**Written Report**

Limit: 5-10 pages

Font: Times New Roman, 12 point

Your written submission should contain the following sections:

1. Executive Summary page (*1 page limit*) – this page contains the overall analysis; could be torn off from the rest of the submission and present understandable results
2. Introduction – description of the problem; importance
3. Methodology – description of the data set, data management and data mining methodologies employed; why were the methods selected?
4. Results – description and reporting of the results obtained; use tables and graphs where appropriate to assist in understanding the results; limited output/code
5. Discussion/Conclusion – summarize and discuss the results; suggestions for future research
6. Appendix – this section does not count against the page limit and should include any relevant output, numerical and graphical summaries

A complete written report includes:

1. Written submission (Word or pdf file)
2. SAS EM project (zipped file) or R script

**Background**

For financial services firms, acquiring new customers is more costly than retaining them. Therefore, retaining existing customers is one of the biggest challenges. Customer churn is everywhere. In unsecured lending, customers cancel credit cards or personal loans, or there’s silent attrition in the form of a slow decrease in customer card spend. In secured lending, mortgages face churn in the form of a loan transfer to other lenders, partial or full payment of loans and loan closure. Customers also can close their bank accounts, resulting in the loss of potentially cheap sources of funds, or they cancel their life or general insurance policies, resulting in the loss of potential future premiums.

Moreover, customers often have multiple relationships across different services (banking, loans, insurance, etc.), and a churn in one relationship can trigger churn in other relationships.

This data set contains details of a bank's customers located in several European countries. The goal of the analysis is to build the best model that identify people at risk of closing a bank account. The data consist of customers’ demographics and some bank account information. The target “Exited” is a binary variable which reflects the fact whether the customer left the bank (closed his account) or continued to be a customer.

Source: *https://www.linkedin.com/pulse/predict-prevent-churn-few-examples-from-financial-yogesh-kumar-sharma/*

**Data**

*Churn\_data.csv*

The data file contains 9,000 rows of data and 12 columns.

*Churn\_score\_data.csv*

The score data file contains 1,000 rows of data and 11 columns.

*Churn\_score\_true\_data.csv*

The file contains 1,000 records of true “Exited” for the “CustomerId” in *Churn\_score\_data.csv*.

**Tasks**

We are interested in the classification analysis related to the variable “Exited”. Can you identify any relationships between the variables (e.g., Which country has more customers that tend to close their accounts? What about tenure? Age? etc.)? A scoring data file (test data) is provided in order to test your model prediction.

**Data Definitions**

**CustomerId**: identification number

**CreditScore:** interval; a credit score of the customer

**Geography:** categorical;locations of the bank in three European countries: France, Germany, and Spain

**Gender:** Female, Male

**Age**: interval

**Tenure:** interval;number of years for which the customer has been with the bank

**Balance:** interval; bank balance of the customer

**NumOfProducts:** interval; number of bank products the customer is utilizing

**HasCrCard:** binary; 1 = if a customer has a credit card with the bank; 0 = otherwise

**IsActiveMember:** binary;1 = if a customer is an active member with the bank; 0 = otherwise

**EstimatedSalary:** interval; estimated salary of the customer in dollars

**Exited:** binary; 1 = if the customer closed the bank account; 0 = if the customer is retained