**Projet data analyste :**

**Prévention et évaluation du risque de crédit des entreprises**

**Plan détaillé du projet :**

Ce projet vise à analyser et prévenir le risque de crédit des entreprises en exploitant le dataset "Corporate Credit Rating with Financial Ratios". Avec de dataset on pourra :

* Analyser les facteurs influençant la notation de crédit des entreprises,
* Déterminer quels ratios financiers sont les plus pertinents pour évaluer le risque,
* Construire un score de risque et un tableau de bord interactif.

Nous utiliserons SQL, Python, Excel et Power BI pour structurer, analyser et visualiser les données.

1. Compréhension et exploration des données

Objectif : Se familiariser avec les variables du dataset et identifier les axes d’analyse.

* Télécharger et explorer le dataset.
* Vérifier la structure des données (types, valeurs manquantes, doublons).
* Explorer les distributions des variables clés (boxplots, histogrammes).
* Identifier les relations entre variables (corrélations, tendances).
* Analyser les variables clés :
  + Ratios financiers (ex : liquidité, rentabilité, endettement).
  + Notation de crédit (ex : AAA, BBB, C, etc.).
    - Visualiser les distributions des variables et identifier des anomalies.
  + Vérifier si certaines variables sont **très corrélées entre elles** (exemple : endettement et notation de crédit).

📊 Outils : Python (pandas, seaborn, matplotlib), Excel.

1. Nettoyage et transformation des données en SQL

Objectif : Préparer des données propres et exploitables.

* Charger le dataset brut dans une base SQL.
* Nettoyer les valeurs manquantes (remplacement, suppression).
* Gérer les valeurs aberrantes (winsorization, suppression).
* Normaliser certaines variables pour une meilleure comparaison.
* Agréger certaines informations utiles (ex : évolution des ratios).

📊 Outils : SQL (PostgreSQL/MySQL/SQLite).

1. Extraction des données transformées vers Python

Objectif : Charger les données nettoyées et effectuer une analyse avancée.

* Écrire des requêtes SQL pour récupérer les données nettoyées.
* Charger ces données en Python pour les manipuler et analyser les corrélations.
* Étudier les relations entre ratios financiers et notation de crédit.
* Visualiser les résultats avec des graphiques (matrice de corrélation, boxplots).

📊 Outils : SQL (connexion via pandas), Python (pandas, matplotlib, seaborn).

1. Création d’un score de risque

Objectif : Construire un indicateur permettant d’évaluer le niveau de risque d’une entreprise.

* Sélectionner les ratios financiers les plus pertinents.
* Pondérer ces ratios en fonction de leur impact sur la notation de crédit.
* Définir des seuils pour classifier les entreprises (faible, moyen, élevé risque).

📊 Outils : Python (numpy, pandas, sklearn), Excel.

1. Modélisation prédictive (optionnel)

Objectif : Automatiser l’évaluation du risque avec un modèle de classification.

* Sélectionner un modèle adapté :
  + Régression logistique (interprétable).
  + Random Forest / XGBoost (plus performant).
    - Entraîner le modèle pour prédire la notation de crédit.
    - Évaluer les performances (accuracy, confusion matrix, AUC-ROC).
    - Comparer le modèle avec des méthodes classiques (ex : Altman Z-score).

📊 Outils : Python (sklearn, XGBoost).

1. Connexion SQL → Power BI pour la visualisation

Objectif : Construire un tableau de bord dynamique permettant d’évaluer le risque de crédit.

* Se connecter à la base SQL depuis Power BI.
* Créer des indicateurs interactifs :
  + Score de risque par entreprise.
  + Répartition des notations de crédit.
  + Évolution des ratios financiers dans le temps.
    - Mettre en place des filtres dynamiques (secteur, niveau de risque).

📊 Outils : SQL, Power BI.

1. Recommandations et conclusions

Objectif : Proposer des mesures pour améliorer la gestion du risque de crédit.

* Identifier les secteurs ou entreprises les plus exposés au risque.
* Définir des stratégies de prévention :
  + Ajustement des critères d’octroi de crédit.
  + Suivi des entreprises à risque élevé.
    - Renforcement des garanties pour certains profils.
      * Rédiger un rapport avec les principaux insights et recommandations.

📊 Outils : PowerPoint / Word pour la documentation.

🡺 **Résultat final** : Un projet professionnalisant et structuré

* Dataset nettoyé et analysé dans SQL et Python.
* Score de risque défini et testé sur les données.
* Modélisation optionnelle pour automatiser l’évaluation.
* Tableau de bord interactif Power BI pour suivre les entreprises à risque.
* Recommandations basées sur les données pour une meilleure gestion du risque.

STRUCTURE GITHUB :

corporate-credit-risk-analysis/

│

├── README.md ← Présentation du projet

├── requirements.txt ← (Optionnel) Librairies nécessaires

│

├── data/ ← Données brutes et nettoyées

│ ├── raw/ ← Données brutes (ex: CSV initial)

│ └── processed/ ← Données nettoyées/exportées (Excel, SQL)

│

├── sql/ ← Requêtes SQL (création table, nettoyage, etc.)

│ ├── create\_schema.sql

│ ├── clean\_data.sql

│ └── extract\_data.sql

│

├── python/ ← Scripts Python d’analyse et de visualisation

│ ├── eda.py ← Analyse exploratoire (EDA)

│ ├── scoring.py ← Score de risque

│ ├── modeling.py ← Modèle prédictif (si fait)

│ └── chargement\_sql.py ← Connexion à SQL Server et extraction

│

├── powerbi/ ← Exports ou captures de ton tableau de bord Power BI

│ └── tableau\_de\_bord.pbix

│

└── report/ ← Recommandations et livrables

├── presentation.pptx

└── rapport\_final.pdf