Team members: Jason Sanchez
Group name: Alpha
Email: jasonfs2@illinois.edu
College: Harold Washington College
Specialization: Data Analyst

Problem Description:

XYZ credit union in Latin America is performing very well in selling the Banking products (eg: Credit card, deposit account, retirement account, safe deposit box etc) but their existing customer is not not buying more than 1 product which means bank is not performing good in cross selling (Bank is not able to sell their other offerings to existing customer). XYZ Credit Union decided to approach ABC analytics to solve their problem.

Type of data analysis:

The development of an effective dashboard is instrumental in providing a concise yet comprehensive view of customer demographics and key performance indicators (KPIs) for making informed business decisions. To achieve this, the initial step is to thoroughly analyze and clean the data. The analysis phase involves an examination of the dataset to understand its structure and contents. This process includes identifying the type of customers, categorizing them into segments such as VIPs, students, and others, and calculating important metrics like average age. Additionally, it is vital to scrutinize the data for anomalies, outliers, and inconsistencies that might affect the accuracy of the insights derived from the dashboard. Cleaning the data is equally crucial. This step involves rectifying any data quality issues, such as missing values or incorrect formats, to ensure the dataset's integrity. Cleaning also includes handling outliers and ensuring that the data is in a format that can be readily utilized for dashboard creation. Once the data has been analyzed and cleaned, the next phase involves designing and developing the dashboard. This dashboard should present a clear and concise overview of customer types, segment-wise distribution, average age, and other KPIs that are pertinent to the business. It should be user-friendly and capable of providing real-time insights, facilitating swift and well-informed decision-making for the organization. In essence, the process begins with data analysis and cleansing to establish a robust foundation upon which an insightful and actionable dashboard can be built. The resulting tool will empower the business with valuable customer insights, ultimately leading to more effective strategic decisions.

Problem with the Data:

When examining the dataset, it becomes evident that several substantial issues warrant immediate attention. Chief among these concerns is the presence of non-numeric data within fields originally designated for numerical values. Rectifying this fundamental data integrity issue is a pivotal first step in our analytical journey, as it sets the stage for a rigorous and insightful examination of the dataset.

The data anomalies, primarily involving strings within numeric fields, necessitate a meticulous data cleansing process. This process involves identifying and rectifying any inconsistencies, inaccuracies, or outliers within the dataset. To achieve this, data transformation techniques, such as data type conversions, filtering, and validation, will be employed. The goal is to ensure that all data points conform to the predefined data types and formats, thereby creating a foundation of reliable and coherent data. Once the data cleansing process is complete, the dataset will be subjected to a series of exploratory data analyses. These analyses will encompass a comprehensive range of statistical, graphical, and computational methods designed to unveil hidden patterns, relationships, and trends within the data. Furthermore, these insights will provide a foundation for subsequent advanced analytics and modeling efforts. In addition to numerical data, textual data within the dataset will also require attention. Text mining and natural language processing (NLP) techniques may be employed to extract valuable information from text fields, offering a holistic view of the dataset's content. This multidimensional analysis, incorporating both numeric and textual elements, will contribute to a more comprehensive understanding of the data and facilitate more nuanced decision-making. In conclusion, addressing the data quality issues, particularly those related to the presence of strings in numeric fields, represents a critical initial phase in our data analysis endeavor. By meticulously cleansing and transforming the data, we will ensure its reliability and integrity, laying the groundwork for in-depth analyses and data-driven insights. This approach not only enhances the quality of our findings but also underscores our commitment to rigorous and expert-level data analysis.

What approaches are you trying to apply on your data set to overcome problems like NA value, outlier etc and why?

Addressing data quality issues within the dataset involves a multifaceted approach aimed at overcoming challenges such as missing values (NA) and outliers. These strategies are instrumental in ensuring the dataset's integrity, reliability, and suitability for subsequent analysis. The first step in this endeavor is data cleansing and transformation. This critical process entails identifying and rectifying anomalies within the dataset, including missing values and outliers. By addressing NA values, we prevent potential distortions in statistical analyses and modeling. Additionally, this step eliminates inconsistencies and inaccuracies, ensuring the dataset is trustworthy and reliable.

To mitigate the presence of non-numeric data within fields originally designated for numerical values, data type conversions are employed. This involves converting string-based data into the appropriate numeric format, promoting data coherence and adherence to predefined data types. Outlier detection and handling represent another pivotal facet of data preparation. Robust techniques are implemented to identify and manage outliers, which can significantly impact analytical results. By addressing outliers, we prevent them from unduly influencing our findings and ensure more accurate insights.

Exploratory Data Analysis (EDA) techniques play a crucial role in understanding the dataset's structure and behavior. Through a comprehensive range of statistical, graphical, and computational methods, EDA uncovers hidden patterns, relationships, and trends. This insight informs decision-making regarding how to effectively address NA values and outliers. Furthermore, text mining and Natural Language Processing (NLP) techniques are employed to extract valuable information from textual data within the dataset. This holistic approach offers a more comprehensive understanding of the dataset's content, enabling nuanced decision-making. In conclusion, these approaches collectively aim to enhance data quality, making it suitable for rigorous and insightful analysis. By addressing data quality issues through data cleansing, transformation, outlier handling, EDA, and text mining/NLP, we ensure the dataset's reliability and underscore our commitment to expert-level data analysis.

Github repo: https://github.com/Kuzma12/VC/tree/Project_week_2