

TP PROJET: RECETTE DE BIÈRE

FATIH EYILI, GILLIAN CATIL, MIKAEL LE DEVEHAT



TABLE DES MATIÈRES

- CONTEXTE
 - BREW
 - PROBLÉMATIQUE
- LES DONNÉES
 - DONNÉES FOURNIS
 - NETTOYAGE
 - ANALYSE EXPLORATOIRE
- COMMENT PRÉDIRE ABV ?
 - OG ET ABV
 - SÉPARATION DES DONNÉES
 - RÉGRESSION LINÉAIRE
 - MSE/MAE
- COMMENT PRÉDIRE IBU ?
 - LES ESSAIS
 - RANDOM FOREST CLASSIFIER
 - HALVING GRID SEARCH
 - PRÉCISION
- L'APPLICATION WEB
 - CRÉATION DU FORMULAIRE
 - PRÉSENTATION
- CONCLUSION

CONTEXTE

BREW



PROBLÉMATIQUE



LES DONNÉES

DONNÉES FOURNIS

- DEUX FICHIERS CSV
- KAGGLE
- SANS DONNÉE SUR LES INGRÉDIENTS

NETTOYAGE

VALEURS MANQUANTES

COLONNES PERTINENTES

BoilGravity has 2990 (4.0%) missing values

MashThickness has 29864 (40.4%) missing values

PitchRate has 39252 (53.1%) missing values

PrimaryTemp has 22662 (30.7%) missing values

PrimingMethod has 67095 (90.8%) missing values

PrimingAmount has 69087 (93.5%) missing values

UserId has 50490 (68.4%) missing values

BeerID is uniformly distributed

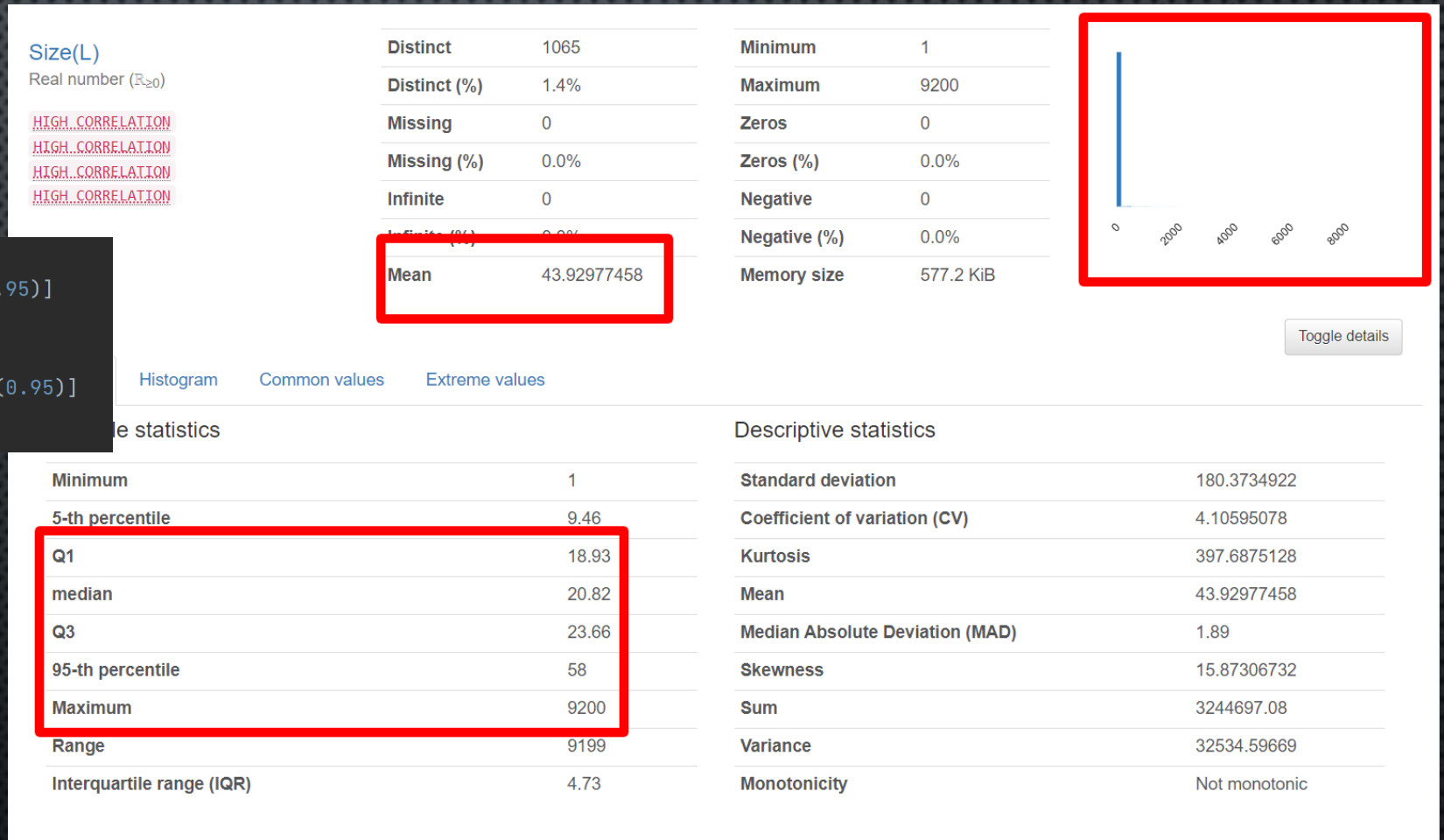
BeerID has unique values

URL has unique values

NETTOYAGE

- GESTION DES OUTLIERS

```
# Gestion des Outliers ++
df = df[df["Size(L)"] <= df["Size(L)"].quantile(0.95)]
df = df[df["OG"] <= df["OG"].quantile(0.95)]
df = df[df["FG"] <= df["FG"].quantile(0.95)]
df = df[df["BoilSize"] <= df["BoilSize"].quantile(0.95)]
df = df[(df["IBU"] <= 150) & (df["IBU"] > 0)]
```



NETTOYAGE

GESTION DES ZÉROS POUR IBU

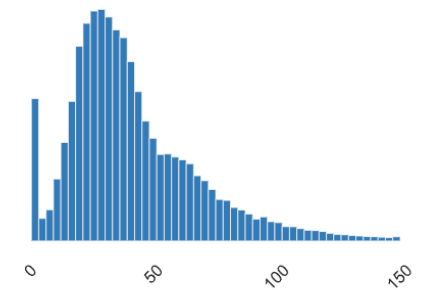
IBU

Real number ($\mathbb{R}_{\geq 0}$)

ZEROS

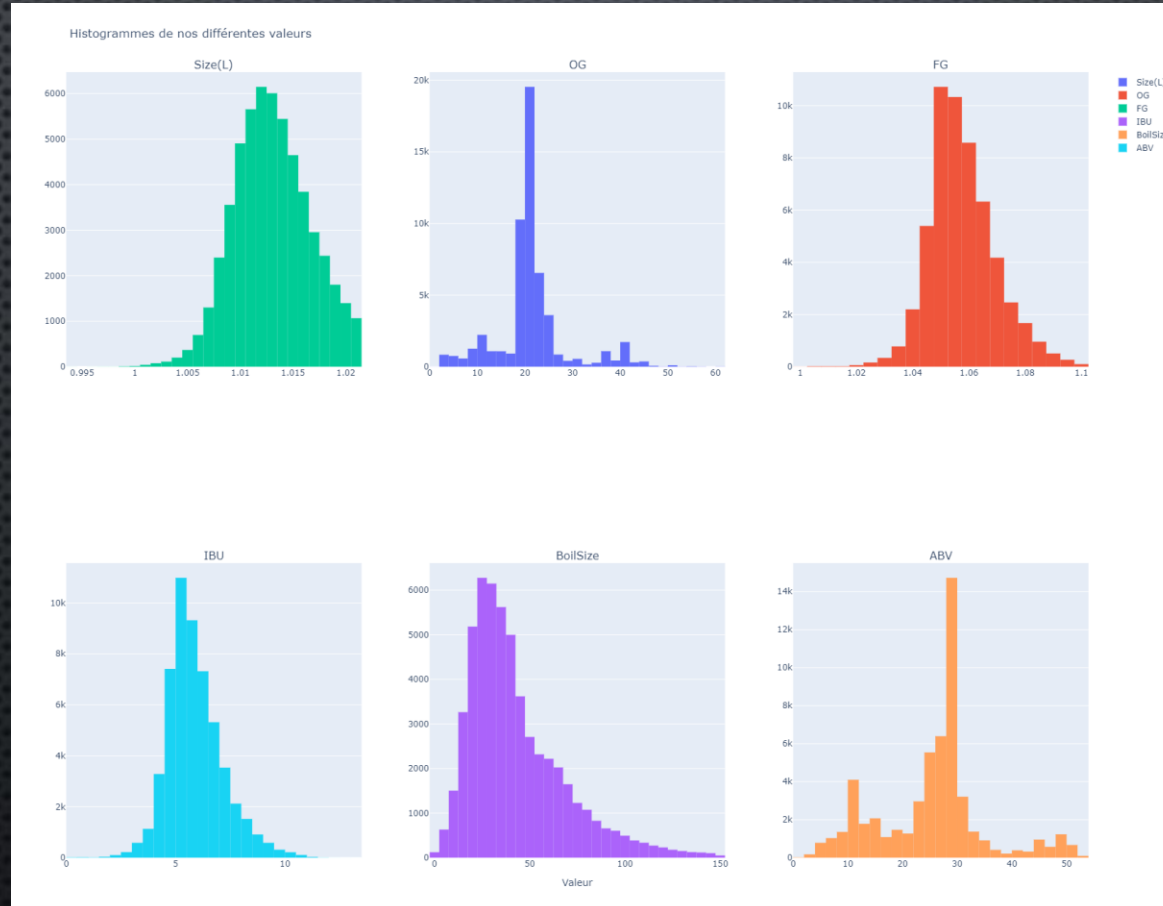
Distinct	10707
Distinct (%)	18.7%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	41.23892535

Minimum	0
Maximum	149.94
Zeros	2147
Zeros (%)	3.8%
Negative	0
Negative (%)	0.0%
Memory size	447.3 KiB

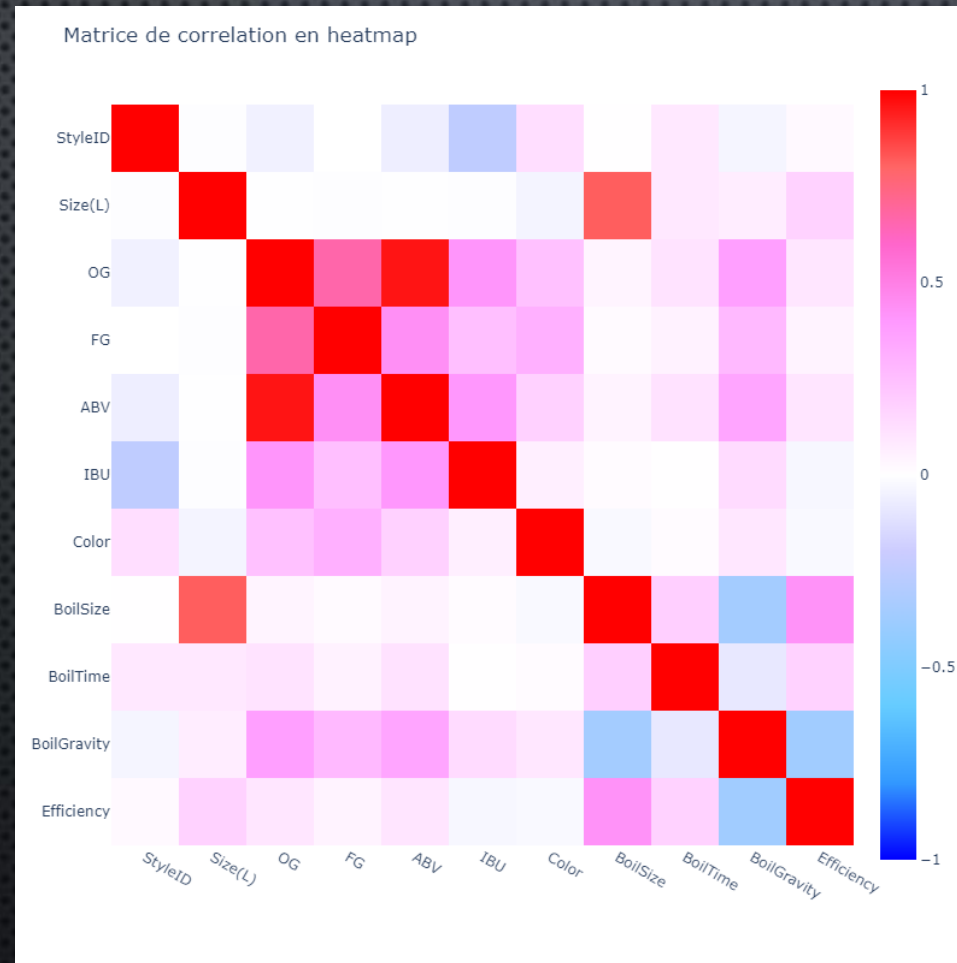


Toggle details

ANALYSE EXPLORATOIRE

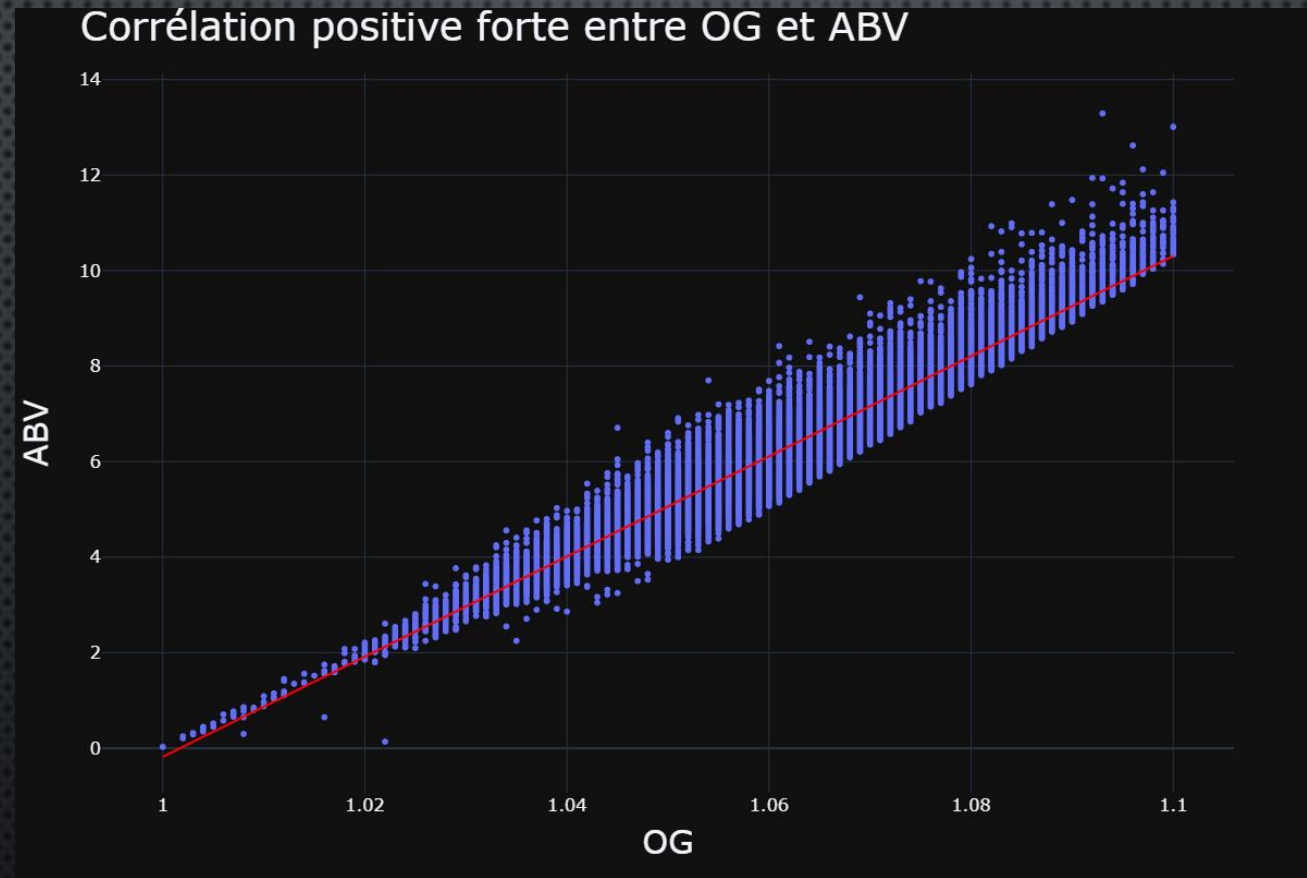


ANALYSE EXPLORATOIRE



COMMENT PRÉDIRE ABV ?

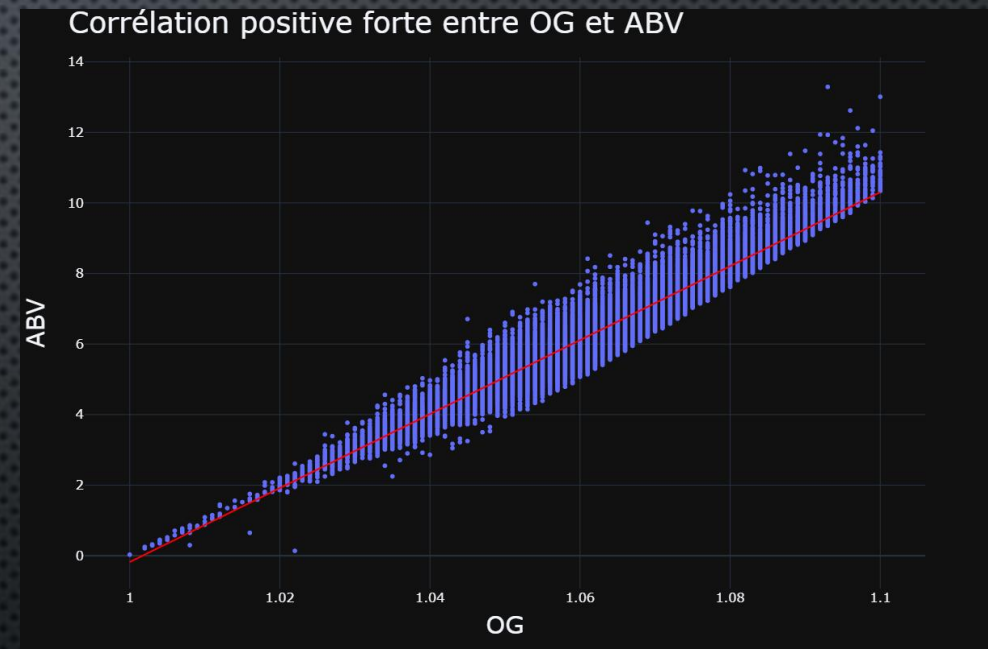
OG ET ABV



RÉGRESSION LINÉAIRE

```
RL = LinearRegression()
```

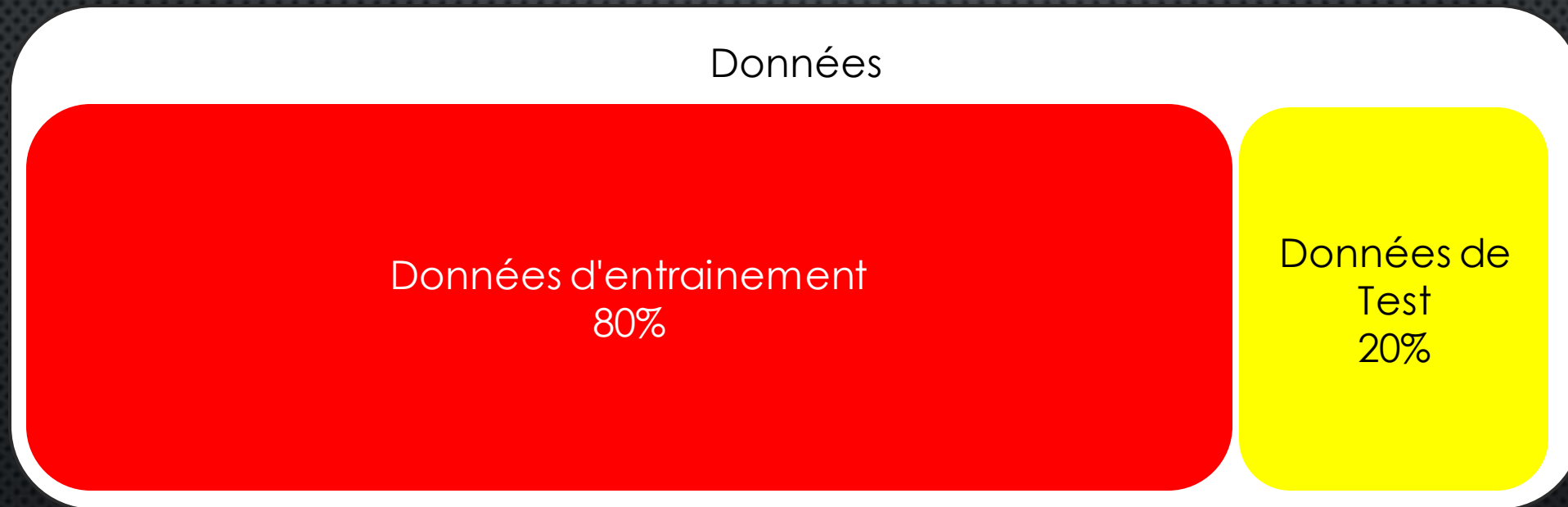
- Technique d'analyse
- $Y = aX + b$
- Définir un maximum de point
- Ecart minimal
- Prédire une valeur inconnue



SÉPARATIONS DES DONNÉES

SCIKIT-LEARN : TRAIN_TEST_SPLIT

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```



MSE/MAE

- MEAN SQUARE VALUE:

$$MSE(y, \hat{y}) = \frac{1}{N} \sum_i^N (\hat{y}_i - y_i)^2$$

$$MSE = 0.129432$$

N le nombre d'exemples du jeu d'entraînement
 y_i la vraie valeur
 \hat{y}_i la valeur prédite = $f(x_i)$

- MEAN ABSOLUTE VALUE:

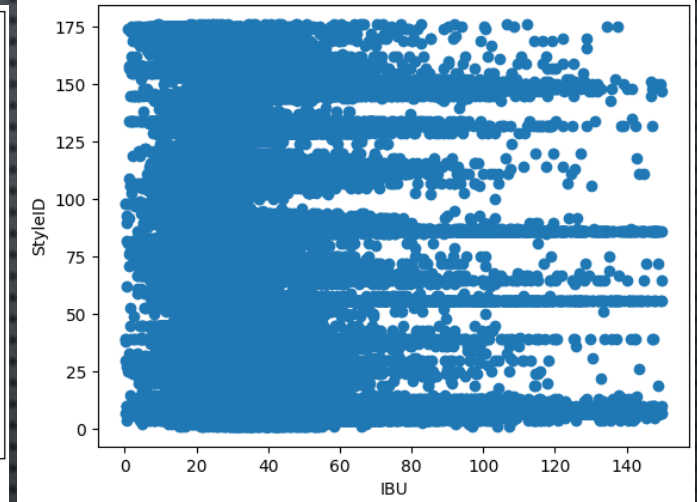
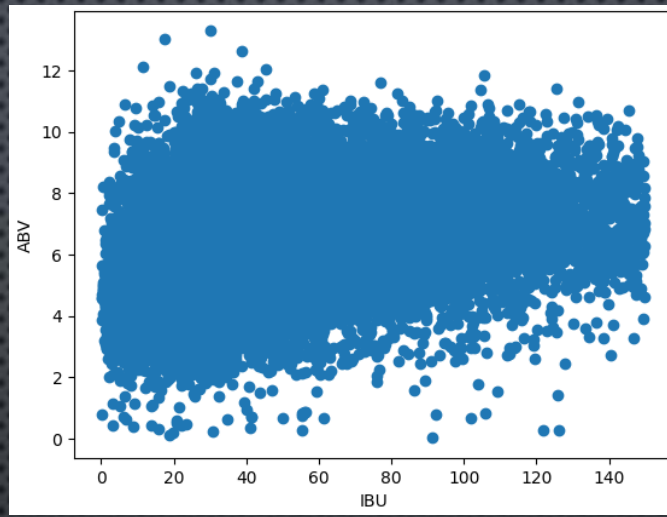
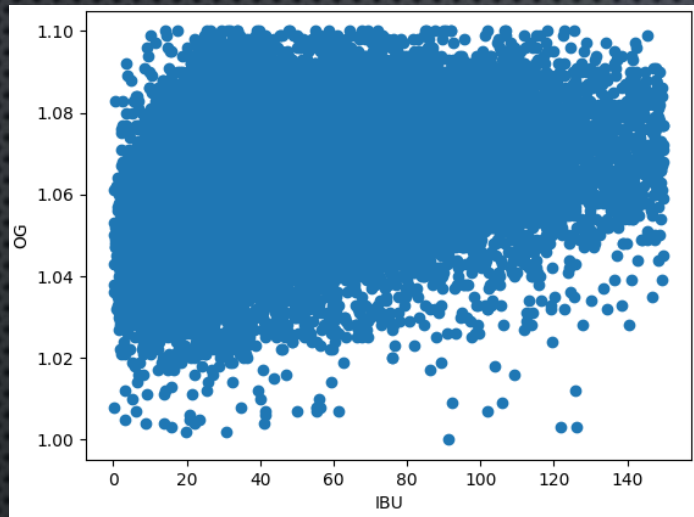
$$MAE(y, \hat{y}) = \frac{1}{N} \sum_i^N |\hat{y}_i - y_i|$$

$$MAE = 0.281369$$

N le nombre d'exemples du jeu d'entraînement
 y_i la vraie valeur
 \hat{y}_i la valeur prédite = $f(x_i)$

COMMENT PRÉDIRE IBU ?

ESSAI



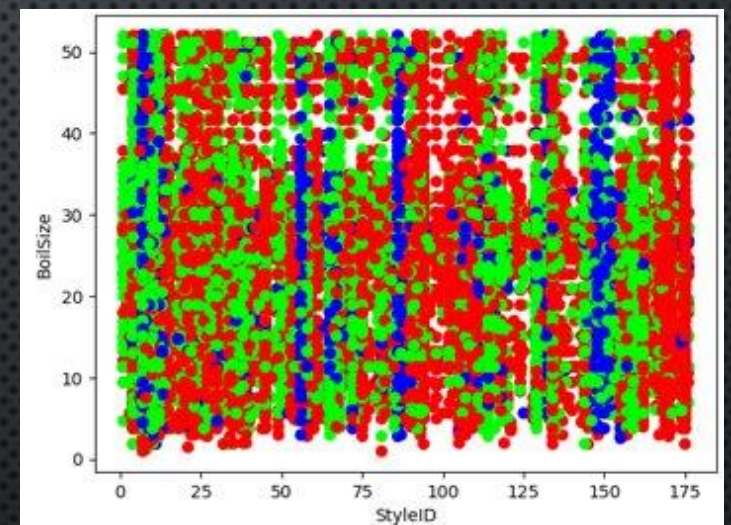
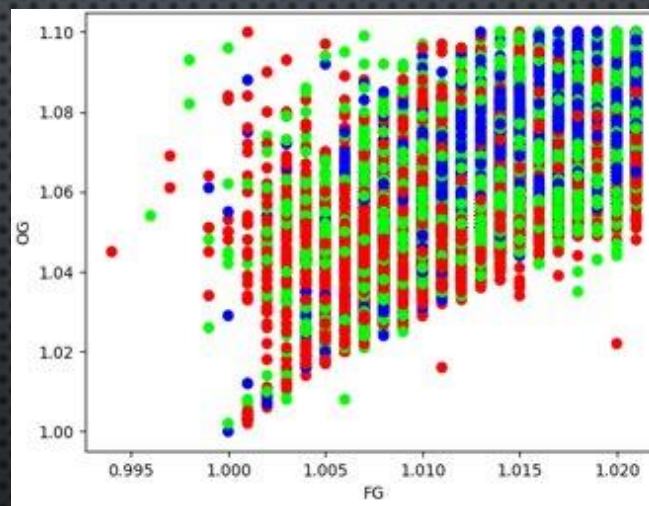
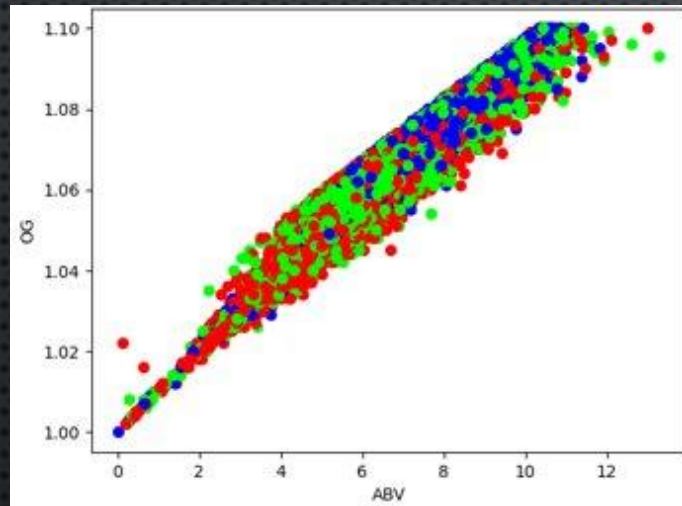
Donnée de IBU

LOW
[0-30]

MEDIUM
[30-60]

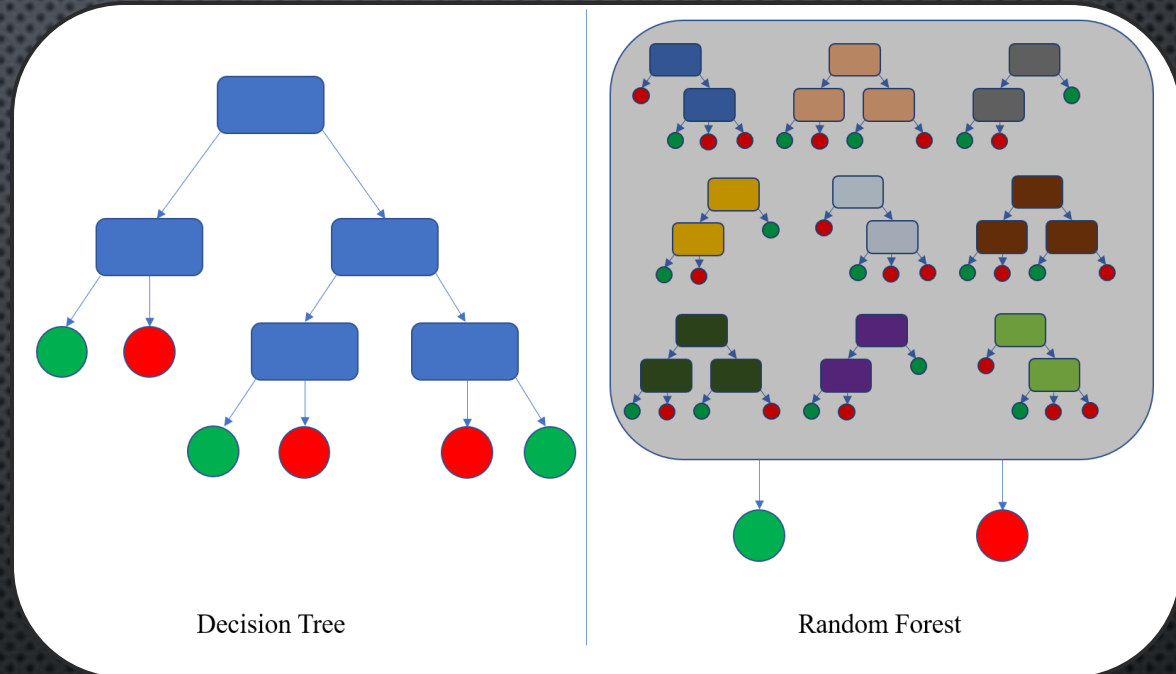
HIGH
[60-150]

ESSAI



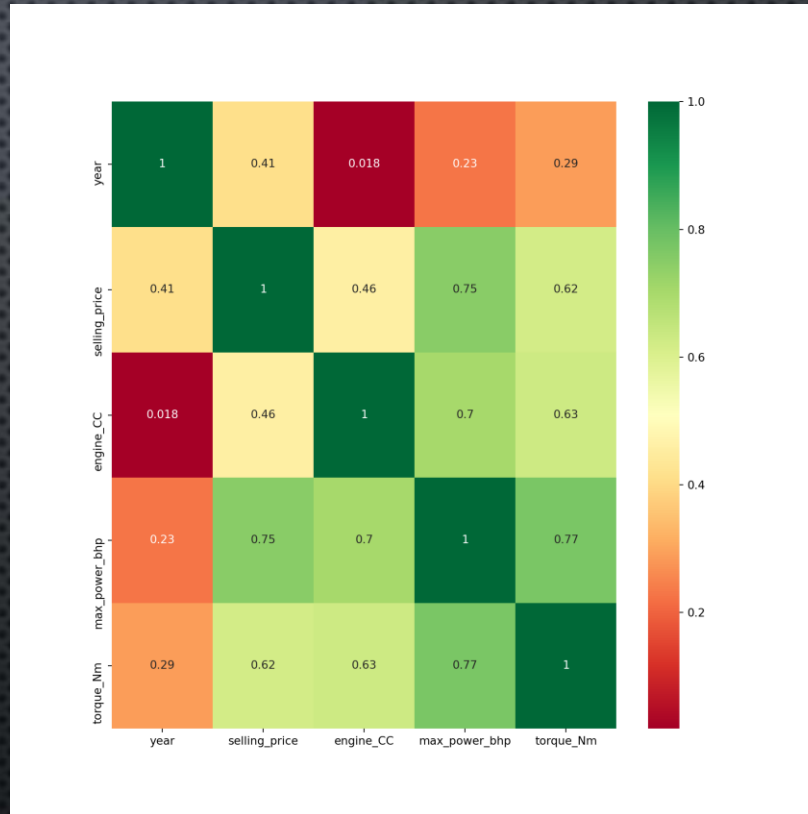
RANDOM FOREST

- STYLEID, EFFICIENCY, OG, BOILGRAVITY, ABV



HALVING GRID SEARCH

- OPTIMISATION DES HYPERPARAMETRES
- MAX_DEPTH,
MIN_SAMPLES_LEAF, MIN_SAMPLES_SPLIT,
MAX_FEATURES
- LA RESSOURCE EST : N_ESTIMATORS



PRÉCISION

- PRÉCISION OBTENUE (EN %):

Accuracy: 64.40

- EXPORT DES DEUX MODÈLES
VIA JOBLIB

```
dump(RL, 'model_Linear_Regression.joblib')
```

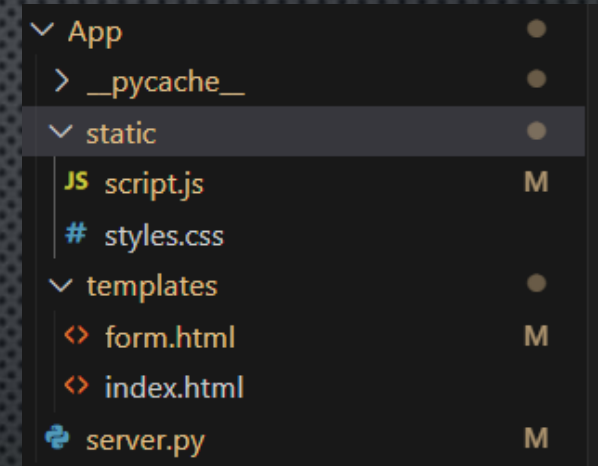
```
dump(rf_best, 'model_Random_Forest.joblib')
```

L'APPLICATION WEB

CRÉATION DU SERVEUR

```
Lr = load('../model_Linear_Regression.joblib')  
Rf = load('../model_Random_Forest.joblib')
```

- FRAMEWORK UTILISÉ: FLASK
- MODÈLES CHARGÉS AVEC JOBLIB



```
@app.route('/')  
def index():  
    return render_template('form.html', columns=columns)  
  
@app.route('/query', methods=['POST'])  
def traiter_requete_ajax():...
```

CRÉATION DU FORMULAIRE

- HTML, CSS, JS NATIF

Projet IA - Mikael, Gillian, Fatih

Formulaire de prédiction pour ABV et IBU

OG :

StyleID :

Efficiency :

BoilGravity :

Prédire

GESTION D'ERREURS

Projet IA - Mikael, Gillian, Fatih

Formulaire de prédiction pour ABV et IBU

OG : 1.07

StyleID : 1

Efficiency : 70

BoilGravity : 1.07

Prédire

Le serveur a rencontré une erreur lors du traitement des données!

Projet IA - Mikael, Gillian, Fatih

Formulaire de prédiction pour ABV et IBU

OG : m

StyleID :

Efficiency :

BoilGravity :

Prédire

Les informations entrées ne sont pas correctes!

- REGEX POUR LES CHAMPS
- CODE 400

AFFICHAGE DES RÉSULTATS

- ABV ET IBU AFFICHÉS
- LÉGENDE CORRESPONDANTE
- AJAX, DONC DYNAMIQUE

Projet IA - Mikael, Gillian, Fatih

Formulaire de prédiction pour ABV et IBU

OG : 1.09

StyleID : 1

Efficiency : 70

BoilGravity : 1.070

Prédire

Voici quelles sont les prédictions :

ABV
9.2586

IBU
High

ABV: Représente le pourcentage d'alcool du volume de la bière

IBU: Représente l'amertume de la bière en PPM (parts par million), divisé en 3 catégories: Low(0-30), Medium(30-60), High(60-150)

CONCLUSION

MERCI DE VOTRE
ATTENTION