Data Glacier – Virtual Internship

Final Project

Name: Isha Panjwani

Email: ishapanjwani5@gmail.com

Country: Canada

Company: Data Glacier

Specialization: Data Analytics

Deliverable: Week 9

Group: Individual

Problem Statement: In this project, our client is a Latin American credit union company XYZ. They are having issues in cross-selling banking products such as credit cards, savings accounts, retirement accounts, and safe deposit boxes. It can take a significant amount of research and business knowledge to increase cross-selling. To succeed in the cross-selling area of the business, Data Analyst at ABC analytics is searching for the best technique to be recommended.

Business statement: The goal of ABC analytics company is to perform Exploratory data analysis on the data provided by the client and gain some meaningful insights. As a data analyst intern, my job was to perform EDA on the credit union's dataset and create visualizations to analyse the data and to provide recommendations to the company to increase effective cross-selling of banking products.

Github Repo link: https://github.com/isha1912/Cross-Selling-Recommendation-EDA

Data Cleaning and Transformation:

(i)   Change of column names: Column names are changed to English words for better understanding of the data.

(ii)   Change of datatypes: Datatypes of columns like – Date, Customer Join Date, Customer Leave Date has been changed from object to Datetime[ns]. Datatypes of columns Employee Index, Country and Gender have been changed from object to category.

(iii)   For reducing redundancy, only the latest records of each customers are kept.

(iv)   Values in some columns have been replaced with full names instead of just initials.

(v)   Some column vales have been changed from numbers to what they represent.

(vi)   Null values in customer leave date represent that the customer is still associated with XYZ Credit Union. The null values are replaced by the last date of updates recorded.

(vii)   Country initials have been replaced by country names.

(viii) Null values in Gross Income column have been replaced by the column mean.

(ix)   Null values in Gender column have been filled with random values from male and female gender options.