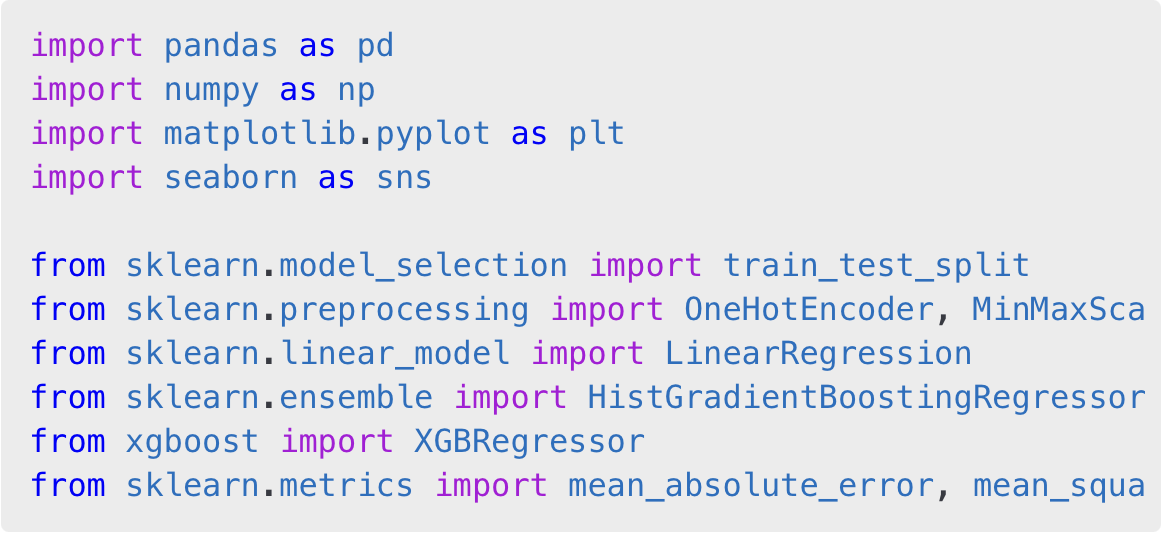
**Documentation Officielle du Projet : Prédiction du Prix des Billets d'Avion**

# 1. Introduction

Ce document fournit une documentation détaillée du code du projet de prédiction des prix des billets d'avion en utilisant des modèles de Machine Learning.

# 2. Importation des Bibliothèques



Ce bloc de code importe les bibliothèques nécessaires pour :

- Manipuler les données (\*\*pandas\*\*, \*\*numpy\*\*)

- Visualiser les données (\*\*matplotlib\*\*, \*\*seaborn\*\*)

- Prétraiter les données (\*\*OneHotEncoder\*\*, \*\*MinMaxScaler\*\*)

- Modéliser et évaluer les performances (\*\*LinearRegression\*\*, \*\*XGBRegressor\*\*, \*\*HistGradientBoostingRegressor\*\*)

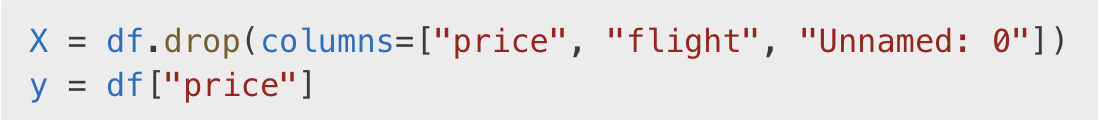
# 3. Chargement et Prétraitement des Données

Chargement du Dataset



- Le fichier CSV contenant les données de vol est chargé dans un DataFrame pandas.

Sélection des Variables Pertinentes

```

- On supprime les colonnes non pertinentes : \*\*flight\*\* et \*\*Unnamed: 0\*\*.

- \*\*X\*\* contient les variables explicatives, et \*\*y\*\* la variable cible (\*\*price\*\*).

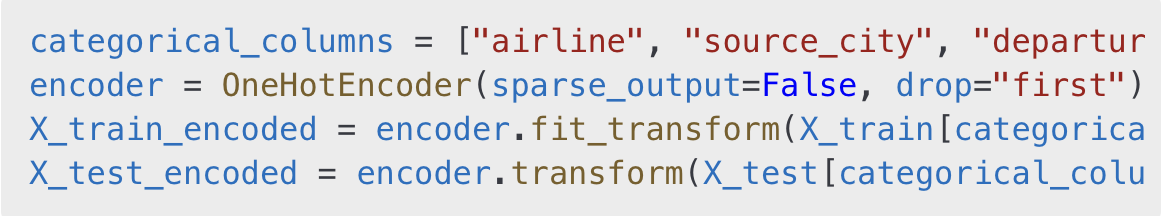
Séparation des Données en Train et Test

```

- Les données sont divisées en 80% pour l'entraînement et 20% pour le test.

- random\_state=42\*\* permet de garantir la reproductibilité des résultats.

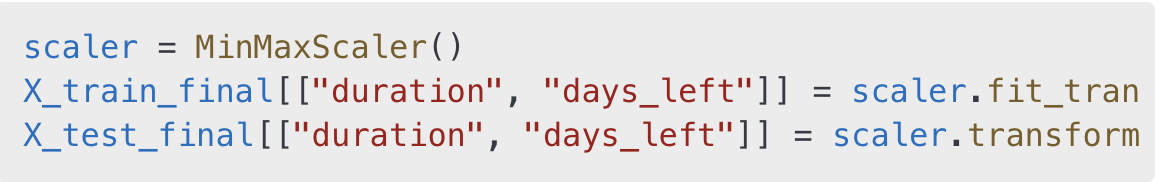
# Encodage des Variables Catégoriques



- Les variables qualitatives sont encodées en variables binaires via \*\*One-Hot Encoding\*\*.

- L'option \*\*drop="first"\*\* permet d'éviter la multicolinéarité.

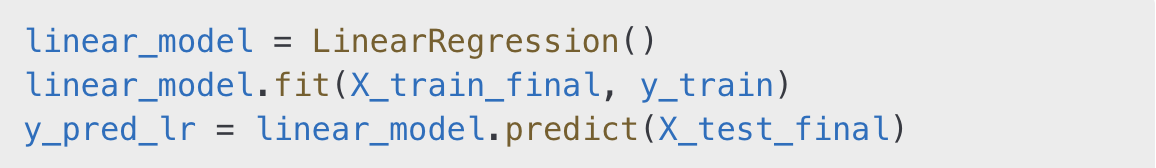
# Normalisation des Variables Numériques

```

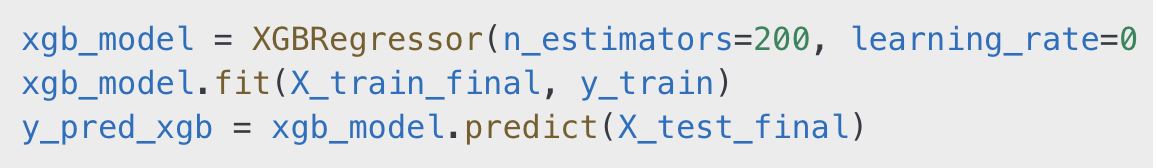
- La normalisation des variables numériques permet d'uniformiser les échelles et d'améliorer l'efficacité des modèles.

# 4. Modélisation et Évaluation

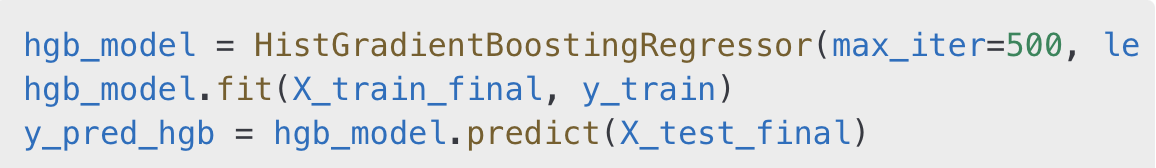
## 4.1 Régression Linéaire

- Utilisation d'une régression linéaire pour obtenir une baseline des performances.

## 4.2 XGBoost

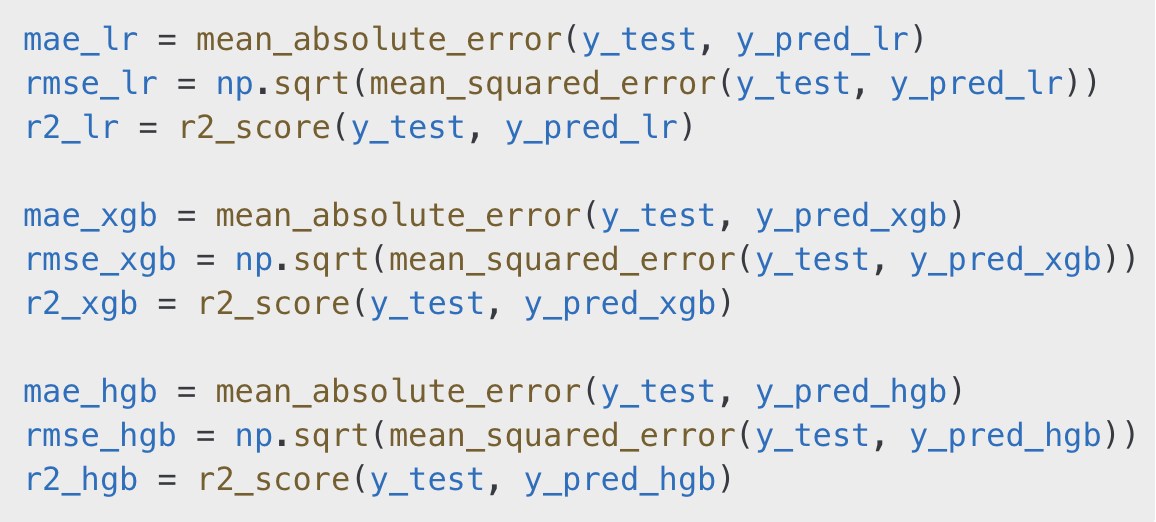
- Modèle de boosting puissant pour capturer les relations non linéaires.

## 4.3 HistGradientBoosting



- Une version optimisée du Gradient Boosting pour de meilleures performances.

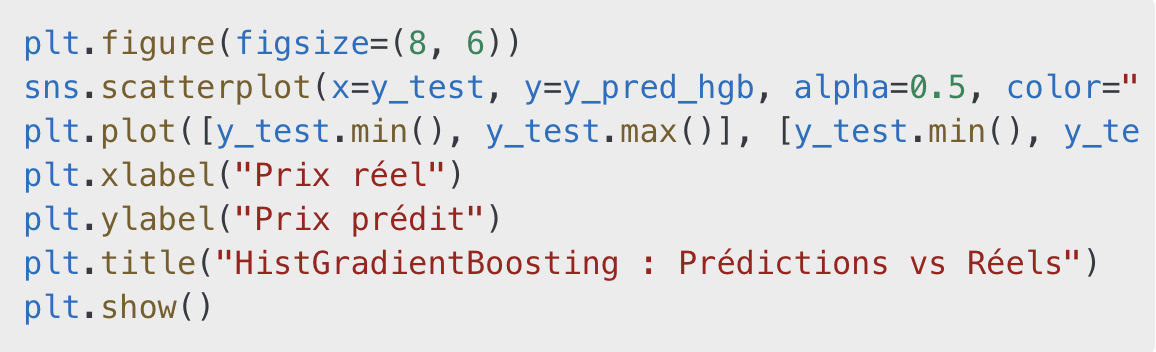
## 4.4 Évaluation des Modèles



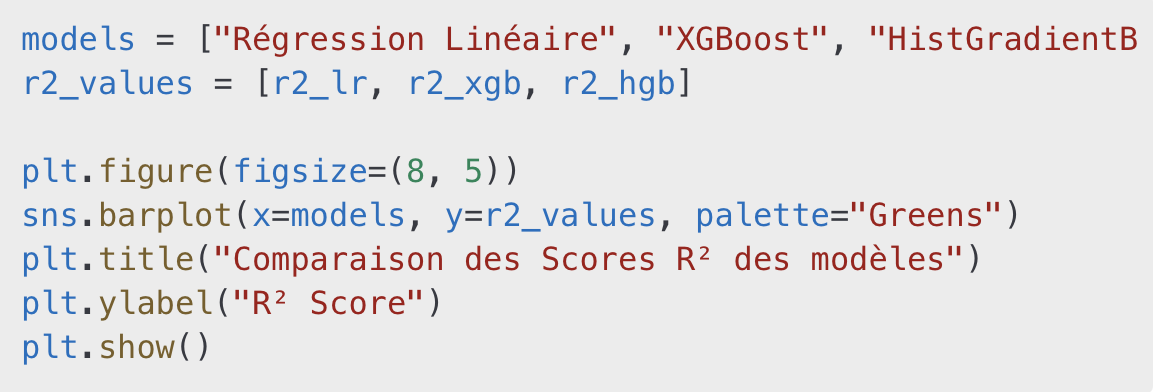
- Comparaison des modèles selon les métriques : \*\*MAE, RMSE, R²\*\*.

# 5. Visualisation des Résultats

## Prédictions vs Valeurs Réelles

- Visualisation des performances du meilleur modèle.

## Comparaison des Scores R²

- Comparaison des performances des modèles en fonction de leur \*\*R²\*\*.

# 6. Conclusion

- HistGradientBoosting s’est avéré être le meilleur modèle pour la prédiction des prix des billets d’avion.

- Les métriques MAE et RMSE les plus faibles indiquent une meilleure précision des prédictions.

- Des améliorations possibles incluent l’ajout de nouvelles variables (météo, événements spéciaux) et l’utilisation de modèles avancés comme le Deep Learning.