**CAR SALES EXPLORATORY DATA ANALYSIS WITH PYTHON**

**Introduction**

The analysis done was to compare and correlate manufacturers of cars and other variables that would determine the rate of purchase and the choice of brand to purchase.

The analysis seeks to proffer inference to why some brands are more purchased than others and some factors that would influence the purchase of those products.

The analysis was done using the PYTHON PROGRAMMING LANGUAGE and the jupyter notebook

**Data Sourcing**

The dataset was sent to an email from the facilitator Mr Joseph Elijah, the email contains two files; which was the dataset and the other was questions and guide used in analyzing and answering the questions relating to the dataset.

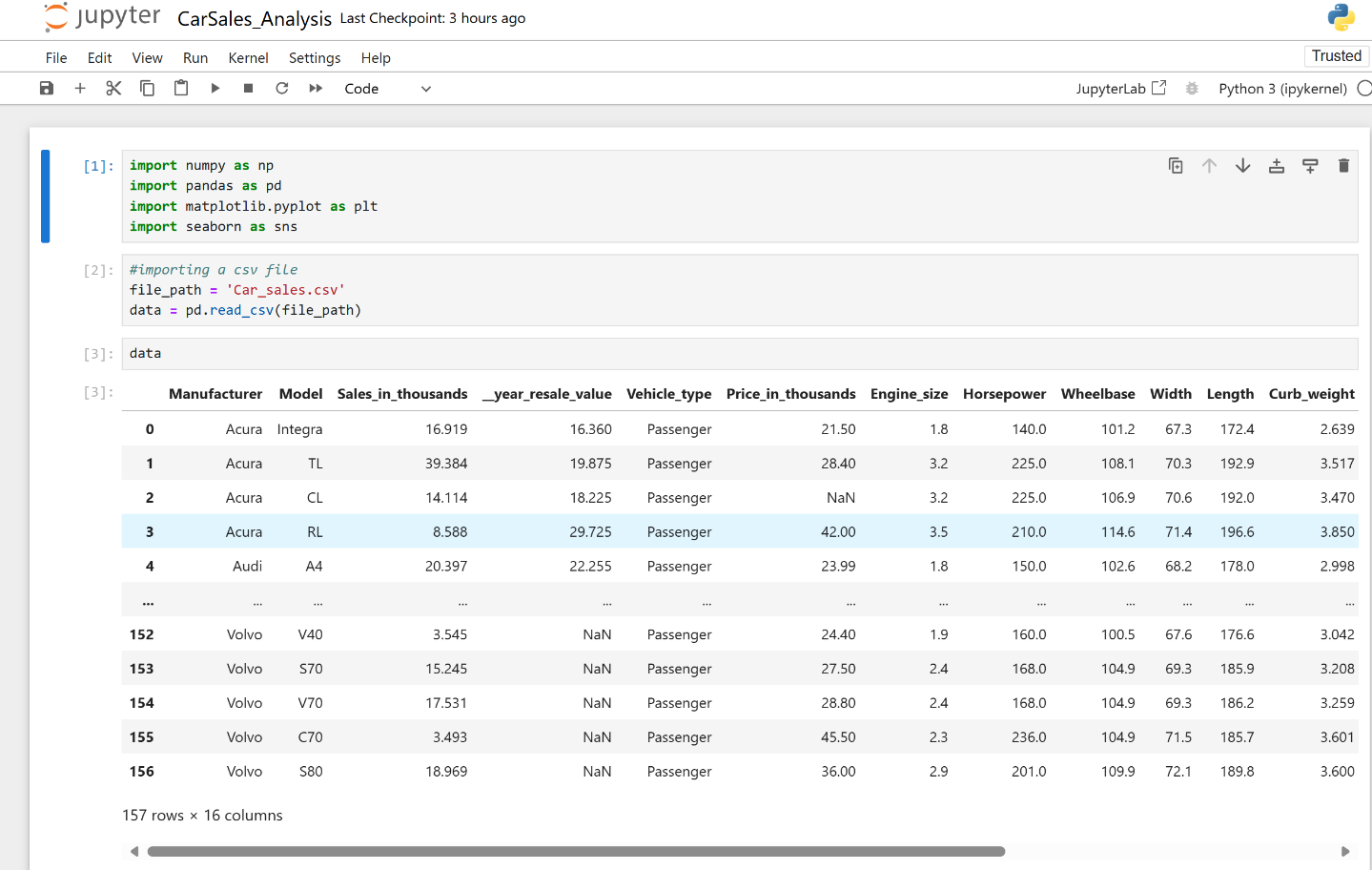
**Data Preprocessing**

The code for importing data libraries and file was imputed to prepare the python workspace

Numpy is imputed to work with numbers/integers, Pandas; to work on Series or Dataframe and also to alter data, Matplotlib and Seaborn are used for data visualization

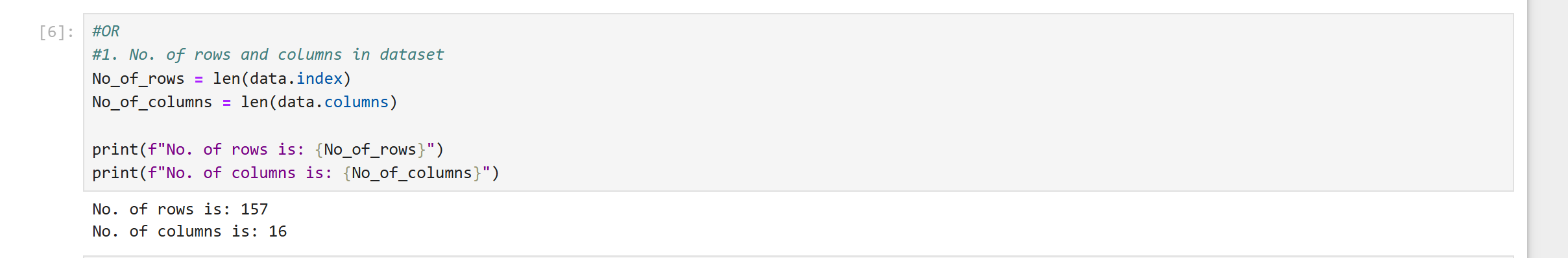
The file was saved in a csv format and was uploaded as such.

And the dataset was displayed as shown below;



**DATA EXPLORATION**

1. How many rows and columns are present in the dataset?



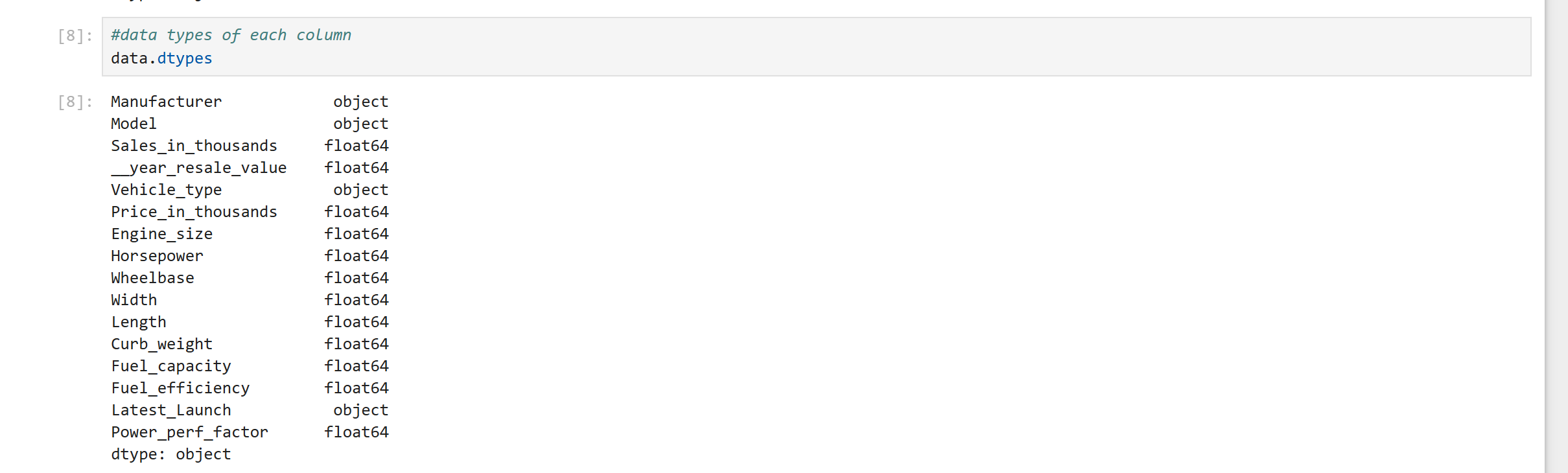
There are 157 rows and 16 columns from the dataset presented above

2. What are the data types of each column? Are the data types for each column

appropriate? If not change the data types of some columns to what they are

supposed to be

Using the “data.dtypes” code, all columns with string values are classified as ‘objects’ and columns with decimals or integers are classified as ‘float’ data types.



**DATA CLEANING**

3. Are there any missing values in the dataset? If so, how many and in which

columns? Fill the missing values in the year\_resale\_value with the average

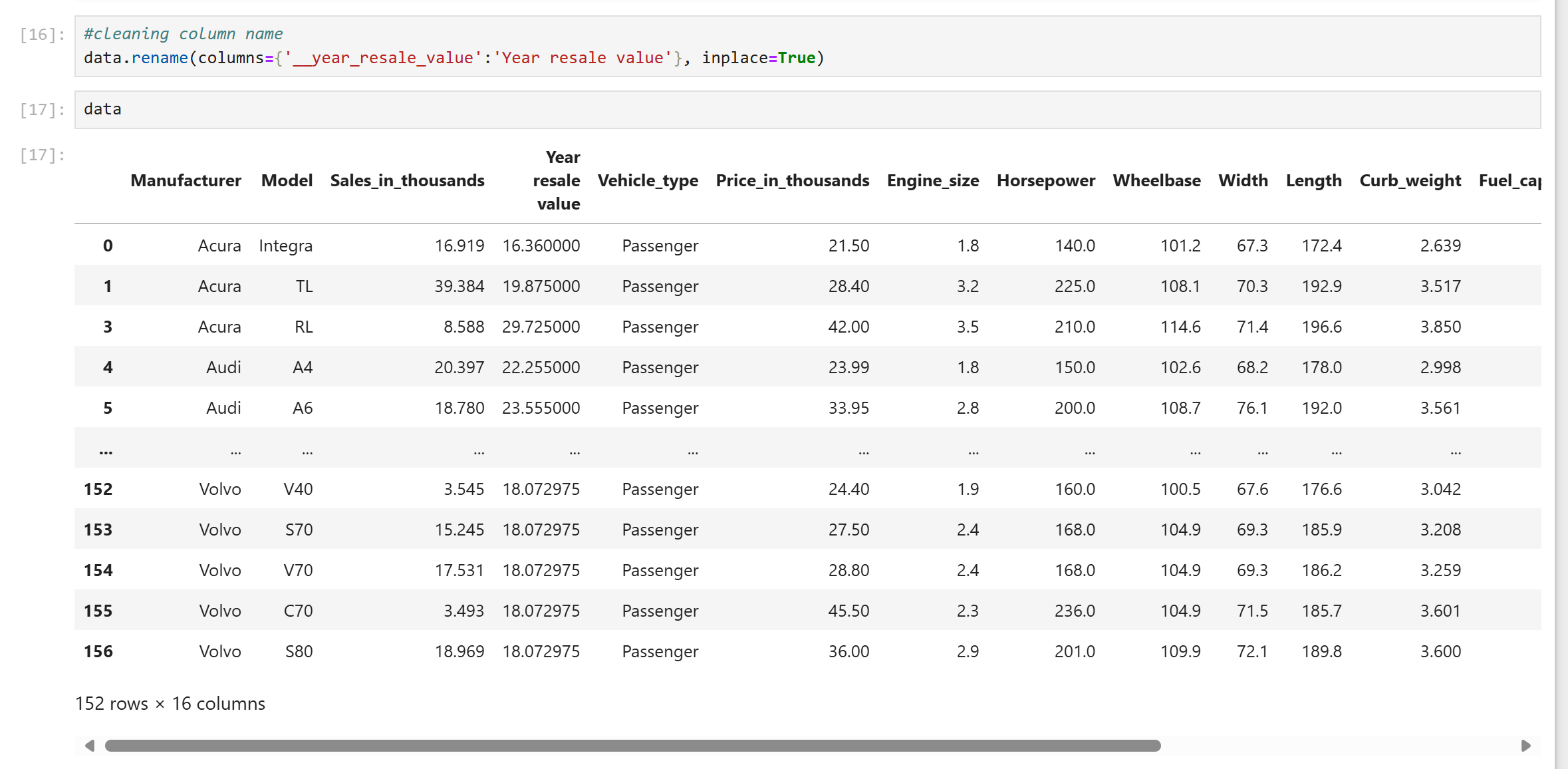
year\_resale\_value, then drop all other rows with missing values in the dataset



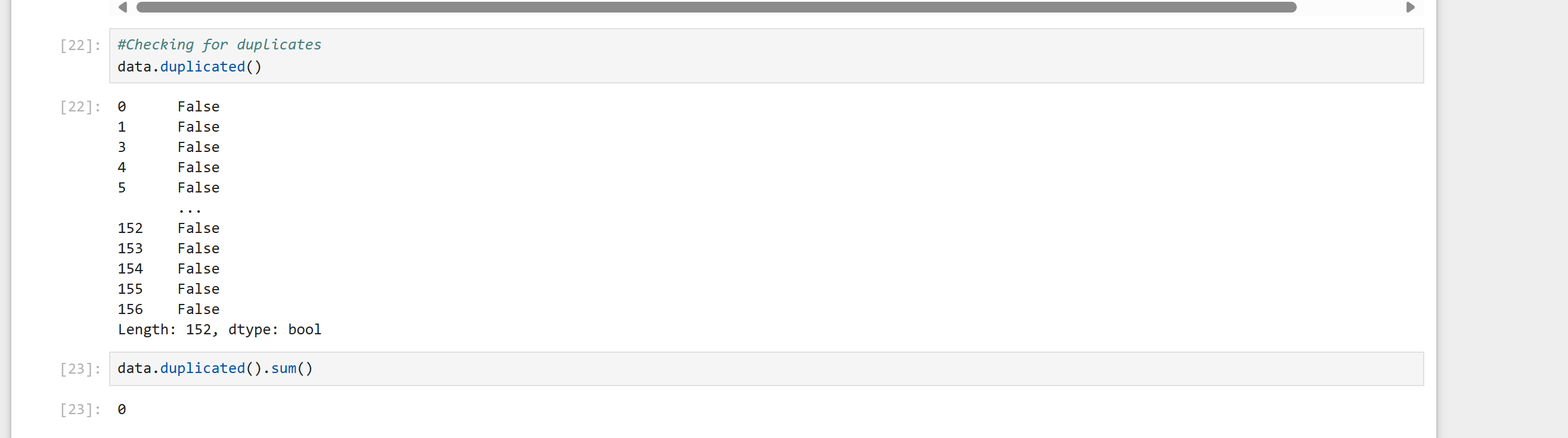
Data was cleaned by checking for and counting missing values using the data.isna().sum() and these were replaced by finding the average of the column affected and replacing the null values with them, afterwhich all other rows with missing values were dropped using the data.dropna()

4. Clean the name of the \_\_year\_resale\_value column by removing the leading

underscore



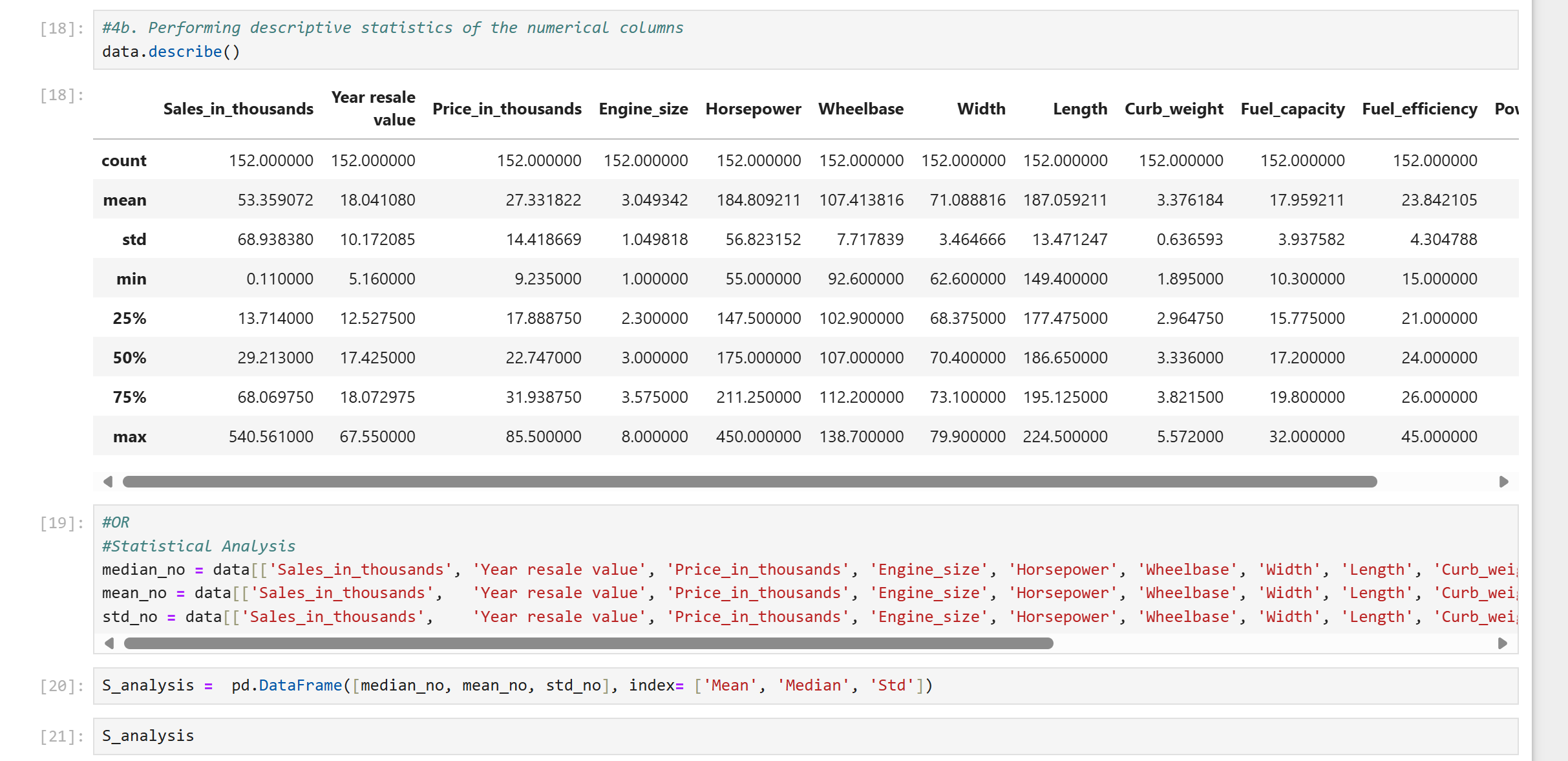
6. Check if the dataset has any duplicates. If there are, drop them



**DESCRIPTIVE STATISTICS**

5. Do descriptive statistics (mean, median, standard deviation) of

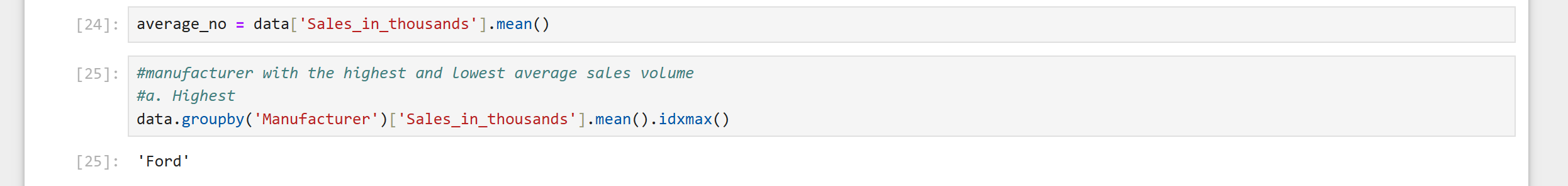
the numerical columns. What do you notice?

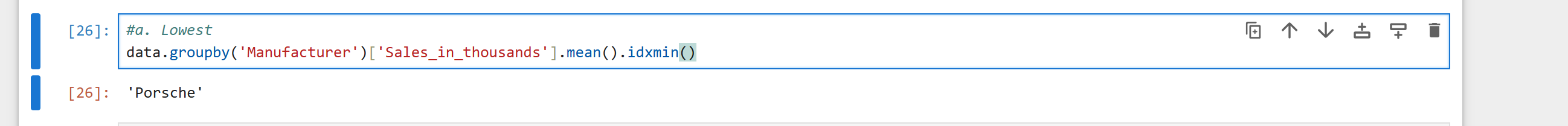


**EXPLORATORY DATA ANALYSIS**

6. Which manufacturer has the highest and lowest average sales volume?

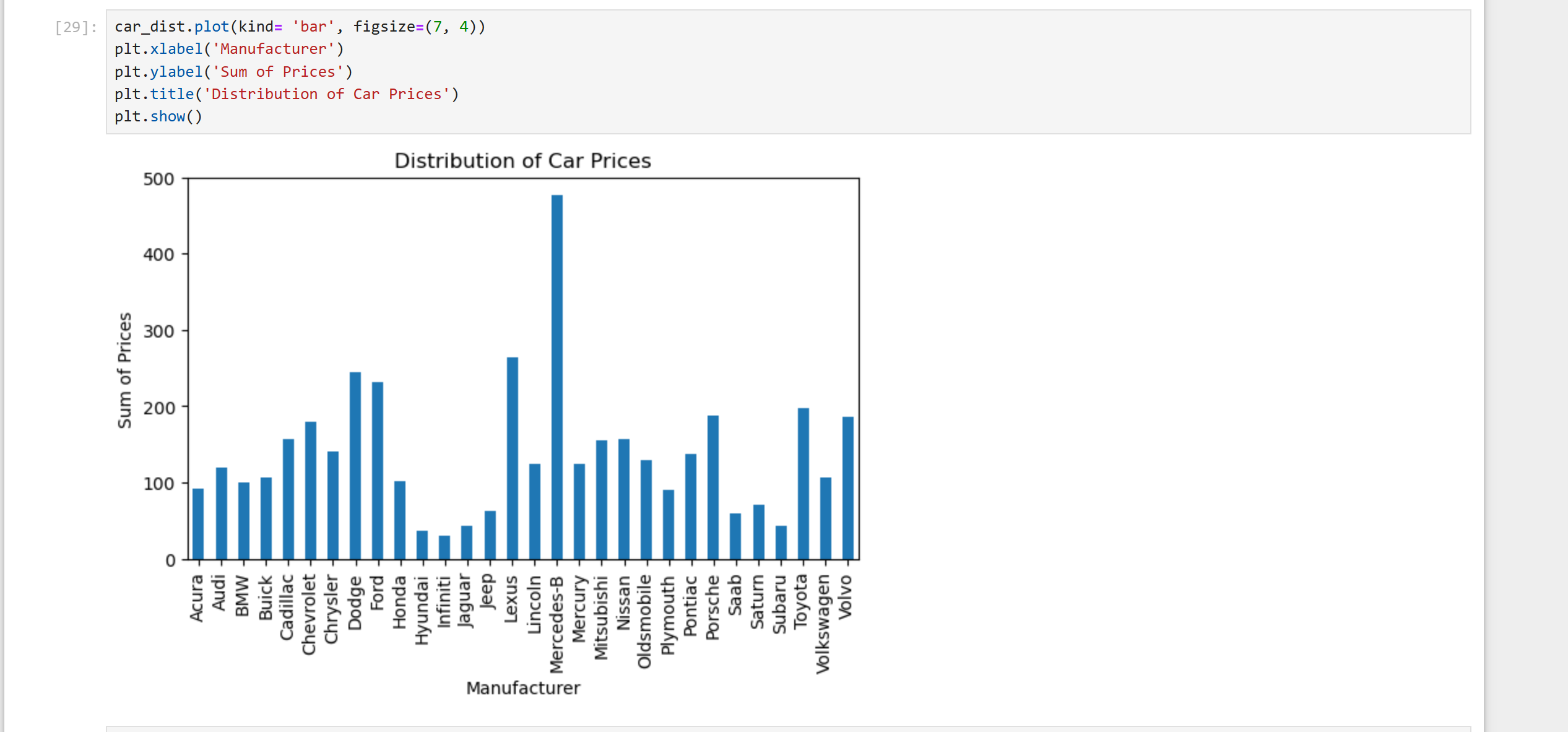
The ‘Manufacturer’ and ‘Sales\_in\_thousands’ are grouped using the function group\_by and the mean is also imputed and idmax()/idmin() is used to determine the highest and lowest sales volume





7. What is the distribution of car prices in the dataset?

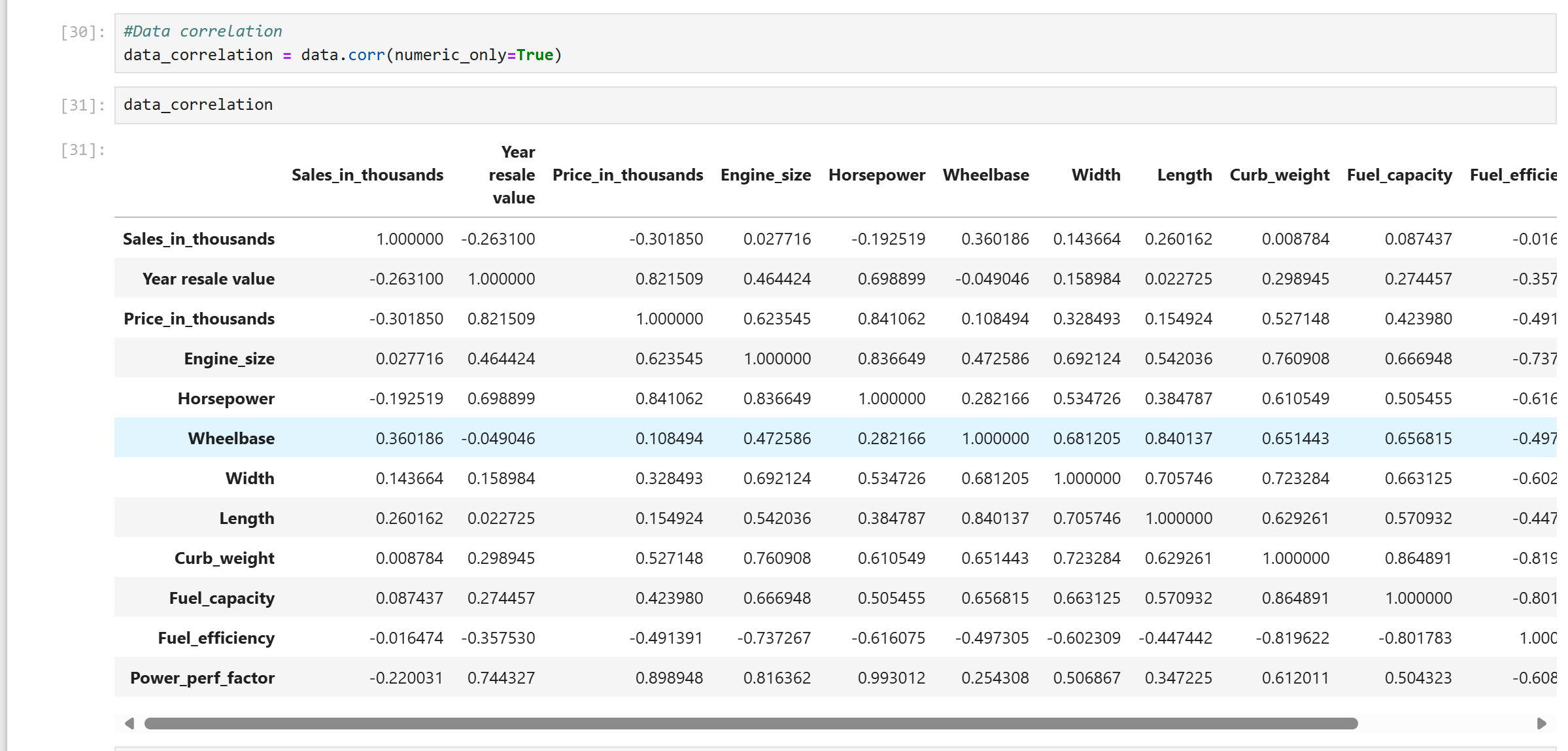




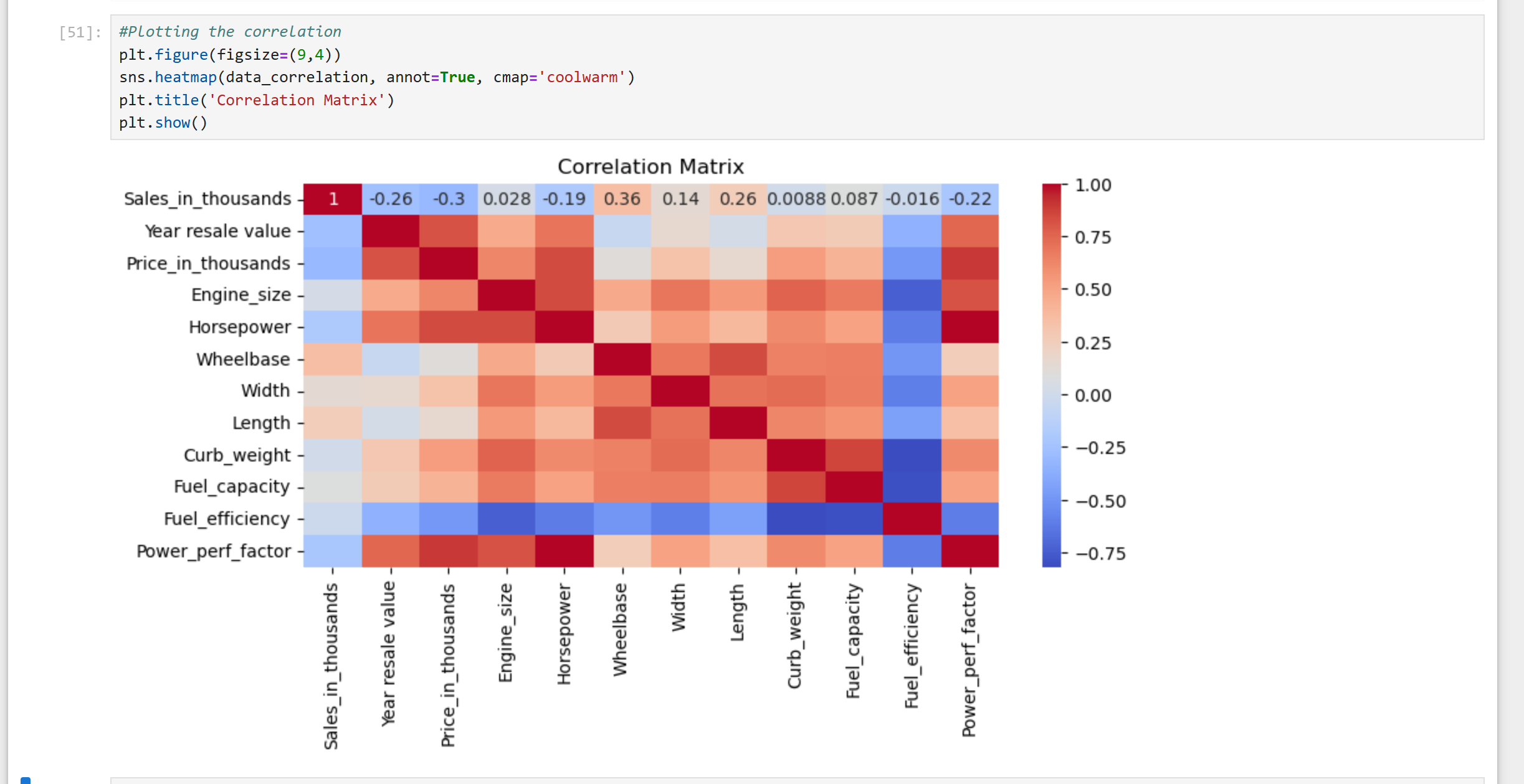
8. Plot a correlation matrix to see if different numerical variables in the dataset

correlate with each other.

Before plotting the matrix, a code to first create the correlation is imputed



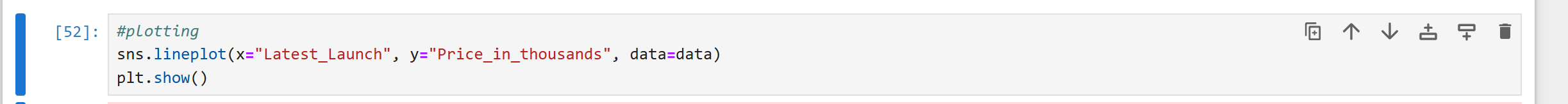
Then the matrix is plotted

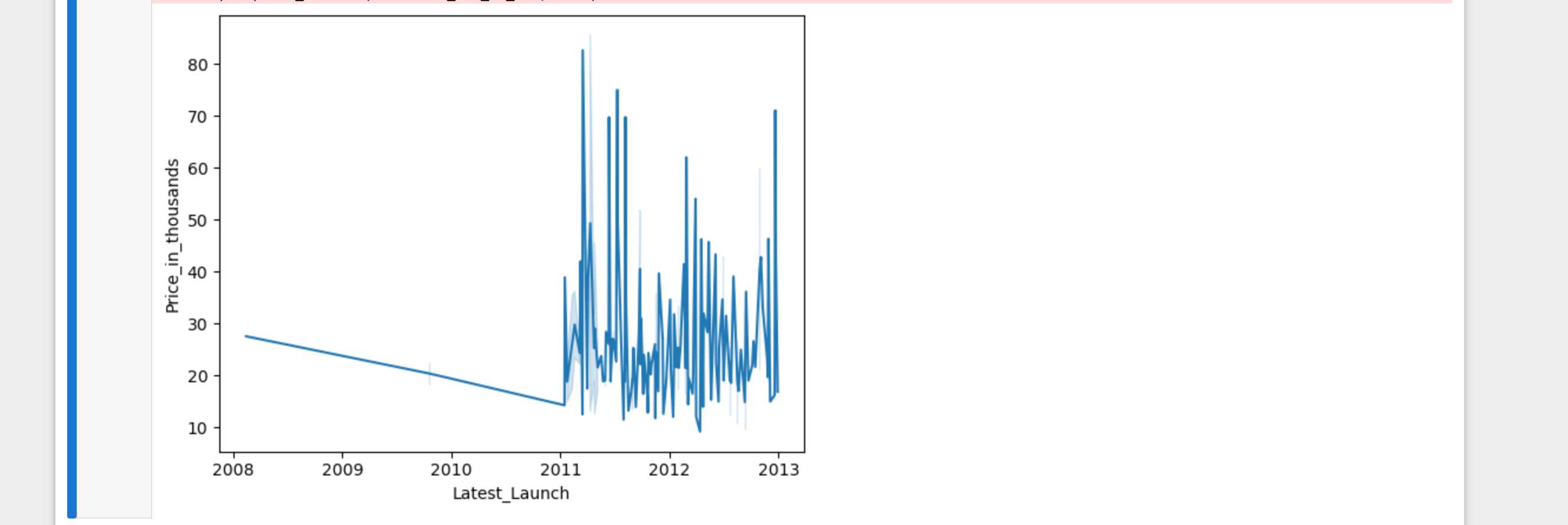


9. Plot the relationship between price and latest launch year.

To plot the relationship between price and launch year, the data type of latest launch had to be changed first

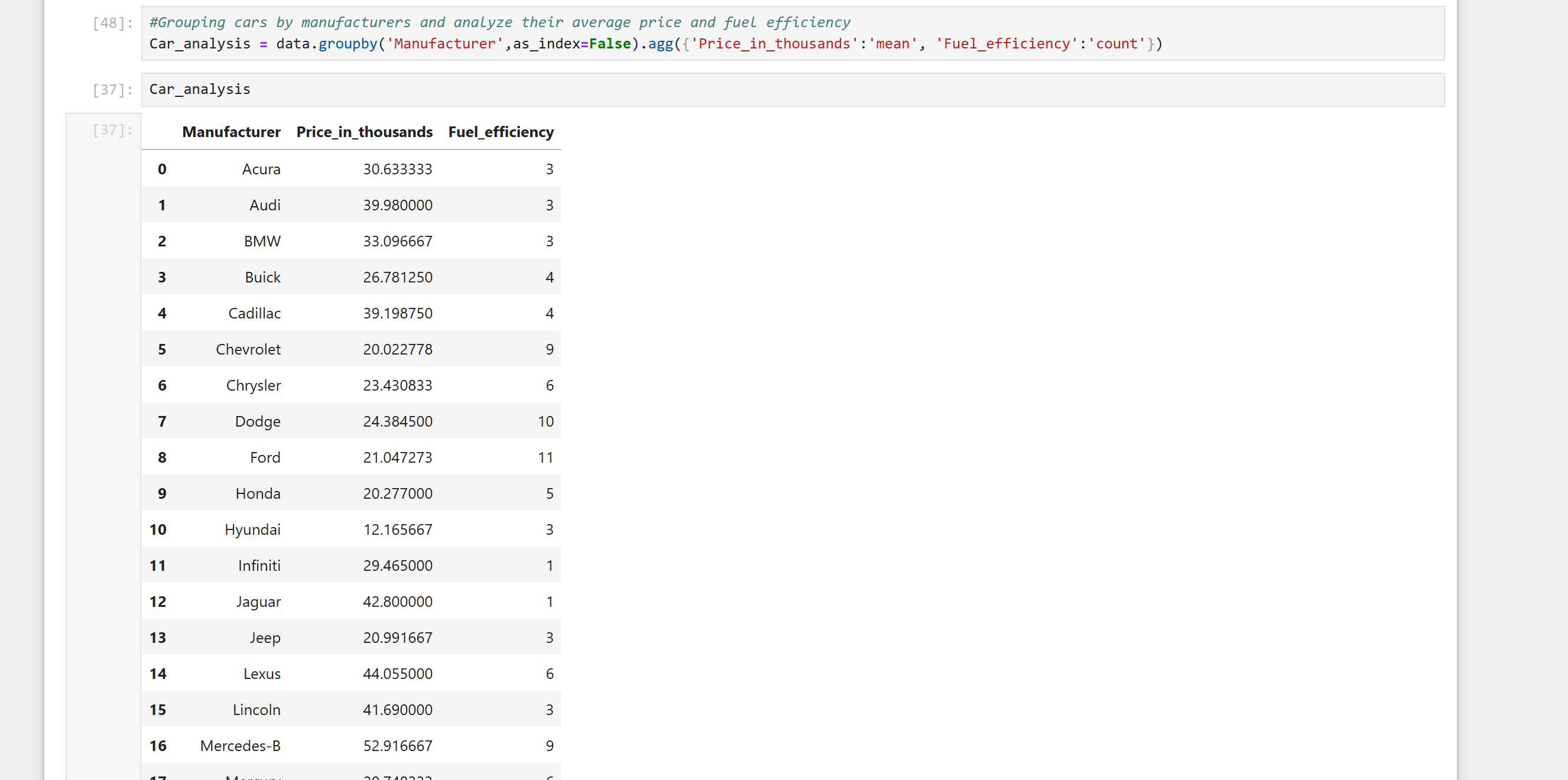




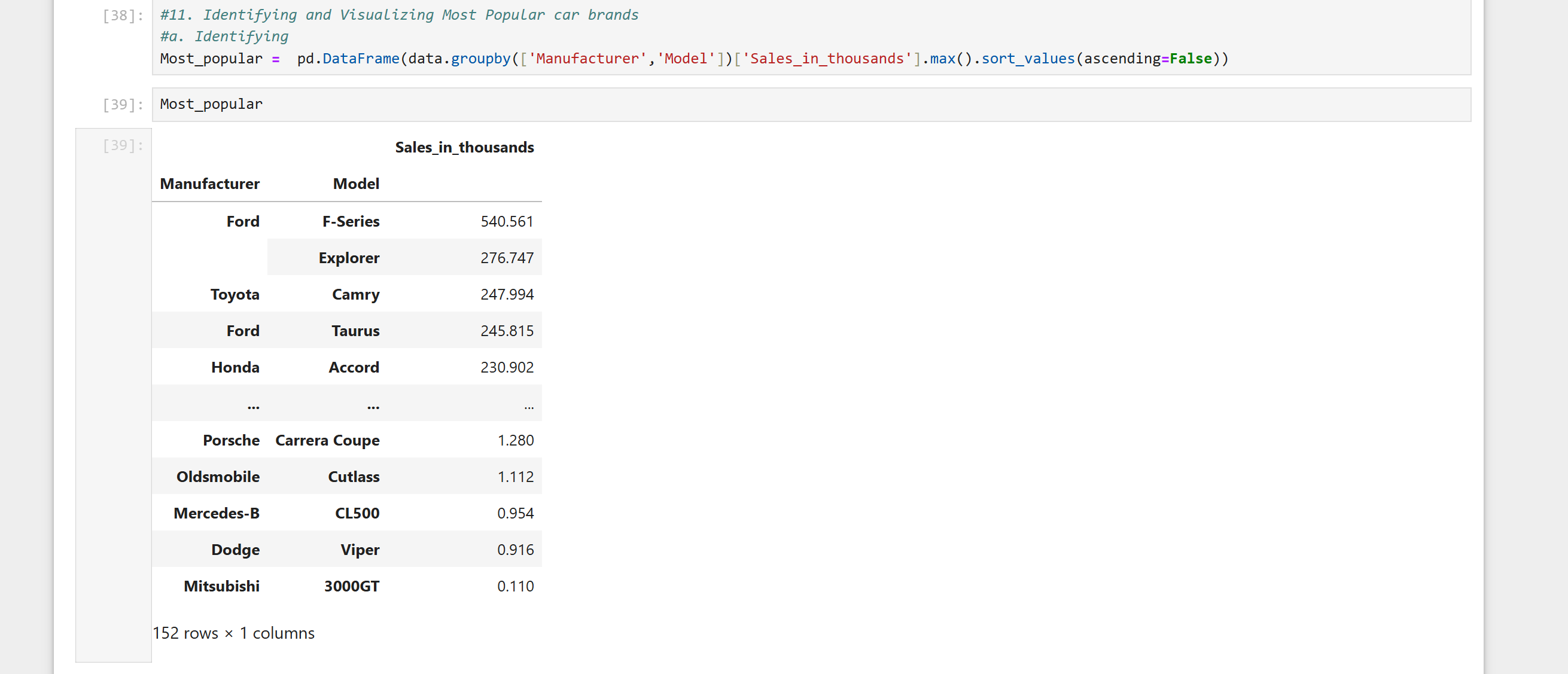


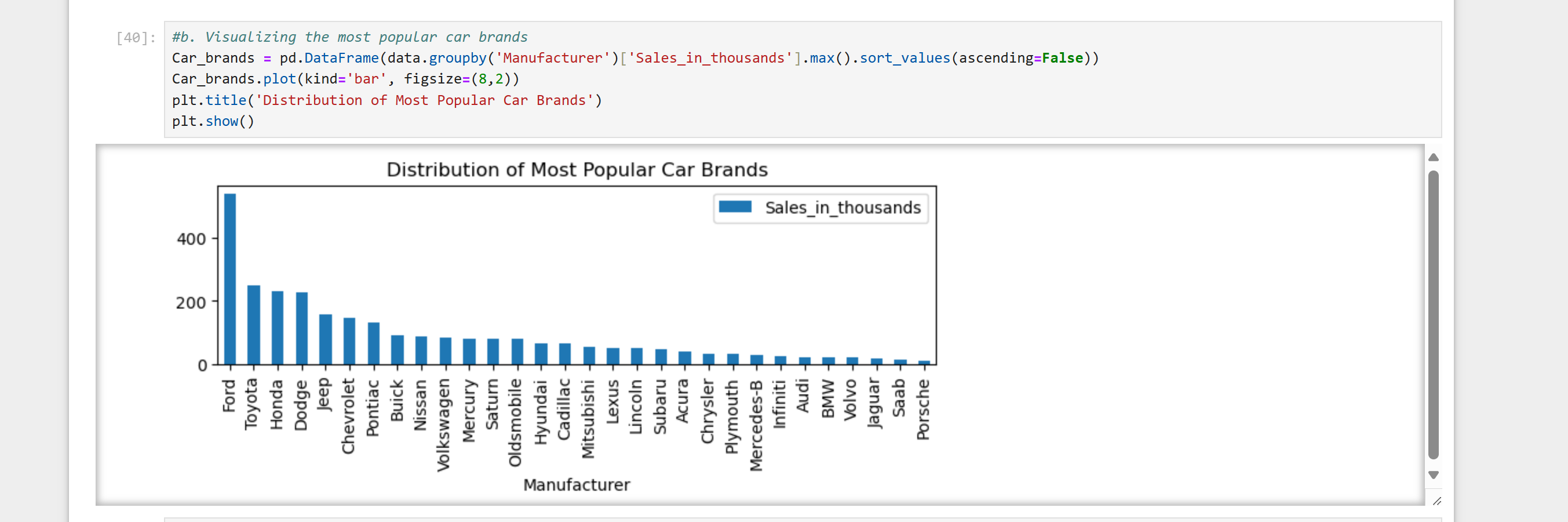
10. Group cars by manufacturers and analyze their average price and fuel

efficiency.



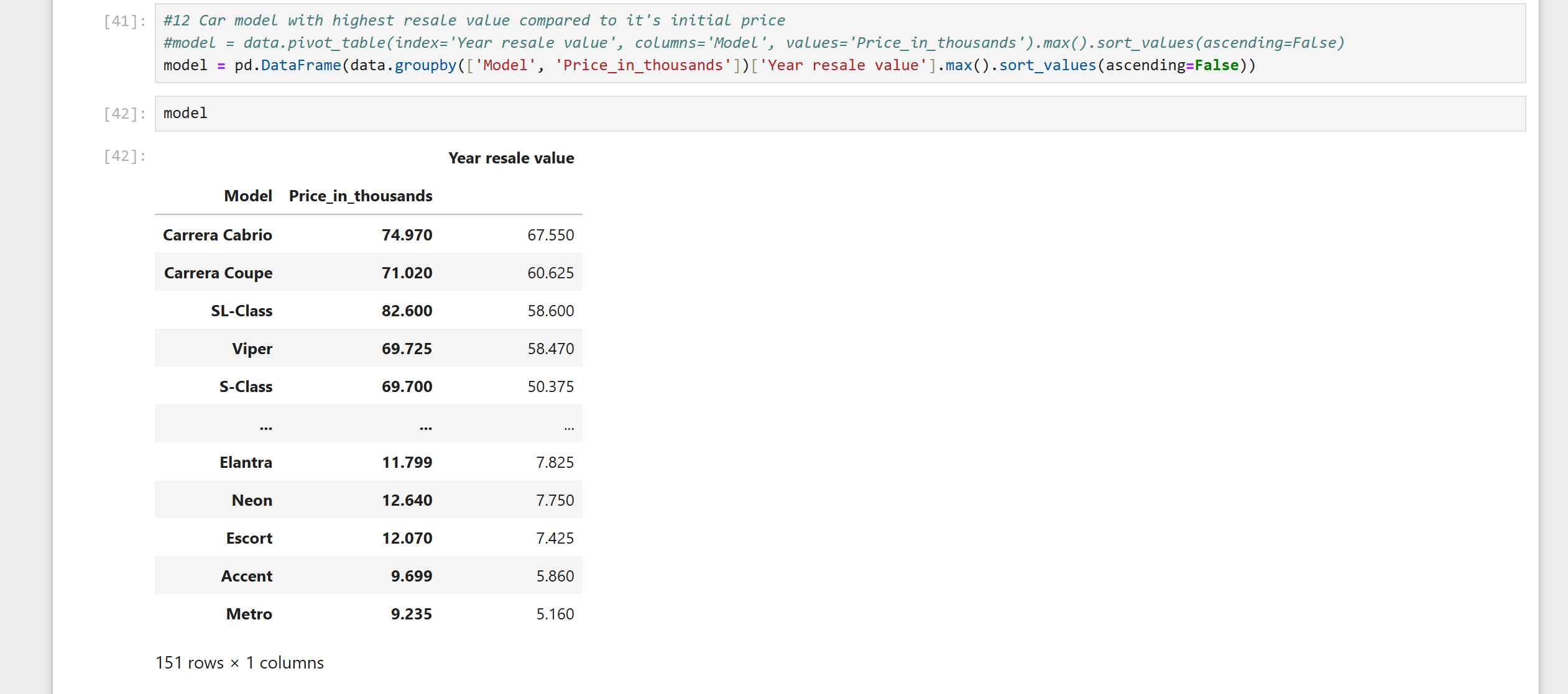
11. Can you identify and visualize the most popular car brands?





From the analysis, the F-series from the Ford company is the most popular car brand

12. Which car model has the highest resale value compared to its initial price?



13. Find the top 3 fuel-efficient cars with an engine size above 2.5 liters.



The top 3 fuel-efficient cars with engine above 2.5 litres are Chevrolet Impala, Chevrolet Monte Carlo and Mercedes-B CLK Coupe

14. Which Lexus model has the highest horsepower?

