Vietnamese Car Data Analysis

by Tran Viet Anh

Table of contents

01
Data and Project overview

02
Data cleaning process

03
Statistical infomation

04
Grafical representation of insights

05
Top collection

06
Conclusion

Data description

- Source dataset about buying and selling used cars in the Vietnamese market at the beginning of 2023.

- Original dataset consists of 30652 rows and 21 columns
- Data was downloaded from **Kaggle.com**
Collection method – scraping.

Link:
https://www.kaggle.com/datasets/nguynthanhlun/vietnamese-car-price

- Fields contain information about: **ad_id**, **origin**, **condition**, **car_model**, **mileage**, **exterior_color**, **interior_color**, **num_of_doors**, **seating_capacity**, **engine**, **fuel_system**, **transmission**, **drive_type**, **fuel_consumption**, **describe**, **brand**, **grade**, **year_of_manufacture**, **car_name**, **price.price** and **url**.

Project objectives

Cleaning

Our project will start by revealing gaps and inaccuracies in the data that can interfere with the analysis and then eliminating them.

Data Analysis

Use Pandas, NumPy and other Python libraries for investigating our data and searching for interesting information and insights.

Statistical

Using Pandas functions to obtain statistical information about data and identify relationships between variables.

Discovering insights

Discovery of interesting points that are worth paying attention to and on which it is possible to conduct further research.
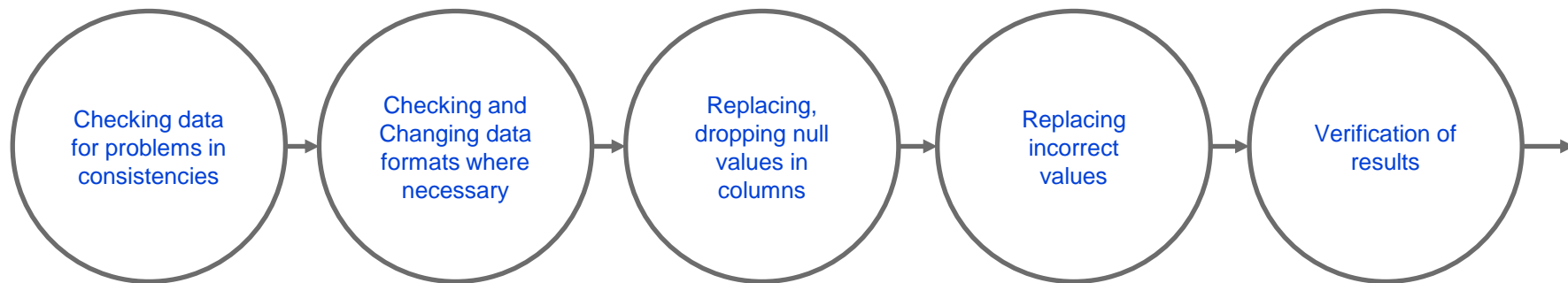
Graphical representation

Using the Python Matplotlib library to represent findings using various types of visualization

Data cleaning

Different Pandas functions were used to **clean** and **prepare** the data for **analysis**. Process followed the **scheme**:

Checking data for problems in consistencies → Checking and Changing data formats where necessary → Replacing, dropping null values in columns → Replacing incorrect values → Verification of results

More details can be found in my **Github** portfolio and the attached **Jupyter** notebook

Dataset stats

Correlation coefficients

|	car_price	n_mileage	year_of_manufacture
car_price	1.00	0.00	0.22
n_mileage	0.00	1.00	-0.01
year_of_manufacture	0.22	-0.01	1.00

- **Car_price** has a strong *positive* correlation with **year_of_manufacture**.
- **Mileage** was *inversely* related to **year_of_manufacture**.

Descriptive statistics

|	car_price	n_mileage	year_of_manufacture
count	30267.00	30267.00	30267.00
mean	1174.57	411464.68	2017.34
std	2070.96	35110338.85	5.29
min	18.00	0.00	1990.00
25%	410.00	0.00	2015.00
50%	620.00	20000.00	2019.00
75%	1015.00	60000.00	2022.00
max	54000.00	4294967295.00	2023.00

Transmission



Transmission type ratio
Manual
Semi-Automatic
car_model
17.8%
0.1%
82.2%
Automatic
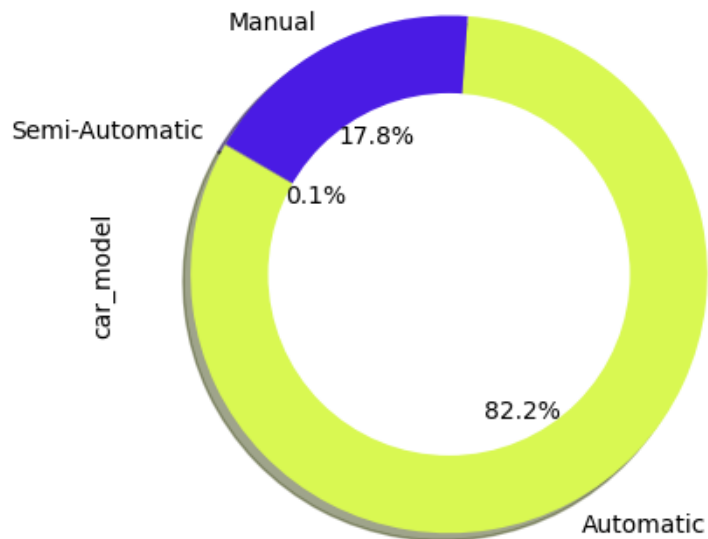
- Most of the offers have **Automatic** transmission type – **25.057** cars.
- **5.377** cars on the market have **Manual** transmission type
- Only **17** offers have **Semi-Automatic** transmission type

Fuel types

- The most of the cars – **24.577** use **Petrol** fuel type.

- **Diesel – 5.323**

- **Electric** and **Hybrid** vehicles are the least represented – **1%** and **0.8%** of the total number (**297** and **254** vehicles)
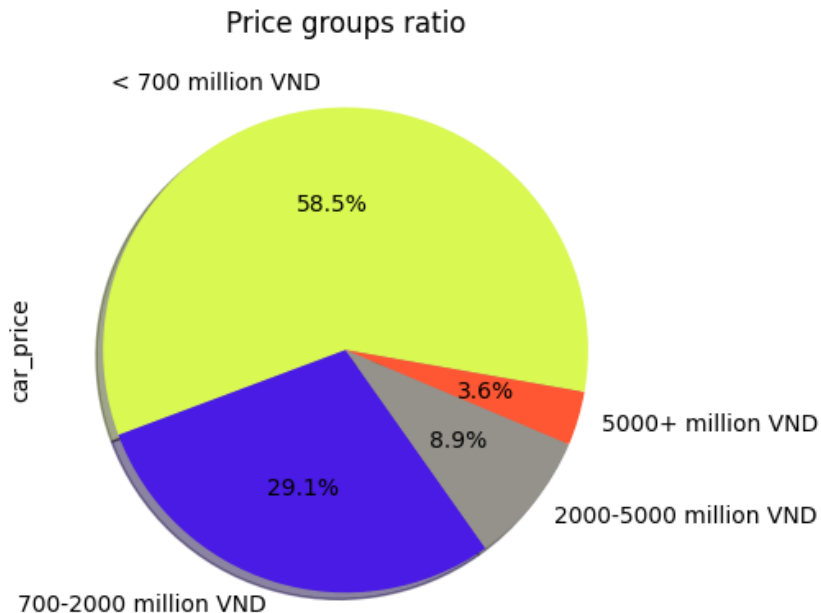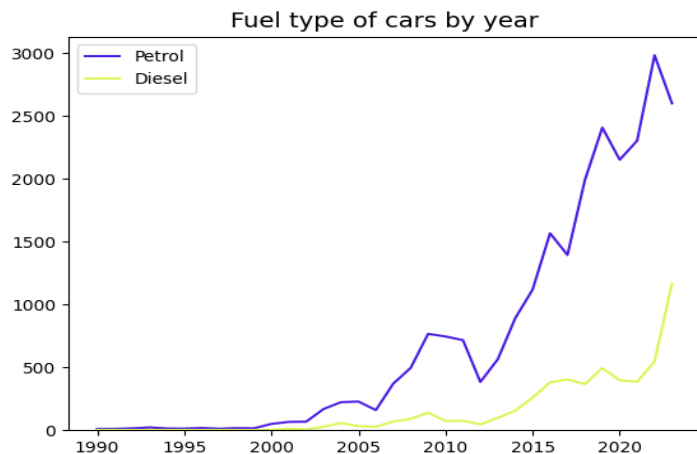

Fuel types ratio

New or used?

Condition Bars
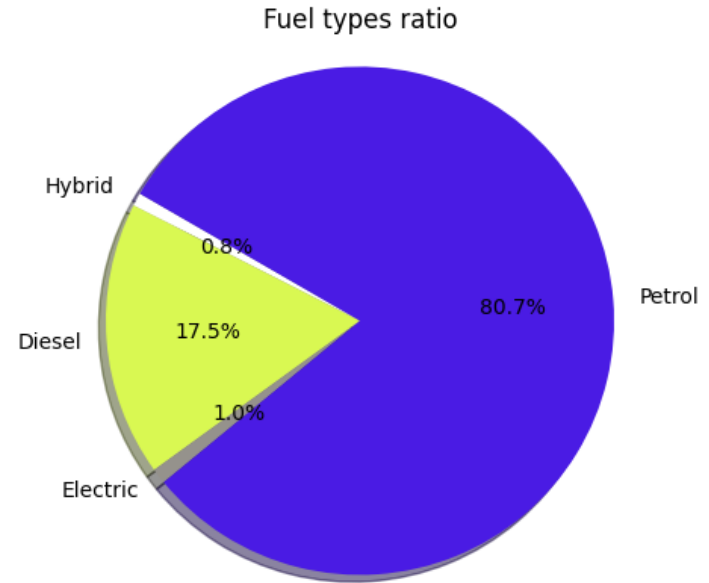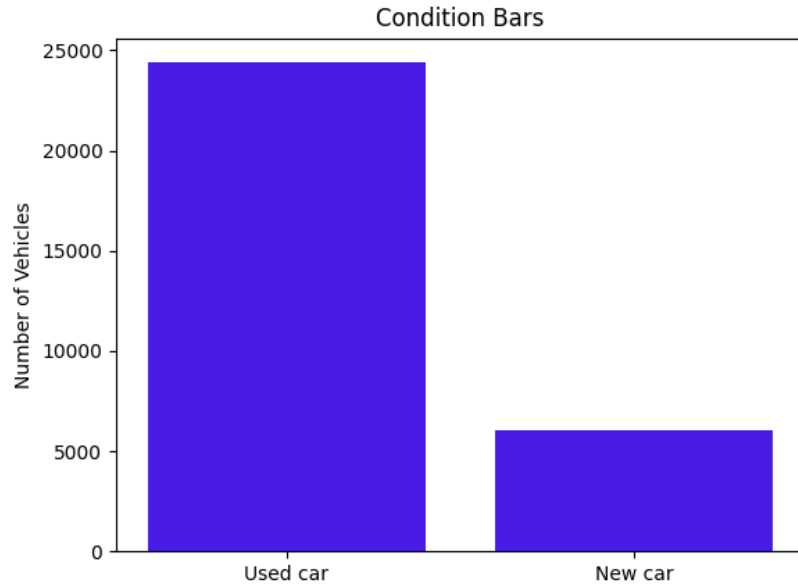
- Most of the vehicles (**24,391**) represented on the market are used.

- **6060** of the cars are new.

Dataset stats

- The **most** of cars – **17,700** vehicles are cheaper than **700** million VND

- **87.6%**, or **26.502** offers, have price less than **2.000** million VND, and **96.6%** or **29.183** offers are priced under **5.000** million VND

- **1.084** vehicles cost more than **5.000** million VND, with the most expensive costing **54.000** million VND.

- The cheapest car priced at **18** million VND

Price groups ratio

< 700 million VND
58.5%
3.6%
5000+ million VND
8.9%
2000-5000 million VND
29.1%
700-2000 million VND
car_price

Fuel types by year

Fuel type of cars by year

Fuel type of cars by year

The number of Hybrid cars tends to be flat with no increase, a significant decrease from **2005** to **2019** and a steady increase from **2019** to **2023**.
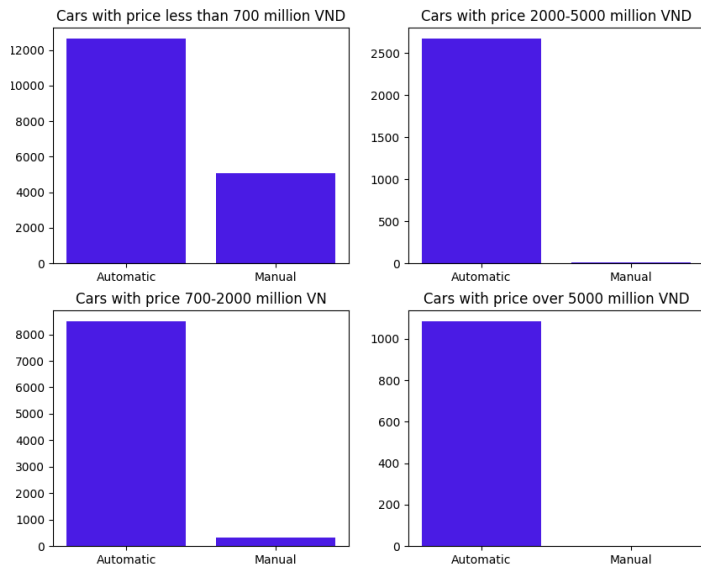
Electric cars have been scarce since the years before **2021**, but from **2021** onwards electric cars have skyrocketed to nearly **25 times** by **2023**.
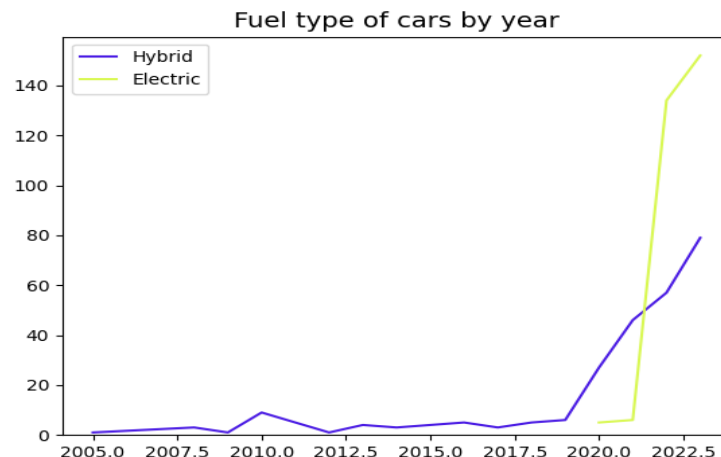
Petrol had many short-term downward fluctuations in **2012**, **2017** and **2019**, but the general trend was upward.

Diesel has many fluctuations and increases slightly until **2021** and increases sharply from **2021** to **2023**, until **2023**.
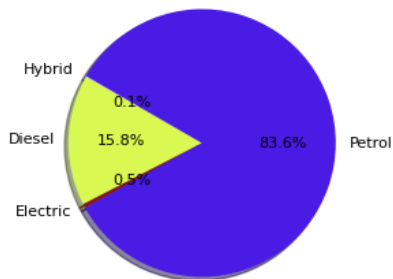
Transmission stats

Cars with price less than 700 million VND

Cars with price 2000-5000 million VND

Cars with price 700-2000 million VN

Cars with price over 5000 million VND

Transmission type by production year

The quantity of vehicles with **Manual** types tends to decrease with growing prices in each price group. **Manual** cars appear mostly at prices below **700 million VND**
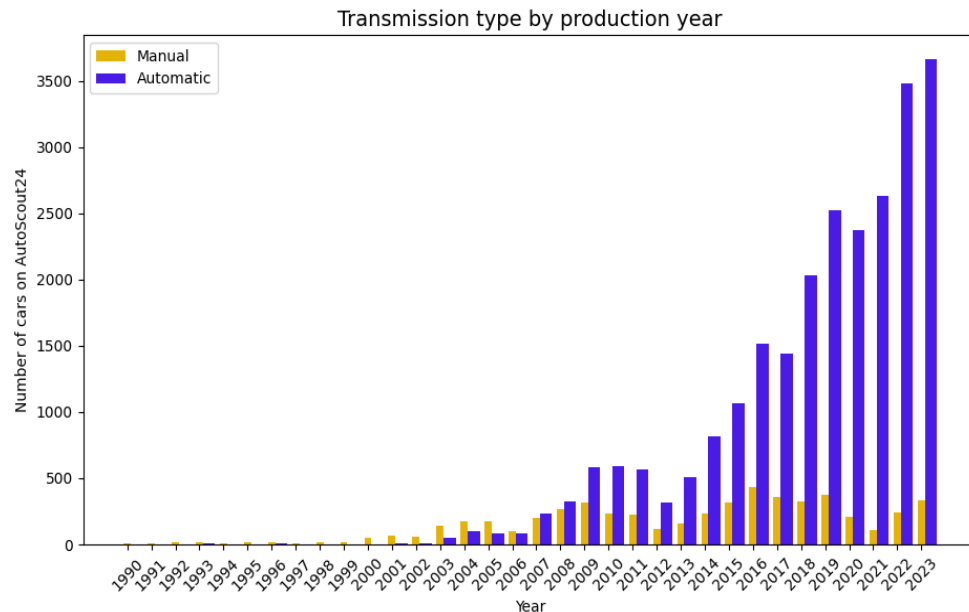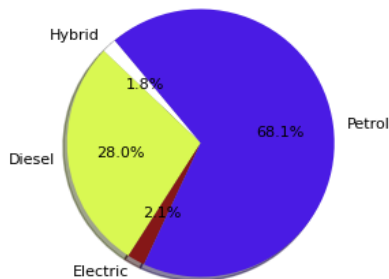
In **2006** and later, the number of cars represented with **Automatic** has tended to increase, and in **2007**, it surpassed the number of cars with **Manual** transmission.
By **2023**, **Automatic** cars will be 9 times that of **Manual** cars
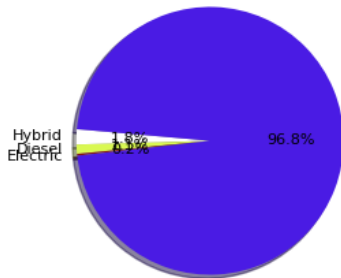
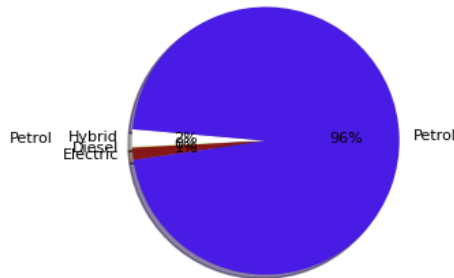# Fuel types in price categories ratio

[1] Price less than 700 million VND

Hybrid 0.1%
Diesel 15.8%
Electric 0.5%
Petrol 83.6%

[2] Price 700-2000 million VND

Hybrid 1.8%
Diesel 28.0%
Electric 2.1%
Petrol 68.1%

[3] Price 2000-5000 million VND

Hybrid 1.8%
Diesel
Electric 0.1%
Petrol 96.8%

[4] Price over 5000 million VND

Hybrid
Diesel 4%
Electric 1%
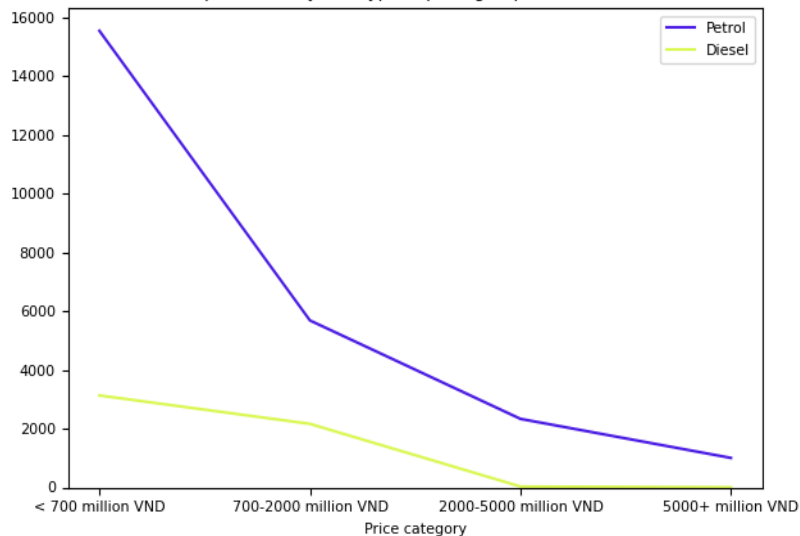Petrol 96%

- **Petrol** cars account for a large proportion **(83.6%)** in the group of cars priced less than **700** million VND. This rate decreased **(16.5%)** in the group of cars priced at **700 - 2000** million VND.

- In the **2000-5000** million VND car group and the over **5000** million VND car group, **Petrol** cars account for almost the entire market share, **96.8%** and **96%** respectively.

- The second most represented trend in the group with prices lower than **2000** million VND is Diesel. At prices more than **2000** million VND, **Hybrid** and **Electric** have higher representation than **Diesel**

Real price categories by fuel types

Cars represented by fuel type in price group (Petrol and Diesel)

Cars represented by fuel type in price group (Hybrid and Electric)

The two types of vehicles with the largest number of vehicles, Petrol and Diesel, both tend to be less popular, while prices increase
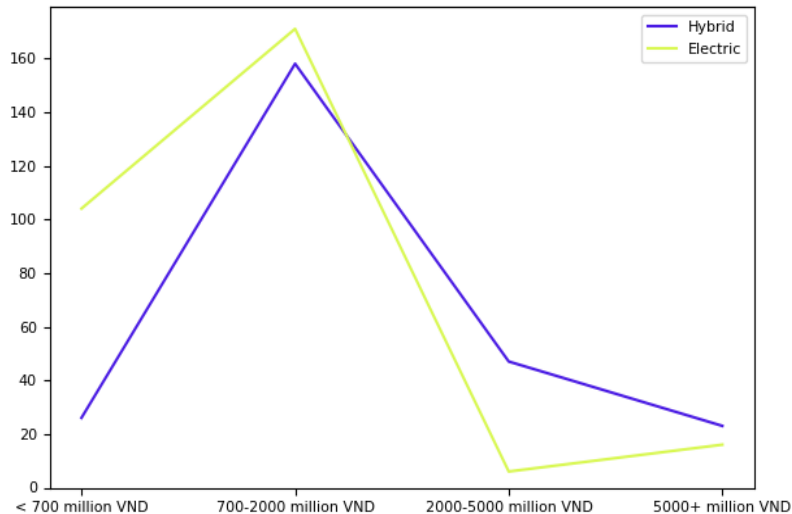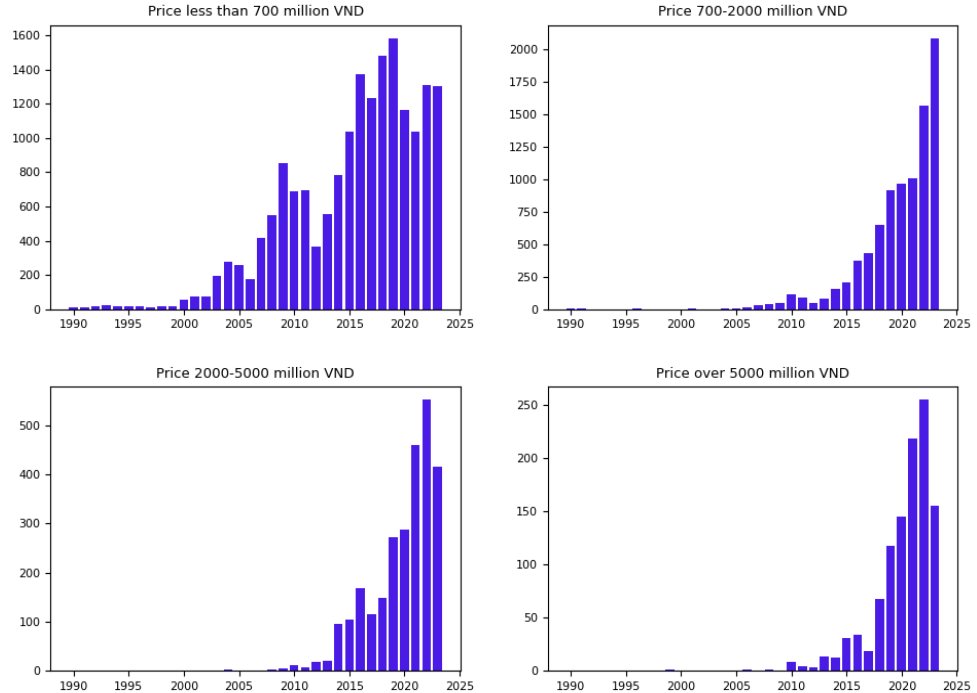
The number of Hybrid and Electric vehicles **increases** in the first two price categories while decreasing in the last two price

Price category trends

Number of cars at the price-bin level each year

In all four segments, there is a tendency to sell more as its year of production increases and there is a decline in the years 2012-2013

Top 10 popular brands

Brand	Average price (million VND)	Cars represented	Average mileage
Toyota	812.75	5544.00	116475.61
Hyundai	585.50	3531.00	281788.19
Ford	787.21	3334.00	103319.84
Mercedes Benz	2654.76	3123.00	40666.76
Kia	505.91	2961.00	1642254.92
Mazda	596.18	2220.00	93872.28
Mitsubishi	562.63	1378.00	118118.08
Honda	595.51	1220.00	59124.23
Lexus	3870.37	929.00	40133.98
VinFast	678.58	790.00	70907.65

Top 10 popular Grades

Brand Grade	Average price (million VND)	Grade represented	Average mileage
Ford Ranger	680.08	1424.00	148695.84
Toyota Vios	403.87	1011.00	109362.47
Mazda 3	532.88	837.00	162392.82
Kia Morning	241.63	782.00	5754189.87
Mercedes Benz GLC	1807.72	759.00	21721.53
Toyota Camry	734.57	753.00	81012.25
Toyota Innova	429.97	727.00	140771.54
Toyota Fortuner	766.73	722.00	349352.41
Ford Everest	964.99	691.00	75915.27
Mercedes Benz C class	1132.68	672.00	62904.51

Top 10 expensive cars

Brand	Model	Price (million VND)	Mileage
Ferrari SF90 Stradale	Coupe	54000.00	0
Rolls Royce Cullinan	SUV	46000.00	0
Rolls Royce Cullinan	SUV	45000.00	0
Rolls Royce Cullinan	SUV	45000.00	0
Rolls Royce Ghost	Sedan	39999.00	6200
Rolls Royce Cullinan	SUV	39990.00	0
Rolls Royce Cullinan	SUV	39900.00	7000
Rolls Royce Ghost	Sedan	39000.00	5000
Rolls Royce Ghost	Sedan	39000.00	5000
Rolls Royce Cullinan	SUV	39000.00	0

Top 10 cheapest cars

Brand Grade	Price	Model	Mileage
Kia Pride	18.00	Sedan	0
Mitsubishi L300	21.00	Van/Minivan	0
Mitsubishi L300	25.00	Van/Minivan	30000
Daewoo Matiz	25.00	Hatchback	0
Nissan Sunny	25.00	Sedan	125000
Kia Pride	27.00	Sedan	0
Mitsubishi Canter	28.00	Sedan	5000000
Toyota Hiace	28.00	Van/Minivan	2563
Toyota Hiace	28.00	Van/Minivan	400000
Daewoo Matiz	28.00	Hatchback	0

Conclusion

In this project, an analysis of data on Vietnamese cars was carried out. Relationships, interesting trends between variables were identifield

An anomaly was discovered in **2017** onwards: **Electronic** cars are sold in significantly increased numbers, while **Hybrid** cars are still the most widely available. Also this year and after, there is a downward trend in the number of cars with a **Manual** transmission and an increase in the number of **Automatic** ones

The analysis was done using Python in Jupyter notebooks. See my GitHub portfolio: github.com/VietAnh2402

Thanks You
for attention!