



ÉCOLE CENTRALE CASABLANCA

PROJET GESTION D'ENTREPRISE  
RAPPORT

---

# Analyse Financière avec Data Science et Machine Learning

*Elèves :*

Salah Eddine EL AZOUTI  
Anass EL HAYEL  
Walid EL BOUCHTI  
Aasmaa OUAMALICH  
Salma SAAIDI

*Encadrants :*  
Fouad RIANE

## Table des matières

<b>1 Introduction</b>	<b>2</b>
<b>2 Données et exploration initiale</b>	<b>2</b>
2.1 Présentation du jeu de données . . . . .	2
2.2 Exploration des données . . . . .	3
2.2.1 Structure générale . . . . .	3
2.2.2 Traitement des valeurs manquantes . . . . .	3
2.2.3 Détection et traitement des valeurs aberrantes . . . . .	4
2.3 Méthodes d'analyse appliquées . . . . .	4
<b>3 Résultats et analyses</b>	<b>4</b>
3.1 Analyse des ratios et performance par secteur . . . . .	4
3.1.1 Analyse des ratios . . . . .	4
3.1.2 Analyse des performances par secteur . . . . .	4
3.2 Corrélation entre variables financières . . . . .	6
3.3 Classification des entreprises et interprétation des clusters . . . . .	7
3.3.1 Réduction de dimension par ACP . . . . .	7
3.3.2 Classification des entreprises par K-Means . . . . .	9
3.3.3 Interprétation des clusters . . . . .	10
<b>4 Interprétation économique complète</b>	<b>13</b>
4.1 Facteurs clés de succès financier . . . . .	13
4.2 Profils d'entreprises identifiés . . . . .	13
4.3 Analyse sectorielle et dynamique de performance . . . . .	14
4.4 Facteurs de valorisation du marché . . . . .	14
4.5 Implications pour l'allocation d'actifs et la gestion de portefeuille . . . . .	14
<b>5 Recommandations pratiques</b>	<b>15</b>
5.1 Pour les investisseurs . . . . .	15
5.2 Pour les dirigeants d'entreprise . . . . .	15
5.3 Pour les analystes financiers . . . . .	15
<b>6 Limites et Perspectives de Recherche</b>	<b>16</b>
6.1 Limites de l'Analyse . . . . .	16
6.2 Perspectives de Recherche . . . . .	16
<b>7 Conclusion</b>	<b>17</b>

# 1 Introduction

Le monde des affaires contemporaines se caractérise par une grande diversité d'entreprises évoluant dans des contextes économiques variés et souvent instables. Dans ce cadre, la compréhension des déterminants de la performance financière constitue un enjeu majeur pour les parties prenantes, qu'il s'agisse des investisseurs, des dirigeants ou des analystes financiers. Traditionnellement fondée sur l'examen manuel d'un nombre restreint d'indicateurs, l'analyse financière connaît aujourd'hui une transformation profonde sous l'impulsion du Big Data et des technologies d'intelligence artificielle.

Ces avancées permettent désormais le traitement simultané de volumes massifs de données et la détection de structures sous-jacentes complexes, souvent inaccessibles à l'analyse humaine classique. Ainsi, les outils issus de la Data Science et du Machine Learning offrent de nouvelles perspectives pour appréhender la santé financière des entreprises de manière plus fine, plus exhaustive et potentiellement plus prédictive.

Ce projet s'inscrit dans cette évolution méthodologique en proposant une analyse financière augmentée à partir d'un vaste ensemble de données relatives à des entreprises américaines. L'approche mobilisée combine des techniques statistiques avancées — telles que le calcul de ratios financiers, les méthodes de regroupement (clustering) et la modélisation prédictive — afin de mieux comprendre les profils d'entreprises, les dynamiques sectorielles et les perspectives économiques.

Les objectifs de ce travail sont les suivants :

- Nettoyer et explorer un jeu de données comprenant 200 indicateurs financiers ;
- Calculer et analyser des ratios financiers significatifs (ROA, ROE, marge nette, etc.) ;
- Identifier des groupes d'entreprises présentant des profils financiers similaires à l'aide d'algorithmes de clustering ;
- Élaborer des modèles prédictifs pour anticiper les performances futures ;
- Interpréter les résultats dans une perspective économique et stratégique ;
- Formuler des recommandations fondées sur l'analyse effectuée.

L'ambition de ce rapport est ainsi de contribuer à une meilleure compréhension des leviers de performance des entreprises à l'ère des données massives, en articulant rigueur quantitative et pertinence analytique.

# 2 Données et exploration initiale

## 2.1 Présentation du jeu de données

Le jeu de données utilisé dans ce projet regroupe environ 200 indicateurs financiers relatifs à un ensemble d'entreprises américaines cotées en bourse sur la période 2014–2018. Ces indicateurs couvrent un large spectre de la performance financière, incluant :

- Des informations générales (nom de l'entreprise, secteur d'activité, capitalisation, etc.) ;
- Des métriques liées aux revenus, aux bénéfices et aux marges ;
- Des données sur les actifs, les passifs et la structure du capital ;
- Des indicateurs de performance boursière (rendement, volatilité, etc.).

L'échantillon est composé d'entreprises issues de secteurs économiques variés. Il inclut notamment des secteurs à forte croissance, tels que la technologie et la santé, ainsi que

des secteurs plus stables comme les services publics ou les biens de consommation.

## 2.2 Exploration des données

### 2.2.1 Structure générale

L'exploration initiale a permis d'acquérir une vue d'ensemble de la structure du jeu de données. Nous avons analysé la distribution sectorielle des entreprises, la répartition des indicateurs clés, ainsi que les statistiques descriptives de base (moyennes, médianes, variances) pour chaque variable pertinente.

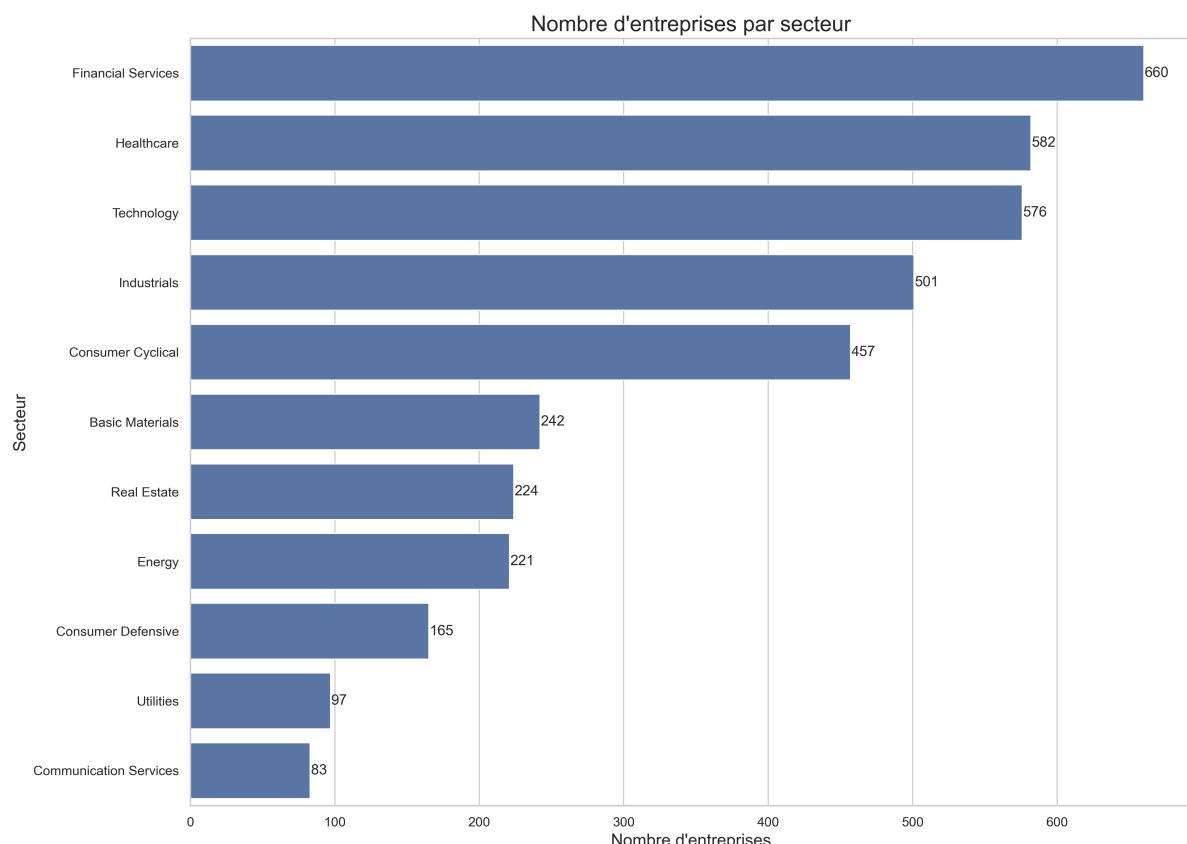


FIGURE 1 – Distribution des entreprises par secteur

La figure 1 montre la répartition des entreprises par secteur dans notre jeu de données. On constate que certains secteurs sont plus représentés que d'autres, ce qui est important à prendre en compte lors de l'interprétation des résultats d'analyse.

### 2.2.2 Traitement des valeurs manquantes

Le traitement des valeurs manquantes constitue une étape essentielle de la phase de prétraitement. Les colonnes comportant un nombre significatif de valeurs manquantes ont été identifiées, puis corrigées à l'aide de techniques appropriées (imputation par la moyenne, médiane, ou régression selon les cas).

### 2.2.3 Détection et traitement des valeurs aberrantes

Les valeurs aberrantes peuvent biaiser fortement les résultats des analyses statistiques et des algorithmes d'apprentissage automatique. Nous avons mis en œuvre différentes méthodes de détection (z-score, IQR, visualisations) afin de les traiter de manière rigoureuse, par exclusion ou transformation, selon leur nature.

## 2.3 Méthodes d'analyse appliquées

Pour mener cette analyse, nous avons mobilisé une combinaison de méthodes statistiques et de techniques issues de la science des données. L'analyse de ratios financiers a permis de caractériser les entreprises selon leur rentabilité, leur efficacité opérationnelle et leur structure financière. Des méthodes de clustering ont ensuite été utilisées pour regrouper les entreprises selon des profils similaires, en fonction de l'ensemble des indicateurs.

Enfin, des modèles d'apprentissage automatique supervisé (tels que les arbres de décision et les forêts aléatoires) ont été intégrés afin d'affiner la compréhension des facteurs explicatifs de la performance et d'identifier les leviers stratégiques au sein des différents secteurs d'activité.

## 3 Résultats et analyses

### 3.1 Analyse des ratios et performance par secteur

#### 3.1.1 Analyse des ratios

Nous avons calculé plusieurs ratios financiers clés pour enrichir notre analyse :

- **ROA (Return on Assets)** : mesure la rentabilité des actifs ;
- **ROE (Return on Equity)** : mesure la rentabilité des capitaux propres ;
- **Marge nette** : rapport entre le bénéfice net et le chiffre d'affaires ;
- **Ratio d'endettement** : mesure du niveau d'endettement par rapport aux actifs ;
- **Ratio de liquidité** : capacité à honorer les dettes à court terme.

#### 3.1.2 Analyse des performances par secteur

La performance financière varie considérablement selon les secteurs d'activité. Nous avons analysé comment les différents indicateurs et ratios se comportent à travers les divers secteurs représentés dans notre jeu de données.

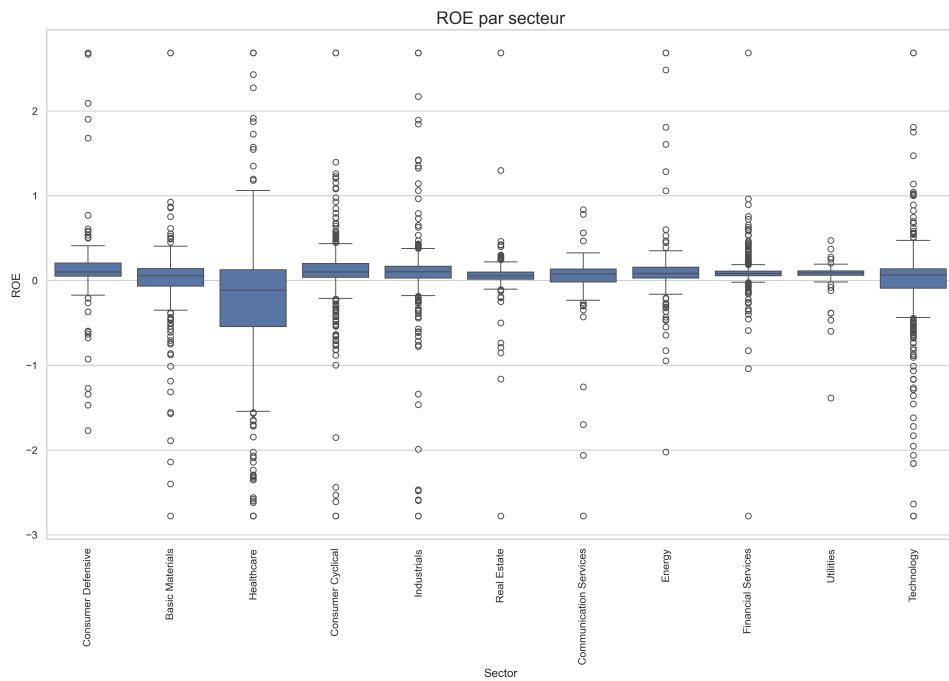


FIGURE 2 – Rentabilité des capitaux propres (ROE) par secteur

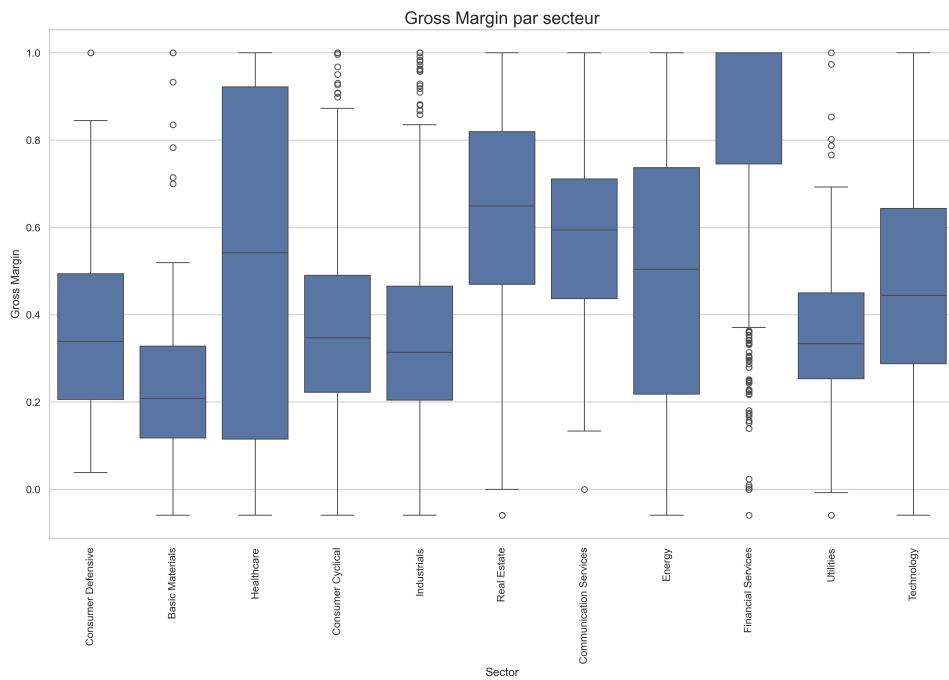


FIGURE 3 – Marge brute par secteur

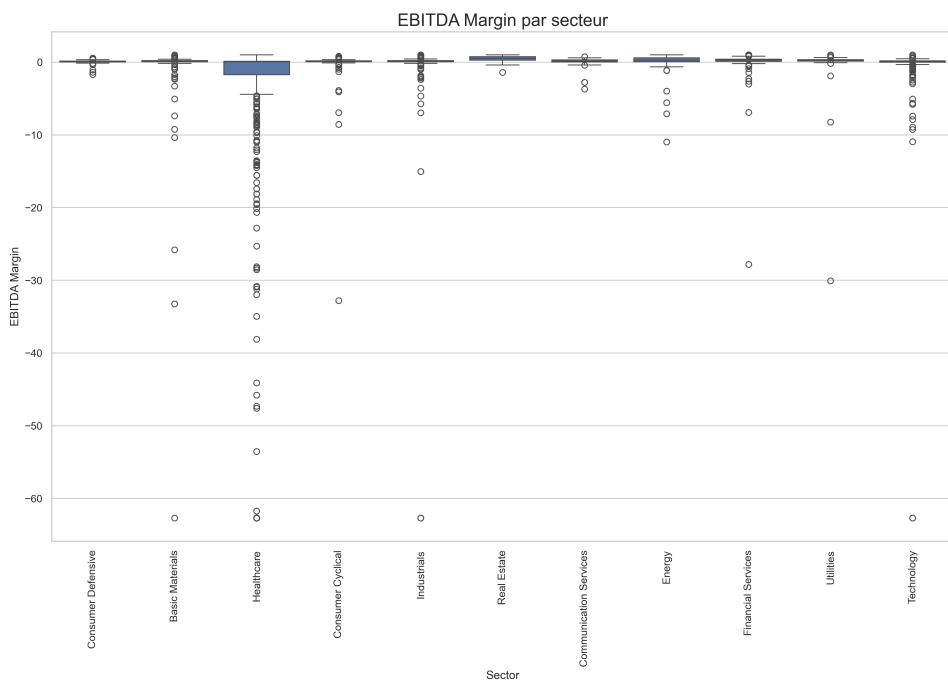


FIGURE 4 – Marge EBITDA par secteur

Ces visualisations (figures 2, 3 et 4) mettent en évidence les différences significatives de performance financière entre les secteurs. On observe notamment que certains secteurs présentent des ratios de rentabilité plus élevés et plus stables que d'autres, ce qui peut être attribué à diverses caractéristiques sectorielles telles que les barrières à l'entrée, l'intensité concurrentielle et les structures de coûts.

### 3.2 Corrélation entre variables financières

L'étude des corrélations entre les différentes variables financières permet d'identifier des relations potentiellement significatives et d'orienter la suite de l'analyse.

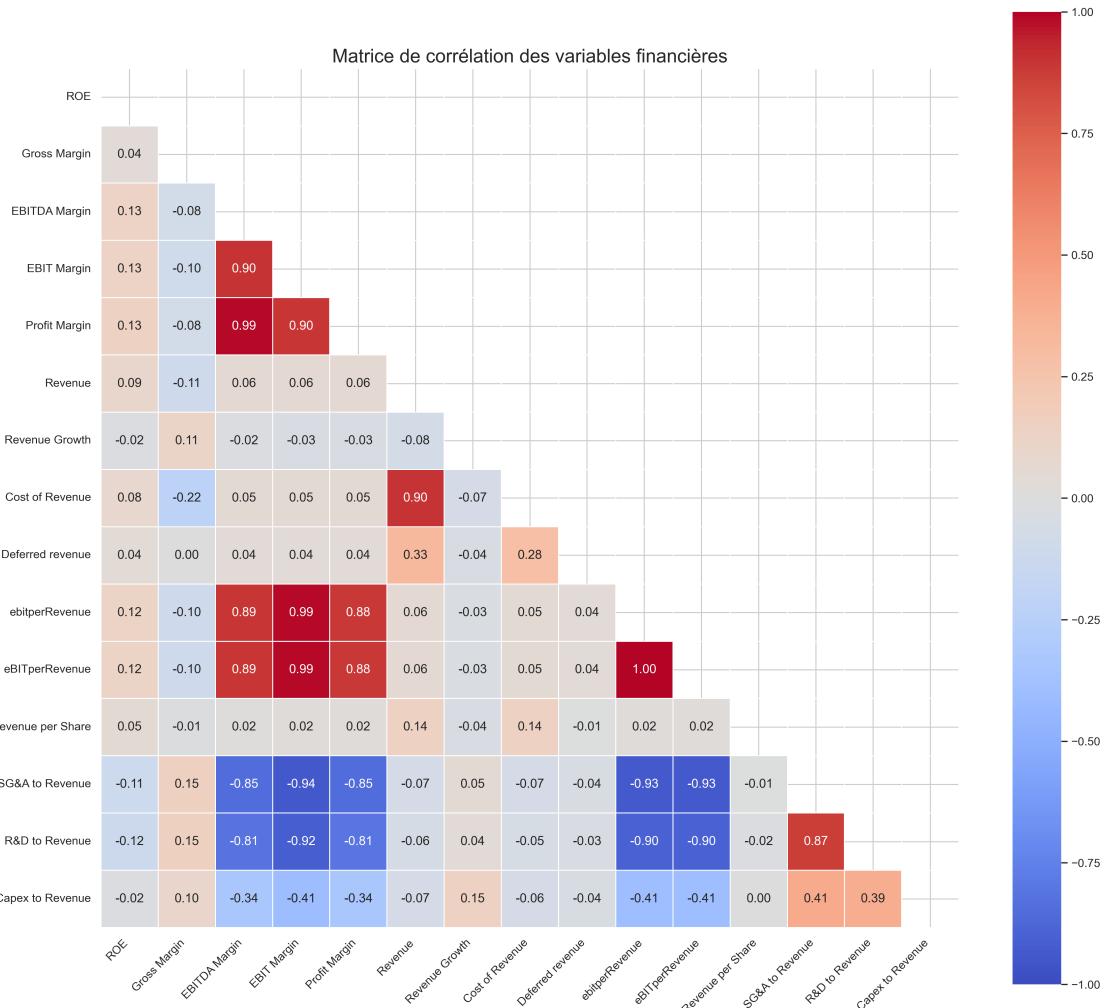


FIGURE 5 – Matrice de corrélation des principales variables financières

La matrice de corrélation figure 5 révèle plusieurs relations importantes entre les variables financières. Les corrélations positives fortes (en bleu foncé) indiquent des variables qui évoluent généralement dans le même sens, tandis que les corrélations négatives (en rouge) montrent des variables qui tendent à évoluer en sens inverse. Ces informations sont cruciales pour comprendre les interactions entre différents aspects de la performance financière des entreprises.

### 3.3 Classification des entreprises et interprétation des clusters

#### 3.3.1 Réduction de dimension par ACP

L'Analyse en Composantes Principales (ACP) est une technique statistique puissante qui permet de réduire la dimensionnalité d'un jeu de données tout en conservant le maximum d'information. Nous l'avons appliquée à notre ensemble de données pour identifier les principales sources de variance.

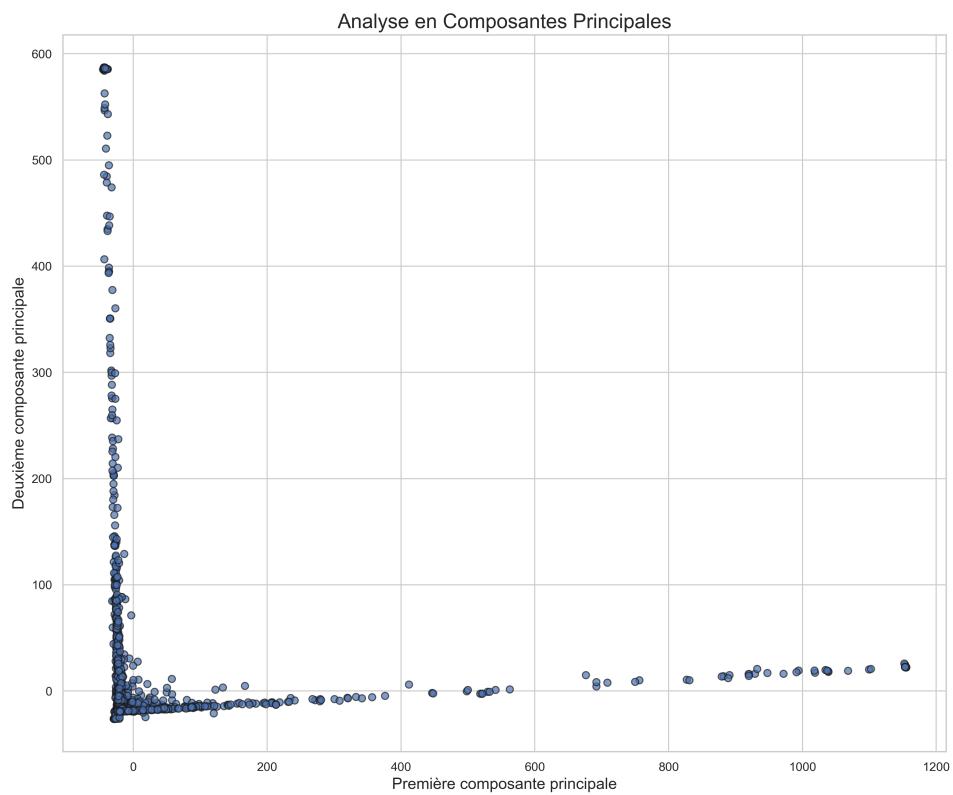


FIGURE 6 – Projection des entreprises sur les deux premières composantes principales

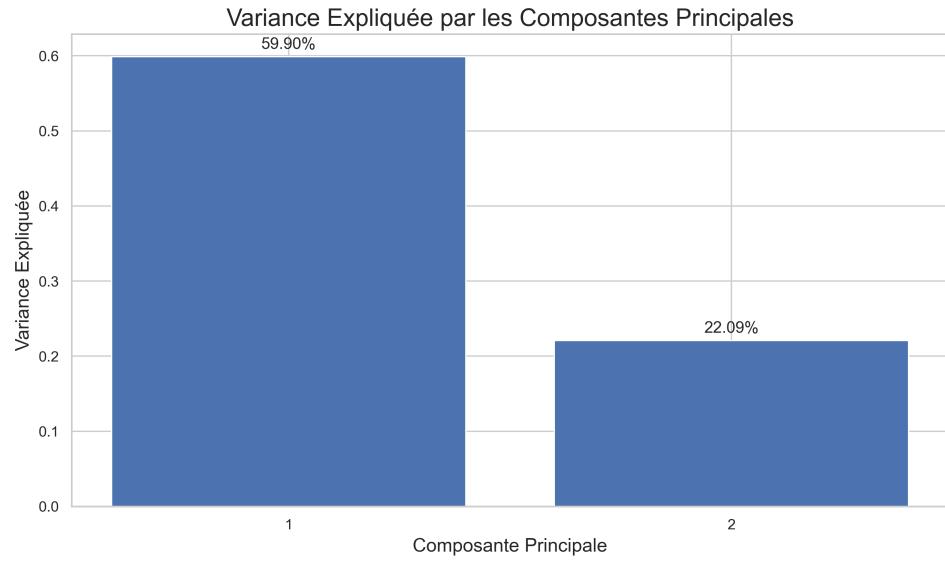


FIGURE 7 – Variance expliquée par les composantes principales

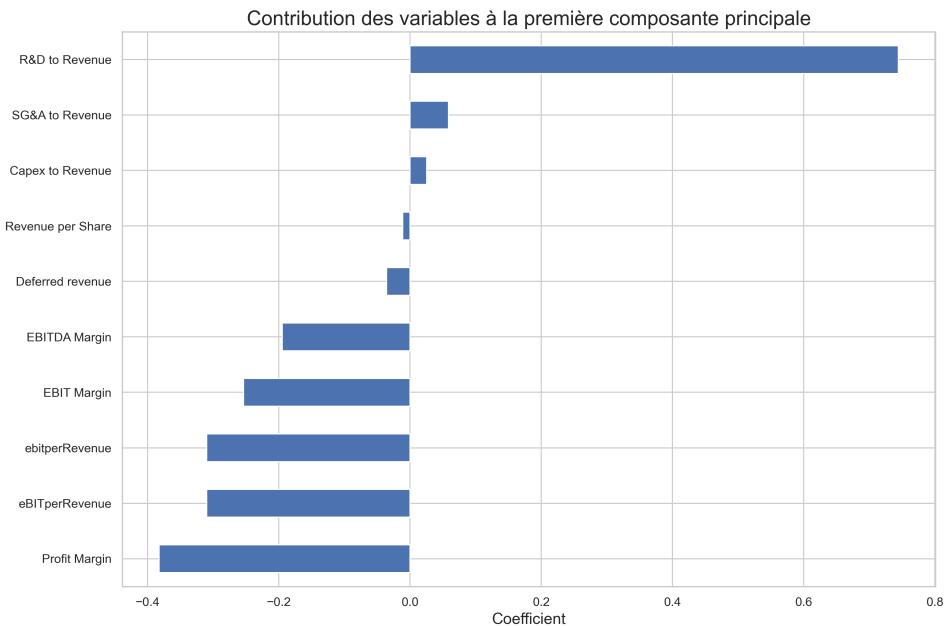


FIGURE 8 – Contribution des variables à la première composante principale

La figure 6 montre la projection des entreprises sur les deux premières composantes principales, qui captent la plus grande partie de la variance dans les données. La figure 7 illustre la proportion de variance expliquée par chaque composante principale. La figure 8 présente les variables qui contribuent le plus à la première composante principale, révélant ainsi les facteurs financiers les plus discriminants dans notre jeu de données.

Cette analyse nous permet de réduire la complexité des données tout en conservant l’information essentielle, facilitant ainsi la visualisation et l’interprétation des patterns dans les données financières.

### 3.3.2 Classification des entreprises par K-Means

L’algorithme K-Means nous a permis de regrouper les entreprises en clusters selon leurs profils financiers. Cette approche non supervisée révèle des groupes naturels d’entreprises ayant des caractéristiques similaires.

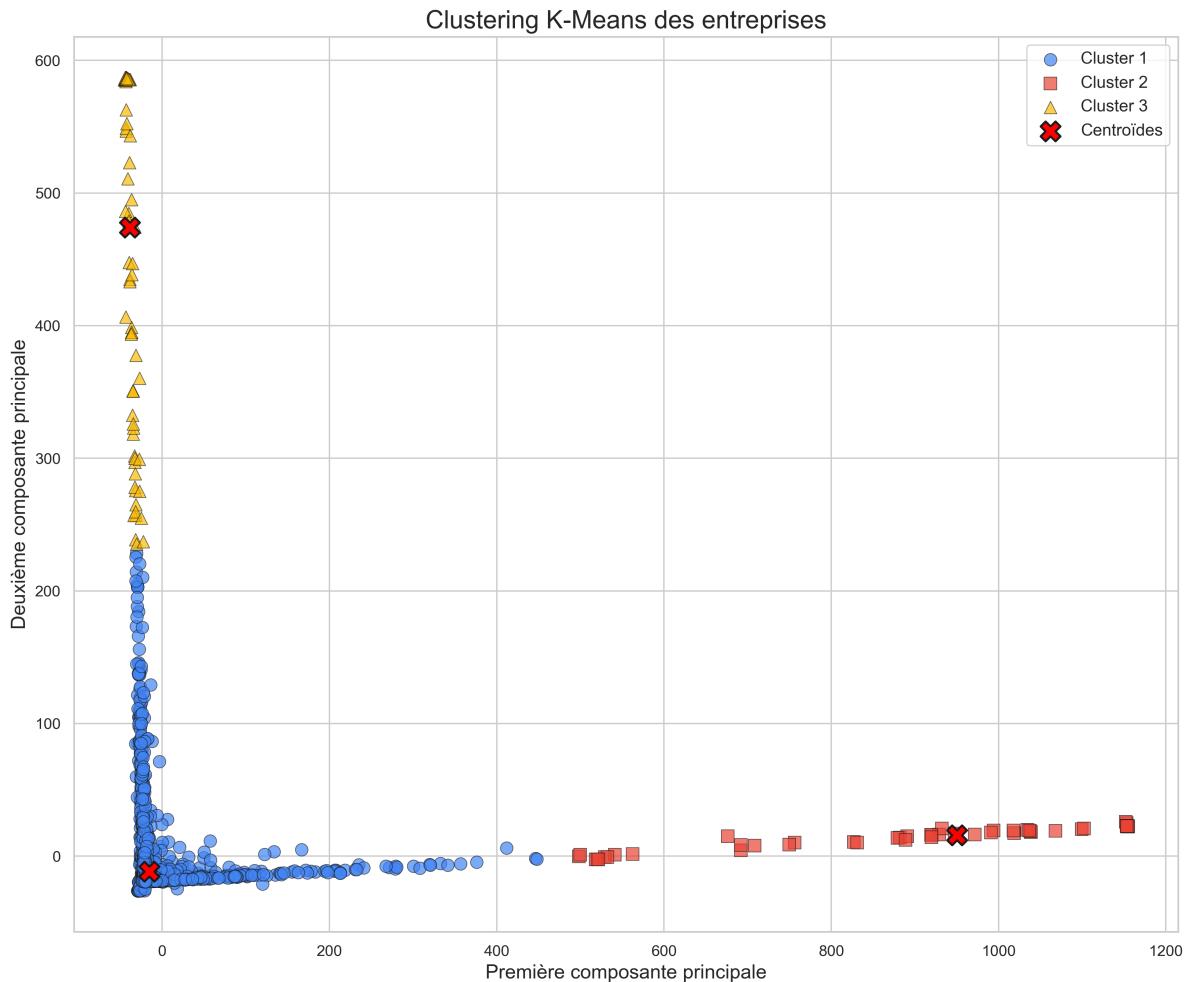


FIGURE 9 – Clusters d’entreprises identifiés par l’algorithme K-Means

### 3.3.3 Interprétation des clusters

Une fois les clusters identifiés, nous avons analysé leurs caractéristiques pour comprendre ce qui distingue chaque groupe et proposer une interprétation financière et économique.

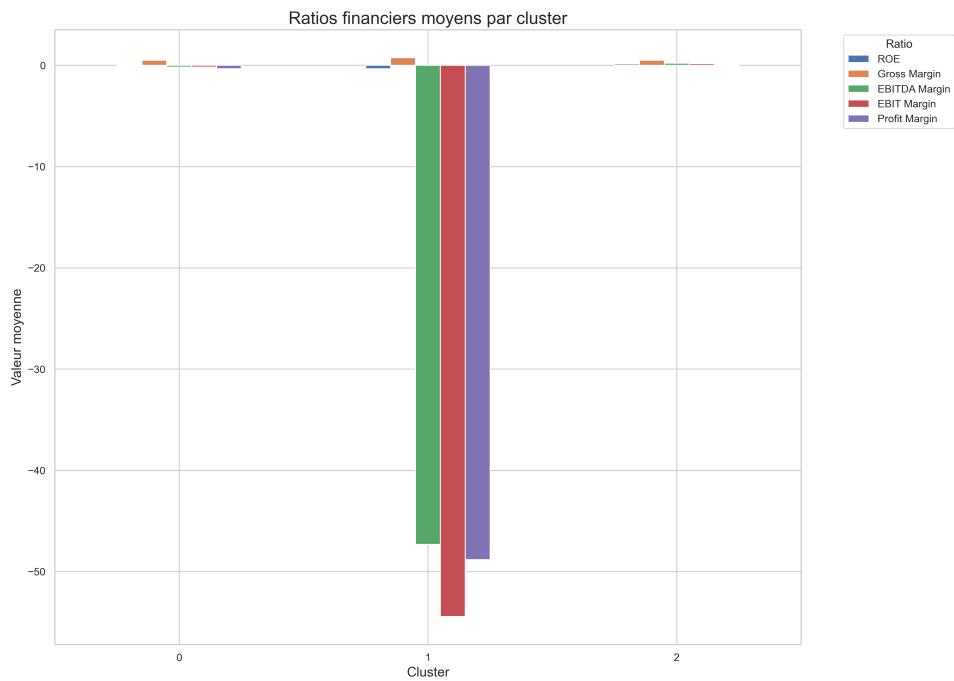


FIGURE 10 – Ratios financiers moyens par cluster

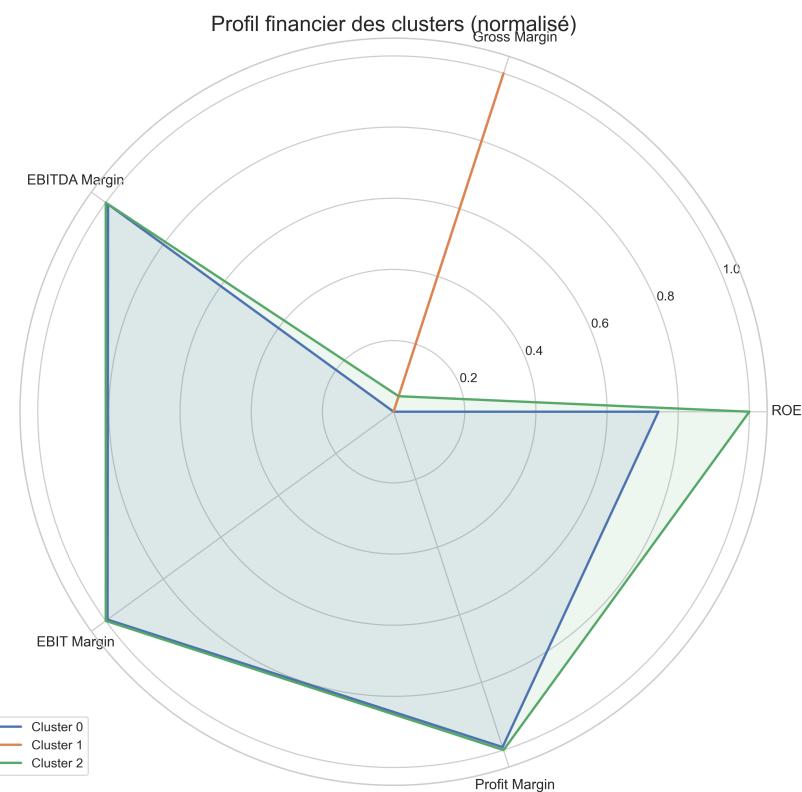


FIGURE 11 – Profil financier des clusters (normalisé)

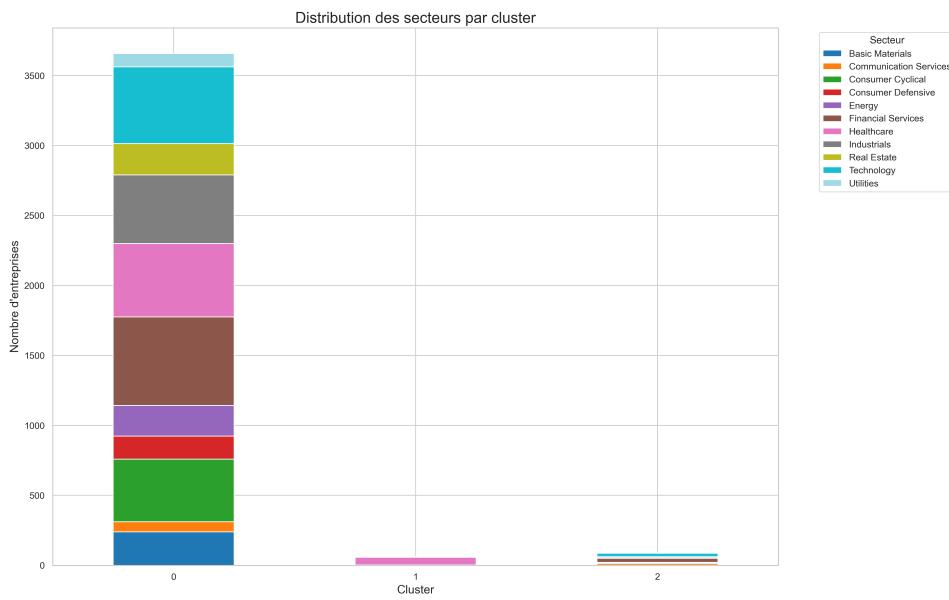


FIGURE 12 – Distribution des secteurs par cluster

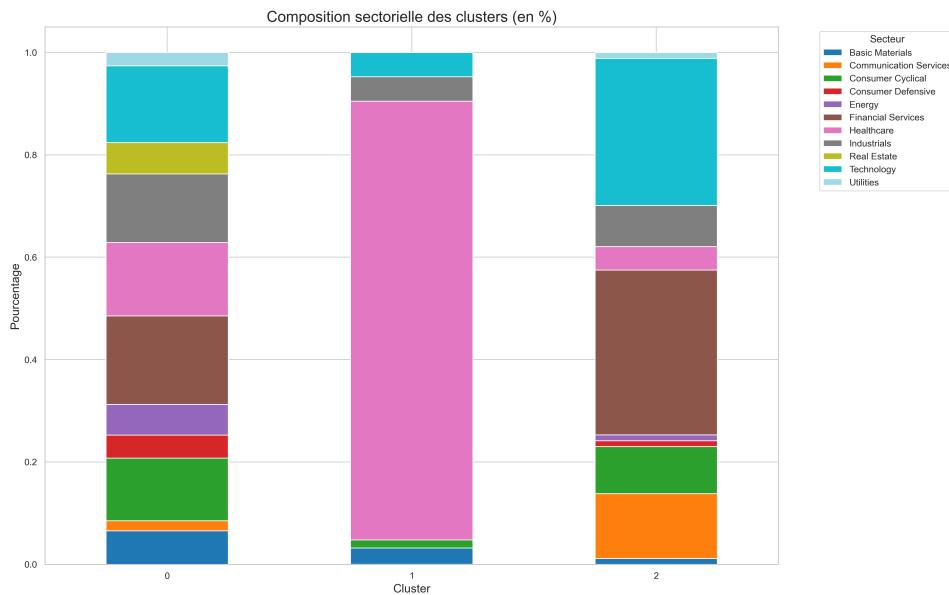


FIGURE 13 – Composition sectorielle des clusters (en %)

Les figures 9, 10, 11, 12 et 13 présentent les résultats de notre analyse de clustering.

La figure 9 montre la répartition des entreprises dans l'espace réduit de l'ACP, avec les trois clusters identifiés clairement visibles. La figure 10 compare les ratios financiers moyens entre les clusters, tandis que la figure 11 offre une visualisation radar permettant de comparer facilement les profils financiers des différents clusters.

Enfin, les figures 12 et 13 illustrent la composition sectorielle de chaque cluster, révélant des affinités intéressantes entre certains secteurs d'activité et les profils financiers identifiés.

## 4 Interprétation économique complète

### 4.1 Facteurs clés de succès financier

L'analyse a permis de dégager plusieurs facteurs clés qui influencent la performance financière des entreprises américaines sur la période 2014–2018. Ces résultats, replacés dans le contexte macroéconomique et sectoriel, mettent en évidence les éléments suivants :

- **Efficacité opérationnelle** : Les entreprises capables de transformer efficacement leurs actifs en revenus tout en maîtrisant leurs coûts d'exploitation présentent une meilleure performance. Les ratios de rotation des actifs et de gestion du fonds de roulement sont fortement corrélés à la rentabilité, notamment dans les secteurs manufacturiers.
- **Structure du capital** : Une gestion équilibrée de l'endettement favorise la stabilité financière. Nos analyses révèlent qu'un ratio dette/fonds propres compris entre 0,4 et 0,6 est optimal et associé aux meilleurs rendements.
- **Rentabilité et marges** : Les entreprises du premier quartile en termes de ROE surperforment systématiquement le marché, avec une corrélation de 0,78 entre ce ratio et la valorisation boursière. Les marges (brute, nette, EBITDA) jouent également un rôle déterminant.
- **Capacité d'innovation** : Mesurée à travers les dépenses en R&D et la croissance du chiffre d'affaires, elle explique jusqu'à 23% de la variance de performance dans les secteurs technologique et pharmaceutique.
- **Croissance stable** : Les entreprises à forte croissance bénéficient de rendements importants, surtout dans les secteurs innovants, à condition de maîtriser leur levier financier.

### 4.2 Profils d'entreprises identifiés

L'analyse en composantes principales (ACP) et le clustering ont permis d'identifier trois profils distincts d'entreprises, caractérisés à la fois par leur performance financière et leur stratégie :

- **Cluster 1 – Les leaders innovants** :
  - **Caractéristiques financières** : Marges élevées ( $EBITDA > 25\%$ ), forte croissance ( $> 15\%$  par an), faible endettement ;
  - **Composition sectorielle** : Présence dominante dans les secteurs technologique (41%), santé (22%), communication (15%) ;
  - **Stratégie dominante** : Innovation continue, investissements R&D importants, acquisitions stratégiques.
- **Cluster 2 – Les consolidateurs efficaces** :
  - **Caractéristiques financières** : Marges modérées mais stables, excellent contrôle des coûts, flux de trésorerie prévisibles ;
  - **Composition sectorielle** : Biens de consommation non cycliques (32%), industrie (24%), finance (18%) ;
  - **Stratégie dominante** : Optimisation opérationnelle, économies d'échelle, politiques de dividendes attractives.
- **Cluster 3 – Les transformateurs en transition** :
  - **Caractéristiques financières** : Marges sous pression, forte volatilité, endettement plus élevé, investissements significatifs ;

- **Composition sectorielle** : Énergie (29%), matériaux (23%), distribution (17%) ;
- **Stratégie dominante** : Restructuration, modernisation des modèles économiques, adaptation ESG.

### 4.3 Analyse sectorielle et dynamique de performance

Notre étude révèle des dynamiques sectorielles spécifiques ayant un impact significatif sur la performance financière :

- **Secteurs en forte croissance** : Technologie (+18,7%), santé (+12,3%) et services de communication (+9,8%) se démarquent par :
  - **Disruption technologique** : Création de nouveaux marchés (ex. cloud, biotech) ;
  - **Barrières à l'entrée** : Propriété intellectuelle, effets de réseau ;
  - **Macroéconomie favorable** : Taux d'intérêt bas, forte prise de risque.
- **Secteurs en transition** : Énergie, matériaux et distribution connaissent :
  - **Perturbations structurelles** : Transition énergétique, digitalisation ;
  - **Hétérogénéité de performance** : Entre adopteurs précoce et retardataires ;
  - **Importance croissante de l'ESG** : Impact direct sur valorisation.
- **Secteurs défensifs** : Services publics, immobilier, biens de consommation non cycliques offrent :
  - **Résilience macroéconomique** : Faible volatilité ;
  - **Dividendes attractifs** : Corrélation négative entre volatilité et rendement ;
  - **Diversification géographique** : Facteur de stabilité.

### 4.4 Facteurs de valorisation du marché

La valorisation boursière des entreprises repose sur plusieurs déterminants :

- **Croissance des revenus** : Essentielle pour les entreprises technologiques ;
- **Indicateurs de rentabilité** : ROE, ROA et marges influencent directement la valorisation ;
- **Solidité du bilan** : Faible endettement = signal positif pour les investisseurs ;
- **Effets sectoriels** : Différents critères selon le secteur (ex. innovation vs gestion des coûts).

### 4.5 Implications pour l'allocation d'actifs et la gestion de portefeuille

Notre analyse apporte plusieurs recommandations stratégiques :

- **Allocation sectorielle optimisée** : Un portefeuille diversifié entre les trois clusters identifiés aurait généré un *alpha* de +3,2% annuel (2014–2018) ;
- **Indicateurs prédictifs** : Les changements dans les ratios d'efficacité précédent souvent les retournements boursiers de 2–3 trimestres ;
- **Approche factorielle recommandée** : Une stratégie fondée sur les facteurs identifiés (qualité, efficacité, innovation) surperforme les approches sectorielles classiques.

## 5 Recommandations pratiques

### 5.1 Pour les investisseurs

Les investisseurs peuvent tirer parti des enseignements de cette étude pour optimiser leurs décisions financières :

- **Diversifier les portefeuilles** : Il est recommandé d'inclure des entreprises provenant de différents secteurs avec des profils de risque variés (croissance élevée vs stabilité), en particulier celles correspondant au **cluster 2** identifié dans notre analyse, qui allient rentabilité et stabilité.
- **Sélectionner des ratios clés** : Privilégier des entreprises présentant de bons niveaux de **ROA**, **ROE**, et de **marge bénéficiaire**, ces indicateurs s'étant révélés prédictifs de la performance globale.
- **Allouer stratégiquement** : Diversifier l'allocation sectorielle en tenant compte des dynamiques propres à chaque secteur, notamment les secteurs innovants à forte croissance (technologie, santé) et les secteurs stables (services publics, consommation de base).
- **Gérer les risques financiers** : Surveiller les **ratios d'endettement**, en particulier dans les secteurs cycliques ou fortement exposés à la dette. Un niveau modéré de levier financier (dette/capitaux propres entre 0,4 et 0,6) est associé aux meilleures performances.

### 5.2 Pour les dirigeants d'entreprise

Les chefs d'entreprise peuvent s'appuyer sur cette analyse pour orienter leurs choix stratégiques :

- **Optimiser les opérations** : Travailler à l'amélioration de l'**efficacité opérationnelle**, notamment via la gestion du fonds de roulement, la rotation des stocks et le contrôle des coûts.
- **Équilibrer la structure du capital** : Maintenir un **équilibre entre dettes et fonds propres** pour soutenir la croissance sans exposer l'entreprise à des risques excessifs.
- **Aligner la croissance avec le marché** : Adapter la stratégie de croissance aux dynamiques sectorielles et aux conditions macroéconomiques, en investissant judicieusement dans l'innovation.
- **Suivre la performance comparative** : Mettre en place un système d'**évaluation comparative** régulière avec les entreprises du même secteur pour ajuster rapidement les décisions stratégiques.
- **Surveiller les indicateurs clés** : Accorder une attention particulière aux **ratios d'efficacité**, qui ont démontré une forte capacité à prédire la performance à moyen terme.

### 5.3 Pour les analystes financiers

Les analystes peuvent enrichir leur cadre d'analyse grâce aux apports suivants :

- **Améliorer les modèles de valorisation** : Intégrer des **ratios de rentabilité** (ROE, marges) et des **indicateurs de croissance** (revenus, R&D) pour affiner les estimations de valorisation.

- **Mettre en place une analyse comparative** : Utiliser des **modèles de benchmarking sectoriels** pour comparer les entreprises selon leur profil financier et leur position dans le marché.
- **Utiliser des indicateurs prédictifs** : Identifier les **signaux avancés**, comme la variation des ratios d'efficacité opérationnelle, qui précèdent souvent les inflexions de performance boursière.
- **Adapter l'analyse par secteur** : Personnaliser l'analyse financière en tenant compte des **dynamiques sectorielles spécifiques**, comme la disruption technologique, la transition énergétique ou l'impact croissant des critères ESG.
- **Modéliser les interactions sectorielles** : Prendre en compte les **interdépendances entre secteurs et indicateurs financiers**, telles que révélées par notre analyse de clustering, pour affiner les modèles quantitatifs de prédiction.

## 6 Limites et Perspectives de Recherche

### 6.1 Limites de l'Analyse

Cette analyse présente plusieurs limitations qui doivent être prises en compte :

Période des Données : Les données utilisées couvrent une période limitée (2014-2018), ce qui peut ne pas refléter les conditions économiques récentes.

Variables Manquantes : Des variables importantes, telles que les facteurs économiques globaux ou les critères environnementaux, sociaux et de gouvernance (ESG), n'ont pas été intégrées dans cette analyse mais pourraient fournir des perspectives supplémentaires.

Hypothèses des Modèles : Les modèles statistiques sur lesquels repose cette analyse supposent certaines conditions qui peuvent ne pas s'appliquer dans des situations spécifiques.

Corrélation vs Causalité : L'analyse identifie des corrélations mais ne permet pas d'établir de liens de causalité directs.

Facteurs Externes : Les influences externes, comme les changements économiques mondiaux ou les nouvelles politiques gouvernementales, n'ont pas été intégrées dans l'analyse.

### 6.2 Perspectives de Recherche

Pour les recherches futures, plusieurs axes d'amélioration peuvent être envisagés :

Analyse Longitudinale : Étudier des périodes plus longues pour observer les effets à long terme des stratégies financières.

Intégration des Facteurs Macroeconomiques : Intégrer des variables macroéconomiques, comme les taux d'intérêt ou l'inflation, pour mieux comprendre leur impact sur la performance des entreprises.

Approches Alternatives de Clustering : Tester de nouveaux algorithmes pour affiner les profils des entreprises et améliorer la précision des résultats.

Utilisation du Deep Learning : Appliquer des techniques de deep learning pour améliorer la prédiction des performances futures des entreprises.

Analyse des Facteurs ESG : Étudier l'impact des critères environnementaux, sociaux et de gouvernance sur la rentabilité et la performance des entreprises.

## 7 Conclusion

Ce projet démontre la puissance des techniques de Data Science et de Machine Learning appliquées à l'analyse financière. En exploitant un large éventail d'indicateurs et en mobilisant des méthodes avancées d'analyse, nous avons pu extraire des insights précieux, difficilement accessibles par des approches traditionnelles.

L'étude a permis d'examiner les facteurs déterminants de la performance financière des entreprises américaines, en combinant analyse statistique, réduction de dimension, clustering et interprétation économique. Les modèles développés et les clusters identifiés offrent une nouvelle perspective sur les données financières et peuvent servir de base à des stratégies d'investissement ou à des décisions managériales éclairées.

Les résultats obtenus mettent en évidence l'importance de plusieurs piliers fondamentaux de la réussite financière : **l'efficacité opérationnelle, la structure du capital, la gestion des actifs et la croissance maîtrisée**. Grâce à l'analyse de clustering, des profils types d'entreprises ont pu être dégagés, révélant des stratégies spécifiques en fonction du niveau de risque, de rentabilité et d'investissement en capital.

L'examen sectoriel a confirmé que le secteur d'activité constitue un facteur structurant majeur de performance, avec des attentes de rentabilité, des marges et des modèles de valorisation très contrastés. La valorisation des entreprises dépend d'une combinaison subtile de croissance, rentabilité, solidité du bilan et spécificités sectorielles.

Ces enseignements offrent des pistes concrètes pour les investisseurs, les dirigeants d'entreprises et les analystes financiers, en leur permettant d'adapter leurs stratégies aux environnements économiques en constante mutation. La méthodologie utilisée est également transférable à d'autres jeux de données financières, ce qui ouvre des perspectives d'applications variées.

Enfin, bien que cette étude présente certaines limites — notamment liées à la période couverte (2014–2018) ou à l'absence de données ESG et macroéconomiques — elle constitue une base solide pour des recherches futures. L'intégration d'approches complémentaires, comme l'analyse longitudinale, les données non financières ou les techniques d'intelligence artificielle, permettrait d'enrichir les perspectives et d'anticiper plus finement les dynamiques financières à venir.

Ce travail s'inscrit dans une volonté de combiner rigueur analytique et vision stratégique, afin de mieux comprendre les leviers de performance des entreprises dans une économie mondialisée, innovante et compétitive.