Name: Deepthika Shiwani Muralikrishnan, Payal Upadhyay, Abida S Bhatti

Group Name: Team Data Analysts

Email: deepthikamk@gmail.com , payalupadhyay456@gmail.com , sarwarabida123@gmail.com

Country: India, South Africa, USA

College/Company: Data Glacier

Specialization: Data Analyst

Github Repository Link: https://github.com/deepthikashiwani/Cross-selling-Recommendation-How-to-increase-cross-selling-of-Banking-Products

Notebook link : https://github.com/deepthikashiwani/Cross-selling-Recommendation-How-to-increase-cross-selling-of-Banking-Products/blob/main/Data%20Cleaning%20Week%2010.ipynb

PROBLEM DESCRIPTION

XYZ credit union in Latin America is performing very well in selling the Banking products (eg: Credit card, deposit account, retirement account, safe deposit box etc) but their existing customer is not buying more than 1 product which means bank is not performing good in cross selling (Bank is not able to sell their other offerings to existing customer). XYZ Credit Union decided to approach ABC analytics to solve their problem. ABC company came up with a framework which will be utilizing machine learning algorithm in the core to increase cross selling. But as a data analyst you need to inspect the data and suggest what action bank can take to increase cross selling (without using ML)
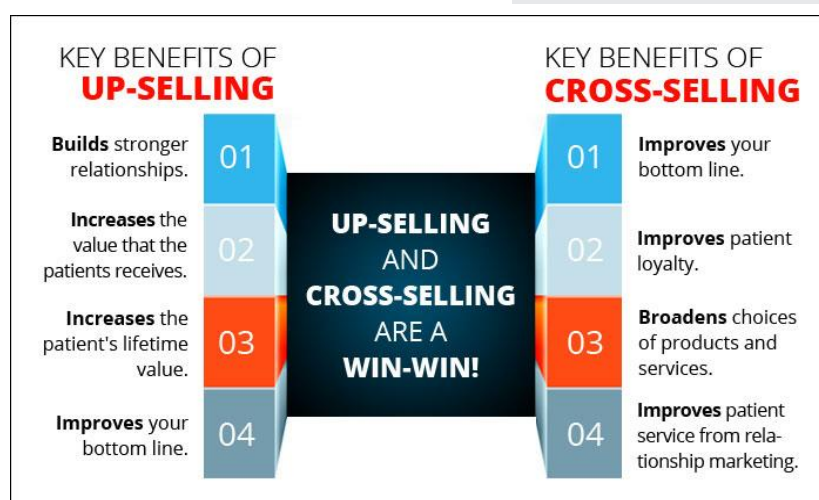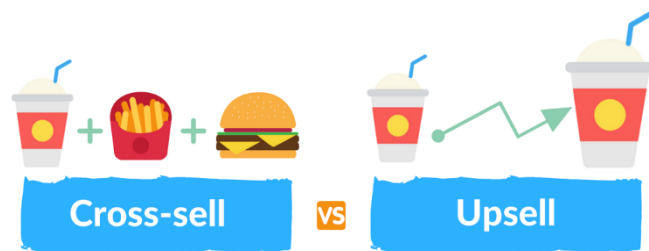
EDA PERFORMED ON DATA

The following approach was used to perform EDA on the dataset

• Understand and clean the data

• Visualize and analyze the data

• Summarize the data

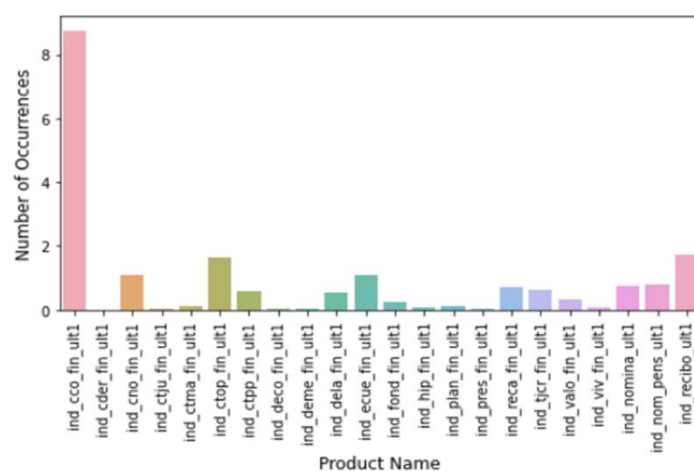• Recommend a solution/actions to increase cross-selling

We have also tried a few automated EDA libraries like DTale, Pandas Profiling and autoviz which provided similar results. The respective files are present in the github repository.
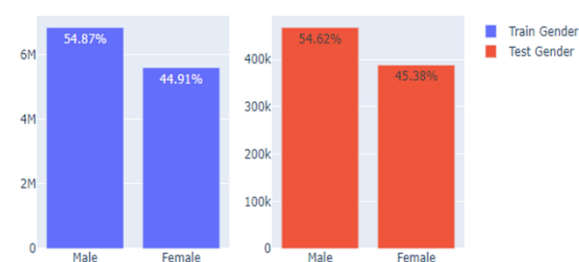
The most popular products appear to be:

❑ ind_cco_fin_ult1 (current account)

❑ ind_ctop_fin_ult1 (particular account)

❑ ind_recibo_ult1 (direct debt)

There are a lot of products that are not popular. Some of them include:

❑ ind_cder_fin_ult1 (Derivada Account)

❑ ind_ctju_fin_ult1 (Junior Account)
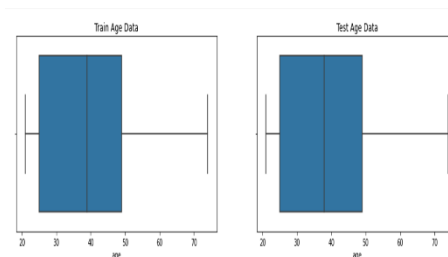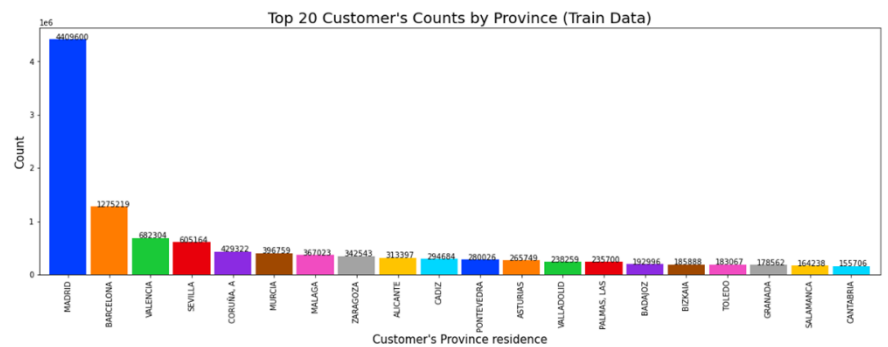
❑ ind_deco_fin_ult1 (Short-term deposits)

Here, we can observe that More than half of the bank's customers are male

Train/Test gender column

All customers are aged between 20 and 80 and this is done by removing outliers (age<19 and age>80)
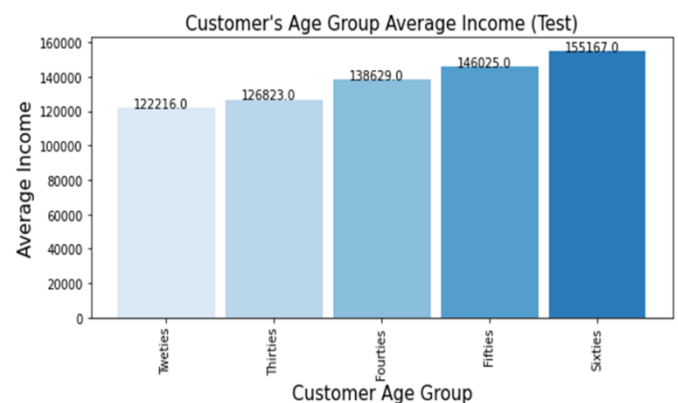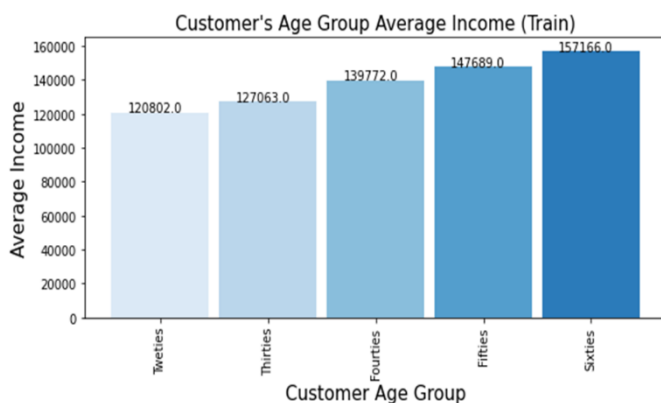
It is observed that most customers are from:

- ❑ Madrid
- ❑ Barcelona
- ❑ Valencia


Top 20 Customer's Counts by Province (Train Data)

Provinces with the least customers are from:

- ❑ Cantabria
- ❑ Salamanca
- ❑ Granada

Top 20 Customer's Counts by Province (Test Data)

Customer's Age Group Average Income (Train)

Customer's Age Group Average Income (Test)
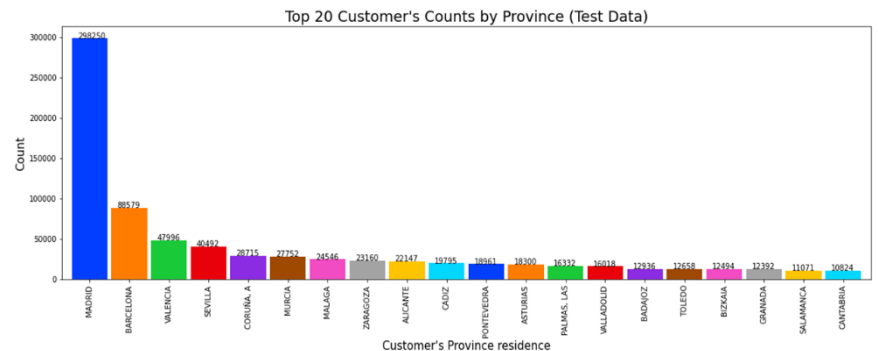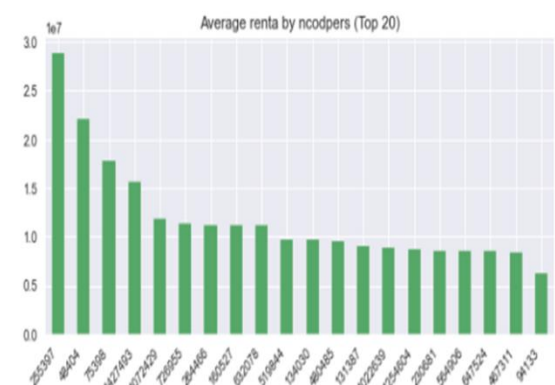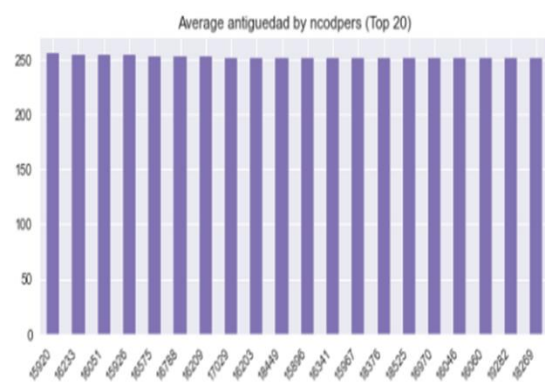
We can see that customers in their sixties and fifties earn the most income. We speculate that this high amount is possible due to them earning from personal businesses as well as pension. Young customers might not be earning as much due to the lack of experience in their field of work.

Customer's Count by Seniority (Test Data) / Customer's Count by Seniority (Train Data)
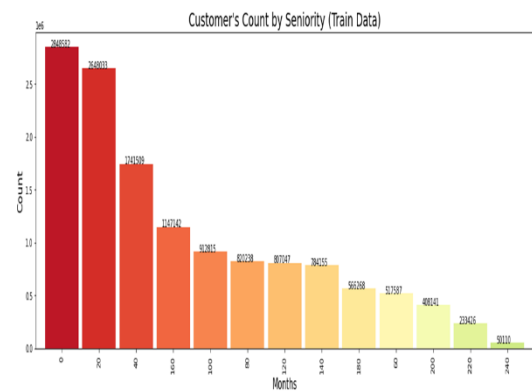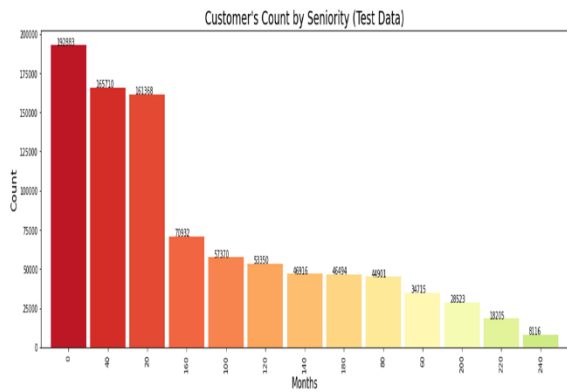
It is clear that there are a lot of new customers but the bank cannot retain their existing customers because the customer count decreases as months and years pass by.

antiguedad | Distplot / antiguedad | Boxplot
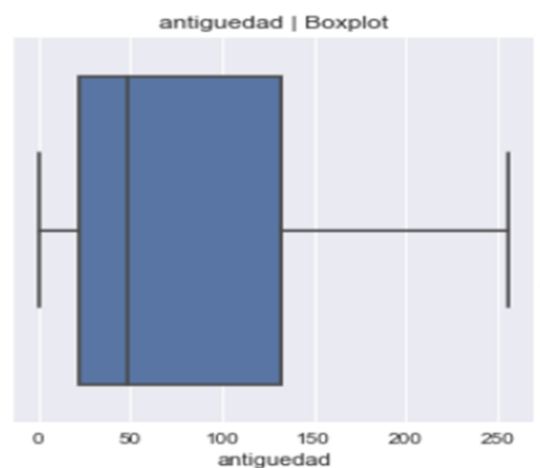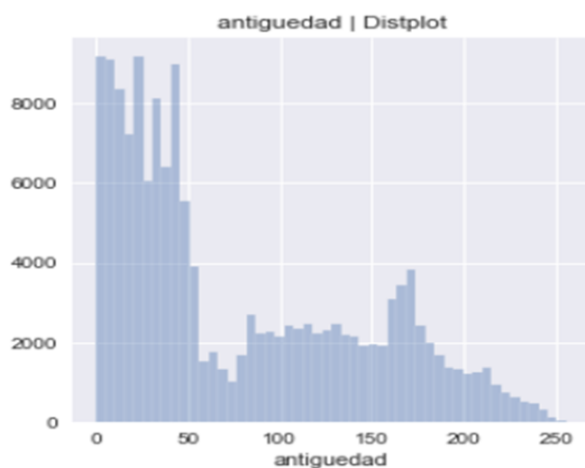
Customer seniority ranges from as low as 0 months to as high as 260 months (21.6 years approx.)

Average antiguedad by ncodpers (Top 20) / Average renta by ncodpers (Top 20)

All top 20 customers have been clients of the bank for at least 250 months (20+ years). The top 20 customers have a household gross income ranging from 0.7 million to 2.8 million.
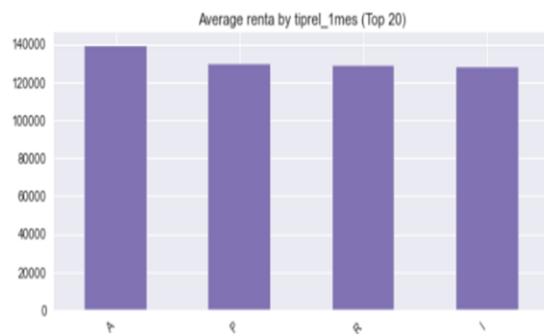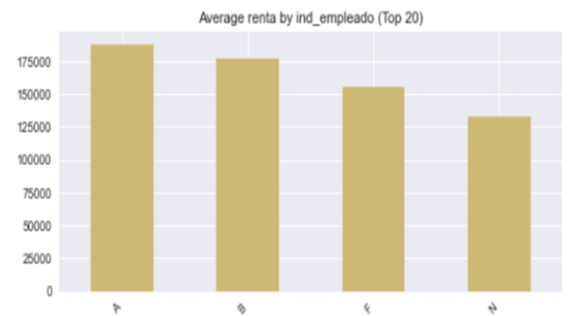
The most senior customers of the bank have an employee index of:
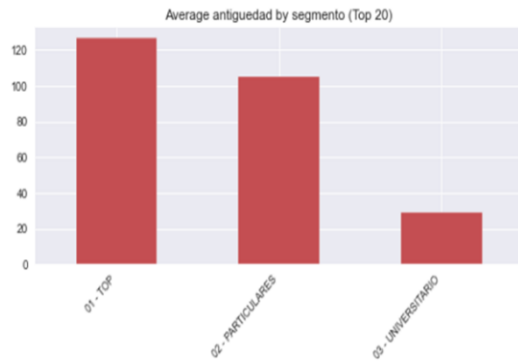
F: Filial and B: Ex-employed
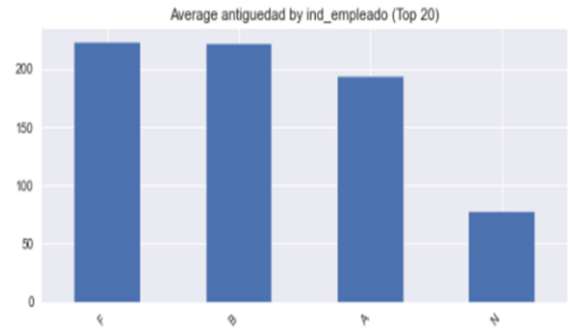
Customers earning the most income have employee index of:

A: Active and B: Ex-employed

The most senior customers are Active members. These are also the ones earning the highest income. Former, Inactive and Potential customers earn the same income in general.

Average antiguedad by segmento (Top 20)

Average renta by segmento (Top 20)

The top/VIP clients are long term clients earning the most income. College graduates earn the least income and are also short term customers for the bank.

Average antiguedad by ind_cco_fin_ult1 (Top 20)

Average renta by ind_cco_fin_ult1 (Top 20)
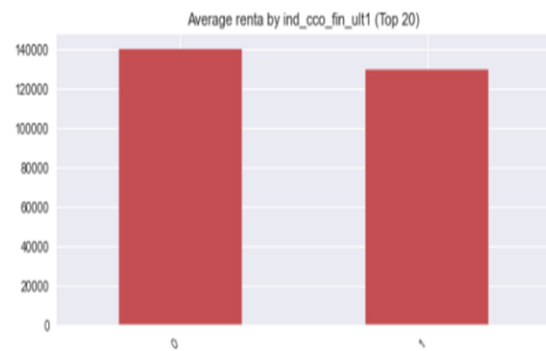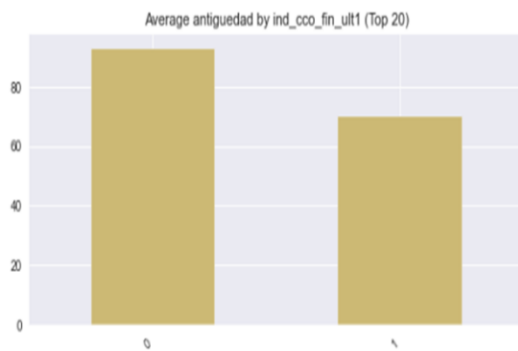
The most popular product (current account) offered by the bank has more inactive members than active member. These inactive members earn a higher income than the active members.

Average antiguedad by ind_ctop_fin_ult1 (Top 20)

Average renta by ind_ctop_fin_ult1 (Top 20)
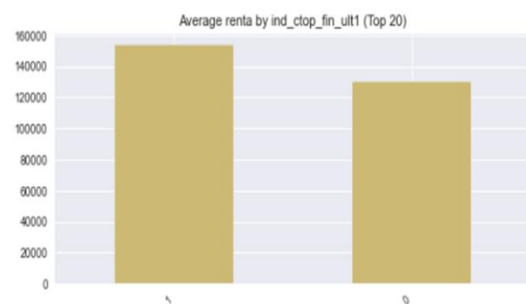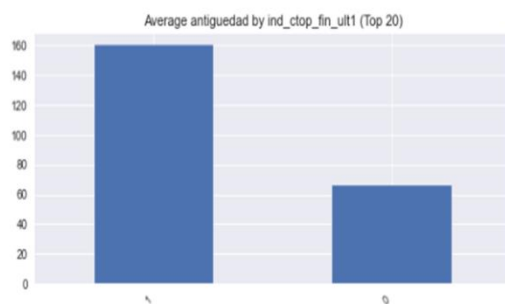
The next most popular product (particular account) has many more active members earning a higher income than the inactive members.

Average antiguedad by ind_recibo_ult1 (Top 20)

Average renta by ind_recibo_ult1 (Top 20)
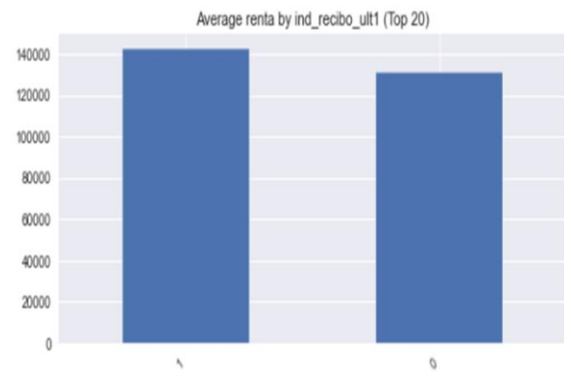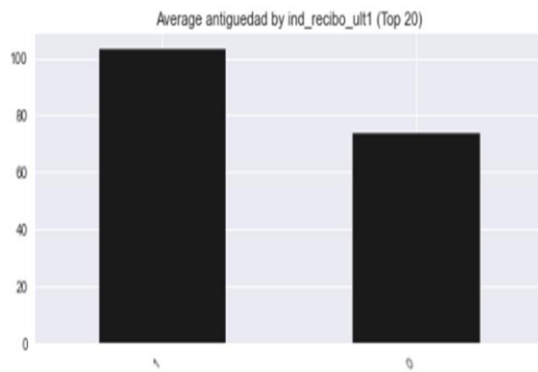
The third most popular product(direct debt) has more active clients than inactive clients. Inactive clients earn a little less than the active clients
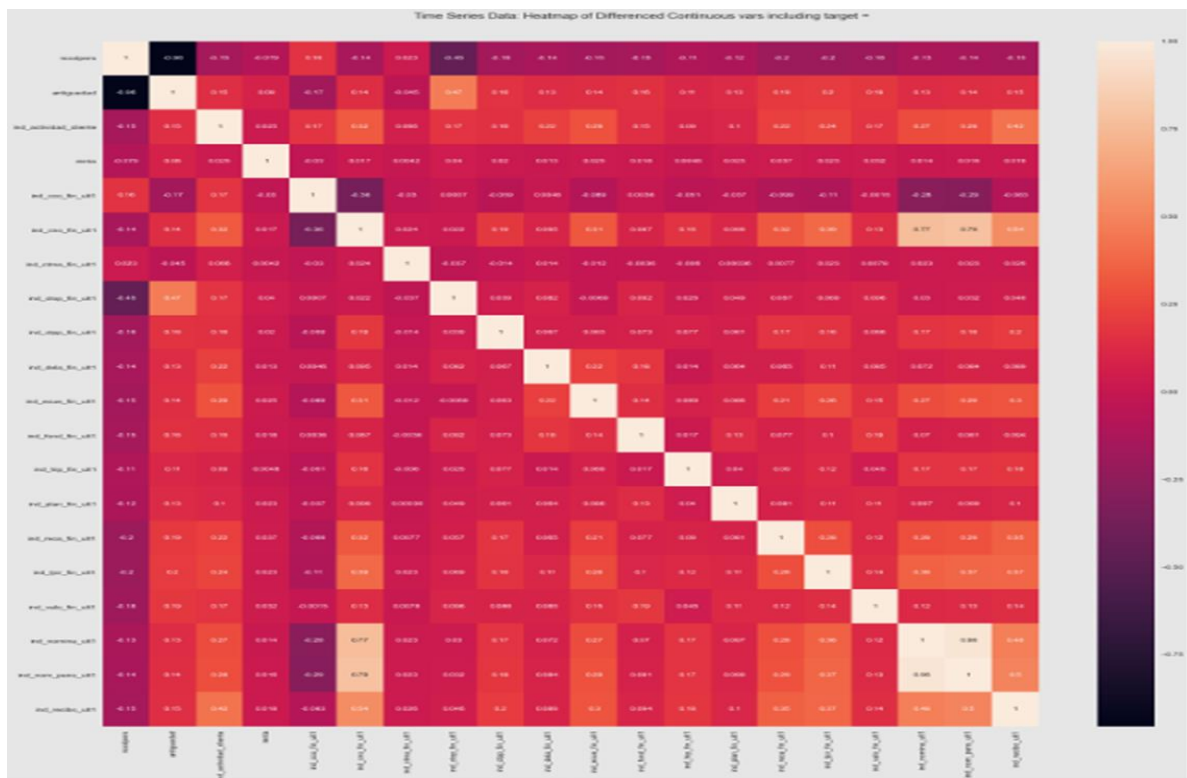
Time Series Data: Heatmap of Differenced Continuous vars including target

The highest correlation exists between ind_nom_pens_ult1 (pensions) and ind_nomina_ult1(payroll).

The pension and payroll also significantly correlated with ind_cno_fin_ult1(payroll account).

EDA SUMMARY

- The three most popular accounts are:
  - ➤ ind_cco_fin_ult1 (current account)
  - ➤ ind_ctop_fin_ult1 (particular account)
  - ➤ ind_recibo_ult1 (direct debt)
- Most customers are male.
- Customers in their sixties earn the most income
- Most customers are from:
  - ➤ Madrid
  - ➤ Barcelona
  - ➤ Valencia
- .The bank has a lot of new customers but fail to retain their existing customers
- College graduates earn the least income and are short term customers.
- The most popular product has more inactive members that earn a higher average income than active members.
- The highest correlation exists between ind_nom_pens_ult1 (pensions) and ind_nomina_ult1(payroll).

FINAL RECOMMENDATION

In order to encourage existing customers to invest in more products the bank can:
- Choose bank service to test for Cross selling as dependent variables.
- Choose best matrices in dataset that can be used in machine learning model.
- Once your matrices are ready, create different machine learning models and choose your model based on performance matrix.
- Provide more interest the longer customers keep their account active. This helps retain customers.
- Depending on age group, offer appropriate gifts/vouchers on opening a combination of accounts such as:
  - ➤ Ages 19-25 : Amazon gift
  - ➤ Ages 26-45 : Voucher for supermarket/essential store
  - ➤ Ages >45 : A free massage session etc
- Keep in touch with customers and inform them of new products/upgrades with their benefits.
- Offer college graduates future benefits once they start earning above a certain threshold.
- Lower monthly fees for those with multiple accounts.
- Personalize offers based on customers' past history and through the use of big data.
- The bank has to finance the benefits offered.