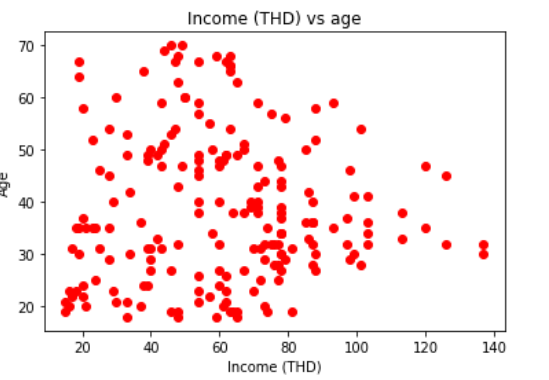
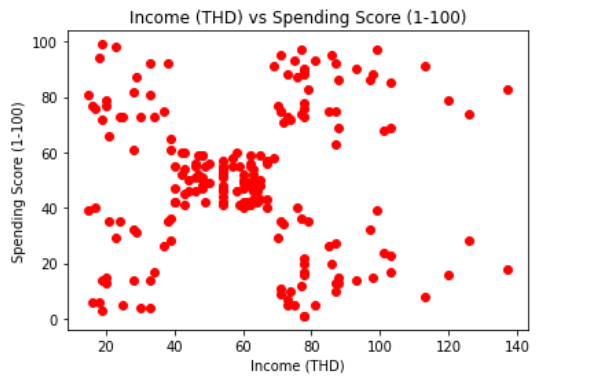
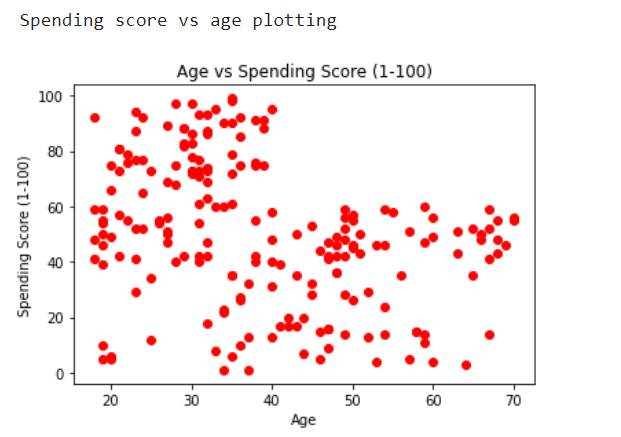
**Report for Assignment 5**

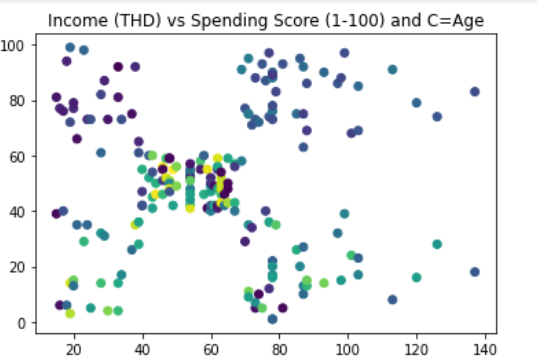
**K Means**

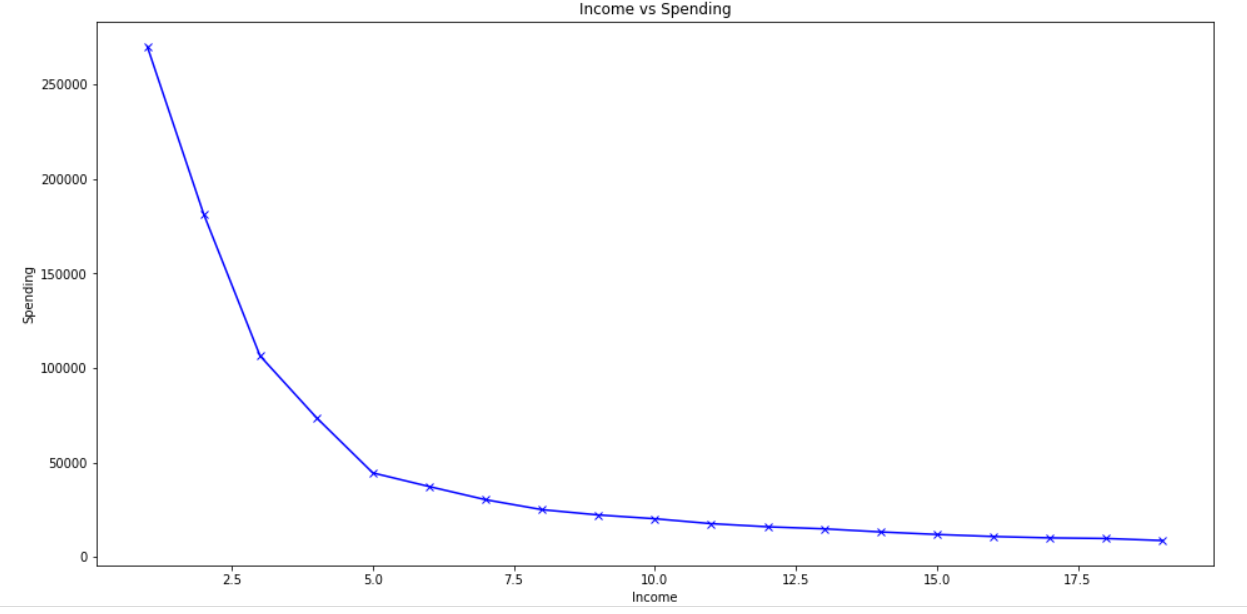
I have taken Mall Customers dataset from Kaggle. Here we had three features Gender, Age, Income and Spending. I plotted these separately.



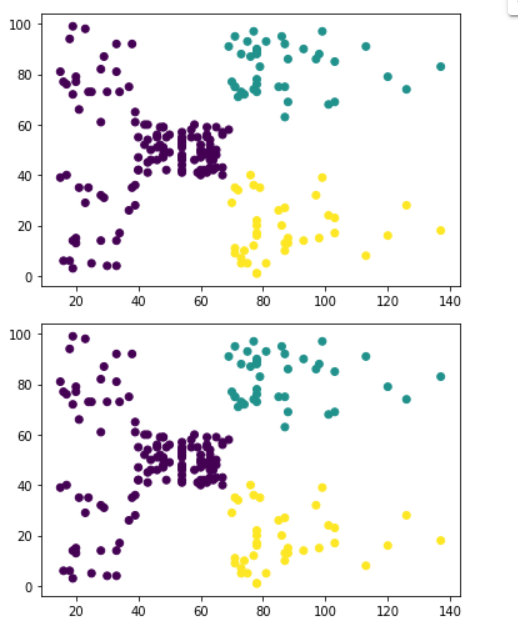


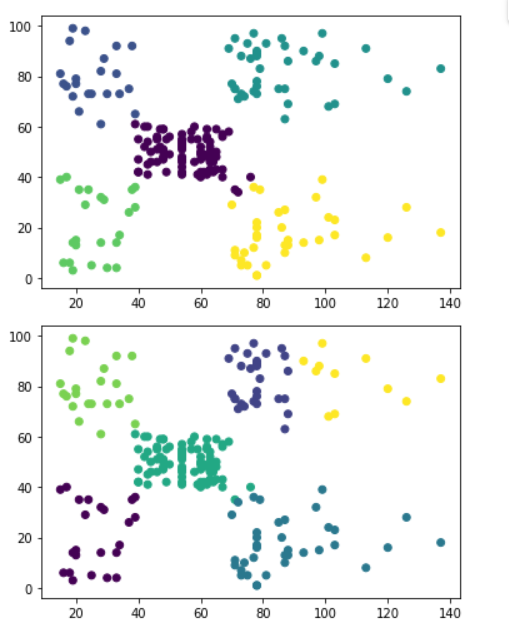






Using K Means we get clusters.





Then we used silhouette\_score to find the best cluster size.

For n\_clusters = 2, silhouette score is 0.2965167311215868

For n\_clusters = 3, silhouette score is 0.4671264942977645

For n\_clusters = 4, silhouette score is 0.49369523511548846

For n\_clusters = 5, silhouette score is 0.5541446796204179

For n\_clusters = 6, silhouette score is 0.5410411871359793

For n\_clusters = 7, silhouette score is 0.5327101345818385

For n\_clusters = 8, silhouette score is 0.45854013950590267

For n\_clusters = 9, silhouette score is 0.4581689704998002

For n\_clusters = 10, silhouette score is 0.4505873675162661

For n\_clusters = 11, silhouette score is 0.44515537320416854

For n\_clusters = 12, silhouette score is 0.4400443397643773

For n\_clusters = 13, silhouette score is 0.42365237726710364

For n\_clusters = 14, silhouette score is 0.4186349372846408

For n\_clusters = 15, silhouette score is 0.4204044838281923

For n\_clusters = 16, silhouette score is 0.4334916895508394

For n\_clusters = 17, silhouette score is 0.43051731261305654

For n\_clusters = 18, silhouette score is 0.41899366690089124

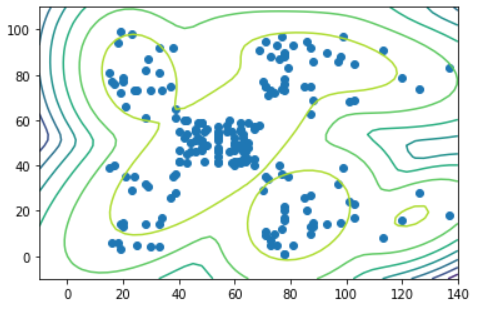
For n\_clusters = 19, silhouette score is 0.4294838009211467

Using silhouette\_score we get silhouette score is maximum for 5 clusters with score 0.5541446796204179.

So, the optimul number of clusters is 5.

**Gaussian Cluster GMM**

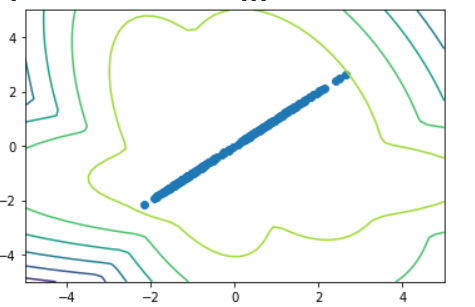
Using inbuilt function from SKlearn we get clusters,



**PCA**

Then we used PCA on the same dataset. We reduced the dimension of the dataset to 2. I sot two feature Principal Component 1 and Principal Component 2. Then we applied K Means and GMM here also.

Using K means I get cluster,



And using GMM on this we get clusters,

