# MODULE 21

1. **Crime Data**

**Package used in python**

Pandas : Used for data manipulation

Matplotlib : Used to create plot for dendrogram

Sklearn

KMeans : Used to run K Mean algorithm

**Loading the data**

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

**EDA**

No NaN Data Found in the dataset.

Removed X Variable and created new dataset for further analysis

Normalizing the dataset with Min Max method in Python

Normalizing the dataset with Scaling Method in R

**K Mean**

Created 2 Clusters. Comparing the 2 cluster.

**Cluster 1:** Cluster 1 has more less crime compared to cluster 2

**Cluster 2:** Cluster 2 has more crime rate compared to cluster 1. All the variable has X2 the crime rate except UrbanPop

1. **Airline Data**

**Package used in python**

Pandas : Used for data manipulation

Matplotlib : Used to create plot for dendrogram

Sklearn

KMeans : Used to run K Mean algorithm

**Loading the data**

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

**EDA**

No NaN Data Found in the dataset.

Removed X Variable and created new dataset for further analysis

Normalizing the dataset with Min Max method in Python

Normalizing the dataset with Scaling Method in R

**K Mean**

Created 3 Clusters. Comparing the 3 cluster.

**Cluster 1**: 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.

**Cluster 2**: Not a active customer on using credit card and not a active travelers , has less Bonus points and Miles , Its possible that they can a tourist.

**Cluster 3**: This Cluster has high Qualifying top status and high number of flight miles and flight transactions and highest award points compared with other cluster.

1. **Insurance Data**

**Package used in python**

Pandas : Used for data manipulation

Matplotlib : Used to create plot for dendrogram

Sklearn

KMeans : Used to run K Mean algorithm

**Loading the data**

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

**EDA**

No NaN Data Found in the dataset.

Removed X Variable and created new dataset for further analysis

Normalizing the dataset with Min Max method in Python

Normalizing the dataset with Scaling Method in R

**K Mean**

Created 2 Clusters. Comparing the 2 cluster.

**Cluster 1**: This cluster has senior citizen and highest claims made compared to other cluster

**Cluster 2**: This Cluster has age group of 40 and policy renewal is around 100 days to pay for next premium.