Tutorial 6 - Clustering

1)
You are given a dataset containing a person's **age,gender,annual income and spending score.** You are required to cluster the persons according to their annual income and spending score. You are required to proceed in the following manner -

a) Do any data cleaning if required (checking for noise, null values etc.)
b) Obtain the count plot for gender and age (histogram)
c) Obtain the plot of age and spending score for both the genders. You can separate the gender and plot them with different colors in the plot to visualize better. What do you infer from this plot ?
d) Obtain the plot of annual income and spending score for both the genders.You can separate the gender and plot them with different colors in the plot to visualize better. What do you infer from this plot ?
e) Build a K-means clustering model and find out the optimal number of clusters using the Elbow method.
f) Plot WCSS vs number of clusters and locate the optimal number of clusters.
g) Build the model with that value of cluster.
h) Visualize and plot the clusters (Annual income vs spending scores) and write down your observations.

2)
You are given a data set that contains evaluation scores provided by students from a certain university. There are a total of 28 course specific questions and additional 5 attributes. You are required to proceed in the following manner -

a) Perform exploratory data analysis. Obtain the count plot of instr and class (histogram).
b) Obtain the correlation matrix.
c) Perform PCA to reduce the dimension to 2.
d) Build a K-means clustering model and use the elbow method to find the optimal number of clusters.
e) Plot WCSS vs number of clusters and build the model for that number of clusters.
f) Visualize and plot the clusters .