# BDA-1

***Car Data analysis***

*Final Project Group:*

# Car database Data analysis

import libraries

load and prepare the dataset

change the values to english

1.Analysis of vehicle price distribution based on model?

2. How many cars are available for sale across a given type's data set?

3. Analysis of the average price of vehicles based on fuel type and vehicle type?

4. Average prices of the brands and types of cars found in the analysis data set?

5. Create a pair of graphs that contain all the columns that contain the metrics?

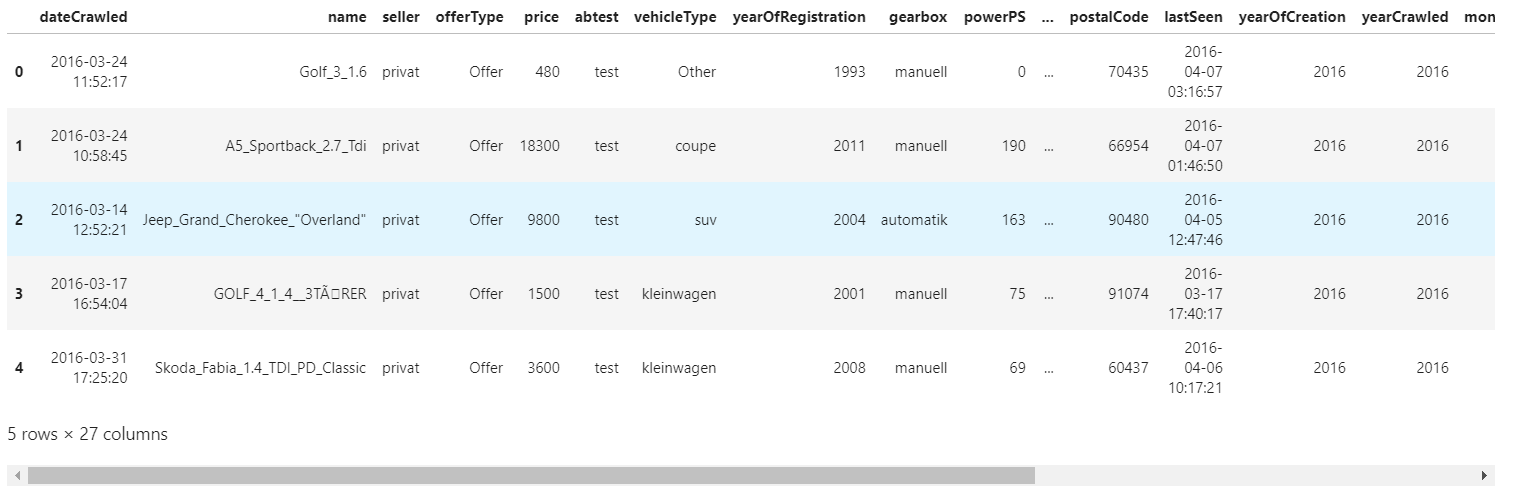
6. According to the data, which brand sells more cars?

Linear regression (car price)

**Data set overview**

The data set comes from the Internet and contains details of used German cars sold on ebay.

Vehicles with a registration year greater than 2015 and less than 1888 are also removed from the data set because the data are inconsistent and produce incorrect results.

**Sample data set**

**Question1**

1.Analysis of vehicle price distribution based on model?

Many outliers with unreasonable data have been removed to prepare the dataset for analysis.Vehicles registered between 1988 and 2015 can be sold at most.The highest in 2000 was 25,536cars.

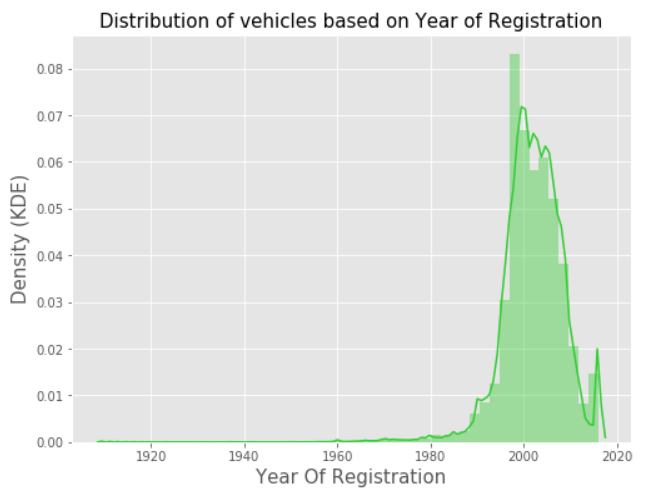
We can see from the price distribution of the model that the average price of SUV is higher than that of other cars, followed by other cars

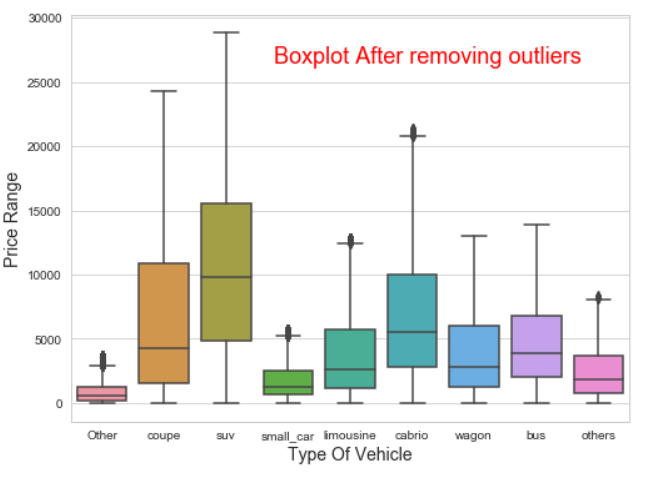
This analysis provides a vehicle price distribution based on vehicle types.

It is clearly seen that the data set contains a lot of outliers and conflicting data, since the year of registration cannot be more than 2015 and less than 1988 .

The prices for a boxed car depend on the type of vehicle after clearing the data set. Depending on the type of vehicle, you can see how prices change out of the box. low, 20, 40 (median), 60th percentile, high can be estimated from this graph.

Many problems with the reporting year, larger than 2015 and less than 1988, are eliminated to prepare a data set for analysis.



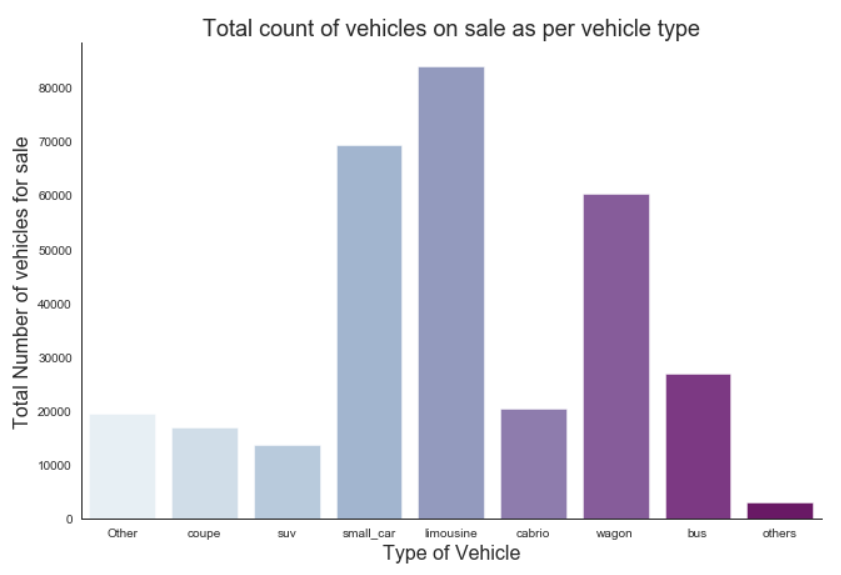


**Question2**

2. How many cars are available for sale across a given type's data set?

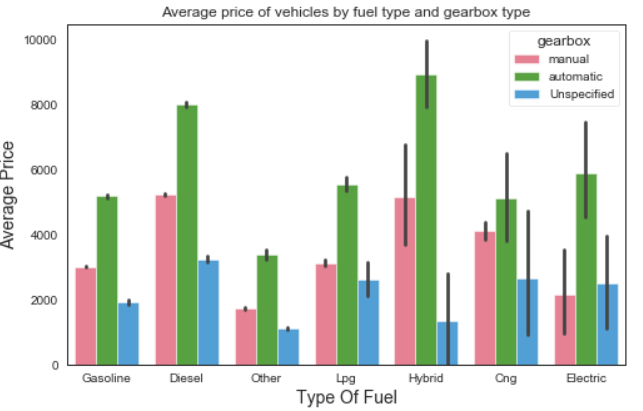
Suvs and Cabrio are the most expensive models at more than $6,000, compared with a mid-range price of $2,650 to $5,000 for Coupe, Bus and others. The cheapest models are the Andere and others with an average price of less than $1,800.

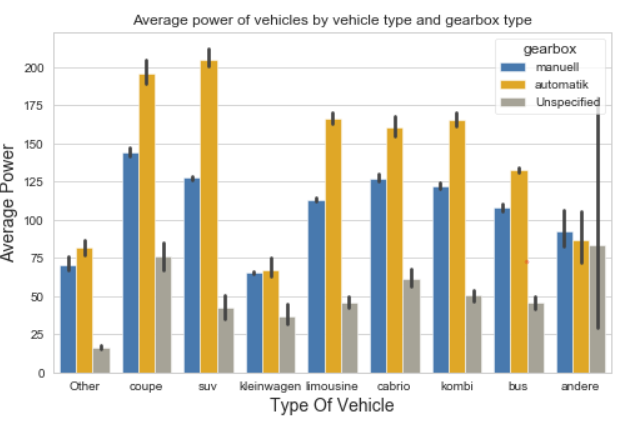
Automotive automation is a trend and a necessary stage in the development of automobiles. Automatic vehicles are the most expensive compared to manual and other unspecified gearbox types.



**Question3**

3. Analysis of the average price of vehicles based on fuel type and vehicle type?





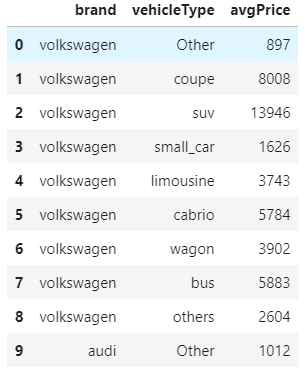
This analysis provides the average price of vehicles based on the type of vehicle fuel, and also depending on the type of vehicle.

Average prices for hybrid cars are the most expensive compared to other fuels like diesel and gasoline.

The type of SUV with gearbox with automatic transmission has the maximum power, and the Kleinwagen with the smallest.

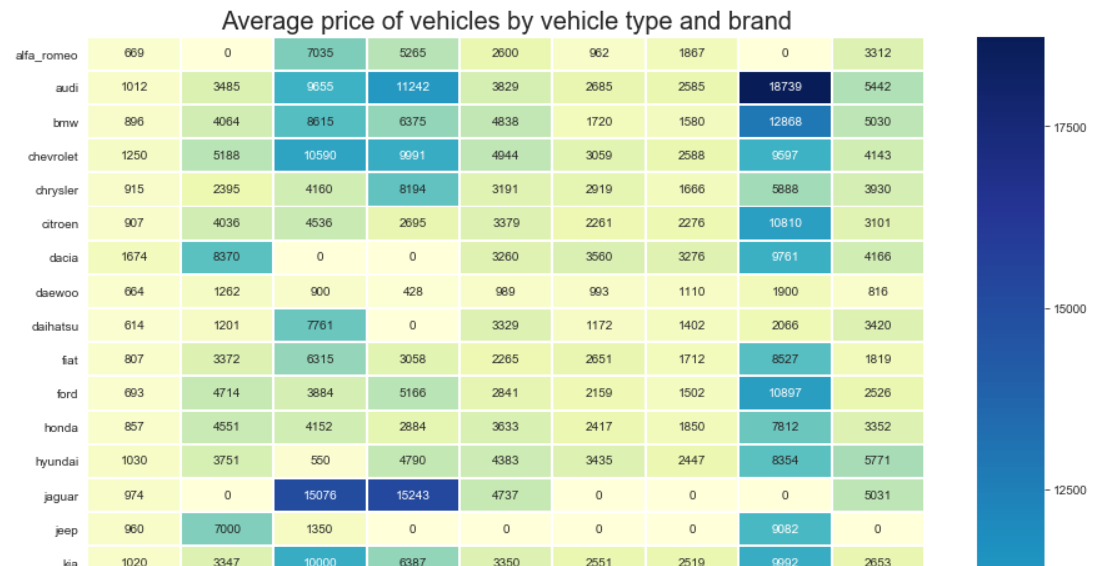
This analysis gives the average price of a vehicle based on fuel type and vehicle type.

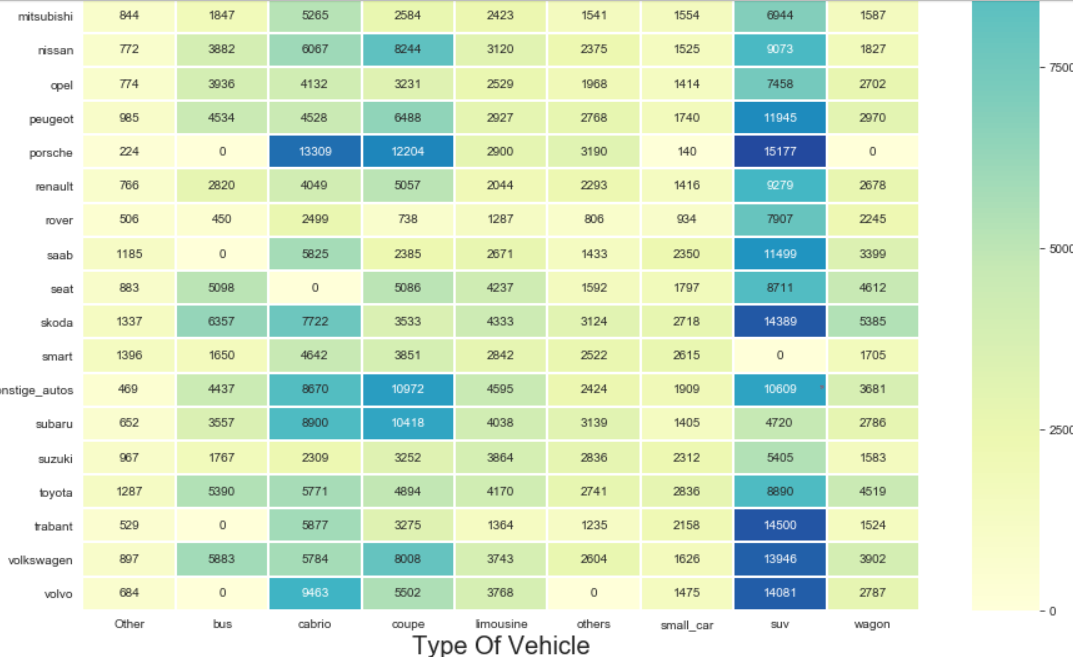
Compared with other fuels such as diesel and gasoline, the hybrid car's average price is the highest SUV model featuring the highest power and smallest Kleinwagen gearbox automatic gearbox.It can be seen that hybrid cars are the future trend but they also have expensive disadvantages.



**Question4**

4. Average prices of the brands and types of cars found in the analysis data set?





Bar graph of vehicle average price based on fuel type and vehicle type.

As you can see, Audi and type suvs are the most expensive drivable cars on the market.Porsche and Kleinwagen are two of the cheapest cars on the market.

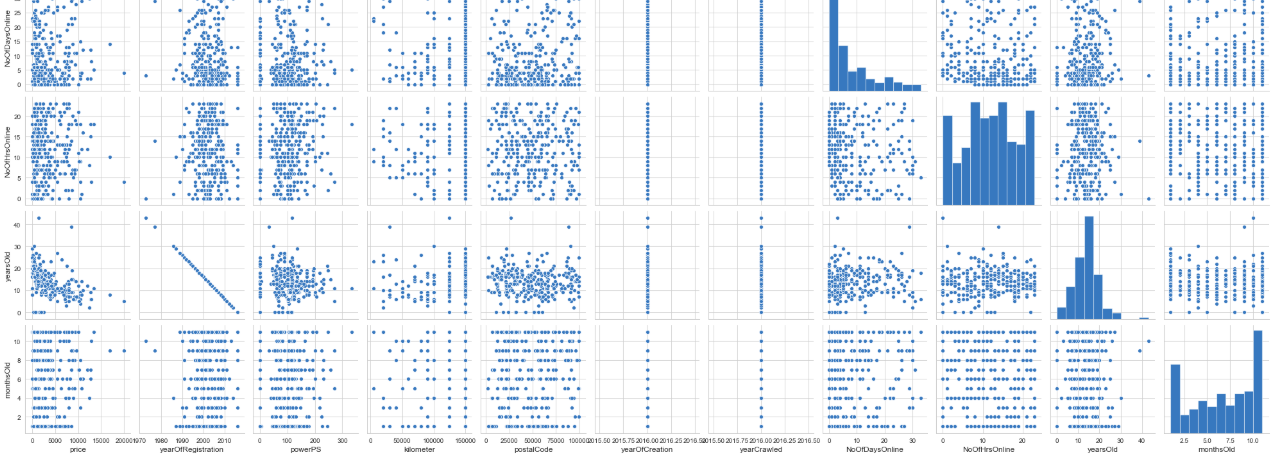
This analysis gives you the average price of the vehicle brand and its types, which can probably be found in the data set.

Automobiles of the brand Audit and SUVs are the most expensive cars available for sale.

Porsche and Kleinwagen are the cheapest cars available for sale.

**Question5**

5. Create a pair of graphs that contain all the columns that contain the metrics?



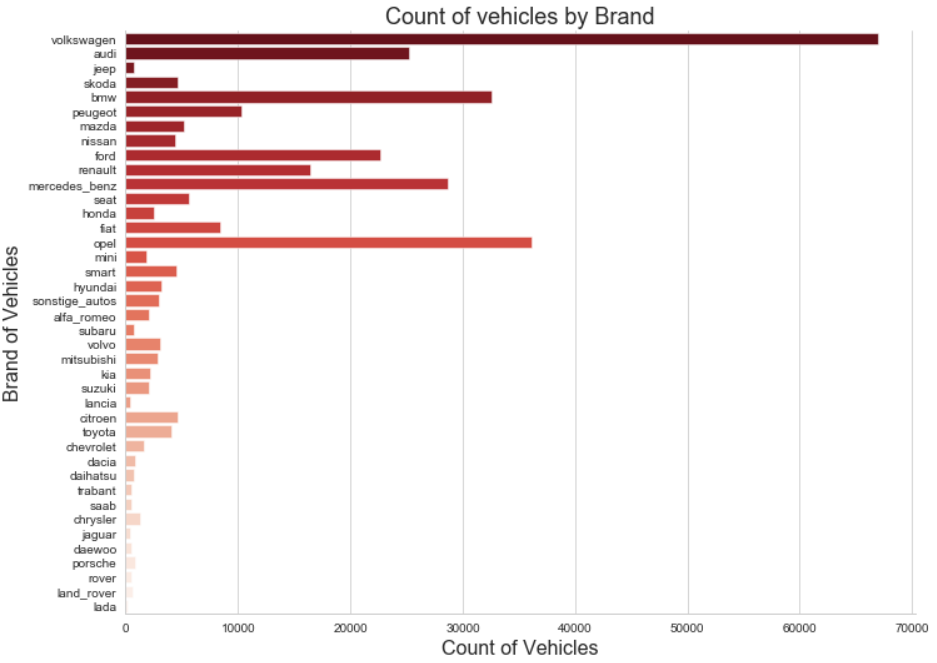
Get a distribution overview and detect the first correlation between the columns

We can get a clearer visualization with sample volume up to 350.Using seaborn to create an analysis we can see the connections between the individual data more clearly.This diagram contains the pairplot for all columns of the measured data

We can see the overall change in the data, such as the price of the vehicle and other factors are gradually decreasing

**Question6**

6. According to the data, which brand sells more cars?



The analysis gives the number of cars available for sale in the entire dataset based on a particular brand.

Sport utility vehicles and cabrios are the most expensive, with coupes over $6,000, buses and so on being moderately expensive, between $2,600 and $6,000, as the cheapest is Andere and other prices average less than $2,500.Among them, Volkswagen, opel and BMW are the three brands with the largest sales volume, decreasing in turn, while Volkswagen has the largest sales volume.

### **Conclusion:**

The higher the p value, the higher the price

The higher the mileage, the lower the price

The higher the registration date/year, the higher the price

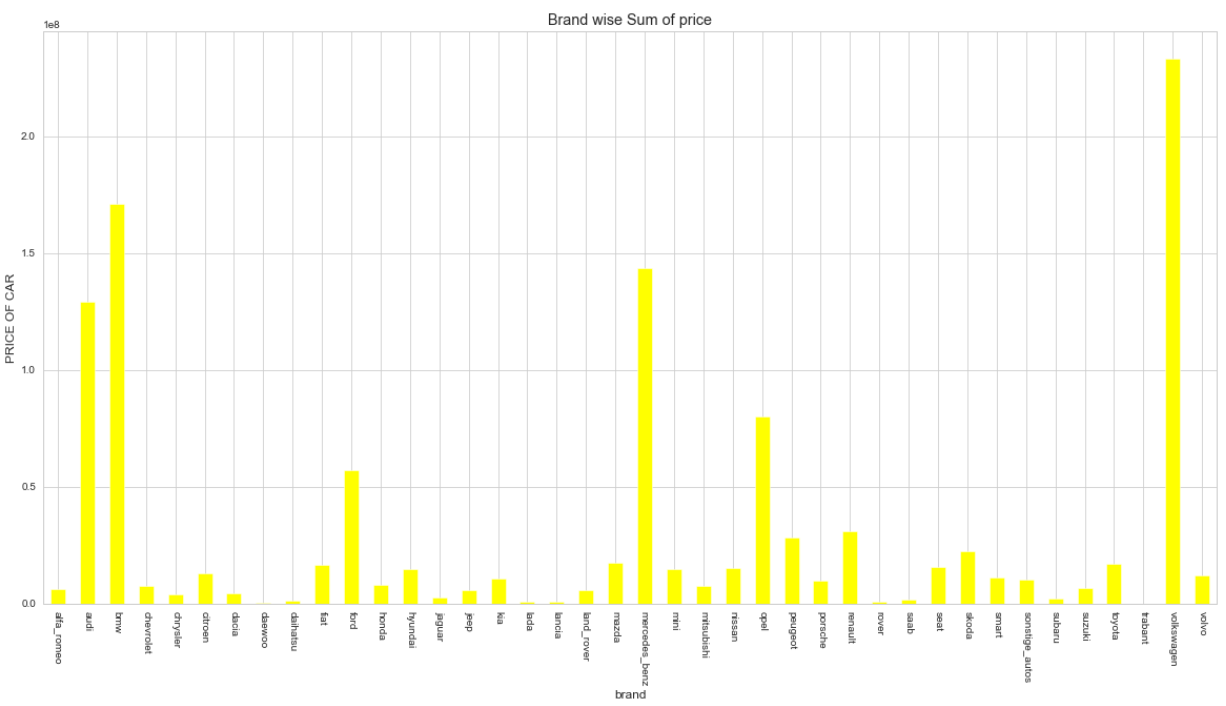
we analyzed the data. On this data, we saw which types of cars are sold more and which cars are sold less. And which cars are good on which machine is most interested in people

**Linear regression (car price)**

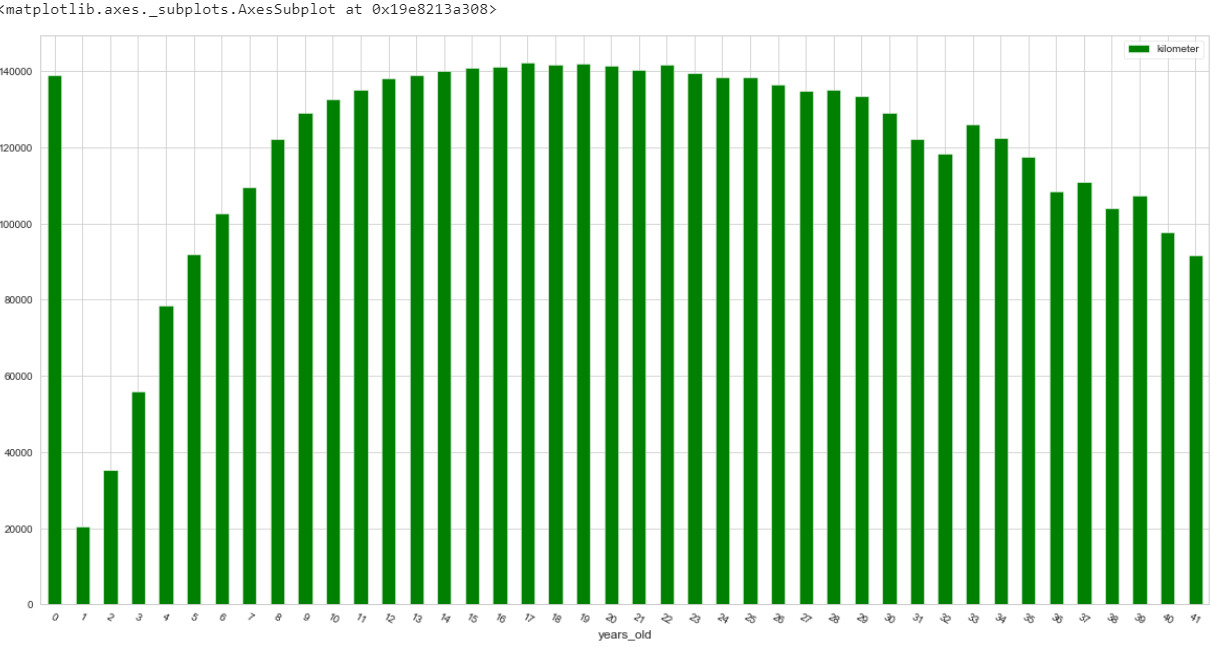
Considering the various characteristics of the car, the car price is predictable.

Given the characteristics/features of the car, the sale price of the car is to be predicted.

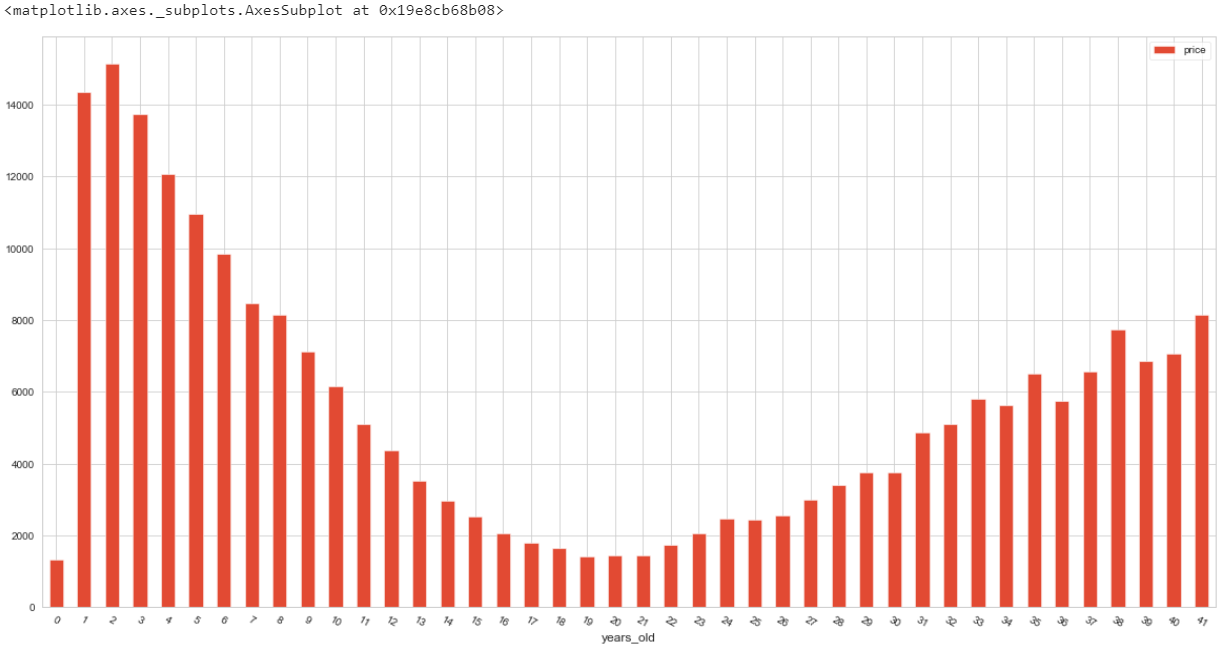
With the addition of various parameters of the vehicle, we finally obtained a good prediction result (MSE and R2, the evaluation indexes of the regression algorithm). We can see more detailed data in the figure



From the total brand price, we can see that Volkswagen has a low price but exceeds other vehicles in terms of quantity



Volkswagen, opel and BMW are the three brands with the largest sales volume, decreasing in turn, while Volkswagen has the largest sales volume.



From the picture we can see that the price of cars will get lower and lower with time