**P2 Exam Data Science**

1. Use the dat set wholesale datacustomers.csv, apply K-means clustering approach to find out 5 clusters. For proper clustering remove outliers. Show the Cluster centers.

2. Use the data mtcars of R and calculate principle component analysis on to that data.

[Hint: Remove categorical data].

3. Use the data USArrests of R, This is a set that contains four variables that represent the number of arrests per 100,000 residents for Assault, Murder, and Rape in each of the fifty US states in 1973. The data set also contains the percentage of the population living in urban areas, UrbanPop:

a. Find 5 cities with highest no. Of murders and rape.

b. Change the name of the columns as “M” ,“A”, “U”, “R”

c. Find 10 cities with least no. Of people living in urban areas.

d. Visualise the data using suitable graphs.

4. Use the data cars from R, create a regression model for predicting distance based on speed. Use 80% of the data to train the model and 20% to test the model. Calculate the accuracy of your model.