## **CREDIT CARD DOCUMENTATION**

#### 1. **Introduction**

* **Objective:** To analyze and visualize credit card data to gain insights into factors that influence credit behavior and status.
* **Tools Used:** Python, Power BI

#### 2. **Data Overview**

 **Dataset:** The dataset includes various credit card usage and demographic information.

 **Key Attributes:**

* Credit status, utilization of unsecured lines, age, gender, region, monthly income, housing status (rented/owned), occupation, and education.
* It also includes details about credit behavior such as the number of times payments were past due, debt ratio, number of open credit lines and loans, number of real estate loans, and the number of dependents.

 **Objective:** The data was cleaned and preprocessed to handle missing values and outliers, and then visualized to uncover patterns and insights related to credit behavior.

#### 3. **Data Cleaning and Preprocessing**

* **Removing Duplicate Columns:** Identified and removed duplicate columns to ensure data consistency.
* **Handling Missing Values:**
  + Used Python to find null values in the dataset.

df.isnull().sum()

* + Filled null values using the mean for numerical columns:

df['MonthlyIncome']= df['MonthlyIncome'].fillna(df['MonthlyIncome'].mean())

df['NumberOfDependents']= df['NumberOfDependents'].fillna(df['NumberOfDependents'].mean())

* **Outlier Treatment:**
  + Identified outliers using the Interquartile Range (IQR) method:

Q1 = df.quantile(0.25)

Q3 = df.quantile(0.75)

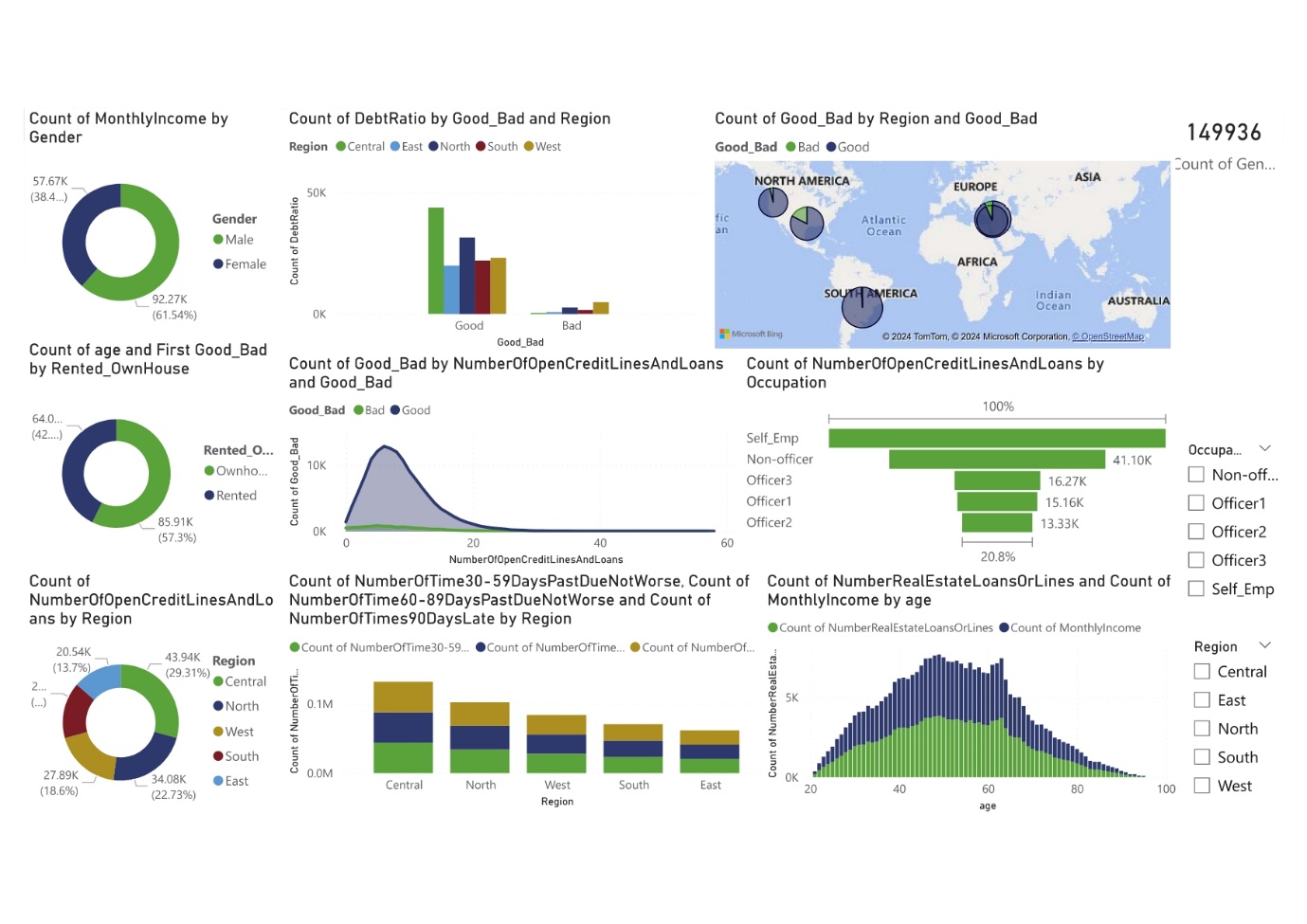
IQR = Q3 - Q1

outliers = ((df < (Q1 - 1.5 \* IQR)) | (df > (Q3 + 1.5 \* IQR)))

* + Treated outliers based on the IQR method to ensure data consistency.

#### 4. **Data Analysis and Visualization**

* **Visualization Tools Used:** Power BI
* **Key Visualizations:**
  + **Pie Charts:**
    - Distribution of Monthly Income by Gender
    - Distribution of Age and First Good/Bad Credit Status by Rented/Owned House Status
    - Distribution of Open Credit Lines and Loans by Region
  + **Bar Charts:**
    - Distribution of Debt Ratio
    - Distribution of Good/Bad Credit Status by Number of Open Credit Lines and Loans
    - Distribution of Number of Times 90 Days Late and Loans by Region
  + **Line Graphs:**
    - Trend of the Number of Times 90 Days Late and Loans over time
  + **Map:**
    - Geographical Distribution of Open Credit Lines and Loans

The dashboard provides a comprehensive overview of the credit card data, allowing users to analyze various aspects such as income distribution, debt ratio, credit status, and geographical distribution of credit lines and loans. This visualization is relevant for understanding patterns and trends in credit card usage and financial behavior. 

#### 5. **Insights and Findings**

 **Gender and Monthly Income:**

The distribution of monthly income revealed that the median income is higher for individuals identifying as male compared to those identifying as female. This suggests a possible gender disparity in income levels among credit card users.

 **Housing Status and Credit Behavior:**

Individuals who own houses tend to have higher revolving utilization of unsecured lines compared to those who rent. This indicates that homeowners may have more credit available or are using more of their available credit.

 **Debt Ratio and Credit Risk:**

Higher debt ratios were observed to be associated with a greater likelihood of being categorized as "Bad" credit status. This underscores the importance of managing debt levels to maintain a good credit status.

 **Age Distribution:**

The age distribution showed a concentration of credit card users in the age range of 30-50 years. Younger and older age groups had fewer users, which may reflect different financial behaviors and credit needs across age groups.

 **Late Payment Patterns:**

The analysis of the number of times individuals were 30-59 days, 60-89 days, and 90 days past due revealed that a higher frequency of late payments is significantly correlated with a "Bad" credit status. This highlights the critical impact of timely payments on credit standing.

#### 6. **Conclusion**

In this project, we set out to analyze and visualize credit card data with the objective of uncovering insights into factors influencing credit behavior and status. Through meticulous data cleaning, outlier handling, and utilizing advanced visualization tools like Power BI, we were able to achieve a comprehensive understanding of the dataset.

**Potential Implications and Recommendations:**

* **Financial Literacy Programs:** Enhancing financial literacy can empower individuals to better manage their credit and debt.
* **Targeted Financial Support:** Providing support for high-risk individuals could improve overall credit health.
* **Policy Interventions:** Addressing income disparities and ensuring equitable access to credit can help bridge gaps across different demographic groups.