

Data Glacier Data Scientist Internship

Batch:LISUM39

Week10: Deliverables

Project: Bank Customer Segmentation

Group name: Apple Analytics

Name: Madoka Fujii

Email address: mdkfji@gmail.com

Country: United States

Company: Omdena

Specialization: Data Analytics

Problem Description:

XYZ Bank plans to enhance its marketing campaign as Christmas offers for its customers. However, instead of offering the same deal to all customers as generic, the bank wants to provide personalized offers to specific customer groups to fit their preferences. Identifying customer categories manually would be inefficient and fail to uncover hidden patterns in the data that could inform better segmentation. To address this, the bank has sought the assistance of Apple Analytics. Additionally, the bank has specified that customer segmentation should result in no more than 5 groups to ensure the campaign's efficiency.

EDA(exploratory data analysis)

Residence

```
*****
Unique values in Residence_index are :
Residence_index
S    993046
N     6954
Name: count, dtype: int64
*****
```

-Residence index shows 994,046 customers are residence while 6,954 customers are not residence. That means 99% of customers are residence.

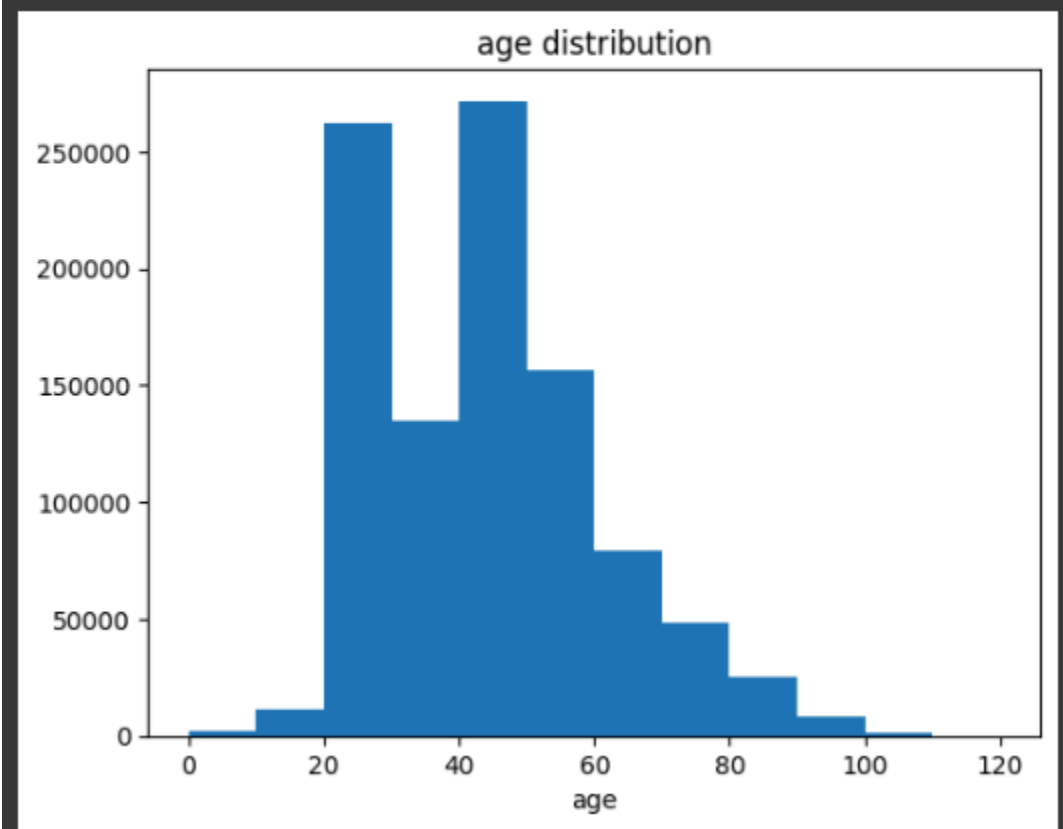
Foreigner

```
*****
Unique values in Foreigner index are :
Foreigner index
N    957110
S     42890
Name: count, dtype: int64
*****
```

-Foreigner index shows 42,890 customers are foreigner while 957,110 customers are not foreigner. That means 95% of customers are not foreigner.

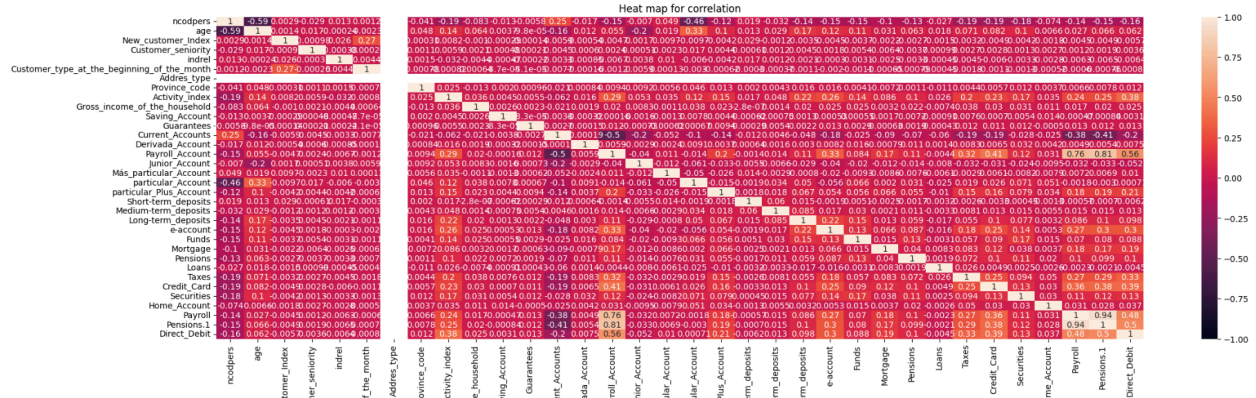
Age distribution

```
plt.hist(df['age'], bins = list(range(0,130,10)))
plt.xlabel('age')
plt.title('age distribution')
plt.show()
```



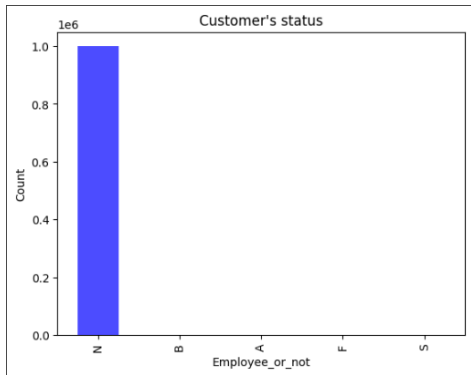
The age distribution of customers between 20 -60 years is shared most.

Correlation



In terms of the correlation of all of the variables, there is no significant relationship found.

Customer's status

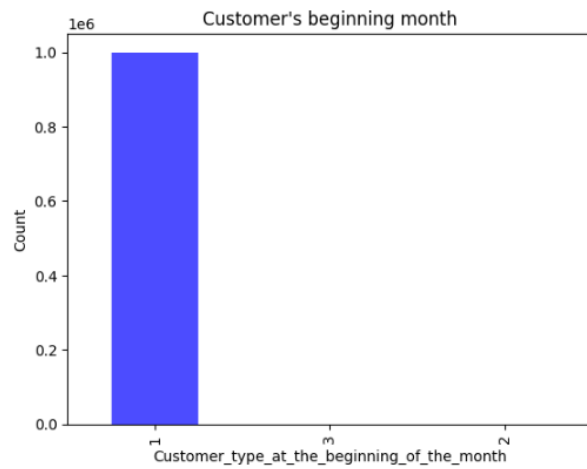


Employee index:

- A active
- B ex employed
- F filial
- N not employee
- P passive

Most of the customers are N as not employees

Customer's beginning month

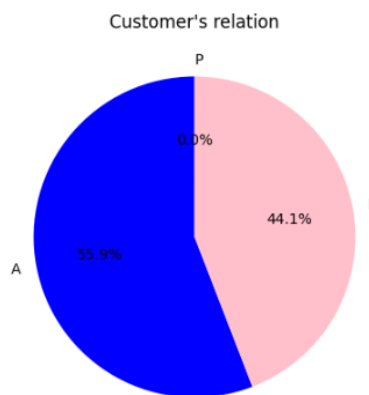


Customer type at the beginning of the month:

- 1 (First/Primary customer)
- 2 (co-owner)
- P (Potential)
- 3 (former primary)
- 4(former co-owner)

Most of the customers are 1 (First/Primary customer)

Customer's relation



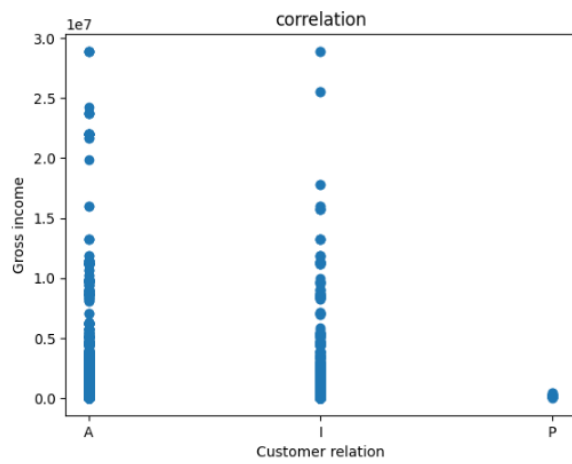
Customer relation type at the beginning of the month:

- A (active)
- I (inactive)
- P (former customer)

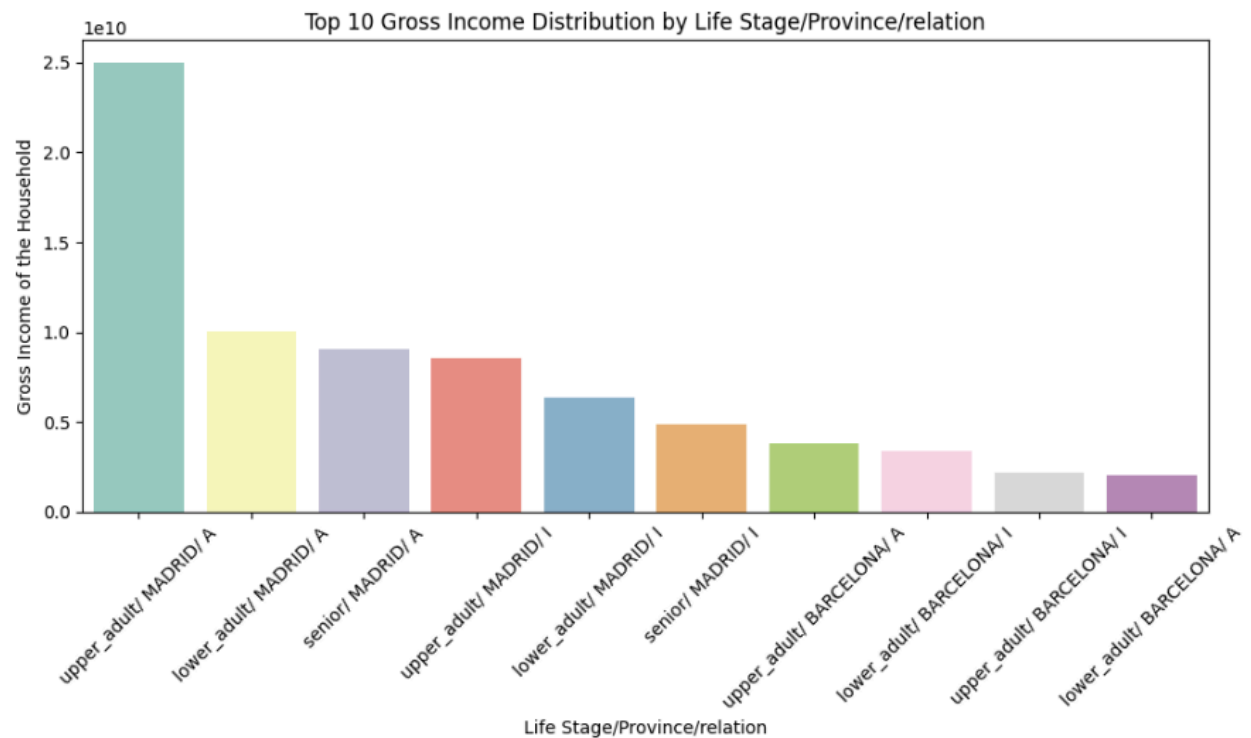
- R (Potential)

A (active) customers: 55.9%

I (inactive) customers: 44.1%

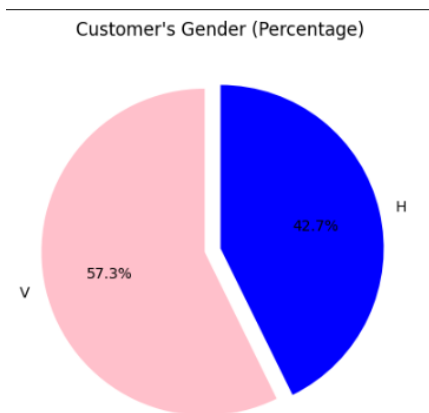


The correlation in terms of customer relations, there is no significant difference in gross income between Activa and Inactive.

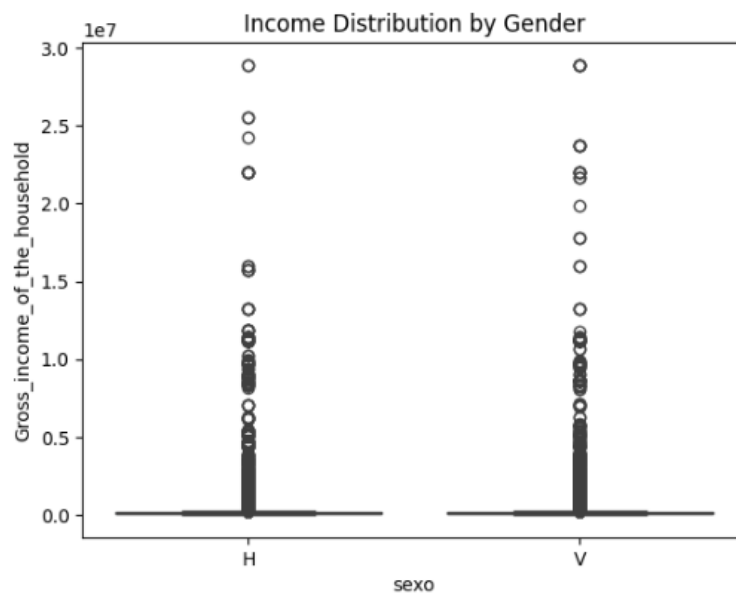


In the provinces of Madrid and Barcelona, there are mixing Active and Inactive customers.

Customer's gender

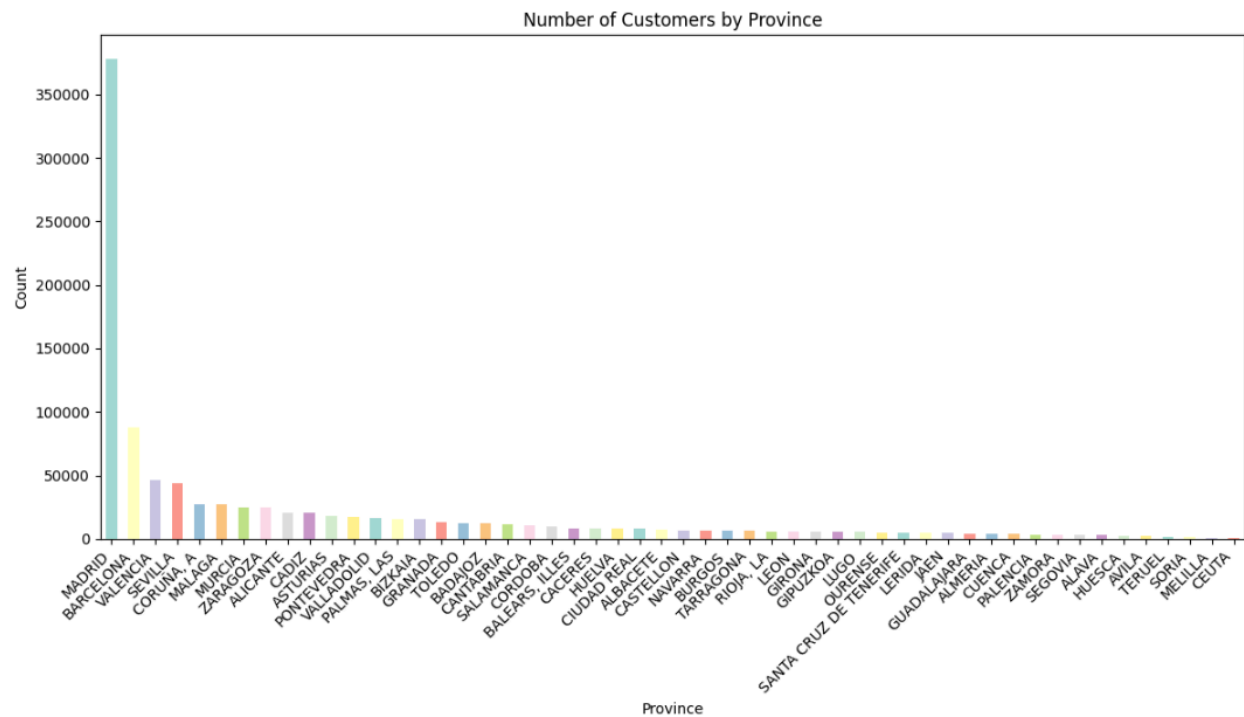


Female(V) is 57.3% which is relatively larger amount compared to male(H) 42.7%.



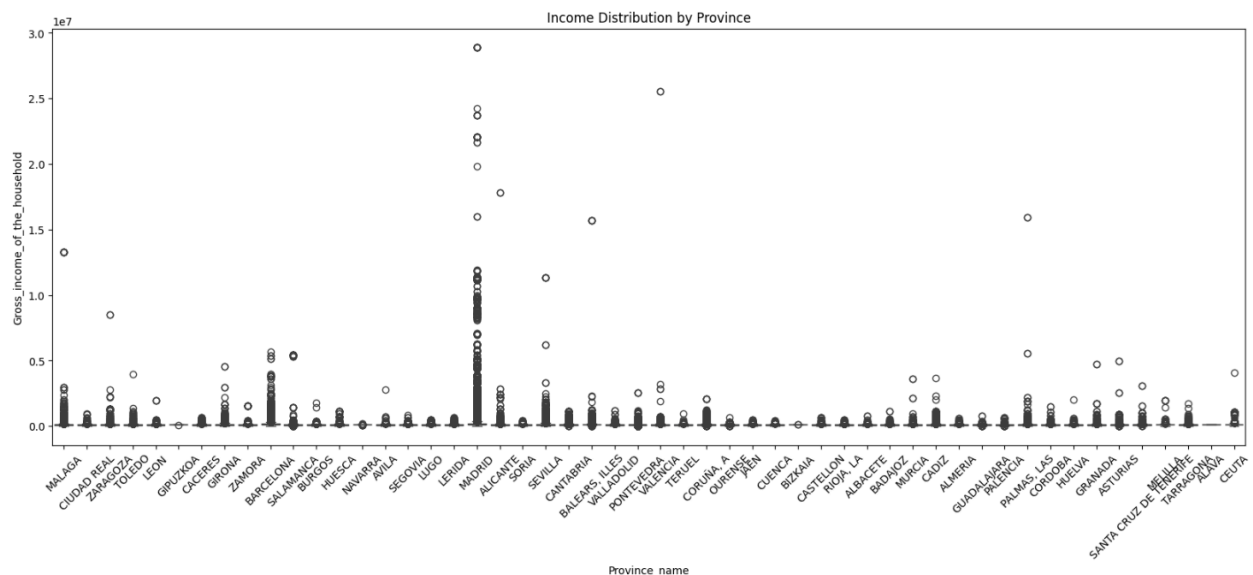
There are no significant differences in the distribution of gross income between Male(H) and Female(V). However, it seems slightly Female has more outliers concentrated above the median.

Number of Customers by Province

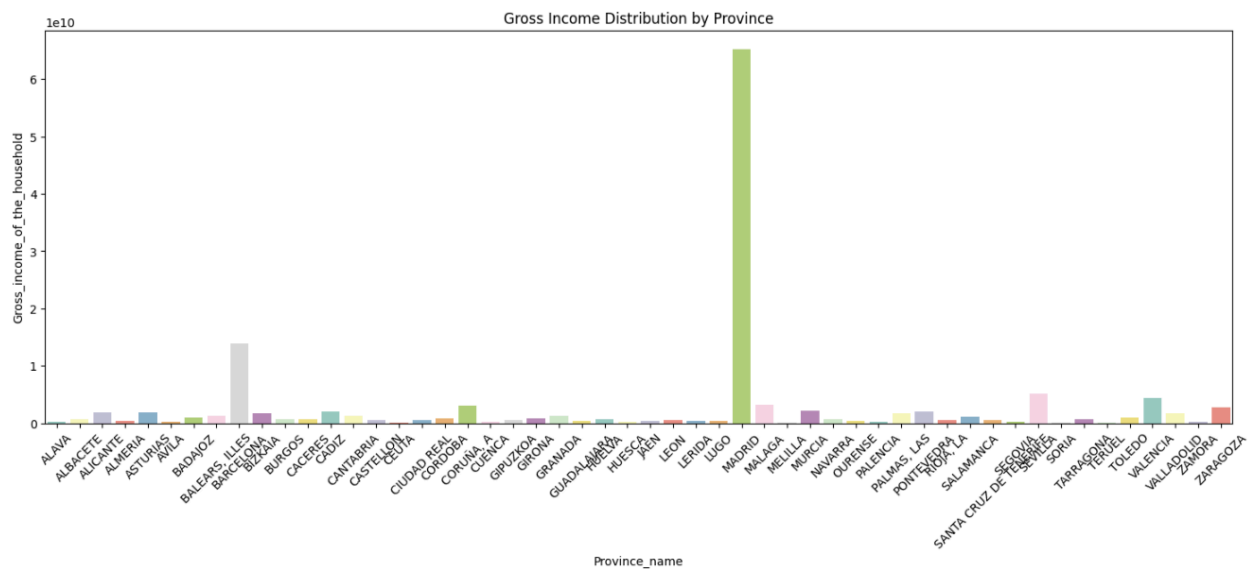


Madrid is the majority of province customers existing. Other provinces are significantly lower compared to Madrid.

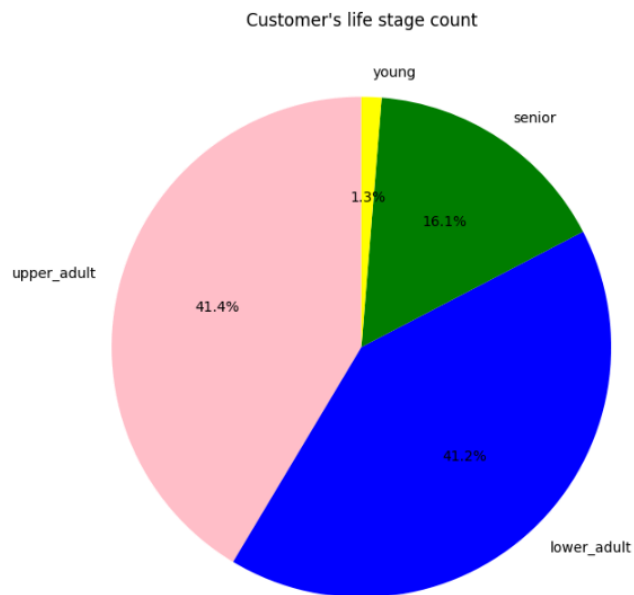
Income Distribution by Province



Madrid has the highest population of customers and the highest gross income household. Especially, a lot of outliers exist in Madrid.



Customer's life stage (grouped by age)



The life stage group is binned as

young: less than 20 years old

lower_adult: more than 20 and less than 40 years old

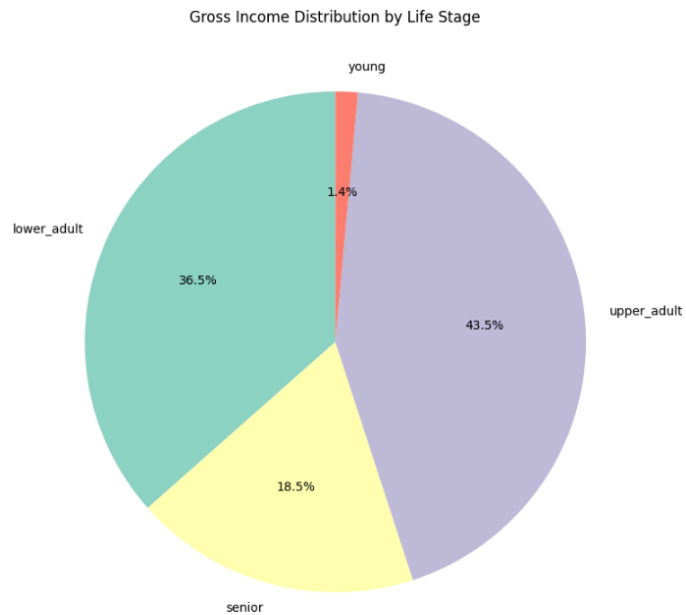
upper_adult: more than 40 and less than 60 years old

senior: more than 60 years old

lower_adult and upper_adult have dominated the share as 82.6%.

senior shares 16.1% and young shared only 1.3%.

Gross income distribution by lifestage



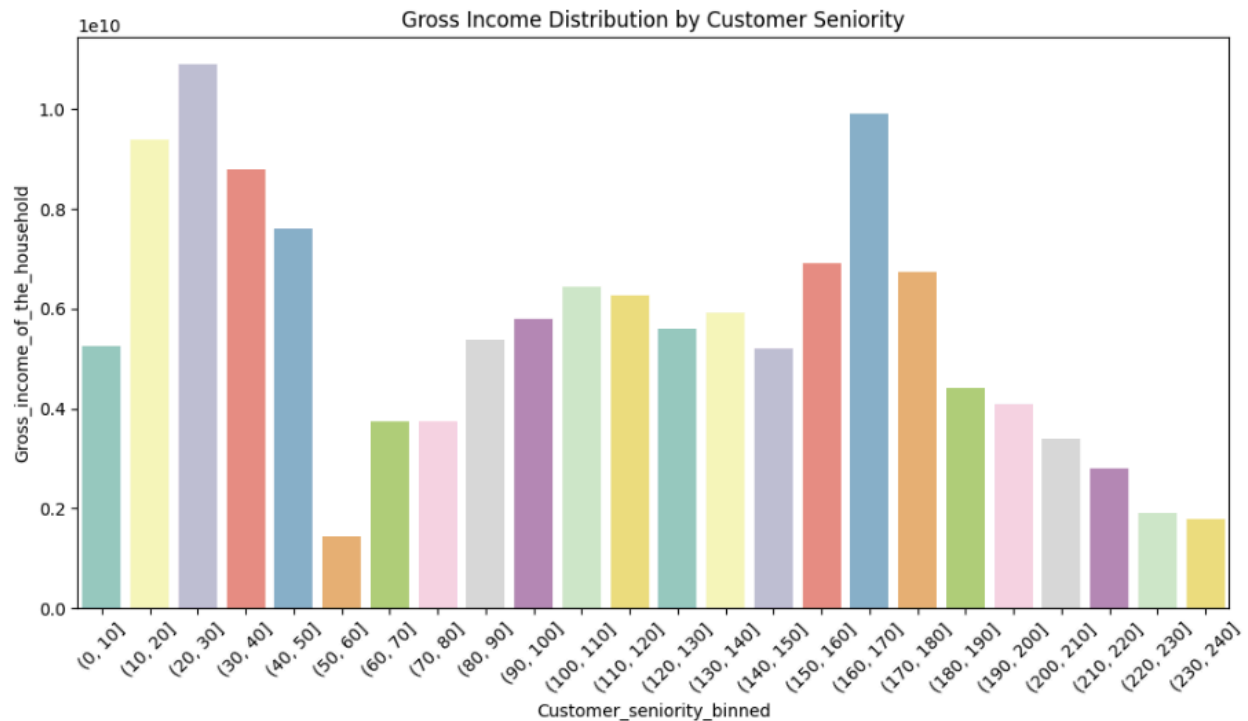
Upper adult: 43.5%

lower adult: 36.5%

senior: 18.5%

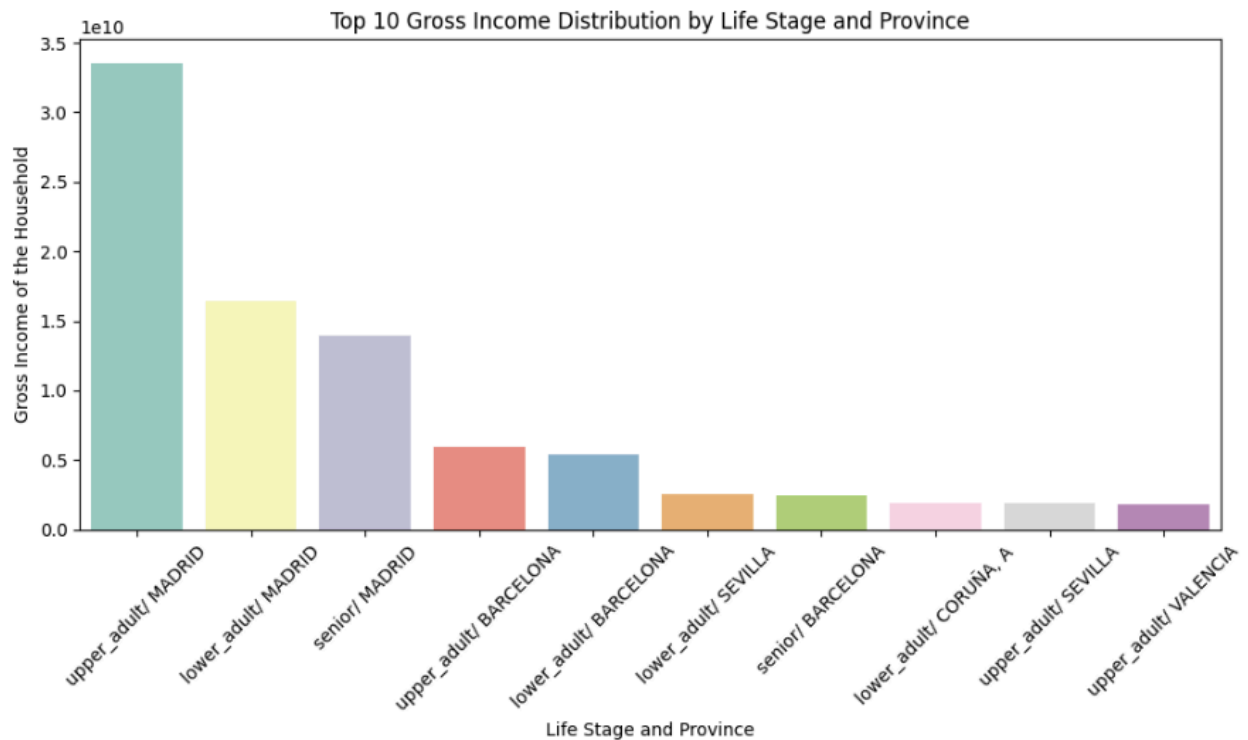
young: 1.4%

Gross income distribution by customer seniority

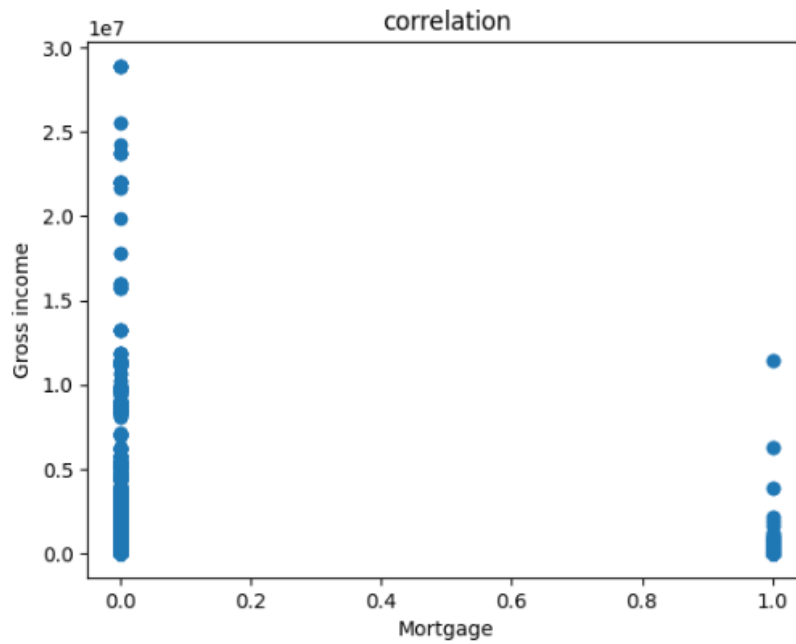
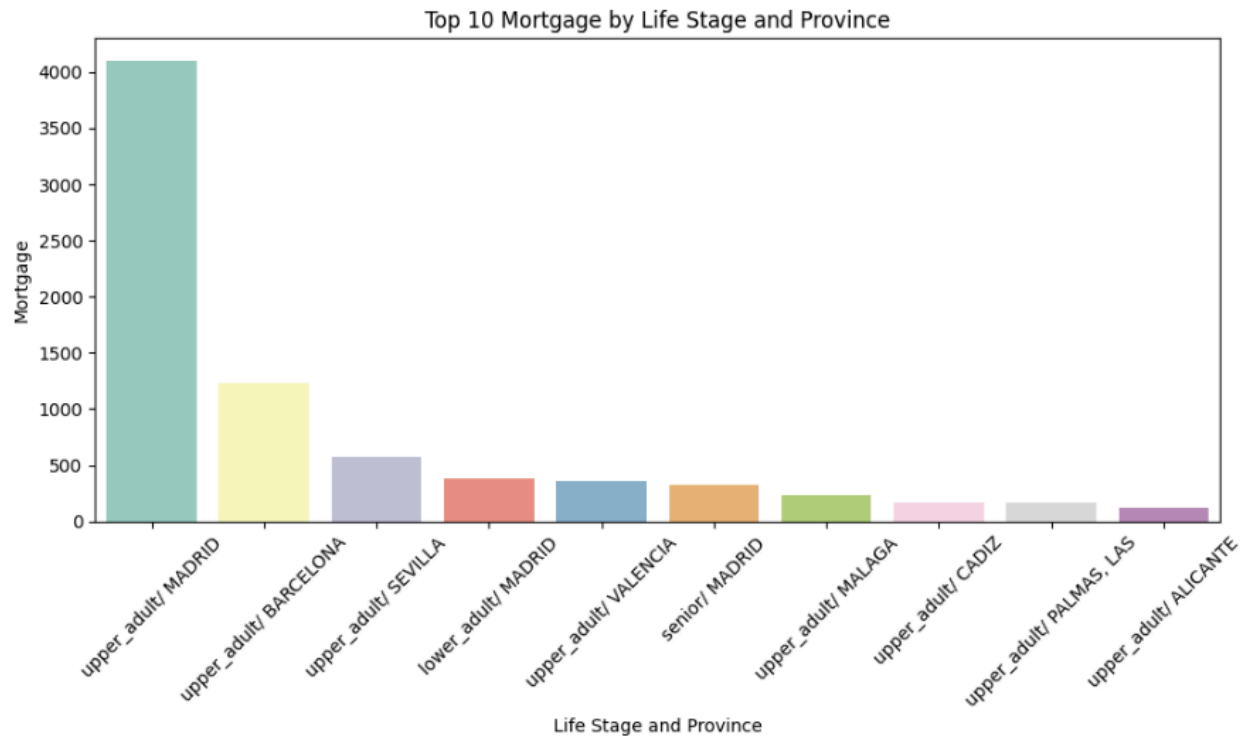


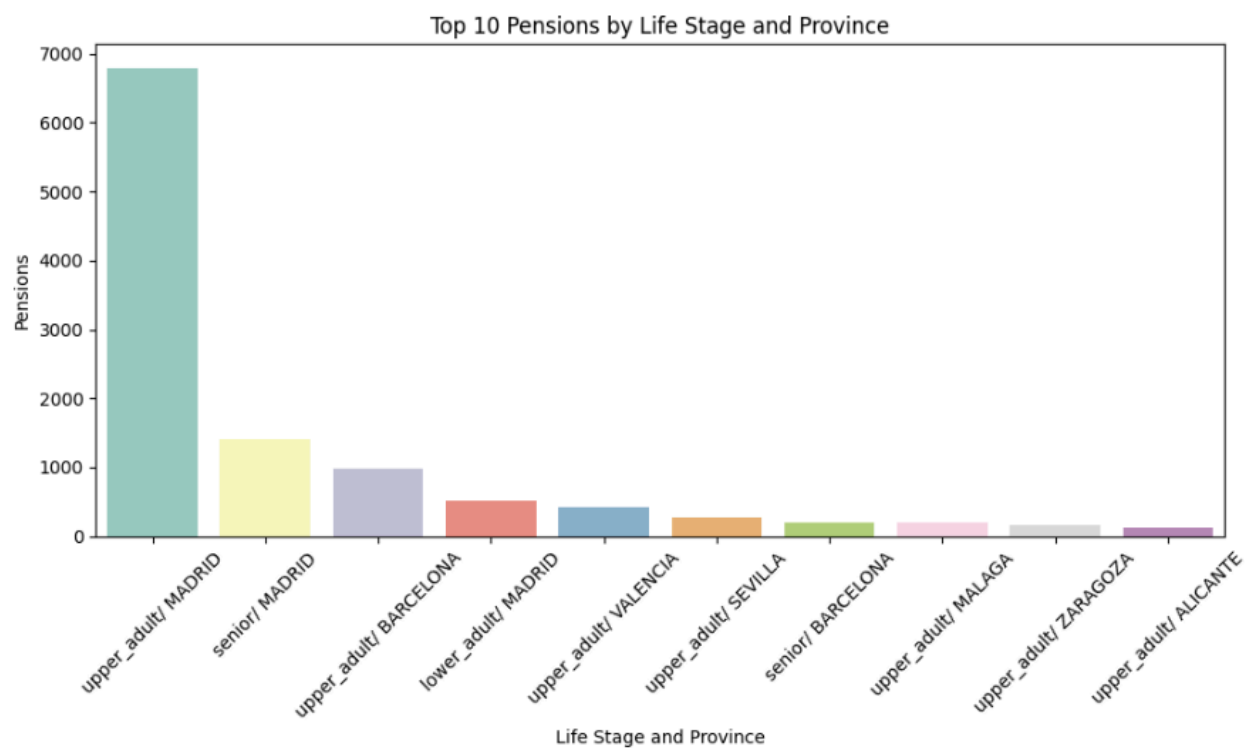
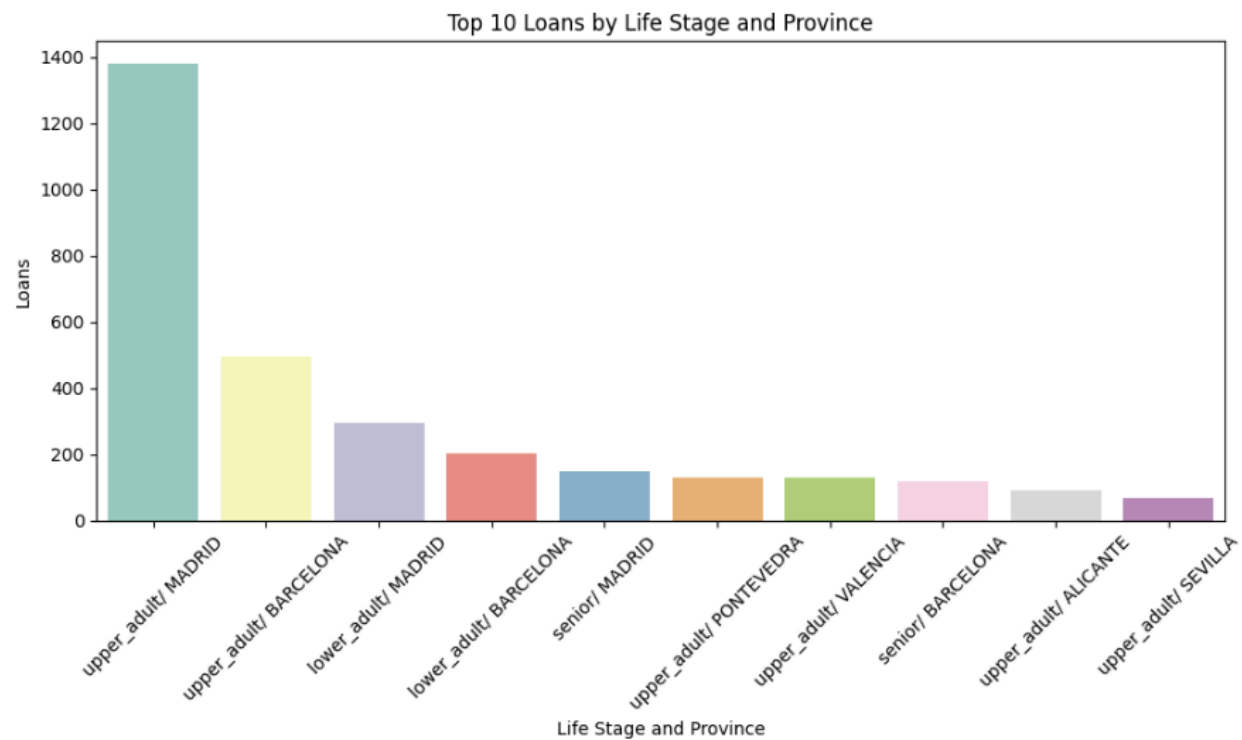
The highest group of customer seniority is 20-30 months. The next highest group is 160-170 months. The lowest group is 50-60 months. It may be some restrictions existing in this term such as some promotions being ended. From 180-240 months(15-20 years), there is a trend to decrease the number of seniority that indicates this bank cannot retain customers for the long term.

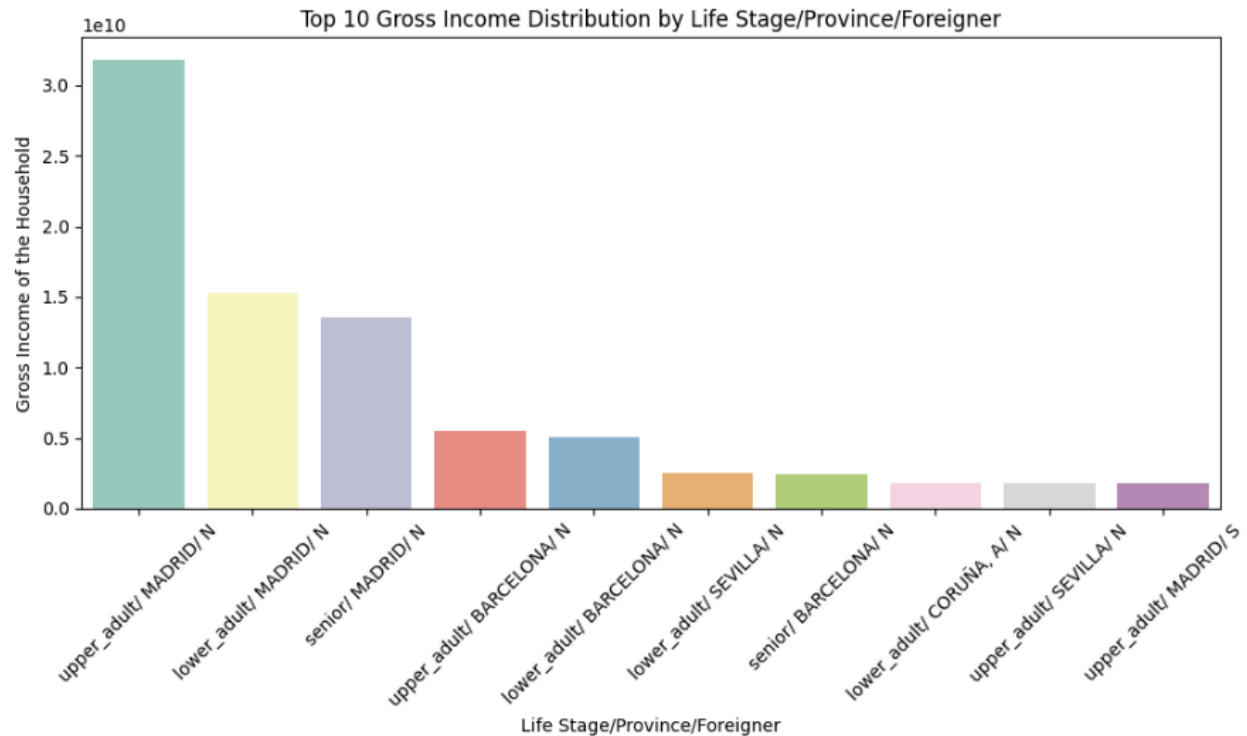
Top 10 Gross income distribution by lifestage and province



Regardless of the life stage group, customers who live in MADRID have the highest income. The customers who live in BARCELONA seem 2nd highest income group. But the life stage of the young is the minority in the any of provinces. There are no young life stage groups in this observation.







The majority of high-income customers are from Residence (no foreigners) while upper_adult/MADRID/S shows they are foreigners.

Final Recommendation

The number of male customers is relatively lower than that of female customers, presenting an opportunity to increase the male customer base.

Currently, inactive customers account for 44.1% of the total, while active customers make up 55.9%. To activate the existing inactive customer base, targeted strategies can be implemented. Notably, a significant proportion of inactive customers resides in high-income regions such as Madrid and Barcelona. By identifying and engaging these customers, the bank could unlock profitable growth opportunities.

The share of foreign customers is relatively small. However, there is a notable group of upper adult foreign residents in Madrid who may belong to high-income brackets. While a more detailed analysis is required, there is potential to expand the foreign high-income customer segment.

In terms of customer seniority, the largest group falls within 20–30 months, followed by 160–170 months. Conversely, the smallest group is within the 50–60 month range. This decline in customer numbers during this period may indicate underlying issues, such as the expiration of certain promotions. Investigating the reasons behind this trend could help address customer attrition and improve retention.

Beyond 180–240 months (15–20 years), there is a noticeable decline in customer seniority, suggesting the bank faces challenges in retaining customers long-term. While promotions and campaigns are valuable, a deeper evaluation of whether the bank’s services are designed to ensure long-term customer satisfaction could yield better results.

Young adults and senior customers represent key demographics for long-term strategic planning. As banking and financial institutions are integral to providing stability and security in people's lives, offering discounted promotions or affordable products tailored to these groups can enhance the bank’s reputation, foster trust, and support sustainable growth.

Project life cycle along with deadline:

Project weeks	Deadline	Lifecycle
Week7	Dec 19, 2024	Problem statement, Pre-process
Week8	Dec 26, 2024	Data process, understanding
Week9	Jan 02, 2025	Data Cleaning, Merge, Review
Week10	Jan 09, 2025	EDA, Final recommendation
Week11	Jan 16, 2025	EDA presentation for business users
Week12	Jan 23, 2025	Model Selection and Model Building/Dashboard
Week13	Jan 30, 2025	Final Project Report and Code

Tabular data details: cust_seg.csv.zip:

Total number of observations	1000000
Total number of files	1
Total number of features	48
Base format of the file	csv.zip
Size of the data	19MB