18/05/2025

Encrypted Scraping and Analysis of Car Listings

Project by: Oumayma Abayed
Eslem Sebri

Abstract

This project involved building a complete data pipeline starting from web scraping to secure storage and interface-based analysis. We scraped car sales listings from the "Véhicules" section of the Tunisian website Tayara.tn, processed and cleaned the data, encrypted it to ensure data confidentiality, and developed a basic interface that allows decryption and displays insights such as car model frequency, pricing trends, and common brands.

Introduction

Why scrape car listings?

We wanted to collect real-world data from an active marketplace to analyze trends in the Tunisian used car market. Car listings offer useful insights such as price ranges, popular brands, , and vehicle conditions, which are valuable for both buyers and sellers.

Why choose Tayara.tn?

Tayara.tn is one of Tunisia's most widely used e-commerce platforms, especially for classified ads. Its "Véhicules" section is frequently updated and contains detailed information about vehicles, making it a rich and reliable source for scraping and analysis.

The problem to solve?

The data available on Tayara.tn is unstructured and only accessible via individual web pages. We needed to collect it in a structured format for analysis. Additionally, we wanted to ensure that the collected data was securely stored and accessed only by authorized users.

Why use encryption?

Encryption ensures that the dataset remains confidential and protected from unauthorized access. Since web-scraped data may be sensitive or subject to privacy concerns, applying encryption demonstrates responsible data handling and adds a security layer to our pipeline.

Project Pipeline

01

SCRAPE CAR SALES DATA FROM TAYARA.TN

02

CLEAN AND STORE THE DATA (CSV)

03

ENCRYPT THE DATASET

04

CREATE AN INTERFACE FOR DECRYPTION AND ANALYSIS

4.1

Web Scraping

Strategies

Tools Used

We used **Selenium WebDriver** with a headless Chrome browser to simulate real user behavior while navigating Tayara.tn. This allowed us to dynamically load JavaScript-rendered content and extract structured data directly from the HTML.

Target

Our scraper focused on the first 18 pages of the "Véhicules" section of Tayara.tn, which contains public listings for used and new vehicles posted by individuals and dealerships.

Data Extracted

- Title
- Price (in Tunisian Dinar)
- Link
- Kilométrage (Mileage)
- Couleur du véhicule (Color)
- Etat du véhicule (State)
- Boite (Gearbox)
- Année (Year)
- Cylindrée (Engine_cap)
- Marque (Brand)
- Modèle (Model)
- Type de carrosserie (Body_type)
- Carburant (Fuel)

Strategies

Pagination Handling

The scraper looped through URLs in the format:
https://www.tayara.tn/ads/c/V%C3%A9hicules/?page=1, ?page=2, ... up to ?page=18,
automatically loading each page and collecting all listing links before visiting them individually.

Error Handling

We implemented several layers of robustness:

- A **try/except** block around each listing scrape to **skip malformed or incomplete entries**
- A delay (**time.sleep**) with randomization to **avoid server-side blocks or rate limits**
- A timeout handler for **driver.get()** using **set_page_load_timeout()** to **handle slow-loading or unresponsive pages**

4.2

Data Cleaning

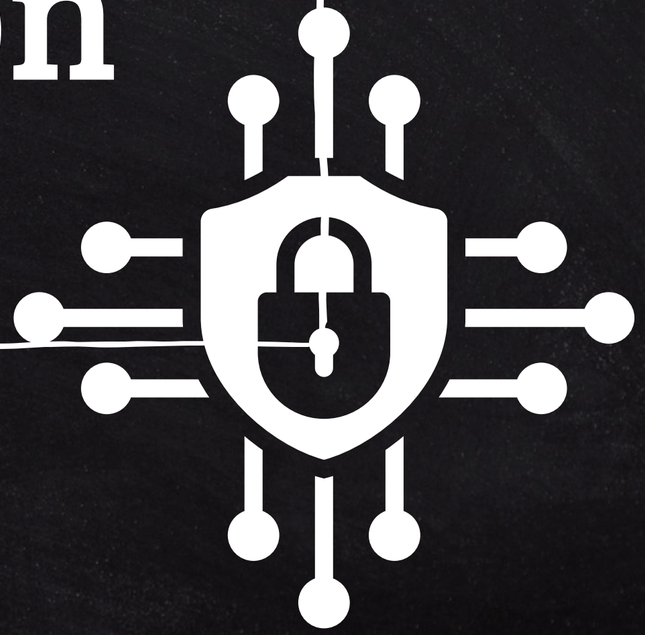
Python – Pandas – Numpy

4.3

CSV Exporting

4.4

Encryption



Encryption Process:

Key Generation

A symmetric encryption key is generated once and saved in **encryption.key** for reuse.

Fernet Initialization

The Fernet object is initialized using the generated key. This object handles the actual encryption and decryption processes, ensuring secure AES encryption with built-in authentication and timestamp features.

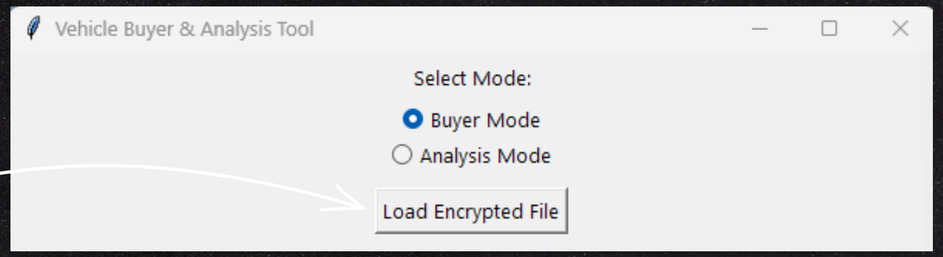
File Encryption

The cleaned CSV file is read in binary mode, encrypted using the Fernet object, saved to a new file named **tayara_vehicles_cleaned_encrypted.csv.**
This encrypted version is unreadable without the correct key, thereby protecting the dataset from unauthorized access

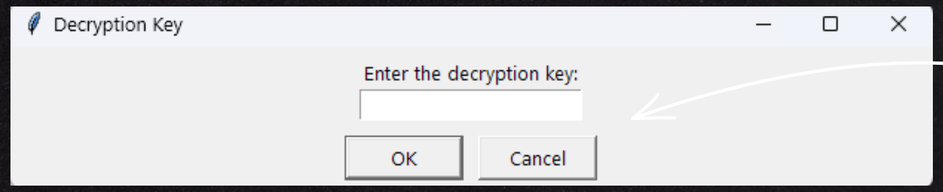
4.5

Interface and Decryption

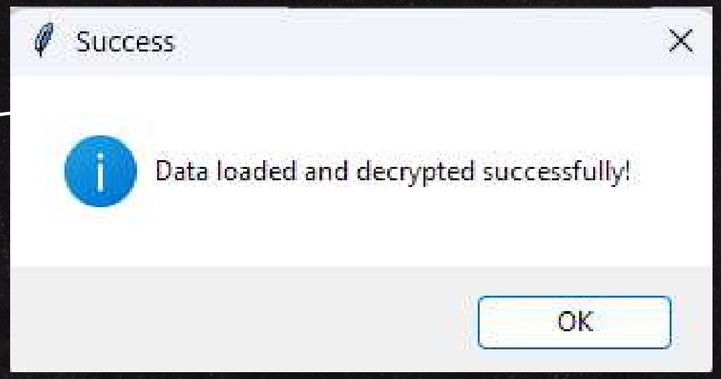
Decryption



Step 1:
upload the encrypted file

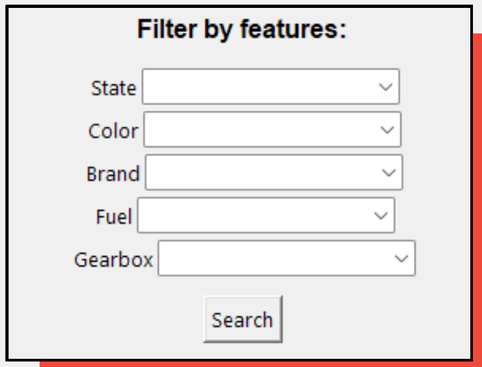


Step 2:
enter the encyption key



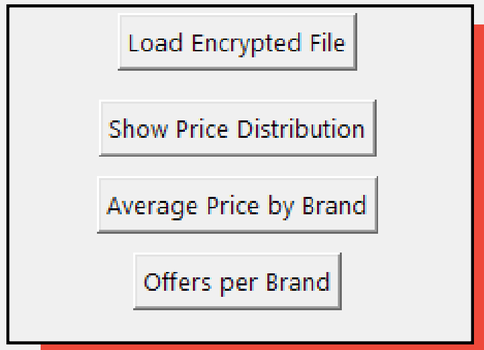
Step 3:
decryption process
using fernet

UI Modes



Mode 1:
User Mode

Users can filter listings by brand, color, state, fuel type, and gearbox.

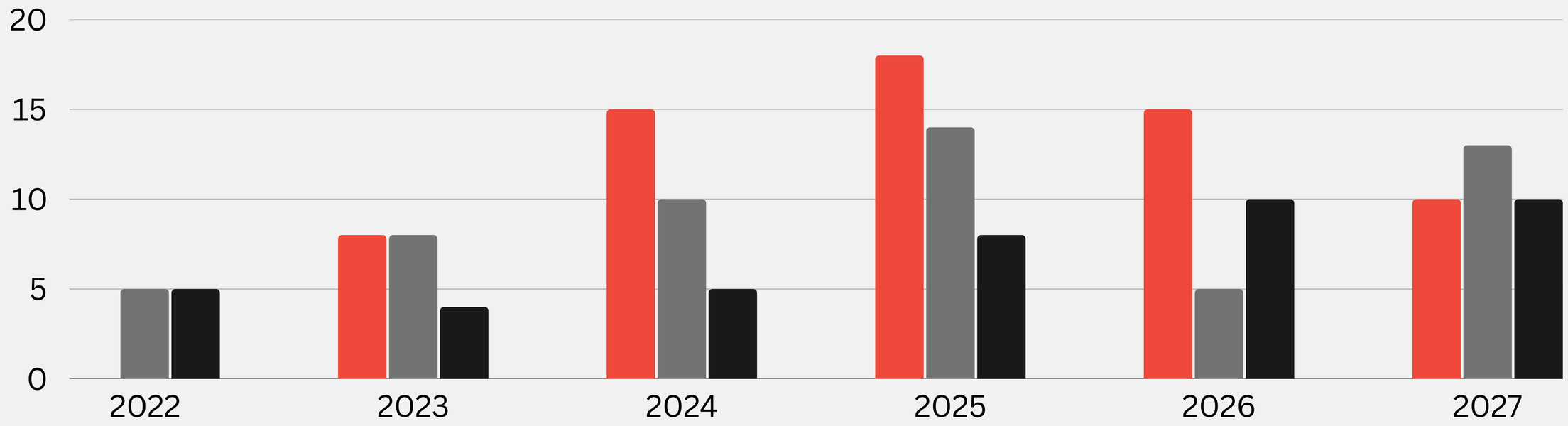


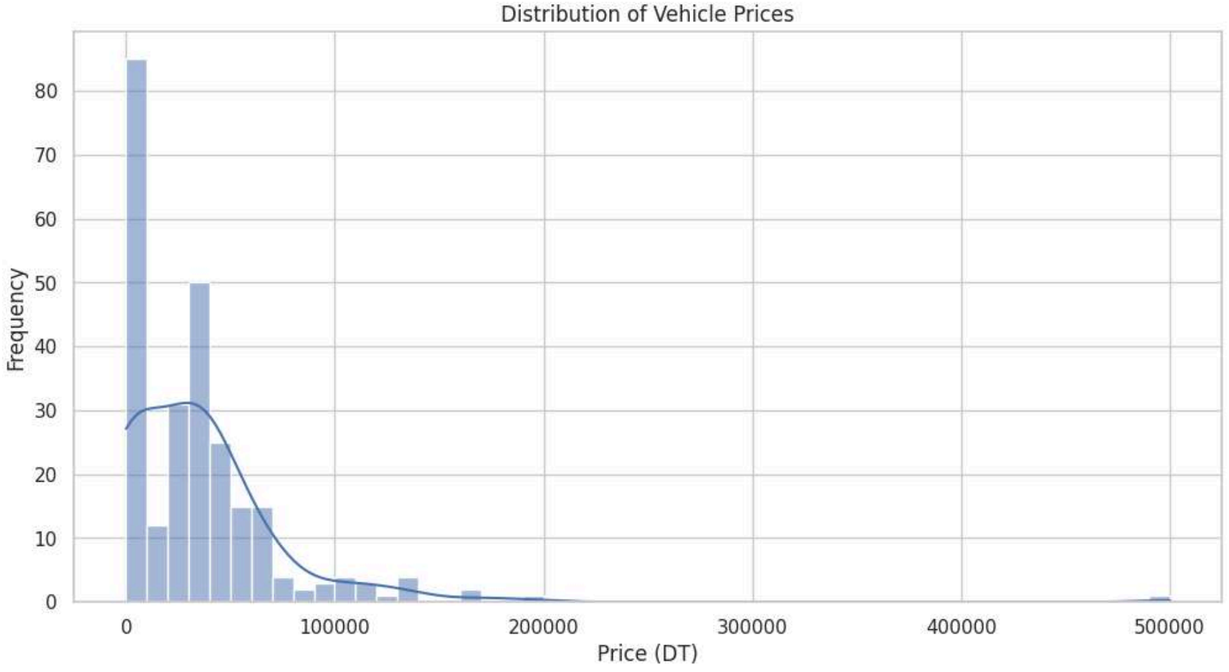
Mode 2:
Data Analysis Mode

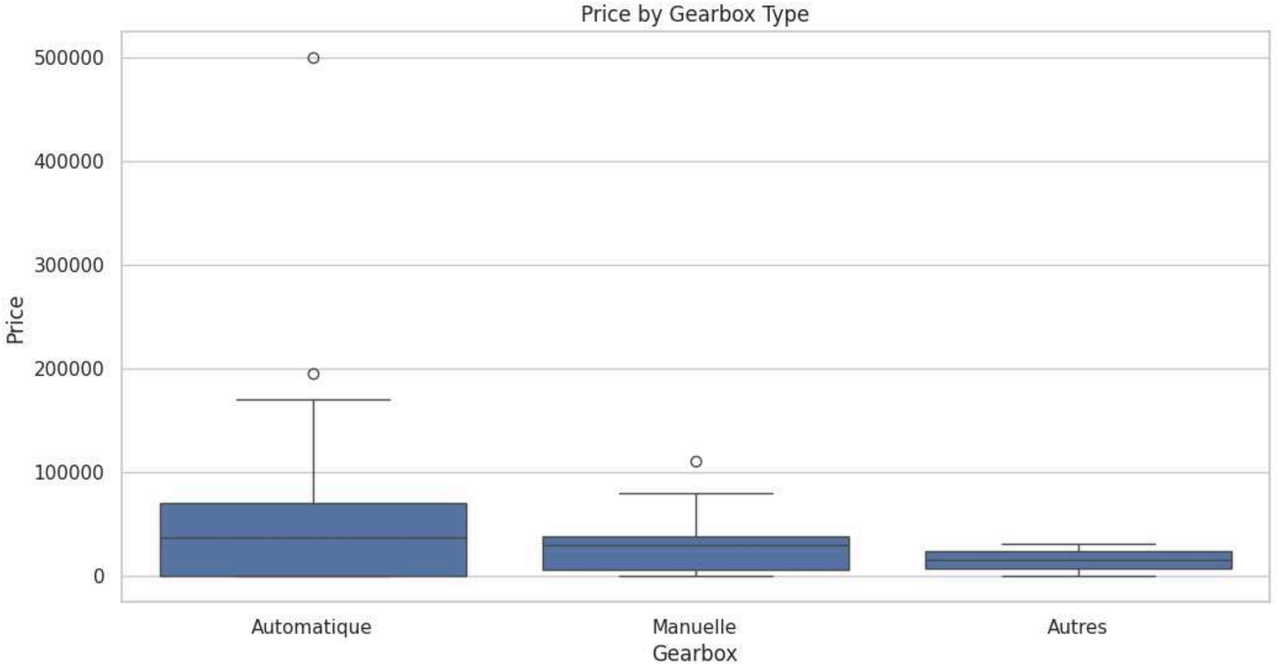
Offers visual insights like price distributions, average prices by brand, and listing frequency.

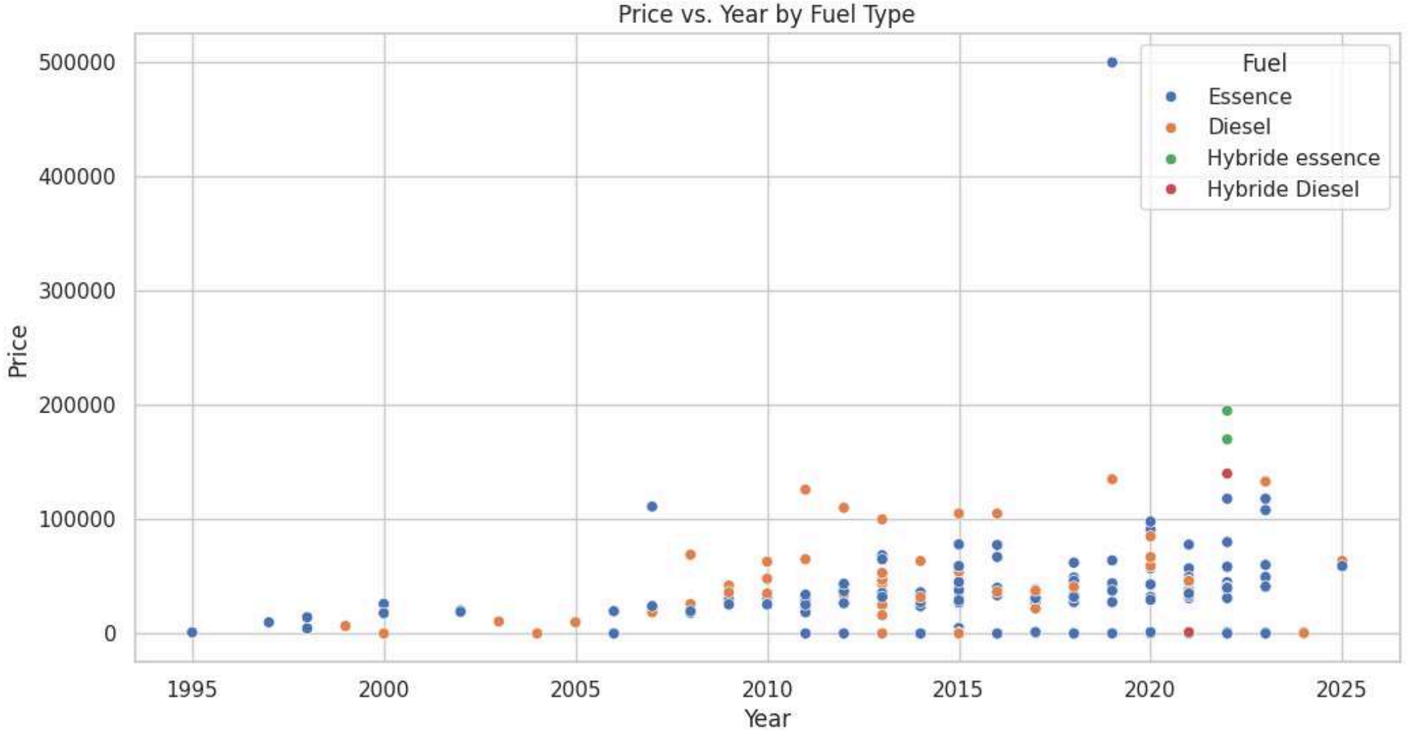
Tool used: Tkinter
a python standard GUI library.

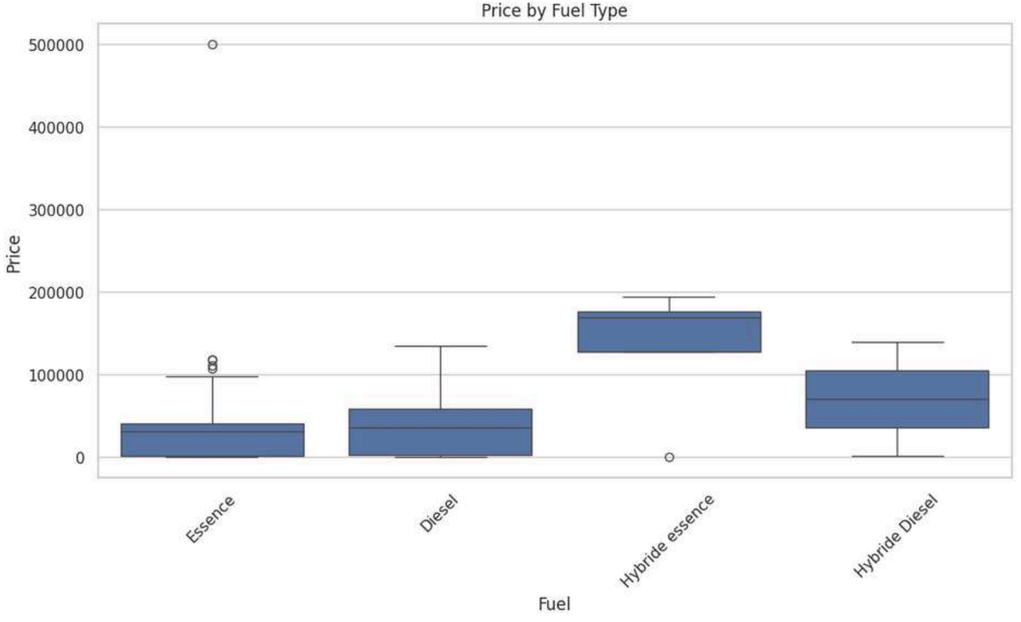
Data Analysis

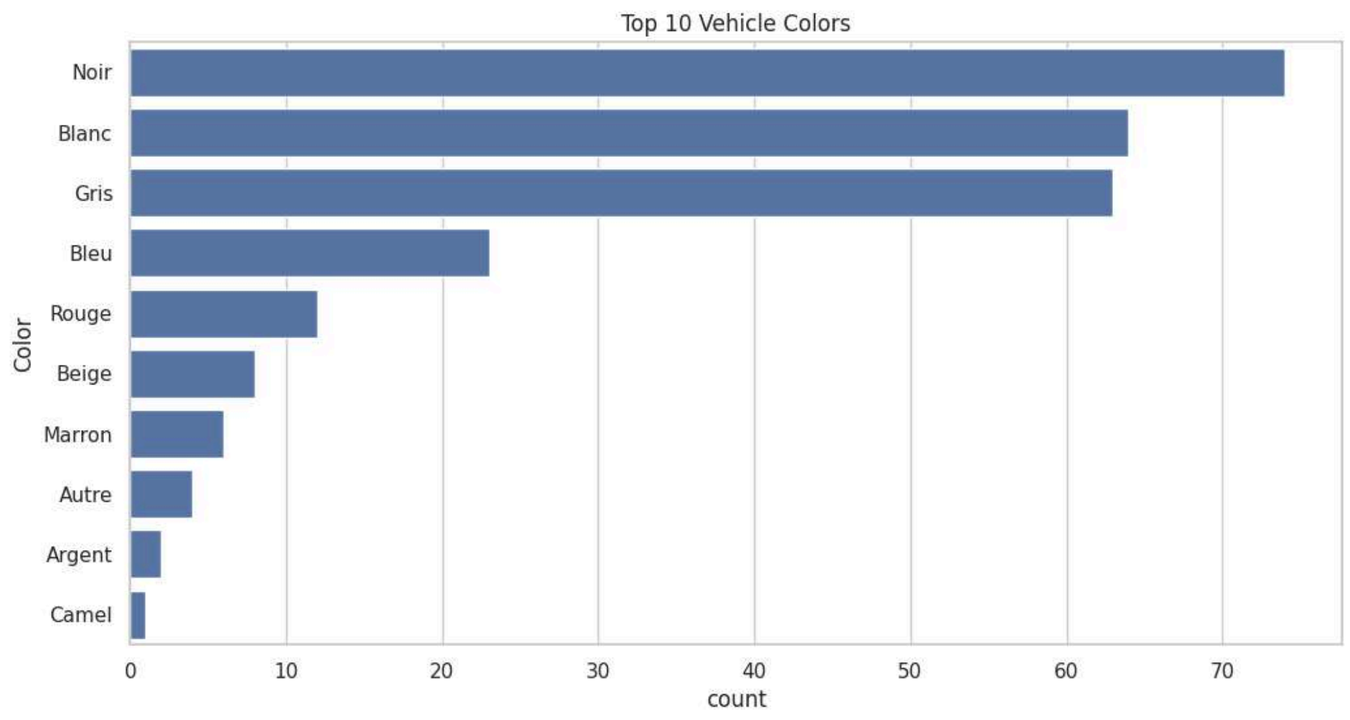


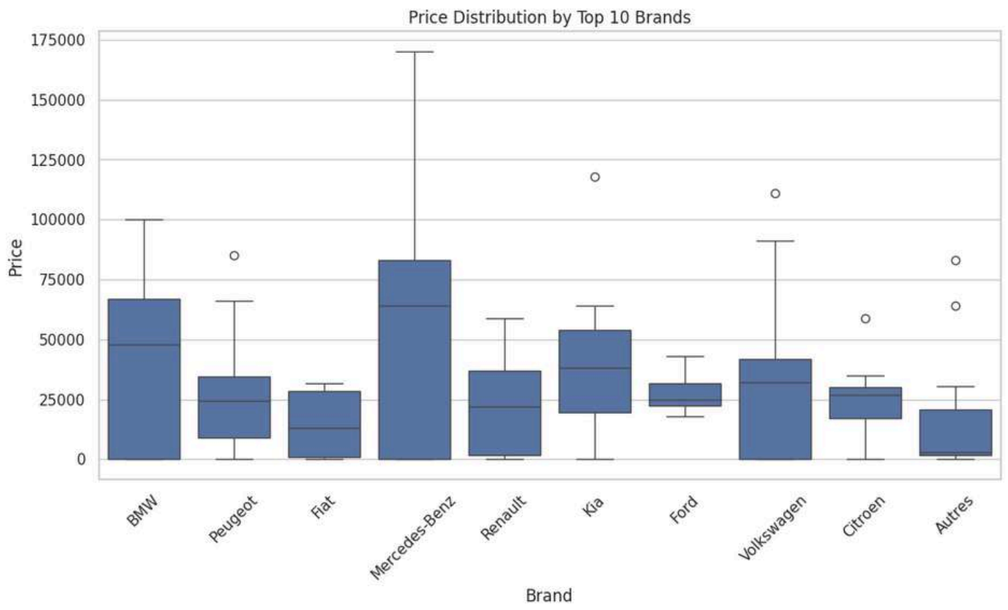

Distribution of Vehicle Prices
Frequency
Price (DT)
0
10
20
30
40
50
60
70
80
0
100000
200000
300000
400000
500000

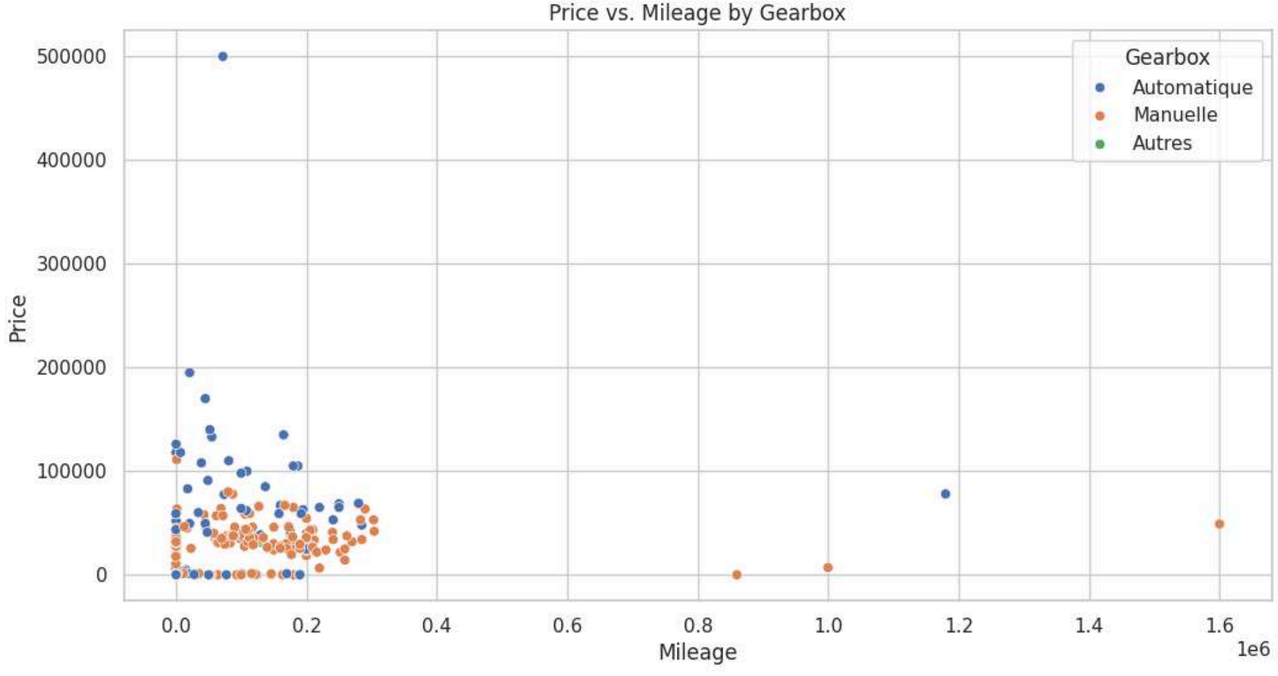

Price by Gearbox Type
Price
Gearbox
0
100000
200000
300000
400000
500000
Automatique
Manuelle
Autres

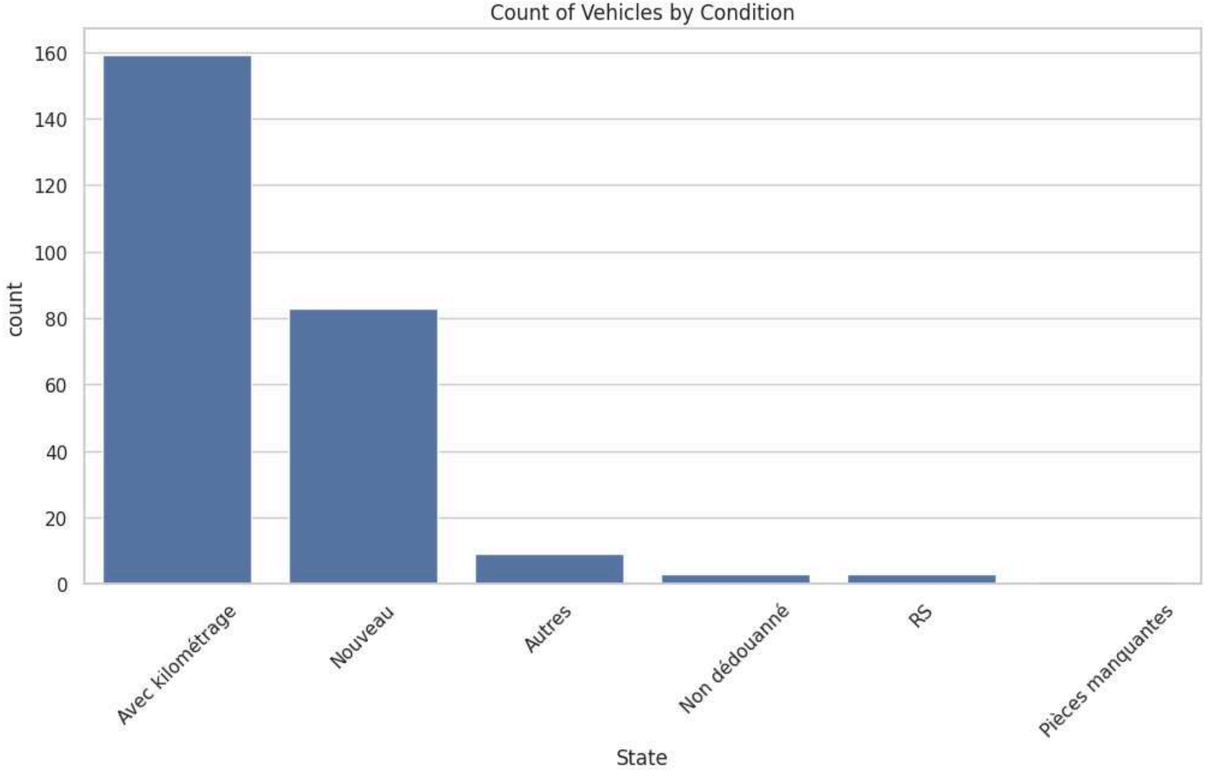

Price vs. Year by Fuel Type
Fuel
Essence
Diesel
Hybride essence
Hybride Diesel
Price
Year
0
100000
200000
300000
400000
500000
1995
2000
2005
2010
2015
2020
2025


Price by Fuel Type
Price
Fuel
0
100000
200000
300000
400000
500000
Essence
Diesel
Hybride essence
Hybride Diesel


Top 10 Vehicle Colors
Noir
Blanc
Gris
Bleu
Rouge
Beige
Marron
Autre
Argent
Camel
Color
0
10
20
30
40
50
60
70
count


Price Distribution by Top 10 Brands
175000
150000
125000
100000
75000
50000
25000
0
Price
BMW
Peugeot
Fiat
Mercedes-Benz
Renault
Kia
Ford
Volkswagen
Citroen
Autres
Brand


Price vs. Mileage by Gearbox
Gearbox
Automatique
Manuelle
Autres
500000
400000
300000
200000
100000
0
Price
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1e6
Mileage


Count of Vehicles by Condition
160
140
120
100
80
60
40
20
0
count
Avec kilométrage
Nouveau
Autres
Non dédouanné
RS
Pièces manquantes
State

Team

Oumayma Abayed

Junior Student

Eslem Sebri

Junior Student