyse, nous devons charger les bibliothèques nécessaires et définir les variables.

- Les bibliothèques quantmod, PerformanceAnalytics et readxl sont utilisées pour importer les données du portefeuille à partir d'un fichier Excel . Celle-ci sont charger avec la fonction library()
- Nous spécifions également les symboles des actions, les pondérations, le chemin du fichier Excel et le nom de la feuille de calcul.

Figure 16 Définitions des variables et bibliothèques



```
1 library(quantmod)
2 library(PerformanceAnalytics)
3 library(readxl)
4 #SET THE VARIABLES
5 tickers <- c("ALL", "BIO", "AUR", "SAI")
6 weights <- c(0.2, 0.4, 0.1, 0.3)
7 file_path <- "C:/Melina/ETUDE 2023/excel/Algiers_stcks2.xls"
8 sheet_name <- "Sheet1"
9
```

Source : Sortie logiciel R

- *Tickers* : Designe les symboles des actions : ALL pour Alliance Assurance , BIO pour Biopharm , AUR pour Egh-El-Aurassi et SAI pour Saidal
- *Weights* : Designe le poids des actions cad la proportion de la valeur totale du portefeuille qui est investie dans cette action spécifique. Il est souvent exprimé en pourcentage et calcule

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

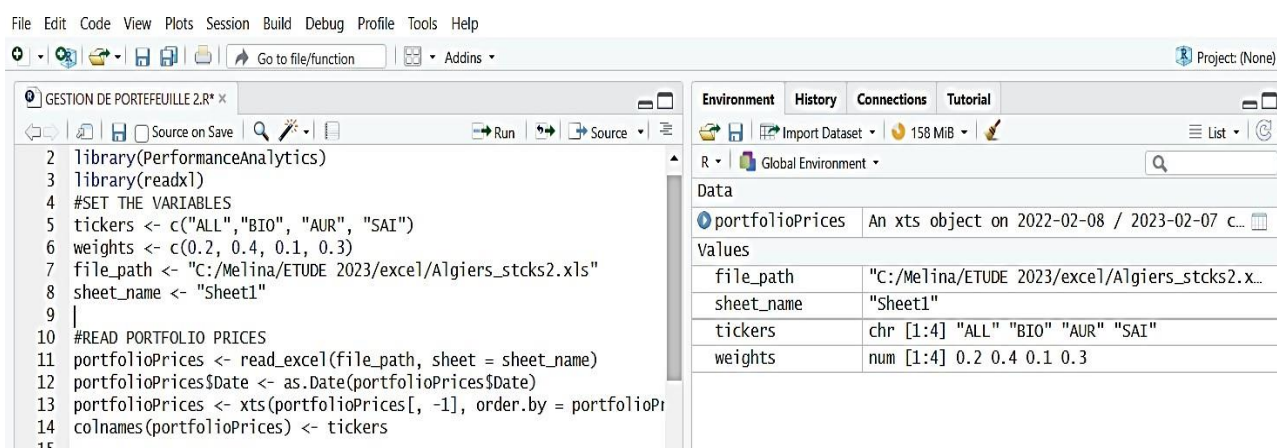
comme suit : **Poids de l'action = (Valeur de l'investissement dans l'action / Valeur totale du portefeuille) * 100** . Dans notre cas l'investisseur investi 10000 DA dans son portefeuille 2000 DA dans ALL , 4000 DA dans BIO ,1000 DA dans AUR et 3000 DA dans SAI . Le poids des actions est ainsi de 0,2 ; 0,4 ; 0,1 et 0,2 respectivement .

- *File_path* : Designe le chemin du fichier Excel et *sheet_name* le no m de la page dans le fichier

2.2 Importation des prix du portefeuille

Nous utilisons la library (readxl)pour importer les données de prix du portefeuille à partir du fichier Excel spécifié. Les prix sont ensuite convertis en une série chronologique (xts) pour faciliter les calculs ultérieurs. Nous renommons également les colonnes avec les symboles des actions correspondantes.

Figure 17 Lecture des fichiers et du portefeuille



```
2 library(PerformanceAnalytics)
3 library(readxl)
4 #SET THE VARIABLES
5 tickers <- c("ALL", "BIO", "AUR", "SAI")
6 weights <- c(0.2, 0.4, 0.1, 0.3)
7 file_path <- "C:/Melina/ETUDE 2023/excel/Algiers_stcks2.xls"
8 sheet_name <- "Sheet1"
9 |
10 #READ PORTFOLIO PRICES
11 portfolioPrices <- read_excel(file_path, sheet = sheet_name)
12 portfolioPrices$Date <- as.Date(portfolioPrices$Date)
13 portfolioPrices <- xts(portfolioPrices[, -1], order.by = portfolioP
14 colnames(portfolioPrices) <- tickers
15
```

file_path	"C:/Melina/ETUDE 2023/excel/Algiers_stcks2.x...
sheet_name	"Sheet1"
tickers	chr [1:4] "ALL" "BIO" "AUR" "SAI"
weights	num [1:4] 0.2 0.4 0.1 0.3

Source : Sortie logiciel R

- la fonction *read_excel* du package *readxl* est utilisée pour importer les données de prix du portefeuille à partir du fichier Excel spécifié. Nous passons le chemin du fichier et le nom de la feuille de calcul en tant qu'arguments. Les données sont stockées dans la variable *portfolioPrices*.
- Pour s'assurer que la colonne Date est reconnue comme un format de date, nous utilisons la fonction *as.Date* pour convertir la colonne Date en format de date. Cela permettra de travailler avec des dates correctement dans les analyses ultérieures.
- la fonction *xts* du package *quantmod* est utile pour créer une série chronologique des prix du portefeuille. Nous spécifions les données à inclure (*portfolioPrices[, -1]*), ce qui exclut la colonne Date) et l'argument *order.by* pour indiquer que les données doivent être

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

ordonnées en fonction de la colonne Date. La série chronologique est ensuite stockée dans la variable *portfolioPrices* qui ensuite s'affiche dans l'environnement

- Nous utilisons ensuite la fonction **colnames** pour attribuer les noms de colonnes appropriés à la série chronologique *portfolioPrices*. Les noms de colonnes sont définis en fonction des symboles des actions spécifiés dans la variable *tickers*.

2.3 Calcul du rendement du portefeuille :

Le rendement du portefeuille fait référence aux gains ou pertes réalisés par un portefeuille d'investissement composé de plusieurs types de placements. Les portefeuilles visent à générer des rendements en fonction des objectifs déclarés de la stratégie d'investissement, ainsi que de la tolérance au risque des investisseurs ciblés par le portefeuille. En général, pour un portefeuille comportant *n* actifs le rendement attendu (espérance) est calculé avec la formule suivante :

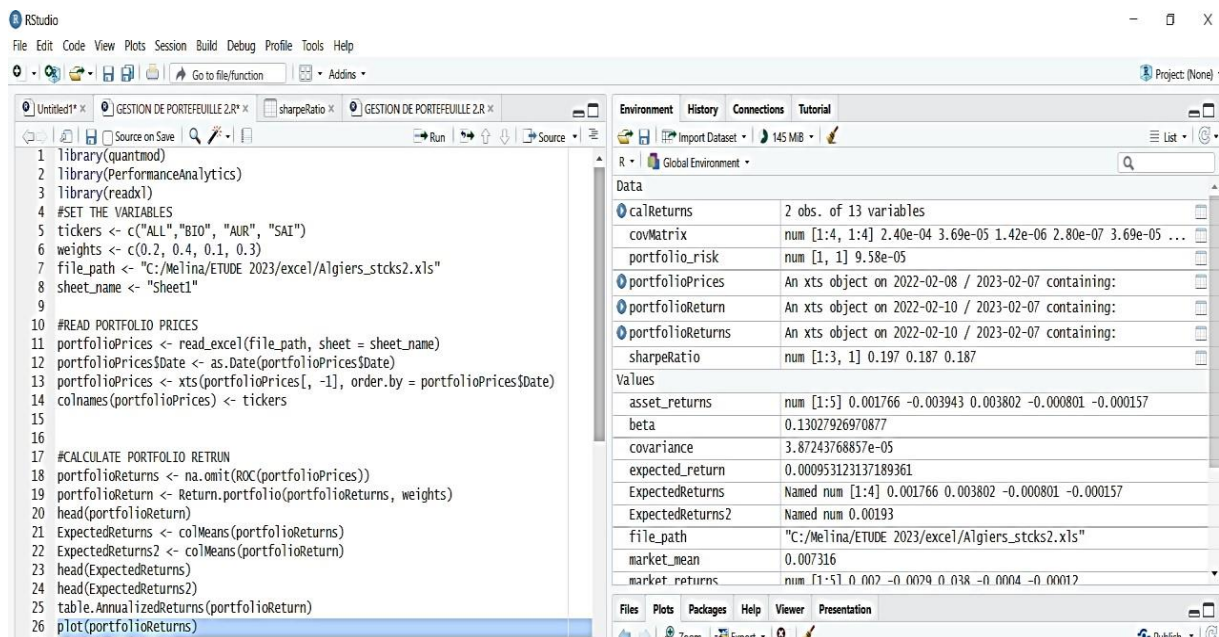
$$E(Rp) = \sum_{k=0}^n w_i E(R_i)$$

Ou *w_i* : représente le poids des actions

R_i : représente le rendement des actifs

Dans le logiciel R, le calcul du rendement attendu peut être effectué en utilisant la fonction *Return.portfolio* du package *PerformanceAnalytics*. Cette fonction prend en compte les poids attribués à chaque actif et les rendements de chaque actif, et retourne le rendement attendu du portefeuille.

Figure 18 Calcul du rendement du portefeuille



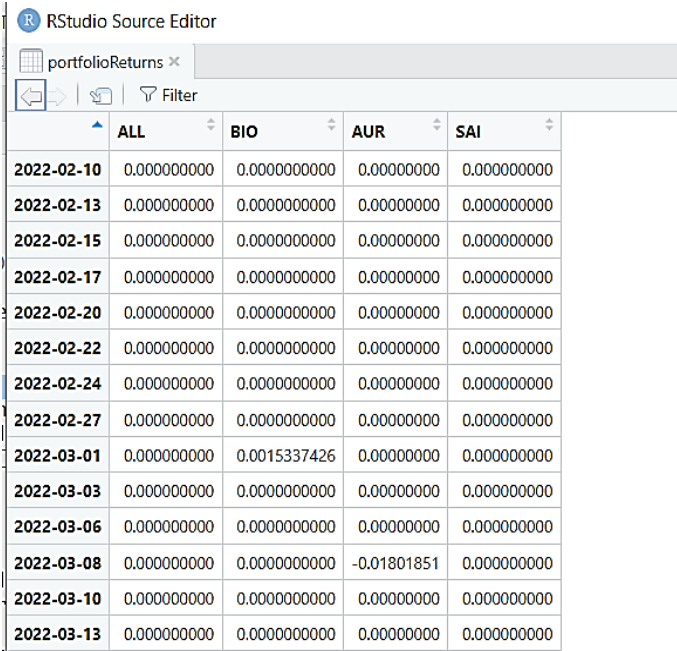
Source : Sortie logiciel R

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

2.3.1 Rendements journaliers des actifs

La fonction *PortfolioReturns* Indique les rendements quotidiens du portefeuille à partir des prix du portefeuille , il est calculé avec la fonction *ROC* du package *quantmod*. La fonction *na.omit* quand a elle est utilisée pour supprimer les valeurs manquantes dans les rendements calculés. Dans l'environnement R amélioré, les calculs effectués nous fournissent un tableau des différents rendements journaliers des actions de chaque actif du portefeuille (ALL, BIO, AUR, SAI). Ce tableau nous permet d'analyser la performance de chaque actif individuellement et de comprendre les variations quotidiennes de leurs rendements. (voir annexe 2)

Figure 19 Rendements journaliers des actions



	ALL	BIO	AUR	SAI
2022-02-10	0.000000000	0.000000000	0.000000000	0.000000000
2022-02-13	0.000000000	0.000000000	0.000000000	0.000000000
2022-02-15	0.000000000	0.000000000	0.000000000	0.000000000
2022-02-17	0.000000000	0.000000000	0.000000000	0.000000000
2022-02-20	0.000000000	0.000000000	0.000000000	0.000000000
2022-02-22	0.000000000	0.000000000	0.000000000	0.000000000
2022-02-24	0.000000000	0.000000000	0.000000000	0.000000000
2022-02-27	0.000000000	0.000000000	0.000000000	0.000000000
2022-03-01	0.000000000	0.0015337426	0.000000000	0.000000000
2022-03-03	0.000000000	0.000000000	0.000000000	0.000000000
2022-03-06	0.000000000	0.000000000	0.000000000	0.000000000
2022-03-08	0.000000000	0.000000000	-0.01801851	0.000000000
2022-03-10	0.000000000	0.000000000	0.000000000	0.000000000
2022-03-13	0.000000000	0.000000000	0.000000000	0.000000000

Source : Sortie logiciel R

2.3.2 Rendement global journalier du portefeuille

La fonction *Return.portfolio* du package *PerformanceAnalytics* calcule le rendement global du portefeuille en utilisant les rendements du portefeuille (*PortfolioReturns*) et les poids attribués à chaque actif. Les rendements du portefeuille sont pondérés selon les poids spécifiés dans la variable *weights* . Les résultat du portefeuille s'affiche dans l'environnement ainsi (voir annexe 3)

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

On remarque que la plupart des rendements sont de zéro, ce qui indique qu'il n'y a pas eu de variation significative de la valeur du portefeuille ces jours-là. Cependant, il y a des jours où les rendements sont différents de zéro (exemple : 01-03-2022 variation de 0.0006132971), ce qui peut être dû à des fluctuations du marché ou à des mouvements spécifiques des actifs détenus dans le portefeuille.

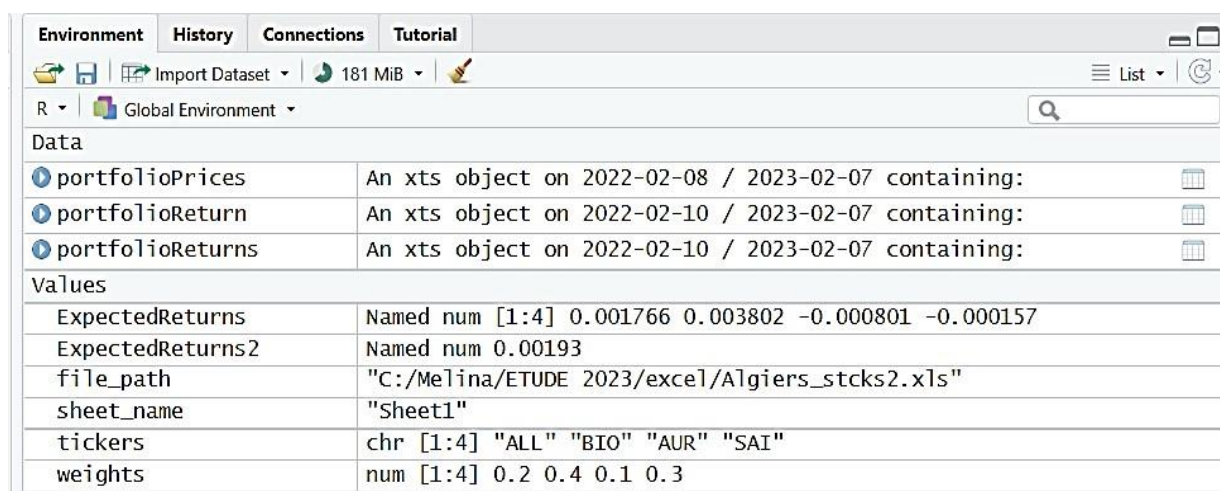
2.3.3 Esperance de rendements des actifs

La variable *ExpectedReturns* stocke les rendements attendus moyens de chaque actif du portefeuille. La fonction *colMeans* est utilisée pour calculer la moyenne des rendements pour chaque colonne (actif) du tableau *portfolioReturns*. Dans la section « *values* » de l'environnement on trouve les rendements espéré de chaque actifs ALL, BIO, AUR, SAI respectivement . (voir figure 18).

2.3.4 Esperance de rendement global du portefeuille

La variable *ExpectedReturns2* stocke les rendements attendus moyens du portefeuille global. Elle calcule le rendement global du portefeuille.

Figure 20 Espérance de rendement du portefeuille



Environment	History	Connections	Tutorial
R Global Environment			
Data			
portfolioPrices	An xts object on 2022-02-08 / 2023-02-07 containing:		
portfolioReturn	An xts object on 2022-02-10 / 2023-02-07 containing:		
portfolioReturns	An xts object on 2022-02-10 / 2023-02-07 containing:		
Values			
ExpectedReturns	Named num [1:4] 0.001766 0.003802 -0.000801 -0.000157		
ExpectedReturns2	Named num 0.00193		
file_path	"C:/Melina/ETUDE 2023/excel/Algiers_stcks2.xls"		
sheet_name	"Sheet1"		
tickers	chr [1:4] "ALL" "BIO" "AUR" "SAI"		
weights	num [1:4] 0.2 0.4 0.1 0.3		

Source : Sortie logiciel R

- Pour l'actif ALL, le rendement est de 0.0017658033, ce qui indique une légère augmentation de la valeur de cet actif
- Pour l'actif BIO, le rendement est de 0.0038018288, ce qui suggère une augmentation plus significative de la valeur de cet actif par rapport aux autres.

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

- Pour l'actif AUR, le rendement est de -0.0008011625, ce qui indique une légère baisse de la valeur de cet actif.
- Pour l'actif SAI, le rendement est de -0.0001573527, ce qui suggère une légère baisse de la valeur de cet actif.

Le résultat de rendement indiqué pour le portefeuille est de 0.001932809, représentant un rendement positif . Il suggère une augmentation de la valeur du portefeuille sur la période donnée (Un rendement positif peut être considéré comme une amélioration par rapport à un rendement négatif ou à une absence de rendement)

2.3.5 Rendement et risque annuel du portefeuille

Table.AnualizedReturns(portfolioReturn) est utilisée pour calculer les rendements annualisés d'un portefeuille d'investissement. Dans notre cas , le rendement annuel est de 0.1029 . Cela signifie que le portefeuille a généré un gain moyen de 10.29% par an au cours de la période étudiée.

Figure 21 Rendement et risque annuel du portefeuille

```
> table.AnualizedReturns(portfolioReturn)
                                portfolio.returns
Annualized Return                  0.1029
Annualized Std Dev                 0.0706
Annualized Sharpe (Rf=0%)         1.4578
> |
```

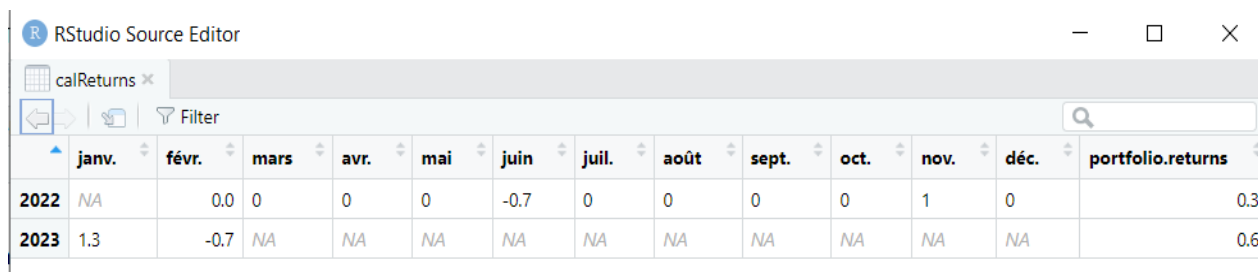
Source : Sortie logiciel R

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

2.3.6 Rendement mensuelle et annuelle du portefeuille

La fonction `table.CalendarReturns(portfolioReturn)` génère un tableau des rendements du portefeuille par période calendaire. Cela permet d'analyser la performance du portefeuille sur une base mensuelle, trimestrielle ou annuelle, en identifiant les périodes de rendement positif ou négatif.

Figure 22 rendement mensuelle et annuelle du portefeuille



	janv.	févr.	mars	avr.	mai	juin	juil.	août	sept.	oct.	nov.	déc.	portfolio.returns
2022	NA	0.0	0	0	0	-0.7	0	0	0	0	1	0	0.3
2023	1.3	-0.7	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0.6

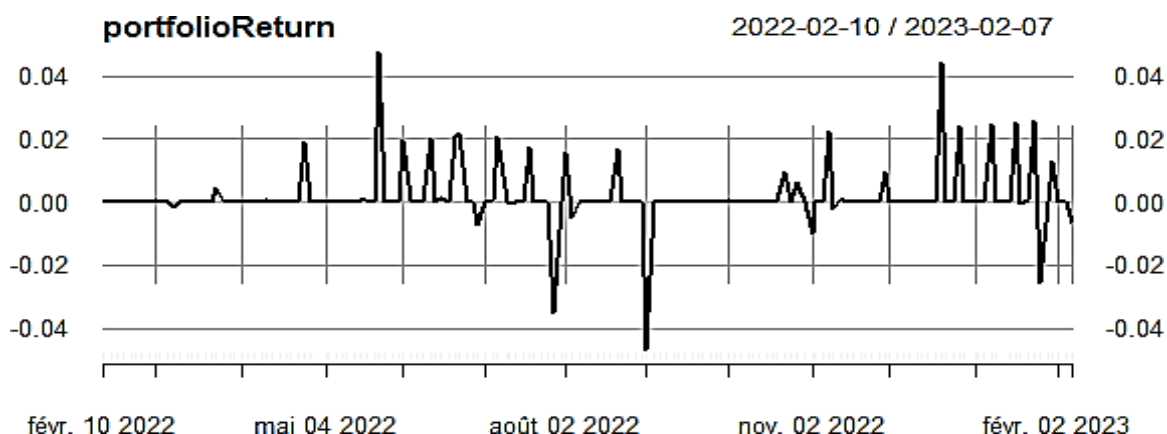
Source : Sortie logiciel R

C'est à dire que pour chaque 10000 DA investi ,nous allons recevoir un rendement de 0.3 pour l'année 2022 et 0.6 pour l'année 2023

2.3.7 Graphique des rendements du portefeuille.

La fonction `plot(portfolioReturn)` permet trace les rendements du portefeuille au fil du temps. L'axe des x représente les dates, et l'axe des y représente les rendements du portefeuille. Chaque point sur le graphique correspond à une observation de rendement à une date spécifique. Comme on peut le voir dans la figure ci-dessus

Figure 23 Graphique des rendements du portefeuille.



Source : Sortie logiciel R

Le graphe montre une période de stagnation des rendements entre février 2022 et mai 2022. Ensuite, on observe des fluctuations positives des rendements entre mai et août, atteignant

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

un maximum d'environ 0,04. Cependant, il y a eu une forte baisse au milieu et à la fin du mois d'août 2022, les rendements chutant jusqu'à environ -0,04. Cette baisse peut être expliquée par des facteurs tels qu'une correction du marché, des événements économiques défavorables ou des informations négatives affectant les actifs du portefeuille.

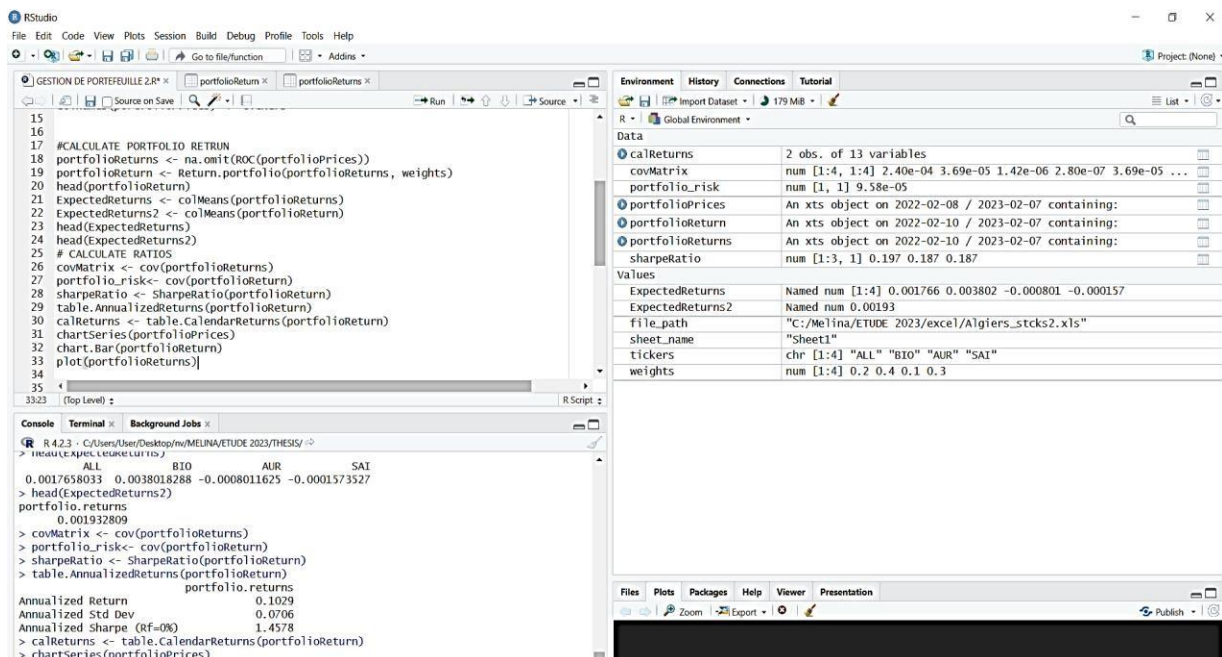
Par la suite, les rendements sont restés relativement stables jusqu'en novembre, puis ont augmenté entre janvier et février. Cependant, vers la fin de février, il y a eu une légère baisse avec les rendements atteignant environ -0,02.

En résumé, les calculs effectués dans cette partie nous permettent d'obtenir les rendements attendus moyens pour chaque actif du portefeuille et pour le portefeuille global. Ces valeurs sont importantes pour évaluer la performance attendue de chaque actif et du portefeuille dans son ensemble, et pour prendre des décisions de gestion basées sur ces estimations de rendements.

2.4 Calcul des ratios :

Dans le code **# CALCULATE RATIOS**, nous effectuons plusieurs calculs et analyses pour évaluer la performance et le risque du portefeuille. Voici une explication des calculs et des notions associées

Figure 24 Calcul des ratios



```
15
16
17 #CALCULATE PORTFOLIO RETURN
18 portfolioReturns <- na.omit(ROC(portfolioPrices))
19 portfolioReturn <- Return.portfolio(portfolioReturns, weights)
20 head(portfolioReturn)
21 ExpectedReturns <- colMeans(portfolioReturns)
22 ExpectedReturns2 <- colMeans(portfolioReturn)
23 head(ExpectedReturns)
24 head(ExpectedReturns2)
25 # CALCULATE RATIOS
26 covMatrix <- cov(portfolioReturns)
27 portfolio_risk <- cov(portfolioReturn)
28 sharpeRatio <- SharpeRatio(portfolioReturn)
29 table.AnualizedReturns(portfolioReturn)
30 caIReturns <- table.CalendarReturns(portfolioReturn)
31 chartSeries(portfolioPrices)
32 chart.Bar(portfolioReturn)
33 plot(portfolioReturns)
34
35
```

Environment

Object	Class	Attributes
caIReturns	num	2 obs. of 13 variables
covMatrix	num	[1:4, 1:4] 2.40e-04 3.69e-05 1.42e-06 2.80e-07 3.69e-05 ...
portfolio_risk	num	[1, 1] 9,58e-05
portfolioPrices	xts	An xts object on 2022-02-08 / 2023-02-07 containing:
portfolioReturn	xts	An xts object on 2022-02-10 / 2023-02-07 containing:
portfolioReturns	xts	An xts object on 2022-02-10 / 2023-02-07 containing:
sharpeRatio	num	[1:3, 1] 0.197 0.187 0.187

Values

ExpectedReturns	Named num [1:4]	0.001766 0.003802 -0.000801 -0.000157
ExpectedReturns2	Named num	0.00193
file_path	chr	"C:/Melina/ETUDE 2023/excel/Algiers_stcks2.xls"
sheet_name	chr	"Sheet1"
tickers	chr [1:4]	"ALL" "BIO" "AUR" "SAI"
weights	num [1:4]	0.2 0.4 0.1 0.3

Console

```
R 4.2.3 > C:/Users/User/Desktop/m/MELINA/ETUDE 2023/THESIS/
> head(ExpectedReturns)
  ALL      BIO      AUR      SAI
0.0017658033 0.0038018288 -0.0008011625 -0.0001573527
> head(ExpectedReturns2)
0.001932809
> covMatrix <- cov(portfolioReturns)
> portfolio_risk <- cov(portfolioReturn)
> sharpeRatio <- SharpeRatio(portfolioReturn)
> table.AnualizedReturns(portfolioReturn)
  portfolio_returns
Annualized Return      0.1029
Annualized Std Dev      0.0706
Annualized Sharpe (Rf=0%) 1.4578
> caIReturns <- table.CalendarReturns(portfolioReturn)
> chartSeries(portfolioPrices)
```

Source : Sortie logiciel R

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

Dans la fenêtre d'environnement et d'historique, on peut observer différents ratios tels que la covariance (covmatrix) et le Sharpe ratio. Ces ratios fournissent des mesures essentielles pour évaluer la performance et le risque d'un portefeuille. Voici une explication détaillée de ces notions

2.4.1 Covariance (covMatrix) :

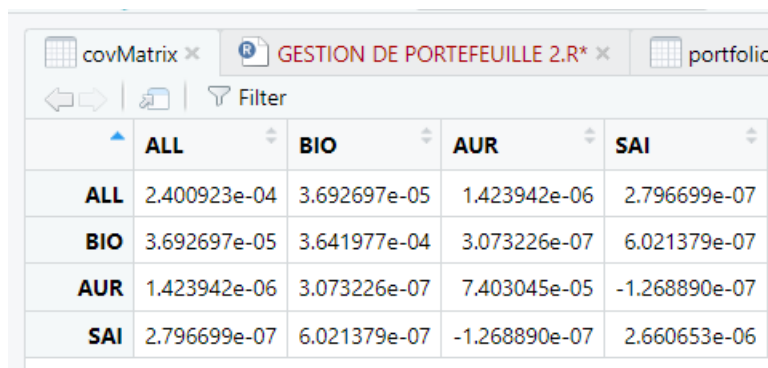
En finance, cette notion permet de mesurer le degré de liaison des fluctuations de deux titres entre eux, ou encore d'un titre avec un indice.³⁷ La covariance peut être vue comme le produit des valeurs de deux variables moins le produit des deux moyennes. Mathématiquement, la formule est la suivante

$$Cov(x, y) = \frac{1}{N} \sum_{i=1}^n (X_i - X)(Y_i - Y) \text{ Avec}$$

X_i = cours de l'actif X à l'instant i
 Y_i = cours de l'actif Y à l'instant i
 X = moyenne du cours de l'actif X
 Y = moyenne du cours de l'actif Y
 N = nombres de périodes

Dans le logiciel R, la fonction `cov {covMatrix <- cov(portfolioReturns)}` est utilisée pour calculer la matrice de covariance des rendements du portefeuille (`portfolioReturns`) et (`portfolioReturn`) comme indiqué ci-dessus .

Figure 25 Covariance des rendements de portefeuille



	ALL	BIO	AUR	SAI
ALL	2.400923e-04	3.692697e-05	1.423942e-06	2.796699e-07
BIO	3.692697e-05	3.641977e-04	3.073226e-07	6.021379e-07
AUR	1.423942e-06	3.073226e-07	7.403045e-05	-1.268890e-07
SAI	2.796699e-07	6.021379e-07	-1.268890e-07	2.660653e-06

Source : Sortie logiciel R

Les résultats affichés dans la figure représentent la matrice de covariance (covMatrix) entre les rendements de différents actifs dans un portefeuille. Par exemple la covariance entre ALL et BIO est positive ce qui indique qu'il y a une corrélation positive entre les rendements de ces deux

³⁷ https://www.abcbourse.com/apprendre/19_variance_covariance.

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

actifs. Cad que lorsque le rendement de l'actif "ALL" augmente, il y a tendance à observer une augmentation du rendement de l'actif "BIO".

En remarque aussi que la covariance entre les rendements des actifs "ALL" et "AUR" est de 1.423942e-06. Cette valeur suggère une faible corrélation entre les rendements de ces deux actifs. La relation entre les rendements de "ALL" et "AUR" est proche de zéro, ce qui indique qu'ils se déplacent indépendamment l'un de l'autre.

2.4.2 Sharpe ratio

Mis en place en 1966 par William Forsyth Sharpe, un économiste américain, le ratio de Sharpe permet de mesurer la rentabilité d'un portefeuille en fonction du risque pris. En d'autres termes c'est le quotient de l'excès de rentabilité par rapport au taux sans risque divisé par le risque total du portefeuille. Il permet de calculer la performance d'un investissement par rapport à celle d'un placement sans risque ³⁸ Sa formule mathématique est :

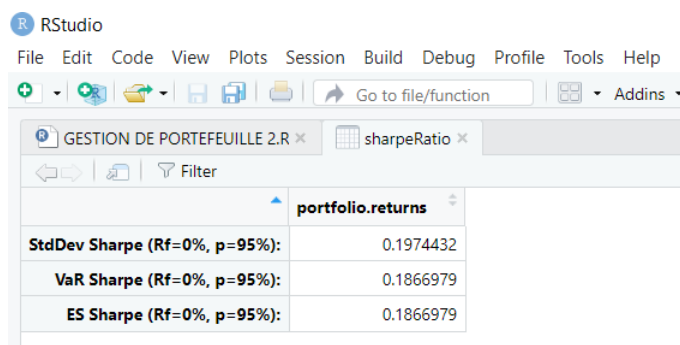
$$Sp = \frac{Rp - Pf}{\sigma_p}$$

Avec

- Sp : Ratio de Sharpe du portefeuille
- Rp : rendement du portefeuille risque P
- Rf : Taux sans risque
- σ_p : volatilité du portefeuille risque P

La fonction `sharpeRatio <- SharpeRatio(portfolioReturn)` est utilisée pour calculer le ratio de Sharpe du portefeuille à partir des rendements globaux du portefeuille. Une fois lancé le résultat s'affiche ainsi

Figure 26 Sharpe ratio



	portfolio.returns
StdDev Sharpe (Rf=0%, p=95%):	0.1974432
VaR Sharpe (Rf=0%, p=95%):	0.1866979
ES Sharpe (Rf=0%, p=95%):	0.1866979

Source : Sortie logiciel R

³⁸ https://www.abcbourse.com/apprendre/19_ratio_de_sharpe.html

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

Les mesures de rendement et de risque fournies pour le portefeuille sont le rendement Sharpe, l'écart-type (StdDev Sharpe), la Value at Risk (VaR) et l'Expected Shortfall (ES), toutes calculées avec un taux sans risque (R_f) de 0% et un niveau de confiance de 95%.

A. La *Standard Deviation* (StdDev Sharpe)

C'est une mesure statistique qui indique la dispersion des rendements d'un portefeuille par rapport à sa moyenne. En d'autres termes, il mesure la variabilité des rendements. Dans ce cas, l'écart-type du portefeuille est de 0.1974432 c'est à dire que pour chaque unité de risque prise dans le portefeuille, le rendement excédentaire généré par le portefeuille est égal à 0.1974432 unités. En somme, le portefeuille offre un rendement supérieur à celui du taux sans risque, compte tenu du niveau de risque mesuré par la volatilité des rendements.

B. La *Value at Risk* (VaR sharpe) ;

Dans sa forme la plus générale, la VaR mesure la perte potentielle de valeur d'un actif ou d'un portefeuille risqué sur une période définie pour un intervalle de confiance donné. Le chiffre "0.1866979" représente le Sharpe ratio calculé en utilisant la Value at Risk (VaR) comme mesure de risque pour le portefeuille ceci indique que, pour chaque unité de risque mesurée par la VaR à un niveau de confiance de 95%, le portefeuille génère une perte maximale de 0.1866979 unités (ou 18.67%) par rapport au taux sans risque. Le Sharpe ratio basé sur la VaR permet d'évaluer la performance ajustée au risque du portefeuille en prenant en compte les pertes potentielles maximales estimées par la VaR.

C. *L'Expected Shortfall* (ES)

Également connu sous le nom de Conditional VaR, est une mesure qui va au-delà de la VaR en fournissant une estimation de la perte moyenne en cas de dépassement de la VaR. L'ES est généralement considéré comme une mesure plus conservatrice, car il prend en compte les pertes qui se produisent au-delà de la VaR. Dans ce cas, l'ES du portefeuille est également de 0.1866979, ce qui indique que si la VaR est dépassée, la perte moyenne du portefeuille serait de 0.1866979.

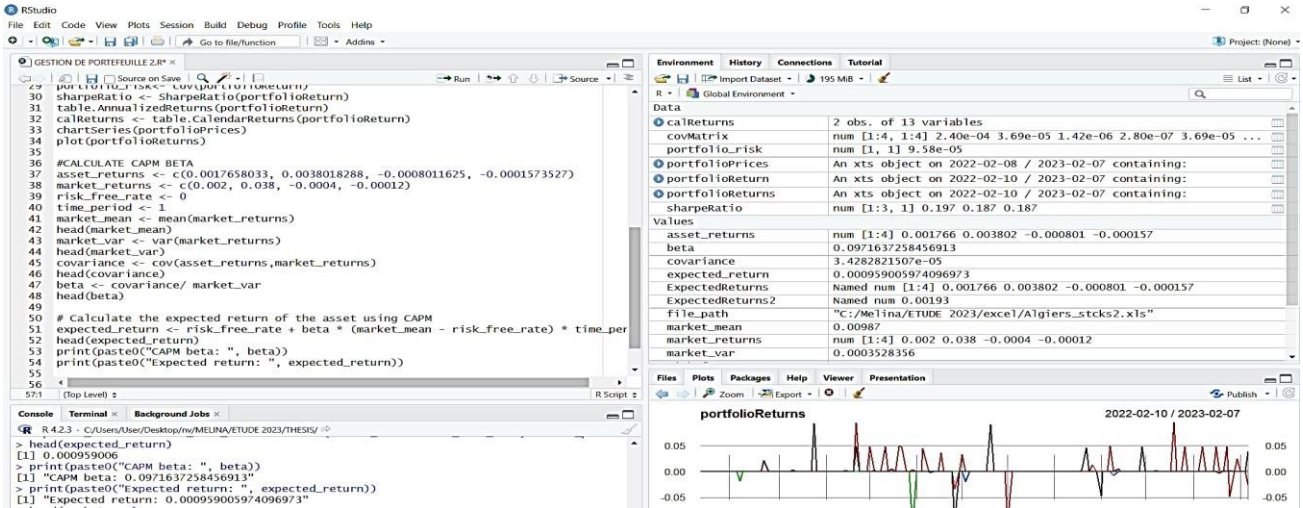
2.5 Calcul du Beta avec la methode de Capital Asset Pricing Model (CAPM)

Le CAPM (Capital Asset Pricing Model) est un modèle utilisé en finance pour estimer le rendement attendu d'un actif financier. Il repose sur l'idée que l'investisseur est récompensé pour le risque supplémentaire qu'il assume en investissant dans un actif risqué par rapport à un actif sans

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

risque. Le principal paramètre du CAPM est le coefficient bêta (beta), qui mesure la sensibilité d'un actif aux fluctuations du marché.

Figure 27 Calcule du CAPM Beta



Source : Sortie logiciel R

Dans cette partie nous allons calculer différents paramètres notamment la moyenne la variance des rendements du marché

Nous allons commencer par assigner les variables dont nous aurons besoin

- `asset_returns <- c(0.0017658033, 0.0038018288, -0.0008011625, -0.0001573527)` et `market_returns <- c(0.002, 0.038, -0.0004, -0.00012)` Ces lignes de code définissent les rendements des actifs individuels (*ExpectedReturn*) et les rendements du marché (*market_returns*). Ces rendements sont généralement exprimés en pourcentage et représentent les variations des prix des actifs ou du marché sur une période donnée.
- `risk_free_rate <- 0` représente le taux sans risque, qui est le taux de rendement garanti sur un investissement considéré comme totalement sûr, généralement représenté par les bons du Trésor d'un pays. Nous avons choisi 0 car toutes les obligations sont arrivées à échéance durant notre période.
- `time_period <- 1` représente la période de temps sur laquelle les rendements sont mesurés, c'est-à-dire une année.

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

2.5.1 La moyenne des rendements du marché.

La moyenne des rendements du marché est une mesure statistique qui peut être utilisée pour évaluer la tendance générale des rendements sur une période donnée. Pour la calculer nous allons utiliser la variable `market_mean` et la fonction `market_mean <- mean(market_returns)` qui représente le rendement moyen du marché sur la période considérée. Dans notre cas la moyenne de rendement de marche est de 0.0987 comme indique dans l'environnement.

2.5.2 La variance du marché

En utilisant la formule `market_var <- var(market_returns)`, nous avons calculé la variance des rendements du marché et obtenu une valeur de 0.0003528356 ainsi il est possible d'interpréter que les rendements du marché ont une certaine volatilité ou dispersion autour de leur moyenne.

La covariance entre rendements de l'actif individuel et les rendements du marché

La fonction `covariance <- cov(asset_returns, market_returns)` calcule la covariance entre les rendements de l'actif individuel et les rendements du marché. La covariance mesure la relation linéaire entre les deux séries de rendements soit 33.428282e-05 pour l'ensemble du portefeuille .

2.5.3 Le Beta

Le coefficient beta est une mesure de la volatilité, ou risque systémique, d'un titre individuel par rapport au risque non systémique du marché dans son ensemble. En termes statistiques, beta représente la pente de la droite par la régression des points de données des rendements d'un titre individuel par rapport à ceux du marché.

Ce coefficient se calcule comme le ratio de la covariance entre la rentabilité d'un portefeuille (R_p) et celle du marché (R_m), par la variance de la rentabilité implicite du marché (R_m) . Sa formule est donc

$$\beta_i = \frac{\text{cov}(R_i, R_m)}{\sigma^2}$$

ou $\text{COV}(R_i, R_M)$ représente la covariance entre le rendement du titre i et celle de marché M .

$\sigma^2 M$: représente la variance de la rentabilité du marché.

En utilisant la formule "`beta <- covariance / market_var`", nous avons calculé un coefficient bêta de 0.09716373 pour notre portefeuille. Ce résultat indique que notre portefeuille est moins sensible aux fluctuations du marché par rapport à la moyenne du marché. En d'autres termes, notre portefeuille est relativement moins volatile que le marché dans son ensemble, ce qui peut être considéré comme un avantage en termes de gestion des risques.

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

2.6 Calculer l'esperance de rendement du portefeuille avec la methode CAPM

Dans cette partie nous allons estimer le rendement attendu du portefeuille en prenant en compte son risque systématique (beta) par rapport au marché dans son ensemble. Cela fournit aux investisseurs une estimation de la rémunération qu'ils pourraient attendre en investissant dans ce portefeuille , en fonction du risque qu'ils sont prêts à assumer.

La ligne de code `Expected_return <- risk_free_rate + beta * (market_mean - risk_free_rate) * time_period` calcule le rendement attendu de l'actif en utilisant la formule CAPM. Le rendement attendu est calculé en ajoutant au taux sans risque (*variable risk_free_rate*) le produit du bêta de l'actif (beta) par l'écart entre le rendement moyen du marché (*market_mean*) et le taux sans risque, le tout multiplié par la période de temps (*time_period*). Le rendement attendu obtenu dans l'environnement de 0.000959006 peut être interprété comme l'estimation du rendement potentiel de l'actif sur une période d'un an, compte tenu de son risque et de sa relation avec le marché

Dans cette section, nous avons développé un algorithme de gestion de portefeuille en utilisant le langage de programmation R. Cet algorithme nous a permis de réaliser des calculs sur différents ratios, tels que le ratio rendement/risque et le ratio de Sharpe.

Le recours à R dans l'analyse du portefeuille s'est avéré extrêmement bénéfique. En effet, R nous a offert une solution puissante et efficace pour effectuer les calculs et les analyses nécessaires à la gestion du portefeuille. Comparé aux outils traditionnels tels qu'Excel, R se distingue par ses bibliothèques spécialisées et sa grande flexibilité.

Dans la section suivante, nous aborderons la gestion de portefeuille en utilisant un autre outil de la *data science* , à savoir Python. Nous effectuerons les mêmes analyses que précédemment, mais cette fois-ci en exploitant les fonctionnalités et les bibliothèques spécifiques de Python

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

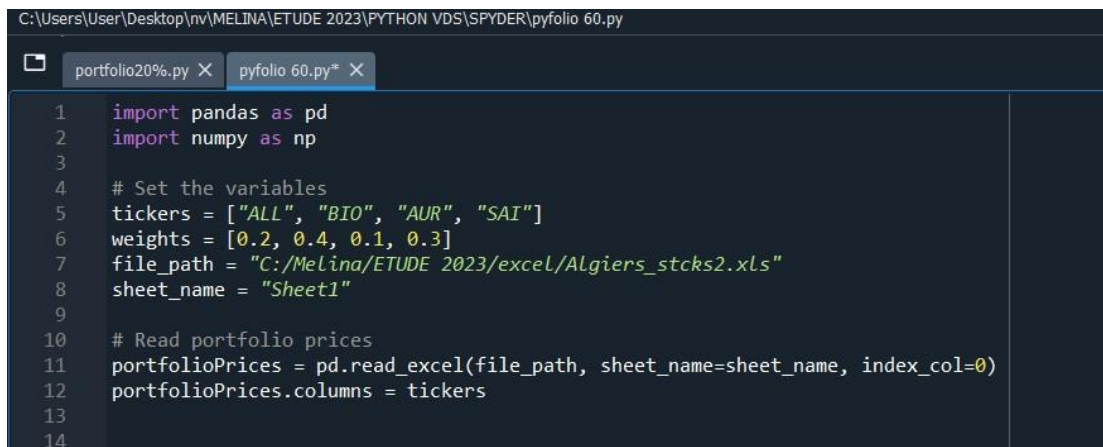
Section 3 : Gestion de portefeuille avec Spyder (Python)

Dans cette section nous allons traiter les mêmes données que celle que nous avons traitée dans la partie précédente avec le logiciel de programmation Python à savoir l'IDE Spyder qui est un environnement de développement pour Python qui intègre de nombreuses bibliothèques importantes dans la finance

3.1 Préparation des données

Dans l'environnement de développement Spyder, les bibliothèques pandas et numpy sont couramment utilisées pour effectuer des analyses de portefeuille. Ainsi, nous commencerons par importer ces bibliothèques, ce qui nous permettra de bénéficier de leurs fonctionnalités avancées pour manipuler et analyser les données du portefeuille.

Figure 28 Importation des fichiers et assignation des variables



```
C:\Users\User\Desktop\nv\MELINA\ETUDE 2023\PYTHON VDS\SPYDER\pyfolio 60.py
portfolio20%.py X pyfolio 60.py X
1 import pandas as pd
2 import numpy as np
3
4 # Set the variables
5 tickers = ["ALL", "BIO", "AUR", "SAI"]
6 weights = [0.2, 0.4, 0.1, 0.3]
7 file_path = "C:/Melina/ETUDE 2023/excel/Algiers_stcks2.xls"
8 sheet_name = "Sheet1"
9
10 # Read portfolio prices
11 portfolioPrices = pd.read_excel(file_path, sheet_name=sheet_name, index_col=0)
12 portfolioPrices.columns = tickers
13
14
```

Source : Notre étude sur le logiciel Python

Une fois les bibliothèques importées le processus est assez similaire à celui du R. En premiers temps il est primordial de définir les variables :

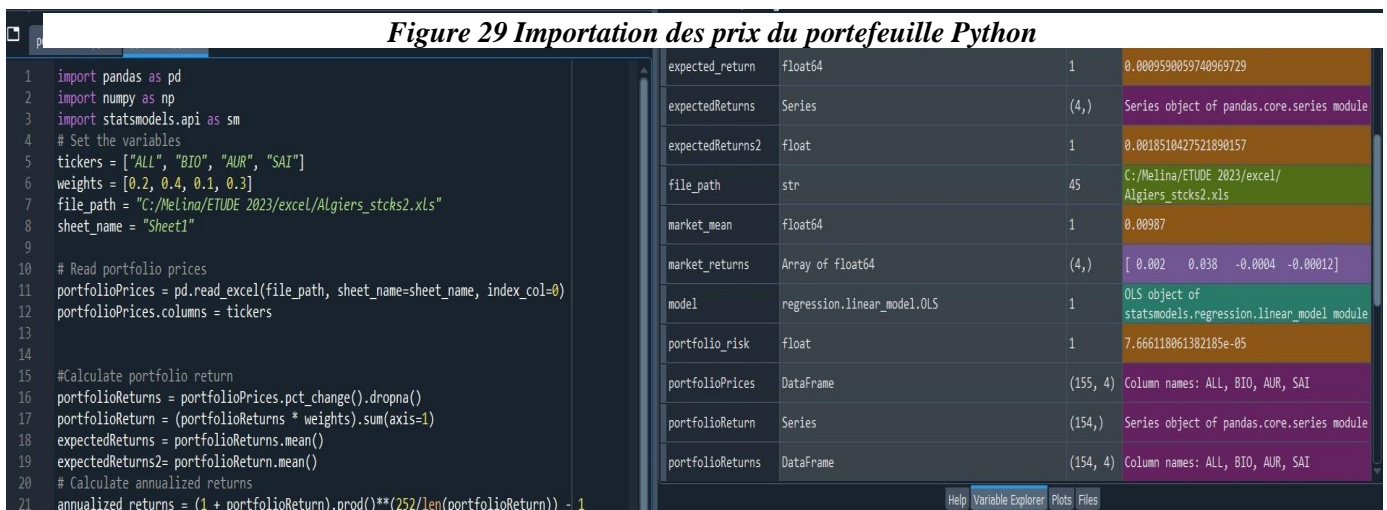
- *tickers* : Liste des symboles des actifs dans le portefeuille. Les symboles sont utilisés pour identifier les colonnes correspondantes dans le fichier de données : ALL pour Alliance Assurance, BIO pour Biopharm, AUR pour Egh-El-Aurassi et SAI pour Saidal
- *weights* : Liste des poids correspondants aux actifs dans le portefeuille. Les poids définissent la proportion de chaque actif dans le portefeuille.
- *file_path* : Chemin vers le fichier Excel contenant les prix du portefeuille. C'est l'emplacement du fichier sur notre système.

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

- *sheet_name* : Nom de la feuille dans le fichier Excel à lire. La feuille spécifiée contient les données des prix du portefeuille

3.2 Importation des prix du portefeuille:

Dans cette Partie nous allons utiliser la fonction `portfolioPrices = pd.read_excel(file_path, sheet_name=sheet_name, index_col=0)`. Cette ligne de code lit le fichier Excel spécifié (voir annexe 1), extrait les données de la feuille spécifiée et stocke les données dans un DataFrame appelé `portfolioPrices`. Ensuite l'argument `index_col=0` est utilisé pour définir la première colonne (colonne des dates) comme index et la ligne de code `portfolioPrices.columns = tickers` assigne les noms des colonnes du DataFrame "portfolioPrices" aux valeurs contenues dans la variable "tickers".



Source : Notre étude sur le logiciel Python

3.3 Calcul des rendements du portefeuille :

Une fois les données téléchargées sur l'environnement de spyder, nous allons calculer le rendement des actifs ainsi que le rendement global du portefeuille en tant que compte des poids :

3.3.1 Rendements journaliers des actifs

Nous utiliserons le code `portfolioReturns = portfolioPrices.pct_change().dropna()` pour calculer les rendements du portefeuille en utilisant la méthode `pct_change()` du DataFrame `portfolioPrices`. La méthode `pct_change()` calcule le pourcentage de changement entre chaque période

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

consécutive .En d`autre terme cette fonction donne les rendements journaliers des actions de chaque actif du portefeuille (ALL, BIO, AUR, SAI).(voir annexe 4)

Figure 30 les rendements journaliers des actions Python

Date	ALL	BIO	AUR	SAI
2022-02-10	0	0	0	0
2022-02-13	0	0	0	0
2022-02-15	0	0	0	0
2022-02-17	0	0	0	0
2022-02-20	0	0	0	0
2022-02-22	0	0	0	0
2022-02-24	0	0	0	0
2022-02-27	0	0	0	0
2022-03-01	0	0.00153492	0	0
2022-03-03	0	0	0	0
2022-03-06	0	0	0	0
2022-03-08	0	0	-0.0178571	0
2022-03-10	0	0	0	0
2022-03-13	0	0	0	0
2022-03-15	0	0	0	0

Source :Notre étude sur le logiciel Python

3.3.2 Rendements globale journalier du portefeuille

L'opération effectuée $portfolioReturn = (portfolioReturns * weights).sum(axis=1)$ est une multiplication élément par élément entre les rendements (*portfolioReturns*) et les poids (*weights*), ce qui attribue à chaque rendement son poids respectif. Cela permet de calculer le rendement du portefeuille pondéré (voir annexe 5)

Figure 31 Rendements journaliers du portefeuille Python

Date	0
2022-02-10	0
2022-02-13	0
2022-02-15	0
2022-02-17	0
2022-02-20	0
2022-02-22	0
2022-02-24	0
2022-02-27	0
2022-03-01	0.000613968
2022-03-03	0
2022-03-06	0
2022-03-08	-0.00178571
2022-03-10	0
2022-03-13	0

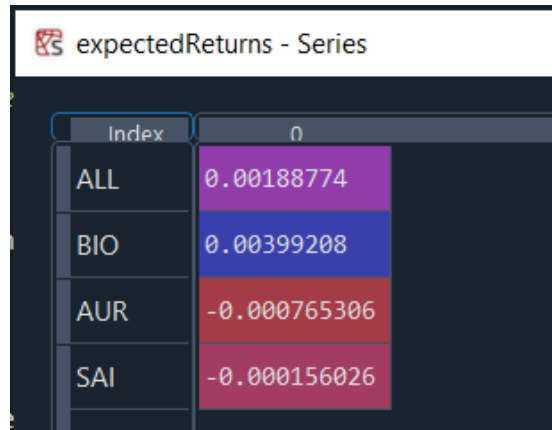
Source :Notre étude sur le logiciel Python

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

3.3.2 Esperance de rendement des actifs :

La méthode `mean()` de la fonction `expectedReturns = portfolioReturns.mean()` nous permet de calculer les rendements moyens de chaque actif dans le portefeuille comme représenté dans la figure suivante:

Figure 32 Espérance de rendement des actifs



Index	0
ALL	0.00188774
BIO	0.00399208
AUR	-0.000765306
SAI	-0.000156026

Source : Notre étude sur le logiciel Python

3.3.3 Esperance de rendement du portefeuille

Représenté par `ExpectedReturns2 = portfolioReturn.mean()` : Dans cette partie on calcule le rendement global du portefeuille en utilisant la méthode `mean()` sur le vecteur `portfolioReturn`. Celle-ci calcule la valeur moyenne de tous les éléments du vecteur. Le résultat obtenu 0.0018510427521890157 représente la performance globale du portefeuille en termes de rendement moyen.

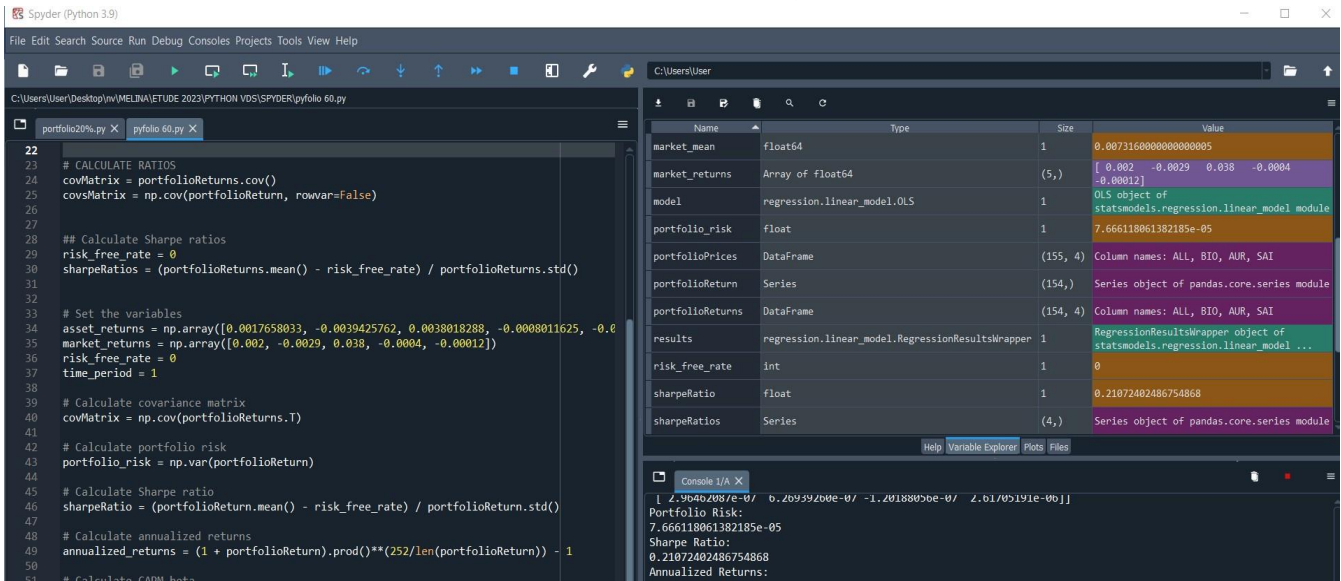
Remarque : On remarque que les résultats obtenus sont généralement similaires avec les logiciels R , avec de légères différences qui peuvent être attribuées aux variations des algorithmes de calcul utilisés dans chaque logiciel.

3.4 Calcule des ratios :

Dans le code `# CALCULATE RATIOS` , nous effectuons plusieurs calculs et analyses pour évaluer la performance et le risque du portefeuille avec Spyder

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

Figure 33 Calcul des ratios Python



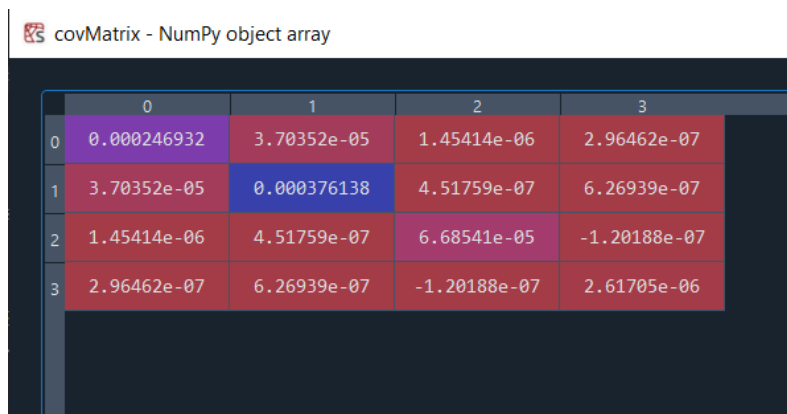
Source :Notre étude sur le logiciel Python

Dans la fenêtre d'environnement et d'historique, on peut observer différents ratios tels que la covariance (covmatrix) et le Sharpe ratio. Ces ratios fournissent des mesures essentielles pour évaluer la performance et le risque d'un portefeuille. Voici une explication détaillée de ces notions

3.4.1 Covariance (covMatrix)

Dans cette fonction nous allons utiliser le code `covMatrix = portfolioReturns.cov()` pour mesurer les relations linéaires entre les rendements de chaque paire d'actifs. Elle fournit des informations sur la façon dont les rendements des actifs varient ensemble.

Figure 34 Covariance des actifs



Source :Notre étude sur le logiciel Python

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

Le 0,1,2,3 représente ALL, BIO, AUR, SAI respectivement tandis que les résultats représentent le degré de liaison entre les actifs

Le code `covsMatrix = np.cov(portfolioReturn, rowvar=False)` quant a lui calcule la relation de covariance entre tous les actifs du portefeuille. Elle représente la façon dont les rendements des actifs varient ensemble en fonction de leurs pondérations respectives dans le portefeuille. Pour notre cas la covariance est de $3.872438e-05$ comme calcule précédemment

3.4.2 Le Sharpe Ratio Le ratio de Sharpe peut être calculé directement comme suit :

Figure 35 Fonction sharpe ratio Python

```
# Calculate Sharpe ratio
asset_returns = np.array([0.0017658033, 0.0038018288, -0.0008011625, -0.0001573527])
market_returns = np.array([0.002, 0.038, -0.0004, -0.00012])
risk_free_rate = 0
time_period = 1
```

Source :Notre étude sur le logiciel Python

On va commencer par assigner les variable (comme dans le R)

- `asset_returns = np.array([0.0017658033, -0.0039425762, 0.0038018288, -0.0008011625, -0.0001573527])`
- `market_returns = np.array([0.002, 0.038, -0.0004, -0.00012])`
- `risk_free_rate = 0`
- `time_period = 1`

Ensuite nous allons utiliser la fonction `sharpeRatio = (portfolioReturn.mean() - risk_free_rate) / portfolioReturn.std()` qui nous donne un Sharpe ratio globale de : 0.21072402486754868

Remarque :

Le calcul du ratio de Sharpe peut varier entre Python et R en raison des différences dans les bibliothèques utilisées et les méthodes de calcul implémentées. Le python utilise la méthode standard de

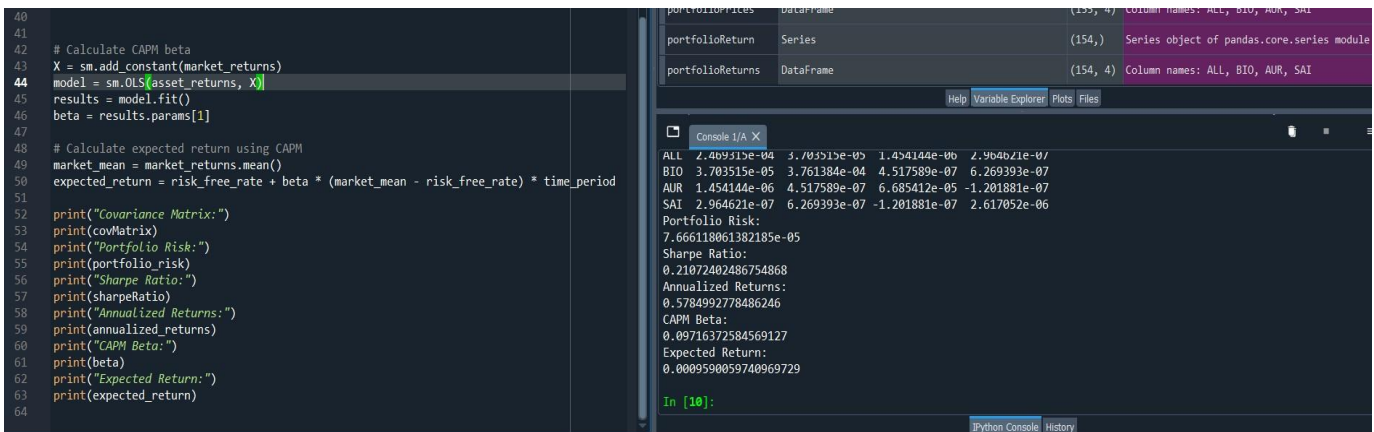
Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

calculé tandis que le R utilise plusieurs méthodes à savoir la *standard déviation*, la *Value at Risk* et la *Expected shortfall*

3.5 Calcul du beta et du rendement espéré avec la méthode CAPM

La dernière partie du programme sera dédiée à calculer le rendement espéré du portefeuille en utilisant la méthode du CAPM

Figure 36 Calcul du rendement espéré avec la méthode



```
40
41
42 # Calculate CAPM beta
43 X = sm.add_constant(market_returns)
44 model = sm.OLS(asset_returns, X)
45 results = model.fit()
46 beta = results.params[1]
47
48 # Calculate expected return using CAPM
49 market_mean = market_returns.mean()
50 expected_return = risk_free_rate + beta * (market_mean - risk_free_rate) * time_period
51
52 print("Covariance Matrix:")
53 print(covMatrix)
54 print("Portfolio Risk:")
55 print(portfolio_risk)
56 print("Sharpe Ratio:")
57 print(sharpeRatio)
58 print("Annualized Returns:")
59 print(annualized_returns)
60 print("CAPM Beta:")
61 print(beta)
62 print("Expected Return:")
63 print(expected_return)
64
```

Output:

```
ALL 2.409319e-04 3.705515e-05 1.454144e-06 2.904621e-07
BIO 3.703515e-05 3.761384e-04 4.517589e-07 6.269393e-07
AUR 1.454144e-06 4.517589e-07 6.685412e-05 -1.201881e-07
SAI 2.964621e-07 6.269393e-07 -1.201881e-07 2.617052e-06
Portfolio Risk:
7.666118061382185e-05
Sharpe Ratio:
0.21072402486754868
Annualized Returns:
0.5784992778486246
CAPM Beta:
0.09716372584569127
Expected Return:
0.0009590059740969729
In [10]:
```

Source :Notre étude sur le logiciel Python

Pour Obtenir les résultats nous devons suivre les étapes suivantes

- A. Ajout de la constante : Nous ajoutons une colonne de constantes à la série de rendements du marché (**market_returns**) en utilisant la fonction **add_constant** de la bibliothèque **statsmodels** (importée sous le nom **sm**). Cela est nécessaire car la régression linéaire avec **OLS**³⁹ nécessite une constante.
- B. Modélisation de la régression linéaire : Nous utilisons la fonction **OLS** de **statsmodels** pour créer un modèle de régression linéaire ordinaire (Ordinary Least Squares) entre les rendements de l'actif (**asset_returns**) et les rendements du marché augmentés de la constante. Le modèle est spécifié en utilisant la syntaxe **sm.OLS(y, X)**, où **y** est la variable dépendante et **X** est la matrice des variables explicatives.
- C. Ajustement du modèle : Nous ajustons le modèle aux données en utilisant la méthode **fit()** sur l'objet du modèle. Cela effectue l'estimation des coefficients de régression et d'autres statistiques associées.

³⁹ La fonction OLS (Ordinary Least Squares) en Python fait partie du module **statsmodels** et est utilisée pour effectuer une régression linéaire ordinaire

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

- D. Obtention du bêta : Une fois le modèle ajusté, nous récupérons le coefficient de pente (bêta) en accédant aux paramètres estimés (**params**) de l'objet **results**. Le bêta est généralement associé au coefficient de la variable explicative correspondant aux rendements du marché, ainsi on trouve un beta de 0.09716372584569127
- E. Calcul du rendement attendu : Nous utilisons le bêta estimé, le rendement moyen du marché (**market_mean**), le taux sans risque (**risk_free_rate**) et la période de temps (**time_period**) pour calculer le rendement attendu selon le modèle CAPM. Le rendement attendu est calculé en utilisant la formule CAPM : rendement attendu = taux sans risque + bêta * (rendement moyen du marché - taux sans risque) * période de temps. On trouve ainsi un rendement de 0.0009590059740969729 (similaire a celui calculer sur le R)

Conclusion :

Voilà qui conclut notre chapitre , mettant en lumière les résultats obtenus et le travail réalisé. Nous avons constaté que les langages de programmation tels que R et Python offrent une large gamme de possibilités en termes de calculs et d'analyses pour la gestion de portefeuilles.

Bien que nous ayons utilisé un portefeuille témoin pour notre exemple, les algorithmes et les méthodes que nous avons présentés peuvent être appliqués à n'importe quel autre type de portefeuille en modifiant simplement les données d'entrée. Cela signifie que les investisseurs , financiers , peuvent utiliser ces outils pour analyser et prendre des décisions concernant leurs portefeuilles d'investissement.

En particulier, sur le marché financier algérien, ces algorithmes peuvent être appliqués par les teneurs de comptes conservateurs travaillant au sein des Intermédiaires en Opérations de Bourse (IOB). Ces derniers pourraient utiliser ces méthodes pour analyser les rendements, évaluer les risques, calculer les ratios de performance tels que le Sharpe ratio, ainsi que pour estimer la VaR (Value at Risk) des portefeuilles et qui sont plus performant , plus précis et plus facile a utiliser que les méthodes traditionnels utilisées a savoir l'Excel.

En utilisant ces outils d'analyse et de gestion de portefeuilles, on peut prendre des décisions éclairées pour optimiser leurs portefeuilles, gérer les risques et maximiser les rendements. Il est important de noter que ces outils doivent être utilisés avec prudence et en tenant compte des caractéristiques spécifiques du marché algérien, ainsi que des réglementations et des contraintes propres à ce marché.

Chapitre 4 : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger

En conclusion, les algorithmes et les méthodes présentés ici offrent des possibilités intéressantes pour l'analyse et la gestion des portefeuilles sur le marché financier algérien.

Conclusion générale

Au terme du présent travail qui porte sur « **La Finance d`entreprise a l`heure de la *data science* Cas : Gestion de portefeuille avec R et Python** », quelques résultats fondamentaux de notre recherche sont mis en évidence . Notre étude s`est fixée au départ l`objectif de présenter comment La *data science* représente une avancée technologique majeure pour la finance moderne.

Dans un premier temps, il a fallu poser le contexte et définir les termes de cette recherche. Le *Big data* est le résultat d`une course effrénée à l`évolution technologique . Aujourd`hui Des milliards de dollars circulent quotidiennement sur les marchés mondiaux et les analystes sont chargés de surveiller ces données avec précision, sécurité et rapidité afin d`établir des prévisions, de découvrir de nouveaux modèles et de créer des stratégies prédictives. Face a ces montagnes de données , se cachent des informations qui, mieux exploitées, peuvent être génératrices de valeur c`est dans ce contexte qu`est née la *data science* sujet principale de notre étude

Au fil de nos recherches, nous avons d`abord exploré les bases de la science des données. Avant d`aborder la *big data* et la *data science* , il est essentiel de comprendre ces concepts dans leur globalité. C`est pourquoi nous avons consacré un chapitre à définir les concepts fondamentaux et à retracer l`évolution de ces technologies. Nous avons également clarifié la distinction entre la *data science* et l`analyse commerciale, deux domaines souvent confondus.

Au préalable de notre développement informatique, nous avons, ensuite, décidé de nous plonger dans le domaine technique. Nous ne voulions pas utiliser des technologies sans les comprendre. C`est pour cette raison que notre deuxième chapitre est consacré aux outils de la *big data*, parmi lesquels Hadoop occupe une place centrale en tant qu`outil principal de collecte de données. De plus, nous abordons également les logiciels R et Python, qui sont largement utilisés par les grandes entreprises pour l`analyse et la gestion des données. Ces outils offrent des fonctionnalités avancées et une flexibilité accrue, permettant ainsi aux professionnels de la finance d`explorer et d`exploiter efficacement les données massives pour prendre des décisions éclairées.

Sans les talents de la *data science* et ces technologies, la finance comme les autres industries dans le monde ne pourrait imaginer extraire des informations désormais vitales pour leur futur. La suite de notre travail s`est attelée sur comment le *big data* s`intègre au milieu de la finance dans un chapitre intitulé *les uses cases de la data science en finance* qui portent sur

les différents cas où la *data science* peut être utilisée au profit des banques, des entreprises et de toutes les institutions.

Dans ce contexte théorique dense, nous avons développé une solution informatique dédiée à la gestion de portefeuille. Notre objectif était de démontrer l'efficacité des technologies *Big data* à

travers un algorithme simple permettant une analyse approfondie des portefeuilles clients composés d'actions d'entreprises cotées en bourse sur une année donnée. Cette étude nous a permis d'analyser des centaines de données en utilisant un jeu de données simple dans un environnement local, mettant en évidence l'efficacité de l'analyse des *big data* dans le domaine de la finance, en particulier en ce qui concerne la gestion de portefeuille.

Cet algorithme peut être utilisé par les teneurs de compte conservateurs au sein des Intermédiaires en Opérations de Bourse (IOB) cotés en bourse algérienne. Il offre une solution informatique efficace pour analyser et gérer les portefeuilles clients, en prenant en compte les performances des actions cotées, les fluctuations du marché et d'autres facteurs pertinents. Grâce à cette approche basée sur les *Big data*, les teneurs de compte conservateurs peuvent prendre des décisions éclairées et optimiser les investissements de leurs clients.

En conclusion, nous pouvons affirmer que la technologie et la science des données sont déjà pleinement intégrées dans la finance moderne. Grâce à ces avancées, de nombreuses applications et solutions émergent pour optimiser la gestion des portefeuilles, améliorer les prévisions financières et permettre une prise de décision plus éclairée. La capacité à exploiter les vastes quantités de données disponibles ouvre de nouvelles opportunités pour les professionnels de la finance, offrant ainsi un avantage concurrentiel. Il est donc crucial pour les acteurs de l'industrie financière de continuer à investir dans la *data science* et à explorer les opportunités qu'elle offre pour rester compétitifs dans un monde de plus en plus axé sur les données.

Références bibliographiques

Ouvrages

1. Babinet Gilles Big data ,Penser l'homme et le monde autrement ,Le Passeur 2014
2. Baesens Bart -Analytics in a Big Data World The Essential Guide to Data Science and Its Applications ,Wiley 2014
3. C. Stratopoulos Theophanis, Gregory P. Shields, -Basic Audit Data Analytics with R , Stratopoulos & Shields, Waterloo - 2018
4. Carver Robert , Practical Data Analysis with JMP , Sas publishing 2010
5. Cielen Davy, Arno Meysman ,Mohamed Ali Introducing Data science ,Manning Publications ,2016
6. Cointot J-C et Y Eychenne La reolution Big data ,DUNOD
7. Consoli Sergio ,Diego Reforgiato Recupero, Michaela Saisana Data Science for Economics and Finance Methodologies and Applications , Springer edition
8. Delort Pierre « Le Big Data », Edition Que sais-je ? 2018
9. Feinleb David - Big Data Bootcamp What Managers Need to Know to Profit from the Big Data Revolution,Apress 2014
10. Grolemund Garrett, Hadley Wickham Hands on Programming with R : write your own functions and simulations , Oreilly Media 2015
11. Harvard Business Review Press HBR Guide to Data Analytics Basics for Managers
12. Hilpisch Yves, Python for Finance : Analyze big financial data , O'Reilly 2015
13. Horton Nicholas J , Ken Kleinman , Chapman and Hall -Using R and R studio for Data management , statistical analysis and graphis , 2015
14. Hull Isaiah Machine Learning for Economics and Finance in TensorFlow 2 , Deep Learning Models for Research and Industry , Apress 2021
15. Humber Max Personal Finance with Python Using pandas, Requests, and Recurrent , Apress
16. Iftikhar Nadeem ,Data Science Essentials, Canterbury Christ Church University 2021
17. LeCun Yann, Rachel Schutt et Catherine Cramer Data Science pour l'entreprise,
18. Ma Weiming James, Mastering Python for Finance *Second Edition* , Packt 2019
19. Magoulas Roger *Data Science* and Business Intelligence: A Practitioner's Guide to Building Intelligent Data Systems
20. Manyika James, Michael Chui et Brad Brown " The Age of Big Data "
21. Mather Bob , ***Basic Python In Finance*** ,2014
22. O'Reilly Radar Big Data now: current perspectives ,O'Reilly media 2011
23. O'Neil Cathy et Rachel Schutt Doing Data Science, O'Reilly 2013

REFERENCES BIBLIOGRAPHIQUES

24. Ozdemir Sinan Principles of Data science ,packet publishing,2016
25. Pierson Lillian Data science for dummies ,John Wiley and sons edition , 2015
26. S.Fletcher , G.Gardner Modelling in Python , Wiley Finance Series ,2009
27. VanderPlas Jake Python Data Science Handbook, O`Reilly 2016
28. Würtz Diethelm Onghow Lam Andrew Ellis Yohan Chalabi , Basic R for finance , rmetrics association & finance online 2010

Articles

1. Adanma Eberendu Cecilia. "Unstructured Data: an overview of the data of Big Data." International Journal of Computer Trends and Technology 3.1 (2016): 46-50.
2. Al-lozia Enas, Amjed Alfityanib, Ayman Abdalmajeed Alsmadic*, Amer Moh'd Al_Hazimehd and Jassim Ahmad Al-Gasawnehe "The role of big data in financial sector: A review paper ", International Journal of Data and Network Science 2022
3. Anastasios Arvanitis, Shivnath Babu, Eric Chu, Adrian Popescu, Alkis Simitsis, Kevin Wilkinson - Automated Performance Management for the Big Data Stack .2022
4. Barbaglia Luca, Sergio Consoli, Sebastiano Manzan, Diego Reforgiato Recupero, Michaela Saisana, and Luca Tiozzo Pezzol , Data Science Technologies in Economics and Finance: A Gentle Walk-In , Research Gate
5. Business Intelligence : place de la BI et des projets décisionnelle _CIGREF_2009
6. Cobbaut Robert Roland Gillet ,Georges Hübner - La gestion de portefeuille Instruments, stratégie et performance ,Debeok 2019
7. Cohen Gil - Algorithmic Trading and Financial Forecasting Using Advanced Artificial Intelligence Methodologies , Maria C. Mariani 2022
8. Eric Stanton - Trading with machine learning and big data -handbook of artificial intelligence and big data applications in investments ,2021
9. Firican George - The history of Big Data
10. Haleem Abid , Mohd Javaid , Mohd Asim Qadri , Ravi Pratap Singh , Rajiv Suman - Artificial intelligence (AI) applications for marketing: A literature-based study - International Journal of Intelligent Networks 2022
11. Kinsey Mc Global Institute- *Big Data* : The next frontier for innovation, competition and productivity –
12. Li1 Yuxin, Jizheng Yi1,2,* , Huanyu Chen1 and Duanxiang Peng , Theory and application of artificial intelligence in financial industry , AIMS Press 2021

REFERENCES BIBLIOGRAPHIQUES

13. Liu Haihua, Shan Huang , Peng Wang and Zejun Li , A review of data mining methods in financial markets , AIMS press , 2021
14. Narvekar Aditya et Debashis Guha Bankruptcy prediction using machine learning and an application to the case of the COVID-19 recession ,AIMS Press 2021
15. Patil Sumer - Artificial Intelligence in Stock Market - International Journal of Advanced Research in Science, Communication and Technology (IJARSCT) , Volume 10, Issue 1, October 2021
16. Raj Rohit , Choudhry Shikha Priya - Analysis of Artificial Intelligence Techniques for Prevention of Financial Fraud, International Journal of Engineering Research & Technology (IJERT)ISSN: 2278-0181Vol. 11 Issue 02, February-2022
17. Regards croisés sur l'économie -Les 5V du *Big Data* Thomas Bourany 2018/2 (n° 23), pages 27 à 31
18. Riecken Hanna - AI in performance management: a game-changing development? , niversity of Twente P.O. Box 217, 7500AE Enschede The Netherlands
19. Ronen Feldman, , and James Sanger. The text mining handbook: advanced approaches in analyzing unstructured data. Cambridge university press, 2007.
20. Souifi Amel Cherfi Boulanger Zohra Zolghadri Marc Barkallah Maher Haddar Mohamed - From Big Data to Smart Data: Application to performance management , IFAC PapersOnLine 54-1 (2021) 857–862
21. Tu Jun , Feng Shi, Li Guo , Textual Analysis and Machine Learning: Crack Unstructured Data in Finance and Accounting
22. Wickham Hadley & Garrett Grolemund R for Data Science import, tidy, transform, visualize, and model data , O`Reilly 2016
23. Yan Yuxing, Python for finance Build real-life Python applications for quantitative finance and financial engineering , Packt 2014

Sites internet

1. <https://www.sgbv.dz/#>
2. <https://everythingcomputerscience.com/CSVideoLessons.html#>
3. <https://randerson112358.medium.com/build-a-killer-tech-stock-portfolio-e57b4ef5e3eb>
4. <https://finance-innovation.org/les-data-scientists-sont-ils-lavenir-de-la-finance/>
5. <https://www.upgrad.com/blog/data-science-use-cases-finance-industrys>

REFERENCES BIBLIOGRAPHIQUES

6. <https://www.projectpro.io/article/hadoop-in-financial-sector/80>
7. <https://www.information-age.com/hadoop-finance-big-data-pursuit-big-bucks-4948/>
8. <https://www.cfainstitute.org/en/programs/cfa/charterholder-careers/roles/data-scientist>
9. <https://www.talend.com/fr/resources/big-data-finance/>
10. <https://www.talend.com/fr/resources/big-data-marketing/>
11. <https://www.projectpro.io/article/hadoop-in-financial-sector/80>
12. <https://www.softkraft.co/how-is-python-used-in-finance/>
13. <https://www.oreilly.com/library/view/python-for-finance/>
14. <https://www.abcbourse.com/>

Annexes

Annexe 1: Cours des action fevrier 2022 -fevrier 2023

Date	ALL	BIO	AUR	SAI
2022-02-08	320	1303	560	543
2022-02-10	320	1303	560	543
2022-02-13	320	1303	560	543
2022-02-15	320	1303	560	543
2022-02-17	320	1303	560	543
2022-02-20	320	1303	560	543
2022-02-22	320	1303	560	543
2022-02-24	320	1303	560	543
2022-02-27	320	1303	560	543
2022-03-01	320	1305	560	543
2022-03-03	320	1305	560	543
2022-03-06	320	1305	560	543
2022-03-08	320	1305	550	543
2022-03-10	320	1305	550	543
2022-03-13	320	1305	550	543
2022-03-15	320	1305	550	543
2022-03-17	320	1305	550	543
2022-03-20	320	1305	550	543
2022-03-22	320	1306	550	543
2022-03-24	327	1306	550	543
2022-03-27	327	1306	550	543
2022-03-29	327	1306	550	543
2022-03-31	327	1306	550	543
2022-04-03	327	1306	550	543
2022-04-05	327	1307	550	543
2022-04-07	327	1307	550	543
2022-04-10	327	1307	550	543
2022-04-12	328	1307	550	543
2022-04-14	328	1307	550	543
2022-04-17	328	1307	550	543
2022-04-19	328	1307	550	543
2022-04-21	328	1307	550	543
2022-04-24	328	1307	550	543
2022-04-26	360	1307	550	543
2022-04-28	360	1307	550	543
2022-05-04	360	1307	550	543
2022-05-05	360	1307	550	543
2022-05-08	360	1307	550	543
2022-05-10	360	1307	550	543
2022-05-12	360	1307	550	543
2022-05-15	360	1307	550	543
2022-05-17	361	1307	550	543
2022-05-19	361	1307	550	544
2022-05-22	361	1307	550	544

2022-05-24	379	1437	550	544
2022-05-26	379	1437	550	544
2022-05-29	379	1437	550	544
2022-05-31	379	1437	550	544
2022-06-02	379	1508	550	544
2022-06-05	380	1508	550	544
2022-06-07	380	1508	550	544
2022-06-09	380	1508	550	544
2022-06-12	380	1582	550	544
2022-06-14	380	1582	550	544
2022-06-16	382	1582	550	544
2022-06-19	382	1582	550	544
2022-06-21	382	1661	550	544
2022-06-23	382	1744	550	544
2022-06-26	382	1744	550	544
2022-06-28	382	1744	550	544
2022-06-30	382	1751	495	544
2022-07-03	382	1751	495	544
2022-07-06	382	1751	495	545
2022-07-07	382	1832	495	545
2022-07-11	382	1832	495	545
2022-07-12	382	1832	495	544
2022-07-14	382	1832	495	544
2022-07-17	382	1832	495	544
2022-07-19	382	1900	495	544
2022-07-21	382	1900	495	544
2022-07-24	382	1900	495	544
2022-07-26	382	1900	495	544
2022-07-28	347	1838	495	544
2022-07-31	347	1838	495	544
2022-08-02	347	1900	495	544
2022-08-04	347	1900	495	534
2022-08-07	347	1900	495	534
2022-08-09	347	1900	495	534
2022-08-11	347	1900	495	534
2022-08-14	347	1900	495	534
2022-08-16	347	1900	495	534
2022-08-18	347	1900	495	534
2022-08-21	380	1900	495	534
2022-08-23	380	1900	495	534
2022-08-25	380	1900	495	534
2022-08-28	380	1900	495	534
2022-08-30	380	1900	495	534
2022-09-01	380	1725	495	534
2022-09-04	380	1725	495	534
2022-09-06	380	1725	495	534
2022-09-08	380	1725	495	534

2022-09-11	380	1725	495	534
2022-09-13	380	1725	495	534
2022-09-15	380	1725	495	534
2022-09-18	380	1725	495	534
2022-09-20	380	1725	495	534
2022-09-22	380	1725	495	534
2022-09-25	380	1725	495	534
2022-09-27	380	1725	495	534
2022-09-29	380	1725	495	534
2022-10-02	380	1728	495	534
2022-10-04	380	1728	495	534
2022-10-06	380	1728	495	534
2022-10-09	380	1728	495	534
2022-10-11	380	1728	495	534
2022-10-13	380	1728	495	534
2022-10-16	380	1728	495	534
2022-10-18	380	1728	495	534
2022-10-20	381	1728	495	534
2022-10-23	399	1728	495	534
2022-10-25	399	1728	495	534
2022-10-27	399	1751	495	534
2022-10-30	399	1751	495	534
2022-11-02	381	1751	495	534
2022-11-03	381	1751	495	534
2022-11-06	381	1751	495	534
2022-11-08	381	1839	495	534
2022-11-10	381	1839	495	530
2022-11-13	383	1839	495	530
2022-11-15	383	1839	495	530
2022-11-17	383	1839	495	530
2022-11-20	383	1839	495	530
2022-11-22	383	1839	495	530
2022-11-24	383	1839	495	530
2022-11-27	383	1839	495	530
2022-11-29	402	1839	495	530
2022-12-01	402	1839	495	530
2022-12-04	402	1839	495	530
2022-12-06	402	1839	495	530
2022-12-08	403	1839	495	530
2022-12-11	403	1839	495	530
2022-12-13	403	1839	495	530
2022-12-15	403	1839	495	530
2022-12-18	403	1839	495	530
2022-12-20	403	2022	495	530
2022-12-22	403	2022	495	530
2022-12-25	403	2022	495	530
2022-12-27	403	2123	495	530

2022-12-29	403	2123	495	530
2023-01-02	403	2123	495	530
2023-01-03	403	2123	495	530
2023-01-05	403	2123	495	530
2023-01-08	403	2229	495	530
2023-01-10	403	2229	495	530
2023-01-15	403	2229	495	530
2023-01-17	403	2340	495	530
2023-01-19	402	2340	495	530
2023-01-22	403	2340	495	530
2023-01-24	403	2457	495	530
2023-01-26	404	2342	495	530
2023-01-29	404	2342	495	530
2023-01-31	404	2401	495	530
2023-02-02	404	2401	495	530
2023-02-05	404	2401	495	530
2023-02-07	420	2340	495	530

Annexe 2 : Rendements journaliers des actifs sur le R

RStudio Source Editor

portfolioReturns

Filter

	ALL	BIO	AUR	SAI
2022-02-10	0.000000000	0.000000000	0.000000000	0.000000000
2022-02-13	0.000000000	0.000000000	0.000000000	0.000000000
2022-02-15	0.000000000	0.000000000	0.000000000	0.000000000
2022-02-17	0.000000000	0.000000000	0.000000000	0.000000000
2022-02-20	0.000000000	0.000000000	0.000000000	0.000000000
2022-02-22	0.000000000	0.000000000	0.000000000	0.000000000
2022-02-24	0.000000000	0.000000000	0.000000000	0.000000000
2022-02-27	0.000000000	0.000000000	0.000000000	0.000000000
2022-03-01	0.000000000	0.0015337426	0.000000000	0.000000000
2022-03-03	0.000000000	0.000000000	0.000000000	0.000000000
2022-03-06	0.000000000	0.000000000	0.000000000	0.000000000
2022-03-08	0.000000000	0.000000000	-0.01801851	0.000000000
2022-03-10	0.000000000	0.000000000	0.000000000	0.000000000
2022-03-13	0.000000000	0.000000000	0.000000000	0.000000000
2022-03-15	0.000000000	0.000000000	0.000000000	0.000000000
2022-03-17	0.000000000	0.000000000	0.000000000	0.000000000
2022-03-20	0.000000000	0.000000000	0.000000000	0.000000000
2022-03-22	0.000000000	0.0007659901	0.000000000	0.000000000
2022-03-24	0.021639175	0.000000000	0.000000000	0.000000000
2022-03-27	0.000000000	0.000000000	0.000000000	0.000000000
2022-03-29	0.000000000	0.000000000	0.000000000	0.000000000
2022-03-31	0.000000000	0.000000000	0.000000000	0.000000000
2022-04-03	0.000000000	0.000000000	0.000000000	0.000000000
2022-04-05	0.000000000	0.0007654038	0.000000000	0.000000000
2022-04-07	0.000000000	0.000000000	0.000000000	0.000000000
2022-04-10	0.000000000	0.000000000	0.000000000	0.000000000
2022-04-12	0.003053437	0.000000000	0.000000000	0.000000000
2022-04-14	0.000000000	0.000000000	0.000000000	0.000000000
2022-04-17	0.000000000	0.000000000	0.000000000	0.000000000
2022-04-19	0.000000000	0.000000000	0.000000000	0.000000000
2022-04-21	0.000000000	0.000000000	0.000000000	0.000000000
2022-04-24	0.000000000	0.000000000	0.000000000	0.000000000
2022-04-26	0.093090423	0.000000000	0.000000000	0.000000000
2022-04-28	0.000000000	0.000000000	0.000000000	0.000000000
2022-05-04	0.000000000	0.000000000	0.000000000	0.000000000
2022-05-05	0.000000000	0.000000000	0.000000000	0.000000000
2022-05-08	0.000000000	0.000000000	0.000000000	0.000000000
2022-05-10	0.000000000	0.000000000	0.000000000	0.000000000
2022-05-12	0.000000000	0.000000000	0.000000000	0.000000000
2022-05-15	0.000000000	0.000000000	0.000000000	0.000000000
2022-05-17	0.002773927	0.000000000	0.000000000	0.000000000
2022-05-19	0.000000000	0.000000000	0.000000000	0.001839927
2022-05-22	0.000000000	0.000000000	0.000000000	0.000000000
2022-05-24	0.048658247	0.0948231725	0.000000000	0.000000000
2022-05-26	0.000000000	0.000000000	0.000000000	0.000000000
2022-05-29	0.000000000	0.000000000	0.000000000	0.000000000
2022-05-31	0.000000000	0.000000000	0.000000000	0.000000000
2022-06-02	0.000000000	0.0482266625	0.000000000	0.000000000
2022-06-05	0.002635048	0.000000000	0.000000000	0.000000000
2022-06-07	0.000000000	0.000000000	0.000000000	0.000000000
2022-06-09	0.000000000	0.000000000	0.000000000	0.000000000

2022-06-09	0.00000000	0.00000000	0.00000000	0.00000000
2022-06-12	0.00000000	0.0479055998	0.00000000	0.00000000
2022-06-14	0.00000000	0.00000000	0.00000000	0.00000000
2022-06-16	0.005249356	0.00000000	0.00000000	0.00000000
2022-06-19	0.00000000	0.00000000	0.00000000	0.00000000
2022-06-21	0.00000000	0.0487299613	0.00000000	0.00000000
2022-06-23	0.00000000	0.0487614949	0.00000000	0.00000000
2022-06-26	0.00000000	0.00000000	0.00000000	0.00000000
2022-06-28	0.00000000	0.00000000	0.00000000	0.00000000
2022-06-30	0.00000000	0.0040057278	-0.10536052	0.00000000
2022-07-03	0.00000000	0.00000000	0.00000000	0.00000000
2022-07-06	0.00000000	0.00000000	0.00000000	0.001836548
2022-07-07	0.00000000	0.0452212129	0.00000000	0.00000000
2022-07-11	0.00000000	0.00000000	0.00000000	0.00000000
2022-07-12	0.00000000	0.00000000	0.00000000	-0.001836548
2022-07-14	0.00000000	0.00000000	0.00000000	0.00000000
2022-07-17	0.00000000	0.00000000	0.00000000	0.00000000
2022-07-19	0.00000000	0.0364456199	0.00000000	0.00000000
2022-07-21	0.00000000	0.00000000	0.00000000	0.00000000
2022-07-24	0.00000000	0.00000000	0.00000000	0.00000000
2022-07-26	0.00000000	0.00000000	0.00000000	0.00000000
2022-07-28	-0.096095829	-0.0331758622	0.00000000	0.00000000
2022-07-31	0.00000000	0.00000000	0.00000000	0.00000000
2022-08-02	0.00000000	0.0331758622	0.00000000	0.00000000
2022-08-04	0.00000000	0.00000000	0.00000000	-0.018553408
2022-08-07	0.00000000	0.00000000	0.00000000	0.00000000
2022-08-09	0.00000000	0.00000000	0.00000000	0.00000000
2022-08-11	0.00000000	0.00000000	0.00000000	0.00000000
2022-08-14	0.00000000	0.00000000	0.00000000	0.00000000
2022-08-16	0.00000000	0.00000000	0.00000000	0.00000000
2022-08-18	0.00000000	0.00000000	0.00000000	0.00000000
2022-08-21	0.090846473	0.00000000	0.00000000	0.00000000
2022-08-23	0.00000000	0.00000000	0.00000000	0.00000000
2022-08-25	0.00000000	0.00000000	0.00000000	0.00000000
2022-08-28	0.00000000	0.00000000	0.00000000	0.00000000
2022-08-30	0.00000000	0.00000000	0.00000000	0.00000000
2022-09-01	0.00000000	-0.0966268357	0.00000000	0.00000000
2022-09-04	0.00000000	0.00000000	0.00000000	0.00000000
2022-09-06	0.00000000	0.00000000	0.00000000	0.00000000
2022-09-08	0.00000000	0.00000000	0.00000000	0.00000000
2022-09-11	0.00000000	0.00000000	0.00000000	0.00000000
2022-09-13	0.00000000	0.00000000	0.00000000	0.00000000
2022-09-15	0.00000000	0.00000000	0.00000000	0.00000000
2022-09-18	0.00000000	0.00000000	0.00000000	0.00000000
2022-09-20	0.00000000	0.00000000	0.00000000	0.00000000
2022-09-22	0.00000000	0.00000000	0.00000000	0.00000000
2022-09-25	0.00000000	0.00000000	0.00000000	0.00000000
2022-09-27	0.00000000	0.00000000	0.00000000	0.00000000
2022-09-29	0.00000000	0.00000000	0.00000000	0.00000000
2022-10-02	0.00000000	0.0017376199	0.00000000	0.00000000
2022-10-04	0.00000000	0.00000000	0.00000000	0.00000000
2022-10-06	0.00000000	0.00000000	0.00000000	0.00000000
2022-10-09	0.00000000	0.00000000	0.00000000	0.00000000
2022-10-11	0.00000000	0.00000000	0.00000000	0.00000000
2022-10-13	0.00000000	0.00000000	0.00000000	0.00000000
2022-10-16	0.00000000	0.00000000	0.00000000	0.00000000

2022-10-16	0.000000000	0.000000000	0.00000000	0.00000000
2022-10-18	0.000000000	0.000000000	0.00000000	0.00000000
2022-10-20	0.002628122	0.000000000	0.00000000	0.00000000
2022-10-23	0.046162042	0.000000000	0.00000000	0.00000000
2022-10-25	0.000000000	0.000000000	0.00000000	0.00000000
2022-10-27	0.000000000	0.0132223829	0.00000000	0.00000000
2022-10-30	0.000000000	0.000000000	0.00000000	0.00000000
2022-11-02	-0.046162042	0.000000000	0.00000000	0.00000000
2022-11-03	0.000000000	0.000000000	0.00000000	0.00000000
2022-11-06	0.000000000	0.000000000	0.00000000	0.00000000
2022-11-08	0.000000000	0.0490348923	0.00000000	0.00000000
2022-11-10	0.000000000	0.000000000	0.00000000	-0.007518832
2022-11-13	0.005235614	0.000000000	0.00000000	0.00000000
2022-11-15	0.000000000	0.000000000	0.00000000	0.00000000
2022-11-17	0.000000000	0.000000000	0.00000000	0.00000000
2022-11-20	0.000000000	0.000000000	0.00000000	0.00000000
2022-11-22	0.000000000	0.000000000	0.00000000	0.00000000
2022-11-24	0.000000000	0.000000000	0.00000000	0.00000000
2022-11-27	0.000000000	0.000000000	0.00000000	0.00000000
2022-11-29	0.048417099	0.000000000	0.00000000	0.00000000
2022-12-01	0.000000000	0.000000000	0.00000000	0.00000000
2022-12-04	0.000000000	0.000000000	0.00000000	0.00000000
2022-12-06	0.000000000	0.000000000	0.00000000	0.00000000
2022-12-08	0.002484473	0.000000000	0.00000000	0.00000000
2022-12-11	0.000000000	0.000000000	0.00000000	0.00000000
2022-12-13	0.000000000	0.000000000	0.00000000	0.00000000
2022-12-15	0.000000000	0.000000000	0.00000000	0.00000000
2022-12-18	0.000000000	0.000000000	0.00000000	0.00000000
2022-12-20	0.000000000	0.0948651750	0.00000000	0.00000000
2022-12-22	0.000000000	0.000000000	0.00000000	0.00000000
2022-12-25	0.000000000	0.000000000	0.00000000	0.00000000
2022-12-27	0.000000000	0.0487430621	0.00000000	0.00000000
2022-12-29	0.000000000	0.000000000	0.00000000	0.00000000
2023-01-02	0.000000000	0.000000000	0.00000000	0.00000000
2023-01-03	0.000000000	0.000000000	0.00000000	0.00000000
2023-01-05	0.000000000	0.000000000	0.00000000	0.00000000
2023-01-08	0.000000000	0.0487228717	0.00000000	0.00000000
2023-01-10	0.000000000	0.000000000	0.00000000	0.00000000
2023-01-15	0.000000000	0.000000000	0.00000000	0.00000000
2023-01-17	0.000000000	0.0485978750	0.00000000	0.00000000
2023-01-19	-0.002484473	0.000000000	0.00000000	0.00000000
2023-01-22	0.002484473	0.000000000	0.00000000	0.00000000
2023-01-24	0.000000000	0.0487901642	0.00000000	0.00000000
2023-01-29	0.000000000	0.000000000	0.00000000	0.00000000
2023-01-31	0.000000000	0.0248800521	0.00000000	0.00000000
2023-02-02	0.000000000	0.000000000	0.00000000	0.00000000
2023-02-05	0.000000000	0.000000000	0.00000000	0.00000000
2023-02-07	0.038839833	-0.0257343879	0.00000000	0.00000000

Annexe 2 : Rendements journaliers des actifs sur le R

Annexe 3 les variations quotidiennes de rendements des actifs sur R

	portfolio.returns
2022-02-10	0.0000000000
2022-02-13	0.0000000000
2022-02-15	0.0000000000
2022-02-17	0.0000000000
2022-02-20	0.0000000000
2022-02-22	0.0000000000
2022-02-24	0.0000000000
2022-02-27	0.0000000000
2022-03-01	0.0006134971
2022-03-03	0.0000000000
2022-03-06	0.0000000000
2022-03-08	-0.0018007458
2022-03-10	0.0000000000
2022-03-13	0.0000000000
2022-03-15	0.0000000000
2022-03-17	0.0000000000
2022-03-20	0.0000000000
2022-03-22	0.0003072311
2022-03-24	0.0043316533
2022-03-27	0.0000000000
2022-03-29	0.0000000000
2022-03-31	0.0000000000
2022-04-03	0.0000000000
2022-04-05	0.0003058120
2022-04-07	0.0000000000
2022-04-10	0.0000000000
2022-04-12	0.0006215694
2022-04-14	0.0000000000
2022-04-17	0.0000000000
2022-04-19	0.0000000000
2022-04-21	0.0000000000
2022-04-24	0.0000000000
2022-04-26	0.0189958967
2022-04-28	0.0000000000
2022-05-04	0.0000000000
2022-05-05	0.0000000000
2022-05-08	0.0000000000
2022-05-10	0.0000000000
2022-05-12	0.0000000000
2022-05-15	0.0000000000
2022-05-17	0.0006072024
2022-05-19	0.0005390002

2022-05-22	0.0000000000
2022-05-24	0.0477995395
2022-05-26	0.0000000000
2022-05-29	0.0000000000
2022-05-31	0.0000000000
2022-06-02	0.0197322447
2022-06-05	0.0005670246
2022-06-07	0.0000000000
2022-06-09	0.0000000000
2022-06-12	0.0201371698
2022-06-14	0.0000000000
2022-06-16	0.0011095774
2022-06-19	0.0000000000
2022-06-21	0.0210179420
2022-06-23	0.0216023718
2022-06-26	0.0000000000
2022-06-28	0.0000000000
2022-06-30	-0.0070446201
2022-07-03	0.0000000000
2022-07-06	0.0004763862
2022-07-07	0.0207855293
2022-07-11	0.0000000000
2022-07-12	-0.0004673204
2022-07-14	0.0000000000
2022-07-17	0.0000000000
2022-07-19	0.0171609345
2022-07-21	0.0000000000
2022-07-24	0.0000000000
2022-07-26	0.0000000000
2022-07-28	-0.0348834727
2022-07-31	0.0000000000
2022-08-02	0.0159456560
2022-08-04	-0.0047271671
2022-08-07	0.0000000000
2022-08-09	0.0000000000
2022-08-11	0.0000000000
2022-08-14	0.0000000000
2022-08-16	0.0000000000
2022-08-18	0.0000000000
2022-08-21	0.0166076748
2022-08-23	0.0000000000

Annexe 3 les variations quotidiennes de rendements des actifs sur R

2022-08-25	0.0000000000
2022-08-28	0.0000000000
2022-08-30	0.0000000000
2022-09-01	-0.0466794973
2022-09-04	0.0000000000
2022-09-06	0.0000000000
2022-09-08	0.0000000000
2022-09-11	0.0000000000
2022-09-13	0.0000000000
2022-09-15	0.0000000000
2022-09-18	0.0000000000
2022-09-20	0.0000000000
2022-09-22	0.0000000000
2022-09-25	0.0000000000
2022-09-27	0.0000000000
2022-09-29	0.0000000000
2022-10-02	0.0007954474
2022-10-04	0.0000000000
2022-10-06	0.0000000000
2022-10-09	0.0000000000
2022-10-11	0.0000000000
2022-10-13	0.0000000000
2022-10-16	0.0000000000
2022-10-18	0.0000000000
2022-10-20	0.0005403465
2022-10-23	0.0095107998
2022-10-25	0.0000000000
2022-10-27	0.0059983179
2022-10-30	0.0000000000
2022-11-02	-0.0097973309
2022-11-03	0.0000000000
2022-11-06	0.0000000000
2022-11-08	0.0226260325
2022-11-10	-0.0018929135
2022-11-13	0.0010486895
2022-11-15	0.0000000000

2022-11-17	0.0000000000
2022-11-20	0.0000000000
2022-11-22	0.0000000000
2022-11-24	0.0000000000
2022-11-27	0.0000000000
2022-11-29	0.0097384698
2022-12-01	0.0000000000
2022-12-04	0.0000000000
2022-12-06	0.0000000000
2022-12-08	0.0005188615
2022-12-11	0.0000000000
2022-12-13	0.0000000000
2022-12-15	0.0000000000
2022-12-18	0.0000000000
2022-12-20	0.0444853035
2022-12-22	0.0000000000
2022-12-25	0.0000000000
2022-12-27	0.0239596717
2022-12-29	0.0000000000
2023-01-02	0.0000000000
2023-01-03	0.0000000000
2023-01-05	0.0000000000
2023-01-08	0.0245294144
2023-01-10	0.0000000000
2023-01-15	0.0000000000
2023-01-17	0.0250442419
2023-01-19	-0.0004628621
2023-01-22	0.0004619260
2023-01-24	0.0257211127
2023-01-26	-0.0253889438
2023-01-29	0.0000000000
2023-01-31	0.0131009551
2023-02-02	0.0000000000
2023-02-05	0.0000000000
2023-02-07	-0.0065460440

Annexe 4 : les variations quotidiennes de rendements des actifs sur Python

Date	ALL	BIO	AUR	SAI	Date	ALL	BIO	AUR	SAI
2022-02-08	320	1303	560	543	2022-04-14	328	1307	550	543
2022-02-10	320	1303	560	543	2022-04-17	328	1307	550	543
2022-02-13	320	1303	560	543	2022-04-19	328	1307	550	543
2022-02-15	320	1303	560	543	2022-04-21	328	1307	550	543
2022-02-17	320	1303	560	543	2022-04-24	328	1307	550	543
2022-02-20	320	1303	560	543	2022-04-26	360	1307	550	543
2022-02-22	320	1303	560	543	2022-04-28	360	1307	550	543
2022-02-24	320	1303	560	543	2022-05-04	360	1307	550	543
2022-02-27	320	1303	560	543	2022-05-05	360	1307	550	543
2022-03-01	320	1305	560	543	2022-05-08	360	1307	550	543
2022-03-03	320	1305	560	543	2022-05-10	360	1307	550	543
2022-03-06	320	1305	560	543	2022-05-12	360	1307	550	543
2022-03-08	320	1305	550	543	2022-05-15	360	1307	550	543
2022-03-10	320	1305	550	543	2022-05-17	361	1307	550	543
2022-03-13	320	1305	550	543	2022-05-19	361	1307	550	544
2022-03-15	320	1305	550	543	2022-05-22	361	1307	550	544
2022-03-17	320	1305	550	543	2022-05-24	379	1437	550	544
2022-03-20	320	1305	550	543	2022-05-26	379	1437	550	544
2022-03-22	320	1306	550	543	2022-05-29	379	1437	550	544
2022-03-24	327	1306	550	543	2022-05-31	379	1437	550	544
2022-03-27	327	1306	550	543	2022-06-02	379	1508	550	544
2022-03-29	327	1306	550	543	2022-06-05	380	1508	550	544
2022-03-31	327	1306	550	543	2022-06-07	380	1508	550	544
2022-04-03	327	1306	550	543	2022-06-09	380	1508	550	544
2022-04-05	327	1307	550	543	2022-06-12	380	1582	550	544
2022-04-07	327	1307	550	543	2022-06-14	380	1582	550	544
2022-04-10	327	1307	550	543	2022-06-16	382	1582	550	544
2022-04-12	328	1307	550	543	2022-06-19	382	1582	550	544
					2022-06-21	382	1661	550	544
					2022-06-23	382	1744	550	544

Annexe 4 : les variations quotidiennes de rendements des actifs sur Python

2022-06-28	0	0	0	0	2022-09-11	0	0	0	0
2022-06-30	0	0.00401376	-0.1	0	2022-09-13	0	0	0	0
2022-07-03	0	0	0	0	2022-09-15	0	0	0	0
2022-07-06	0	0	0	0.00183824	2022-09-18	0	0	0	0
2022-07-07	0	0.0462593	0	0	2022-09-20	0	0	0	0
2022-07-11	0	0	0	0	2022-09-22	0	0	0	0
2022-07-12	0	0	0	-0.00183486	2022-09-25	0	0	0	0
2022-07-14	0	0	0	0	2022-09-27	0	0	0	0
2022-07-17	0	0	0	0	2022-09-29	0	0	0	0
2022-07-19	0	0.0371179	0	0	2022-10-02	0	0.00173913	0	0
2022-07-21	0	0	0	0	2022-10-04	0	0	0	0
2022-07-24	0	0	0	0	2022-10-06	0	0	0	0
2022-07-26	0	0	0	0	2022-10-09	0	0	0	0
2022-07-28	-0.091623	-0.0326316	0	0	2022-10-11	0	0	0	0
2022-07-31	0	0	0	0	2022-10-13	0	0	0	0
2022-08-02	0	0.0337323	0	0	2022-10-16	0	0	0	0
2022-08-04	0	0	0	-0.0183824	2022-10-18	0	0	0	0
2022-08-07	0	0	0	0	2022-10-20	0.00263158	0	0	0
2022-08-09	0	0	0	0	2022-10-23	0.0472441	0	0	0
2022-08-11	0	0	0	0	2022-10-25	0	0	0	0
2022-08-14	0	0	0	0	2022-10-27	0	0.0133102	0	0
2022-08-16	0	0	0	0	2022-10-30	0	0	0	0
2022-08-18	0	0	0	0	2022-11-02	-0.0451128	0	0	0
2022-08-21	0.0951009	0	0	0	2022-11-03	0	0	0	0
2022-08-23	0	0	0	0	2022-11-06	0	0	0	0
2022-08-25	0	0	0	0	2022-11-08	0	0.050257	0	0
2022-08-28	0	0	0	0	2022-11-10	0	0	0	-0.00749064
2022-08-30	0	0	0	0	2022-11-13	0.00524934	0	0	0
2022-09-01	0	-0.0921053	0	0	2022-11-15	0	0	0	0
2022-09-04	0	0	0	0	2022-11-17	0	0	0	0
2022-09-06	0	0	0	0					
2022-09-08	0	0	0	0					

Annexe 4 : les variations quotidiennes de rendements des actifs sur Python

2022-11-20	0	0	0	0
2022-11-22	0	0	0	0
2022-11-24	0	0	0	0
2022-11-27	0	0	0	0
2022-11-29	0.0496084	0	0	0
2022-12-01	0	0	0	0
2022-12-04	0	0	0	0
2022-12-06	0	0	0	0
2022-12-08	0.00248756	0	0	0
2022-12-11	0	0	0	0
2022-12-13	0	0	0	0
2022-12-15	0	0	0	0
2022-12-18	0	0	0	0
2022-12-20	0	0.0995106	0	0
2022-12-22	0	0	0	0
2022-12-25	0	0	0	0
2022-12-27	0	0.0499505	0	0
2022-12-29	0	0	0	0
2023-01-02	0	0	0	0
2023-01-03	0	0	0	0
2023-01-05	0	0	0	0
2023-01-08	0	0.0499293	0	0
2023-01-10	0	0	0	0
2023-01-15	0	0	0	0
2023-01-17	0	0.0497981	0	0
2023-01-19	-0.00248139	0	0	0
2023-01-22	0.00248756	0	0	0
2023-01-24	0	0.05	0	0
2023-01-26	0.00248139	-0.046805	0	0
2023-01-29	0	0	0	0
2023-01-31	0	0.0251921	0	0
2023-02-02	0	0	0	0
2023-02-05	0	0	0	0
2023-02-07	0.039604	-0.0254061	0	0

Annexe 5 : les variations quotidiennes de rendements des actifs sur Python

Date	0	Date	0	Date	0
2022-02-10	0	2022-04-07	0	2022-06-05	0.000527704
2022-02-13	0	2022-04-10	0	2022-06-07	0
2022-02-15	0	2022-04-12	0.000611621	2022-06-09	0
2022-02-17	0	2022-04-14	0	2022-06-12	0.0196286
2022-02-20	0	2022-04-17	0	2022-06-14	0
2022-02-22	0	2022-04-19	0	2022-06-16	0.00105263
2022-02-24	0	2022-04-21	0	2022-06-19	0
2022-02-27	0	2022-04-24	0	2022-06-21	0.0199747
2022-03-01	0.000613968	2022-04-26	0.0195122	2022-06-23	0.019988
2022-03-03	0	2022-04-28	0	2022-06-26	0
2022-03-06	0	2022-05-04	0	2022-06-28	0
2022-03-08	-0.00178571	2022-05-05	0	2022-06-30	-0.0083945
2022-03-10	0	2022-05-08	0	2022-07-03	0
2022-03-13	0	2022-05-10	0	2022-07-06	0.000551471
2022-03-15	0	2022-05-12	0	2022-07-07	0.0185037
2022-03-17	0	2022-05-15	0	2022-07-11	0
2022-03-20	0	2022-05-17	0.000555556	2022-07-12	-0.000550459
2022-03-22	0.000306513	2022-05-19	0.000552486	2022-07-14	0
2022-03-24	0.004375	2022-05-22	0	2022-07-17	0
2022-03-27	0	2022-05-24	0.0497581	2022-07-19	0.0148472
2022-03-29	0	2022-05-26	0	2022-07-21	0
2022-03-31	0	2022-05-29	0	2022-07-24	0
2022-04-03	0	2022-05-31	0	2022-07-26	0
2022-04-05	0.000306279	2022-06-02	0.0197634	2022-07-28	-0.0313772

Annexe 5 : les variations quotidiennes de rendements des actifs sur Python

Date	0	Date	0	Date	0
2022-07-31	0	2022-09-25	0	2022-11-20	0
2022-08-02	0.0134929	2022-09-27	0	2022-11-22	0
2022-08-04	-0.00551471	2022-09-29	0	2022-11-24	0
2022-08-07	0	2022-10-02	0.000695652	2022-11-27	0
2022-08-09	0	2022-10-04	0	2022-11-29	0.00992167
2022-08-11	0	2022-10-06	0	2022-12-01	0
2022-08-14	0	2022-10-09	0	2022-12-04	0
2022-08-16	0	2022-10-11	0	2022-12-06	0
2022-08-18	0	2022-10-13	0	2022-12-08	0.000497512
2022-08-21	0.0190202	2022-10-16	0	2022-12-11	0
2022-08-23	0	2022-10-18	0	2022-12-13	0
2022-08-25	0	2022-10-20	0.000526316	2022-12-15	0
2022-08-28	0	2022-10-23	0.00944882	2022-12-18	0
2022-08-30	0	2022-10-25	0	2022-12-20	0.0398042
2022-09-01	-0.0368421	2022-10-27	0.00532407	2022-12-22	0
2022-09-04	0	2022-10-30	0	2022-12-25	0
2022-09-06	0	2022-11-02	-0.00902256	2022-12-27	0.0199802
2022-09-08	0	2022-11-03	0	2022-12-29	0
2022-09-11	0	2022-11-06	0	2023-01-02	0
2022-09-13	0	2022-11-08	0.0201028	2023-01-03	0
2022-09-15	0	2022-11-10	-0.00224719	2023-01-05	0
2022-09-18	0	2022-11-13	0.00104987	2023-01-08	0.0199717
2022-09-20	0	2022-11-15	0	2023-01-10	0
2022-09-22	0	2022-11-17	0	2023-01-15	0

Table des matières

Table des matières

Remerciements	I
Liste des abréviations	IV
Liste des tableaux et figures.....	VI
Sommaire	VII
Introduction générale.....	1
Chapitre I : Le Big Data et la Data Science : Un domaine en plein essor.....	4
Introduction.....	5
Section 1 : Qu'est ce que le <i>Big Data et la Data science</i> ?.....	6
1.1. Historique et origine de l'évolution du Big Data	6
1.2. Définitions des termes et contexte	8
1.3. Les 5 « V » de la Big Data	10
1.3.1. Le volume.....	11
1.3.2. La vitesse.....	11
1.3.3. La variété.....	12
1.3.4. La véracité.....	12
1.3.5. La valeur.....	12
1.4. Les Types de Données.....	12
1.4.1. Les données structurées	13
1.4.2. Les données non structurées	13
1.4.3. Les données semi-structurées :	14
1.5. Du Big Data a la <i>Data science</i>	15
1.6. Les grandes tendances <i>Big Data</i>	16
Section 2 : Business intelligence VS <i>Data science</i>	18
2.1 Définition de la Business Intelligence.....	18
2.2 Différence entre le <i>Business Intelligence</i> et <i>Data science</i>	19
2.2.1. Perspective.....	19
2.2.2. Types de données	20
2.2.3. Livrables	20
2.2.4. Processus.....	20
Section 3 :Les grands défis en matière de <i>Big Data</i> dans la finance.....	21
3.1 Défis en termes de données	21

3.1.1. Qualité des données.....	21
3.1.2. Intégration des données	22
3.1.3. Analyse des données.....	22
3.2 Autres défis.....	22
3.2.1 Réglementation.....	22
3.2.2 Gestion de la capacité.....	23
3.2.3 Etique.....	23
Conclusion.....	24
Chapitre II : Les technologies de la data science applique à la finance.....	25
Introduction.....	26
Section 01 : APACHE HADOOP	26
1.1. Définition de Hadoop.....	26
1.2. Historique de Hadoop.....	27
1.3. Evolution du système HADOOP.....	28
1.3.1 Le système de fichier distribué d'Hadoop HDFS.....	28
1.3.2 MapReduce.....	30
1.4 L'écosystème Hadoop :.....	31
1.5 Hadoop dans la banque et la finance.....	32
1.5.1 Analyse des données	32
1.5.2 La détection de la fraude	33
1.5.3 Gestion des risques.....	33
1.5.4 Stockage et sécurité des données.....	33
1.5.5 Segmentation de la clientèle.....	34
1.6 Hadoop et les grands acteurs financiers	35
Section 02 : Le logiciel Python	37
2.1 Définition de Python	37
2.2 Historique de Python.....	37
2.3 Caractéristique du Python	38
2.3.1 Open source.....	38
2.3.2 Interprété.....	38
2.3.3 Multiparadigme.....	38
2.3.4 Multi-usages.....	38
2.3.5 Multiplateforme.....	38

2.3.6 Typage dynamique.....	39
2.3.7 Sensible à l'indentation.....	39
2.3.8 Collecte des déchets.....	39
2.4 Les concepts de bases de python :.....	40
2.4.1 Variable.....	40
2.4.2 Types de données.....	41
2.4.3 Structures de contrôle de flux.....	42
2.4.4 Les fonctions.....	42
2.4.5 Les modules.....	43
2.5 Fonctionnement et environnement technologiques (Écosystème de Python).....	43
2.6 Les différentes bibliothèques et frameworks disponibles pour la finance	44
2.6.1 Numpy.....	44
2.6.2 Pandas.....	44
2.6.3 Scipy.....	44
2.6.4 Scikit-learn.....	45
2.6.5 Matplotlib.....	45
2.6.6 QuantLib.....	45
2.6.7 Pyfolio.....	46
2.6.8 TensorFlow.....	46
2.7 Les avantages de l'écosystème Python pour la finance	46
2.8 Python pour la finance.....	47
2.8.1 Développement de modèles prédictifs.....	47
2.8.2 Trading algorithmique.....	48
2.8.3 Gestion de portefeuille.....	48
2.8.4 Logiciel bancaires.....	49
2.9 Les obstacles à l'utilisation de Python en finance	49
Section 03 : Le logiciel « R ».....	51
3.1 Initiation au « R » :.....	51
3.1.1 Origines Le logiciel.....	51
3.1.2 Description sommaire de R	52
3.2 Installation de R et RStudio	53
3.2.1 Mécanismes de fonctionnement du "R".....	56

3.3 Utilisation du langage "R" dans la science de gestion.....	63
3.3.1 Utilisation de R dans l'analyse de données de marché	63
3.3.2 La Gestion des risques avec R.....	64
3.3.3 Optimisation de portefeuille.....	65
3.3.4 Analyse de segmentation de la clientèle avec R (clustering)	65
3.3.5 R dans l'analyse des séries temporelles	66
3.4 Python ou R : quel est le meilleur pour la <i>data science</i> en finance.....	67
Conclusion.....	68
Chapitre III : Les <i>uses cases</i> de la <i>data science</i> en finance.....	69
Introduction.....	70
Section 1 : Applications de la <i>data science</i> pour l'optimisation des processus et la croissance des entreprises	70
1.1 Applications de la <i>data science</i> pour le marketing	70
1.2 Applications de la <i>data science</i> pour la gestion de performance	72
1.3 . Applications de la <i>data science</i> dans la détection de fraude	74
1.4 Applications de la <i>data science</i> pour le trading algorithme.....	76
Section 2 : la <i>Data science</i> pour les entreprise moderne et Startups.....	78
2.1 Les « Machines Learning » au profit de l'innovation et des Startup	78
2.2 Les <i>Data scientists</i> des « <i>assets</i> » pour les entreprises moderne	79
2.3 Comment le <i>Big Data</i> aide-t-il les entreprises numériques ?.....	80
Conclusion.....	81
Chapitre IV : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger.....	82
Introduction.....	83
Section 01 : Présentation du marché financier algérien SGBV.....	84
1.1.Présentation de la SGBV :.....	84
1.2.Fonctionnement de la bourse.....	84
1.2.1.Le marché des titres de capital	84
1.2.2. Le marché des titres des créance.....	85
1.3. Structure et organisation de la Bourse d'Alger :	85
1.3.1 La commission d'organisation & de surveillance des opérations de bourse (COSOB) .86	
1.3.2 Les intermédiaires en opération de bourse (IOB).....	87
1.3.3 Les organismes de placement collectifs en valeurs mobilières OPCVM.....	88
1.3.4 Algérie clearing.....	88

1.4 Le processus de cotations et négociation des titres	88
1.4.1 Ouverture de la séance	90
1.4.2 Détermination du cours	90
1.4.3 Allocation des ordres	90
1.4.4 Effectuer les transactions en bloc	90
1.4.5 Clôture de la séance de négociation	90
Section 2 : Gestion de portefeuille avec R studio.....	93
2.1 Préparation des données	93
2.2 Importation des prix du portefeuille.....	94
2.3 Calcul du rendement du portefeuille :	95
2.3.1 Rendements journaliers des actifs	96
2.3.2 Rendements globale journalier du portefeuille.....	96
2.3.3 Esperance de rendement des actifs.....	97
2.3.4 Esperance de rendements globale du portefeuille.....	97
2.3.5 Rendement et risque annuel du portefeuille.....	98
2.3.6 Rendement mensuelle et annuelle du portefeuille.	99
2.3.7 Graphique des rendements du portefeuille.....	99
2.4 Calcul des ratios :	100
2.4.1 Covariance (covMatrix) :	101
2.4.2 Sharpe ratio	102
2.5 Calcul du Beta avec la methode de Capital Asset Pricing Model (CAPM).....	103
2.5.1 La moyenne des rendements du marché.....	105
2.5.2 La variance du marché	105
2.5.3 Le beta.....	105
2.6 Calculer l'esperance de rendement du portefeuille avec la methode CAPM.....	106
Section 3 : Gestion de portefeuille avec Spyder (Python).....	107
3.1 Préparation des données	107
3.2 Importation des prix du portefeuille	108
3.3 Calcul des rendements du portefeuille	108
3.3.1 Rendements journaliers des actifs.....	108
3.3.2 Rendements globale journalier du portefeuille.....	109
3.3.3 Espérance de rendement des actifs	110
3.3.4 Esperance de rendements globale du portefeuille.....	110

3.4 Calcule des ratios :	110
3.4.1 Covariance (covMatrix).....	111
3.4.2 Sharpe ratio.....	112
3.5 Calcule du beta et du rendement espéré avec la méthode CAPM.....	113
Conclusion.....	114
Conclusion générale.....	116

Bibliographie

Annexes

Table des matières...

Annexe 1: Cours des action fevrier 2022 -fevrier 2023

Date	ALL	BIO	AUR	SAI
2022-02-08	320	1303	560	543
2022-02-10	320	1303	560	543
2022-02-13	320	1303	560	543
2022-02-15	320	1303	560	543
2022-02-17	320	1303	560	543
2022-02-20	320	1303	560	543
2022-02-22	320	1303	560	543
2022-02-24	320	1303	560	543
2022-02-27	320	1303	560	543
2022-03-01	320	1305	560	543
2022-03-03	320	1305	560	543
2022-03-06	320	1305	560	543
2022-03-08	320	1305	550	543
2022-03-10	320	1305	550	543
2022-03-13	320	1305	550	543
2022-03-15	320	1305	550	543
2022-03-17	320	1305	550	543
2022-03-20	320	1305	550	543
2022-03-22	320	1306	550	543
2022-03-24	327	1306	550	543
2022-03-27	327	1306	550	543
2022-03-29	327	1306	550	543
2022-03-31	327	1306	550	543
2022-04-03	327	1306	550	543
2022-04-05	327	1307	550	543
2022-04-07	327	1307	550	543
2022-04-10	327	1307	550	543
2022-04-12	328	1307	550	543
2022-04-14	328	1307	550	543
2022-04-17	328	1307	550	543
2022-04-19	328	1307	550	543
2022-04-21	328	1307	550	543
2022-04-24	328	1307	550	543
2022-04-26	360	1307	550	543
2022-04-28	360	1307	550	543
2022-05-04	360	1307	550	543
2022-05-05	360	1307	550	543
2022-05-08	360	1307	550	543
2022-05-10	360	1307	550	543
2022-05-12	360	1307	550	543
2022-05-15	360	1307	550	543
2022-05-17	361	1307	550	543
2022-05-19	361	1307	550	544
2022-05-22	361	1307	550	544

2022-05-24	379	1437	550	544
2022-05-26	379	1437	550	544
2022-05-29	379	1437	550	544
2022-05-31	379	1437	550	544
2022-06-02	379	1508	550	544
2022-06-05	380	1508	550	544
2022-06-07	380	1508	550	544
2022-06-09	380	1508	550	544
2022-06-12	380	1582	550	544
2022-06-14	380	1582	550	544
2022-06-16	382	1582	550	544
2022-06-19	382	1582	550	544
2022-06-21	382	1661	550	544
2022-06-23	382	1744	550	544
2022-06-26	382	1744	550	544
2022-06-28	382	1744	550	544
2022-06-30	382	1751	495	544
2022-07-03	382	1751	495	544
2022-07-06	382	1751	495	545
2022-07-07	382	1832	495	545
2022-07-11	382	1832	495	545
2022-07-12	382	1832	495	544
2022-07-14	382	1832	495	544
2022-07-17	382	1832	495	544
2022-07-19	382	1900	495	544
2022-07-21	382	1900	495	544
2022-07-24	382	1900	495	544
2022-07-26	382	1900	495	544
2022-07-28	347	1838	495	544
2022-07-31	347	1838	495	544
2022-08-02	347	1900	495	544
2022-08-04	347	1900	495	534
2022-08-07	347	1900	495	534
2022-08-09	347	1900	495	534
2022-08-11	347	1900	495	534
2022-08-14	347	1900	495	534
2022-08-16	347	1900	495	534
2022-08-18	347	1900	495	534
2022-08-21	380	1900	495	534
2022-08-23	380	1900	495	534
2022-08-25	380	1900	495	534
2022-08-28	380	1900	495	534
2022-08-30	380	1900	495	534
2022-09-01	380	1725	495	534
2022-09-04	380	1725	495	534
2022-09-06	380	1725	495	534
2022-09-08	380	1725	495	534

2022-09-11	380	1725	495	534
2022-09-13	380	1725	495	534
2022-09-15	380	1725	495	534
2022-09-18	380	1725	495	534
2022-09-20	380	1725	495	534
2022-09-22	380	1725	495	534
2022-09-25	380	1725	495	534
2022-09-27	380	1725	495	534
2022-09-29	380	1725	495	534
2022-10-02	380	1728	495	534
2022-10-04	380	1728	495	534
2022-10-06	380	1728	495	534
2022-10-09	380	1728	495	534
2022-10-11	380	1728	495	534
2022-10-13	380	1728	495	534
2022-10-16	380	1728	495	534
2022-10-18	380	1728	495	534
2022-10-20	381	1728	495	534
2022-10-23	399	1728	495	534
2022-10-25	399	1728	495	534
2022-10-27	399	1751	495	534
2022-10-30	399	1751	495	534
2022-11-02	381	1751	495	534
2022-11-03	381	1751	495	534
2022-11-06	381	1751	495	534
2022-11-08	381	1839	495	534
2022-11-10	381	1839	495	530
2022-11-13	383	1839	495	530
2022-11-15	383	1839	495	530
2022-11-17	383	1839	495	530
2022-11-20	383	1839	495	530
2022-11-22	383	1839	495	530
2022-11-24	383	1839	495	530
2022-11-27	383	1839	495	530
2022-11-29	402	1839	495	530
2022-12-01	402	1839	495	530
2022-12-04	402	1839	495	530
2022-12-06	402	1839	495	530
2022-12-08	403	1839	495	530
2022-12-11	403	1839	495	530
2022-12-13	403	1839	495	530
2022-12-15	403	1839	495	530
2022-12-18	403	1839	495	530
2022-12-20	403	2022	495	530
2022-12-22	403	2022	495	530
2022-12-25	403	2022	495	530
2022-12-27	403	2123	495	530

2022-12-29	403	2123	495	530
2023-01-02	403	2123	495	530
2023-01-03	403	2123	495	530
2023-01-05	403	2123	495	530
2023-01-08	403	2229	495	530
2023-01-10	403	2229	495	530
2023-01-15	403	2229	495	530
2023-01-17	403	2340	495	530
2023-01-19	402	2340	495	530
2023-01-22	403	2340	495	530
2023-01-24	403	2457	495	530
2023-01-26	404	2342	495	530
2023-01-29	404	2342	495	530
2023-01-31	404	2401	495	530
2023-02-02	404	2401	495	530
2023-02-05	404	2401	495	530
2023-02-07	420	2340	495	530

Annexe 2 : Rendements journaliers des actifs sur le R

RStudio Source Editor

portfolioReturns ×

Filter

	ALL	BIO	AUR	SAI
2022-02-10	0.000000000	0.000000000	0.000000000	0.000000000
2022-02-13	0.000000000	0.000000000	0.000000000	0.000000000
2022-02-15	0.000000000	0.000000000	0.000000000	0.000000000
2022-02-17	0.000000000	0.000000000	0.000000000	0.000000000
2022-02-20	0.000000000	0.000000000	0.000000000	0.000000000
2022-02-22	0.000000000	0.000000000	0.000000000	0.000000000
2022-02-24	0.000000000	0.000000000	0.000000000	0.000000000
2022-02-27	0.000000000	0.000000000	0.000000000	0.000000000
2022-03-01	0.000000000	0.0015337426	0.000000000	0.000000000
2022-03-03	0.000000000	0.000000000	0.000000000	0.000000000
2022-03-06	0.000000000	0.000000000	0.000000000	0.000000000
2022-03-08	0.000000000	0.000000000	-0.01801851	0.000000000
2022-03-10	0.000000000	0.000000000	0.000000000	0.000000000
2022-03-13	0.000000000	0.000000000	0.000000000	0.000000000
2022-03-15	0.000000000	0.000000000	0.000000000	0.000000000
2022-03-17	0.000000000	0.000000000	0.000000000	0.000000000
2022-03-20	0.000000000	0.000000000	0.000000000	0.000000000
2022-03-22	0.000000000	0.0007659901	0.000000000	0.000000000
2022-03-24	0.021639175	0.000000000	0.000000000	0.000000000
2022-03-27	0.000000000	0.000000000	0.000000000	0.000000000
2022-03-29	0.000000000	0.000000000	0.000000000	0.000000000
2022-03-31	0.000000000	0.000000000	0.000000000	0.000000000
2022-04-03	0.000000000	0.000000000	0.000000000	0.000000000
2022-04-05	0.000000000	0.0007654038	0.000000000	0.000000000
2022-04-07	0.000000000	0.000000000	0.000000000	0.000000000
2022-04-10	0.000000000	0.000000000	0.000000000	0.000000000
2022-04-12	0.003053437	0.000000000	0.000000000	0.000000000
2022-04-14	0.000000000	0.000000000	0.000000000	0.000000000
2022-04-17	0.000000000	0.000000000	0.000000000	0.000000000
2022-04-19	0.000000000	0.000000000	0.000000000	0.000000000
2022-04-21	0.000000000	0.000000000	0.000000000	0.000000000
2022-04-24	0.000000000	0.000000000	0.000000000	0.000000000
2022-04-26	0.093090423	0.000000000	0.000000000	0.000000000
2022-04-28	0.000000000	0.000000000	0.000000000	0.000000000
2022-05-04	0.000000000	0.000000000	0.000000000	0.000000000
2022-05-05	0.000000000	0.000000000	0.000000000	0.000000000
2022-05-08	0.000000000	0.000000000	0.000000000	0.000000000
2022-05-10	0.000000000	0.000000000	0.000000000	0.000000000
2022-05-12	0.000000000	0.000000000	0.000000000	0.000000000
2022-05-15	0.000000000	0.000000000	0.000000000	0.000000000
2022-05-17	0.002773927	0.000000000	0.000000000	0.000000000
2022-05-19	0.000000000	0.000000000	0.000000000	0.001839927
2022-05-22	0.000000000	0.000000000	0.000000000	0.000000000
2022-05-24	0.048658247	0.0948231725	0.000000000	0.000000000
2022-05-26	0.000000000	0.000000000	0.000000000	0.000000000
2022-05-29	0.000000000	0.000000000	0.000000000	0.000000000
2022-05-31	0.000000000	0.000000000	0.000000000	0.000000000
2022-06-02	0.000000000	0.0482266625	0.000000000	0.000000000
2022-06-05	0.002635048	0.000000000	0.000000000	0.000000000
2022-06-07	0.000000000	0.000000000	0.000000000	0.000000000
2022-06-09	0.000000000	0.000000000	0.000000000	0.000000000

2022-06-09	0.00000000	0.00000000	0.00000000	0.00000000
2022-06-12	0.00000000	0.0479055998	0.00000000	0.00000000
2022-06-14	0.00000000	0.00000000	0.00000000	0.00000000
2022-06-16	0.005249356	0.00000000	0.00000000	0.00000000
2022-06-19	0.00000000	0.00000000	0.00000000	0.00000000
2022-06-21	0.00000000	0.0487299613	0.00000000	0.00000000
2022-06-23	0.00000000	0.0487614949	0.00000000	0.00000000
2022-06-26	0.00000000	0.00000000	0.00000000	0.00000000
2022-06-28	0.00000000	0.00000000	0.00000000	0.00000000
2022-06-30	0.00000000	0.0040057278	-0.10536052	0.00000000
2022-07-03	0.00000000	0.00000000	0.00000000	0.00000000
2022-07-06	0.00000000	0.00000000	0.00000000	0.001836548
2022-07-07	0.00000000	0.0452212129	0.00000000	0.00000000
2022-07-11	0.00000000	0.00000000	0.00000000	0.00000000
2022-07-12	0.00000000	0.00000000	0.00000000	-0.001836548
2022-07-14	0.00000000	0.00000000	0.00000000	0.00000000
2022-07-17	0.00000000	0.00000000	0.00000000	0.00000000
2022-07-19	0.00000000	0.0364456199	0.00000000	0.00000000
2022-07-21	0.00000000	0.00000000	0.00000000	0.00000000
2022-07-24	0.00000000	0.00000000	0.00000000	0.00000000
2022-07-26	0.00000000	0.00000000	0.00000000	0.00000000
2022-07-28	-0.096095829	-0.0331758622	0.00000000	0.00000000
2022-07-31	0.00000000	0.00000000	0.00000000	0.00000000
2022-08-02	0.00000000	0.0331758622	0.00000000	0.00000000
2022-08-04	0.00000000	0.00000000	0.00000000	-0.018553408
2022-08-07	0.00000000	0.00000000	0.00000000	0.00000000
2022-08-09	0.00000000	0.00000000	0.00000000	0.00000000
2022-08-11	0.00000000	0.00000000	0.00000000	0.00000000
2022-08-14	0.00000000	0.00000000	0.00000000	0.00000000
2022-08-16	0.00000000	0.00000000	0.00000000	0.00000000
2022-08-18	0.00000000	0.00000000	0.00000000	0.00000000
2022-08-21	0.090846473	0.00000000	0.00000000	0.00000000
2022-08-23	0.00000000	0.00000000	0.00000000	0.00000000
2022-08-25	0.00000000	0.00000000	0.00000000	0.00000000
2022-08-28	0.00000000	0.00000000	0.00000000	0.00000000
2022-08-30	0.00000000	0.00000000	0.00000000	0.00000000
2022-09-01	0.00000000	-0.0966268357	0.00000000	0.00000000
2022-09-04	0.00000000	0.00000000	0.00000000	0.00000000
2022-09-06	0.00000000	0.00000000	0.00000000	0.00000000
2022-09-08	0.00000000	0.00000000	0.00000000	0.00000000
2022-09-11	0.00000000	0.00000000	0.00000000	0.00000000
2022-09-13	0.00000000	0.00000000	0.00000000	0.00000000
2022-09-15	0.00000000	0.00000000	0.00000000	0.00000000
2022-09-18	0.00000000	0.00000000	0.00000000	0.00000000
2022-09-20	0.00000000	0.00000000	0.00000000	0.00000000
2022-09-22	0.00000000	0.00000000	0.00000000	0.00000000
2022-09-25	0.00000000	0.00000000	0.00000000	0.00000000
2022-09-27	0.00000000	0.00000000	0.00000000	0.00000000
2022-09-29	0.00000000	0.00000000	0.00000000	0.00000000
2022-10-02	0.00000000	0.0017376199	0.00000000	0.00000000
2022-10-04	0.00000000	0.00000000	0.00000000	0.00000000
2022-10-06	0.00000000	0.00000000	0.00000000	0.00000000
2022-10-09	0.00000000	0.00000000	0.00000000	0.00000000
2022-10-11	0.00000000	0.00000000	0.00000000	0.00000000
2022-10-13	0.00000000	0.00000000	0.00000000	0.00000000
2022-10-16	0.00000000	0.00000000	0.00000000	0.00000000

2022-10-16	0.000000000	0.000000000	0.00000000	0.00000000
2022-10-18	0.000000000	0.000000000	0.00000000	0.00000000
2022-10-20	0.002628122	0.000000000	0.00000000	0.00000000
2022-10-23	0.046162042	0.000000000	0.00000000	0.00000000
2022-10-25	0.000000000	0.000000000	0.00000000	0.00000000
2022-10-27	0.000000000	0.0132223829	0.00000000	0.00000000
2022-10-30	0.000000000	0.000000000	0.00000000	0.00000000
2022-11-02	-0.046162042	0.000000000	0.00000000	0.00000000
2022-11-03	0.000000000	0.000000000	0.00000000	0.00000000
2022-11-06	0.000000000	0.000000000	0.00000000	0.00000000
2022-11-08	0.000000000	0.0490348923	0.00000000	0.00000000
2022-11-10	0.000000000	0.000000000	0.00000000	-0.007518832
2022-11-13	0.005235614	0.000000000	0.00000000	0.00000000
2022-11-15	0.000000000	0.000000000	0.00000000	0.00000000
2022-11-17	0.000000000	0.000000000	0.00000000	0.00000000
2022-11-20	0.000000000	0.000000000	0.00000000	0.00000000
2022-11-22	0.000000000	0.000000000	0.00000000	0.00000000
2022-11-24	0.000000000	0.000000000	0.00000000	0.00000000
2022-11-27	0.000000000	0.000000000	0.00000000	0.00000000
2022-11-29	0.048417099	0.000000000	0.00000000	0.00000000
2022-12-01	0.000000000	0.000000000	0.00000000	0.00000000
2022-12-04	0.000000000	0.000000000	0.00000000	0.00000000
2022-12-06	0.000000000	0.000000000	0.00000000	0.00000000
2022-12-08	0.002484473	0.000000000	0.00000000	0.00000000
2022-12-11	0.000000000	0.000000000	0.00000000	0.00000000
2022-12-13	0.000000000	0.000000000	0.00000000	0.00000000
2022-12-15	0.000000000	0.000000000	0.00000000	0.00000000
2022-12-18	0.000000000	0.000000000	0.00000000	0.00000000
2022-12-20	0.000000000	0.0948651750	0.00000000	0.00000000
2022-12-22	0.000000000	0.000000000	0.00000000	0.00000000
2022-12-25	0.000000000	0.000000000	0.00000000	0.00000000
2022-12-27	0.000000000	0.0487430621	0.00000000	0.00000000
2022-12-29	0.000000000	0.000000000	0.00000000	0.00000000
2023-01-02	0.000000000	0.000000000	0.00000000	0.00000000
2023-01-03	0.000000000	0.000000000	0.00000000	0.00000000
2023-01-05	0.000000000	0.000000000	0.00000000	0.00000000
2023-01-08	0.000000000	0.0487228717	0.00000000	0.00000000
2023-01-10	0.000000000	0.000000000	0.00000000	0.00000000
2023-01-15	0.000000000	0.000000000	0.00000000	0.00000000
2023-01-17	0.000000000	0.0485978750	0.00000000	0.00000000
2023-01-19	-0.002484473	0.000000000	0.00000000	0.00000000
2023-01-22	0.002484473	0.000000000	0.00000000	0.00000000
2023-01-24	0.000000000	0.0487901642	0.00000000	0.00000000
2023-01-29	0.000000000	0.000000000	0.00000000	0.00000000
2023-01-31	0.000000000	0.0248800521	0.00000000	0.00000000
2023-02-02	0.000000000	0.000000000	0.00000000	0.00000000
2023-02-05	0.000000000	0.000000000	0.00000000	0.00000000
2023-02-07	0.038839833	-0.0257343879	0.00000000	0.00000000

Annexe 2 : Rendements journaliers des actifs sur le R

Annexe 3 les variations quotidiennes de rendements des actifs sur R

	portfolio.returns
2022-02-10	0.0000000000
2022-02-13	0.0000000000
2022-02-15	0.0000000000
2022-02-17	0.0000000000
2022-02-20	0.0000000000
2022-02-22	0.0000000000
2022-02-24	0.0000000000
2022-02-27	0.0000000000
2022-03-01	0.0006134971
2022-03-03	0.0000000000
2022-03-06	0.0000000000
2022-03-08	-0.0018007458
2022-03-10	0.0000000000
2022-03-13	0.0000000000
2022-03-15	0.0000000000
2022-03-17	0.0000000000
2022-03-20	0.0000000000
2022-03-22	0.0003072311
2022-03-24	0.0043316533
2022-03-27	0.0000000000
2022-03-29	0.0000000000
2022-03-31	0.0000000000
2022-04-03	0.0000000000
2022-04-05	0.0003058120
2022-04-07	0.0000000000
2022-04-10	0.0000000000
2022-04-12	0.0006215694
2022-04-14	0.0000000000
2022-04-17	0.0000000000
2022-04-19	0.0000000000
2022-04-21	0.0000000000
2022-04-24	0.0000000000
2022-04-26	0.0189958967
2022-04-28	0.0000000000
2022-05-04	0.0000000000
2022-05-05	0.0000000000
2022-05-08	0.0000000000
2022-05-10	0.0000000000
2022-05-12	0.0000000000
2022-05-15	0.0000000000
2022-05-17	0.0006072024
2022-05-19	0.0005390002

2022-05-22	0.0000000000
2022-05-24	0.0477995395
2022-05-26	0.0000000000
2022-05-29	0.0000000000
2022-05-31	0.0000000000
2022-06-02	0.0197322447
2022-06-05	0.0005670246
2022-06-07	0.0000000000
2022-06-09	0.0000000000
2022-06-12	0.0201371698
2022-06-14	0.0000000000
2022-06-16	0.0011095774
2022-06-19	0.0000000000
2022-06-21	0.0210179420
2022-06-23	0.0216023718
2022-06-26	0.0000000000
2022-06-28	0.0000000000
2022-06-30	-0.0070446201
2022-07-03	0.0000000000
2022-07-06	0.0004763862
2022-07-07	0.0207855293
2022-07-11	0.0000000000
2022-07-12	-0.0004673204
2022-07-14	0.0000000000
2022-07-17	0.0000000000
2022-07-19	0.0171609345
2022-07-21	0.0000000000
2022-07-24	0.0000000000
2022-07-26	0.0000000000
2022-07-28	-0.0348834727
2022-07-31	0.0000000000
2022-08-02	0.0159456560
2022-08-04	-0.0047271671
2022-08-07	0.0000000000
2022-08-09	0.0000000000
2022-08-11	0.0000000000
2022-08-14	0.0000000000
2022-08-16	0.0000000000
2022-08-18	0.0000000000
2022-08-21	0.0166076748
2022-08-23	0.0000000000

Annexe 3 les variations quotidiennes de rendements des actifs sur R

2022-08-25	0.0000000000
2022-08-28	0.0000000000
2022-08-30	0.0000000000
2022-09-01	-0.0466794973
2022-09-04	0.0000000000
2022-09-06	0.0000000000
2022-09-08	0.0000000000
2022-09-11	0.0000000000
2022-09-13	0.0000000000
2022-09-15	0.0000000000
2022-09-18	0.0000000000
2022-09-20	0.0000000000
2022-09-22	0.0000000000
2022-09-25	0.0000000000
2022-09-27	0.0000000000
2022-09-29	0.0000000000
2022-10-02	0.0007954474
2022-10-04	0.0000000000
2022-10-06	0.0000000000
2022-10-09	0.0000000000
2022-10-11	0.0000000000
2022-10-13	0.0000000000
2022-10-16	0.0000000000
2022-10-18	0.0000000000
2022-10-20	0.0005403465
2022-10-23	0.0095107998
2022-10-25	0.0000000000
2022-10-27	0.0059983179
2022-10-30	0.0000000000
2022-11-02	-0.0097973309
2022-11-03	0.0000000000
2022-11-06	0.0000000000
2022-11-08	0.0226260325
2022-11-10	-0.0018929135
2022-11-13	0.0010486895
2022-11-15	0.0000000000

2022-11-17	0.0000000000
2022-11-20	0.0000000000
2022-11-22	0.0000000000
2022-11-24	0.0000000000
2022-11-27	0.0000000000
2022-11-29	0.0097384698
2022-12-01	0.0000000000
2022-12-04	0.0000000000
2022-12-06	0.0000000000
2022-12-08	0.0005188615
2022-12-11	0.0000000000
2022-12-13	0.0000000000
2022-12-15	0.0000000000
2022-12-18	0.0000000000
2022-12-20	0.0444853035
2022-12-22	0.0000000000
2022-12-25	0.0000000000
2022-12-27	0.0239596717
2022-12-29	0.0000000000
2023-01-02	0.0000000000
2023-01-03	0.0000000000
2023-01-05	0.0000000000
2023-01-08	0.0245294144
2023-01-10	0.0000000000
2023-01-15	0.0000000000
2023-01-17	0.0250442419
2023-01-19	-0.0004628621
2023-01-22	0.0004619260
2023-01-24	0.0257211127
2023-01-26	-0.0253889438
2023-01-29	0.0000000000
2023-01-31	0.0131009551
2023-02-02	0.0000000000
2023-02-05	0.0000000000
2023-02-07	-0.0065460440

Annexe 4 : les variations quotidiennes de rendements des actifs sur Python

Date	ALL	BIO	AUR	SAI	Date	ALL	BIO	AUR	SAI
2022-02-08	320	1303	560	543	2022-04-14	328	1307	550	543
2022-02-10	320	1303	560	543	2022-04-17	328	1307	550	543
2022-02-13	320	1303	560	543	2022-04-19	328	1307	550	543
2022-02-15	320	1303	560	543	2022-04-21	328	1307	550	543
2022-02-17	320	1303	560	543	2022-04-24	328	1307	550	543
2022-02-20	320	1303	560	543	2022-04-26	360	1307	550	543
2022-02-22	320	1303	560	543	2022-04-28	360	1307	550	543
2022-02-24	320	1303	560	543	2022-05-04	360	1307	550	543
2022-02-27	320	1303	560	543	2022-05-05	360	1307	550	543
2022-03-01	320	1305	560	543	2022-05-08	360	1307	550	543
2022-03-03	320	1305	560	543	2022-05-10	360	1307	550	543
2022-03-06	320	1305	560	543	2022-05-12	360	1307	550	543
2022-03-08	320	1305	550	543	2022-05-15	360	1307	550	543
2022-03-10	320	1305	550	543	2022-05-17	361	1307	550	543
2022-03-13	320	1305	550	543	2022-05-19	361	1307	550	544
2022-03-15	320	1305	550	543	2022-05-22	361	1307	550	544
2022-03-17	320	1305	550	543	2022-05-24	379	1437	550	544
2022-03-20	320	1305	550	543	2022-05-26	379	1437	550	544
2022-03-22	320	1306	550	543	2022-05-29	379	1437	550	544
2022-03-24	327	1306	550	543	2022-05-31	379	1437	550	544
2022-03-27	327	1306	550	543	2022-06-02	379	1508	550	544
2022-03-29	327	1306	550	543	2022-06-05	380	1508	550	544
2022-03-31	327	1306	550	543	2022-06-07	380	1508	550	544
2022-04-03	327	1306	550	543	2022-06-09	380	1508	550	544
2022-04-05	327	1307	550	543	2022-06-12	380	1582	550	544
2022-04-07	327	1307	550	543	2022-06-14	380	1582	550	544
2022-04-10	327	1307	550	543	2022-06-16	382	1582	550	544
2022-04-12	328	1307	550	543	2022-06-19	382	1582	550	544
					2022-06-21	382	1661	550	544
					2022-06-23	382	1744	550	544

Annexe 4 : les variations quotidiennes de rendements des actifs sur Python

2022-06-28	0	0	0	0	2022-09-11	0	0	0	0
2022-06-30	0	0.00401376	-0.1	0	2022-09-13	0	0	0	0
2022-07-03	0	0	0	0	2022-09-15	0	0	0	0
2022-07-06	0	0	0	0.00183824	2022-09-18	0	0	0	0
2022-07-07	0	0.0462593	0	0	2022-09-20	0	0	0	0
2022-07-11	0	0	0	0	2022-09-22	0	0	0	0
2022-07-12	0	0	0	-0.00183486	2022-09-25	0	0	0	0
2022-07-14	0	0	0	0	2022-09-27	0	0	0	0
2022-07-17	0	0	0	0	2022-09-29	0	0	0	0
2022-07-19	0	0.0371179	0	0	2022-10-02	0	0.00173913	0	0
2022-07-21	0	0	0	0	2022-10-04	0	0	0	0
2022-07-24	0	0	0	0	2022-10-06	0	0	0	0
2022-07-26	0	0	0	0	2022-10-09	0	0	0	0
2022-07-28	-0.091623	-0.0326316	0	0	2022-10-11	0	0	0	0
2022-07-31	0	0	0	0	2022-10-13	0	0	0	0
2022-08-02	0	0.0337323	0	0	2022-10-16	0	0	0	0
2022-08-04	0	0	0	-0.0183824	2022-10-18	0	0	0	0
2022-08-07	0	0	0	0	2022-10-20	0.00263158	0	0	0
2022-08-09	0	0	0	0	2022-10-23	0.0472441	0	0	0
2022-08-11	0	0	0	0	2022-10-25	0	0	0	0
2022-08-14	0	0	0	0	2022-10-27	0	0.0133102	0	0
2022-08-16	0	0	0	0	2022-10-30	0	0	0	0
2022-08-18	0	0	0	0	2022-11-02	-0.0451128	0	0	0
2022-08-21	0.0951009	0	0	0	2022-11-03	0	0	0	0
2022-08-23	0	0	0	0	2022-11-06	0	0	0	0
2022-08-25	0	0	0	0	2022-11-08	0	0.050257	0	0
2022-08-28	0	0	0	0	2022-11-10	0	0	0	-0.00749064
2022-08-30	0	0	0	0	2022-11-13	0.00524934	0	0	0
2022-09-01	0	-0.0921053	0	0	2022-11-15	0	0	0	0
2022-09-04	0	0	0	0	2022-11-17	0	0	0	0
2022-09-06	0	0	0	0					
2022-09-08	0	0	0	0					

Annexe 4 : les variations quotidiennes de rendements des actifs sur Python

2022-11-20	0	0	0	0
2022-11-22	0	0	0	0
2022-11-24	0	0	0	0
2022-11-27	0	0	0	0
2022-11-29	0.0496084	0	0	0
2022-12-01	0	0	0	0
2022-12-04	0	0	0	0
2022-12-06	0	0	0	0
2022-12-08	0.00248756	0	0	0
2022-12-11	0	0	0	0
2022-12-13	0	0	0	0
2022-12-15	0	0	0	0
2022-12-18	0	0	0	0
2022-12-20	0	0.0995106	0	0
2022-12-22	0	0	0	0
2022-12-25	0	0	0	0
2022-12-27	0	0.0499505	0	0
2022-12-29	0	0	0	0
2023-01-02	0	0	0	0
2023-01-03	0	0	0	0
2023-01-05	0	0	0	0
2023-01-08	0	0.0499293	0	0
2023-01-10	0	0	0	0
2023-01-15	0	0	0	0
2023-01-17	0	0.0497981	0	0
2023-01-19	-0.00248139	0	0	0
2023-01-22	0.00248756	0	0	0
2023-01-24	0	0.05	0	0
2023-01-26	0.00248139	-0.046805	0	0
2023-01-29	0	0	0	0
2023-01-31	0	0.0251921	0	0
2023-02-02	0	0	0	0
2023-02-05	0	0	0	0
2023-02-07	0.039604	-0.0254061	0	0

Annexe 5 : les variations quotidiennes de rendements des actifs sur Python

Date	0	Date	0	Date	0
2022-02-10	0	2022-04-07	0	2022-06-05	0.000527704
2022-02-13	0	2022-04-10	0	2022-06-07	0
2022-02-15	0	2022-04-12	0.000611621	2022-06-09	0
2022-02-17	0	2022-04-14	0	2022-06-12	0.0196286
2022-02-20	0	2022-04-17	0	2022-06-14	0
2022-02-22	0	2022-04-19	0	2022-06-16	0.00105263
2022-02-24	0	2022-04-21	0	2022-06-19	0
2022-02-27	0	2022-04-24	0	2022-06-21	0.0199747
2022-03-01	0.000613968	2022-04-26	0.0195122	2022-06-23	0.019988
2022-03-03	0	2022-04-28	0	2022-06-26	0
2022-03-06	0	2022-05-04	0	2022-06-28	0
2022-03-08	-0.00178571	2022-05-05	0	2022-06-30	-0.0083945
2022-03-10	0	2022-05-08	0	2022-07-03	0
2022-03-13	0	2022-05-10	0	2022-07-06	0.000551471
2022-03-15	0	2022-05-12	0	2022-07-07	0.0185037
2022-03-17	0	2022-05-15	0	2022-07-11	0
2022-03-20	0	2022-05-17	0.000555556	2022-07-12	-0.000550459
2022-03-22	0.000306513	2022-05-19	0.000552486	2022-07-14	0
2022-03-24	0.004375	2022-05-22	0	2022-07-17	0
2022-03-27	0	2022-05-24	0.0497581	2022-07-19	0.0148472
2022-03-29	0	2022-05-26	0	2022-07-21	0
2022-03-31	0	2022-05-29	0	2022-07-24	0
2022-04-03	0	2022-05-31	0	2022-07-26	0
2022-04-05	0.000306279	2022-06-02	0.0197634	2022-07-28	-0.0313772

Annexe 5 : les variations quotidiennes de rendements des actifs sur Python

Date	0	Date	0	Date	0
2022-07-31	0	2022-09-25	0	2022-11-20	0
2022-08-02	0.0134929	2022-09-27	0	2022-11-22	0
2022-08-04	-0.00551471	2022-09-29	0	2022-11-24	0
2022-08-07	0	2022-10-02	0.000695652	2022-11-27	0
2022-08-09	0	2022-10-04	0	2022-11-29	0.00992167
2022-08-11	0	2022-10-06	0	2022-12-01	0
2022-08-14	0	2022-10-09	0	2022-12-04	0
2022-08-16	0	2022-10-11	0	2022-12-06	0
2022-08-18	0	2022-10-13	0	2022-12-08	0.000497512
2022-08-21	0.0190202	2022-10-16	0	2022-12-11	0
2022-08-23	0	2022-10-18	0	2022-12-13	0
2022-08-25	0	2022-10-20	0.000526316	2022-12-15	0
2022-08-28	0	2022-10-23	0.00944882	2022-12-18	0
2022-08-30	0	2022-10-25	0	2022-12-20	0.0398042
2022-09-01	-0.0368421	2022-10-27	0.00532407	2022-12-22	0
2022-09-04	0	2022-10-30	0	2022-12-25	0
2022-09-06	0	2022-11-02	-0.00902256	2022-12-27	0.0199802
2022-09-08	0	2022-11-03	0	2022-12-29	0
2022-09-11	0	2022-11-06	0	2023-01-02	0
2022-09-13	0	2022-11-08	0.0201028	2023-01-03	0
2022-09-15	0	2022-11-10	-0.00224719	2023-01-05	0
2022-09-18	0	2022-11-13	0.00104987	2023-01-08	0.0199717
2022-09-20	0	2022-11-15	0	2023-01-10	0
2022-09-22	0	2022-11-17	0	2023-01-15	0

Table des matières

Remerciements	I
Liste des abréviations	IV
Liste des tableaux et figures.....	VI
Sommaire	VII
Introduction générale.....	1
Chapitre I : Le Big Data et la Data Science : Un domaine en plein essor.....	4
Introduction.....	5
Section 1 : Qu'est ce que le <i>Big Data et la Data science</i> ?.....	6
1.1. Historique et origine de l'évolution du Big Data	6
1.2. Définitions des termes et contexte	8
1.3. Les 5 « V » de la Big Data	10
1.3.1. Le volume.....	11
1.3.2. La vitesse.....	11
1.3.3. La variété.....	12
1.3.4. La véracité.....	12
1.3.5. La valeur.....	12
1.4. Les Types de Données.....	12
1.4.1. Les données structurées	13
1.4.2. Les données non structurées	13
1.4.3. Les données semi-structurées :	14
1.5. Du Big Data a la <i>Data science</i>	15
1.6. Les grandes tendances <i>Big Data</i>	16
Section 2 : Business intelligence VS <i>Data science</i>	18
2.1 Définition de la Business Intelligence.....	18
2.2 Différence entre le <i>Business Intelligence</i> et <i>Data science</i>	19
2.2.1. Perspective.....	19
2.2.2. Types de données	20
2.2.3. Livrables	20
2.2.4. Processus.....	20
Section 3 :Les grands défis en matière de <i>Big Data</i> dans la finance.....	21
3.1 Défis en termes de données	21

3.1.1. Qualité des données.....	21
3.1.2. Intégration des données	22
3.1.3. Analyse des données.....	22
3.2 Autres défis.....	22
3.2.1 Réglementation.....	22
3.2.2 Gestion de la capacité.....	23
3.2.3 Etique.....	23
Conclusion.....	24
Chapitre II : Les technologies de la data science applique à la finance.....	25
Introduction.....	26
Section 01 : APACHE HADOOP	26
1.1. Définition de Hadoop.....	26
1.2. Historique de Hadoop.....	27
1.3. Evolution du système HADOOP.....	28
1.3.1 Le système de fichier distribué d'Hadoop HDFS.....	28
1.3.2 MapReduce.....	30
1.4 L'écosystème Hadoop :.....	31
1.5 Hadoop dans la banque et la finance.....	32
1.5.1 Analyse des données	32
1.5.2 La détection de la fraude	33
1.5.3 Gestion des risques.....	33
1.5.4 Stockage et sécurité des données.....	33
1.5.5 Segmentation de la clientèle.....	34
1.6 Hadoop et les grands acteurs financiers	35
Section 02 : Le logiciel Python	37
2.1 Définition de Python	37
2.2 Historique de Python.....	37
2.3 Caractéristique du Python	38
2.3.1 Open source.....	38
2.3.2 Interprété.....	38
2.3.3 Multiparadigme.....	38
2.3.4 Multi-usages.....	38
2.3.5 Multiplateforme.....	38

2.3.6 Typage dynamique.....	39
2.3.7 Sensible à l'indentation.....	39
2.3.8 Collecte des déchets.....	39
2.4 Les concepts de bases de python :.....	40
2.4.1 Variable.....	40
2.4.2 Types de données.....	41
2.4.3 Structures de contrôle de flux.....	42
2.4.4 Les fonctions.....	42
2.4.5 Les modules.....	43
2.5 Fonctionnement et environnement technologiques (Écosystème de Python).....	43
2.6 Les différentes bibliothèques et frameworks disponibles pour la finance	44
2.6.1 Numpy.....	44
2.6.2 Pandas.....	44
2.6.3 Scipy.....	44
2.6.4 Scikit-learn.....	45
2.6.5 Matplotlib.....	45
2.6.6 QuantLib.....	45
2.6.7 Pyfolio.....	46
2.6.8 TensorFlow.....	46
2.7 Les avantages de l'écosystème Python pour la finance	46
2.8 Python pour la finance.....	47
2.8.1 Développement de modèles prédictifs.....	47
2.8.2 Trading algorithmique.....	48
2.8.3 Gestion de portefeuille.....	48
2.8.4 Logiciel bancaires.....	49
2.9 Les obstacles à l'utilisation de Python en finance	49
Section 03 : Le logiciel « R ».....	51
3.1 Initiation au « R » :.....	51
3.1.1 Origines Le logiciel.....	51
3.1.2 Description sommaire de R	52
3.2 Installation de R et RStudio	53
3.2.1 Mécanismes de fonctionnement du "R".....	56

3.3 Utilisation du langage "R" dans la science de gestion.....	63
3.3.1 Utilisation de R dans l'analyse de données de marché	63
3.3.2 La Gestion des risques avec R.....	64
3.3.3 Optimisation de portefeuille.....	65
3.3.4 Analyse de segmentation de la clientèle avec R (clustering)	65
3.3.5 R dans l'analyse des séries temporelles	66
3.4 Python ou R : quel est le meilleur pour la <i>data science</i> en finance.....	67
Conclusion.....	68
Chapitre III : Les <i>uses cases</i> de la <i>data science</i> en finance.....	69
Introduction.....	70
Section 1 : Applications de la <i>data science</i> pour l'optimisation des processus et la croissance des entreprises	70
1.1 Applications de la <i>data science</i> pour le marketing	70
1.2 Applications de la <i>data science</i> pour la gestion de performance	72
1.3 . Applications de la <i>data science</i> dans la détection de fraude	74
1.4 Applications de la <i>data science</i> pour le trading algorithme.....	76
Section 2 : la <i>Data science</i> pour les entreprise moderne et Startups.....	78
2.1 Les « Machines Learning » au profit de l'innovation et des Startup	78
2.2 Les <i>Data scientists</i> des « <i>assets</i> » pour les entreprises moderne	79
2.3 Comment le <i>Big Data</i> aide-t-il les entreprises numériques ?.....	80
Conclusion.....	81
Chapitre IV : La gestion d'un portefeuille avec les langages R et Python cas : Bourse d'Alger.....	82
Introduction.....	83
Section 01 : Présentation du marché financier algérien SGBV.....	84
1.1.Présentation de la SGBV :.....	84
1.2.Fonctionnement de la bourse.....	84
1.2.1.Le marché des titres de capital	84
1.2.2. Le marché des titres des créance.....	85
1.3. Structure et organisation de la Bourse d'Alger :	85
1.3.1 La commission d'organisation & de surveillance des opérations de bourse (COSOB) .86	
1.3.2 Les intermédiaires en opération de bourse (IOB).....	87
1.3.3 Les organismes de placement collectifs en valeurs mobilières OPCVM.....	88
1.3.4 Algérie clearing.....	88

1.4 Le processus de cotations et négociation des titres	88
1.4.1 Ouverture de la séance	90
1.4.2 Détermination du cours	90
1.4.3 Allocation des ordres	90
1.4.4 Effectuer les transactions en bloc	90
1.4.5 Clôture de la séance de négociation	90
Section 2 : Gestion de portefeuille avec R studio.....	93
2.1 Préparation des données	93
2.2 Importation des prix du portefeuille.....	94
2.3 Calcul du rendement du portefeuille :	95
2.3.1 Rendements journaliers des actifs	96
2.3.2 Rendements globale journalier du portefeuille.....	96
2.3.3 Esperance de rendement des actifs.....	97
2.3.4 Esperance de rendements globale du portefeuille.....	97
2.3.5 Rendement et risque annuel du portefeuille.....	98
2.3.6 Rendement mensuelle et annuelle du portefeuille.	99
2.3.7 Graphique des rendements du portefeuille.....	99
2.4 Calcul des ratios :	100
2.4.1 Covariance (covMatrix) :	101
2.4.2 Sharpe ratio	102
2.5 Calcul du Beta avec la methode de Capital Asset Pricing Model (CAPM).....	103
2.5.1 La moyenne des rendements du marché.....	105
2.5.2 La variance du marché	105
2.5.3 Le beta.....	105
2.6 Calculer l'esperance de rendement du portefeuille avec la methode CAPM.....	106
Section 3 : Gestion de portefeuille avec Spyder (Python).....	107
3.1 Préparation des données	107
3.2 Importation des prix du portefeuille	108
3.3 Calcul des rendements du portefeuille	108
3.3.1 Rendements journaliers des actifs.....	108
3.3.2 Rendements globale journalier du portefeuille.....	109
3.3.3 Espérance de rendement des actifs	110
3.3.4 Esperance de rendements globale du portefeuille.....	110

3.4 Calcule des ratios :	110
3.4.1 Covariance (covMatrix).....	111
3.4.2 Sharpe ratio.....	112
3.5 Calcule du beta et du rendement espéré avec la méthode CAPM.....	113
Conclusion.....	114
Conclusion générale.....	116

Bibliographie

Annexes

Table des matières...