# MODULE 20

# **Airline Data**

**Package used in R**

Readxl : Used to load Excel Data

Readr : Used to Extract the final output

**Package used in python**

Pandas : Used for data manipulation

Matplotlib : Used to create plot for dendrogram

Sklearn AgglomerativeClustering : Used to run hierarchical Cluster

Scipy

cluster.hierarchy : Used to create cluster on dendrogram

Linkage : Used as a linkage function for the dendrogram

**Loading the data**

Loading the EastWestAirlines dataset from Data Sheet in R and Python .

**EDA**

No NaN Data Found in the dataset.

Normalizing the dataset with Min Max method.

## **Dendrogram**

Created Dendrogram with Euclidean Distance with method complete.

Created 3 clusters to on the Dendrogram.

**Cluster 1**: Not a active customer on using credit card and not a active travelers. Its possible that they can a tourist.

**Cluster 2**: Most of the customer have enrolled long before the other cluster customer. They are less active compared with cluster3.

**Cluster 3**: customer are more active on using the credit card and has more rewards and miles points and have high balance and frequent plane and non plane travelers. Its possible that this cluster can be Business Class people.

1. **Crime Data**

**Package used in R**

readxl- Used to load Excel Data

Readr- Used to Extract the final output

**Package used in python**

Pandas - Used for data manipulation

Matplotlib - Used to create plot for dendrogram

Sklearn AgglomerativeClustering - Used to run hierarchical Cluster

Scipy

cluster.hierarchy - Used to create cluster on dendrogram

Linkage - Used as a linkage function for the dendrogram

**Loading the data**

Loading the crime\_data dataset in R and Python .

**EDA**

No NaN Data Found in the dataset.

Normalizing the dataset with Min Max method.

## **Dendrogram**

Created Dendrogram with **Manhattan** Distance with Method Complete.

Created 3 Clusters to on the Dendrogram

**Cluster 1:** Has more 2x assault compared to cluster 2 people are getting assaulted on streets and Rape is very high compared to other clusters. This cities has very high crime rate.

**Cluster 2**: Urban Pop is very high on cluster 2 compared to other clusters Murder and Assault is less compared to compared to cluster 1.

**Cluster 3**: Crime rate in Cluster 3 is less compared to other 2 cluster. This cities are safe compared to other cities.