Your Deep Learning Partner

Exploratory Data Analysis

<Repeat Sales>

Sept 9 2024

Agenda

Executive Summary
Problem Statement
Approach
EDA
EDA Summary
Recommendations

Data Glacier
Your Deep Learning Partner

Executive Summary

Repeat Sales presents a data-driven plan for XYZ Credit Union to enhance cross-selling strategies. By analyzing customer behavior and transaction data, our recommendations focus on optimizing product offerings and personalized marketing efforts to increase engagement and drive revenue growth through targeted cross-sell opportunities.

Problem Statement

XYZ Credit Union in Latin America are selling business products such as Credit Cards, Deposit Accounts, and Retirement Accounts, but they are not performing well in cross-selling: persuading customers to acquire more than one of their banking products. This is resulting in reduced customer satisfaction, loyalty, and opportunities for revenue growth.

Impact: Reduced revenue from business products, missed opportunities for enhanced client engagement, and potential loss of market share to competitors who offer more comprehensive solutions.

Objectives: To enhance the effectiveness of cross-selling business products, thereby increasing client uptake and revenue.

Approach

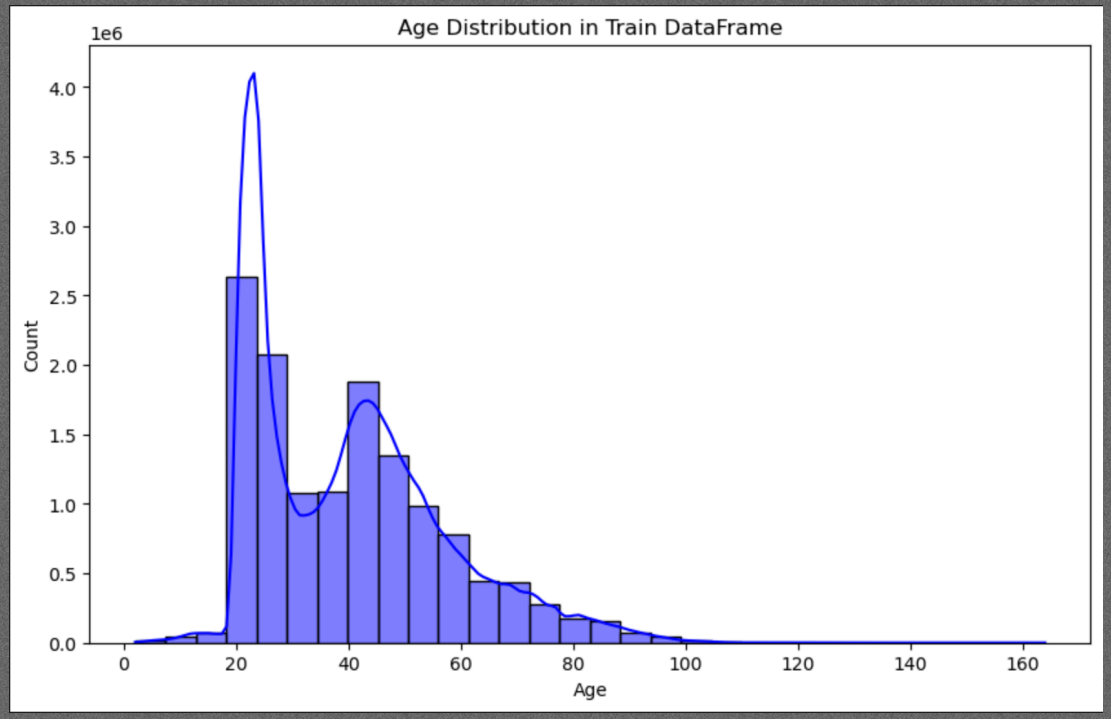
- For mixed data types, sort through the affected columns and make sure they contain strictly the data types that are expected. If they do not, either convert them to the same type or omit them if they are unable to do so. We want to do this to make our data more accessible and readable, and to ensure that we can process each column without any difficulties or errors.
- For missing or null values, first check all the columns that are affected to ensure they have been properly calculated. If there is a mistake, try to convert the objects in each column into something readable. For some numerical columns, we used mean or median-based imputation approaches, and for categorical columns used mode-based imputation. For columns with very few missing values, we often dropped them.
- For outliers, locate them using the z-score and IQR tests. Collect and observe where most outliers are found, and remove them from the dataset if they do not fit in with the rest of the values. While it is difficult to closely inspect a dataset of this size, this process must be selective. After all, it is entirely possible, given the size of this file, that the 6 million 'outlier' values are not worth omitting, but rather provide invaluable insight into some of the strengths and weaknesses of the company. Therefore, it is important to both locate and inspect these outliers and program them into our EDA and understanding of the company.
- For skewness, inspect the columns in which the number is far from 0. For numbers greater than 0, understand that the data is right, or positively skewed, meaning that most of the data points are concentrated on the left side. For a column like 'age', a high positive skew would mean that the customer base is young, on average. For skewness numbers less than 0, be aware that the data is left, or negatively skewed. This means that most of the data points are concentrated on the right side of the distribution. For a column like 'seniority', this would mean that the bulk of the customer base has been there for plenty of months. To reiterate, we don't want to manipulate or mitigate skewness, but rather be cognizant of it while conducting our Exploratory Data Analysis.

Exploratory Data Analysis

- We have conducted 7 individual hypotheses to inform an insightful, detail-oriented proposal on how to improve cross-selling for XYZ Credit Union.

Hypothesis #1:

If we target an older demographic, then there is more time for customers to acquire other banking products



Age data is right (or positively) skewed, meaning the majority of the data is clumped on the left.

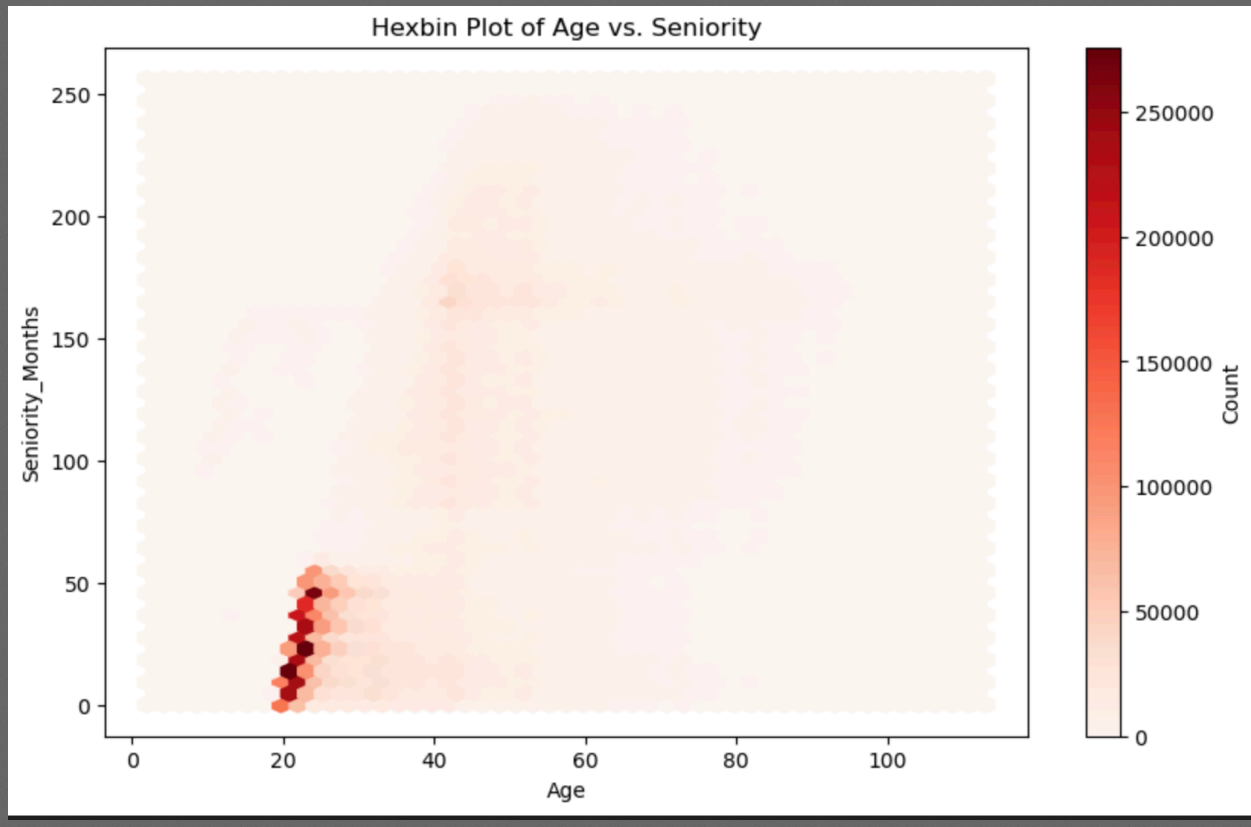
In general, this means that the credit union has targeted a younger audience.

Let's now look at the relationship between age and seniority– how long a customer stays with the credit union

Hypothesis #1

There is no obvious trend or correlation between Age and Seniority.

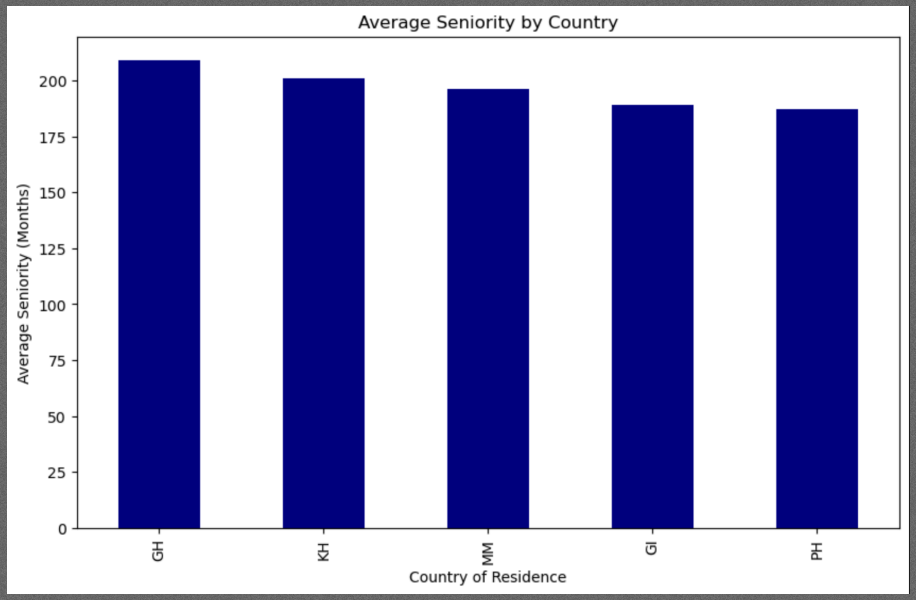
The only noticeable pattern we can recognize is the dark clump around age 20 and a slightly lighter clump around age 40 on the x-axis. On average, the average seniority for a 20-year-old is between 0-50 months, while the average seniority for a 40-year-old is around 150-170 months.



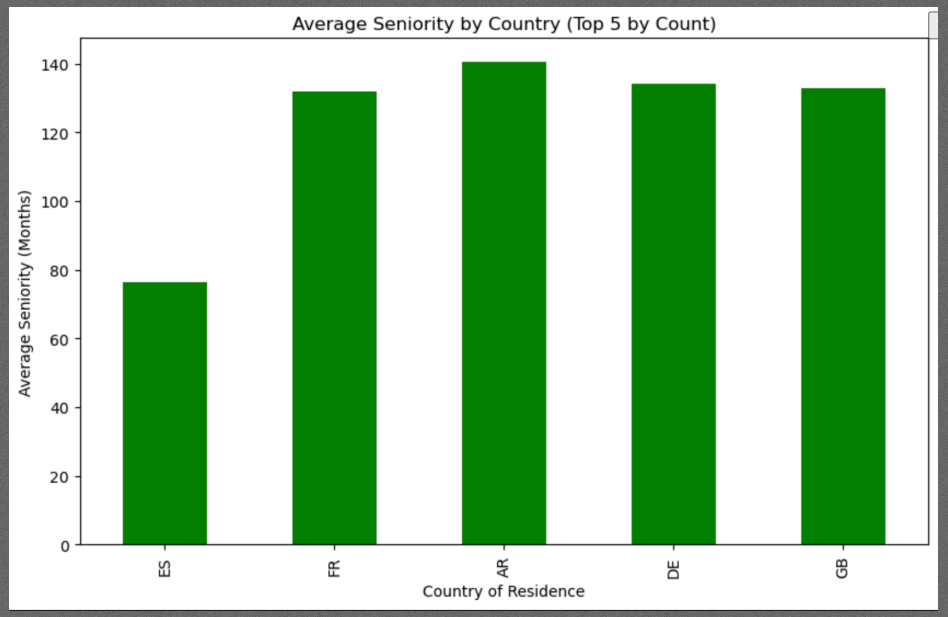
But this still doesn't tell us whether we should target younger clients who can develop seniority or older people who already have proven loyalty.

Hypothesis #2

If we develop stronger loyalty in the locations that support us most frequently, then cross-selling will improve as a result



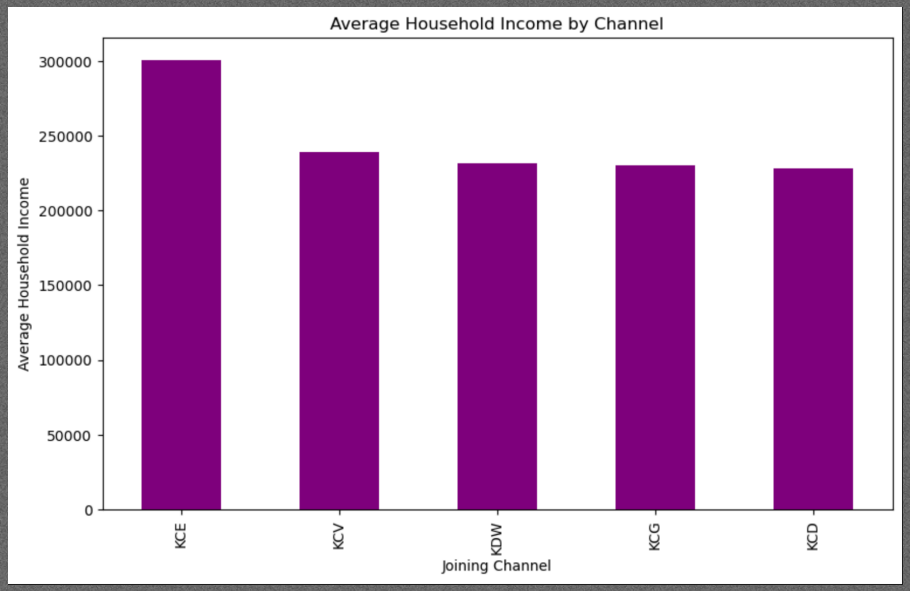
On the left is a graph that shows the countries with the strongest loyalty



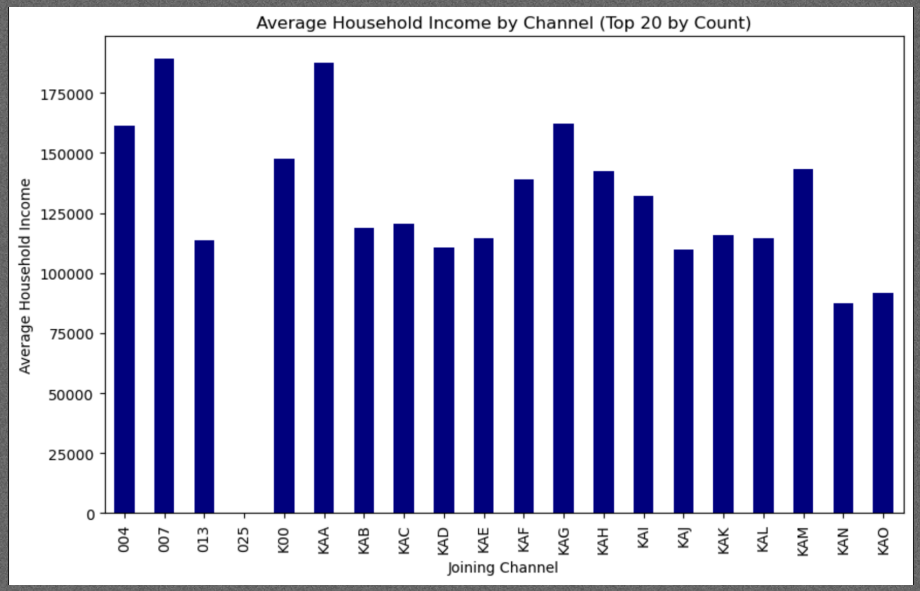
On the right is a graph that shows the top 5 countries that comprise the most clientele. With that being said, all five of these countries have less average seniority.

Hypothesis #3

If a channel recruits higher households of income, then we should target these channels more frequently



The graph on the left shows the Joining Channels that have attracted the highest Household incomes



The graph on the right shows the 20 Joining Channels that are most frequentlt used to attract clients. None of these channels, however, attract household income as high as the top 5 channels.

Hypothesis #4

If we target a more diverse market and adhere to different perspectives, then we can enable much more room for growth in sales and potentially cross sales.

Currently, we know that …

99.9% of customers are NOT current employees

99.1% of customers live in the same region as their bank

95% of customers were born in the same place they bank

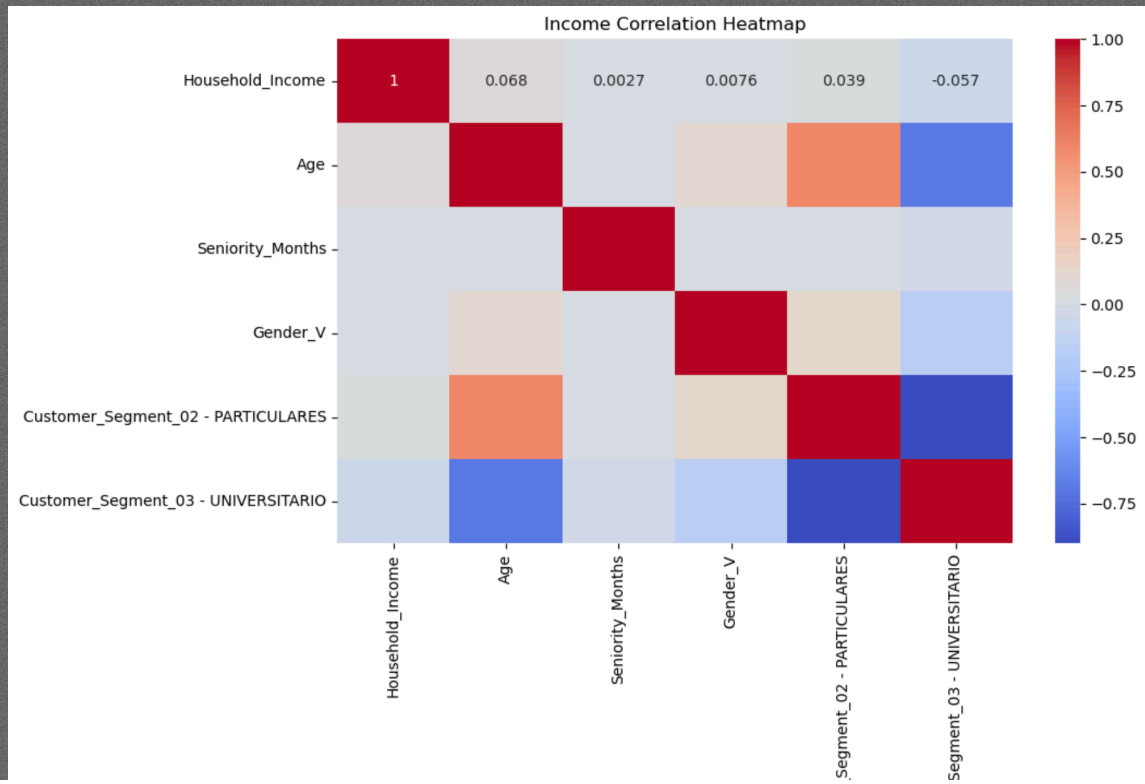
What would happen if we were to diversify our customer base?

Hypothesis #5

If we target households with higher levels of annual income, then they are more likely to purchase more than someone with less income

This heatmap shows the correlation between Age, Customer Seniority (measured in months), Gender (Female reference point), and Education in relationship to Household Income

None of these metrics show a significant relationship to Household Income. Therefore, in pursuit of higher income individuals, we cannot target–for example– college graduates or individuals in their late 50's… it would have little statistical benefit.

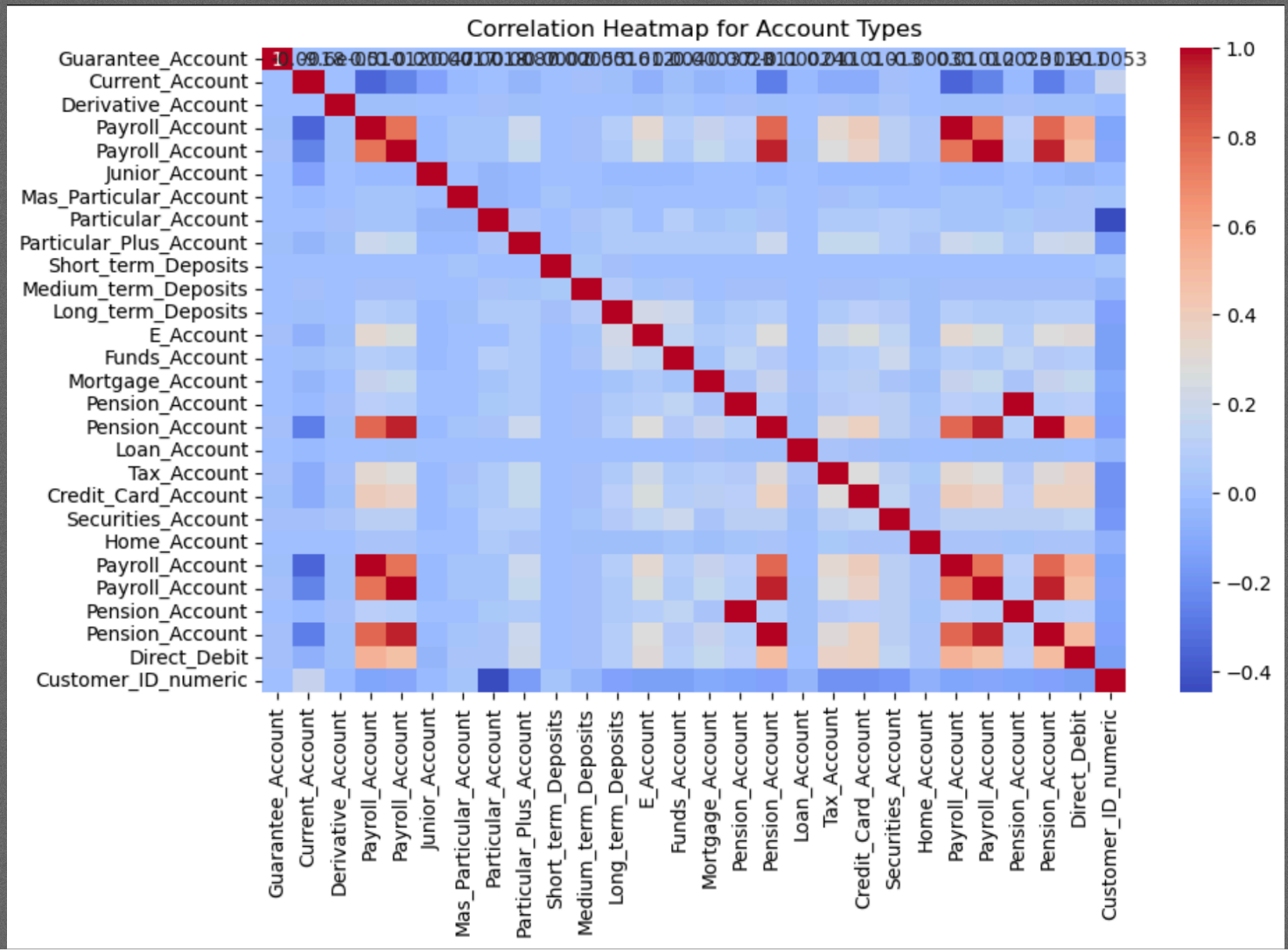


Hypothesis #6

If an individual account has a consistent correlation to another account, then it is more likely to cross-sell when you target that particular account

In this display, ‘Payroll Account’, ‘Pension Account’, and ‘Direct Debit’ Accounts appear to be the most strongly correlated.

Therefore, in cultivating a marketing strategy, a company should direct customers towards these accounts should they hope to improve cross-selling projections.

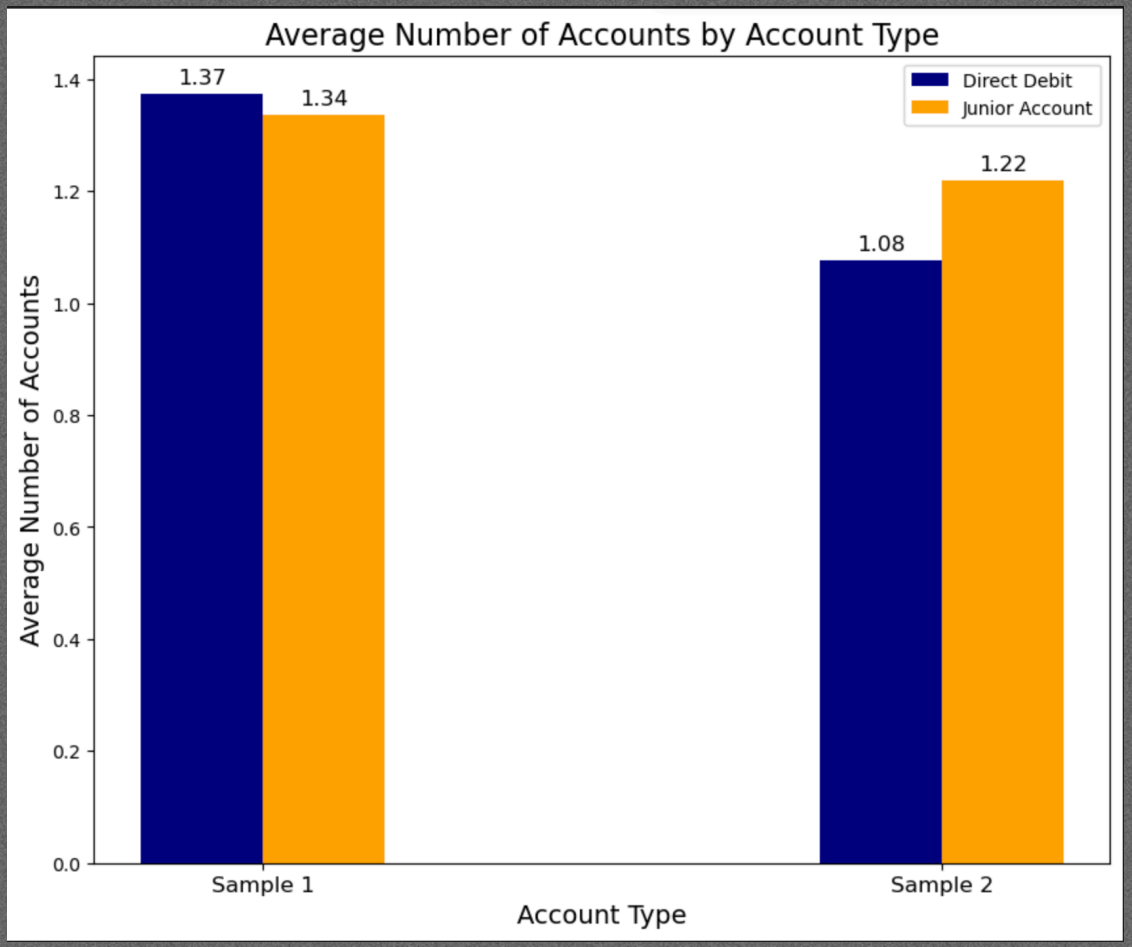


Hypothesis #6.5

Debunking the last slide

To test this notion–that marketing should focus on accounts that are more correlated to other accounts rather than those that are purchased individually– we compared the average number of accounts for customers who had: a) a Direct Debit account and b) a Junior Account. For reference, the Junior Account was amongst the least correlated accounts of the 24.

In the first sample of data we looked at, labelled 'Sample 1', it shows that the average number of accounts a customer will have GIVEN that one of them is a Direct Debit account, is only 0.04% higher than that of a customer. In the second sample of data–a slightly smaller set– the average number of accounts is .14% higher for someone with a Junior Account.



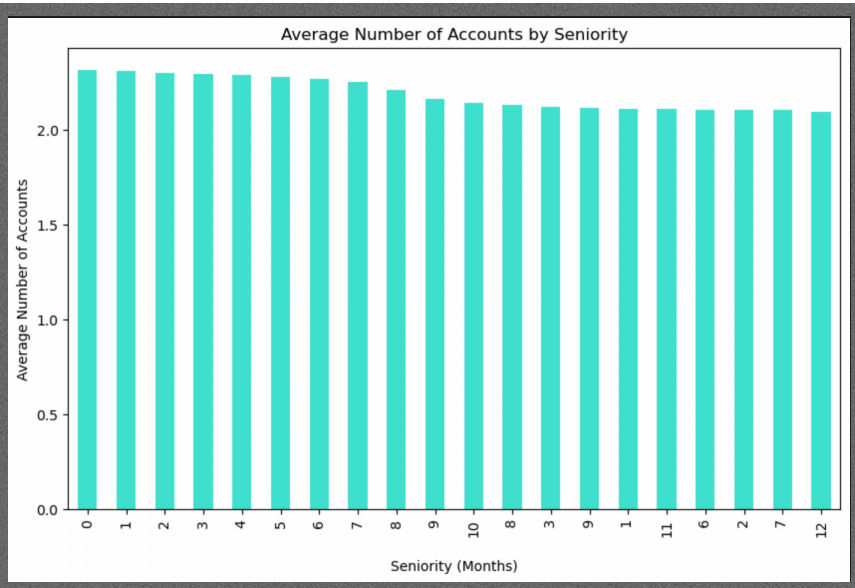
Hypthosis #7

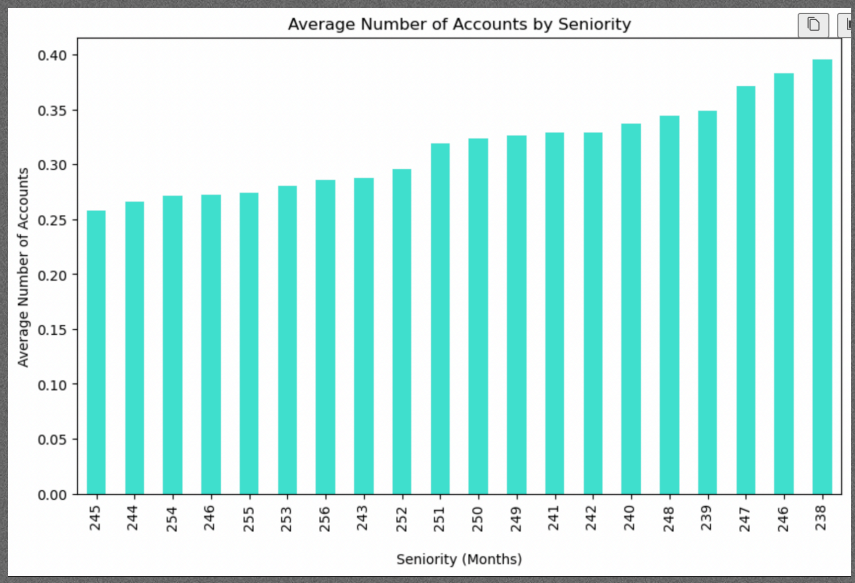
If we prioritize cross-selling for first-time sales, that will be more effective than waiting.

The graphs on the right show the average number of accounts a customer has given their seniority– the number of months they have spent as a customer. Amongst the highest of averages are customers who have spent 0-10 months with the credit union. This means at least one of two things.

1. Customers are more inclined to purchase several accounts within the first year of membership

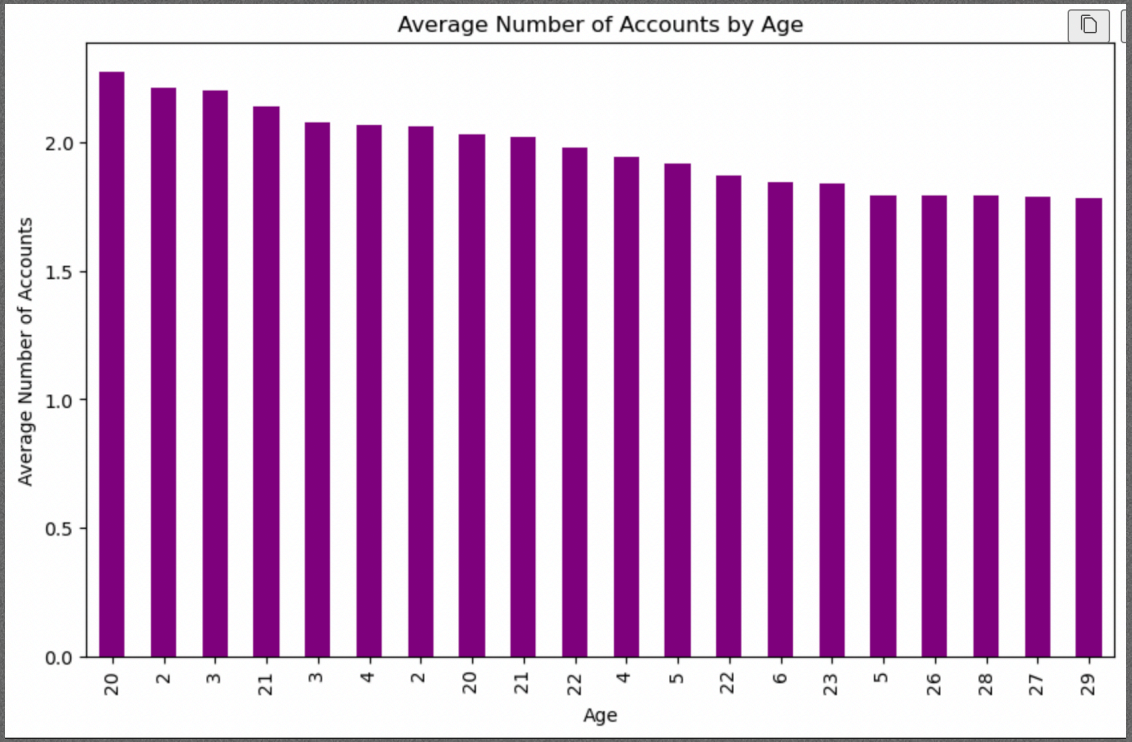
2. Customers are more likely to cancel accounts as their seniority grows.





Hypothesis #7

Revisit



This graph demonstrates a similar concept. It shows the top 20 ages linked to having the highest average number of accounts. The distribution ranges from 2 to 28 years old.

At first glance, we may want to consider 2-6-year-olds as outliers, when in reality, parents are likely setting up multiple accounts for their children at the youngest age possible.

EDA Summary

- We just analyzed 7 individual hypotheses regarding strategies to improve cross-selling.
- These 7 hypotheses aimed to analyze on a deeper level the relationship between demographics like age and gender, customer retention and location analysis, resource manipulation, diversity, income, account correlation, and seniority to cross-selling success.
- Now let’s make some suggestions.

Recommendations

- Our clientele is largely comprised of relatively young individuals. Instead of shifting towards an older crowd, focus on improving customer loyalty in a younger demographic.
- Of the five countries that comprise the largest proportion of our customer base, none of them are among the most loyal. Spain for example, accounts for greater than 95% of our clientele. Instead of searching for new clients, work on improving customer retention through markets that you have already attracted.
- Currently, out of the 20 Channels that are used most frequently to join as customers, none of them are among the top 5 in recruiting the highest-income individuals. We should use more frequently the channels that produce customers who would be willing to pay more money for our services.
- Our customer base is not inclusive or diverse. To improve cross-selling, we should target people from different backgrounds with different perspectives and ambitions.
- Ideally, we should target more households with higher incomes. Given that there is almost no link between age, gender, education, seniority and household income, it would be difficult to derive such a path to do so.
- Some accounts are more heavily correlated to the purchase of another account than others. Despite an example that proved that this correlation doesn't make a significant difference, it is still advisable to target customers towards accounts that are more likely to be purchased with another account.
- Prioritize cross-selling in the first sale. Customers are more likely to purchase multiple accounts at the beginning of a subscription as opposed to later, where they are more likely to cancel subscriptions. Also, aim to improve customer retention and avoid having senior clients cancel their subscriptions.

Thank you