**Car Price Prediction**

**Submitted By:**

**V.V.S.N.SAI MOHAN - AM.EN.U4CSE20275**

M. VISHAL REDDY - AM.EN.U4CSE20248

S. MANIKANTA KUMAR - AM.EN.U4CSE20268

K. NAVEEN KUMAR - AM.EN.U4CSE20238

B. JAGADANANDA THARAKA - AM.EN.U4CSE20217

**ABSTRACT**

I need a model which predicts the price of a used cars with given constraints.

Accurate car price prediction involves expert knowledge, because price usually depends on many distinctive features and factors.

Typically, the most significant ones are present price, brand and model, age, mileage etc. The fuel type used in the car as well as fuel consumption per mile highly affect the price of a car due to a frequent change in the price of a fuel. Different features like exterior colour, type of transmission, safety, air condition, etc. will also influence the car price.

* ***Task*** *(T): Classify the price of that used car using the previous data knowledge.*
* ***Experience*** *(E): A corpus of data with the features to be learned and predict the price.*
* ***Performance*** *(P): Classification accuracy, the number of car prices predicted correctly out of all car data considered as a percentage.*

**Assumptions:**

Some ML models won’t predict 100% correct there is chance of wrong prediction. What if it predicts wrong…?

A) We should train model to get max correct prediction.

B) Finding the best regression algorithm, among Linear Regression, Lasso Regression, Random Forest Regression, Ridge regression, etc., for our problem would be a challenge.

C) Reconstructing the given dataset. Changing the categorical values into numerical form. And removing the unnecessary features.

**1. INTRODUCTION**

**Motivation:**

Some car buyers will suffer a lot of problems like dealer may charge amount more than the car value.

This model can help them to predict the actual and reasonable price of those used cars.

**Benefits of solution:**

If this model generates more than 0.9 accuracy, then it’ll be more helpful to people. Because when your model accuracy is more than 0.9 then it’ll almost give correct decision. This can be useful for dealers as well as individual people to buy used cars.

For example: Some dealers will set more amount of car value then the individuals may suffer a lot. So by using this individuals can get the car by the reasonable price.

**Solution Use**:

This solution may be a temporary solution for the given data. Because in future there may be some extra features added like for this data set. We have fuel\_type with petrol and diesel and in future electric vehicles may add to this so there must be some extra features considered to set the value of the used cars. The accuracy may drop down in future. So, this solution is for only petrol and diesel vehicles.

This solution is mainly for individuals who want to buy a car from owners directly without dealers. After completing this project we’ll post this in GitHub and LinkedIn so that it’ll be helpful for those who want to do projects and also the actual use like implementing this model for car value.

**2. Dataset finalization**

**80% of data set for training and 20% of data for testing. We collected 3 data sets suitable for our problem, we’ll do on any one of them.**

1.



- It is about the used cars description that given.

- There are 9 features:

Car\_name, year, selling\_price, present\_price, Kms\_driven, fuel\_type, seller\_type, transmission, owner.

- It has been used to predict the used cars price prediction.

**Source: Kaggle.com**

2.



- It is about the used cars description that given.

- There are 9 features,

Brand, price, body, Mileage, EngineV, Engine\_type, Registration, Year, Model.

**Source: Github.com**

3.



- It is about the used cars description that given.

- There are 13 features (with empty cells),

Name, Location, Year, Kilometers\_Driven, Fuel\_type, Transmission, Owner\_type, Mileage,Engine, power, seats, New\_price, price.

**Source: Github.com**

**Data Pre-Processing:**

S. Manikanta kumar – AM.EN.U4.CSE20268

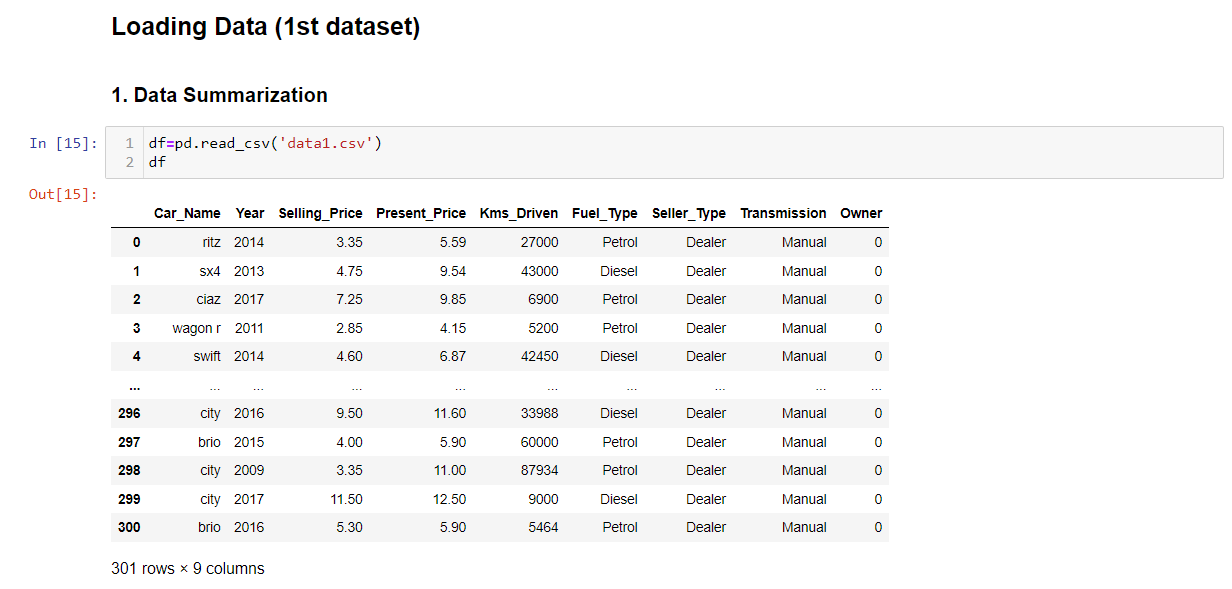
K. Naveen Kumar -AM.EN.U4CSE20238

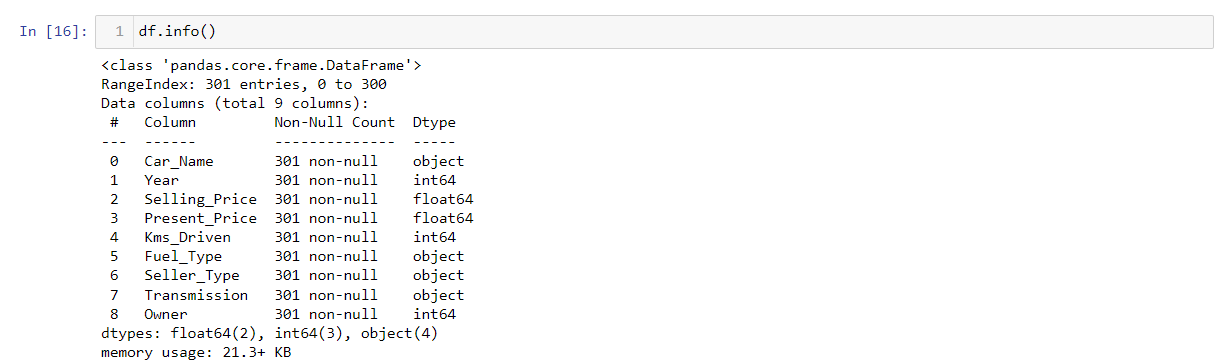
In this Project, we are going to predict the Price of Used Cars using various features like 'Name', 'Location', 'Year', 'Kilometers Driven' etc. The data used in this project was downloaded from Kaggle.

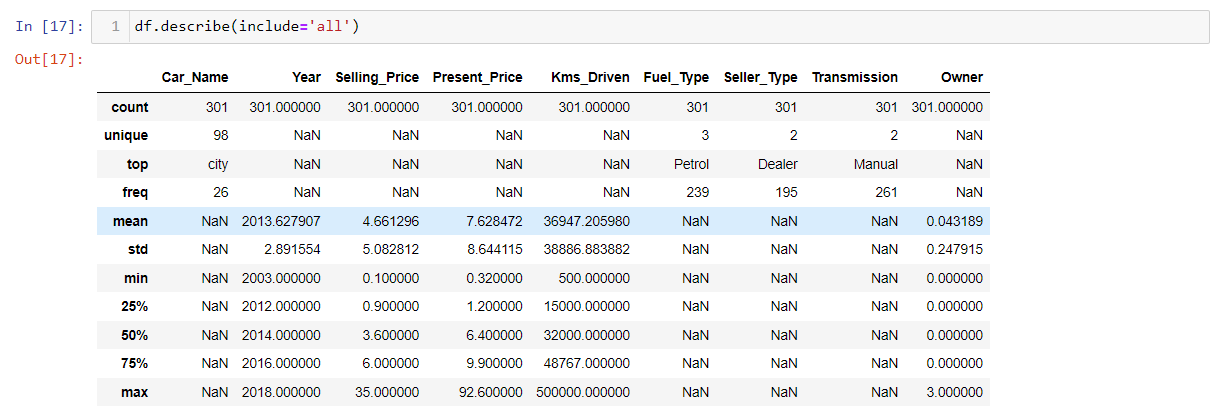
We have taken dataset 1.

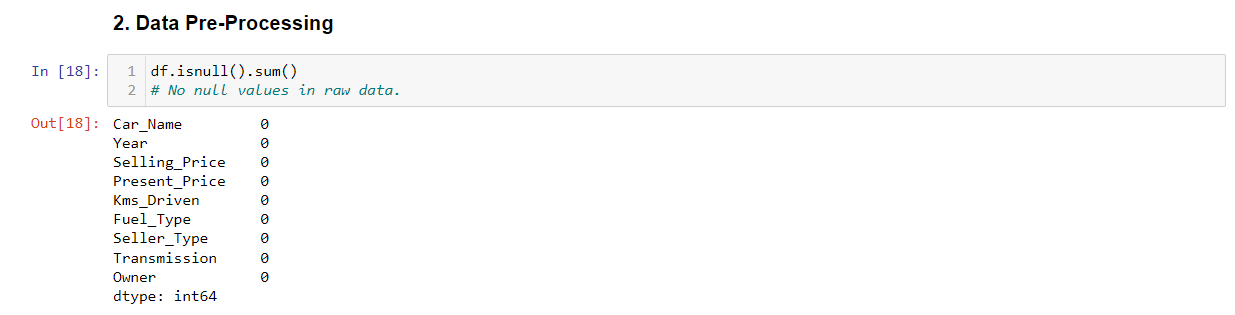
Graphical user interface, application

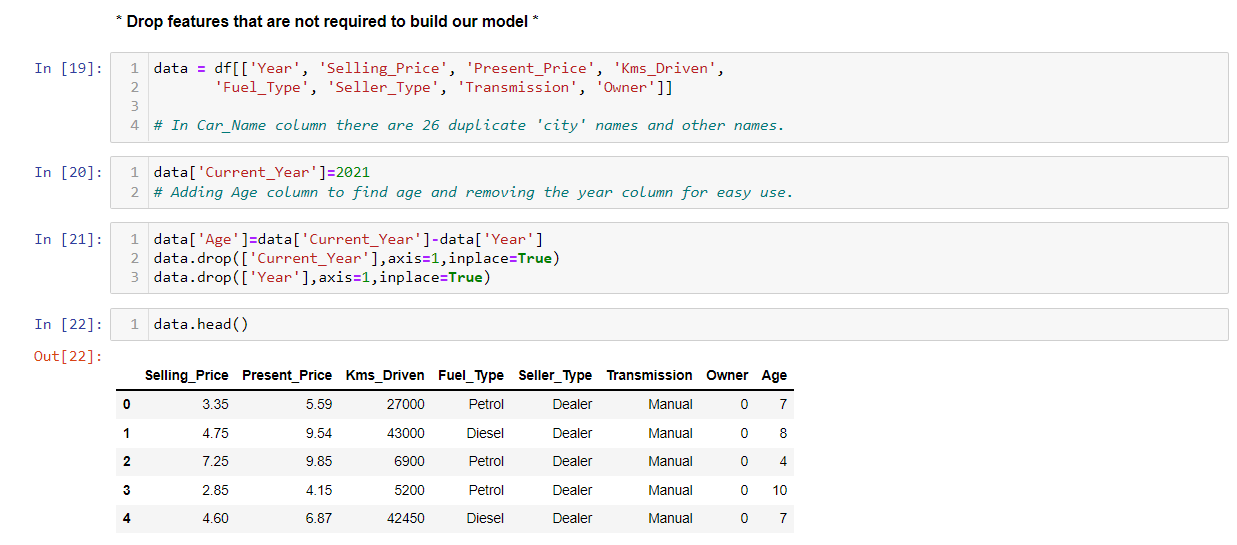
Description automatically generated





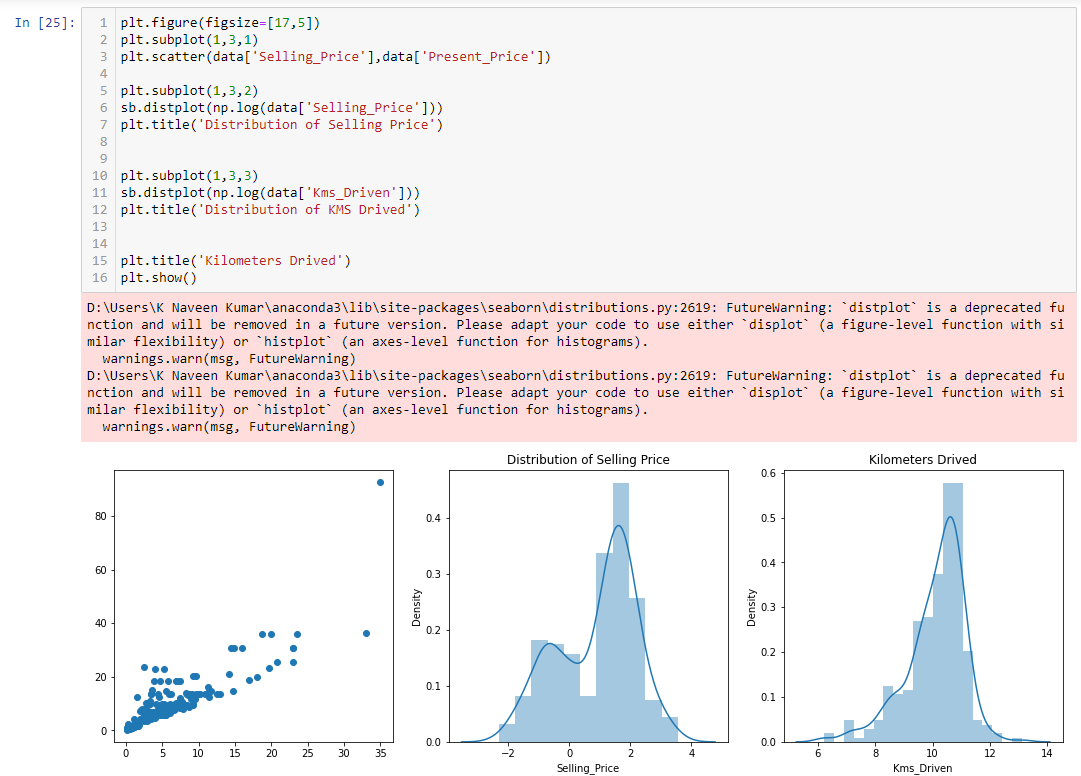


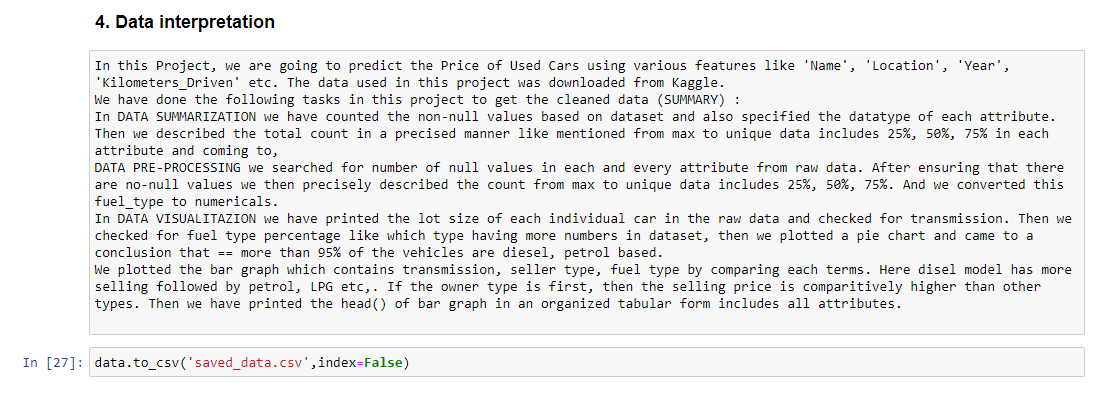










****

In this Project, we are going to predict the Price of Used Cars using various features like 'Name', 'Location', 'Year', 'Kilometers Driven' etc. The data used in this project was downloaded from Kaggle.

We have done the following tasks in this project to get the cleaned data **(SUMMARY)** :

In **DATA SUMMARIZATION** we have counted the non-null values based on dataset and specified the datatype of each attribute. Then we described the total count in a précised manner like mentioned from max to unique data includes 25%, 50%, 75% in each attribute and coming to,

**DATA PRE-PROCESSING**, we searched for number of null values in each attribute from raw data. After ensuring that there are no-null values we then precisely described the count from max to unique data includes 25%, 50%, 75%. And we converted this fuel type to numerical.

In **DATA VISUALITAZION** we have printed the lot size of each individual car in the raw data and checked for transmission. Then we checked for fuel type percentage like which type having more numbers in dataset, then we plotted a pie chart and concluded that == more than 95% of the vehicles are diesel, petrol based.

We plotted the bar graph which contains transmission, seller type, fuel type by comparing each term. Here diesel model has more selling followed by petrol, LPG etc., If the owner type is first, then the selling price is comparatively higher than other types. Then we have printed the head() of bar graph in an organized tabular form includes all attributes.

**ORIGINAL DATA:**



**CLEANED DATA:**

****