

Customer Churn Analysis

The story Behind The Data

A bank is concerned that more and more customers are leaving its credit card services. They would really appreciate if someone could analyze it for them, in order to understand the main reasons for leaving the services, and to come up with recommendations for how the bank can mitigate that. Eventually, the bank would like to proactively implement these recommendations in order to keep their customers happy.

A full ERD can be found [here](#)

Data Description

In this task, few datasets are provided:

- BankChurners.csv** - this file contains basic information about each client (10 columns). The columns are:
 - CLIENTNUM** - Client number. Unique identifier for the customer holding the account;
 - Attrition Flag** - Internal event (customer activity) variable - if the client had churned (attrited) or not (existing).
 - Dependent Count** - Demographic variable - Number of dependents
 - Card_Category** - Product Variable - Type of Card (Blue, Silver, Gold, Platinum)
 - Months_on_book** - Period of relationship with bank
 - Months_Inactive_12_mon** - No. of months inactive in the last 12 months
 - Contacts_Count_12_mon** - No. of Contacts in the last 12 months
 - Credit_Limit** - Credit Limit on the Credit Card
 - Avg_Open_To_Buy** - Open to Buy Credit Line (Average of last 12 months)
 - Avg_Utilization_Ratio** - Average Card Utilization Ratio
- basic_client_info.csv** - this file contains some basic client info per each client

(6 columns) - - **CLIENTNUM** - Client number. Unique identifier for the customer holding the account - **Customer Age** - Demographic variable - Customer's Age in Years - **Gender** - Demographic variable - M=Male, F=Female - **Education_Level** - Demographic variable - Educational Qualification of the account holder (example: high school, college graduate, etc. - **Marital_Status** - Demographic variable - Married, Single, Divorced, Unknown - **Income_Category`** - Demographic variable - Annual Income Category of the account holder (< 40K,40K - 60K, 60K–80K, 80K–120K, > \$120K, Unknown)

- enriched_churn_data.csv** - this file contains some enriched data about each client (7 columns) -
 - CLIENTNUM** - Client number. Unique identifier for the customer holding the account
 - Total_Relationship_Count** - Total no. of products held by the customer
 - Total_Revolving_Bal** - Total Revolving Balance on the Credit Card
 - Total_Amt_Chng_Q4_Q1** - Change in Transaction Amount (Q4 over Q1)
 - Total_Trans_Amt** - Total Transaction Amount (Last 12 months)
 - Total_Trans_Ct** - Total Transaction Count (Last 12 months)
 - Total_Ct_Chng_Q4_Q1** - Change in Transaction Count (Q4 over Q1)

```
In [44... # import necessary libraries
import csv
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sns
import xgboost as xgb
import plotly.express as px
import plotly.io as pio
import plotly.graph_objs as go
pio.renderers.default = 'notebook'
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report
from sklearn.model_selection import cross_val_score
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score, confusion_matrix, precision_score, recall_score, f1_score
```