



Data Glacier Group Project



YUHONG CHEN
HAOYUE CHANG
GIFTY OSEI

Group Name: DATA ALCHEMISTS

Team Members:

Name: Yuhong Chen

Email: ryechenn@gmail.com

Country: Canada

College/Company: McGill University

Name: Haoyue Chang

Email: kerrychy1215@gmail.com

Country: USA

College/Company: Northeastern University

Name: Gifty Osei

Email: gifty18osei@gmail.com

Country: USA

College/Company: Montana State University

Specialization: Data Analyst

CONTENT

Problem Description

Data Information

Exploratory Data
Analysis

Recommended Modeling
& Evaluation

Project Introduction

- We are determined to help XYZ Bank improve its cross-selling strategies and enhance customer engagement. The bank offers a wide array of financial products and services, including savings accounts, credit cards, mortgages, loans, and investment options. However, we've observed that many of our customers have limited product adoption and aren't fully utilizing the range of services available to them.

- To tackle this challenge head-on, we plan to implement customer segmentation techniques to gain deeper insights into our customer base. By dividing the customers into distinct groups based on their demographics, financial behavior, and product usage patterns, we hope to identify specific customer segments that are more likely to use products and services. Armed with this valuable information, we aim to create personalized marketing strategies and tailored cross-selling initiatives to boost customer satisfaction and encourage higher product adoption. As part of our data analysis team, the objective is to thoroughly analyze the extensive customer dataset provided by XYZ Bank and conduct a comprehensive customer segmentation analysis. The dataset includes detailed information about each customer, such as age, gender, income, transaction history, product holdings, and tenure with our bank.

Cross-Selling



Data Overview

fecha_datos The table is partitioned for this column

ncodpers Customer code

ind_empleado Employee index: A active, B ex employed, F filial, N not employee, P pasive

pais_residencia Customer's Country residence

sexo Customer's sex

age Age

fecha_alta The date in which the customer became as the first holder of a contract in the bank

ind_nuevo New customer Index. 1 if the customer registered in the last 6 months.

antiguedad Customer seniority (in months)

indrel 1 (First/Primary), 99 (Primary customer during the month but not at the end of the month)

ult_fec_cli_1t Last date as primary customer (if he isn't at the end of the month)

indrel_1mes Customer type at the beginning of the month ,1 (First/Primary customer), 2 (co-owner),P (Potential),3 (former primary), 4(former co-owner)

tiprel_1mes Customer relation type at the beginning of the month, A (active), I (inactive), P (former customer),R (Potential)

indresi Residence index (S (Yes) or N (No) if the residence country is the same than the bank country)

indext Foreigner index (S (Yes) or N (No) if the customer's birth country is different than the bank country)

conyuemp Spouse index. 1 if the customer is spouse of an employee

canal_entrada channel used by the customer to join

indfall Deceased index. N/S

tipodom Address type. 1, primary address

cod_prov Province code (customer's address)

nomprov Province name

ind_actividad_cliente Activity index (1, active customer; 0, inactive customer)

renta Gross income of the household

segmento segmentation: 01 - VIP, 02 - Individuals 03

- college graduated

ind_ahor_fin_ult1 Saving Account

ind_aval_fin_ult1 Guarantees

ind_cco_fin_ult1 Current Accounts

ind_cder_fin_ult1 Derivada Account

ind_cno_fin_ult1 Payroll Account

ind_ctju_fin_ult1 Junior Account

ind_ctma_fin_ult1 Más particular Account

ind_ctop_fin_ult1 particular Account

ind_ctpp_fin_ult1 particular Plus Account

ind_deco_fin_ult1 Short-term deposits

ind_deme_fin_ult1 Medium-term deposits

ind_dela_fin_ult1 Long-term deposits

ind_ecue_fin_ult1 e-account

ind_fond_fin_ult1 Funds

ind_hip_fin_ult1 Mortgage

ind_plan_fin_ult1 Pensions

ind_pres_fin_ult1 Loans

ind_reca_fin_ult1 Taxes

ind_tjcr_fin_ult1 Credit Card

ind_valo_fin_ult1 Securities

ind_viv_fin_ult1 Home Account

ind_nomina_ult1 Payroll

ind_nom_pens_ult1 Pensions

ind_recibo_ult1 Direct Debit

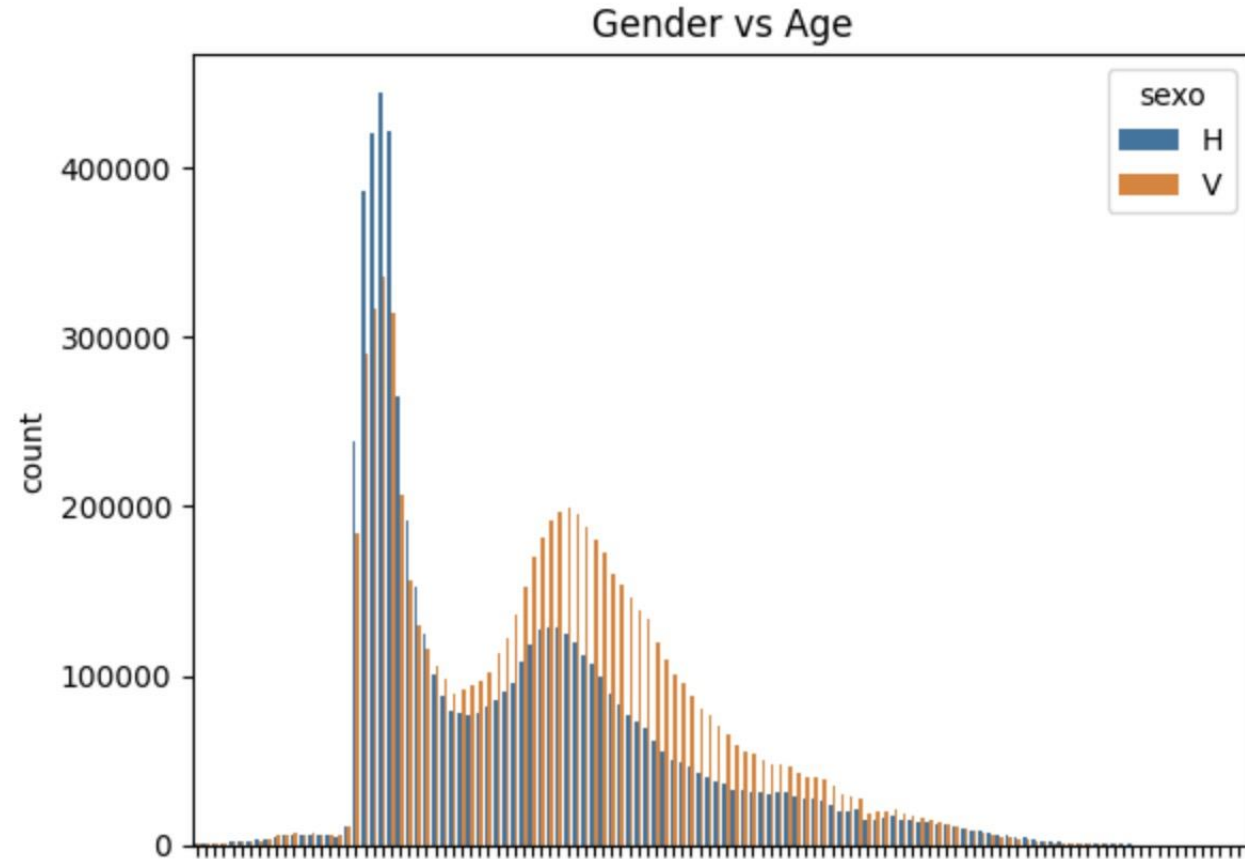
Descriptive Statistics

- Table 1 shows the summary statistics of some selected variables in our dataset.
- The selection was at random.

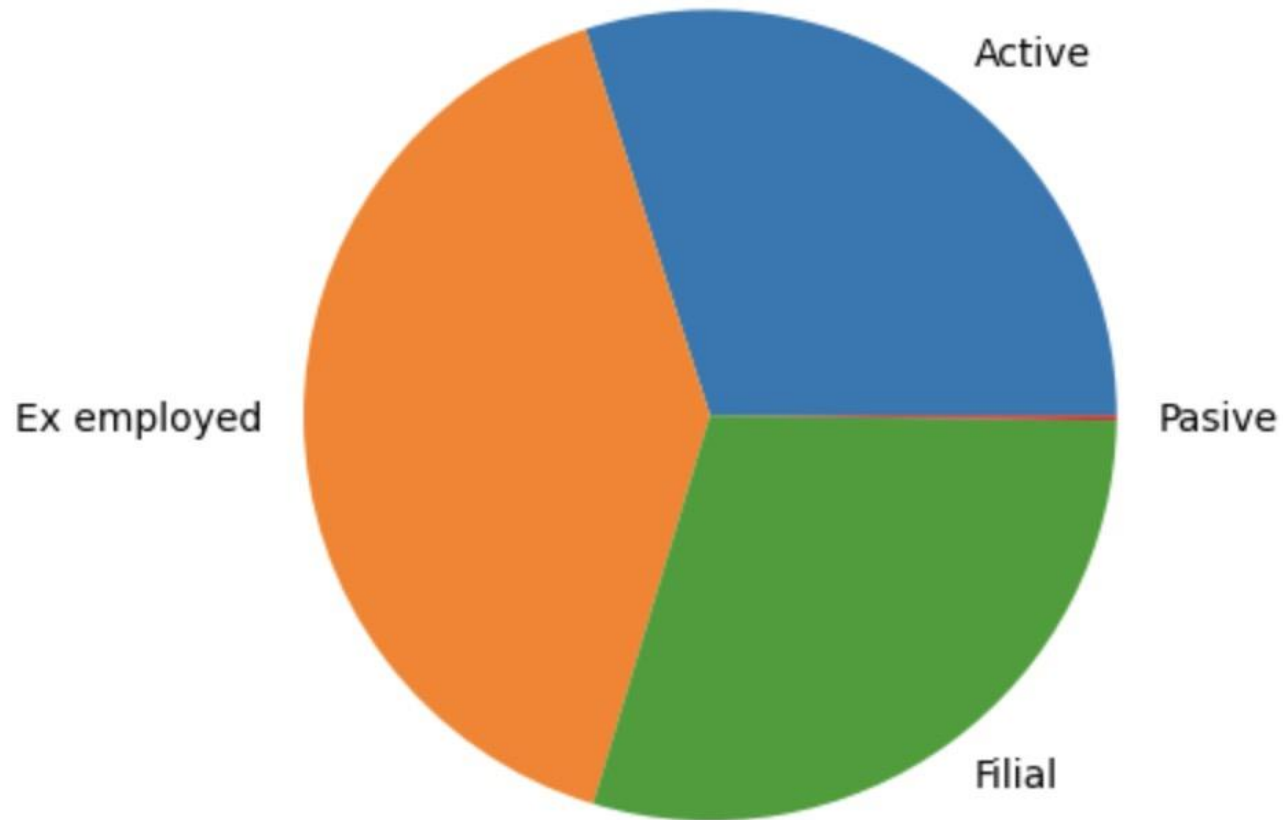
	ncodpers	ind_nuevo	indrel	tipodom	cod_prov	ind_actividad_cliente	renta
count	13647309.0	13647309.0	13647309.0	13647309.0	13647309.0	13647309.0	13647309.0
mean	834904.0	0.0	1.0	1.0	27.0	0.0	134254.0
std	431565.0	0.0	4.0	0.0	13.0	0.0	205659.0
min	15889.0	0.0	1.0	1.0	1.0	0.0	1203.0
25%	452813.0	0.0	1.0	1.0	15.0	0.0	76437.0
50%	931893.0	0.0	1.0	1.0	28.0	0.0	124680.0
75%	1199286.0	0.0	1.0	1.0	34.0	1.0	137452.0
max	1553689.0	1.0	99.0	1.0	52.0	1.0	28894396.0

EDA

- This chart plots the relationship between Gender and Age of customers.
- We could see the count of customers reaches highest point in younger age range(20) in 2 genders. Higher in H sexo type.
- And for another account which is second highest in middle age range(50) but is higher in another sexo type V.



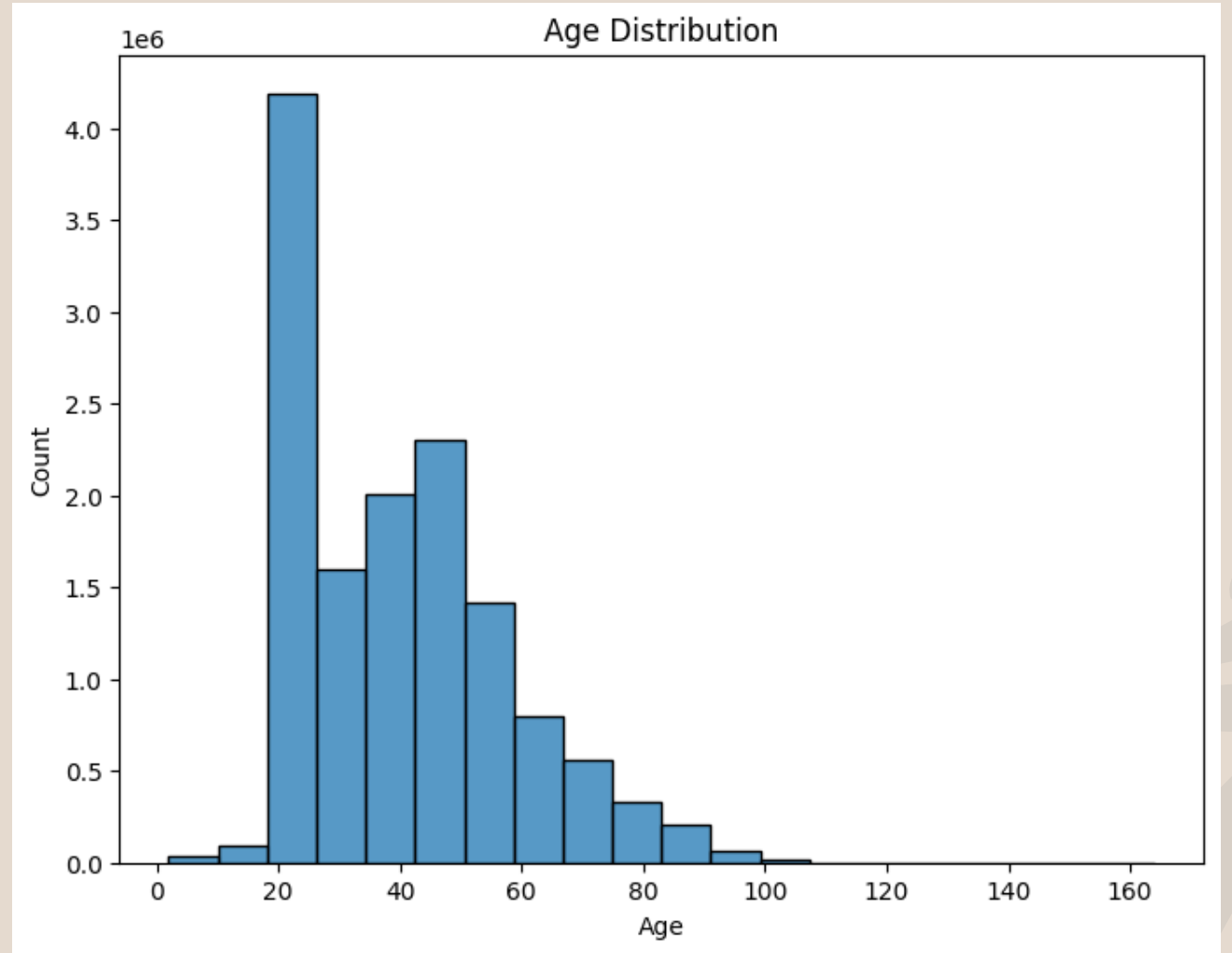
EDA



- The pie chart shows the percentage of employment index distribution.
- Unemployed ranks the 1st and then the Ex employed. Pasive is least here in the plot.
- So we could see that the employment situation is not positive for the unemployment percentage and Ex employed count.

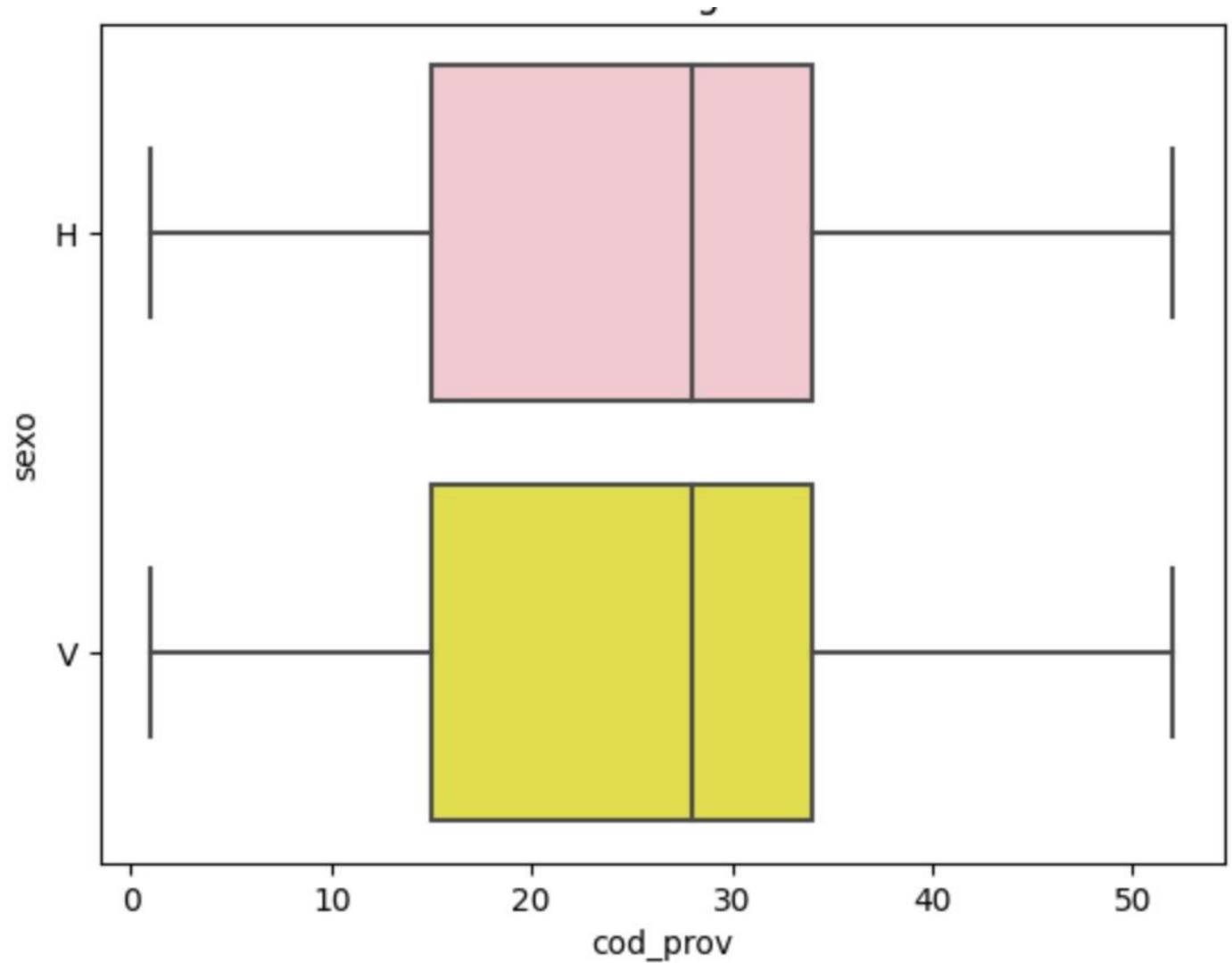
EDA

- We can see from the Age distribution plot that the majority of customers are between ages 20 and 50 meaning the products are patronized by the working force.



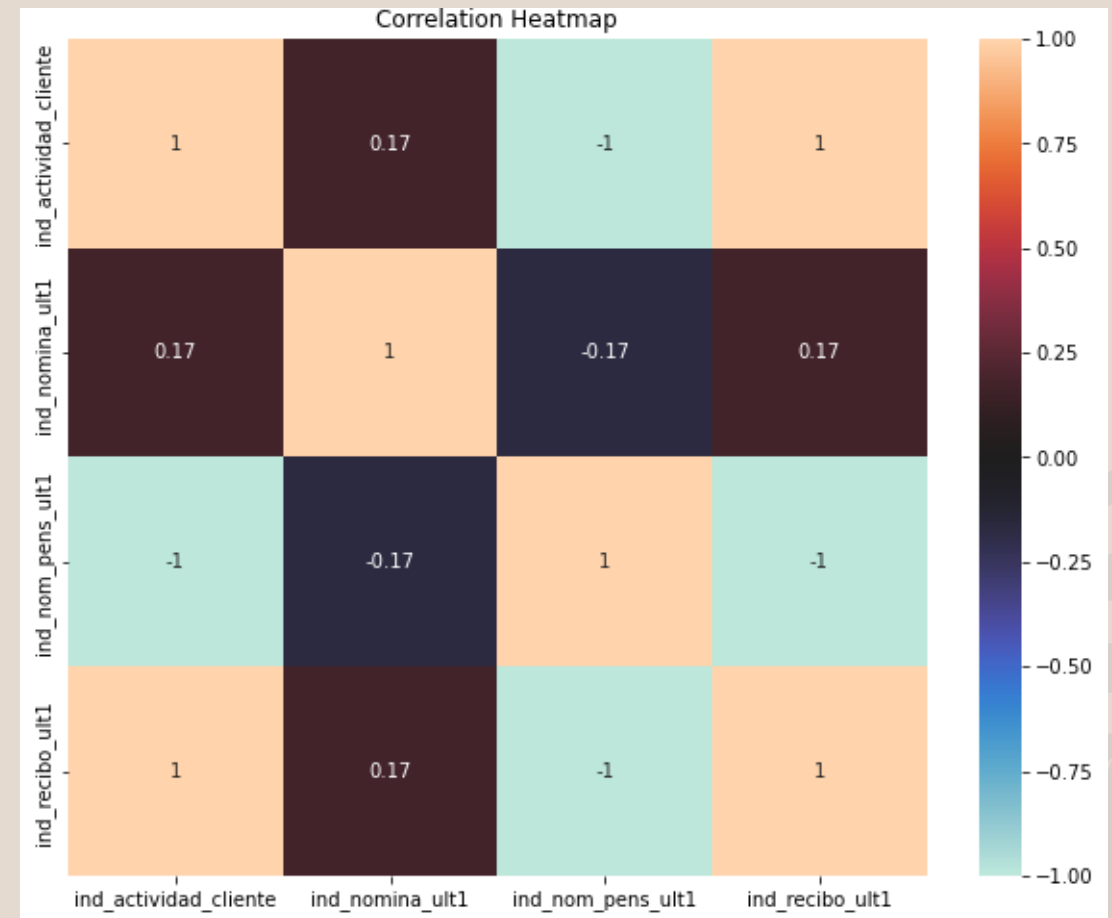
EDA

- We can see the Province code and gender distribution from this plot.
- For different gender, Province code (customer's address) are spreading nearly the same.



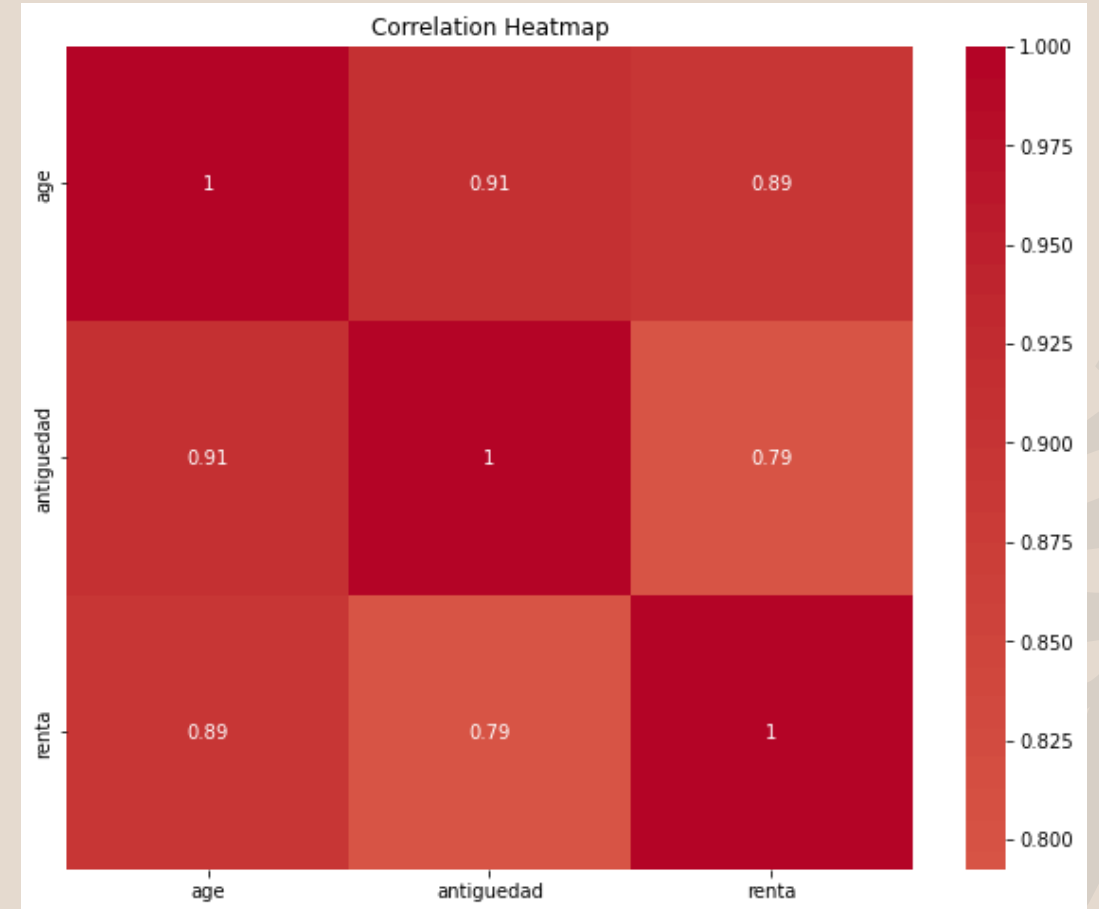
EDA

- The x-axis and y-axis display the variable of `ind_actividad_cliente`, `ind_nomina_ult1`, `ind_nom_pens_ult1`, and `ind_recibo_ult1`.
- "`ind_actividad_cliente`" and "`ind_nomina_ult1`": The correlation coefficient between "`ind_actividad_cliente`" and "`ind_nomina_ult1`" is close to zero. This suggests that there is little to no linear relationship between a customer's activity index and whether they received a payroll payment.
- "`ind_actividad_cliente`" and "`ind_nom_pens_ult1`": The correlation between "`ind_actividad_cliente`" and "`ind_nom_pens_ult1`" is -1, indicating a negative linear association between a customer's activity index and whether they received a pension payment.
- "`ind_actividad_cliente`" and "`ind_recibo_ult1`": The correlation between "`ind_actividad_cliente`" and "`ind_recibo_ult1`" (direct debit) indicates that there is a strong linear relationship between a customer's activity index and their participation in direct debit transactions.
- "`ind_nomina_ult1`" and "`ind_nom_pens_ult1`": Depending on the sample data, you might observe a correlation coefficient close to zero. This could indicate a possible connection between receiving payroll and receiving pension payments, though further analysis would be needed to establish causation.



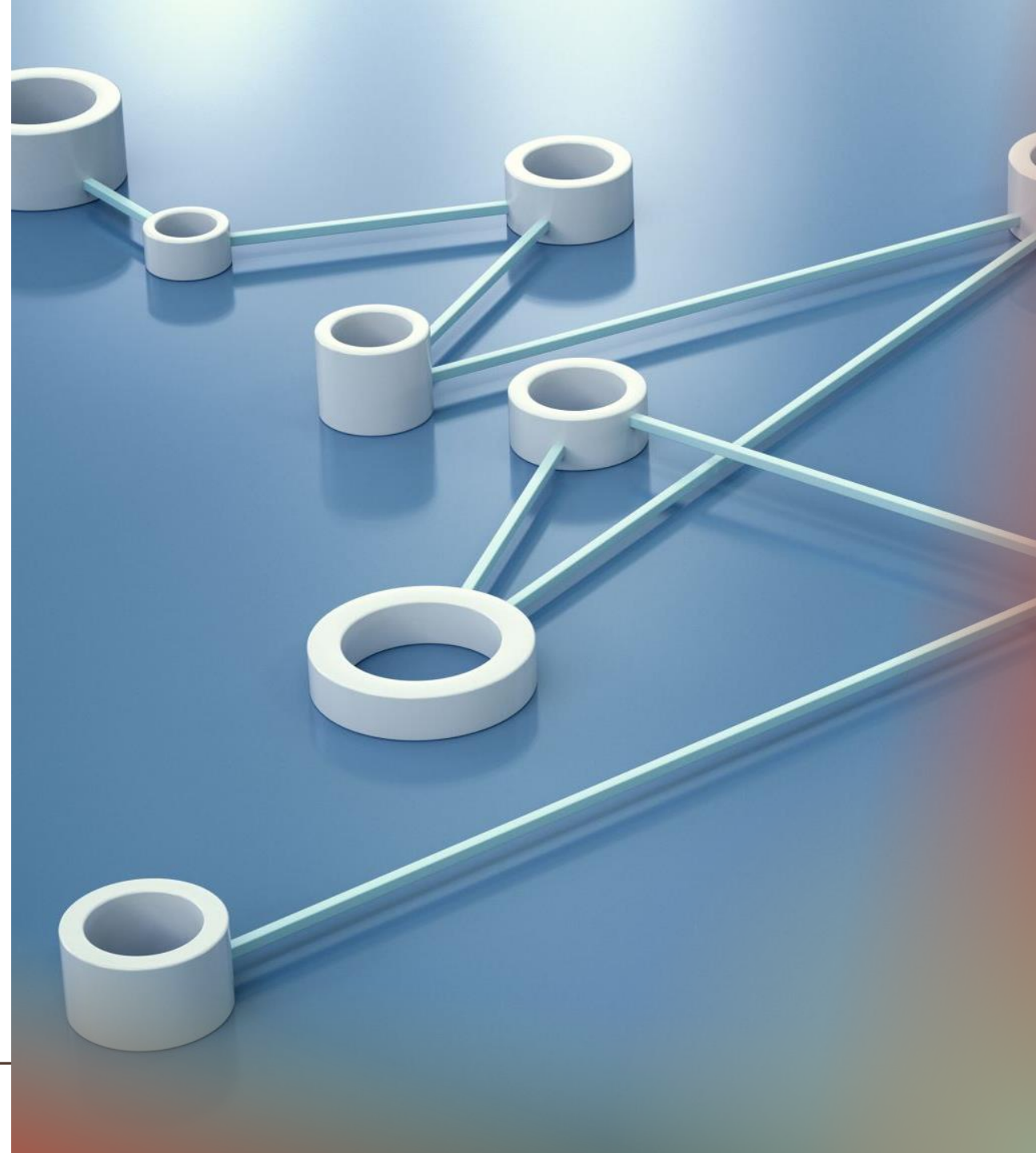
EDA

- The x-axis and y-axis display the variable names age , antigüedad, and renta. A positive correlation between "age" and "antigüedad" because older customers tend to have longer relationships with the bank. There has a strong correlation between "age" and "renta". And there is a strong correlation between "antigüedad" and "renta" because the length of time a customer has been with the bank may necessarily be strongly related to their income.



Future model recommendation

- In the future, we can use the customer segmentation model to conduct the analysis by using clustering algorithms like K-Means or hierarchical clustering to group customers based on using clustering algorithms like K-Means or hierarchical clustering to group customers based on their attributes. This could help identify different customer segments with distinct behaviors, and preferences, attributes. This could help identify different customer segments with distinct behaviors and preferences.
- Meanwhile, we can use product recommendation model to build a recommendation system using collaborative filtering or matrix factorization techniques to suggest financial products to customers based on their historical interactions and preferences.



THANK YOU!

Data Alchemist

