Cross-Selling Data Analysis

<Virtual Internship>

**<20-Sep-2023>**

Agenda

Problem Statement

General EDA

Further EDA

Recommendations

Problem Statement

- Background: XYZ, a US-based private firm, intends to explore investment opportunities in Cab Industry, which has experienced remarkable growth recently. As part of their Go-to-Market (G2M) strategy, they aim to gain comprehensive market insights before making any final decisions.

- Objective : Leverage data to provide actionable business insights for XYZ firm, supporting their decision-making on which cab company to invest.
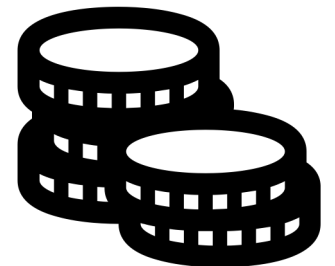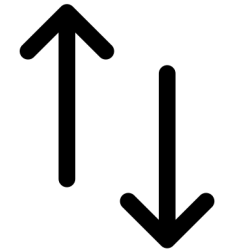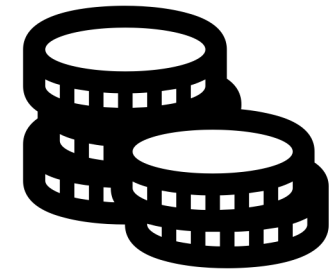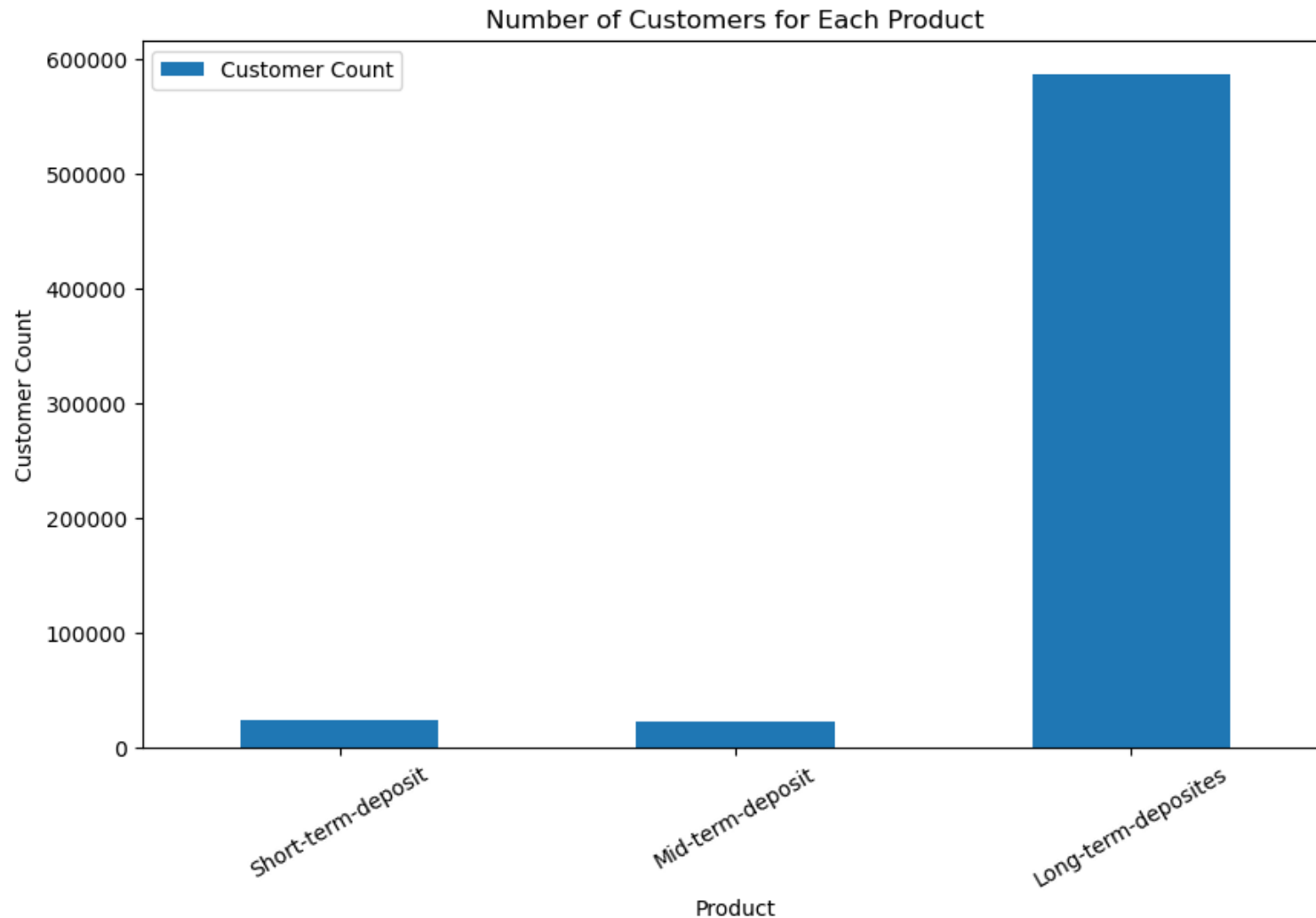
Approach

- Data Preprocessing

- Hypothesis Generation

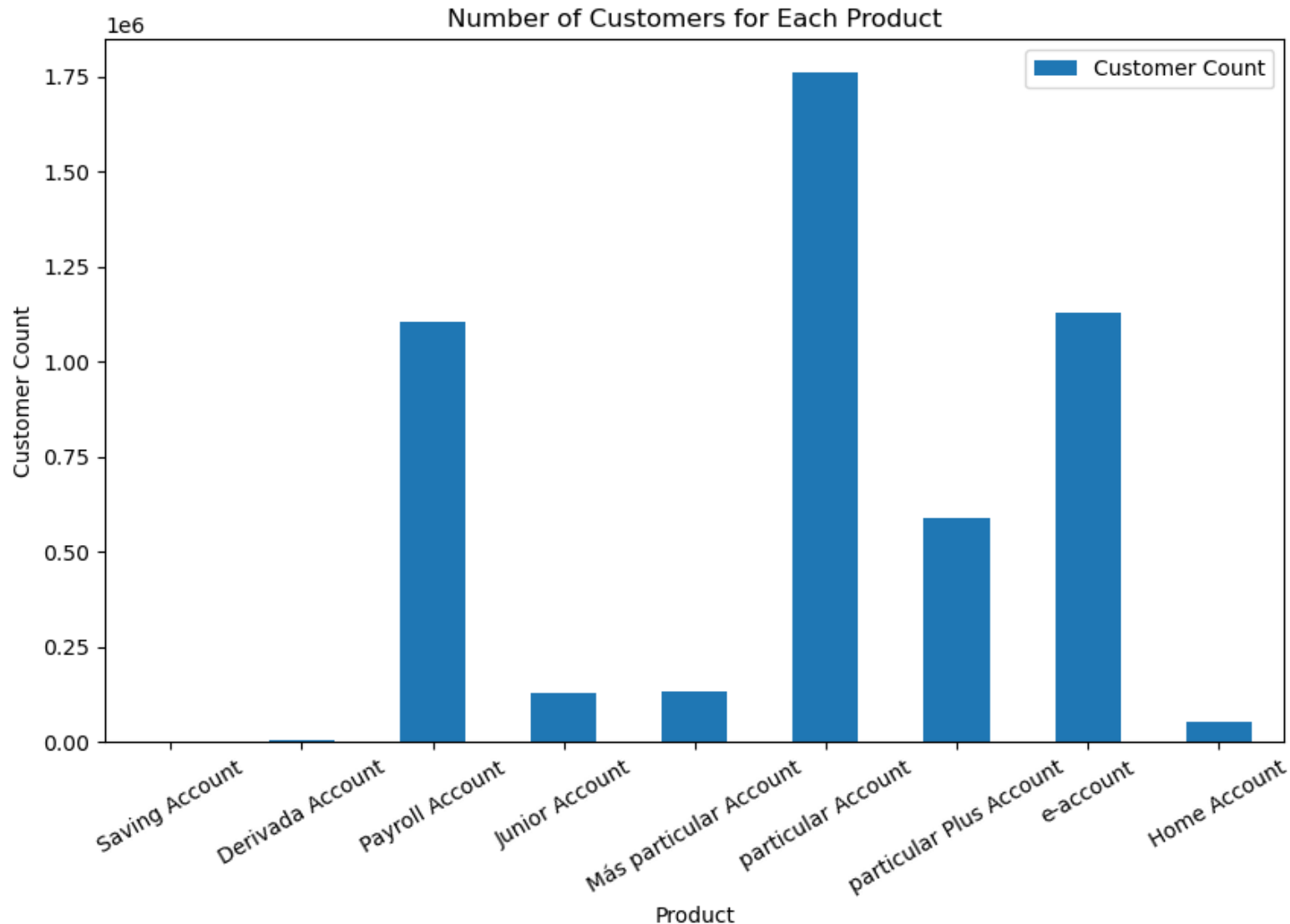- Explorative Data Analysis

- Data Visualization

General EDA for different products

For short-term, mid-term and long-term deposit, the long-term deposit outrages the other two

General EDA for different products

Number of Customers for Each Product

For bank account products, the Payroll account and e-account have largest number of users.

General EDA for different products

Number of Customers for Each Product
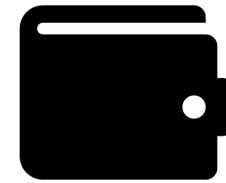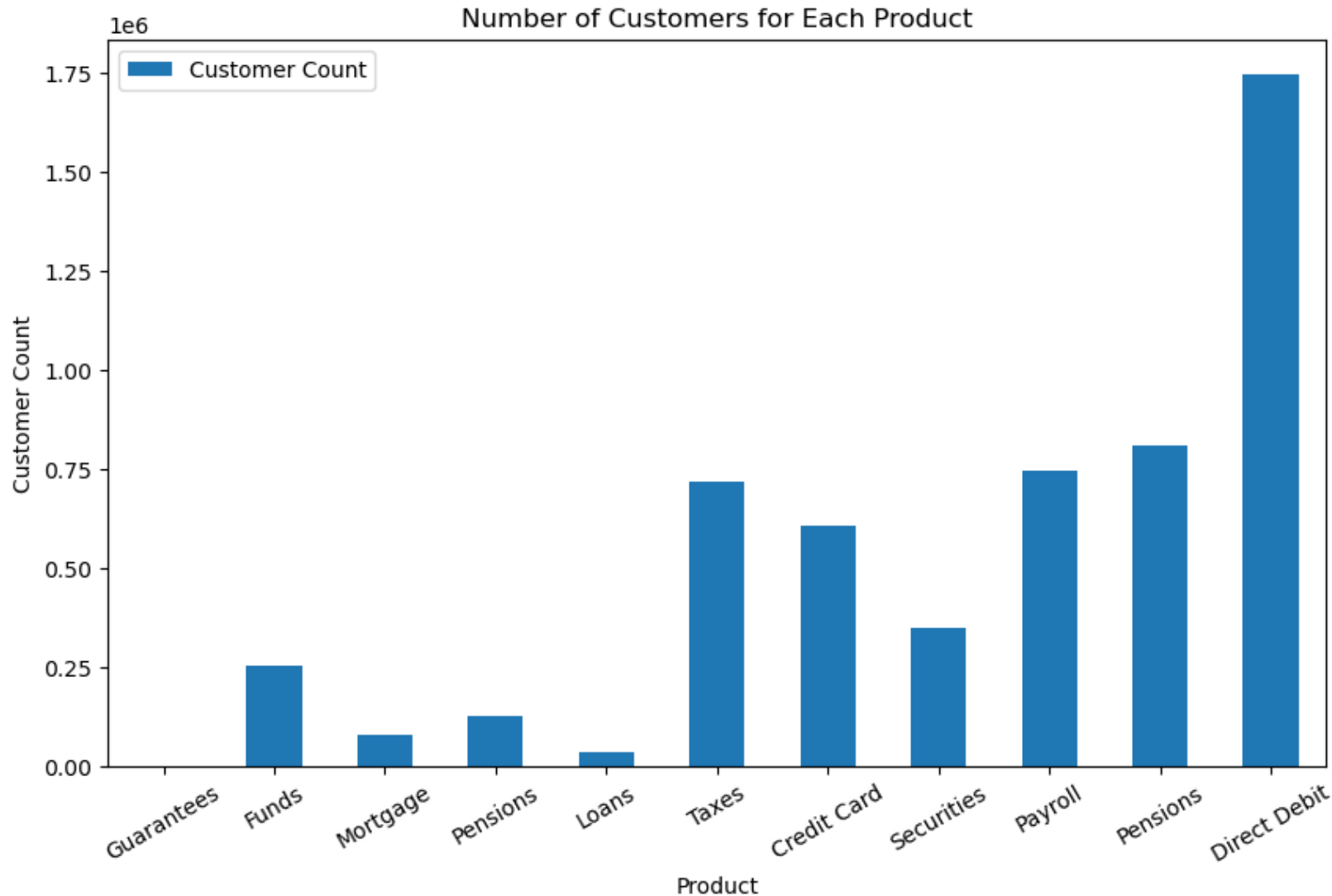
For bank services product, Direct Debit has largest number of user.

Further Analysis

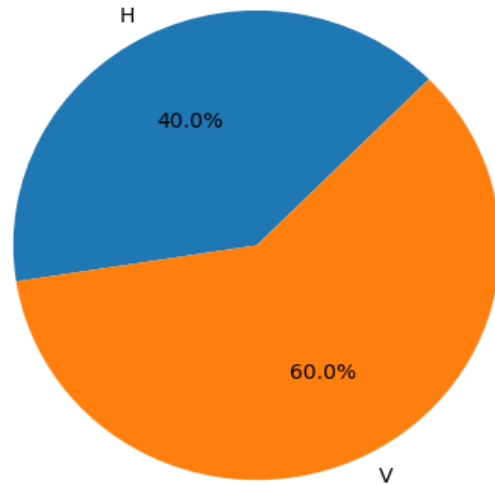- Based on previous analysis, we have selected 4 most popular products for further analysis. The purpose for this is to paint a 'User portrait' (identifying features of customers).

Long-term Deposit

age_group Distribution
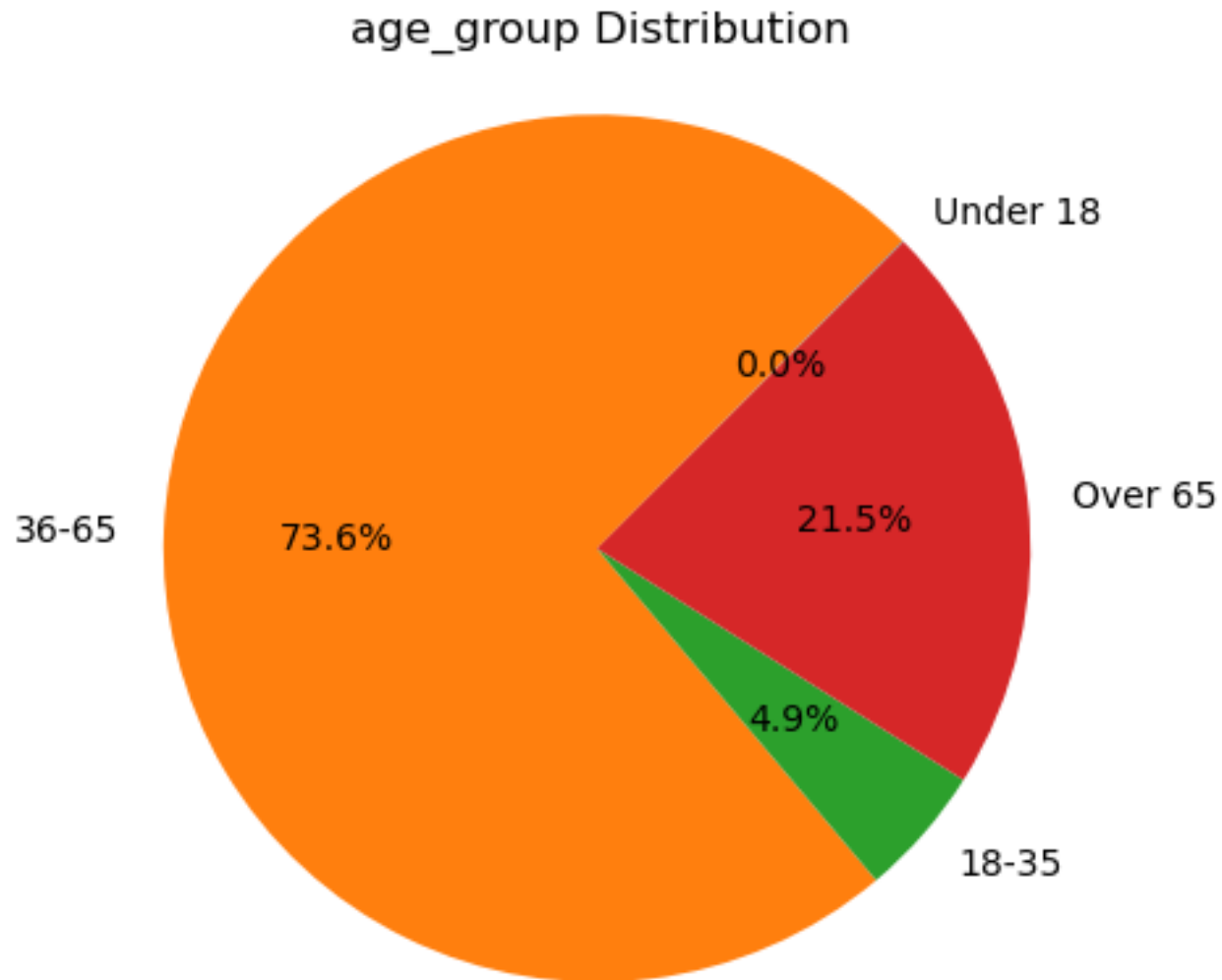
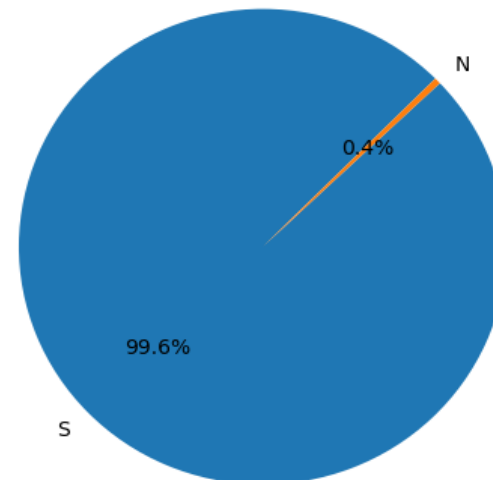Age group 36-65, and over 65 years old are the largest age group choosing long-term deposit
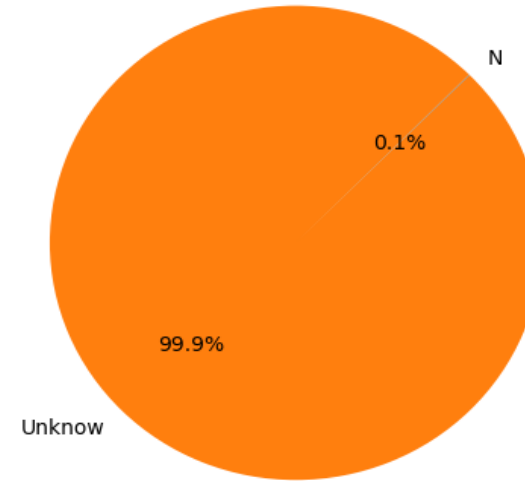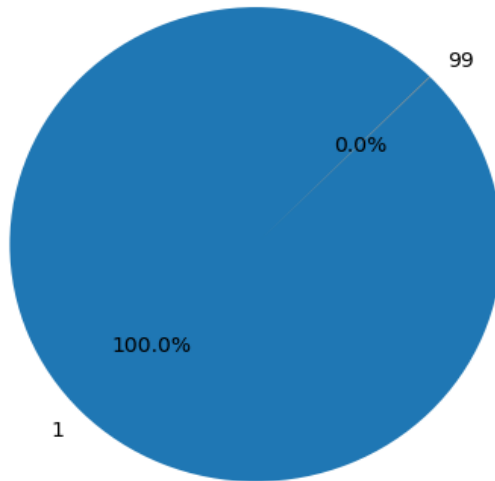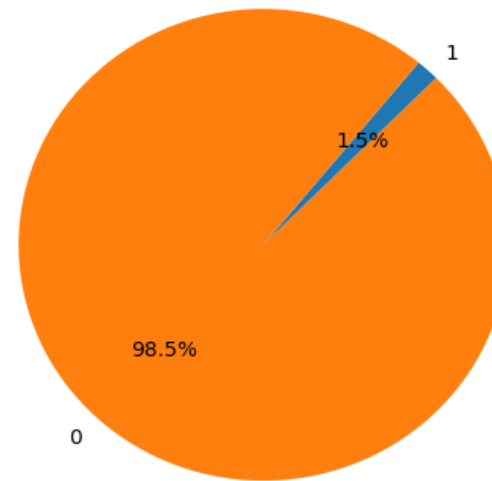
Long-term Deposit
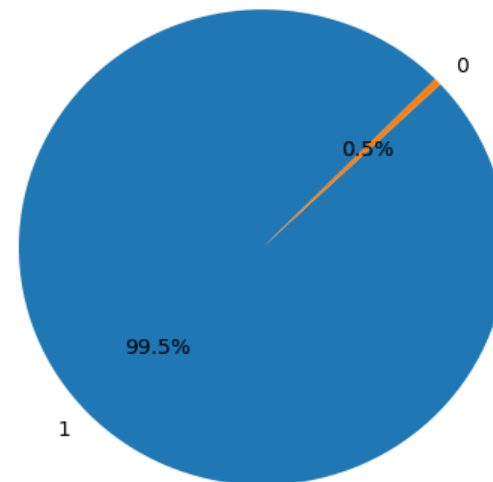
Long-term Deposit

Foreigner_index Distribution
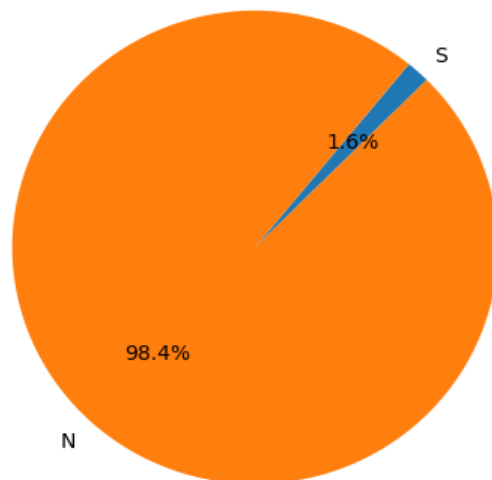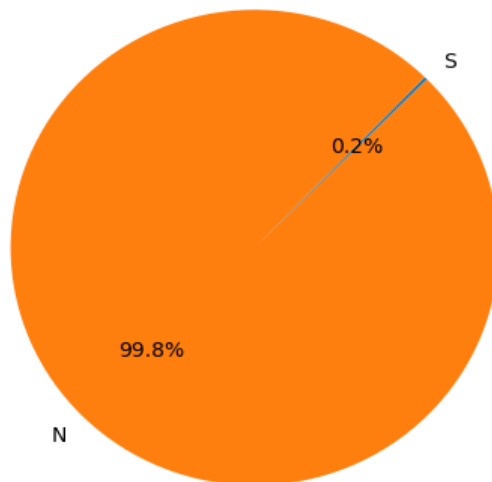
Residence_index Distribution

Decreased_index Distribution

Active Distribution

Long-term Deposit
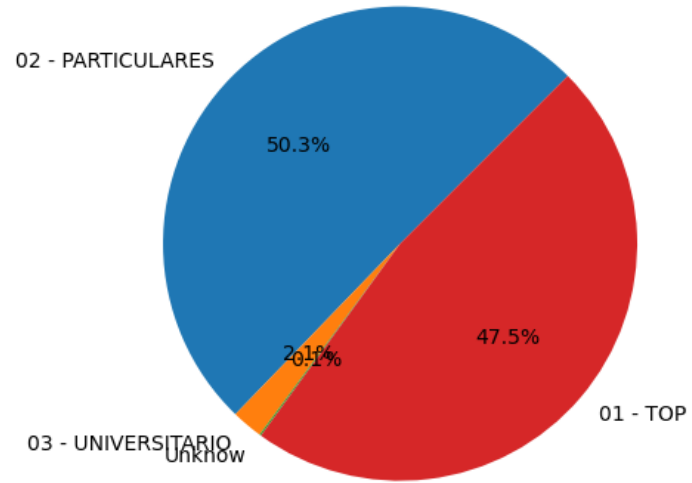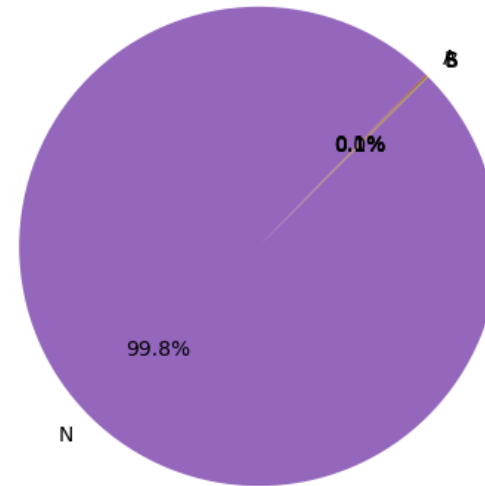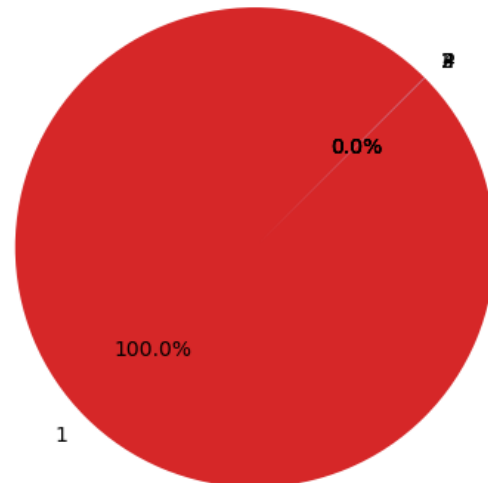
Long-term Deposit

Gross Income Distribution
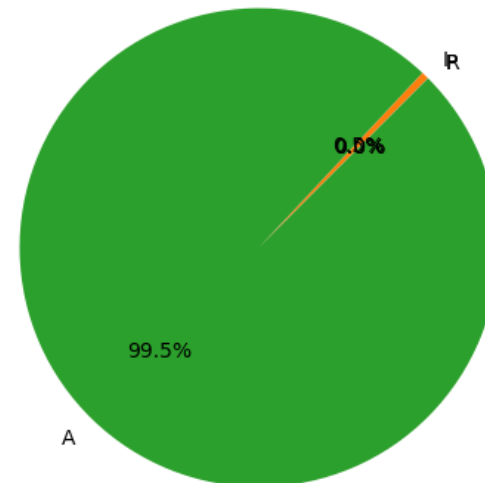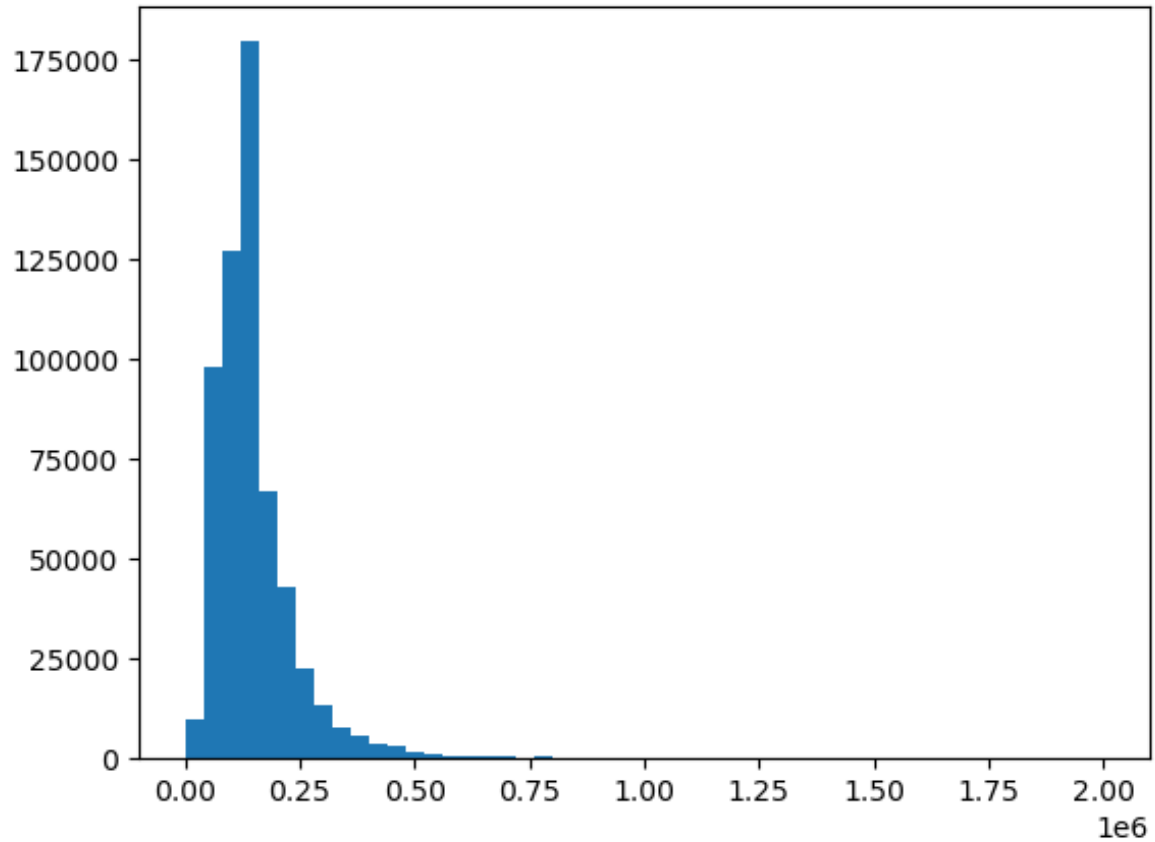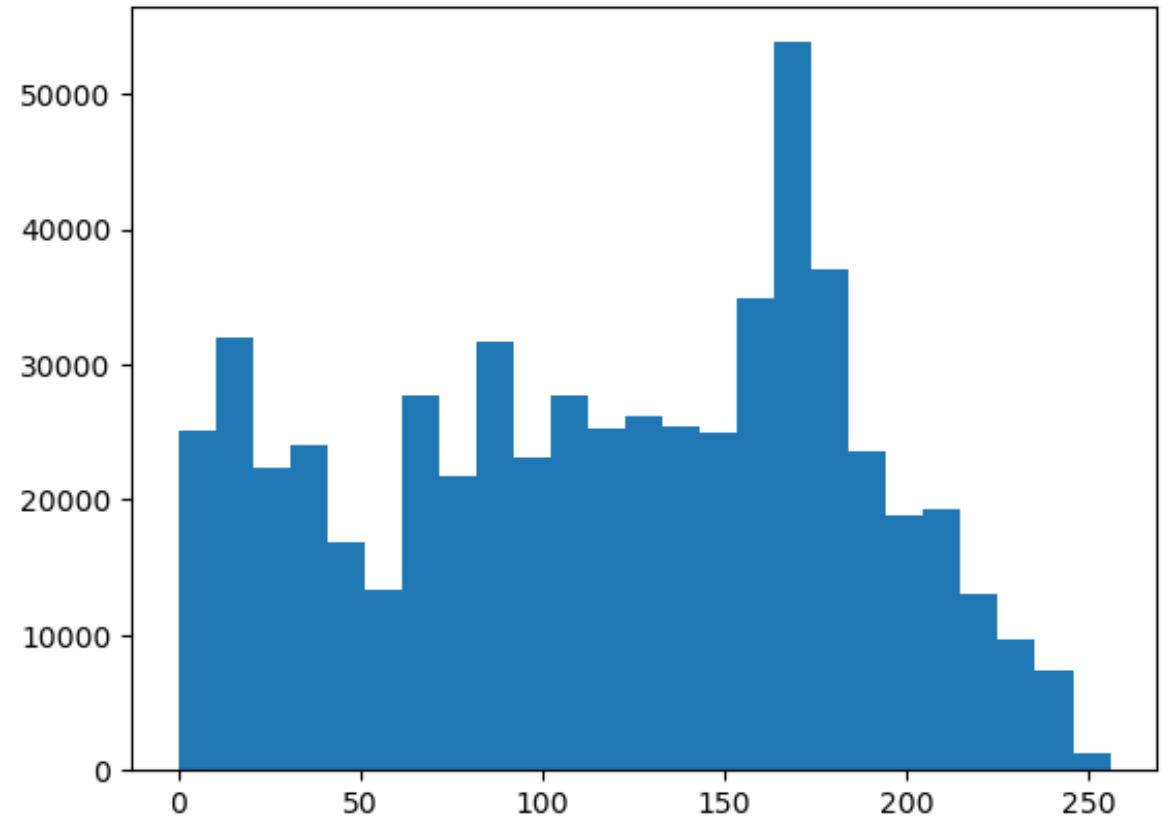
Customer Seniority Distribution

Long-term Deposit

Count of Channel_join (Ordered)
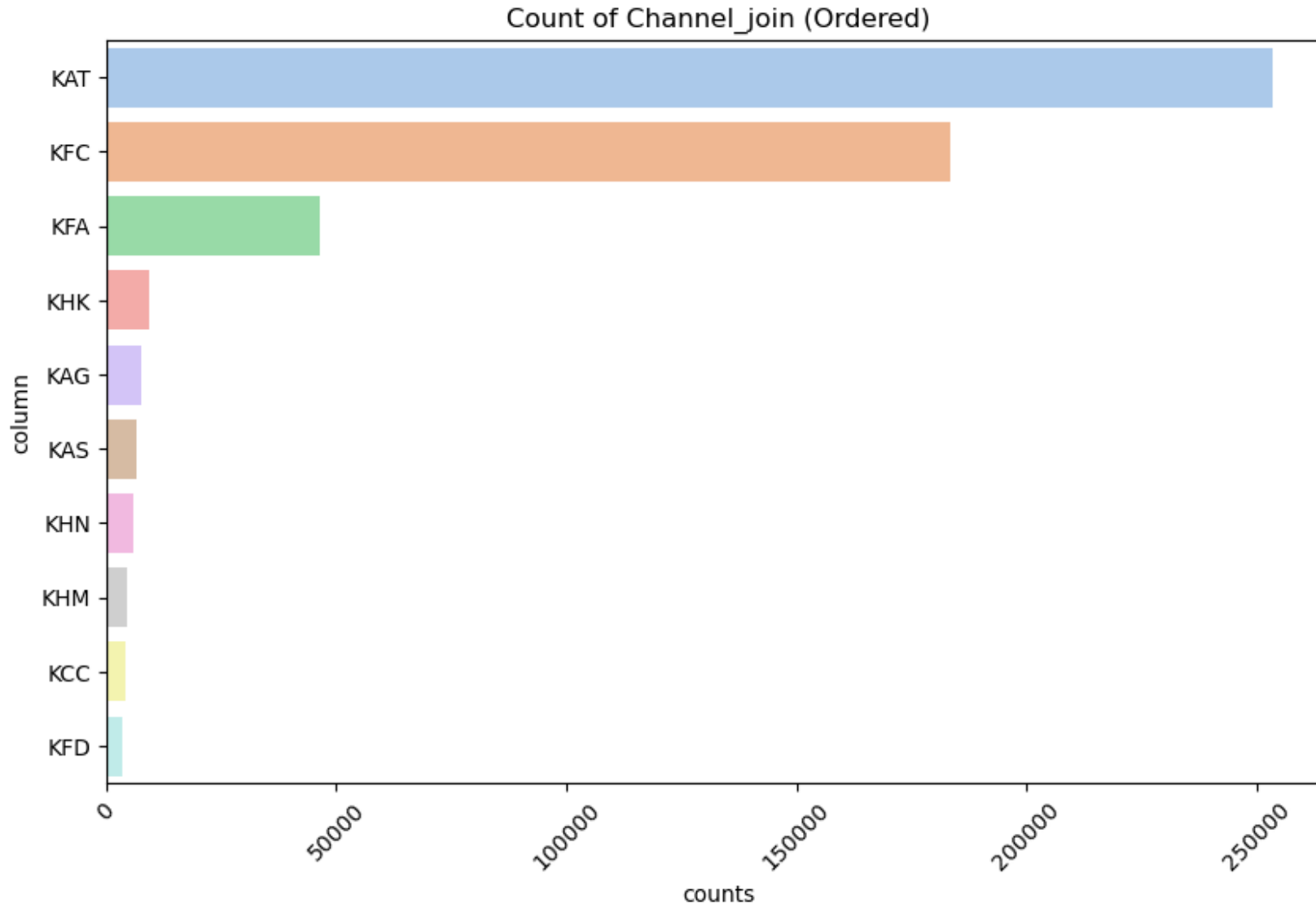
Customers from KAT, KFC and KFA should be our target

Long-term Deposit

Count of Province name (Ordered)
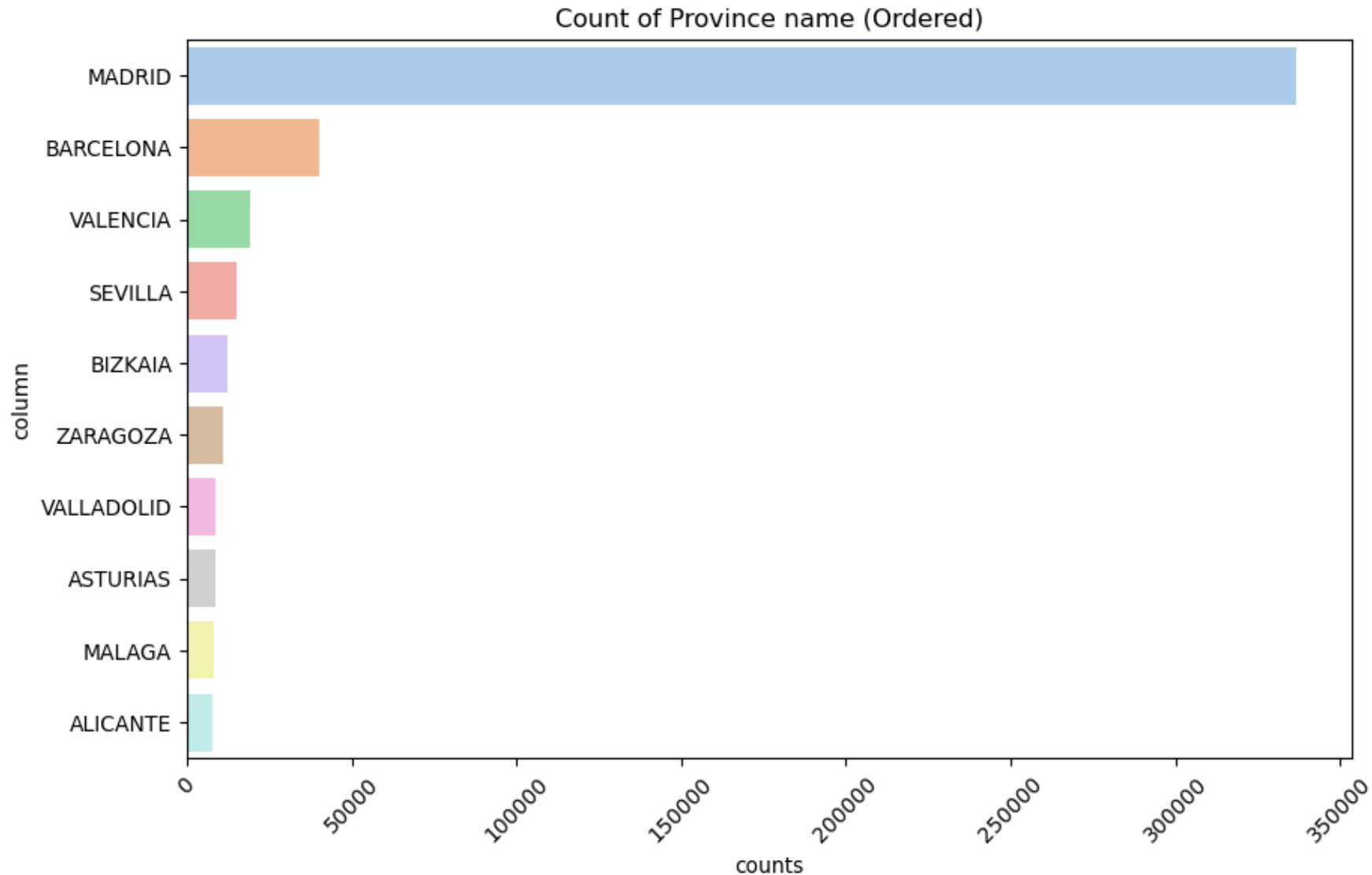
Customers from Madrid, Barcelona and Valencia should be our target
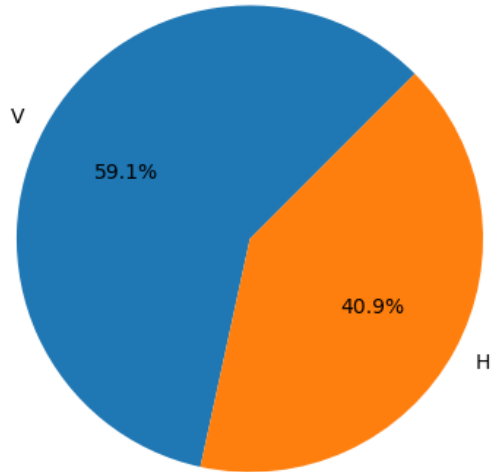
Payroll Account

age_group Distribution
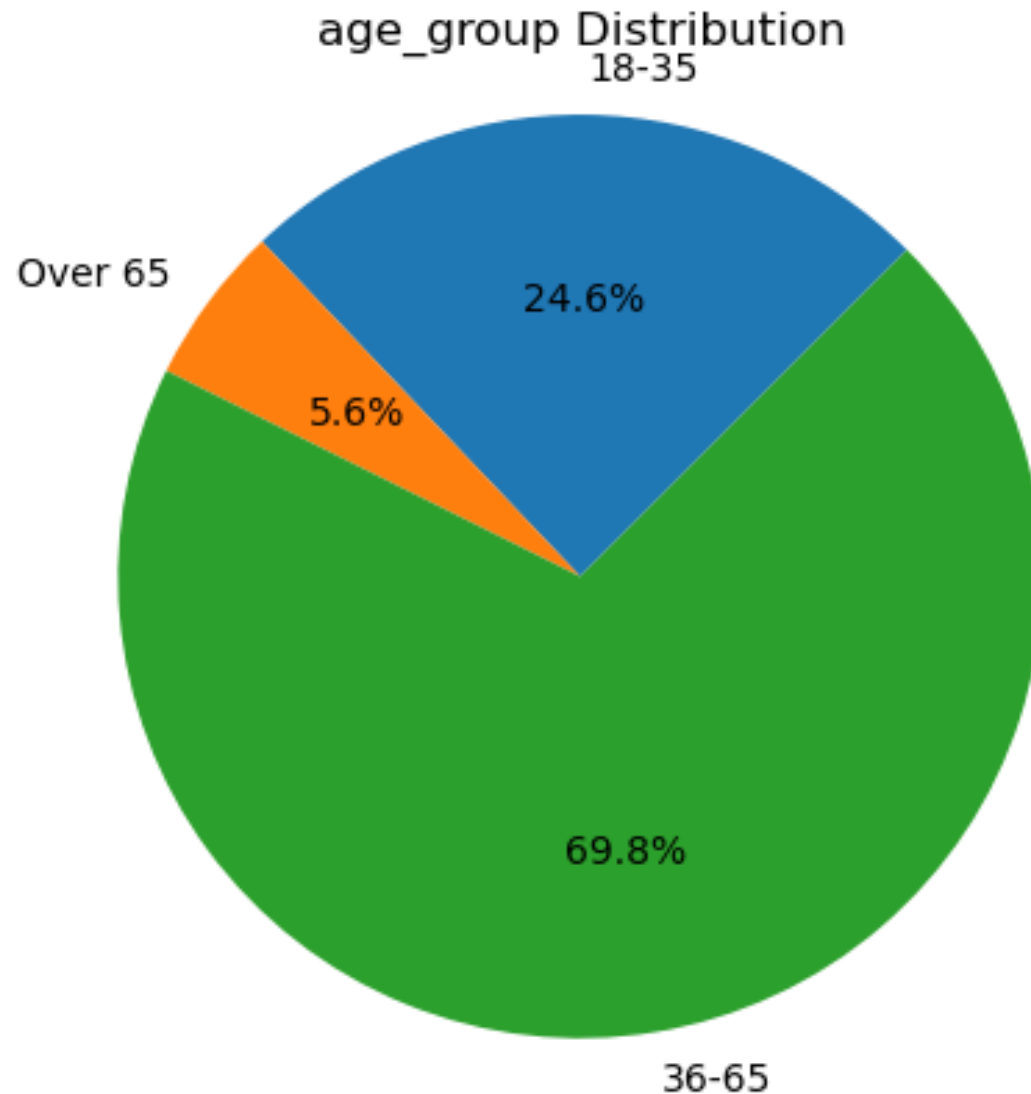
Age group 18-35, and 36-65 years old are the largest age group choosing long-term deposit
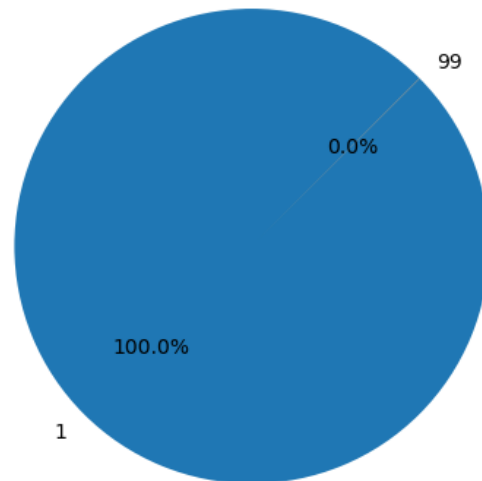
Payroll Account
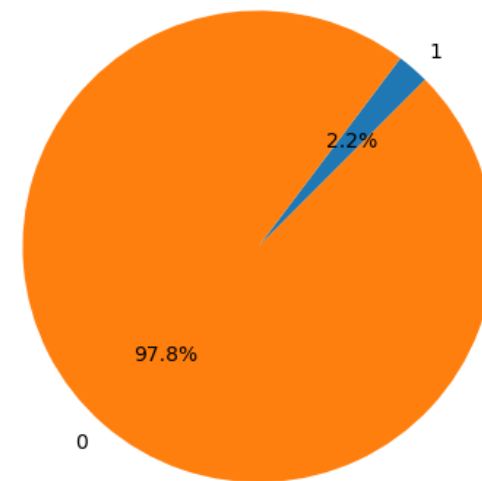
Payroll Account

Foreigner_index Distribution

S 4.4%

N 95.6%

Residence_index Distribution

N 0.1%

S 99.9%

Decreased_index Distribution

S 0.1%

N 99.9%

Active Distribution

0 1.3%

1 98.7%
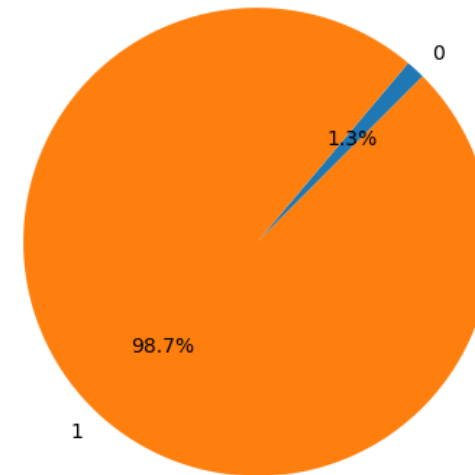
Payroll Account

segmento Distribution
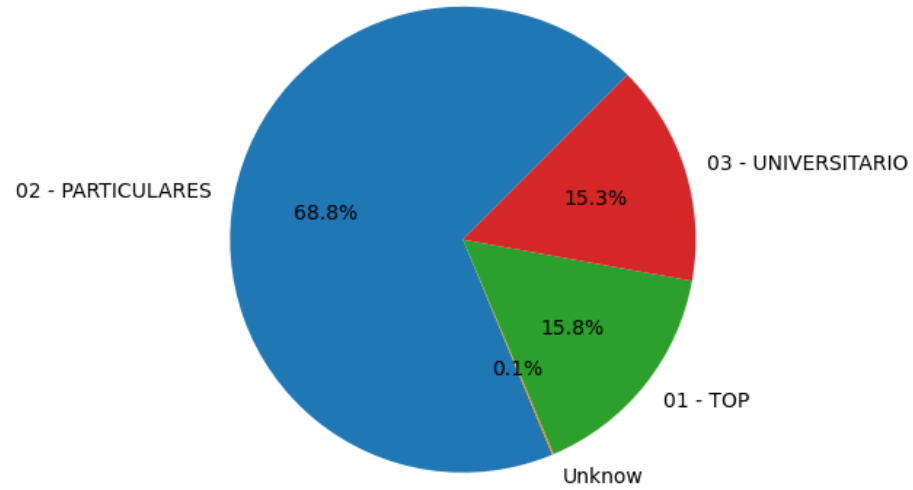
02 - PARTICULARES 68.8%
03 - UNIVERSITARIO 15.3%
15.8%
0.1%
01 - TOP
Unknow

Employee_index Distribution

0.0%
99.8%
N

MonthBeginCustomerType Distribution

0.0%
100.0%
1

MonthBeginCustomerRelation Distribution

R P
0.0% 0.0%
98.7%
A
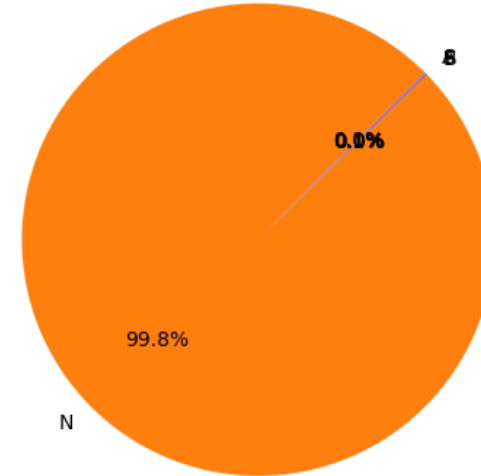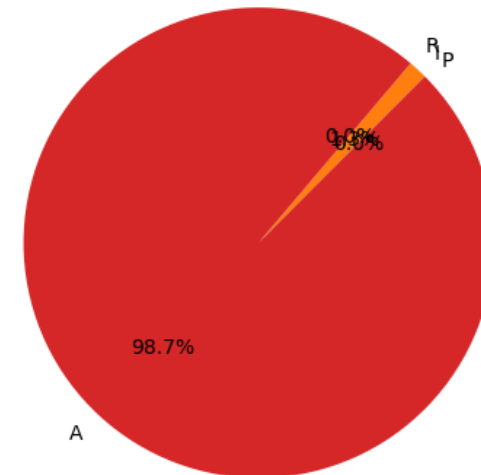
Payroll Account

Gross Income Distribution
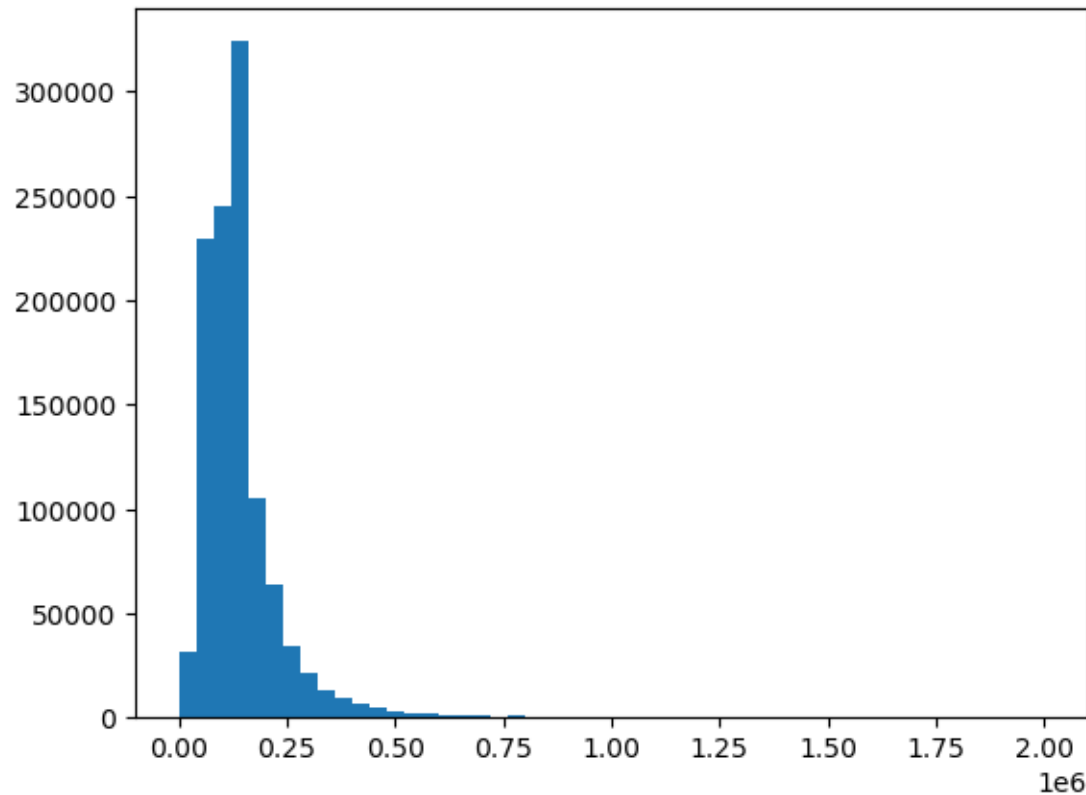
Customer Seniority Distribution

Payroll Account

Count of Channel_join (Ordered)
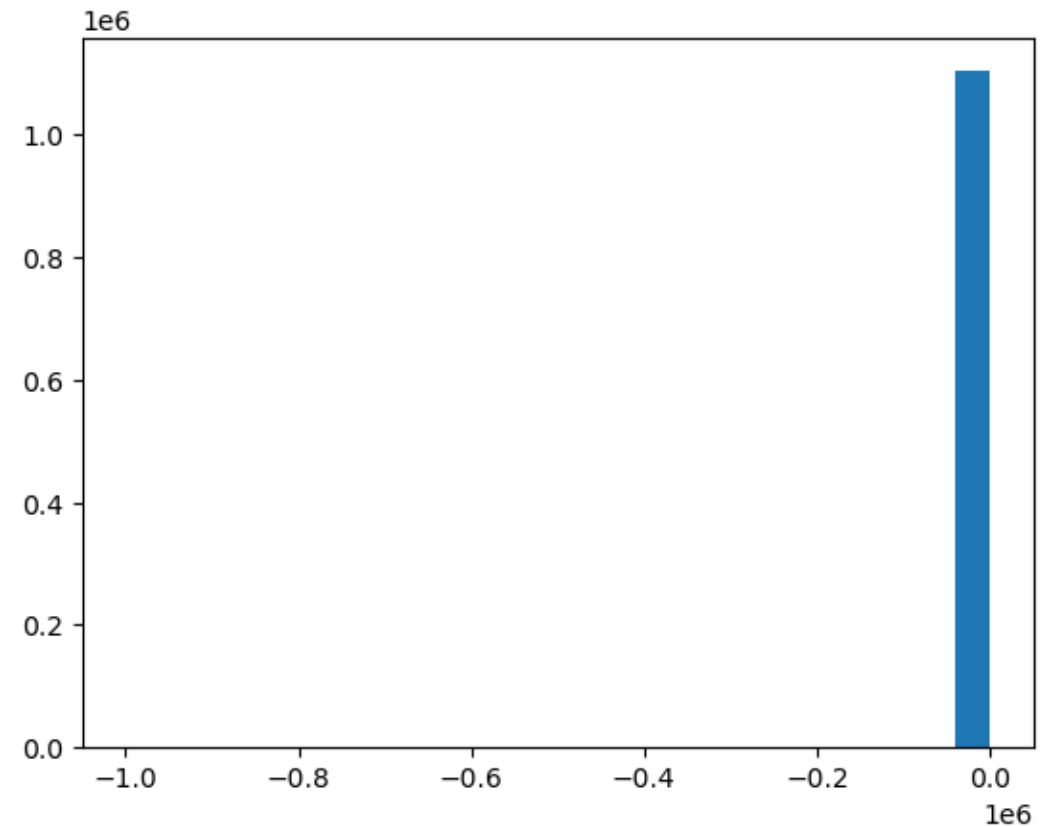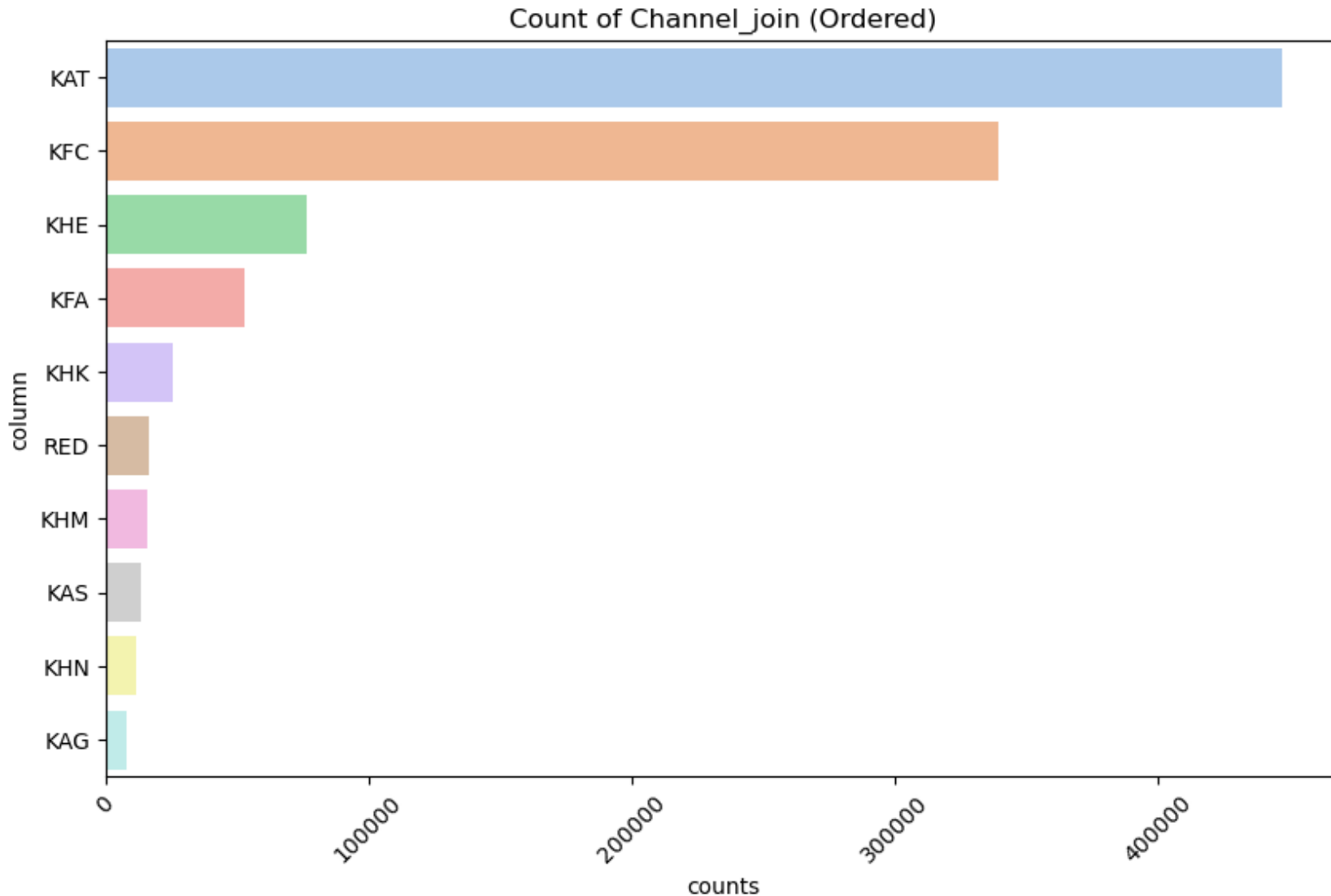
Customers from
KAT, KFC, KHE and KFA
should be our target

Payroll Account

Count of Province name (Ordered)
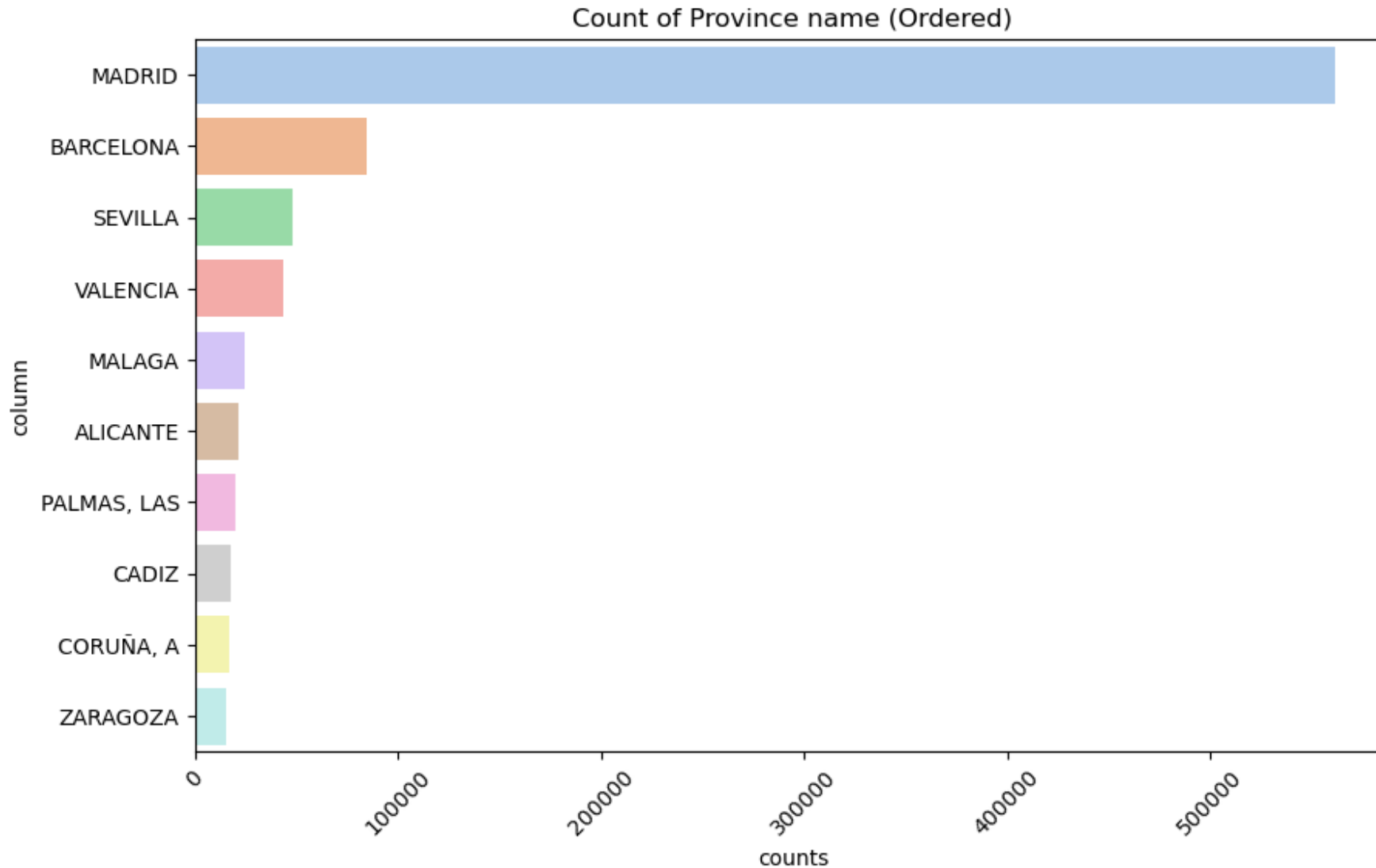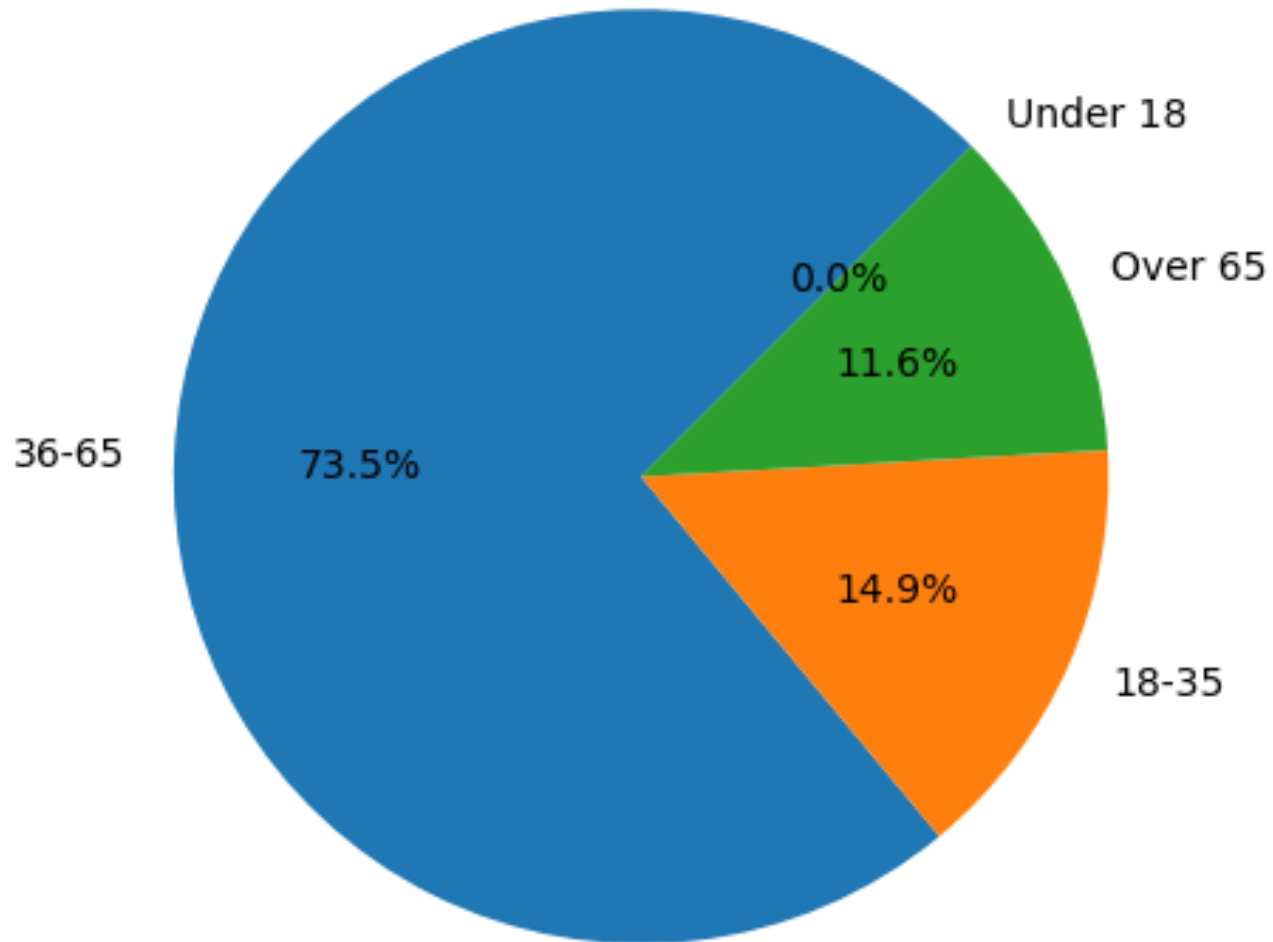
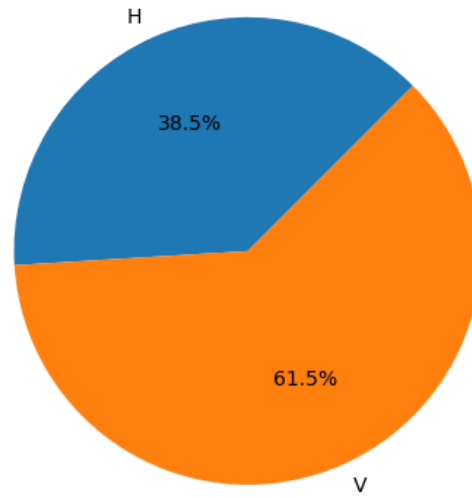Customers from Madrid, Barcelona, Sevilla, and Valencia should be our target
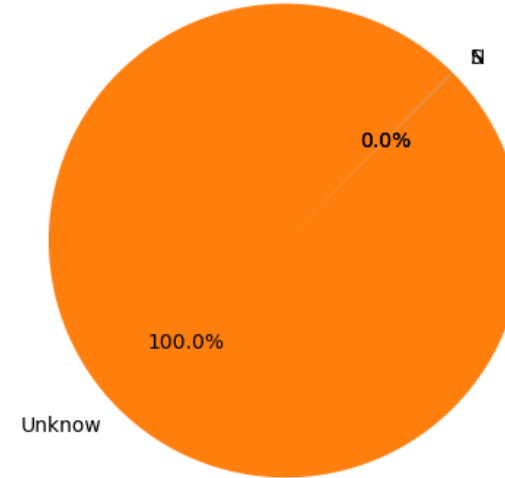
E-Account

age_group Distribution

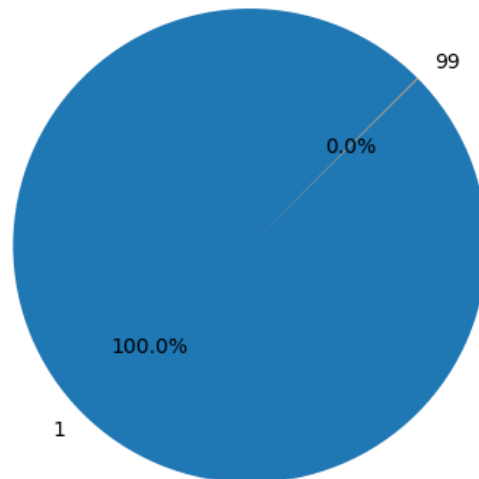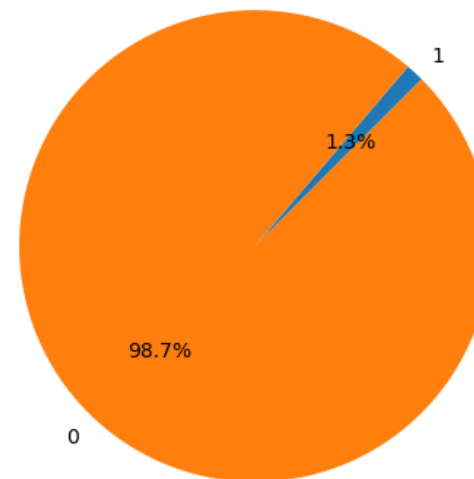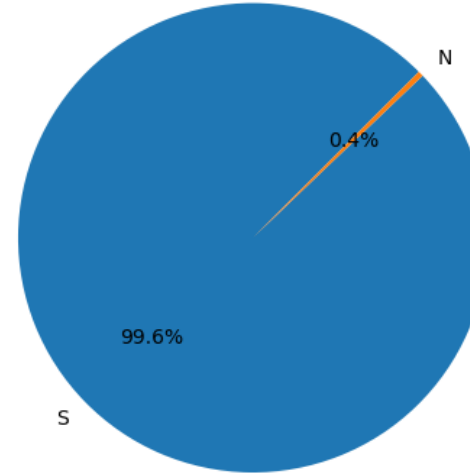Age group 36-65 years old are the largest age group choosing long-term deposit
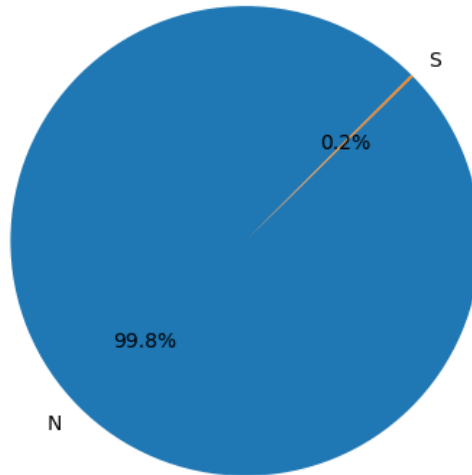
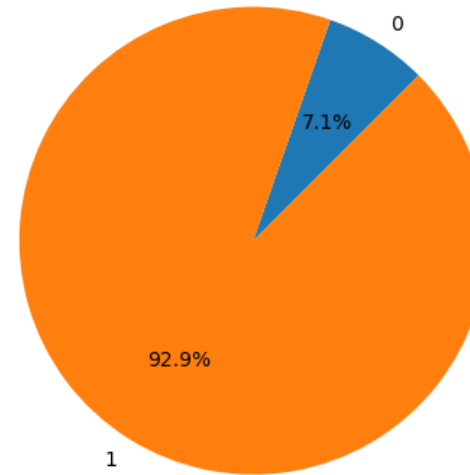# E-Account

E-Account

Foreigner_index Distribution
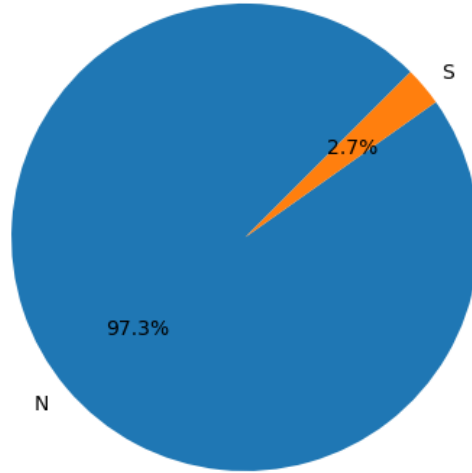
Residence_index Distribution

Decreased_index Distribution

Active Distribution

E-Account

segmento Distribution
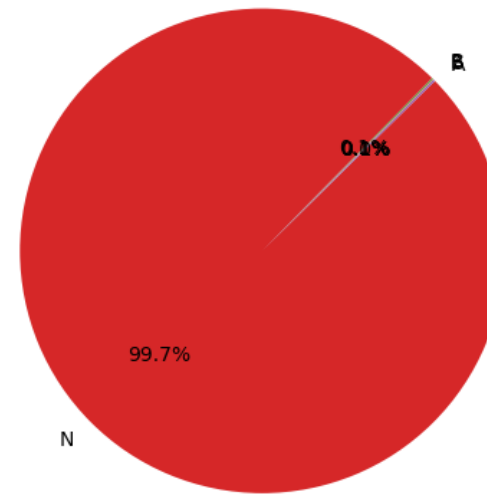
01 - TOP
03 - UNIVERSITARIO
Unknow
23.2%
6.9%
0.1%
69.8%
02 - PARTICULARES
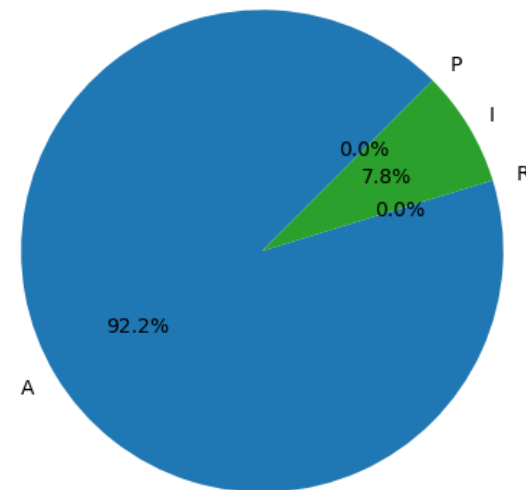
Employee_index Distribution

0.1%
99.7%
N

MonthBeginCustomerType Distribution

0.0%
100.0%
1

MonthBeginCustomerRelation Distribution

P
I
R
0.0%
7.8%
0.0%
92.2%
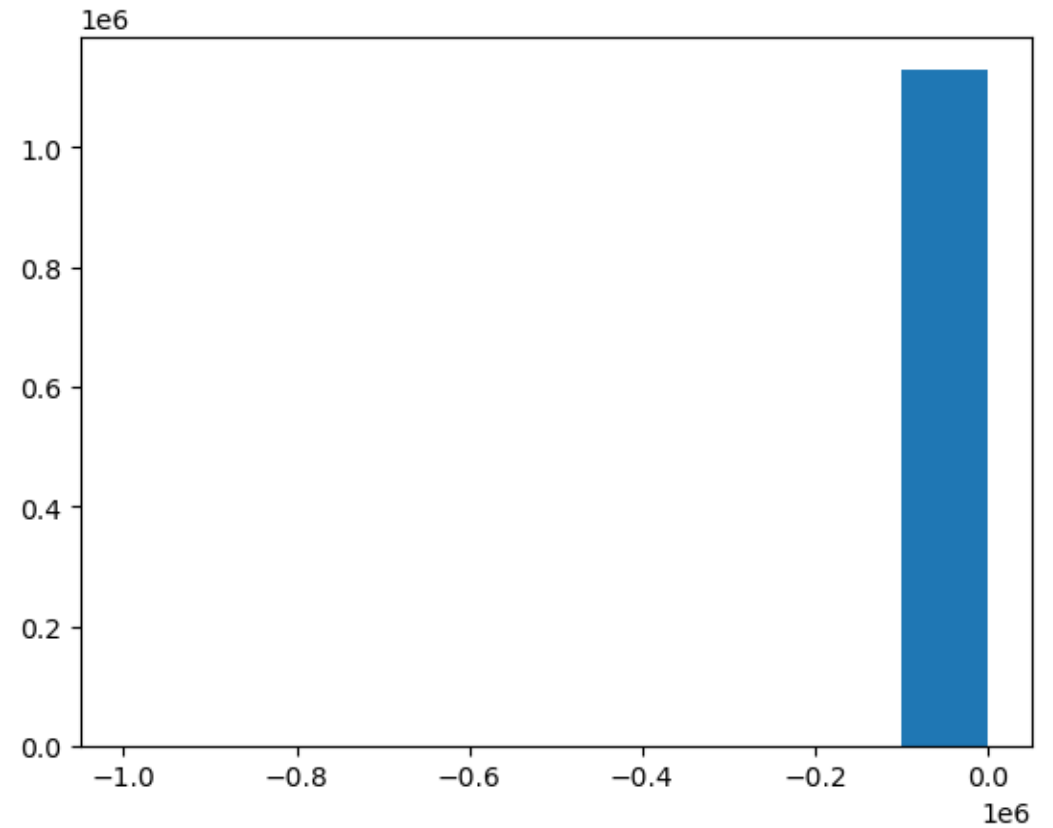A

E-Account

Gross Income Distribution

Customer Seniority Distribution

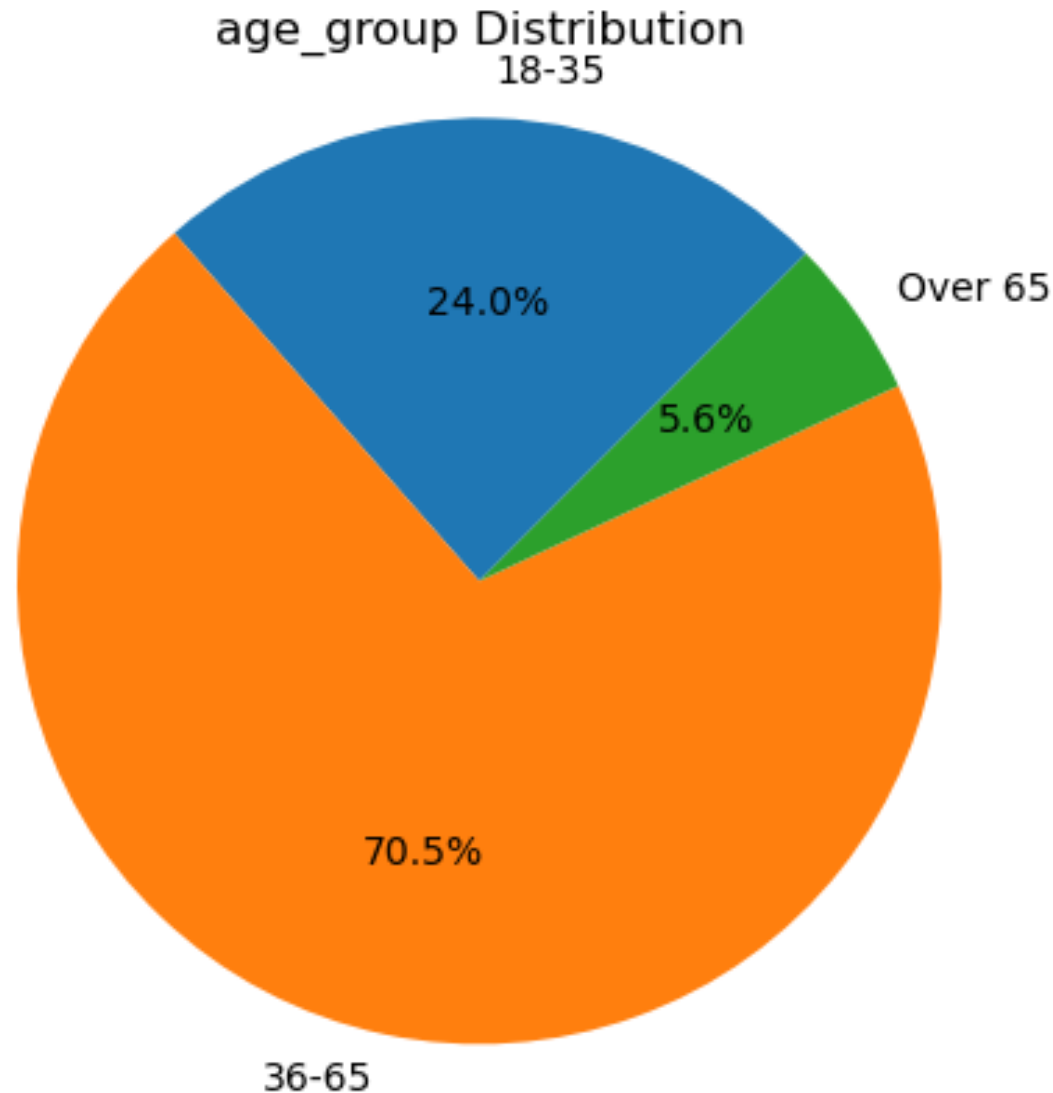# E-Account

Count of Channel_join (Ordered)

Customers from KAT, KFC and KFA should be our target

E-Account

Count of Province name (Ordered)

Customers from
Madrid, Barcelona,
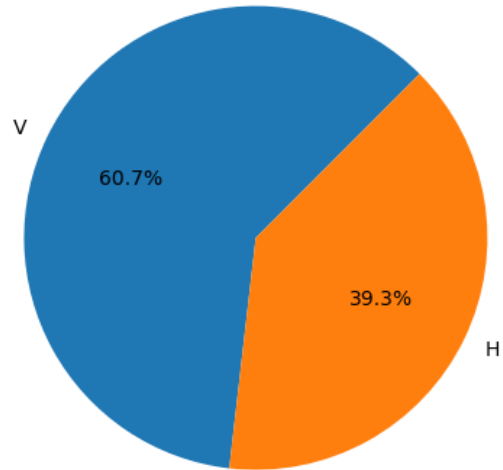and Valencia
should be our
target

Direct Debit

age_group Distribution

Age group 36-65 years old are the largest age group choosing long-term deposit

Direct Debit

Direct Debit

Direct Debit

Direct Debit

Gross Income Distribution
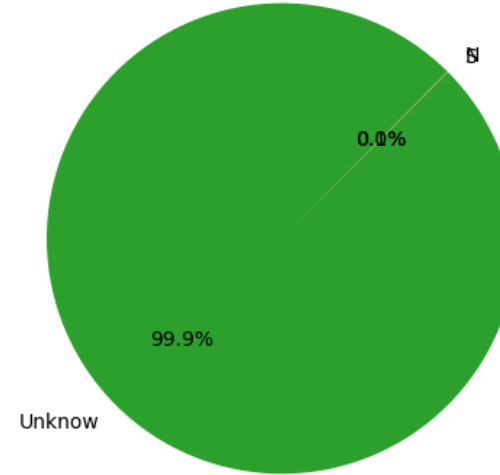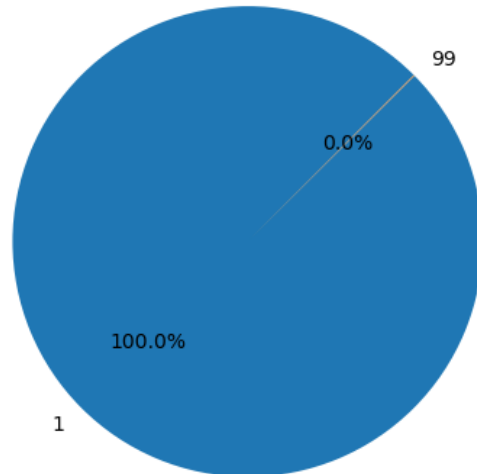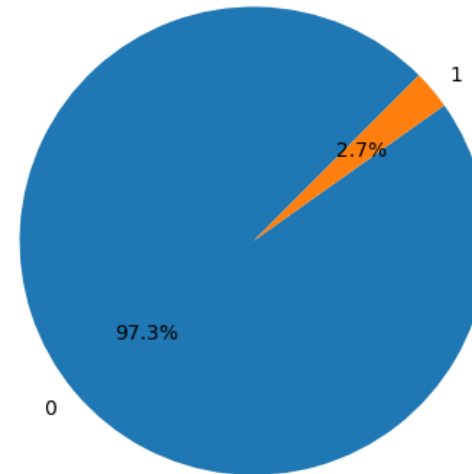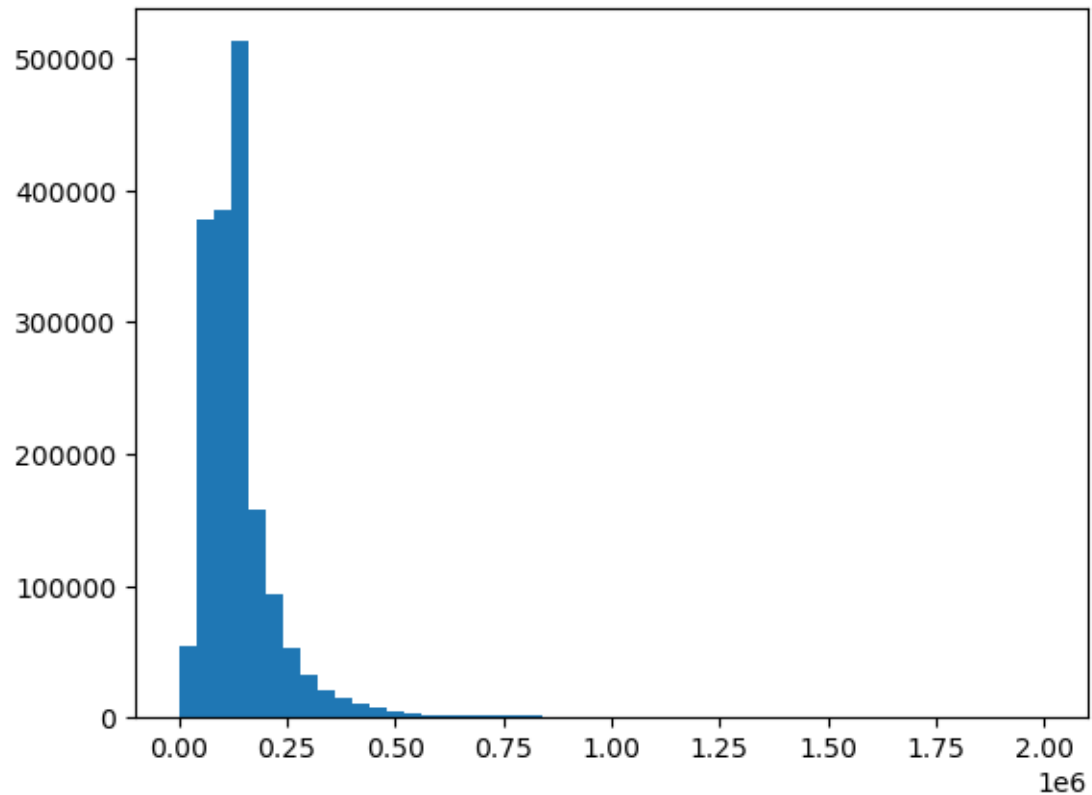
Customer Seniority Distribution

Direct Debit

Count of Channel_join (Ordered)
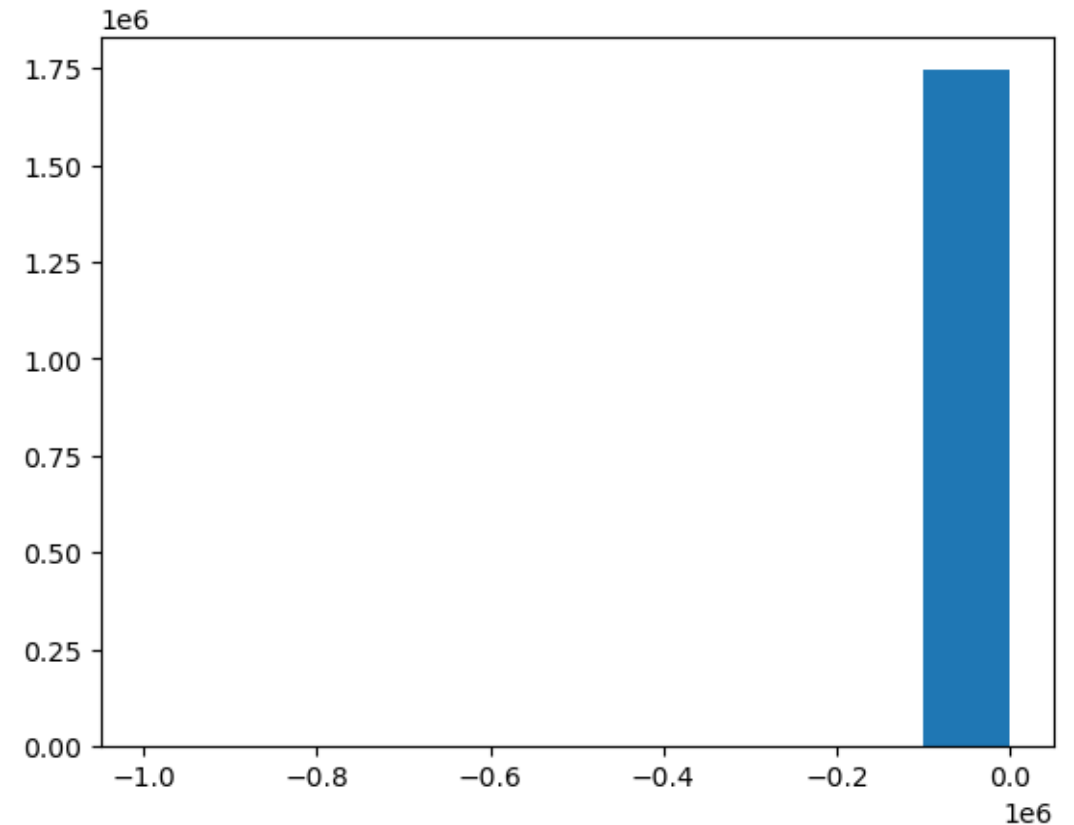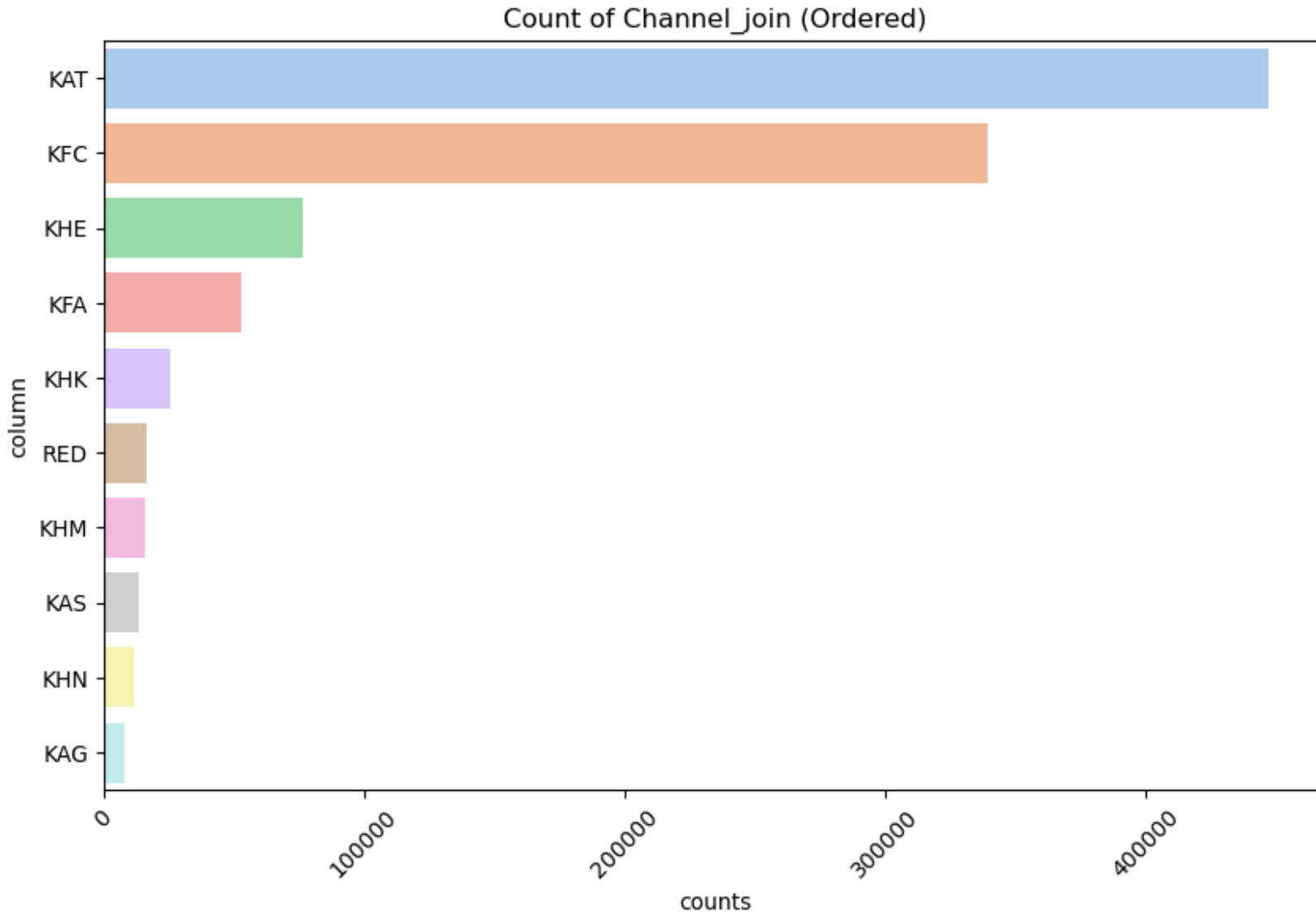
Customers from KAT, KFC and KHE should be our target

Direct Debit

Count of Province name (Ordered)
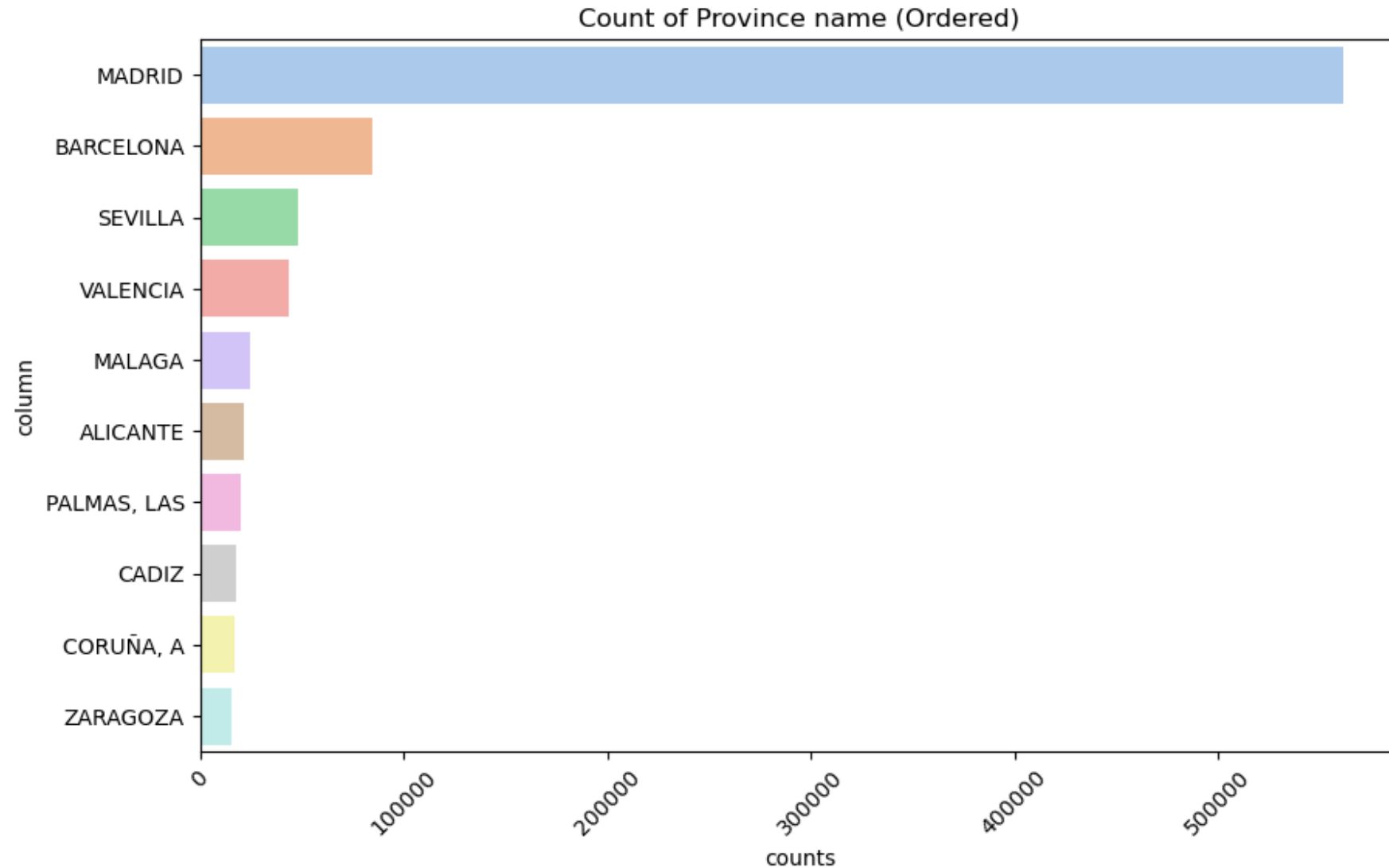
Customers from Madrid, Barcelona, and Sevilla should be our target

EDA Summary

The exploratory data analysis reveals valuable insights:

- Dominant products

- Product based Customer analysis

- Cross-selling insights on target customers

Thank You

Recommended Model: GBM

- **Robustness to Outliers and Irregularities**: GBM can handle noises and outliers the dataset without compromising its performance.

- **Feature Importance Analysis**: GBM provides insights into feature importance, helping us understand the most influential variables in our data.

- **Handling Mixed Data Type**: GBM can effectively handle mixed data types, including both numeric and categorical variables.

- **Efficiency on Large Datasets**: GBM implementations like XGBoost is designed to be memory-efficient and capable of handling large datasets with millions of observations.