Group Name: Data Detective

Name: Syed Ahmad Sohail

Email: syedahmad.sohail@ucalgary.ca

Country: Canada

College/Company: Data Glacier

Specialization: Data Analyst

Problem description:

XYZ Credit Union is a banking institution located in Latin America, and it's doing well in selling different types of banking products such as credit cards, deposit accounts, retirement accounts, and safe deposit boxes to its customers. However, despite their success in selling these products, the bank is facing a challenge of not being able to sell more than one product to their existing customers. This means that the bank is not performing well in cross-selling, which is an important strategy to increase revenue and customer loyalty. Cross-selling is the practice of encouraging customers to purchase additional products or services beyond their initial purchase. In this case, the bank is missing out on the opportunity to sell more of their products to their existing customers, which could lead to increased revenue and stronger customer relationships. As a result, the bank has approached ABC analytics to help them solve this problem. The goal is to identify ways to increase cross-selling and provide actionable insights that the bank can implement to improve its cross-selling strategy. By doing so, the bank can improve its overall performance and maintain its competitive edge in the market.

Data Understanding:

The dataset contains information related to banks and customers, and all the data is in Spanish. It is noticed that there are a few missing values in the dataset. Before performing any data analysis, it is necessary to address the missing values. The data has a few columns that contain NA values. We need to identify which columns have missing values and fill them appropriately. Additionally, it is crucial to understand the data and its relationship with the business context. This will help in making informed decisions and performing effective data analysis. Overall, to ensure a thorough understanding of the data, we need to explore the dataset in detail, clean the data, and process it accordingly.

Type of Data:

The type of data being dealt here is Structured data which refers to data that is organized in a fixed format with a known schema.

Problems in Data:

There are several problems with the data provided. Firstly, the 'Spouse Index' column is not defined properly, it is unclear what the values in this column represent. Additionally, all the column names are in Spanish which might make it difficult for English speakers to work with the data. Furthermore, the 'New Customer Index' column only has the value '0' which may be incorrect. Lastly, the gender of the customers is defined by the letters 'H' and 'V' which are not commonly used or recognized in English or other languages. These issues can make it difficult to

analyze and draw meaningful insights from the data. It can be seen from the data primarily that a lot of columns look completely empty like the mortgage plan and pensions. The data also only comprised to be from spain and almost all the

Cleaning Data:

- First the column names in Spanish to be converted to English to make it easier to analyze using dictionary.
- NA or Nan values to be either Averaged or rows to be removed with missing fields using dropna().
- Dropping unwanted columns using drop function.
- Merge and unique for the required required columns.