Cross Selling EDA

Navodith Shankar
ns229@njit.edu
United States
NJIT
Data Analyst
04/16/2023

Agenda

Problem Statement

Datasets

EDA

Recommendations

Problem Statement

Problem Statement

- XYZ credit union in Latin America is performing very well in selling the Banking products (eg: Credit card, deposit account, retirement account, safe deposit box etc) but their existing customer is not not buying more than 1 product which means bank is not performing good in cross selling (Bank is not able to sell their other offerings to existing customer). XYZ Credit Union decided to approach ABC analytics to solve their problem.
- Solve the problem, the credit union is looking to leverage data and analytics to develop a better understanding of their customers and improve their cross-selling efforts

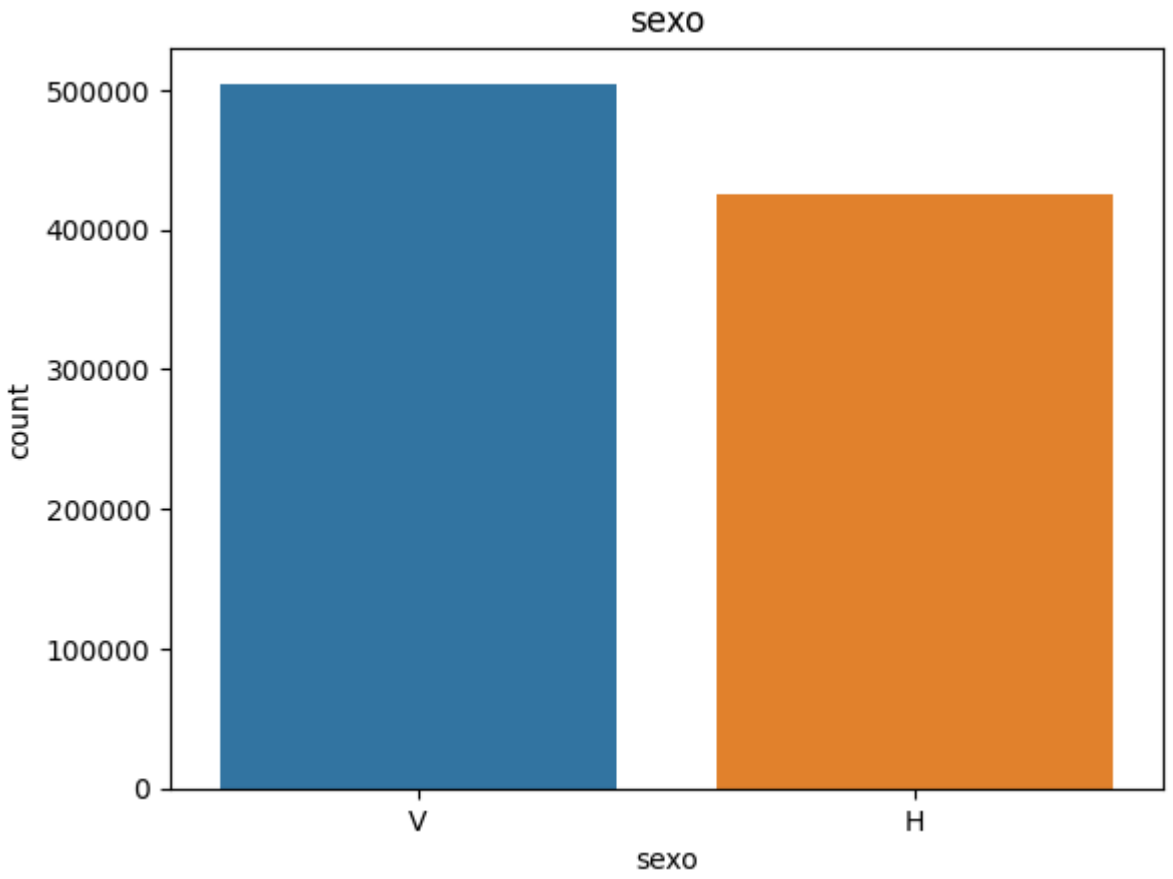
Data Sets

Data Sets

- Test.csv
- Train.csv

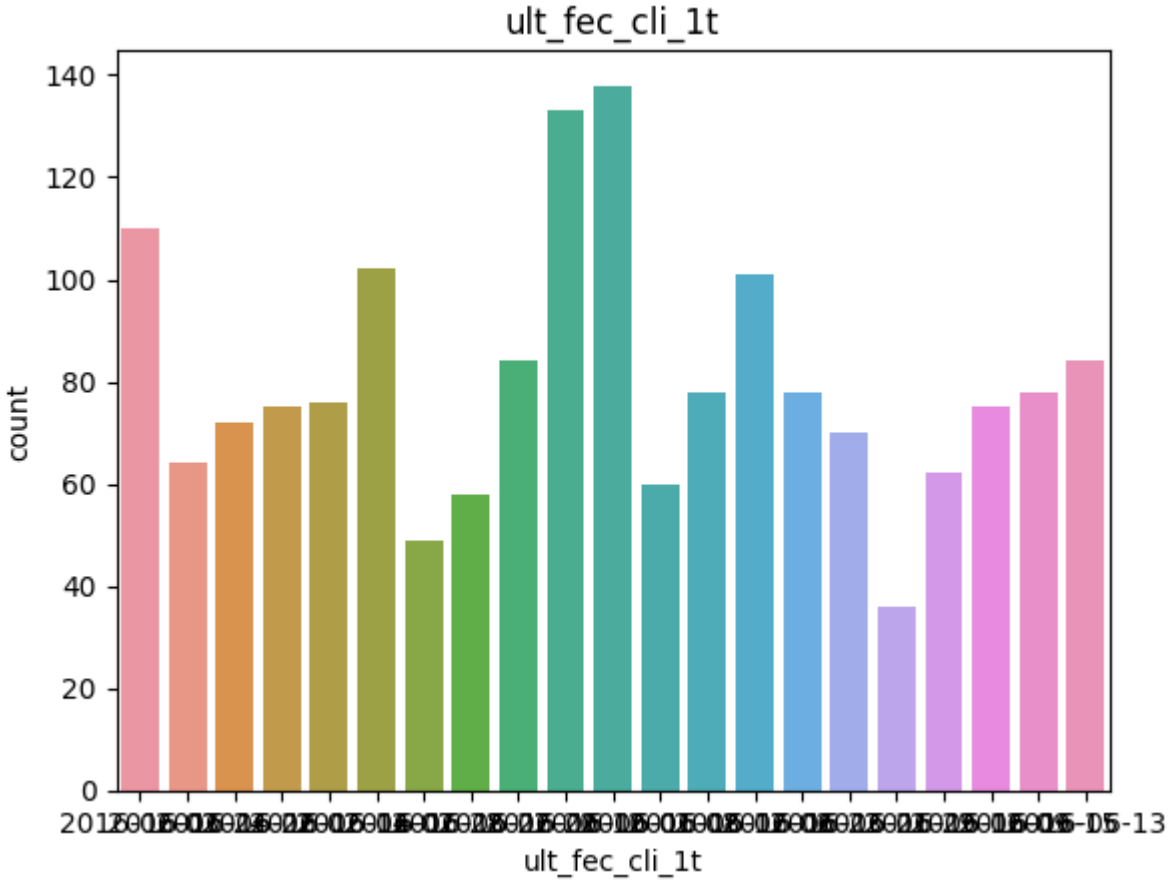
EDA & Imputation

Gender Distribution



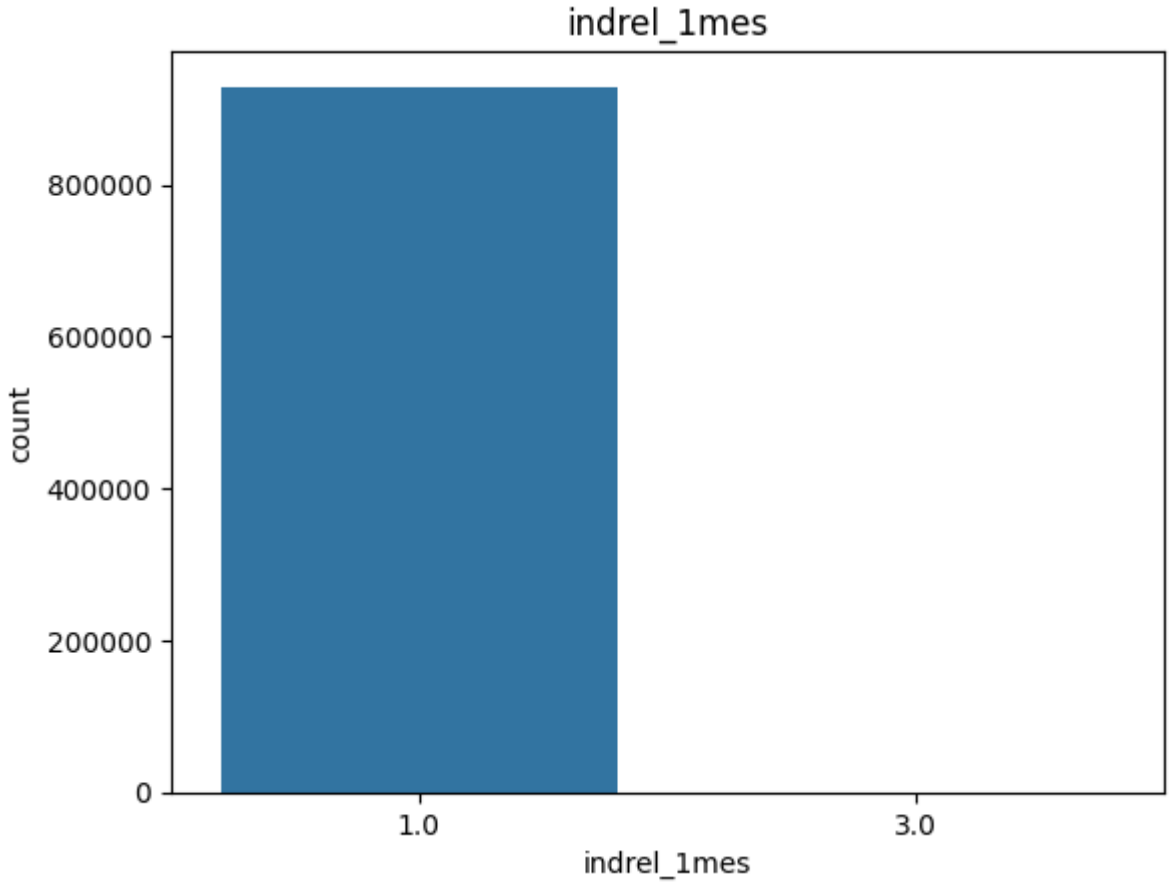
- As we can see, V is the most popular with the count being 500,000 and H having a count of over 400,000

Primary User Analysis



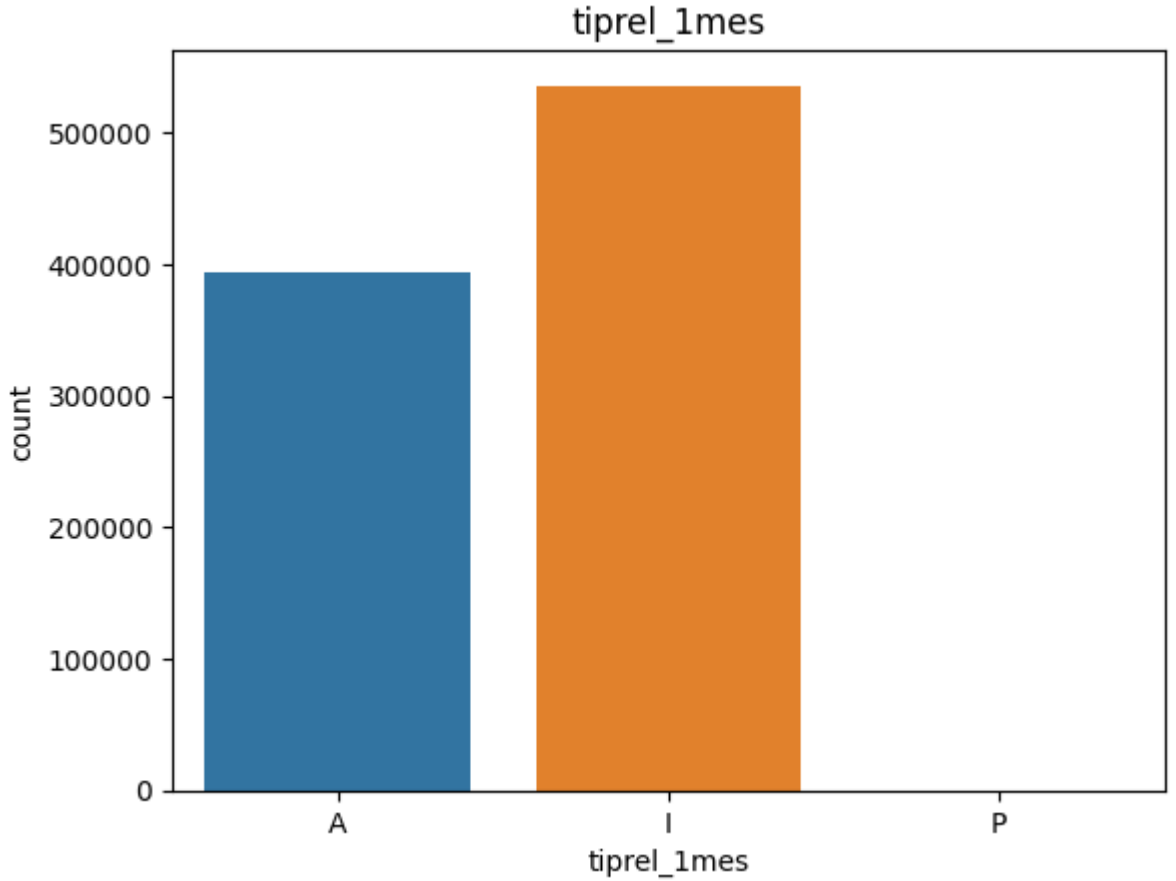
- This is a very big dataset, displaying all the dates clearly is tough but by going through the datasets, most customers stop being a primary customer during the middle of the year.

Customer Type



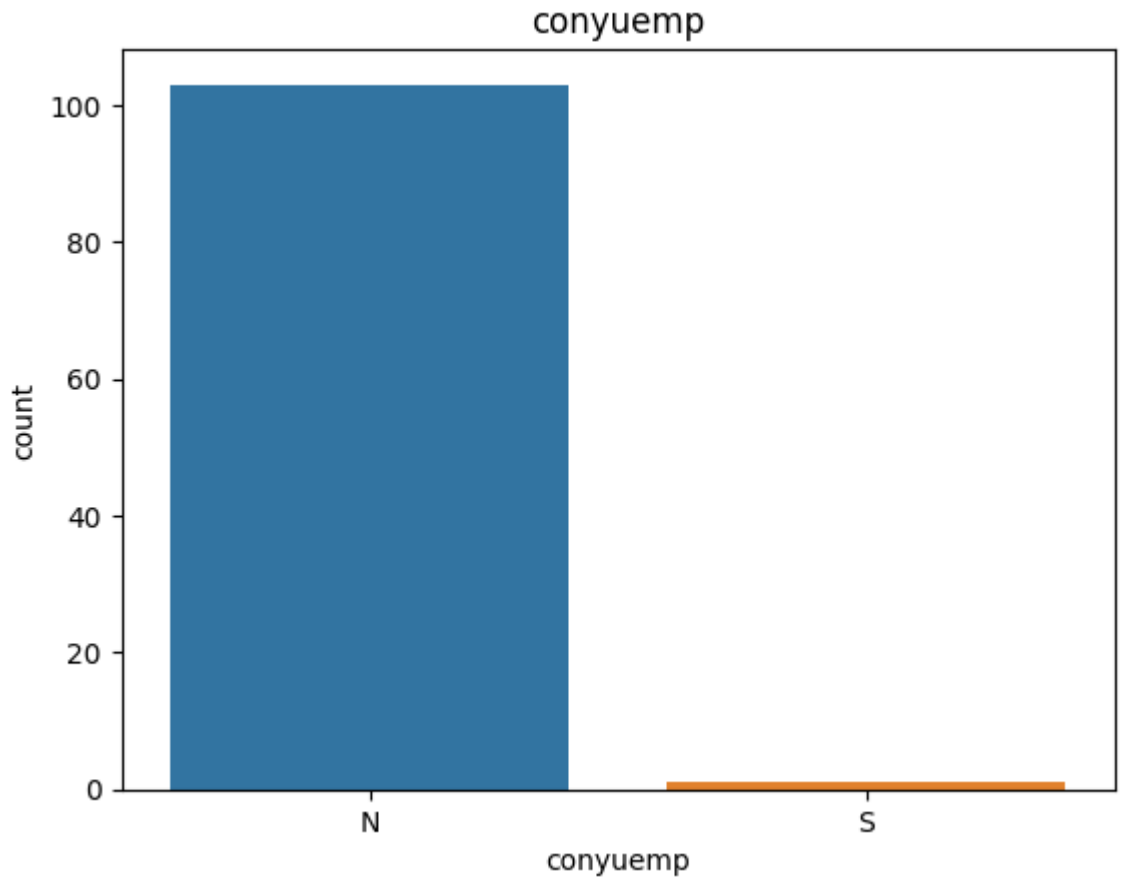
- As we can see, the only customer types are 1 and 3 which is First/Primary and Former Primary with a lot of customers being primary customers with count of over 800,000. In the dataset, there is a very small amount of Former Primary customers but it is so small it cannot be illustrated.

Customer Relation



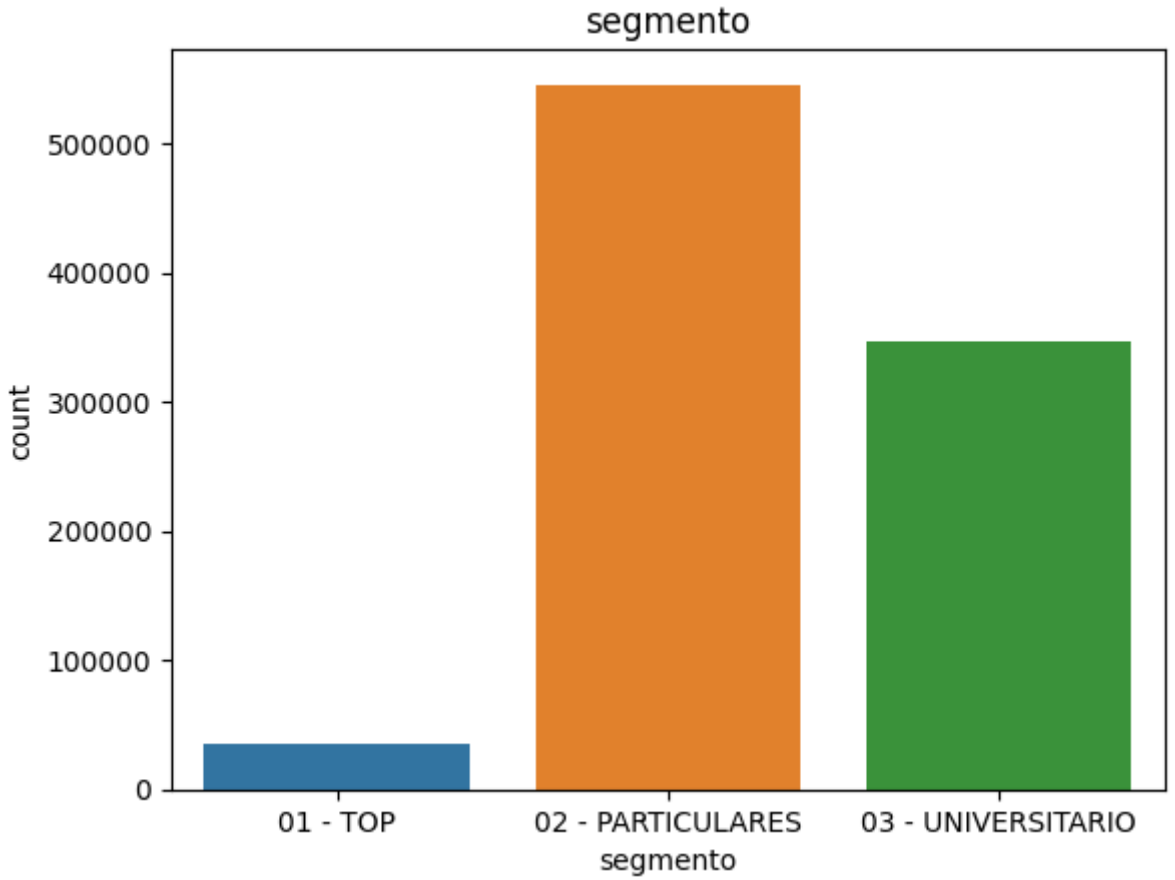
- As we can see, there are only three types customer relations, Active, Inactive and Former customer. We can see that there is a high amount of Inactive users compared to Active users and almost no former customers. The exact number of Former customers can be seen in the dataset description. The numbers are so small it cannot be illustrated.

Customer-Spouse Relation



- As we can see, most customers are the spouse of an employee (N) whereas some customers are not the spouse of an employee (S)

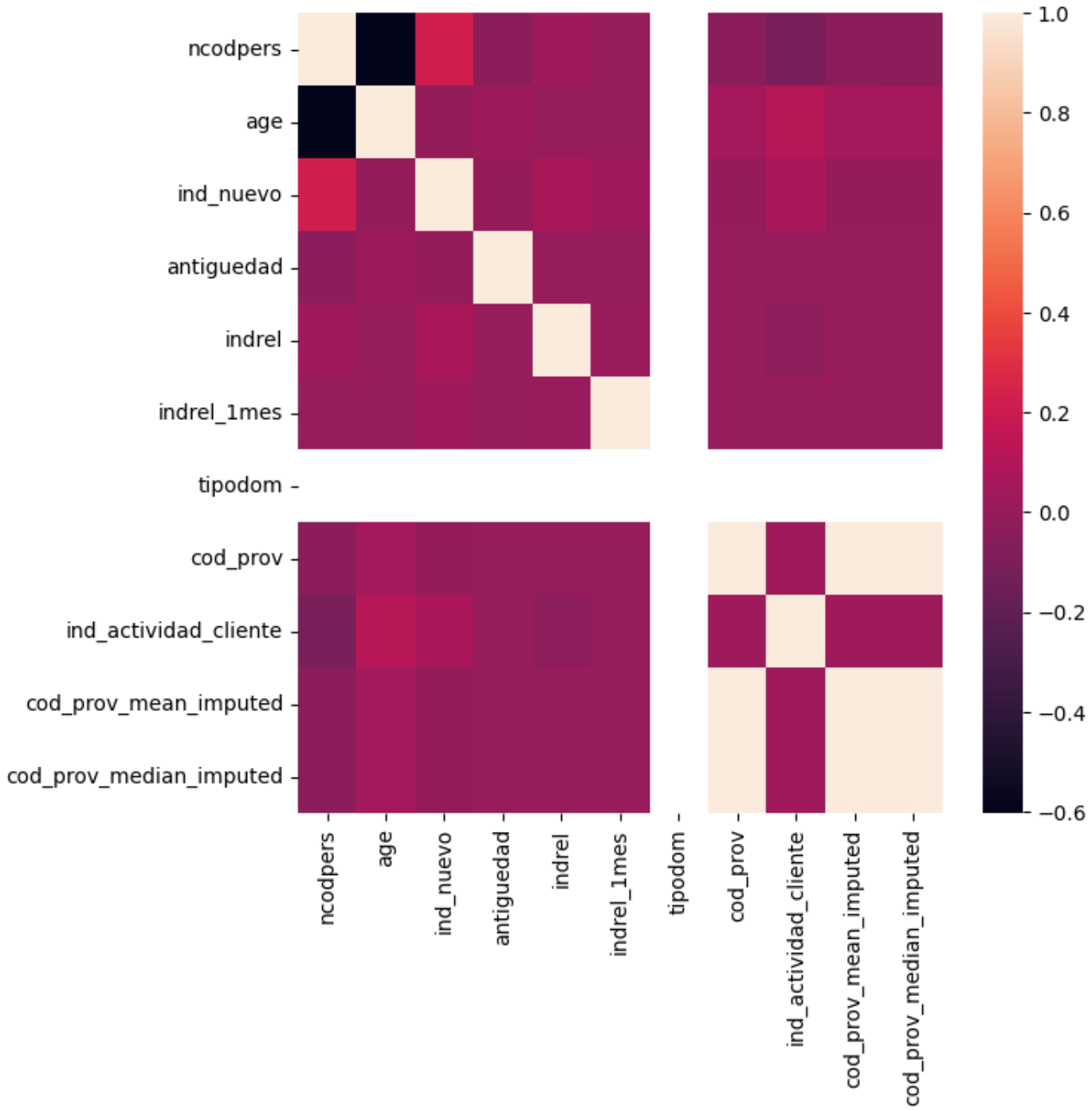
Segmentation Analysis



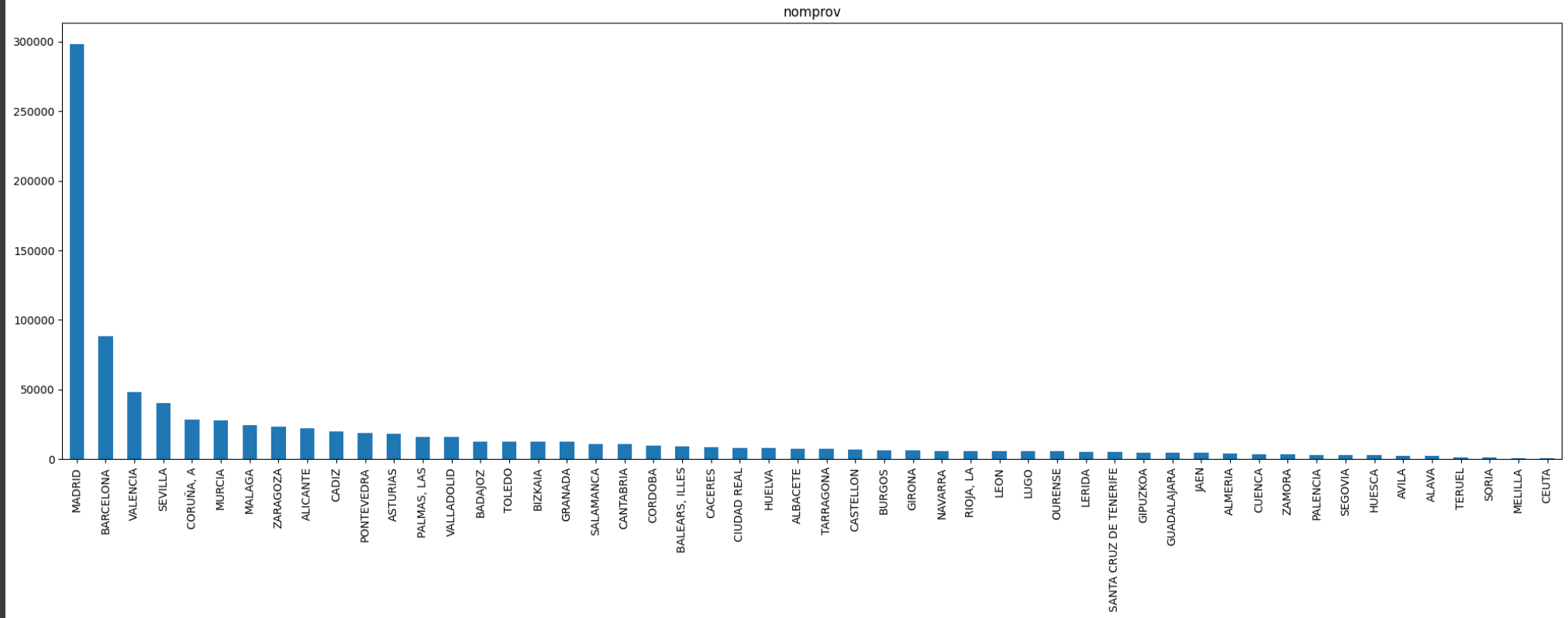
- As we can see, most customers are Individuals (02) with a count of over 500,000 and some customers are college graduates (03) and a small amount of customers are VIPS (01)

Data Glacier
Your Deep Learning Partner

Correlation Heat Map



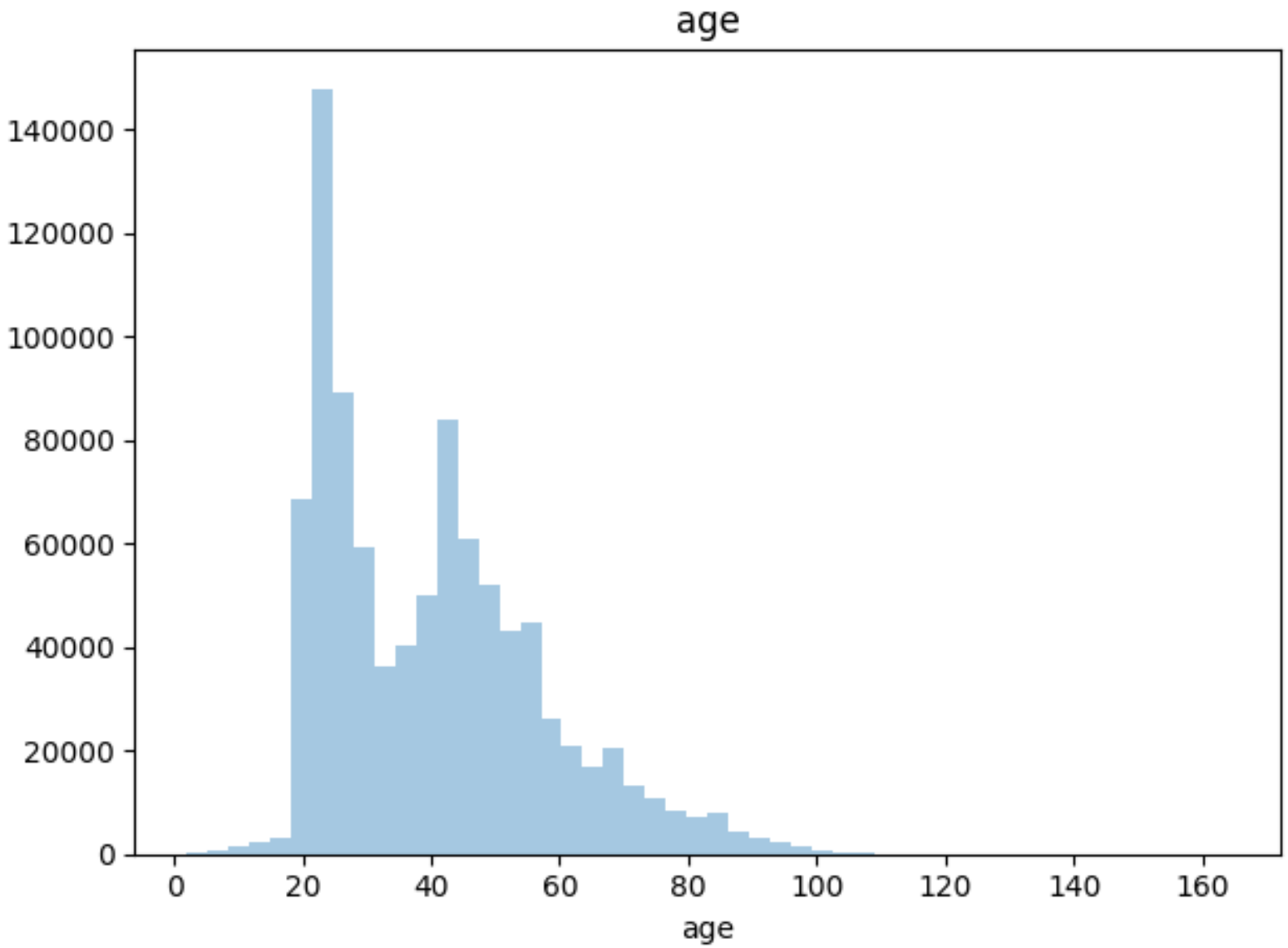
Province Names



- As we can see, the most popular province is Madrid with a count of almost 300,000 which the credit union can take advantage of. Most of the other provinces are not as popular as Madrid

Data Glacier
Your Deep Learning Partner

Age

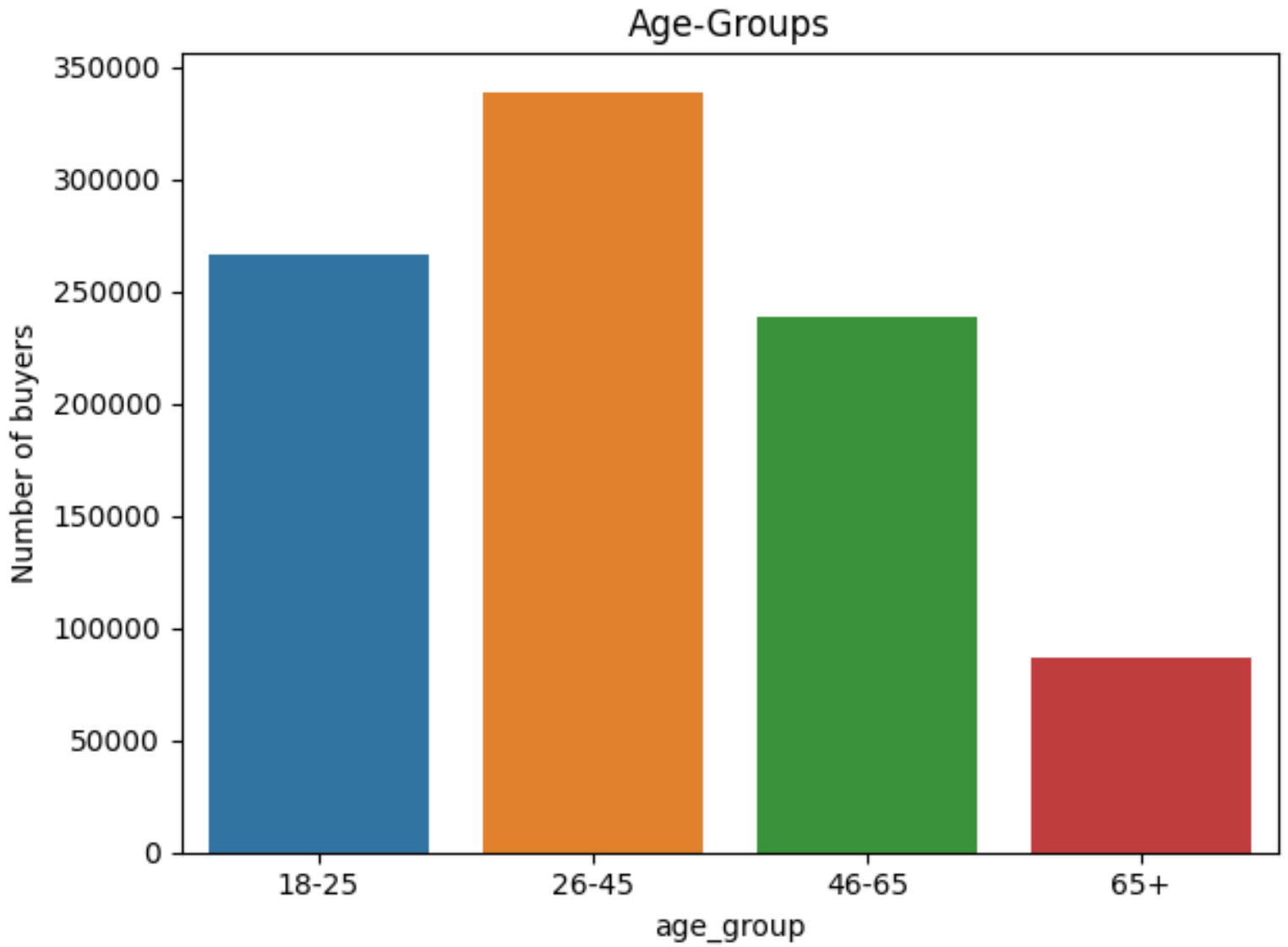


- As we can see, the age of most customers lie in between 20-60 with the most popular range being from 18 – 30.

Data Glacier
Your Deep Learning Partner

Age (cont.)

	age_group	count
0	18-25	265964
1	26-45	338802
2	46-65	238322
3	65+	86527



- As we can see more clearly now, the most popular age group is 26-45 with a count of almost 350,000

Data Glacier
Your Deep Learning Partner

Recommendation

- Based on the exploratory data analysis, it is recommended that XYZ credit union focuses on improving their customer retention strategies and offering personalized and targeted promotions to their customers to increase cross-selling opportunities. Additionally, the bank should consider providing better customer service and investing in digital channels to enhance the overall customer experience. Finally, the bank should closely monitor customer behavior and use data-driven insights to continuously optimize their cross-selling efforts.

Recommendation for models

- Logistic Regression
- Decision Tree
- Random Forest
- Ada Boost

Cross Selling EDA

Navodith Shankar

04/17/2023

Data Glacier

Your Deep Learning Partner

Thank You