

12 MAI 2025

# ANALYSE COMPARATIVE DES ANNONCES AUTOMOBILES EN LIGNE

*Étude basée sur les données de  
AutoScout24 et Automobile.fr*

SOFIANE HAMMA  
GOVINDARAJEN COODIEN  
LYES AIT TAYEB  
RAYANE DZIRI

# SOMMAIRE

**Contexte et objectifs 01**

**Méthodologie 02**

**Présentation des sites 03**

**Données collectées 04**

**Traitement et préparation des données 05**

**Analyses 06**

**Conclusion 07**

# CONTEXTE ET OBJECTIFS

## Contexte

Le marché des véhicules d'occasion est en plein essor, avec une digitalisation croissante des ventes. Les plateformes comme AutoScout24 et Automobile.fr sont devenues des références incontournables pour les acheteurs et les vendeurs.

## Problématique

Face à la multitude d'annonces en ligne, il est difficile d'avoir une vision claire des tendances du marché : quels sont les prix moyens ? Quelles marques sont les plus représentées ? Quelles variables influencent le prix ?

## Objectifs

Notre objectif est de collecter automatiquement les données de ces deux sites, les structurer proprement, puis les analyser afin d'identifier les grandes tendances, comparer les plateformes, et mettre en lumière les facteurs influençant les prix.

# MÉTHODOLOGIE



## **Etape 1**

Scraping avec  
Python (  
BeautifulSoup,  
Selenium)



## **Etape 2**

Stockage et  
nettoyage des  
données (pandas)



## **Etape 3**

Analyse statistique  
et visualisation  
(Seaborn,  
Matplotlib)



## **Etape 4**

Comparaison entre  
les plateformes

# PRÉSENTATION DES SITES

## **AutoScout24**

AutoScout24 est l'un des plus grands sites européens de vente de véhicules d'occasion. Il couvre plusieurs pays et propose un large choix de voitures avec des filtres détaillés : marque, modèle, prix, kilométrage, carburant, etc.

## **Automobile.fr**

Automobile.fr est la version française du site mobile.de, très populaire en Allemagne. Il s'adresse principalement au marché français et permet d'accéder à des milliers d'annonces avec des informations précises sur chaque véhicule.

# DONNÉES COLLECTÉES

*Colonnes récupérées*

✦ **Marque, modèle**

✦ **Prix**

✦ **Année, kilométrage**

✦ **Energie, Boite**

✦ **Code postal, Département**



	A	B	C	D	E
1	Marque,Modèle,Prix (â,~),Année,Boîte,KM,Consommation,Code Postal,Département				
2	Renault,Grand Scenic,,03/2008,,, "Autre, Diesel" ,,				
3	BMW,116 Baureihe 1 Lim. 116i NUR FUR TEILE,,10/2004,"Couleur extérieure: Noir, Nombre de portes: 4/5",85 kW (116 Ch DIN),"Citadine, Essence, Boîte manuelle" ,,				
4	Renault,Kangoo Privilege,,07/2007,"Couleur extérieure: Gris, Nombre de portes: 4/5",62 kW (84 Ch DIN),"Break, Diesel, Boîte automatique" ,,				
5	Peugeot,Expert Kasten L1H1 Kasten,,04/2013,"Couleur extérieure: Blanc, Nombre de portes: 4/5",94 kW (128 Ch DIN),"Autre, Diesel, Boîte manuelle" ,,				
6	Peugeot,308 SW Allure 130 PureTech EAT6 STOP & START,,05/2014,Nombre de portes: 4/5,96 kW (131 Ch DIN),"Break, Essence, Boîte automatique" ,,				
7	Peugeot,308 Style 120 VTi,,12/2014,Nombre de portes: 4/5,88 kW (120 Ch DIN),"Berline, Essence, Boîte manuelle" ,,				
8	Renault,Megane III Coupe Dynamique,,02/2010,"Couleur extérieure: Noir, Nombre de portes: 2/3",81 kW (110 Ch DIN),"Coupé Sport, Diesel, Boîte manuelle" ,,				
9	Citroën,DS3 THP 155 SportChic,,09/2013,Nombre de portes: 2/3,115 kW (156 Ch DIN),"Citadine, Essence, Boîte manuelle" ,,				
10	Peugeot,"208 STE 1,2 PUERETECH 75" ,,,03/2022,"Couleur extérieure: Blanc, Nombre de portes: 4/5",55 kW (75 Ch DIN),"Citadine, Essence, Boîte manuelle" ,,				
11	Land,Rover Range Rover Sport V6 TD SE,,05/2006,"Couleur extérieure: Noir, Nombre de portes: 4/5",140 kW (190 Ch DIN),"SUV / Off-road Vehicle / Pickup Truck, Diesel, Boîte automatique" ,,				
12	Audi,Q5 3.0 TDI quattro NUR FUR TEILE,,12/2010,"Couleur extérieure: Noir, Nombre de portes: 4/5",176 kW (239 Ch DIN),"SUV / Off-road Vehicle / Pickup Truck, Diesel, Boîte automatique" ,,				
13	Renault,"Kangoo Maxi 1,5 dci 90" ,,,03/2014,"Couleur extérieure: Blanc, Nombre de portes: 4/5",66 kW (90 Ch DIN),"Monospace / Ludospace, Diesel, Boîte manuelle" ,,				
14	Citroën,"DS3 Cabrio 1,6 Thp 156 SportChic" ,,,01/2013,"Couleur extérieure: Noir, Nombre de portes: 2/3",115 kW (156 Ch DIN),"Cabriolet / Roadster, Essence, Boîte manuelle" ,,				
15	Volkswagen,"Touran Comfortline 1,6tdi 105ps DSG" ,,,12/2012,"Couleur extérieure: Beige, Nombre de portes: 4/5",77 kW (105 Ch DIN),"Citadine, Diesel, Boîte automatique" ,,				
16	Audi,A1 1.4 TFSI S-TRONIC,,02/2012,Nombre de portes: 2/3,90 kW (122 Ch DIN),"Citadine, Essence, Boîte automatique" ,,				
17	Citroën,"C5 Lim. Exclusive 2,0hdi 163ps Automatik" ,,,08/2015,"Couleur extérieure: Brun, Nombre de portes: 4/5",120 kW (163 Ch DIN),"Berline, Diesel, Boîte automatique" ,,				
18	Mercedes-Benz,B 200 B -Klasse cdi 6 Gang Navi Camera,,02/2015,"Couleur extérieure: Blanc, Nombre de portes: 4/5",100 kW (136 Ch DIN),"Citadine, Diesel, Boîte manuelle" ,,				
19	DS,Automobiles DS5 1.6hdi panoramique roof eat6,,04/2016,"Couleur extérieure: Blanc, Nombre de portes: 4/5",88 kW (120 Ch DIN),"Berline, Diesel, Boîte automatique" ,,				
20	Alfa,Romeo Giulia 1300 SUPER NUOVA-Origine France/Oldtimer,,02/1976,Nombre de portes: 4/5,, "Berline, Essence, Boîte manuelle" ,,				
21	Opel,Astra J 1.4l 140ch,,12/2010,"Couleur extérieure: Blanc, Nombre de portes: 4/5",103 kW (140 Ch DIN),"Citadine, Essence, Boîte manuelle" ,,				

# TRAITEMENT ET PRÉPARATION DES DONNÉES

## **Nettoyage des données**

- Conversion des dates en format standard (pd.to\_datetime).
- Nettoyage des champs kilométrage (KM) et prix (Prix) en supprimant les caractères non numériques.

## **Préparation pour l'analyse**

- Création de la colonne Annee\_Only pour extraire l'année.
- Conversion des colonnes en types exploitables (int, float, datetime).
- Sélection des variables d'intérêt : Prix, KM, Annee\_Only, Modèle, Marque, Département.




```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler


df = pd.read_csv('voitures.csv')
df['Année'] = pd.to_datetime(df['Année'], format='%m-%Y', errors='coerce')
df['Annee_Only'] = df['Année'].dt.year
df['KM'] = pd.to_numeric(df['KM'].astype(str).str.replace(r'\D', '', regex=True))
df['Prix'] = pd.to_numeric(df['Prix'], errors='coerce')
```

# ANALYSE UNIVARIÉE


*Objectif : Étudier la distribution individuelle de chaque variable.*



Histogrammes de  
Prix, KM et  
Annee\_Only.

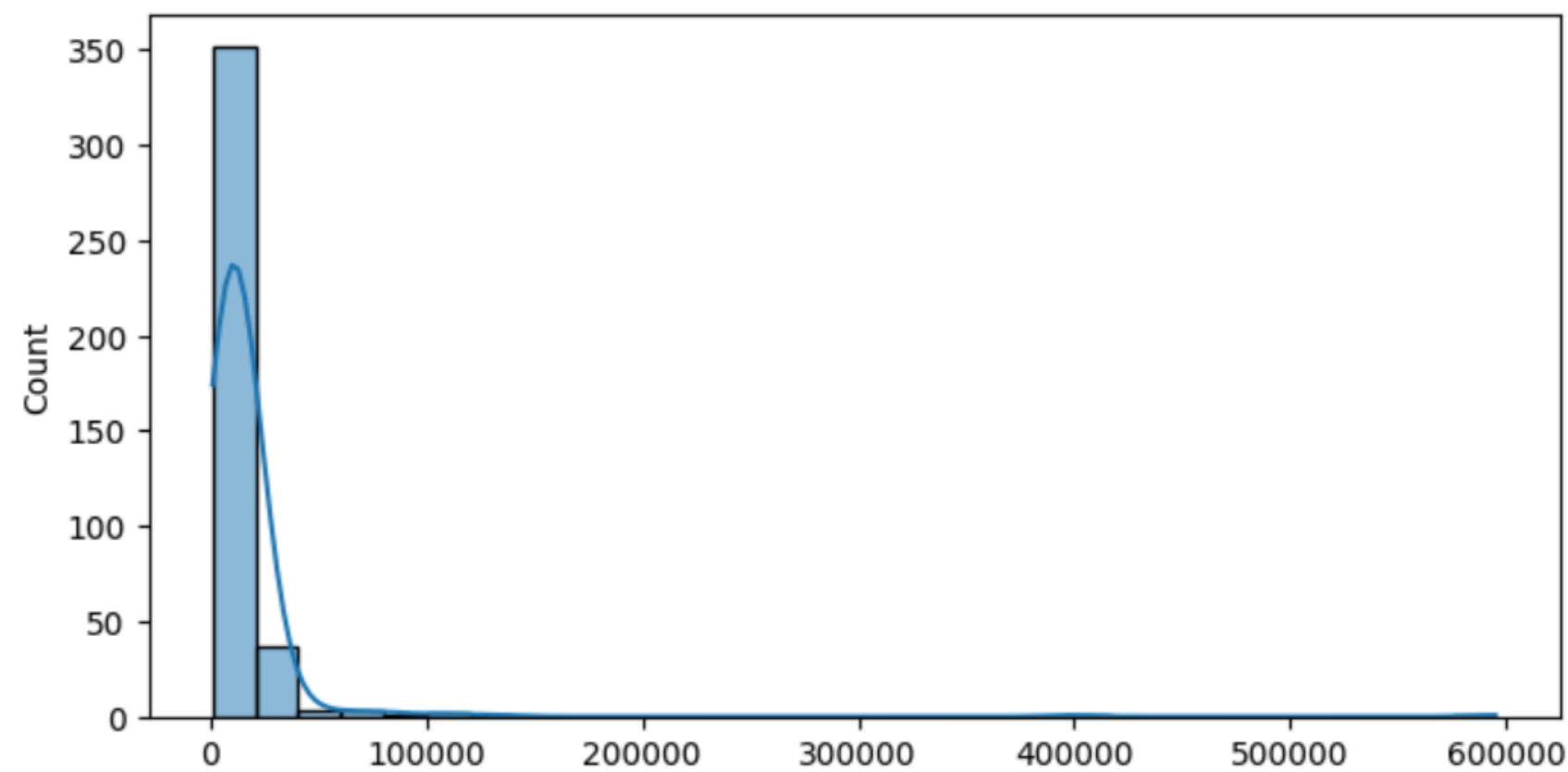


Utilisation de  
Seaborn avec  
histplot + courbe KDE  
pour visualiser la  
densité.

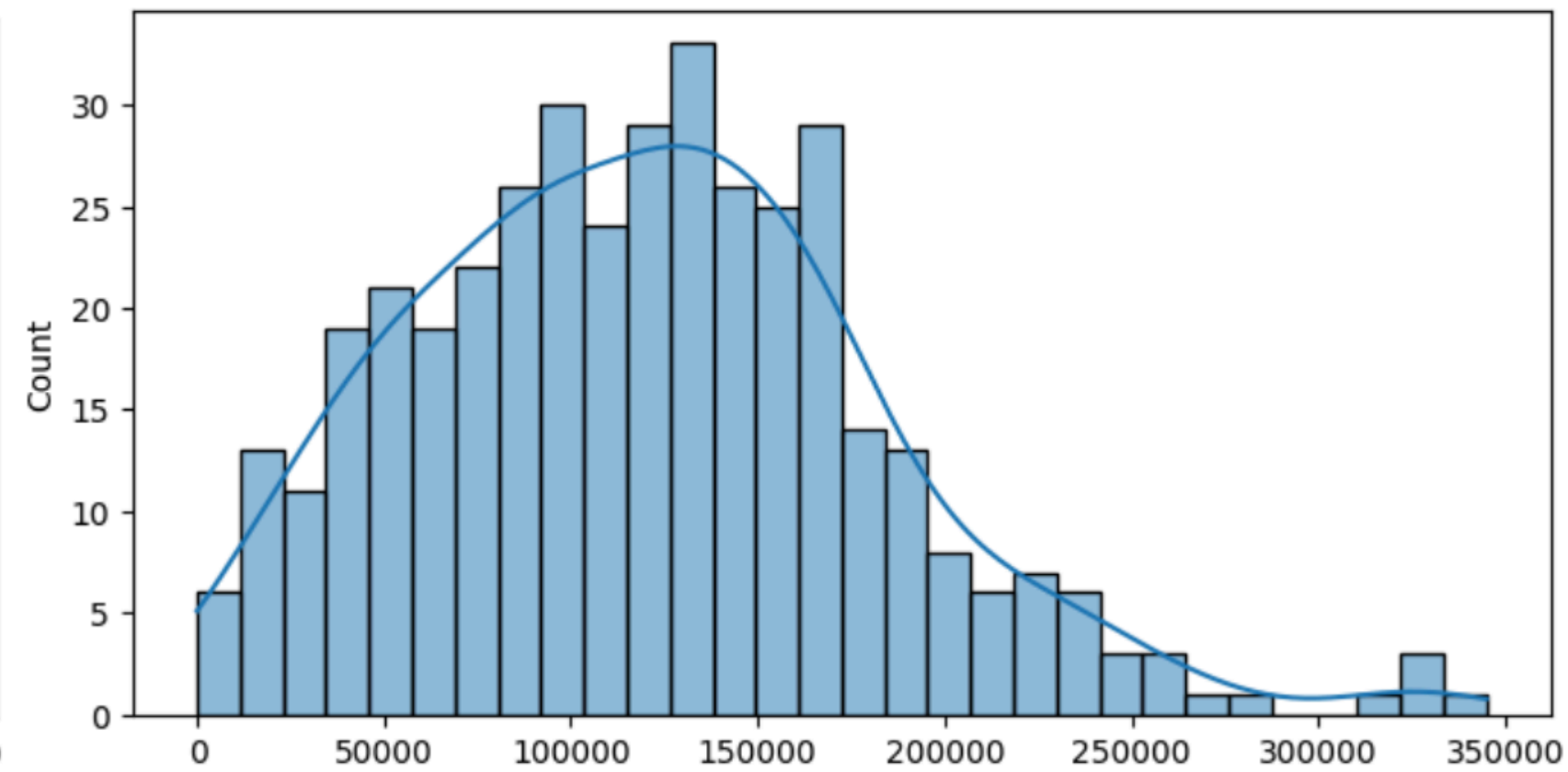


Identification des  
valeurs extrêmes,  
asymétries et  
tendances globales.

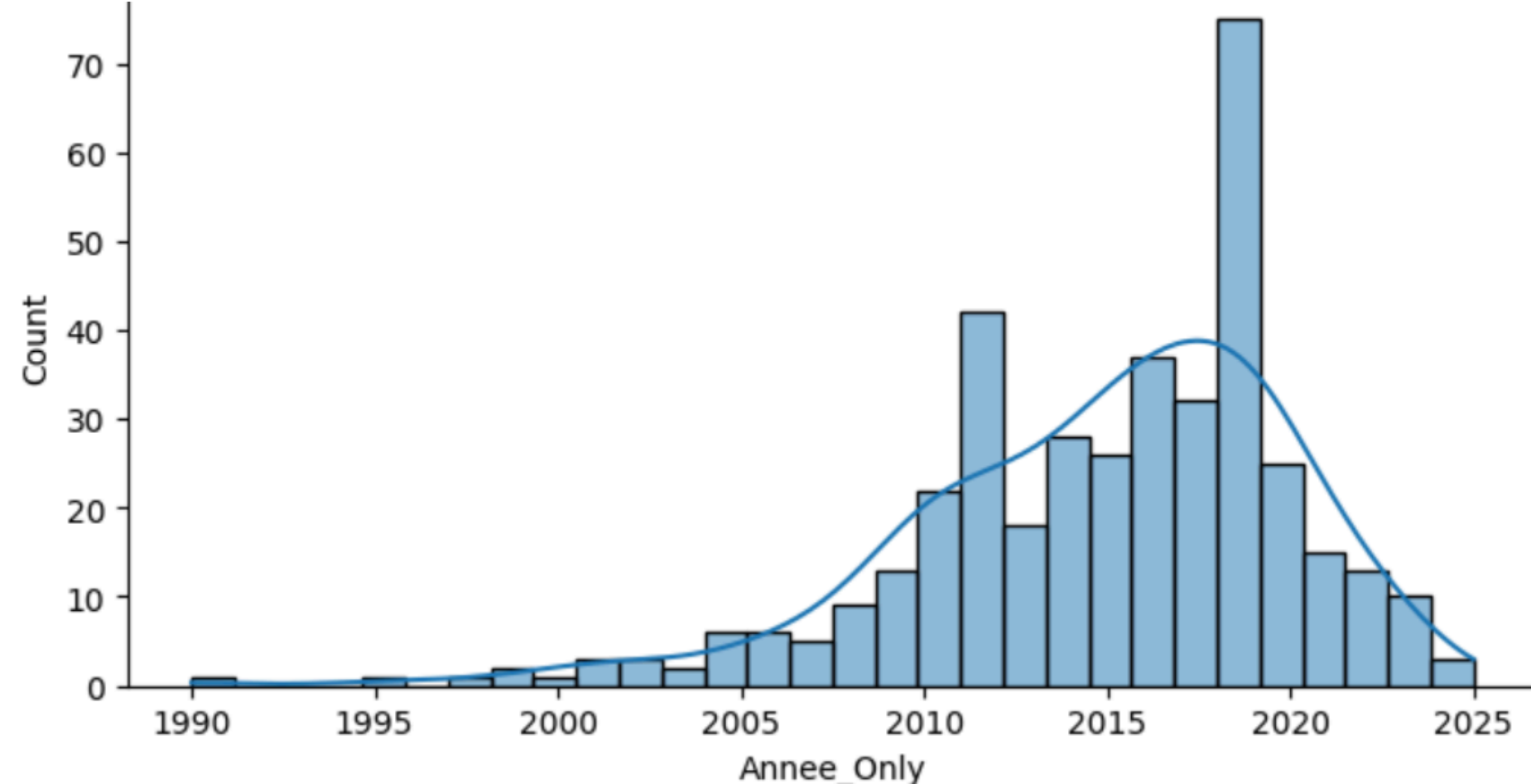
Distribution de Prix



Distribution de KM

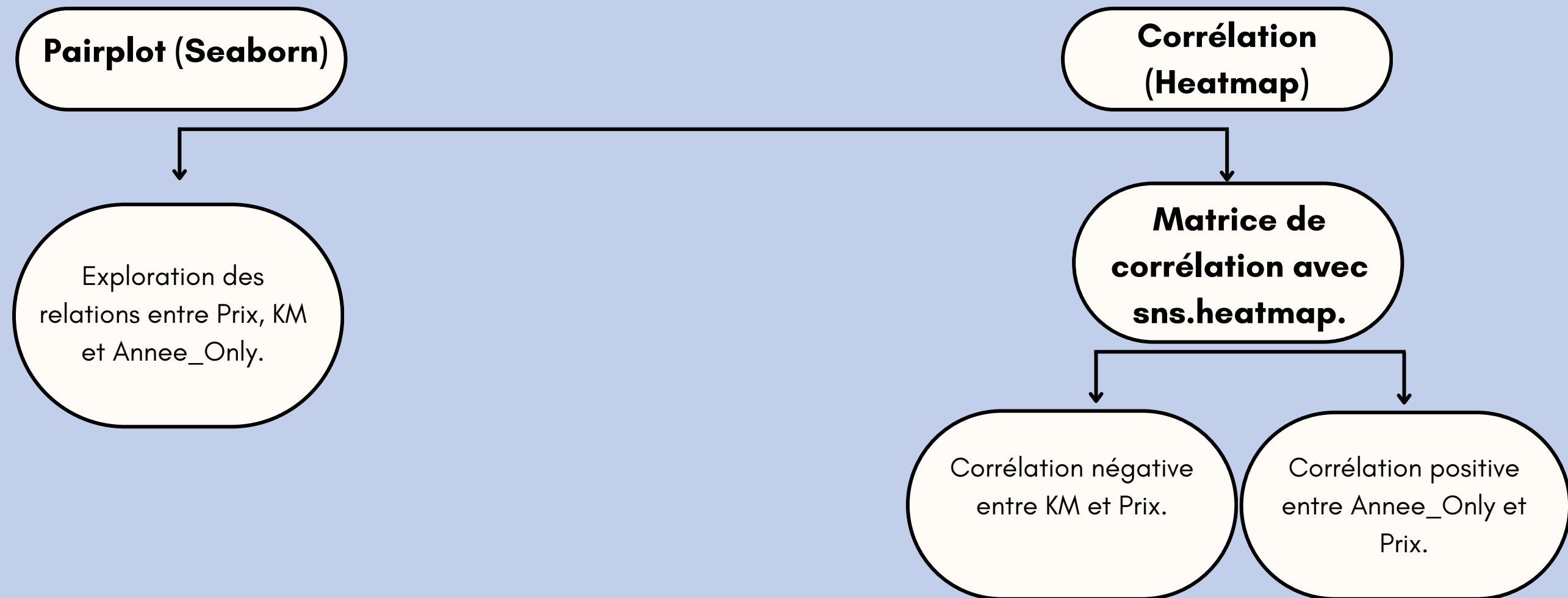


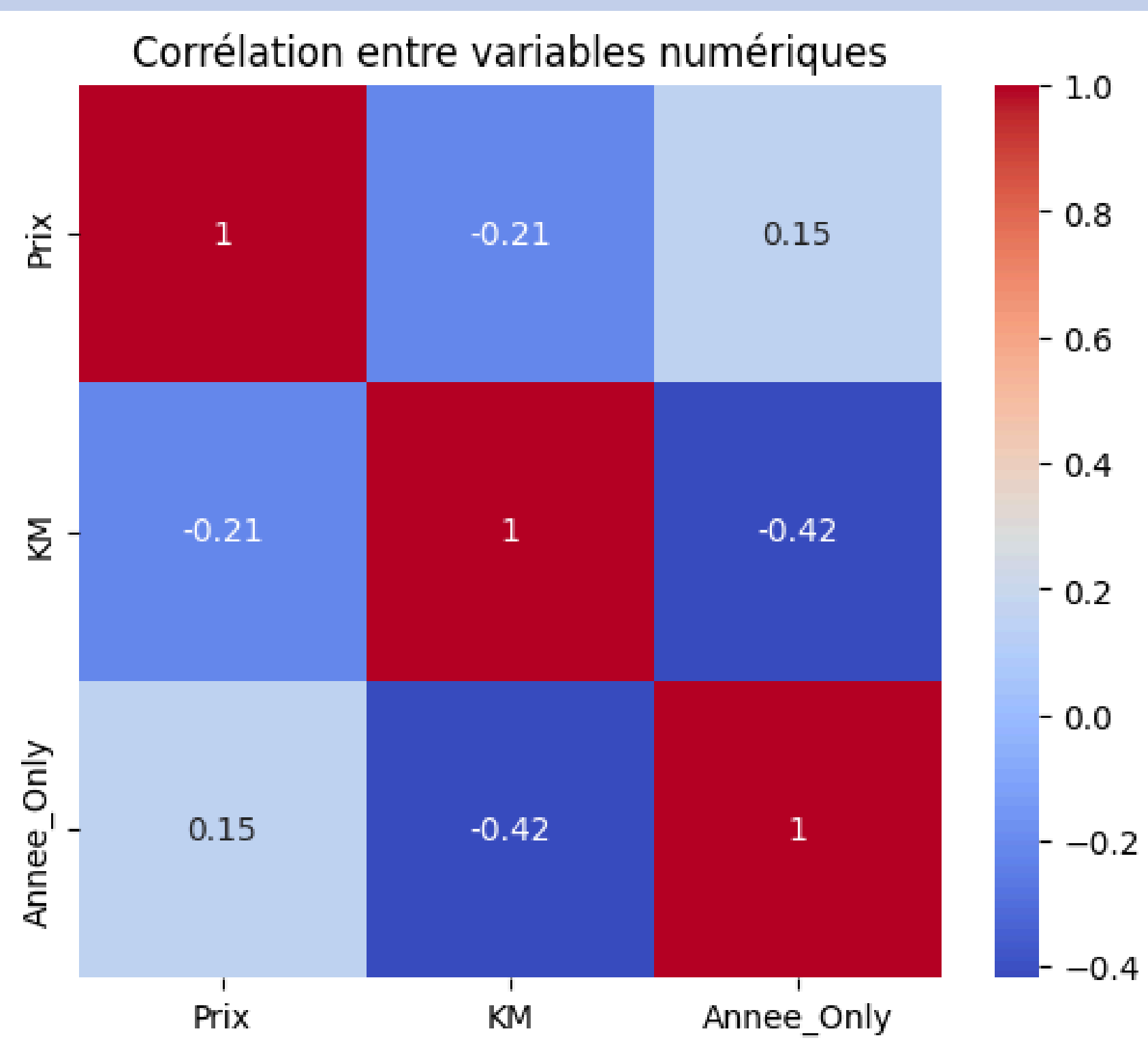
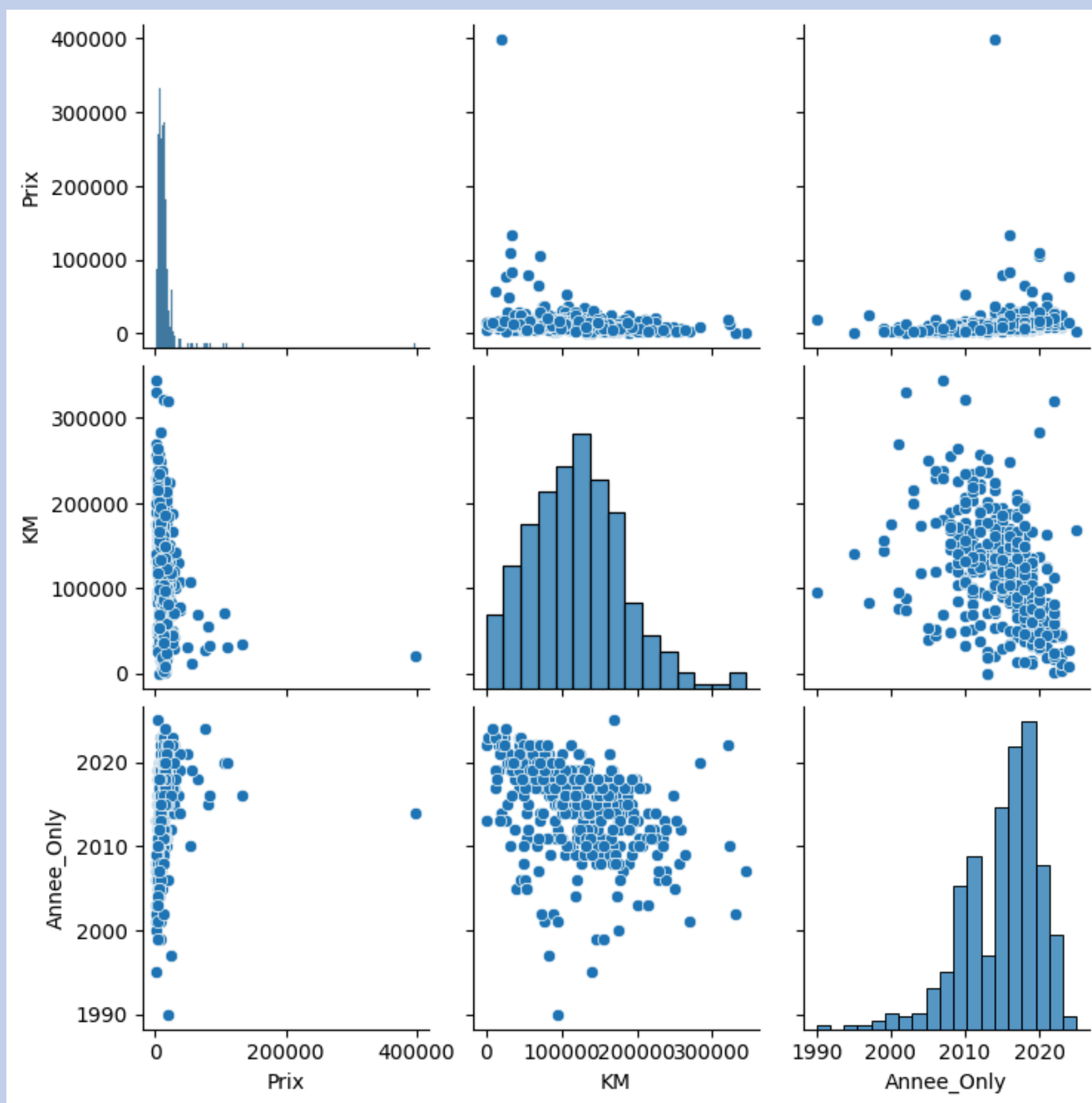
Prix



KM

# ANALYSE BIVARIÉE





# ANALYSE DES MODÈLES ET DÉPARTEMENTS

## **Modèles les plus populaires par département**

Nous avons identifié, pour chaque département, le modèle de voiture le plus fréquemment mis en vente. Cela permet de repérer les préférences régionales et les tendances locales du marché automobile.

## **Les 10 modèles les plus fréquents**

Une analyse globale montre que certains modèles dominent largement le marché. Ce sont souvent des véhicules réputés pour leur fiabilité, leur prix abordable ou leur popularité.

## **Prix moyen par modèle**

En étudiant le prix moyen de chaque modèle, nous avons pu repérer ceux qui se positionnent sur le haut de gamme et ceux qui restent plus accessibles. Ces données permettent de mieux comprendre la structure tarifaire du marché.



# ANALYSE PAR MARQUE

*Objectif : comprendre quelles marques se positionnent comme "premium" ou "économiques".*



## **Comparaison des marques**

- Moyenne des prix et kilométrages par marque.
- Groupement via `groupby('Marque')[['Prix', 'KM']].mean()`.

	Département	Modèle	Nombre
0	3	kadjar	1
1	4	500	1
18	6	q3	3
20	8	2008	1
22	10	308	1
24	12	note	1
26	13	208	2
41	14	3008	1
43	15	duster	1
44	17	a 250	1

```
Modèle
3008      13
polo      12
208       11
clio      11
tiguan    10
2008      10
golf       9
5008       7
sandro    7
308        7
Name: count, dtype: int64
```

```
Modèle
12 cylindri  595000.0
458          399000.0
570s        132900.0
m8           109900.0
rs6          105900.0
continental gt   82900.0
cayman        79900.0
m3            59450.0
mustang       56900.0
a 35 amg      49980.0
Name: Prix, dtype: float64
```

# CLUSTERING DES MARQUES

## **KMeans clustering (4 groupes)**

Application d'un KMeans sur les marques, basé sur Prix moyen et KM moyen

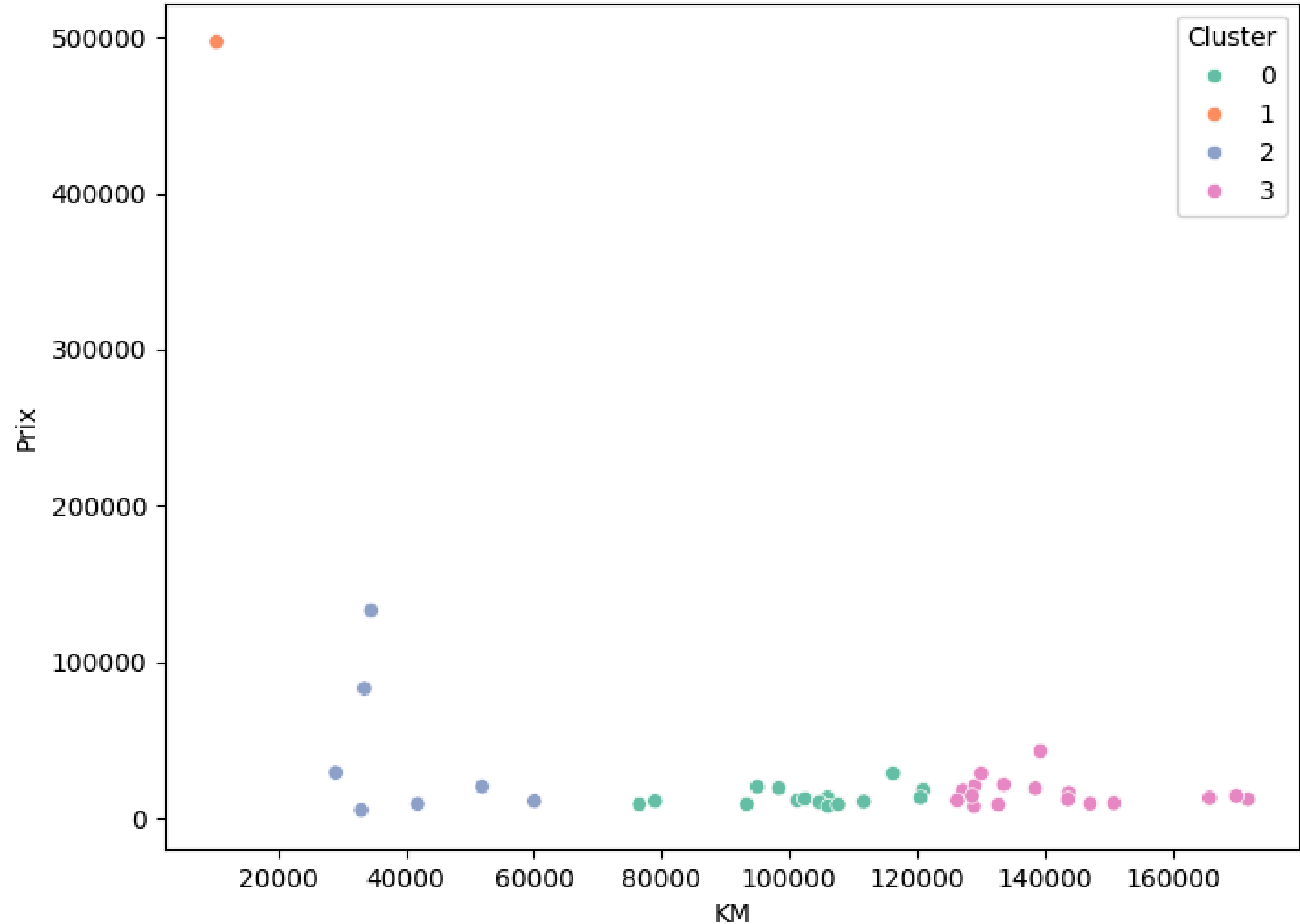
## **Prétraitement**

Standardisation des données avec StandardScaler.

## **Visualisation**

- Scatterplot des marques, colorées par cluster (hue='Cluster').
- Interprétation possible :
  - Cluster 0 = Marques haut de gamme (haut prix, faible kilométrage).
  - Cluster 1 = Marques économiques, etc.

Clustering des marques selon prix et kilométrage



# CONCLUSION DE L'ANALYSE EXPLORATOIRE

## Enseignements clés

Notre analyse a révélé des tendances claires dans le marché automobile d'occasion. Le prix des véhicules est fortement influencé par leur kilométrage et leur année de mise en circulation. De plus, certaines marques et modèles se démarquent selon leur positionnement : haut de gamme, économique, ou grand public. Les disparités entre départements soulignent aussi l'impact de l'offre locale.

## Apports de l'analyse

Grâce à la visualisation des données et à l'application de méthodes statistiques (histogrammes, corrélation, clustering), nous avons obtenu une compréhension approfondie du marché. Le clustering a permis de regrouper les marques selon leur profil type, facilitant l'interprétation des différences de positionnement.

## Perspectives futures

Cette analyse constitue une première étape. Elle peut être enrichie par des modèles prédictifs de prix, une comparaison dynamique entre AutoScout24 et Automobile.fr, ou encore l'intégration de critères supplémentaires comme le type de carburant, la transmission ou les options. Cela ouvrirait la voie à des outils d'aide à l'achat plus performants.

MERCI  
*à vous !*