



APPRENTISSAGE ARTIFICIEL

PROJET FINAL 2025

Analyse et Prédiction des Prix Dior sur
le Marché Chinois

Chloé Makoundou Master 1 IBD

SOMMAIRE

- 1) Introduction
- 2) Dataset Analyse Exploratoire
- 3) Nettoyage des données
- 4) Modélisation
- 5) Evaluation
- 6) Conclusion



INTRODUCTION

Les consommateurs chinois façoneront l'avenir du luxe. » a déclaré Andrew Wu, Président du Groupe LVMH Chine.



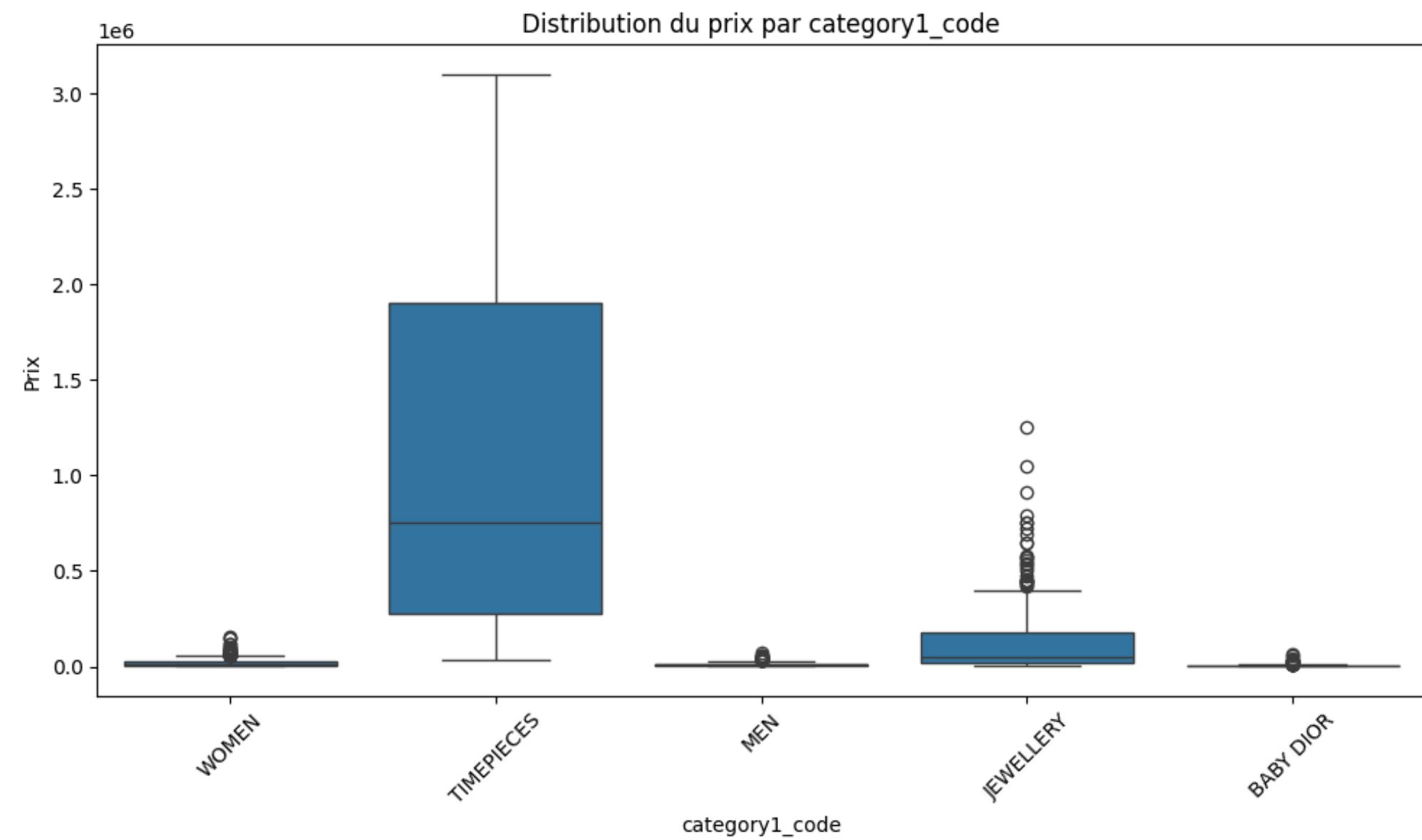
DATASETS

Comment modéliser et prédire efficacement les prix des produits Dior de Chine, et quels sont les facteurs déterminants qui influencent le pricing dans le secteur du luxe ?



ANALYSE EXPLORATOIRE

```
RangeIndex: 4499 entries, 0 to 4498
Data columns (total 17 columns):
 #   Column           Non-Null Count  Dtype  
 ---  --  
 0   website_name     4499 non-null   object  
 1   competence_date  4499 non-null   object  
 2   country_code     4499 non-null   object  
 3   currency_code    4499 non-null   object  
 4   brand            4499 non-null   object  
 5   category1_code   4499 non-null   object  
 6   category2_code   4499 non-null   object  
 7   category3_code   4498 non-null   object  
 8   product_code     4499 non-null   object  
 9   title             4499 non-null   object  
 10  itemurl          4499 non-null   object  
 11  imageurl         4499 non-null   object  
 12  full_price       4499 non-null   float64 
 13  price             4499 non-null   float64 
 14  full_price_eur   4499 non-null   float64 
 15  price_eur        4499 non-null   float64 
 16  flg_discount     4499 non-null   int64  
dtypes: float64(4), int64(1), object(12)
memory usage: 597.7+ KB
```



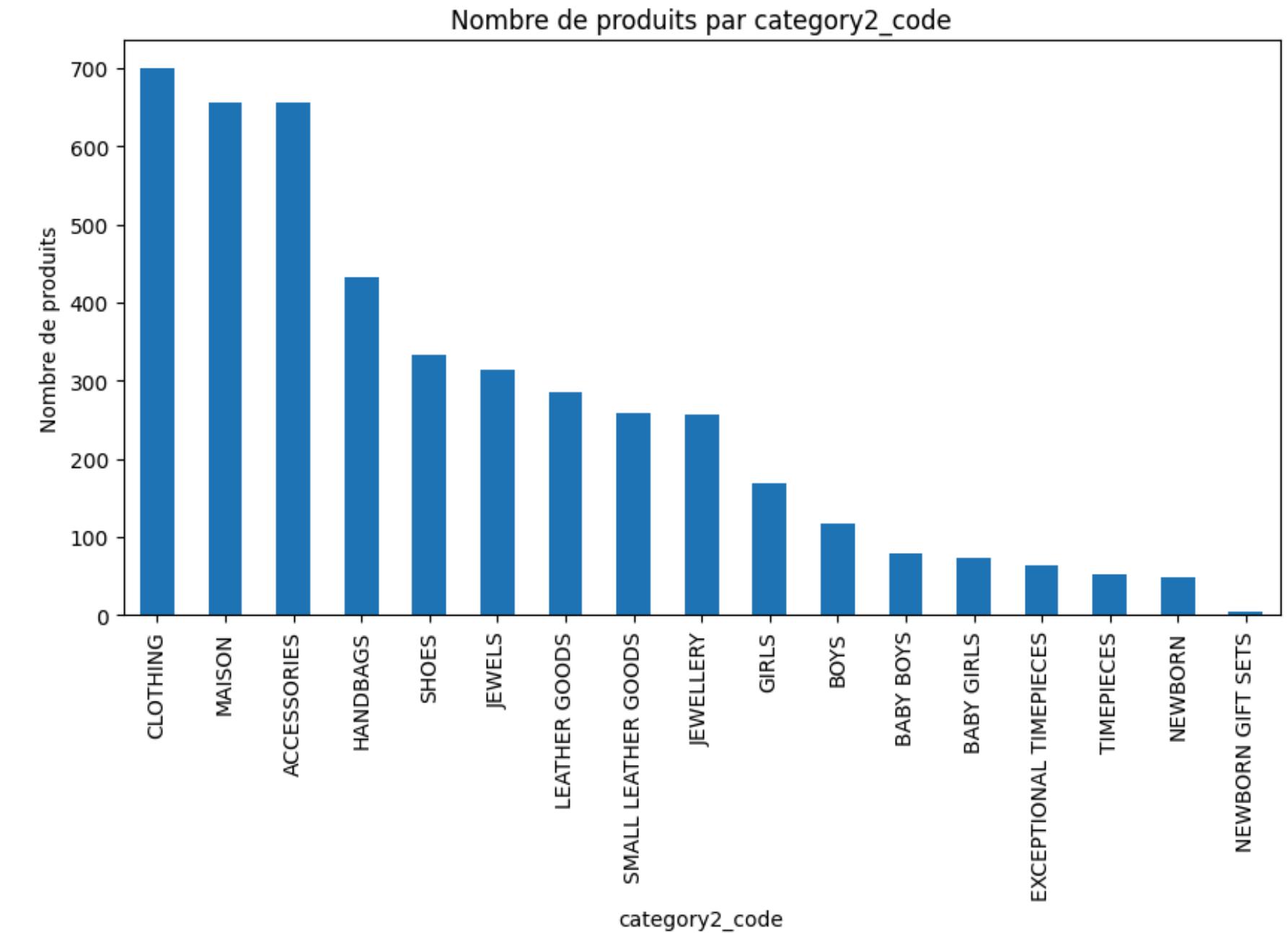
NETTOYAGE DES DONNÉES

Features

- Category1_code
- Category2_code
- Category3_code

Target

- Price (En CYN)



MODÉLISATION

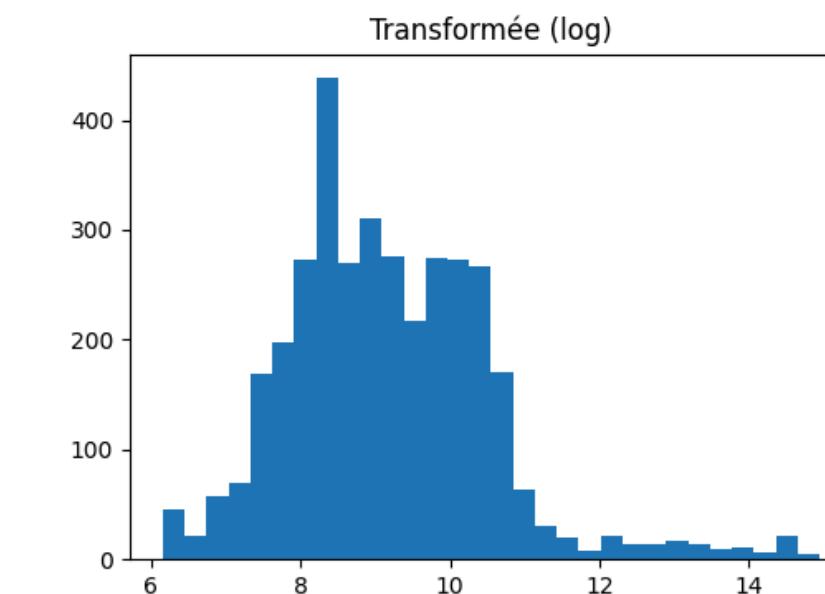
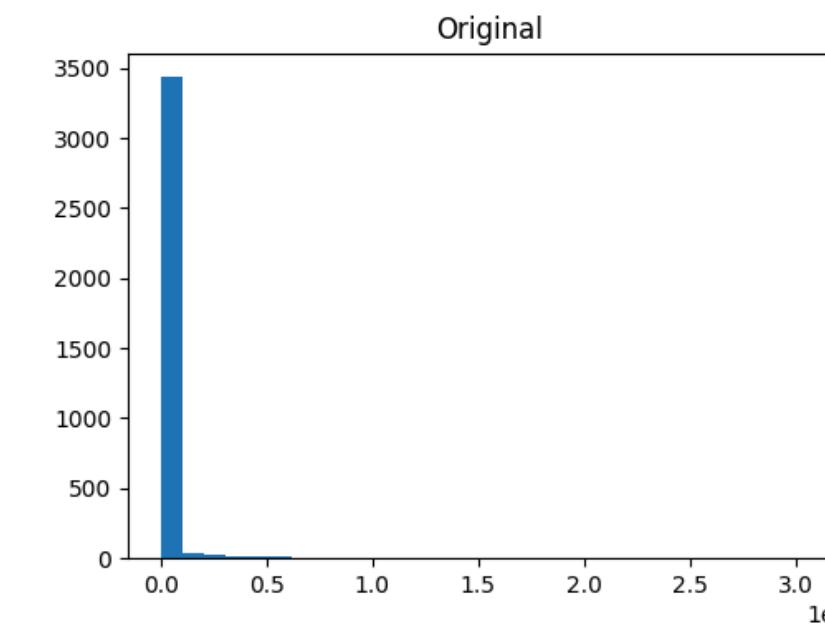
APPRENTISAGE SUPVISE

REGRESSION

Linear Regression

Random Forest Regressor

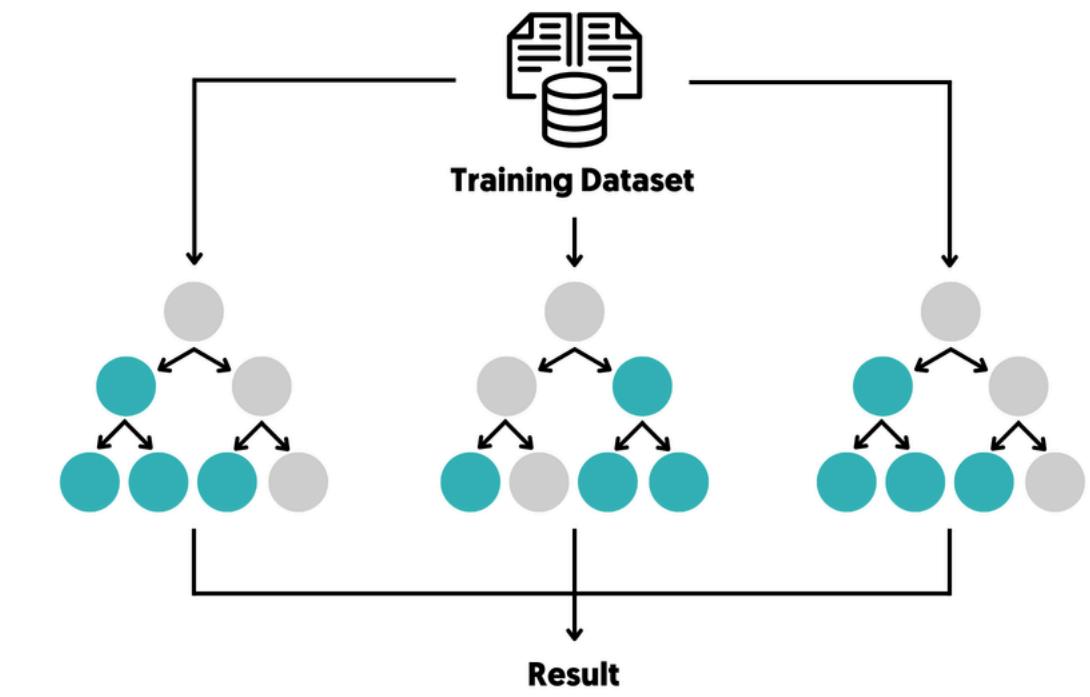
XGBoost Regressor



$$Y_i = \beta_0 + \beta_1 X_i$$

Constant/Intercept
↓
Dependent Variable
↑
Independent Variable
↓
Slope/Coefficient

This diagram illustrates the linear regression equation $Y_i = \beta_0 + \beta_1 X_i$. It shows the components of the equation: the dependent variable Y_i , the constant/intercept β_0 , the independent variable X_i , the slope/coefficient β_1 , and the labels for the variables.



MODÉLISATION(BONUS)

APPRENTISSAGE NON SUPERVISE

CLUSTERING

Kmeans

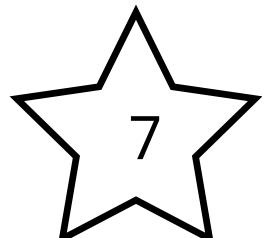
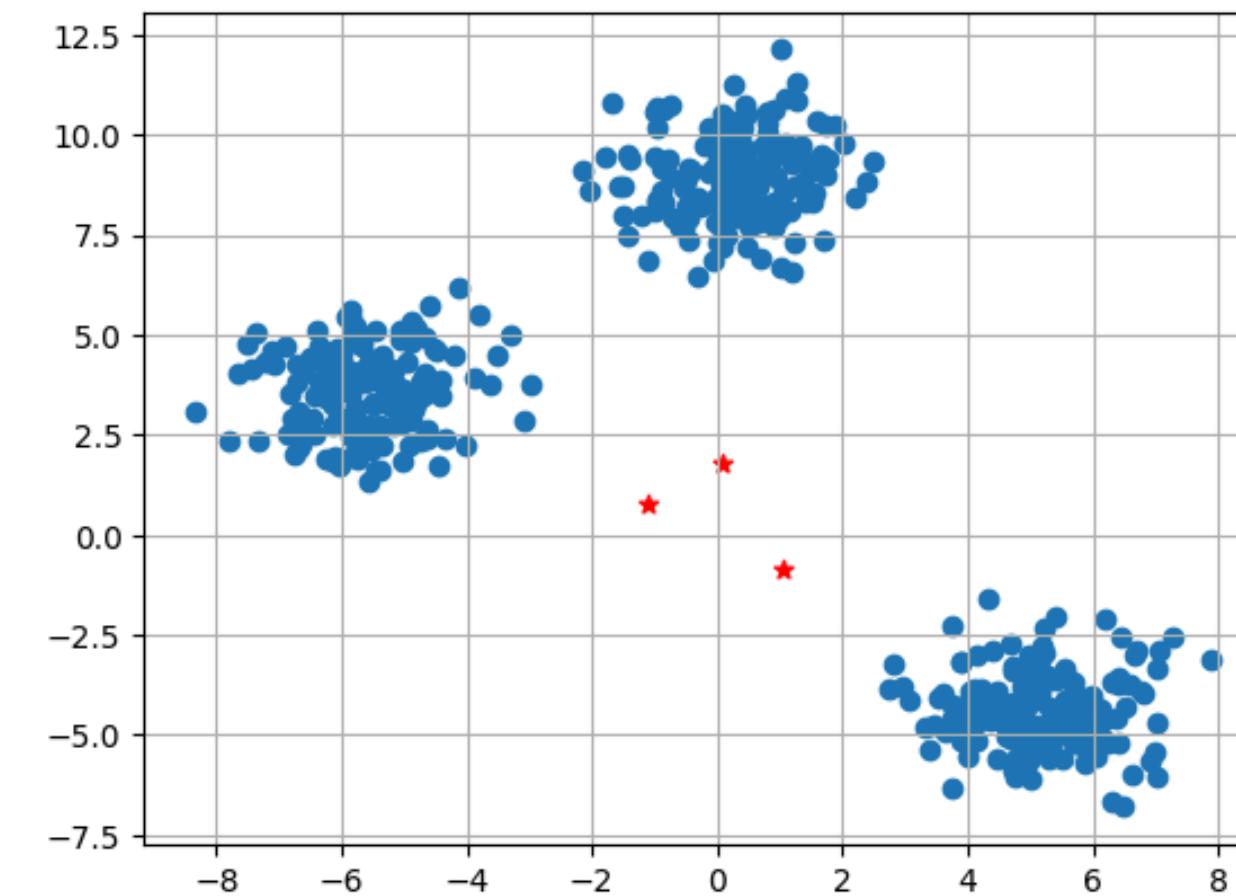
APPRENTISSAGE SUPERVISE

REGRESSION

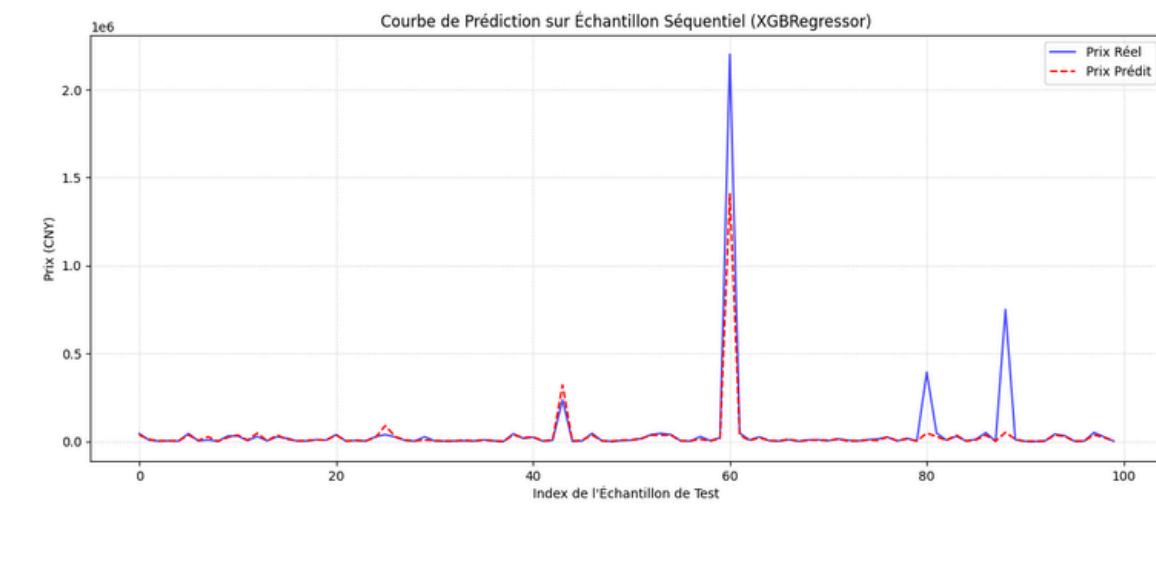
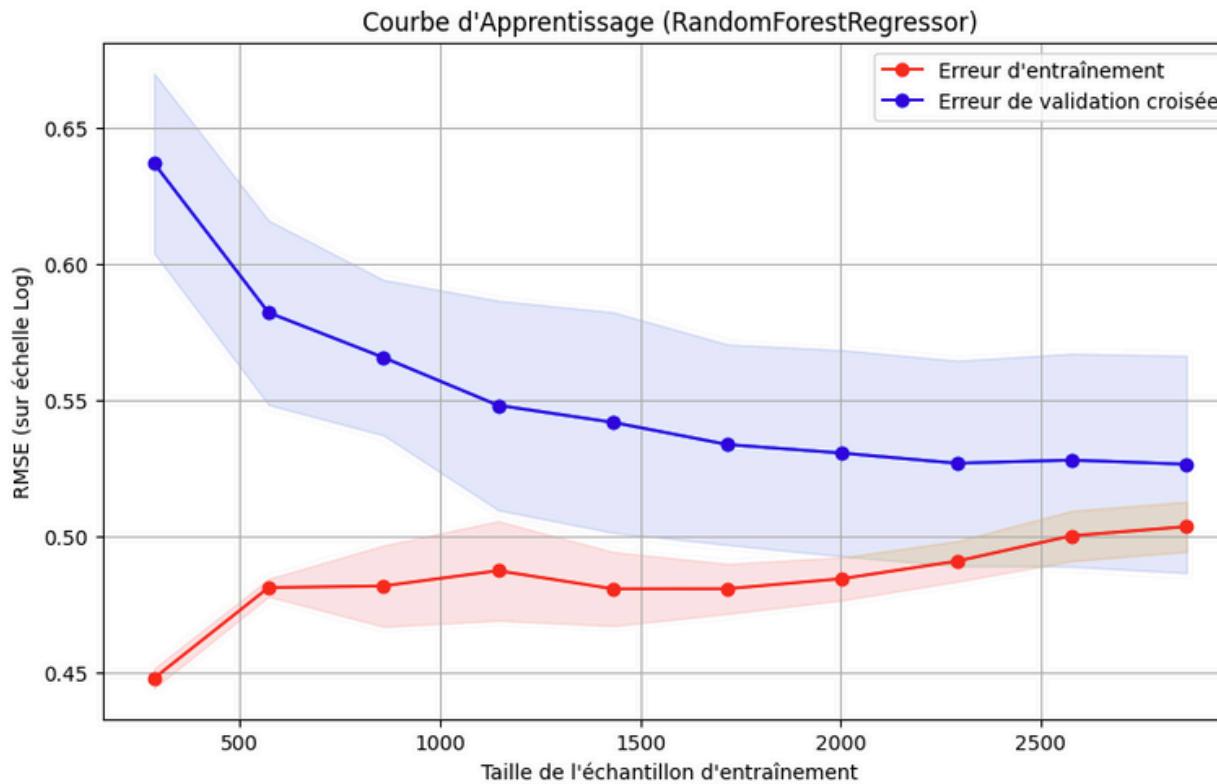
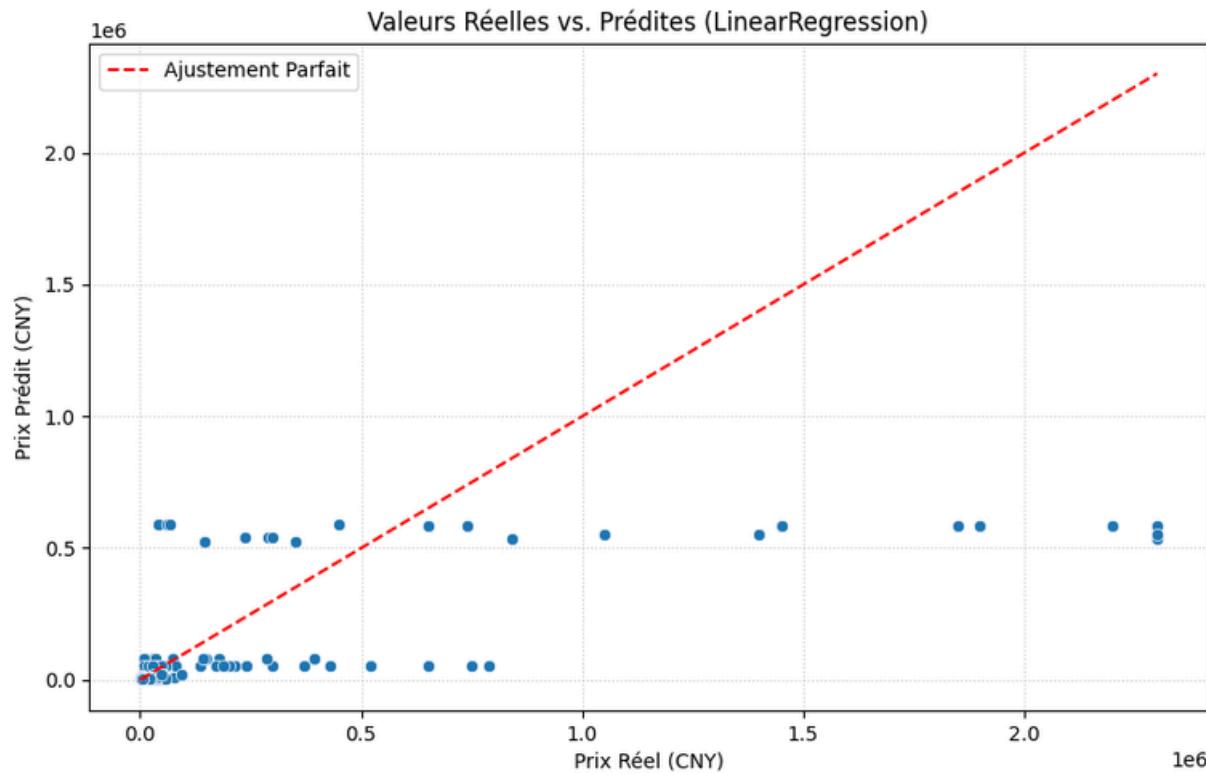
Linear Regression

Random Forest Regressor

XGBoost Regressor



EVALUATION



Modèle : LinearRegression
RMSE : 152,278.96
MAE : 31,959.76
R² : 0.4335

Modèle : RandomForestRegressor
RMSE : 90,045.92
MAE : 18,062.48
R² : 0.8019

Modèle : XGBRegressor
RMSE : 90,653.85
MAE : 18,190.06
R² : 0.7992



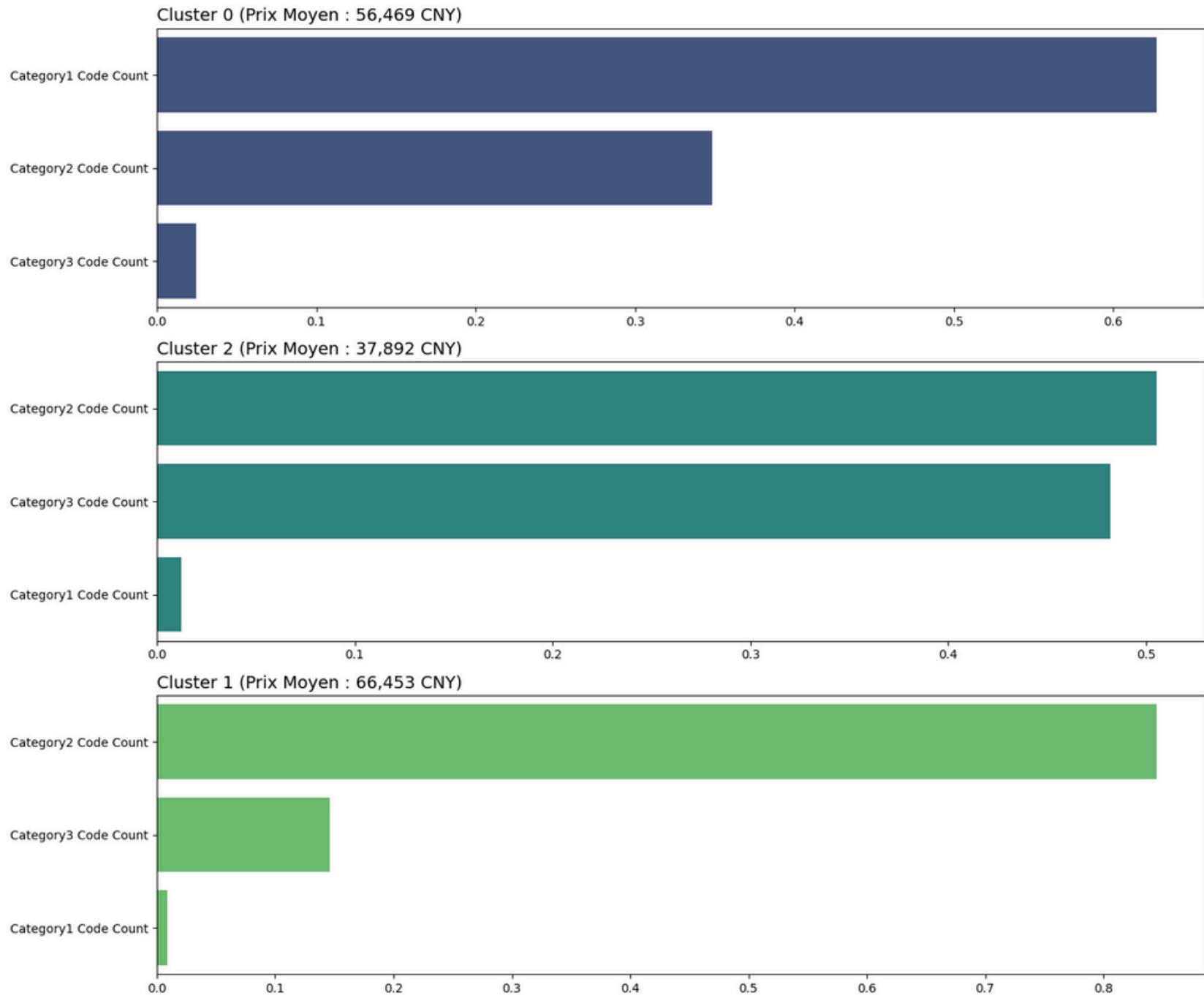
EVALUATION (BONUS)

cluster_id	Prix Moyen	Taille du Segment
0	56469.300613	815
1	66453.293173	498
2	37891.988304	1539

R^2 (Test) : 0.8109
RMSE (Test) : 87986.68
MAE (Test) : 19368.40



Top 5 des Caractéristiques les Plus Importantes par Segment de Produits



CONCLUSION

- Contrer la Fast-Fashion
- Intégrer le Guócháo
- Saisir l'Opportunité de l'Épargne

