

Projet

Dans le cadre du projet, vous devrez choisir des données sur lesquelles vous appliquerez les principes du data cleansing, pour ensuite faire une analyse à l'aide des statistiques descriptives. Le rendu pourra se faire via un jupyter notebook pour faciliter la visualisation des données.

L'évaluation sera réalisée avec :

- 1) Une présentation orale en binôme des méthodes utilisées, des analyses et des résultats lors de la sixième séance
- 2) Un rendu comprenant un rapport et les codes pythons commentés, avec les sources.
- 3) Un examen/contrôle sur une des séances de TP lors de la séance 4
- 4) Une évaluation par séance du travail du binôme avec des réponses à des questions. La note du binôme sera la note donnée à celui qui doit répondre aux questions.
- 5) Une évaluation par les pairs du travail en début de la séance 5

Avec ces évaluations, vous devez montrer :

- 1) Que vous savez utiliser Python, Pandas et des bibliothèques graphiques pour l'analyse de donnée
- 2) Que vous savez manipuler des fichiers de grande taille
- 3) Que vous savez mettre en œuvre les principes du Data Cleansing
- 4) Que vous savez analyser les données et les représentez judicieusement.

Pour le rapport, je vous demande de rédiger un journal de bord :

Vous noterez systématiquement pour chacune des séances, en précisant les dates :

- ce que vous avez fait et pourquoi ;
- quels sont les problèmes rencontrés pour le faire ;
- les solutions que vous avez testé en expliquant pourquoi certaines ne fonctionnaient pas ;

Si vous avez utilisé une IA pour faire une action :

- indiquez le prompt que vous lui avez donné ;
- précisez pourquoi vous l'avez utilisé
- précisez comment vous avez utilisé l'information : vous avez interdiction de faire un copier-coller. Vous devez vous approprier les commandes, les commenter et les expliquer dans votre journal de bord

Vous devez indiquer au minimum trois entrées pour chacune des séances.

Voici des exemples de sites où vous pouvez aller chercher des données :

- Données sur Kaggle : <https://www.kaggle.com/datasets>
 - <https://www.kaggle.com/datasets/yanghu583/electric-vehicle-population-data-2025>
 - <https://www.kaggle.com/datasets/yashdevladdha/uber-ride-analytics-dashboard/code>
 - <https://www.kaggle.com/datasets/shreyanshdangi/co-emissions-across-countries-regions-and-sectors>
 - <https://www.kaggle.com/datasets/sulthonagthoris/mobile-phone-specifications-dataset/data>
 - <https://www.kaggle.com/code/devraai/urban-flood-risk-analysis-and-prediction>
 - <https://www.kaggle.com/code/devraai/job-posting-data-analysis-and-experience-predic>
 - <https://www.kaggle.com/datasets/rohitgrewal/global-superstore-data>
 - <https://www.kaggle.com/datasets/elahehgoirokh/data-science-job-postings-with-salaries-2025>
- Enedis : <https://data.enedis.fr/explore/?sort=modified>
- RATP : <https://data.ratp.fr/explore/?sort=modified>
 - <https://data.ratp.fr/explore/dataset/trafic-annuel-entrant-par-station-du-reseau-ferre-2021/table/?sort=-rang>
 - <https://data.ratp.fr/explore/dataset/commerces-de-proximite-agrees-ratp/table/>
- SNCF : <https://ressources.data.sncf.com/explore/?sort=modified>
 - <https://ressources.data.sncf.com/explore/dataset/ponctualite-mensuelle-transilien/information/>
 - <https://ressources.data.sncf.com/explore/dataset/regularite-mensuelle-intercites/information/>
 - <https://ressources.data.sncf.com/explore/dataset/regularite-mensuelle-tgv-agst/information/>
- Ville de Paris :
 - <https://opendata.paris.fr/explore/dataset/respirons-mieux-dans-le-20eme-donnees-mini-stations/information/>
 - https://opendata.paris.fr/explore/dataset/belib-points-de-recharge-pour-vehicules-electriques-disponibilite-temps-reel/table/?disjunctive.statut_pdc&disjunctive.arrondissement&sort=code_insee_commune
 - <https://opendata.paris.fr/explore/dataset/velib-emplacement-des-stations/information/>

- https://opendata.paris.fr/explore/dataset/comptage-velo-donnees-compteurs/information/?disjunctive.id_compteur&disjunctive.nom_compteur&disjunctive.id&disjunctive.name
- <https://opendata.paris.fr/explore/dataset/comptage-multimodal-comptages/information/?disjunctive.mode&disjunctive.trajectoire&disjunctive.voie&disjunctive.sens&disjunctive.label>
- Data.gouv : <https://www.data.gouv.fr/datasets/search>
 - <https://www.data.gouv.fr/datasets/search?format=csv>
- UC Irvine : <https://archive.ics.uci.edu/datasets>
- Word bank : <https://data.worldbank.org/home>