



14/ 05/ 2024

# **UTILISATION D'UN MODÈLE DE RÉGRESSION LINÉAIRE POUR PRÉDIRE LE PRIX DES VOITURES D'OCCASION**

ENCADRER PAR :  
• MR SAAD BENJELLOUN

RESLISER PAR :  
• FARID ABDELAJALIL  
• AHMED TAHA LAMMRANI  
• MOHAMED BOUHMID



# INTRODUCTION

**500K**

Plus de 500 000  
voitures d'occasion  
ont été vendues au  
Maroc en 2021 !

**X 3**

Expansion du marché de la  
voiture d'occasion : trois fois  
le volume du marché du  
neuf.



Volatilité du prix

# SOMMAIRE



**1.** Présentation de la base de données

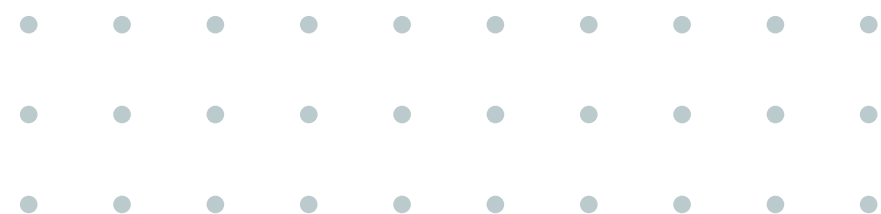
**2.** Pré-processing

**3.** Analyse des données

**4.** Les modèles

3

**5.** Conclusion



01.

# Presentation de la base de données

4



# Présentation de la base de données

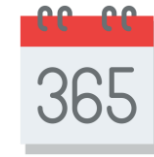


Importé neuf	Oui	Description luxury auto met en vente audi rs q3 (importée neuve) , importée neuve avec beaucoup d'options intéressantes.
Année	2023	
Boite de vitesses	Automatique	
Carburant	Essence	
Date	13-05-2024	
Nombre de portes	5	
Couleur	Noir	
Carrosserie	Crossover	
Véhicule en garantie	24 Mois	

# Présentation de la base de données

17028  
lignes

14  
colonnes



Année



Carrosserie



Date



modèle



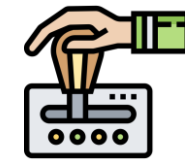
Première main



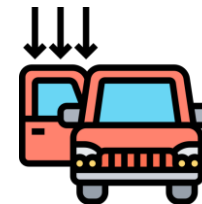
Carburant



Couleur



Boîte de  
vitesses



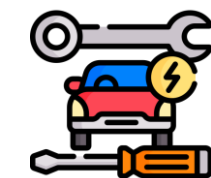
Nombre de  
portes



Prix



Puissance  
fiscale



Voiture  
personnalisée



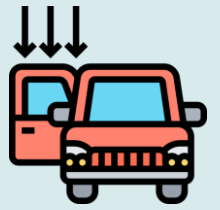
Véhicule en  
garantie



Kilométrage

# Présentation de la base de données

## Colonnes numériques



Nombre de  
portes



Prix



Kilométrage



Puissance  
fiscale

## Colonnes catégoriques



Première main



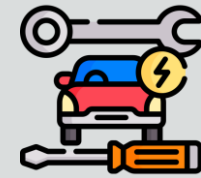
Carrosserie



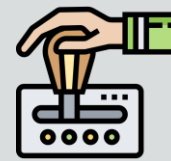
Carburant



Couleur



Voiture  
personnalisée



Boîte de  
vitesses



modèle



Véhicule en  
garantie

## Colonne temporelle



Année



Date

# Présentation de la base de données

Année	Boîte de vitesses	...	Nombre de portes	Couleur
Année\n\t\t\t\t 2022	Boîte de vitesses\n\t\t\t\t\tAut...	...	Véhicule dédouané\n\t\t\t\t\t...	NaN
Kilométrage\n\t\t\t\t\t34000	Année\n\t\t\t\t 2022	...	Puissance fiscale\n\t\t\t\t 14	Nombre de portes\n\t\t\t\t\t 5
Année\n\t\t\t\t 2022	Boîte de vitesses\n\t\t\t\t\tAut...	...	Nombre de portes\n\t\t\t\t\t 5	Première main\n\t\t\t\t\tOui
Année\n\t\t\t\t 2022	Boîte de vitesses\n\t\t\t\t\tAut...	...	Nombre de portes\n\t\t\t\t\t 5	Première main\n\t\t\t\t\tOui
Année\n\t\t\t\t 2022	Boîte de vitesses\n\t\t\t\t\tMan...	...	Nombre de portes\n\t\t\t\t\t 5	Première main\n\t\t\t\t\tOui
Année\n\t\t\t\t 2022	Boîte de vitesses\n\t\t\t\t\tMan...	...	Nombre de portes\n\t\t\t\t\t 5	Véhicule dédouané\n\t\t\t\t\t...
Kilométrage\n\t\t\t\t\t39000	Année\n\t\t\t\t 2022	...	Puissance fiscale\n\t\t\t\t 12	Nombre de portes\n\t\t\t\t\t 5





# Présentation de la base de données

Kilométrage	84 000
Année	2020
Boite de vitesses	Automatique
Carburant	Diesel
Date	13-05-2024
Puissance fiscale	6
Nombre de portes	5
Première main	Oui
Véhicule dédouané	2020

Kilométrage	118 300
Année	2020
Boite de vitesses	Automatique
Carburant	Diesel
Date	14-05-2024
Puissance fiscale	12
Nombre de portes	5
Couleur	Gris
Carrosserie	Suv et 4x4
Première main	Oui

02.

# Pré-processing

10







# Pré-processing

Année	Boite de vitesses	...	Nombre de portes	Couleur
Année\n\t\t\t\t 2022	Boite de vitesses\n\t\t\t\t\tAut...	'''	Véhicule dédouané\n\t\t\t\t\t...	NaN
Kilométrage\n\t\t\t\t\t34000	Année\n\t\t\t\t 2022	...	Puissance fiscale\n\t\t\t\t 14	Nombre de portes\n\t\t\t\t 5
Année\n\t\t\t\t 2022	Boite de vitesses\n\t\t\t\t\tAut...	...	Nombre de portes\n\t\t\t\t 5	Première main\n\t\t\t\t\tOui
Année\n\t\t\t\t 2022	Boite de vitesses\n\t\t\t\t\tAut...	...	Nombre de portes\n\t\t\t\t 5	Première main\n\t\t\t\t\tOui
Année\n\t\t\t\t 2022	Boite de vitesses\n\t\t\t\t\tMan...	...	Nombre de portes\n\t\t\t\t 5	Première main\n\t\t\t\t\tOui
Année\n\t\t\t\t 2022	Boite de vitesses\n\t\t\t\t\tMan...	...	Nombre de portes\n\t\t\t\t 5	Véhicule dédouané\n\t\t\t\t\t...
Kilométrage\n\t\t\t\t\t39000	Année\n\t\t\t\t 2022	...	Puissance fiscale\n\t\t\t\t 12	Nombre de portes\n\t\t\t\t 5

type	object
price	object
Kilométrage	object
Année	object
Boite de vitesses	object
Carburant	object
Date	object
Puissance fiscale	object
Nombre de portes	object
Couleur	object
Carrosserie	object
Véhicule en garantie	object
Voiture personnalisée (tuning)	object
Première main	object

# Pré-processing

---

Parcourir chaque ligne et colonne pour permuter les cellules.

---

Supprimer les préfixes et les espaces supplémentaires.

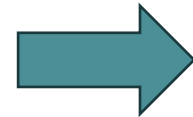
---

Convertir les données aux types appropriés.



Date

= **Date de publication du poste**

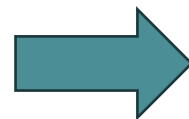


**Mois et l'année de publication du poste**



Année

= **Année de sortie**



**Age = 2025 – Année**



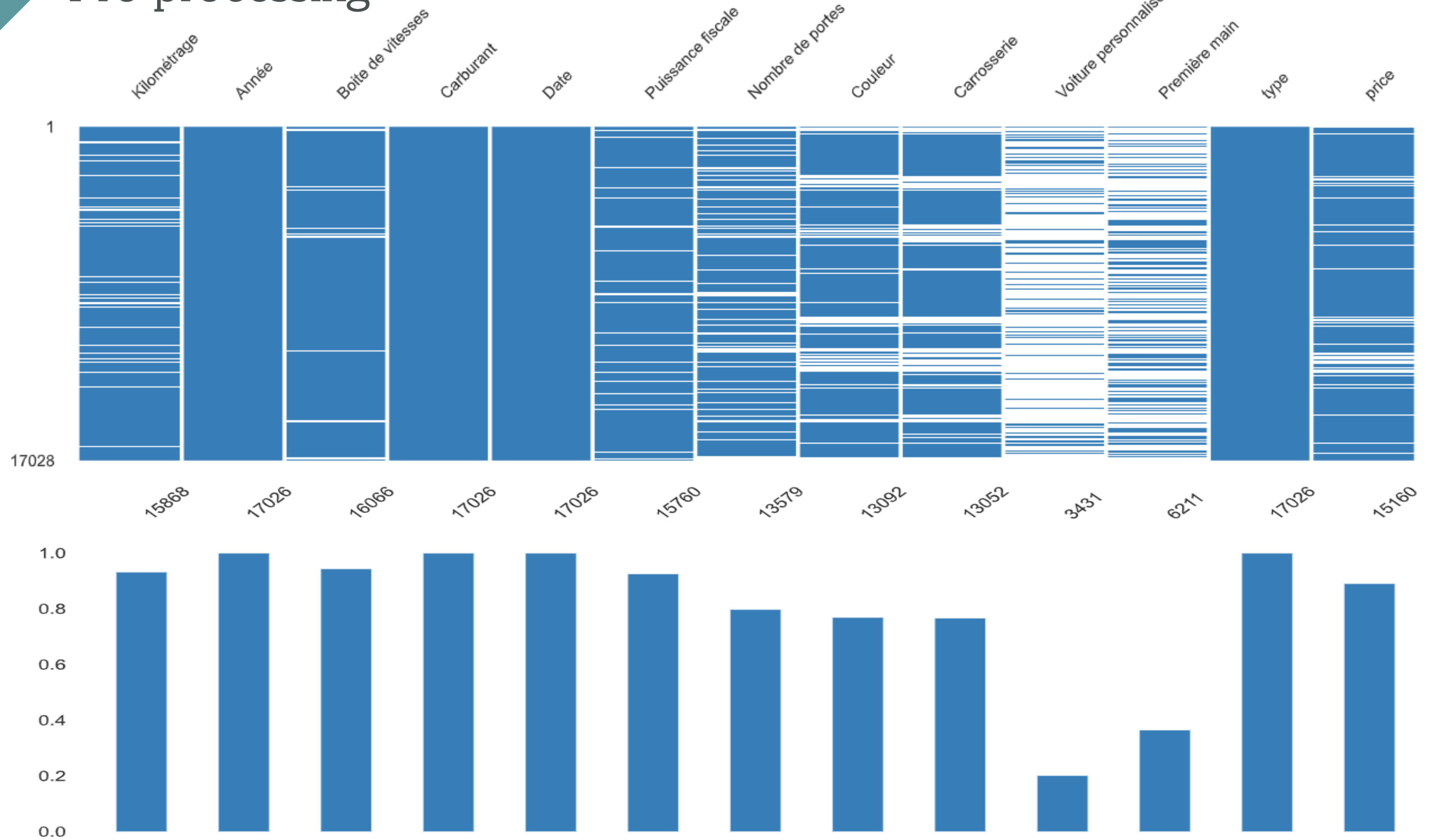


# Pré-processing

Boîte de vitesses	Carburant	Date	Puissance fiscale	Nombre de portes	Couleur	Carrosserie	car_model	price	car_age
Automatique	Diesel	2024-04-21	6	5	Noir	Citadine	PEUGEOT 208	195000	3
Automatique	Diesel	2024-04-21	12	5	Noir	Suv et 4x4	VOLKSWAGEN Touareg	145000	12
Manuelle	Diesel	2024-04-08	6	5	Bleu	Coupé	DACIA Sandero	150000	3
Automatique	Diesel	2024-04-08	8	3	Noir	Coupé	BMW Serie 2	320000	5
Automatique	Diesel	2024-04-08	11	4	Autre	Suv et 4x4	JEEP Wrangler	310000	7

Kilométrage	float64
Année	float64
Boîte de vitesses	object
Carburant	object
Date	datetime64[ns]
Puissance fiscale	float64
Nombre de portes	float64
Couleur	object
Carrosserie	object
Voiture personnalisée (tuning)	object
Première main	object
type	object
price	float64
dtype:	object

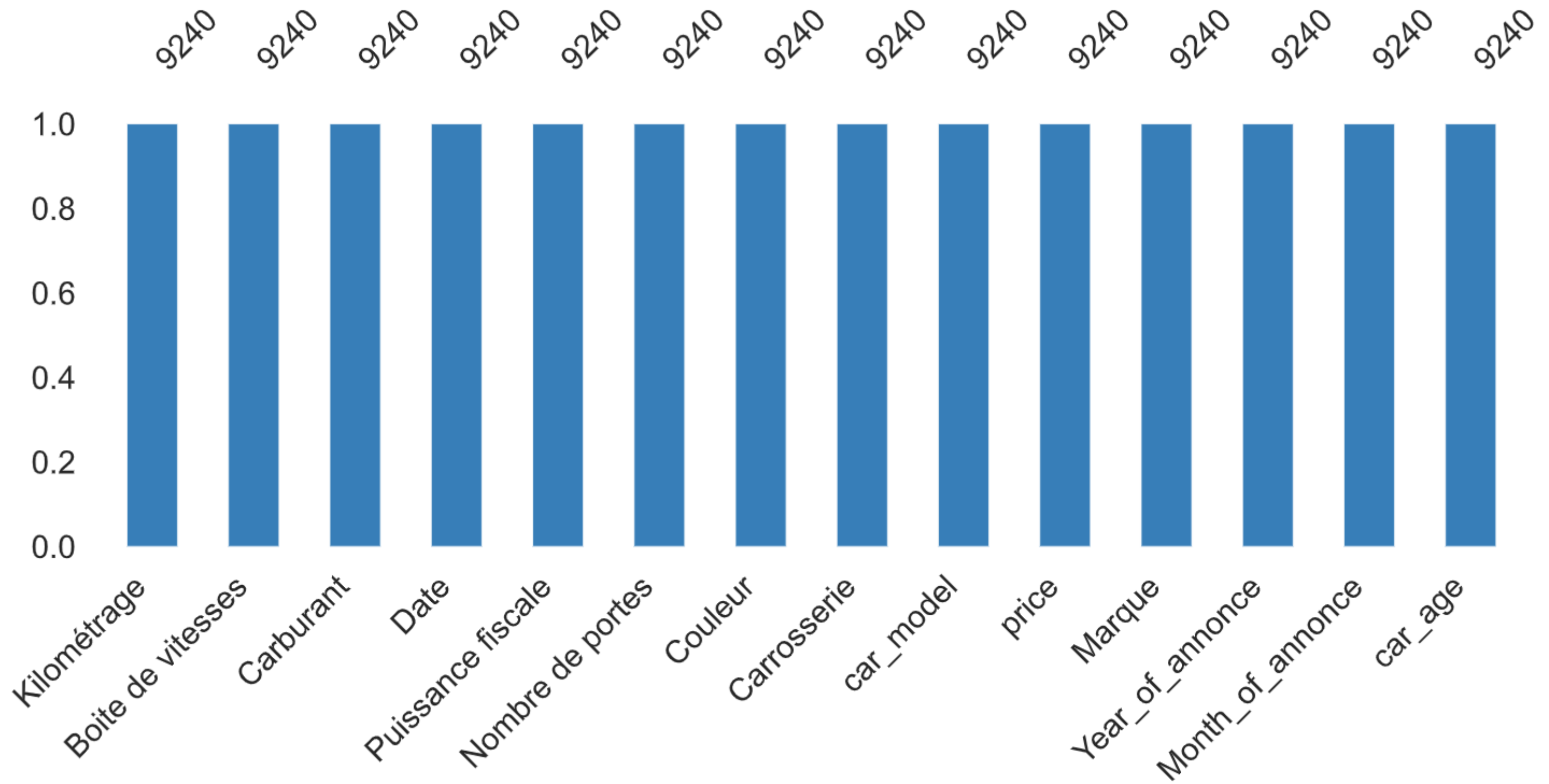
# Pré-processing







## Pré-processing





## Pré-processing

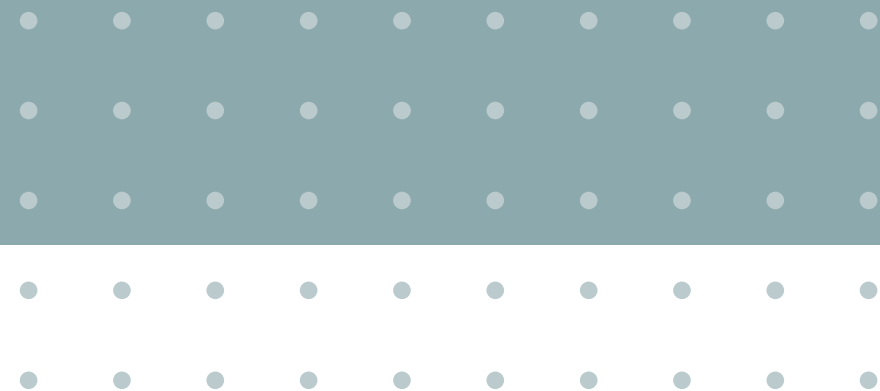
**17028**

**9240**

**45% de  
réduction**

03.

# Analyse des données





# Analyse Exploratoire des Données (EDA)

Vue d'ensemble des différentes variables et de leurs caractéristiques:

	Kilométrage	Date	Puissance fiscale	Nombre de portes	price	car_age
count	9.240000e+03	9240	9240.000000	9240.000000	9.240000e+03	9240.000000
mean	9.412440e+05	2023-01-08 00:41:36.623376640	7.673701	4.798052	1.588810e+05	12.019481
min	1.000000e+00	2021-04-21 00:00:00	2.000000	2.000000	1.000000e+00	1.000000
25%	8.900000e+04	2022-07-24 00:00:00	6.000000	5.000000	8.100000e+04	8.000000
50%	1.450000e+05	2022-12-15 12:00:00	8.000000	5.000000	1.250000e+05	11.000000
75%	2.000000e+05	2023-08-02 00:00:00	8.000000	5.000000	1.900000e+05	16.000000
max	2.147484e+09	2024-04-21 00:00:00	60.000000	10.000000	2.700000e+06	85.000000
std	3.888787e+07	NaN	2.567905	0.597027	1.342143e+05	6.159939

Suppression des  
outliers à 5 %

	Kilométrage	Date	Puissance fiscale	Nombre de portes	price	car_age
count	7574.000000	7574	7574.000000	7574.000000	7574.000000	7574.000000
mean	148257.185899	2022-12-30 21:11:44.409822976	7.516768	4.807367	144137.347769	11.577106
min	24000.000000	2021-04-21 00:00:00	2.000000	2.000000	46000.000000	2.000000
25%	96000.000000	2022-07-18 00:00:00	6.000000	5.000000	87000.000000	8.000000
50%	146000.000000	2022-12-05 00:00:00	7.000000	5.000000	126000.000000	11.000000
75%	196000.000000	2023-07-26 00:00:00	8.000000	5.000000	180000.000000	15.000000
max	303000.000000	2024-04-21 00:00:00	60.000000	8.000000	375000.000000	85.000000
std	67797.541783	NaN	2.315740	0.584806	73536.722148	5.046809

# Analyse Exploratoire des Données (EDA)

Vue d'ensemble des différentes variables et de leurs caractéristiques:

	Kilométrage	Date	Puissance fiscale	Nombre de portes	price	car_age
count	9.240000e+03	9240	9240.000000	9240.000000	9.240000e+03	9240.000000
mean	9.412440e+05	2023-01-08 00:41:36.623376640	7.673701	4.798052	1.588810e+05	12.019481
min	1.000000e+00	2021-04-21 00:00:00	2.000000	2.000000	1.000000e+00	1.000000
25%	8.900000e+04	2022-07-24 00:00:00	6.000000	5.000000	8.100000e+04	8.000000
50%	1.450000e+05	2022-12-15 12:00:00	8.000000	5.000000	1.250000e+05	11.000000
75%	2.000000e+05	2023-08-02 00:00:00	8.000000	5.000000	1.900000e+05	16.000000
max	2.147484e+09	2024-04-21 00:00:00	60.000000	10.000000	2.700000e+06	85.000000
std	3.888787e+07	NaN	2.567905	0.597027	1.342143e+05	6.159939

Réduction de la  
base de données  
par 18%

	Kilométrage	Date	Puissance fiscale	Nombre de portes	price	car_age
count	7574.000000	7574	7574.000000	7574.000000	7574.000000	7574.000000
mean	148257.185899	2022-12-30 21:11:44.409822976	7.516768	4.807367	144137.347769	11.577106
min	24000.000000	2021-04-21 00:00:00	2.000000	2.000000	46000.000000	2.000000
25%	96000.000000	2022-07-18 00:00:00	6.000000	5.000000	87000.000000	8.000000
50%	146000.000000	2022-12-05 00:00:00	7.000000	5.000000	126000.000000	11.000000
75%	196000.000000	2023-07-26 00:00:00	8.000000	5.000000	180000.000000	15.000000
max	303000.000000	2024-04-21 00:00:00	60.000000	8.000000	375000.000000	85.000000
std	67797.541783	NaN	2.315740	0.584806	73536.722148	5.046809

# Analyse Exploratoire des Données (EDA)

Vue d'ensemble des différentes variables et de leurs caractéristiques:

Application du filtre :  
Prix > 30k MAD  
Kilométrage < 1M km

	Kilométrage	Date	Puissance fiscale	Nombre de portes	price	car_age
count	9.240000e+03	9240	9240.000000	9240.000000	9.240000e+03	9240.000000
mean	9.412440e+05	2023-01-08 00:41:36.623376640	7.673701	4.798052	1.588810e+05	12.019481
min	1.000000e+00	2021-04-21 00:00:00	2.000000	2.000000	1.000000e+00	1.000000
25%	8.900000e+04	2022-07-24 00:00:00	6.000000	5.000000	8.100000e+04	8.000000
50%	1.450000e+05	2022-12-15 12:00:00	8.000000	5.000000	1.250000e+05	11.000000
75%	2.000000e+05	2023-08-02 00:00:00	8.000000	5.000000	1.900000e+05	16.000000
max	2.147484e+09	2024-04-21 00:00:00	60.000000	10.000000	2.700000e+06	85.000000
std	3.888787e+07	NaN	2.567905	0.597027	1.342143e+05	6.159939

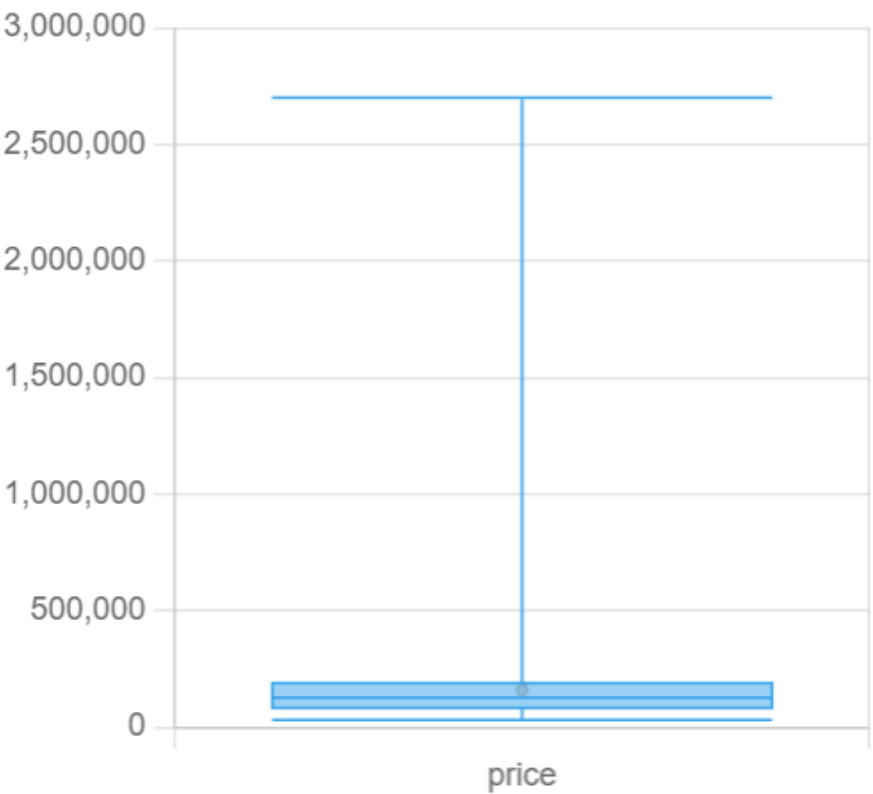
	Kilométrage	Année	Date	Puissance fiscale	Nombre de portes	price	car_age
count	9113.000000	7484.000000	9113	9113.000000	9113.000000	9.113000e+03	9113.000000
mean	152549.454954	2013.420230	2023-01-05 07:09:19.771754752	7.677933	4.798749	1.601954e+05	11.88489
min	1.000000	1940.000000	2021-04-21 00:00:00	2.000000	2.000000	3.100000e+04	1.000000
25%	88600.000000	2010.000000	2022-07-21 00:00:00	6.000000	5.000000	8.300000e+04	8.000000
50%	144000.000000	2014.000000	2022-12-12 00:00:00	8.000000	5.000000	1.260000e+05	11.000000
75%	200000.000000	2017.000000	2023-07-31 00:00:00	8.000000	5.000000	1.900000e+05	15.000000
max	999000.000000	2023.000000	2024-04-21 00:00:00	60.000000	10.000000	2.700000e+06	85.000000
std	95315.513040	5.049394	NaN	2.577289	0.597488	1.323749e+05	5.99237



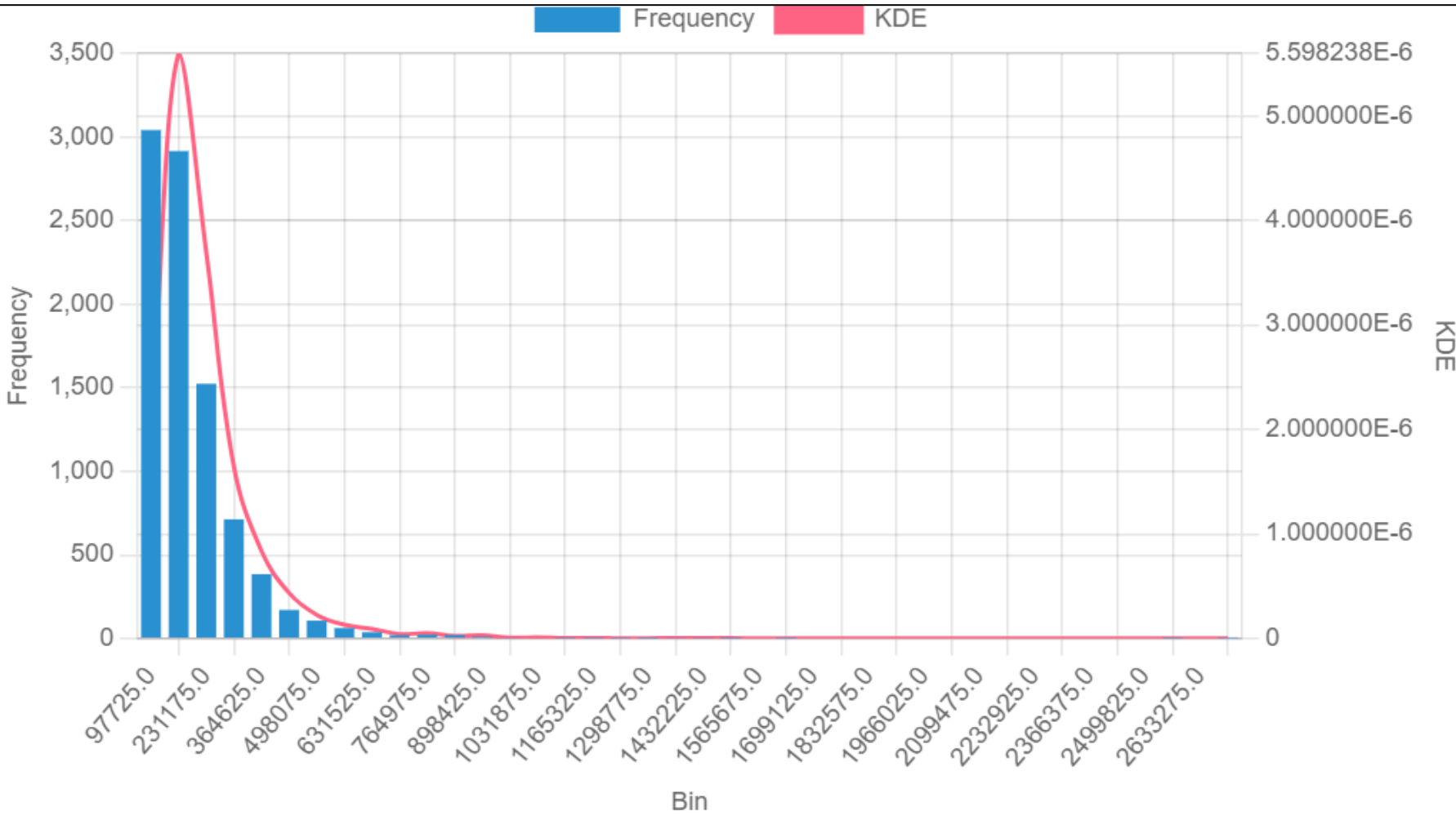
# Analyse Exploratoire des Données (EDA)

## Profiling : Prix (Price)

- **Total Rows:** 9,064
  - **Count (non-nan):** 9,064
  - **Count (missing):** 0
  - **% Missing:** 0
- **Most Frequent:** 120,000 (2.03%)
  - **Frequency:** 184
- **25%:** 83,000
- **50%:** 127,000
- **75%:** 190,000
- **max:** 2,700,000
- **mean:** 160,439.2577
- **median:** 127,000
- **min:** 31,000
- **mode:** 120,000
- **sem:** 1,392.3266
- **std:** 132,556.5166
- **sum:** 1,454,221,432
- **Unique:** 472
- **var:** 17,571,230,080.5135



Description détaillée de la colonne "price"

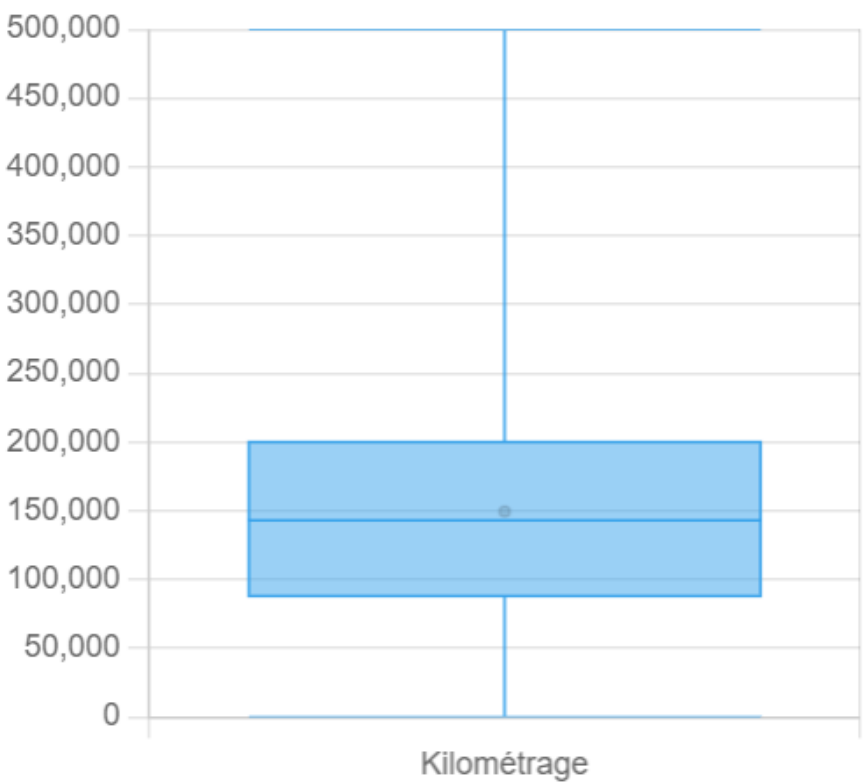


Histogramme de la colonne "price"

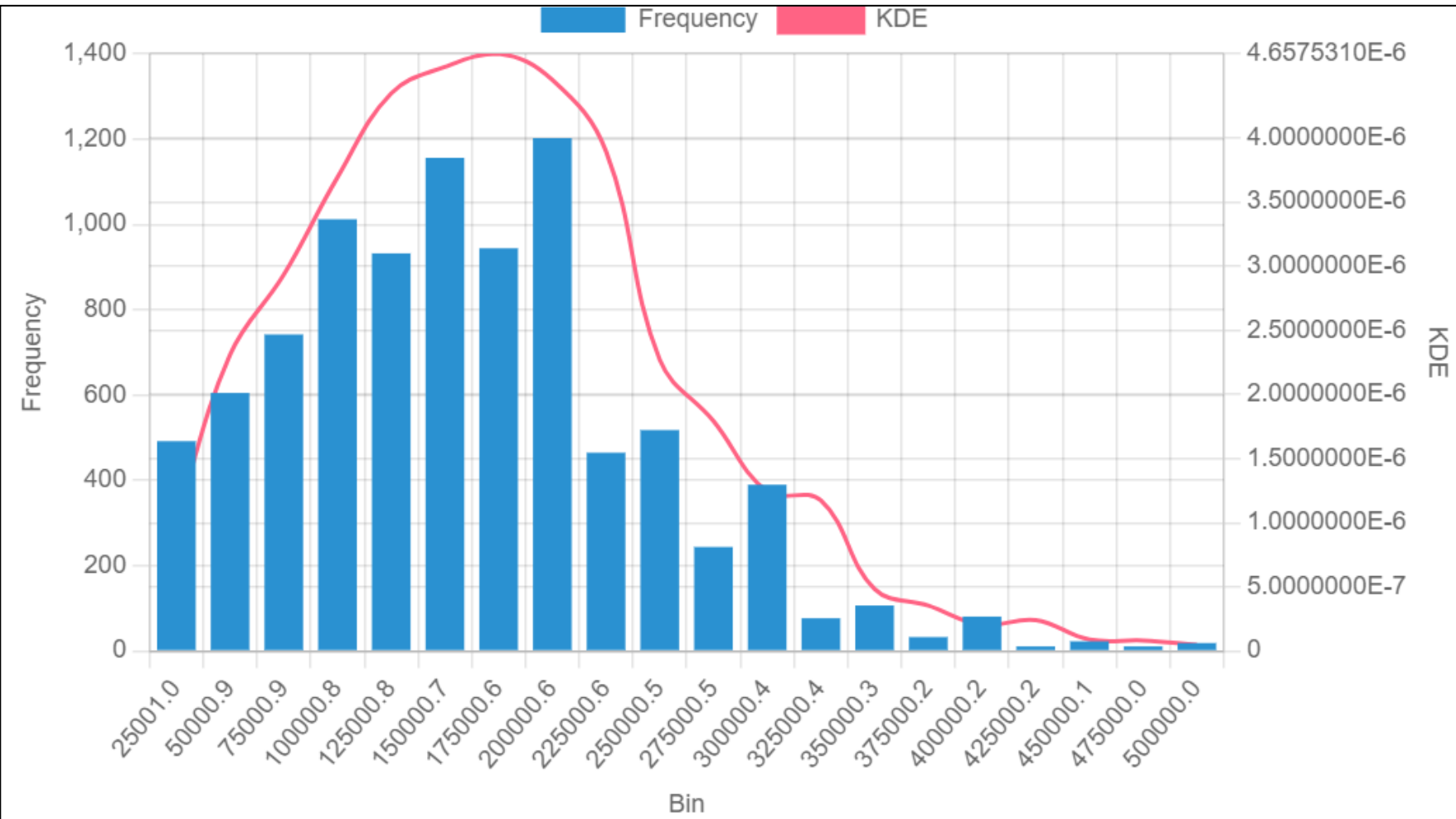
# Analyse Exploratoire des Données (EDA)

## Profiling : Kilométrage (Kilometrage)

- **Total Rows:** 9,064
  - **Count (non-nan):** 9,064
  - **Count (missing):** 0
  - **% Missing:** 0
- **Most Frequent:** 200,000 (3.76%)
  - **Frequency:** 341
- **25%:** 88,000
- **50%:** 143,000
- **75%:** 200,000
- **max:** 500,000
- **mean:** 149,460.6465
- **median:** 143,000
- **min:** 1
- **mode:** 200,000
- **sem:** 893.2395
- **std:** 85,040.9062
- **sum:** 1,354,711,300
- **Unique:** 1,231
- **var:** 7,231,955,718.944



Description détaillée de la colonne "kilometrage"

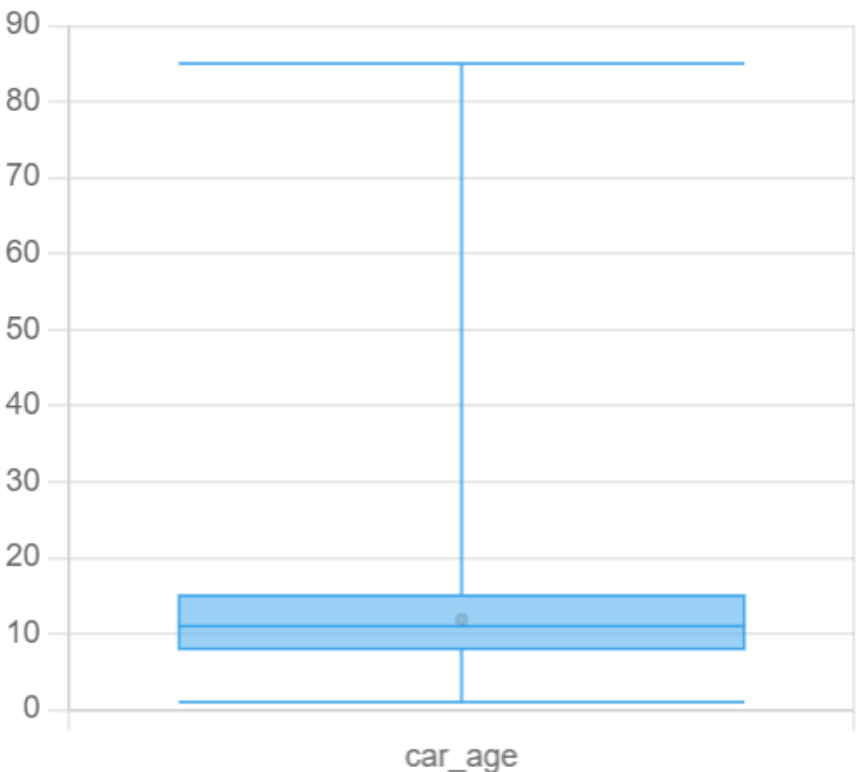


Histogramme de la colonne "kilometrage"

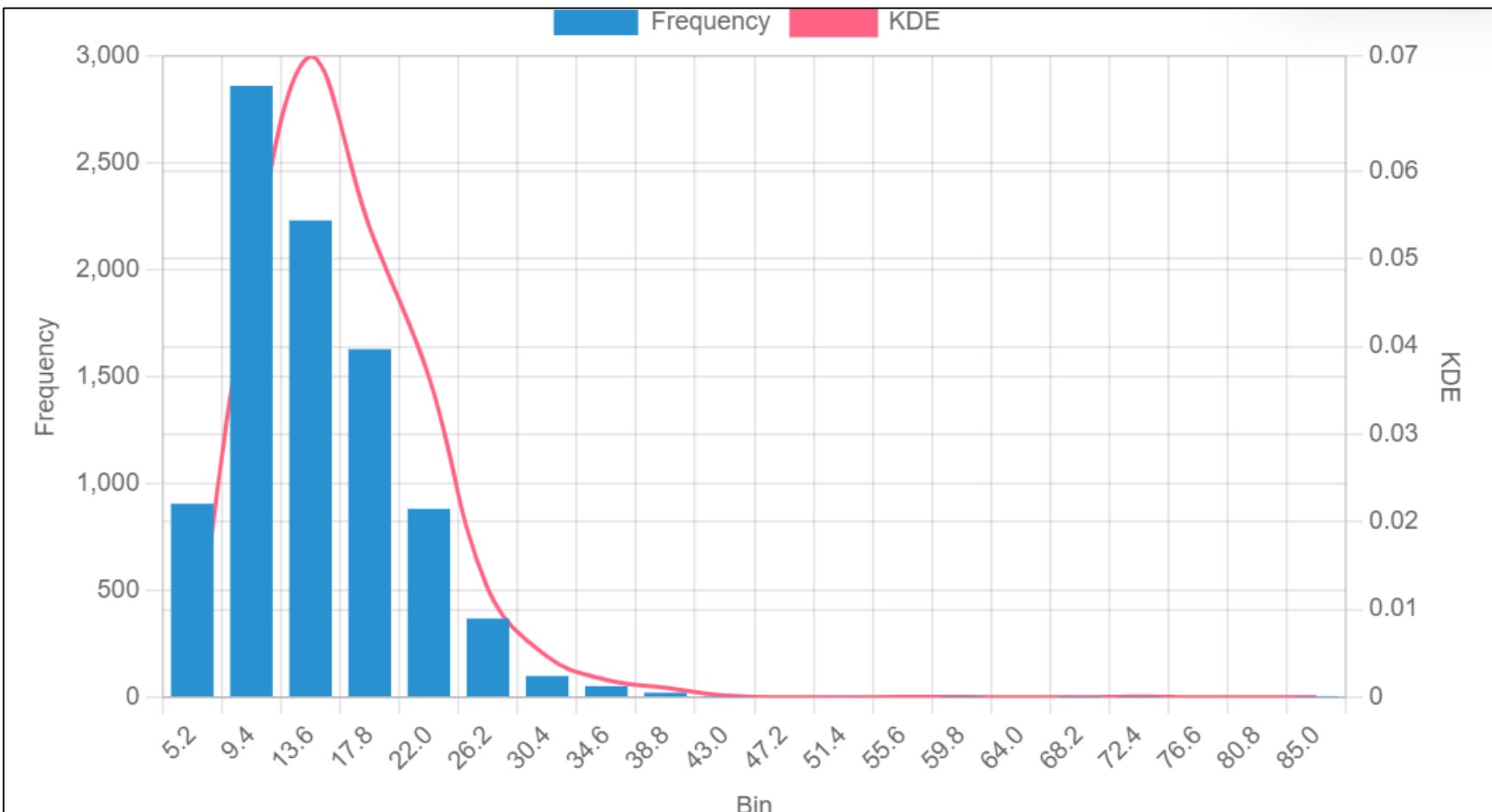
# Analyse Exploratoire des Données (EDA)

## Profiling : Âge du véhicule (Car Age)

- **Total Rows:** 9,064
  - **Count (non-nan):** 9,064
  - **Count (missing):** 0
  - **% Missing:** 0
- **Most Frequent:** 8 (9.05%)
  - **Frequency:** 820
- **25%:** 8
- **50%:** 11
- **75%:** 15
- **max:** 85
- **mean:** 11.8654
- **median:** 11
- **min:** 1
- **mode:** 8
- **sem:** 0.0625
- **std:** 5.9504
- **sum:** 107,548
- **Unique:** 45
- **var:** 35.4077



Description détaillée de la colonne "car\_age"



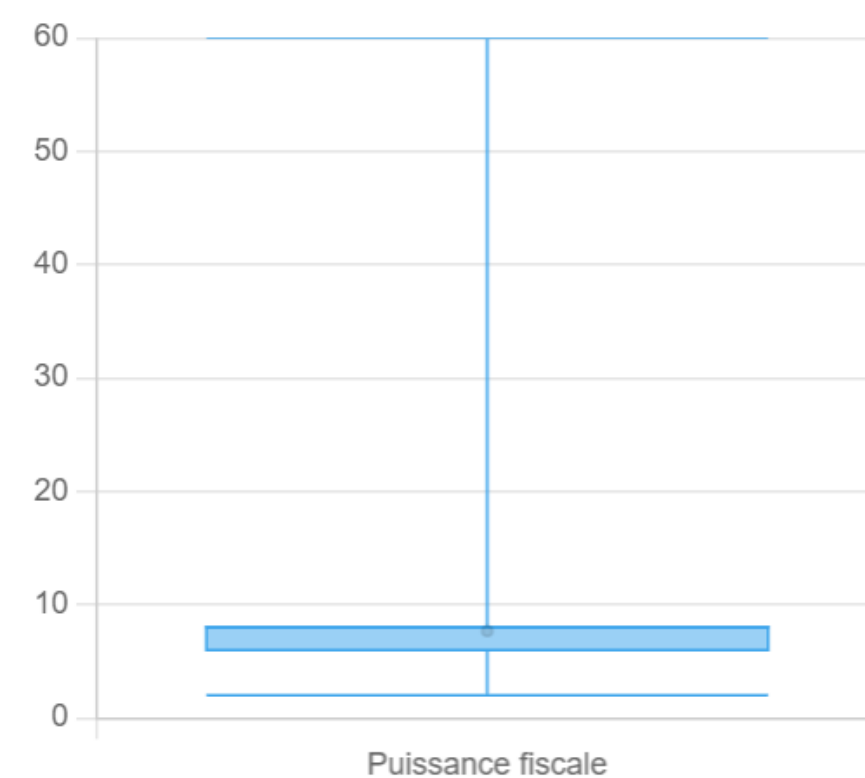
Histogramme de la colonne "car\_age"



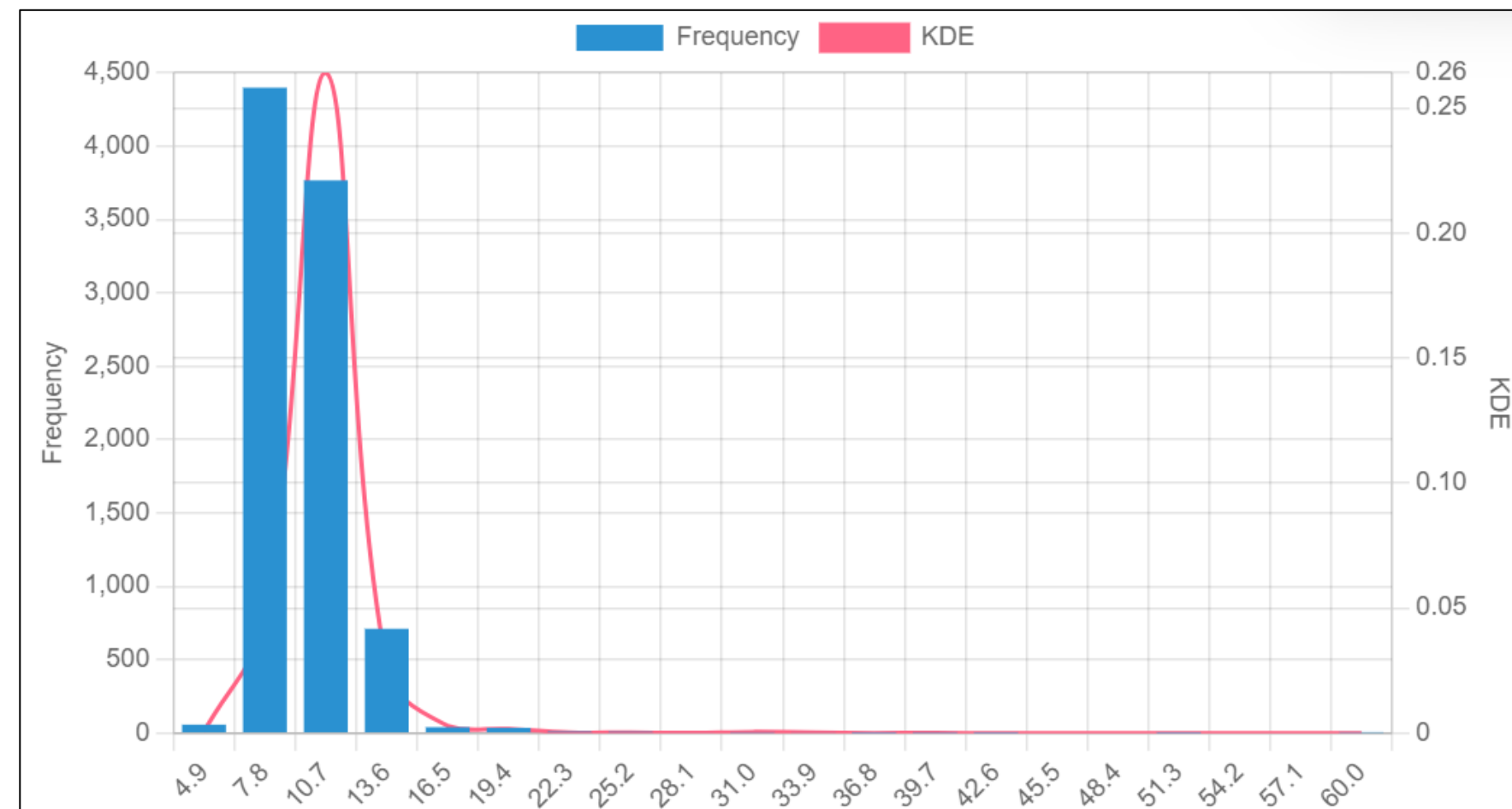
# Analyse Exploratoire des Données (EDA)

## Profiling : Puissance fiscale (Puissance fiscale)

- **Total Rows:** 9,064
  - **Count (non-nan):** 9,064
  - **Count (missing):** 0
  - **% Missing:** 0
- **Most Frequent:** 6 (35.85%)
  - **Frequency:** 3249
- **25%:** 6
- **50%:** 8
- **75%:** 8
- **max:** 60
- **mean:** 7.6796
- **median:** 8
- **min:** 2
- **mode:** 6
- **sem:** 0.0271
- **std:** 2.5814
- **sum:** 69,608
- **Unique:** 36
- **var:** 6.6638



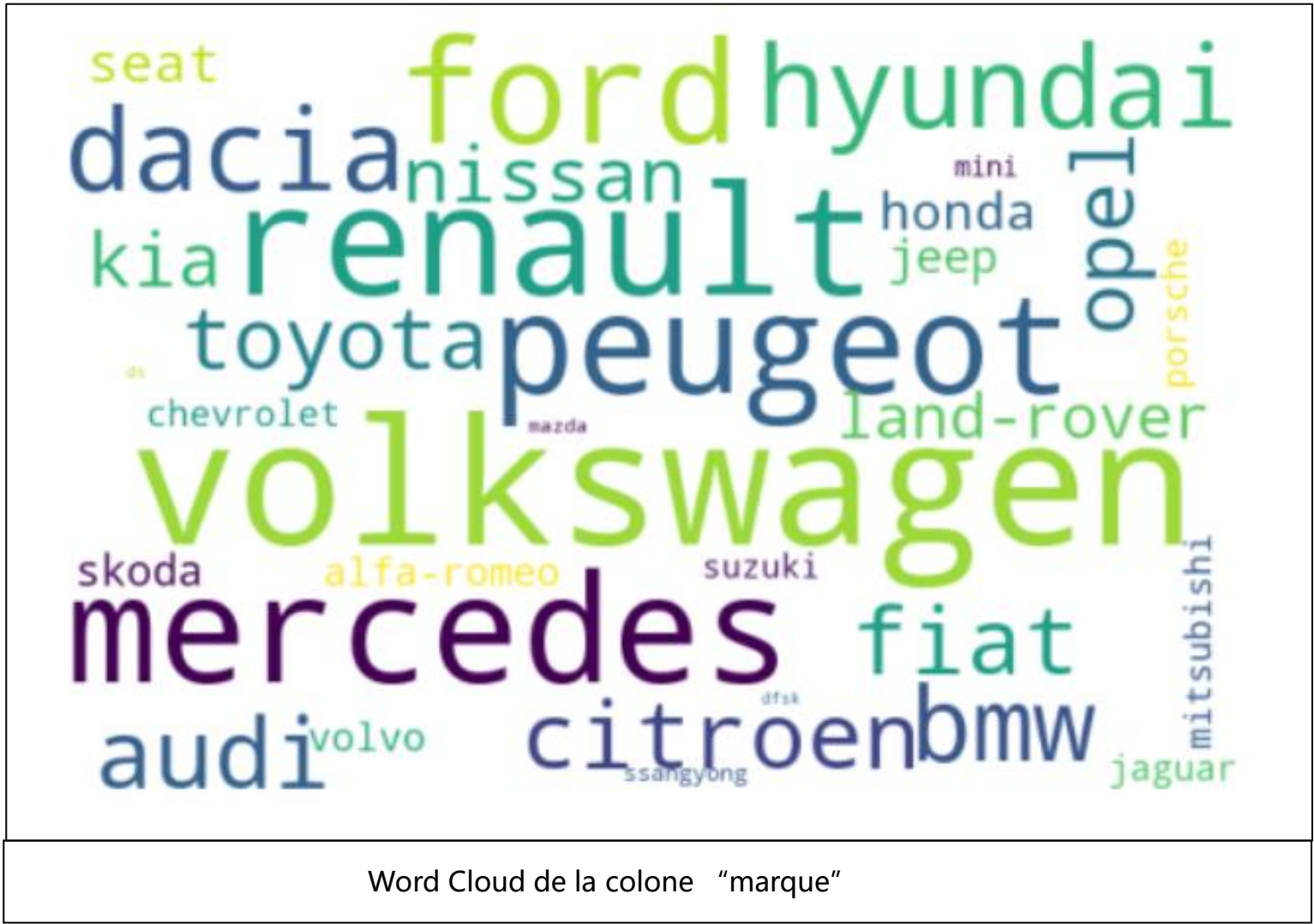
Description détaillée de la colonne "Puissance fiscale"



Histogramme de la colonne "Puissance fiscale"

# Analyse Exploratoire des Données (EDA)

Profiling : Marque (Marque)

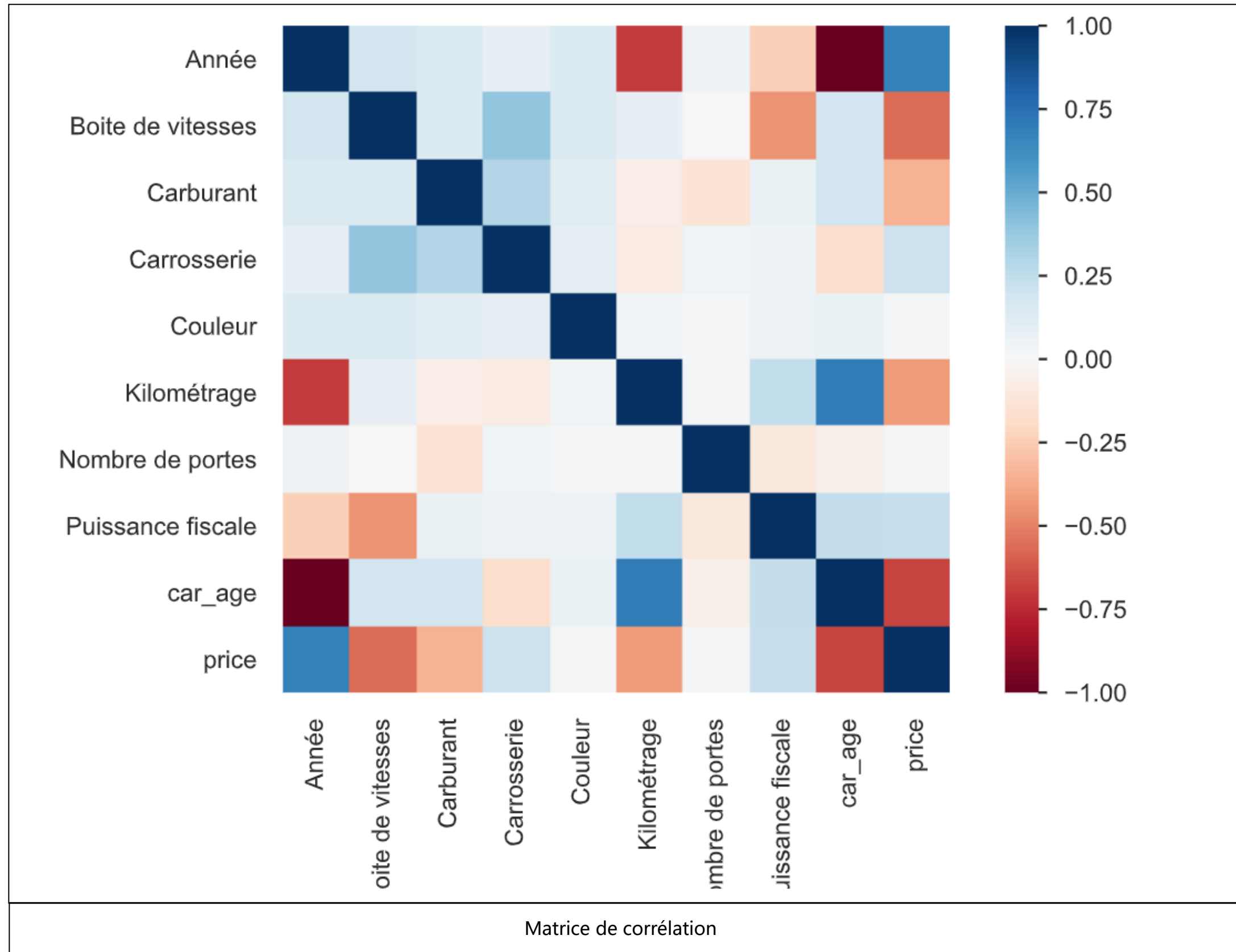


Value	Count	Frequency (%)
volkswagen	1128	12.4%
renault	809	8.9%
mercedes	754	8.3%
ford	655	7.2%
peugeot	633	7.0%
hyundai	542	6.0%
dacia	533	5.9%
bmw	432	4.8%
audi	387	4.3%
citroen	355	3.9%
Other values (52)	2836	31.3%

Tableau des fréquences de la colonne "marque"

# Analyse Exploratoire des Données (EDA)

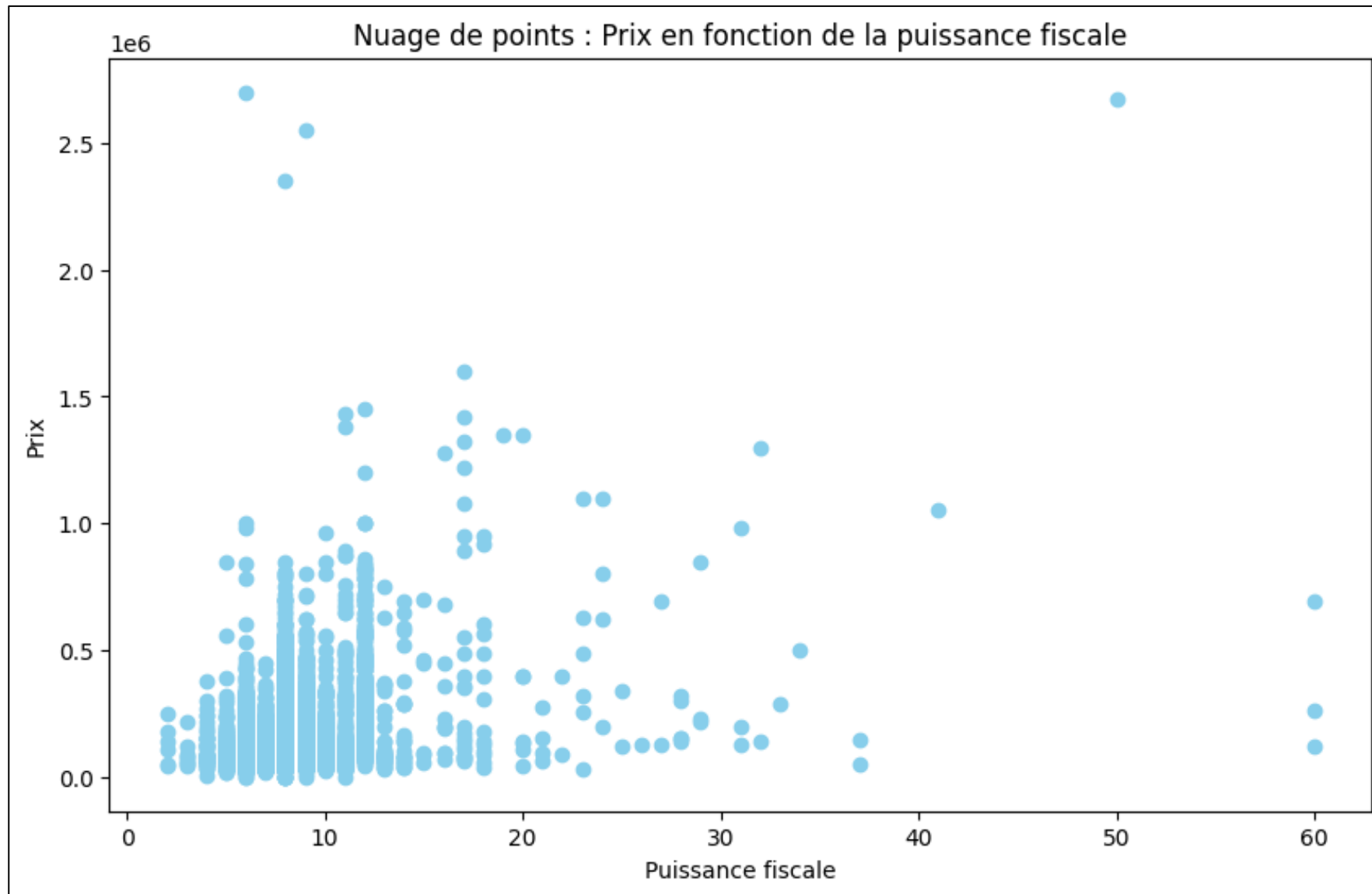
## Corrélations



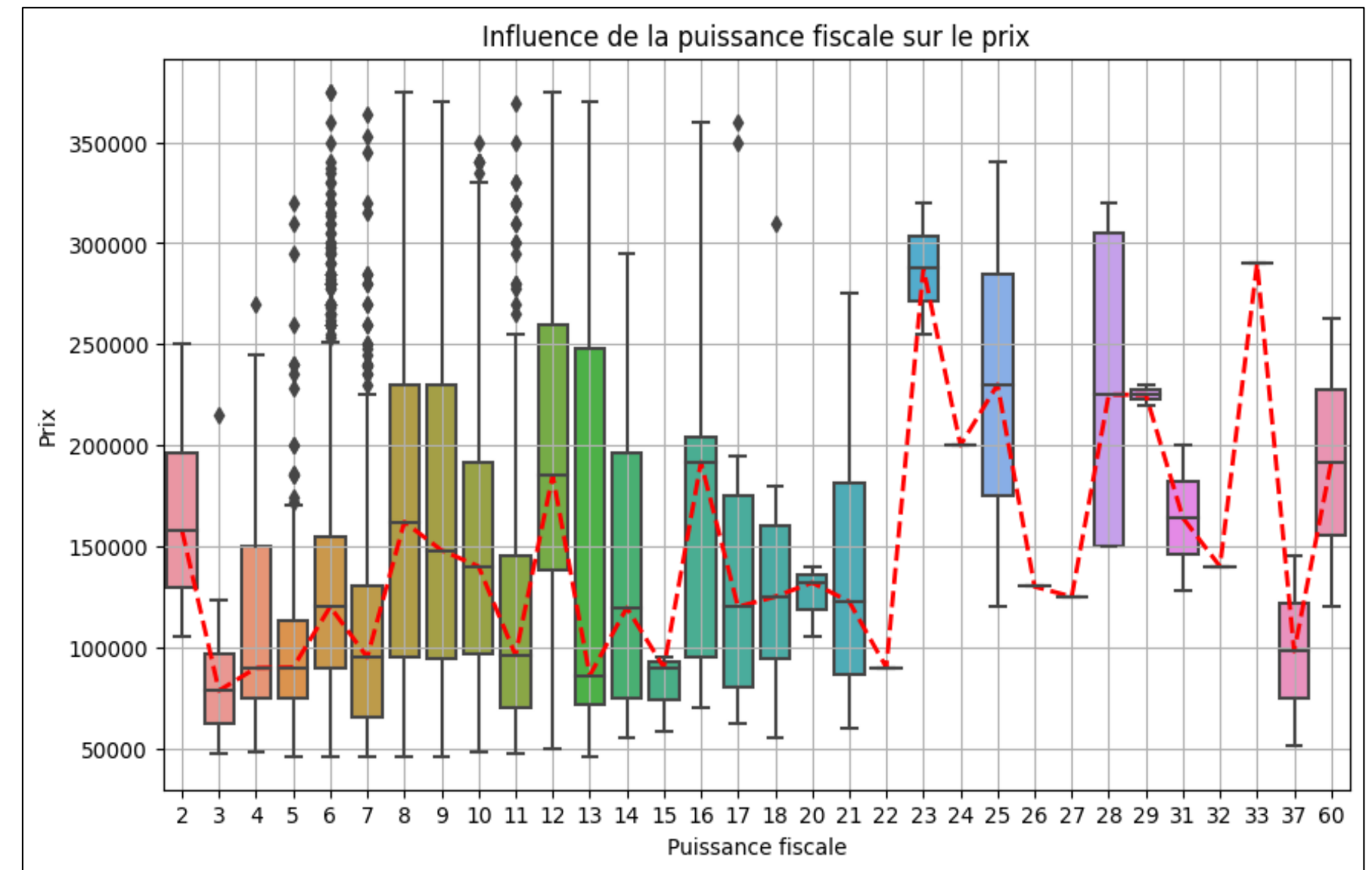


# Analyse Exploratoire des Données (EDA)

## Profiling : Puissance fiscale (Puissance fiscale)



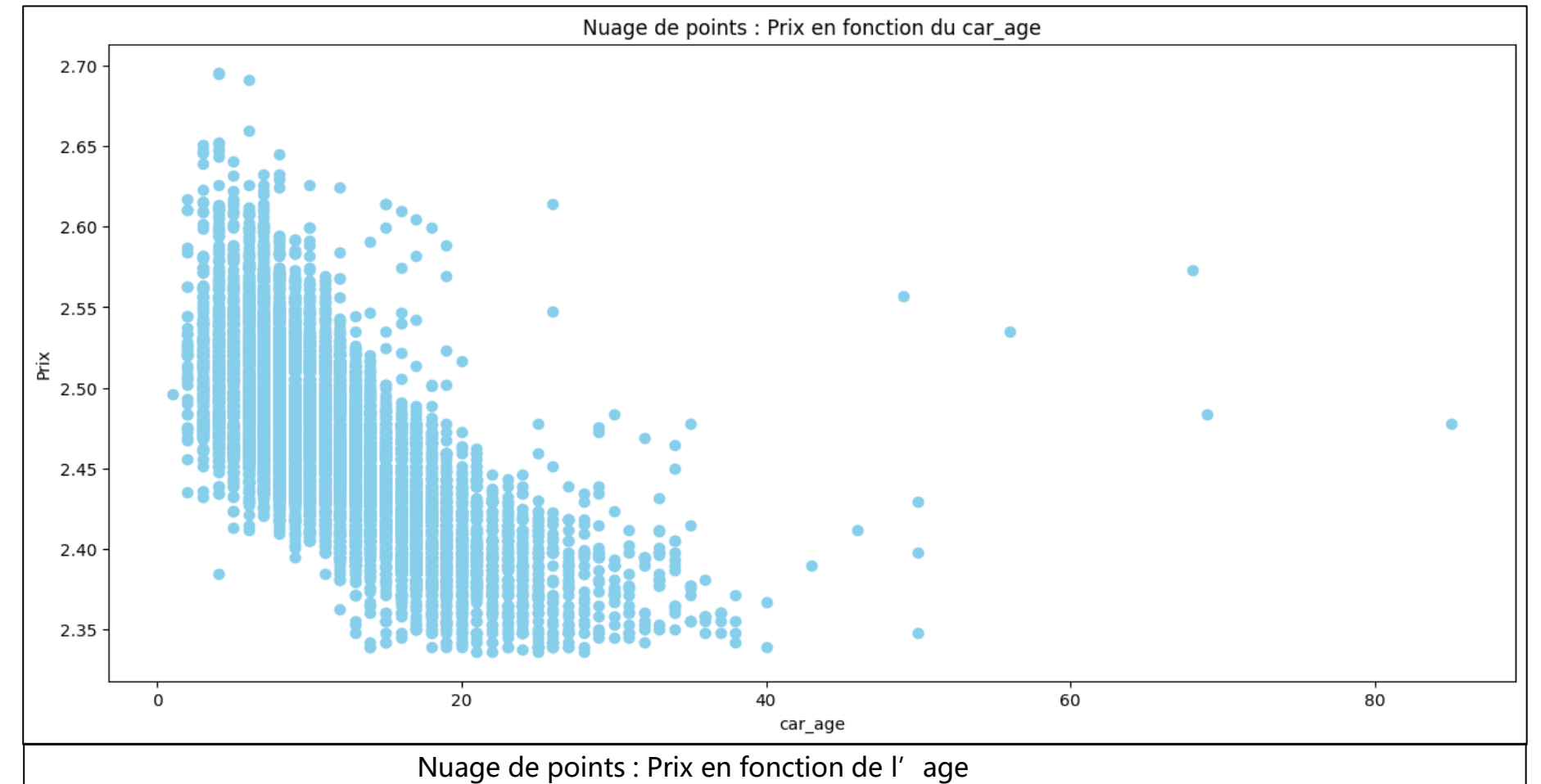
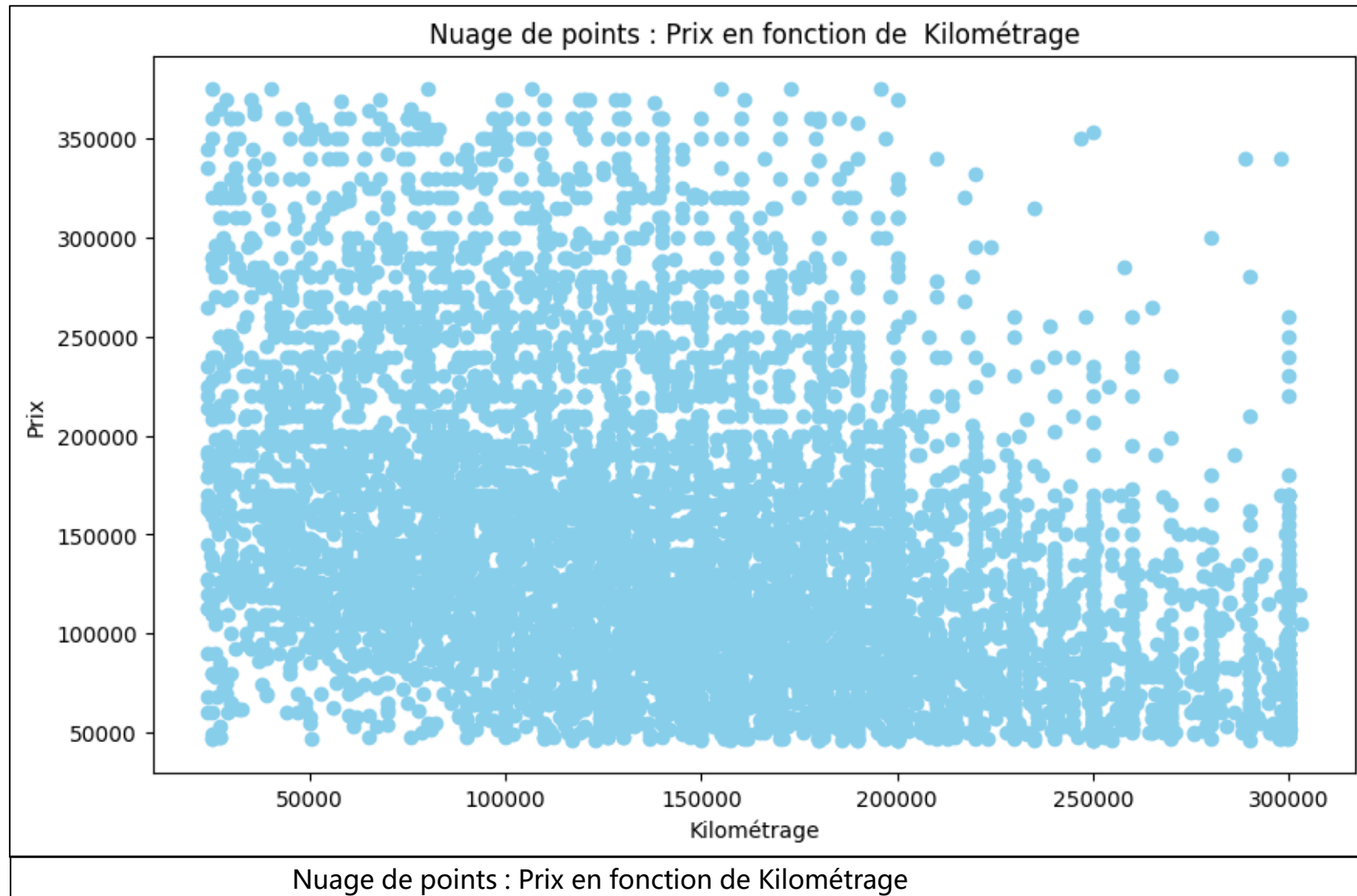
Nuage de points : Prix en fonction de la puissance fiscale



Box plot: Prix en fonction de la puissance fiscale

# Analyse Exploratoire des Données (EDA)

## Corrélations



04.

# Les modèles



# Les modèles

Régression linéaire Multilinéaire:

Les variables indépendantes:

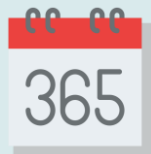
## Colonnes numériques



Kilométrage



Puissance fiscale



Âge

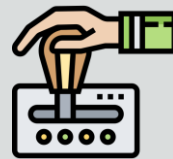
## Colonnes catégoriques



Carburant



Modèle



Boîte de vitesses



Année de l'annonce

La variable dépendante: **Prix des voitures**

Application de modèle pour Prédire:

Prix

- R-squared pour train\_set: 0.79236517831107
- R-squared pour test\_set: 0.5750159421316576

Log(prix)

- R-squared pour train\_set: 0.9011683015625314
- R-squared pour test\_set: 0.8703317984965688

Log(log(prix))

- R-squared pour train\_set: 0.9048258728676629
- R-squared pour test\_set: 0.8774805095645392



# Les modèles

Régression linéaire Multilinéaire:

Les variables indépendantes:

OLS Regression Results

=====

Dep. Variable:

price

R-squared:

0.905

Model:

OLS

Adj. R-squared:

0.897

Method:

Least Squares

F-statistic:

114.0

Date:

Tue, 14 May 2024

Prob (F-statistic):

0.00

Time:

08:28:42

Log-Likelihood:

17188.

No. Observations:

6379

AIC:

-3.339e+04

Df Residuals:

5887

BIC:

-3.007e+04

Df Model:

491

Covariance Type:

nonrobust

=====

coef

std err

t

P>|t|

[0.025

0.975]

-----

const

4.5090

0.587

7.676

0.000

3.358

5.661

car\_age

-0.0054

6.2e-05

-87.587

0.000

-0.006

-0.005

Kilométrage

-4.607e-08

3.03e-09

-15.185

0.000

-5.2e-08

-4.01e-08

Puissance fiscale

-0.0006

0.000

-3.857

0.000

-0.001

-0.000

Year\_of\_annonce

-0.0010

0.000

-3.311

0.001

-0.002

-0.000

Boite de vitesses\_Manuelle

-0.0118

0.001

-16.872

0.000

-0.013

-0.010

Carburant\_Electrique

0.0357

0.018

2.032

0.042

0.001

0.070

Carburant\_Essence

-0.0164

0.001

-17.797

0.000

-0.018

-0.015

Carburant\_Hybride

0.0214

0.004

5.232

0.000

0.013

0.029

car\_model\_ABARTH 595 turismo 165ch

3.27e-11

1.09e-11

2.990

0.003

1.13e-11

5.41e-11

car\_model\_ABARTH 695

-0.0079

0.020

-0.402

0.688

-0.046

0.031

# Les modèles

## Régression linéaire généralisée

```
Dep. Variable:          price  No. Observations:          6379
Model:                GLM      Df Residuals:          5887
Model Family:          Gamma   Df Model:            491
Link Function:          log     Scale:              4.7832e-05
Method:                IRLS     Log-Likelihood:       17206.
Date:                  Tue, 14 May 2024  Deviance:           0.27822
Time:                  09:28:02  Pearson chi2:         0.282
No. Iterations:         10      Pseudo R-squ. (CS):   0.9998
Covariance Type:        nonrobust

=====
                        coef      std err          z      P>|z|      [0.025      0.975]
-----
const                   1.7334        0.239        7.263      0.000        1.266        2.201
car_age                -0.0022      2.52e-05     -87.351      0.000       -0.002       -0.002
Kilométrage           -1.869e-08    1.23e-09    -15.160      0.000     -2.11e-08    -1.63e-08
Puissance fiscale      -0.0002      6.16e-05     -3.922      0.000       -0.000       -0.000
Year_of_annonce        -0.0004        0.000     -3.314      0.001       -0.001       -0.000
Boite de vitesses_Manuelle -0.0048        0.000    -16.798      0.000       -0.005       -0.004
Carburant_Electrique    0.0136        0.007        1.907      0.056       -0.000        0.028
Carburant_Essence       -0.0068        0.000    -18.234      0.000       -0.008       -0.006
Carburant_Hybride        0.0081        0.002        4.874      0.000        0.005        0.011
car_model_ABARTH 595 turismo 165ch -3.228e-13  1.37e-13     -2.353      0.019     -5.92e-13    -5.39e-14
car_model_ABARTH 695     -0.0032        0.008     -0.404      0.686       -0.019        0.012
...

=====
Mean Squared Error for LOG_log_price is: 0.0003526221772965861
R-squared (R2) on test set: 0.8780249448772938
Mean Squared Error for the price is: 7731981250.593452
```



# Les modèles

Comparaison avec d'autres modèles:

Linear Regression MSE: 0.000355674431137125

Ridge Regression MSE: 0.0003578920252191233

Lasso Regression MSE: 0.002364703637756404

Random Forest MSE: 0.00037049914324493354

Support Vector Machine MSE: 0.002328583735843007

Linear Regression R-squared: 0.876969143925401

Ridge Regression R-squared: 0.876202058989152

Lasso Regression R-squared: 0.18202859849735908

Random Forest R-squared: 0.8718411480336188

Support Vector Machine R-squared: 0.19452278437270676

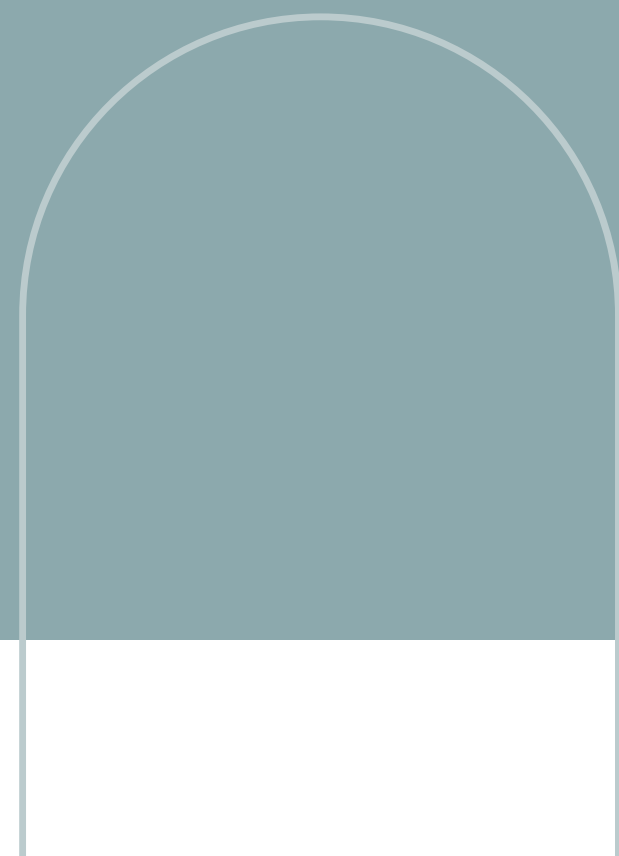
A decorative grid of 40 small, light gray dots arranged in 5 rows and 8 columns, located in the top left corner of the slide.

# Conclusion





Merci de  
votre  
attenssion



02.

# PRODUCT

*Liceria & Co.*

36



**“JUST KEEP  
MOVING  
FORWARD.”**



03.

# MARKET & TREND

*Liceria & Co.*

38



03.

# BUSINESS MODEL

*Liceria & Co.*

39