DATA-MINING

ASSINGMENT-2

Course Instructor : Sir Zain Mirza

SYED MUHAMMAD
RIMMEL ASGHAR
20B-005-SE   SE-A

1) Which dataset you've selected?

I have selected car price Dataset to perform EDA (Exploratory data Analysis) on it.

2) What analysis you've done in the starter code?

I have imported the required libraries for analysis which are Seaborn , Matplotlib ,Pandas, numpy and Plotly . After importing first I load the dataset into a variable called df. Then I have examined does the dataset contains Null values. After that I looked the summary of the dataset and examined the data type of each column in my DataSet. Then I analyse how many columns with data type object are present in the dataset. This part was all my Preprocessing Part. Now I have started Doing Cleaning on the dataset. First I examined the number of rows and columns. To handle noisy data in the data set I have dropped the rows that have any Null Values. I also observed that some columns have int values in str , I modify it and change the data type to int. After getting done with the cleaning Part its now time to start EDA (Exploratory Data Analysis). The Analysis done by me contains:

1) Analysis of Fuel type used by different cars.
2) Does the car Contains Leather Interior.
3) Analysis of Gearbox in Cars.
4) Analysis of Manufacture of Cars.
5) Analysis on Car categories.
6) Analysis on Yearly Car Production.
7) Analysis of Engine Volume.
8) Analysis of Cylinder.
9) Analysis of Mileage.
10) Analysis of Airbag.
11) Factor Affecting Price.
12) Analysis of Popularity of Manufacture.

3) What information you got?

The information I've got from my Analysis is given further below:

1) Petrol was the most famous fuel type in this Dataset.
2) 72.54% of cars contains Leather Interior.
3) Automatic gear box was the most famous in this Dataset.
4) Most of the cars is Manufactured by Hyundai and Toyota.
5) Most of the cars in this Dataset belongs to Sedan category.
6) Most of the car Manufactured was in the year 2012 about 2110 units.
7) Most of the car consists of 2.0 engine volume.
8) Most of the cars are 4 cylinder.
9) Most of the cars are on a 0 mileage.
10) Most of the cars contains 4 airbags.
11) Factors affecting price are given as follows
      - Manufacture vs price : Lamborghini manufactures car on a high price.
      - Gear box vs Price: Triptronic is the most expensive gearbox.
      - Fuel type vs price : it is examined that diesel is the most expensive fuel type.
12) Hyundai Sonata is the most famous car manufactured by Hyundai.