Minor Project

Exploratory Data Analysis – 70 marks

## Domain:

Banking

## Problem Statement:

Happy Bank provides various credit cards to customers. The manager of Happy Bank is disturbed by more and more customers leaving their credit card services. The team did a customer survey to check customer attrition. Various customer attributes like Customer\_Age, Credit\_Limit, Dependent\_Count. The team would really appreciate it if one could predict for them who is gonna get churned so they can proactively go to the customer to provide them better services and turn customers' decisions in the opposite direction.

# Dataset Description:

Attrition\_Flag: Internal event (customer activity) variable - if the account is closed then 1 else 0(Target)

Customer\_Age: Demographic variable - Customer's Age in Years Dependent\_Count:Demographic variable - Number of dependents Credit\_Limit:Credit Limit on the Credit Card

And so on….

.

# Steps and Tasks:

* 1. Import Necessary Libraries.(2 points)
  2. Display a sample of five rows of the data frame.(1 point)
  3. Check the shape of the data (number of rows and columns).(1 point)
  4. Check the percentage of missing values in each column of the data frame.(1 point)
  5. Check if there are any duplicate rows.(1 point)
  6. Check the distribution of the Customer\_Age column. Check the basic statistics like mean, median, and standard deviation of the age column.(6 points)
  7. Plot 2 box plots and 2 pie chart of the parameter of your on choice and write your intuition about it (8 marks)
  8. Plot a Box-plot of Total\_Revolving\_Bal and Card\_Category by characterizing with Attrition\_Flag. Write your intuitions about it.(10 points)
  9. Plot a percentage segment bar graph between Education\_Level and Attrition\_Flag of the customers.(10 point)
  10. Plot a percentage segment bar graph between Income\_Category and Attrition\_Flag of the customers.(10 point)
  11. Drop CLIENTNUM column.Make a sub data frame which consists of all the numerical columns(i.e.int64,float64) along with the Attrition\_Flag column. Plot a clear heatmap to view the correlation using seaborn. (10 points)
  12. Plot a boxplot for the Credit\_Limit column and check if it contains any outlier or not.(5 point)
  13. Map the Attrition\_Flag values to 0 and 1(i.e. Existing Customer=0 and Attrited Customer=1. Standardize the columns.(5 point)