Name: Avinash

Batch ID: DSWEMCOB270221

Topic : K-Means clustering

PFA R and Python codes for clustering

* Imported dataset to analyze.
* Renamed columns to maintain uniformity
* Inspected duplicate rows and missing values in the dataset.
* Inspected variance in each column.
* If 90% of same value repeating, it does not suit for analysis.
* Remove unwanted or not useful columns for analysis.
* Check for outliers and apply required outlier treatment.
* Convert categorical columns into numerical.
* Check for normal distribution and apply transformation if needed.
* Apply normalization/standardization to the dataset.
* Apply algorithm to the dataset and export as a report.

Inference:

* Number of clusters is pre-determined in k-means clustering.
* From elbow curve, we can conclude the suitable number of clusters.
* k-means clustering best suitable for large dataset.