



# **DATAFOREVER**

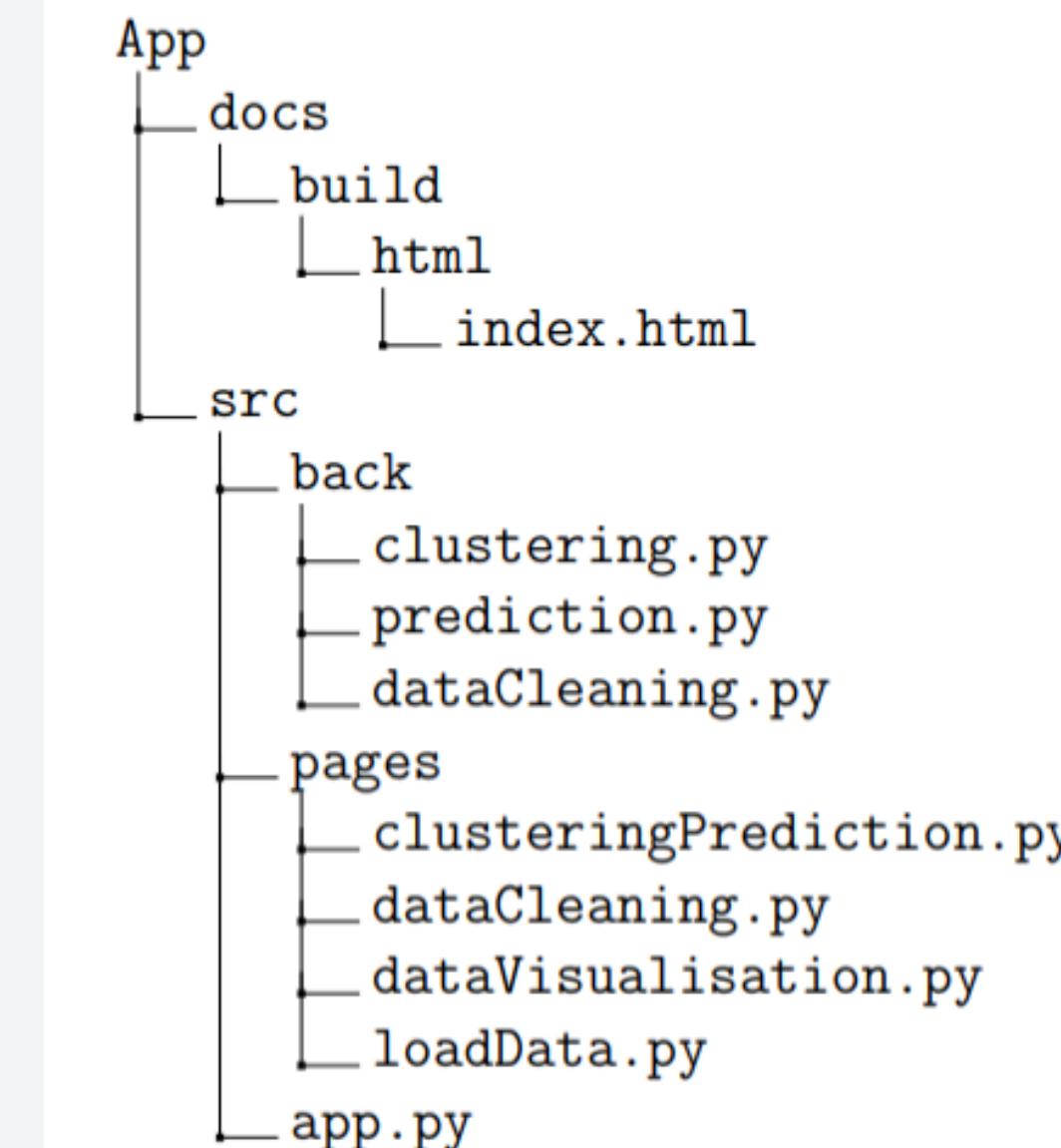
**PAR ALEXIS CARREAU, AUDRIC SABATIER ET HUGO ROCHELLE**

# SOMMAIRE

- 
- 01** ARCHITECTURE
  - 02** CONCEPTION DU BACK-END
  - 03** DESIGN DU FRONT-END
  - 05** VERSIONING GIT
  - 06** BILAN COLLECTIF
  - 07** BILAN INDIVIDUEL
  - 08** POINTS D'AMÉLIORATIONS

# ARCHITECTURE

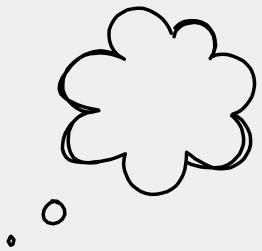
- Séparation :
  - Docs :
    - Documentation module par module des fonction dans un fichier HTML
  - Back :
    - Logique/traitement des données
  - Front:
    - Affichage des résultats du back via appels



# CONCEPTION DU BACK-END



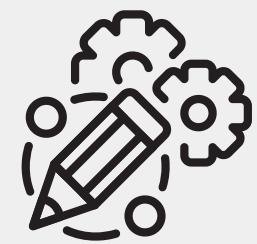
- Séparation des fonctionnalités par fichier
- Logique métier
- Réalisation des calculs



- Ensemble de fonctions
- Développement générique
- Pensé pour être réutilisé



# DESIGN DU FRONT-END



UI



- Ergonomie -> peu de boutons
- Utilisation widget de base streamlit (moderne et épuré)



UX

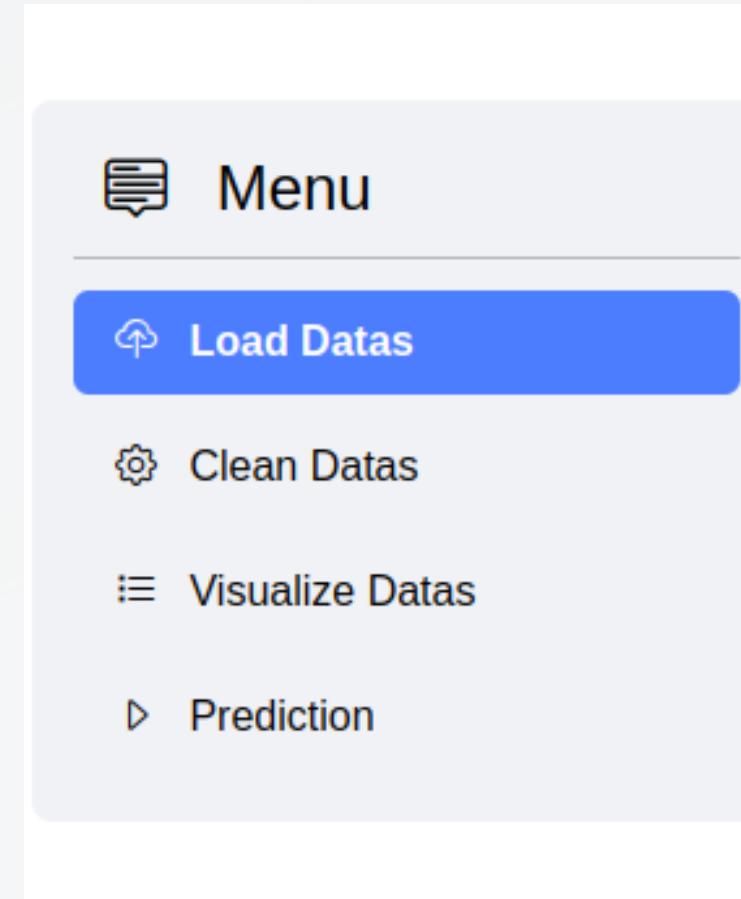


- Gestion des erreurs
- Pop up de confirmation

Data has been cleaned successfully!

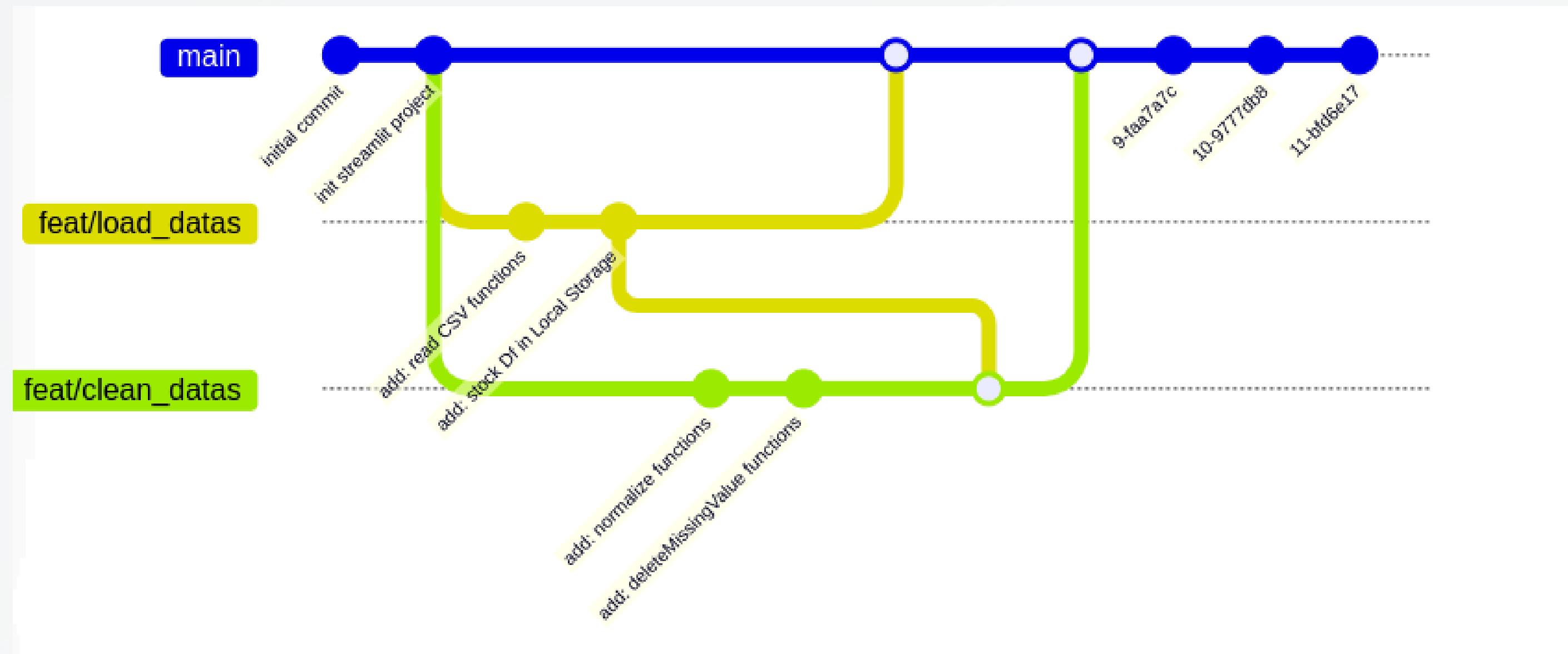
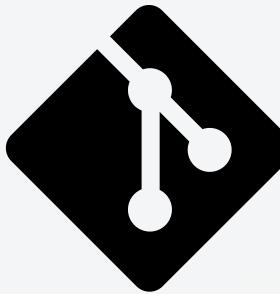
No data loaded yet. Please upload a CSV file first.

# DESIGN DU FRONT-END



- Toutes les pages sont accessible
- 1 page = 1 fonctionnalité
- Mise en cache du dataset entre les pages
- Vérification si les données ont été chargées/nettoyés

# VERSIONING GIT



# BILAN COLLECTIF



## RÉCUPÉRATION

- Mise en cache des données au sein du projet

## NETTOYAGE

- Proposer différentes solutions (Valeur manquante, ...)
- Grande liberté de choix pour l'utilisateur

## CLUSTERING/ PREDICTION

- Mise en place de plusieurs algorithmes de clustering et prédictions

## VISUALISATION

- Visualisation des données après nettoyage
- Visualisation des données après clustering

# BILAN INDIVIDUEL - ALEXIS

- Découverte de l'outil Streamlit
- Approfondissement de l'analyse de données
- Réalisation des visualisation des données et résultats

# BILAN INDIVIDUEL - AUDRIC

- Application du cours -> pratique vs théorie
- Travail sur loading datas
  - Réflexion passage dataframe entre page
- Défis -> affichage (cluster, centroïds...)
- Pas de carrière possible en tant que DataAnalyst pour le moment
  - A améliorer en Mastère (cours de BigData, Machine Learning)

# BILAN INDIVIDUEL - HUGO

- Découverte de Streamlit pour faire application Python
- Mise en place de fonctions de nettoyage & normalisation des données
- Mise en place du clustering avec DBSCAN
- Utile pour ma poursuite d'étude en Mastère (Big Data, Machine Learning, Intelligence Artificielle), progrès à faire

# POINTS D'AMÉLIORATIONS



O1



O2



O3



## SILHOUETTE SCORE FAIBLE

Choisir les composantes principales à utiliser

## MEILLEURE GESTION

Choix des paramètres proposés à l'utilisateur

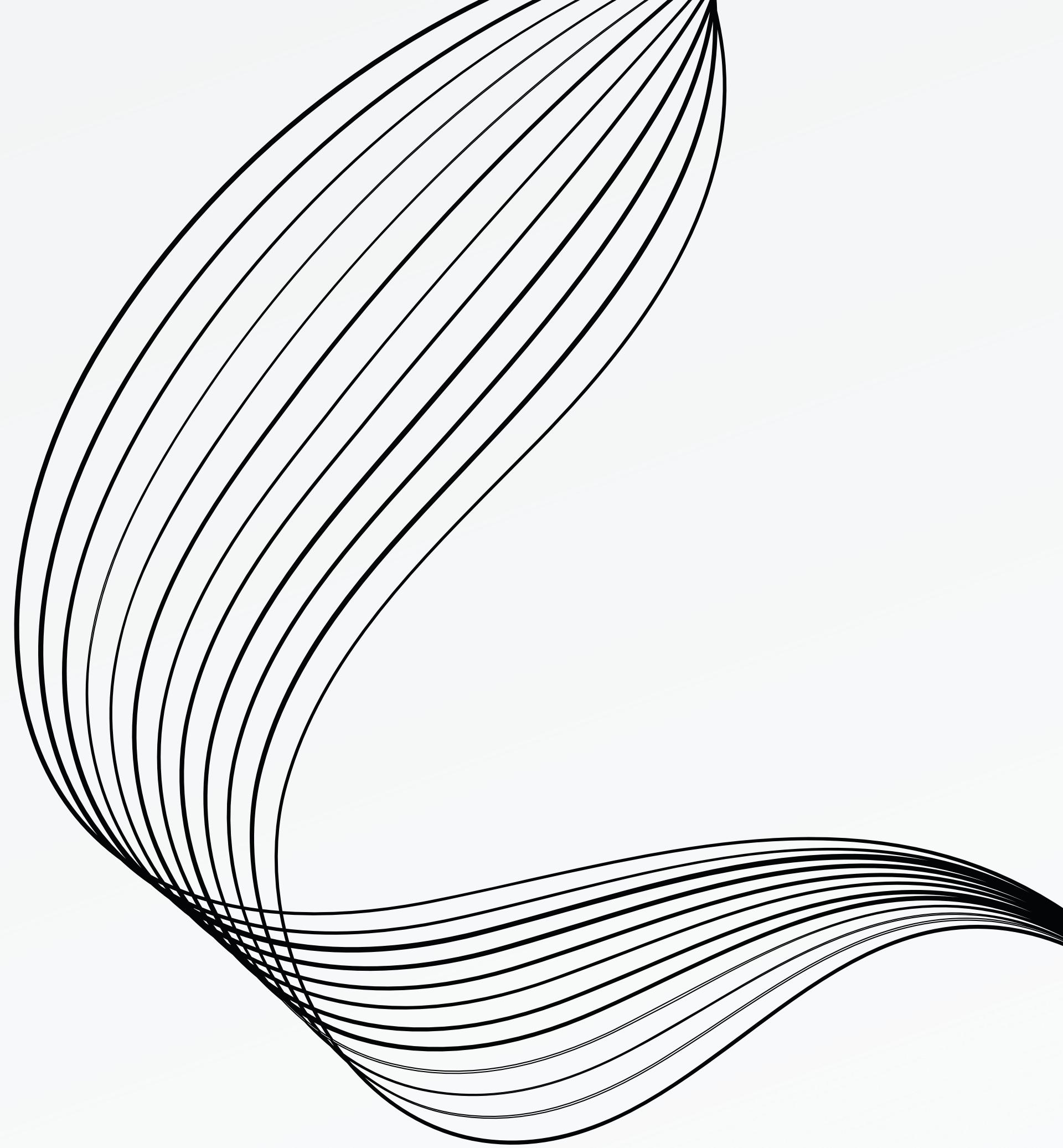
## DIVERSIFICATION

Proposer plus d'algorithmes et de choix à l'utilisateur

# DEMONSTRATION

The screenshot shows a user interface for data analysis. On the left, a sidebar menu includes 'Menu', 'Load Data' (which is highlighted in blue), 'Clean Data', 'Visualize Data', and 'Prediction'. The main area is titled 'Upload your CSV data' and features a central input field with the placeholder 'Choose a CSV file'. Below this is a section with a cloud icon and the text 'Drag and drop file here' and 'Limit 200MB per file • CSV'. To the right of this section is a 'Browse files' button with a small hand cursor icon indicating it is interactive. A blue banner at the bottom of the main area contains the text 'Veuillez télécharger un fichier CSV.' (Please download a CSV file.). In the top right corner of the main area, there are 'Deploy' and settings icons.

**MERCI DE  
NOUS AVOIR  
ÉCOUTÉ**



# BIBLIOGRAPHIE

## Documentations

- Documentation officiel de Streamlit (<https://streamlit.io/>)

## Blog personnel

- [Dr. Ernesto Lee](#)

## Intelligence artificielle

- ChatGPT (<https://chatgpt.com/>)
- Phind (<https://www.phind.com/>)
- Mistral (<https://chat.mistral.ai/>)
- Copilot

*Utilisation d'outils génératifs au sein du projet à hauteur de 80% (correction de bug + documentation sur les méthodes à utiliser).*