Objective: Performing clustering on dataset having categorical and numerical data.

Constraints:

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:

