



**Data Glacier**

## **Week 13: Final Report**

**Group Name:** Solo Analyst

**Name:** Munirah Alfahaid

**Email:** munirah9hamad@gmail.com

**Country:** Saudi Arabia

**Specialization:** Data Analyst

## **1-Problem Description**

XYZ Credit Union in Latin America has been successful in selling individual banking products like credit cards, deposit accounts, and retirement accounts. However, they are facing a challenge in cross-selling, as existing customers are typically not purchasing more than one product. The goal of this project is to analyze the current situation and suggest strategies to increase cross-selling opportunities among existing customers without relying on machine learning solutions.

## **2-Business Understanding**

The objective of this project is to identify actionable strategies to enhance the cross-selling of banking products to existing customers. Cross-selling is crucial for increasing customer lifetime value, enhancing customer satisfaction by providing comprehensive financial solutions, and improving the overall profitability of XYZ Credit Union. By understanding the customer segments, their behaviors, and the factors influencing their purchasing decisions, we can recommend targeted approaches to promote multiple product ownership among customers.

## **3-Project Lifecycle Along with Deadline**

### **1. Business Understanding (1-2 days):**

- Review the problem statement and clarify objectives.
- Understand the key business goals and constraints.

### **2. Data Understanding (3-4 days):**

- Explore the dataset to understand its structure and contents.
- Identify key variables related to customer demographics, product ownership, and purchasing behavior.
- Assess data quality, including missing values, outliers, and inconsistencies.

### **3. Data Cleansing and Transformation (4-5 days):**

- Clean the data by handling missing values, correcting inconsistencies, and addressing outliers.
- Transform the data into a suitable format for analysis, including encoding categorical variables and normalizing numerical variables.
- Create new features that may provide additional insights into customer behavior.

### **4. Exploratory Data Analysis (5-6 days):**

- Perform analyses to uncover patterns and relationships.
- Segment customers based on relevant factors such as age, income, and product ownership.
- Identify trends and correlations that may impact cross-selling potential.

### **5. EDA Recommendation (ppt) (3-4 days):**

- Summarize the key findings from the EDA.
- Develop actionable recommendations for increasing cross-selling.
- Prepare a presentation with clear visualizations and insights.

### **6. Dashboard Creation (5-7 days):**

- Develop a dashboard to track key performance indicators (KPIs) such as customer segments, product ownership, and customer demographics.
- Ensure the dashboard provides actionable insights for business decision-makers.

## 7. Final Presentation Preparation (2-3 days):

- Compile all findings, recommendations, and visualizations into a comprehensive final presentation.
- Review and refine the presentation to ensure clarity and impact.

**Total Duration:** Approximately 3-4 weeks, depending on the complexity of the data and the depth of analysis required. Adjustments to the timeline may be necessary based on the project's progress and any additional data requirements.

## 4-Data Collection

### Cross Selling data

<b>Total number of observations</b>	929615
<b>Total number of files</b>	1
<b>Total number of features</b>	24
<b>Base format of the file</b>	csv
<b>Size of the data</b>	110.3 MB

## **5-Data Cleansing and Transformation**

To address the issue of missing values in our dataset, the following approaches were employed:

### **Categorical Data Imputation:**

For the categorical column `sexo`, which contains missing values, the most frequent value (mode) was used to fill in the missing entries. This is a common practice for categorical variables where the most frequent category is assumed to be the best replacement for missing values.

### **Numeric Data Imputation:**

For the numeric column `cod_prov`, which had missing values, the median of the column was used to fill in the missing entries. The median is preferred over the mean because it is less affected by outliers, providing a more robust central tendency measure.

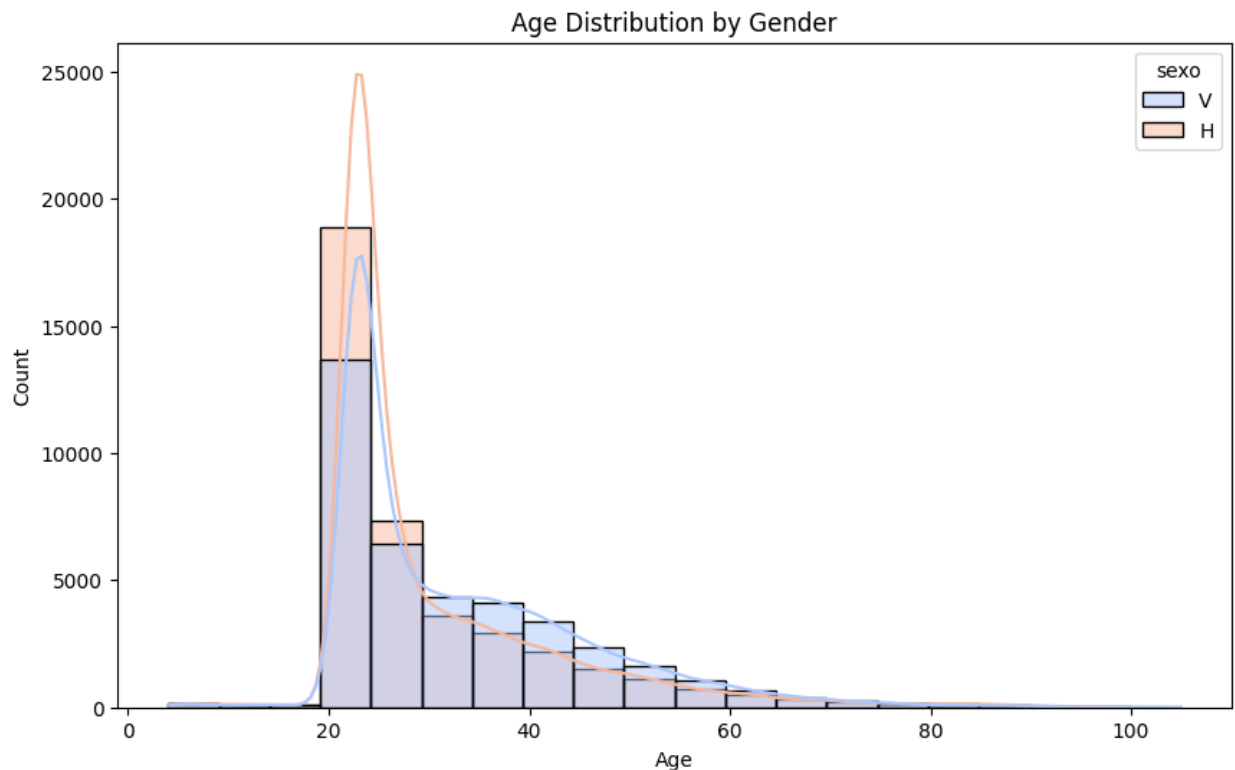
### **KNN Imputation for Numeric Columns:**

To handle missing values in other numeric columns, the K-Nearest Neighbors (KNN) imputation method was applied. This method imputes missing values by finding the nearest neighbors (in terms of other features) and averaging their values. This technique is useful for capturing the underlying structure in the data.

## Handling Non-Numeric Columns Separately:

For non-numeric columns that were not suitable for KNN imputation, the mode was used to fill in the missing values. This ensured that categorical data were treated appropriately, maintaining the integrity of the dataset.

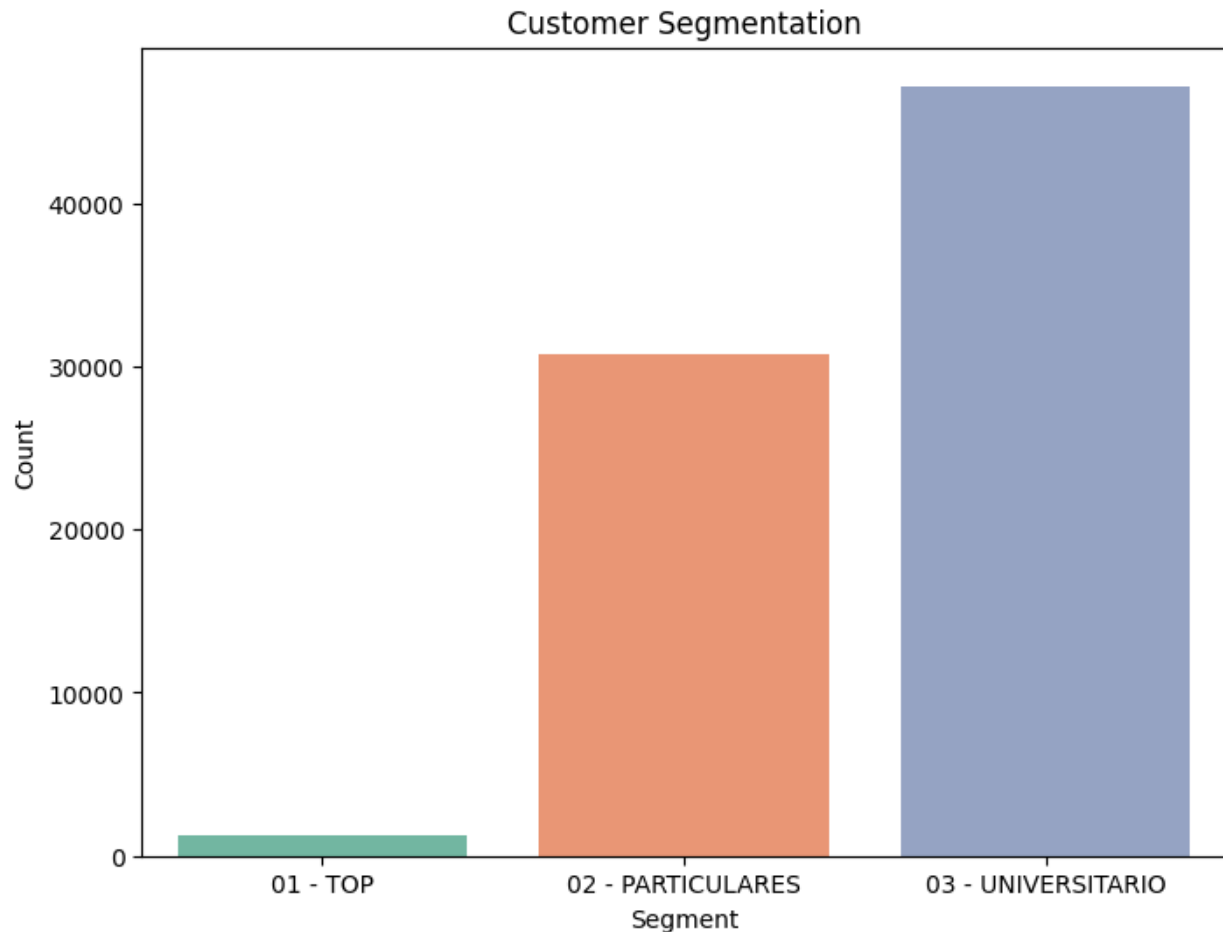
## 6- EDA



- Most customers fall in the 20 to 40-year-old range, with a sharp peak around the 25-30 age group.
- There are more male customers in this age range than females, as seen by the taller blue bars.
- The chart shows that as age increases, the number of customers decreases for both genders.

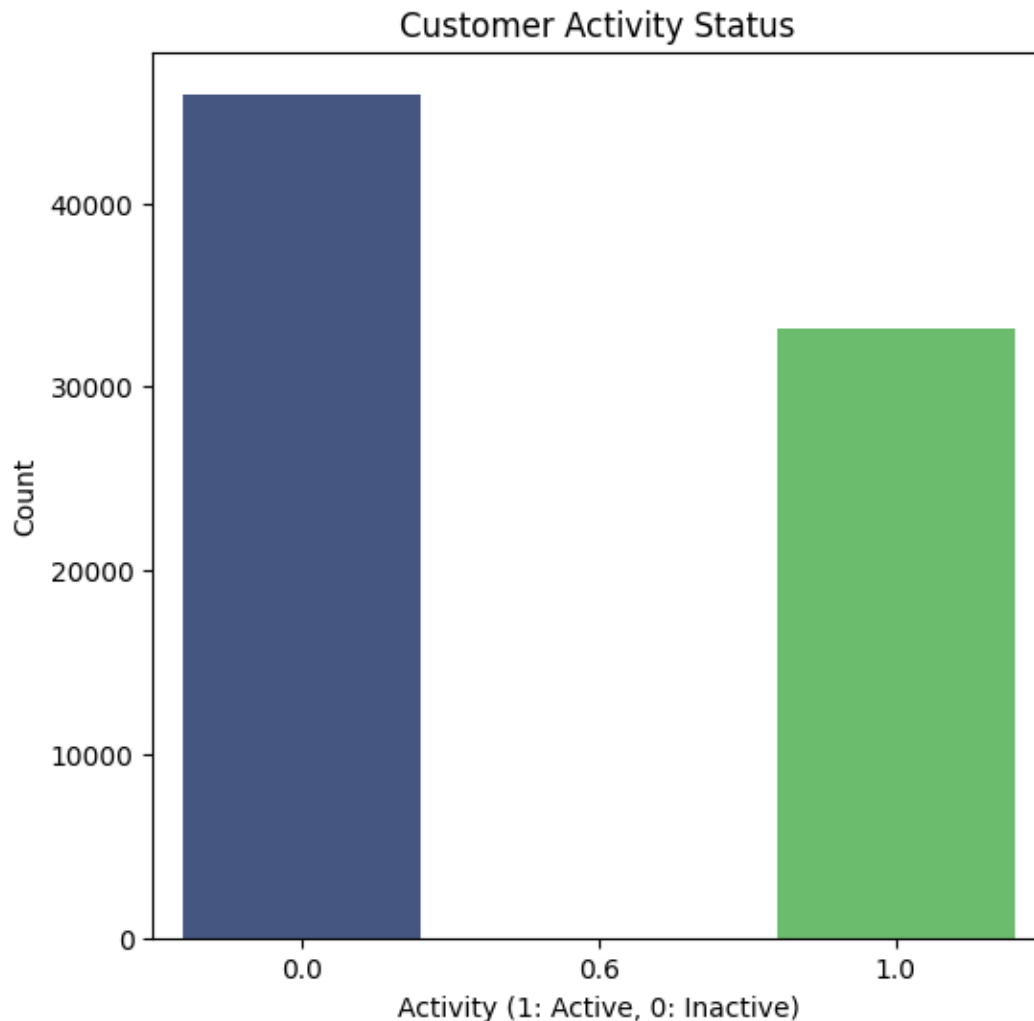
There are outliers beyond 100 years of age, but they are very few.

This visualization suggests that the bank's customer base is concentrated in the younger population (20-40), which could be important in understanding customer behavior and how to market different products for cross-selling.



- **PARTICULARES:** The majority of the customers fall into this segment, indicating that regular banking services are the most commonly used among the bank's clients.
- **UNIVERSITARIO:** This group is also substantial, suggesting the bank has a significant number of young or student customers.
- **TOP:** This segment is quite small, meaning that there are relatively few high-value or priority customers.

This distribution is useful for tailoring marketing strategies and cross-selling products based on segment size and customer needs.

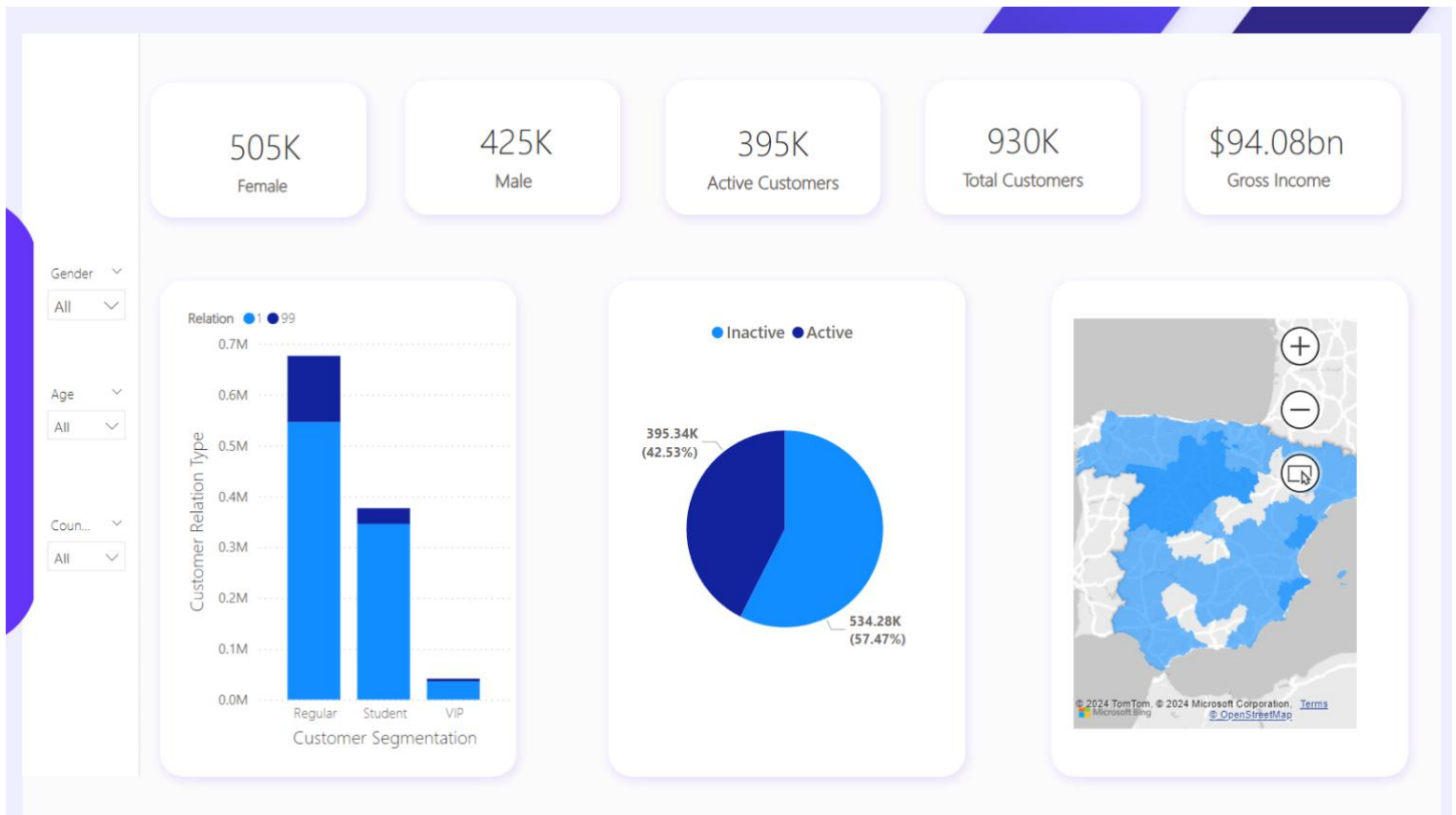


- **Inactive Customers (0.0):** There are more inactive customers than active ones, with over 500,000 customers in this category. This could suggest that a large portion of the bank's customers are no longer using its products or services.
- **Active Customers (1.0):** Although fewer in number compared to inactive customers, a substantial number (over 400,000) are still actively engaging with the bank.



## 7- Dashboard

I designed this dashboard with a focus on visualizing the key performance indicators (KPIs) that are critical to understanding customer segmentation and activity patterns. I used Figma for the initial design, which helped me focus on the overall structure and presentation before implementing it in Power BI for interactivity and deeper data insights.



## Key Performance Indicators (KPIs)

- **Female Customers:** 505K
- **Male Customers:** 425K
- **Active Customers:** 395K
- **Total Customers:** 930K
- **Gross Income:** \$94.08bn

## Dashboard Overview

The dashboard provides an intuitive view of:

- **Customer Segmentation:** This bar chart shows the distribution of customer types (Regular, Student, VIP) segmented by active and inactive customers.
- **Customer Activity:** The pie chart visualizes the percentage of active vs inactive customers, giving a quick glance at the customer engagement.
- **Geographical Distribution:** The map on the right highlights the spread of customers by region, providing insights into where most of the customer base is located.

## Interactivity

The dashboard is fully interactive, allowing users to filter data dynamically using the slicers on the left:

- **Gender:** Users can filter data based on gender.
- **Age:** Allows slicing the data to see customer segmentation and activity patterns by age.
- **Country:** Filters customers based on their country of residence.

Each slicer dynamically updates the KPIs and visualizations, enabling a comprehensive and user-friendly exploration of the dataset. By interacting with these slicers, you can observe changes in customer activity and gross income, providing insights to optimize cross-selling strategies.