Data Analysis Project -1

Report

This Report provides a comprehensive analysis of used cars in Germany which are on sale on ebay.

DATA CLEANING :

- Data cleaning is a critical phase in any data analysis project, serving as the foundation upon which accurate and reliable insights are built.

- It involves the process of identifying, correcting, and handling inaccuracies, inconsistencies, and anomalies within a dataset.

- This report will provide a detailed account of the methodologies, techniques, and tools employed during the data cleaning process. Additionally, it will showcase specific challenges encountered, the corresponding strategies employed, and the outcomes achieved.

- By offering transparency into the data cleaning process, this report aims to establish a solid foundation for the subsequent stages of analysis

ANALYSIS – 1

1. Perform General Analysis :

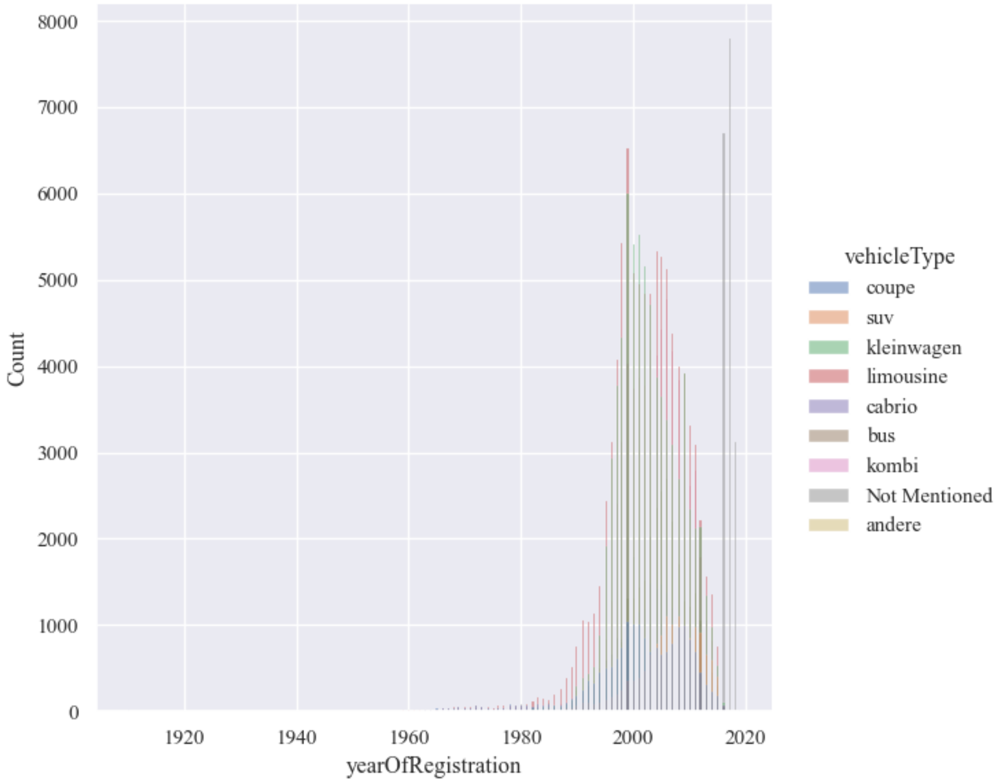
After the process of Data Cleaning is completed one needs to analyze the data.

Performing General Data Analysis includes having the knowledge of different aggregate functions, filtering techniques, visualization techniques that can be developed from the dataset.

We must go through the datatypes of the columns in accordance for the visualization.

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

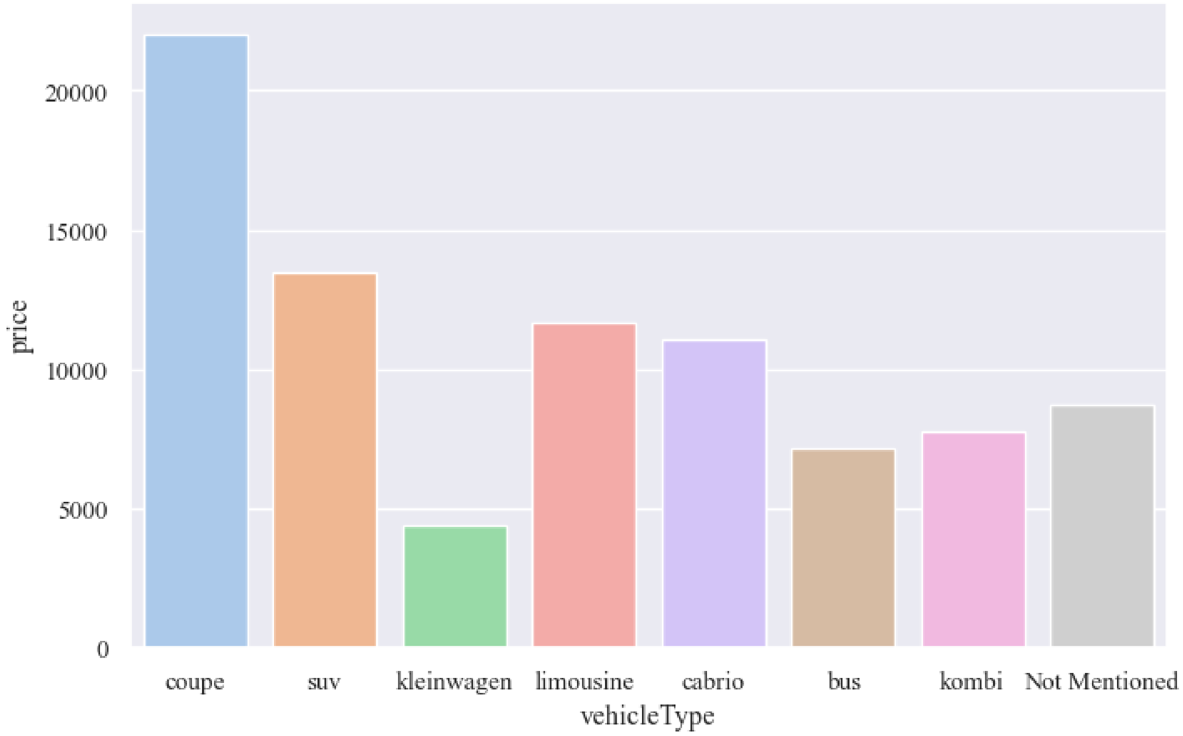
For knowing the distribution of vehicles across the years we can use plot.



From the above plot we can say that in the year 2000 more number of vehicles were purchased and limousine and kleinwagen are the type of vehicles which were of high demand.

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

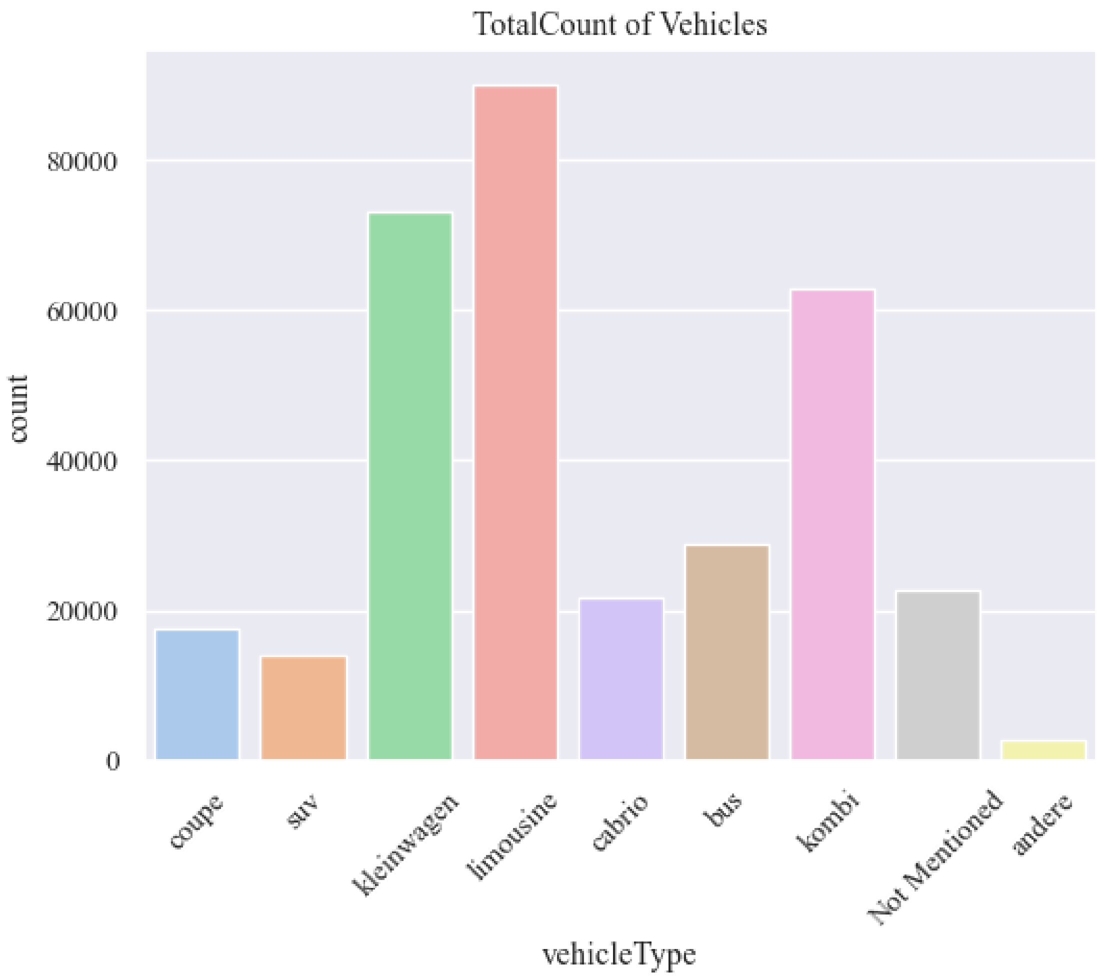
This type of analysis can be done by finding the average price of each vehicle type using aggregate functions like mean.



The above plot visualizes the Variation of price range based on the type of vehicle.

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

To find the Counts of types of vehicles available on ebay we can use the built-in function of value.counts() which gives us the count of each unique element in the specified column.

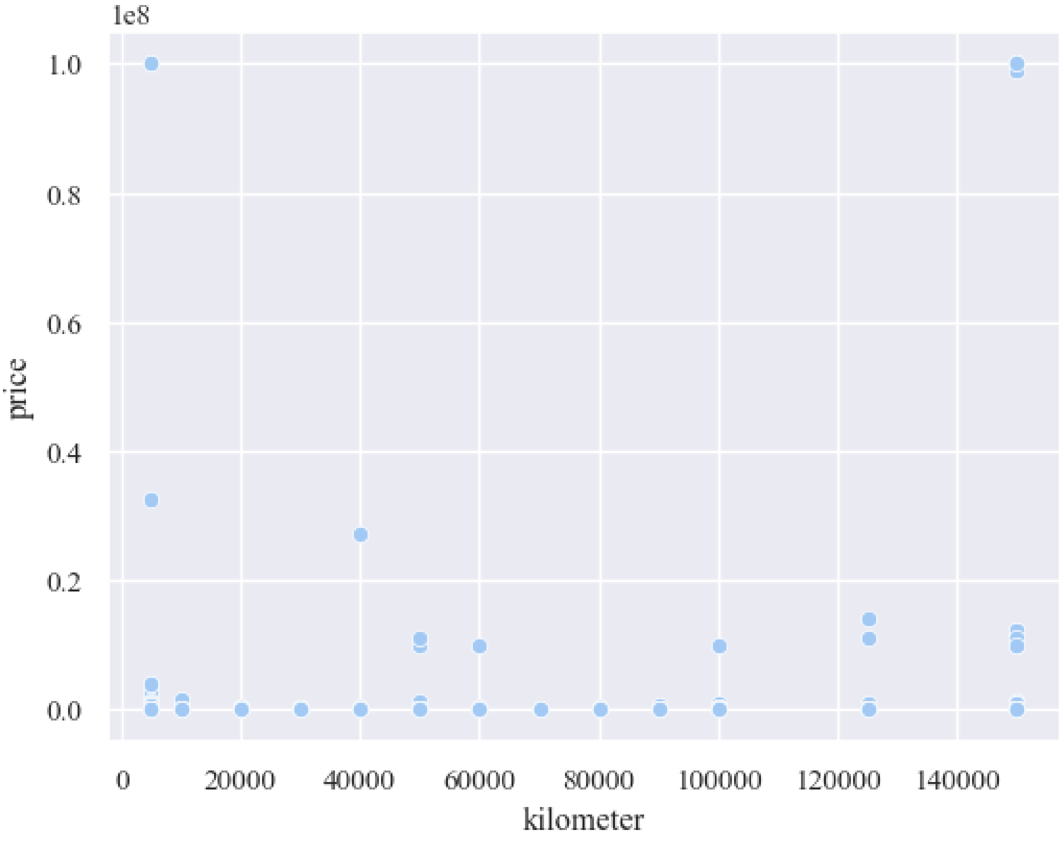


A count plot can be used to visualize this type of data where we show the frequency of the values of x-axis.

5. Is there any relationship between dollar_price and kilometer? (Explain with appropriate analysis)

The relation between two numerical values can be given by Pearson correlation coefficient/Spearman correlation coefficient.

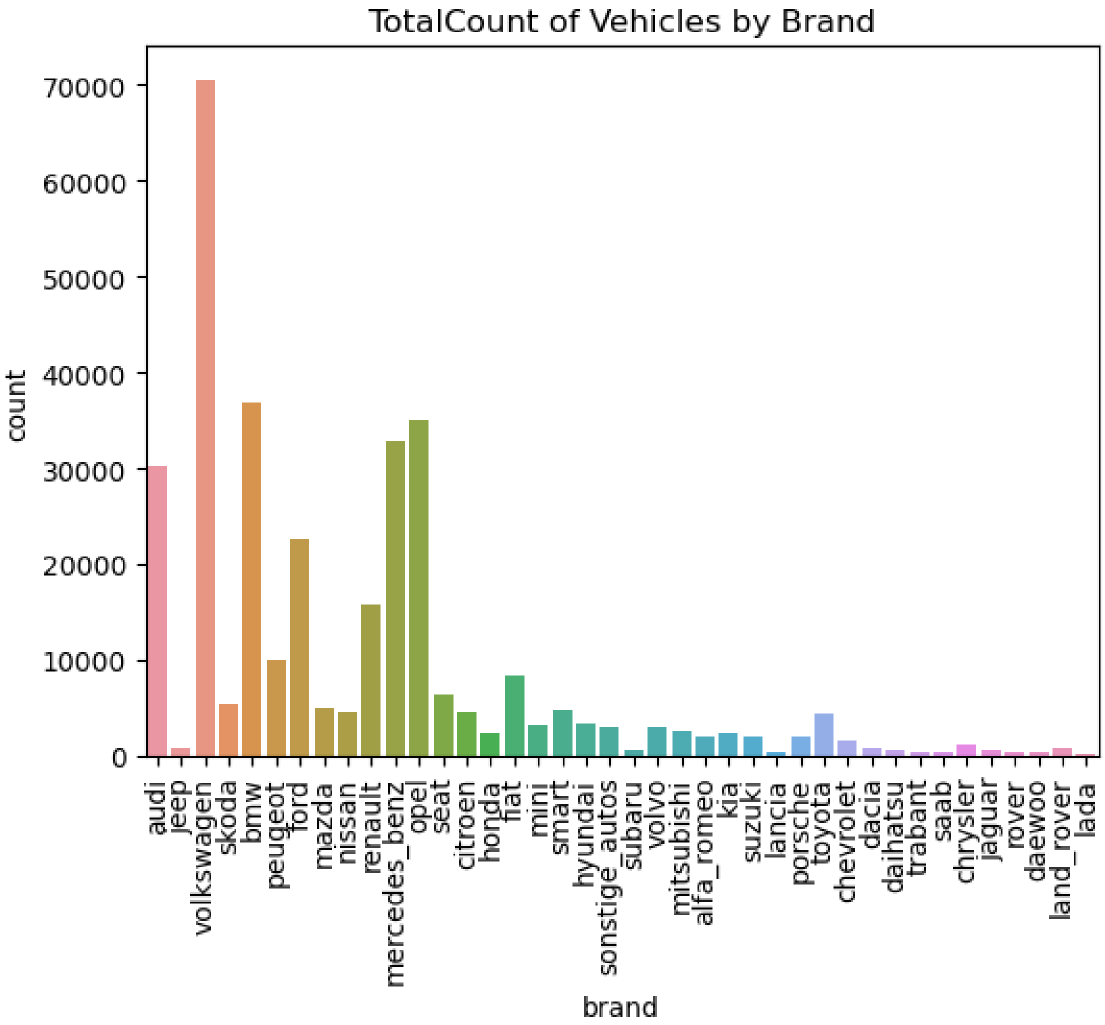
Scatter plot can be used for representing the relationship between the dollar_price and kilometer in the dataset.



ANALYSIS – 2

1) Can you tell me No of Vehicles by Brand Available on ebay for sale with help of visualization.

To visualize the number of vehicles available on eBay for sale by brand, you can create a bar plot using Seaborn or any other plotting library of your choice.



we first count the number of vehicles for each brand using the ‘value_counts’ method. Then, we create a bar plot using Seaborn to visualize the counts for each brand. Replace the sample data with your actual dataset to visualize the number of vehicles available by brand on eBay based on your specific data.

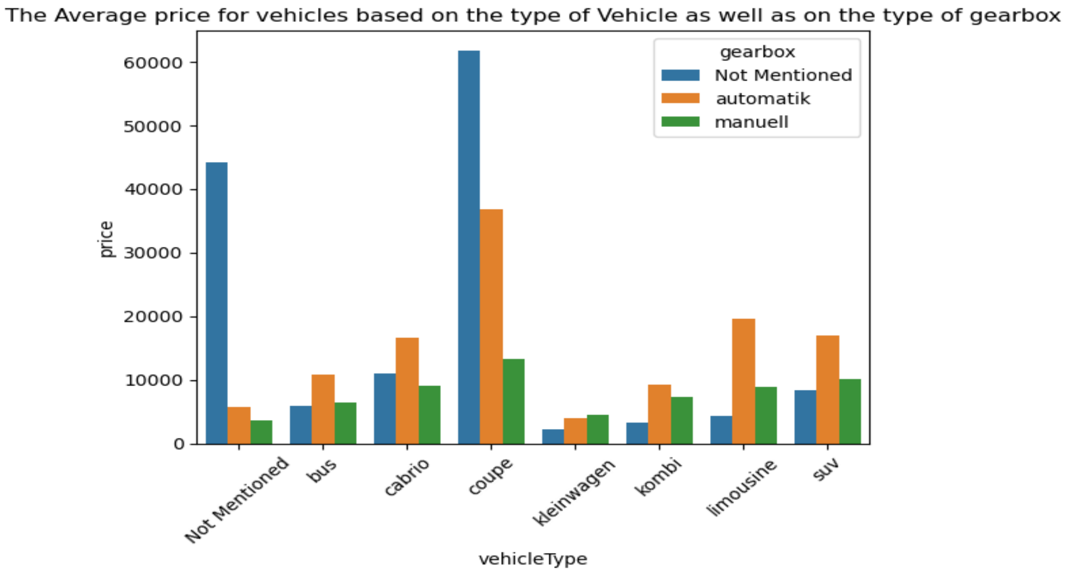
2) What is the Average price for vehicles based on the type of vehicle as well as on the type of gearbox. Explain me with both numerical and visualization analysis.

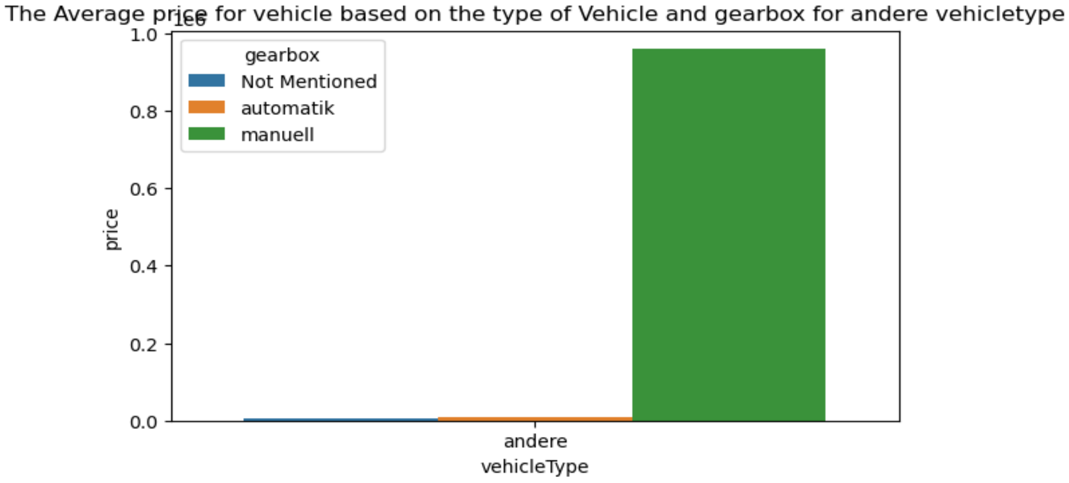
Numerical Analysis:

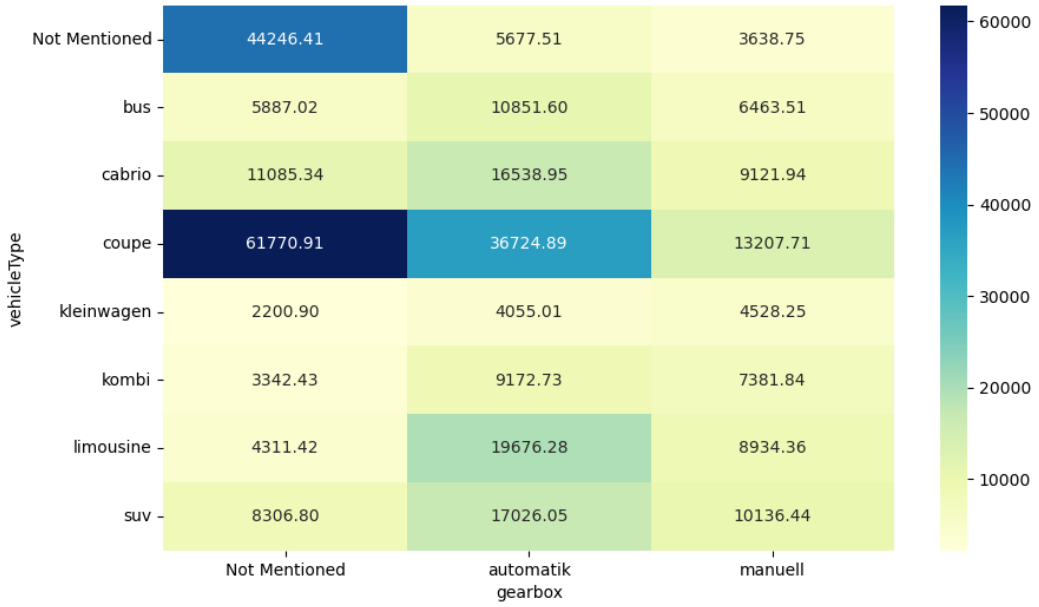
You can calculate the average price for vehicles by grouping the data based on vehicle type and gearbox type and then computing the mean price within each group.

Visualization:

To create a visualization to complement the numerical analysis, you can use a heatmap or barplot to show the average prices with Seaborn.







The cell colors represent the average prices, and the annotations display the numerical values. Replace the sample data with your actual dataset for a more meaningful analysis and visualization.

3) What is the marginal probability of private seller.

The marginal probability of "private seller" is the probability of the event "private seller" occurring without considering any other conditions or events related to other variables.

we calculate the marginal probability of "private seller" by counting the number of times "private seller" appears in the "Seller Type" column and dividing it by the total number of observations in the dataset.

ANALYSIS – 3

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

Reducing the memory usage of a dataset can be important when working with large datasets to optimize performance and avoid memory-related issues. Here are some strategies to reduce the memory usage of your dataset in Python:

1.Select Appropriate Data Types: One of the most effective ways to reduce memory usage is to use more memory-efficient data types for your columns. For example:

- Use **int8**, **int16**, or **int32** instead of **int64** for integer columns if the data range allows.
- Use **float32** instead of **float64** for floating-point columns if the precision is not critical.

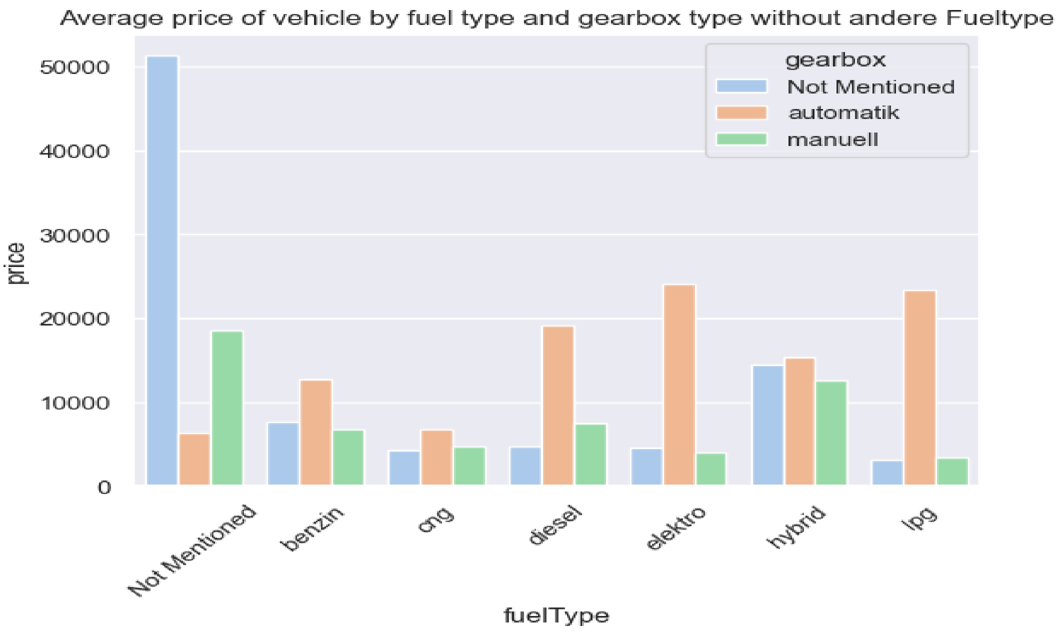
2.Remove Unnecessary Columns: If there are columns that are not needed for your analysis, consider dropping them from the dataset.

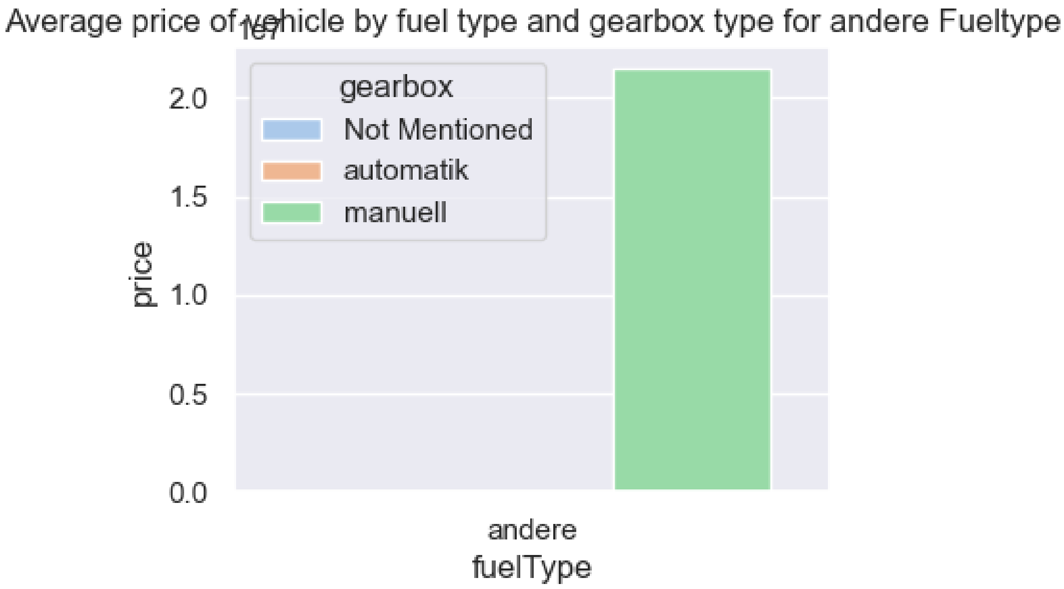
3.Filter Rows: If you are only interested in a subset of your data, filter the rows you need and create a new DataFrame with the filtered data.

By following these strategies, you can reduce the memory usage of your dataset while maintaining the integrity of your data for analysis. Be cautious about reducing precision or data loss when changing data types, and make sure the changes are appropriate for your specific use case.

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

we first calculate the average price of vehicles by grouping the data based on both "Fuel Type" and "Gearbox Type."


Average price of vehicle by fuel type and gearbox type without andere Fueltype

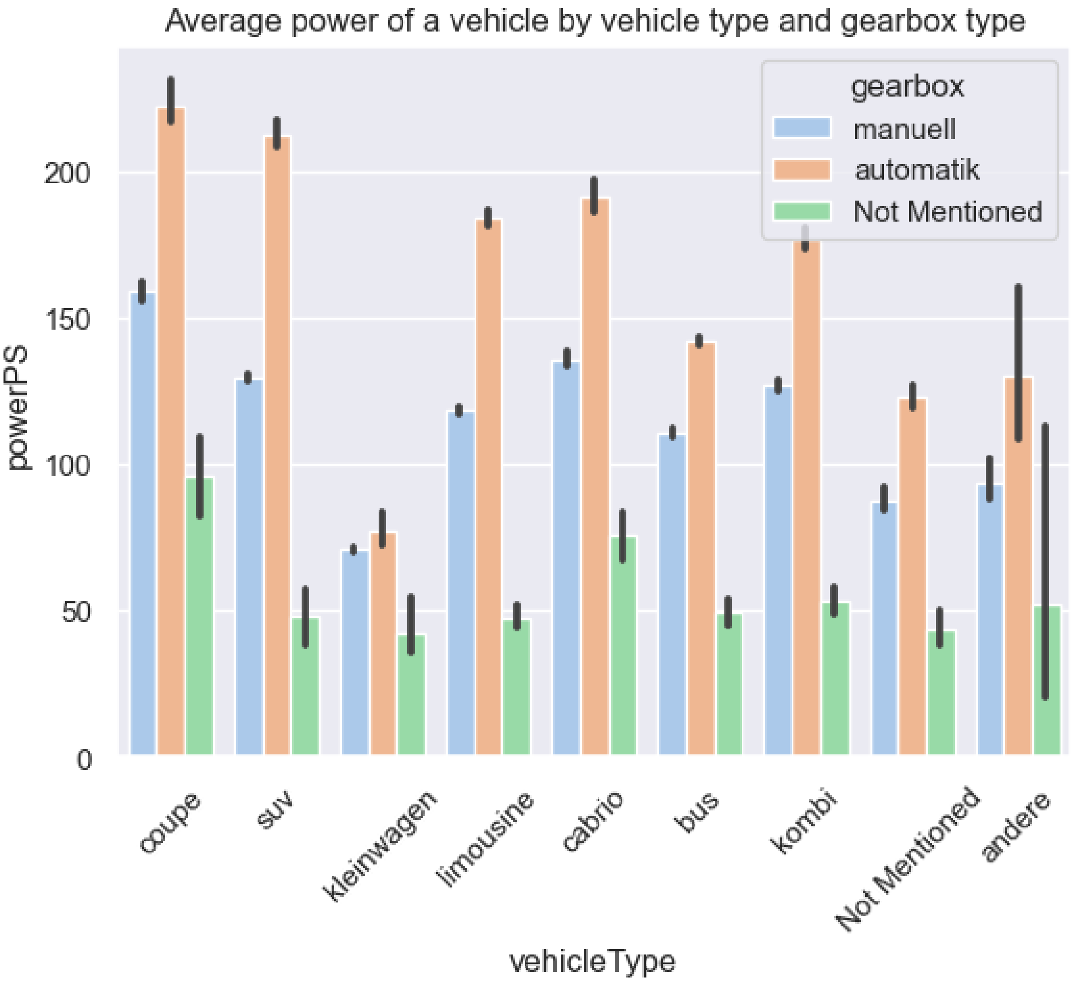

Average price of vehicle by fuel type and gearbox type for andere Fueltype

We then create a bar plot using Seaborn to visualize the average prices, and the hue parameter is used to differentiate the data by "Gearbox Type."

Replace the sample data with your actual dataset to calculate the average price by fuel type and gearbox type for your specific data

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

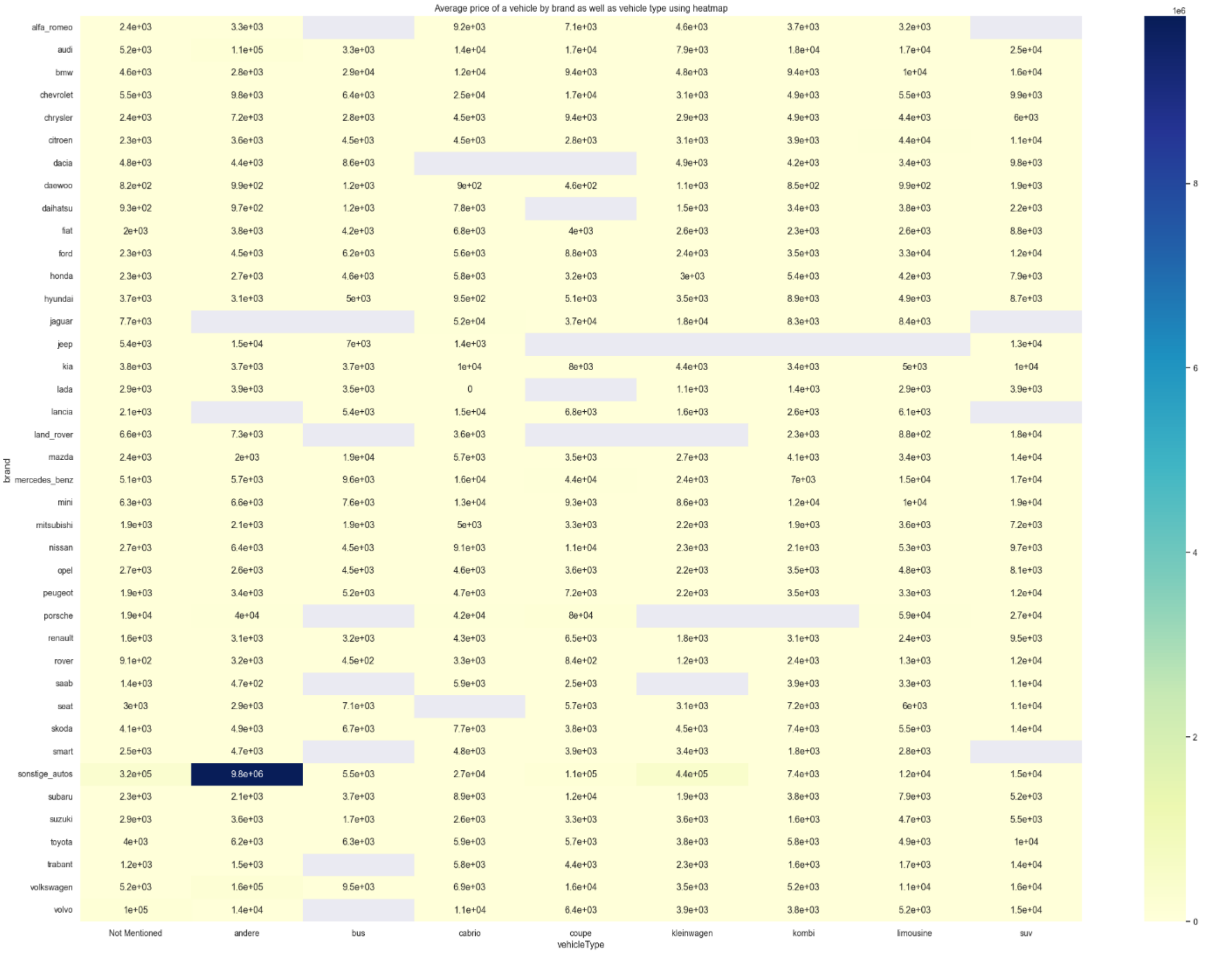
we first calculate the average power of vehicles by grouping the data based on both "Vehicle Type" and "Gearbox Type." We then create a bar plot using Seaborn to visualize the average power, and the hue parameter is used to differentiate the data by "Gearbox Type."



Replace the sample data with your actual dataset to calculate the average power by vehicle type and gearbox type for your specific data.

4) What is the Average price of a vehicle by brand as well as vehicle type using heatmap.

A heatmap is a great way to represent such a two-dimensional relationship.we calculate the average price by using a pivot table, and then we create a heatmap using Seaborn to visualize the average prices.


Average price of a vehicle by brand as well as vehicle type using heatmap

brand	Not Mentioned	andere	bus	cabrio	coupe	kleinwagen	kombi	limousine	suv
alfa_romeo	2.4e+03	3.3e+03		9.2e+03	7.1e+03	4.6e+03	3.7e+03	3.2e+03	
audi	5.2e+03	1.1e+05	3.3e+03	1.4e+04	1.7e+04	7.9e+03	1.8e+04	1.7e+04	2.5e+04
bmw	4.6e+03	2.8e+03	2.9e+04	1.2e+04	9.4e+03	4.8e+03	9.4e+03	1e+04	1.6e+04
chevrolet	5.5e+03	9.8e+03	6.4e+03	2.5e+04	1.7e+04	3.1e+03	4.9e+03	5.5e+03	9.9e+03
chrysler	2.4e+03	7.2e+03	2.8e+03	4.5e+03	9.4e+03	2.9e+03	4.9e+03	4.4e+03	6e+03
citroen	2.3e+03	3.6e+03	4.5e+03	4.5e+03	2.8e+03	3.1e+03	3.9e+03	4.4e+04	1.1e+04
dacia	4.8e+03	4.4e+03	8.6e+03			4.9e+03	4.2e+03	3.4e+03	9.8e+03
daewoo	8.2e+02	9.9e+02	1.2e+03	9e+02	4.6e+02	1.1e+03	8.5e+02	9.9e+02	1.9e+03
daihatsu	9.3e+02	9.7e+02	1.2e+03	7.8e+03		1.5e+03	3.4e+03	3.8e+03	2.2e+03
fiat	2e+03	3.8e+03	4.2e+03	6.8e+03	4e+03	2.6e+03	2.3e+03	2.6e+03	8.8e+03
ford	2.3e+03	4.5e+03	6.2e+03	5.6e+03	8.8e+03	2.4e+03	3.5e+03	3.3e+04	1.2e+04
honda	2.3e+03	2.7e+03	4.6e+03	5.8e+03	3.2e+03	3e+03	5.4e+03	4.2e+03	7.9e+03
hyundai	3.7e+03	3.1e+03	5e+03	9.5e+02	5.1e+03	3.5e+03	8.9e+03	4.9e+03	8.7e+03
jaguar	7.7e+03			5.2e+04	3.7e+04	1.8e+04	8.3e+03	8.4e+03	
jeep	5.4e+03	1.5e+04	7e+03	1.4e+03					1.3e+04
kia	3.8e+03	3.7e+03	3.7e+03	1e+04	8e+03	4.4e+03	3.4e+03	5e+03	1e+04
lada	2.9e+03	3.9e+03	3.5e+03	0		1.1e+03	1.4e+03	2.9e+03	3.9e+03
lancia	2.1e+03		5.4e+03	1.5e+04	6.8e+03	1.6e+03	2.6e+03	6.1e+03	
land_rover	6.6e+03	7.3e+03		3.6e+03			2.3e+03	8.8e+02	1.8e+04
mazda	2.4e+03	2e+03	1.9e+04	5.7e+03	3.5e+03	2.7e+03	4.1e+03	3.4e+03	1.4e+04
mercedes_benz	5.1e+03	5.7e+03	9.6e+03	1.6e+04	4.4e+04	2.4e+03	7e+03	1.5e+04	1.7e+04
mini	6.3e+03	6.6e+03	7.6e+03	1.3e+04	9.3e+03	8.6e+03	1.2e+04	1e+04	1.9e+04
mitsubishi	1.9e+03	2.1e+03	1.9e+03	5e+03	3.3e+03	2.2e+03	1.9e+03	3.6e+03	7.2e+03
nissan	2.7e+03	6.4e+03	4.5e+03	9.1e+03	1.1e+04	2.3e+03	2.1e+03	5.3e+03	9.7e+03
opel	2.7e+03	2.6e+03	4.5e+03	4.6e+03	3.6e+03	2.2e+03	3.5e+03	4.8e+03	8.1e+03
peugeot	1.9e+03	3.4e+03	5.2e+03	4.7e+03	7.2e+03	2.2e+03	3.5e+03	3.3e+03	1.2e+04
porsche	1.9e+04	4e+04		4.2e+04	8e+04			5.9e+04	2.7e+04
renault	1.6e+03	3.1e+03	3.2e+03	4.3e+03	6.5e+03	1.8e+03	3.1e+03	2.4e+03	9.5e+03
rover	9.1e+02	3.2e+03	4.5e+02	3.3e+03	8.4e+02	1.2e+03	2.4e+03	1.3e+03	1.2e+04
saab	1.4e+03	4.7e+02		5.9e+03	2.5e+03		3.9e+03	3.3e+03	1.1e+04
seat	3e+03	2.9e+03	7.1e+03		5.7e+03	3.1e+03	7.2e+03	6e+03	1.1e+04
skoda	4.1e+03	4.9e+03	6.7e+03	7.7e+03	3.8e+03	4.5e+03	7.4e+03	5.5e+03	1.4e+04
smart	2.5e+03	4.7e+03		4.8e+03	3.9e+03	3.4e+03	1.8e+03	2.8e+03	
sonstige_autos	3.2e+05	9.8e+06	5.5e+03	2.7e+04	1.1e+05	4.4e+05	7.4e+03	1.2e+04	1.5e+04
subaru	2.3e+03	2.1e+03	3.7e+03	8.9e+03	1.2e+04	1.9e+03	3.8e+03	7.9e+03	5.2e+03
suzuki	2.9e+03	3.6e+03	1.7e+03	2.6e+03	3.3e+03	3.6e+03	1.6e+03	4.7e+03	5.5e+03
toyota	4e+03	6.2e+03	6.3e+03	5.9e+03	5.7e+03	3.8e+03	5.8e+03	4.9e+03	1e+04
trabant	1.2e+03	1.5e+03		5.8e+03	4.4e+03	2.3e+03	1.6e+03	1.7e+03	1.4e+04
volkswagen	5.2e+03	1.6e+05	9.5e+03	6.9e+03	1.6e+04	3.5e+03	5.2e+03	1.1e+04	1.6e+04
volvo	1e+05	1.4e+04		1.1e+04	6.4e+03	3.9e+03	3.8e+03	5.2e+03	1.5e+04

The cell colors represent the average prices, and the annotations display the numerical values. Replace the sample data with your actual dataset to calculate and visualize the average price by brand and vehicle type for your specific data.