Problem: customers leaving a bank's credit card services.

Background

A manager at the bank is disturbed with more and more customers leaving their credit card services. They would really appreciate if one could predict who is going to get churned or attrited  so they can proactively go to these customers to provide them better services and turn customers' decisions in the positive direction

The manager has prepared a sample data set with customer information.

Now, this dataset consists of 10,115 customers mentioning their age, salary, marital\_status, credit card limit, credit card category, etc. There are twenty one features, including a flag that indicates whether the customer has stopped doing business with the bank or not.

**Project Description**

 Your data mining team has been asked by the manager to provide a mechanism to predict which customers are at risk of leaving the bank.

   In other words you will create an automated prediction process of customer churn by training, validating and deploying a data mining model using a machine learning tool, such as RapidMiner.

 The data set containing the sample of records will be provided to your team for EDA in a separate folder with the name "BankChurners set for EDA".

**1 - Exploratory Data Analysis phase.**

   The team will analyze the data and fully understand it based on the EDA.

1.     The team will use any software that can help in the EDA phase, preferably Tableau and Rapidminer

2.     Each member will choose a data likely predictor variable in the sample file and provide a detailed analysis of it in a separate slides. (Reference the EDA examples used in class for Churning telephone company

3.     The team will explore and recommend at least three possible machine learning models that could be used in to provide a solution based on the results of the EDA.

4.     Each team will create a consolidated PowerPoint deck to present the EDA findings.

5.     Each Powerpoint slide will have the name of the member  that will be describing the work in the particular slide.

6.     The presentation deck must have a slide for the introduction of the team members at the beginning of the deck and a summary slide at the end of the deck with the EDA findings and model recommendations.

Column meaning:

1.     Clientnum - Client number. Unique identifier for the customer holding the account

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

3.     Customer\_Age - Demographic variable - Customer's Age in Years

4.     Gender - Demographic variable - M=Male, F=Female

5.     Dependent\_count - Demographic variable - Number of dependents

6.     Education\_Level - Demographic variable - Educational Qualification of the account holder (example: high school, college graduate, etc.)

7.     Marital\_Status - Demographic variable - Married, Single, Divorced, Unknown

8.     Income\_Category - Demographic variable - Annual Income Category of the account holder (< $40K, $40K - 60K, $60K - $80K, $80K-$120K, > $120K, Unknown)

9.     Card\_Category - Product Variable - Type of Card (Blue, Silver, Gold, Platinum)

10.  Months\_on\_book - Period of relationship with bank

11.  Total\_Relationship\_Count - Total no. of products held by the customer

12.  Months\_Inactive\_12\_mon - No. of months inactive in the last 12 months

13.  Contacts\_Count\_12\_mon - No. of Contacts in the last 12 months

14.  Credit\_Limit - Credit Limit on the Credit Card

15.  Total\_Revolving\_Bal - Total Revolving Balance on the Credit Card

16.  Avg\_Open\_To\_Buy - Open to Buy Credit Line (Average of last 12 months)

17.  Total\_Amt\_Chng\_Q4\_Q1 - Change in Transaction Amount (Q4 over Q1)

18.  Total\_Trans\_Amt - Total Transaction Amount (Last 12 months)

19.  Total\_Trans\_Ct - Total Transaction Count (Last 12 months)

20. Total Ct Chng Q4 Q1 – Rate of counts from Q1 to Q4

21.  Avg\_Utilization\_Ratio - Average Card Utilization Ratio