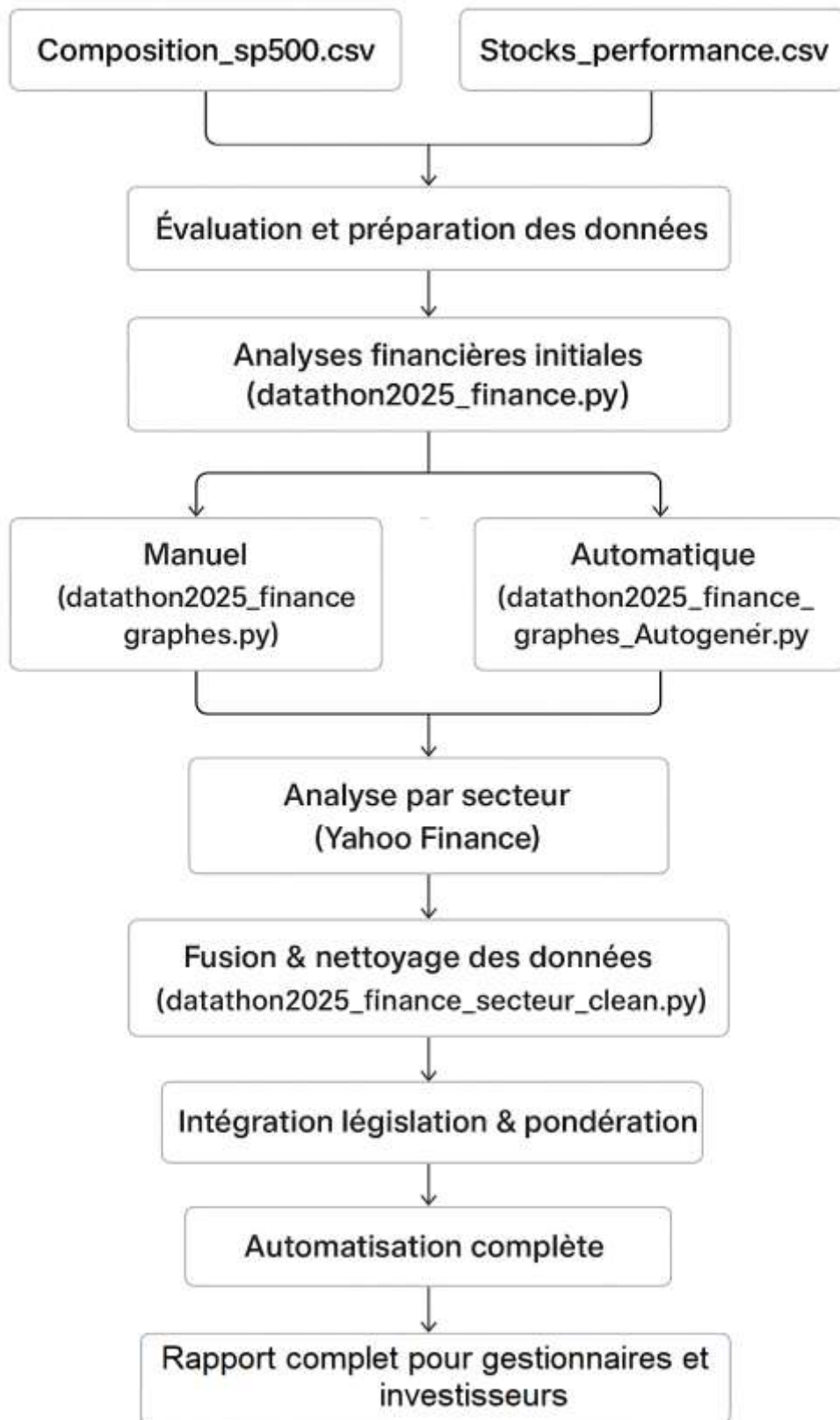


- ◆ Analyse du **poids des entreprises** dans le S&P 500.
- ◆ Évaluation du **Net Income** pour mesurer la performance.
- ◆ Calcul du **Risk Score** pour identifier les pertes potentielles.
- ◆ Pondération du **risque par secteur** et exposition législative.
- ◆ Priorisation des **entreprises et secteurs à fort impact**.

Le résumé de la démarche suivante pour notre analyse financière :

1. **Évaluation et préparation des fichiers sources**
 - Analyse du contenu des fichiers Composition_sp500.csv et Stocks_performance.csv fournis par les sources de compétition.
 - Vérification de la cohérence des données et enregistrement des fichiers préparés.
2. **Analyses financières initiales**
 - Création du fichier datathon2025_finance.py pour effectuer des analyses sur les revenus (Income) : top 10 des meilleures performances et top 10 des pertes.
3. **Visualisation des données**
 - Création du fichier datathon2025_finance_graphes.py pour générer des graphiques un par un (enregistrement manuel).
 - Optimisation avec datathon2025_finance_graphes_Autogénére.py pour enregistrer automatiquement tous les graphiques.
4. **Analyse par secteur**
 - Utilisation de Yahoo Finance pour identifier les différents secteurs de chaque enregistrement.
 - Fusion (inner join) des fichiers Composition_sp500.csv et Stocks_performance.csv pour conserver uniquement les enregistrements communs.
 - Nettoyage des colonnes pour réduire les redondances.
 - Création du fichier datathon2025_finance_secteur_clean.py.
5. **Intégration des données législatives**
 - Analyse le H.R.1 - One Big Beautiful Bill Act via le site du gouvernement américain pour identifier les secteurs impactés directement ou indirectement par la législation.
 - Pondération des secteurs et normalisation des données.
6. **Automatisation de l'analyse complète**
 - Développement d'un script Python automatisé pour :
 - Appliquer toutes les étapes précédentes en intégrant la législation et la pondération normalisée.
 - Générer automatiquement tous les graphiques nécessaires.
 - Calculer le risque de chaque secteur et identifier les trois entreprises les plus risquées par secteur.
7. **Résultat final**

Grâce à cette démarche d'amélioration continue et à l'application de méthodes analytiques industrielles basées sur l'analyse financière, nous pouvons générer un **rapport financier complet pour les gestionnaires et investisseurs en moins de quelques minutes.**



Rapport d'Analyse Financière — Datathon 2025

Projet : Automatisation de l'analyse financière sectorielle avec intégration législative

1. Introduction

Ce projet a été réalisé dans le cadre du Datathon 2025, une compétition visant à développer des outils d'analyse financière automatisés et intelligents à partir de données réelles du marché boursier.

L'objectif principal de notre travail était de :

- Construire un modèle de traitement et d'analyse des données boursières à partir de fichiers CSV fournis.
- Identifier les secteurs les plus performants et les plus à risque.
- Évaluer l'impact potentiel d'une législation récente sur la structure du marché.
- Générer automatiquement des graphiques et des indicateurs pour un usage décisionnel rapide.

Ce projet a été conçu pour être entièrement automatisé, reproductible et adaptable à d'autres jeux de données financiers.

2. Données sources et préparation

Deux fichiers principaux ont été utilisés :

- `Composition_sp500.csv` : contenant la liste des entreprises de l'indice S&P 500, leur symbole boursier, leur poids (pondération dans l'indice) et leur prix.
- `Stocks_performance.csv` : présentant des informations financières détaillées telles que le chiffre d'affaires, le revenu net, le bénéfice par action (EPS) et le flux de trésorerie libre (FCF).

2.1. Évaluation initiale

Une première étape a consisté à vérifier la cohérence des données (colonnes, types, valeurs manquantes) et à corriger les anomalies de formatage (symboles, séparateurs, encodage).

2.2. Fusion des fichiers

Les deux fichiers ont été fusionnés à l'aide de méthode *inner join*, afin de ne conserver que les enregistrements communs aux deux sources.

Cette fusion a permis de créer un fichier consolidé `SP500_with_Sector.csv`, enrichi par la suite d'informations sectorielles.

3. Analyse financière initiale

Un premier script, `datathon2025_finance.py`, a permis d'effectuer des analyses fondamentales :

- Identification du Top 10 des entreprises les plus rentables selon le Net Income.
- Identification du Top 10 des entreprises les plus déficitaires.

Ces premiers résultats ont permis de visualiser la distribution des performances financières individuelles.

4. Visualisation automatisée

Pour faciliter l'interprétation, un second script, `datathon2025_finance_graphes.py`, a été développé pour générer des graphiques comparatifs.

Une version améliorée, `datathon2025_finance_graphes_Autogener.py`, a ensuite été conçue pour enregistrer automatiquement toutes les figures générées, garantissant la reproductibilité et l'efficacité du processus.

Les graphiques produits incluent :

- Top 10 par poids dans le S&P 500
- Top 10 par revenu net positif (profits)
- Top 10 par revenu net négatif (pertes)
- Graphique de score de risque basé sur la pondération et le revenu net

5. Analyse sectorielle

Afin d'approfondir l'analyse, nous avons introduit une dimension sectorielle.

L'API Yahoo Finance (`yfinance`) a été utilisée pour récupérer le secteur d'activité de chaque entreprise à partir de son symbole. Les données nettoyées sont sauvegardées dans le fichier `SP500_with_Sector_Clean.csv`.

Le script `datathon2025_finance_secteur_clean.py` a permis :

- De grouper les entreprises par secteur.
- De calculer les poids, revenus et scores de risque moyens par secteur.
- D'identifier les 3 entreprises les plus risquées de chaque secteur.

Cette étape a rendu possible une analyse macroéconomique à partir des données microéconomiques.

6. Intégration de la dimension législative

Dans le cadre de l'analyse, le texte législatif H.R. 1 (Lower Energy Costs Act) a été étudié à partir du site officiel [congress.gov](https://www.congress.gov).

Cette loi vise principalement à :

- Favoriser la production d'énergie nationale.
- Réduire les contraintes réglementaires sur les secteurs énergétiques et industriels.
- Impacter indirectement les coûts de production dans d'autres secteurs.

6.1. Traduction analytique

Nous avons attribué un score d'exposition législative à chaque secteur :

- 1.0 : Exposition élevée (Energy, Utilities, Materials, Industrials)
- 0.5 : Exposition moyenne (Technology, Consumer Discretionary, Consumer Staples)
- 0.0 : Exposition faible (Financials, Health Care, Real Estate, Communication Services)

Ces pondérations ont été intégrées dans le calcul du `Risk_Score` global de chaque entreprise et secteur.

7. Automatisation complète

Le script final `datathon2025_finance_secteur_rev2.py` combine toutes les étapes :

1. Lecture et préparation des données consolidées.
2. Normalisation des pondérations.
3. Calcul du score de risque et du risque législatif.
4. Agrégation et visualisation sectorielle.
5. Génération automatique de tous les graphiques :
 - Top 10 par poids, par risque et par revenu net.
 - Top 3 entreprises les plus risquées par secteur.

L'ensemble du processus est exécutable en moins de 5 minutes et produit un rapport graphique complet.

8. Résultats et conclusions

Grâce à cette approche structurée et itérative :

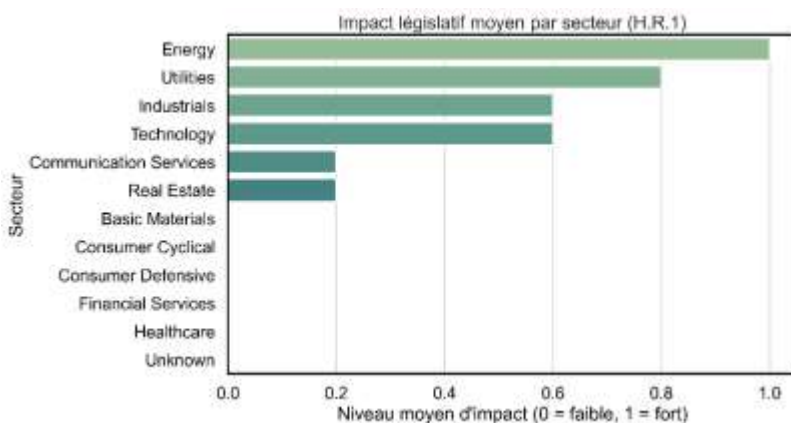
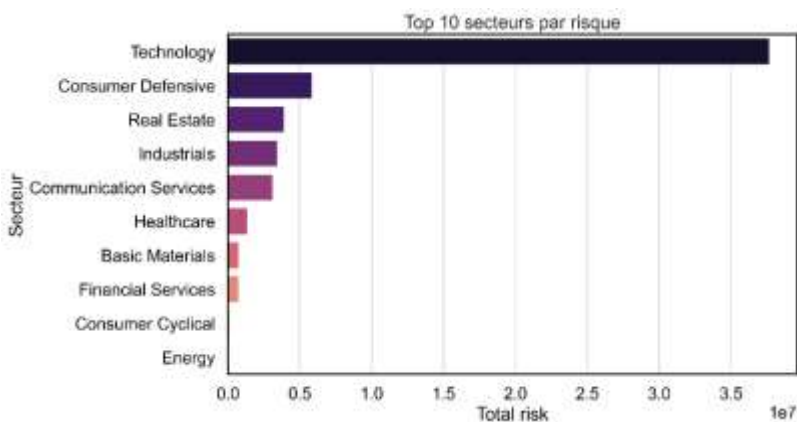
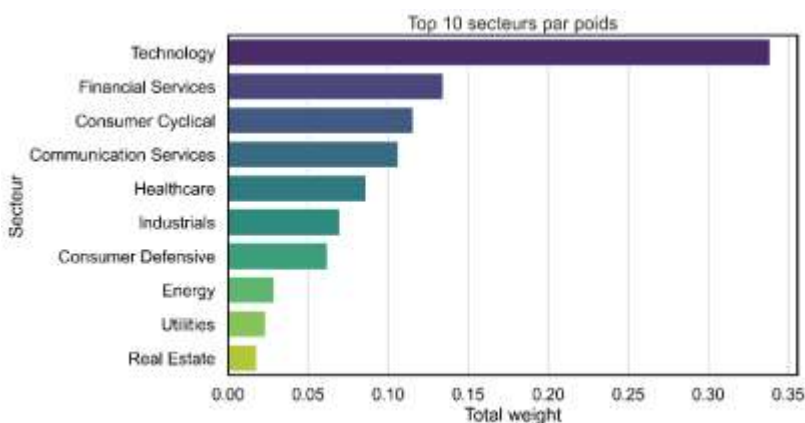
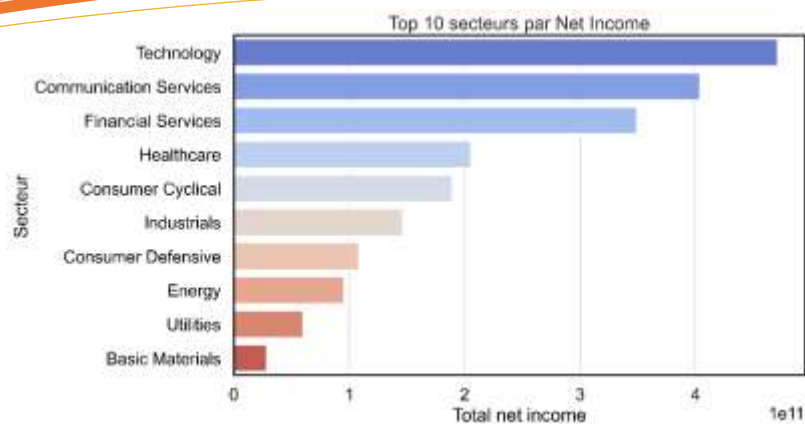
- Les données financières du S&P 500 ont été nettoyées, enrichies et analysées de façon automatique.
- L'impact potentiel d'une législation majeure a été intégré dans la modélisation du risque.
- Les graphiques générés permettent de visualiser la répartition du risque et de la performance par secteur, offrant un outil d'aide à la décision rapide pour les gestionnaires et investisseurs.

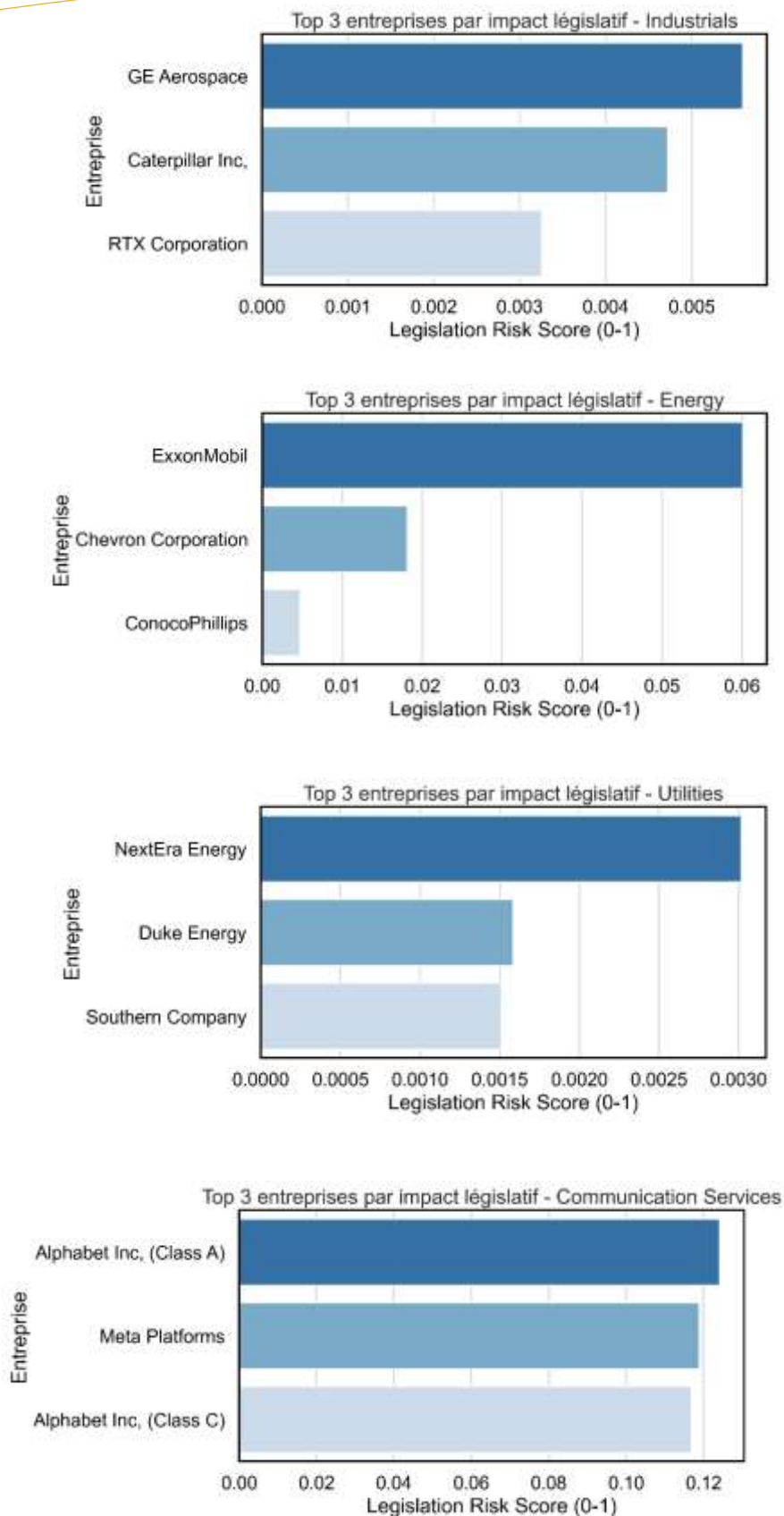
Ce projet démontre la valeur de la data science appliquée à la finance, combinant automatisation, rigueur analytique et intégration contextuelle de facteurs externes tels que la législation.

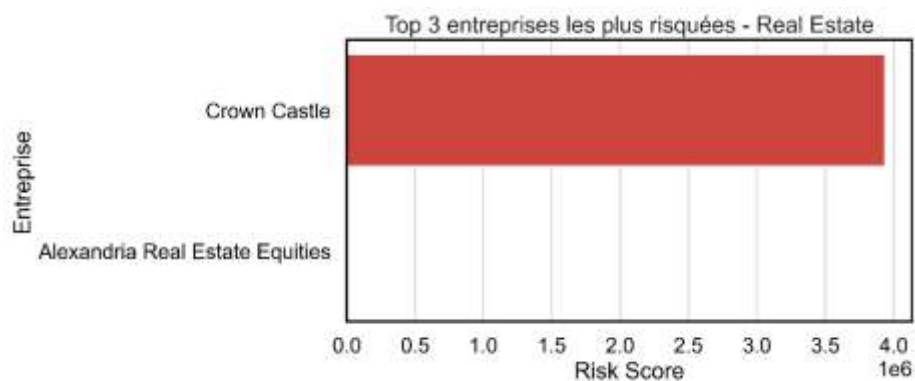
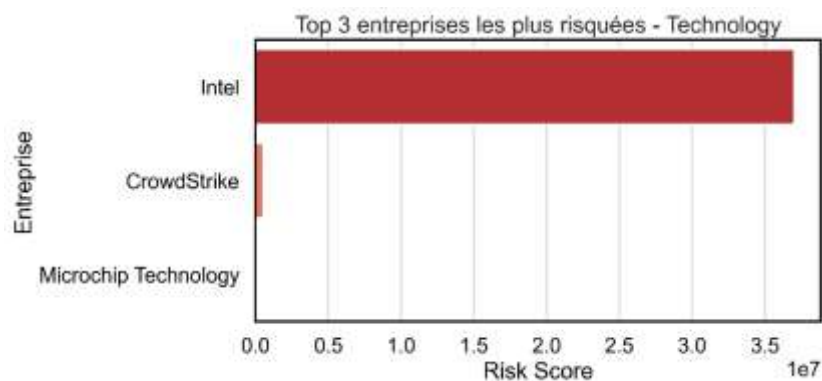
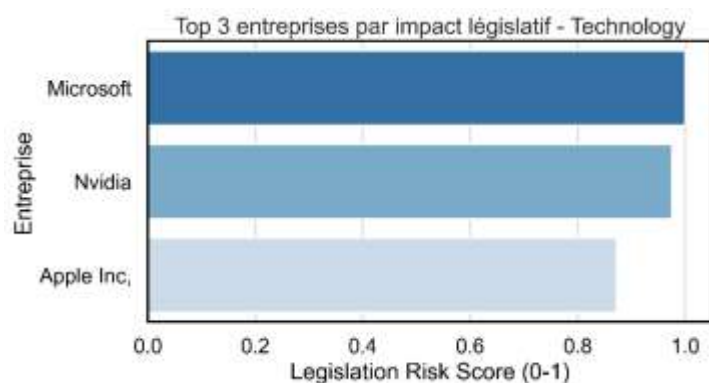
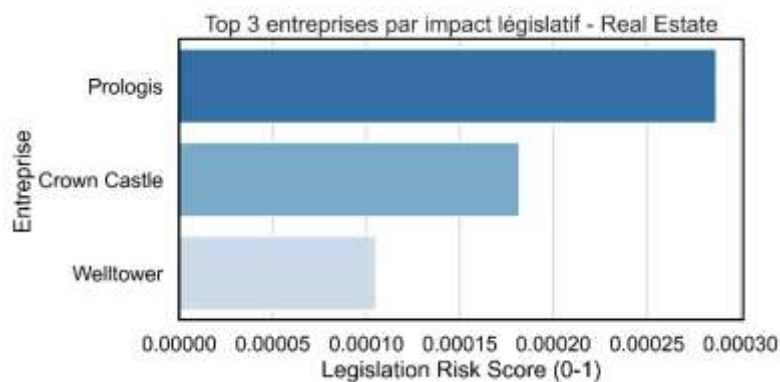
9. Perspectives

Les extensions possibles incluent :

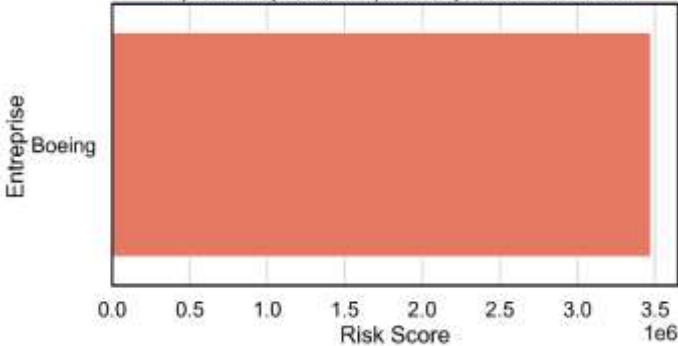
- L'intégration d'autres sources législatives ou économiques (Fed, SEC, ESG).
- L'ajout d'indicateurs de performance avancés (Sharpe, Beta sectoriel, volatilité).
- L'automatisation complète du rapport en format PDF ou web interactif.



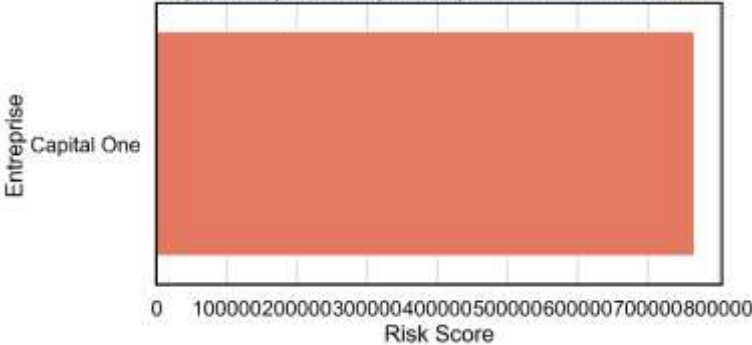




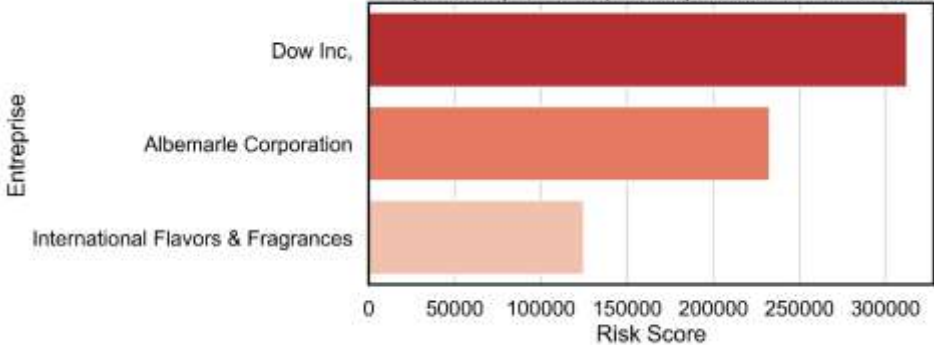
Top 3 entreprises les plus risquées - Industrials



Top 3 entreprises les plus risquées - Financial Services



Top 3 entreprises les plus risquées - Basic Materials



Top 3 entreprises les plus risquées - Healthcare

