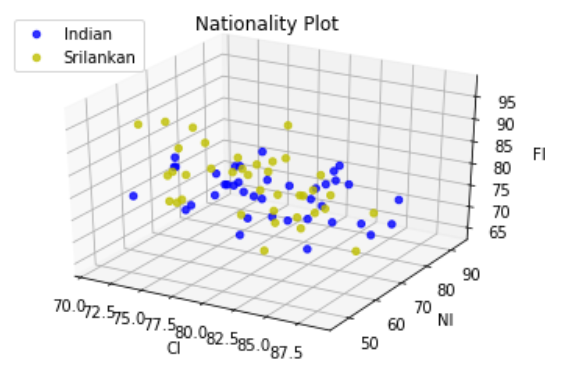
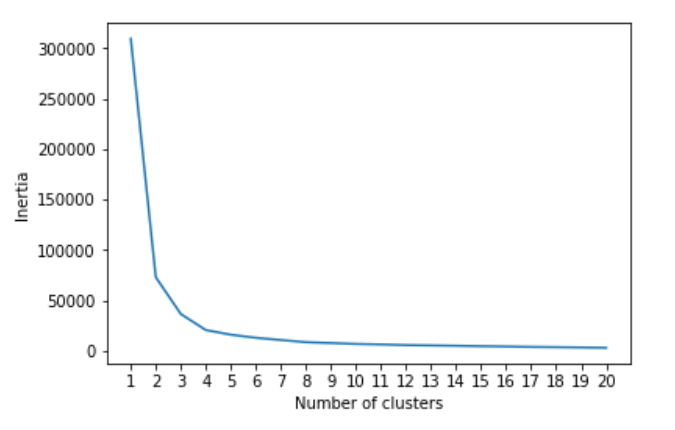
Dataset: Nationality Malaysian, ethinicity all

Number of Samples = 77

Test Dataset Size = 25%

Clustering the data based on Nationality (Indian, Srilankan)



Graph of model inertia vs number of model clusters.

From this graph using elbow method this is evident that classifying Nationality from given dataset is highly possible by clustering as the optimum number of clusters is 2. For cluster size of 2 the inertia is 73080.33896565421

which can be reduced increasing sample size.

Attributes used(7): PIN, Age, Nationality, Sex, HI, FI, NI

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Predicted Class | |
|  |  | Indian | Srilankan |
| Actual Class | Indian | 9 | 3 |
| Srilankan | 3 | 5 |

Accuracy = 70%

Precision for predicting Indian =75%

Precision for predicting Srilankan =62.5%

Recall for predicting Indian = 75%

Recall for predicting Female = 62.5%