Data Science Career Track - Bank Churn Rate Project Proposal

Problem Statement:
Provide churn rate mitigation strategy to a bank who uploaded the dataset to reduce churn rate by 30% and increase overall bank revenue by 20% as well as provide key insights for the upcoming 2024 financial quarter.

Context:
This bank has a large amount of information about their customers. However, their goal is to help analysts predict the churn rate of these customers in order to reduce it.

Criteria for Success:
Criteria for success would be determining the exact or the combination of factors causing customer churn for the bank developing a plan to either reduce or remove these factors prior to next 2024 financial quarter.

Scope of Solution Space:
Scope of this solution space or the churn mitigation plan would be applied to the incoming financial quarter in order to assess effectiveness before permanence.

Constraints within Solution Space:
Foreseeable constraints include time to be able to implement this reduction plan as would be best to roll out the mitigation strategy prior to next financial quarter. Another would be having access to the location of customers as location might also be a factor affecting customer churn but is sensitive information.

Stakeholders:
CEO, CFO, Database Admin, Head of Data Analysis

DataSources:
Dataset url:
https://www.kaggle.com/datasets/celocruz/bank-database?select=supply_chain_train.csv
Data is in .csv format

From the Website:
The database includes demographic information such as age, gender, marital status and income category. It also contains information on card type, number of months in portfolio and inactive periods. In addition, it has key data on the spending behavior of customers approaching their cancellation decision. Among the latter information are the total renewable balance, the credit limit, the average open-to-buy rate and analyzable metrics such as the total amount of change from the fourth quarter to the first quarter or the average utilization rate.

Data Science Career Track - Bank Churn Rate Project Proposal

Outline of Problem Solving Steps:

1. Problem Identification:
    a. Develop Project Proposal
    b. Identify Scope and Criteria for Success

2. Data Wrangling:
    a. Organize .csv in pandas Dataframe
    b. Determine “na” values
    c. Validate features of dataset and determine constraints (ex. No negative income, gender does not equal blue, etc.)
    d. Validate Data Definitions of Features

3. EDA:
    a. Develop correlation heatmap to assess feature relationships
    b. Develop individual scatter plots to further assess correlations
    c. Determine top features leading to customer churn and determine target feature

4. Modeling:
    a. Split “train” to train test split
    b. Scale features if necessary and impute ‘na’ values
    c. Assess model effectiveness based on mean value error metric
    d. Determine effective predictive model and test on ‘test.csv’ provided
    e. Test several business case scenarios to determine potential mitigation strategy

5. Documentation:
    a. Create report documenting key insights and model effectiveness
    b. Create slide deck for executive and head of Data Analysis proving model effectiveness and use.
    c. Code for the project and steps will be stored in Github Repo