

CERTIFICATION BIG DATA

POEC OPCO ATLAS

Architecte Big Data et Science de données

BC06 du titre RNCP niv.7 n°32123

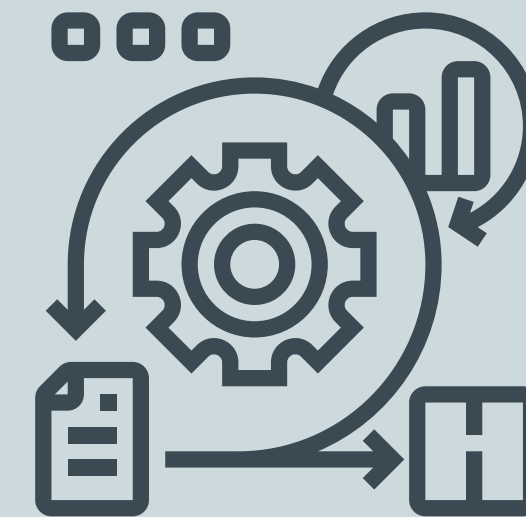


Systeme de
versioning et
mise en relation
de code :
GitHub

Points
biquotidiens
d'avancement

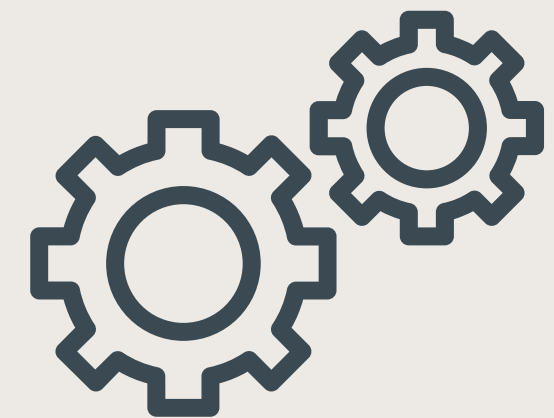
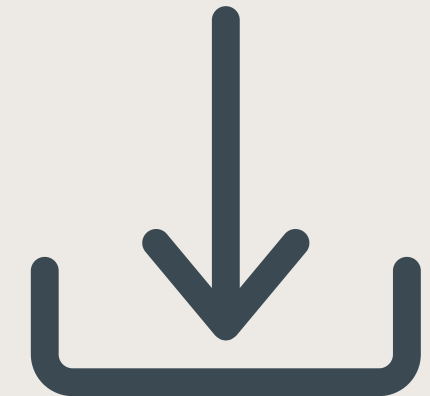
Modélisation
visuelle sur
support
physique

Répartition de
certaines tâches
en fonction des
compétences et
lacunes



OUTILS

- Python version 3.11.4
- Utilisation de plusieurs librairies
 - Pandas
 - Scikit-learn
 - Matplotlib
- Différentes bases de données
 - Structurées
 - Non structurées
- Outils de visualisation variés
 - Librairie python
 - Interface Web
 - PowerBI

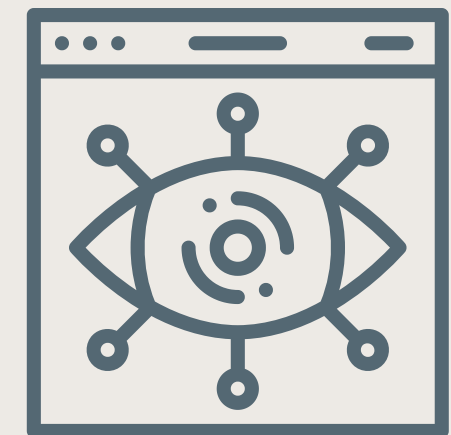
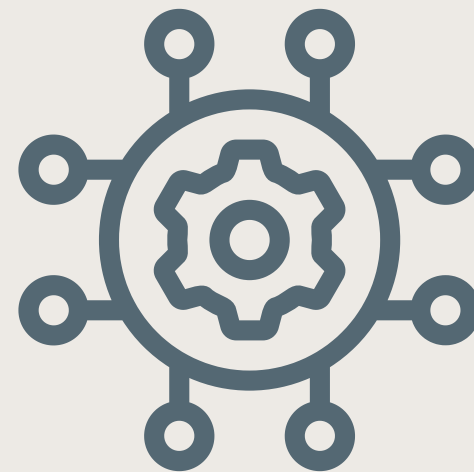
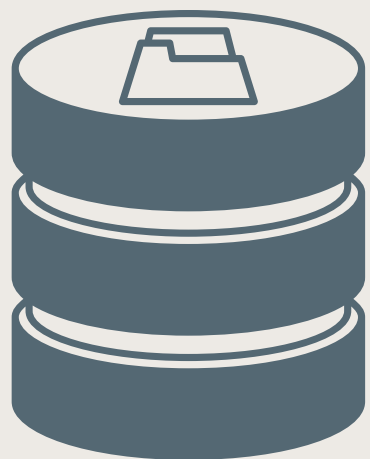


ÉTAPES DE DÉVELOPPEMENT

Données et
pré-traitement

Traitement
algorithmique

Interface utilisateur
et visualisation



PROJET MONGO DB

<https://github.com/charlottebttr/ProjetMongoPython.git>



COMPÉTENCES VALIDÉES

- Définir l'architecture des données
- Ecrire des algorithmes d'analyse de données
- Mettre en place des procédures de sauvegarde et de restauration
- Maîtriser l'analyse et la science de données
- Développer des requêtes SQL et NoSQL pour traiter des données volumineuses

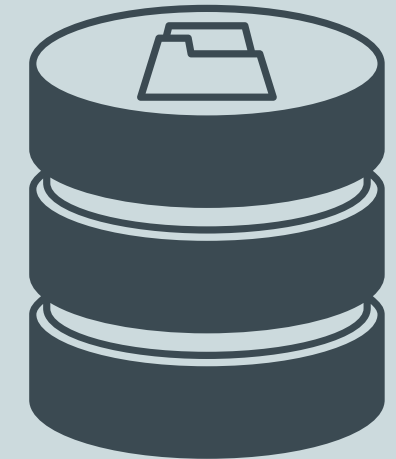
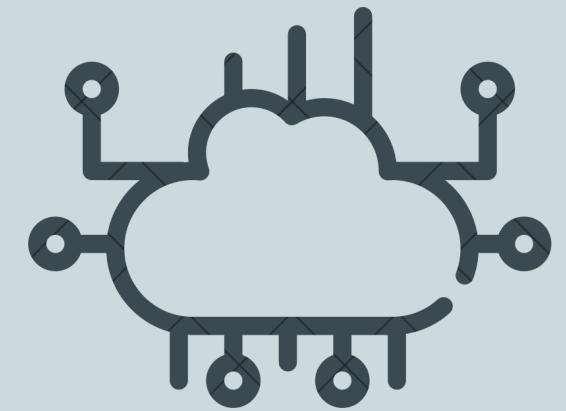


PROJET 1

mongoDB

PRÉSENTATION DES DONNÉES

- Données salariales
- Etats-Unis
- Sources diverses et libre d'accès provenant d'enquêtes et de sites d'offres d'emploi
- 6704 entrées
- 6 variables :
 - âge
 - genre
 - niveau d'éducation
 - années d'expérience
 - poste
 - salaire





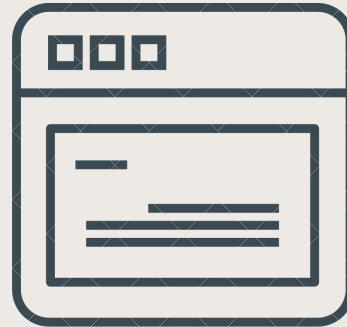
Salaires en fonction du niveau
d'études et de l'expérience
professionnelle



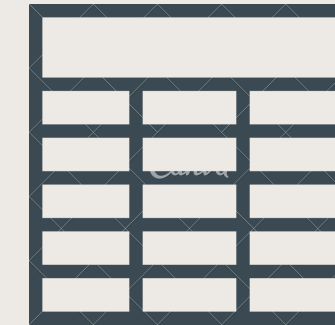
Comparaison des profils types
des 50 hommes les mieux
rémunérés avec les 50 femmes
les mieux rémunérées

FONCTIONNEMENT DE L'APPLICATION

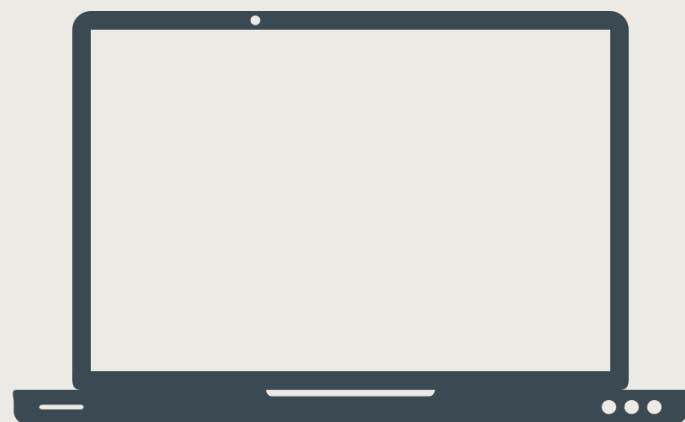
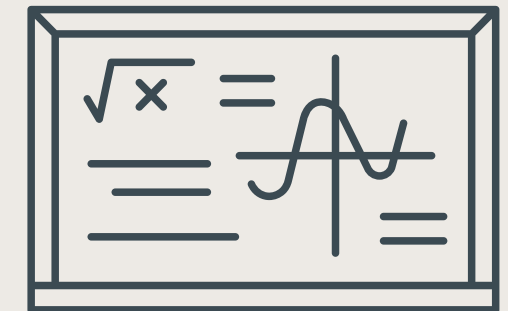
UI.py
→ interface utilisateur



problemSolver.py
→ construction et gestion des requêtes



graphDesiner.py
→ dessine des graphes



MODÈLE DE NOTRE BASE NOSQL

Salaries



Complete profiles

- _id : ObjectId
- Age
- Gender
- Education Level
- Job Title
- Years of Experience
- Salary

Top profiles

- _id : Gender
- AvgSal
- AvgAge
- Experience
- Education

REQUÊTES MONGO DB



```
def solveTop100Salaries(self):
    returned_data = []
    #Récupération des données de notre collection via le Singleton
    db = MongoDBSingleton.get_instance().db["complete_profiles"]
    #Requête permettant de récupérer les 100 individus les mieux rémunérés, de les regrouper par
    # genre et de calculer leur salaire moyen et leur année d'expériences moyenne
    pipeline = [{"$sort": {"Salary": -1 }},
                {'$limit': 100},
                { "$group": {"_id": "$Gender",
                             "AvgSal": {"$avg": "$Salary"},
                             'Genre': {'$sum':1},
                             'Experience': {'$avg': '$Years of Experience'}}},
                {'$sort': {'_id':1}}]
    #Aggregation de la requête
    results = db.aggregate(pipeline)
```

```
def solveBottom100Salaries(self):
    returned_data = []
    #Récupération des données de notre collection via le Singleton
    db = MongoDBSingleton.get_instance().db["complete_profiles"]
    #Requête permettant de récupérer les 100 individus les moins rémunérés, de les regrouper par
    # genre et de calculer leur salaire moyen et leur année d'expériences moyenne
    pipeline = [{"$sort": {"Salary": 1 }},
                {'$limit': 100},
                { "$group": {"_id": "$Gender",
                             "AvgSal": {"$avg": "$Salary"},
                             'Genre': {'$sum':1},
                             'Experience': {'$avg': '$Years of Experience'}}},
                {'$sort': {'_id':1}}]
    #Aggregation de la requête
    results = db.aggregate(pipeline)
```

VISUALISATION DE L'APPLICATION

Projet Mongo - ACA

Page actuelle : 0

Identifiant	Age	Genre	Niveau d'éducation	Emploi	Années d'expérience	
649da739328d09be6e9b3e60	32.0	Male	Bachelor	Software Engineer	5.0	90000.0
649da739328d09be6e9b3e61	28.0	Female	Master	Data Analyst	3.0	65000.0
649da739328d09be6e9b3e62	45.0	Male	PhD	Senior Manager	15.0	150000.0
649da739328d09be6e9b3e63	36.0	Female	Bachelor	Sales Associate	7.0	60000.0
649da739328d09be6e9b3e64	52.0	Male	Master	Director	20.0	200000.0
649da739328d09be6e9b3e65	29.0	Male	Bachelor	Marketing Analyst	2.0	55000.0
649da739328d09be6e9b3e66	42.0	Female	Master	Product Manager	12.0	120000.0
649da739328d09be6e9b3e67	31.0	Male	Bachelor	Sales Manager	4.0	80000.0
649da739328d09be6e9b3e68	26.0	Female	Bachelor	Marketing Coordinator	1.0	45000.0
649da739328d09be6e9b3e69	38.0	Male	PhD	Senior Scientist	10.0	110000.0

Page précédente

Page suivante

SalaryByGenderByEduL

Résoud ma problématique

Page précédente

Page suivante

SalaryByGenderByEduL

SalaryByGenderByEduLeve

SalaryByGenderByYearsOf

Top100Salaries

Bottom100Salaries

Top50SalariesByGender


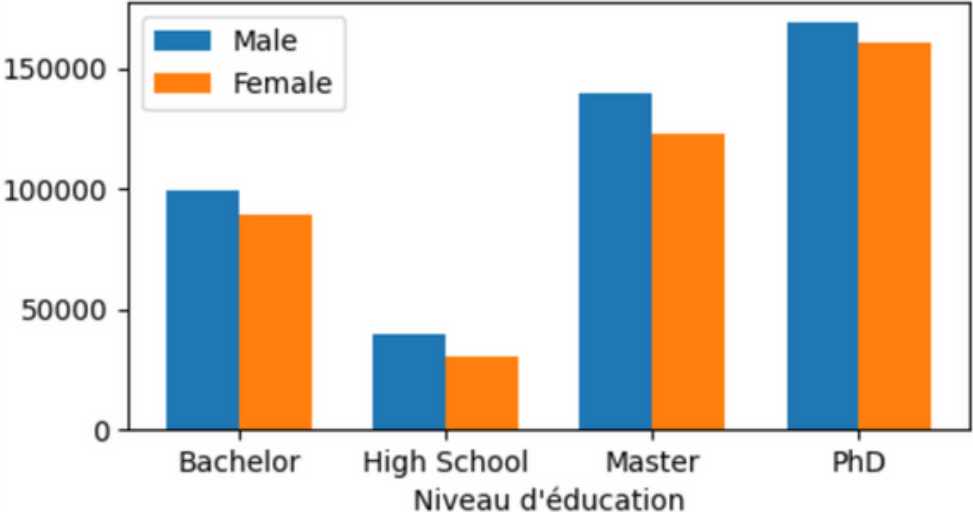


Figure 1

Moyenne des salaires par niveau d'éducation et genre



Niveau d'éducation	Male	Female
Bachelor	90000.0	65000.0
High School	45000.0	30000.0
Master	120000.0	110000.0
PhD	150000.0	140000.0

Home

Previous

Next

Fullscreen

Search

Settings

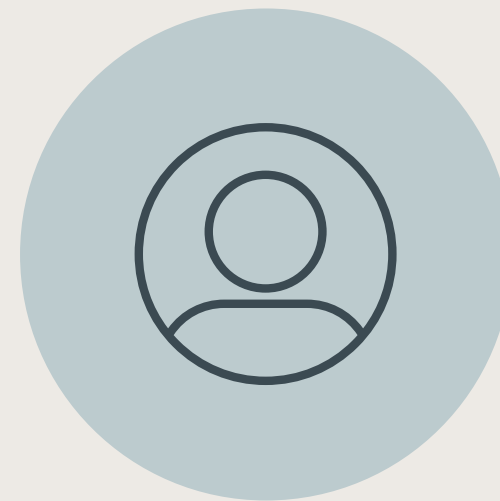
Save

RESULTATS



Profil type homme

```
_id: "Male"  
AvgSal: 223400  
AvgAge: 49.06  
Experience: 21.46  
Education: 7.74
```

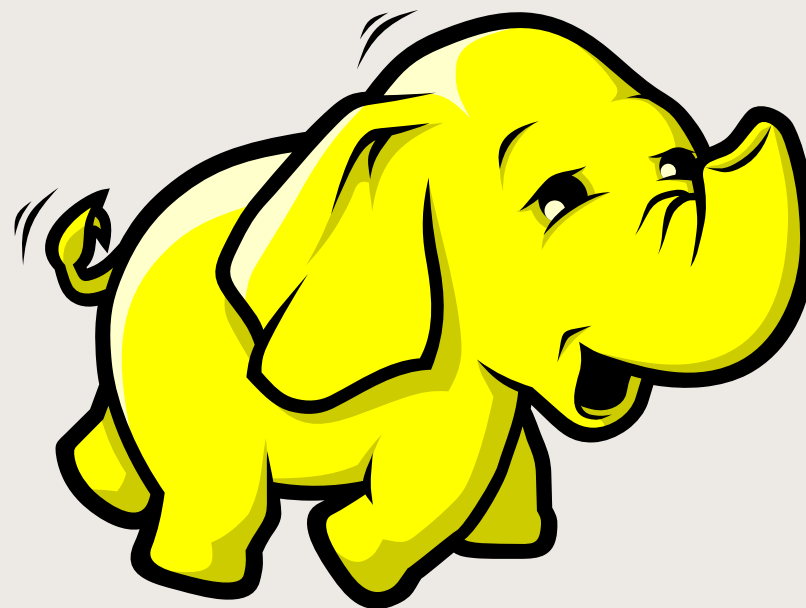


Profil type femme

```
_id: "Female"  
AvgSal: 206900  
AvgAge: 40.6  
Experience: 16.18  
Education: 4.3
```

PROJET BIG DATA

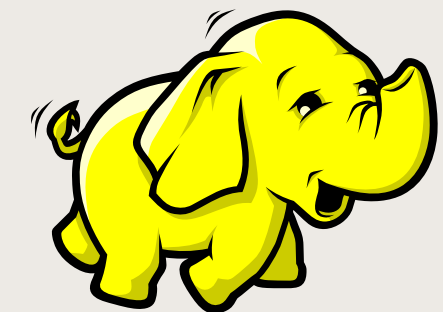
<https://github.com/charlottebtttr/ProjetBigData.git>



COMPÉTENCES VALIDÉES

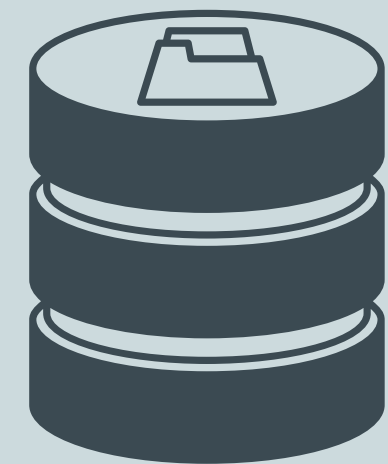
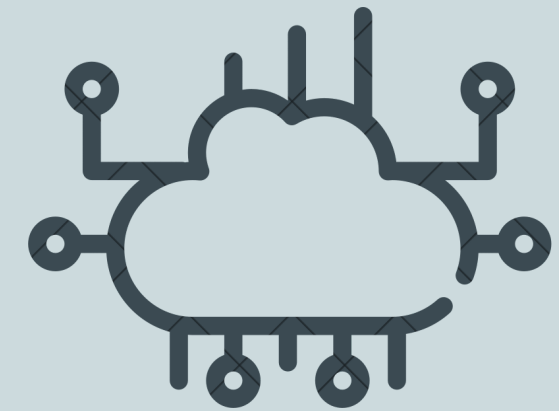
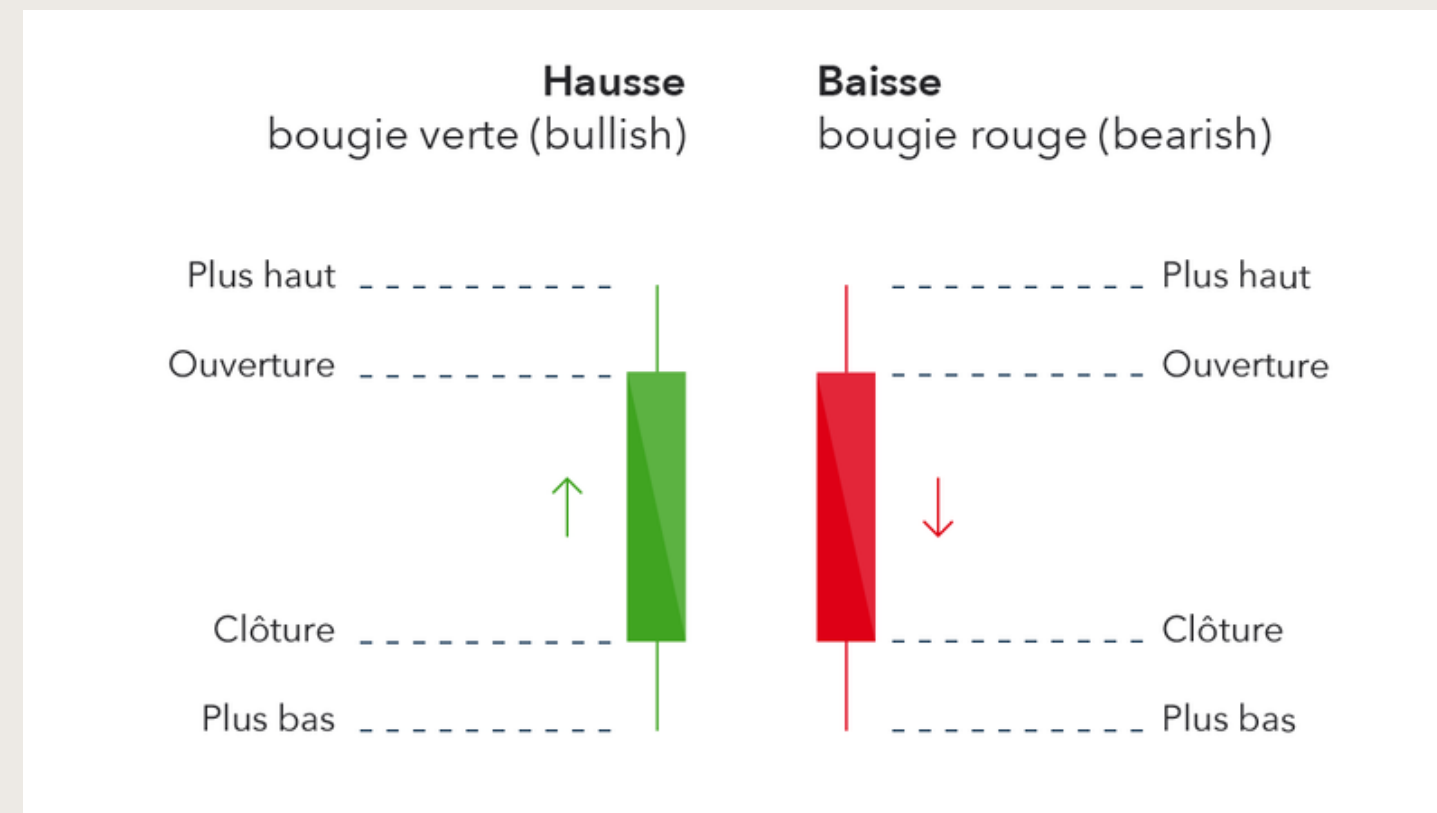
- Mettre en place une solution d'analyse des données volumineuses
- Installer et configurer l'écosystème Hadoop
- Concevoir et déployer un système d'entrepôt de données structurées et non-structurées
- Sécuriser les bases de données

PROJET 2



PRÉSENTATION DES DONNÉES

- Analyse de données boursières
- Plateforme de trading : MetaTrader5
- Temps réel et historique
- Paires de devises majeures
- Données :
 - Temps : date et heure
 - Open
 - Low
 - High
 - Close



Comment extraire, transformer et charger dans une base de données, des données boursières, à travers un système de calcul distribué, dans l'objectif de les visualiser et d'en proposer une analyse stratégique ?



STRUCTURE

MetaTrader 5



Mapper.py

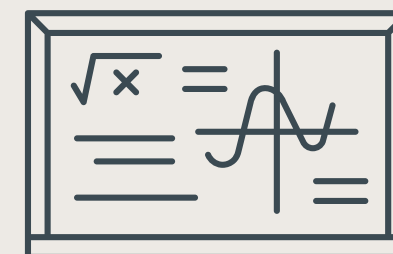


Clé : temps



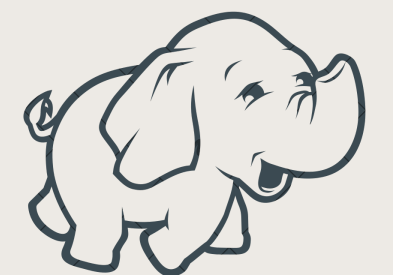
Valeur : close
ou différence

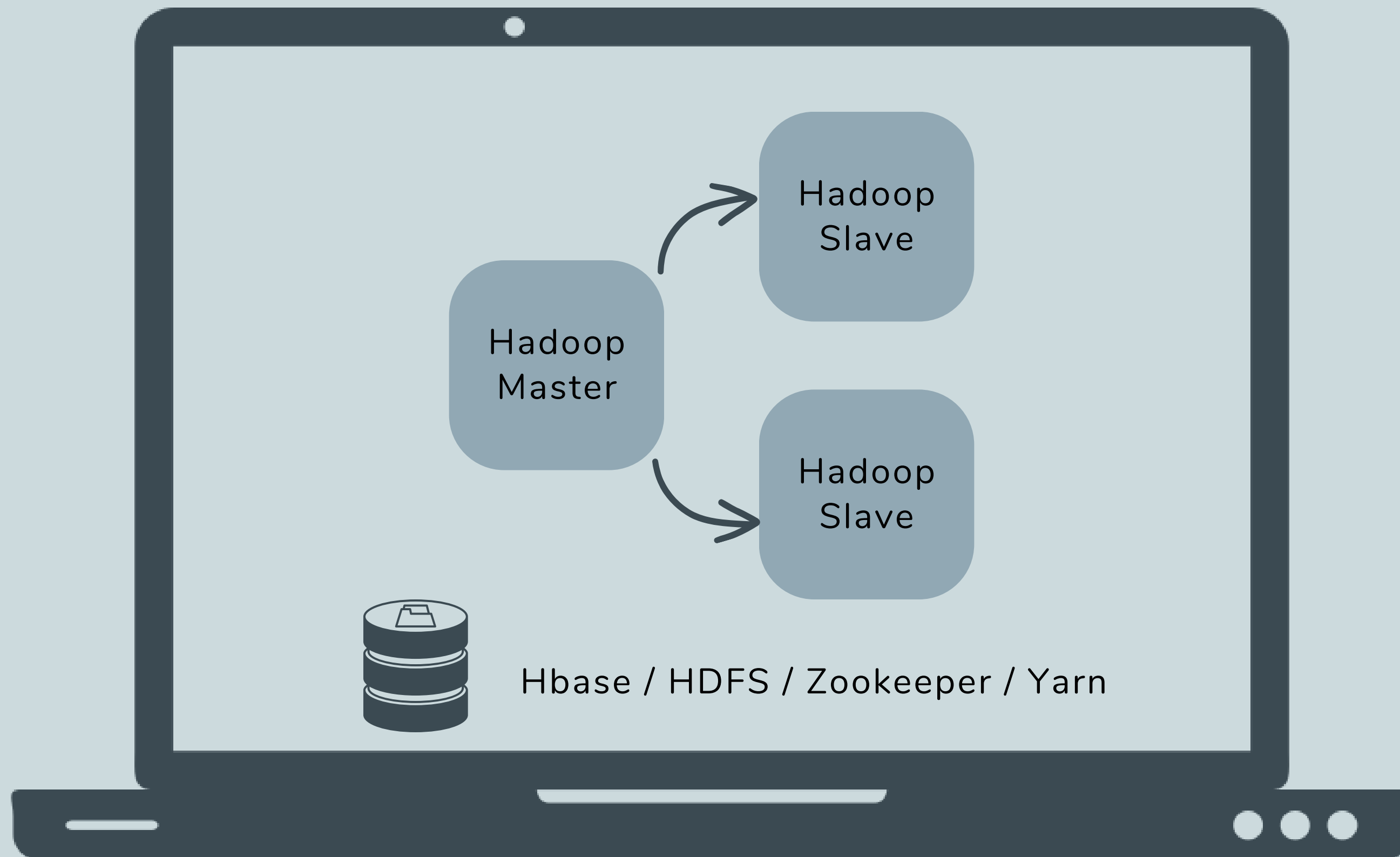
Reducer.py

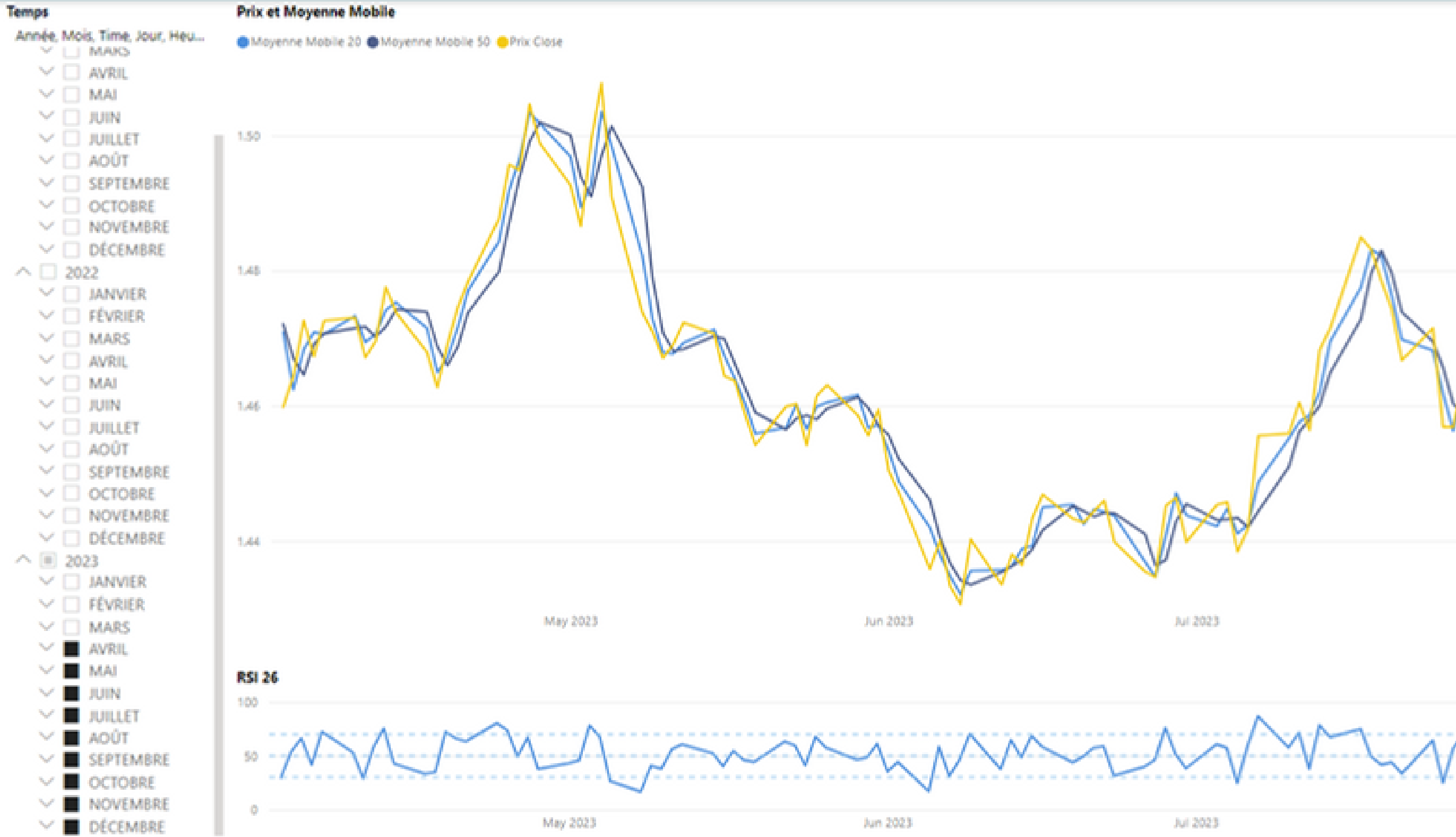


Calculs
Indicateurs

Ecosystème
Hadoop







PROJET MACHINE LEARNING

<https://github.com/charlottebtttr/ProjetMachineLearning.git>



COMPÉTENCES VALIDÉES

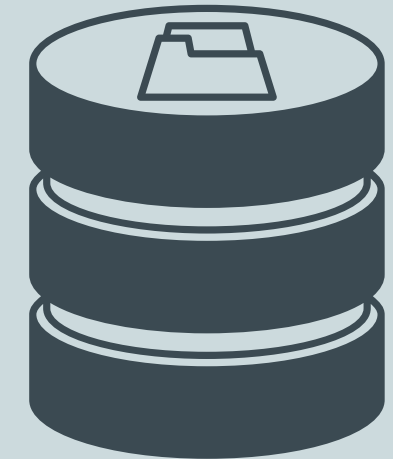
- Concevoir un système d'intelligence artificielle et d'apprentissage automatique
- Rédiger un cahier des charges techniques et répondre à la demande
- Maîtriser la recherche étendue

PROJET 3



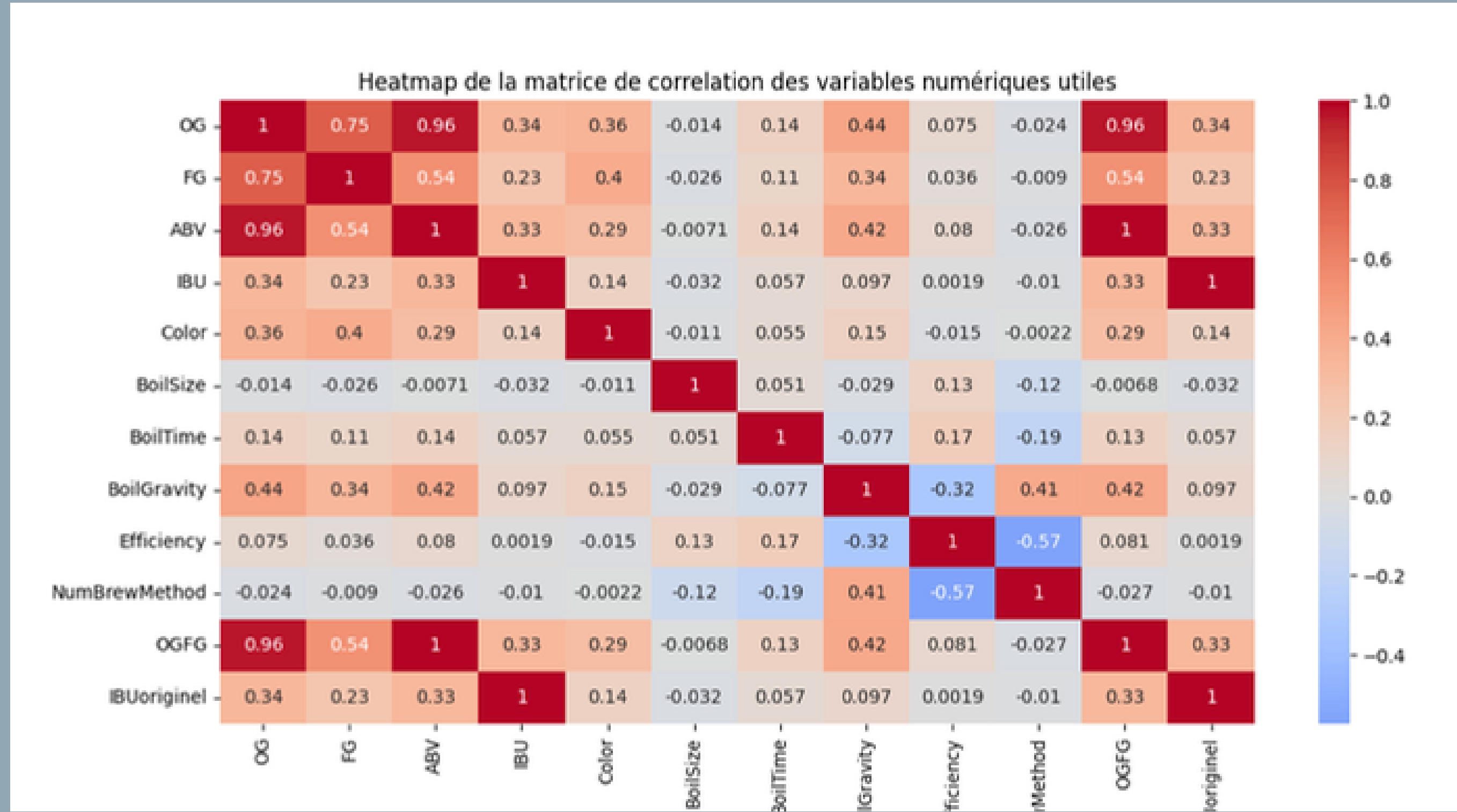
PRÉSENTATION DES DONNÉES

- Données américaines extraites de Kaggle
- Recettes de bières des utilisateurs de Brewer's Friend
- 73861 lignes et 23 colonnes
- Données :
 - Identification des bières (ID, Nom, Brasseur, ID Brasseur, URL)
 - Caractéristiques des bières (Couleur, Style, ABV, IBU)
 - Données propres à la recette et à la méthode (Densités, Volumes, Temps d'ébullition, Températures, Quantités de sucres ou de levures, etc.)

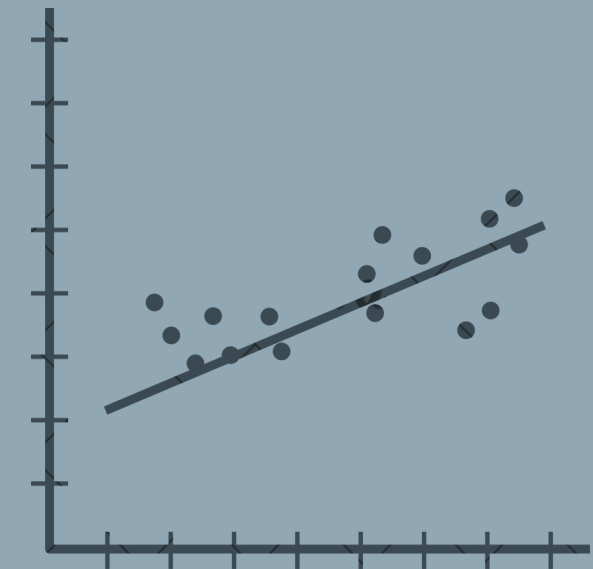


Comment concevoir un système d'intelligence artificielle et d'apprentissage automatique afin de prédire l'amertume et la teneur en alcool d'une bière à partir des autres éléments de la recette ?



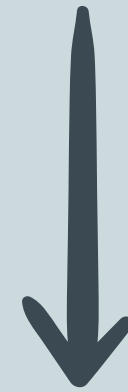


- Tests de différents modèles :
 - Régression linéaire
 - Régression logistique
 - Régression par forêt aléatoire
 - Classification par forêt aléatoire
- Plusieurs jeux de variables différents
 - Mesure des modèles selon MSE, MAE, R^2 et coefficients
 - Optimisation de différents paramètres
- Modèles conservés :
 - ABV et IBU : Régressions par forêt aléatoire



- Implémentation des données et pré-traitement

recipeData.csv → setup.py → cleanData.csv

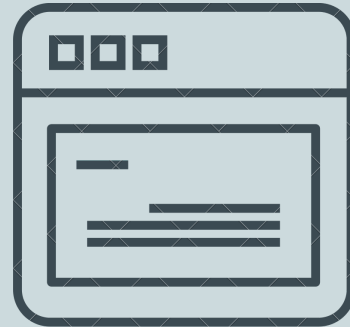


classifieurColor.py
classifieurBrewMethod.py
classifieurStyle.py

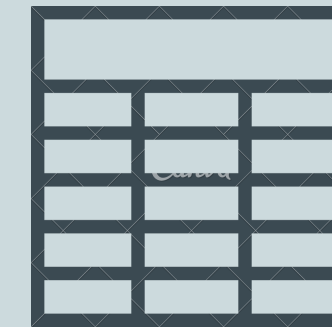
- Implémentation de classifieurs transformant des données textuelles en numériques pour faciliter les prédictions

FONCTIONNEMENT DE L'APPLICATION

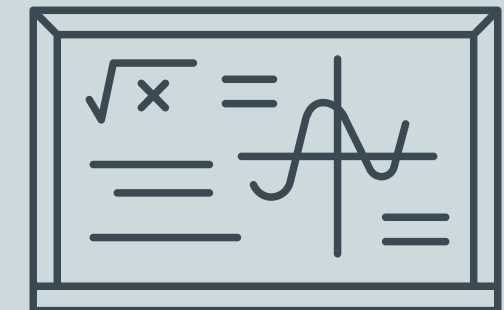
main.py
→ interface



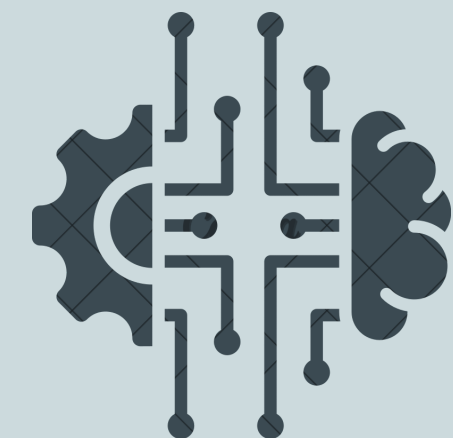
columnMapper.py
→ sélectionne les colonnes utiles



graphDesiner.py
→ dessine des graphes



predicteur.py
→ prédit l'ABV ou l'IBU



Application Web Projet ML

Dessiner un graphique

Formulaire pour dessiner un graphique :

Uniquement pour les graphiques 'bar' et 'pie' :

Prédire une caractéristique

Les caractéristiques de ma bière :

Volume de moût à faire bouillir (L)

Temps d'ébullition du moût (minutes)

Densité du moût avant ébullition (≥ 1)

Densité spécifique du moût avant fermentation (≥ 1)

Densité spécifique du moût après fermentation (≥ 1)

Méthode de brassage utilisée (All grain, BIAB, extract ou Partial Mash)

Caractéristique de la bière à prédire :

Prédire une caractéristique

Les caractéristiques de ma bière :

Volume de moût à faire bouillir (L)

Temps d'ébullition du moût (minutes)

Densité du moût avant ébullition (≥ 1)

Densité spécifique du moût avant fermentation (≥ 1)

Densité spécifique du moût après fermentation (≥ 1)

Méthode de brassage utilisée (All grain, BIAB, extract ou Partial Mash)

Caractéristique de la bière à prédire :

Résultats de la prédiction : [8.372]

Résultats de la prédiction : [39.94]

DÉMONSTRATIONS





**MERCI DE VOTRE ÉCOUTE !
AVEZ-VOUS DES QUESTIONS ?**

