Cross selling Recommendation Report

Name: Shreya Ramachandra

Email: ramacha4@uwindsor.ca

Country: Canada

Company: Data Glacier

Specialization: Data Analytics

Problem Description

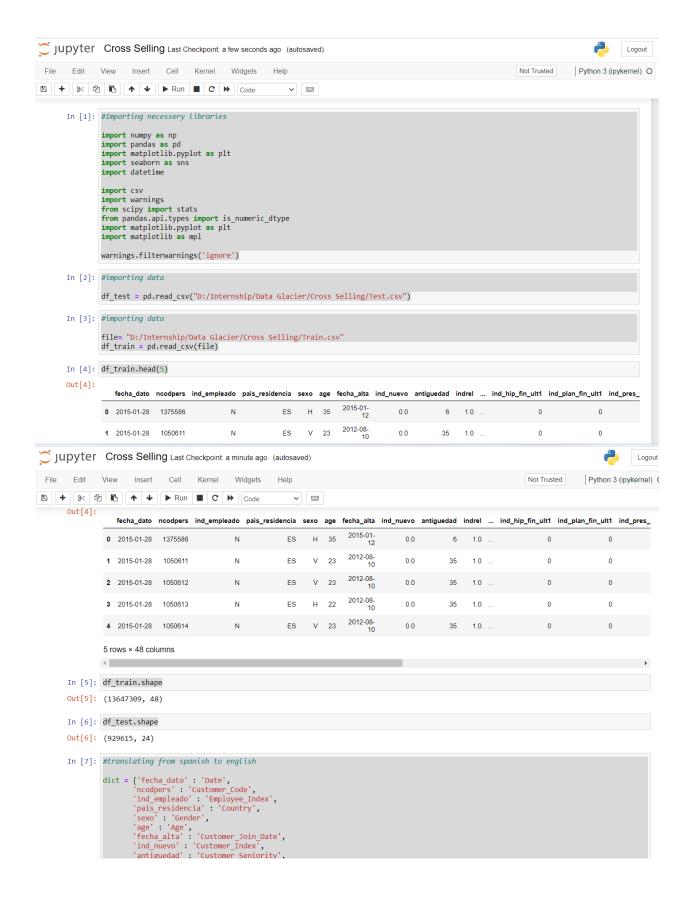
Despite having a strong track record of selling various banking products, such as credit cards, deposit accounts, and retirement accounts, the XYZ credit union in Latin America is struggling to persuade its current customers to purchase additional products or services. This lack of cross-selling is impacting the credit union's overall performance in terms of revenue, profits, and customer retention.

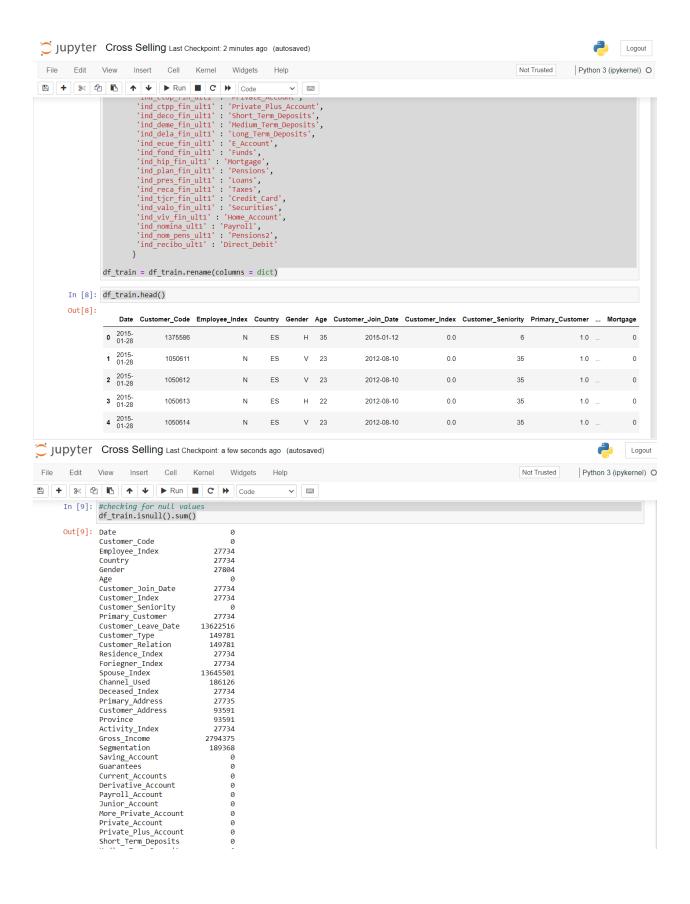
Data Understanding

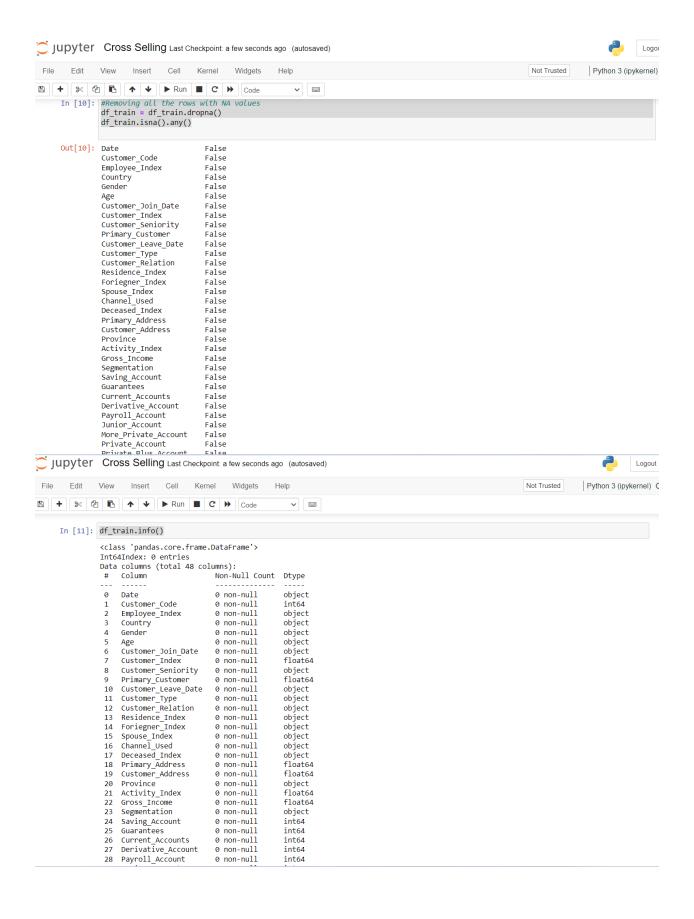
- 1. **Used Pandas and other libraries to load the data into a Pandas DataFrame**: This step involves loading the raw data into a structure called a DataFrame, which is a 2-dimensional labeled data structure with columns of potentially different types. Pandas is a popular library for data analysis and manipulation, and it provides convenient methods for loading data from various sources into a data frame.
- Checked for and handled duplicates: Duplicate records can occur in data sets due to various reasons such as human error, system errors, or data import issues. To handle duplicates in the data, a unique identifier such as customer code is used to identify and remove duplicate records.
- 3. **Ensured data types are correct and consistent**: It is important to ensure that data types are correct and consistent in every column. This can be done by checking the data types of each column and converting them to the correct type.
- 4. **Translated data from Spanish to English**: The translation of data from Spanish to English can improve the understanding of the data.
- 5. **Updated the customer leave date to the correct format**: Dates are commonly stored in different formats in data sets, so it is important to convert them to a consistent format to ensure that the data can be easily analyzed and processed.
- 6. **Corrected the data type of gender**: The gender column may contain various text representations for males and females, so it is important to standardize these values to two consistent categories (e.g. male, female).
- 7. **Check for missing values**: Missing values in data can impact the accuracy of the analysis, so it is important to check for missing values and decide on a strategy for handling them. One common strategy is to drop the missing values.
- 8. Save the cleaned and transformed data for future use: After cleaning and transforming the data, it is a good practice to save the processed data for future use to avoid repeating the process each time the data is needed.

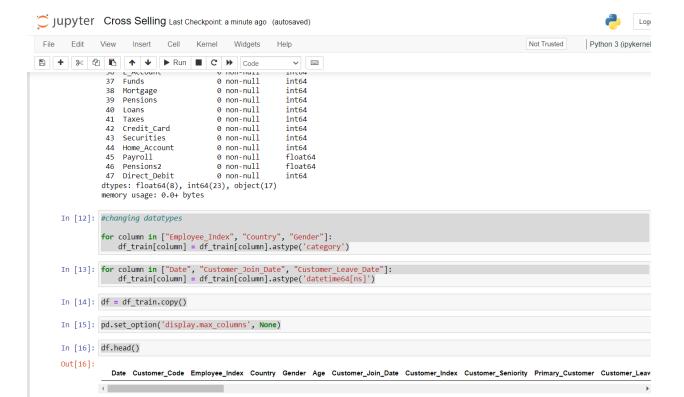
Github:

https://github.com/ShreyaRamachandra/Portfolio Project Cross selling Recommendation-Bank-









In [17]: #dropping duplicates

df = df.drop_duplicates(subset='Customer_Code', keep="last")