

1) Exploration.

- * Matrice de Correlation
- * Description des stats base
- * Valeurs manquantes (NaN).
- * Evaluer Dist de prob X_i
 - Outliers (Normal -
- * Box plot.

2) Pre-traitement:

Label Encoder

OHE

- Nettoyage.
- Elimination des outliers.
- Remplacement de valeurs manquantes
- Encodage.
- Feature Engineering.

3) 2^e model -

Algo principal.

Regression lineaire.

↳

Model

Classification - DT.

↳ Calcul des métriques du modèle
itératif sur le pré-traitement.

Evaluer 2 ou 3 autres Algo secondaires

Regression : Ridge / Lasso.

Classification : NB / RF

Comparison der Modelle

Mod \ met.	R ²	RMSE	T-value
R-L			
Ridge			
Lasso			

Model \ m.f	Accuracy	Precis	Recall	F1-Score	T-value
DT					
NB					
R.F					

Interpretation

Label Encoder.

^x Marque.	Marque
Subaru.	3
Toyota.	4
Ford.	2
Bmw.	1
Alfa-Ro	0

Alfa Romeo	0
Bmw	1
Ford	2
Subaru	3
Toyota	4

OHE

^x Marque.
Subaru-
Toyota-
Ford
Toyota
Ford

Marque-Subaru	Marque-Toyota	Marque-Ford
1	0	0
0	1	0
0	0	1
0	1	0
0	0	1