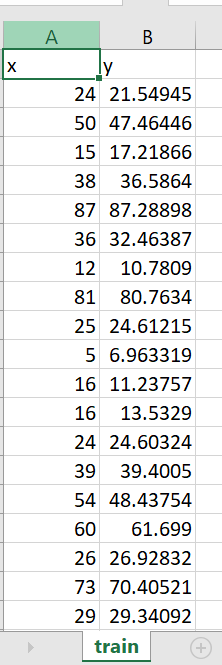
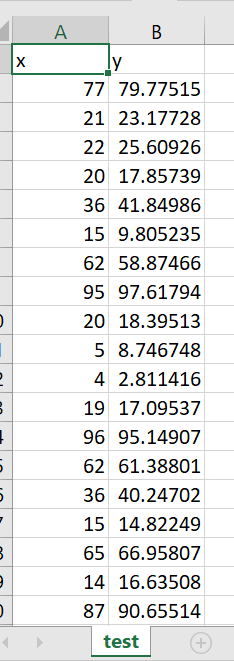
**The Machine Learning & Deep Learning Show**  
**Assignment 01**Submission Date: 9th June 2022

Name: **SADIA ANZUM** RID: **23**

**Problem 1: Implement a Univariate Linear Regression after finding your own data**

**Task 1: Dataset**  
*Screenshot of the dataset.*

* *

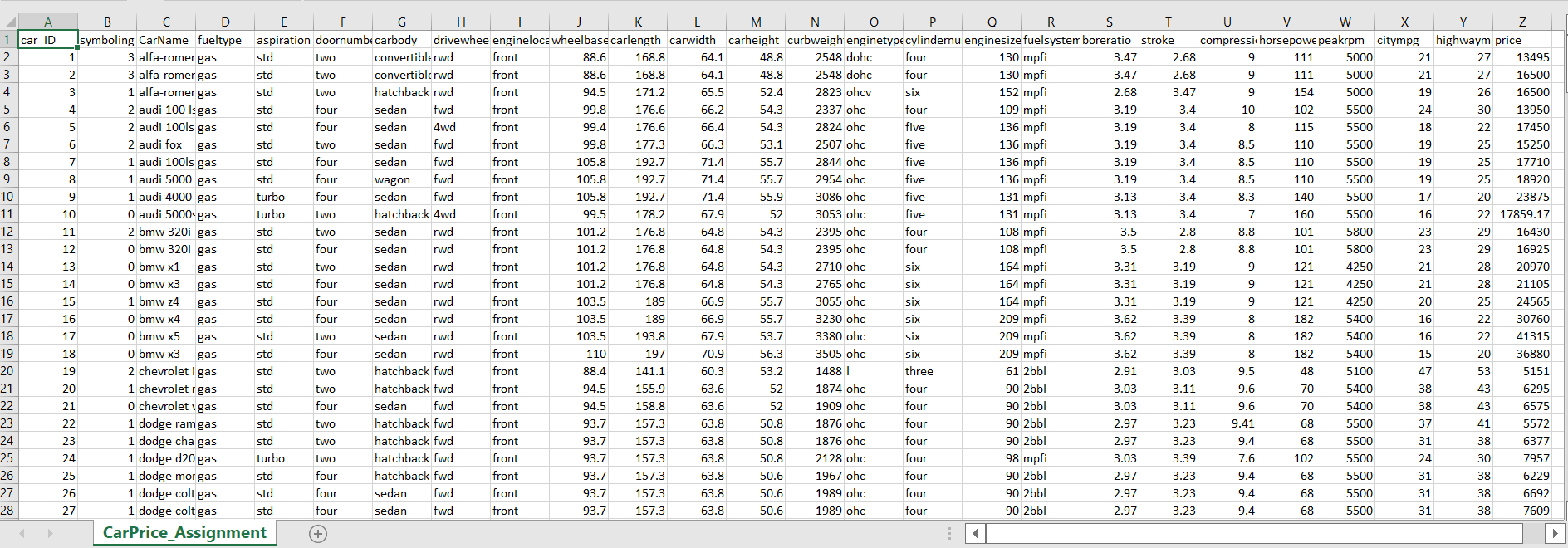
*Link to the source of the dataset:* [*https://www.kaggle.com/datasets/andonians/random-linear-regression*](https://www.kaggle.com/datasets/andonians/random-linear-regression)**Task 2: Coding**   
 *Link:* <https://colab.research.google.com/drive/1N5LQiEsASaUHLpNSKWuAOaCDKnCyJMJD?usp=sharing>

**Task 3: Code Explanation**

After uploading the train dataset (train.csv), I checked for any missing values. There was 1 row having Y column as null so I dropped the particular row. Hence, number of rows went from 700 to 699. Since there were separate csv files for train and test sets, I imported them separately and split them using iloc. X and y signify independent and dependent variables respectively. Visualization has been done by scatter plot using Matplotlib. Model used is Simple Linear Regression.

**Problem 2: Implement a Multi-variate Linear Regression after finding your own data**

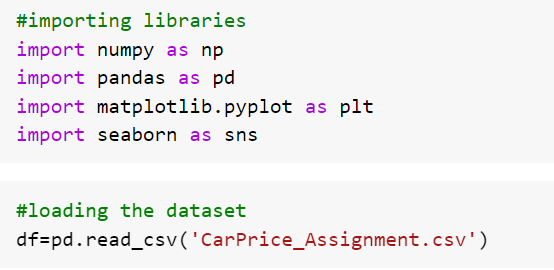
**Task 1: Dataset**



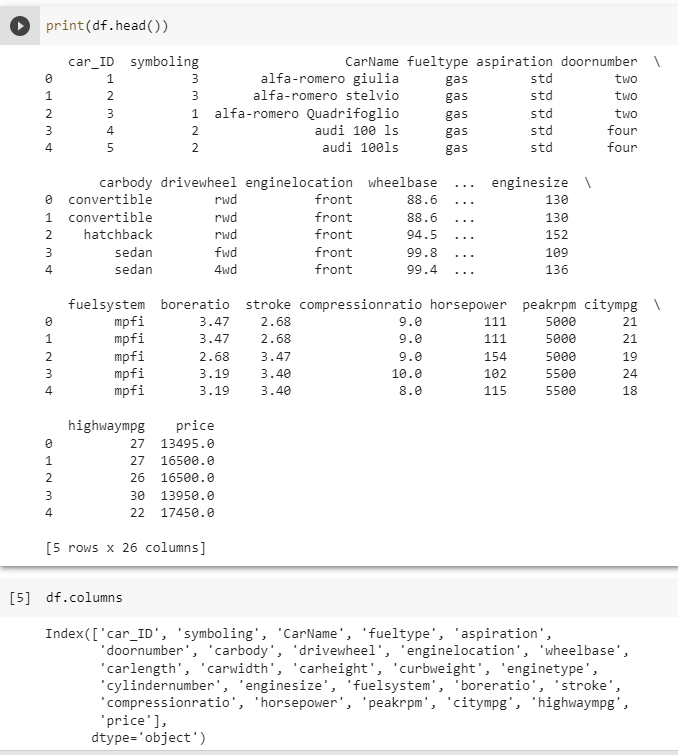
*Link to the source of the dataset:* [*https://www.kaggle.com/datasets/hellbuoy/car-price-prediction*](https://www.kaggle.com/datasets/hellbuoy/car-price-prediction)**Task 2: Coding**   
 *Link for the drive:* [*https://colab.research.google.com/drive/1KtnPwp9kbEs6kzivwMmfXzelPxWN4f4p#scrollTo=m2-6Ts87c1F1*](https://colab.research.google.com/drive/1KtnPwp9kbEs6kzivwMmfXzelPxWN4f4p#scrollTo=m2-6Ts87c1F1)

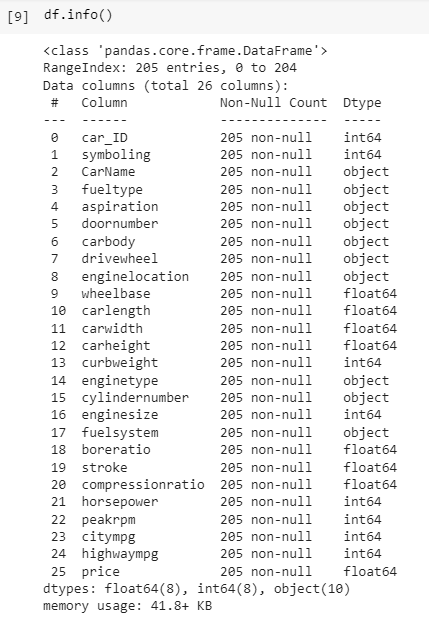
**Task 3: Code Explanation**

* Importing necessary libraries and loading the dataset after uploading the csv file

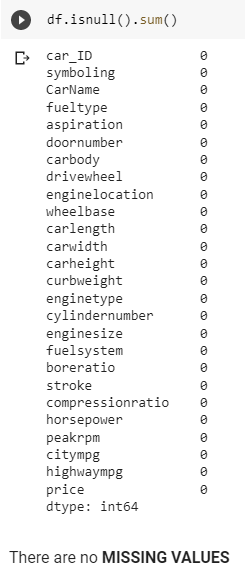


* Getting an overview of the dataset.

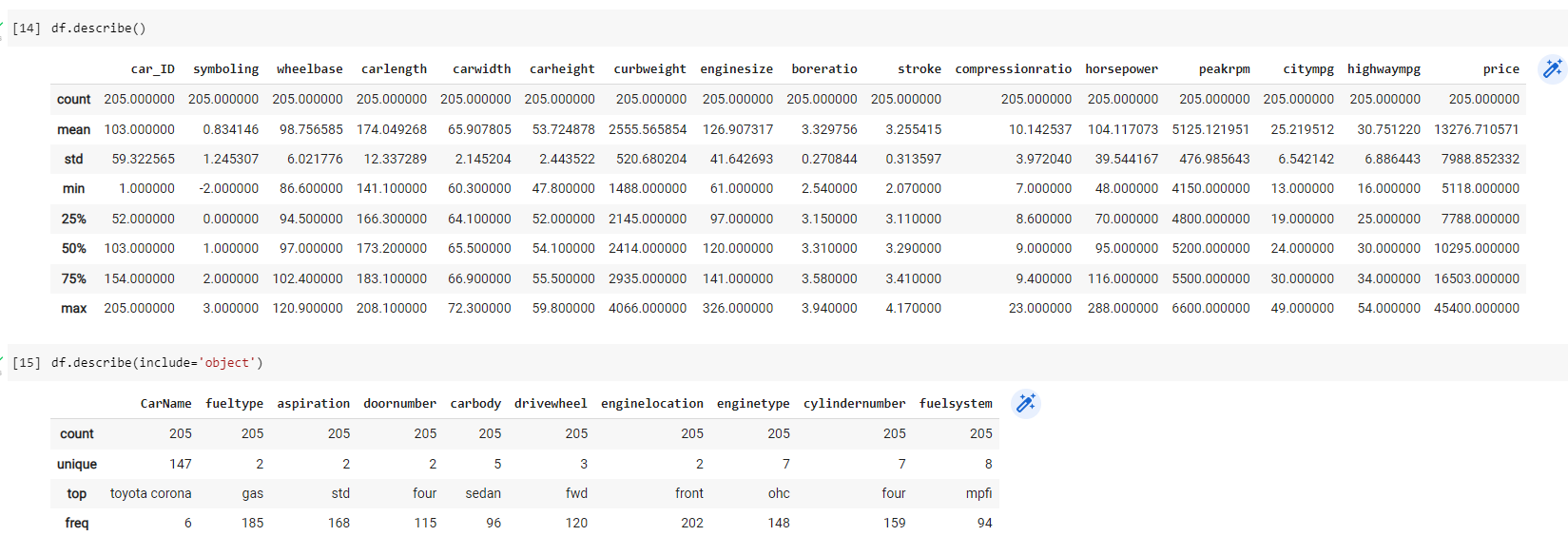




* Checking for Missing Values

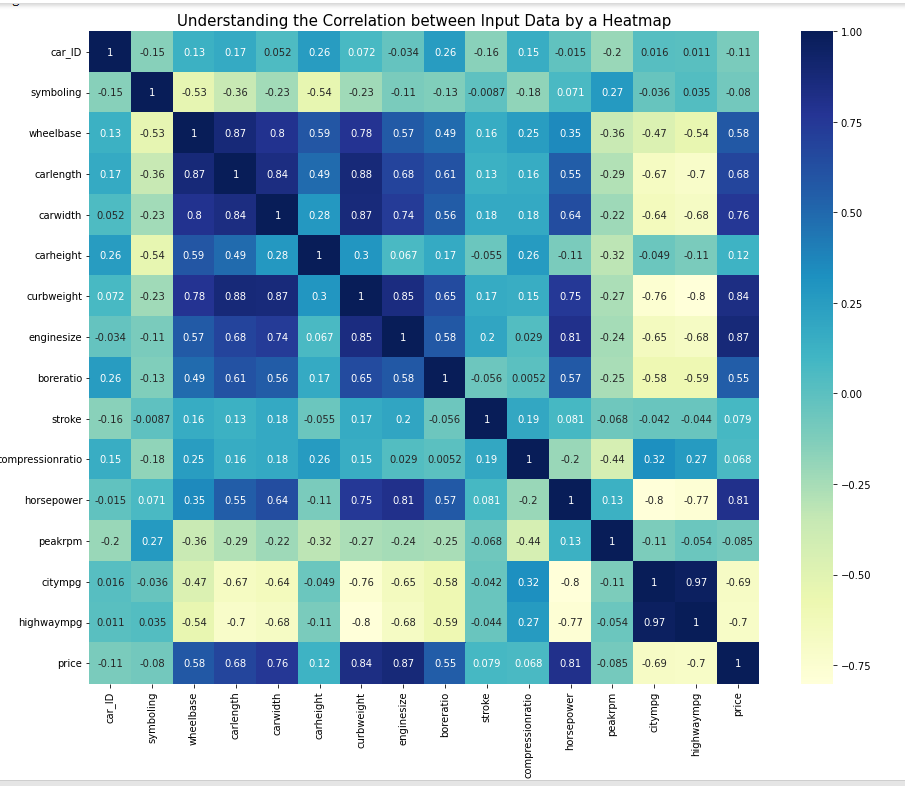


* To understand the distribution of the data



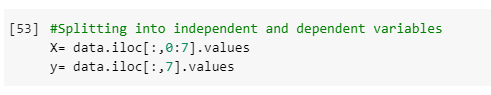
* I decided to drop the columns that have less than 5 unique values or those having low data range or some features that in practical would not affect the price of the car, for example- car\_ID.

For confirmation, I plotted a heatmap to understand how all the variables are correlated with one another and also, the target variable ‘price’.



* After analysing the heatmap, the appropriate columns are dropped and then the data is split into Dependent Variable ‘y’ and Independent Variable ‘X’.





* Data is split into train and test sets and then trained on the model.

