

FINAL PROJECT

CAR PRICE PREDICTION

VEBRIANTO Z. A

TABLE OF CONTENTS

- O1** DATA PREPARATION
- O2** EXPLORATORY DATA ANALYSIS (EDA)
- O3** MODELLING AND EVALUATION
- O4** CONCLUSION AND RECOMMENDATION

DATA PREPARATION



DATASET

- Terdapat 18 kolom dan 19237 baris
- Data Description
- ID : Primary key
- Price : Harga mobil (Kolom Target)
- Levy : Pajak impor dan ekspor mobil
- Manufacturer : Perusahaan yang membuat mobil
- Model : Tipe mobil
- Prod. year : Tahun pembuatan mobil
- Category : Jenis mobil
- Engine volume : Kapasitas mesin
- Mileage : Jarak yang sudah ditempuh
- Cylinders : Jumlah cilinder
- Gear box type : Jenis gear box
- Drive wheels : Sistem penggerak mobil
- Doors : Jumlah pintu
- Wheel : Letak stir kemudi
- Color : Warna mobil
- Airbags : Jumlah airbags
- Fuel type : Jenis bahan bakar
- Leather interior : Menggunakan kulit atau tidak

DATA CLEANING



Missing Values
0



Duplicated Values
313

Drop Duplicated

Before :

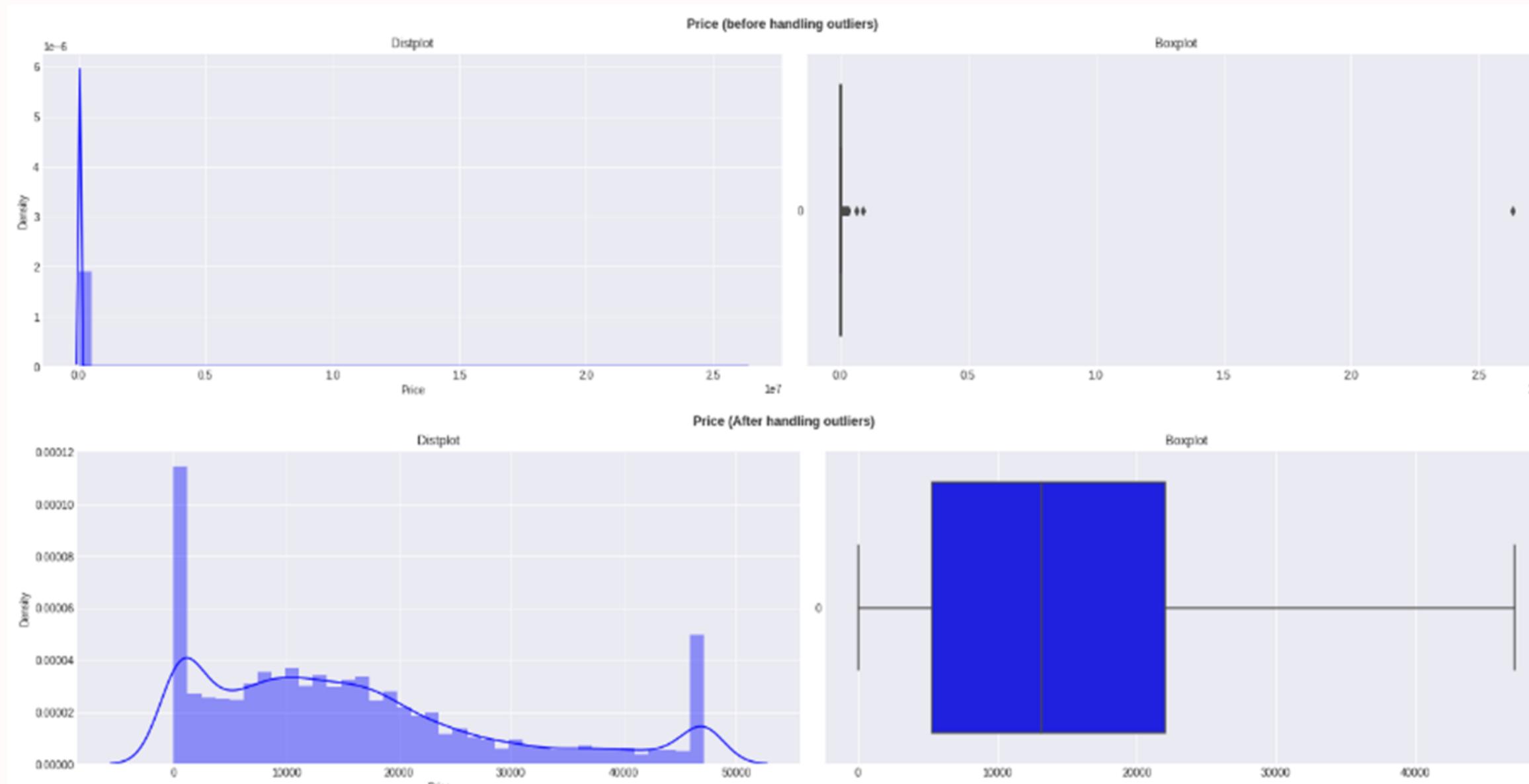
- 18 Kolom
- 19237 Baris

After :

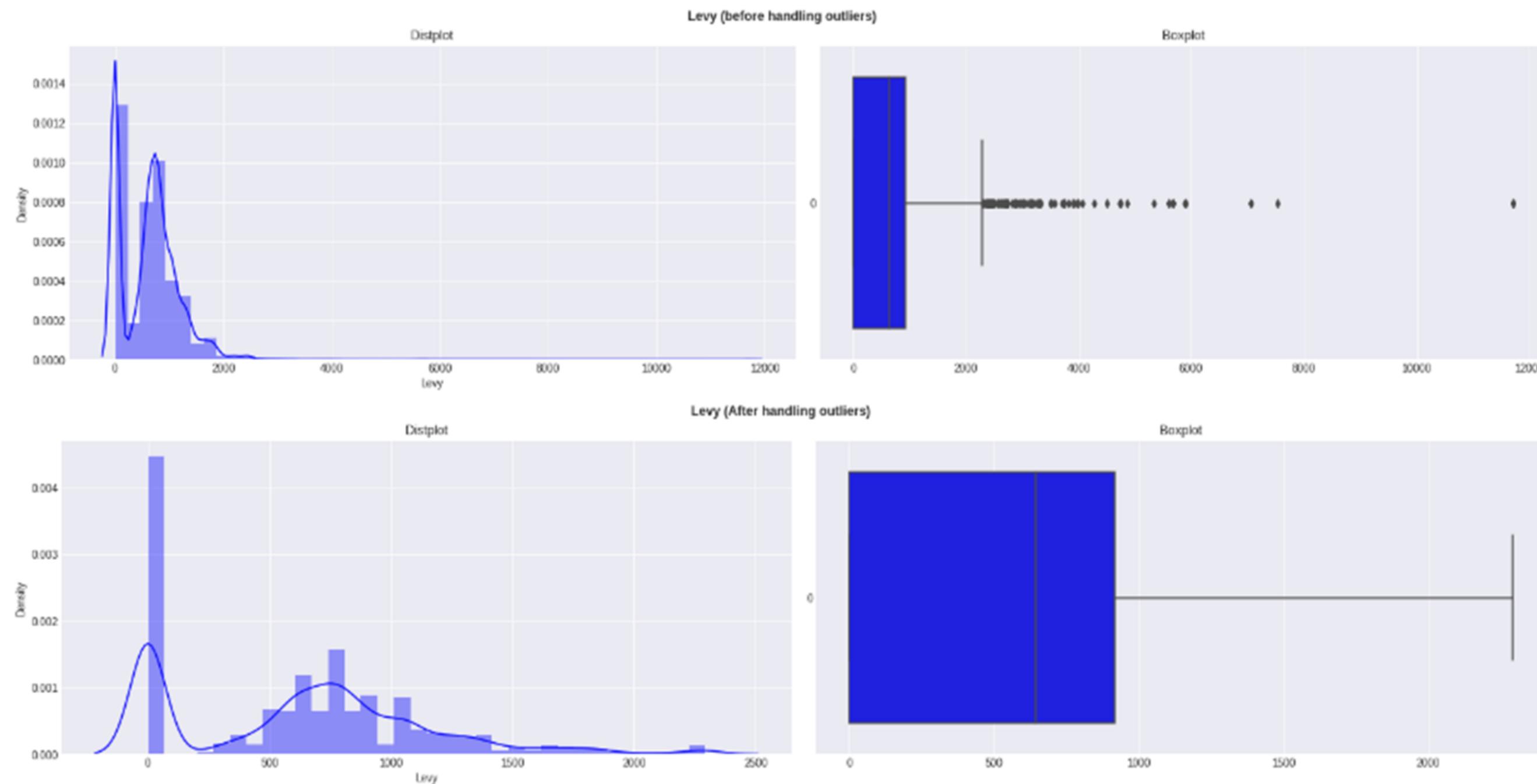
- 18 Kolom
- 18924 Baris

EXPLORATORY DATA ANALYSIS (EDA)

DETECT OUTLIER



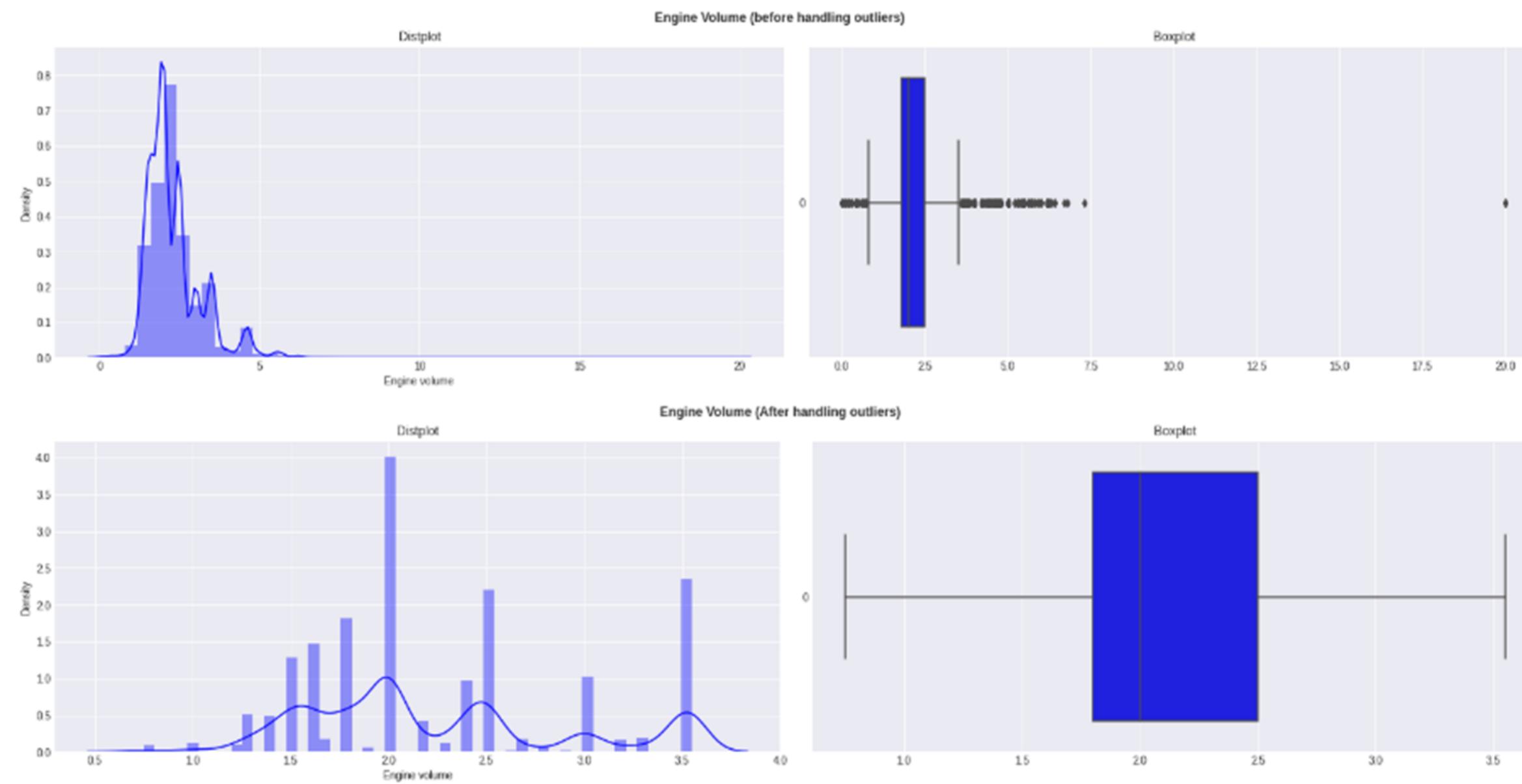
► Total outliers in Price : 1055 (5.57%)



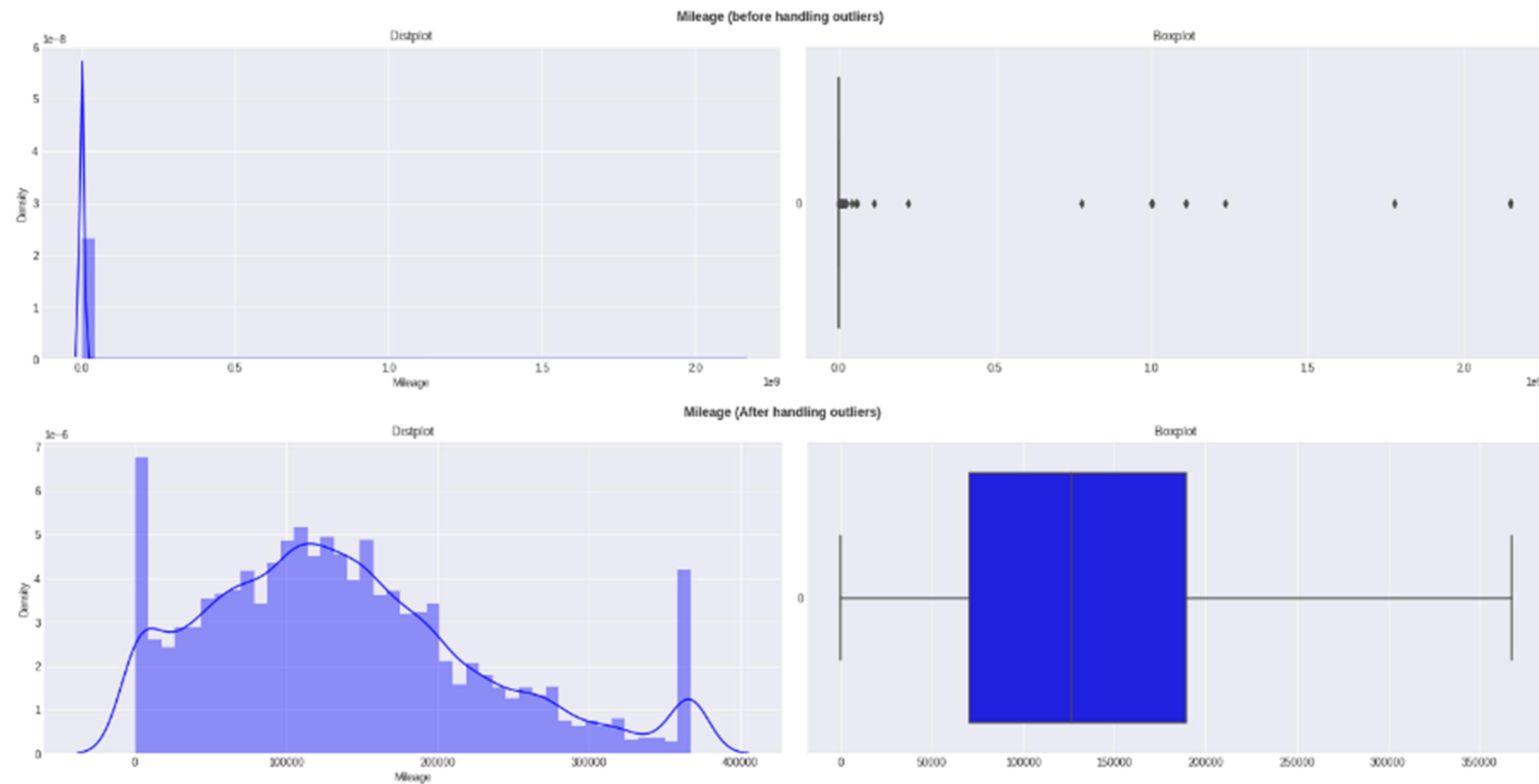
► Total outlier in Levy : 160 (0,85%)



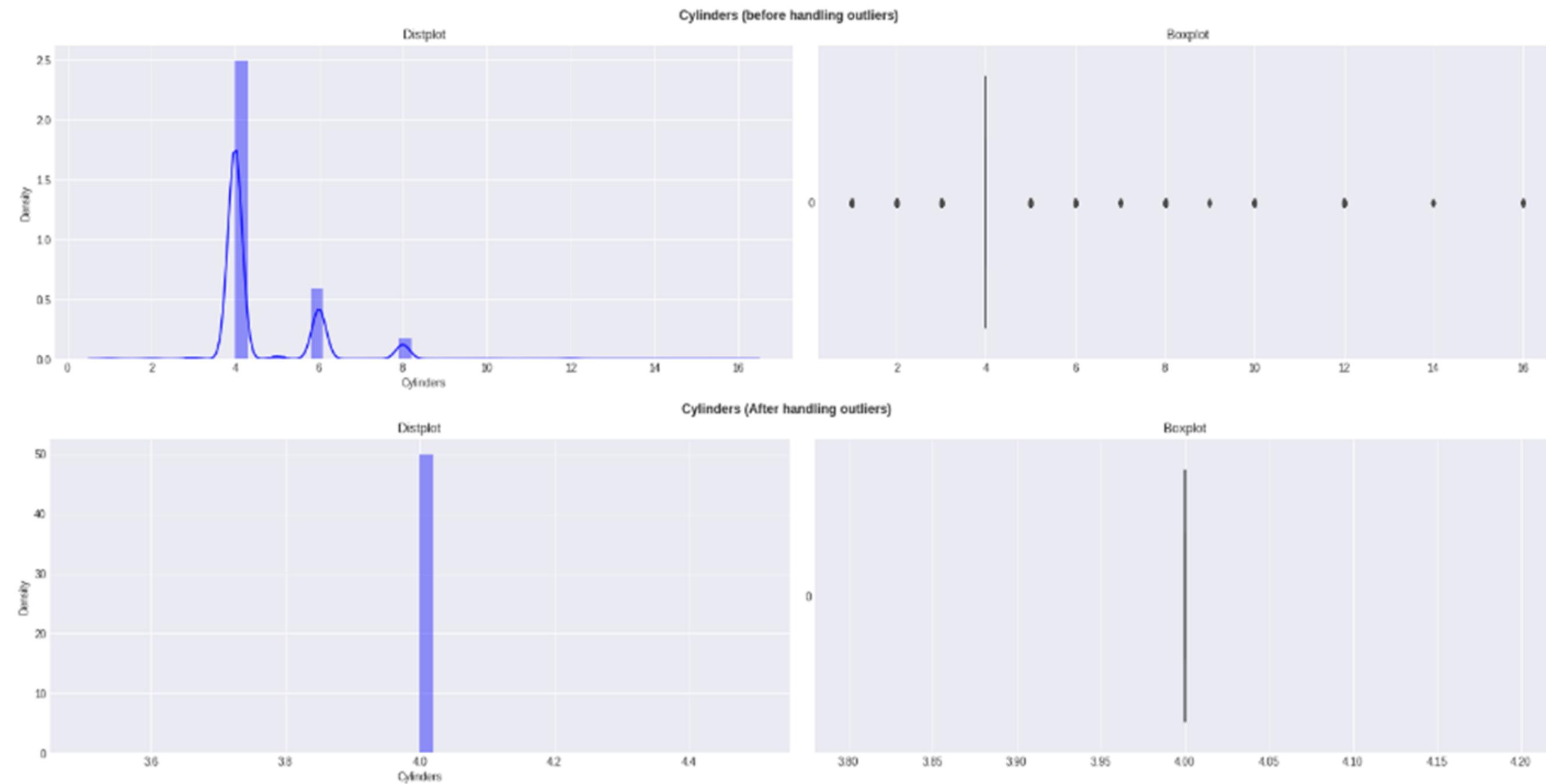
► Total outliers in Prod. Year : 962 (5.08%)



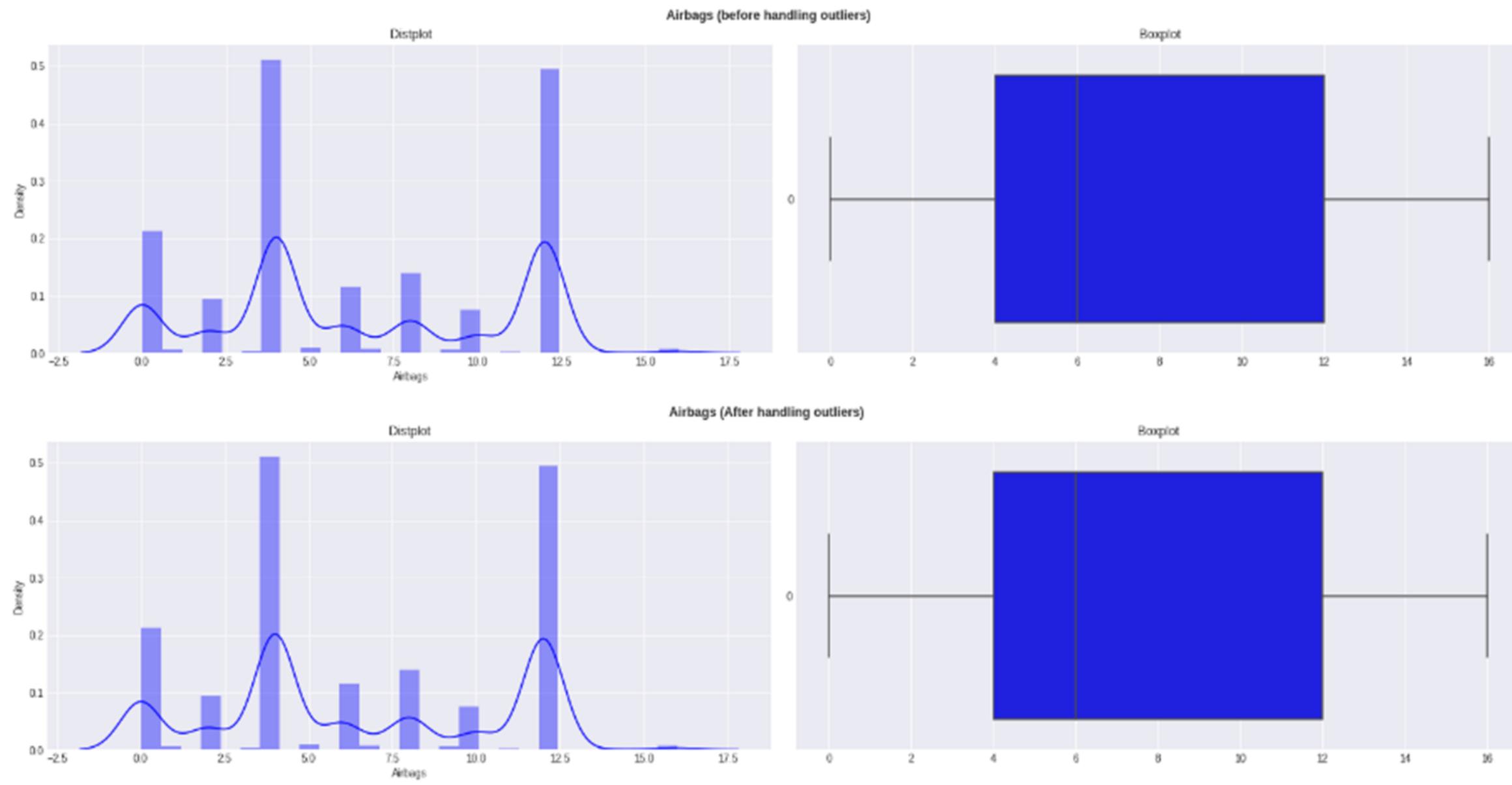
► Total outlier in Engine volume : 1358 (7,18%)



► Total outliers in Mileage : 635 (3,36%)

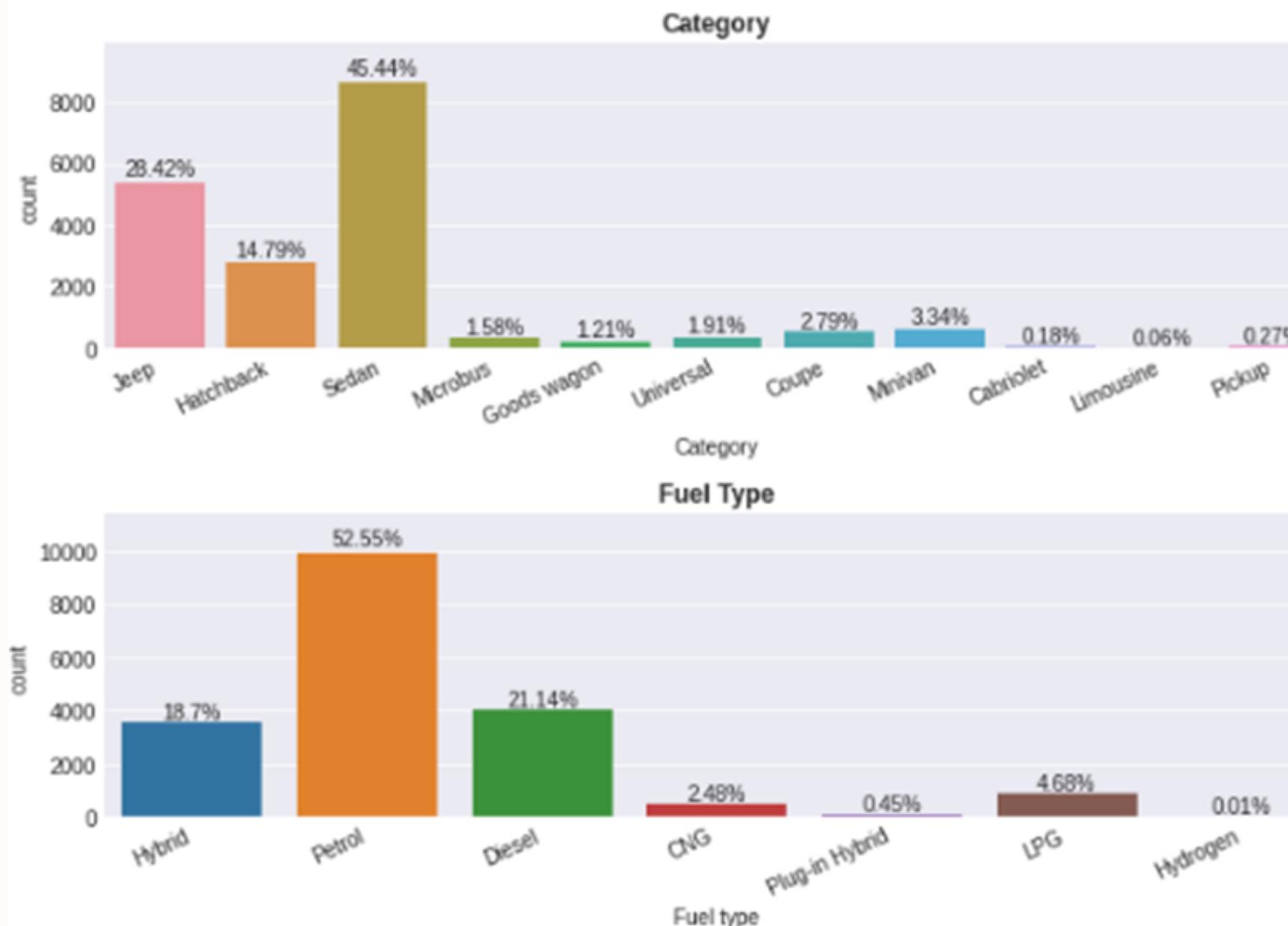


► Total outlier in Cylinders : 4765 (25,18%)



► Total outliers in Airbags : 0 (0,0%)

VISUALISASI CATEGORY DAN FUEL TYPE



Category

Pada kolom Category, sekitar 45.44% adalah sedan, 17.7% adalah hybrid, 14.79% adalah hatchback dan lainnya memiliki pangsa pasar dibawah 5%.

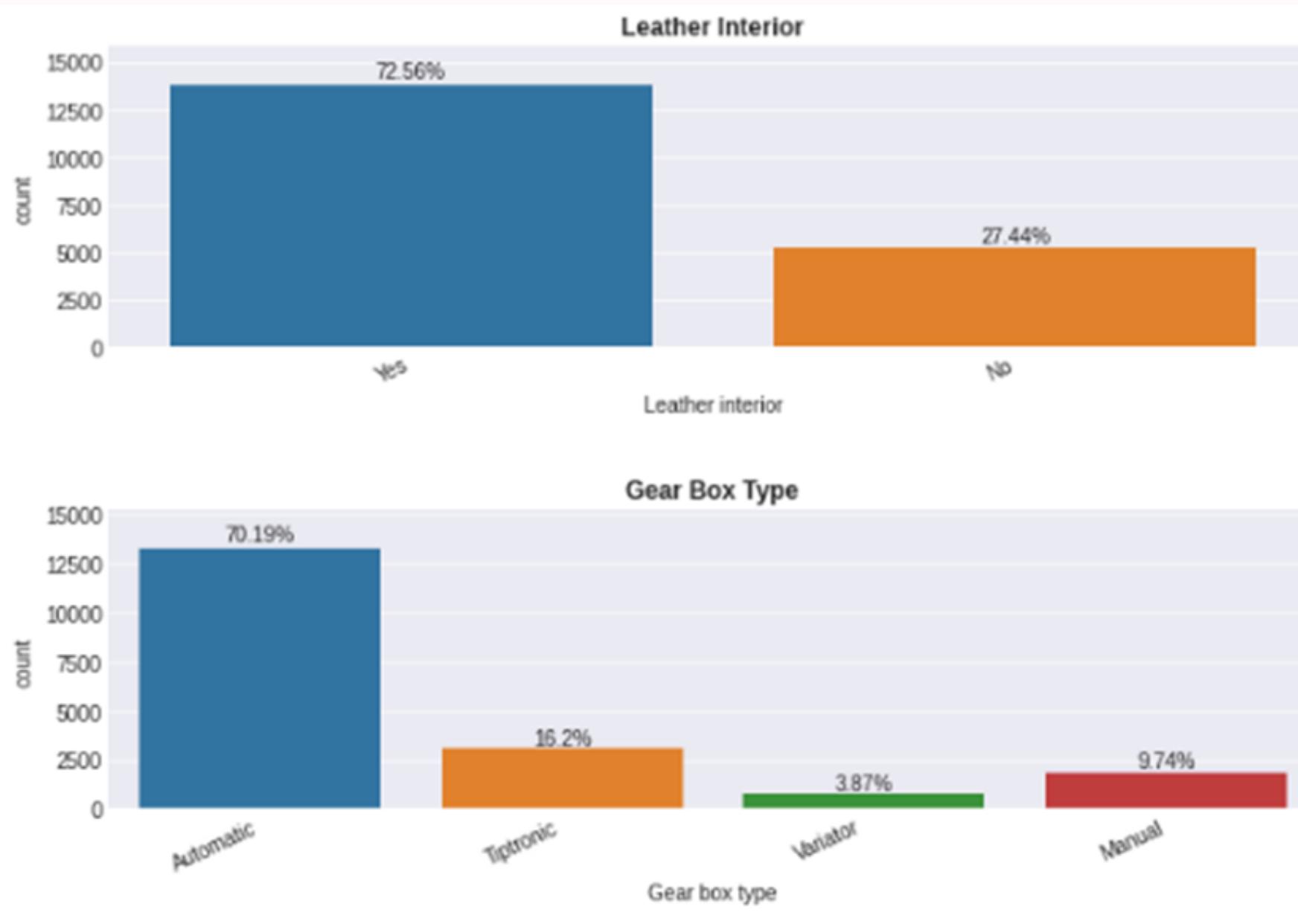


Fuel Type

Pada kolom Fueltype, 52.55% jenis bahan bakarnya di dominasi oleh petrol, 21.14% adalah diesel, 18.7% adalah hybrid dan jenis bahan bakar lain seperti CNG, LPG, dan Hidrogen, serta plug-in hybrid sangat rendah.



VISUALISASI LEATHER INTERIOR DAN GEAR BOX TYPE

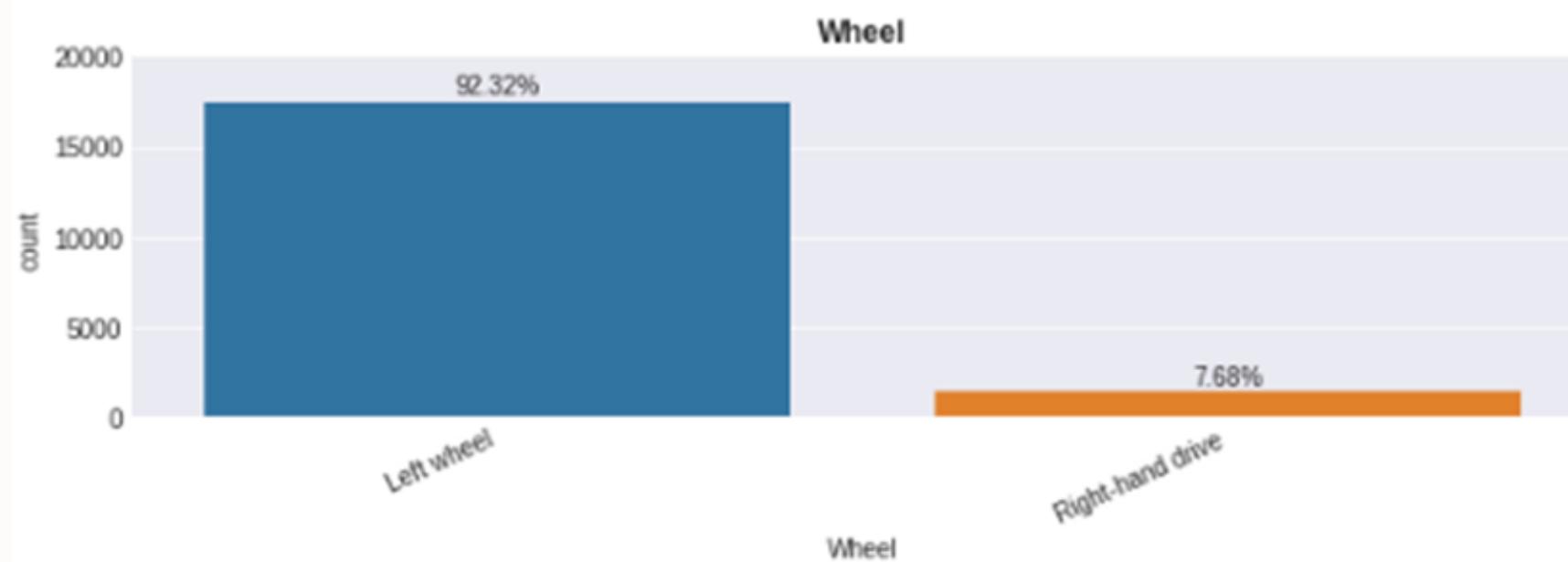
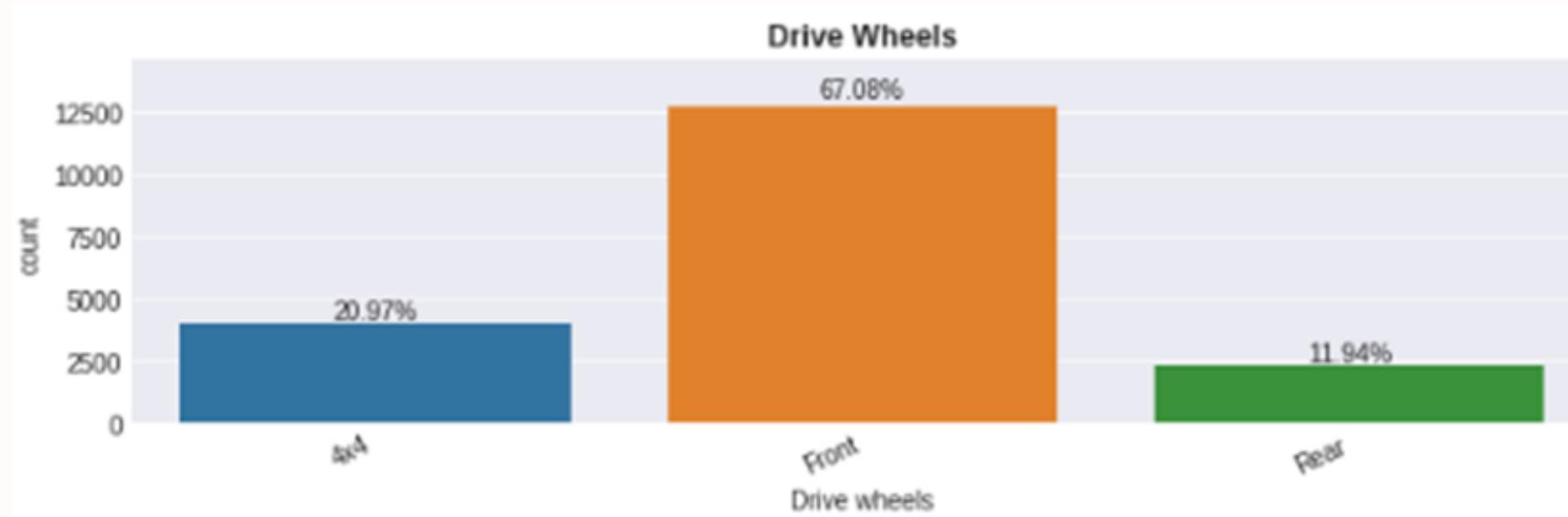


Leather Interior
72.56% varian mobil menggunakan kulit pada interiornya.



Gear Box Type
70.19% varian mobil menggunakan gear box type automatic.

VISUALISASI DRIVE WHEELS DAN WHEEL



Drive Wheels

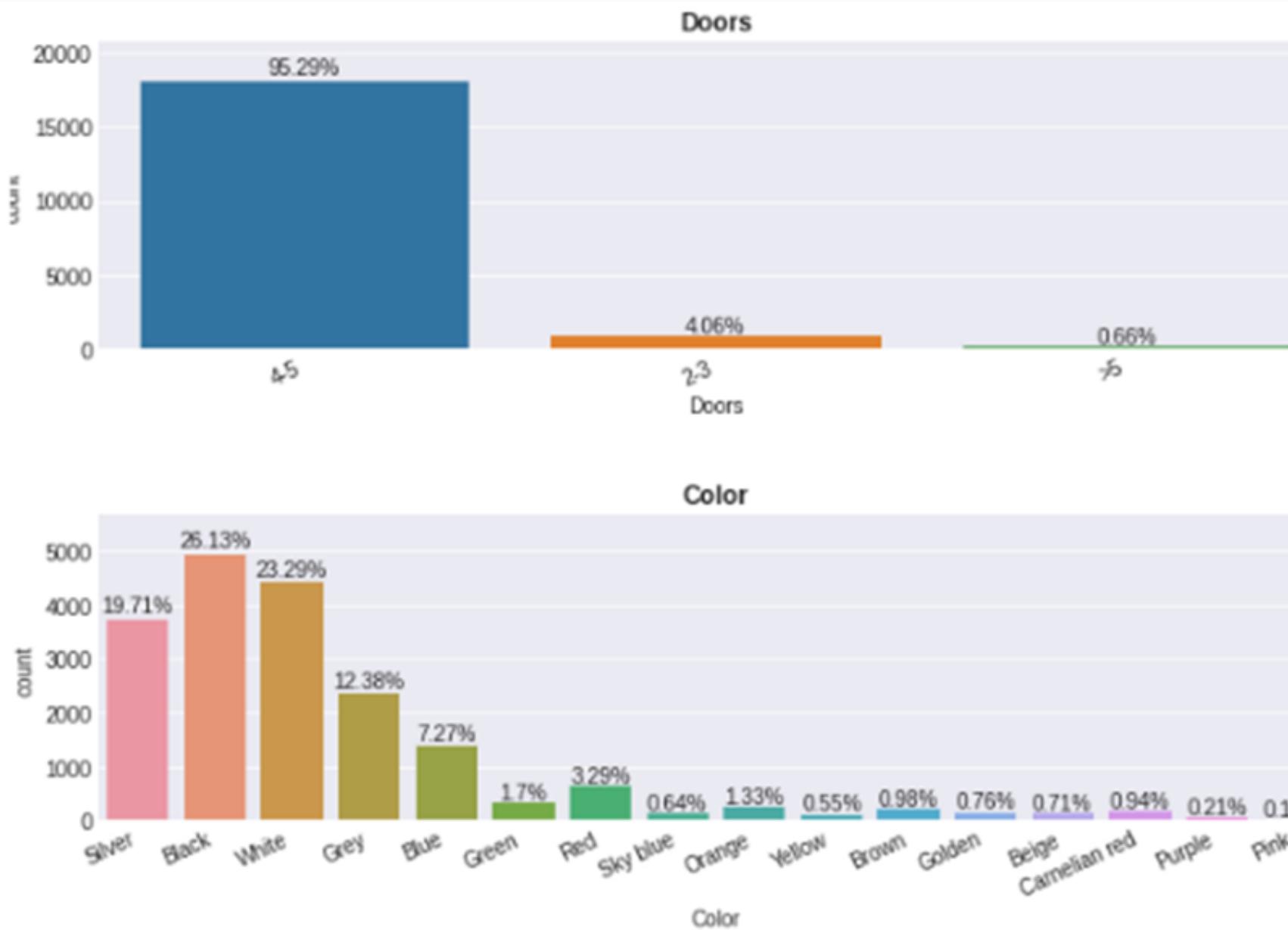
67.08% varian mobil menggunakan penggerak roda depan (Drive wheels).



Wheel

92.32% varian mobil merupakan setir kiri (Left wheel).

VISUALISASI DOORS DAN COLOR



Doors

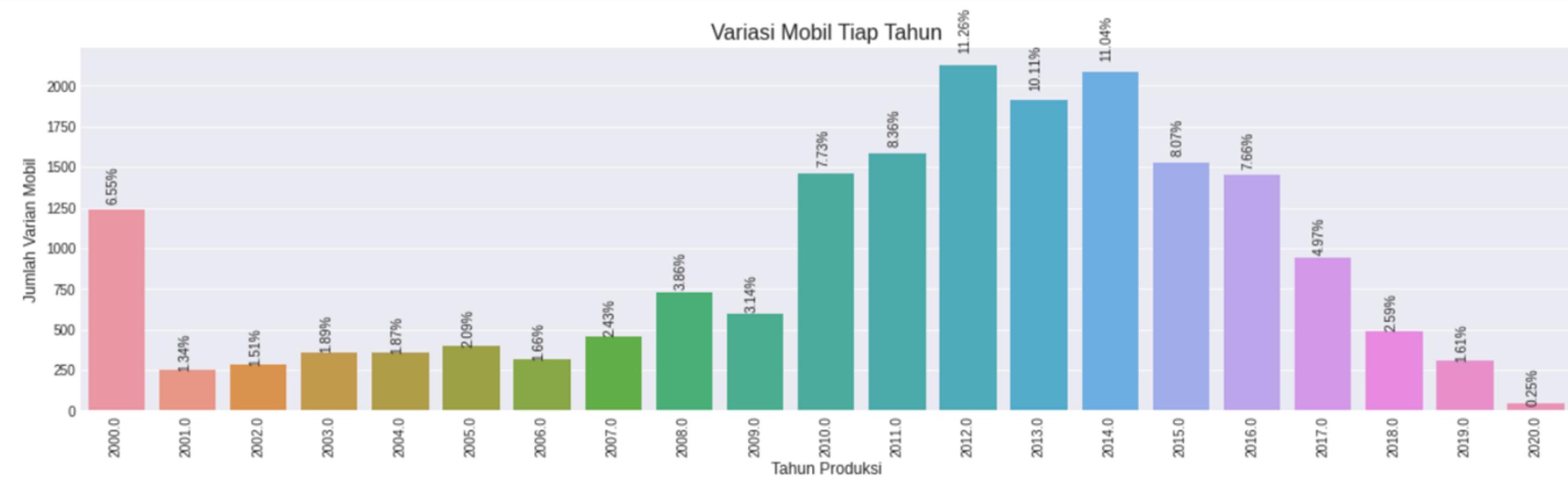
95.29% varian mobil memiliki 4-5 pintu.



Color

Varian mobil terbanyak menggunakan warna (color) black 26.13%, white 23.29% dan silver 19.71%

VISUALISASI JUMLAH PRODUKSI MOBIL TIAP TAHUN



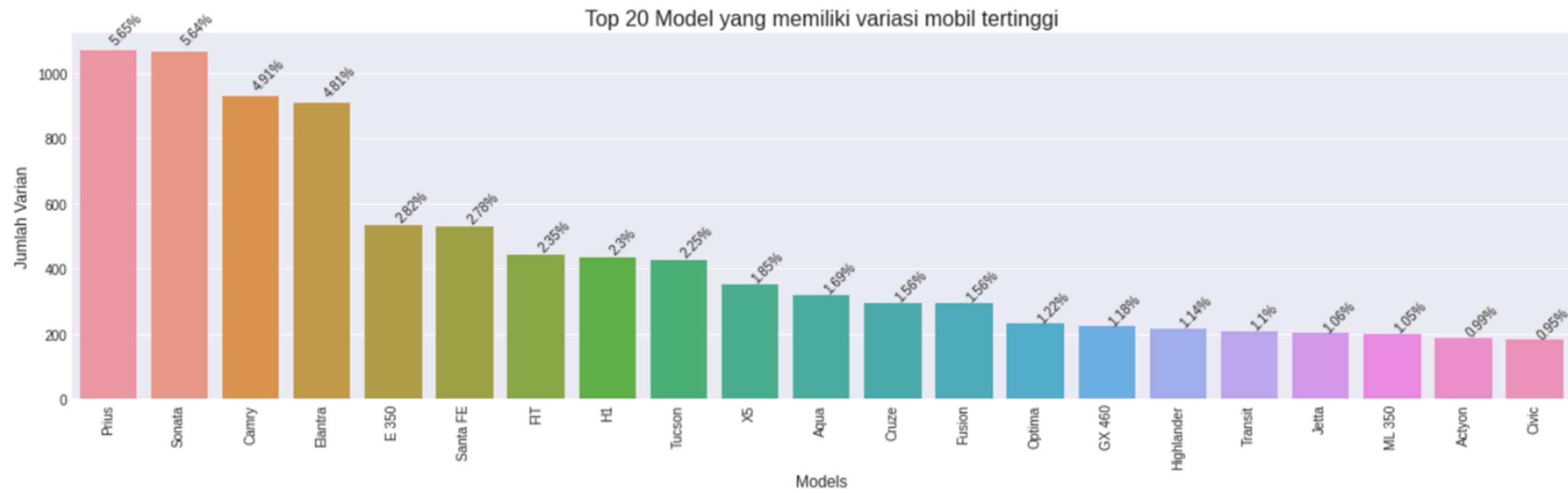
01

Jumlah varian mobil tiap tahun bertahap meningkat dan kemudian menurun tajam sejak tahun 2015

02

Jumlah varian mobil terbaik di tahun 2012 (11.26%), 2013 (10.11%) dan 2014 (11.04%).

VISUALISASI TOP 20 MODEL MOBIL



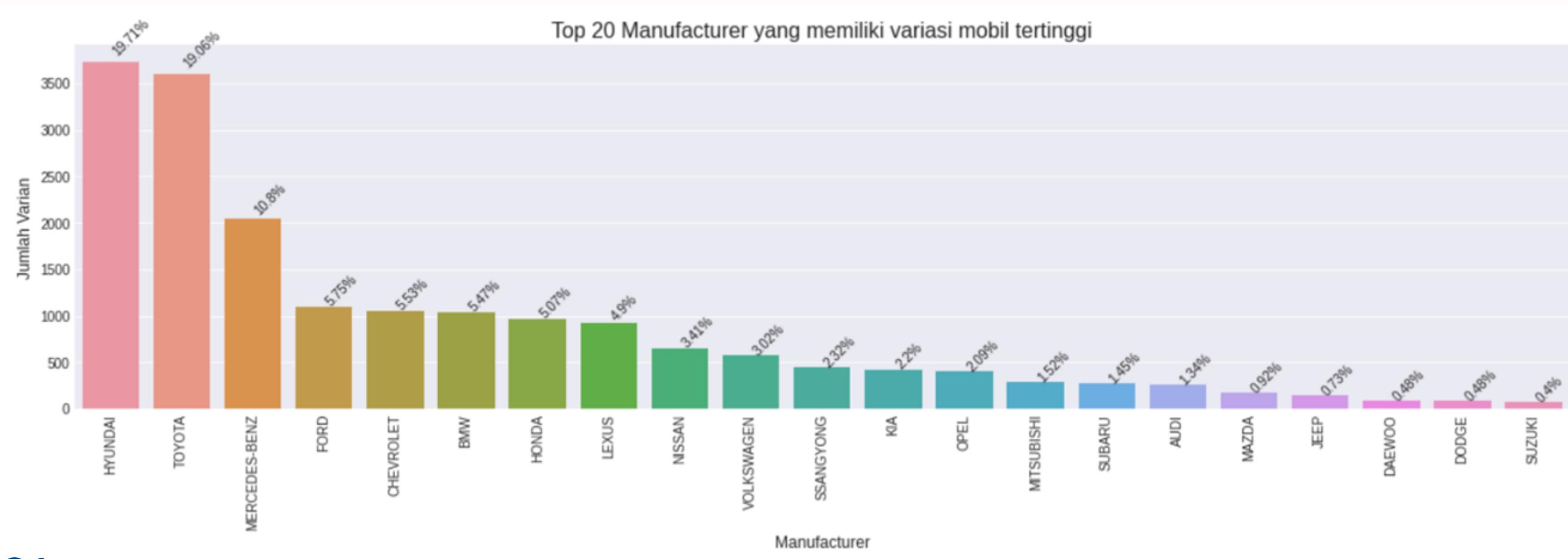
01

Model Prius, Sonata, Camry, Elantra memiliki varian mobil tertinggi.

02

Dari 1590 model, hanya dari keempatnya saja sudah menyumbang 21% varian mobil.

VISUALISASI TOP 20 MANUFACTURER

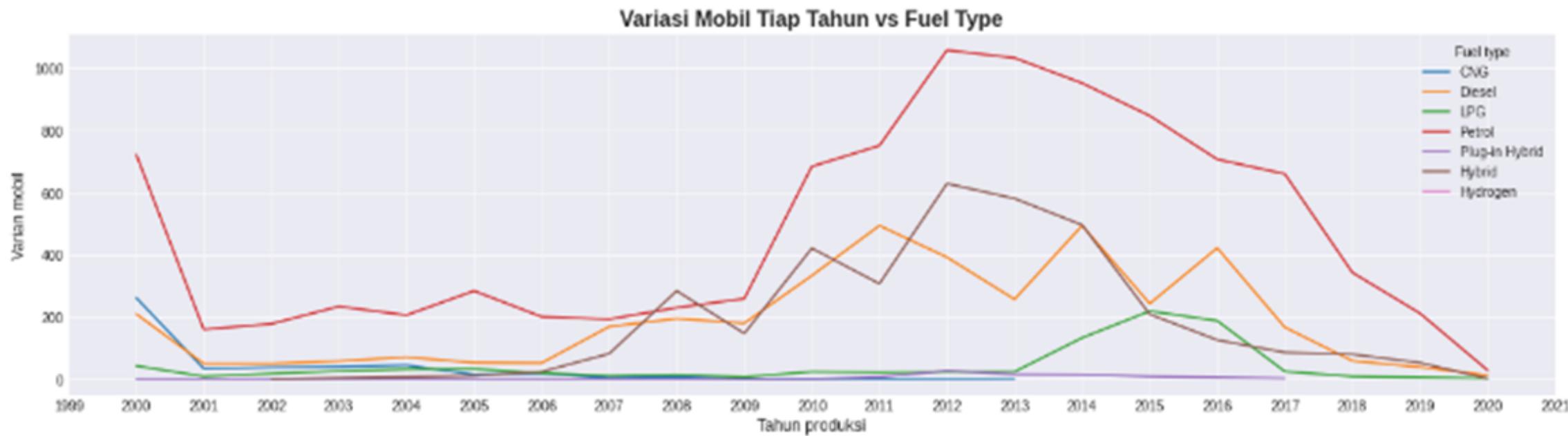
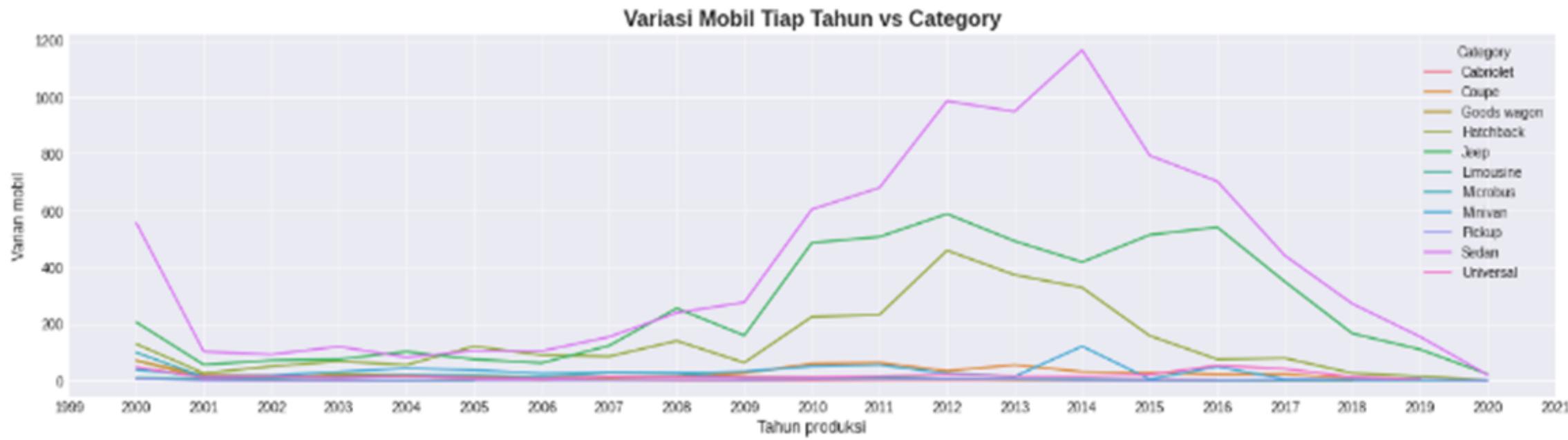


01

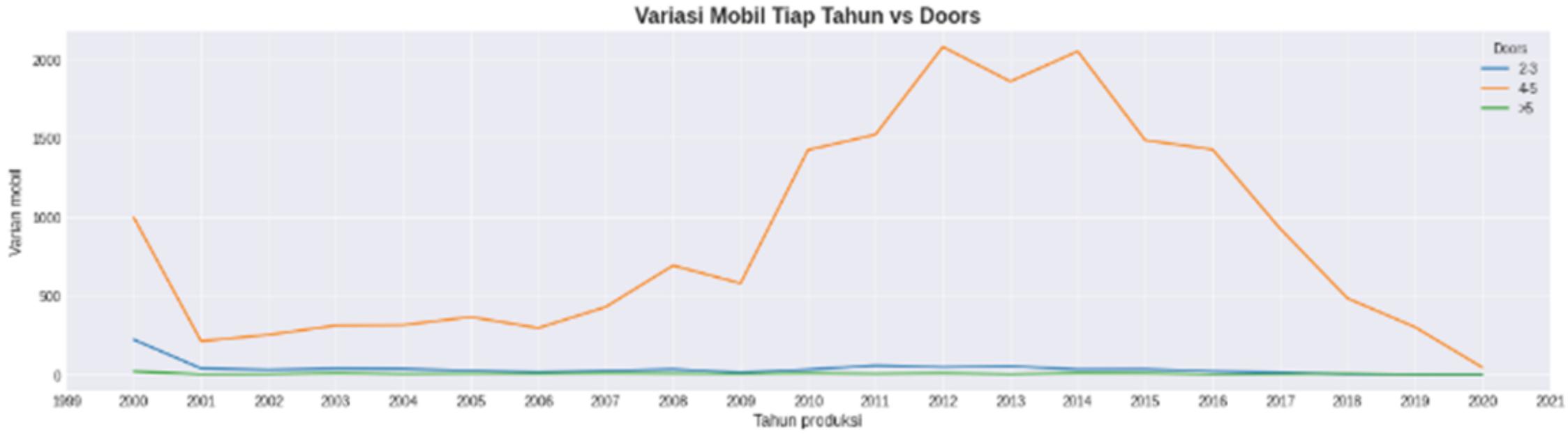
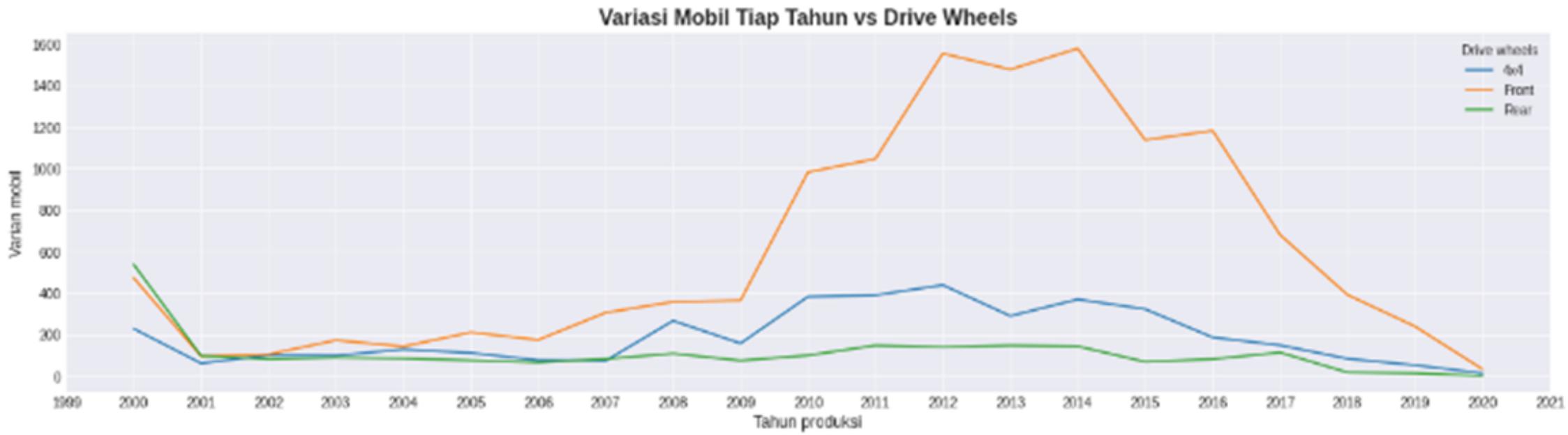
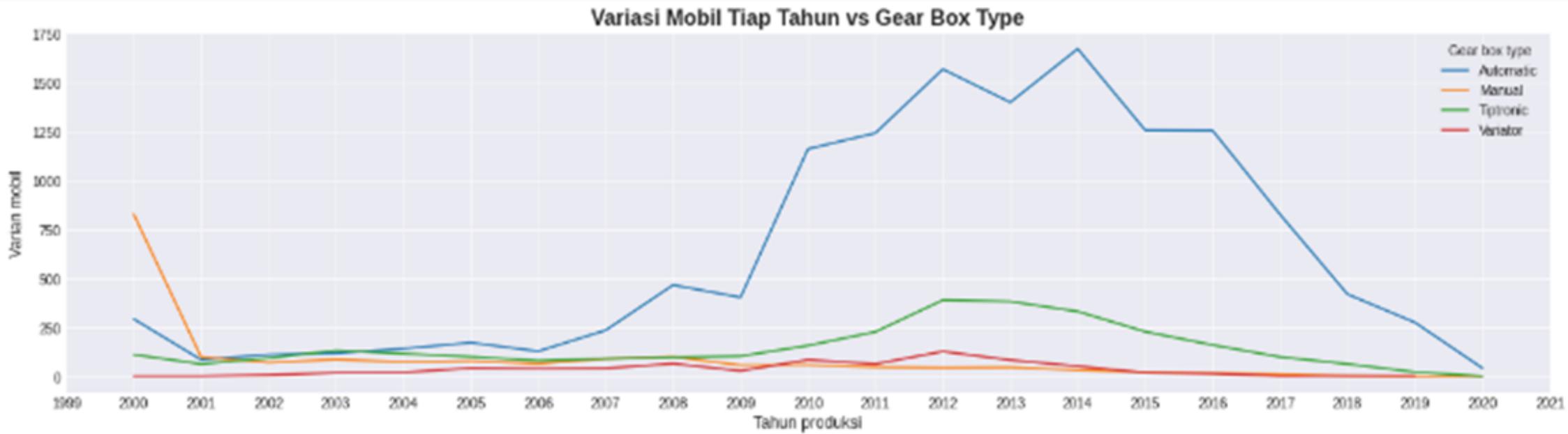
Manufacturer Hyundai, Toyota, Mercedes-Benz memiliki varian mobil tertinggi.

02

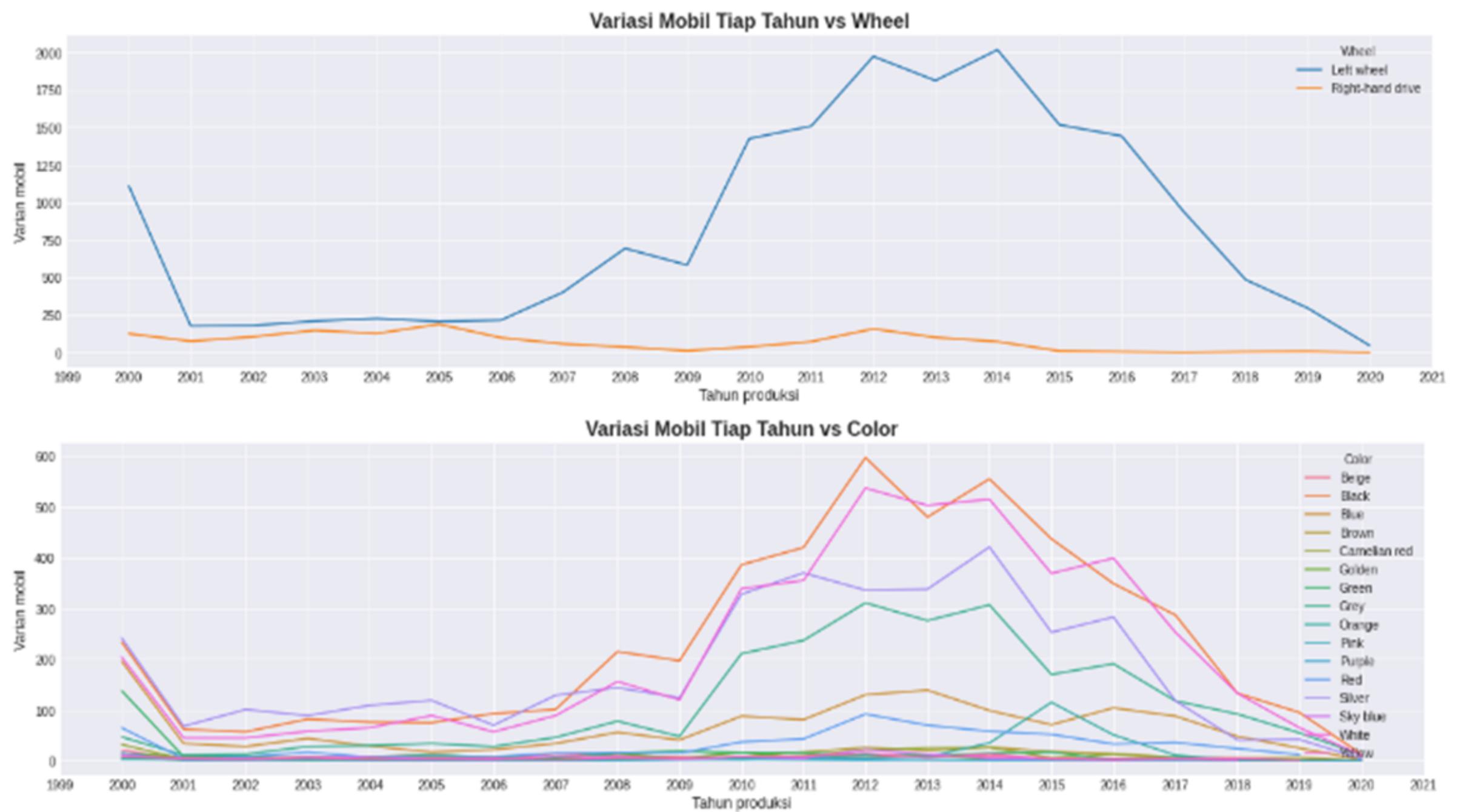
Dari 65 Manufacturer, hanya dari ketiganya saja sudah menyumbang 48.8% varian mobil.



- Category sedan mendominasi di tiap tahunnya, disusul oleh jeep dan hatchback.
- Sebagian besar varian mobil memiliki interior kulit(Leather interior).
- Jenis bahan bakar petrol selalu menjadi pilihan.



- Transmisi (Gearbox type) automatic mengalami kenaikan signifikan sekitar tahun 2007 – 2014, lalu mengalami penurunan sampai tahun 2020.
- Mobil dengan penggerak roda depan (Front) dan jumlah pintu (Doors) 4-5 selalu mendominasi pasar.



- Setir kemudi sebelah kiri (Left Wheels) selalu mendominasi pasar.
- Warna (Color) Black, White dan Silver selalu menjadi pilihan di tiap tahunnya.

MODELLING AND EVALUATION

MACHINE LEARNING MODEL



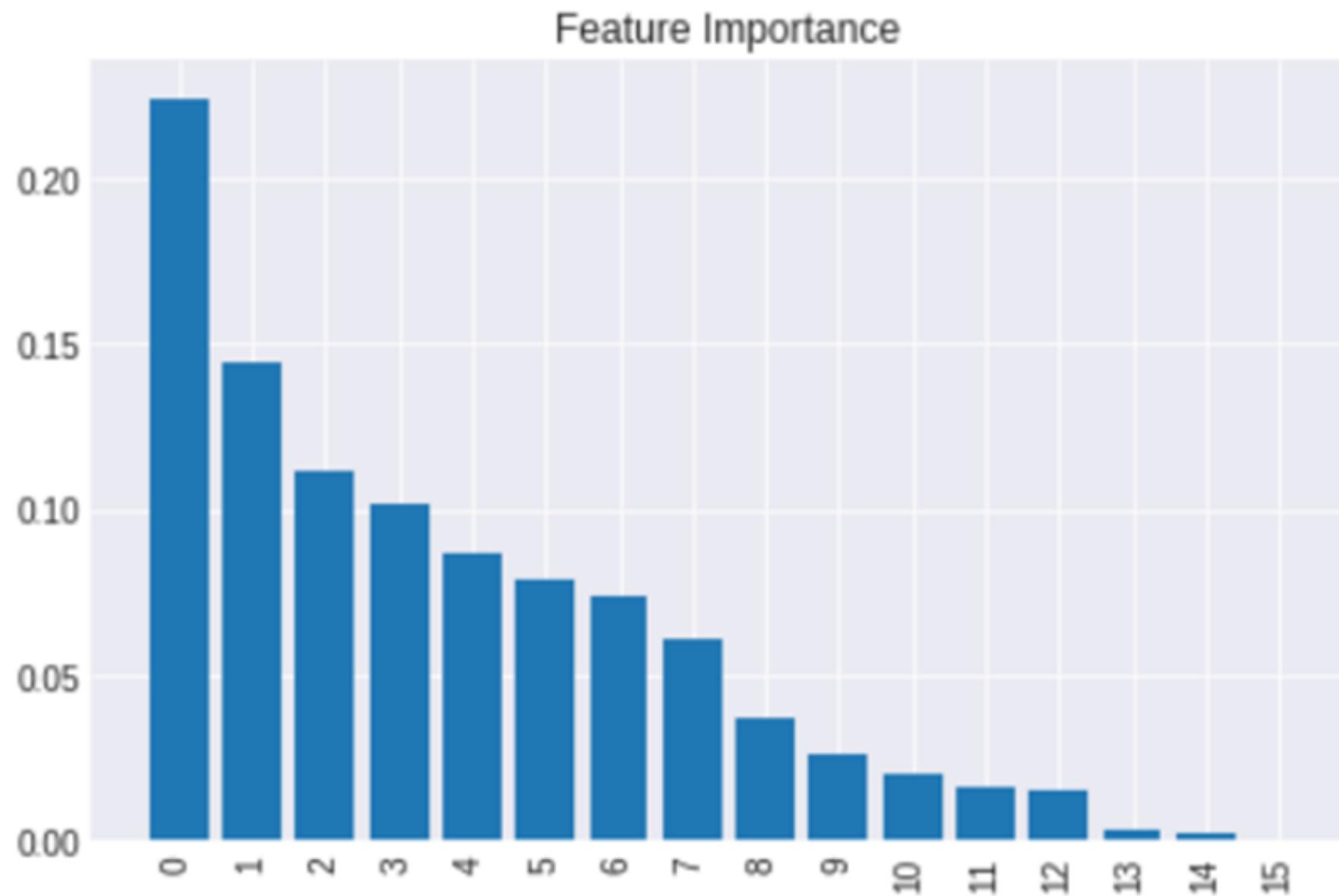
ML Model	Test Accuracy	Train Accuracy
Linear Regression	0,27	0,28
DecisionTree Regression	0,609	0,986
Random Forest Regression	0,782	0,959
Gradient Boost	0,619	0,641
Xgboost	0,763	0,892
K-Neighbors Regression	0,685	0,798

HYPERPARAMETER TUNING

ML Model	R2		RMSE		MAE		MAPE	
	Train	Test	Train	Test	Train	Test	Train	Test
Decision Tree	0.74	0.64	6805.75	8023.85	4519.66	5217.89	10.55	10.03
Random Forest	0.93	0.76	3522.44	6480.48	2169.79	3941.17	5.23	8.78

RANDOM FOREST THE BEST MODEL

FEATURE IMPORTANCES



Prod. year	0.224366
Airbags	0.144220
Mileage	0.111694
Engine volume	0.101543
Fuel type	0.086687
Levy	0.078202
Gear box type	0.073434
Model	0.061065
Manufacturer	0.036817
Color	0.025869
Drive wheels	0.019655
Category	0.016213
Leather interior	0.015062
Doors	0.002906
Wheel	0.002269
Cylinders	0.000000

CONCLUSION AND RECOMMENDATION



Conclusion

- Untuk Pemodelan Random Forest adalah yang terbaik untuk memprediksi harga mobil.
- Umur mobil / tahun produksi mobil mempengaruhi harga sebuah mobil.
- Jumlah Airbags mempengaruhi harga sebuah mobil.
- Jarak tempuh mobil mempengaruhi harga sebuah mobil.



Recommendation

Jika kita ingin memulai bisnis dalam bidang jual beli mobil :

- Gunakanlah Random Forest untuk pemodelan machine learning.
- Gunakan Tahun Produksi, Jumlah Airbags dan Jarak tempuh mobil untuk memprediksi kenaikan atau penurunan harga sebuah mobil.

THANK YOU

Script :
<https://github.com/VebriantoZubaidi/CarPricePrediction/blob/main/Vebrianto%20Zubaidi%20Arrahman%20-%20Script%20Progress%20Final%20Project.ipynb>