EAST WEST UNIVERSITY

Department of Computer Science & Engineering

Assignment 01
Title: Statistics for Data Science;
Course Code: CSE520

Instruction:

- **This is an individual assignment. You cannot allow anyone to copy the content and style of your assignment. Violation of this rule may result zero mark in the assignment.**
- **Use Python/R/or other software to solve the questions.**
- **Attach the programming codes at the end of the assignment.**

Answer the following questions based on the 'Motor Trend Car Road Tests Data'.
[data set is provided]

Motor Trend Car Road Tests Data Description: The data was extracted from the 1974 *Motor Trend* US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models).

A data frame with 32 observations on 11 (numeric) variables.

[1]	mpg	Miles/(US) gallon	[7]	qsec	1/4 mile time
[2]	cyl	Number of cylinders	[8]	vs	Engine (0 = V-shaped, 1 = straight)
[3]	disp	Displacement (cu.in.)	[9]	am	Transmission (0 = automatic, 1 = manual)
[4]	hp	Gross horsepower	[10]	gear	Number of forward gears
[5]	drat	Rear axle ratio	[11]	carb	Number of carburetors
[6]	wt	Weight (1000 lbs)			

Questions:

i. Identify the qualitative and quantitative variables.
ii. Give the appropriate graphical representation for qualitative variables. Comment the graphs.
iii. Give a histogram for the variable 'mpg (Miles/(US) gallon)' and comment the plot.
iv. Give stem-leaf-plot for all numerical variables and comment the plots.
v. Find the boxplot for the variable 'wt(Weight (1000 lbs))'. Do you identify outliers from this data? If so, how will you interpret those car models?
vi. Which car models have less 'mpg' than the first quarter in terms of less 'mpg' . Also, find car models that have more 'mpg' than the third quarter.