

## Règle générale :

Pour l'ensemble des sujets proposés, il faudra présenter :

- Une **exploration des données** bien construite, avec à l'appui une description des variables, leurs relations, les points remarquables.
- Il faudra que cette exploration soit appuyée de **visualisation soignées** et pertinentes. (Incluez au moins une visualisation avec **plotly**, une bibliothèque en vogue pour la visualisation des données.)
- N'hésitez pas à procéder à une phase de "feature engineering" (si possible) --> création / sélection de variables d'intérêt.
- Pourquoi pas faire évoluer ou approfondir la problématique.
- Des cycles de modélisation : explorez plusieurs modèles possibles.
- Au moins un cycle d'**optimisation**.
- Une modélisation finale avec scoring complet du/des modèles retenus.
- Optionnel : pourquoi pas s'intéresser à l'interprétabilité du modèle ?
- Une analyse des résultats obtenus.

Vous mettrez en place un repository github pour me proposer un rendu avec :

- Un rapport de projet qui doit contenir :
  - Une présentation des données, leur contexte et la position du problème.
  - Un descriptif de la méthodologie choisie et mise en place, des différentes étapes (et aussi des modèles choisis).
  - Une description des résultats obtenus (y compris de la phase d'optimisation).
  - Une critique des résultats.
  - Les problèmes rencontrés et comment vous y avez fait face.
  - Des perspectives (que feriez-vous en plus).
  - Une conclusion.
- Les codes (langage python) du projet (et aussi de la visualisation éventuellement).

La présentation d'une durée de 25 à 30 minutes devra reprendre les étapes principales de votre rapport. Vous êtes libres concernant le support de présentation. Les supports interactifs (de type streamlit ou dash) peuvent être envisagés. Tout le monde doit prendre la parole.

Je sais :

- Que les projets existent sur internet... ne les reproduisez pas à l'identique, ce serait une (très) mauvaise idée !
- Qu'il y a différents niveaux de difficultés → je le prendrai en compte lors de l'évaluation du projet.

L'ensemble codes + rapport est attendu au plus tard le **01 Avril** La présentation aura lieu le **04 Avril**.