Data Science Internship at Data Glacier

Project: Cross-selling Recommendations for Banking Products

Name: Jesumbo Joseph Oludipe

Email: oludipejesumboj@gmail.com

Country: United Kingdom

College: University of Greenwich, UK

Specialization: Data Analyst

Batch code: LISUM19

Submitted to: Data Glacier

Table of Contents

1. Problem Description

XYZ Credit Union, a prominent financial institution in Latin America, has been successful in selling various banking products including credit cards, deposit accounts, retirement accounts, and safe deposit boxes. However, despite the initial success, the existing customers do not show any inclination to purchase more than one product. This indicates that the bank is not effectively utilizing the opportunity of cross-selling to its customers. To address this issue, XYZ Credit Union has sought the assistance of ABC Analytics to analyze the data and provide actionable insights to increase cross-selling.

2. Business Understanding

Based on the given problem statement, while XYZ Credit Union has been successful in selling a variety of banking products in Latin America, they have not been able to capitalize on cross-selling opportunities. Despite having a large customer base, the bank's existing customers do not purchase more than one product, which is indicative of a lack of effective cross-selling strategies. This project aims to identify the factors that affect cross-selling, understand customer behavior, and propose strategies that can improve the bank's overall performance in selling multiple products to its existing customers.

3. Project Lifecycle

The project lifecycle and expected completion date are given below.

S/N	Task	Completion Date
1	Problem Statement, Business Understanding, Data Intake Report	April 19, 2023
2	Data Understanding	April 26, 2023
3	Data Cleaning and Transformation	May 2, 2023
4	Exploratory Data Analysis (EDA)	May 9, 2023
5	EDA Presentation	May 16, 2023
6	Dashboard	May 23, 2023
7	Final Submission (Report + Presentation)	May 30, 2023

4. Data Understanding

The dataset that has been provided includes comprehensive customer information such as age, gender, and country of residence, alongside the various bank products that they currently own, such as credit cards, deposit accounts, retirement accounts, safe deposit boxes, and more. In total, the dataset comprises 48 features (columns) and 13647308 observations (rows), offering a large volume of data for analysis and providing valuable insights into the bank's customer base and product preferences.

Table 1: Data Intake Report

Total number of observations	13647308
Total number of files	1
Total number of features	48
Base format of the file	CSV
Size of the data	2.13 GB

5. Data Cleaning

As a data analyst, I recognize the importance of data cleaning to ensure that the data is accurate and reliable for any downstream analysis or use. In this case, the dataset required several cleaning steps. The following data-cleaning steps were taken:

- The column names were changed to make them more descriptive and understandable.
- The gender values were converted from "H" and "V" to "M" and "F" to standardize the data.
- The values "S" and "N" were also converted to "Yes" and "No" to make them more explicit.
- The rows filled entirely with missing values (NA) were removed.
- Finally, any Spanish words present in the dataset were translated to English to ensure consistency in the language used throughout the dataset.

These cleaning steps were necessary to improve the quality of the data and enable accurate analysis and interpretation of the results.