Name: Shrey Prajapati

Student Number: 8779344

Course: INFO8496-24S: Analysis for Financial Technology

Case Study 1

Date: 05/29/2024

1. **Load data from all tabs into Power Query, preprocess the data and save as connection only.**

A screenshot of a computer

Description automatically generated

1. **Create a data model by adding this data to the Power Pivot data model**.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Preprocessing of the data:**

1. Implemented Promoted Headers and changed type in the customer code table.

A screenshot of a computer

Description automatically generated

1. Implemented promoted headers, changed type, removed blank rows, removed columns with a null value, removed top rows that are null, and split the characters in the product column.

A screenshot of a computer

Description automatically generated

1. Applied to Remove blank rows and null value columns in the rate table.

A screenshot of a computer

Description automatically generated

**3 a) Using the Statistics function to determine the highest amount owed.**

A screenshot of a computer

Description automatically generated

**3b). Apply new discounts to new customers and calculate the "New YTD Balance for Q1" (refer to the "read me" sheet for a series of calculations)**

Performed the Subtraction and multiplication in the rate table to calculate the Discounted amount and new customer rate.

A screenshot of a computer

Description automatically generated

Merged and expanded the rate table.

A screenshot of a computer

Description automatically generated

Added conditional column to calculate the Final Discounted rate.

A screenshot of a computer

Description automatically generated

Calculated New Fees

A screenshot of a computer

Description automatically generated

Calculated New total.

A screenshot of a computer

Description automatically generated

Calculated Balance Due

A screenshot of a computer

Description automatically generated

Calculated Last Quarter

A screenshot of a computer

Description automatically generated

Calculate New YTD

A screenshot of a computer

Description automatically generated

**3c) Calculate the average and median of the "YTD balances" owed by customer status**.

Found Average and Median of New YTD

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**3d) Split the Account ID column by the number of characters/positions; determine the count of Regular, High Risk, Government Affiliate, and Premium customers, respectively, and display YTD balances owed in each category.**

Split the Account column.

A screenshot of a computer

Description automatically generated

Merged it with customer code

A screenshot of a computer

Description automatically generated

Expanded the customer code by amount.

A screenshot of a computer

Description automatically generated

Applied count and sum for ytd balances by applying count rows and sum using group by command.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**3e) Determine the total amount lost by the company as discounts during this period.**

Calculated difference of fees using a custom column

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**3f) Identify customers who purchased additional licenses and specify the associated products.**

Applied group by to the products and account.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

1. **Describe the four steps in data preprocessing and explain how/if they apply to the practice data set you are working with.**

**Ans.** The four steps used in data preprocessing are data cleaning, Date integration, data reduction, and data transformation.

1. **Data cleaning**

For the cleaning of the data, we have removed the unnecessary rows and columns in the above step which have null values and are blank or empty.

1. **Data Integration**

By removing the unnecessary data from the dataset and providing accurate data in the dataset which is relevant to each column and made it proper available for the analysis of the data as per requirements.

1. **Data Reduction**

The data was not required and possessing the space in the dataset for that the data redundancy to reduce the acquired available space in which there was no requirement of the data.

1. **Data transformation**

The data transformation is applied to the dataset by calculating the sum, and maximum of the data, splitting the columns, and applying group by to transform the data for further processing of the data.