

République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche
Scientifique



Université Abou Bakr Belkaid – Tlemcen

Faculté de Technologie

Département de Génie Electrique et Electronique

Filière : Génie Industriel

Projet de Fin d'Etudes

Master : Génie Industriel

Intitulé :

Big DATA : Conceptions, architectures, fonctionnements et applications.

Présenté le Mercredi 14 Juin 2017 par :

- HADJARI Imane - Spécialité : Ingénierie des Systemes
- BENBACHIR Meriem - Spécialité : Management de l'ingénierie
- BOUKHATEM Fatima Zohra - Spécialité : Management de l'ingénierie

Devant le Jury :

Président :

- Mr. BENNEKROUF Mohamed - MCB - Ecole Supérieur en Sciences Appliquées - Tlemcen.

Examineurs :

- Mr. MALIKI Fouad - MAA - Ecole Supérieur en Sciences Appliquées - Tlemcen.
- Mr. HADRI Abdelkader - MAA - Université Abou-Bakr Belkaid - Tlemcen.

Encadrant :

- Mr. HASSAM Ahmed - MCB - Université Abou-Bakr Belkaid - Tlemcen.

Année Universitaire : 2016 /2017

Remerciements

*Au nom d'Allah le tout puissant, le plus grand
Merci lui revient de nous avoir donné la foi, de
nous avoir guidé vers le droit chemin, de nous avoir aidé
tout au long de nos années d'étude et nous avoir permis
d'en arriver là.*

*Nous tenons tout d'abord à adresser nos remerciements les plus sincères
à notre promoteur **M^r HASSAM Ahmed ; MCB**, pour avoir accepté de
diriger ce travail. Nous vous exprimons toute notre gratitude pour la
bonne humeur, la patience, l'encadrement, les conseils ainsi que pour
avoir patiemment corrigé ce document.*

*Nous exprimons toute notre reconnaissance aux membres de jury :
M^r. BENNKROUF ; MCB, qui nous a fait l'honneur de présider ce jury.*

***M^r MALIKI** Maître assistant de classe ; **MCA**, voulu consacrer à lire et à
juger ce travail en tant qu'examineurs.*

*Ainsi, nous adressons notre remerciement les plus chaleureux à toutes les
personnes qui ont aidé de près ou de loin par le fruit de leur connaissance
pendant toute la durée de notre travail.*

Dédicace

Je dédie ce modeste travail à ma famille avec tous mes sentiments de respect, d'amour, de gratitude et de reconnaissance pour tous les sacrifices déployés pour m'élever dignement et assurer mon éducation dans les meilleurs conditions.

A ma mère et mon père pour l'éducation qu'ils m'ont prodigué ; avec tous les moyens et au prix de toutes les sacrifices qu'ils ont consentis à mon égard, pour le sens du devoir qu'ils m'ont enseigné depuis mon Enfance.

A mes chères sœurs, Asmae et son ange Assil, Sara, Alae et la belle Doha qui je souhaite tout le bonheur du monde. Vous avez toujours été pour moi d'une aide très précieuse.

Je clos ce dédicace en dédiant ce travail au peu d'amis que j'ai eus la chance d'avoir les connaître en particulier: mes amies et binômes Meriem et Fatima, Meryem, Dihya, Assia, Soad, Samira et Brahim.

A toute la promotion du 5^{ème} Année Génie Industriel chacun par son nom, je vais ne jamais vous oublier.

Imane

Dédicac

A ma très chère mère Affable, honorable, aimable :

Tu représentes pour moi le symbole de la bonté par excellence, la source de tendresse et l'exemple du dévouement qui n'a pas cessé de m'encourager et de prier pour moi.

A mon Père Abdelkrim : Aucune dédicace ne saurait exprimer l'amour, l'estime, le dévouement et le respect que j'ai toujours eu pour vous. Rien au monde ne vaut les efforts fournis jour et nuit pour mon éducation et mon bien être.

A Mon très cher frère Younes, et mon ange Khadîdja, voue êtes la lumière de mon chemin et ma vie.

A tous les membres de ma famille, Hanna, MMima, Fadila, Naima, Samira, Zahira, Fatima, Latifa, Mon superbe oncle Fouad, petits et grands : Veuillez trouver dans ce modeste travail l'expression de mon affection.

A mes fidèles KHATER Rahma et KHater Aissa : les mots ne suffisent guère pour exprimer l'attachement, et le respect que je porte pour vous.

A notre cher et dynamique professeur assistant HASSAM Ahmed : Un remerciement particulier et sincère pour tous vos efforts fournis. Vous avez toujours été présent. Que ce travail soit un témoignage de ma gratitude et mon profond respect.

A mes chères amies BOUKHATEM Fatima, HADJARI Imane, DAHMAN Dihya et AOUS Assia : Je ne peux trouver les mots justes et sincères pour vous exprimer mon affection et mes pensées, vous êtes pour moi des sœurs et des amies sur qui je peux compter.

Un profond respect et un remerciement particulier pour tous les membres de la famille génie industriel

« MERIEM BENBACHIR »

Dédicace

✿ Je dédie cette thèse à ... ✍

A mon très cher père Mr LAKHDAR

Aucune dédicace ne saurait exprimer l'amour, l'estime, le dévouement et le respect que j'ai toujours eu pour vous.

Rien au monde ne vaut les efforts fournis jour et nuit pour mon éducation et mon bien être.

Je te dédie ce travail qu'est le fruit de tes sacrifices que tu as consentis pour mon éducation et ma formation en témoignage de mon profond amour.

Ce travail Que Dieu te préserver et t'accorder santé, longue vie et bonheur.

A ma très chère mère KHDIDJA

Aucune dédicace ne saurait être assez satisfaisante pour exprimer ce que tu mérites pour tous les sacrifices que tu n'as cessé de me donner depuis ma naissance, durant mon enfance et même à l'âge adulte.

Ta prière et ta bénédiction m'ont été d'un grand secours pour mener à bien mes études.

Je te dédie ce travail en témoignage de mon profond amour.

Que Dieu te préserver et t'accorder santé, longue vie et bonheur.

A mes très chères sœurs KHOULOUD et MOKHTARIA.

*En témoignage de l'attachement, de l'amour et de l'affection que je porte
Vous êtes toujours dans mon cœur. Je vous remercie pour votre soutien,
amour et votre affection si sincère.*

*Je vous dédie ce travail avec tous mes vœux de bonheur, de santé et de
réussite.*

A mon très cher frère AKRAM

*Mon cher petit frère présent dans tous mes moments d'examens par son
soutien moral et ses belles surprises sucrées.*

Je te souhaite un avenir plein de joie, de bonheur, de réussite et de sérénité.

Je t'exprime à travers ce travail mes sentiments de fraternité et d'amour.

A tous les membres de ma très chère famille, petits et grands

Vous avez toujours été présents pour les bons conseils.

*Votre affection et votre soutien m'ont été d'un grand secours au long de ma
vie professionnelle et personnelle.*

*Veillez trouver dans ce modeste travail ma reconnaissance pour tous vos
efforts.*

A mes chers binômes et meilleurs amies MERJEM et IMANE

*Je ne peux trouver les mots justes et sincères pour vous exprimer mon
affection et mes pensées, vous êtes pour moi des sœurs et des amies sur qui je
peux compter.*

En témoignage de l'amitié qui nous uni et des souvenirs de tous les moments que nous avons passé ensemble, je vous dédie ce travail et je vous souhaite une vie pleine de santé et de bonheur.

*A mes meilleurs amies **DIHIA, ASSIA, WAFAA et WISSEM***

Je vous dédie ce travail en témoignage de l'amitié qui nous uni et des souvenirs de tous les moments que nous avons passé ensemble.

Je vous remercie pour être présentes à tout moment, pour votre soutien moral et votre amour.

Je vous souhaite une vie pleine de santé et de bonheur.

*A ma cousine Zeineb, et mes chères amies et collègues surtout **IMANE, CHAIMAA et AHLEM...***

Je vous dédie ce travail en témoignage de notre amitié.

*A notre encadreur **Mr HASSAM Ahmed***

Vous avez bien voulu nous confier ce travail riche d'intérêt et nous guider à chaque étape de sa réalisation.

Vous nous avez toujours réservé le meilleur accueil, malgré vos obligations professionnelles.

Nous saisissons cette occasion pour vous exprimer notre profonde gratitude tout en vous témoignant notre respect.

Merci

BOUKHATEM

Fatima Zohra

Table des matières

<i>Remerciements</i>	2
LISTE DE FIGURES	10
LISTE DES TABLEAUX	12
CHAPITRE I.....	14
1. BIG DATA GENERALITE ET DIMENSIONNEMENT.....	14
1.1. INTRODUCTION	14
1.2. Définition et caractéristiques d’une « information ».....	14
1.3. La différence entre une donnée et une information :	15
1.4. Les types des données.....	15
1.5. Le stockage des données.....	15
1.6. Définition des BIG DATA.....	16
1.7. Petite Data est devenue grande ... c’est le BIG DATA	17
1.8. Historique des BIG DATA.....	18
1.9. L’intérêt du BIG DATA.....	21
1.10. Les caractéristiques de BIG DATA.....	23
1.11. BIG DATA en chiffres	24
1.12. Importance et défis techniques des BIG DATA	25
1.13. Les acteurs de BIG DATA	26
1.14. Les domaines d’application des BIG DATA.....	27
1.15. Les enjeux des BIG DATA.....	28
1.16. Limite des BIG DATA.....	30
1.17. Conclusion	31
CHAPITRE II.....	32
2. TECHNOLOGIES DES BIG DATA ET LEURS ARCHITECTURES.....	32
2.1. INTRODUCTION	32
2.2. TECHNOLOGIES DES BIG DATA.....	32
2.2.1. Hadoop	32
2.2.3. MAPREDUCE	38
2.2.4. NOSQL.....	38
2.2.5. IN-MEMORY	38
2.3. Architectures de la technologie de BIG DATA	38
2.3.1. La collecte des données	38
2.3.2. Stockage des données.....	40
2.3.3. Traitement des données	41
2.3.4. Organisation et gestion des données.....	43

2.3.5.	Analyse des données	44
2.4.	Conclusion	46
3.	OUTILS ET APPLICATION	47
3.1.	Introduction.....	47
3.2.	Algorithme de collecte et le traitement des données :.....	47
3.3.	Présentation de l'application :	48
3.4.	Intérêt d'application	48
3.5.	Interface d'application.....	48
3.6.	Résultats Analyses des données.....	54
3.7.	CONCLUSION	92
	CONCLUSION GENERALE	93
	REFERENCES BIBLIOGRAPHIE	94
	ANNEXE	96

LISTE DE FIGURES

Figure 1.1 : Les types de données	15
Figure 1.2 : C'est quoi le BIG DATA.....	16
Figure 1.3 : Petite Data est devenue grande ... c'est le BIG DATA.....	18
Figure 1.4 : Le Hadoop.....	20
Figure 1.5 : Le coup d'accélérateur des entreprises en matière de stockage des données	21
Figure 1.6 : les caractéristiques de BIG DATA	23
Figure 1.7 : les acteurs de BIG DATA	26
Figure 1.8 : Les domaines d'applications de BIG DATA	28
Figure 1.9 : les enjeux de BIG DATA	30
Figure 1.10 : les limites du BIG DATA	31
Figure 2.1 : Ecosystem Hadoop	33
Figure 2.2 : Architecture Hadoop.....	34
Figure 2.3 Le processus d'un traitement Mapreduce	36
Figure 3.1 L'interface de l'application	49
Figure 3.2 Espace d'affichage de l'information	49
Figure 3.3 espace de code source	50
Figure 3.4 espace d'afficher les liens des produits.....	50
Figure 3.5 Espace d'affichage des marques	50
Figure 3.6 espace de recherche et navigation	51
Figure 3.7 espaces de recherche des sites web	51
Figure 3.8 Exemple d'une recherche sur Samsung S8.....	51
Figure 3.9 Exemple de résultat de recherche	52
Figure 3.10 Exemple de résultat d'un code source recherchés	52
Figure 3.11 le resultat de la bouton " search"	53
Figure 3.12 le bouton copier	53
Figure 3.13 représente les informations utiles	53
Figure 3.14 Le bouton " clear" pour supprimer.....	54
Figure 3.15 : La moyenne Prix des Smartphone selon le site Amazon.com	55
Figure 3.16 La moyenne des notes des Smartphones selon le site Amazon.com.....	56
Figure 3.17 La moyenne des nombres de vue des Smartphones selon le site Amazon.com	57
Figure 3.18 La moyenne des prix des Smartphones selon le site Amazon.co.uk.....	58
Figure 3.19 La moyenne des notes des Smartphones selon le site Amazon.co.uk	59
Figure 3.20 La moyenne des nombres de vue des Smartphones selon le site Amazon.co.uk.....	60
Figure 3.21 La moyenne des prix des Smartphones selon le site Amazon.fr	61
Figure 3.22 La moyenne des notes des Smartphones selon le site Amazon.fr	62
Figure 3.23 La moyenne de nombre de vue des Smartphones selon le site Amazon.fr	63
Figure 3.24 La moyenne des prix des Smartphones selon le site Amazon.de	64
Figure 3.25 La moyenne des notes des Smartphones selon le site Amazon.de	64
Figure 3.26 La moyenne des nombres de vue des Smartphones selon le site Amazon.de	65
Figure 3.27 : La moyenne des prix des Smartphones selon le site Amazon.es	66
Figure 3.28 La moyenne des notes des Smartphones selon le site Amazon.es	67
Figure 3.29 La moyenne des nombres de vue des Smartphones selon le site Amazon.es	68
Figure 3.30 La moyenne des prix des Smartphones selon le site Amazon.ca	69
Figure 3.31 La moyenne des notes des Smartphones selon le site Amazon.ca :.....	70
Figure 3.32 La moyenne des nombres de vue Smartphones selon le site Amazon.ca	71
Figure 3.33 : La moyenne des prix des Smartphones selon la haute gamme	72

Figure 3.34: La moyenne des notes Smartphones selon la haute gamme.....	73
Figure 3.35 La moyenne des nombres de vue Smartphones selon la haute gamme	74
Figure 3.36 La moyenne des prix des Smartphones selon la moyenne gamme	76
Figure 3.37 La moyenne des notes des Smartphones selon la moyenne gamme	77
Figure 3.38 La moyenne des nombres de vue des Smartphones selon la moyenne gamme	78
Figure 3.39 La moyenne des prix des Smartphones selon l'entrée de gamme	80
Figure 3.40 La moyenne des notes des Smartphones selon l'entrée de gamme	81
Figure 3.41 La moyenne des nombres de vue des Smartphones selon l'entrée de gamme	82
Figure 3.42 La moyenne des prix des Smartphones selon la marque SAMSUNG.....	83
Figure 3.43 La moyenne des notes des Smartphones selon la marque SAMSUNG	84
Figure 3.44 La moyenne des nombres de vue des Smartphones selon la marque SAMSUNG.....	85
Figure 3.45 La moyenne des prix des Smartphones selon la marque LG.....	87
Figure 3.46 La moyenne des notes des Smartphones selon la marque LG.....	87
Figure 3.47 La moyenne des nombres de vue des Smartphones selon la marque LG.....	88
Figure 3.48 La moyenne des prix des Smartphones selon la marque APPLE.....	90
Figure 3.49 La moyenne des notes des Smartphones selon la marque APPLE.....	90
Figure 3.50 La moyenne des nombres de vue Smartphones selon la marque APPLE	91

LISTE DES TABLEAUX

Tableau 3-1 Les valeurs moyennes des prix des Smartphones selon le site Amazon.com	55
Tableau 3-2 Les valeurs moyennes des notes des Smartphones selon le site Amazon.com	55
Tableau 3-3 Les valeurs moyennes des nombres de vue des Smartphones selon le site Amazon.com	56
Tableau 3-4 Les valeurs moyennes des prix des Smartphones selon le site Amazon.co.uk	57
Tableau 3-5 Les valeurs moyennes des notes des Smartphones selon le site Amazon.co.uk	58
Tableau 3-6 Les valeurs moyennes des nombres des vues des Smartphones selon le site Amazon.com	59
Tableau 3-7 Les valeurs moyennes des prix des Smartphones selon le site Amazon.fr	60
Tableau 3-8 Les valeurs moyennes des notes des Smartphones selon le site Amazon.fr	61
Tableau 3-9 Les valeurs moyennes des nombre de vue des Smartphones selon le site Amazon.fr	62
Tableau 3-10 Les valeurs moyennes des prix des Smartphones selon le site Amazon.de	63
Tableau 3-11 Les valeurs moyennes des notes des Smartphones selon le site Amazon.de	64
Tableau 3-12 Les valeurs moyennes des nombres de vue des Smartphones selon le site Amazon.de	65
Tableau 3-13 Les valeurs moyennes des prix des Smartphones selon le site Amazon.es	66
Tableau 3-14 Les valeurs moyennes des notes des Smartphones selon le site Amazon.es	67
Tableau 3-15 Les valeurs moyennes des nombres de vue des Smartphones selon le site Amazon.de	68
Tableau 3-16 Les valeurs moyennes des prix des Smartphones selon le site Amazon.ca	69
Tableau 3-17 Les valeurs moyennes des notes des Smartphones selon le site Amazon.ca	69
Tableau 3-18 Les valeurs moyennes des nombres de vue des Smartphones selon le site Amazon.ca	70
Tableau 3-19 Les valeurs moyennes des prix des Smartphones selon la haute gamme	72
Tableau 3-20 Les valeurs moyennes des notes des Smartphones selon la haute gamme	73
Tableau 3-21 Les valeurs moyennes des notes des Smartphones selon la haute gamme	74
Tableau 3-22 Les valeurs moyennes des prix des Smartphones selon la moyenne gamme	75
Tableau 3-23 Les valeurs moyennes des notes des Smartphones selon la moyenne gamme	77
Tableau 3-24 Les valeurs moyennes de nombre de vue des Smartphones selon la moyenne gamme	78
Tableau 3-25 Les valeurs moyennes des prix des Smartphones selon l'entrée gamme	79
Tableau 3-26 Les valeurs moyennes des notes des Smartphones selon l'entrée gamme	80
Tableau 3-27 Les valeurs moyennes de nombre de vue des Smartphones selon l'entrée gamme	82
Tableau 3-28 Les valeurs moyennes des prix des Smartphones selon la marque SAMSUNG	83
Tableau 3-29 Les valeurs moyennes des notes des Smartphones selon la marque SAMSUNG	84
Tableau 3-30 : Les valeurs moyennes de nombre de vue des Smartphones selon la marque SAMSUNG	85
Tableau 3-31 Les valeurs moyennes des prix des Smartphones selon la marque LG	86
Tableau 3-32 Les valeurs moyennes des notes des Smartphones selon la marque LG	87
Tableau 3-33 Les valeurs moyennes de nombre de vue des Smartphones selon la marque LG	88
Tableau 3-34 Les valeurs moyennes des prix des Smartphones selon la marque APPLE	89
Tableau 3-35 Les valeurs moyennes des notes des Smartphones selon la marque APPLE	90
Tableau 3-36 : Les valeurs moyennes de nombre de vue des Smartphones selon la marque APPLE	91

INTRODUCTION GENERALE

Dans le monde de l'internet, on trouve différents types de données sur des produits différents : vidéos, photos, messages, annonces...etc. Ces données à leurs tours donnent des informations sur les prix, nombre d'achat, nombre de vue, nombre de lien et l'évolution de marché de ce produit en général.

Avec l'essor des réseaux Internet et Wi-Fi, des smartphones, des objets connectés et des réseaux sociaux, de plus en plus de données de formes variées sont générées. En parallèle, le développement d'outils de stockage et d'analyse – notamment le cloud computing – ainsi que de nouveaux outils de visualisation permettent la valorisation de ces données non structurées, variées et en très grande quantité : c'est un phénomène que l'on nomme le big data.

Les applications du big data sont nombreuses « pour optimiser la logistique, le marketing via la personnalisation, le développement et le suivi de produits plus proches des besoins des clients, comme le suivi de marché des Smartphone, le big data est donc source d'innovations dans tous les secteurs d'activité, mais aussi de performance et de productivité.

L'objectif de ce mémoire est de présenter les démarches de big data dans le service de marketing, en montrant que les algorithmes des big data jouent le rôle de faciliter la recherche aux clients, car les informations deviennent rapides et dans un seul clic d'un part, et d'autre part les entreprises peuvent s'avoir la tendance de vente de leurs produits grâce aux informations requise « le nombre de vue, la note donnée par ces clients » pour arriver à une bonne prédiction par la suite.

Ce mémoire est organisé comme suit :

- Le premier chapitre est consacré à des généralités et dimensionnement du Big data.
- Dans le deuxième chapitre on s'intéresse aux technologies du big data.
- Le dernier chapitre est réservé à la présentation de notre contribution proprement dite sous forme d'un exemple très simplifié de l'application du big data, où nous allons développer une application qui va collecter et filtrer les données de certains sites web du géant du commerce électronique Amazon.
- Les données seront visualisées à l'aide du tableur de Microsoft Excel, pour construire des graphes qui seront par la suite interprétés et analysés.

CHAPITRE I

1. BIG DATA GENERALITE ET DIMENSIONNEMENT

1.1. INTRODUCTION

L'évolution du Système informatique amène les entreprises à traiter de plus en plus de données issues de sources toujours plus variées.

Les prévisions de taux de croissance des volumes de données traitées dépassent les limites des technologies traditionnelles. On parle de pétaoctet « milliard d'octets » voir de zettaoctet « trillion d'octets » Quelle solution pour exploiter ces données ?

D'abord Il n'est pas aisé de définir ce qu'est une donnée, mais il est plus facile de comprendre ce qu'est une information et l'intérêt qu'il y a à la détenir. Il est important de saisir la différence avant de comprendre les enjeux de la collecte des données.

1.2. Définition et caractéristiques d'une « information »

- Si l'on consulte le dictionnaire (1) on apprend que ce concept traduit plusieurs notions qui ont conduit au terme « informatique ».
- On nous dit que l'information, dans un sens très courant, représente des renseignements sur quelqu'un ou quelque chose. On voit que cette notion est assez subjective et n'est pas susceptible de variations fréquentes ou importantes.
- L'information est aussi décrite comme l'action d'informer le public ou l'opinion. Là encore, cette notion couvre un champ qui peut être large sur un domaine particulier, donc elle sera reçue de manière individuelle ou collective avec un degré élevé de subjectivité.
- Autrement dit, l'information n'est pas comprise nécessairement de la même manière selon la personne qui la reçoit.
- Dans une acceptation plus récente et à caractère scientifique, l'information est aussi décrite comme ce qui est transmis dans un signal ou une combinaison de signaux (2): par exemple : l'information génétique.
- Il est nécessaire de recevoir des informations tout au long de la journée parce que l'Homme a toujours cherché à enrichir et renouveler sa connaissance, de façon à s'adapter à son environnement pour maîtriser son avenir, proche ou lointain.

1.3. La différence entre une donnée et une information :

A la différence d'une information, Une donnée est « Une information numérique ou alphanumérique, codée, lisible par la seule machine, en vue de son enregistrement, traitement, conservation et communication... ».

Une donnée présente plusieurs ou toutes les caractéristiques suivantes : unicité, instabilité, objectivité, accessibilité, persistance, mobilité (3)

- Unicité : une donnée a une valeur unique à un moment donné.
- Instabilité : une donnée est variable dans le temps.
- Objectivité : une donnée est vraie ou fausse.
- Accessibilité : une donnée est facilement accessible et n'a pas besoin d'être publique.
- Persistance : une donnée peut être stockée sur un support durable.
- Mobilité : une donnée peut être combinée, calculée, comparée, déduite etc.
- Ces caractéristiques rendent une donnée sans grande valeur si elle n'est pas combinée à d'autres données.

1.4. Les types des données

Types de données	Caractéristiques
Données structurées	Organisation logique Formats identiques Facilité de recherche Appréhensible par un ordinateur
Données non structurées	Entassement sans logique Formats différents Difficulté de recherche Nécessite un traitement humain

Figure 1.1 : Les types de données

1.5. Le stockage des données

L'activité humaine ne peut pas exploiter utilement des données isolées. Par extension, de très nombreuses données combinées peuvent générer de l'information très utile.

Nous avons mis en lumière la spécificité d'une donnée : elle doit être captée et éventuellement stockée en vue d'être combinée avec d'autres données. Cette collecte est une opération essentielle dans la chaîne de valeur et elle peut mobiliser des moyens considérables ou extrêmement performants.

Le besoin d'analyses de données a permis l'émergence de systèmes de bases de données, dans les années 1980, puis les systèmes d'infocentres dit « **data warehouse** ».

Ces systèmes permettent de stocker les données avec des règles d'accès très puissantes, dans un modèle où les données sont structurées et bien décrites.

Les concepteurs de ces systèmes savent à l'avance l'usage qui pourra être fait des données et ont donc imaginé une structuration permettant ces usages. Ces données ont des caractéristiques connues à l'avance, ce sont des données chiffrées communes à tout système comptable.

La collecte permet de les « ranger » de manière certaine dans des tables bien décrites et selon des règles de stockage dans ce qu'on appelle un modèle de données.

A l'inverse, les données du « BIG DATA » ne présentent pas nécessairement de caractéristiques connues à l'avance.

A l'inverse, les données du « BIG DATA » ne présentent pas nécessairement de caractéristiques connues à l'avance. Alors c'est quoi « BIG DATA » (2)

1.6. Définition des BIG DATA



Figure 1.2 : C'est quoi le BIG DATA

1.6.1. Définition (1)

A l'origine du concept de "BIG DATA" se trouve l'explosion du volume de données informatiques, conséquence de la flambée de l'usage d'Internet, au travers des réseaux sociaux, des appareils mobiles, des objets connectés, etc.

Littéralement (1): ces termes signifient méga données, grosses données ou encore données massives. Ils désignent un ensemble très volumineux de données qu'aucun outil classique de gestion de base de données ou de gestion de l'information ne peut vraiment travailler. En effet, on procède environ 2,5 trillions d'octets de données tous les jours. Ce sont les informations provenant

de partout : messages que nous nous envoyons, vidéos que nous publions, informations climatiques, signaux GPS, enregistrements transactionnels d'achats en ligne et bien d'autres encore.

Ces données sont baptisées BIG DATA ou volumes massifs de données. (4)

1.6.2. Définition (2)

Selon le CXP (5) : les BIG DATA désignent des méthodes et des technologies « pas seulement des outils » pour des environnements évolutifs « augmentation du volume de données, augmentation du nombre d'utilisateurs, augmentation de la complexité des analyses, disponibilité rapide des données » pour l'intégration, le stockage et l'analyse des données multi-structurées « structurées, semi structurées et non structurées ».

1.6.3. Définition (3)

Les BIG DATA est une démarche (ou un ensemble de technologies, d'architectures, d'outils et de procédures) qui consiste à collecter puis à traiter en temps réel des énormes volumes proviennent de sources, diverses structurées et non structurées, difficilement gérables avec des solutions classiques de stockage et de traitement. (6)

1.7. Petite Data est devenue grande ... c'est le BIG DATA

L'accroissement des données produites par les entreprises, les particuliers, les scientifiques et les acteurs publics et l'avènement du web, des réseaux, la diffusion des capteurs, des Smartphones et autres objets connectés sont couplés d'outils informatiques de nouvelles perspectives d'analyse (7).

Cette surcharge d'information ou « infobésité » est parfois difficile à gérer (6) car ici on ne parle plus de peu de giga-octets (10^9 octets) on parle maintenant plutôt de téraoctets (10^{12} octets), de péta-octets (10^{15} octets), d'exa-octets (10^{18} octets) et de zetta-octets (10^{21} octets) (7)

Donc il devient nécessaire d'utiliser de nouvelles démarches et moyens pour faciliter la gestion de ces données et on a choisi les « BIG DATA ».

L'information, une fois lisible par une machines, nourrie d'algorithmes toujours plus performants, éveille de plus en plus d'appétit, d'inventivité, fait naitre de plus en plus d'innovations.

Les ordinateurs doivent très vite démultiplier leurs capacités d'accueil et de traitements de ces données numériques. Les disques durs, qu'ils soient internes ou externes, doivent prendre du volume.

Les disques durs, qu'ils soient internes ou externes, doivent prendre du volume, de plus en plus de volume ! Idem pour les serveurs destinés à stocker, donner accès, partager ces données.

Cette croissance exponentielle des volumes de data génère la multiplication des serveurs et oblige à penser des supercalculateurs capables stocker et de traiter ces volumes de données. C'est ainsi que s'impose également le Cloud, cette solution « d'informatique en nuage », cherchant à répondre aux problématiques croissantes de stockage et d'accessibilité des données.

Parallèlement, le nouveau terme (anglo-saxon) prend place dans le langage informatique, celui de « BIG DATA ». Il qualifie donc indifféremment les méga, giga, téra, pétaoctets... (sans limite) que l'on produit et qui s'entassent en continu dans les systèmes d'information des organisations, les moteurs de recherche, les plateformes de réseaux sociaux...

Les problématiques liées à ce concept de BIG DATA enflamment les conversations, nourrissent les réflexions de l'écosystème informatique. Comment gérer ce BIG DATA aux contours sans limites, fonctionnalités évolutives, et les questions en tous genres qu'il soulève comme sa consommation d'énergie, la sécurité et la confidentialité des données, les notions d'éthique de legacy...

Autre spécificité du BIG DATA, l'hétérogénéité des données qu'il recouvre. Ce ne sont plus seulement des données triées, classées, très structurées, ce sont des données qui déferlent de partout, issues des devices mobiles, des réseaux sociaux, des comportements clients...

La gestion de ces briques de données protéiformes devient une préoccupation largement médiatisée.

(8)

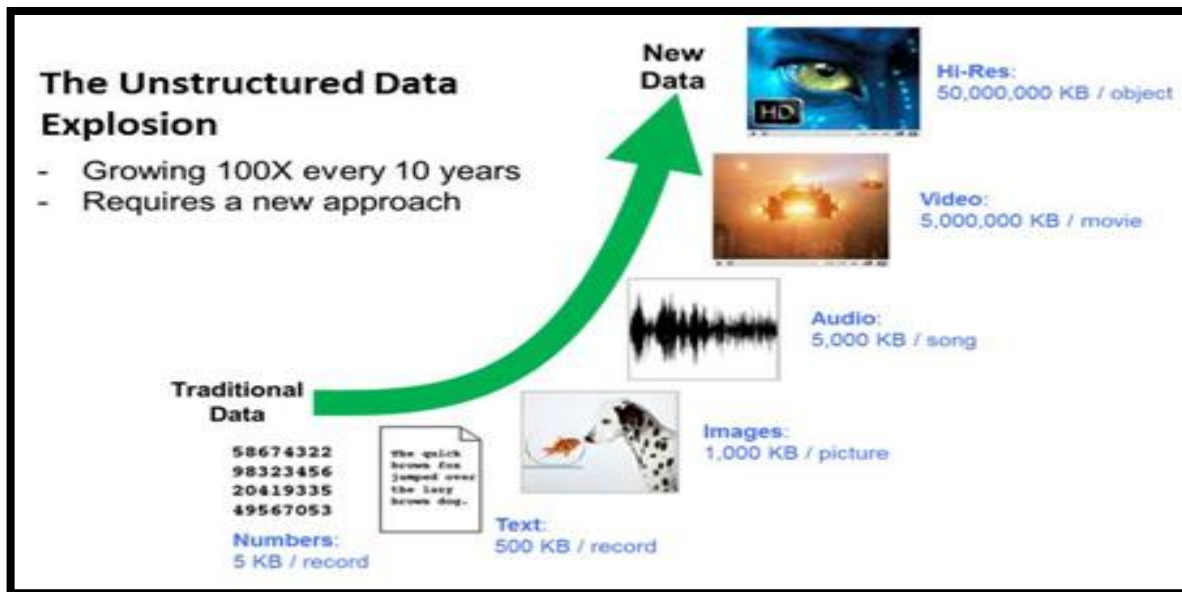


Figure 1.3 : Petite Data est devenue grande ... c'est le BIG DATA

1.8. Historique des BIG DATA

L'utilisation du terme BIG DATA pour qualifier les études de données en grande quantités, c'est à dire, nécessitant le concours de plusieurs ordinateurs actuels performants est très récente, elle n'a été popularisée que vers 2009 et les premières recherches à ce sujet datent de la fin des années 90. La notion de BIG DATA est cependant à relativiser selon les époques.

Pour faire un bref historique du BIG DATA

Il y a des milliers d'années, l'homme utilisait les données directes de la nature afin de décrire les phénomènes qu'il observait. On parle dans ce cas de science empirique.

Il y a quelques centaines d'années, les mathématiques et la physique ont été théorisées et formalisées donnant un ensemble de formules.

Depuis quelques dizaines d'années, la simulation de phénomènes complexes est rendue possible grâce à l'évolution des technologies et de l'informatique. On parle alors de calculs numériques et de simulations.

L'unification des théories, de l'expérience et des simulations nous amène actuellement à la data science. Les données sont collectées, simulées, analysées et vérifiées.

Pour une histoire plus détaillée du BIG DATA, il est très instructif de se reporter à l'article de Gil Press daté du 5 mai 2013 : "une très courte histoire du BIG DATA" (Forbes.com).

En une trentaine de dates de 1944 à 2012, il retrace la prise de conscience précoce de l'explosion des données (dans les rayons des bibliothèques universitaires dans un premier temps, puis via les NTIC), avec comme préoccupation centrale d'en quantifier la croissance tout en apportant quelques considérations critiques autour du déluge informationnel.

L'explosion des données est en effet d'abord perçue comme une menace sur la vie privée (« Assault on privacy - 1971- Arthur Miller).

L'expression « BIG DATA » fait finalement son apparition en octobre 1997 dans la bibliothèque numérique de l'ACM (association for computing machinery) :

En effet, lors d'une conférence en octobre 1997 qui porte sur les défis technologiques futurs, deux chercheurs de la NASA, Michael Cox et David Ellsworth utilisent le terme de « BIG DATA » pour la première fois pour décrire un véritable challenge et les défis des prochaines années. C'est dans un article qu'ils expliquent et définissent clairement cette notion : « *Le BIG DATA est un volume massif de données, c'est-à-dire l'ensemble des données qui ont été récoltées depuis toujours* ».

[11]

Une seconde conférence en octobre 1999, entièrement sur le thème de la visualisation des données, par l'IEEE (Institute of Electronic and Electronics Engineers) permet à un public plus grand d'accéder à cette nouvelle notion et de se poser des questions sur ses impacts.

Dans les années 2000, alors que l'exabytes (10^{18} bytes) entrent en jeu dans la quantification des données produites annuellement, la valeur du BIG DATA est mise en avant, d'abord pour les bénéfices que peuvent en tirer la recherche dans les secteurs de la physique, de la biologie ou des sciences sociales.

La montée en puissance des sites Facebook, LinkedIn, Amazon et Twitter dans les années 2000 et plus particulièrement à partir de 2005, révèle tout le potentiel des données publiques disponibles sur internet.

EN MARS 2007 : Une solution ouverte face à l'explosion du BIG DATA : le Hadoop :

Le Hadoop : a été créé en 2006 pour que de nouveaux systèmes sachent gérer l'explosion des données Web. Pouvant être téléchargé, utilisé et amélioré gratuitement, Hadoop constitue un moyen entièrement ouvert de stockage et de traitement des données qui « permet de traiter de manière parallèle et distribuée de grands volumes de données sur des serveurs standard et économiques, capables à la fois de stocker et de traiter les données, et d'évoluer à l'infini ». Source: HADOOP: Scalable, Flexible Data Storage and Analysis par Mike Olson.



Figure 1.4 : Le Hadoop

En 2012 : Le coup d'accélérateur des entreprises en matière de stockage des données

Des scientifiques du McKinsey Global Institute ont publié un article intitulé BIG DATA : The Next Frontier for Innovation, Competition, and Productivity, dans lequel ils estiment qu'« en 2009, presque tous les secteurs de l'économie des États-Unis possédaient au moins 200 téraoctets de données stockées en moyenne (soit le double de l'entrepôt de données du détaillant américain Wal-Mart en 1999) par entreprise comptant plus de 1 000 employés ». Toujours d'après cette source, les secteurs des valeurs mobilières et des placements étaient les premiers en termes de données stockées par entreprise. Les scientifiques ont calculé que 7,4 exaoctets de données d'origine avaient été sauvegardés par les entreprises, tandis que 6,8 exaoctets de données avaient été sauvegardés par les consommateurs pour la seule année 2010.



Figure 1.5 : Le coup d'accélérateur des entreprises en matière de stockage des données

Décembre 2013 SAP HANA

Les entreprises ont commencé à mettre en oeuvre de nouvelles technologies in-memory, comme SAP HANA, afin d'analyser et d'optimiser des quantités massives de données. Les entreprises se sont encore davantage appuyées sur l'utilisation des données en tant qu'actif métier pour rester compétitives, alors

Néanmoins, les sociétés sont encore peu informées. Des livres et articles sont donc publiés, afin que le monde prenne conscience du nombre de données qu'il génère et les impacts que celles-ci pourraient avoir. (9)

Que le BIG DATA représente la nouvelle technologie la plus importante à comprendre et à exploiter pour rester à la pointe d'un marché très versatile. Source : CIO, 12 BIG DATA Predictions for 2014.

2020 L'avenir du BIG DATA

La production de données augmente à un rythme effréné. Les spécialistes penchent aujourd'hui pour une hausse de 4 300 % de la génération annuelle des données d'ici à 2020. Les facteurs de croissance incluent la transition des technologies analogiques vers le numérique, ainsi que la hausse rapide de la génération des données par les entreprises comme par les particuliers. Source : (10)

1.9. L'intérêt du BIG DATA

Le sujet du BIG DATA est très ancré dans l'actualité. Celui-ci est considéré comme un levier majeur de la digitalisation de l'entreprise : le traitement des données est déjà utilisé déjà dans la finance pour vérifier par exemple la solvabilité des clients. Le marketing l'utilise pour établir des prédictions sur le comportement des consommateurs, mieux les comprendre pour personnaliser l'offre.

Le BIG DATA représente un futur outil de poids tant pour le recrutement et pour la gestion des effectifs de l'entreprise. S'il est amené à avoir une importance majeure, son utilisation soulève aussi certaines problématiques pour l'entreprise et en particulier pour les RH.

Le BIG DATA : Une source de nouvelles pratiques déjà en marche : Aujourd'hui, le volume des données à disposition est gigantesque, leur analyse et leur exploitation constitue donc un impact stratégique majeur pour les entreprises. Ces données sont aussi bien celles qui sont internes à l'entreprise, formelles (évaluation des performances, parcours du salarié dans l'entreprise...) ou informelles (messageries internes des entreprises) et enfin totalement externes à l'entreprise. (Web, réseaux sociaux ...)

Un Outil à fort potentiel : De plus en plus au centre de l'attention, le BIG DATA démultiplie les possibilités de la fonction RH. Il permet d'optimiser le sourcing en recueillant les données issues de

vos offres d'emploi, comparer ces données par rapport aux concurrents, les utiliser dans la prise de décision: ce que proposent E-quest ou encore Monster. Son utilisation a donc aussi un impact en termes de gestion des coûts.

Pour le marketing : Dans le monde d'aujourd'hui, le pouvoir est entre les mains du consommateur. Il agit selon ses envies et quand il le souhaite. Il zappe la pub, se désengage et navigue dans la direction qu'il a lui-même choisie pour effectuer ses achats. Il refuse d'être marketé, ce qui ne simplifie pas la tâche des spécialistes du marketing ! Mais grâce à l'arrivée du BIG DATA, les prises de décisions de ces spécialistes sont facilitées. En effet, l'analyse des données avec des technologies comme Hadoop, fait gagner le marketing en précision et en efficacité.

Quand on a une bonne connaissance du profil de son client, il est plus facile d'identifier le message marketing adapté. Les clients sont tous uniques et l'utilisation du BIG DATA peut devenir un partenaire utile pour décortiquer leurs besoins et attentes. (11)

La construction de la relation client : La forme la plus efficace de marketing est la construction d'une relation client solide. Le développement de cette relation se base sur des échanges réguliers et engendre la confiance du consommateur.

En utilisant le BIG DATA, les entreprises identifient ce que les acheteurs attendent de leurs produits. Bien sûr, leur but final est d'augmenter leur chiffre d'affaires. Cependant, pour être à la hauteur de leur clientèle, elles doivent répondre au mieux à leurs besoins, à travers l'envoi d'informations, l'établissement d'un lien constant, etc.

Les analyses de données et la connexion digitale sont un pont entre les consommateurs et l'entreprise. En connaissant leurs besoins, les entreprises peuvent créer des communautés autour de leurs centres d'intérêts.

Les succès économiques des grands du web commencent alors à nourrir deux idées principales :

Les données brutes accumulées ont une valeur intrinsèque de par les fameuses 4 caractéristiques en V « pour volume essentiellement mais aussi variété, véracité et vitesse de leur production donc leur fraîcheur » qu'on va les expliquer par la suite.

Ces données brutes ont une autre valeur liée aux analyses et corrélations auxquelles elles s'offrent et l'information qui en découle participe à la chaîne de valeur de l'entreprise exemple. : L'algorithme de recommandation d'Amazon.

1.10. Les caractéristiques de BIG DATA

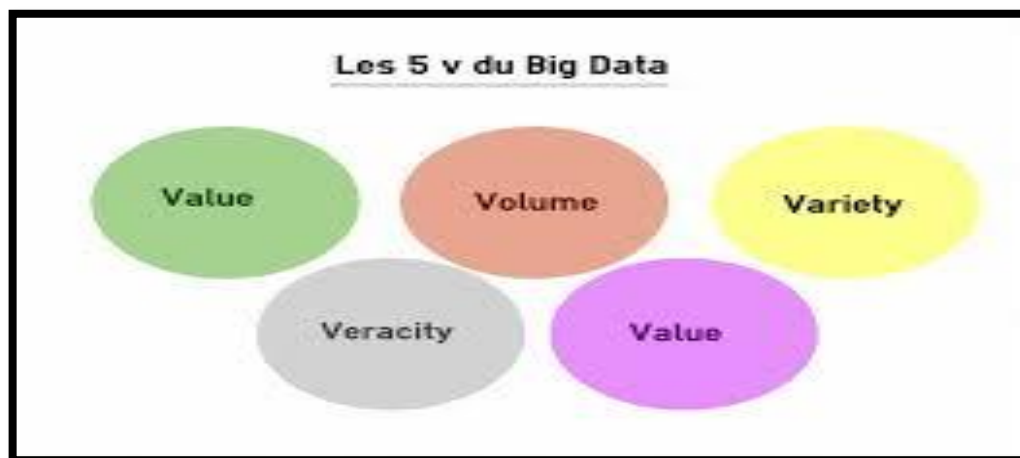


Figure 1.6 : les caractéristiques de BIG DATA

Le BIG DATA est une démarche « un ensemble de technologies, d'architectures, d'outils et de procédures » consistant à collecter puis à traiter en temps réel, ou presque, des données à la fois très nombreuses et très hétérogènes.

Depuis les débuts du BIG DATA, 3 V permettent de définir les attentes des utilisateurs :

Volume

Le volume de données créé et géré par les entreprises est en constante augmentation. Aujourd'hui, les données se comptent en téraoctets, voire en pétaoctets.

Vélocité

Une des grandes forces du BIG DATA, c'est de pouvoir utiliser les données à mesure qu'elles sont collectées. Mais pour cela, il faut une puissance de calcul et des outils d'analyse taillés sur mesure. Car l'immense majorité de données collectées par le BIG DATA ont une valeur ajoutée forte qui s'étirole rapidement avec le temps. Il faut les utiliser rapidement, ou elles n'ont aucune valeur.

Variété

Données structurées ou non structurées ; texte, données de capteurs, son, vidéo, données de géolocalisation, fichiers journaux ; les données collectées par le BIG DATA sont de nature très diverse. Il faut donc que les outils de traitement soient à même de prendre en charge cette diversité, pour donner du sens.

Deux nouveaux V

Reste que deux nouveaux "V" permettent de mieux préciser les contours de cette activité en pleine explosion.

Véracité

La notion de véracité met en avant la dimension qualitative des données nécessaire au fonctionnement des outils de BIG DATA. Car au delà de l'intégrité des données, la véracité des données permet aux directions métiers de mieux les utiliser au quotidien.

Visibilité

Enfin, la visibilité proposée par les plates-formes de BIG DATA efficaces permette aux utilisateurs de bien utiliser les données et informations publiées en effectuant des recherches pertinentes.

Au final, ces 5 V permettent de définir le BIG DATA. Un outil qui permet de gérer des données qualitatives et volumineuses, très diverses, et traitées en temps réel. Surtout, un outil qui assure à l'utilisateur une visibilité des données, via des outils de reporting, qui lui permette de prendre de bonnes décisions. (12)

1.11. BIG DATA en chiffres

Après le siècle du pétrole, nous entrons dans l'ère de la donnée. Les chiffres ci-dessous permettent de présenter la quantité de données générées jusqu'ici et la croissance dans les prochaines années.

« 12 zettaoctets de données ont été créés dans le monde en 2011 118 milliards d'emails sont envoyés chaque jour 235 téraoctets de données ont été collectés par The Library of Congress en avril 2011 30 fois plus de données seront générées d'ici 2020 Le télescope "Square kilometers away" produira plus d'1 téraoctet de données par minute en 2024 Twitter génère 7 téraoctets de données par jour Facebook génère 10 téraoctets de données par jour

Facebook traite 50 milliards de photos 30 milliards de contenus sont échangés chaque mois sur Facebook » (Saulem, Définition de BIG DATA trois V variété, volume, vitesse - Le BIG DATA.htm, 2013)

Nous pouvons bien constater que nous nageons dans un océan de donnée où le niveau de la mer augmente rapidement. (13)

1.12. Importance et défis techniques des BIG DATA

- Dans la technologie de l'information : améliorer la sécurité, diagnostiquer les anomalies et le dépannage en analysant les structures dans les logs existants.
- Au service chargé de la clientèle : en utilisant des informations des centres d'appels afin d'obtenir les modèles de clientèle et donc d'améliorer la satisfaction du client par la personnalisation des services.
- Dans l'amélioration des services et des produits : à travers l'utilisation du contenu des médias sociaux. En connaissant les préférences des clients potentiels, l'entreprise peut modifier son produit afin de répondre à une plus large gamme de personnes.
- Dans la détection de la fraude : dans les transactions en ligne pour toute type d'industrie.
- Dans l'évaluation des risques en analysant les informations provenant des transactions sur le marché financier.

Les défis du BIG DATA

- Afin de déterminer la meilleure stratégie pour une entreprise, il est essentiel que les données qu'on compte sur soient correctement analysées.
- Le laps de temps de cette analyse est important parce que certaines d'entre elles doivent être effectuées fréquemment afin de déterminer rapidement tout changement dans l'environnement des affaires.
- Nouvelles technologies ⇒ problème organisationnel.
- La nécessité des spécialistes de l'informatique : pour qu'une entreprise prend l'initiative du Big Data, elle doit soit engager des experts ou former les employés existants dans ce nouveau domaine.
- **La confidentialité et la sécurité** : Comme le Big Data englobe une grande quantité de données complexes, il est très difficile pour une entreprise de trier ces données selon des niveaux privés et d'appliquer la sécurité adéquate. En outre, la plupart des entreprises aujourd'hui font des affaires à travers des pays et continents et les différences entre les lois privées sont considérables et doivent être pris en considération lors du démarrage de l'initiative Big Data.

- Si on résume tout ça, il y a essentiellement trois types de défis autour du BIG DATA :
- Le stockage et la gestion des données massives, de l'ordre de la centaine de téraoctets ou du pétaoctet, qui dépassent les limites courantes des bases de données relationnelles classiques du point de vue du stockage et de la gestion des données.
- **La gestion des données non-structurées** (qui constituent souvent l'essentiel des données dans les scénarios BIG DATA), c'est-à-dire comment organiser du texte, des vidéos, des images, etc...
- **L'analyse de ces données massives**, à la fois pour le reporting et la modélisation prédictive avancée, mais également pour le déploiement.

1.13. Les acteurs de BIG DATA

Les grands acteurs du web tel que Google, Yahoo, Facebook, Twitter, LinkedIn, ont été les premiers à être confrontés à des volumétries de données extrêmement importantes et ont été à l'origine des premières innovations en la matière portées principalement sur deux types de technologies :

- Les bases de données (NoSql).
- Les plateformes de développement et de traitement des données.

La majorité de ces entreprises ont décidés d'ouvrir ces développements internes au monde Open Source.

Un certain nombre de ces technologies comme « hadoop » font partie de la fondation Apache et ont été intégrés aux offres de « BIG DATA » des grands acteurs tel que IBM, Oracle, Microsoft, EMC ... [16]



Figure 1.7 : les acteurs de BIG DATA

1.14. Les domaines d'application des BIG DATA

En marketing

C'est tout le secteur qui se trouve renouvelé : le BIG DATA permet en effet aux professionnels du secteur de connaître leur client « à 360° », c'est-à-dire à la fois par son parcours internet mais également par ses achats en magasin ou ses préférences affichées sur les réseaux sociaux. Anticiper les besoins de celui-ci et cibler des offres personnalisées est devenu le credo du marketing « data-driven », qui met en avant des techniques inédites : le RealTime Bidding pour l'achat d'espaces en temps réel, le Retargeting pour le ciblage personnalisé ou encore l'analyse de sentiment pour la détection de comportements sur les réseaux sociaux. Le marketing se fait de plus en plus prédictif avec le BIG DATA, et l'on assiste à une éclosion de nouveaux modèles statistiques davantage inductifs. (14)

Dans le domaine du pilotage de L'entreprise

Les usages sont également nombreux et porteurs d'innovation : en assurant une circulation immédiate et généralisée de l'information sur l'activité, le BIG DATA laisse entrevoir une optimisation complète des processus et des ressources métiers. Il réduit de facto le temps de réaction face à des erreurs ou des pannes et permet d'ajuster en permanence les équilibres offre-demande et temps-ressource. C'est une promesse importante dans des secteurs comme ceux de l'énergie ou des transports qui sont constamment portés par la logique de flux ; outre une réduction importante des coûts, le BIG DATA permet ici d'identifier au plus près les moteurs de l'activité, ce qui n'était pas possible avec les indicateurs traditionnels, soumis à des délais de latence bien plus importants. (14)

Pour la Recherche

Domaine d'application originel du BIG DATA, l'apport de celui-ci est assez évident : en autorisant le traitement de multitudes de données, le BIG DATA permet à la science de réaliser des avancées importantes, lorsqu'il s'agit d'explorer l'infiniment petit (ex : exploration géologique), de croiser des données complexes (ex : imagerie) ou d'effectuer des simulations (ex : domaine spatial). C'est d'ailleurs en génétique que le BIG DATA a fait ses premières armes car ce secteur réclamait une approche à la fois quantitative et qualitative avancée. [(14)

Dans le domaine de l'Information

Tout simplement, qu'elle soit corporate ou grand public, le traitement des BIG DATA a profondément modifié la donne : pour une requête donnée, il est désormais possible d'accéder à un croisement d'informations très disparates, issues de sources jusque-là négligées. L'instantanéité des réseaux sociaux est à ce titre une innovation de taille : l'analyse des tweets est devenue une source

de renseignements courante pour comprendre les comportements ou les goûts de populations segmentées. De plus, au-delà de la compréhension de phénomènes, la data s'avère un outil de communication efficace pour faire passer des messages factuels aux publics : c'est tout le rôle du data journalisme, qui vise à mettre en lumière des événements complexes au travers de la représentation de données multiples. (14)

On le voit, le BIG DATA a donc eu un effet profondément novateur sur de nombreuses activités : il a ouvert de nouveaux champs d'exploration, avec à la clé des informations et fonctions inédites. Mais, au-delà, le BIG DATA a également permis de créer intégralement de nouveaux services, de nouvelles filières fondées sur des données inexplorées jusqu'alors. (14)

Les données mobiles

Par exemple, dont le traitement temps réel est devenu un enjeu stratégique pour de nombreux secteurs, sont un important gisement d'innovation : en effet, la possibilité de géo localiser des utilisateurs a ouvert de nombreuses perspectives aux annonceurs qui y voient une opportunité unique de placer leurs publicités à proximité des lieux d'achat. De la même façon, les opérateurs télécom travaillent avec les autorités publiques pour comprendre les flux de populations grâce à ces données et, dans le secteur du tourisme, des services inédits de localisation sont mis en place. (14)

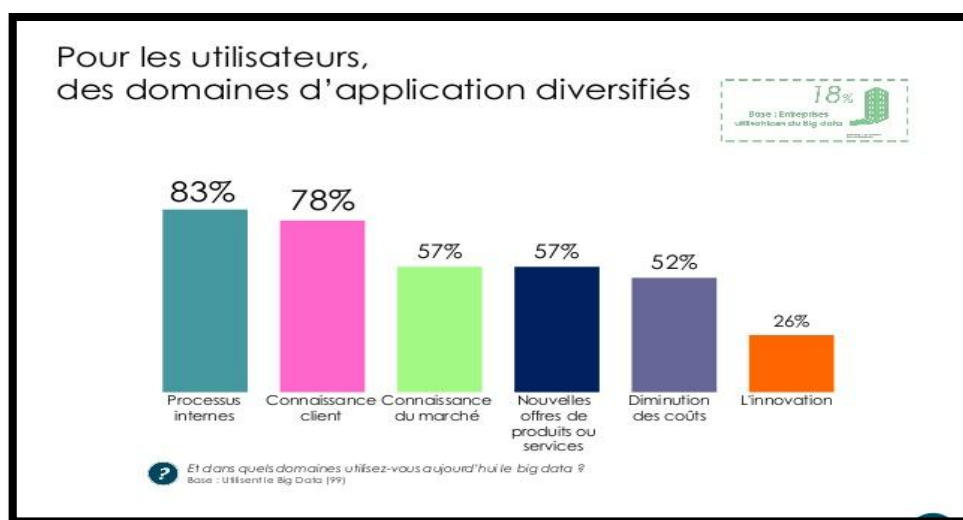


Figure 1.8 : Les domaines d'applications de BIG DATA

1.15. Les enjeux des BIG DATA

Le BIG DATA apparaît comme le challenge technologique des années 2010-2020. Dépassant les domaines techniques et informatiques, le BIG DATA suscite un vif intérêt auprès des politiciens, des scientifiques et des entreprises. Les enjeux du BIG DATA touchent plusieurs secteurs d'activités. (13)

. Enjeux techniques

Les enjeux techniques s'articulent autour de l'intégration, le stockage, l'analyse, l'archivage, l'organisation et la protection des données. (13)

. Enjeux économiques

D'après le cabinet de conseil dans le marketing IDC, « le marché du BIG DATA représentera 24 milliards de dollars e 2016, avec une part de stockage estimée à 1/3 de ce montant » (*Saulem, Enjeux économiques du BIG DATA à l'échelle mondiale - Le BIG DATA.htm, 2013*). Il va sans dire que la « donnée » est le nouvel or noir du siècle présent, les spécialistes s'accordent déjà sur le fait que le BIG DATA sera l'arme économique de demain pour les entreprises et se présentera comme un levier qui fera la différence.

Les entreprises collectent de plus en plus d'information en relation avec leurs activités (production, stockage, logistique, ventes, clients, fournisseurs, partenaires, etc), toutes ces informations peuvent être stockées et exploitées pour stimuler leur croissance. Les BIG DATA permettent :

- D'améliorer les stratégies marketing et commerciale

- D'améliorer et entretenir la relation client
- De fidéliser la clientèle
- De gagner de nouvelles parts de marché
- De réduire les coûts logistiques
- De favoriser la veille concurrentielle

Le client est un acteur majeur dans ce contexte. Jusqu'à présent, la vente consistait à se demander « J'ai un produit, à qui vais-je pouvoir le vendre ? ». A l'ère du BIG DATA, nous devons changer le paradigme pour dire « J'ai un client, de quoi a-t-il besoin aujourd'hui ? ». En connaissant mieux son public, à travers ses achats, ses activités sur Internet, son environnement, les commerçants peuvent améliorer l'expérience-client, exploiter la recommandation, imaginer le marketing prédictif (le marketing prédictif regroupe les techniques de traitement et de modélisation des comportements clients qui permettent d'anticiper leurs actions futures à partir du comportement présent). (13)

. Enjeux juridiques

Le principal enjeu juridique dans un contexte où les utilisateurs sont souvent des « produits », reste la protection de la vie privée (13)



Figure 1.9 : les enjeux de BIG DATA

1.16. Limite des BIG DATA

Si le terrain de jeu du BIG DATA est loin d'être restreint, il n'est pas sans limites. Elles tiennent, en premier lieu, à la nature des données et aux traitements envisagés, et lorsqu'il est question des données personnelles, la vigilance est nécessaire. Dans certains pays, le traitement des données à caractère personnel est régi par des dispositions particulières, ce qui n'est pas le cas dans la majorité des pays.

Nous sommes arrivés à un point où la protection des données personnelles, portée à la défense des libertés fondamentales de l'individu, est en train de devenir un argument économique. L'enjeu étant dorénavant, d'élaborer le cadre normatif le plus attractif pour le développement de l'économie numérique et des échanges de données, ceci nécessite d'être vigilant dans un contexte de forte concurrence entre les puissances économiques.

L'autre préoccupation provient de la sécurité : une faille minuscule peut menacer des quantités de données considérables. Si les utilisateurs perdent confiance dans l'utilisation de leurs informations, c'est donc tout l'édifice du BIG DATA qui risque de s'écrouler. Pour éviter cela, par exemple en Europe, la commission européenne a présenté, en début 2012, un règlement qui vise à protéger davantage les utilisateurs. Ce texte devrait être voté en 2014 pour une application en 2016, il obligera les entreprises à demander le consentement explicite de l'utilisateur avant de collecter ses données. (13)

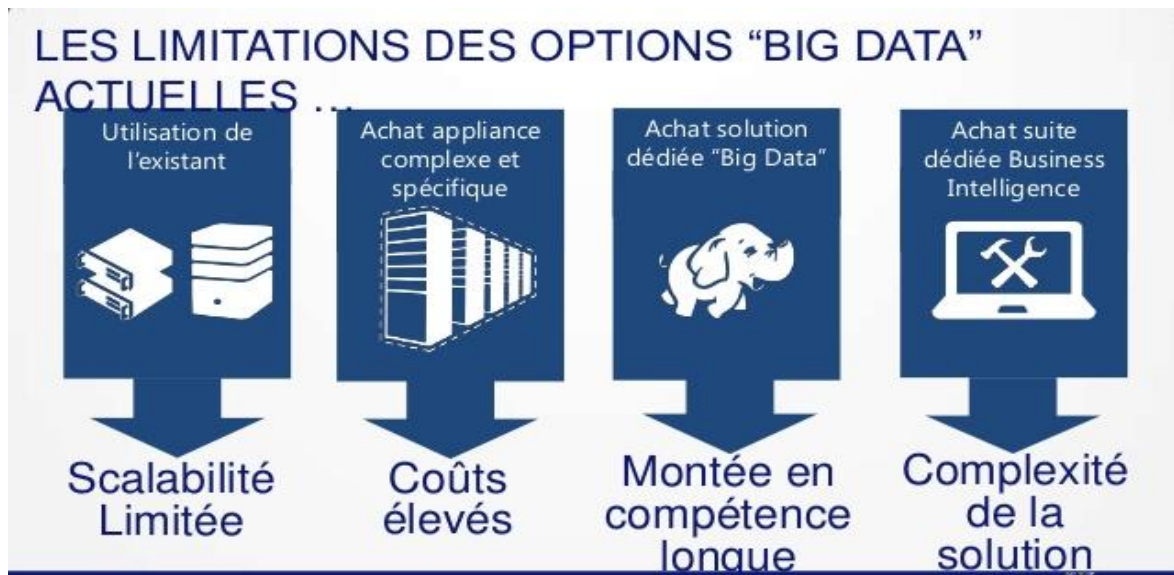


Figure 1.10 : les limites du BIG DATA

1.17. Conclusion

Le BIG DATA est né d'une réponse à un problème de temps de traitement des données, les massives quantités de données sont compressées, pour qu'elles prennent moins d'espace de stockage : par le système binaire et le codage Huffman. Ces données sont mises en relation et des calculs sont effectués grâce aux ordinateurs, (l'ENIAC et l'ABC se disputent la première place) et aux logiciels & programmes.

Les données collectées dans le cadre d'une étude BIG DATA peuvent avoir des origines et des formes très différentes, elles sont ensuite traitées à travers des algorithmes et des structures adaptées dans le but de produire une information. Aux fondements des algorithmes des solutions de BIG DATA se trouvent des lois de statistiques et de probabilités, telles que la loi normale où les moindres carrés.

CHAPITRE II

2. TECHNOLOGIES DES BIG DATA ET LEURS ARCHITECTURES

2.1. INTRODUCTION

Face à l'explosion du volume d'informations, le BIG DATA vise à proposer une alternative aux solutions traditionnelles de bases de données et d'analyse (serveur SQL, plateforme de Business Intelligence...). Confrontés très tôt à des problématiques de très gros volumes, les géants du web, au premier rang desquels Yahoo (mais aussi Google et Facebook), ont été les premiers à déployer ce type de technologies. Selon le Gartner, le BIG DATA (en français mégadonnées ou "Grandes données") regroupe une famille d'outils qui répondent à un triple problématique : un Volume de données important à traiter, une grande Variété d'informations (en provenance de plusieurs sources, non-structurées, structurées, Opendata...), et un certain niveau de Vitesse à atteindre - c'est-à-dire de fréquence de création, collecte, traitement/analyse et partage de ces données (8)

2.2. TECHNOLOGIES DES BIG DATA

L'univers technologique du BIG DATA s'appuie sur des outils bien identifiés qui constituent la base innovante de ce mode de traitement. A eux seuls, ces outils résument le vocabulaire technologique du BIG DATA et en constituent la référence. Mais de quoi parle-t-on exactement lorsque l'on évoque ces innovations ? Voici un petit lexique simplifié qui permettra même aux moins technophiles de comprendre les fondements techniques du BIG DATA. (3)

2.2.1. Hadoop

Emblème par excellence du BIG DATA, Hadoop est une architecture spécifique de bases de données, permettant de traiter en grand nombre tous types de données « compris les données non structurées ».

On dit qu'elle est organisée sur un mode non-relationnel, c'est-à-dire que les données ne sont pas triées en fonction de leurs relations entre elles. Elle constitue en quelque sorte la librairie à partir de laquelle on pourra effectuer des calculs parallèles « via MapReduce », sachant que ces données sont distribuées c'est-à-dire qu'Hadoop les « récupère » (3)

2.2.2. Ecosystème Hadoop

Pour certains, l'avenir appartiendra à ceux qui seront capable d'analyser les vastes volumes de données qu'ils ont collectés.

En écologie, un écosystème est l'ensemble formé par une association ou communauté d'êtres vivants et son environnement biologique, géologique, édaphique, hydrologique, climatique, etc. l'écosystème Hadoop est l'ensemble des projets Apache ou non, liés à Hadoop et qui sont appelés à se cohabiter.

Hadoop désigne un framework Java libre ou un environnement d'exécution distribuée, performant et scalable, dont la vocation est de traiter des volumes des données considérables. Il est le socle d'un vaste écosystème constitué d'autres projets spécialisés dans un domaine particulier parmi lesquels on compte les entrepôts de données, le suivi applicatif (monitorring) ou la persistance de données.

(9)

Le schéma ci-après présente les différents éléments de l'écosystème Hadoop en fonction du type d'opération.

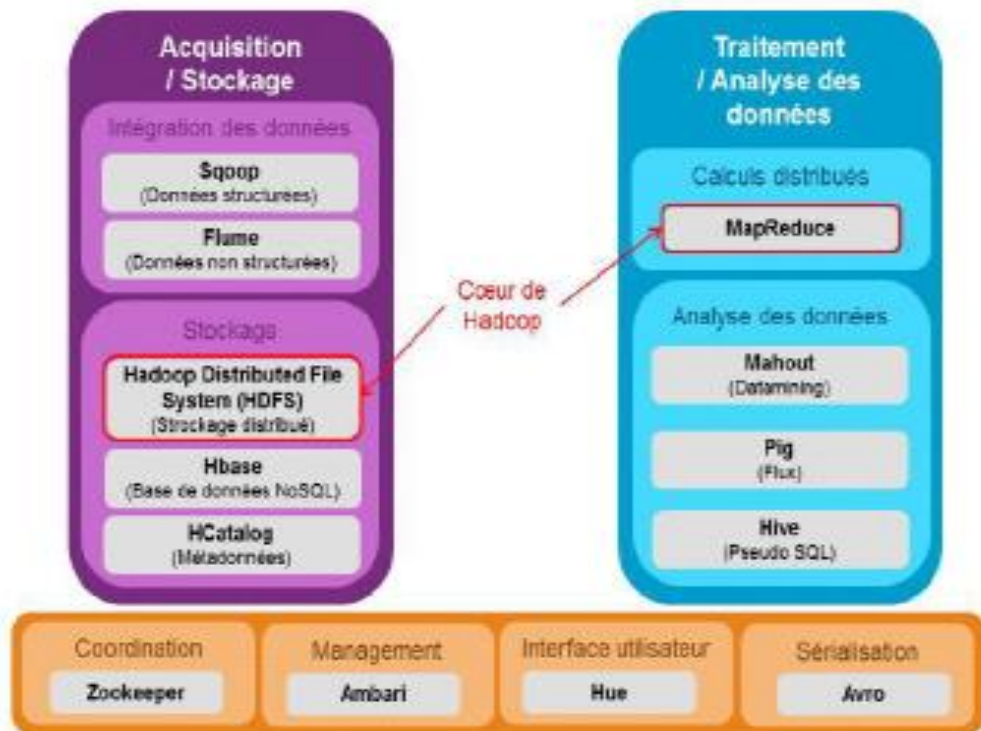


Figure 2.1 : Ecosystem Hadoop

2.2.2.1. Hadoop kernel

Hadoop Kernel est le coeur de l'écosystème Hadoop. Ce framework est actuellement le plus utilisé pour faire du BIG DATA.

Hadoop est écrit en Java et a été créé par Doug Cutting et Michael Cafarella en 2005, (*Le framework Apache Hadoop*).

Le noyau est composé de 2 composants. Hadoop présente l'avantage d'être issu de la communauté open source, de ce fait un porte un message exprimant une opportunité économique. En revanche, il affiche une complexité qui est loin de rendre accessible au commun des DSI.

Nous continuerons par une présentation de quelques grands concepts d'Hadoop.

2.2.2.2. Architecture Hadoop

- Le schéma ci-dessous présente l'architecture distribuée dans le contexte Hadoop.

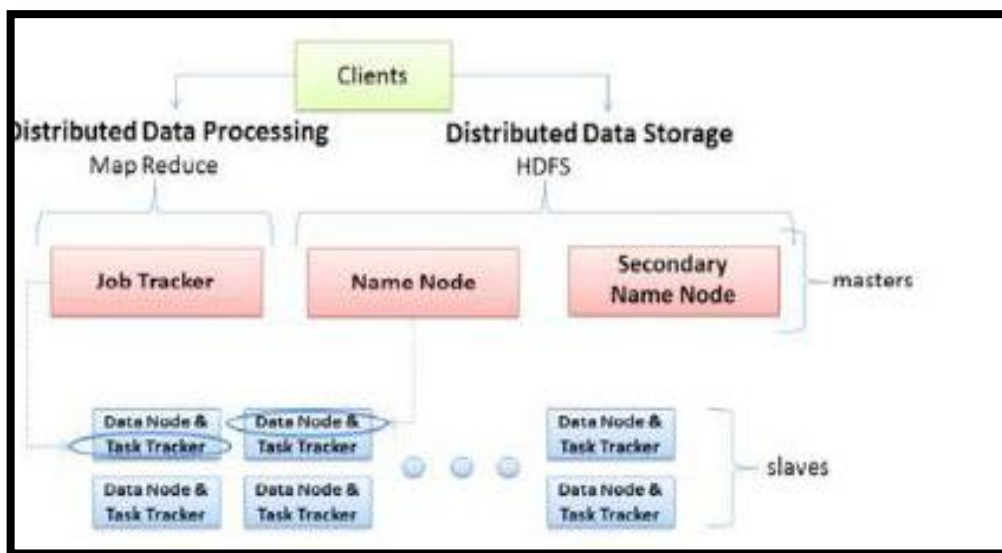


Figure 2.2 : Architecture Hadoop

Il est primordial de savoir qu'une architecture Hadoop est basée sur le principe maître/esclave, représentant les deux principaux rôles des machines.

Les sous rôles relatifs au système de fichiers et à l'exécution des tâches distribuées sont associés à chaque machine de l'architecture.

Les ordinateurs maîtres ont trois principaux rôles qui leur sont associées :

- JobTracker** : c'est le rôle qui permet à l'ordinateur maître de lancer des tâches distribuées, en coordonnant les esclaves. Il planifie les exécutions, gère l'état des machines esclaves et agrège les résultats des calculs.
- NameNode** : ce rôle assure la répartition des données sur les machines esclaves et la gestion de l'espace de nom du cluster. La machine qui joue ce rôle contient des métadonnées qui lui permettent de savoir sur quelle machine chaque fichier est hébergé.

- c. **SecondaryNameNode** : ce rôle intervient pour la redondance du NameNode. Normalement, il doit être assuré par une autre machine physique autre que le NameNode car il permet en cas de panne de ce dernier, d'assurer la continuité de fonctionnement du cluster.

Deux rôles sont associés aux machines esclaves :

- a. **TaskTracker** : ce rôle permet à un esclave d'exécuter une tâche MapReduce sur les données qu'elle héberge. Le TaskTracker est piloté par JobTracker d'un ordinateur maître qui lui envoie la tâche à exécuter.
- b. **DataNode** : dans le cluster, c'est une machine qui héberge une partie des données. Les noeuds de données sont généralement répliqués dans le cadre d'une architecture Hadoop dans l'optique d'assurer la haute disponibilité des données.

Lorsqu'un client veut accéder aux données ou exécuter une tâche distribuée, il fait appel à la machine maître qui joue le rôle de JobTracker et de Namenode.

Maintenant que nous avons vu globalement l'articulation d'une architecture Hadoop, nous allons voir deux principaux concepts inhérents aux différents rôles que nous avons présentés.

A. HDFS

HDFS est le système de fichiers Java, permettant de gérer le stockage des données sur des machines d'une architecture Hadoop. Il s'appuie sur le système de fichier natif de l'OS (unix) pour présenter un système de stockage unifié reposant sur un ensemble de disques et de systèmes de fichiers.

La consistance des données réside sur la redondance ; une donnée est stockée sur au moins n volumes différents.

Le NameNode rendrait le cluster inaccessible s'il venait à tomber en panne, il représente le SPOF (maillon faible) du cluster Hadoop. Actuellement, la version 2.0 introduit le failover automatisé (capacité d'un équipement à basculer automatiquement vers un équipement alternatif, en cas de panne). Bien qu'il y ait plusieurs NameNodes, la promotion d'un NameNode se fait manuellement sur la version 1.0.

Dans un cluster les données sont découpées et distribuées en blocks selon la taille unitaire de stockage (généralement 64 ou 128 Mo) et le facteur de répllication (nombre de copie d'une donnée, qui est de 3 par défaut).

Un principe important de HDFS est que les fichiers sont de type « write-one » ; ceci est lié au fait que lors des opérations analytiques, la lecture des données est beaucoup plus utilisée que l'écriture.

B. MapReduce

Mapreduce qui est le deuxième composant du noyau Hadoop permet d'effectuer des traitements distribués sur les noeuds du cluster.

Il décompose un job « unité de traitement mettant en œuvre un jeu de données en entrée, un programme MapReduce (packagé dans un JAR (Java Archive : fichier d'archive, utilisé pour distribuer un ensemble de classes Java)) et des éléments de configuration » en un ensemble de tâche plus petites qui vont produire chacune un sous ensemble du résultat final ; ce au moyen de la fonction **Map**.

L'ensemble des résultats intermédiaires est traité (par agrégation, filtrage), ce au moyen de la fonction **Reduce**.

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

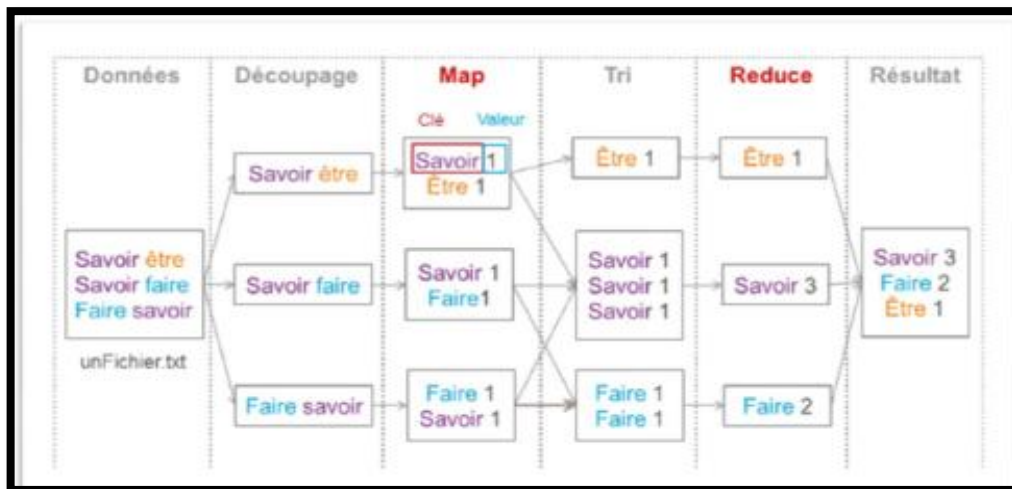


Figure 2.3 Le processus d'un traitement Mapreduce

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.

B.1. Modes d'utilisation

Hadoop peut être sous trois modes différents :

- **§ Mode Standalone** : ce mode a pour objectif de tester le fonctionnement d'une tâche MapReduce. Ici, la tâche est exécutée sur le poste client dans la seule machine virtuelle Java

(JVM), pas besoin d'une configuration particulière car c'est la mode de fonctionnement de base de Hadoop.

- **Mode Pseudo distributed** : ce mode permettra de tester l'exécution d'une tâche MapReduce sur une seule machine tout en simulant le fonctionnement d'un cluster Hadoop. Le job est exécuté sur la machine et les opérations de stockage et de traitement du job seront gérées par des processus Java différents. L'objectif de ce mode est de tester le bon fonctionnement d'un job sans besoin de mobiliser toutes les ressources du cluster.
- **Mode Fully distributed** : c'est le mode réel d'exécution d'Hadoop. Il permet de mobiliser le système de fichier distribué et les jobs MapReduce sur un ensemble de machines ; ceci nécessite de disposer de plusieurs postes pour héberger les données et exécuter les tâches.

Dans la suite, d'autres composants qui entrent dans l'écosystème Hadoop sont présentés. (9)

B.2. Les composants de l'écosystème Hadoop

La version courante du projet Apache Hadoop inclut ces modules :

- **Hadoop Common** :

Les utilitaires communs qui supportent les autres modules d'Hadoop. (4)

- **Hadoop Distributed File System (HDFS):**

Un système de fichiers distribués qui fournit un accès haut-débit aux données de l'application. (4)

- **Hadoop YARN** :

Un framework pour la planification des tâches et la gestion des ressources du cluster. (4)

- **Hadoop MapReduce** :

Un système basé sur YARN pour le traitement parallèle des gros volumes de données. Cependant, l'écosystème Hadoop ne contient pas uniquement Hadoop, mais beaucoup d'autres projets Apache tels que : **Pig, Hive, Hbase, Sqoop, Flume, Zookeeper** et bien d'autres. (4)

- 1) **Pig** : une plate-forme pour analyser des ensembles de gros volumes de données. Cela consiste en un langage haut-niveau pour l'expression de programmes d'analyse de données, couplé à une infrastructure pour évaluer ces programmes. (4)
- 2) **Hive**: un système d'entrepôt de données pour Hadoop qui offre un langage de requête de type SQL pour faciliter les agrégations, le requêtage ad-hoc et l'analyse de gros volumes de données stockés dans des systèmes de fichiers compatibles Hadoop. (4)
- 3) **Hbase**: un stockage de données distribué et scalable dédié au BIG DATA avec un accès direct et une lecture/écriture temps réel. (4)
- 4) **Sqoop**: un outil conçu pour transférer efficacement une masse de données entre Apache Hadoop et un stockage de données structuré tel que les bases de données relationnelles. (4)

- 5) **Flume**: un service distribué, fiable et disponible pour collecter efficacement, agréger et déplacer une grande quantité de logs. (4)
- 6) **Zookeeper**: un service centralisé pour maintenir les configurations, la nomenclature, pour fournir une synchronisation distribuée et des services groupés (4)

2.2.3. MAPREDUCE

Couplé à Hadoop (qui est en fait son implémentation principale), MapReduce est le mode de calcul permettant de traiter les BIG DATA. Il présente une fonction Map (distribution des données sur plusieurs clusters parallèles où les calculs intermédiaires seront effectués) et une fonction Reduce (les résultats des calculs intermédiaires distribués sont recentralisés en vue du calcul final). MapReduce est issu de la recherche Google et a été développé au même moment qu'Hadoop ce qui explique l'association traditionnelle que l'on fait des deux outils. (3)

2.2.4. NOSQL

On parle de NoSQL lorsque des requêtes sont effectuées dans un langage de programmation qui n'est pas limité au SQL. Ce dernier est en effet le langage le plus utilisé pour des traitements de données mais ses normes de définition, parfois trop strictes, ne sont pas adaptées aux données massives et non-structurées. Du coup, certaines requêtes sont effectuées indépendamment de ces normes et donc de ce langage propre, ce qui fait dire à certains que l'on bascule sur du NoSQL (littéralement : « au-delà du SQL »). Pour autant il est quasiment impossible de définir le NoSQL selon un corpus de règles puisqu'il représente précisément l'affranchissement par rapport à un ordre établi. (3)

2.2.5. IN-MEMORY

On parle de traitement in-memory pour évoquer les traitements qui sont effectués dans la mémoire vive de l'équipement informatique, plutôt que sur des serveurs externes. L'avantage du traitement in-memory est celui de la vitesse puisque les données sont immédiatement accessibles. En revanche, ces données ne sont pas stockées sur le long terme, ce qui peut poser des problèmes d'historisation. (3)

2.3. Architectures de la technologie de BIG DATA

2.3.1. La collecte des données

La collecte des données : l'expression est paradoxale car si l'on prend la peine de collecter quelque chose, c'est justement parce que ce n'est pas donné !

En amont de la démarche statistique se trouvent les données (*data*) et en aval, la connaissance, puis l'action. Encore qu'il existe un contrecourant dans ce long fleuve impétueux puisque c'est la bonne connaissance d'un sujet qui permet de définir les caractères à observer.

S'il est un volume qui connaît un accroissement exponentiel depuis des décennies, c'est bien celui des données et le ralentissement n'est pas annoncé pour les semaines à venir... Aujourd'hui, on parle de *BIG DATA* pour définir des bases si colossales que les outils informatiques classiques doivent être repensés. (15)

- **Fiabilité**

En amont de la collecte, il faut pouvoir estimer si le niveau d'exactitude des données est suffisant. Inutile de se lancer dans une étude pour laquelle les observations sont réputées non fiables.

En effet, qu'elles soient rares ou surabondantes, elles sont le matériau qui sera traité pour être transformé en information. Et l'information, c'est le nerf du business et de la recherche... Aussi ne faut-il pas oublier que la fiabilité des données est sans doute le critère n°1 de la qualité d'une étude. On a coutume de résumer cette évidence par *Garbage In Garbage Out* (GIGO). Malheureusement, l'exactitude des réponses n'est pas la motivation principale de l'enquêteur mal payé ou de l'enquêteur qui n'a que faire des résultats de l'étude. Même une bonne traçabilité ne garantit pas la justesse des données.

Comme l'exprime B. Delmas, « *les sources d'erreurs [...] sont en effet nombreuses et rien ne prouve que ces erreurs s'annulent comme on le suppose souvent pour se débarrasser à bon compte du problème.* » (*Statistique descriptive*, Nathan 1996). [13]

- **Traitement**

Une fois collectées, les données subissent quelques traitements avant même d'être exploitées. Elles sont « nettoyées », ce nettoyage consistant à éliminer ou imputer les valeurs manquantes ou aberrantes.

Puis elles changent de statut. Elles deviennent les modalités de variables dans une base d'analyse (sous forme de fichier CVS ou Excel, table SAS, etc.). Concrètement, chaque ligne d'un fichier informatique est affectée à une unité statistique et chaque colonne indique les valeurs ou les modalités prises par une variable en particulier.

Les données sont habituellement qualifiées par leur nature (continues, discrètes, qualitatives, textuelles) et par le type de collecte utilisé. Voyons ces types de plus près. [13]

- **Les données primaires**

Il s'agit de celles qui sont spécialement collectées pour répondre à une problématique, c'est-à-dire pour réaliser une étude statistique *ad hoc*. Donc, pas de *BIG DATA* par ici ! La collecte a un coût

important (coûts variables très élevés par rapport au coût fixe, contrairement aux données secondaires internes évoquées plus bas). Le commanditaire délègue en principe cette activité à un prestataire sous-traitant qui utilise des moyens particuliers tels qu'enquêtes à base de questionnaires ou méthodes expérimentales.

Les questionnaires sont les outils habituels des études de marché voire d'audits sociaux. Avant de les mettre en oeuvre, il convient de déterminer la taille de l'échantillon, son mode de sélection (aléatoire, par quotas, stratifié, par panel...) et la formulation des questions qui se traduira par tel type de variable (échelles non comparatives...).

Une expérimentation invoquée est une recherche d'explication de phénomènes sur lesquels l'observateur n'a pas de prise. Exemple : corrélation constatée entre la météo et la vente de certains produits alimentaires. Au contraire, une expérimentation provoquée reconstitue des situations en laboratoire (marché-test simulé, par exemple). NB : cette typologie est celle de Madeleine Grawitz, « *Méthodes des sciences sociales* », Dalloz 1981, p. 413.

Une suite d'expériences provoquées avec modifications successives des variables susceptibles d'expliquer un résultat est un plan d'expérience. (5)

2.3.2. Stockage des données

Les BIG DATA sont généralement stockées et organisées dans des systèmes de fichiers distribués. S'il existe différentes approches et stratégies de mise en oeuvre, l'information est stockée sur plusieurs disques durs (parfois plusieurs milliers) et ordinateurs classiques. Un index « map » permet de savoir à quel endroit (sur quel ordinateur/disque) se situe une information particulière. En fait, dans un souci de sécurité et de robustesse, chaque information est généralement stockée plusieurs fois, par exemple sous forme de triplets. Supposons par exemple que vous ayez collecté des transactions individuelles pour une enseigne de la grande distribution. Le détail de chaque transaction peut alors être stocké sous forme de triplets sur différents serveurs et disques durs, avec une table-maître ou « map » indiquant précisément l'endroit où est stocké le détail de chaque transaction. Grâce à l'utilisation de matériel informatique classique et de logiciels libres pour la gestion de ce système de fichiers distribué (de type Hadoop), il est possible de créer assez facilement des entrepôts de données fiables de l'ordre du pétaoctet, et ces systèmes de stockage deviennent de plus en plus courants. Vous trouverez davantage d'informations techniques sur la mise en place de ces systèmes en recherchant sur Internet des informations relatives aux systèmes de fichiers distribués ou à Hadoop. (15)

2.3.3. Traitement des données

Nous venons de voir que si la distribution physique des données est assurée par la plateforme de stockage, et donc transparente pour l'application, ce n'est pas le cas de sa structure.

Lorsque la donnée stockée est exploitable directement (par exemple elle est déjà assez structurée, et le traitement consiste en un calcul de statistiques simples comme un comptage ou une moyenne), le code du traitement peut directement opérer dessus. Selon les plateformes, l'implémentation devra tenir compte ou non d'une exécution parallèle du traitement par le framework.

C'est le cas des plateformes s'appuyant sur un framework MapReduce, par exemple, qui imposent une structure particulière aux programmes qui l'utilisent.

Cependant, dans de nombreux cas, la donnée issue du système de stockage n'est pas exploitable en l'état par le coeur du traitement. Il faut d'abord la faire passer dans un pipeline d'interprétation, pour lui donner une forme consommable par l'algorithme.

Ce pipeline peut bien sûr être implémenté par le programme lui-même, ou faire appel à des bibliothèques pré-existantes que le programme se contentera de coordonner.

Avant de citer quelques unités de traitement possibles pour un tel pipeline, notons que ce dernier peut s'exécuter de différentes manières : en tâche de fond sur de gros volumes (mode batch), sur des volumes faibles ou avec des temps de réponse garantis faibles (mode interactif), ou au fur et à mesure de l'arrivée des données unitaires dans le système (mode événementiel). (6)

LES ÉTAPES DU PIPELINE

- Une première étape quasiment obligatoire, avec la variété des sources de données à laquelle on est confrontée, est la **préparation des données**. Il faut en général les nettoyer, les enrichir (avec des données de référence ou des valeurs par défaut pour les portions manquantes), afin qu'elles atteignent une qualité suffisante pour être réellement traitées.
- **L'indexation** des données est une étape nécessaire lorsqu'un besoin de recherche dans les données est identifié. Il n'est en effet pas question de faire une recherche brutale sur des volumes aussi importants, pour retrouver une donnée précise. Les solutions de type NoSQL, qui stockent des objets, offrent des index de type clef primaire. On observe cependant que les besoins plus élaborés (index secondaires, recherches plein-texte) sont en général confiés à des moteurs de recherche tiers (type Solr, Lucene) intégrés avec le stockage, et qui opèrent directement sur leurs données.
- Pour les **données non ou faiblement structurées** (texte, images, vidéo...), il est nécessaire de faire une extraction de métadonnées, ou des traitements spécifiques au type de document. À titre d'exemple, pour du texte on trouvera une extraction de mots, des statistiques sur les

paragraphes, voire du traitement élaboré de langage naturel (NLP --*Natural Language Processing*) et une interprétation sémantique. Pour des images, on peut envisager d'en calculer la taille, de faire de l'OCR, de la détection de visages ou de logos... Cette première étape du pipeline permet de doter la donnée brute d'une structure exploitable par les étapes suivantes. Elle peut aussi alimenter un moteur de recherche.

- Dans de nombreux cas, on est amené à appliquer des méthodes statistiques plus poussées que de simples sommes, afin d'extraire plus d'information ou de construire des modèles statistiques. C'est le domaine du **machine learning**, qui n'est ni plus ni moins que l'incarnation moderne et (partiellement) automatisée des techniques traditionnelles de data mining. Il sert trois problématiques : décrire, expliquer et prédire. Les algorithmes font appel à des combinaisons de techniques statistiques, d'intelligence artificielle, et bien sûr au flair de l'analyste de données (le fameux data scientist) qui va choisir les meilleures stratégies en fonction du contexte.
- On peut voir les **traitements des graphes** comme un cas particulier du machine learning, mais il a aussi ses caractéristiques propres. De nombreux problèmes peuvent être formulés avec une représentation en graphes. C'est une représentation compacte d'objets et de leurs relations, qui offre l'avantage d'être outillée par une théorie mature. La panoplie offerte par cette théorie est vaste, et couvre un spectre d'analyse très large : analyse descriptive globale sur les propriétés du graphe, calcul d'influence des sommets au sein du réseau, calcul de chemins, etc. Les défis, avec les graphes, viennent des arcs dans les scénarios de parcours. Lorsque le graphe doit être distribué sur plusieurs noeuds de stockage, l'efficacité du partitionnement dépend directement de la répartition de ces arcs. (6)

LES ÉTAPES de traitement de données

- **Contrôle** : cette étape permet de s'assurer que les données brutes collectées sont valides. (7)
- **Correction** : sur certains types de données, des mécanismes d'estimation ou de correction peuvent être mis en oeuvre. Ces dispositifs sont concertés avec l'ensemble des parties prenantes. Par exemple, en cas de dysfonctionnement d'un dispositif de comptage d'un client particulier, l'énergie consommée est estimée à l'aide de méthodes reconnues par les organismes compétents (cf. article 20 du cahier des charges fonctionnelles sur le comptage électrique joint à la Communication de la CRE du 29 janvier 2004). De même, les corrections climatiques sont réalisées à l'aide d'une méthode dite « des gradients » partagée avec l'ensemble des acteurs et approuvée par le régulateur. (7)

- **Stockage** : ERDF procède ensuite à un stockage sécurisé des données validées et corrigées dans ses systèmes informatiques. Ces systèmes sont sécurisés conformément au cadre légal et réglementaire. Ainsi, la chaîne de traitement des données issues des compteurs communicants Linky respecte notamment le référentiel de sécurité rédigé par ERDF, évalué et certifié par l'Agence nationale de la sécurité des systèmes d'information (ANSSI). (7)
- **-Traitements adaptés** : en fonction des finalités, des traitements adaptés peuvent être effectués sur les données. Par exemple, il est possible de retraiter les données de consommation afin d'en déduire des données de facturation à destination des fournisseurs titulaires des contrats de fourniture, d'effectuer des agrégations de données de consommation et de production sur le périmètre d'une concession, etc. (7)
- **-Mise à disposition** : les données sont ensuite utilisées par ERDF pour l'exploitation, la maintenance et le développement du réseau public de distribution. ERDF transmet également aux acteurs externes les données nécessaires pour l'exercice de leurs activités ou missions (clients finaux, fournisseurs, collectivités, tiers, régulateur, etc.) dans le respect du cadre juridique, notamment concernant la protection des données commercialement sensibles et à caractère personnel. (7)

2.3.4. Organisation et gestion des données

L'ère de la gestion des importants volumes de données (BIG DATA) et de leur analyse ouvre de nouvelles opportunités significatives de différenciation concurrentielle en utilisant des analyses avancées. Il y a dix bonnes raisons pour lesquelles les organisations se tournent vers de nouvelles solutions de gestion de données pour gérer leur volume croissant de données et l'évolution des besoins en terme d'analyses. De nouvelles plateformes combinant le stockage et l'analyse des données au sein d'un même système permettent de relever le nouveau défi proposé par le phénomène des BIG DATA. Le stockage des données est géré par une base de données disposant d'une architecture permettant un important degré de parallélisations. Le traitement des données est géré par un moteur intégré d'analyse, ainsi l'analyse est opérée directement en base de données. [10]

❖ Une gestion rentable des données – BIG DATA sur du matériel standard

Tous les 18 mois, le volume de données d'une organisation double. La plupart des organisations sont maintenant confrontées à des téraoctets de données disponibles pour analyse. Le stockage de volume de données important (du Téraoctet au Pétaoctet) met en exergue les limites des modèles classiques de base de données et de Data Warehouse (DW), qui ne sont plus rentables pour faire face à ces quantités de données énormes. Les coûts du matériel pour optimiser le traitement des BIG

DATA sur les DW traditionnels, car ils sont poussés au-delà des téraoctets, deviennent étonnamment difficile à justifier. Plus grave pour l'entreprise, l'exécution de requêtes critiques pour l'activité peut nécessiter plusieurs heures sur des systèmes traditionnels au lieu de quelques secondes sur de nouvelles plateformes. Par conséquent, les coûts matériels importants imposés par les architectures traditionnelles contraignent les organisations à supprimer des données qui pourraient être conservées pour une analyse plus approfondie de leur activité. Une nouvelle approche technologique est nécessaire. Il devient alors nécessaire de disposer d'une architecture de gestion de données avec un traitement massivement parallélisable (MPP) qui parallélise les requêtes sur un cluster de serveurs plutôt que de se limiter aux performances d'un seul serveur. Cette architecture MPP devrait paralléliser toute les tâches de gestion des données, y compris les requêtes, les chargements, les sauvegardes, et les exportations. Tous les systèmes MPP ne présentent pas forcément ce type de fonctionnement. Ceux qui parallélisent les tâches fournissent de meilleures performances (entre 10 et 1000 fois supérieures selon les cas) que les systèmes traditionnels de bases de données relationnelles. (16)

❖ **Concurrence et performance**

Les DW et plateformes analytiques se sont imposés comme jamais auparavant pour suivre la croissance de nouveaux utilisateurs. Dans une enquête récente d'IDC, plus de 50% des personnes interrogées ont déclaré qu'après avoir ajouter plus d'utilisateurs à leur DW, les performances se sont considérablement dégradées. Gérer les besoins d'accès concurrents pour garantir la satisfaction des utilisateurs signifie une capacité à anticiper les performances et des niveaux de service garantis quelque soit les traitements réalisés. Un système qui prend en charge une forte concurrence sur des traitements simples ne nécessite pas nécessairement un même niveau de performance que pour des traitements qui deviennent plus complexes. Afin de s'assurer que les requêtes critiques sont traitées immédiatement, vous devriez être en mesure de réaffecter dynamiquement les ressources CPU et celles de stockage en se basant sur les transactions en cours. Cette gestion dynamique de charges doit travailler sur des centaines de traitements mixtes simultanés. Une mesure ne doit pas être basé sur les utilisateurs connectés, mais sur ceux actifs, et sur les traitements simultanés. (16)

2.3.5. Analyse des données

- ❖ Le volume de données produit et à traiter par les entreprises ne cesse d'augmenter. Ce constat n'est pas récent. Mais à l'ère d'Internet et de la multiplication des appareils connectés, l'analyse

de ces informations hétérogènes est devenue complexe. Un challenge par des technologies très pointues. (17)

- ❖ Le Web croule sous les données textuelles, audiovisuelles et graphiques. En 2008, environ 480 milliards de gigaoctets étaient disponibles sur Internet. Deux ans plus tard, ce volume a presque doublé ! Et cette évolution semble sans fin : selon le cabinet d'analyses Gartner, à l'horizon 2015, quelque trois milliards d'individus seront connectés et partageront des données, personnelles ou non, pour atteindre un total de huit Zo (Zettaoctets). Pour information, un Zo est égal à un milliard de To (Teraoctet), soit l'équivalent du volume stocké sur 250 milliards de DVD. Bienvenue dans l'ère du BIG DATA !
- ❖ Il s'agit d'analyser ces milliards de données afin de dénicher l'information pertinente permettant de prendre la bonne décision. Cette activité représente un marché évalué à 3,2 milliards de dollars en 2010. Elle devrait croître de 39,4% par an pour atteindre 16,9 milliards en 2015 (source : étude IDC « Digital Universe ») ! Cette croissance ne se limite pas à l'analyse d'énormes volumes. Le marché de la gestion et de l'analyse de données est estimé à plus de 100 milliards de dollars et croît de près de 10% par an, soit environ deux fois plus vite que le marché global des logiciels.
- ❖ Cette expression BIG DATA a été évoquée pour la première fois par le cabinet d'études Gartner en 2008. Quatre ans plus tard, la situation a considérablement évolué. Pour les entreprises, ces milliards de données sont à la fois une formidable source d'informations permettant de mieux connaître ses clients et les tendances de leur secteur d'activité (être capable de faire la différence entre le buzz ou l'épiphénomène et des « signaux forts ») et aussi de suivre l'évolution de ses concurrents.
- ❖ Dans l'univers de la gestion de l'information, trouver la donnée pertinente dans d'importants volumes structurées (bases de données, applications métier) est une activité maîtrisée. Mais la montée en puissance d'internet et la multiplication des appareils connectés à des réseaux (smartphones, tablettes mais aussi capteurs, compteurs électriques, bornes...) génèrent des données non structurées (emails, documents, fichiers, audio, vidéo, etc.).
- ❖ Cette évolution génère des coûts élevés. Une étude d'IDC a indiqué que les employés perdent au minimum cinq heures par semaine à rechercher les informations utiles pour leurs activités quotidiennes, ce qui coûte aux entreprises entre 4 000 € et 16000 € par an et par employé. Les solutions de BIG DATA doivent donc relever ces défis en tenant compte de la règle dite des 5 V. (17)

2.4. Conclusion

Face à l'explosion du volume d'informations, le BIG DATA vise à proposer une alternative aux solutions traditionnelles de bases de données et d'analyse (serveur SQL, plateforme de Business Intelligence...). Confrontés très tôt à des problématiques de très gros volumes, les géants du web, au premier rang desquels Yahoo (mais aussi Google et Facebook), ont été les premiers à déployer ce type de technologies. Selon le Gartner, le BIG DATA regroupe une famille d'outils qui répondent à un triple problématique : un **V**olume de données important à traiter, une grande **V**ariété d'informations (en provenance de plusieurs sources, non-structurées, structurées, Opendata...), et un certain niveau de **V**élocité à atteindre - c'est-à-dire de fréquence de création, collecte, traitement/analyse et partage de ces données...

En résumé, les principales technologies de BIG DATA sont nombreuses. Pour optimiser les temps de traitement sur des bases de données géantes, plusieurs solutions peuvent entrer en jeu

Des bases de données NoSQL qui implémentent des systèmes de stockage considérés comme plus performants que le traditionnel SQL pour l'analyse de données en masse (orienté clé/valeur, document, colonne ou graphe).

Des infrastructures de serveurs pour distribuer les traitements sur des dizaines, centaines, voire milliers de nœuds. C'est ce qu'on appelle le traitement massivement parallèle. Le framework Hadoop est sans doute le plus connu d'entre eux. Il combine le système de fichiers distribué HDFS, la base NoSQL HBase et l'algorithme MapReduce. D'autres technologies visant à tendre vers des traitements plus "temps réel" ont émergé dans la foulée (c'est le cas d'Apache Spark).

Le stockage des données en mémoire (Memtables) permet d'accélérer les temps de traitement des requêtes. (18)

CHAPITRE III

3. OUTILS ET APPLICATION

3.1. Introduction

Dans ce chapitre nous allons présenter un exemple très simplifié de l'utilisation des big data, pour cela on a choisi de collecter, traiter puis analyser des données du web. Nous avons choisi de prendre comme exemple quelques sites web du géant du commerce électronique Amazon (amazon.com, amazon.fr, amazon.co.uk, amazon.de, amazon.es et amazon.ca). Tous d'abord nous allons collecter une très petite partie des données de ces sites web, nous allons ensuite les traiter et enfin les analyser.

3.2. Algorithme de collecte et le traitement des données :

Dans cette partie du chapitre nous avons présenté les différentes étapes à suivre (algorithme) pour collecter et traiter les données de l'exemple pratique étudié.

Dans cet exemple nous allons collecter et traiter les données correspondantes aux smartphones à partir des sites web du commerce électronique Amazon (amazon.com, amazon.fr, amazon.co.uk, amazon.de, amazon.es et amazon.ca). Dans les sites Amazon il y'a beaucoup de types de produits chaque produit contient plusieurs types d'informations (une grande variation d'information) d'une part, d'autre part il y'a une énorme quantité d'informations (plusieurs exaoctets), pour rendre notre travail (la collecte et le traitement de données) faisable dans un temps raisonnable nous avons choisi les smartphones comme produit et quatre informations : les désignations des produits, les prix des produits, le nombre de vue pour chaque produit et les notes données pour chaque produit par les consommateurs. Dans la partie suivante nous allons montrer les différentes démarches que nous allons suivre pour faire la collecte des données.

3.2.1. Algorithme de collecte de données :

Dans cette section nous allons présenter l'algorithme à suivre pour collecter les données à partir des différents sites web d'Amazon cité dans la section précédente.

Algorithme :

Etape1 : Choisir un site web Amazon à traiter parmi les six sites (amazon.com, amazon.fr, amazon.co.uk, amazon.de, amazon.es ou amazon.ca).

Etape2 : Choisir une marque de smartphone à traiter (Samsung, Apple ou LG).

Etape 3 : Donner le type du smartphone à chercher (selon la marque choisie dans l'étape précédente).

Etape 4 : Lancer la recherche sur le site web choisi dans l'étape 1.

Etape 5 : collecter les informations à partir des codes source des pages html des sites web Amazon.

Etape 6 : Exécuter les étapes de 1 à 5 jusqu'à ce que toutes les pages html de la recherche effectuée soient terminées.

Etape 7 : Filtrer les données collectées pour garder que les données utiles (désignation du produit, prix du produit, nombre de vue du produit et notes données aux produits).

3.2.2. Méthode de traitement des données :

Afin d'analyser les données collectées nous allons d'abord les traiter. Le traitement des données est effectué en utilisant le tableur Excel de Microsoft.

Après filtrage des données dans la phase précédente (phase de collecte), les données filtrées seront copiées dans un fichier texte (txt) pour qu'on puisse les formater et les ouvrir par la suite dans un fichier Excel pour créer les différents tableaux qu'on va utiliser pour tracer des courbes en variant les différents critères qu'on a choisis et par la suite analyser les résultats trouvés.

3.3. Présentation de l'application :

Dans cette partie il est très important de bien comprendre le principe de fonctionnement de notre application, alors pour remédier à l'application le travail s'articule sur les points suivants :

- ✓ La description des étapes de l'application :
- ✓ La collecte des données sur les différents sites
- ✓ Le filtrage des données
- ✓ Explication du rôle de chaque champ dans l'interface

3.4. Intérêt d'application

L'intérêt de notre application est le suivant :

- ✓ Collecter le maximum de données sur les différents produits voulus
- ✓ Garder juste les données utiles en filtrant les données collectées.

3.5. Interface d'application

Dans cette partie on va expliquer notre interface d'application et le rôle de chaque champ

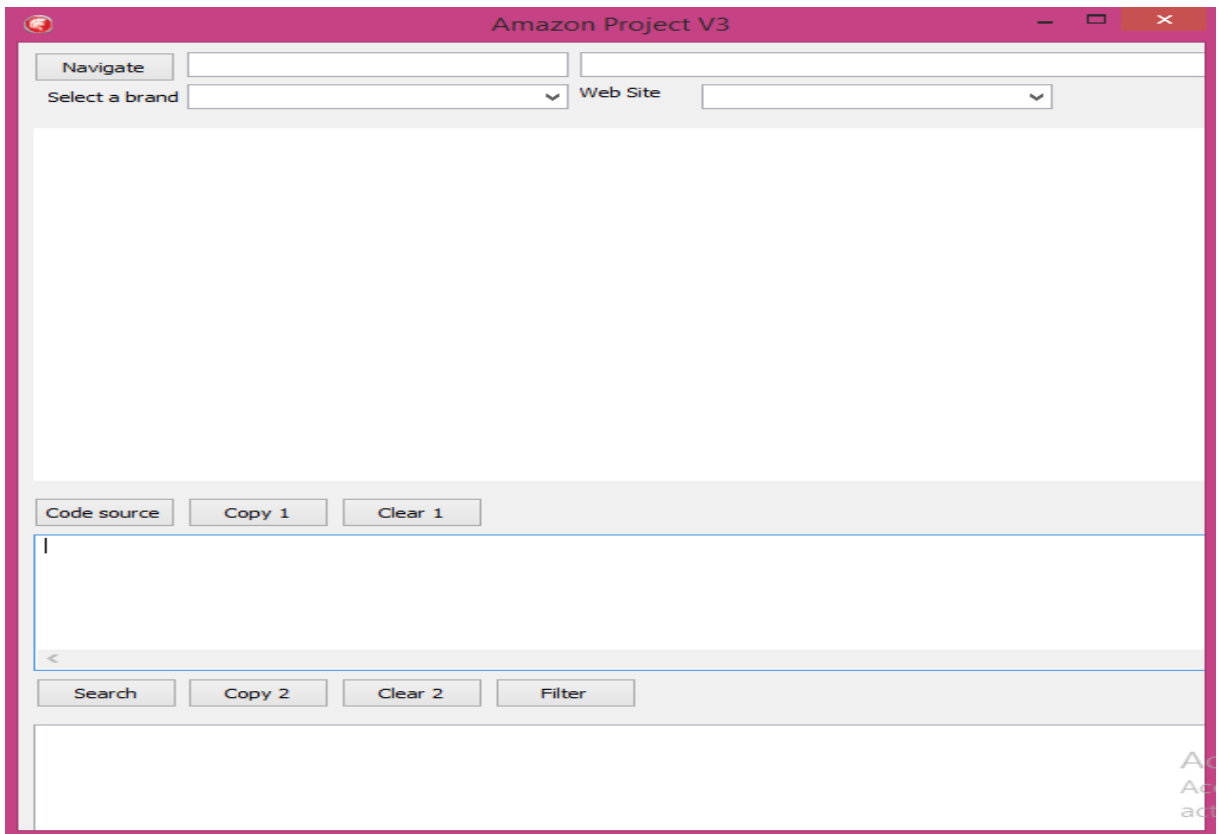


Figure 3.1 L'interface de l'application

Notre interface contient deux parties :

- La 1^{ère} partie consiste à collecter les données qui sont les différents liens des smart phones elle se compose des champs où on fait choisir le site, la marque, un champ où on fait entrer le nom de produit, un espace où sa sera afficher les différentes informations sur les produits.

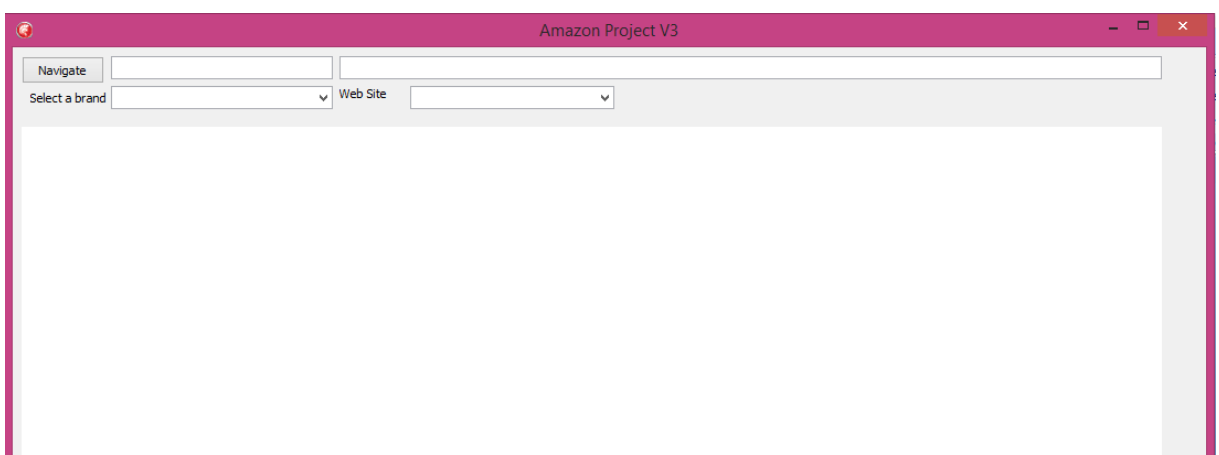


Figure 3.2 Espace d'affichage de l'information

Et un autre espace où sa sera afficher les codes source de ces informations.

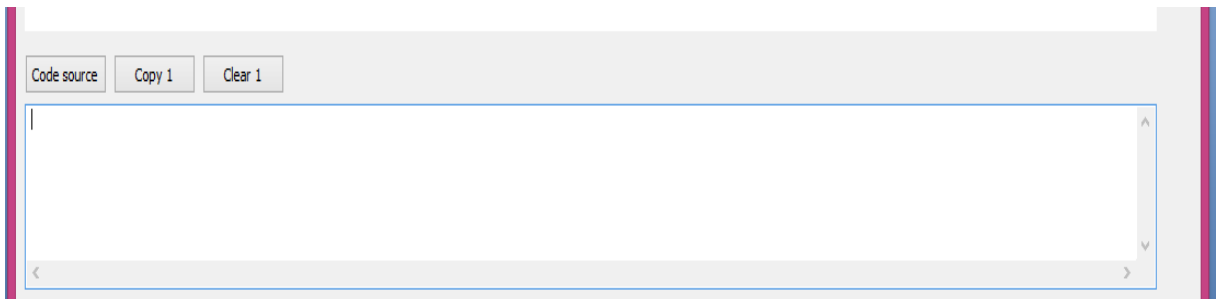


Figure 3.3 espace de code source

- La 2ième partie consiste à filtrer ces données c'est-à-dire elle consiste à supprimer toute information inutile et tout doublant.

Elle contient un espace où sa sera afficher les différents liens des produits.

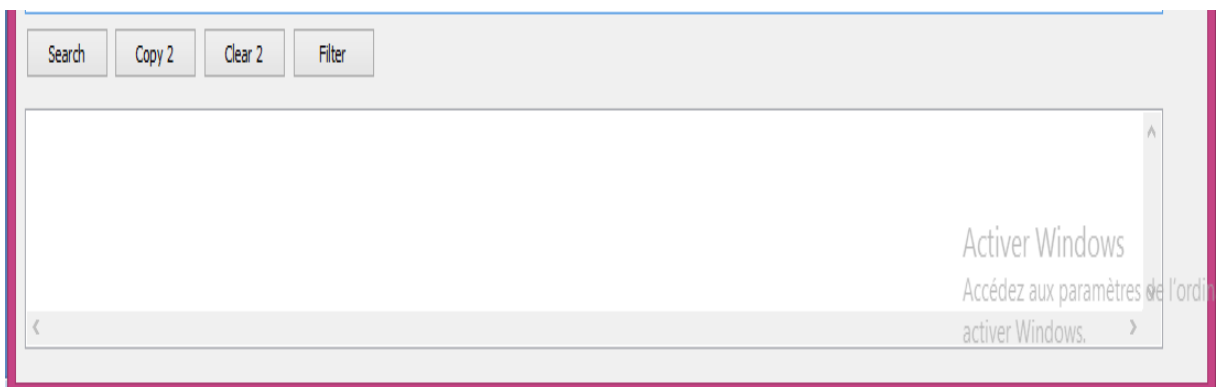


Figure 3.4 espace d'afficher les liens des produits

3.5.1. Explication des différents champs de l'interface :

- ✓ Pour chercher un produit dans une marque des smart phones il faut :
 - ✓ Tout d'abord sélectionner la marque

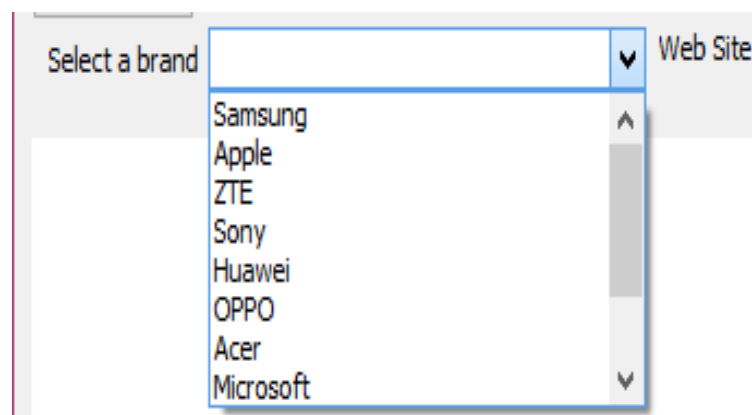


Figure 3.5 Espace d'affichage des marques

- ✓ Puis faire entrer le nom de produit
 - Par exemple : si on a sélectionné la marque « Samsung » on fait entrer « Samsung Galaxy S8 » comme nom de produit

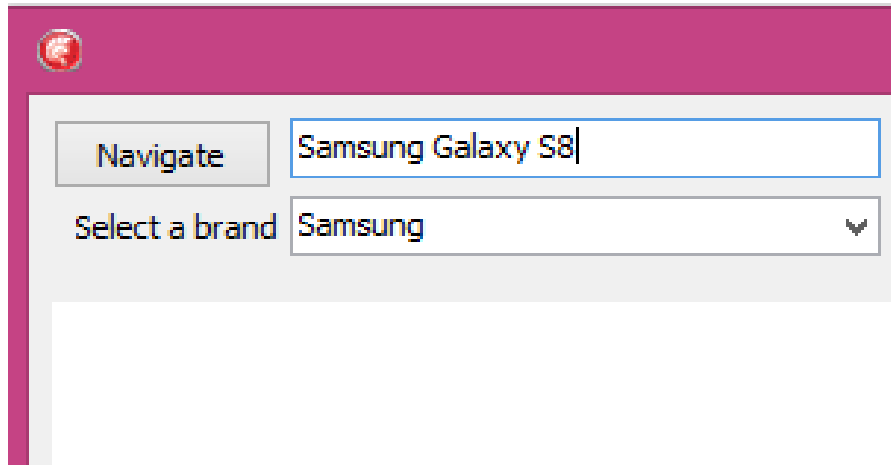


Figure 3.6 espace de recherche et navigation

- ✓ Ensuite il faut sélectionner un site

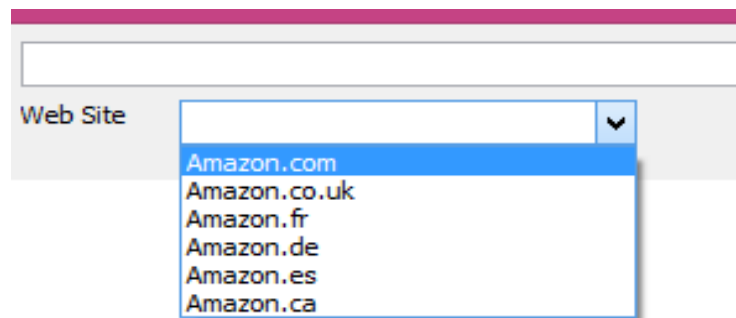


Figure 3.7 espaces de recherche des sites web

- ✓ En fin il faut cliquer sur le bouton « Navigate » pour chercher la page

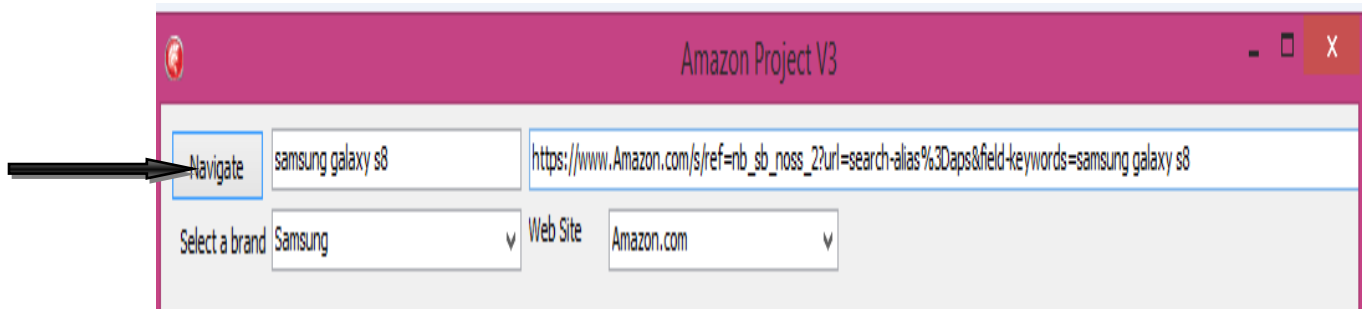


Figure 3.8 Exemple d'une recherche sur Samsung S8

Après ces étapes une page sera affichée

- Par exemple : si on a sélectionné la marque « Samsung », on a fait entrer « Samsung Galaxy S8 » comme nom de produit et on a choisi « Amazon.com » comme site puis on a cliqué sur le bouton « Navigate » : la page suivante sera affichée.

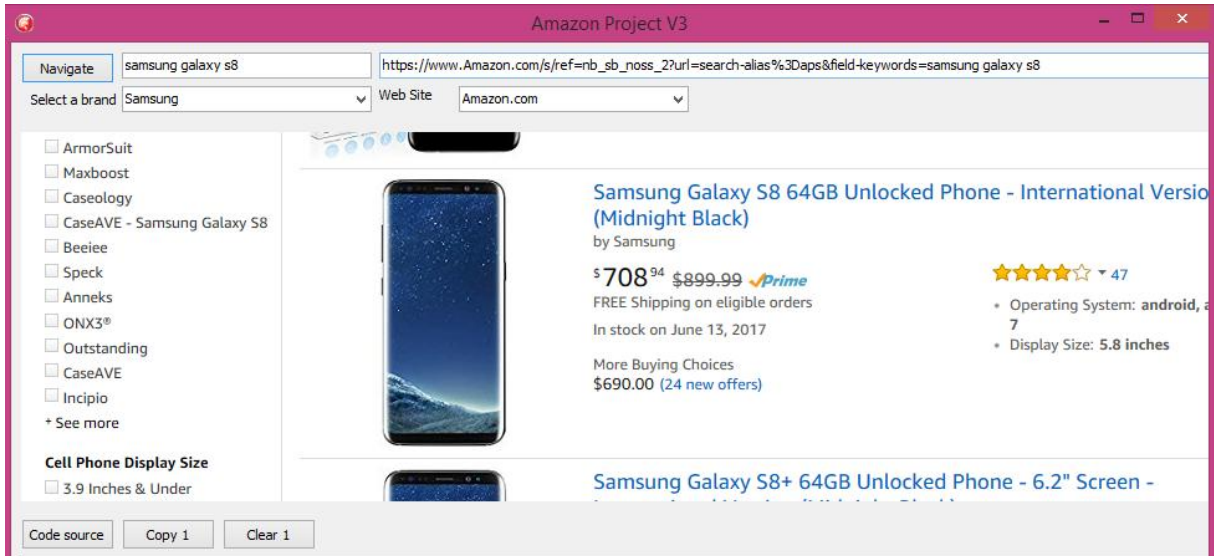
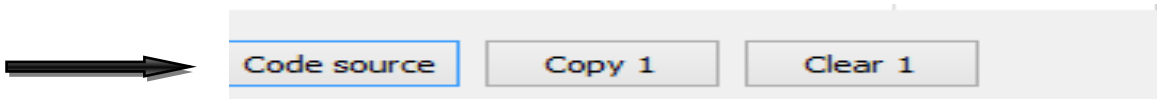


Figure 3.9 Exemple de résultat de recherche

Cette page contient les différentes informations sur le produit demandé de la marque choisie sur le site choisi.

- ✓ Pour collecter ces différentes informations il faut :
 - ✓ Cliquer sur le bouton « Code source » à chaque fois qu’une nouvelle page est affichée.



Après qu’on clique sur le bouton « Code source » les différents codes source seront affichés

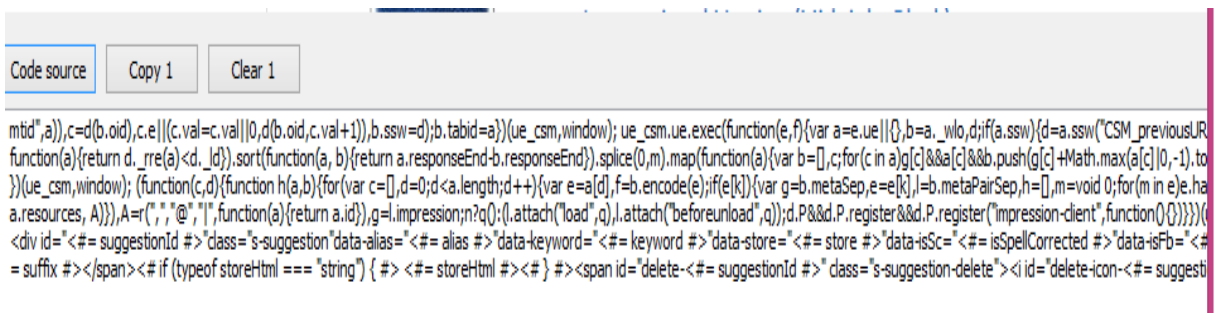


Figure 3.10 Exemple de résultat d’un code source recherchés

- ✓ Puis il faut cliquer sur le bouton « Search » pour que les différents liens des Produits soient affichés

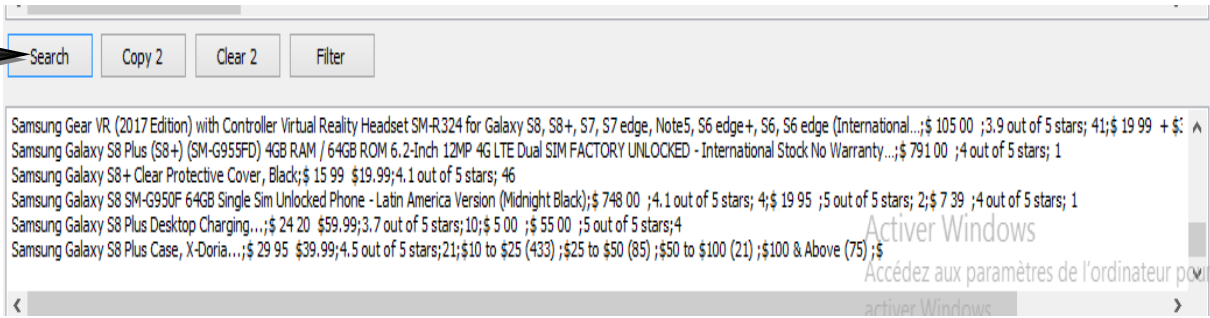
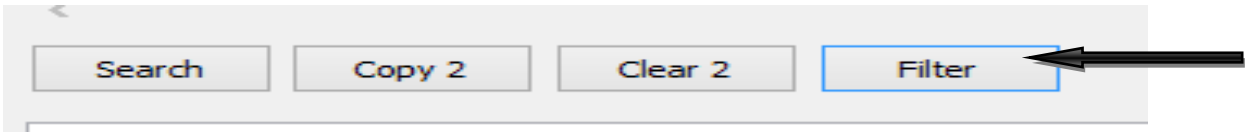


Figure 3.11 le resultat de la bouton " search"

- ✓ Pour filtrer ces données qui sont les liens des différents produits il faut :
 - ✓ Cliquer sur le bouton « **Filtre** »



Ce filtrage nous permet de garder juste les informations utiles et sans doublant.

- ✓ Pour copier ces informations (liens) dans une base données (Excel) il faut :
 - ✓ Cliquer sur le bouton « **Copy 2** »

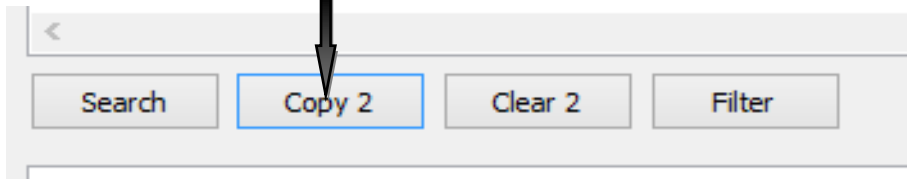


Figure 3.12 le bouton copier

Parce qu'on a besoin de liens pas des codes sources pour faire notre traitement car ces liens contiennent les informations utiles qu'on va les utiliser dans notre étude « **Big Data** » qui sont « **La description** », « **Le prix** », « **La note** » et « **Le N° de vue** »

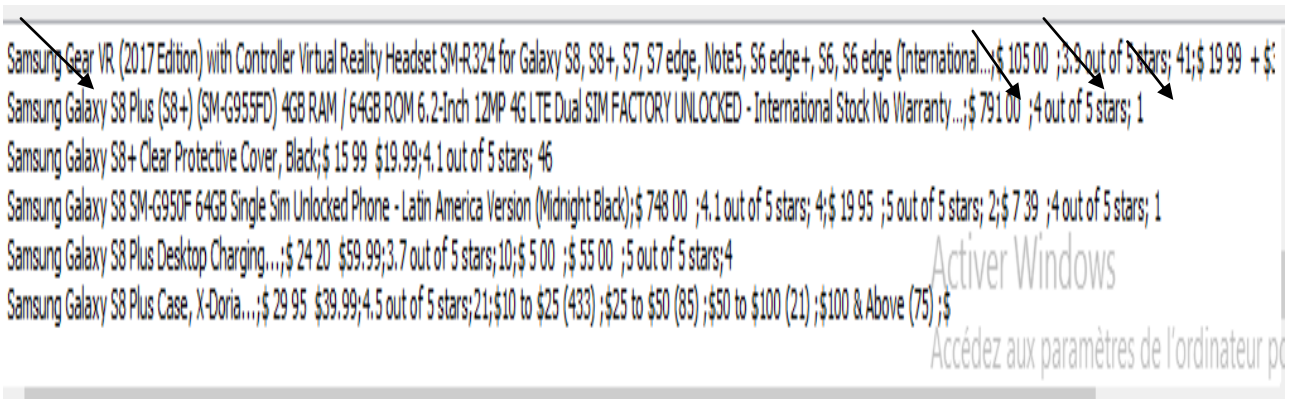


Figure 3.13 représente les informations utiles

- ✓ Pour collecter les données d'un autre produit il faut :

- ✓ Tous d’abord cliquer sur les boutons « **Clear 1** » et « **Clear 2** » pour effacer les données de produit précédent et garder juste les données de nouveau Produit.

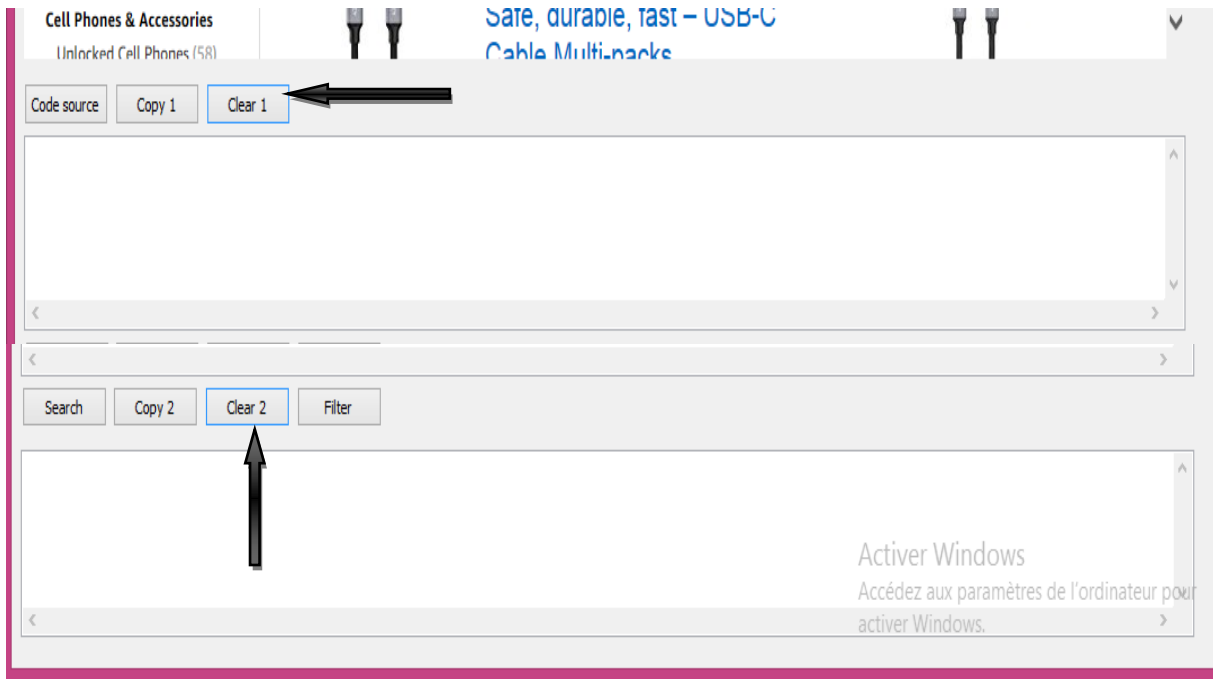


Figure 3.14 Le bouton " clear" pour supprimer

3.6. Résultats Analyses des données

3.6.1. Analyse des données selon les sites web

L’analyse selon les sites se fait en comparons les données concernant le prix, la note et le nombre de vue de chaque produit de chaque marque dans les différents 6 plateformes utilisées (Amzon.com, Amazon.co.uk, Amazon.fr, Amazon.de, Amazon.es et Amazon.ca) et ça nous permet par la suite de visualiser les résultats sous forme des histogrammes afin d’arriver à la prédiction.

A. Selon le site Amazon.com

L’analyse se fait en fixant le site Amazon.com puis en fixe la marque des produits (Apple, Samsung, LG) et la catégorie (prix, note, nombre de vue) et en fait l’analyse en variant les gamme (haute, moyenne, entrée de gamme)

a. Analyse des prix :

Les valeurs des prix sont des valeurs moyennes.

	HG	MG	EG
SAMSUNG	566,07	314,70	196,55
LG	535,61	309,6	141,36
APPL	582,94	406,25	

Tableau 3-1 Les valeurs moyennes des prix des Smartphones selon le site Amazon.com

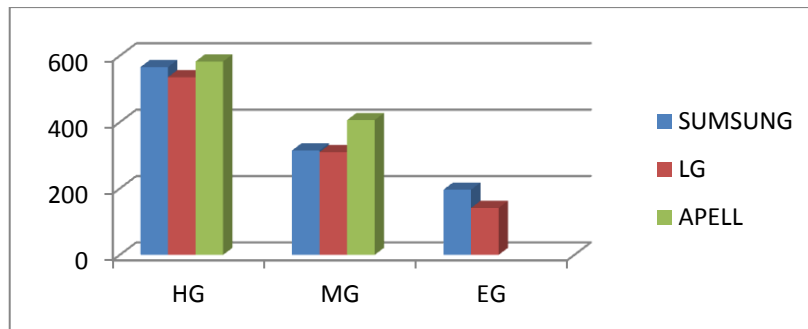


Figure 3.15 : La moyenne Prix des Smartphone selon le site Amazon.com

Dans cette figure en représente Les prix des Smartphone en fixant le site web amazon.com, on remarque que cette figure se composent de 3 parties :

1ère partie : représente l’histogramme des prix des Smartphone selon la haute gamme, cette partie montre que les Smartphone de marque APPLE ont les prix les plus élevés ensuite les Smartphone SAMSUNG et en dernière partie on a les Smartphone LG

2^{ème} partie : La figure illustre la représentation des prix des Smartphone de la moyenne gamme. On remarque que toujours que les Smartphones APPLE ont les prix les plus élevés par rapport aux autres marques LG et SAMSUNG qui donne presque la même tendance de valeur des prix.

3^{ème} partie : cette dernière partie de figure vient a représenté des Smartphones entrée de gamme. Dans cette catégorie des Smartphones, on ne trouve pas la marque APILLE par contre on remarque que les prix des Smartphones SMANSUNG sont toujours plus grand que la marque LG.

b. Analyse des notes

Les valeurs des notes à prendre sont des valeurs moyennes.

	HG	MG	EG
SAMSUNG	4,11	4,05	4,09
LG	4,2	4,27	3,54
APPLE	4,10	4,13	

Tableau 3-2 Les valeurs moyennes des notes des Smartphones selon le site Amazon.com

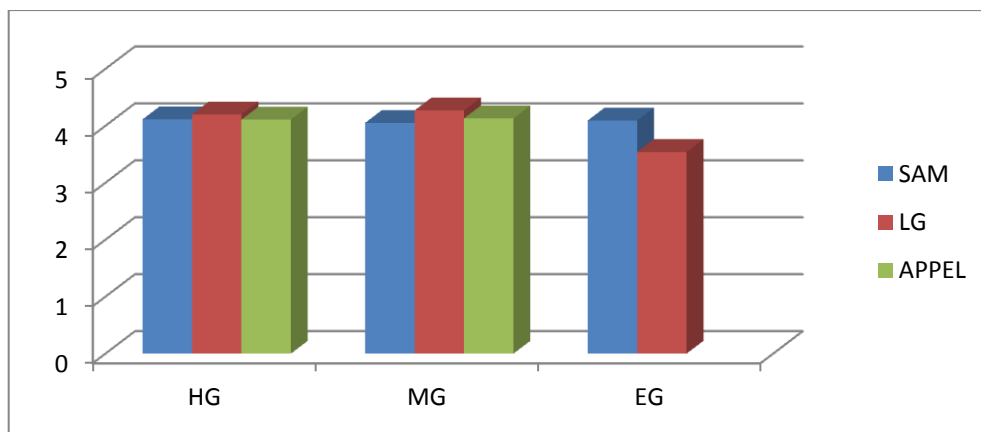


Figure 3.16 La moyenne des notes des Smartphones selon le site Amazon.com

La figure représente La moyenne des notes des Smartphone en fixant toujours le site web amazon.com, on remarque que cette figure se compose aussi de 3 parties :

1^{ère} partie : représente l'histogramme des notes des Smartphone selon la haute gamme, cette partie montre que les Smartphone SAMSUNG, LG et APPLE ont presque la même moyenne de note qui est 4 et nous montre que les trois marques ont une bonne réputation et ça dans le site Amazon.com.

2^{ème} partie : Cette partie donne la représentation des prix des Smartphone de la moyenne gamme. Elle montre une petite différence de note entre les trois marques des Smartphone ou on remarque que les Smartphone LG prend la grande valeur.

3^{ème} partie : cette dernière partie de figure vient à représenter des Smartphones entrée de gamme. Dans cette catégorie des Smartphones, on ne trouve pas la marque APPLE par contre on remarque que les valeurs des notes des Smartphones SAMSUNG sont plus élevées que la marque LG.

c. Analyse des nombres des vue

Les valeurs des notes à prendre sont des valeurs moyennes

	HG	MG	EG
SAMSUNG	130,3	117,3	148,48
LG	110,6	50,9	30,62
APPL	125,27	74,73	

Tableau 3-3 Les valeurs moyennes des nombres de vue des Smartphones selon le site Amazon.com

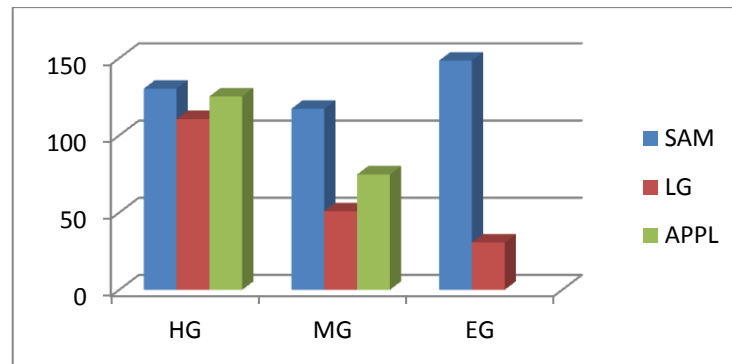


Figure 3.17 La moyenne des nombres de vue des Smartphones selon le site Amazon.com

Cette figure se compose de 3 parties :

1^{ère} partie : représente l'histogramme des nombres de vue des Smartphone selon la haute gamme, cette partie montre que les Smartphone de marque SAMSUNG ont le maximum des vues ensuite les Smartphone APPLE et en dernière partie on a les Smartphone LG.

2^{ème} partie : La figure montre la représentation des nombres de vue des Smartphone de la moyenne gamme. On remarque que les Smartphones SAMSUNG ont toujours les prix les plus élevés par rapport aux autres marques LG et SAMSUNG qui donne presque la même tendance de valeur des prix.

3^{ème} partie : Dans cette partie de figure on représente les Smartphones entrée de gamme. Toujours pour la marque APPLLE il n ya pas d'entrée de gamme des produits par contre Les Smartphones SAMSUNG ont la valeur des vues la plus grand ou les Smartphones LG ont le minimum de vue dans ce cas.

B. Selon le site Amazon.co.uk

L'analyse se fait en fixant le site Amazon.co.uk puis en fixe la marque des produits (Apple, Samsung, LG) et la catégorie (prix, note, nombre de vue) et en fait l'analyse en variant les gammes (haute, moyenne, entrée de gamme)

a. Analyse des prix

Les valeurs des prix, notes et nombre de vues a prendre sont des valeurs moyennes

	HG	MG	EG
SAMSUNG	502,92	232,69	150,15
LG	406,39	220,3	105,96
APPL	492,05	352,61	

Tableau 3-4 Les valeurs moyennes des prix des Smartphones selon le site Amazon.co.uk

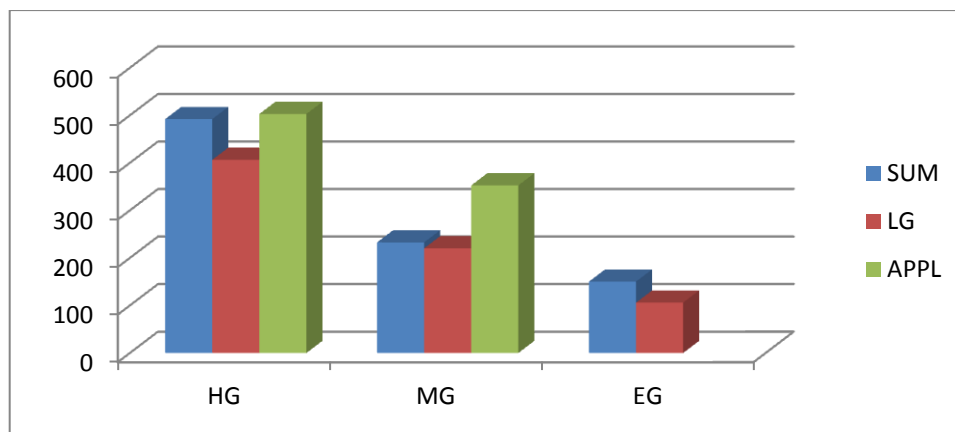


Figure 3.18 La moyenne des prix des Smartphones selon le site Amazon.co.uk

La figure est divisée en 3 parties :

1^{ère} partie : illustre l'histogramme des prix des Smartphone selon la haute gamme, cette partie montre que les Smartphone de marque APPLE ont les prix les plus élevés ensuite les Smartphone SAMSUNG et puis les Smartphone LG

2^{ème} partie : est la représentation des prix des Smartphone de la moyenne gamme. On remarque toujours que les Smartphones APPLE ont les prix les plus élevés par rapport aux autres marques LG et SAMSUNG qui donne presque les mêmes valeurs moyennes des prix.

3^{ème} partie : la dernière partie de figure représentée des Smartphones entrée de gamme. Dans cette catégorie des Smartphones, on ne trouve pas la marque APPLLE par contre on remarque que les prix des Smartphones SMANSUNG sont un peu plus grand que la marque LG.

b. Analyse des notes

Les valeurs des notes a prendre sont des valeurs moyennes

	HG	MG	EG
SAMSUNG	4,43	4,48	4,46
LG	4,32	4,43	4,32
APPLE	4,06	4,01	

Tableau 3-5 Les valeurs moyennes des notes des Smartphones selon le site Amazon.co.uk

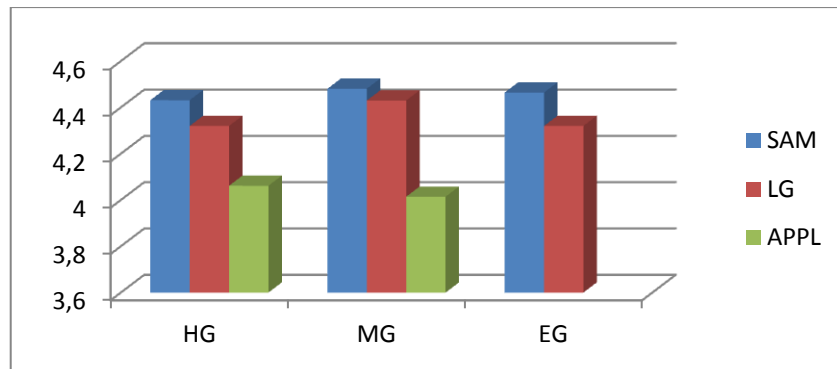


Figure 3.19 La moyenne des notes des Smartphones selon le site Amazon.co.uk

La figure représente Les notes des Smartphone en fixant le site web amazon.co.uk, cette figure se compose de 3 parties :

1^{ère} partie : l'histogramme des notes des Smartphone selon la haute gamme, cette partie montre que les Smartphone de marque SAMSUNG ont les notes les plus élevés ensuite les Smartphone LG et en dernière partie on a les Smartphone APPLE

2^{ème} partie : C'est une illustration des prix des Smartphone de la moyenne gamme. Dans cette partie on trouve que les Smartphones SAMSUNG ont les notes par rapport aux autres marques LG et SAMSUNG qui donne presque la même tendance de valeur des prix.

3^{ème} partie : cette partie de figure vient a montré des Smartphones entrée de gamme. Dans cette catégorie des Smartphones, on ne trouve pas la marque APPLLE par contre on remarque que les Smartphones SMANSUNG ont toujours une moyenne du note plus grand que la marque LG

c. Analyse des nombres de vue

Les valeurs des notes a prendre sont des valeurs moyennes

	HG	MG	EG
SAMSUNG	127,68	21	94,92
LG	137,25	52,5	33,4
APPLE	84,84	84,57	

Tableau 3-6 Les valeurs moyennes des nombres des vues des Smartphones selon le site Amazon.com

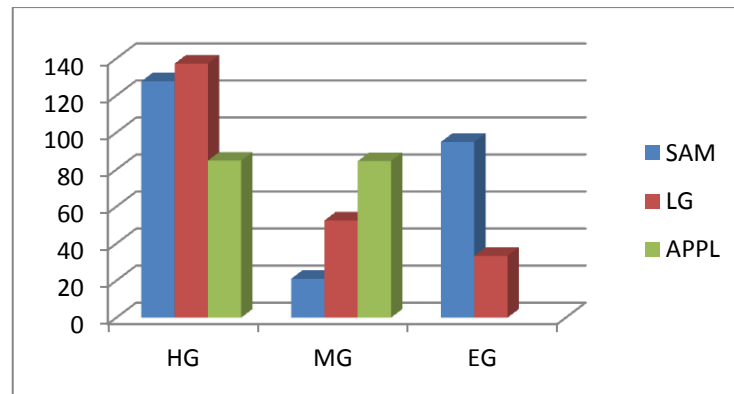


Figure 3.20 La moyenne des nombres de vue des Smartphones selon le site Amazon.co.uk

La figure est divisée en 3 parties :

1^{ère} partie : est La visualisation de la moyenne des vues de la haute gamme. Elle montre dans ce cas que les Smartphones LG ont le maximum de vue puis en classe les Smartphones SAMSUNG avec moyen de vue assez grand et par la suite on a les Smartphones de la marque APPLE.

2^{ème} partie : est la représentation des prix des Smartphone de la moyenne gamme. On remarque toujours que les Smartphones APPLE ont les prix les plus élevés par rapport aux autres marques LG et SAMSUNG qui donne presque les mêmes valeurs moyennes des prix.

3^{ème} partie : représente les Smartphones entrée de gamme. Les Smartphones APPLE n'ont pas toujours des produits dans cette catégorie. En outre, on remarque que les Smartphones SAMSUNG ont les prix les plus grands par rapport à LG

C. Selon le site Amazon.fr

L'analyse se fait en fixant le site Amazon.fr puis en fixe la marque des produits (Apple, Samsung, LG) et la catégorie (prix, note, nombre de vue) et en fait l'analyse en variant les gammes (haute, moyenne, entrée de gamme)

a. Analyse des prix

Les valeurs des prix à prendre sont des valeurs moyennes

	HG	MG	EG
SAMSUNG	589,39	329,96	165,04
LG	520	324	95,47
APPLE	623,41	444,50	

Tableau 3-7 Les valeurs moyennes des prix des Smartphones selon le site Amazon.fr

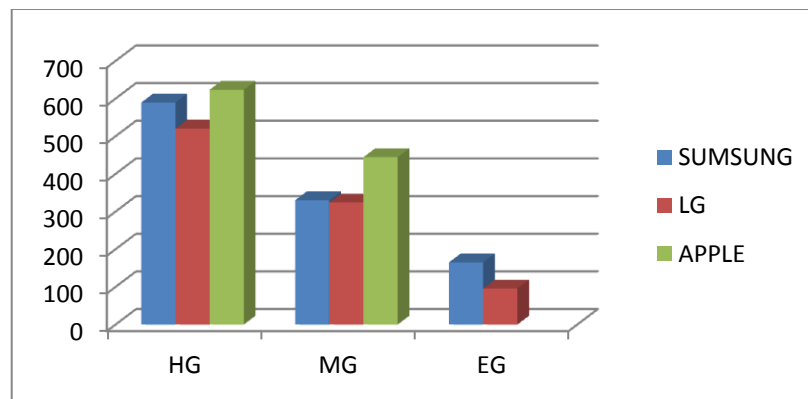


Figure 3.21 La moyenne des prix des Smartphones selon le site Amazon.fr

La figure représente Les notes des Smartphone en fixant le site web amazon.fr. La figure se compose de 3 parties :

1^{ère} partie : Dans cette partie de la figure, on classe les Smartphones de la haute gamme. La marque APPLE dans cette partie est la plus couteuse ensuite les Smartphone de marque SAMSUNG qui ont de mêmes des prix élevés ensuite les Smartphone LG avec des prix assez grande.

2^{ème} partie : est les prix des Smartphone de la catégorie moyenne. La partie montre que la gamme APPLE est toujours la plus couteuse par rapport aux deux autres gammes qui ont des valeurs des proches dans ce cas.

3^{ème} partie : cette partie de figure vient a montré des Smartphones entrée de gamme ou on ne trouve pas des produits de marque APPLE. Les Smartphones dans cette catégorie sont moins chers ou on remarque aussi que les prix des Smartphones SAMSUNG sont un peu plus élevés par rapport aux autres de LG.

b. Analyse des notes

Les valeurs des notes prendre sont des valeurs moyennes

	HG	MG	EG
SAMSUNG	3,92	3,55	3,87
LG	4,2	4,09	3,78
APPLE	3,82	3,77	

Tableau 3-8 Les valeurs moyennes des notes des Smartphones selon le site Amazon.fr

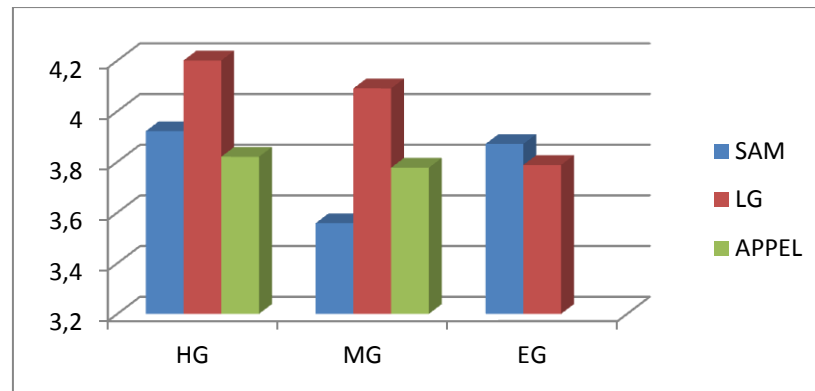


Figure 3.22 La moyenne des notes des Smartphones selon le site Amazon.fr

L'histogramme est divisé en 3 parties :

1^{ère} partie : est la partie de la haute gamme des Smartphones. Vue la tendance de la figure, on remarque que les Smartphones LG sont notés comme les meilleurs produit par rapport aux marque SAMSUNG et LG

2^{ème} partie : représente les notes de la gamme moyenne de chaque marque des Smartphones. Toujours, on remarque dans ce cas que les Smartphones de la marque LG sont les préférés puis on classe les produits APLE avec un résultat plus grand que SAMSUNG.

3^{ème} partie : Dans la catégorie d'entrée de gamme, les Smartphones SAMSUNG sont les préférable par rapport aux autres produits de LG

c. Analyse des nombres de vue

Les valeurs des nombres de vue à prendre sont des valeurs moyennes

	HG	MG	EG
SAMSUNG	126,66	121,42	253,58
LG	21,44	33,3	8,43
APPLE	124,55	97,52	

Tableau 3-9 Les valeurs moyennes des nombre de vue des Smartphones selon le site Amazon.fr

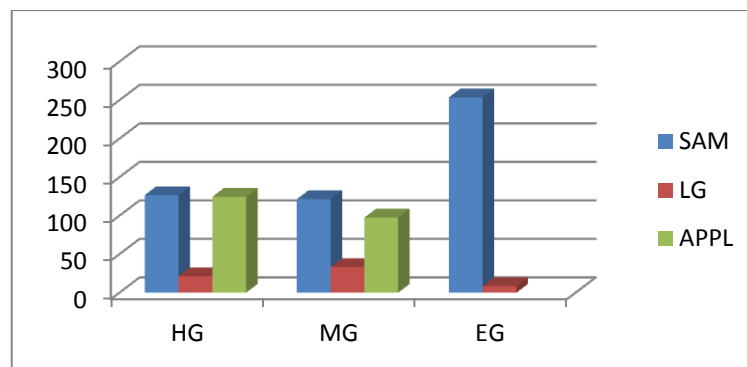


Figure 3.23 La moyenne de nombre de vue des Smartphones selon le site Amazon.fr

1^{ère} partie : Dans cette partie de la catégorie haute de gamme, les Smartphone SAMSUNG et APPLE sont les plus vue par la population qui utilise le site Amazon.fr.

2^{ème} partie : est les valeurs moyennes des nombre de vue des Smartphone de la catégorie moyenne. On remarque toujours que les marques SMASUNG et APPLE ont le nombre de vue le plus grand par rapport au LG avec un petit baissment de la valeur dans les produits APPLE.

3^{ème} partie : cette partie de figure vient a montré des Smartphones entrée de gamme. Les Smartphones SAMSUNG prend la valeur maximum de nombre de vue par contre les Smartphones LG ont une moyenne de vue presque nulle et ça toujours en absence des produits APPLE pour cette catégorie.

D. Selon le site Amazon.de

L'analyse se fait en fixant le site Amazon.de puis en fixe la marque des produits (Apple, Samsung, LG) et la catégorie (prix, note, nombre de vue) et en fait l'analyse en variant les gammes (haute, moyenne, entrée de gamme)

a. Analyse des prix

Les valeurs des prix, a prendre sont des valeurs moyennes

	HG	MG	EG
SAMSUNG	568,75	266,66	183,17
LG	230,33	230,66	76,01
APPLE	450,6	304,87	

Tableau 3-10 Les valeurs moyennes des prix des Smartphones selon le site Amazon.de

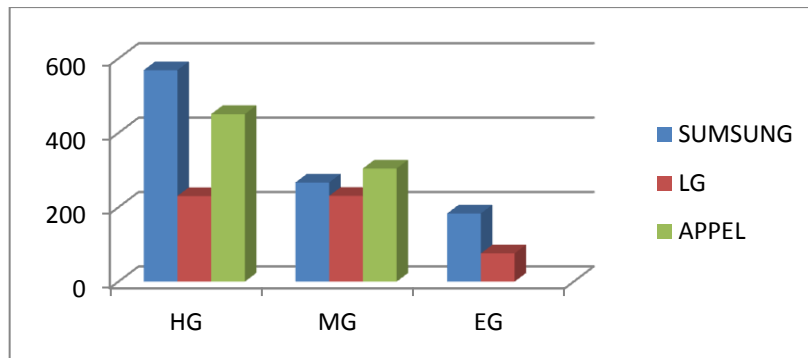


Figure 3.24 La moyenne des prix des Smartphones selon le site Amazon.de

LA figure représente Les prix des Smartphone en fixant le site web amazon.de. La figure se compose de 3 parties :

1^{ère} partie : représente le classement des Smartphones de la haute gamme. La marque SAMSUNG dans cette partie les produits APPLE sont les plus chères, puis on note les prix des Smartphone SAMSUNG ensuite les Smartphones LG qui sont les moins chers.

2^{ème} partie : Dans la partie des produits de moyenne gamme les valeurs prix des Smartphones sont plus proches mais les produits APPLE restent toujours celle avec des prix élevés.

3^{ème} partie : de même dans cette partie on ne peut pas définir une gamme d’entrées pour les produits APPLE mais c’est le cas ou toujours les Smartphones SAMSUNG restent les plus chers.

b. Analyse des notes

	HG	MG	EG
SAMSUNG	3,8	4,31	3,8375
LG	3,71	4,45	4,26
APPLE	3,84	4	

Tableau 3-11 Les valeurs moyennes des notes des Smartphones selon le site Amazon.de

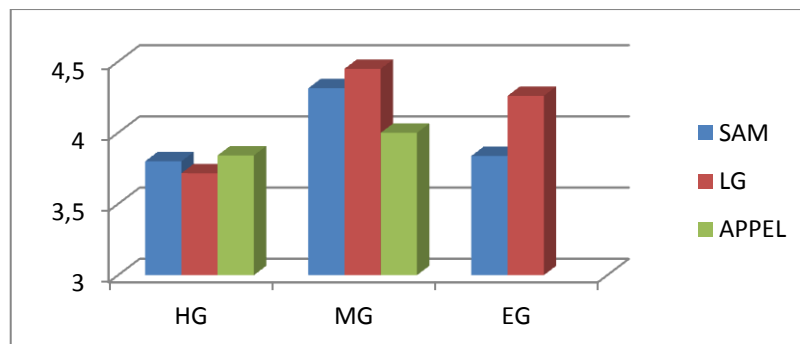


Figure 3.25 La moyenne des notes des Smartphones selon le site Amazon.de

La figure est divisée en 3 parties :

1ère partie : est la partie de la haute gamme de figure où on remarque que la valeur moyenne des notes de trois marques est très proche.

2^{ème} partie : est cette partie de moyenne gamme montre une différence par rapport à la ou on trouve que les Smartphones LG sont notés les meilleurs par rapport aux produits de SAMSUNG puis la marque APPLE arrive à la fin.

3^{ème} partie : la figure vient à montré l'entrée gamme des Smartphones ou la marque LG reste celle avec une valeur moyenne de note plus élevée que SAMSUNG.

c. Analyse des nombres de vue

Les valeurs des et nombre de vues à prendre sont des valeurs moyennes

	HG	MG	EG
SAM	426,92	240,63	312,37
LG	209,22	250,5	19
APPLE	244,22	354,86	

Tableau 3-12 Les valeurs moyennes des nombres de vue des Smartphones selon le site Amazon.de

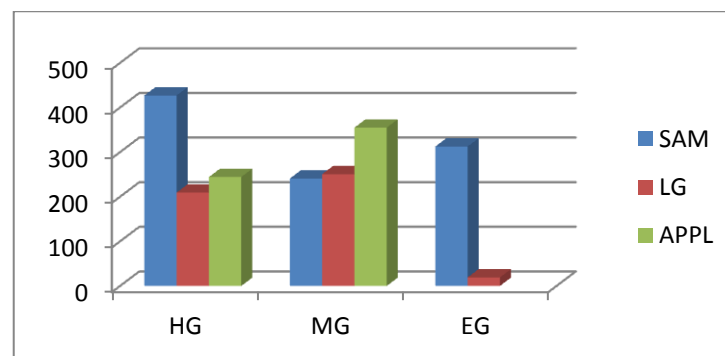


Figure 3.26 La moyenne des nombres de vue des Smartphones selon le site Amazon.de

1ère partie : Dans cette partie de la catégorie haute de gamme, les Smartphone SAMSUNG et APPLE sont les plus vue par les populations qui utilisent le site Amazon.de.

2^{ème} partie : représente les Smartphone de la catégorie moyenne. Dans cette figure, les produits APPLE sont les plus vue par rapport aux autres produits SAMSUNG et LG qui ont presque la même valeur moyenne des vue

3^{ème} partie : La partie entrée de gamme de la figure montre que les Smartphone SAMSUNG donne la plus grande valeur des nombres de vue

E. Selon le site Amazon.es

L'analyse se fait en fixant le site Amazon.es puis en fixe la marque des produits (Apple, Samsung, LG) et la catégorie (prix, note, nombre de vue) et en fait l'analyse en variant les gamme (haute, moyenne, entrée de gamme)

a. Analyse des prix

Les valeurs des prix a prendre sont des valeurs moyennes

	HG	MG	EG
SAM	583,77	273,60	167,96
LG	450,3	240,3	41,65
APPL	594,40	408,43	

Tableau 3-13 Les valeurs moyennes des prix des Smartphones selon le site Amazon.es

En remarque que les moyennes de SAMSUNG, LG et Apple et dans les trois gammes sont des chiffres raisonnables en terme des prix

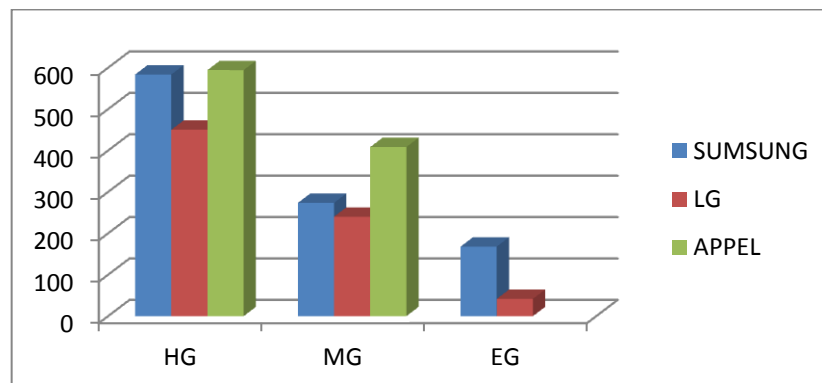


Figure 3.27 : La moyenne des prix des Smartphones selon le site Amazon.es

La figure représente Les prix des Smartphone en fixant le site web amazon.es. La figure se compose de 3 parties :

1ère partie : représente la visualisation des Smartphones de la haute gamme. Les deux marques APPLE et SAMSUNG ont les prix les plus chers avec une valeur un peu plus grande pour les produits APPLE de même la marque LG montre une valeur assez grande mais elle reste moins chère que les autres marques.

2^{ème} partie : Dans la partie des produits de moyenne gamme les valeurs prix des Smartphones sont plus proches mais les produits APPLE restent toujours celle avec des prix élevés.

3^{ème} partie : de même dans cette partie on ne peut pas définir une gamme d'entrées pour les produits APPLE mais c'est le cas où toujours les Smartphones SAMSUNG restent plus cher.

a. Analyse des notes

Les valeurs des notes à prendre sont des valeurs moyennes

	HG	MG	EG
SAMSUNG	4,03	4,205	4,25
LG	3,82	4,2	4,11
APPLE	4,06	4,05	

Tableau 3-14 Les valeurs moyennes des notes des Smartphones selon le site Amazon.es

- En remarque que les moyennes de Samsung, LG et Apple et dans les trois gammes sont des chiffres raisonnables en terme des notes.

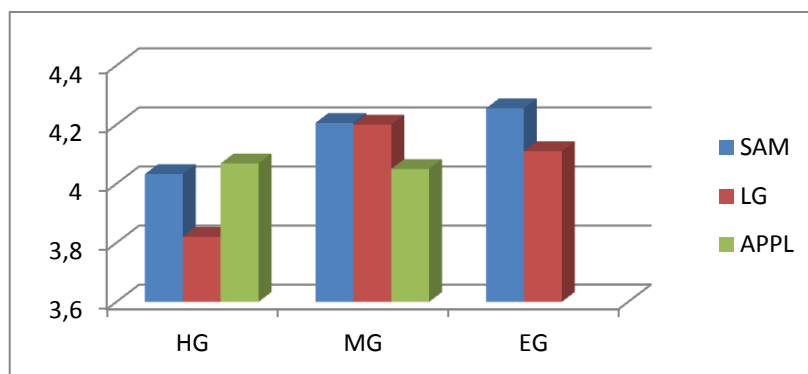


Figure 3.28 La moyenne des notes des Smartphones selon le site Amazon.es

La figure est divisée en 3 parties :

1^{ère} partie : est la partie de la haute gamme de figure où on remarque que les produits APPLE sont notés étant que la plus préférés comparant aux produits SAMSUNG et LG arrive avec la valeur des notes la plus petite.

2^{ème} partie : cette partie de moyenne gamme montre une différence par rapport à la première où on trouve que les Smartphones SAMSUNG et LG sont notés les meilleurs par rapport aux produits de la marque APPLE.

3^{ème} partie : la figure vient à montré l'entrée gamme des Smartphones ou la marque SAMSUNG prend une valeur moyenne de note plus élevé que la marque LG.

b. Analyse des nombres de vue

Les valeurs des nombres de vues à prendre sont des valeurs moyennes

	HG	MG	EG
SAMSUNG	99,20	34,35	31,72
LG	47	55,99	21,11
APPLE	74,93	42,2	

Tableau 3-15 Les valeurs moyennes des nombres de vue des Smartphones selon le site Amazon.de

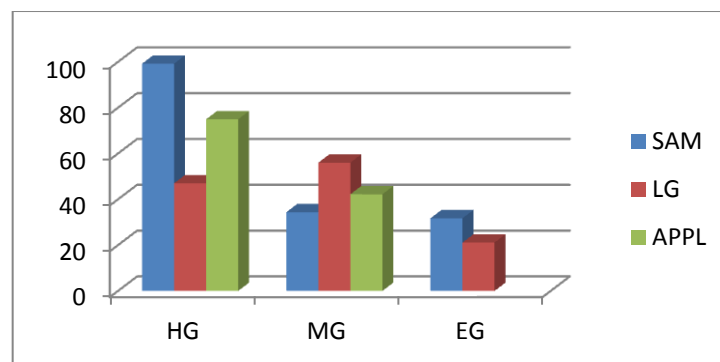


Figure 3.29 La moyenne des nombres de vue des Smartphones selon le site Amazon.es

1^{ère} partie : Dans cette partie de la catégorie haute de gamme, les Smartphone SAMSUNG et APPLE sont les plus vue sur le site Amazon.es.

2^{ème} partie : représente les Smartphone de la catégorie moyenne. Dans cette figure les produits LG sont les plus vue par rapport aux autres produits SAMSUNG et APPLE.

3^{ème} partie : La partie entrée de gamme de la figure montre que les Smartphone SAMSUNG donnent la plus grande valeur des moyennes de nombre de vue .

F. Selon le site Amazon.ca

L'analyse se fait en fixant le site Amazon.ca puis en fixe la marque des produits (Apple, Samsung, LG) et la catégorie (prix, note, nombre de vue) et en fait l'analyse en variant les gamme (haute, moyenne, entrée de gamme)

a. Analyse des prix

Les valeurs des prix à prendre sont des valeurs moyennes

	HG	MG	EG
SAMSUNG	708,82	468,85	292,75
LG	440,3	402,19	140,3
APPLE	854,66	500,81	

Tableau 3-16 Les valeurs moyennes des prix des Smartphones selon le site Amazon.ca

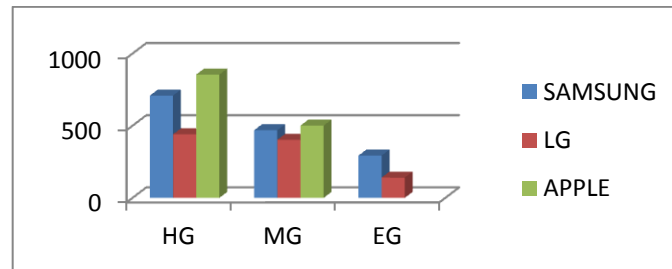


Figure 3.30 La moyenne des prix des Smartphones selon le site Amazon.ca

. La figure est divisée en 3 parties :

1^{ère} partie : illustre l'histogramme des prix des Smartphone selon la haute gamme, cette partie montre que les Smartphone de marque APPLE ont les prix les plus élevés ensuite les Smartphone SAMSUNG et puis les Smartphone LG

2^{ème} partie : est la représentation des prix des Smartphone de la moyenne gamme. On remarque toujours que les Smartphones APPLE ont les prix les plus élevés par rapport aux autres marques LG et SAMSUNG qui donne presque les mêmes valeurs moyennes des prix.

3^{ème} partie : la dernière partie de figure représentée des Smartphones entrée de gamme. Dans cette catégorie des Smartphones, on ne trouve pas la marque APPLE par contre on remarque que les prix des Smartphones SAMSUNG sont un plus grand que la marque LG.

c. Analyse des notes

Les valeurs des notes à prendre sont des valeurs moyennes

	HG	MG	EG
SAMSUNG	4,26	4,36	4,44
LG	4,26	3,96	4,77
APPLE	3,54	3,97	

Tableau 3-17 Les valeurs moyennes des notes des Smartphones selon le site Amazon.ca

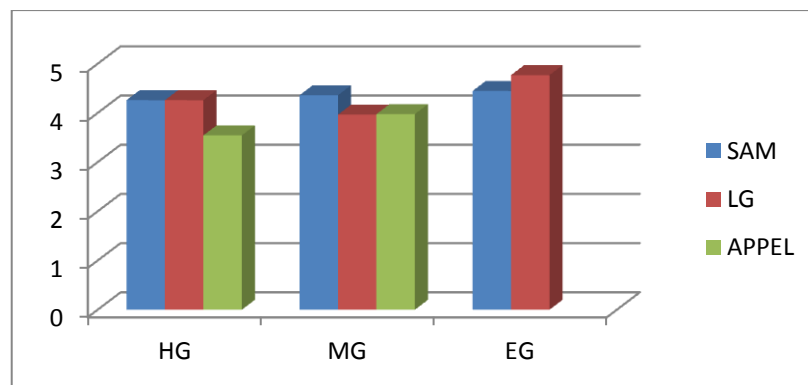


Figure 3.31 La moyenne des notes des Smartphones selon le site Amazon.ca :

La figure représente Les notes des Smartphone en fixant le site web amazon.ca, cette figure se compose de 3 parties :

1^{ère} partie : l'histogramme des notes des Smartphone selon la haute gamme, cette partie montre que les Smartphone de marque SAMSUNG et LG ont les valeurs moyennes des notes les plus élevés par rapport aux produits APPLE

2^{ème} partie : C'est une illustration des notes des Smartphone de la moyenne gamme. Dans cette partie on trouve que les Smartphones SAMSUNG sont notés comme les meilleurs produits par rapport aux autres marques LG et APPLE qui donne presque la même tendance de valeur des notes.

3^{ème} partie : cette partie de figure vient a montré des Smartphones entrée de gamme. Dans cette catégorie des Smartphones, on ne trouve pas la marque APPLLE par contre on remarque que les Smartphones LG ont une moyenne du note plus grand que la marque SAMSUNG.

d. Analyse des nombres de vue

Les valeurs des nombres de vue à prendre sont des valeurs moyennes

	HG	MG	EG
SAM	35,13	2,1	2,25
LG	40,6	3,96	35,99
APPL	16,05	17,94	

Tableau 3-18 Les valeurs moyennes des nombres de vue des Smartphones selon le site Amazon.ca

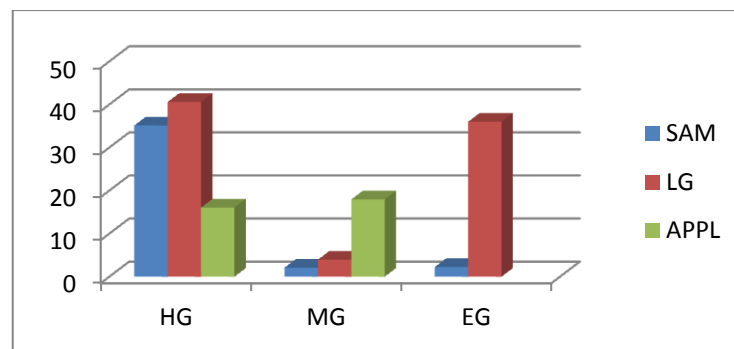


Figure 3.32 La moyenne des nombres de vue Smartphones selon le site Amazon.ca

La figure est divisée en 3 parties :

1ère partie : est La visualisation de la moyenne des vues de la haute gamme. Elle montre dans ce cas que les Smartphones LG ont le maximum de vue puis en classe les Smartphones SAMSUNG avec moyen de vue assez grand et par la suite on a les Smartphones de la marque APPLE.

2^{ème} partie : est la représentation des nombres de vue des Smartphone de la moyenne gamme. Dans cette figure on remarque que les produits APPLE sont les plus vue par rapport aux produits de la marque LG et la marque SAUMSUNG qui est presque nulle.

3^{ème} partie : représente les Smartphones entrée de gamme. Les Smartphones APPLE n'ont pas toujours des produits dans cette catégorie. En outre, on remarque que les Smartphones LG présente une moyenne de vue grande dans ce site et de la même façon les Smartphones SAMSUNG ont une moyenne de vue presque nulle.

3.6.2. Analyse des données selon les gammes

L'analyse selon les gammes se fait en comparant les données concernant le prix, la note et le nombre de vue de chaque produit de chaque marque dans les différents gammes (haut, moyenne, entrée de gamme) et ça nous permet par la suite de visualiser les résultats sous forme des histogrammes afin d'arriver à la prédiction.

A. Selon le haute de gamme

L'analyse se fait en fixant le la gamme(Haute) puis en fixe la marque des produits (Apple, Samsung, LG) et la catégorie (prix, note, nombre de vue) et en fait l'analyse en variant les sites (Amzon.com, Amazon.co.uk, Amazon.fr, Amazon.de, Amazon.es, Amazon.ca).

a. Analyse des prix

Les valeurs des prix sont des valeurs moyennes

	amazon.com	amazon.fr	amazon.de	amazon.ca	amazon.es	amazon.co.uk
SAMSUNG	566,07	589,39	568,76	708,82	583,774	502,92
LG	535,61	520	230,33	440,3	450,3	406,39
APPELE	582,94	623,41	854,66	594,40	594,40	492,05

Tableau 3-19 Les valeurs moyennes des prix des Smartphones selon la haute gamme

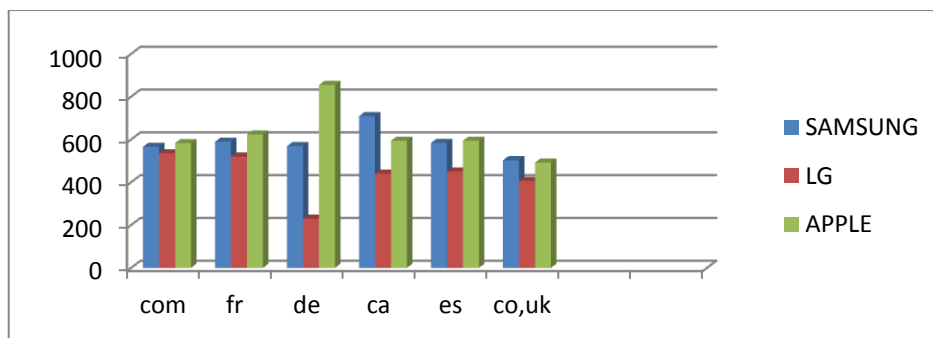


Figure 3.33 : La moyenne des prix des Smartphones selon la haute gamme

La figure représente Les prix des Smartphone en fixant la gamme haute. La figure se compose de 6 parties :

1ère partie : l'histogramme de moyenne des prix selon le site Amazon.com La figure montre que les produits APPLE ont une tendance plus grande en terme de prix par rapport aux autres produits de SAMSUNG et LG

2ème partie : La figure présente la moyenne des prix selon le site Amazon.fr. Dans cette partie de figure les prix des Smartphones APPLE sont toujours les plus élevés par rapport aux autres marques SAMSUNG et LG qui présentent aussi des valeurs des moyennes importantes.

3ème partie : Cette partie ou il est présenté les prix moyennes selon le site Amazon.de, les produits APPLE éteindre une valeur maximale des prix de presque 900, on outre les valeurs de la marque SAMSUNG sont assez grande par rapport aux produits LG.

4ème partie : Cette des parties des moyennes des prix selon le site Amazon.ca montre qu'il ya une déférence dans ce cas ou on trouve que les prix moyennes des Smartphone SAMSUNG sont plus élevé de celle de APPLE et LG

5ème partie : La figure des prix moyenne selon le site Amazon.es présente les marques SAMSUNG et APPLE sont couteuses contrairement aux produits LG qui sont un peu moins chers.

6^{ème} partie : La classification des prix moyenne selon le site Amazon.co.uk nous montre que les Smartphones SAMSUNG sont plus chers que les Smartphones APPLE ainsi que les Smartphones LG qui présentent une valeur un peu plus petite.

b. Analyse des notes

Les valeurs des notes sont des valeurs moyennes

	Amazon.com	Amazon.fr	Amazon.de	Amazon.ca	Amazon.es	Amazon.co.uk
SAMSUNG	4,11	3,92	3,8	4,26	4,03	4,43
LG	4,2	4,2	3,71	4,26	3,82	4,32
APPLE	4,10	3,82	3,84	3,54	4,06	4,06

Tableau 3-20 Les valeurs moyennes des notes des Smartphones selon la haute gamme

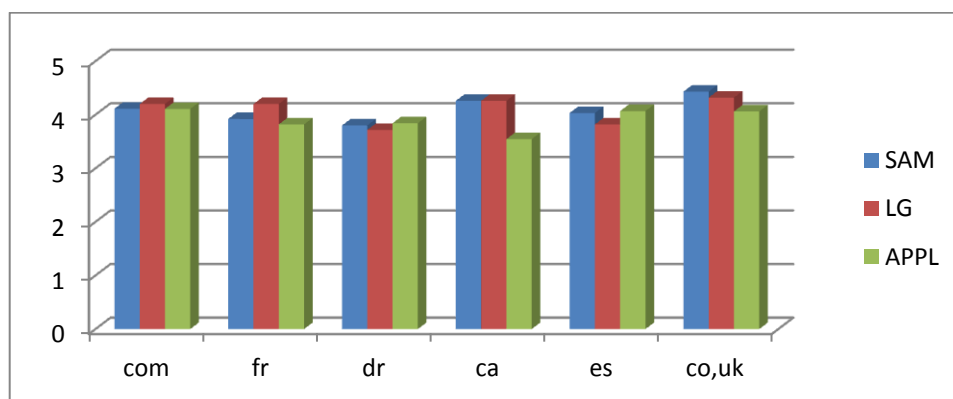


Figure 3.34: La moyenne des notes Smartphones selon la haute gamme

La figure est divisée en 6 parties :

1ère partie : la partie de la moyenne des notes selon le site Amazon.com montre que La marque LG a la tendance plus grande en termes de note par rapport aux autres produits de SAMSUNG et LG qui sont presque proches

2^{ème} partie : La visualisation présente la moyenne des notes selon le site Amazon.fr. Dans cette partie de figure les notes des Smartphones APPLE sont un peu plus grand que les autres marques SAMSUNG et LG.

3ème partie : Cette partie ou il est présenté les prix moyennes selon le site Amazon.de, les produits SAMSUNG sont notés comme les meilleurs produits comparant aux deux autres marques LG et APPLE. Cette dernière prend une valeur moyenne plus petite que celle de la marque LG.

4ème partie : Cette partie des moyennes des prix selon le site Amazon.ca montre que les Smartphones APPLE donne une valeur de notes plus élevés que la marque SAMSUNG qui a aussi une valeur importante par rapport à la marque LG.

5ème partie : La figure des prix moyenne selon le site Amazon.es présente presque la même tendance pour les produits APPLE et SAMSUNG avec des valeurs moyennes plus grandes que celle du LG.

6ème partie : La classification des prix moyenne selon le site Amazon.co.uk nous montre dans ce cas que les Smartphones LG ont les valeurs moyenne des notes les plus grandes comparant aux SAMSUNG et APPLE qui sont presque les mêmes.

c. Analyse des nombres des vue

Les valeurs des notes sont des valeurs moyennes

	Amazon.com	Amazon.fr	Amazon.de	Amazon.ca	Amazon.es	Amazon.co.uk
SAMSUNG	130,3	126,66	426,92	35,13	99,20	127,68
LG	110,6	21,44	209,22	40,6	47	137,25
APPLE	125,27	124,55	244,22	16,05	74,9375	84,84

Tableau 3-21 Les valeurs moyennes des notes des Smartphones selon la haute gamme

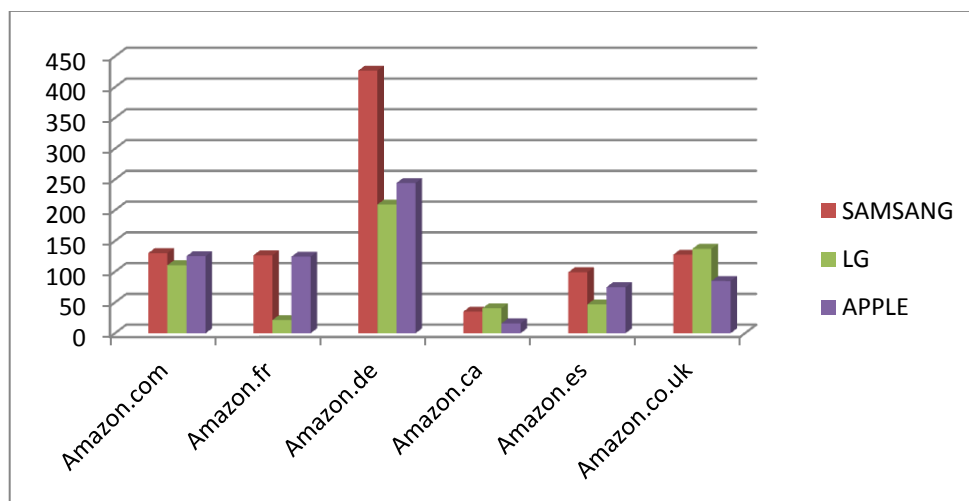


Figure 3.35 La moyenne des nombres de vue Smartphones selon la haute gamme

La figure représente Les nombre de vue des Smartphone en fixant la gamme haute. La figure toujours se compose de 6 parties :

1ère partie : la figure des moyennes nombre de vue selon le site Amazon.com montre que les produits SAMSUNG sont les plus vue avec une petite différence des produits APPLE puis en classe la marque LG avec une valeur très proches des deux premiers.

2ème partie : partie des moyennes de nombre de vue selon le site Amazon.fr. Dans cette partie de figure on note que les Smartphones SAMSUNG ont un maximum moins de vue par rapport aux Smartphones APPLE et LG qui ont aussi des valeurs moyennes importantes.

3ème partie : Cette partie ou il est présenté les nombres de vue moyennes selon le site Amazon.de, on remarque que les produits SAMASUNG et APLLE ont des valeurs moyennes de vue très proches et on trouve aussi que les produits LG présente aussi une moyenne de vue remarquable.

4ème partie : Cette des parties des moyennes des nombre de vue selon le site Amazon.ca vient a montré la moyenne des vues qui est importante pour les Smartphones APPLE et SAMSUNG et puis en classe les Smartphones LG qui ont aussi des valeurs assez grandes.

5ème partie : L'histogramme des moyennes de vue selon le site Amazon.es présente nous montres que les produits APPLE sont vus par la grande majorité qui visite le site Amazon.es et les autres produits SAMSUNG et LG sont classés dans le même niveaux en terme de moyenne de vue

6ème partie : La classification des moyennes de vue selon le site Amazon.co.uk nous montre toujours que les Smartphones APPLE sont les plus vue par rapport aux autres marques puis en notes les produits LG et par la suite les produits SAMSUNG.

B. Selon la moyenne de gamme

L'analyse se fait en fixant le la gamme(Moyenne) puis en fixe la marque des produits (Apple, Samsung, LG) et la catégorie (prix, note, nombre de vue) et en fait l'analyse en variant les sites (Amzon.com, Amazon.co.uk, Amazon.fr, Amazon.de, Amazon.es, Amazon.ca).

a. Analyse des prix

Les valeurs des prix sont des valeurs moyennes

	amazon.com	amazon.fr	amazon.de	amazon.ca	amazon.es	amazon.co.uk
SAMSUNG	314,70	329,96	266,68	468,85	273,61	232,69
LG	486,83	448,37	142,97	402,19	282	313,26
APPLE	406,25	444,50	250,30	500,81	408,43	352,61

Tableau 3-22 Les valeurs moyennes des prix des Smartphones selon la moyenne gamme

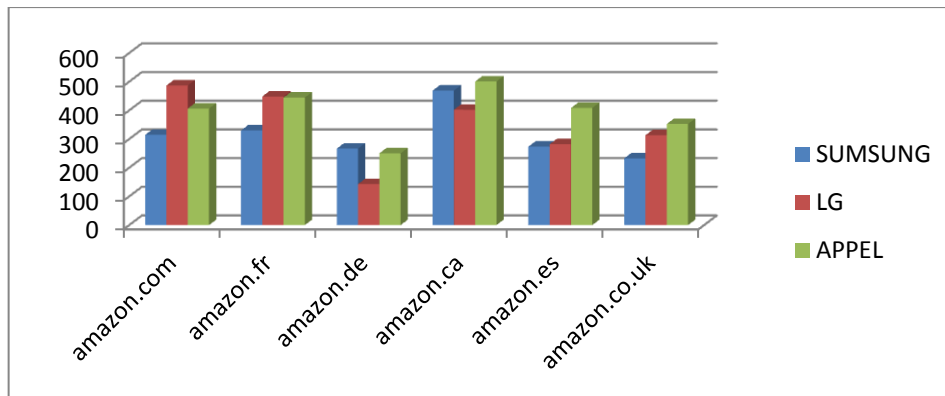


Figure 3.36 La moyenne des prix des Smartphones selon la moyenne gamme

La figure se compose de 6 parties :

1ère partie : La visualisation de moyenne des prix selon le site Amazon.com La figure montre que les produits LG sont les plus chers par rapport aux produits APPLE qui présente aussi des prix importants par rapport aux produits.

2^{ème} partie : La figure présente la moyenne des prix selon le site Amazon.fr. Dans cette partie de figure montre que les prix des Smartphones APPLE et LG sont les plus élevés par rapport à la marque SAMSUNG et LG qui présentent aussi des valeurs importantes.

3^{ème} partie : Cette partie ou il est présenté les prix moyennes selon le site Amazon.de, les produits APPLE et SAMSUNG éteindre des valeurs des prix moyennes, en d'autres partie les produits LG présente des prix aussi des prix bons.

4^{ème} partie : Dans cette partie de site Amazon.ca on remarque que la marque APPLE est toujours la plus couteuse puis en classe la marque SAMSUNG avec une petite différence de prix ensuite on note la marque LG avec aussi des prix importants.

5^{ème} partie : Dans le site Amazon.es les Smartphones APPLE sont toujours les plus chers avec des prix élevés par rapport aux ix des autres marques SAMSUNG et LG qui présente presque la moyenne des prix

6^{ème} partie : les prix moyenne des produits APPLE sont toujours les plus élevés selon le site Amazon.co.uk ensuite en note les prix des Smartphones LG ensuite SMASUNG.

b. Analyse des notes

Les valeurs des notes sont des valeurs moyennes

	Amazon.com	Amazon.fr	Amazon.de	Amazon.ca	Amazon.es	Amazon.co.uk
SAMSUNG	4,05	3,55	4,31	4,36	4,20	4,48
LG	4,27	4,09	4,45	3,966	4,2	4,43
APPLE	4,13	3,77	4	3,97	4,05	4,01

Tableau 3-23 Les valeurs moyennes des notes des Smartphones selon la moyenne gamme

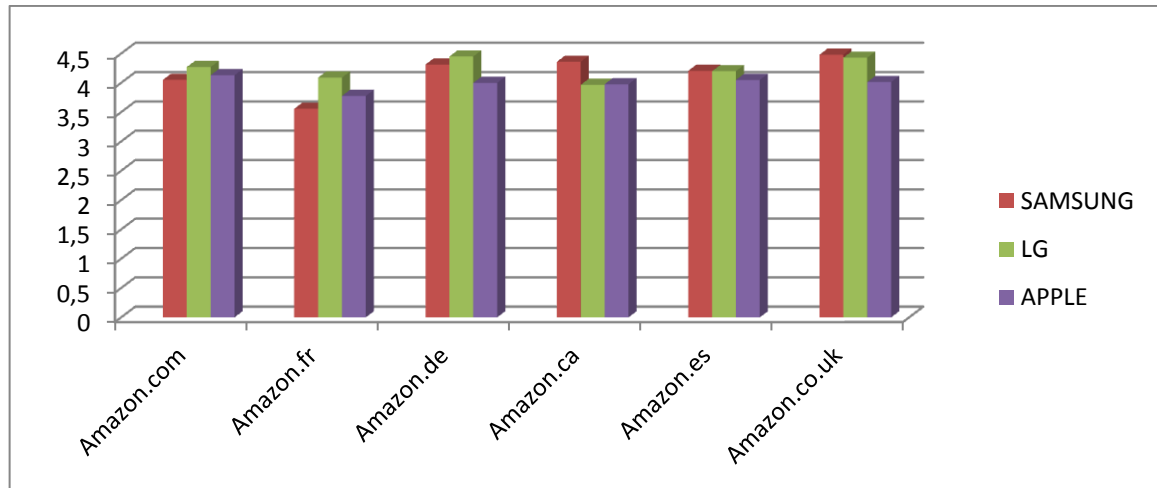


Figure 3.37 La moyenne des notes des Smartphones selon la moyenne gamme

- La figure est composée de 6 parties :

1ère partie : la partie de la moyenne des notes selon le site Amazon.com montre que La marque LG a la tendance la plus grande en terme de note par rapport aux autres produits de SAMSUNG et LG qui sont presque proches.

2ème partie : La visualisation présente la moyenne des notes selon le site Amazon.fr. Dans cette partie de figure les notes des Smartphones sont toujours les plus grands sont que les autres marques SAMSUNG et LG qui présente aussi des moyennes des notes importantes.

3ème partie : Selon le site Amazon.de,on marque toujours que les produits LG sont notés comme les meilleurs produits comparant aux deux autres marques SAMSUNG et APPLE qui ont aussi des très bonne notes .

4ème partie : Dans cette partie et selon le site Amazon.ca la figure nous montre que les Smartphones SAMSUNG donne la valeur de notes la plus élevés que les deux marque APPLE et LG qui ont la même valeur moyenne des notes.

5ème partie : Cette partie vient a montré les valeur moyenne de notes selon le site Amazon.es on trouve que les deux marques SAMNSUNG et LG présente la même valeur moyenne de notes et

elles sont notés comme les meilleurs produits par rapport à la marque APLE qui présente une valeur plus petite.

6ème partie : La classification des valeurs moyenne de notes selon le site Amazon.co.uk nous montre dans ce cas que les Smartphones SAMSUNG donnent une valeur un peu plus élevée que les Smartphones LG et APPLE aussi mais elles toujours des valeurs importantes.

c. Analyse des nombre de vue

Les valeurs de nombre de vue sont des valeurs moyennes

	Amazon.com	Amazon.fr	Amazon.de	Amazon.ca	Amazon.es	Amazon.co.uk
SAMSUNG	17,22	121,42	240,63	2,1	34,35	21
LG	188,88	33,3	250,5	3,96	241,63	52,5
APPLE	74,73	97,53	354,86	17,94	42,2	84,57

Tableau 3-24 Les valeurs moyennes de nombre de vue des Smartphones selon la moyenne gamme

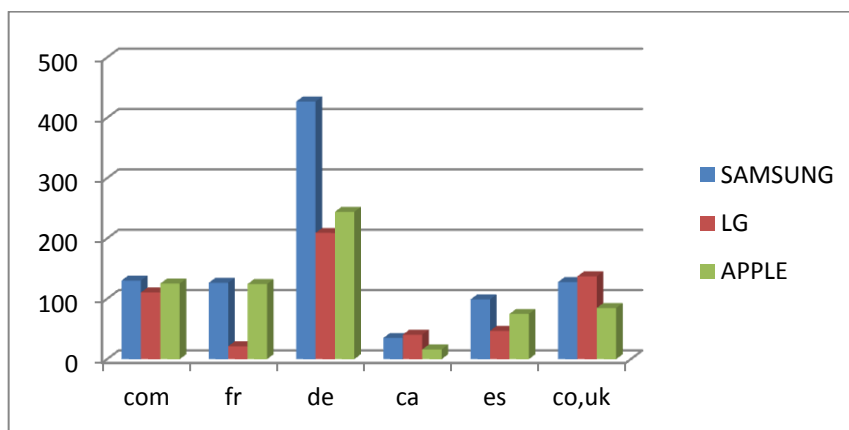


Figure 3.38 La moyenne des nombres de vue des Smartphones selon la moyenne gamme

La figure toujours se compose de 6 parties :

1ère partie : la figure des moyennes nombre de vue selon le site Amazon.com montre que les produits SAMSUNG sont les plus vue avec une petite différence des produits APPLE puis en classe la marque LG avec une valeur très proches des deux premiers.

2ème partie : Dans cette partie de figure selon le site Amazon.fr on remarque que les deux marque SAMSUNG et APLE ont presque les mêmes valeurs moyennes de vue et qui sont importantes par rapport aux Smartphones LG.

3^{ème} partie : partie des moyennes de nombre de vue selon le site Amazon.fr. Dans cette partie de figure on note que les Smartphones SAMSUNG ont un maximum de vue par rapport aux Smartphones APPLE et LG qui ont aussi des valeurs moyennes importantes.

4^{ème} partie : Cette partie ou il est présenté les nombres de vue moyennes selon le site Amazon.de, on remarque que les valeurs moyennes des vues des trois marques sont petites selon le site de

5^{ème} partie : L'histogramme des moyennes de vue selon le site Amazon.es nous montres que les produits SAMSUNG sont vus par la majorité qui visite le site Amazon.ca puis on classe les Smartphones APPLE une valeur remarquable et en place les produits LG par la suite.

6^{ème} partie : La classification des moyennes de vue selon le site Amazon.co.uk vient a montré que les Smartphones LG ont plus vue par rapport aux autres marques puis en notes les produits SAMSUNG et par la suite les produits APPLE.

C. Selon l'entrée de gamme

L'analyse se fait en fixant le la gamme(Entrée) puis en fixe la marque des produits (Apple, Samsung, LG) et la catégorie (prix, note, nombre de vue) et en fait l'analyse en variant les sites (Amzon.com, Amazon.co.uk, Amazon.fr, Amazon.de, Amazon.es, Amazon.ca).

a. Analyse des prix

Les valeurs des prix sont des valeurs moyennes

	amazon.com	amazon.fr	amazon.de	amazon.ca	amazon.es	amazon.co.uk
SAMSUNG	196,55	165,04	183,46	292,75	167,96	150,15
LG	141,36	95,47	76,01778	140,3	41,65	105,96
APPELE	Pas de produit	Pas de produit	Pas de produit	Pas de produit	Pas de produit	Pas de produit

Tableau 3-25 Les valeurs moyennes des prix des Smartphones selon l'entrée gamme

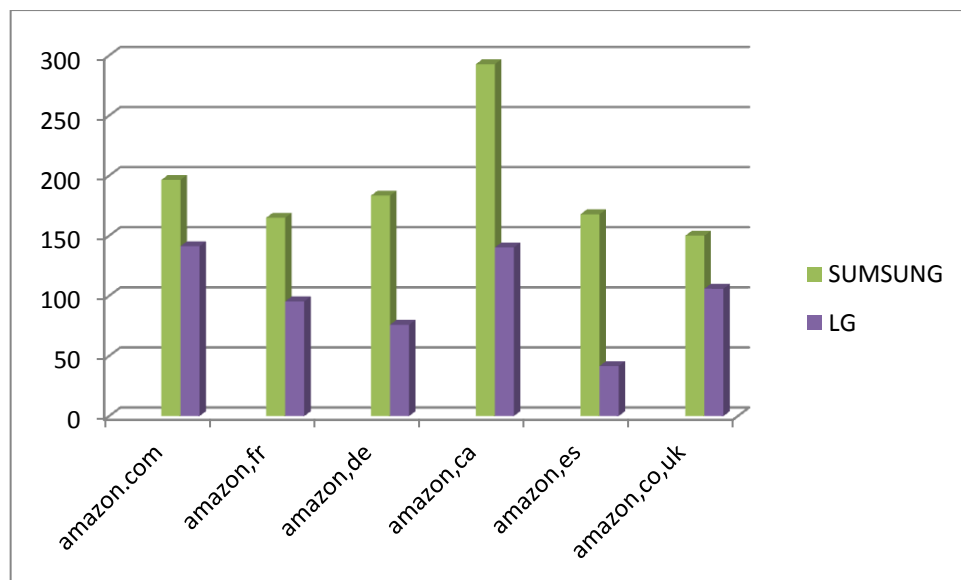


Figure 3.39 La moyenne des prix des Smartphones selon l'entrée de gamme

Pour les produits APPLE on note qu'il n'y a pas une gamme d'entrée dans les six plateformes (Site). L'historique de moyenne des prix selon nous montre que les produits SAMSUNG sont toujours plus chers selon les six sites (com,fr,de,co.uk,ca,es) avec une variation des prix moyennes de la marque LG selon chaque site.

b. Analyse des notes

Les valeurs des notes sont des valeurs moyennes

	Amazon.com	Amazon.fr	Amazon.de	Amazon.ca	Amazon.es	Amazon.co.uk
SAMSUNG	4,09	3,87	3,8375	4,44	4,25	4,46
LG	3,54	3,78	4,26	4,77	4,11	4,32
APPLE	Pas de produit	Pas de produit	Pas de produit	Pas de produit	Pas de produit	Pas de produit

Tableau 3-26 Les valeurs moyennes des notes des Smartphones selon l'entrée gamme

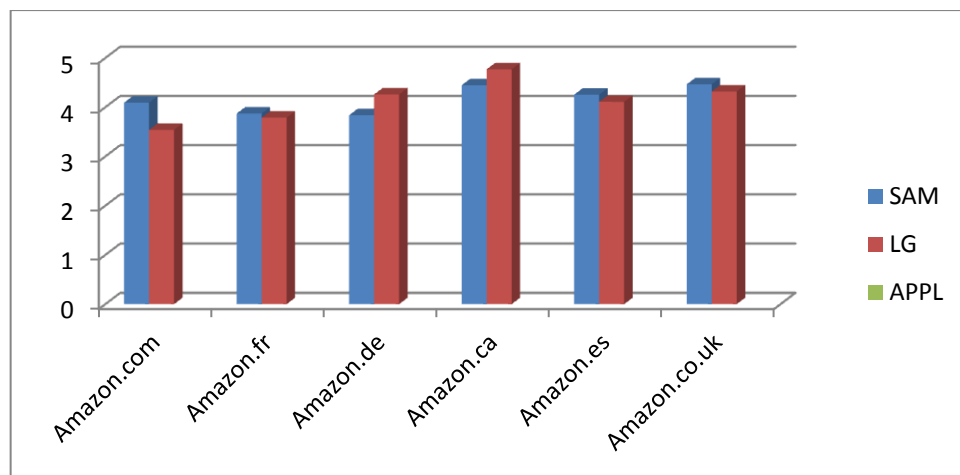


Figure 3.40 La moyenne des notes des Smartphones selon l'entrée de gamme

Pour les produits APPLE on note qu'il n'y a pas une gamme d'entrée dans les six plateformes (Site)

1ère partie : la partie de la moyenne des notes selon le site Amazon.com montre que la marque SAMSUNG a la tendance la plus grande en terme de note par rapport aux autres produits de LG qui

2^{ème} partie : La visualisation présente moyenne des notes selon le site Amazon.fr. Dans cette partie des valeurs moyennes de notes des Smartphones APPLE sont un peu plus grand que les autres produits de la marque SAMSUNG.

3^{ème} partie : Cette partie ou il est présenté les prix moyennes selon le site Amazon.de, les produits LG sont notés comme les meilleurs produits comparant aux Smartphones de la marque SMASUNG.

4^{ème} partie : Dans cette partie des moyennes des prix selon le site Amazon.ca on trouve que les Smartphones donnent toujours une valeur de notes plus élevés que la marque SAMSUNG qui a aussi une valeur importante.

5^{ème} partie : La figure des prix moyenne selon le site Amazon.es présente presque la même tendance pour les produits des deux marques LG et SAMSUNG.

6^{ème} partie : La classification des prix moyenne selon le site Amazon.co.uk nous montre dans ce cas que les Smartphones SAMSUNG ont les valeurs moyennes des notes les plus grandes comparant aux produits de la marque SAMSUNG.

c. Analyse des nombres de vue

Les valeurs des nombres de vue sont des valeurs moyennes

	Amazon.com	Amazon.fr	Amazon.de	Amazon.ca	Amazon.es	Amazon.co.uk
SAMSUNG	148,48	253,58	312,37	2,25	31,72	94,92
LG	30,625	8,4375	19	55,72	21,11	33,4
APPLE	Pas de produit	Pas de produit	Pas de produit	Pas de produit	Pas de produit	Pas de produit

Tableau 3-27 Les valeurs moyennes de nombre de vue des Smartphones selon l'entrée gamme

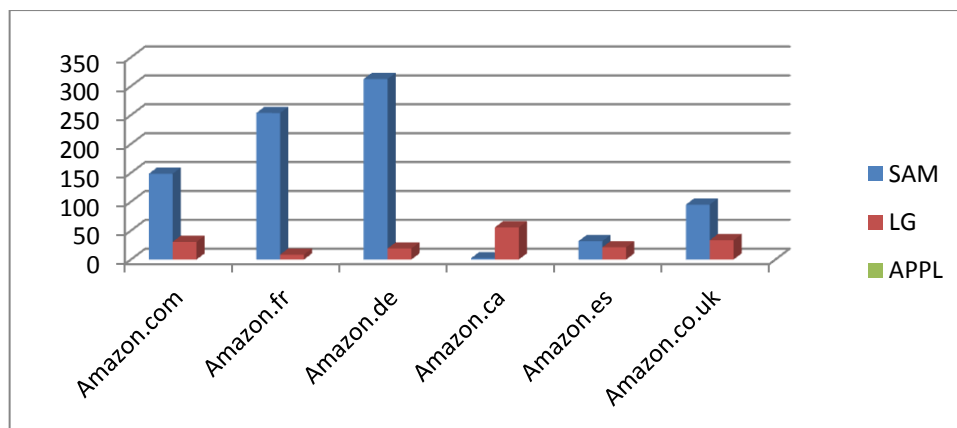


Figure 3.41 La moyenne des nombres de vue des Smartphones selon l'entrée de gamme

Pour les produits APPLE on note qu'il n'y a pas une gamme d'entrée dans les six plateformes (Site).

La figure est divisée en 6 parties :

1ère partie : Dans cette figure selon le site Amazon.com on trouve que les produits SAMSUNG sont les plus comparant aux produits de la marque LG qui ont une moyenne de vue négligeable.

2ème partie : L'histogramme selon le site Amazon.fr montre que la marque SAMSUNG a toujours les la moyennes des vues la plus élevés par rapport aux produits LG qui ont presque une valeur nulle.

3ème partie : partie des moyennes de nombre de vue selon le site Amazon.de. Dans cette partie de figure on note que les Smartphones SAMSUNG ont un maximum de vue par rapport aux Smartphones LG ont des valeurs moyennes très petites.

4ème partie : Cette partie ou il est présenté les nombres de vue moyennes selon le site Amazon.de, on remarque les Smartphones LG sont les plus vue que SAMSUNG par les canadiens.

5ème partie : L'histogramme des moyennes de vue selon le site Amazon.es nous montres que les produits SAMSUNG et LG ont des valeurs moyennes baisse dans ce site par rapport aux autre sites.

6^{ème} partie : La classification des moyennes de vue selon le site Amazon.co.uk vient à montré que les Smartphones SAMSUNG sont plus vus par rapport aux autres produits de marque LG.

3.6.3. Analyse des données selon les marques

L'analyse selon les sites se fait en comparons les données concernant le prix, la note et le nombre de vue de chaque produit de chaque marque dans les différents 6 plateformes utilisées (Amzon.com, Amazon.co.uk, Amazon.fr, Amazon.de, Amazon.es, Amazon.ca) et ça nous permet par la suite de visualiser les résultats sous forme des histogrammes afin d'arriver à la prédiction.

A. Selon la marque SAMSUNG

L'analyse se fait fixant la marque des produits (Apple, Samsung, LG) et la catégorie (prix, note, nombre de vue) et en fait l'analyse en variant les sites (com,fr,co.uk,de,es,ca) .

a. Analyse des prix

	amazon.com	amazon.fr	amazon.de	amazon.ca	amazon.es	amazon.co.uk
Haute Gamme	566,07	589,39	568,76	708,82	583,77	502,92
Moyenne Gamme	314,70	329,96	266,68	468,85	273,60	232,69
Entrée Gamme	196,55	165,04	183,46	292,75	167,96	150,15

Tableau 3-28 Les valeurs moyennes des prix des Smartphones selon la marque SAMSUNG

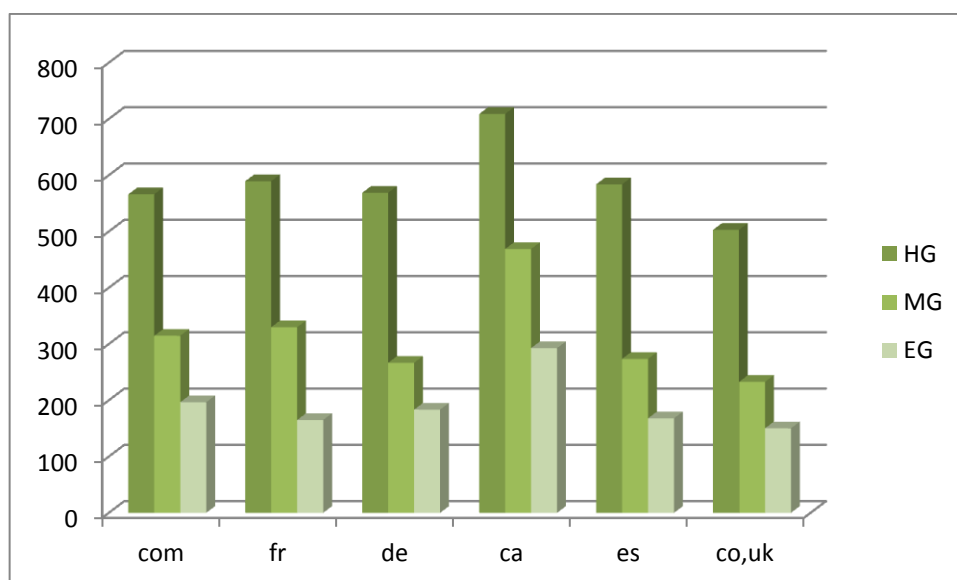


Figure 3.42 La moyenne des prix des Smartphones selon la marque SAMSUNG

L’histogramme des moyennes des prix selon la marque SAMSUNG montre que les prix de la gamme haute sont toujours plus élevés que les prix dans la catégorie des moyenne et d’entrée gamme selon les différents six plateformes(Site).

b. Analyse des notes

	Amazon.com	Amazon.fr	Amazon.de	Amazon.ca	Amazon.es	Amazon.co.uk
Haute Gamme	4,11	3,92	3,8	4,26	4,03	4,43
Moyenne Gamme	4,05	3,55	4,31	4,36	4,20	4,48
Entrée Gamme	4,09	3,87	3,8375	4,44	4,25	4,46

Tableau 3-29 Les valeurs moyennes des notes des Smartphones selon la marque SAMSUNG

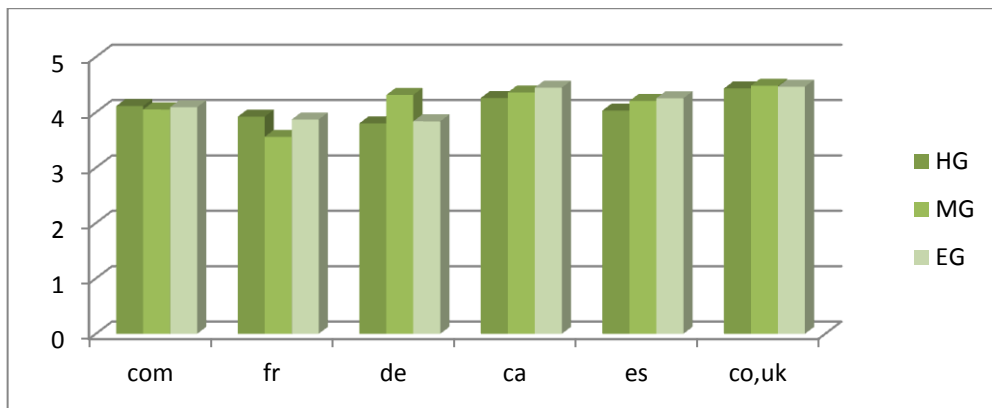


Figure 3.43 La moyenne des notes des Smartphones selon la marque SAMSUNG

La figure est divisée en 6 parties :

1^{ère} partie : Dans cette figure selon le site Amazon.com pour la marque SAMSUNG on trouve que les moyenne des notes sont presque les même dans les trois gammes (haute, moyenne, entrée).

2^{ème} partie : L’histogramme selon le site Amazon.fr montre que la marque SAMSUNG a des valeurs moyenne des notes assez élevées dans la haute et l’entrés gamme par rapport à la moyenne gamme.

3^{ème} partie : partie des moyennes de nombre de vue selon le site Amazon.de. Dans cette partie on note que la figure des Smartphones SAMSUNG a la même tendance dans l’entrée et la haute gamme comparant de la moyenne qui a une tendance plus grande et donc des moyennes des notes élevées.

4^{ème} partie : Les trois gammes selon le site Amason.ca pour la marque SAMSUNG ont presque les même notes e avec des valeurs importantes.

5^{ème} partie : Selon le site Amazon.es, on remarque que c'est le même de site Amazon.ca sauf qu'il ya un abaissement des valeurs dans les trois gammes.

6^{ème} partie : La classification des moyennes des notes selon le site Amazon.co.uk vient a montré que les Smartphones SAMSUNG sont notés de mêmes façons pour les trois gammes.

c. Analyse des nombres de vue

	Amazon.com	Amazon.fr	Amazon.de	Amazon.ca	Amazon.es	Amazon.co.uk
Haute Gamme	51,09	126,66	426,92	35,13	99,20	127,68
Moyenne Gamme	17,22	121,42	240,63	2,1	34,35	21
Entrée Gamme	148,48	253,58	312,37	2,25	31,72	94,92

Tableau 3-30 : Les valeurs moyennes de nombre de vue des Smartphones selon la marque SAMSUNG

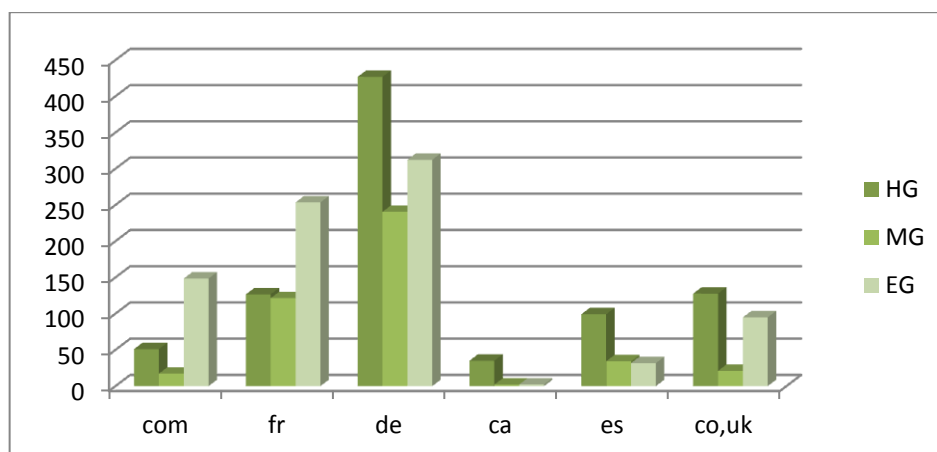


Figure 3.44 La moyenne des nombres de vue des Smartphones selon la marque SAMSUNG

La figure est divisée en 6 parties :

1ère partie : Dans cette figure selon le site Amazon.com on trouve que les Smartphones SAMSUNG on trouve que les produits d'entrée gamme sont les plus vue par rapport aux autres gammes puis on classe les produits de la haute gamme ensuite la moyenne gamme avec une valeur baisse.

2^{ème} partie : Les valeurs moyenne des vues selon le site Amazon pour la marque SAMSUNG sont presque les même pour la haute et la moyenne gamme par contre l'entrée de gamme présente une grande moyenne de vue

3^{ème} partie : partie des moyennes de nombre de vue selon le site Amazon.de. Dans cette partie de figure on note que les Smartphones SAMSUNG ont des moyennes des vues élevés dans les tris gamme avec une valeur maximale pour la haute gamme et des valeur importantes pour l'entrées et la moyenne gamme.

4^{ème} partie : Cette partie ou il est présenté les nombres de vue moyennes selon le site Amazon.de, on remarque la moyenne et l'entrée de gamme des Smartphones SAMSUNG sont presque n'ont pas vue par la population canadienne par contre on remarque que la haute gamme présente une valeur moyenne des vues baisse.

5^{ème} partie : Les vue des produits SAMSUNG selon le site Amazon.es présente des valeurs proche et faibles pour la moyenne et l'entrée gamme de la marque mais en trouve pour la gamme haute une moyenne des vue plus grandes que celles des autres.

6^{ème} partie : La classification des moyennes de vue selon le site Amazon.co.uk vient a montré que les Smartphones SAMSUNG sont plus vus parlons de la gamme haute et aussi l'entrée de gamme par rapport à la moyenne gamme qui donne une moyenne des vues très faible.

B. Selon la marque LG

L'analyse se fait en fixant le site Amazon.com puis en fixe la marque des produits (Apple, Samsung, LG) et la catégorie (prix, note, nombre de vue) et en fait l'analyse en variant les gamme (haute, moyenne, entrée de gamme)

a. Analyse des prix

	amazon.com	amazon.fr	amazon.de	amazon.ca	amazon.es	amazon.co.uk
Haute Gamme	535,61	520	230,33	440,3	450,3	406,39
Moyenne Gamme	309,6	324	230,66	402,19	240,3	220,3
Entrée Gamme	141,36	95,47	76,01	140,3	41,65	105,96

Tableau 3-31 Les valeurs moyennes des prix des Smartphones selon la marque LG

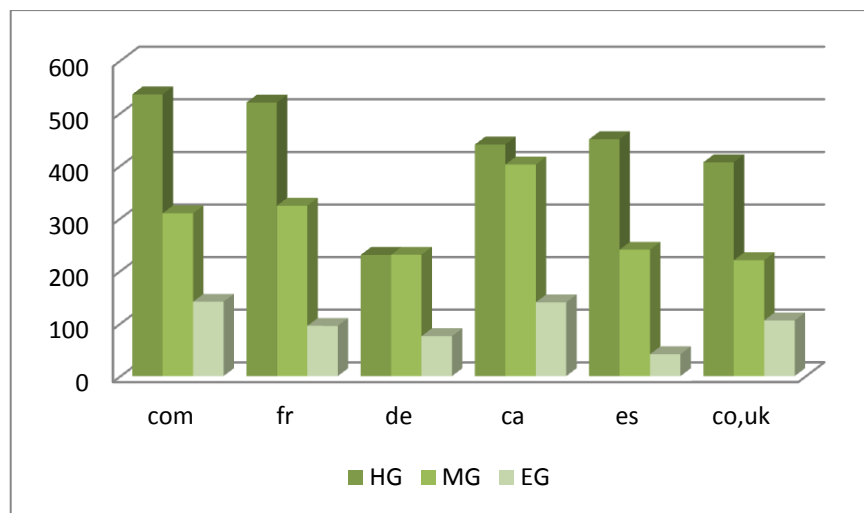


Figure 3.45 La moyenne des prix des Smartphones selon la marque LG

L’histogramme des moyennes des prix selon la marque LG vient a montré que la haute gamme est toujours la plus cher par rapport aux autres gamme selon les différents plateformes(Site) sauf pour le site Amazon.es ou on trouve qu’il y a un équilibre entre la haute et la moyenne gamme dans la valeur moyenne des prix.

b. Analyse des notes

	Amazon.com	Amazon.fr	Amazon.de	Amazon.ca	Amazon.es	Amazon.co.uk
Haute Gamme	4,2	4,2	3,71	4,26	3,82	4,32
Moyenne Gamme	4,27	4,09	4,45	3,966	4,2	4,43
Entrée Gamme	3,54	3,7875	4,26	4,77	4,11	4,32

Tableau 3-32 Les valeurs moyennes des notes des Smartphones selon la marque LG

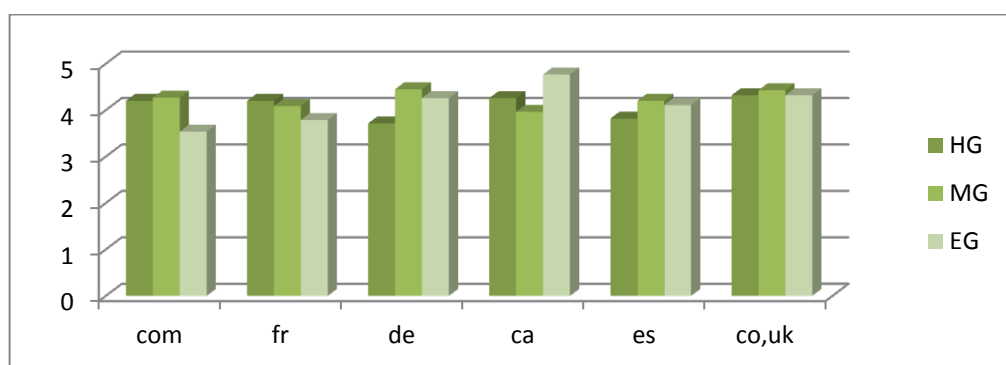


Figure 3.46 La moyenne des notes des Smartphones selon la marque LG

La figure est divisée en 6 parties :

1ère partie : Selon le site Amazon.com pour la marque LG on trouve que la haute gamme et la moyenne gamme présente presque la même valeur moyenne des notes qui sont plus grande de la valeur moyenne de l’entrée gamme

2^{ème} partie : L’histogramme selon le site Amazon.fr montre que la marque LG a des valeurs moyenne des notes assez élevées dans la haute et l’entrés gamme par rapport à la moyenne gamme.

3^{ème} partie : Dans cette partie on note que les notes pour la moyenne et l’entrée de gamme sont plus grandes de celle de la haute gamme concernant la marque LG.

4^{ème} partie : Les trois gammes selon le site Amason.ca pour la marque LG présentent des valeurs importantes en notant que l’entrée de gamme prend la meilleure valeur de note.

5^{ème} partie : Selon le site Amazon.es, on remarque que les valeurs des notes de d’entrés et de la moyenne gamme sont très proche et aussi plus élevées par rapport à la moyenne gamme toujours pour la marque LG

6^{ème} partie : La classification des moyennes des notes selon le site Amazon.co.uk vient a montré que les Smartphones LG ont presque la même valeur moyenne des notes pour les trois gammes.

c. Analyse des nombres de vue

	Amazon.com	Amazon.fr	Amazon.de	Amazon.ca	Amazon.es	Amazon.co.uk
Haute Gamme	110,6	21,44	209,22	40,6	47	137,25
Moyenne Gamme	50,9	33,3	250,5	3,96	55,99	52,5
Entrée Gamme	30,62	8,43	19	35,99	21,11	33,4

Tableau 3-33 Les valeurs moyennes de nombre de vue des Smartphones selon la marque LG

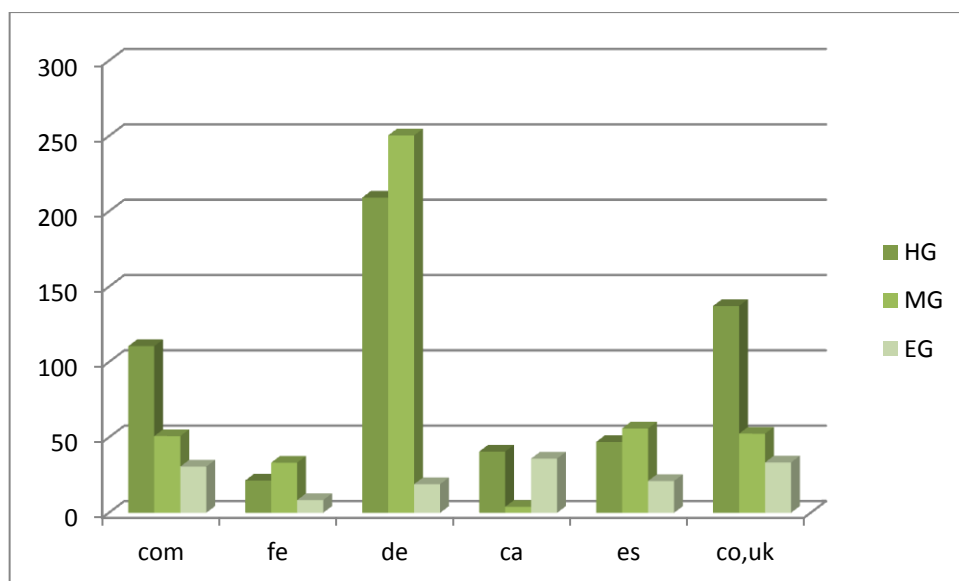


Figure 3.47 La moyenne des nombres de vue des Smartphones selon la marque LG

1ère partie : Dans cette figure selon le site Amazon.com on trouve la haute gamme de la marque LG est le plus vue par les clients puis en classe la moyenne gamme par la suite en note l'entrée de gamme

2^{ème} partie : Les valeurs moyennes des vues selon le site Amazon.fr pour la marque LG sont remarquables dans la haute gamme par rapport au moyenne gamme et d'entrée de gamme qui est très baisse

3^{ème} partie : la moyenne gamme selon le site Amazon.de pour la marque LG atteindre une valeur maximale des moyennes de vues puis en classe la haute gamme avec une valeur importante aussi mais l'entrée de gamme montre une valeur très faible comparant aux gammes précédentes

4^{ème} partie : Cette partie ou il est présenté les nombres de vue moyennes selon le site Amazon.de, la moyenne gamme présente une valeur très faible de moyenne de vue par contre la haute de gamme et l'entrés de gamme présente des valeurs assez grandes.

5^{ème} partie : Les vue des produits LG selon le site Amazon.es sont plus grandes dans la moyenne gamme par rapport a la haute gamme et l'entrés de gamme qui donne une valeur très faible

6^{ème} partie : La classification des moyennes de vue selon le site Amazon.co.uk vient a montré que les Smartphones LG sont plus vus parlons de la gamme haute et aussi la moyenne gamme par rapport à la l'entrée de gamme.

C. Selon la marque APPLE

L'analyse se fait en fixant la marque APPLE et la catégorie (prix, note, nombre de vue) et en fait l'analyse en variant les sites (com,fr,co.uk,de,es,ca).

a. Analyse des prix

	amazon.com	amazon.fr	amazon.de	amazon.ca	amazon.es	amazon.co.uk
Haute Gamme	582,94	623,41	450,6	854,66	594,40	492,05
Moyenne Gamme	406,25	444,50	304,34	500,81	408,43	352,61
Entrée Gamme	Pas de produit	Pas de produit	Pas de produit	Pas de produit	Pas de produit	Pas de produit

Tableau 3-34 Les valeurs moyennes des prix des Smartphones selon la marque APPLE

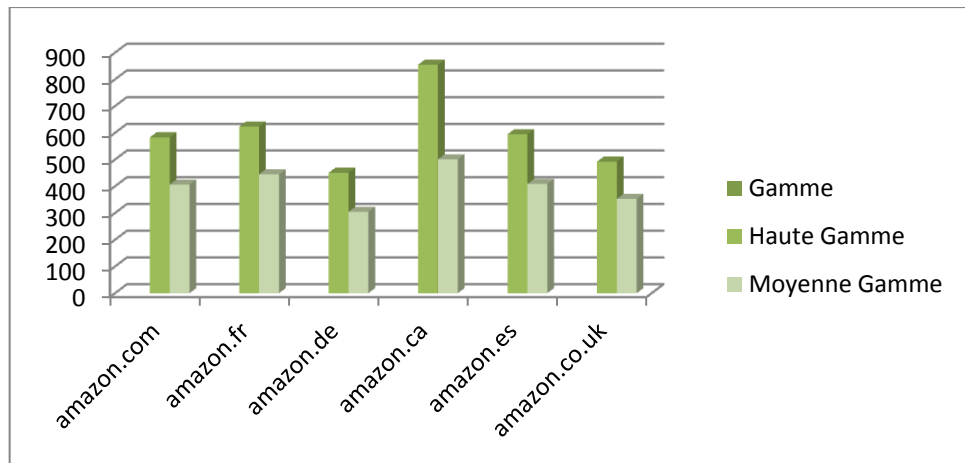


Figure 3.48 La moyenne des prix des Smartphones selon la marque APPLE

Pour les produits APPLE on note qu’il n’y a pas une gamme d’entrée dans les six plateformes (Site)

	Amazon.com	Amazon.fr	Amazon.de	Amazon.ca	Amazon.es	Amazon.co.uk
Haute Gamme	4,10	3,82	3,84	3,54	4,0	4,06
Moyenne Gamme	4,13	3,77	4	3,97	4,05	4,01
Entrée Gamme	Pas de produit	Pas de produit	Pas de produit	Pas de produit	Pas de produit	Pas de produit

L’histogramme de moyenne des prix selon nous montre que les produits APPLE sont toujours couteux selon les six sites avec une variation des prix moyens de la marque selon chaque site.

Tableau 3-35 Les valeurs moyennes des notes des Smartphones selon la marque APPLE

b. Analyse des notes

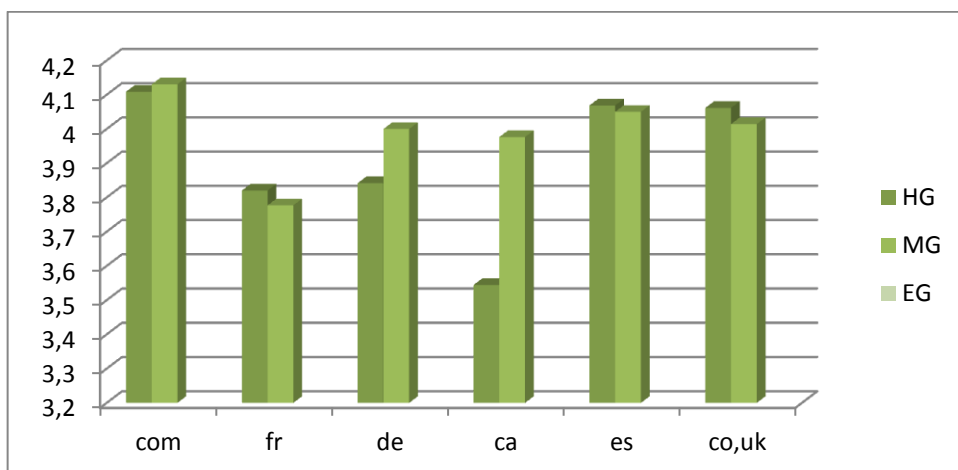


Figure 3.49 La moyenne des notes des Smartphones selon la marque APPLE

1ère partie : Selon le site Amazon.com pour la marque LG on trouve que la haute gamme et la moyenne gamme présente presque la même valeur moyenne des notes et qui sont des valeurs importantes.

2^{ème} partie : L’histogramme selon le site Amazon.fr montre que la marque LG a des valeurs moyennes des notes assez élevées dans la haute gamme par rapport à la moyenne gamme.

3^{ème} partie : Dans cette partie et selon le site Amazon .es on note que les notes pour la moyenne gamme sont plus grandes de celle de la haute gamme concernant la marque LG.

4^{ème} partie : La classification des valeurs moyennes des notes selon le site Amason.ca pour la marque LG montre que la moyenne gamme est plus grandes que la haute de gamme

5^{ème} partie : Selon le site Amazon.es, on remarque que les valeurs de la note de la haute gamme et la moyenne gamme sont très proche et aussi plus élevées mais on note que la haute gamme présente les valeurs plus grandes toujours pour la marque APPLE.

6^{ème} partie : La figure dans cette partie selon le site Amazon.co.uk pour la marque APPLE vient a montrés que la haute gamme est notée toujours comme la meilleurs avec des valeurs importantes puis en classe la moyenne gamme

c. Analyse des nombre de vue

	Amazon.com	Amazon.fr	Amazon.de	Amazon.ca	Amazon.es	Amazon.co.uk
Haute Gamme	125,27	124,55	244,22	16,05	74,93	84,84
Moyenne Gamme	74,73	97,53	354,86	17,94	42,2	84,57
Entrée Gamme	Pas de gamme	Pas de gamme	Pas de gamme	Pas de gamme	Pas de gamme	Pas de gamme

Tableau 3-36 : Les valeurs moyennes de nombre de vue des Smartphones selon la marque APPLE

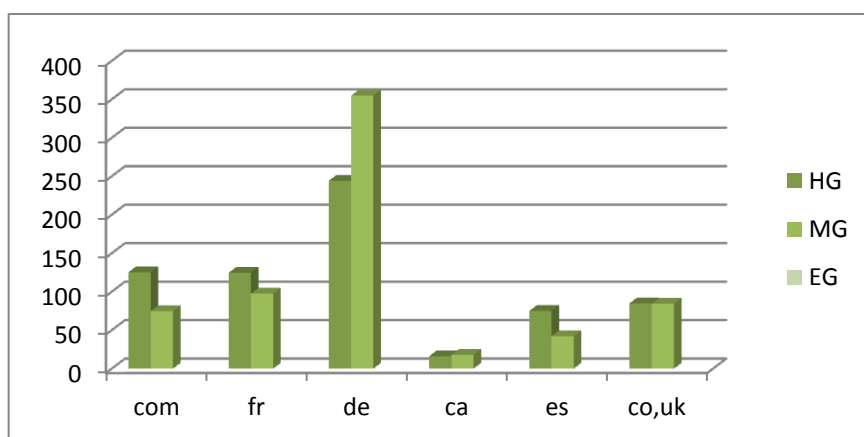


Figure 3.50 La moyenne des nombres de vue Smartphones selon la marque APPLE

1ère partie : Dans cette figure selon le site Amazon.com on trouve que la haute gamme de la marque APPLE est le plus vue par les clients ensuite on trouve la moyenne gamme.

2^{ème} partie : Les valeurs moyennes des vues selon le site Amazon.fr pour la marque APPLE sont remarquables dans la haute gamme par rapport à la moyenne gamme

3^{ème} partie : La moyenne gamme selon le site Amazon.de pour la marque APPLE atteindre une valeur maximale des moyennes de vues puis on trouve la haute gamme avec une valeur importante aussi.

4^{ème} partie : Cette partie où il est présenté les nombres de vue moyennes selon le site Amazon.de, pour la marque APPLE le nombre moyenne de vue pour les deux gammes est très faible ça nous montre que les produits APPLE ne sont pas vus par les canadiens.

5^{ème} partie : Les vues des produits APPLE selon le site Amazon.es sont plus grandes pour la haute gamme que la moyenne gamme

6^{ème} partie : La classification des moyennes de vue selon le site Amazon.co.uk montre que les Smartphones APPLE ont presque les mêmes valeurs moyennes des vues pour la haute et la moyenne gamme.

3.7. CONCLUSION

Dans ce chapitre nous avons présenté en petit exemple afin de donner une idée globale sur quelques techniques utilisées dans les bigdata, nous avons développé une application afin de faire la collecte des données à partir de six sites web que nous avons sélectionnés à partir des nombreux sites web de l'un des géants de l'internet et des nouvelles technologies et le leader du commerce électronique Amazon. Nous avons choisi les smartphones comme produit parmi les divers produits qui sont représentés sur ces sites web. Comme critères d'analyse des produits nous avons choisi le prix des produits, le nombre de vues des produits et les notes données par les clients. Dans la deuxième partie du chapitre nous avons montré l'interface de l'application développée avec des explications sur son fonctionnement, en dernière partie du chapitre nous avons présenté les résultats de la récolte des données avec des analyses et des explications de chaque situation.

CONCLUSION GENERALE

« Avec les BIG DATA, les données d'une organisation deviennent stratégiques. C'est la découverte d'une mine d'or gigantesque dont les ressources étaient jusque-là insoupçonnées. »

Gilles Grapinet, directeur général adjoint d'Atos2

Avec les BIG DATA, les grandes entreprises qui veulent faire la course en tête doivent beaucoup investir dans leurs systèmes décisionnels, et en particulier utiliser les BIG DATA pour mieux comprendre les marchés, maîtriser leurs actions et faire vivre à leurs clients les expériences qu'ils attendent.

Si pratiquement toutes les fonctions de l'entreprise peuvent être concernées par les BIG DATA, actuellement c'est surtout la fonction Marketing / Vente qui s'en préoccupe. Pour elle il s'agit principalement de trouver de nouvelles opportunités, à travers la traque des usages et des expériences des clients, de mettre en lumière des attentes, des besoins, auxquels on peut répondre avec des offres innovantes, d'exploiter au mieux les relations, de savoir positionner, pousser l'offre auprès des personnes à potentiel en exploitant au mieux les moments de contact, d'optimiser leurs investissements marketing en fonction des résultats qu'ils ont pu observer de leurs programmes antérieurs.

Tout ceci est rendu possible grâce aux démarches des BIG DATA qui permettent de collecter, stocker et faire des calculs massivement parallèles avec un traitement de données sur des profondeurs d'historiques, des séries temporelles très longues. Il s'agit aussi de mettre en place des profils et d'activer des campagnes dynamiques, grâce à des analyses multidimensionnelles, permettant d'aligner les besoins des clients et les intérêts de l'entreprise avec une mise en priorité et une personnalisation des offres...

« LE BIG DATA EST UN MEILLEURE OFFRE A FAIRE A UN CLIENT »

REFERENCES BIBLIOGRAPHIE

1. Dictionnaire HACHETTE. 2010.
2. /, **Gérard-Michel Cochard/Technologies des réseaux de communication/(en ligne)/ 2007/.** technologie_des_reseaux_de_communication.pdf. Université Virtuelle de Tunis : s.n., consulté le 16 Décembre 2014. pp. Disponible sur <http://pf-mh.uvt.rnu.tn/444/1/>.
3. **White, Colin.** /Converting Hype into value with Big Data and Analytics. / Octobre 2013/ Consulté le 16 décembre 2014.
4. **Jade AMELINE, François DESAULLE, Violette DONNEAUD.** <http://www.tpe-bigdata.fr/2/ii>. [En ligne]
5. **Group, CXP.** Le Forum du . Organisé par le CXP Group, reconnu pour son expertise, il vous permet de faire le point sur votre Systè. cabinet européen indépendant : s.n., 11 juin 2015 est un événement unique consacré aux tendances et aux évolutions du marché du logiciel.
6. **Press., Jil.** A Very Short History of Big Data . sur le site FORBES. [En ligne] 5 septembre 2013. [Citation : [Consulté le 19 novembre 2014].]
7. <http://www.atelier.net/trends/articles/nombre-de-donnees-croit-plus-vite-protection-analyse> . <http://www.atelier.net/>. [En ligne] [Publié le 17 décembre 2012.
8. **Eureka Presse.** <http://www.cnetfrance.fr/news/ces-2015-samsung-veut-un-internet-des-objetSamsung> veut un Internet des objets ouvert. CNETFrance.fr. [En ligne] CES 2015, 06 janvier mardi . [Citation : 04 01 2015.]
9. **KONE, Angeline Mastère spécialisé SI.** http://www.memoireonline.com/05/14/8890/m_Big-data-rapport-de-stage.html. s.l. : INSA LYON - , 2013.
10. **o, KerryNoone.** <https://visual.ly/community/infographic/technology/big-data-just-beginning-explode>. [En ligne] 13 03 2012 .
11. http://www.decideo.fr/De-l-interet-du-big-data-pour-le-marketing_a7204.html. **Xavier Guerin.** MapR le 15 Juillet 2014.
12. |, **Par Guillaume Serries.** <http://actualites/big-data-3-4-ou-5-v-39832210.htm>. www.zdnet.fr. [En ligne] 04 Février 2016.
13. <https://bigdatafrance.wordpress.com/2016/03/21/quoi-de-neuf-au-big-data-paris/>. <https://bigdatafrance.wordpress.com/>. [En ligne] 21 03 2016.
14. **PEAUDECERF, Marion.** http://www.bigdataparis.com/guide/2016-2017/BD_guide_16-17_BD_16119.pdf. www.bigdataparis.com. [En ligne] 2016.
15. **JOUANNOT, Olivier.** résentation Générale Big Data. France : s.n., 2013.
16. **Michel, Gerad.**
17. **Wite, Coline.** Using Big Data for Making. s.l. : IBM, july 2011.
18. <http://marketinghits.com/blog/big-data-just-beginning-to-explode/>. marketinghits.com/blog/big-data-just-beginning-to-explode. [En ligne]

في هذا العمل قمنا بالتطرق لموضوع مثير للاهتمام ومهم جدا في الآونة الأخيرة، . هو مجال Big Data ، هذه الطريقة الجديدة للنظر في البيانات والمعلومات، والتي تستخدم حاليا خاصة مع انفجار بيانات الإنترنت (الفيديو، الصور، الرسائل، والشبكات الاجتماعية، ، استخدام الهواتف الذكية على نحو متزايد، والألواح الإلكترونية والأشياء المتصلة بشكل خاص. تحدثنا في هذه المذكرة عن بنية، هيكل، عمل ومجالات تطبيق Big Data ...
الكلمات المفتاحية: Cloud Computing ، BigData ، هيكل BigData ، جمع البيانات، تحليل البيانات.

Résumé

Dans ce travail nous avons abordé un sujet très intéressant et très important actuellement et surtout dans le future, c'est le bigdata. Cette nouvelle manière de voir les données et les informations. Elle est très utilisée actuellement notamment avec l'explosion des données sur internet (vidéos, photos, messages, réseaux sociaux, ...), l'utilisation de plus en plus des smartphones, des tablettes et particulièrement des objets connectés. Nous avons parlé dans ce mémoire des bigdata de leurs conceptions, architectures, fonctionnements et applications.

Mots clé : Bigdata, Cloud computing, architecture bigdata, collecte de données, analyse de données.

Abstract

In this project we tackled a very interesting and very significant subject of the now time and especially in the future. It is the big data, this new way of seeing the data and information. It is used very much currently, in particular with the explosion of the data on Internet (videos, photos , messages, social media...), the increasingly large use of the smart phones, the tablets and particularly the connected objects. We spoke in this graduation essay about the Bigdata, their designs, architectures, operations and applications.

Key words: Bigdata, Cloud computing, big data architecture, data collect, analyze data.

ANNEXE

