Amrita School of Computing
Department of Computer Science & Engineering
19CSE304 Foundations of Data Science

LAB SHEET-2

Exercise 1:

1. Read the csv file from the given URL to a dataframe.
   https://archive.ics.uci.edu/ml/machine-learning-databases/autos/imports-85.data.

2. Include the following headers to the above csv file
   ["symboling","normalized-losses","make","fuel-type","aspiration", "num-of-doors","body-style","drive-wheels","engine-location","wheel-base","length","width","height","curb-weight","engine-type","num-of-cylinders", "engine-size","fuel-system","bore","stroke","compression-ratio","horsepower", "peak-rpm","city-mpg","highway-mpg","price"]

3. Display the first five rows of the dataset.

4. Replace the "?" in the above file with NaN.

5. Find the missing values in the dataset.

6. Count the missing values in each column.

7. Identify the column(s) of a given Data Frame which have at least one missing value.

8. Find the Indexes of missing values of column "normalized-losses" in the given DataFrame.

9. Replace the missing values in "normalized-losses" ,"stroke", "bore",with the mean.

11.Replace the missing values in "num_doors" with the maximum frequency value

12. Replace the missing values in "horse_power",peak_rpm with the values in the next row.

13. Drop the rows of "price", if value is missing.

14. List the datatypes of each column.

15. Convert the columns to appropriate datatype.

For Example: Numerical variables should have type 'float' or 'int', and variables with strings such as categories should have type 'object'. In the above csv file'bore' and 'stroke' variables are numerical values that describe the engines, so we should expect them to be of the type 'float' or 'int'; however, they are shown as type 'object'.

16. Normalize the columns "length", "width" and "height" so their value ranges from 0 to 1.

Exercise 2:

https://www.datacamp.com/tutorial/types-of-data-plots-and-how-to-create-them-in-python

Exercise 3:

1. Make use of the mpg dataset and do the various visualization plots which is given in the following link:

   Exploration and analysis - Auto-MPG | Kaggle