



Data Glacier

Your Deep Learning Partner

Cross Selling Project

Virtual Internship

Vishnu Priya Malchetti

Project	Cross Selling
Submitted By	Vishnu Priya Malchetti
Email	Vishnupriyam.sapbasis@gmail.com
Country	Ireland
Specialization	Data Analyst
Internship Batch	LISUM41



Data Glacier

Your Deep Learning Partner

Agenda

- **Problem Statement**
- **Data Overview**
- **Business Understanding**
- **Data Understanding**
- **Exploratory Data Analysis (EDA)**
- **Model Recommendation**

Problem Statement

XYZ Bank aims to enhance customer engagement and improve cross-selling strategies, as many customers have limited product adoption despite the bank's diverse portfolio of financial products and services. To address this challenge, the data analysis team will leverage customer segmentation techniques to analyze the extensive customer dataset, which includes demographic, financial, and product usage information. By identifying distinct customer groups with unique behaviors and preferences, XYZ Bank seeks to develop personalized marketing strategies and tailored cross-selling initiatives to boost customer satisfaction and increase the adoption of financial products and services.



Data Glacier

Your Deep Learning Partner

Data Overview

fecha_dato 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)

Data Overview

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 (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 Addres type. 1, primary address

Data Overview

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

Business Understanding

This project's primary goal is to suggest what action the bank can take to increase cross selling by focusing on customer satisfaction and action.

1. Enhanced Customer Loyalty: Satisfying customers fosters loyalty, encouraging them to return for future purchases and strengthening their long-term relationship with our brand.
2. Improved Conversion Rates: Elevating the shopping experience leads to higher conversion rates, resulting in increased sales and revenue generation.
3. Reduced Cart Abandonment: Identifying and resolving factors contributing to cart abandonment will reduce bounce rates and enhance overall conversion rates.
4. Enhanced Brand Reputation: Delivering positive customer experiences stimulates word-of-mouth marketing, bolstering our brand reputation and attracting new customers.
5. Broaden choices of products and services: With more choices of products and services, different types of customers can choose and increase customer transaction elasticity

Data Understanding

Table 1 shows the summary statistics of some selected variables in our dataset.

	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

Exploratory Data Analysis (EDA)

```
[46]: df1.describe().round()
```

[46]:	ncodpers	ind_nuevo	indrel	tipodom	cod_prov	ind_actividad_cliente	renta	ind_ahor_fin_ult1	ind_aval_fin_ult1	ind_cco_fin_ult1	...	ind_hip_fin
count	13647309.0	13647309.0	13647309.0	13647309.0	13647309.0	13647309.0	13647309.0	13647309.0	13647309.0	13647309.0	...	13647
mean	834904.0	0.0	1.0	1.0	27.0	0.0	134254.0	0.0	0.0	1.0	...	
std	431565.0	0.0	4.0	0.0	13.0	0.0	205659.0	0.0	0.0	0.0	...	
min	15889.0	0.0	1.0	1.0	1.0	0.0	1203.0	0.0	0.0	0.0	...	
25%	452813.0	0.0	1.0	1.0	15.0	0.0	76437.0	0.0	0.0	0.0	...	
50%	931893.0	0.0	1.0	1.0	28.0	0.0	124680.0	0.0	0.0	1.0	...	
75%	1199286.0	0.0	1.0	1.0	34.0	1.0	137452.0	0.0	0.0	1.0	...	
max	1553689.0	1.0	99.0	1.0	52.0	1.0	28894396.0	1.0	1.0	1.0	...	

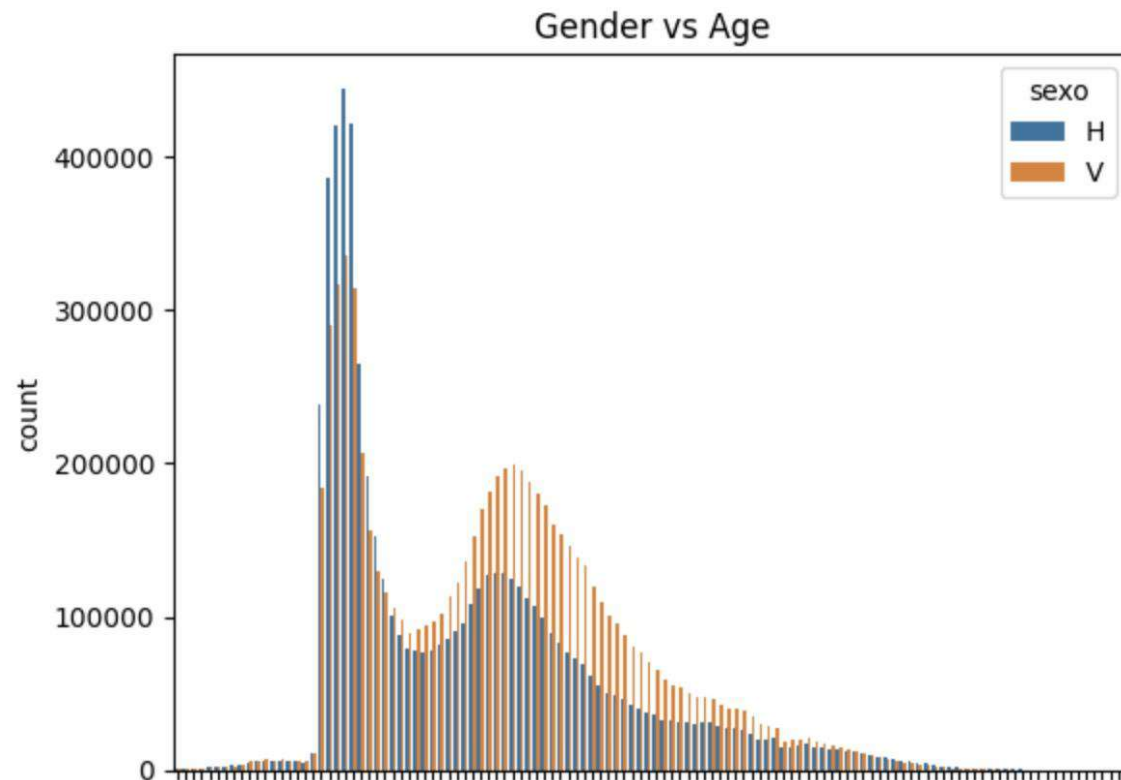
8 rows × 31 columns

Exploratory Data Analysis (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

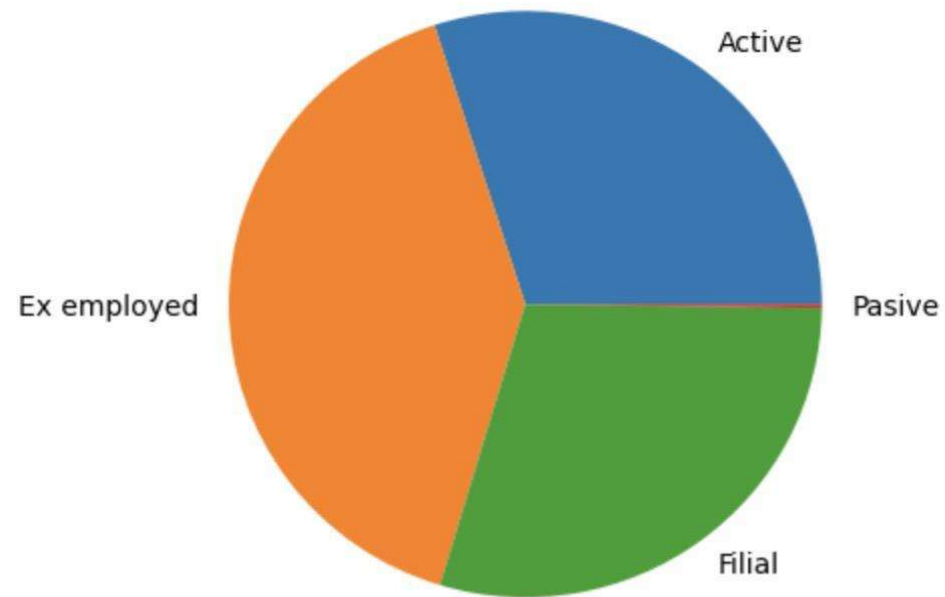


Exploratory Data Analysis (EDA)

The pie chart shows the percentage of employment index distribution

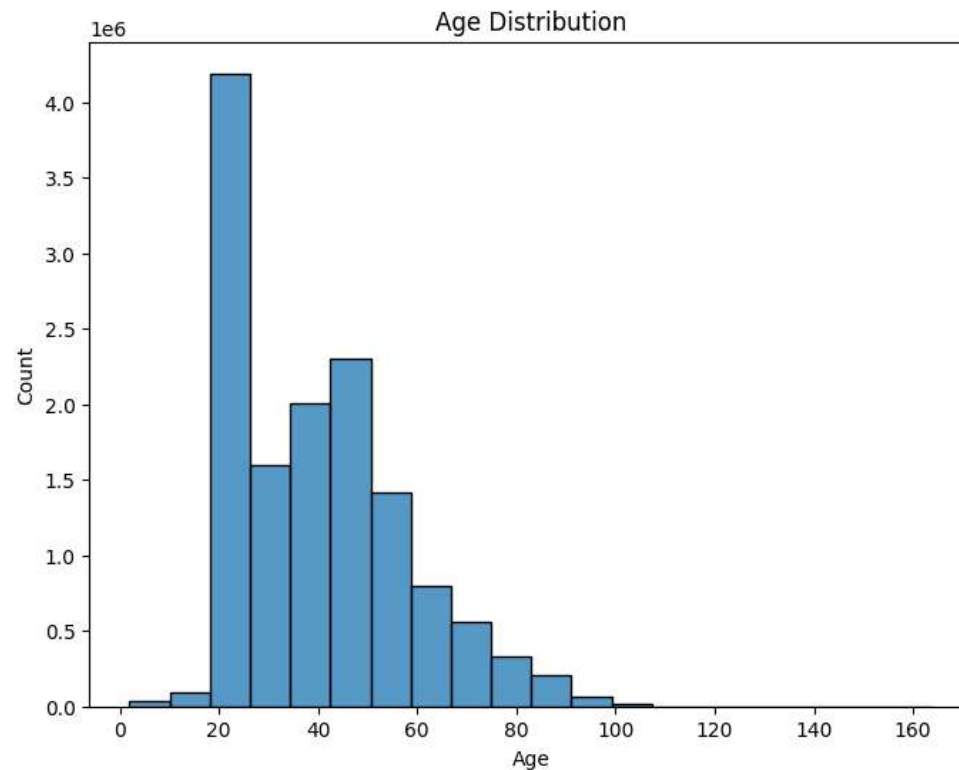
Unemployed ranks the 1st and then the Ex-employed. Passive is least here in the plot

So we could see that the employment situation is not positive for the unemployment percentage and Ex-employed count



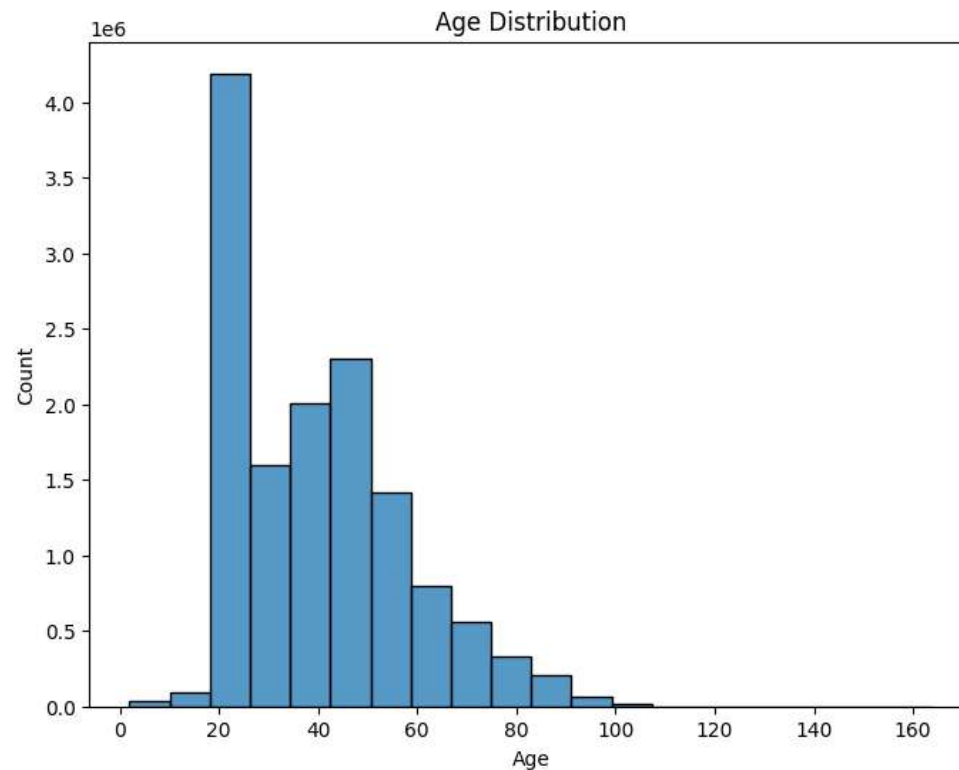
Exploratory Data Analysis (EDA)

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.



Exploratory Data Analysis (EDA)

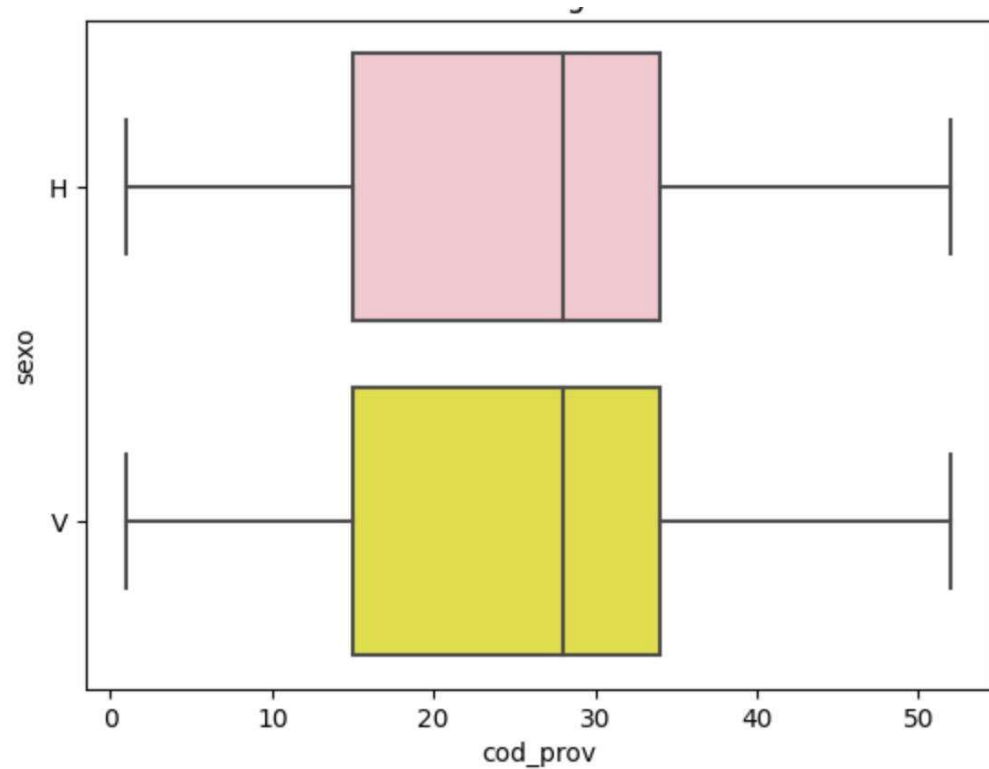
We can see from the Age distribution plot that the majority of customers are between ages 20 and 50 meaning the products are patronized by the working force.



Exploratory Data Analysis (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



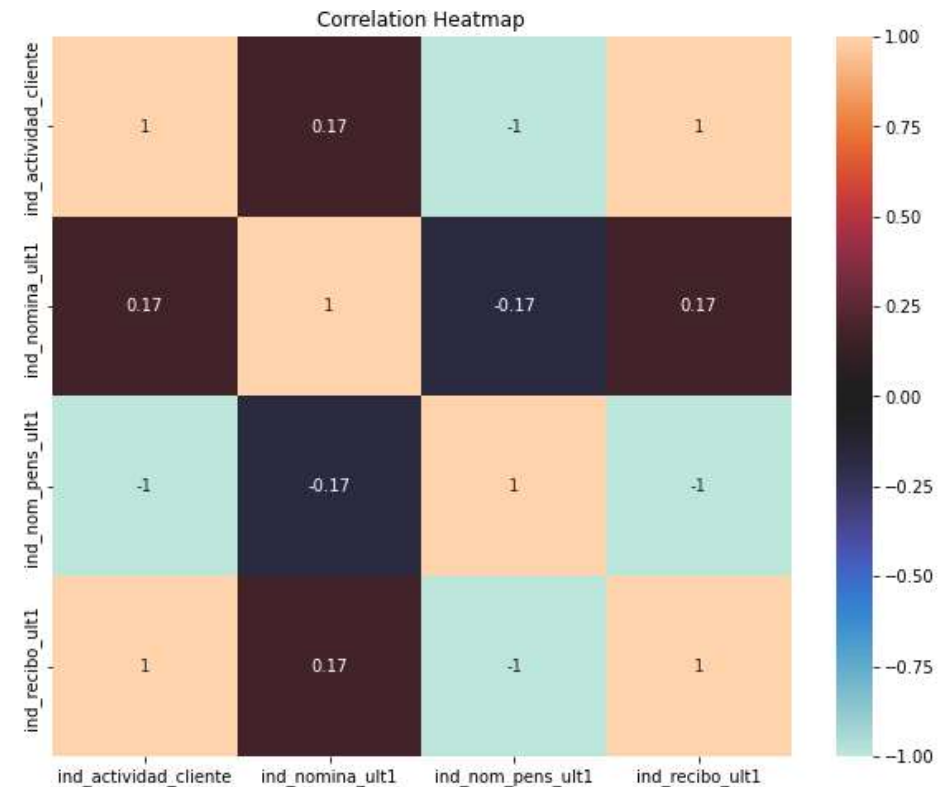
Exploratory Data Analysis (EDA)

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.

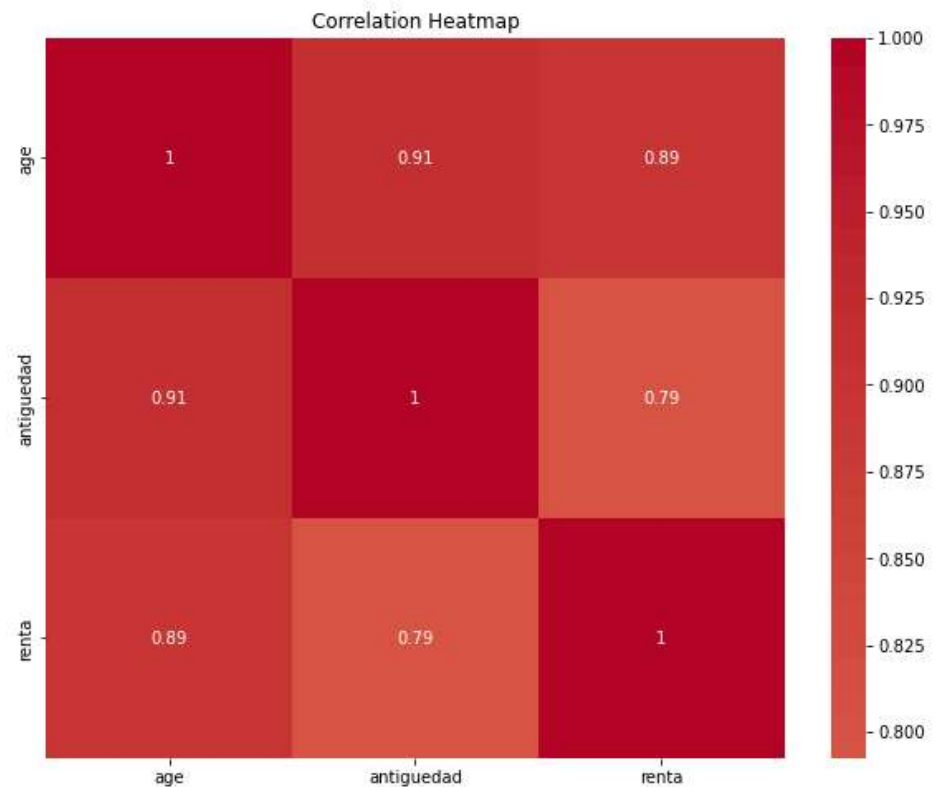
"`ind_actividad_cliente` " and "`ind_nom_pens_ult1`":The correlation between "`ind_actividad_cliente` " and "`ind_nom_pens_ult1`" is 1, indicating a negative linear association between a customer's activity index and whether they received a pension payment.

"`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



Exploratory Data Analysis (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.



Modelling

Linear model on target variable 'ind_ahor_fin_ult1'

Accuracy: The achieved accuracy of 1.00 (100%) might seem impressive at first glance. However, it's crucial to recognize that this high accuracy is largely due to the substantial class imbalance within the dataset, where the majority class is dominant. Relying solely on accuracy can be misleading in scenarios like this, and it should not be the sole determinant of model performance assessment.

Precision and Recall: The classification report underscores a significant contrast in the model's performance for the two classes:

The precision for class 0 is elevated (1.00), indicating that when the model predicts class 0, it's usually accurate. However, the precision for class 1 is exceedingly low (0.00), signifying the model's difficulty in making precise predictions for class 1. The recall for class 0 is high (1.00), suggesting that the model captures most instances of class 0.

F1 Score: The F1 score, a harmonized measure of precision and recall, is substantial for class 0 (1.00), while being strikingly low for class 1 (0.00). This confirms that the model's performance is heavily skewed towards the majority class.

Accuracy: 1.00					
Classification Report:					
	precision	recall	f1-score	support	
0	1.00	1.00	1.00	4085	454
1	0.00	0.00	0.00	4	19
accuracy			1.00	4085	873
macro avg			0.50	0.50	4085
weighted avg			1.00	1.00	4085

Logistic Regression

Logistic Regression:

Accuracy: 0.9998974515360609

Confusion Matrix: $\begin{bmatrix} 4085454 & 0 \\ 419 & 0 \end{bmatrix}$

Classification Report:		precision	recall	f1-score	support
0	1.00	1.00	1.00	4085454	
1	0.00	0.00	0.00	419	
accuracy		1.00		4085873	
macro avg		0.50	0.50	0.50	4085873
weighted avg		1.00	1.00	1.00	4085873

Final Recommendation

According to the model, we recommend business managers and business consultants to collect and analyze the characteristics of customers, obtain the most suitable model among different products, and select repeated independent variables. Moreover, the overlapping independent variables of the company's existing products are used to classify customers. For different products, finding the same customer attributes in the process of modelling and confirm the plan for cross selling products could increase the profit of cross product sales.



Data Glacier

Your Deep Learning Partner

Thank You