**Problem:** Why are some customers leaving the Great One Bank Inc. credit card services?

**Background**

The CEO at the Great One Bank is concerned with the news that some customers are cancelling their credit card services accounts. He is looking forward to the possibility that one could predict which of the current customers is most likely to cancel - to churn or be attrited - so the bank can proactively go to these t-risk customers to provide them better services and turn their attitudes into a more loyal mode.

The CEO has asked the CIO to engage the Data Science team to tackle this issue. She has given a sample data set with historical customer information to the Data Science director to work with and he has passed it on to the Data Science team.

Now, this dataset consists of 5,998 customer records out of an entire population of 3,100,111 credit card customers. There are twenty one features, including a flag variable that indicates whether the customer has stopped doing business with the bank (attrited), or not.

**Project Description**

The Data Science team has been given the task to provide a mechanism to predict which customers are at risk of leaving the bank.    In other words, as part of the team you will collaborate and 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 done.

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

2.     Each team member will choose **two** different predictor variables in the sample file and provide a detailed analysis of it in separate slides indicating the findings as to why they are or are not good predictor variables. Note: Reference the EDA examples used in class for the Churn 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. **IMPORTANT: No machine learning training, testing and validation of any models is part of this process. Those tasks are to be done in phase 2 of the project.**

4.     Each team member will create a PowerPoint deck to present in class the individual’s EDA work with the name of the team member that will be presenting the work in detail for the individual slides.

6.     Each individual team member presentation deck must have an introduction of the team member names at the beginning of the deck, and a summary slide at the end of the deck with the individual EDA findings and the three (3) model recommended by the team.

**Attributes meaning**:

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

2.     Attrition\_Flag - Internal event (customer activity) variable – Existing or Attrited

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