

Age et finance

Etudes de statistiques descriptive

Projet de recherche réalisé dans le cadre du master 1 MECEN



Université de Tours
Par Alexis VINCENT, Jawad GRIB et Abdul BALOGOUN
2024-2025

Table des Matières

1	Introduction	2
1.1	Contexte	2
1.2	Problématique	2
1.3	Objectifs	2
1.4	Méthodologie	3
1.5	Plan	3
2	Revue de la littérature	3
3	Hypothèses	4
4	Préparation des données et analyse	4
4.1	Les variables	4
4.2	ACP	5
4.3	ACM	8
4.4	Clustering	10
5	Discussion	13
6	Conclusion	13
7	Annexe	14
7.1	Récupération des données	14

1 Introduction

L'existence des commerces et la création de nouveaux commerces en France se sont intensifiées ces dernières années. En effet, selon l'INSEE, 1 027 831 commerces différents ont été créés en 2022, contre 547 531 en 2013. Ainsi, en 10 ans, on observe une forte augmentation du nombre de commerces. Cela soulève alors deux questions : quels secteurs sont principalement concernés par cette croissance et comment ces commerces sont-ils répartis sur le territoire français ? Nous avons tenté de répondre à ces questions lors d'un travail précédent, avec la réalisation d'un tableau de bord accessible à ce [lien](#).

Maintenant que nous avons une vue d'ensemble du tissu commercial français, une nouvelle question se pose : comment ces commerces évoluent-ils individuellement ? Existe-t-il des commerces en France présentant des caractéristiques financières et structurelles similaires ? Dans cette étude, nous allons nous intéresser aux liens entre les variables financières et structurelles d'un échantillon d'entreprises présentes en France et rechercher des dynamiques qui pourraient regrouper certaines d'entre elles.

1.1 Contexte

Dans un travail précédent, nous avons développé un tableau de bord interactif permettant aux collectivités locales de visualiser facilement la répartition et la concentration des types de commerces les plus représentés sur le territoire français (hors DOM/TOM). Ce tableau de bord repose sur une base de données ouverte, fournissant des informations sur plus de 650 000 entreprises, mise à disposition par OpenStreetMap.

Dans le cadre de cette étude, nous nous concentrerons sur un échantillon d'une trentaine d'entreprises françaises pour analyser les liens entre quatre variables financières (chiffre d'affaires, bénéfice, capital, actifs immobilisés) et deux variables structurelles (nombre d'employés et date de création). Nous chercherons également à identifier des similitudes et des dynamiques communes entre ces commerces à partir de ces caractéristiques.

1.2 Problématique

Nous allons donc nous poser les deux questions suivantes :

- Existe-t-il un lien entre les différentes variables financières ?
- La date de création d'une entreprise a-t-elle un impact sur ses performances financières ?

1.3 Objectifs

Pour répondre à cette problématique, nous utiliserons des méthodes d'analyse et de classification non supervisée, telles que :

- **L'Analyse en Composantes Principales (ACP)** : Cette méthode permet de réduire la dimensionnalité des données tout en conservant la variance maximale, facilitant ainsi l'interprétation des relations entre les variables.
- **L'Analyse des Correspondances Multiples (ACM)** : Elle permet d'analyser des jeux de données comprenant plusieurs variables qualitatives, en identifiant des groupes d'individus ou d'éléments partageant des caractéristiques similaires.
- **La Classification Hiérarchique** : Cette méthode permet de regrouper les individus en classes hiérarchisées en fonction de leur similarité, ce qui permet de découvrir des structures sous-jacentes dans les données.

Ces méthodes nous permettront d'explorer et de comprendre les relations complexes entre les variables financières et d'identifier des tendances et des regroupements significatifs.

1.4 Méthodologie

Pour répondre à cette problématique, une approche quantitative sera utilisée. L'analyse des données financières d'entreprises permettra d'identifier les relations entre ces variables au travers d'outils statistiques.

1.5 Plan

Dans un premier temps, nous explorerons diverses recherches effectuées sur ces sujets afin de comparer nos hypothèses avec celles des études existantes.

Ensuite, nous analyserons nos données en appliquant les méthodes statistiques suivantes :

- **ACP** : Étude des liens entre les variables financières et la date de création de l'entreprise.
- **ACM** : Étude des liens entre les variables financières et la date de création de l'entreprise sous forme catégorielle (pour réduire l'impact des chocs temporels et renforcer la robustesse de l'étude).
- **Clustering** : Application sur les résultats de l'ACM afin d'observer des similitudes sous-jacentes parmi les individus de notre jeu de données.

Et pour terminer, nous conclurons sur les différents résultats.

2 Revues de la littératures

"Le profit découle directement de l'excédent des recettes sur les coûts." - Jean-Baptiste Say (1803)

Jean-Baptiste Say nous explique qu'il existe un lien entre le revenu et le chiffre d'affaires assez basique : sans chiffre d'affaires suffisant, les profits sont impossibles. En d'autres termes, cela indique explicitement que le profit est lié positivement au chiffre d'affaires.

"Le chiffre d'affaires d'une entreprise est déterminé par la demande effective, et c'est cette même demande qui conditionne les niveaux de profit." - John Maynard Keynes (1936)

Keynes nous explique que la demande des consommateurs génère le chiffre d'affaires, et que c'est cette même demande qui est au cœur de tout problème de maximisation de profit, montrant ainsi un lien entre profit et chiffre d'affaires.

"Les capitaux propres correspondent à l'ensemble des ressources de la société, ils reflètent sa valeur financière. Il s'agit des fonds apportés par les associés ou actionnaires lors de la création de la société auxquels s'ajoutent les fonds générés par son activité." – Définition des capitaux propres

On voit donc que les capitaux propres sont incrémentés des fonds générés par l'activité de la société. En d'autres termes, ces deux variables sont censées être liées.

"L'âge de l'entreprise est généralement corrélé positivement à sa rentabilité, car les entreprises plus anciennes bénéficient souvent d'une meilleure stabilité financière, de processus plus affinés et d'une plus grande réputation sur le marché, ce qui leur permet de générer des profits plus élevés." – Penrose, E.T. (1959), *The Theory of the Growth of the Firm*

Selon Penrose, les entreprises plus anciennes ont souvent acquis une stabilité, une réputation et des compétences qui leur permettent de mieux gérer les défis économiques et d'optimiser leur rentabilité. Cette maturité leur offre des avantages concurrentiels qui peuvent se traduire par une croissance des

profits à long terme.

"Les entreprises qui disposent d'un nombre élevé d'employés, en particulier ceux ayant des compétences techniques spécifiques, sont souvent capables de mieux gérer et d'exploiter les actifs immobilisés, augmentant ainsi leur rentabilité." – Barney, J.B. (1991), *Firm Resources and Sustained Competitive Advantage*

On note ici un lien entre actifs immobilisés, nombre d'employés et profit engendré par ces derniers.

Dans la littérature économique, plusieurs théories mettent en évidence les relations entre les variables financières et les performances des entreprises. Jean-Baptiste Say (1803) souligne que le profit est positivement lié au chiffre d'affaires, affirmant que sans un chiffre d'affaires suffisant, les profits sont impossibles. Keynes (1936) confirme cette idée en expliquant que la demande des consommateurs détermine directement le chiffre d'affaires, et donc les niveaux de profit. De plus, les capitaux propres, qui représentent les ressources de l'entreprise, sont intrinsèquement liés à la rentabilité et au développement de l'entreprise. Par ailleurs, Penrose (1959) et Barney (1991) suggèrent qu'un âge avancé d'entreprise, associé à une base de capital humain plus large et spécialisée, permet une meilleure gestion des actifs immobilisés, ce qui contribue à des profits plus élevés. Ces éléments soulignent non seulement la corrélation entre le capital propre et le profit, mais aussi l'impact de l'âge et de la taille de l'entreprise sur sa rentabilité financière.

Ainsi, notre étude visera à observer si ces constatations économiques peuvent être vérifiées dans le cadre de notre analyse des liens entre les variables financières et structurelles des entreprises.

3 Hypothèses

- H1 : Les variables financières des entreprises sont corrélées.
- H2 : L'âge d'une entreprise impacte positivement sa puissance financière.

4 Préparation des données et analyse

4.1 Les variables

Nous nous intéressons à un jeu de données composé de 31 entreprises présentes sur le territoire français, chacune associée aux six variables suivantes :

- Chiffre d'affaires (Revenue)
- Profit
- Capitaux propres (Capital)
- Actifs immobilisés (Assets)
- Date de création (Inceptions)
- Nombre d'employés (Employees)

Voici un extrait de ce jeu de données :

	revenue	inception	employees	profit	capital	assets
Accor	4.22400e+09	1967	14466.0	402000000	4549000000	1.066900e+10
Adidas	2.14270e+10	1924	59030.0	-75000000	4580000000	1.802000e+10
American Express	5.28620e+10	1850	64000.0	7514000000	19584069000	1.665030e+11
BP	2.41392e+11	1909	70100.0	27653000000	82990000000	2.881200e+11
Bank of America	9.85810e+10	1784	208666.7	26515000000	282283000000	3.180151e+12
Caterpillar	5.94270e+10	1925	104000.0	6705000000	14080000000	7.850900e+10

Pour l'ACM, nous utiliserons le tableau suivant :

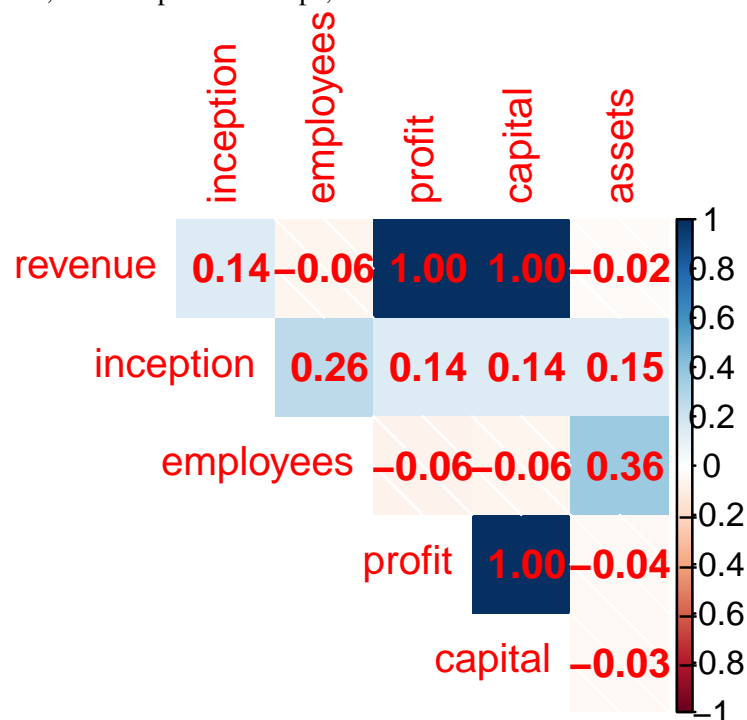
	revenue	inception	employees	profit	capital	assets
Accor	Faible	Récente	Peu	Faible	Faible	Faible
Adidas	Faible	Intermédiaire	Peu	Faible	Faible	Faible
American Express	Moyenne	Ancienne	Peu	Moyenne	Moyenne	Moyenne
BP	Élevée	Ancienne	Peu	Élevée	Élevée	Moyenne
Bank of America	Moyenne	Ancienne	Beaucoup	Élevée	Élevée	Élevée
Caterpillar	Moyenne	Intermédiaire	Moyenne	Moyenne	Faible	Moyenne

Les données de ce tableau ont été transformées en données qualitatives en utilisant les terciles comme séparateurs.

Les données sont récupérables avec une requête SparkSQL (en annexe), [avec notre jeu de données initial](#) sur les codes de la colonne Wikidata.

4.2 ACP

Nous allons étudier, dans un premier temps, les corrélations entre les variables :



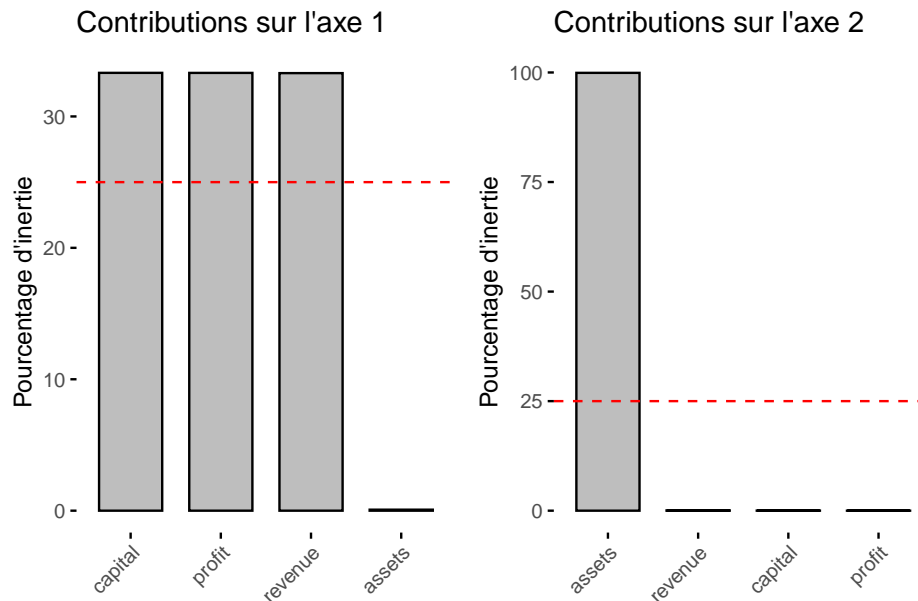
Nous observons un lien étroit entre les variables financières telles que le profit, le capital et le revenu. De plus, l'année de création de l'entreprise semble également corrélée à ces variables. Nous allons donc réaliser une analyse en composantes principales (ACP) pour examiner ces différentes relations.

Etudes des inerties:

	F1	F2	F3	F4
Inertie	3.00	1.00	0e+00	0
Pourcentage d'inertie	75.02	24.96	1e-02	0
Pourcentage cumulé d'inertie	75.02	99.99	1e+02	100

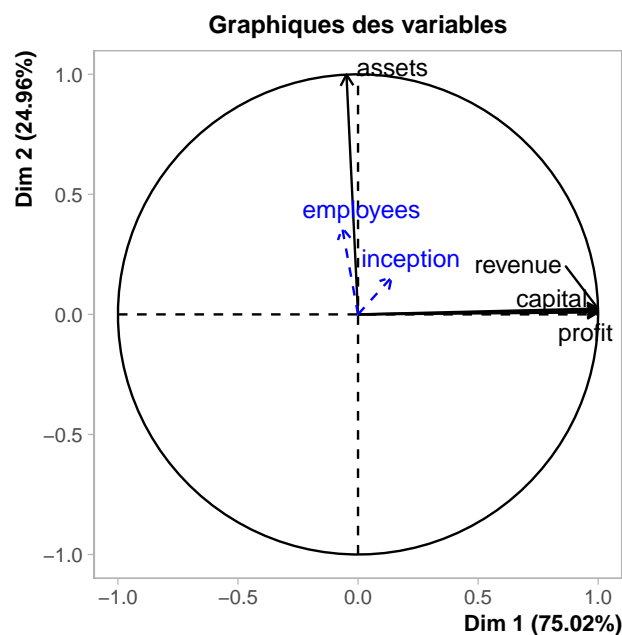
Nous étudierons uniquement les 2 premiers axes qui conservent à la fois 99% de la part d'inertie cumulée, avec des valeurs propres sur l'axe 1 de 3 et de 1 sur l'axe 2.

Étudions les contributions des variables :



On observe que l'axe 1 semble capter les variables plutôt financières, tandis que l'axe 2 semble quant à lui capter les variables structurelles de l'entreprise.

Analysons maintenant le graphique des variables :



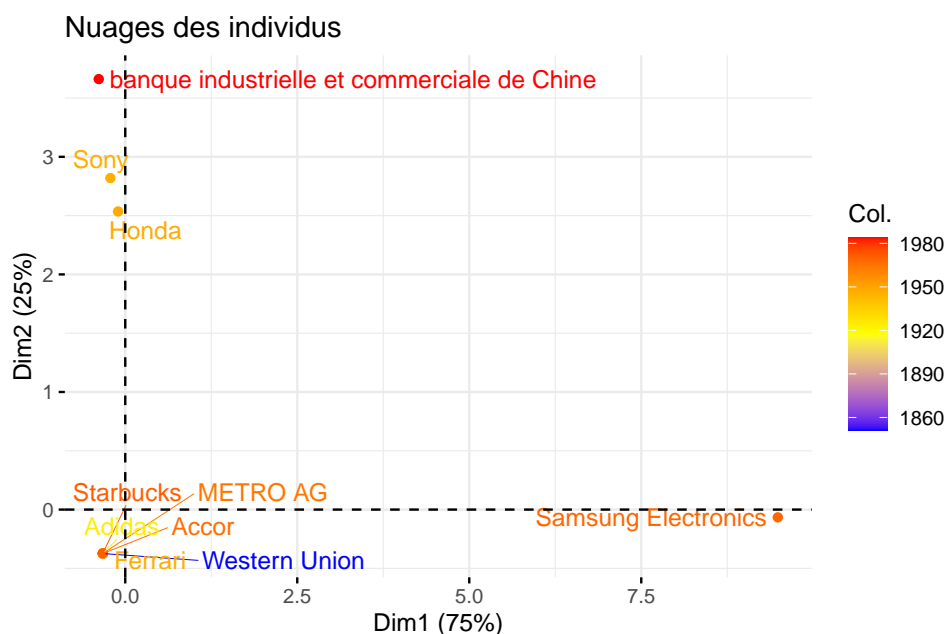
On observe que les variables financières (revenus, capital, profit) semblent corrélées lorsqu'on compare les entreprises en fonction de leur niveau d'actifs. De plus, les variables quantitatives supplémentaires révèlent une relation entre le nombre d'employés et les actifs. Enfin, la date de création de l'entreprise semble être légèrement liée aux variables financières.

L'axe 1 semble discriminer les entreprises selon leur puissance financière, avec les moins puissantes financièrement situées dans la partie négative et les plus puissantes dans la partie positive. On peut donc supposer que cet axe représente, en composante principale, la puissance financière des entreprises.

L'axe 2, quant à lui, semble traduire une différenciation basée sur le niveau des actifs.

Cette analyse en composantes principales (ACP) nous fournit un premier élément de réponse quant aux corrélations entre les variables financières. Il est également intéressant de noter que les actifs semblent légèrement corrélés, à la fois au nombre d'employés et à la date de création de l'entreprise.

Il reste maintenant à déterminer si la puissance financière est positivement corrélée à l'ancienneté des entreprises. Pour visualiser cela, nous allons représenter les individus en les habillant avec la variable inception (date de création).



On observe que quelques individus surclassent la majorité de la population, notamment la Banque industrielle et commerciale de Chine, Sony, Honda et Samsung Electronics.

Cela est probablement dû au fait que ces individus possèdent des valeurs extrêmes par rapport au reste de la population. Nous remarquons également que la date de création d'une entreprise ne semble pas liée à sa puissance financière.

Pour atténuer l'impact de ces valeurs extrêmes, nous allons transformer les variables quantitatives en variables qualitatives. Ensuite, nous procéderons à une Analyse des Correspondances Multiples (ACM) afin d'explorer la répartition des entreprises selon leurs caractéristiques communes.

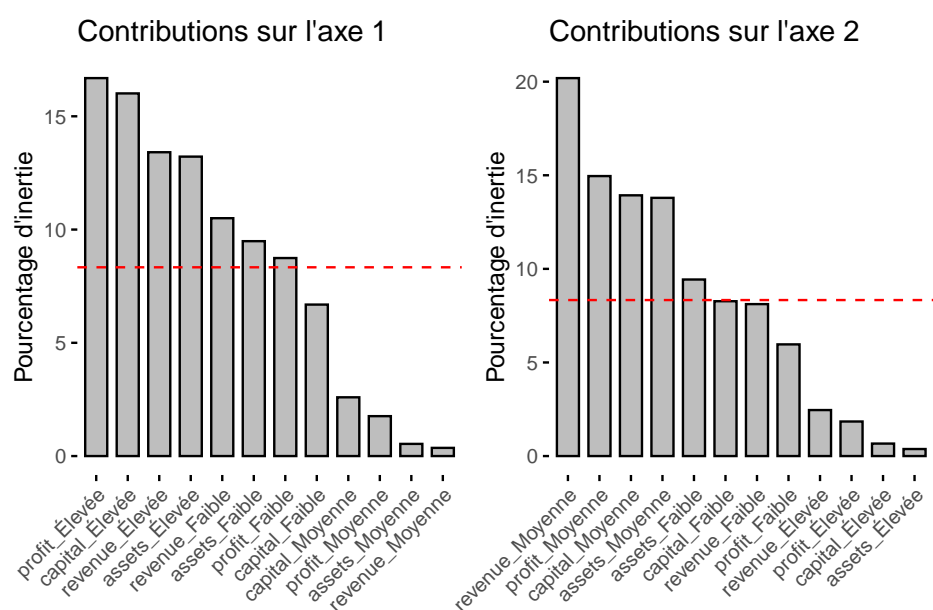
4.3 ACM

Nous allons donc, comme pour l'ACP, commencer par analyser le tableau des inerties.

	F1	F2	F3	F4	F5	F6	F7	F8
Inertie	0.83	0.58	0.25	0.12	0.09	0.07	0.05	0.02
Pourcentage d'inertie	41.57	29.12	12.37	5.96	4.52	3.31	2.35	0.82
Pourcentage cumulé d'inertie	41.57	70.69	83.06	89.01	93.53	96.84	99.18	100.00

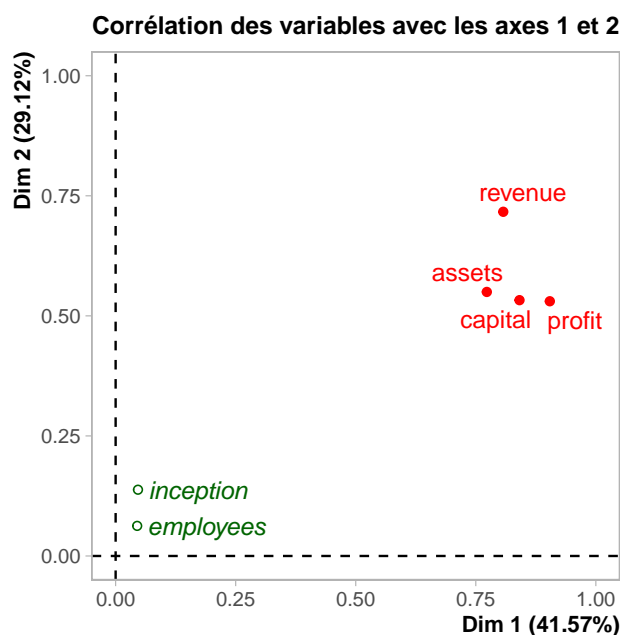
Nous allons étudier ici l'axe 1 et l'axe 2, qui expliquent à eux seuls plus de 70% de l'inertie.

Voici les variables qui contribuent le plus à chaque axes:



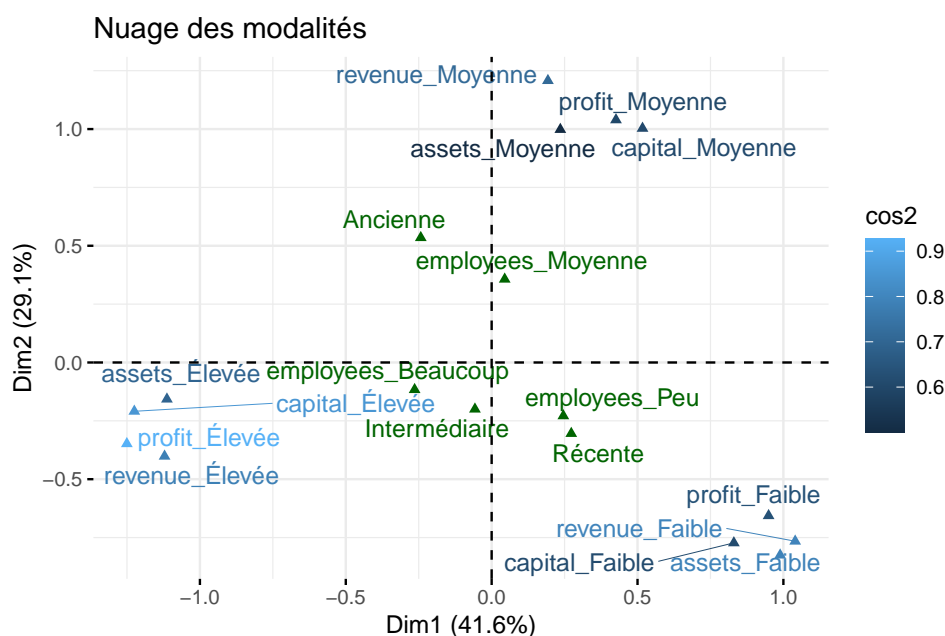
Nous ne pouvons observer de schéma particulier entre les modalités des variables. Le plan semble regrouper les individus selon des profils précis.

Nous allons maintenant étudier le graphique des corrélations des variables.



Ce graphique nous montre que les deux axes semblent expliquer en même temps les variables financières. On suppose par ailleurs que, grâce à leur proximité, elles semblent être liées les unes aux autres.

En revenant sur les résultats des graphiques précédents, l'axe 1 semble davantage capter les valeurs extrêmes des individus (élevées et faibles), tandis que l'axe 2 semble plutôt représenter leur niveau de profil moyen entre les variables financières. Pour mieux visualiser cela, nous allons afficher le graphique des variables.

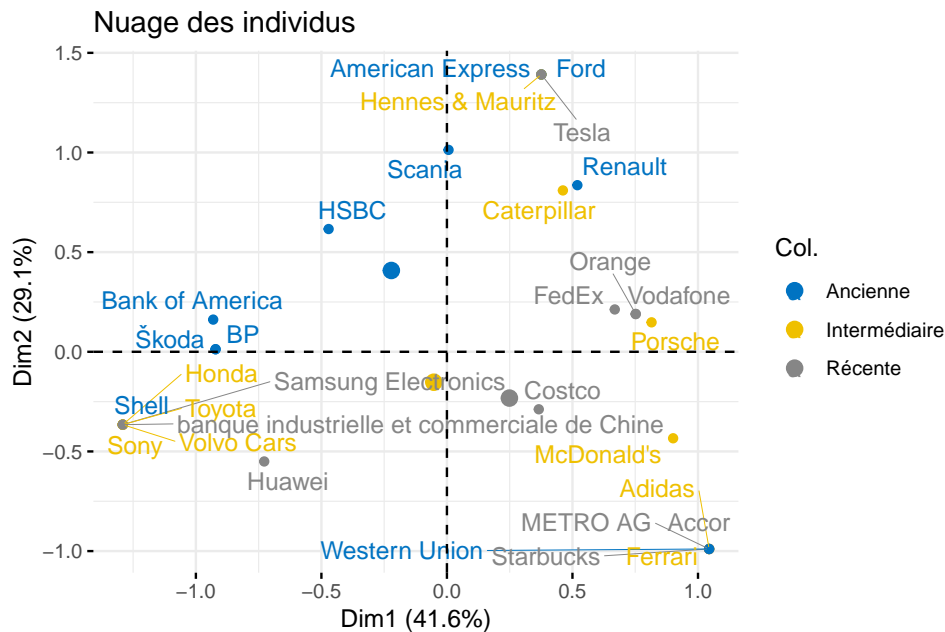


Nous observons ici la séparation des individus en 3 groupes principaux : les profils faibles (à droite), moyens (en haut) et élevés (à gauche). Par ailleurs, on peut noter que le nombre d'employés semble tendre vers une valeur élevée selon la puissance financière (ce qui est assez logique, car avec une puissance financière faible, on ne peut pas payer beaucoup d'employés). Cependant, l'année de création ne semble pas suivre notre hypothèse de départ, à savoir que plus une entreprise est ancienne, plus elle a

une puissance financière élevée.

Grâce à ce graphique et aux résultats de l'ACP précédemment effectuée, nous pouvons confirmer que les variables financières, à savoir les actifs, le capital, le profit et le chiffre d'affaires, sont liées entre elles.

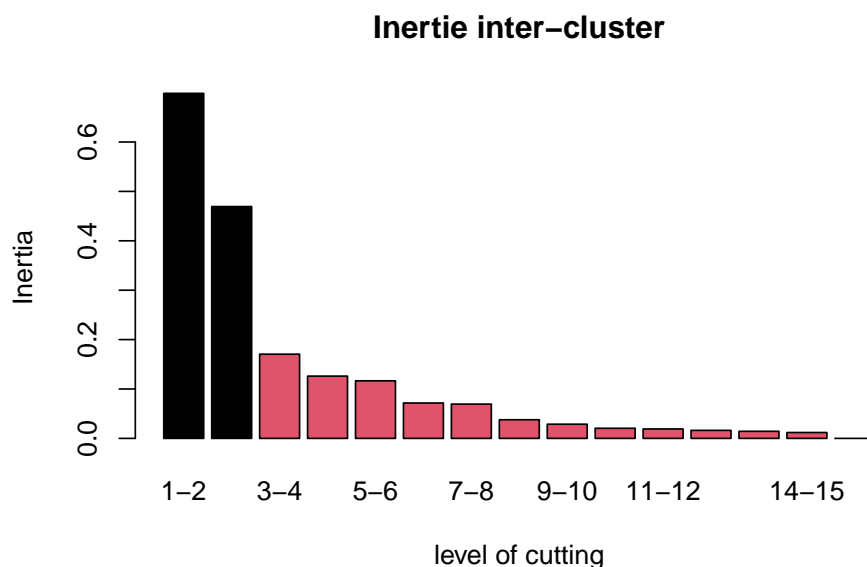
Pour pouvoir valider notre seconde hypothèse, nous allons visualiser les individus sur ce plan en les habillant par années de création.



Nous pouvons voir que les entreprises anciennes se répartissent davantage dans la partie haute/gauche du plan. On pourrait donc supposer que notre hypothèse de départ sur l'âge d'une entreprise et sa puissance financière est vérifiée. Cependant, comme on peut observer, les entreprises récentes et intermédiaires sont, quant à elles, réparties uniformément sur tout le plan. Nous sommes donc dans l'impossibilité de généraliser cette hypothèse au vu des résultats de l'ACP et de l'ACM. Elle est donc rejetée.

4.4 Clustering

Dans cette section, nous allons classer les individus selon leurs caractéristiques communes. Pour cela, nous utiliserons une méthode de clustering hiérarchique sur notre ACM. Afin de choisir le nombre de clusters optimal, nous étudierons l'inertie inter-cluster.



Nous observons que passer de 2 à 3 clusters génère un gain d'inertie encore très satisfaisant. Nous allons donc opter pour une classification en 3 classes.

Ainsi, voici les modalités qui construisent ces clusters :

	Cla/Mod	Mod/Cla	Global	p.value	v.test
profit=profit_Élevée	100.00	100.00	35.48	0.00	5.70
capital=capital_Élevée	90.91	90.91	35.48	0.00	4.70
revenue=revenue_Élevée	90.91	90.91	35.48	0.00	4.70
assets=assets_Élevée	81.82	81.82	35.48	0.00	3.83
capital=capital_Faible	10.00	9.09	32.26	0.05	-1.96
revenue=revenue_Moyenne	10.00	9.09	32.26	0.05	-1.96
assets=assets_Faible	0.00	0.00	32.26	0.00	-2.87
capital=capital_Moyenne	0.00	0.00	32.26	0.00	-2.87
profit=profit_Moyenne	0.00	0.00	32.26	0.00	-2.87
profit=profit_Faible	0.00	0.00	32.26	0.00	-2.87
revenue=revenue_Faible	0.00	0.00	32.26	0.00	-2.87

	Cla/Mod	Mod/Cla	Global	p.value	v.test
assets=assets_Moyenne	80.00	80	32.26	0.00	3.69
capital=capital_Moyenne	80.00	80	32.26	0.00	3.69
revenue=revenue_Moyenne	80.00	80	32.26	0.00	3.69
profit=profit_Moyenne	70.00	70	32.26	0.00	2.88
employees=employees_Moyenne	60.00	60	32.26	0.04	2.10
capital=capital_Élevée	9.09	10	35.48	0.05	-1.96
assets=assets_Faible	0.00	0	32.26	0.01	-2.65
profit=profit_Élevée	0.00	0	35.48	0.00	-2.87
revenue=revenue_Élevée	0.00	0	35.48	0.00	-2.87

	Cla/Mod	Mod/Cla	Global	p.value	v.test
assets=assets_Faible	100.00	100	32.26	0.00	5.59
capital=capital_Faible	80.00	80	32.26	0.00	3.69
revenue=revenue_Faible	80.00	80	32.26	0.00	3.69
profit=profit_Faible	70.00	70	32.26	0.00	2.88
revenue=revenue_Élevée	9.09	10	35.48	0.05	-1.96
assets=assets_Moyenne	0.00	0	32.26	0.01	-2.65
assets=assets_Élevée	0.00	0	35.48	0.00	-2.87
capital=capital_Élevée	0.00	0	35.48	0.00	-2.87
profit=profit_Élevée	0.00	0	35.48	0.00	-2.87

Nous remarquons que cette répartition suit principalement un schéma de groupes aux puissances financières élevées, moyennes dans un autre et faibles dans le dernier.

Voici comment se répartissent les individus dans les clusters :

	Cluster1	Cluster2	Cluster3
Honda	0.26	American Express	0.54
Samsung Electronics	0.26	Ford	0.54
Shell	0.26	Hennes & Mauritz	0.54
Sony	0.26	Tesla	0.54
Toyota	0.26	Renault	0.85
		Adidas	0.49
		Ferrari	0.49
		METRO AG	0.49
		Starbucks	0.49
		Western Union	0.49

En se concentrant sur le secteur de l'automobile, étant prédominant dans notre jeu de données. Nous allons essayer de comprendre comment et pourquoi ces entreprises ont été classées ensemble d'un point de vue économique léger.

Dans le cluster à la structure financière la plus solide, on retrouve Honda. Ce choix est cohérent, car Honda est l'un des leaders mondiaux de l'automobile, avec une production de masse destinée à un large public. Sa stratégie consiste à proposer des voitures accessibles au plus grand nombre, ce qui en fait une marque "tout public", représentative d'une structure financière robuste.

Dans le cluster à la structure financière plutôt moyenne, on retrouve Ford, qui adopte une approche similaire à celle de Honda : proposer des voitures accessibles, mais en ciblant principalement le marché européen. Ford se distingue par son ancrage local et son positionnement concurrentiel dans la région, ce qui lui permet de maintenir une rentabilité stable, bien que sa structure financière ne soit pas aussi solide que celle de Honda.

Enfin, dans le cluster à la structure financière la plus faible, nous trouvons Ferrari. Cette marque, bien que très profitable, ne suit pas le modèle de production de masse. Ferrari se distingue par son approche de niche, visant un public plus restreint mais prêt à payer un prix premium pour des véhicules exclusifs. Cette stratégie, axée sur des marges élevées plutôt que sur un volume de production élevé, peut expliquer sa structure financière moins développée en comparaison avec des géants comme Honda ou Ford. Il est important de noter que cela ne signifie pas que Ferrari ait des profits faibles ; au contraire, ses marges par véhicule sont très élevées, mais la production limitée influe sur sa structure financière globale.

5 Discussion

Nous souhaiterions tout de même revenir sur la méthode de récupération des données. En effet, le fait d'ajouter des variables dans la requête SQL de Wikidata nous contraint à réduire le nombre d'individus récupérés. Pour régler ce problème, nous pourrions peut-être réduire le nombre de variables choisies, comme par exemple le chiffre d'affaire et le profit, qui sont très liés, afin d'inclure plus d'individus. La méthode utilisée pour récupérer la base de données sera détaillée en annexe.

Il existe également un problème concernant la temporalité des variables. En effet, les variables sont récupérées selon différentes temporalités, car elles ne sont pas forcément renseignées pour toutes les entreprises à la même année. Ainsi, un profit réalisé en 2019 ou 2020 a pu subir un choc économique dû au COVID, comparé à des entreprises avec un revenu renseigné à des périodes plus sereines.

Pour finir, il est possible que sur le site wikidata, certaines variables étaient renseignées plusieurs fois à des dates différentes. Pour avoir une observation par entreprises nous avons donc fait la moyennes pour toutes les variables.

6 Conclusion

Notre étude visait à examiner les relations entre certaines variables financières et structurelles des entreprises, en s'appuyant sur des hypothèses issues de la littérature économique. Plus spécifiquement, nous avons cherché à valider les hypothèses suivantes :

- **H1** : Les variables financières des entreprises sont corrélées.
- **H2** : L'âge d'une entreprise impacte positivement sa puissance financière.

Pour répondre à ces questions, nous avons mobilisé des méthodes quantitatives, notamment l'Analyse en Composantes Principales (ACP), l'Analyse des Correspondances Multiples (ACM) et la Classification Hiérarchique. Ces outils nous ont permis d'explorer les relations entre les variables financières et structurelles des entreprises, tout en identifiant des regroupements significatifs.

Les résultats obtenus montrent que :

- Les variables financières des entreprises présentent effectivement des corrélations significatives, confirmant **H1**. Ces corrélations soulignent l'interdépendance entre différents indicateurs économiques, comme le chiffre d'affaires, le profit et les capitaux propres.
- En revanche, l'hypothèse **H2**, qui postulait une influence positive de l'âge sur la puissance financière des entreprises, n'a pas été confirmée. Nos analyses révèlent que l'âge d'une entreprise n'est pas systématiquement corrélé à une meilleure performance financière ou à une stabilité accrue.

Ces résultats apportent des nuances aux théories économiques étudiées. Bien que certains liens économiques classiques, tels que ceux entre le chiffre d'affaires et le profit, ou entre les capitaux propres et la rentabilité, soient validés, d'autres facteurs, comme l'âge de l'entreprise, semblent jouer un rôle plus complexe ou contextuel dans la performance financière.

Ainsi, cette recherche souligne l'importance d'une approche empirique pour vérifier la pertinence des modèles théoriques. Dans une perspective future, il serait intéressant d'explorer les facteurs modérateurs, tels que le secteur d'activité ou les contextes économiques, qui pourraient influencer la relation entre l'âge et la performance des entreprises. De plus, l'intégration de données longitudinales permettrait d'affiner ces analyses et d'examiner les évolutions temporelles de ces relations.

7 Annexe

7.1 Récupération des données

Pour récupérer les données issues de notre base de données, nous nous sommes appuyés sur un code Wikidata disponible pour une partie des entreprises présentes sur le territoire français. Wikidata est une plateforme similaire à Wikipédia, mais dédiée aux données structurées, notamment celles concernant les entreprises. Elle recense des entreprises identifiables grâce à un code unique, ainsi que des informations les concernant, telles que le chiffre d'affaires, le profit, etc., également identifiables par des codes spécifiques.

Ces codes nous permettent, grâce au package WikidataQueryServiceR, d'effectuer des requêtes en SparkSQL pour récupérer les données nécessaires.

Nous avons ainsi développé un script permettant de tester, pour chaque code Wikidata, si les variables utilisées dans cette étude sont disponibles pour chaque société. Si une seule variable est manquante, la société n'est pas ajoutée au jeu de données.