Data Glacier

Week 11: Deliverable

Data Analyst: Cross Selling Recommendation Project

Team Member Details

Group Name: Cosmic Analyst

Name: Bilgan Kiris

Email: bilgan2001@gmail.com

Country: Canada

College : Durham College

Specialization : Data Analyst

Problem Description

XYZ Credit Union in Latin America excels in selling individual banking products (e.g., credit cards, deposit accounts, retirement accounts). However, their customers rarely purchase multiple products, indicating low cross-selling performance. This project aims to analyze customer data and recommend actionable strategies to improve cross-selling for their products.

EDA Report

Overview

The purpose of the Exploratory Data Analysis (EDA) is to gain insights into the dataset by understanding the distributions, relationships, and potential anomalies within the data. This report outlines the steps performed and key findings during EDA.

Data Review and Cleaning

- The dataset comprises 929,615 rows and 27 columns.

- **Missing Values**: There were no missing values in the dataset.

- **Data Types**: The dataset contains both numerical and categorical features, with appropriate encoding applied where necessary.

- **Outlier Handling**: Outliers in the `renta` column were identified using Z-scores, but no rows were removed as all Z-scores were within the acceptable range (-3 to 3).

Future Engineering

- Log transformations were applied to numerical columns such as `renta`, `age`, and `indrel` to reduce skewness and normalize their distributions.

- The `segmento` column was cleaned to simplify customer segmentation categories into `PARTICULARES`, `TOP`, and `UNIVERSITARIO`.

Data Visualization

Age Distribution Across Customer Segments

Customer Segmentation Distribution
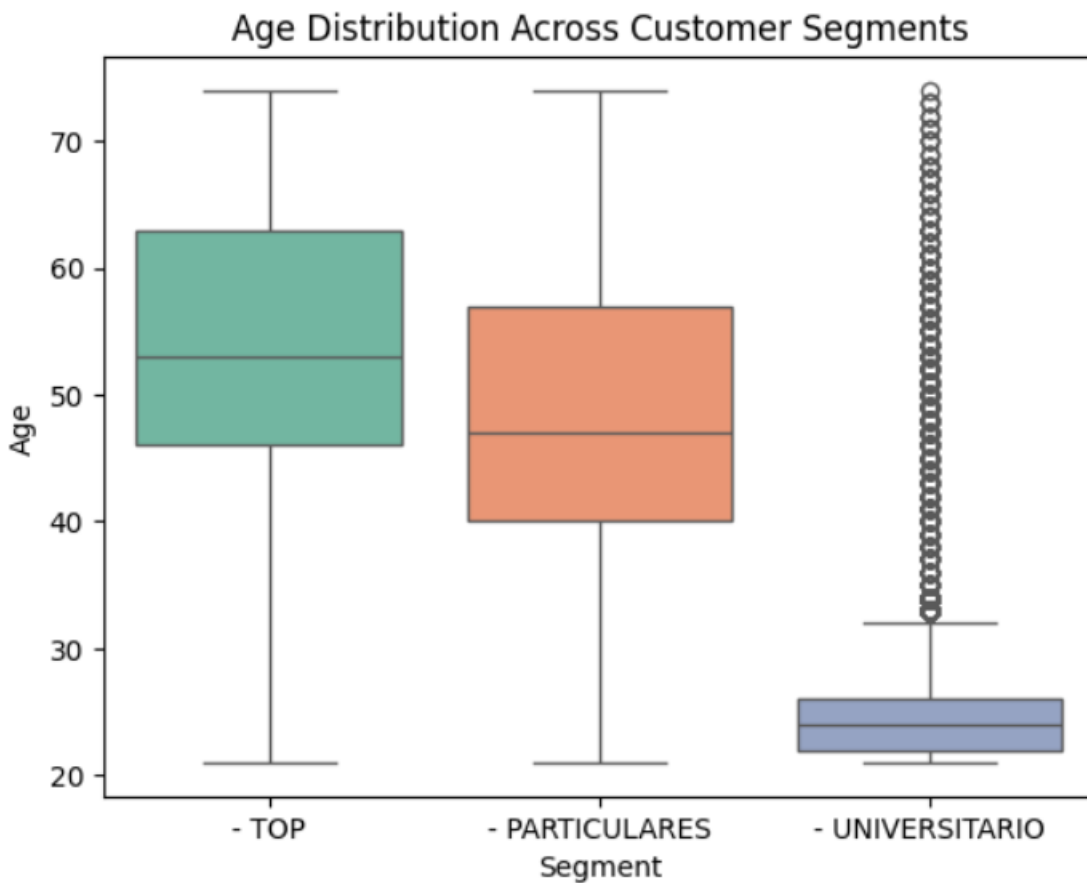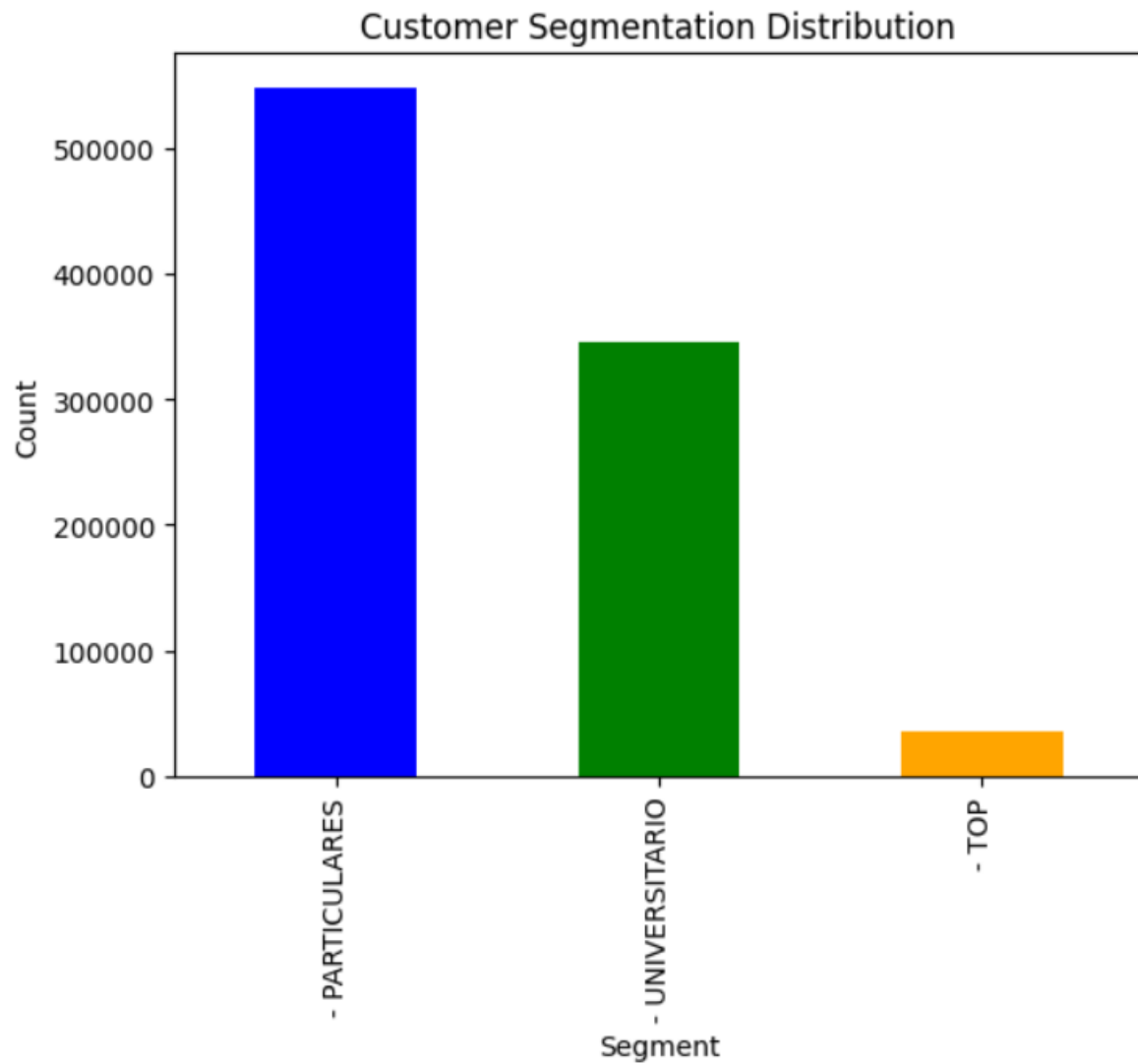
The cleaned `segmento` column indicates the distribution of customers across three main segments:

Segment	Count	Percentage
PARTICULARES	547,626	58.9%
UNIVERSITARIO	346,028	37.2%
TOP	35,961	3.9%

- The majority of customers (58.9%) fall into the `PARTICULARES` segment, representing individual customers.

- `UNIVERSITARIO` accounts for 37.2% of customers, primarily young or student accounts.

- A small proportion (3.9%) belongs to the `TOP` segment, representing high-value customers.

3.2 Age Distribution Across Customer Segments

The average age distribution within each customer segment provides insights into the demographic focus of each group:

Segment	Average Age (years)
PARTICULARES	48.7
UNIVERSITARIO	24.5
TOP	54.6

- `PARTICULARES` customers tend to be middle-aged.

- `UNIVERSITARIO` customers are predominantly young, likely indicating students or recent graduates.

- `TOP` customers are typically older, likely reflecting their established financial standing.

- The income (`renta`) column was standardized using Z-scores to detect and handle outliers.

- Log transformations revealed that the `renta` column exhibited a normalized distribution post-transformation.

- The majority of customers have a moderate income level, with a small proportion in higher-income brackets.

Conclusion

The EDA process provided key insights into customer segmentation, age distribution, and income levels. These findings can guide further analysis and model building to target specific customer groups effectively. Future steps include feature extraction, correlation analysis, and predictive modeling.