Data Glacier – Virtual Internship

Name: Safawat Al Naser
Email: snaser@alliant.edu
Country: USA
Company: Data Glacier
Specialization: Data Analytics
Deliverable: Week 8
Group: One_group

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 analyze the data and to provide recommendations to the company to increase effective cross-selling of banking products.

Data Understanding:

(i) Attributes of data:

- File types': .csv
- Size:257 mb
- more than 13M observations
- nearly 937k customers of the credit union.
- Data include all the attributes which is needed for EDA analysis.
- Some data types are wrong which need to be fixed.
- Need to convert the data as their type.
-

Working with missing values:

1. Using Pandas Jupiter notebook to work with data.
2. Checked the percentage of missing values.
3. What is the use of this missing values need to be understand?
4. If missing values is less, then 25% we have to try to find a way to fill up.

5. For Outliers we simply removed it because it will affect the analysis. For that we need to use boxplot or check the value counts.

some columns have more than 99% of null values which is not needed for the analysis. so, I removed them. Gross income missing values are filled by the average values of the column.

For checking the skewness of the data need to work more on the data. As skewness mostly related to integer column.

For the easiness of analysis some works have done like:

1. Rename the columns.
2. Changed the data types as required.
3. Changed column values so that in can be easily understandable.