**Data Information: -**

The goal of the dataset is to help the auditors by building a classification model that can predict the fraudulent firm on the basis of the present and historical risk factors. The information about the sectors and the counts of firms are listed respectively as Irrigation (114), Public Health (77), Buildings and Roads (82), Forest (70), Corporate (47), Animal Husbandry (95), Communication (1), Electrical (4), Land (5), Science and Technology (3), Tourism (1), Fisheries (41), Industries (37), Agriculture (200).

This research work is a case study of an external government audit company which is also the external auditor of government firms in India. During audit planning, auditors examine the business of different government offices but the target is to visit the offices with a very-high likelihood and significance of misstatements. This is calculated by assessing the risk relevant to the financial reporting goals.

**Problem Statement:**

Use the clustering techniques to identify the similarities of the data and cluster them based on the similarities and then perform classification algorithms to build a model to predict the cluster labels.

**Assignment Expectations/Steps -**

1. Load the dataset
2. Perform Exploratory Data Analysis on the data
3. Build a K- Means Clustering algorithm and share your findings
4. Build an Agglomerative Clustering algorithm and share your findings
5. Apply PCA to the original Dataset
6. Build classification algorithms and choose the best model
7. Share your business insights.