Analyzing Credit Card Customer Churn Behaviour

**Problem Statement:**

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 for them who is considering leaving the bank so they can proactively go to the customer to provide them better services and reverse the customers' decision in their favour.

**Data Source and Description:**

[](file:///C:\Users\malar\Downloads\portfolio\BankChurnerscc.csv)

The dataset consists of records of 10,127 bank customers (rows) and 20 columns describing various features viz.- 'Attrition\_Flag', 'Customer\_Age', 'Gender', 'Dependent\_count', 'Education\_Level', 'Marital\_Status', 'Income\_Category', 'Card\_Category', 'Months\_on\_book', 'Total\_Relationship\_Count', 'Months\_Inactive\_12\_mon', 'Contacts\_Count\_12\_mon', 'Credit\_Limit', 'Total\_Revolving\_Bal', 'Avg\_Open\_To\_Buy', 'Total\_Amt\_Chng\_Q4\_Q1', 'Total\_Trans\_Amt', 'Total\_Trans\_Ct', 'Total\_Ct\_Chng\_Q4\_Q1', 'Avg\_Utilization\_Ratio'. We have both categorical as well as numerical features. We select the best of these features and model customer attrition behaviour using these features. Such a prediction system can act as an early warning system for the bank and incentivize them to do the needful to retain customers.

**Columns:**

* **Clientnum** Num Client number. Unique identifier for the customer holding the account
* **Attrition\_Flag** char Internal event (customer activity) variable
* **Customer\_Age** Num Demographic variable - Customer's Age in Years
* **Gender** Char Demographic variable - M=Male, F=Female
* **Dependent\_count** Num Demographic variable - Number of people dependents
* **Education\_Level** Char Demographic variable - Educational Qualification of the account holder (example: high school, college graduate, etc.)
* **Marital\_Status** Char Demographic variable - Married, Single, Unknown
* **Income\_Category** Char Demographic variable - Annual Income Category of the account holder (< 40K, 40K - 60K, 60K - 80K, 80K-120K, > 120K, Unknown)
* **Card\_Category** Char Product Variable - Type of Card (Blue, Silver, Gold, Platinum)
* **Months\_on\_book** Num Months on book (Time of Relationship)
* **Total\_Relationship\_Count** Num Total no. of products held by the customer
* **Months\_Inactive\_12\_mon** Num No. of months inactive in the last 12 months
* **Contacts\_Count\_12\_mon** Num No. of Contacts in the last 12 months
* **Credit\_Limit** Num Credit Limit on the Credit Card
* **Total\_Revolving\_Bal** Num Total Revolving Balance on the Credit Card
* **Avg\_Open\_To\_Buy Num** Open to Buy Credit Line (Average of last 12 months)
* **Total\_Amt\_Chng\_Q4\_Q1** Num Change in Transaction Amount (Q4 over Q1)
* **Total\_Trans\_Amt Num** Total Transaction Amount (Last 12 months)
* **Total\_Trans\_Ct Num** Total Transaction Count (Last 12 months)
* **Total\_Ct\_Chng\_Q4\_Q1** Num Change in Transaction Count (Q4 over Q1)
* **Avg\_Utilization\_Ratio** Num Average Card Utilization Ratio

### Data Exploration & Cleaning

* Before beginning any kind of data exploration or analysis, I changed all the "Unknown" values to be recognized as null values.
* I also removed 2 columns of irrelevant data.
* I split the data into 2 tables, attrited customers and existing customers using the "Attrition\_Flag" variable. The attrited customer table and existing customer table had 1,627 and 8,500 observations, respectively.
* I created histograms for the 11 categorical variables and 8 numerical variables in the 2 new tables and compared the peaks and trends.

### Analysis

The data visualizations showed me that the categorical data did not vary between the attrited and existing customers considerably. However,  some histograms of the numerical variables followed different trends and peaked at different points between the 2 tables. These variables were considered the dependent variables of becoming an attrited customer. Those variables and their peaks are as follows:

|  |  |
| --- | --- |
| **Dependent Variables** | **Peaks** |
| Total\_Revolving\_bal | 0 |
| Total\_Amt\_Chng\_Q4\_Q1 | 0.701 |
| Total\_Trans\_Amt | 2,329 |
| Total\_Trans\_Ct | 43 |
| Total\_Ct\_Chng\_Q4\_Q1 | 0.531 |

**Tools Used:** PowerBI for Visualization,Microsoft Excel for Data Cleaning and Preprocessing

**Technologies used:** SQL Server 2019 , python

INPUT Dataset:

A screenshot of a computer

Description automatically generated

Bank\_Churners SQL Server

A computer screen with a white screen

Description automatically generated

Python Script to connect with SQL Server:

A screenshot of a computer

Description automatically generated

PowerBI Visualization (InActive Customers)  [LiveDemo](file:///C:\Users\malar\OneDrive\Documents\pfp-CC_churn.pbix)

A screenshot of a computer

Description automatically generated

PowerBI Visualization (Existing Customers)

A screenshot of a computer

Description automatically generated

I have verified the Results in Excel.