Assignment 5
CE 787A: Computational Tools for Transportation Engineering
Total Marks: 35

Instructions: Use Jupyter notebook to work on this assignment using Python language. All cell outputs and visualizations (if any) should be visible and necessary comments should be put to make the code readable. Once the code is ready, convert it to HTML (File > Download as HTML), save the html as pdf, and submit the pdf. The file name should be AssignmentNumber_RollNo_FirstName.pdf (only one pdf for the assignment) Also, submit the *.ipynb file* of the jupyter notebook separately.

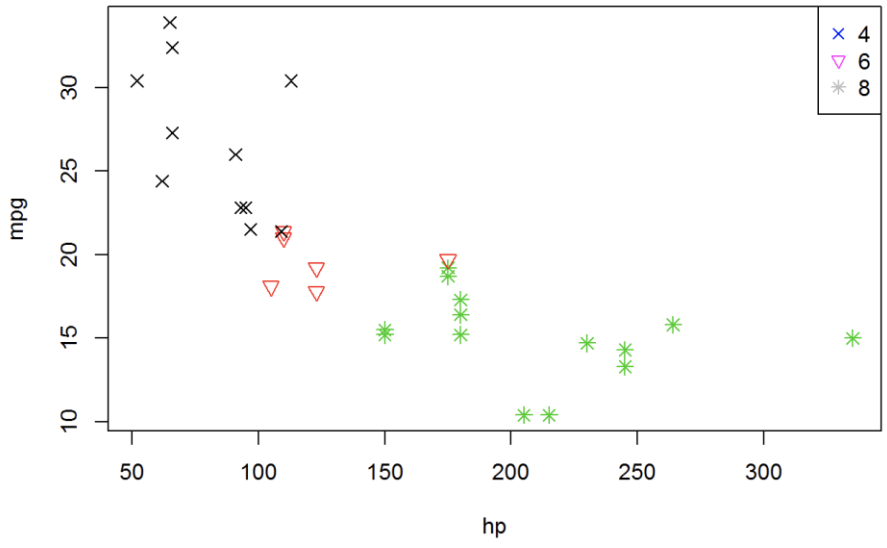
1. The file ***car_info.csv*** contains information regarding different car models. The data is extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and other aspects of automobile design and performance for 32 automobiles (1973–74 models). The data schema and description is as follows:

- mpg: Miles/(US) gallon
- cyl: Number of cylinders
- hp: Gross horsepower

With this required information, perform the following tasks:

a) Read the csv file. Find out the number of classes of cylinders (cyl) present in the dataset. 2

b) Plot a scatter plot of "*mpg*" vs "*hp*". Use different point types for different class of cylinder. Provide axis title, plot title, legend appropriately. A sample plot is shown below (The point types shown here are for representational purpose only. You can choose your own point types). 5+3



c) Now use different sub plots to plot "*mpg*" vs "*hp*" for different classes of cylinders. Each sub plot should be for each unique value of cylinders. Provide all axis titles, plot title, sub plot title appropriately. 6+4

2. The file *model-accuracy.csv* contains *accuracy* of 4 different *models* (Model 1, Model 2, etc.) for different number of *samples* (training points). The schema is provided below:

- *Samples*: number of samples. As number of samples increase, the accuracy in general is expected to increase.
- *Accuracy*: Model accuracy which can vary between 0 to 1. 0 means worst accuracy and 1 means best accuracy.
- *Model*: Model name. There are 4 different models in this dataset (Model 1, Model 2, Model 3, Model 4).

With the above information, perform the following tasks:

a) Read the csv file. Find out the highest and lowest accuracy of each model. You can output this as a dataframe. 1+3

b) Pivot the dataframe to keep the model accuracy of each model as a separate column. Output the pivot dataframe. Sample is shown below. 3

Samples	Model 1	Model 2	Model 3	Model 4
10				
12				
14				

c) Plot the accuracy vs samples as line plot with different line types/color for different model. Provide axis title, plot title, legend appropriately. Comment on which model performs best according to you. 4+2+2

A sample plot is shown below for illustration The line types, color, etc. shown here are for representational purpose only. You can choose your own style. Note, the data shown in the sample plot is different from the data provided in this assignment.