**BDAT 1009 Enterprise Analytics**

**Final Assignment**

**Total Mark – 25 Points**

**Due Date – April 21st 2023 12:00 pm**

**Dataset**

For your Final Assignment you will use the following dataset in Kaggle:

[https://www.kaggle.com/datasets/blastchar/telco-customer-churn](https://can01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.kaggle.com%2Fdatasets%2Fblastchar%2Ftelco-customer-churn&data=05%7C01%7Csabbir.ahmed%40georgiancollege.ca%7C0ed1e4f9d6864e2ff8ea08dac7133bfe%7Cda9a94b6468149bcbd7cbab9eac0ad3c%7C0%7C0%7C638041182698249456%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=D8b1j76IhPajtvKfri87PDc6cVgu%2F5tvyoBnODSh%2F4o%3D&reserved=0)

Each row represents a customer, each column contains customer’s attributes described on the column Metadata.

The data set includes information about:

* Customers who left within the last month – the column is called Churn
* Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
* Customer account information – how long they’ve been a customer, contract, payment method, paperless billing, monthly charges, and total charges
* Demographic info about customers – gender, age range, and if they have partners and dependents

**Your Assignment**

Using this data, the company wants you to analyze overall Churn last month. Goal is to understand main driver of Churn and come up with strategy for retention. As an analyst your job is to use tools like Excel, PowerBI or Python to perform following tasks:

**1) Churn Analytics (15%)**

For each question below you can use Excel, PBI or Python.

* Provide brief description of method you used to bring data to your chosen tool. Briefly provide your observation. Highlight any cleaning job you needed to perform and what was your approach. (1 Mark)
* You are asked to analyze various aspect of customer and its impact on Churn. Chart or Graph for Churn Rate (%) by (Note: Looking for % and NOT count of Churn): (5 Marks)
  + Tenure
  + Product (Internet service – DSL vs Fiber Optic vs Other)
  + Contract
  + Monthly Charges ( Use range. E.g. High is Monthly charges > 100 etc.)

Explore the data and provide a summary of Impact of Tenure, Product, Contract and Monthly charges on Churn referring to the charts you created. Make at least 4 observations and refer to the charts above to demonstrate the observations. (3 Marks)

* Analyze the data and determine if customers with High monthly charges and lower tenure churn more than customer with lower monthly charges and high tenure. Demonstrate your analysis with a chart. (2 Marks)
* Given the company's primary objective is to reduce overall Churn - Come up with 1 example of ANY SMART Objectives. Explain why it is a SMART Objective. **Refer to Lecture # 3.** (2 Marks)
* Provide 1 chart/graph to track or monitor your SMART Objective identified above. (2 Marks)

**2) Balance Scorecard and Visualization (5%) -**

* You are asked to design a balance scorecard to track churn. Provide a “mock up” of a Balance Scorecard to track Churn. (2 Marks)

***Hint: Refer to Lesson#3 examples of Balance Scorecard:***

A picture containing graphical user interface

Description automatically generated

* Identify what dataset is required or needs to be sourced to complete your proposed Balance Scorecard.

***Hint: Refer to Lesson#3.***

* Build a visualization dashboard which will provide an overview of Telco churn. Minimum of 4 and Maximum 6 Charts. You can re-use charts from Part 1. You can use up to 2 dummy chart (E.g. Financial, Training etc.) (3 Marks)

**3) Churn Prediction (5%)**

Use WEKA, RapidMiner or SAS to run experiment on the given data to see if we can predict Churn using the parameters given. (2 Marks)

Experiment using at least 2 Classification techniques. Discuss Accuracy and provide recommendation. (2 Marks)

Provie a summary report (1 Mark) –

1) Observation of the data and attributes (E.g. You chose to ignore a column/attribute due to missing values).

2) How you prep the data and import in your choice of prediction tool.

**OR**

**Association Mining (5%)**

In order to increase retention effort, the company wants to understand which of the following value added services are closely associated to each other among customers who do not churn and have internet service.

* OnlineSecurity
* OnlineBackup
* DeviceProtection
* StreamingTV
* StreamingMovies

Filter your dataset (Churn = “No”, InternetService <> “No”) and prepare a transaction type dataset with following columns (2 marks):

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **customerID** |  |  | **OnlineSecurity** | **OnlineBackup** | **DeviceProtection** | **StreamingTV** | **StreamingMovies** |

Use Aporiri association in a tool of your choice (Weka, SAS, RapidMiner) and see if there are any association. Play with support and confidence to find a suitable association. Briefly explain your experiment. (1 marks)

List at least 2 association you find with confidence and support. (2 marks)

***Hint: Refer to Lecture # 5 and Lab#4.***

**Submit -**

1) Final Report in Word or PPT or PDF - Ensure all screen shots are provided. Max 10 Slides/Pages. (Please make it simple to read for business users).

2) Charts For part 1 & 2 in Python, Excel or PBI File.

3) For Part 3 – RapidMiner, SAS or WEKA Screenshot.