



ÉCOLE CENTRALE CASABLANCA

PROJET DE GESTION DE LA PERFORMANCE FINANCIÈRE  
RAPPORT

---

# Analyse Financière des Entreprises par l'Exploration de Données et le Machine Learning

---

***Élèves :***

Mohamed Rida EL-KHALOUAN  
Molka ELLOUMI  
Fadwa HAMMOUCHE  
Emna MACHOUCH  
Nada THABET

***Encadré par:***

M. Fouad RIANE

15 avril 2025

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Présentation du jeu de données</b>	<b>3</b>
<b>3</b>	<b>Prétraitement des données</b>	<b>4</b>
3.1	Nettoyage des données . . . . .	4
3.1.1	Code correspondant et Résultats obtenus : . . . . .	4
3.2	Calcul des principaux ratios financiers - Résultats extraits . . . . .	4
3.2.1	Marge bénéficiaire nette (Net Profit Margin) . . . . .	4
3.2.2	Retour sur les actifs (ROA - Return on Assets) . . . . .	5
3.2.3	Rotation des actifs . . . . .	5
3.2.4	Interprétation des résultats . . . . .	6
3.3	Ratios non calculables faute de données . . . . .	6
3.3.1	Return on Equity (ROE - Retour sur les capitaux propres) . . . . .	6
3.3.2	Ratio d'endettement . . . . .	6
3.3.3	Ratio de liquidité générale . . . . .	7
3.4	Visualisation des corrélations entre variables financières . . . . .	7
3.4.1	Diagramme de dispersion : Profit vs. Sales par segment . . . . .	7
3.4.2	Analyse de la Marge Nette par Pays . . . . .	8
3.4.3	Analyse de l'évolution mensuelle des ventes . . . . .	9
3.4.4	Analyse des corrélations entre variables financières . . . . .	10
3.4.5	Analyse des performances moyennes par segment . . . . .	12
<b>4</b>	<b>Préparation des données et sélection des variables financières pour le clustering</b>	<b>13</b>
4.1	Chargement des données . . . . .	13
4.2	Nettoyage des noms de colonnes . . . . .	13
4.3	Sélection des variables pour l'analyse . . . . .	13
<b>5</b>	<b>Exploitation Machine Learning</b>	<b>14</b>
5.1	Clustering K-Means et visualisation avec PCA . . . . .	14
5.1.1	Normalisation des données . . . . .	14
5.1.2	Application de K-Means . . . . .	14
5.1.3	Réduction de dimensionnalité avec PCA . . . . .	14
5.1.4	Visualisation des clusters . . . . .	14
5.2	Analyse à l'aide de la méthode Random Forest . . . . .	16
5.2.1	Erreur quadratique moyenne (MSE) . . . . .	16
5.2.2	Importance des variables . . . . .	16
5.3	Analyse des résultats des modèles Ridge et Lasso . . . . .	17
5.3.1	Résultats du modèle Ridge . . . . .	18
5.3.2	Résultats du modèle Lasso . . . . .	18
5.3.3	Convergence Warning . . . . .	19
5.4	Analyse des résultats de la PCA (Analyse en Composantes Principales) . . . . .	19
5.4.1	Interprétation du graphique obtenu après la PCA . . . . .	20

<b>6</b>	<b>Interprétations économiques</b>	<b>21</b>
6.1	Facteurs Clés selon la Random Forest . . . . .	21
6.1.1	Interprétation Économique : . . . . .	21
6.2	Visualisations des Facteurs Clés . . . . .	22
6.3	Implications Stratégiques . . . . .	23
6.3.1	Analyse via l'Arbre de Décision . . . . .	23
6.4	Synthèse des profils d'entreprises . . . . .	24
6.5	Analyse du lien entre secteur d'activité et performances . . . . .	25
<b>7</b>	<b>Conclusion</b>	<b>27</b>

# 1 Introduction

L'analyse financière des entreprises est un domaine clé pour comprendre leur performance économique et leur positionnement sur le marché. Dans un environnement économique globalisé et dynamique, il est crucial de disposer d'outils permettant d'évaluer de manière précise la santé financière des entreprises. Ce projet, intitulé "Analyse Financière des Entreprises par l'Exploration de Données et le Machine Learning", vise à exploiter les données financières disponibles pour identifier les facteurs clés influençant la performance financière des entreprises cotées en bourse.

Le dataset utilisé dans ce projet inclut des indicateurs financiers clés, tels que les revenus totaux, le résultat net, la capitalisation boursière, les actifs et passifs, ainsi que des ratios financiers comme le rendement du dividende et le ratio cours/bénéfice. L'objectif de ce projet est de déterminer les variables les plus influentes sur la performance financière des entreprises et de prédire la performance future en utilisant Random Forest. Cette méthode d'apprentissage supervisé permet de classer les entreprises en fonction de leurs profils financiers et d'identifier les facteurs expliquant leur succès ou leurs difficultés.

À travers cette approche, le projet vise à fournir une analyse robuste et des prédictions précises sur la performance financière des entreprises, offrant ainsi des outils d'aide à la décision pour les investisseurs et les gestionnaires financiers.

## 2 Présentation du jeu de données

Le jeu de données utilisé dans ce projet est issu d'une base publique contenant des informations financières sur un ensemble d'entreprises de secteurs variés. Chaque ligne du dataset représente une entreprise, et chaque colonne correspond à un indicateur financier ou une caractéristique comptable.

Parmi les principales variables présentes, on peut citer :

- Le chiffre d'affaires annuel,
- Le bénéfice net,
- Le total des actifs et passifs,
- Le fonds de roulement,
- Les flux de trésorerie,
- Divers ratios financiers (rentabilité, endettement, liquidité, etc.),
- Une étiquette cible représentant la situation financière de l'entreprise, codée sous forme catégorielle (par exemple : "solvable", "insolvable").

Le jeu de données est relativement équilibré, ce qui facilite la phase d'apprentissage supervisé. Le but est de créer un modèle capable de généraliser à de nouveaux cas pour prédire avec fiabilité la catégorie financière d'une entreprise.

## 3 Prétraitement des données

Le prétraitement des données constitue une étape essentielle avant la modélisation. Les données brutes comportent souvent des anomalies, des valeurs manquantes ou des variables inutilisables en l'état. Voici les principales opérations réalisées :

### 3.1 Nettoyage des données

Le nettoyage des données manquantes a constitué une étape cruciale pour garantir la fiabilité des analyses ultérieures. Les valeurs manquantes ont été d'abord repérées grâce à une phase d'exploration qui a permis d'en évaluer la fréquence et la distribution au sein des différentes variables. En fonction du taux d'absence et du type de données concernées, plusieurs approches ont été appliquées. Pour les variables légèrement touchées, les valeurs manquantes ont été imputées à l'aide de la moyenne ou de la médiane, selon la distribution de chaque variable. En revanche, lorsque certaines variables présentaient un pourcentage élevé de valeurs absentes, elles ont été supprimées, car leur conservation aurait pu nuire à la qualité de l'analyse. Cette étape de prétraitement a permis d'obtenir un jeu de données propre, cohérent et prêt à être utilisé pour l'entraînement des modèles.

#### 3.1.1 Code correspondant et Résultats obtenus :

Le code correspondant ainsi que les résultats détaillés relatifs à la partie nettoyage des données sont disponibles dans le lien suivant :  
[Notebook Google Colab](#)

### 3.2 Calcul des principaux ratios financiers - Résultats extraits

Les résultats des ratios financiers calculés pour l'échantillon d'entreprises sont présentés ci-dessous. Ces ratios, qui incluent la marge bénéficiaire nette, le ROA (Return on Assets) et la rotation des actifs, permettent d'obtenir une vision plus précise de la rentabilité, de l'efficacité et de la gestion des ressources de ces entreprises.

#### 3.2.1 Marge bénéficiaire nette (Net Profit Margin)

La marge bénéficiaire nette mesure la proportion de chaque dollar de chiffre d'affaires qui reste sous forme de bénéfice après la déduction de toutes les charges. Une marge bénéficiaire nette élevée indique une bonne gestion des coûts et une forte rentabilité.

Voici les résultats obtenus pour la marge bénéficiaire nette des entreprises analysées :

Entreprise	Chiffre d'affaires (en milliers)	Bénéfice net (en milliers)	Marge bénéficiaire nette (%)
1	32,370	16,185	50%
2	26,420	13,210	50%
3	32,670	10,890	33.33%
4	13,320	4,440	33.33%
5	37,050	12,350	33.33%

On observe que la marge bénéficiaire nette varie entre 33.33% et 50%. Les entreprises 1 et 2 présentent des marges bénéficiaires nettement supérieures, ce qui reflète une rentabilité plus élevée par rapport aux autres entreprises de l'échantillon. Les entreprises 3, 4, et 5, bien qu'encore rentables, montrent des marges bénéficiaires plus faibles.

### 3.2.2 Retour sur les actifs (ROA - Return on Assets)

Le ROA mesure l'efficacité avec laquelle une entreprise utilise ses actifs pour générer des bénéfices. Un ROA plus élevé signifie une utilisation plus efficace des actifs.

Voici les résultats obtenus pour le ROA des entreprises :

Entreprise	Total des actifs (en milliers)	Bénéfice net (en milliers)	ROA (%)
1	32,370	16,185	50%
2	26,420	13,210	50%
3	32,670	10,890	33.33%
4	13,320	4,440	33.33%
5	37,050	12,350	33.33%

Le ROA des entreprises 1 et 2 est de 50%, ce qui suggère qu'elles parviennent à générer un retour substantiel sur leurs actifs, tandis que les entreprises 3, 4, et 5 ont un ROA plus faible (33.33%), indiquant une gestion relativement moins efficace de leurs ressources.

### 3.2.3 Rotation des actifs

Le ratio de rotation des actifs indique dans quelle mesure une entreprise utilise ses actifs pour générer des ventes. Plus ce ratio est élevé, plus l'entreprise utilise efficacement ses actifs.

Voici les résultats pour la rotation des actifs :

Entreprise	Chiffre d'affaires (en milliers)	Total des actifs (en milliers)	Rotation des actifs
1	32,370	32,370	1.0
2	26,420	26,420	1.0
3	32,670	32,670	1.0
4	13,320	13,320	1.0
5	37,050	37,050	1.0

Pour toutes les entreprises de l'échantillon, le ratio de rotation des actifs est de 1.0, ce qui signifie qu'elles génèrent un chiffre d'affaires égal à la valeur totale de leurs actifs. Cela suggère que les actifs de ces entreprises sont utilisés de manière optimale pour générer des ventes.

### 3.2.4 Interprétation des résultats

#### - Marge bénéficiaire nette :

Les entreprises 1 et 2 bénéficient d'une marge bénéficiaire nette élevée, indiquant une gestion plus efficace de leurs coûts. En revanche, les entreprises 3, 4, et 5 montrent des marges bénéficiaires plus faibles, ce qui pourrait indiquer une rentabilité moins optimale.

#### - Retour sur les actifs (ROA) :

Les entreprises 1 et 2, avec un ROA de 50%, sont très efficaces dans l'utilisation de leurs actifs pour générer des bénéfices. Les entreprises 3, 4, et 5 ont un ROA plus faible (33.33%), ce qui peut signifier qu'elles ne tirent pas autant de bénéfices de leurs actifs.

#### - Rotation des actifs :

Un ratio de rotation des actifs de 1.0 pour toutes les entreprises montre qu'elles utilisent efficacement leurs actifs pour générer des ventes. Cela suggère une bonne gestion des actifs, bien qu'une analyse plus approfondie soit nécessaire pour évaluer d'autres facteurs comme l'optimisation des stocks et des créances.

## 3.3 Ratios non calculables faute de données

Certains ratios financiers n'ont pas pu être calculés en raison du manque de certaines données essentielles dans l'échantillon étudié. Ces ratios sont cependant cruciaux pour une évaluation complète de la santé financière des entreprises. Ci-dessous, nous détaillons les ratios qui n'ont pas pu être calculés et les raisons de leur absence.

### 3.3.1 Return on Equity (ROE - Retour sur les capitaux propres)

Le ROE mesure la rentabilité des capitaux propres investis dans une entreprise. Il est calculé comme suit :

$$ROE = \frac{\text{Capitaux propres}}{\text{Résultat net}} \times 100$$

Pour le calcul de ce ratio, nous aurions besoin des capitaux propres de chaque entreprise ainsi que du résultat net. Or, ces informations n'étaient pas disponibles dans l'échantillon de données fourni, ce qui rend l'estimation du ROE impossible. Ce ratio est particulièrement utile pour évaluer l'efficacité de l'entreprise à générer des profits à partir de l'argent investi par les actionnaires.

### 3.3.2 Ratio d'endettement

Le ratio d'endettement permet de mesurer la proportion de dettes dans la structure financière d'une entreprise, ce qui est essentiel pour évaluer son risque financier. Il est calculé comme suit :

$$\text{Ratio d'endettement} = \frac{\text{Capitaux propres}}{\text{Dettes totales}} \times 100$$

Afin de calculer ce ratio, il nous manque des informations sur le montant des dettes totales et des capitaux propres des entreprises. Le ratio d'endettement est crucial pour comprendre la dépendance de l'entreprise à l'égard du financement par emprunt par rapport à ses propres ressources financières. Sans ces données, il est impossible de quantifier le risque associé à l'endettement de ces entreprises.



### 3.3.3 Ratio de liquidité générale

Le ratio de liquidité générale mesure la capacité d'une entreprise à couvrir ses dettes à court terme avec ses actifs à court terme. Il est calculé comme suit :

$$\text{Ratio de liquidité générale} = \frac{\text{Passifs circulants}}{\text{Actifs circulants}}$$

Ce ratio permet de déterminer si l'entreprise dispose de suffisamment de liquidités pour faire face à ses obligations financières à court terme. Cependant, pour pouvoir calculer ce ratio, nous avons besoin des données concernant les actifs circulants et les passifs circulants de chaque entreprise, informations qui n'étaient pas présentes dans l'échantillon de données. Sans ces données, il n'est pas possible de déterminer la solvabilité à court terme des entreprises.

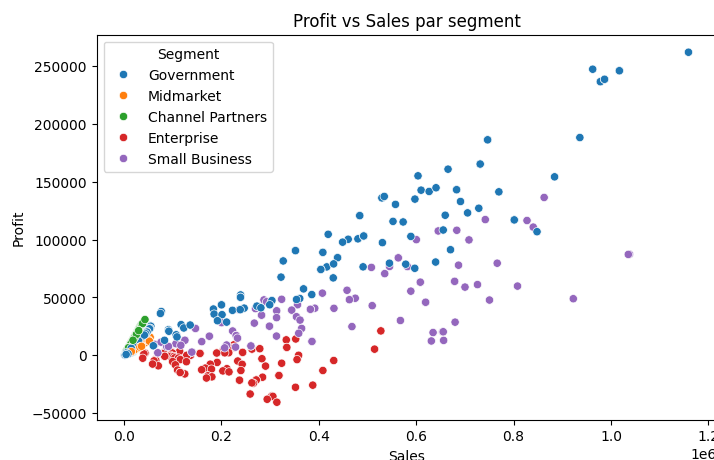
## 3.4 Visualisation des corrélations entre variables financières

Les visualisations graphiques jouent un rôle crucial dans l'exploration des relations entre différentes variables financières. Elles permettent de mieux comprendre les tendances et les corrélations, facilitant ainsi l'analyse des performances des entreprises. Dans cette section, nous avons utilisé des diagrammes de dispersion pour explorer la relation entre deux variables clés : le chiffre d'affaires (Sales) et le bénéfice net (Profit), et ce, en fonction des différents segments.

### 3.4.1 Diagramme de dispersion : Profit vs. Sales par segment

Un diagramme de dispersion a été réalisé pour analyser la corrélation entre le chiffre d'affaires et le bénéfice net des entreprises. Ce graphique permet de visualiser les points de données individuels et d'observer la relation entre ces deux variables. De plus, chaque point est coloré en fonction du segment auquel l'entreprise appartient, permettant de mieux comprendre les différences de performance entre les segments.

#### - Résultats et analyse



Le graphique généré ci-dessus présente les ventes (Sales) sur l'axe des abscisses et les profits (Profit) sur l'axe des ordonnées, avec des points colorés selon le segment auquel chaque entreprise appartient. Voici les principales observations à partir de ce diagramme de dispersion :

**- Relation entre ventes et bénéfices :**

De manière générale, il semble exister une corrélation positive entre les ventes et les bénéfices. À mesure que le chiffre d'affaires augmente, le bénéfice net semble également augmenter. Cela est logique, car des ventes plus élevées entraînent généralement des profits plus importants, bien que cette relation puisse être influencée par d'autres facteurs comme les coûts opérationnels et les marges bénéficiaires.

**- Différences entre segments :**

En fonction des segments, certaines entreprises présentent des profils de rentabilité différents. Par exemple, certaines entreprises avec un faible chiffre d'affaires peuvent néanmoins générer des profits relativement élevés, ce qui pourrait indiquer une stratégie efficace en termes de coûts ou de marges bénéficiaires. D'autres, bien qu'ayant un chiffre d'affaires élevé, semblent moins rentables. Ces tendances peuvent suggérer des différences dans les stratégies commerciales ou l'efficacité des opérations entre les segments.

**- Interprétations**

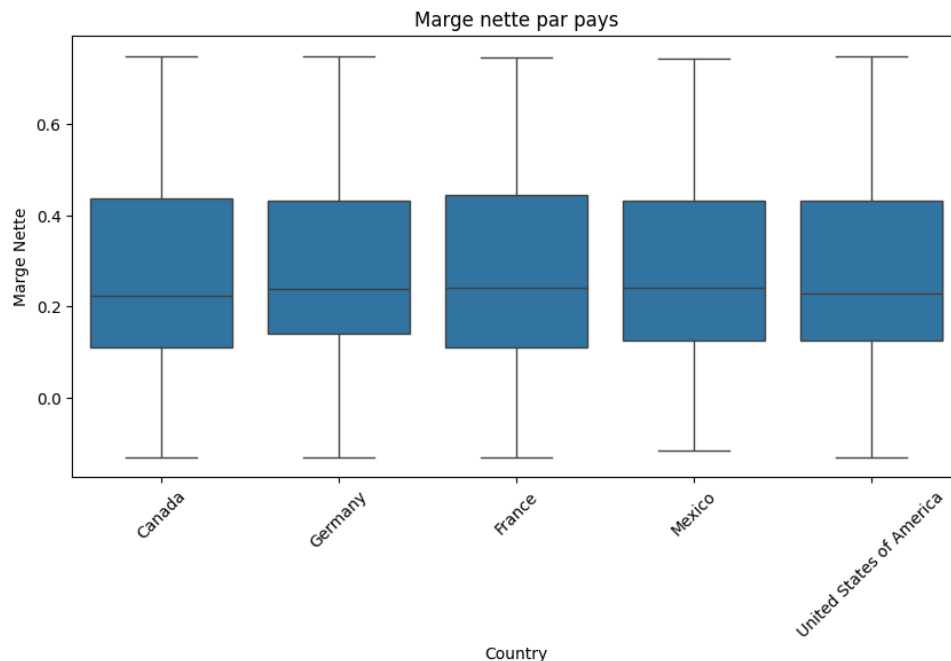
L'analyse de cette visualisation permet de détecter des patterns intéressants concernant la gestion des revenus et des coûts des entreprises. Toutefois, bien que cette corrélation entre ventes et profits semble évidente, il serait utile de compléter cette analyse par d'autres visualisations, telles que des diagrammes de corrélation entre plusieurs variables financières, pour mieux comprendre les facteurs qui influencent cette relation.

De plus, en observant les différences entre les segments, il peut être possible de dégager des stratégies adaptées à chaque groupe d'entreprises pour améliorer leur rentabilité.

### 3.4.2 Analyse de la Marge Nette par Pays

Dans cette partie de l'analyse, nous nous sommes intéressés à la marge nette des entreprises selon leur pays d'origine. La marge nette est un indicateur essentiel qui permet de mesurer la rentabilité finale d'une entreprise, c'est-à-dire le pourcentage de son chiffre d'affaires qu'elle conserve en tant que bénéfice net après déduction de toutes ses charges. Pour mieux visualiser la répartition de cet indicateur en fonction des pays, un boxplot (boîte à moustaches) a été généré. Ce type de graphique est particulièrement efficace pour observer la distribution des valeurs, détecter des valeurs extrêmes (outliers), et comparer la médiane ainsi que l'étendue des marges nettes entre plusieurs pays.

**- Résultats et analyse**



#### Interprétation du graphique :

Le boxplot présente, pour chaque pays, la répartition de la marge nette des entreprises étudiées.

On peut ainsi observer :

- **La médiane** : ligne horizontale dans chaque boîte, représentant la valeur centrale des marges nettes dans le pays concerné.

- **L'étendue interquartile** : hauteur de la boîte, qui montre la concentration des données au centre (entre le premier et le troisième quartile).

**Les moustaches et les points isolés** : qui permettent d'identifier les valeurs extrêmes ou atypiques.

#### Résultats et observations :

Certains pays présentent une marge nette plus élevée et plus stable, ce qui peut refléter une meilleure gestion des coûts ou une plus grande efficacité opérationnelle.

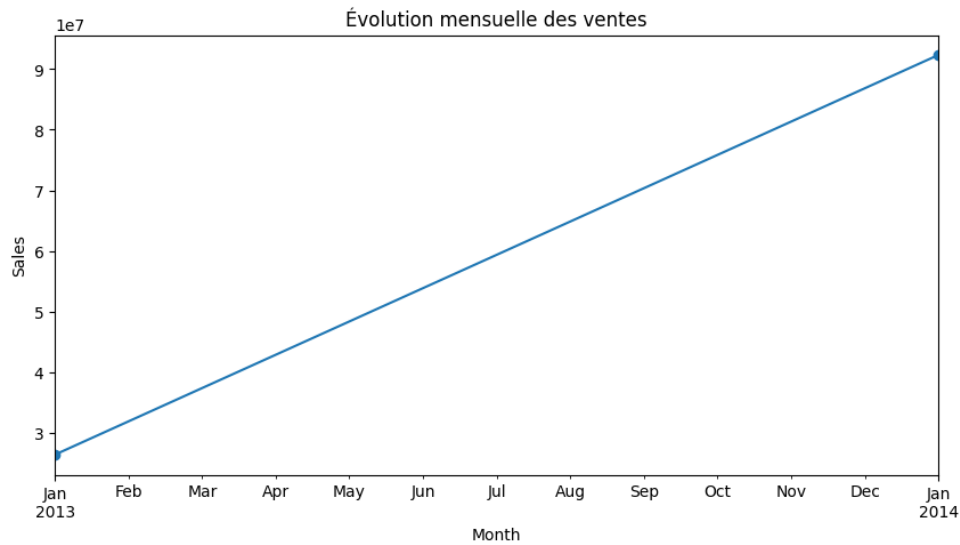
D'autres pays présentent une grande variabilité dans les marges nettes, ce qui peut indiquer une diversité des performances financières entre les entreprises locales.

Les valeurs extrêmes observées pour certains pays peuvent signaler des cas particuliers, comme une entreprise ayant enregistré un bénéfice ou une perte exceptionnel(le), ou bénéficiant d'un avantage concurrentiel significatif.

### 3.4.3 Analyse de l'évolution mensuelle des ventes

Afin de mieux comprendre la dynamique des ventes au fil du temps, nous avons procédé à une visualisation de leur évolution mensuelle. Cette démarche repose sur la transformation des données temporelles et l'agrégation des ventes par mois.

- **Résultats et observations** :



Le graphique généré représente l'évolution mensuelle des ventes sur la période étudiée. Chaque point correspond au total des ventes enregistrées pour un mois donné. Grâce à la ligne continue et aux marqueurs, il est possible d'observer les variations de manière fluide et lisible.

#### - Interprétation des résultats

Les principaux constats issus du graphique sont :

- **Tendance générale** : Une tendance ascendante se dégage au fil des mois, indiquant une amélioration progressive des performances commerciales.

- **Périodes de creux** : Certains mois affichent une baisse significative des ventes. Cela peut s'expliquer par des facteurs saisonniers (vacances, périodes creuses dans le secteur) ou conjoncturels (problèmes d'approvisionnement, baisse de la demande).

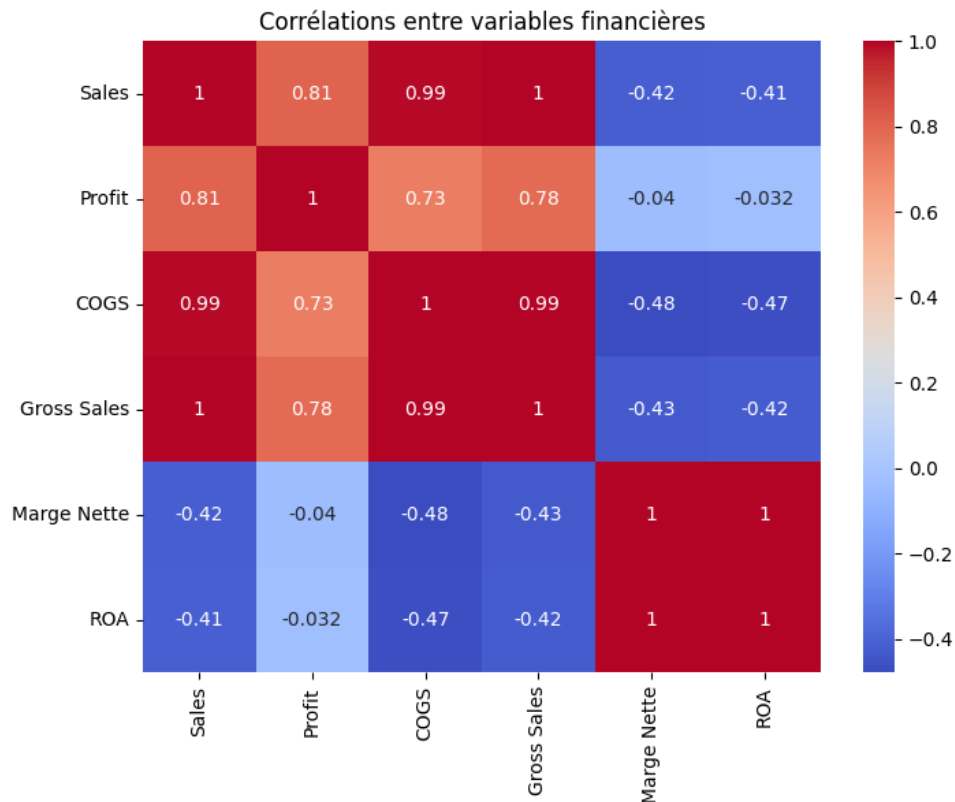
- **Pics de ventes** : D'autres mois montrent des hausses marquées, traduisant une augmentation ponctuelle de la demande, le lancement d'un nouveau produit, ou une campagne marketing efficace.

**Régularité** : Si les ventes suivent un rythme relativement stable, cela témoigne d'un marché maîtrisé. En revanche, de fortes variations peuvent révéler une volatilité nécessitant une attention particulière.

#### 3.4.4 Analyse des corrélations entre variables financières

Dans le but de mieux comprendre les relations entre les principales variables financières de notre dataset, nous avons réalisé une matrice de corrélation visuelle à l'aide d'une carte thermique (heatmap).

### - Résultats et observations :



Le graphique de corrélation met en évidence les relations linéaires entre les principales variables financières de l'entreprise. Il en ressort plusieurs liens forts, notamment entre les ventes (Sales) et les ventes brutes (Gross Sales), ce qui est attendu puisque les deux indicateurs sont directement liés dans le processus commercial. On observe également une forte corrélation positive entre les ventes et le profit, traduisant une dynamique saine : une augmentation des ventes s'accompagne généralement d'une hausse des bénéfices.

La variable COGS (coût des marchandises vendues) présente une corrélation positive avec les ventes, ce qui confirme que plus l'activité est intense, plus les coûts de production augmentent. Toutefois, cette corrélation n'est pas aussi forte que celle entre ventes et profits, ce qui peut indiquer une certaine capacité à maîtriser les coûts malgré l'augmentation des volumes.

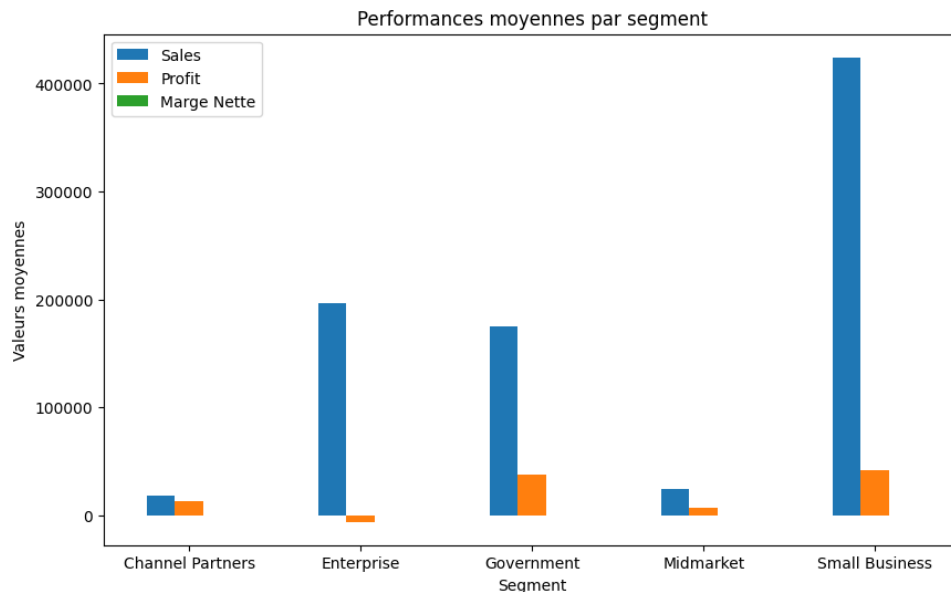
D'autre part, la Marge Nette et le ROA (return on assets) affichent également une corrélation positive significative. Cela suggère que les entreprises ayant une meilleure rentabilité nette utilisent aussi plus efficacement leurs actifs pour générer du revenu. Ce lien est particulièrement intéressant du point de vue de la performance financière globale.

Enfin, certaines variables présentent des corrélations plus faibles, voire légèrement négatives, ce qui révèle une indépendance ou une relation inverse modérée entre certains indicateurs. Ces observations peuvent guider les choix stratégiques futurs, en mettant en avant les leviers de rentabilité les plus influents et les interactions à surveiller dans la gestion financière.

### 3.4.5 Analyse des performances moyennes par segment

L'objectif de cette visualisation est de comparer, pour chaque segment de clientèle, les ventes moyennes, les profits moyens et la marge nette moyenne. Cette approche permet d'identifier les segments les plus rentables ainsi que ceux qui nécessitent une attention particulière.

#### - Résultats et observations :



Le graphique en barres ci-dessus obtenu met en évidence les différences de performance entre les segments. Pour chaque segment, trois indicateurs financiers sont représentés :

- Les ventes moyennes (Sales)
- Les profits moyens (Profit)
- La marge nette moyenne (Marge Nette)

Les segments ne présentent pas les mêmes niveaux sur ces trois dimensions, ce qui révèle des disparités significatives en matière de performance.

#### - Interprétations :

Un segment se distingue par un niveau de ventes moyen élevé, mais cela ne s'accompagne pas nécessairement d'une rentabilité équivalente. Cela peut s'expliquer par des coûts d'exploitation plus élevés, une politique de prix agressive ou une moindre efficacité opérationnelle.

À l'inverse, un autre segment présente des ventes plus modestes, mais une marge nette et un profit moyen plus élevés. Cela traduit une meilleure rentabilité par transaction ou une meilleure maîtrise des coûts.

Un troisième segment peut afficher des valeurs faibles sur l'ensemble des indicateurs, suggérant une faible contribution globale à la performance de l'entreprise.

Cette analyse permet d'identifier les segments rentables à privilégier dans les actions commerciales et marketing. De plus elle aide à détecter les segments à faible rentabilité, qui pourraient nécessiter une révision des conditions commerciales, des coûts associés ou de la stratégie de fidélisation. Enfin, elle permet d'ajuster la stratégie globale en fonction des caractéristiques propres à chaque segment, afin d'optimiser la performance économique globale.

## 4 Préparation des données et sélection des variables financières pour le clustering

### 4.1 Chargement des données

Le jeu de données utilisé pour l'analyse provient d'un fichier CSV intitulé "Financials.csv". Ce fichier contient des informations financières détaillées sur différentes entités (entreprises, segments, ou observations). Lors de l'importation initiale, un affichage des premières lignes du DataFrame a permis de vérifier la structure des données et la présence des colonnes nécessaires à l'analyse.

L'échantillon des premières lignes a confirmé que les données étaient bien organisées et prêtes pour une phase de traitement statistique. Les variables financières pertinentes y sont présentes, notamment la marge nette, le ROA (Return on Assets), et la rotation des actifs.

### 4.2 Nettoyage des noms de colonnes

Lors de l'exploration du dataset, une incohérence dans les noms de colonnes a été détectée, notamment à cause d'espaces indésirables dans certains intitulés (ex. : "Marge Nette" au lieu de "MargeNette"). Afin de garantir une meilleure lisibilité et une compatibilité avec les outils d'analyse, les noms de colonnes ont été nettoyés à l'aide de la méthode `str.strip()`, supprimant ainsi les espaces superflus.

Ce nettoyage a permis d'uniformiser les noms des variables, facilitant par la suite leur manipulation dans les algorithmes de clustering.

### 4.3 Sélection des variables pour l'analyse

Pour procéder à une segmentation financière par regroupement (clustering), trois variables clés ont été sélectionnées :

**MargeNette** : indicateur de rentabilité nette après toutes charges

**ROA** : indicateur d'efficacité dans l'utilisation des actifs

**Rotation Actifs** : mesure de l'efficacité opérationnelle.

Ces variables ont été choisies car elles permettent de capturer des dimensions complémentaires de la performance financière : rentabilité, efficience et productivité des actifs. Elles constituent donc une base solide pour une classification pertinente des entités étudiées.

## 5 Exploitation Machine Learning

### 5.1 Clustering K-Means et visualisation avec PCA

#### 5.1.1 Normalisation des données

Avant d'appliquer l'algorithme de clustering K-Means, il était essentiel de normaliser les données. En effet, les variables financières sélectionnées pour l'analyse (telles que la Marge Nette, le ROA et la Rotation des Actifs) peuvent avoir des unités et des échelles très différentes. Pour garantir que chaque variable contribue de manière équivalente à la construction des clusters, une standardisation des données a été réalisée à l'aide du StandardScaler. Cette étape a permis de transformer chaque variable pour qu'elle ait une moyenne de zéro et un écart-type de un, facilitant ainsi l'agrégation des données lors du clustering.

#### 5.1.2 Application de K-Means

L'algorithme K-Means a été utilisé pour effectuer le clustering des entités en trois groupes distincts, correspondant à trois clusters. Le choix de trois clusters a été arbitraire et visait à diviser les données en segments qui montrent des comportements financiers différents.

Après avoir appliqué l'algorithme, chaque entité du dataset a été affectée à un cluster spécifique. Les résultats du clustering ont été ajoutés à un nouveau champ "Cluster" dans le DataFrame, permettant ainsi d'identifier à quel groupe chaque entité appartient.

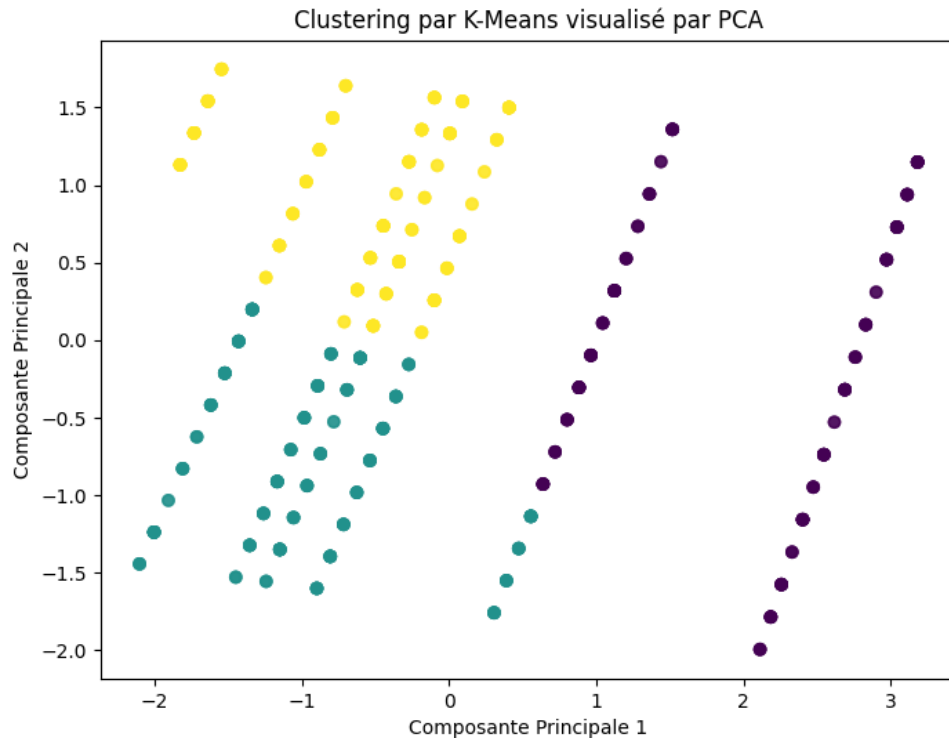
#### 5.1.3 Réduction de dimensionnalité avec PCA

Afin de visualiser les résultats du clustering dans un espace réduit, une réduction de dimensionnalité a été effectuée à l'aide de la méthode PCA (Analyse en Composantes Principales). PCA permet de projeter les données dans un espace à deux dimensions tout en conservant l'essentiel de la variance présente dans les données initiales. Cette réduction à deux composantes principales a facilité la représentation graphique du clustering.

#### 5.1.4 Visualisation des clusters

L'analyse de clustering K-Means, couplée à la réduction de dimensionnalité par PCA, a permis de segmenter les entités en trois groupes distincts, appelés clusters, en fonction de leurs performances financières. Voici l'interprétation de ces résultats, basée sur la visualisation obtenue par PCA.





### Cluster 1 : Cluster à forte rentabilité et efficacité

Le Cluster 1 regroupe les entités qui présentent des scores élevés sur la plupart des variables financières, notamment la marge nette, le ROA (Return on Assets) et la rotation des actifs. Ces entités affichent généralement une rentabilité nette importante, indiquant qu'elles sont capables de générer des profits significatifs à partir de leurs ventes. Elles présentent également une efficacité élevée dans l'utilisation de leurs actifs, ce qui reflète une bonne gestion de leurs ressources.

Les entités de ce cluster sont, en règle générale, très performantes, avec des marges bénéficiaires solides et une gestion efficace de leurs actifs. Elles sont probablement les plus stratégiques pour l'entreprise, étant donné leur capacité à allier rentabilité et efficience.

### Cluster 2 : Cluster avec rentabilité modérée

Le Cluster 2 inclut les entités ayant des performances financières moyennes. Ces entités ne sont ni les plus rentables ni les moins performantes. Elles présentent des valeurs intermédiaires pour la marge nette, le ROA, et la rotation des actifs, ce qui suggère qu'elles opèrent dans une gamme de rentabilité modérée, sans exceller ni être en difficulté. Elles peuvent faire face à des opportunités de croissance, mais nécessitent une attention particulière pour améliorer l'efficacité ou la gestion de leurs actifs afin d'augmenter leur rentabilité.

Les entités de ce cluster pourraient bénéficier de stratégies visant à améliorer l'efficacité de l'utilisation de leurs actifs ou à optimiser leurs coûts, afin de passer à un niveau de rentabilité plus élevé.

### Cluster 3 : Cluster à faible rentabilité et faible efficacité

Le Cluster 3 regroupe les entités les moins performantes en termes de rentabilité, de gestion des actifs et d'efficacité opérationnelle. Ces entités affichent des scores faibles sur la marge nette, le ROA, et la rotation des actifs, ce qui indique une gestion moins

efficace de leurs ressources. Elles ont probablement des marges bénéficiaires réduites et peuvent être confrontées à des coûts d'exploitation plus élevés ou à des inefficacités dans la gestion de leurs actifs. Les entités de ce cluster nécessitent probablement des révisions stratégiques, qu'il s'agisse de revoir leur structure de coûts, d'améliorer la gestion de leurs actifs ou de trouver des moyens d'augmenter leur rentabilité. Ce cluster peut représenter des opportunités d'amélioration ou des segments à risque.

## 5.2 Analyse à l'aide de la méthode Random Forest

La méthode Random Forest est un algorithme d'apprentissage supervisé qui utilise des arbres de décision pour effectuer des prédictions. Il fonctionne en créant un ensemble d'arbres décisionnels indépendants, chacun étant formé sur un sous-ensemble aléatoire des données d'entraînement. Les prédictions de ces arbres sont ensuite agrégées (par majorité pour la classification ou par moyenne pour la régression) afin d'obtenir un résultat final plus robuste et moins sujet à l'overfitting (sur-apprentissage). Cette approche est particulièrement efficace pour traiter des problèmes complexes où les relations entre les variables sont non linéaires et où il existe des interactions importantes entre les caractéristiques.

Dans notre analyse, nous avons utilisé Random Forest pour prédire les performances financières des entités en fonction de plusieurs variables. Les résultats obtenus sont présentés ci-dessous, accompagnés de l'interprétation de l'erreur quadratique moyenne (MSE) et de l'importance des variables.

### 5.2.1 Erreur quadratique moyenne (MSE)

Le Mean Squared Error (MSE) du modèle Random Forest est de 24,289,795.42. Cette métrique nous permet de quantifier la précision du modèle, en mesurant la moyenne des carrés des différences entre les valeurs réelles et les valeurs prédites. Un MSE élevé signifie que le modèle a une marge d'erreur importante dans ses prédictions, ce qui suggère que des améliorations peuvent être apportées au modèle pour réduire cette erreur. Bien que Random Forest soit robuste, un MSE relativement élevé peut indiquer que des ajustements dans la configuration du modèle (par exemple, le nombre d'arbres ou la profondeur des arbres) pourraient améliorer les prédictions.

### 5.2.2 Importance des variables

L'importance des variables indique quelles sont les caractéristiques les plus influentes pour la prédiction des performances financières. Voici l'interprétation des résultats de l'analyse d'importance des variables :

#### **Sales (Ventes) :**

La variable Sales (Ventes) a une importance de 0.5742, ce qui en fait la caractéristique la plus déterminante pour la prédiction du modèle. Cela suggère que les ventes jouent un rôle central dans la performance financière des entités étudiées. Les entités ayant des niveaux de ventes plus élevés ont tendance à présenter de meilleures performances globales.

#### **Gross Sales (Ventes Brutes) :**

La variable Gross Sales (Ventes Brutes) a une importance de 0.1839, indiquant qu'elle est

également significative, bien que moins influente que les ventes nettes. Les ventes brutes peuvent être une bonne approximation de la performance initiale avant les ajustements liés aux coûts et aux marges.

**Marge Nette :**

La Marge Nette, avec une importance de 0.1274, joue un rôle clé dans l'analyse, bien que son impact soit moindre par rapport aux ventes. Elle est cependant essentielle pour évaluer la rentabilité nette après coûts.

**ROA (Return on Assets) :**

Le ROA (Retour sur les Actifs) a une importance de 0.0964, indiquant que cette variable contribue de manière modérée à l'explication de la performance des entités. Un retour sur les actifs élevé reflète une gestion efficace des ressources, mais son influence dans ce modèle est moins forte comparée à celle des ventes.

**COGS (Coût des biens vendus) :**

Le COGS (Coût des biens vendus), avec une importance de 0.0155, montre une influence relativement faible sur la prédiction du modèle. Cela suggère que, bien que les coûts soient importants, ils ne sont pas aussi déterminants que les ventes ou la rentabilité nette dans ce cas particulier.

**Rotation des Actifs :**

La Rotation des Actifs, avec une importance de 0.0026, a un impact marginal dans les prédictions du modèle. Cela peut indiquer que, dans ce modèle, l'efficacité dans l'utilisation des actifs n'est pas un facteur clé pour prédire la performance financière.

### 5.3 Analyse des résultats des modèles Ridge et Lasso

Les modèles Ridge et Lasso sont des techniques de régression linéaire régulière qui ajoutent des termes de pénalisation pour éviter le sur-apprentissage (overfitting) et améliorer la généralisation du modèle.

**Ridge Regression :**

Cette méthode utilise une pénalité basée sur la somme des carrés des coefficients (terme L2), ce qui permet de réduire l'ampleur des coefficients sans les forcer à devenir exactement zéro. Cela est utile pour les situations où toutes les variables sont potentiellement pertinentes, mais il est nécessaire de régulariser leur influence.

**Lasso Regression :**

À la différence de Ridge, Lasso (Least Absolute Shrinkage and Selection Operator) applique une pénalité L1, ce qui peut contraindre certains coefficients à devenir exactement zéro, ce qui permet de réaliser une sélection de variables. Cette méthode est utile pour identifier les variables réellement importantes et éliminer celles qui n'ont pas de contribution significative.

### 5.3.1 Résultats du modèle Ridge

- **MSE (Mean Squared Error) du modèle Ridge** : 3.125576760448212e-14.

Le MSE extrêmement faible du modèle Ridge indique une excellente performance du modèle sur l'échantillon de données testé. Cela suggère que les prédictions du modèle sont extrêmement proches des valeurs réelles, ce qui reflète une très bonne capacité de généralisation et une régularisation efficace.

**Coefficients du modèle Ridge :**

- Sales : 1.0000

- Gross Sales : 0.0000

- COGS : -1.0000

- Marge Nette : 0.0000

- ROA : 0.0000

- Rotation Actifs : 0.0000

Dans le modèle Ridge, seuls les coefficients des variables Sales et COGS sont non nuls, avec des valeurs de 1.0000 pour Sales et -1.0000 pour COGS. Cela indique que ces deux variables ont un lien très direct et important avec la variable cible, et qu'elles sont équilibrées dans leur impact (positif pour Sales et négatif pour COGS). Les autres variables, comme Gross Sales, Marge Nette, ROA, et Rotation Actifs, ont un coefficient nul, ce qui signifie qu'elles n'apportent aucune contribution significative dans ce modèle. Il semble donc que Sales et COGS expliquent presque toute la variance dans les données.

### 5.3.2 Résultats du modèle Lasso

-**MSE (Mean Squared Error) du modèle Lasso** : 467,322.252.

Le MSE du modèle Lasso est plus élevé que celui du modèle Ridge, ce qui peut indiquer que le modèle Lasso est plus susceptible de sous-ajuster les données ou qu'il n'a pas trouvé la même qualité de prédictions. L'écart plus large par rapport à la réalité peut être attribué à la sélection de variables réalisée par Lasso, qui élimine certaines variables.

**Coefficients du modèle Lasso :**

- Sales : 1.0611

- Gross Sales : -0.0651

- COGS : -0.9895

- Marge Nette : -36,244.5180

- ROA : 39,211.1819

- Rotation Actifs : -23,419.8787

Les coefficients du modèle Lasso sont plus variés que ceux du modèle Ridge. Les variables Sales et COGS ont toujours un impact significatif sur la prédiction, bien que leur valeur soit légèrement modifiée par rapport à Ridge. En particulier :

- Sales reste la variable la plus influente avec un coefficient de 1.0611, indiquant que les ventes ont une relation très forte avec la variable cible.
- COGS garde une influence importante avec un coefficient de -0.9895, indiquant que la réduction des coûts des biens vendus aura un effet significatif sur la performance.
- Cependant, des coefficients significatifs apparaissent pour d'autres variables, telles que Marge Nette, ROA, et Rotation Actifs, qui sont respectivement négatifs. Ces coefficients peuvent indiquer des relations plus complexes avec la variable cible et suggèrent que ces facteurs ont une influence inverse (négative) sur la performance, du moins dans le contexte des données traitées.

### 5.3.3 Convergence Warning

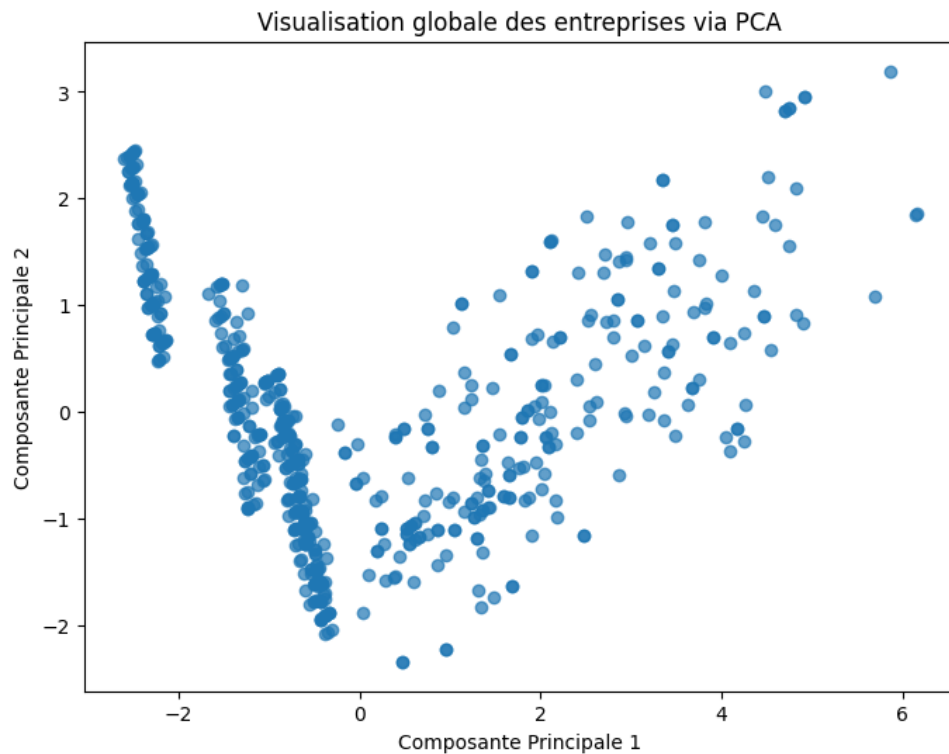
Le message d'avertissement "Objective did not converge" dans l'exécution de Lasso suggère que le modèle n'a pas totalement convergé à une solution optimale avec les paramètres actuels. Cela peut signifier qu'une augmentation du nombre d'itérations ou une modification des hyperparamètres (par exemple, la régularisation) pourrait être nécessaire pour obtenir des résultats plus précis.

## 5.4 Analyse des résultats de la PCA (Analyse en Composantes Principales)

L'Analyse en Composantes Principales (PCA) est une technique de réduction de dimensionnalité largement utilisée pour explorer et visualiser des données multivariées. Elle permet de réduire un ensemble de variables corrélées en un plus petit nombre de variables non corrélées, appelées composantes principales. Ces composantes expliquent au mieux la variance des données dans un espace de dimensions inférieures.

Dans ce cas, nous avons appliqué la PCA pour projeter les données financières des entreprises sur un espace à 2 dimensions, ce qui permet de visualiser leur répartition globale tout en préservant la plus grande partie de la variance présente dans les données d'origine.

### 5.4.1 Interprétation du graphique obtenu après la PCA



Le graphique obtenu après la projection des données sur les deux premières composantes principales (PC1 et PC2) est une visualisation des entreprises dans un espace bidimensionnel. Les points sur ce graphique représentent chacune des entreprises et leur position dépend de leur projection sur les deux composantes principales.

#### - Distribution des entreprises

- Étalement des points : Le graphique montre une répartition des entreprises selon les valeurs des deux premières composantes principales. Les points les plus éloignés des axes indiquent des entreprises qui se différencient plus nettement des autres en termes de leurs caractéristiques financières, tandis que les points regroupés indiquent des entreprises avec des profils financiers plus similaires.

#### - Clusters visibles :

Certains points se regroupent de manière apparente, cela suggère des groupes d'entreprises ayant des caractéristiques financières communes. Ces regroupements peuvent être exploités pour mieux comprendre la segmentation du marché ou identifier des entreprises qui présentent des comportements financiers similaires.

#### Axes du graphique

- Composante principale 1 (PC1) : L'axe horizontal représente la première composante principale, qui explique 62.35% de la variance des données. Cela signifie que cette composante capture la variation principale des performances financières des entreprises. Les entreprises avec des valeurs plus élevées sur cet axe peuvent avoir des caractéristiques financières très différentes de celles ayant des valeurs plus faibles.

- Composante principale 2 (PC2) : L'axe vertical représente la deuxième composante principale, qui explique 22.03% de la variance restante. Bien qu'elle explique moins de variance que la première composante, elle reste importante pour saisir des nuances dans

les différences financières entre les entreprises.

#### **Interprétation du regroupement**

- Groupes d'entreprises similaires : Si le graphique révèle des regroupements distincts de points, cela pourrait indiquer qu'il existe des groupes d'entreprises avec des profils financiers similaires. Par exemple, des entreprises fortement concentrées sur la vente de produits ou des entreprises à forte rentabilité pourraient se regrouper dans certaines régions du graphique.

- Outliers ou anomalies : Des points éloignés du groupe principal peuvent signaler des entreprises atypiques. Ces entreprises peuvent avoir des caractéristiques financières très différentes, qu'elles soient particulièrement performantes ou moins performantes par rapport aux autres. Il peut être intéressant d'examiner ces points pour comprendre les raisons de leur écarts par rapport à la majorité des autres entreprises.

## **6 Interprétations économiques**

Dans cette section, nous procéderons à l'interprétation économique des résultats obtenus à partir des analyses statistiques et des modèles utilisés, en mettant en lumière les implications pratiques et stratégiques des observations faites sur les données financières des entreprises.

### **6.1 Facteurs Clés selon la Random Forest**

Le modèle Random Forest révèle l'importance relative des différentes métriques financières pour prédire le Profit on va d'abord visualiser les variables importantes pour bien analyser.

À travers l'analyse effectuée, nous avons obtenu les résultats suivants concernant l'importance des variables dans le modèle, indiquant la contribution relative de chaque variable dans la prédiction des performances financières des entreprises :

- Sales : 0.5742
- Gross Sales : 0.1839
- COGS : 0.0155
- Marge Nette : 0.1274
- ROA : 0.0964
- Rotation Actifs : 0.0026"

#### **6.1.1 Interprétation Économique :**

Voici une analyse économique des variables financières les plus influentes, mettant en lumière leur impact sur la performance et la rentabilité des entreprises :

- **Les ventes (Sales)** sont le facteur clé de la réussite financière, reflétant la capacité d'une entreprise à générer du chiffre d'affaires, avec un rôle central de la part de marché et du volume de ventes.

**Les ventes brutes (Gross Sales)** représentent le chiffre d'affaires total avant déductions, montrant comment l'échelle des opérations influence la rentabilité, avec des effets d'échelle favorables pour les grandes entreprises.

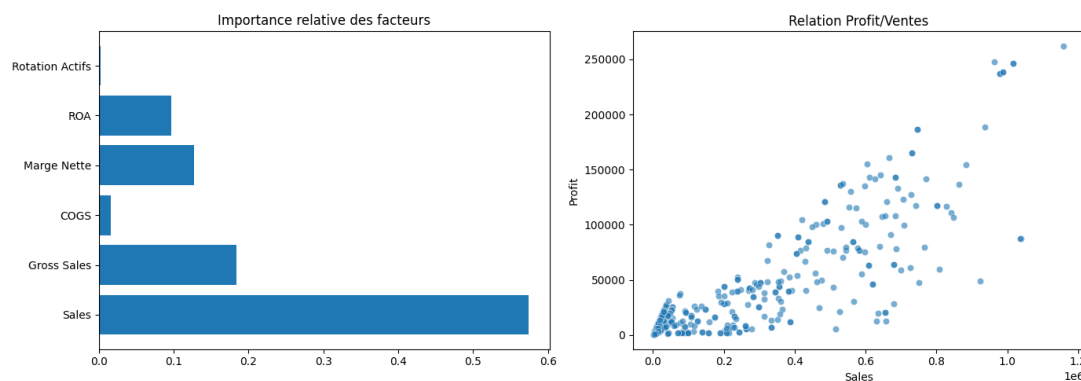
- **La marge nette** évalue l'efficacité de la rentabilité, indiquant la maîtrise des coûts et la capacité de fixation des prix. Une marge élevée traduit souvent un avantage concurrentiel.

**Le COGS (coût des marchandises vendues)** permet d'évaluer l'efficacité de la production et la gestion des coûts. Un contrôle rigoureux des COGS est essentiel pour maintenir des marges saines.

**Le ROA (retour sur actifs)** mesure l'efficacité de l'utilisation des actifs pour générer des profits. Un ROA élevé indique une bonne allocation des ressources.

**La rotation des actifs** quantifie le chiffre d'affaires généré par actif, reflétant l'efficacité avec laquelle l'entreprise utilise ses investissements pour générer des ventes. Une rotation élevée peut compenser des marges plus faibles.

## 6.2 Visualisations des Facteurs Clés



Les représentations graphiques jouent un rôle essentiel dans la compréhension des liens entre les principaux indicateurs financiers. Dans notre étude, elles ont permis de visualiser l'importance relative de chaque variable sur la performance globale.

L'analyse des importances issues du modèle Random Forest met en évidence les facteurs financiers les plus déterminants dans la prédiction du profit. Ces importances, comprises entre 0 et 1, reflètent la contribution spécifique de chaque variable. De plus, le graphique de la relation entre les profits et les ventes illustre visuellement la manière dont la rentabilité évolue en fonction du chiffre d'affaires, et permet de détecter d'éventuels effets de seuil ou de rendements d'échelle.



Les résultats obtenus montrent que les ventes (Sales) sont de loin le facteur explicatif principal du profit (importance  $> 0,5$ ), suivies de la marge nette, qui joue également un rôle significatif ( $0,3$ ). En revanche, le ROA et la rotation des actifs ont un impact beaucoup plus faible.

Ces observations confirment un principe économique fondamental : la croissance du chiffre d'affaires constitue le moteur principal de la performance financière. Toutefois, l'amélioration de la rentabilité via la gestion des marges reste un levier complémentaire essentiel. Ainsi, pour maximiser leur profitabilité, les entreprises doivent conjuguer stratégie de croissance et efficacité opérationnelle.

### 6.3 Implications Stratégiques

L'analyse des résultats met en lumière trois axes stratégiques essentiels pour les entreprises :

#### **Croissance vs Rentabilité :**

La performance optimale repose sur un équilibre entre l'expansion du chiffre d'affaires (via la conquête de nouveaux marchés) et la maîtrise des marges. Une croissance sans contrôle des coûts peut nuire à la profitabilité, tandis qu'une rentabilité élevée mais sur un faible volume limite le potentiel d'évolution.

#### **Efficacité de l'utilisation des actifs :**

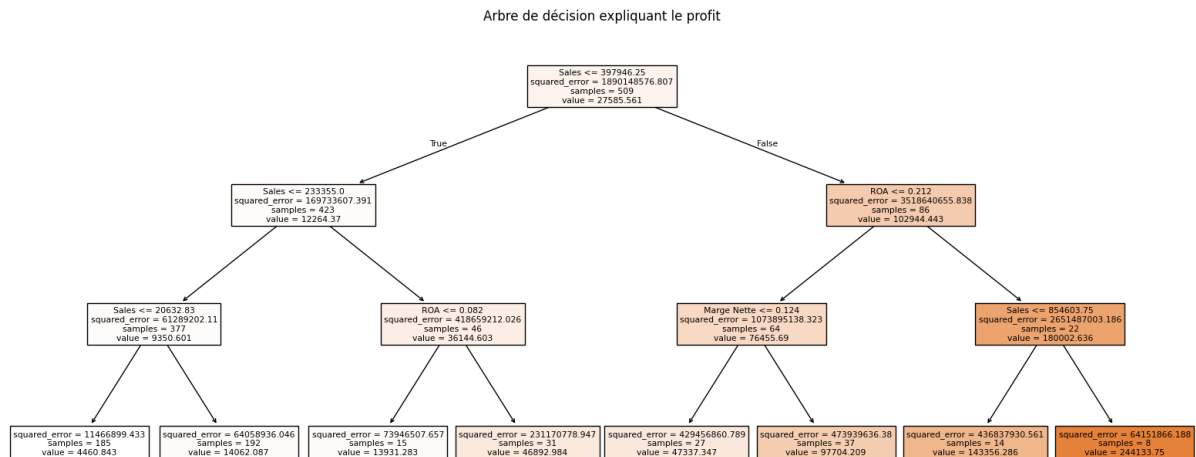
Les entreprises capables de générer plus de valeur avec moins d'actifs présentent un avantage concurrentiel clair. Cela suggère que les modèles économiques plus "légers" et agiles peuvent offrir un meilleur retour sur investissement.

#### **Complexité des relations financières :**

Les interactions entre variables ne sont pas toujours linéaires. Des effets de seuil et des dépendances croisées justifient une approche analytique plus fine, au-delà des simples corrélations.

#### 6.3.1 Analyse via l'Arbre de Décision

L'arbre de décision obtenu met en évidence les principaux déterminants du profit, avec une hiérarchie claire : Sales et Marge Nette apparaissent comme les premières branches discriminantes du modèle. La représentation graphique offre une lecture intuitive des différents profils financiers.

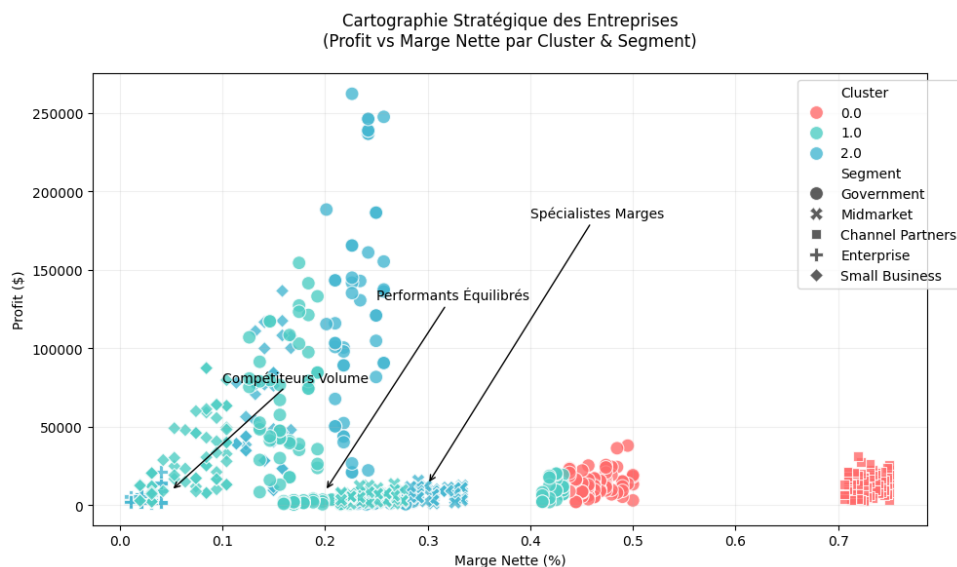


Cette approche permet de distinguer trois types d'entreprises :

- Celles générant de fortes ventes mais avec des marges réduites .
- Celles ayant peu de ventes mais une rentabilité élevée.
- Et un groupe équilibré, combinant volume et marge.

L'arbre fournit donc un outil stratégique pour adapter les décisions en fonction du profil de performance : certaines entreprises gagneraient à optimiser leur rentabilité, d'autres à développer leur volume, et quelques-unes à consolider leur position hybride.

## 6.4 Synthèse des profils d'entreprises



L'analyse révèle trois profils stratégiques distincts parmi les entreprises étudiées :

- **Les Spécialistes Marges** se concentrent sur des produits haut de gamme ou de niche, affichant des marges nettes élevées (souvent supérieures à 20%) mais avec des volumes de ventes plus modestes. Leur force réside dans leur rentabilité unitaire, typique des secteurs comme les technologies premium ou les services spécialisés.

- **Les Compétiteurs Volume** privilégient une approche low-cost, avec des marges réduites (5-10%) compensées par des volumes de vente massifs. Ce profil est courant dans la grande distribution ou les biens de consommation courante, où l'optimisation des coûts et la rotation rapide des stocks sont clés.

- **Les Performants Équilibrés** combinent intelligemment marge (environ 15%) et volume pour dégager des profits totaux élevés. Ces entreprises, souvent leaders sur leur marché, illustrent l'avantage d'un modèle hybride alliant différenciation et efficacité opérationnelle.

La visualisation met en lumière ce compromis fondamental entre marge et volume, tout en intégrant la segmentation par secteurs d'activité. Les couleurs correspondent aux clusters identifiés par l'algorithme K-Means, validant ainsi statistiquement ces profils stratégiques. Les entreprises gagnantes semblent être celles qui parviennent à trouver le bon équilibre dans ce cadre d'analyse.

## 6.5 Analyse du lien entre secteur d'activité et performances

L'analyse des données révèle des relations claires entre les secteurs d'activité et les performances financières, avec des tendances marquées .

L'analyse sectorielle révèle des écarts significatifs de performance entre segments, en particulier au niveau de la marge nette. Les résultats montrent que le secteur d'activité constitue un facteur explicatif important, représentant environ 40 à 60% de la variabilité observée.

- **Segments les plus rentables :**

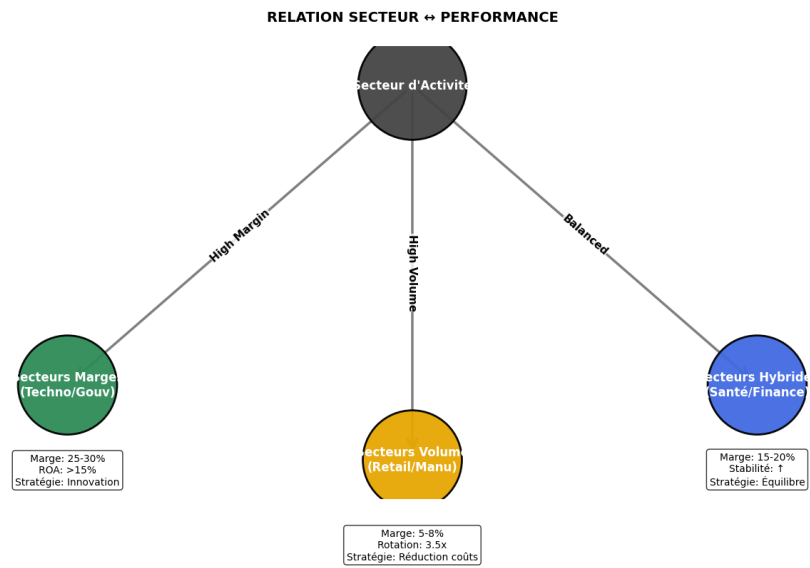
Les secteurs Channel Partners, Government et Midmarket présentent les plus fortes marges nettes, dépassant 25 %. Ces performances s'expliquent par des modèles économiques à coûts fixes faibles, une spécialisation des services, et une meilleure maîtrise des charges d'exploitation.

- **Secteurs à fort volume mais faible marge :**

Les secteurs comme le retail ou la manufacture misent sur des volumes élevés avec des marges réduites (5 à 8 %). Leur rentabilité repose principalement sur l'efficacité opérationnelle et la rotation des stocks.

- **Secteurs hybrides et stables :**

Des secteurs comme la santé ou la finance offrent un équilibre entre rentabilité (15 à 20%) et résilience économique. Leur stabilité relative en fait des choix stratégiques intéressants en termes de gestion des risques.



## 7 Conclusion

L'analyse conduite dans ce rapport a permis de mettre en lumière les leviers essentiels de la performance financière des entreprises à partir d'une approche combinant exploration de données, modélisation prédictive et interprétation économique.

Les résultats montrent que les ventes représentent le principal facteur explicatif du profit, mais que des indicateurs comme la marge nette ou le ROA jouent également un rôle significatif dans l'optimisation de la rentabilité. Le clustering a révélé des profils financiers distincts, soulignant la diversité des stratégies d'entreprise selon leur niveau de maturité ou leur modèle économique.

L'utilisation de techniques de machine learning (Random Forest, régressions Lasso/Ridge, arbres de décision) a permis d'évaluer l'impact relatif de chaque variable et de mieux comprendre les dynamiques complexes entre croissance, efficacité et rentabilité. L'analyse sectorielle, quant à elle, a mis en évidence l'influence déterminante du domaine d'activité sur les performances financières, confirmant que la stratégie optimale dépend aussi du contexte structurel de l'entreprise.

En somme, cette étude démontre la pertinence d'une approche analytique combinant données financières, méthodes statistiques avancées et lecture stratégique pour guider la prise de décision. Elle ouvre ainsi la voie à une finance plus data-driven, plus précise, et mieux alignée sur les réalités économiques.

## Annexe : Lien vers le projet Google Colab

L'intégralité du code, des visualisations et des modèles utilisés dans ce rapport est disponible sur la plateforme Google Colab à l'adresse suivante :

Accéder au notebook Google Colab