Data Glacier

Your Deep Learning Partner

Customer Segmentation Project

Virtual Internship

30-Aug-2022

Group Information

Group Name: M.A.S

Specialization: Data Science

Submitted to: Data Glacier canvas platform

Internship Batch: LISUM10: 30

Group Members	Three members		
Name	**Email**	**Country**	**Collage/Company**
Moath Mohammed Bin Musallam	moathmusallam@gmail.com	Saudi Arabia	University of East Anglia
Andrew Kojo Mensah-Onumah	kojoakmo@gmail.com	Ghana	Data Glacier
Shaimaa Saleh Obad Al-khawlani	s.khawlany@gmail.com	Yemen	Data Glacier

Link of the EDA file

https://github.com/kojomensahonums/Customer-segmentation-with-Data-Glacier/blob/master/Week10-EDA/Customer-Segmentation-Project-EDA.ipynb

Problem description

Most banks around the world have variant large customer base with different income levels, ages, characteristics, values and lifestyles. XYZ bank wants to increase the production and the satisfactions of all customers categories by roll out Christmas offers to their customers. But Bank does not want to roll out same offer to all customers instead they want to roll out personalized offer to particular set of customers. If they manually start understanding the category of customer then this will be not efficient and also, they will not be able to uncover the hidden pattern in the data (pattern which group certain kind of customer in one category).

Objectives:

To increase the production and the satisfactions of all customers categories by roll out Christmas offers to their customers.

Points to go through

- Total data count :1000000

The main goal is the bank's customers and product from Different Aspects:

How many customers the bank have per (Country /Province / Gender /Age /
Channel used by the customer/ Customer type)?

What is the Total Gross income of the household per (Country/ Province/ Residence index/
Foreigner index/ Deceased index/ Customer type)?

What is the most used products in general?
What is the most channel used by the customer?
What bank products are used by people of different age brackets?
What are the popular products per city?
What bank products are used by people of different income brackets?
What are the least and most used products?
What are the popular products by gender?

Customer age distribution

- Since our data is about bank products, we should see who are our customer.

- From the graph, we can see that observations customer age range varies between 20 to 70 but the most are in 20 to 22.

- and the pie Char shows the customer age range varies between 20 to 100 but the most are in 20 to 22.
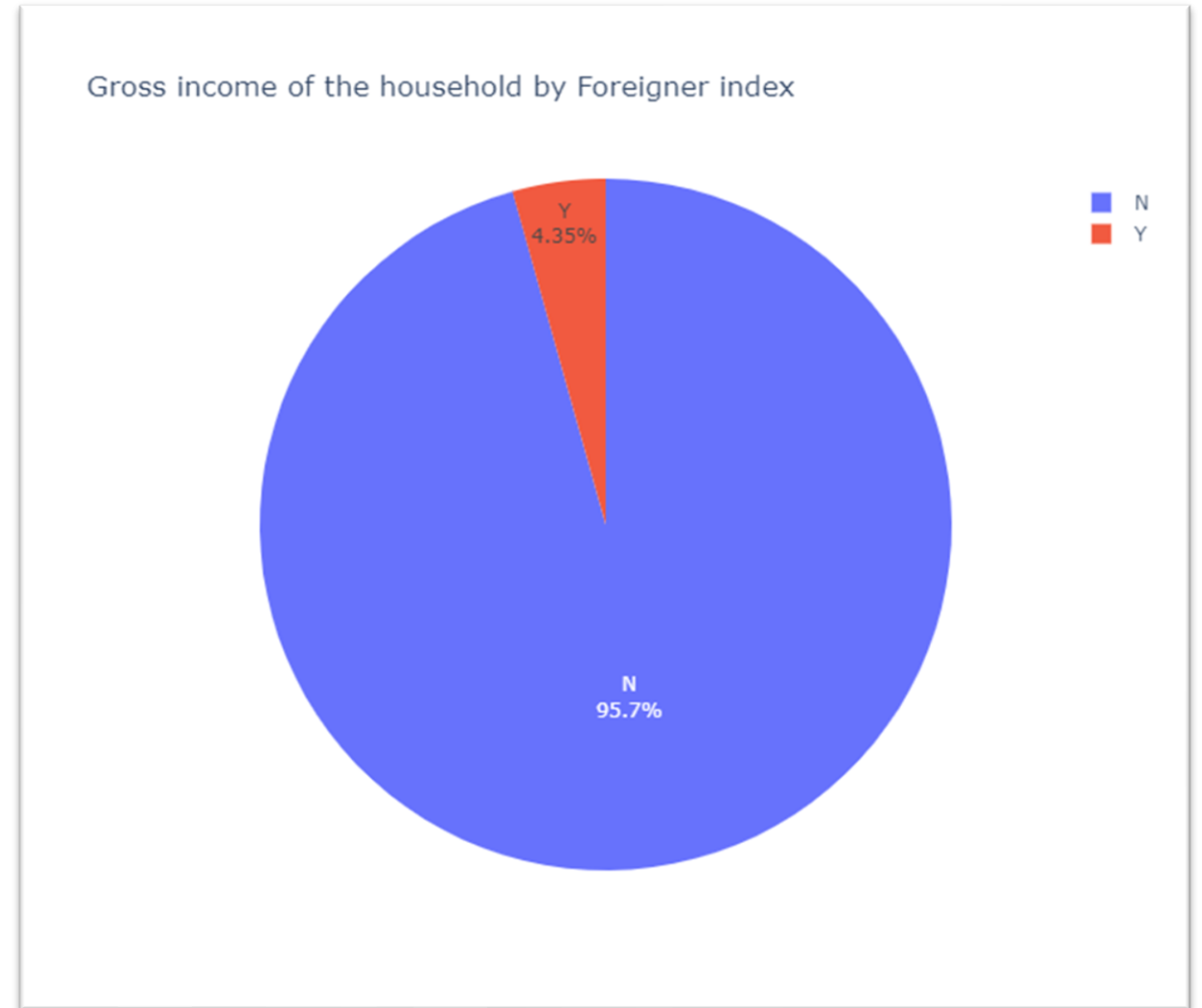

Number of customers by Age
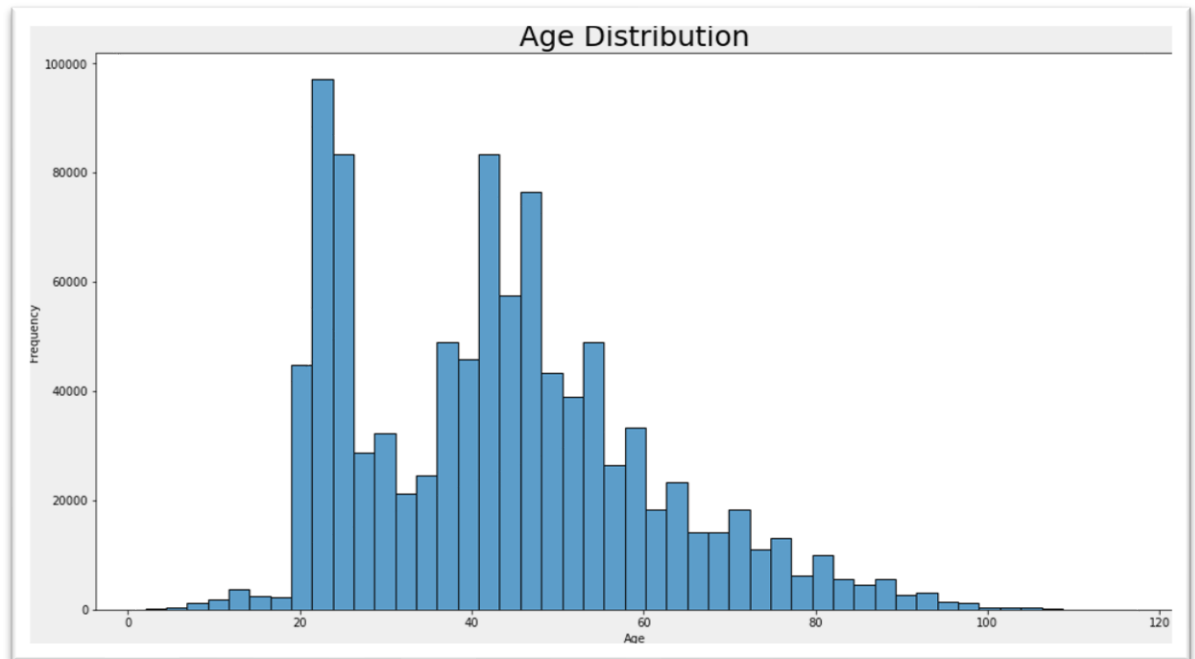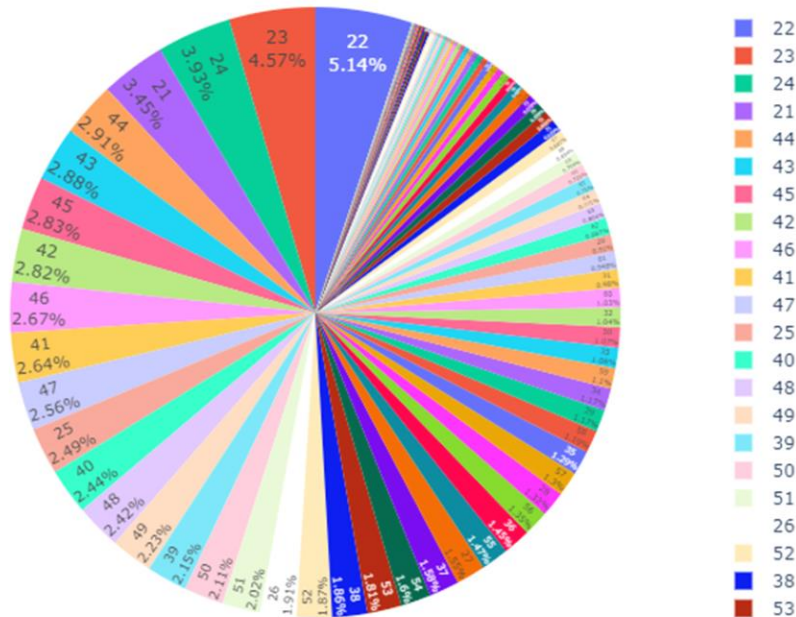
Age Distribution

Customer per gender

As per the pie chart, the number of customer by gender shows the Female customers are more than Male.

Number of Customers by Gender

M
43.2%

F
56.8%

F
M
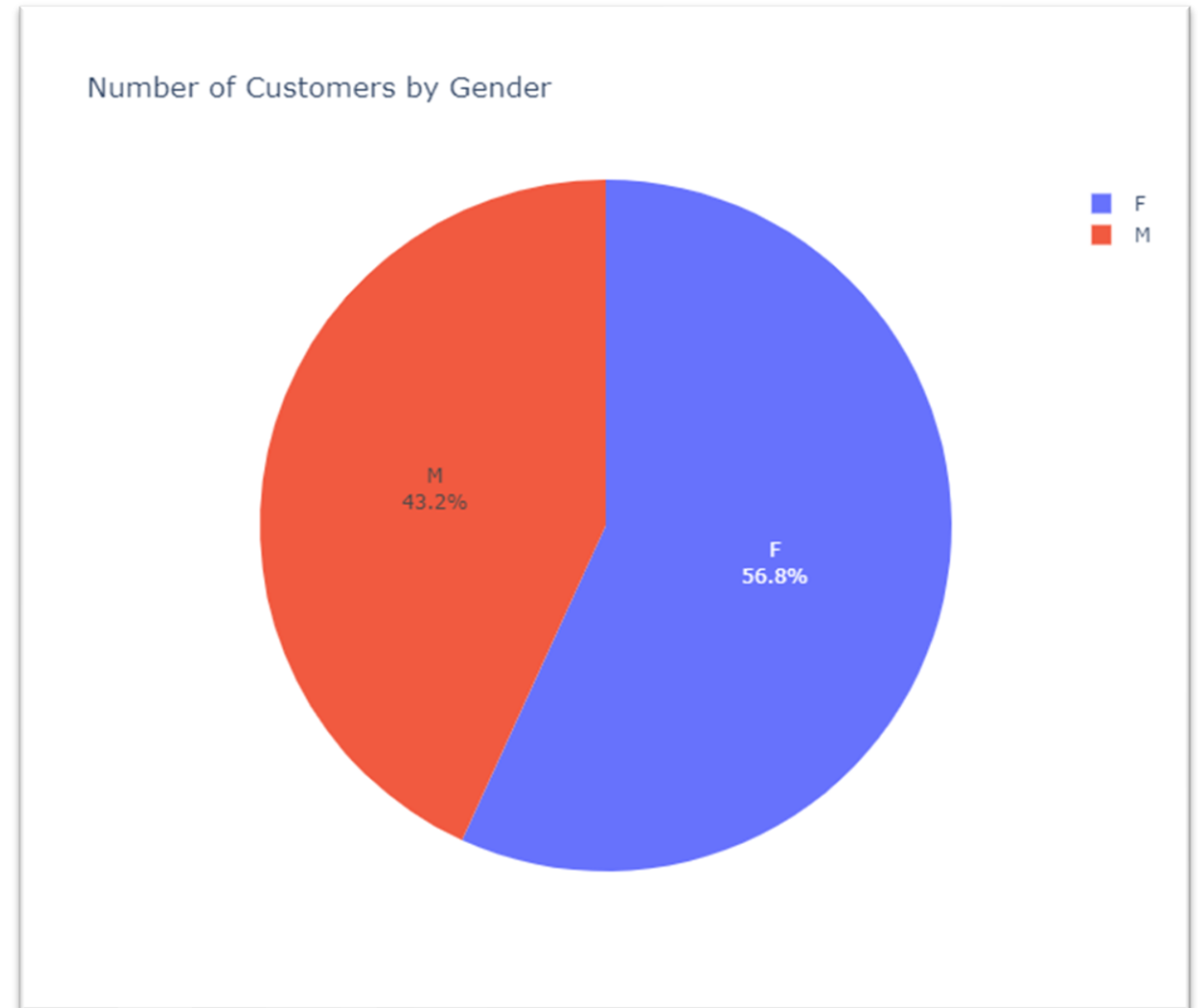
Customer citizenship distribution

The pie chart shows the number of Citizens and foreigners' customers.

Gross income of the household by Foreigner index

Customer by Nationality

As per the pie chart, the number of customer by Nationality shows the customers are from Spain.

Total Customers by Nationality

ES
99.3%

ES
FR
AR
DE
GB
US
CO
RO
IT
VE
EC
MX
BR
CH
BO
BE
PY
PT
CL
CU
PE
NL
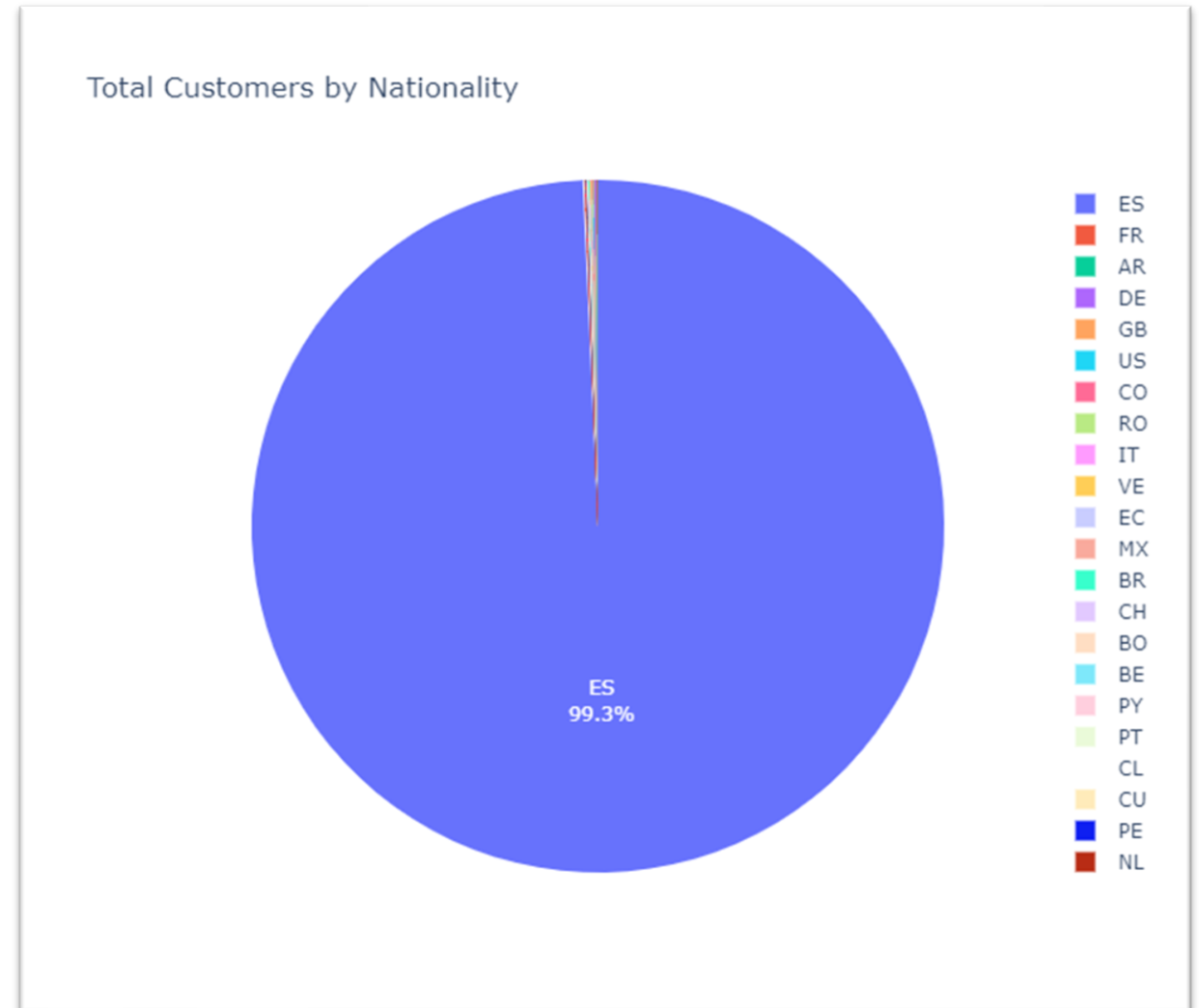
Customer by cities distribution

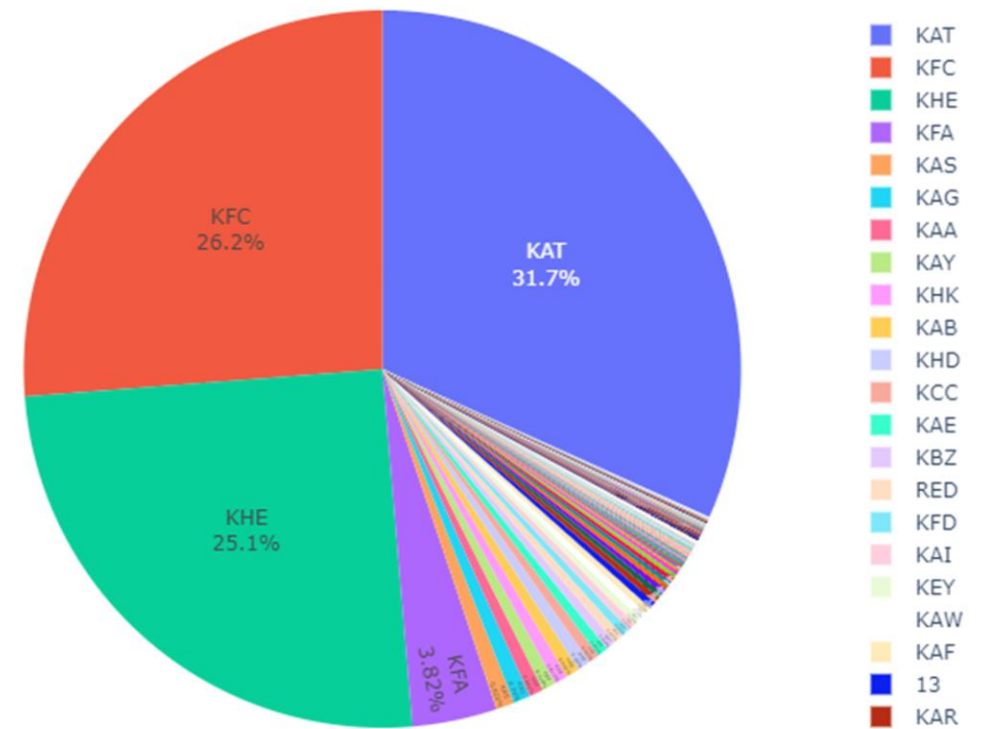As per the pie chart, the number of customer by Cities shows the most customers are in Madrid then Barcelona.

Total Customers by Country residence
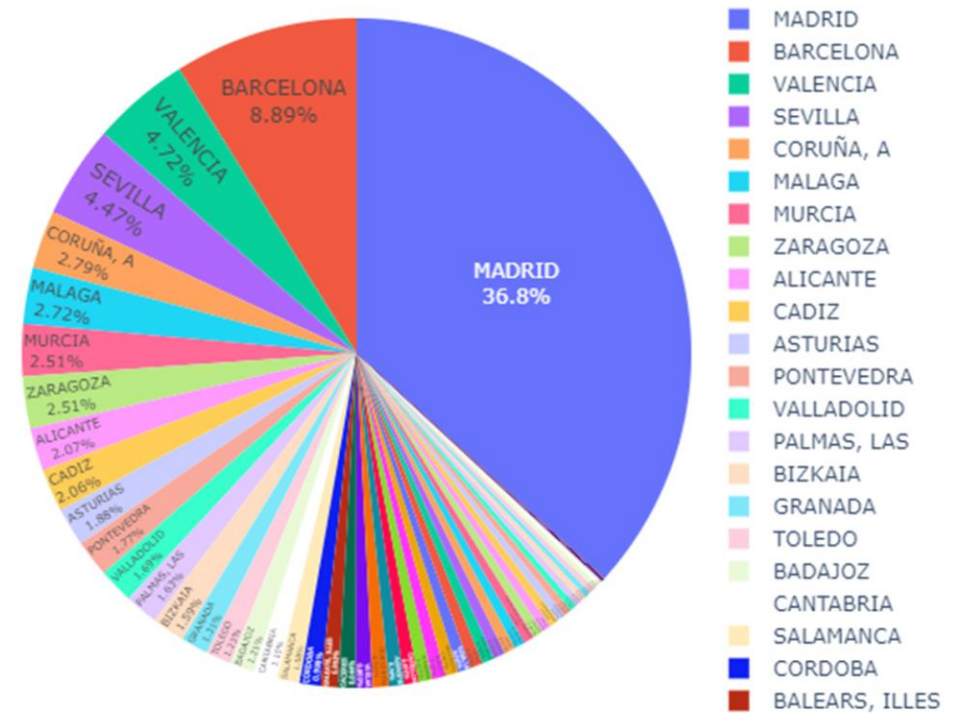
Number of Channels used by the customers

From the chart the most used channels are (KAT, KFC and KHE).
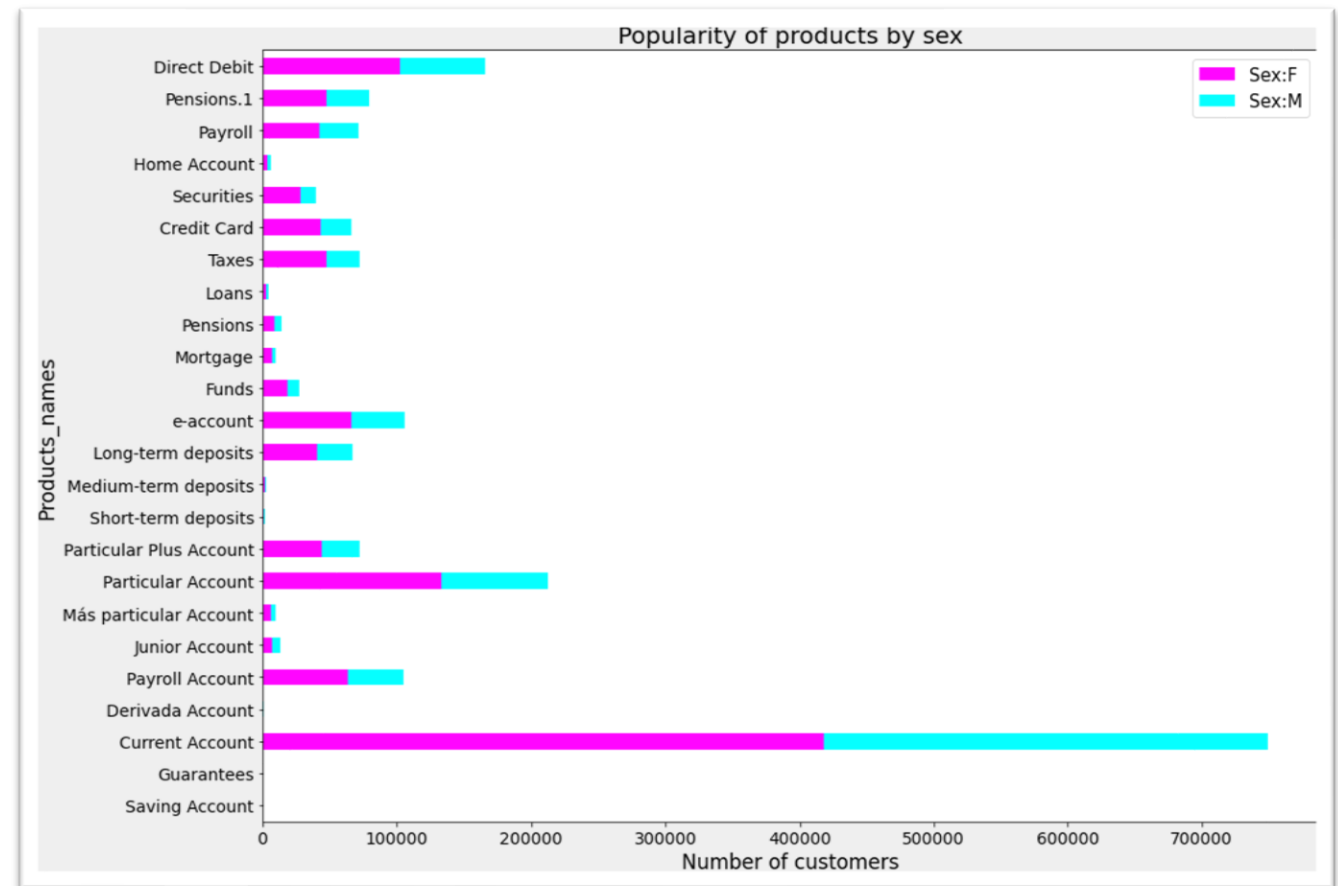
Products Analysis

The bank most products used was as follow:

Current Account: 749626
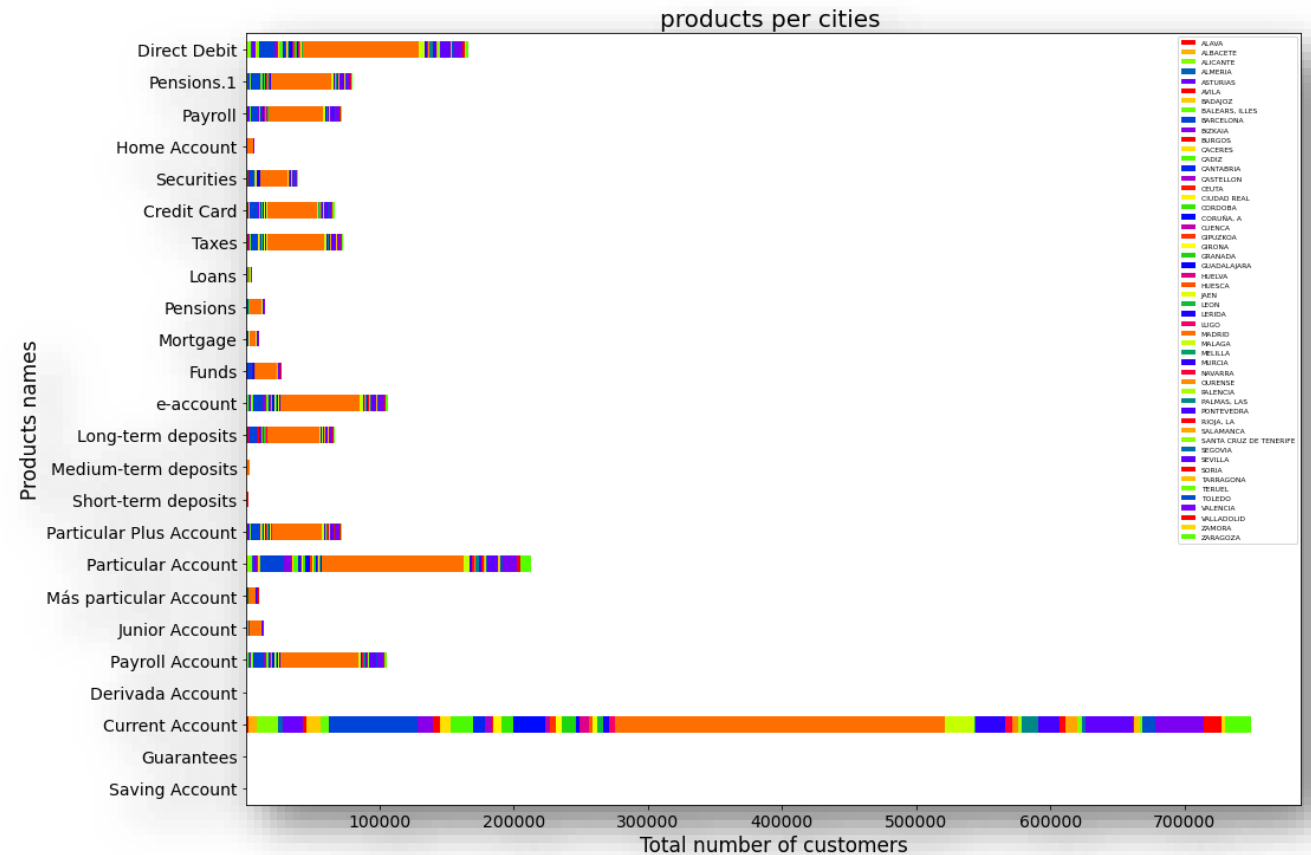Particular Account: 212486
Direct Debit: 166275

The graph show the distribution of the products by gender

Popularity of products by sex

Products Analysis

The bank most products used per cities was as it's shown in the graph.

However, The graph show the distribution of the products by cities, and the most city using bank products is Madrid the color code is orange.

Products Analysis

The bank products used per age as it's shown in the graph.

The graph show the distribution of the products by age, and the most people using bank products is youth and meddle age group of people.

Number of products by age

Products Analysis

The bank products used per age as it's shown in the graph.

The graph show the distribution of the products by income class, and the most people using the bank product "current account".

Distribution of products among customers by income group
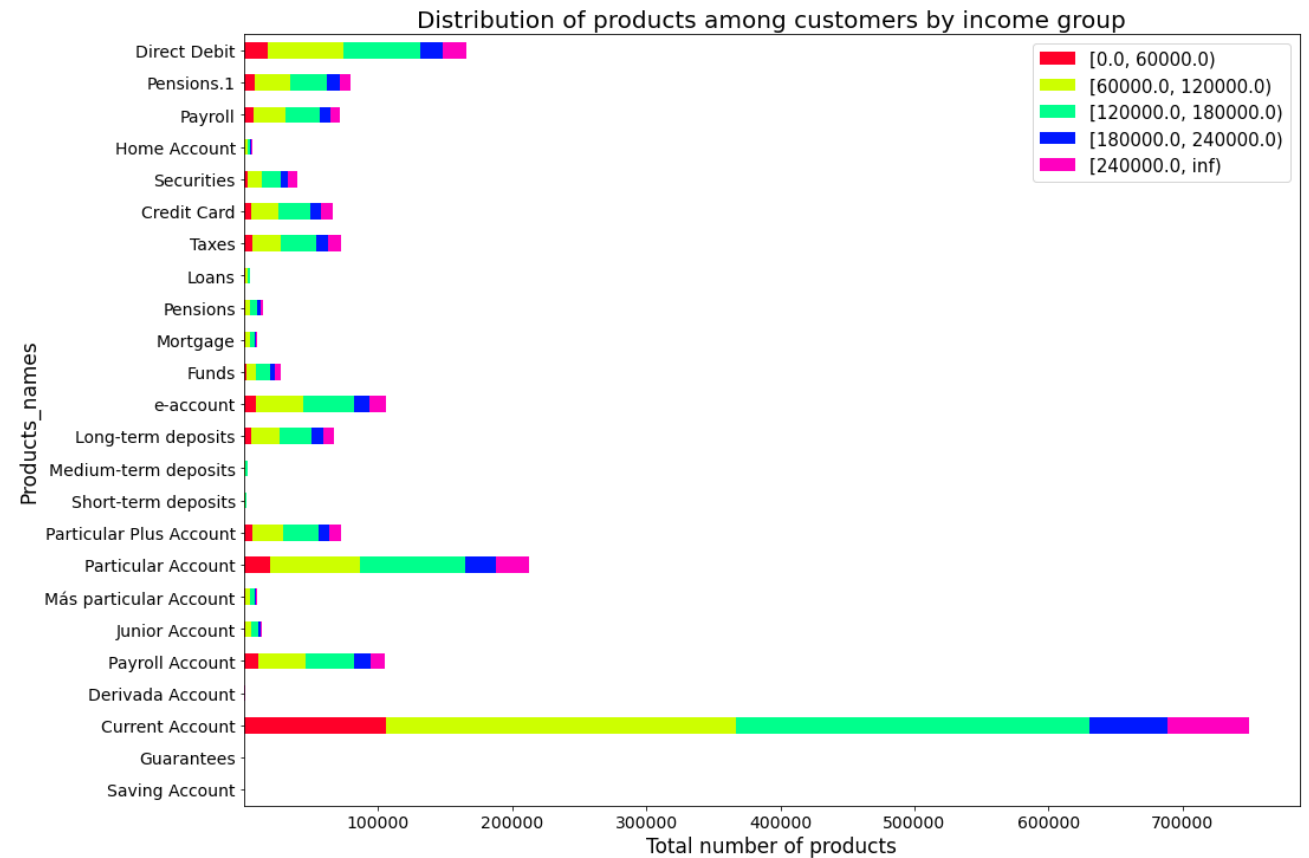
Conclusion

At the end from the above analysis, we summarize the following:-

- Most of the customers and products used are in Spain

- Madrid has the highest costomer numbers and product used.

- Females are the most customer number in the bank in term of genders.

- Regarding to the ages, all observations mentions that customer in twenties are most bank's customers.

- The most account/product used is the Current Account.

- The most channel used by customer is KAT channel.

Recommendations

1. According to the age insights it is highly recommended to focus on ages that is younger than and older than twenties ages.

2. Males has lower product usage than Females, so we recommend to give special offes for Males.

3. We can say that Spain is the only Country which uses bank's prosucts, the bank has to foucs on other countries.

4. As per the insights most of the customers are in Madrid then Berclona, the other Cities are very low so special offers has to proposed to these cities to increase the customer numbers.

5. the bank has to give spcial offers for the prodcuts rather than current account since it is the most used product.

Tech Recommendations

The technical recommendations will be to use the following methods. Since the data have no **dependent value**, **unsupervised learning** techniques should be applied.

- **The elbow method:**

- The elbow method uses the sum of squared distance (SSE) to choose an ideal value of k based on the distance between the data points and their assigned clusters

- **PCA:**

- Principal component analysis (PCA) is a widely used statistical technique for unsupervised dimension reduction. It is believed that it improves the clustering results in practice (noise reduction)

- **Clustering:**

- Clustering is the task of dividing data points into a number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups.

M.A.S

Thank You