

Clustering Workshop

1. In this workshop, we shall perform Customer Segmentation using K-Means clustering.
2. Please use the dataset ***cust_seg.csv*** for the questions that follow.
3. Plot 4 histograms to show the distributions for the following attributes:
 - a. Age
 - b. Income
 - c. Years Employed
 - d. DebtIncomeRatio
4. Plot the number of Default and Non-Default cases.
5. There are missing entries under the *Defaulted* column. Use a KNN Classifier to determine if those missing entries should be classified as Default or Non-Default cases.
6. Insert the results of your Default/Non-Default classification results into your dataset such that there are no more missing entries under the *Defaulted* column.
7. Cluster your updated dataset into 4 clusters using K-Means.
8. Create a 3D plot of your clustering results using these 3 attributes:
 - a. Age
 - b. Income
 - c. DebtIncomeRatio
9. What can we infer from the 3D plot?