

Group Name : Dream Crushers

Name : Aiman Lameesa

Email: st122876@ait.asia

Country :Bangladesh

College: Asian Institute of Technology

Specialization (Data Science, NLP, Data Analyst): Data Science

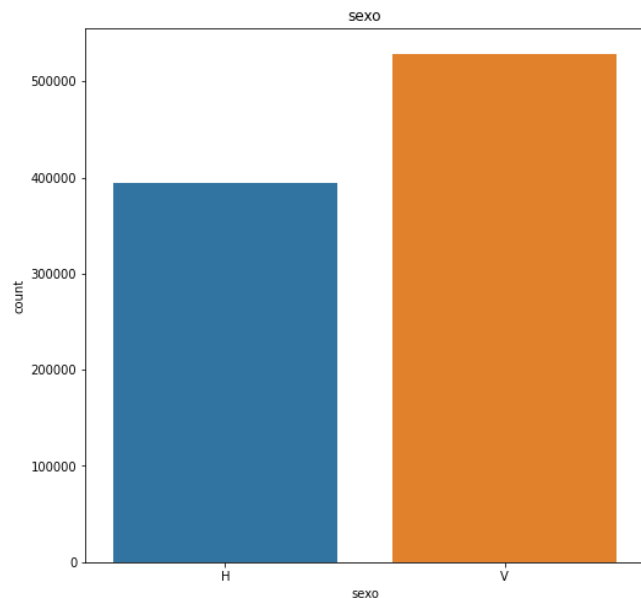
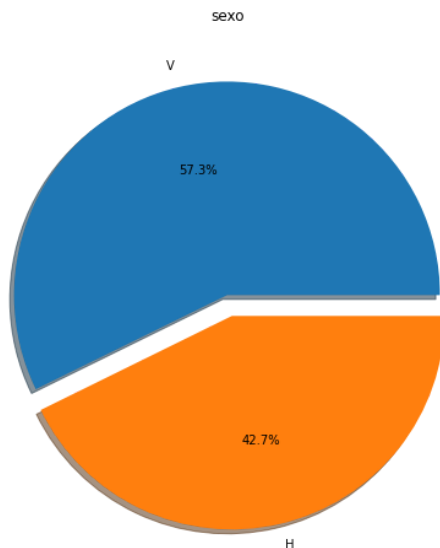
“Customer Segmentation” dataset is selected for my group project.

Problem description:

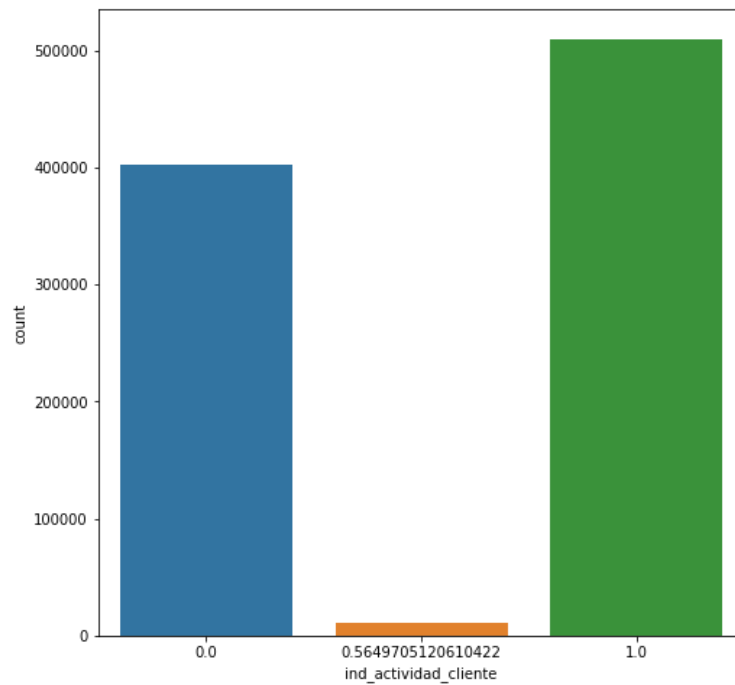
XYZ bank wants to introduce their Christmas offers to their customers. But they do not want to roll out the same offer to all their customers since it will not be profitable to introduce the same offer for different types of customers. Instead, they decide to initiate personalized offers to different sets of customers. Moreover, it is not efficient and beneficial for them to manually understand the hidden patterns in their customer data. That is why, they approached an analytics company, ABC, to help them to understand their customer behaviors in order to introduce Christmas offers. They mentioned to the company that they prefer to have at most 5 groups of customers to maintain the efficiency of their campaign.

Exploratory Data Analysis:

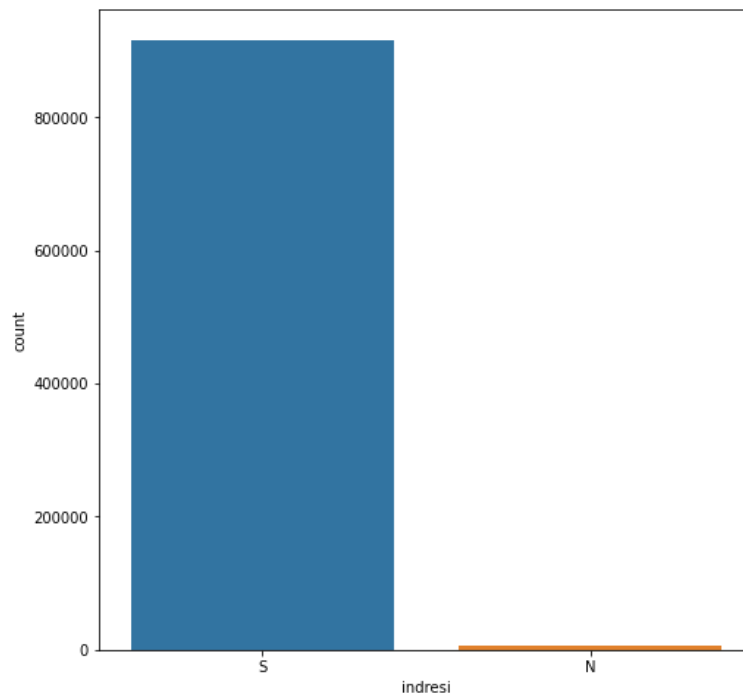
1. Percentage and Total Count of Customers' Gender: A pie chart and a countplot are given below to show the ratio and the total count of the customers' gender respectively.



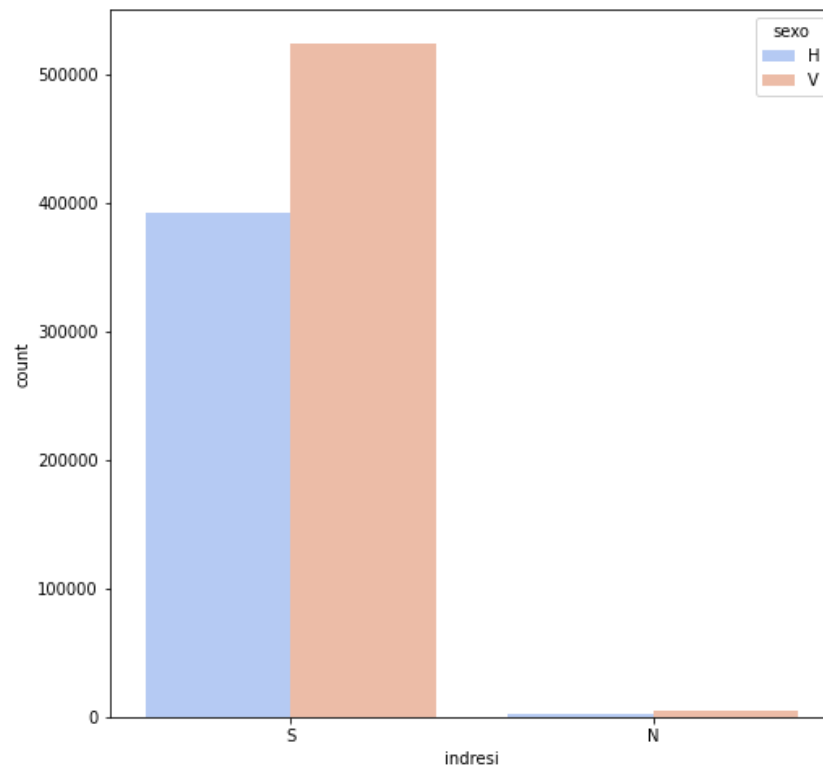
2. Countplot of Active and Inactive Clients: A countplot is given based on the activity index where there are more active clients than the inactive ones.



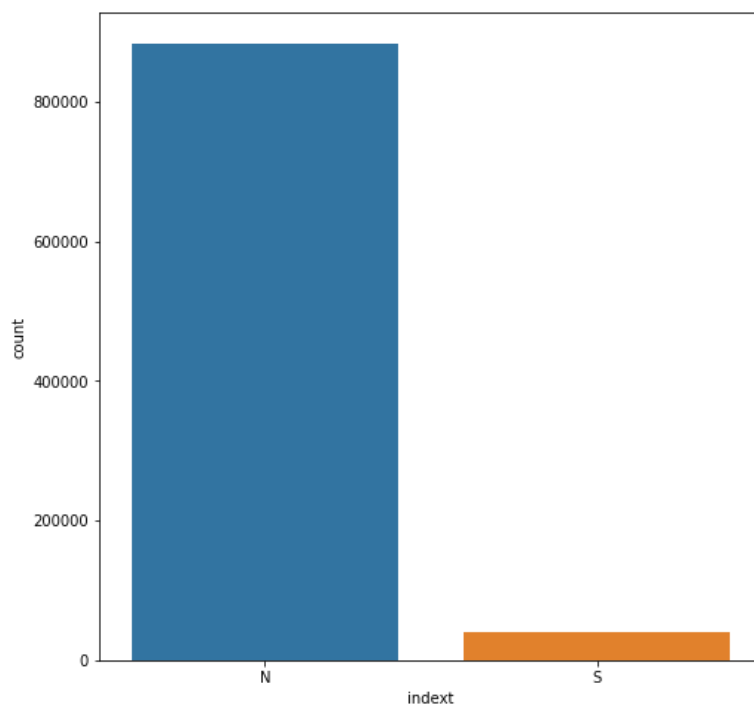
3. Countplot based on Residence Index: The countplot shows that there are comparatively higher numbers of resident clients than the non-resident ones.



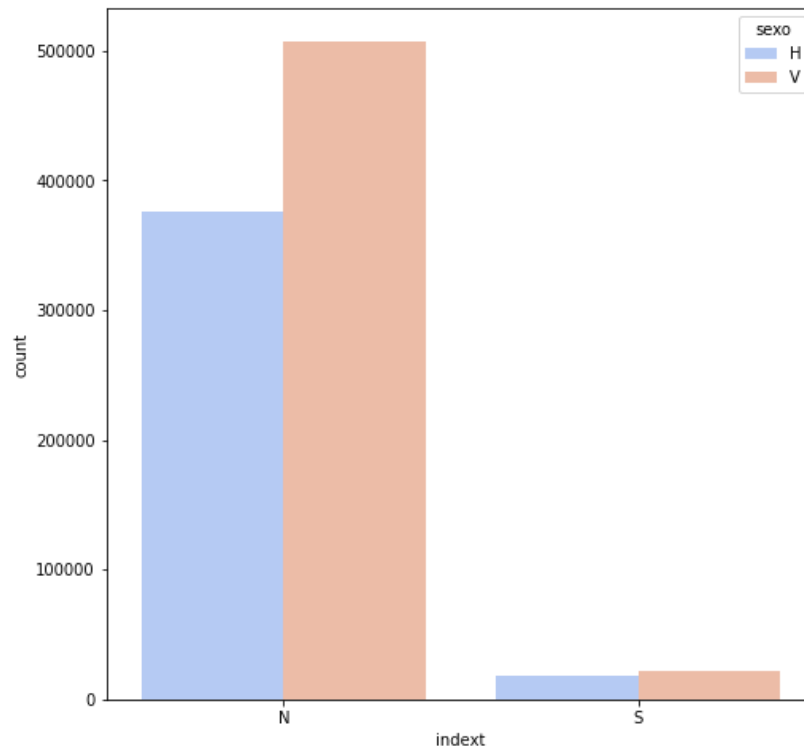
The countplot of residence and non-residence clients labeled by their gender is shown below:



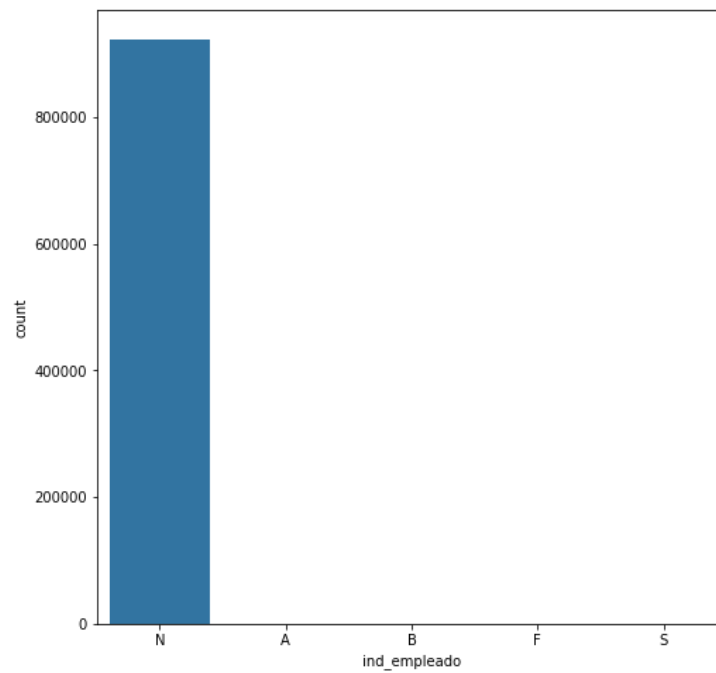
4. Countplot based on Foreigner Index: The countplot shows that there is a high number of clients whose birth country is not different from the bank country.



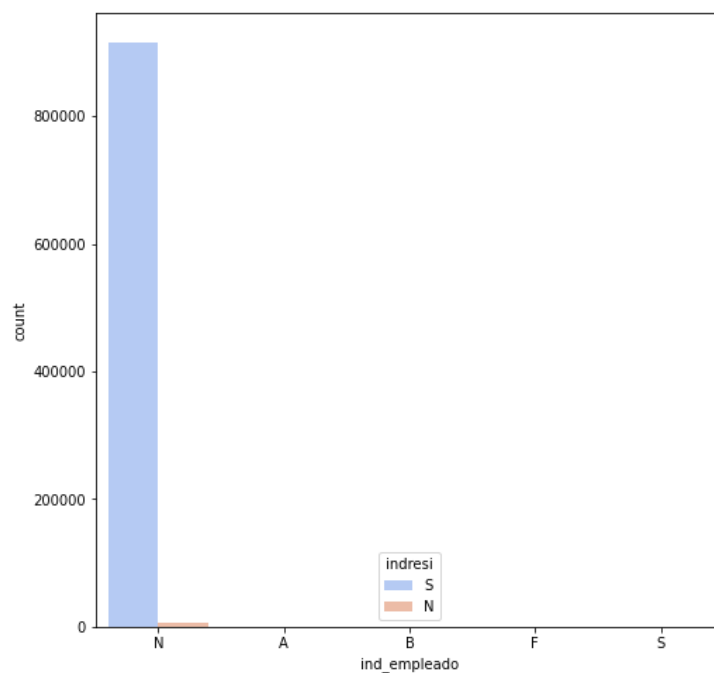
The countplot of non-foreigner and foreigner clients labeled by their gender is shown below:



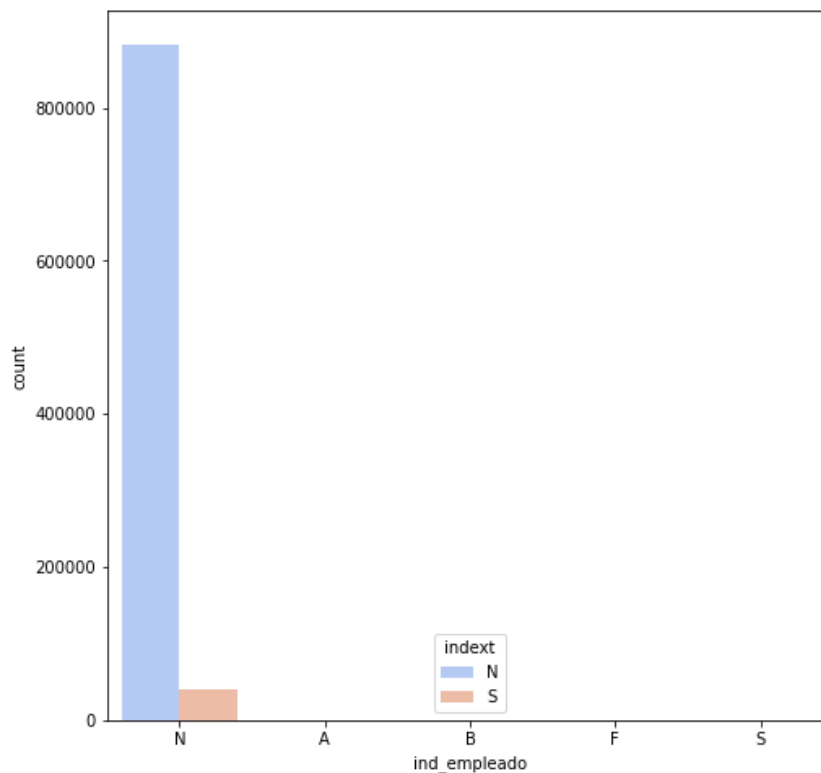
5. Countplot based on Employee Index: The countplot shows that most of the customers are not employed.



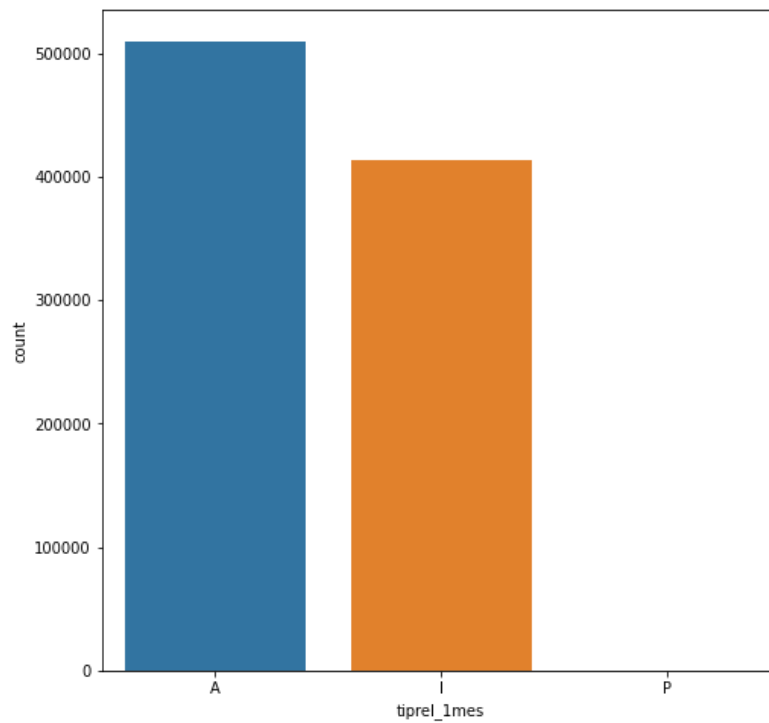
Moreover, the majority of the non-employed customers are residents, which means that their residence country is the same as the bank country.



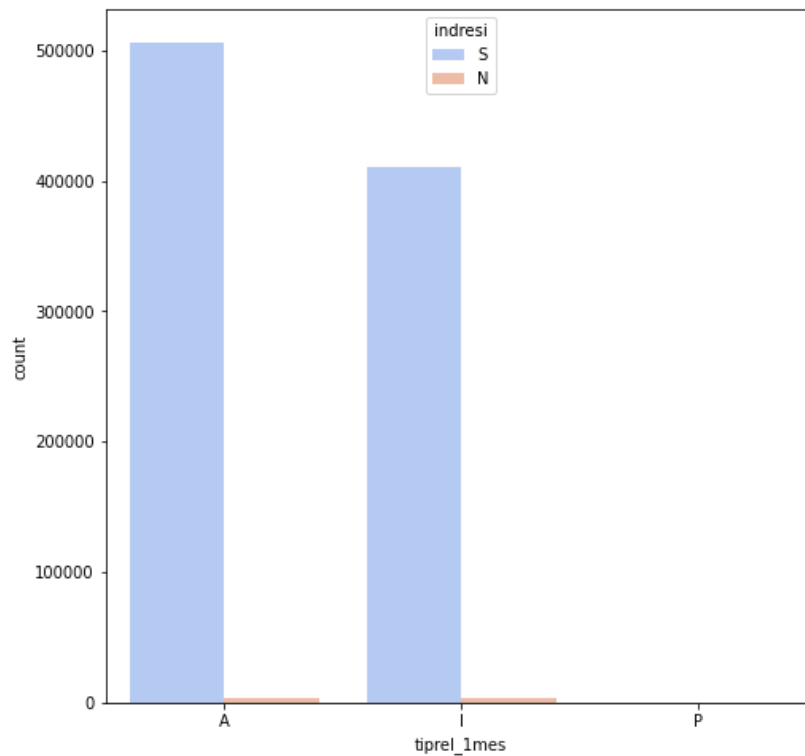
Based on the foreigner index, most of the non-employed customers are not foreigners, but there are a few people foreign customers as obtained from the countplot given below:



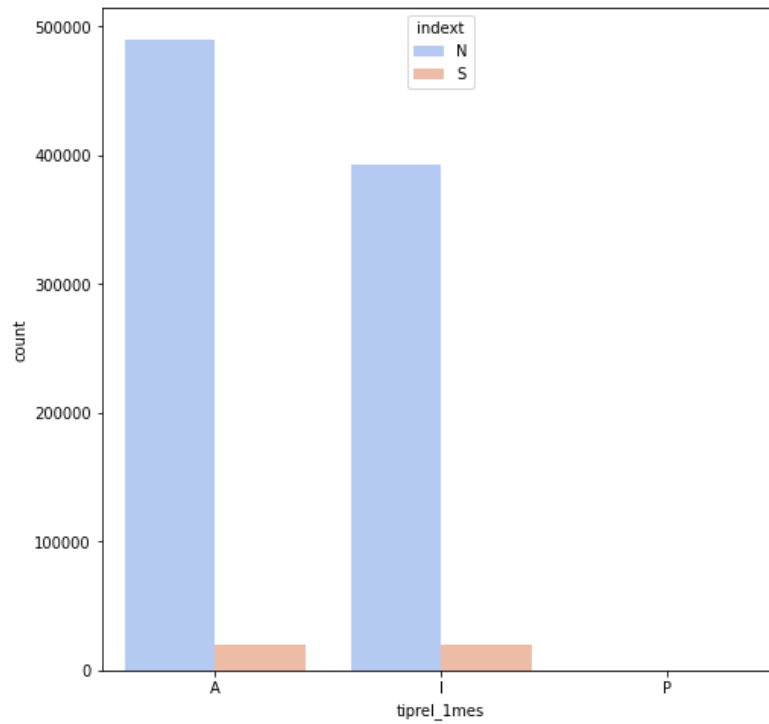
6. Countplot based on Type of Customer at the Beginning of the Month: The countplot shows that there are many active customers than inactive ones.



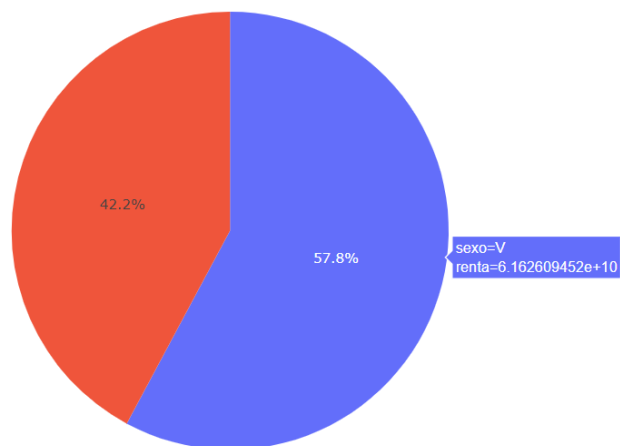
Most of the active and inactive customers are residents.



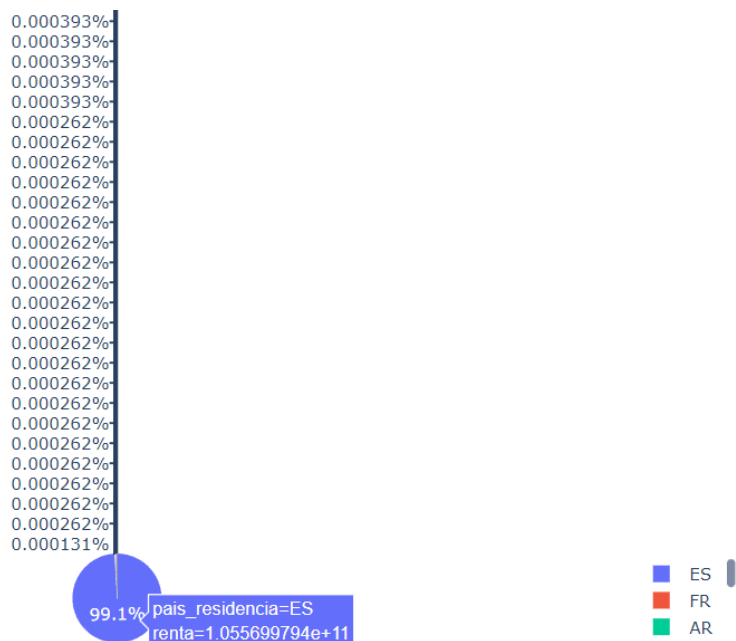
Also, there are a few foreigner customers among the active and inactive ones.



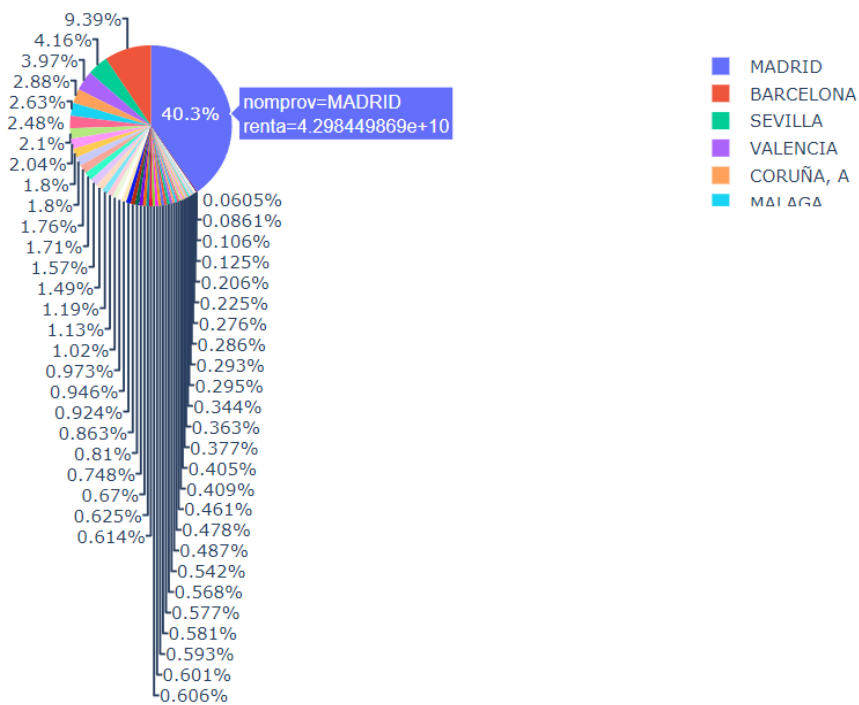
7. Interactive Pie Chart on Gross income of the Customers: The piechart shows that there is many customers of sex “V” than “H” who has a higher number of gross income of the household.



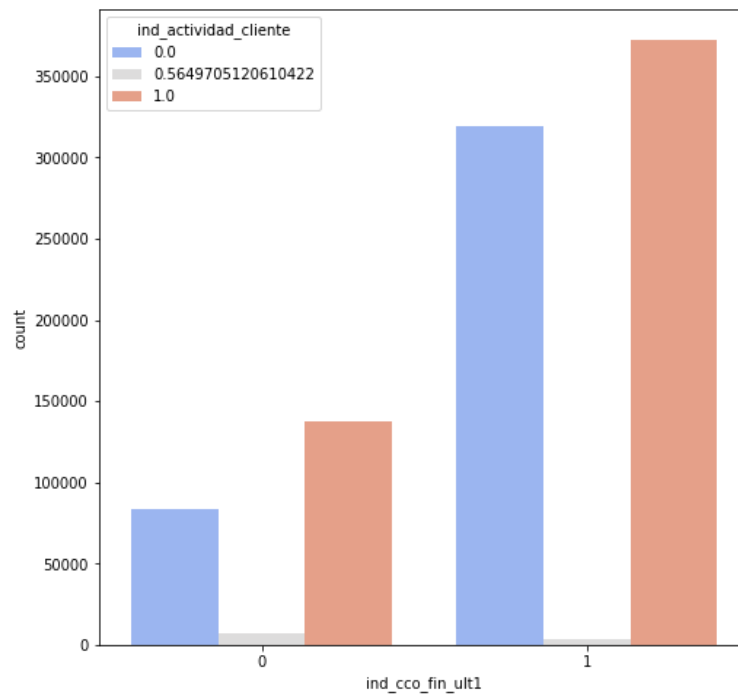
Also, 99.1% of the customers are from “ES” who have a high gross income than the other customers.



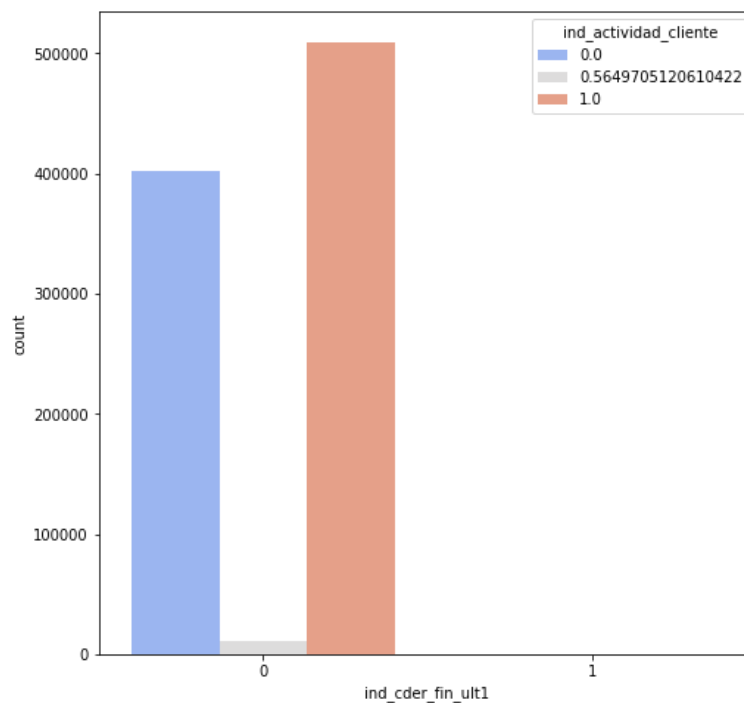
The majority of the customers are from “Madrid” province with high gross income of the household.



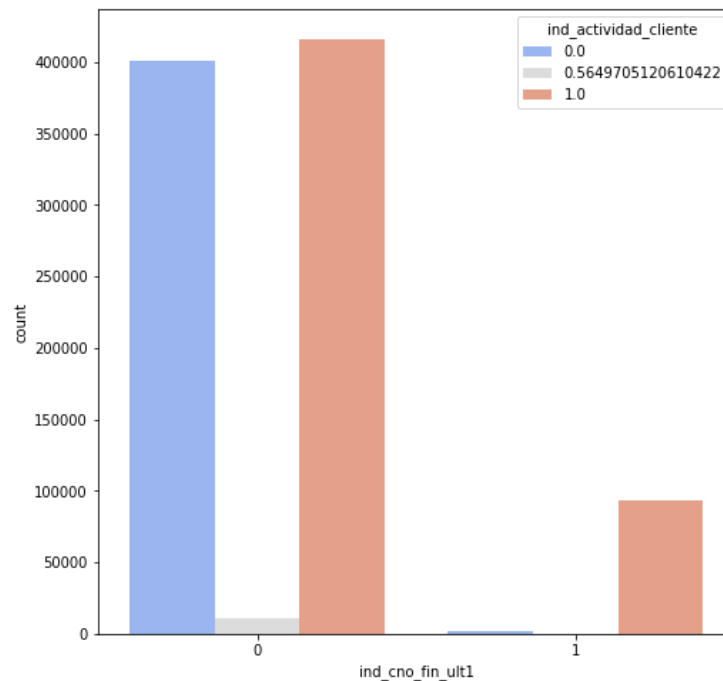
8. Countplot of Current Accounts based on Activity Index of the Customers: The countplot shows that the majority of the active and inactive customers have current accounts.



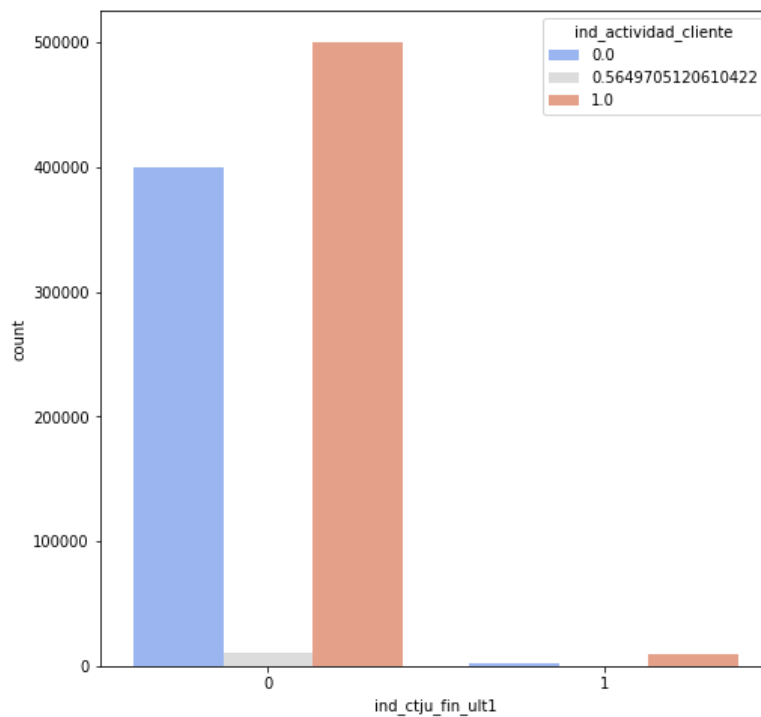
9. Countplot of Derivada Account based on Activity Index of the Customers: The countplot shows that the majority of the customers don't have a derivada account.



10. Countplot of Payroll Account based on Activity Index of the Customers: Almost all the inactive customers have payroll accounts than the active ones.

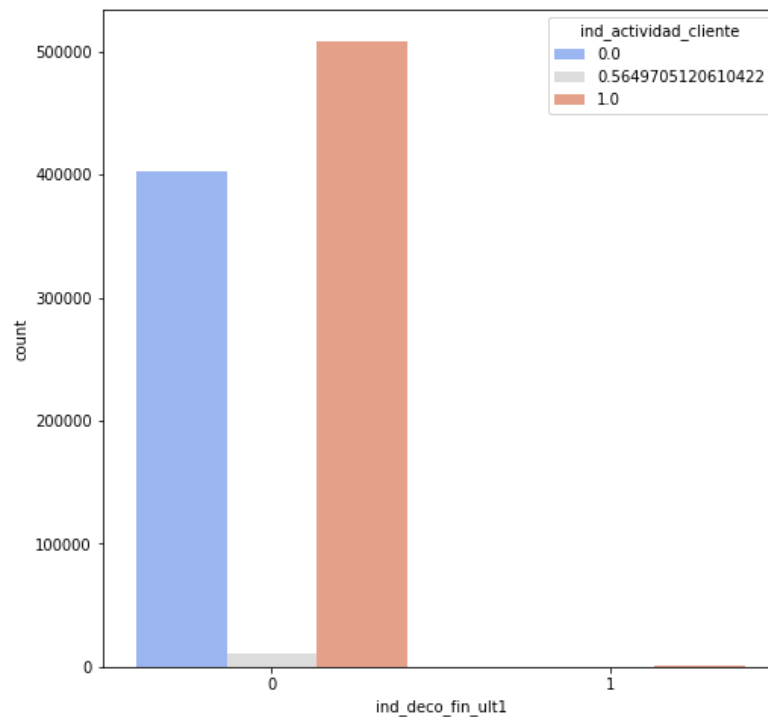


11. Countplot of Junior Account based on Activity Index of the Customers: Almost all the inactive customers have junior accounts, whereas a very few active customers have them.



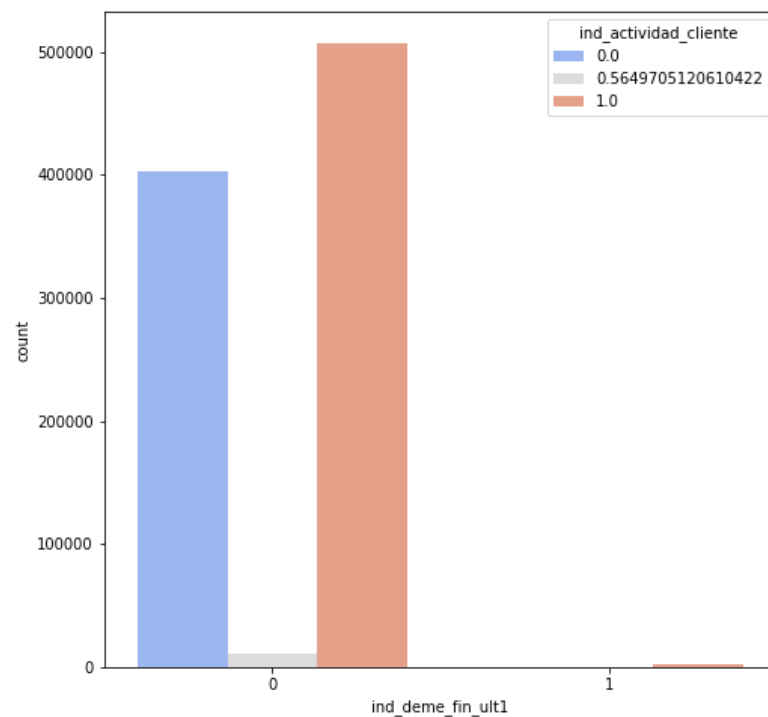
12. Countplot of Short-term Deposits based on Activity Index of the Customers:

Mostly the inactive customers have short-term deposits.



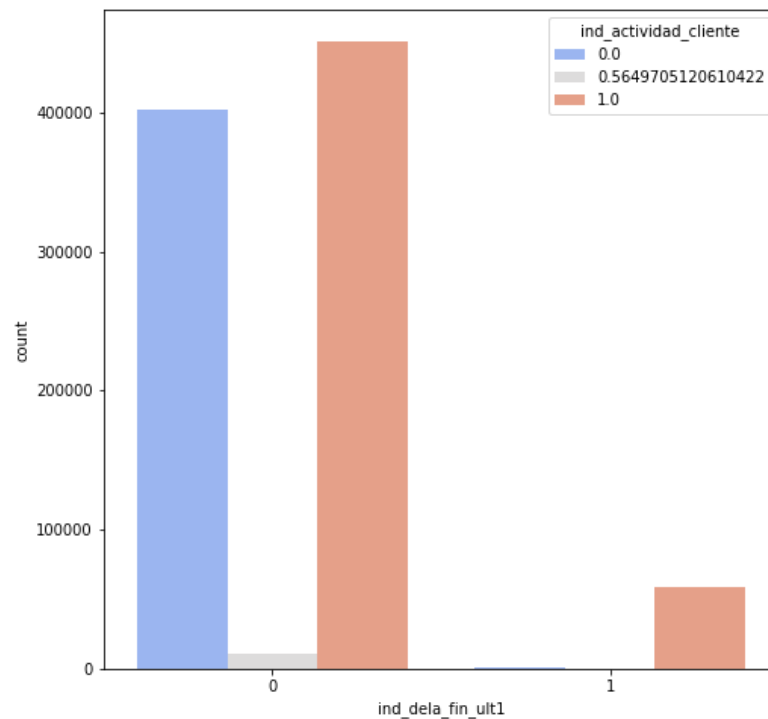
13. Countplot of Medium-term Deposits based on Activity Index of the Customers:

Mostly the inactive customers have medium-term deposits.



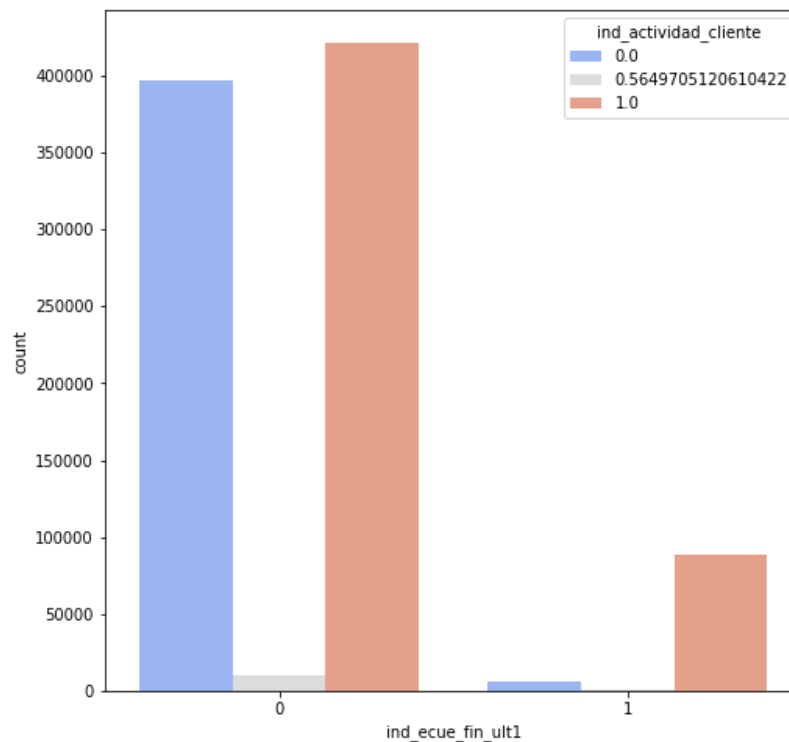
14. Countplot of Long-term Deposits based on Activity Index of the Customers:

Mostly the inactive customers and a few active customers have long-term deposits.

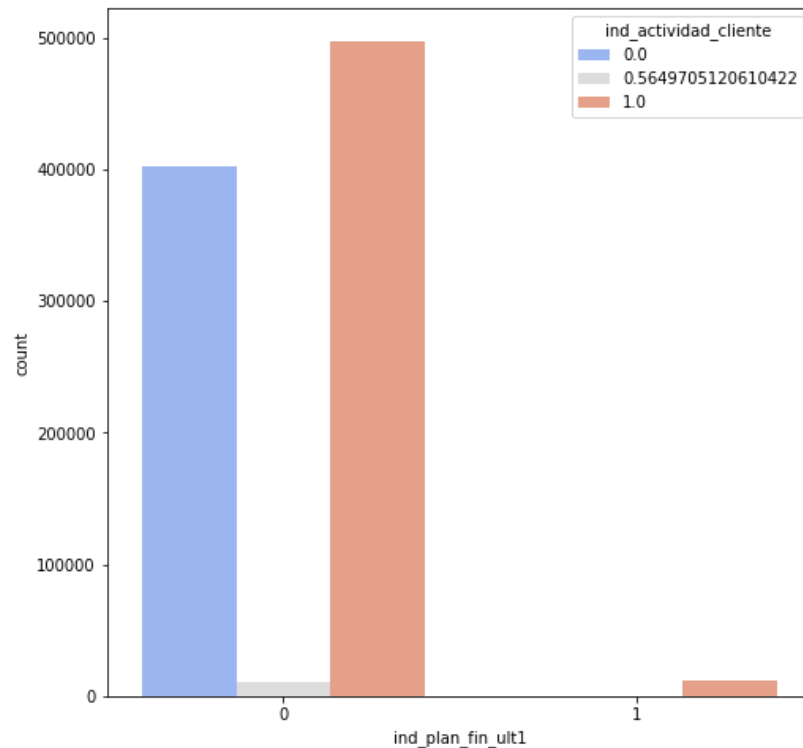


15. Countplot of E-account based on Activity Index of the Customers: Mostly the inactive customers and some active customers have an e-account.

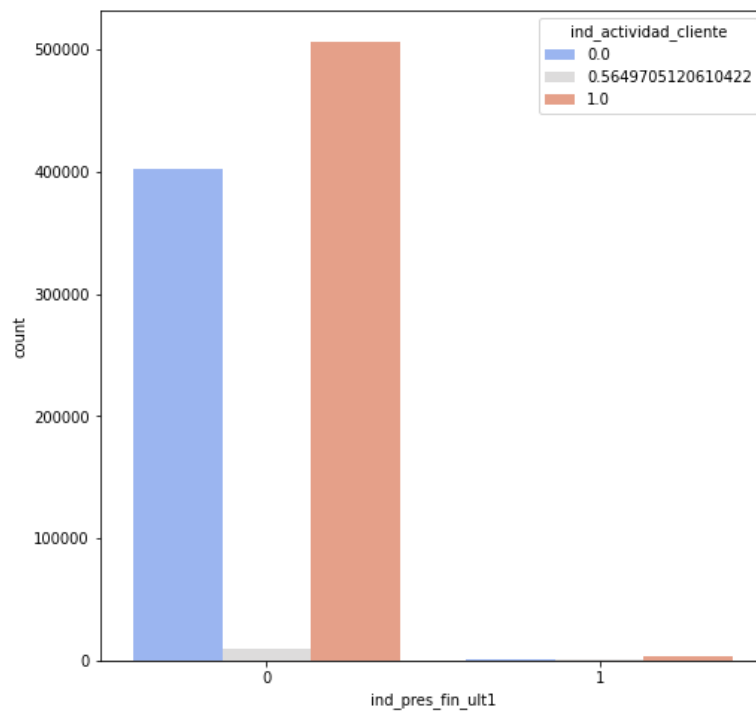
Mostly the inactive customers and some active customers have an e-account.



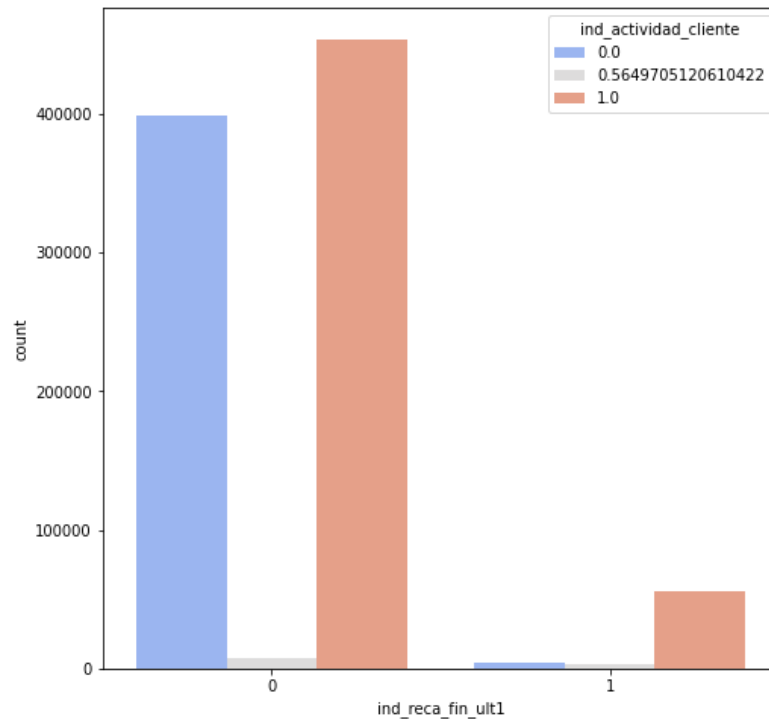
16. Countplot of Pensions on Activity Index of the Customers: Mostly the inactive customers and a few active customers have pensions.



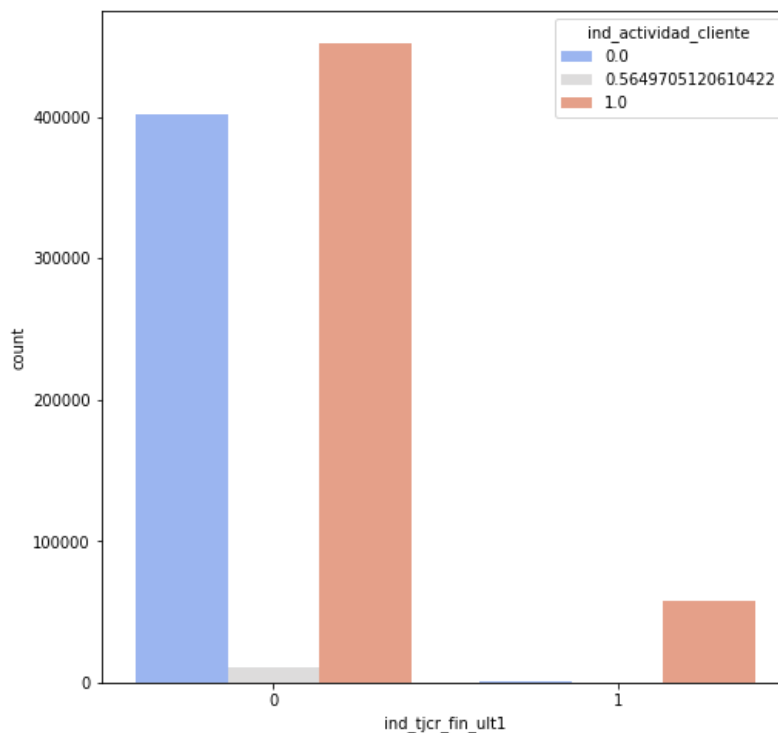
17. Countplot of Loans on Activity Index of the Customers: Almost all the inactive customers have loans.



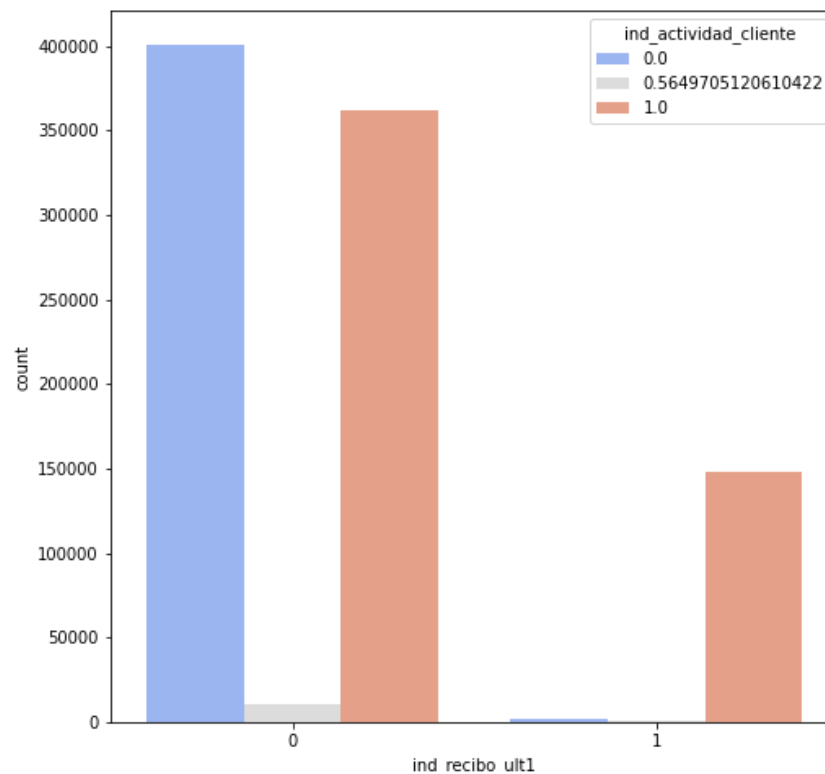
18. Countplot of Taxes on Activity Index of the Customers: Almost all the inactive customers and a few active customers have taxes.



19. Countplot of Credit Card on Activity Index of the Customers: Almost all the inactive customers and a few active customers have credit cards.



20. Countplot of Direct Debit on Activity Index of the Customers: Inactive customers have more direct debit than the active ones. A quite number of active customers have direct debit too.

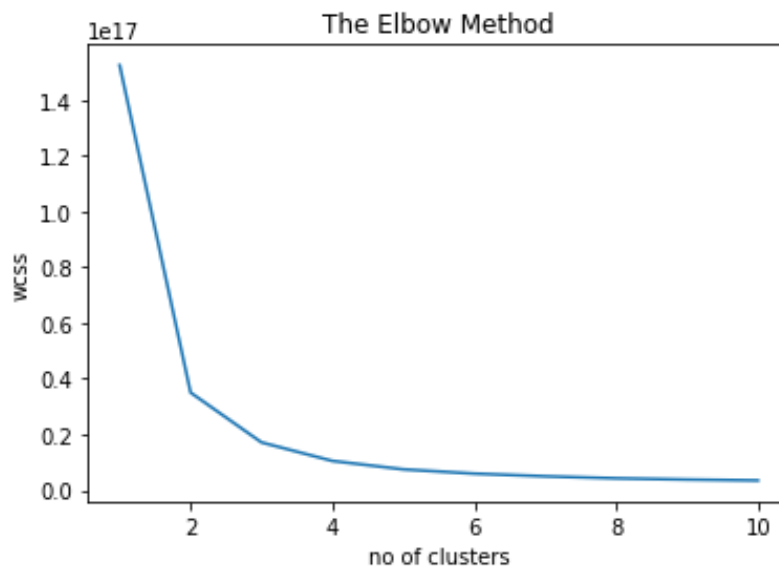


Final Recommendation: From the exploratory data analysis, we can see that, we can get around 2 or 3 groups of customers using the given features. To segment the groups of the customers, we can use unsupervised machine learning algorithms, such as, K-means clustering in this given dataset. There are other clustering algorithms to segment the customers into different groups but K-means clustering is the most popular one to solve segmentation problems given some features. It utilizes some clusters to calculate the features for each customer. Thus, it comes to a conclusion on how the customers should be segmented into different groups. So, this algorithm should be implemented in the next step for customer segmentation in this project. We can also implement other algorithms, i.e., hierarchical clustering and other algorithms to compare the results and come to a conclusion.

Model Selection and Model Building:

We selected K-means clustering, where we used the Elbow method to find out the proper number of clusters we should use to segment the customer into different groups. We experimented with the Elbow method from 1 to 10 clusters. In the obtained result, we get two

sharp curves at clusters 2 and 3, which indicates that the number of clusters or groups for the customers given in the dataset should be either 2 or 3.



Conclusion:

It is recommended to the bank to introduce Christmas offers to customers considering that their customers can be segmented into either 2 or 3 groups.

Github Repo Link:

[aimanlameesa/Week-12 \(github.com\)](https://github.com/aimanlameesa/Week-12)