ADVANCED DATA ANALYSIS

Data Analysis on Germany cars

Prepared By: Thota Chandra Sekhar

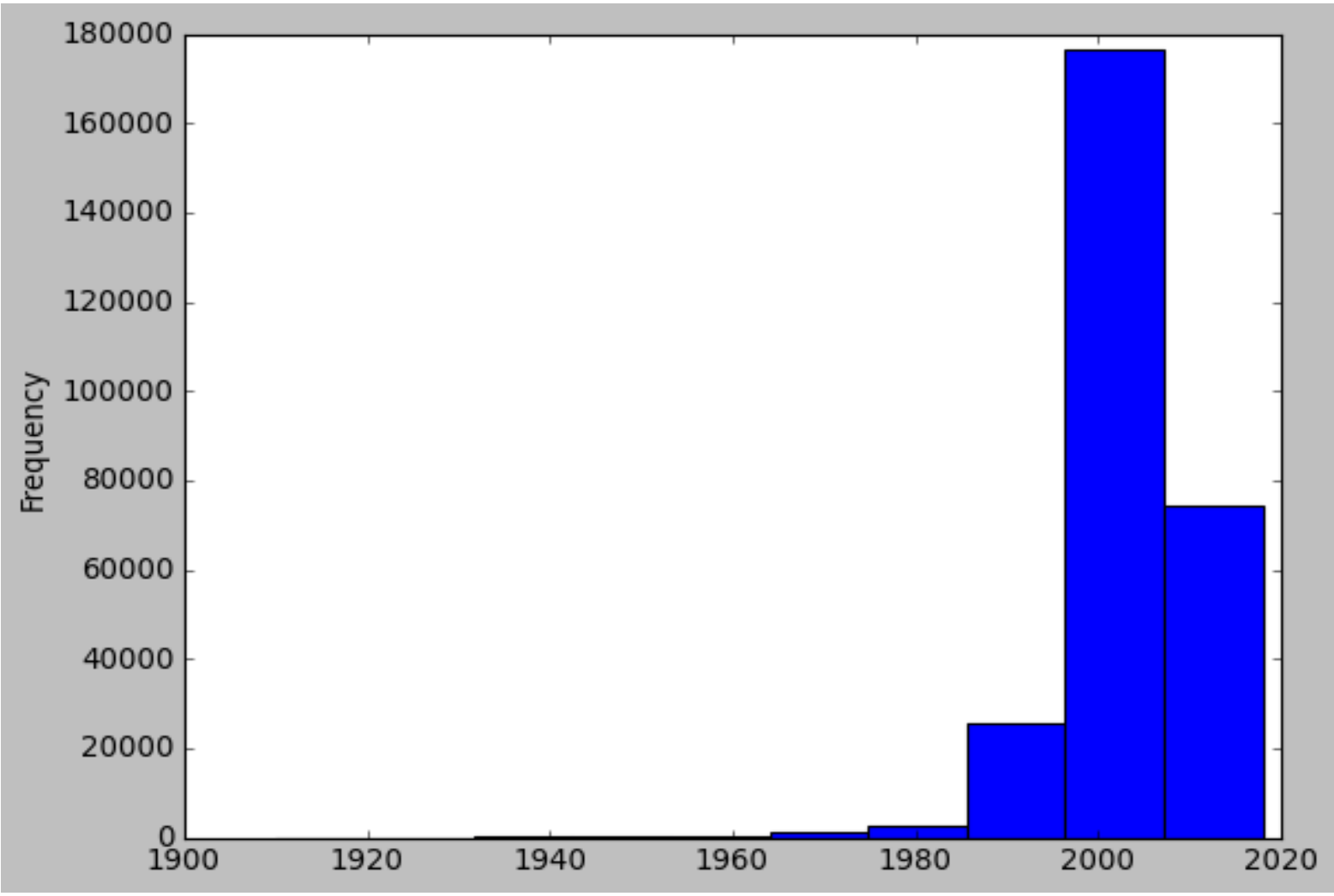
Date: 23-12-2023

Analysis 1:

➢ PERFORM GENERAL DATA ANALYSIS:

- Firstly we need to import the necessary libraries which is useful for the data analysis like: “NumPy ” , ”pandas ”, ”matplotlib”, ”seaborn”.
- Later we need to extract the Data from the site and assign to a variable.
- Now we need to start the “Data Cleaning”.
- **Missing Values:** Here we need to Remove the Missing values because here the missing values above 13% so, if we use replace the values the data will be changed , so that we need drop all the missing values rows.
- **Duplicates:** Later we need to check the Duplicate Values and here we have 4 duplicate values so we need to remove those duplicated values.
- **Type Casting :** Now we need to check the type of the column and change to actual type of the column ,we got 3 columns of errors in object type now we changed them to Datetime type.
- **Outliers**: Here we have outliers like price with “0”,”1”,”2” and more than actual price of the car , so we are removing the Outliers .And we have Power PS columns with outliers we need to remove even those.
- **Structural Errors** : In this data we have Month of Registration with “0” , so we need to replace with the mode value.

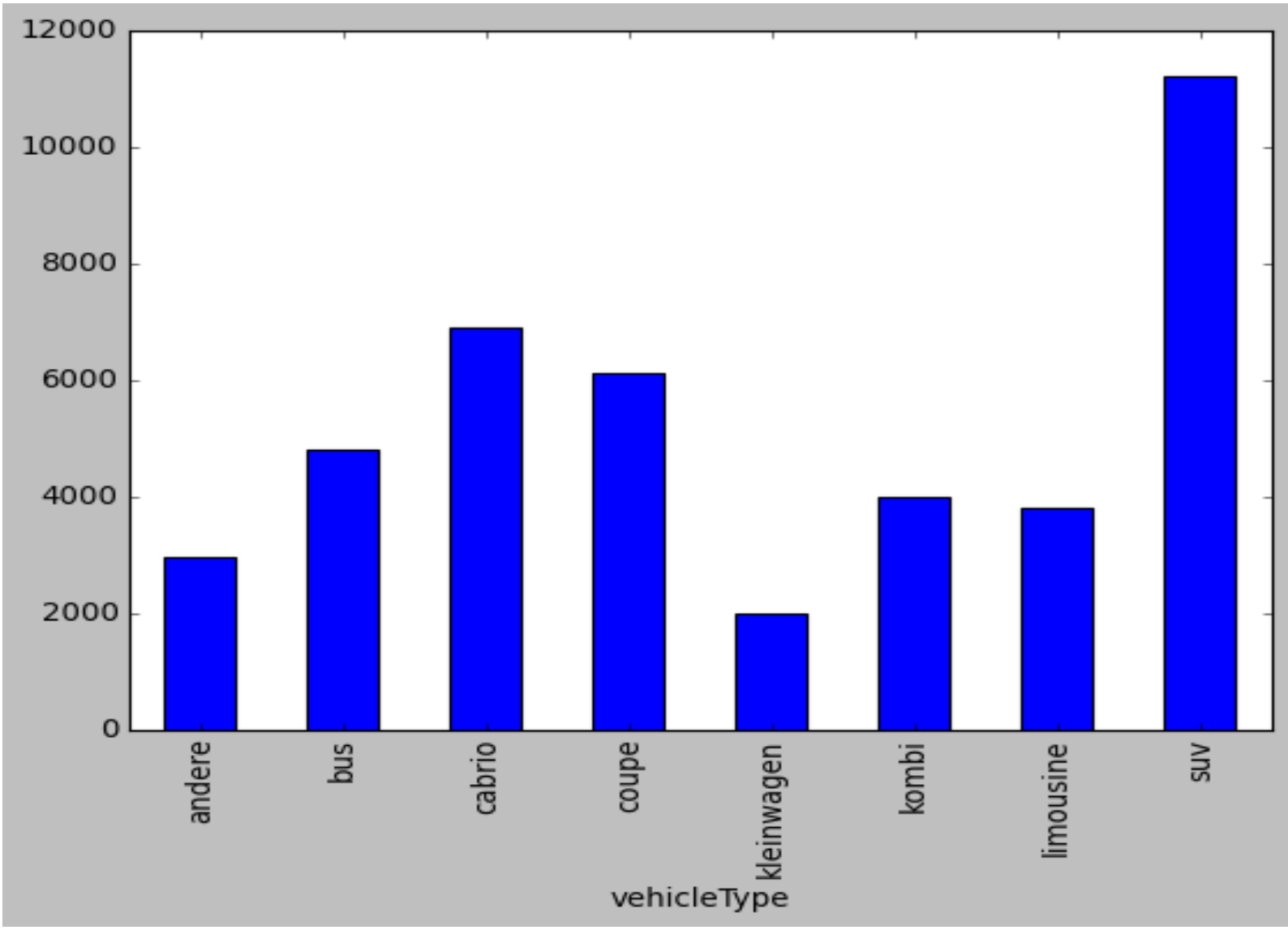
➢ Can you tell me the Distribution of Vehicles based on Year of Registration with the help of a plot



Summary :

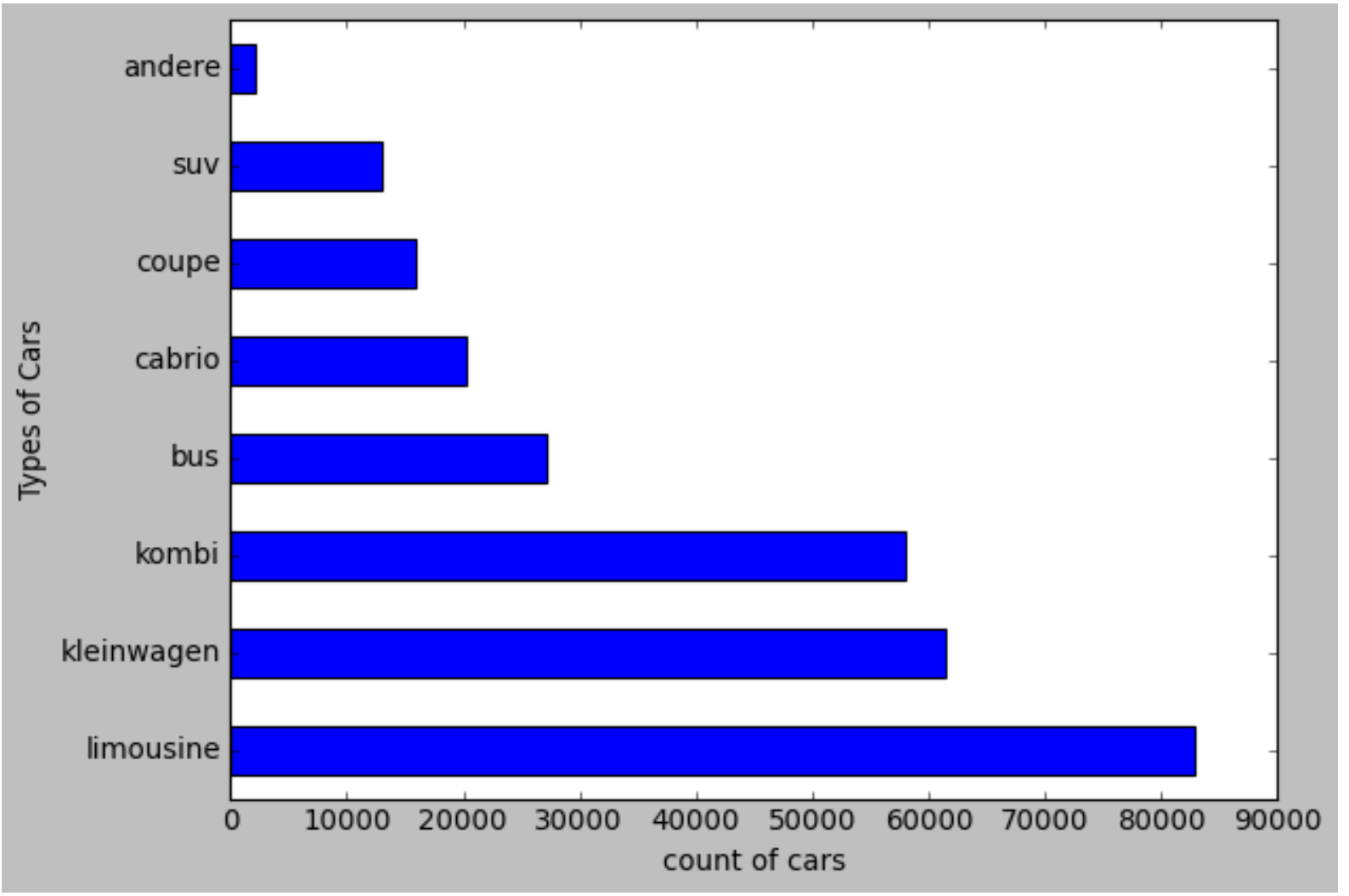
- Here the number of cars is frequency and the x-axis is registered years
- Here we observed that the most of the people did not register the cars at beginning at 1940,1950,1960.
- Gradually the registration of the cars is increased by 1970 from there the registration of cars are increased gradually up to 25000
- The Registration of cars are between 2000 to 2015 up to 180000.
- Later the graph was decreased gradually from the year 2015 till 2020 up to 80000.

➢ Create a plot based on the Variation of the price range by the vehicle type



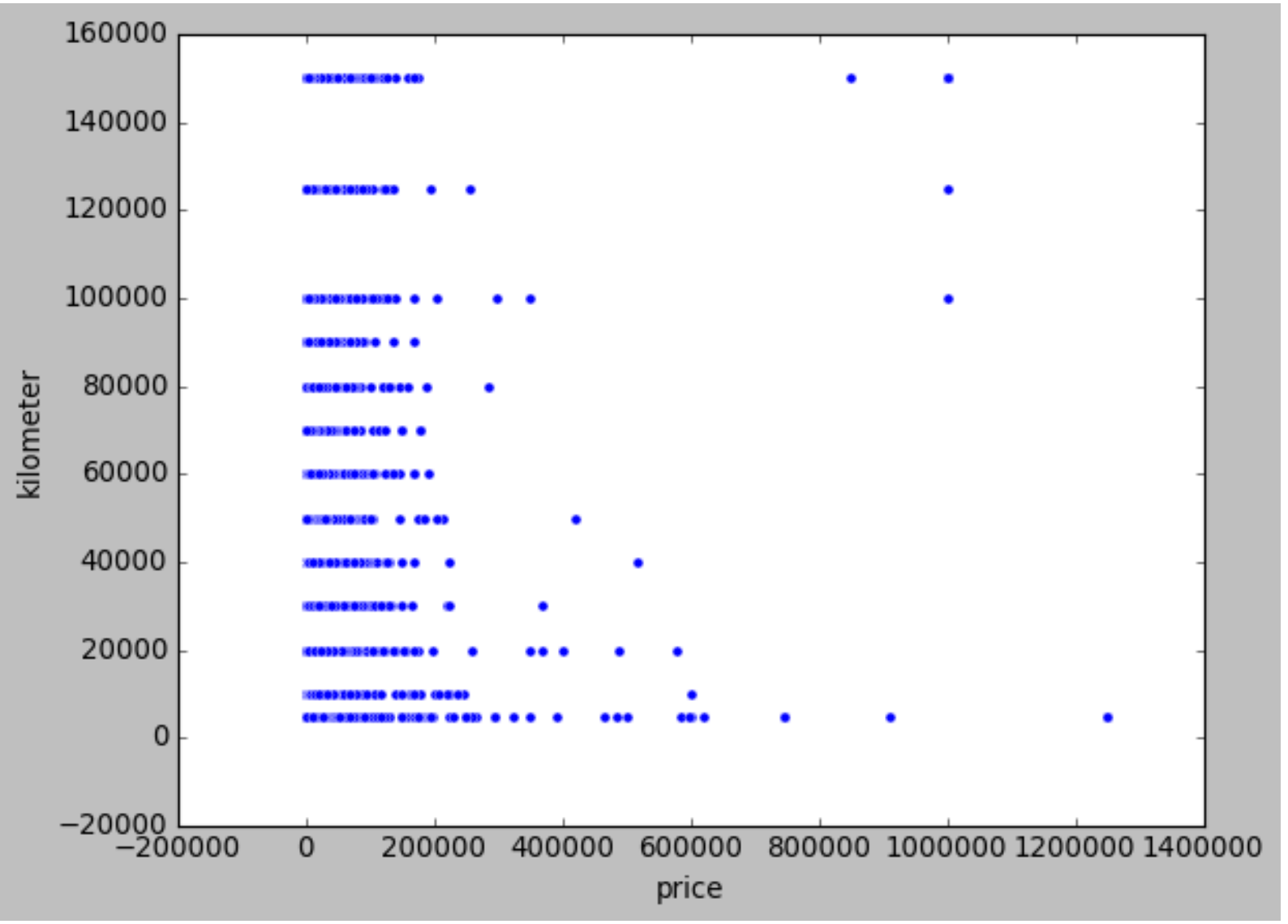
Summary : Here are the vehicle types and there prices in the range.

- **Find out Total count of vehicles by type available on eBay for sale. As well as create a visualization for the client**



- Summary: These are the total number of vehicles and the number of cars by vehicle type

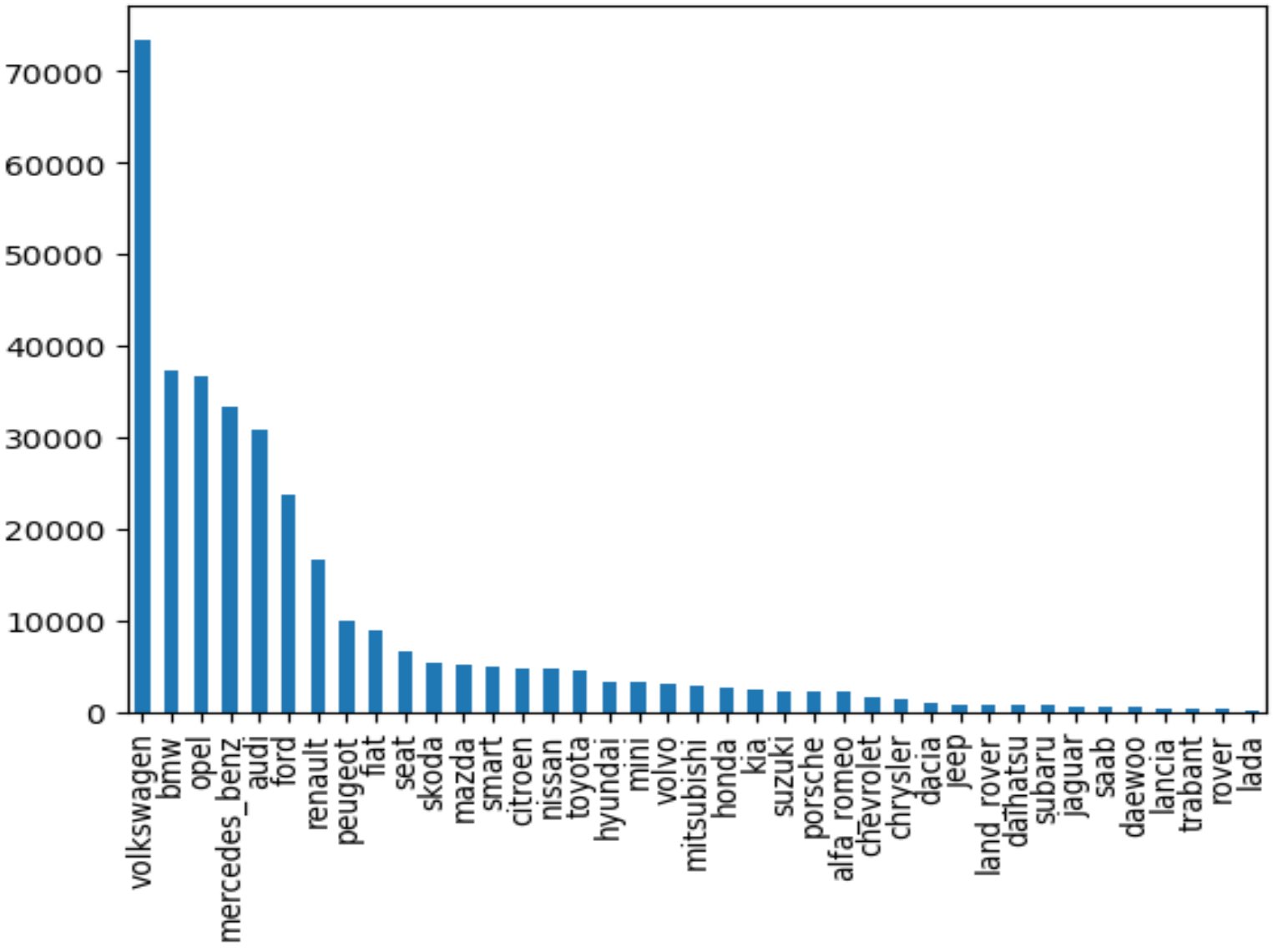
- **Is there any relationship between dollar price and kilometre? (Explain with appropriate analysis)**



- Summary: There is no relationship between dollar price and the kilometre because if there is relationship between any two columns then the scatter plot goes left to right from minimum value to maximum value and if one column value increases then other column value also increases with a single line dotted plot.

Analysis **2:**

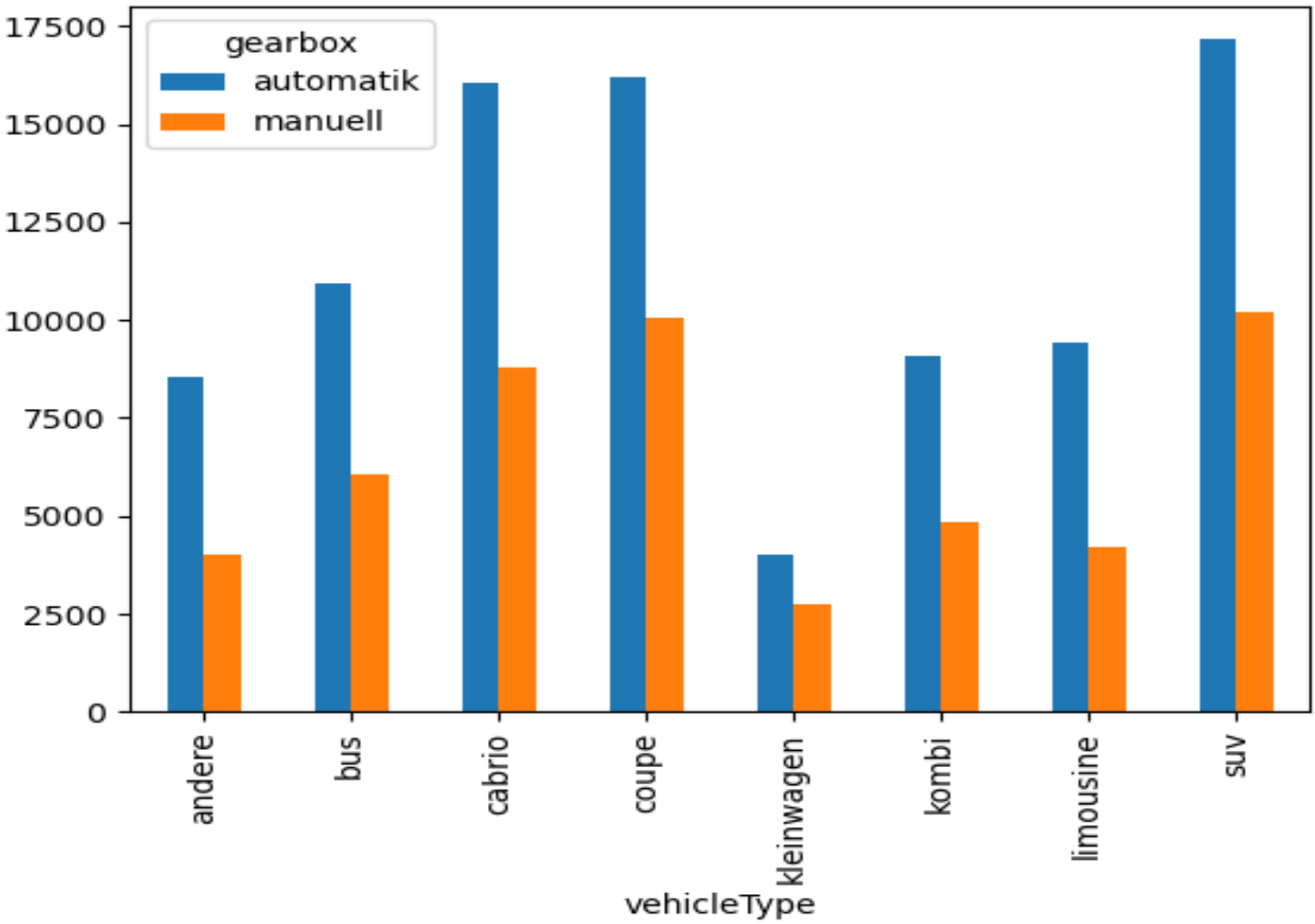
- **Can you tell me No of Vehicles by Brand Available on eBay for sale with the help of visualization**



Summary: These are the number of vehicles by brand available on eBay for sale, From this we can conclude that “Volkswagen” brand have more Sales.

- ***What is the Average price for vehicles based on the type of vehicle as well as on the type of gearbox. Explain me with both numerical and visualization analysis***

gearbox	automatik	manuell
vehicleType		
andere	8535.707819	4040.197784
bus	10926.346774	6047.074835
cabrio	16058.176796	8804.126836
coupe	16209.865955	10064.038935
kleinwagen	4004.249388	2758.509936
kombi	9075.676357	4856.891430
limousine	9419.215696	4204.549341
suv	17145.914836	10214.561145



Summary:

This are the Average price for vehicles based on the type of vehicle as well as on the type of gearbox.

➢ **What is the marginal probability of private seller**

```
privat          0.999991
gewerblich      0.000009
Name: seller, dtype: float64
```

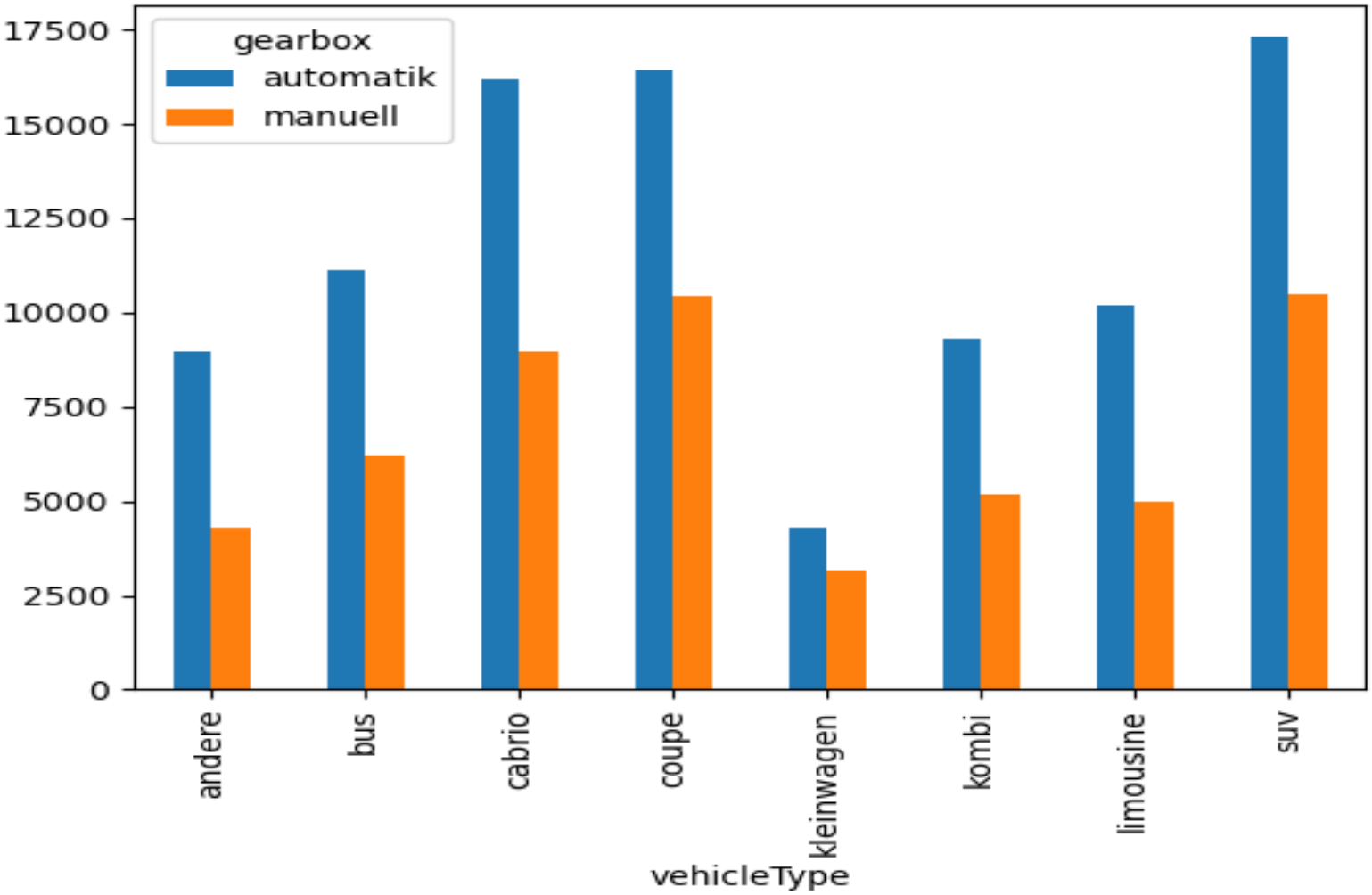
Analysis 3 :

➢ **The memory usage of the data is around 6.1 mb .How can we reduce the memory usage of the data set?**

Summary:

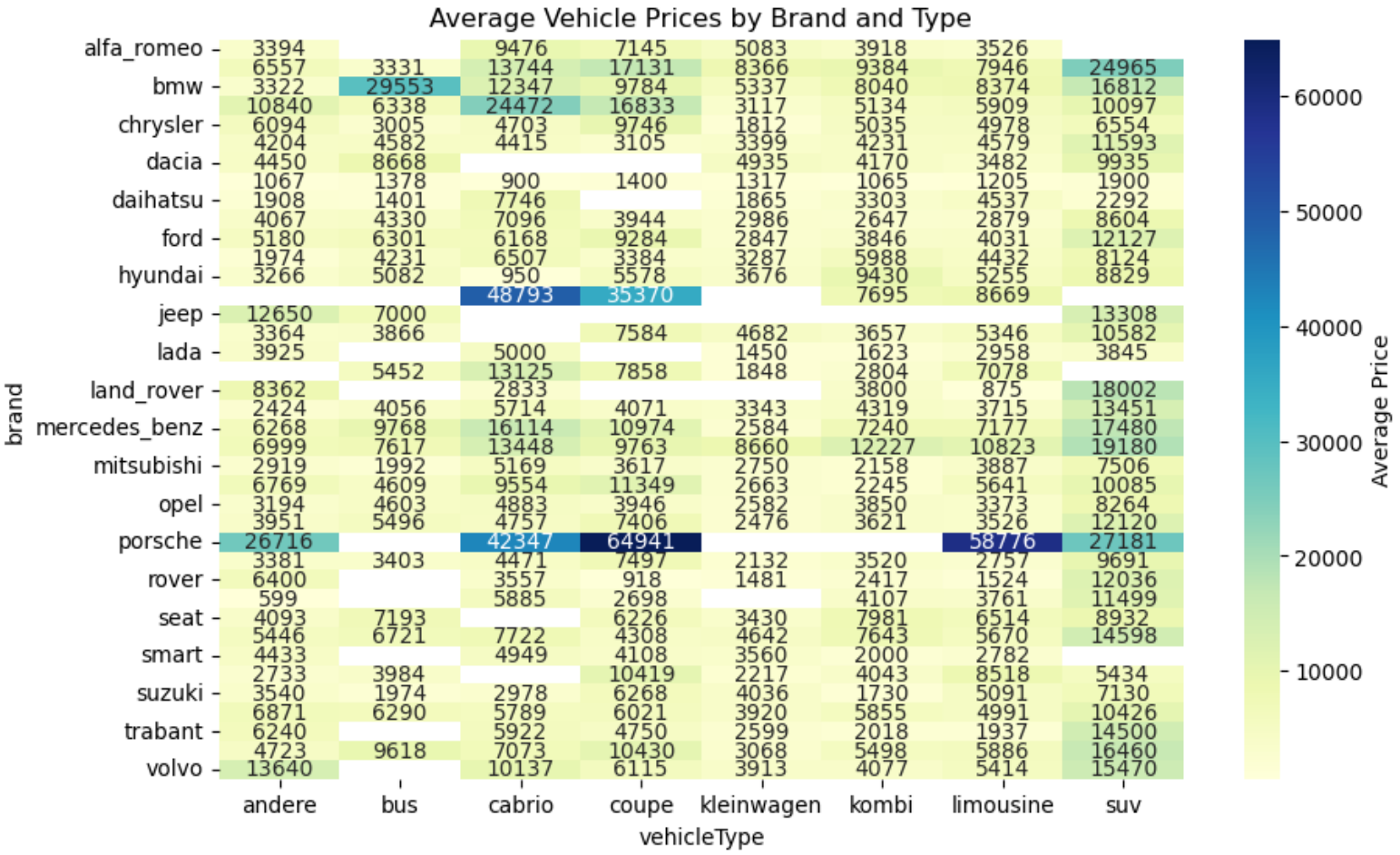
- Here we can reduce memory usage by dropping the data by:
- Removing unwanted columns.
- changing the type of the data as int to float and float to int.
- Optimize testate.

➢ **What is the Average price of vehicle by fuel type and gearbox type. Give a plot**



- **Summary:**
- This is the Average price of Vehicle by fuel type and Gearbox Type.

➢ What is the Average power of a vehicle by vehicle type and gearbox type. Give a plot



• Summary:

- *Here we classified the vehicle type by the brand with the price.*
- *Later we did Data preparation.*
- *Data Transformation.*
- *Visualization.*
- *Customization.*
- *Display*