**Analyse Python :**

1. ANALYSE DU RATING

A et BBB les notes qui regroupent le plus d’entreprises pour le reste les plus gros ensembles sont répartis entre BBB et BB-.

1. ANALYSE DES RATIOS

Pour l’analyse, on a besoin de ratio de liquidité, d’endettement et de rentabilité, on a choisi 4 ratios suite à l’analyse de leur corrélation avec le Rating score.

- NPM et ROE : écart-type est supérieur à la moyenne, distribution assez dispersée. 12 % de valeurs aberrantes pour ROE et le NPM n’a pas de valeurs aberrantes.

- Current ratio : écart-type assez élevé et asymétrie à droite.

- Debt/Equity ratio : Moyenne proche de l’écart type, mais écart type assez élevé ce qui montre un ratio dispersé et la distribution nous montre qu’il y a un grand nombre d’entreprise avec un endettement élevé.

1. ANALYSE DE LA CORRELATION

Résultat des corrélations entre Rating score et les ratios financiers :

* Ratio de Liquidité :
  + Current Ratio -0.16), corrélation négative. Une meilleure liquidité est associée à un score plus bas, ratio à approfondir car non intuitif.
* Ratio d’endettement :
  + Debt/Equity Ratio (-0.29) : corrélation négative importante, un ratio d’endettement important = mauvais signe pour la solidité financière.
  + Long-term Debt / Capital (0.04) ; Corrélation très faible voire nulle. Cela signifie il n’y a pas de relation linéaire forte entre ces ratios d’endettement et la notation de crédit.
* Ratio de rentabilité :
  + Net Profit Margin (0.41), ROE et ROA (0.3) EBITDA Margin (0.20) : Corrélation positive importante. On s'attend à ce qu’une meilleure rentabilité → meilleure note.
  + Il se peut que certaines entreprises très rentables soient paradoxalement mal notées (startups, secteurs volatiles, effets comptables)
* Les ratios restants ne sont pas pertinents car leur corrélation faible.

Quel que soit le résultat de l’analyse de la corrélation de la notation de crédit avec l’un des ratios, un seul ratio financier ne suffit pas pour avoir une image complète de situation d’une entreprise. Nécessité ici de ramener une approche sectorielle, pour segmenter notre analyse par secteur.

1. ANALYSE SECTORIELLE

* Répartition par notation de crédit :
  + Forte granularité des notations de crédit donc pas pertinent
* Répartition par Investment Grade
  + Les secteurs Utilities, Manufacturing, Health, et Business Equipment on tune grande part d’entreprise “Investment Grade”.
* Heatmap de la moyenne par secteur :
  + Current Ratio (liquidité à court terme) :
    - BusEq (Business Equipment), Other, Manuf et Chemicals ont les ratios les plus élevés → ces entreprises conservent plus d’actifs liquides.
    - Utils et Telcm ont des ratios supérieur mais proche de 1, probablement car ce sont des secteurs très régulés ou stables.
  + Debt/Equity Ratio (structure financière) :
    - Energy, BusEq et Health ont un faible endettement.
    - Money, Durables et Telcm ont un endettement élevé.
    - Manuf, Chems et Utils ont des ratios élevés → typique des secteurs à forte intensité capitalistique.
  + Net Profit Margin
    - Money, Health, BusEq et Telcm sont en haut du classement avec une marge nette forte. Cette analyse est en partie confirmer par la répartition des Investment Grade par secteur.
  + Return On Equity
    - Money, Energy et Utilities ont un niveau faible de ROE donc moins de création de valeur pour les actionnaires.
    - Health, Non-Durables et Chemicals ont un niveau de ROE élevé.

Etude des valeurs aberrantes :

* + Current Ratio globalement stable, sauf quelques secteurs :
    - Secteurs avec peu voir pas d’outliers (≤ 5%) : Manuf, Non-Durbl, Chems, Money, Energy, BusEq, Durables et Others
      * Ces secteurs semblent avoir une gestion de trésorerie assez homogène
    - Secteurs avec outliers modérés (~7–11%) : Utilities (7.85%), Shops (5.58%), Health (10.71%)
      * Diversité des modèles économiques dans ces secteurs (ex : shops = retail vs e-commerce vs luxe) = écarts plus marqués.
  + Debt/Equity Ratio (Structure financière) :
    - Secteurs avec le plus d’outliers (>10%) : Health (12.05%), Manufacturing (11.47%), Business Equipment (11.80%), Other (11.19%)
      * L’endettement varie beaucoup entre les entreprises de ces secteurs
    - Secteurs sans outliers : Telecommunications, Money, Durables
  + Net Profit Margin (Rentabilité nette). Ratio relativement stable dans la plupart des secteurs.
    - Secteurs avec le plus d’outliers : Chemicals (16.81%), Manufacturing (6.13%), Utilities (7.17%) et Non-Durables (8.19%)
      * Une rentabilité très inégale dans ces secteurs.
    - Secteurs sans outliers : Health, Energy, Business Equipment, Money, Other
      * Rentabilité très stable entre secteurs
  + ROE
    - Secteurs avec le plus d’outliers (>10%) : Shops (17.93%), Utilities (14.68%) et Chemicals (10.92%)
      * Ces secteurs ont des écarts extrêmes de performance financière.
    - Secteurs moyennement dispersés (10-4%) : Manufacturing, Health, BusEq et Other.
    - Secteurs sans outliers : Money, Durables, Non-Durables, Telecommunications

Pour éviter des erreurs d'interprétation causées par des valeurs aberrantes, on va analyser les médianes des ratios financiers par secteur :

En résumé :

* Concernant les ratios financiers, le current ratio n’a pas d’écart extrême entre la médiane et la moyenne.
* Le NPM n’a pas eu de gros écart à part pour le secteur Energy avec une moyenne à -17 contre une médiane à 3.8, les rentabilités sont très dispersées entre les entreprises de ce secteur.
* Le ratio d’endettement D/E Ratio montre des écarts sur certains secteurs. Comme Tlcm -7.7 (moyenne) vs 0.91 (médiane), fortement influencé par valeurs négatives extrêmes. Également Durables, Shops et Money qui passe d’une moyenne négative à un une médiane proche de 1.
* Enfin ce sont le ratios de rentabilité ROE a été le plus touché par les écarts entre la moyenne et la médiane avec plus de 5 secteurs touchés. Durables avec un passage de -48 à 13, Money -28 à 3.3 ou encore Energy 39 à 8.5.

En analysant la distribution du ROE par secteur nos interprétations se confirment pour les secteurs sensibles aux valeurs extrêmes on remarque une asymétrie de la distribution. Dans l’étape suivante sur Power BI on aura un visuel plus clair qui nous permettra de voir les différence entre secteur.

De même, en analysant les moyennes des ratios on se rend compte que la winsorisation a eu un impact plus significatif sur les ratios Debt/Equity et ROE ce qui indique que ces ratios contenaient des valeurs extrêmes influençant fortement la moyenne. Ces changements montrent que la winsorisation a efficacement réduit l'impact des valeurs aberrantes dans ces ratios.

1. SCORE RISQUE

Score risque pour évaluer le niveau de risque d’une entreprise :

* On définit le score risque à partir du poids des ratios basés sur leur corrélation avec la notation
* On normalise les ratios pour le score des secteurs pour pouvoir comparer les ratios entre secteur et entreprises
* Pour le score risque des secteurs on prend la médiane des ratios par secteur et à partir de cette médiane on calcul la corrélation avec la notation. On obtient alors le poids des ratios basés sur leur corrélation avec la notation.

Pondération du score risque avec des modèles de machine learning :

Utilisation de la validation croisée Kfold pour juger la robustesse et la performance des modèles. Nos prenons les ratios normalisés en variables explicatives et le Rating score en variable cible. Après études des modèles :

* L’arbre de décision est plus performant mais légèrement moins stable que la régression linéaire.

On rajoute d’autres ratios financiers dans nos variables explicatives pour voir si la performance du modèle change :

* L’ajout de nouvelles variables rend le modèle moins pertinent

Réduction de variables explicatives et ajout de ratios inverse pour l’interprétation des ratios d’endettement :

* Meilleure performance globale : baisse du MSE et RMSE, hausse du R²
* Modèle plus stable baisse des écarts-types

Etudes avec les modèles de boosting (plus précis que les algorithme simple) :

* GradientBoosting() : Meilleure performance globale du modèle avec baisse du MSE et RMSE, hausse du R²et légèrement plus performant que XGBoost

Création du score :

* Nous choisissons la pondération des variables (ratios) ici du modèle GradientBoost pour le calcul du nouveau score risque.
* Multiplication par 10 pour avoir un score compris dans l’intervalle [0 ;10]
* Classification (Risk level) des entreprises selon leur score risque, de « Very Low » à « Very High ».
* Création du score risque secteur