Data Glacier

Your Deep Learning Partner

Data Glacier – Final Project Week 11: EDA Presentation

Data Analyst: Cross-Selling Recommendation

-by Asha K C
LISUM43

Date: 20-May-2025

Agenda

- Team Details
- Problem Statement
- Business Understanding
- GitHub Repo Link
- EDA Summary
- Final Recommendations

Team Details

Team Member Details: Individual project (no team)

Name: Asha K. C.

Email: aasha.kc@outlook.com

Country: India

Company: Data Glacier

Specialization: Data Analyst

Problem Statement

Data Glacier
Your Deep Learning Partner

The XYZ bank is having difficulty cross-selling its products to existing customers. Customers are not buying additional products sold by their bank. Hence, data analysts must provide the information required to enhance their cross-selling methods.

Dataset provided:

1. Train.csv (1,36,47,309 entries) – This dataset is 2.13 MB. There are a total of 48 features. Features mainly consist of customer code, age, gender, income, address, products held, seniority, active/passive index,…

Business Understanding

Data Glacier
Your Deep Learning Partner

The problem statement indicates that XYZ Bank excels at acquiring new customers but struggles with customer retention and service diversification. Cross-selling involves offering complementary or related products to current customers.

For example:

- A customer with a checking account might be offered a credit card or a retirement savings plan.
- A customer with a home loan could benefit from the option to acquire a safe deposit box or an insurance product.

Cross-sell vs Upsell

	Definition	Use case
Cross-sell	Selling additional products to customers	Someone has a checking account → offers a credit card
Upsell	Selling a higher-ed or upgraded version of a product	Customer has a basic savings account → suggest a premium account with benefits

GitHub Repository link

Data Glacier
Your Deep Learning Partner

https://github.com/Asha-KC-07/Data-Glacier-Internship-2025---LISUM43/blob/main/Week%2010%20-%20Data%20Analyst_Cross-selling%20recomendation/EDA_v1.ipynb

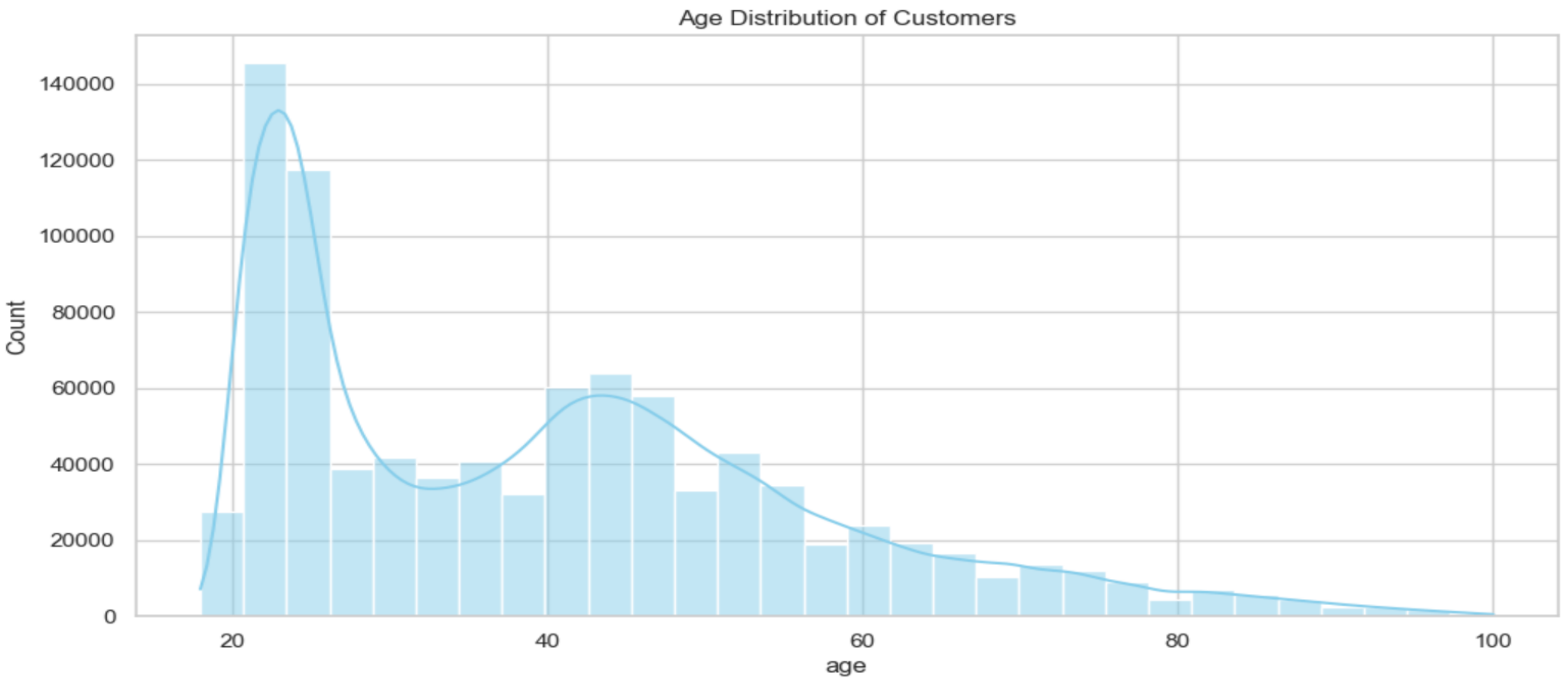
Data Glacier

Your Deep Learning Partner

EDA Summary

1. Demographics Analysis

Data Glacier
Your Deep Learning Partner

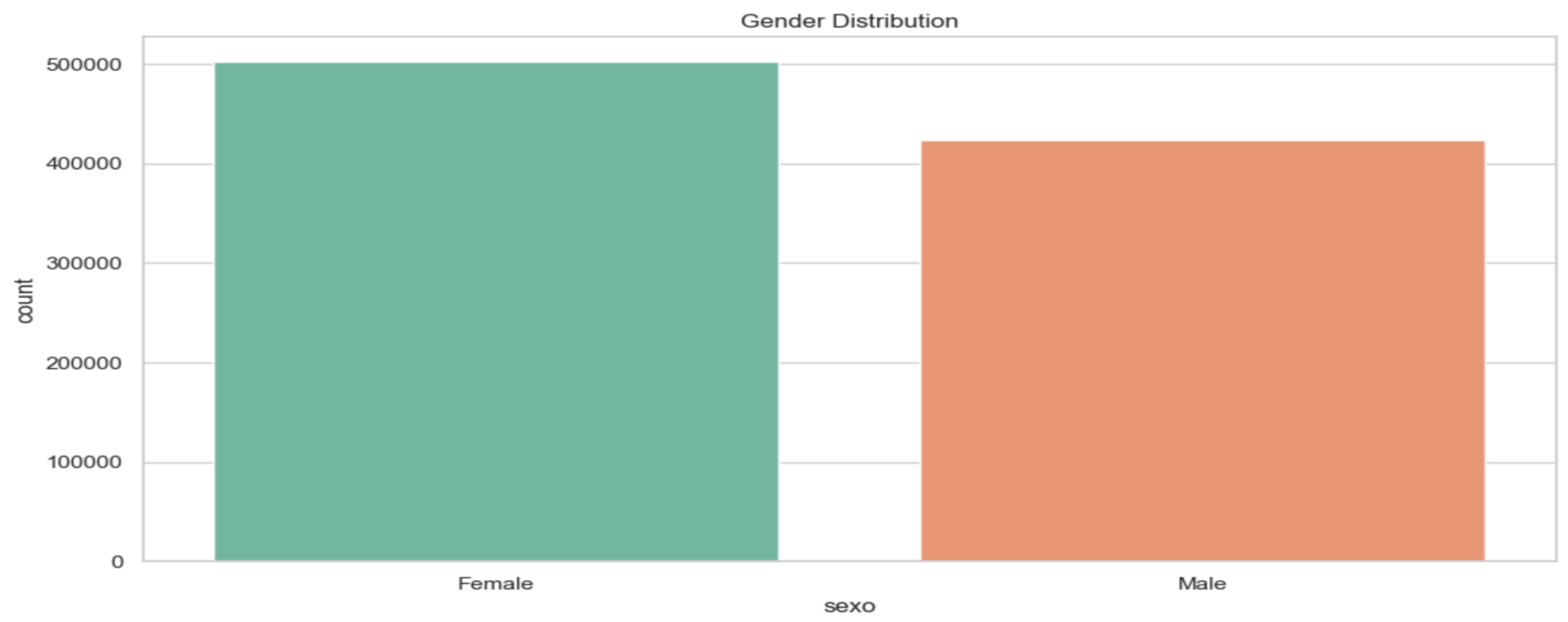


1. Age Distribution Histogram

- Most customers fall between **25 to 60 years**.
- There's a visible right skew due to older age values.
- Minor counts exist for extremely old ages (over 90).

1. Demographics Analysis

Data Glacier
Your Deep Learning Partner

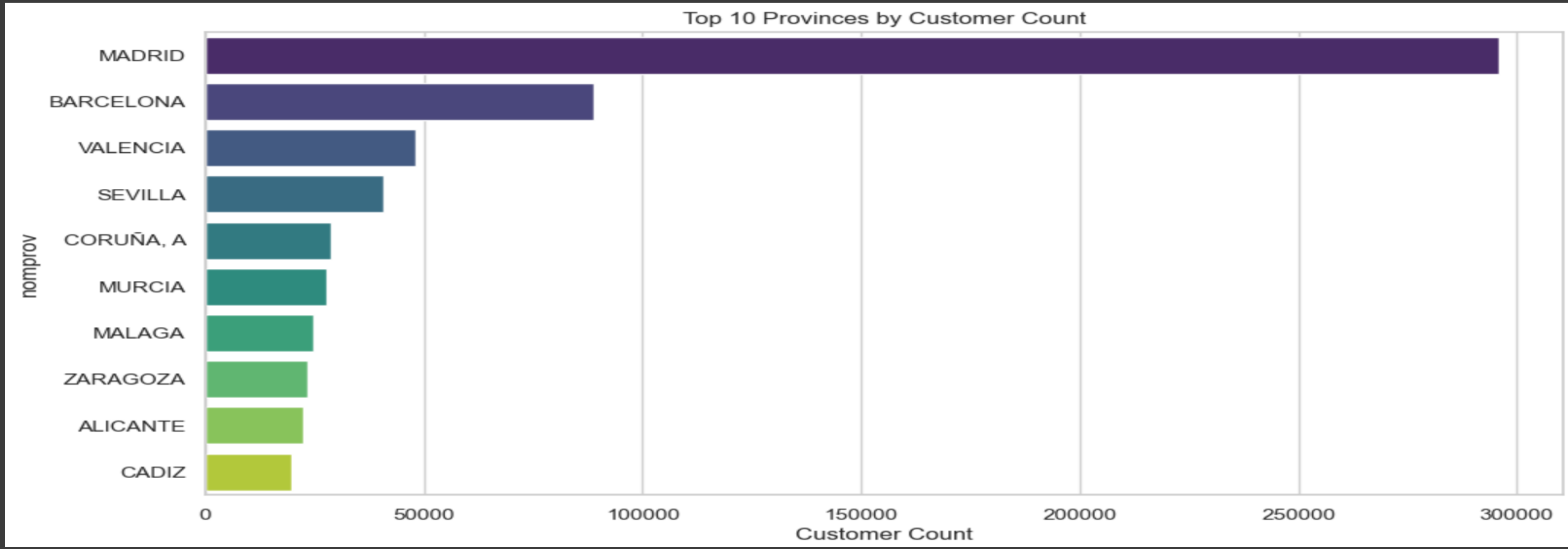


2. Gender Distribution

- **Slightly more females** than males.
- Imbalance isn't significant but could affect segment-specific targeting.

1. Demographics Analysis

Data Glacier
Your Deep Learning Partner



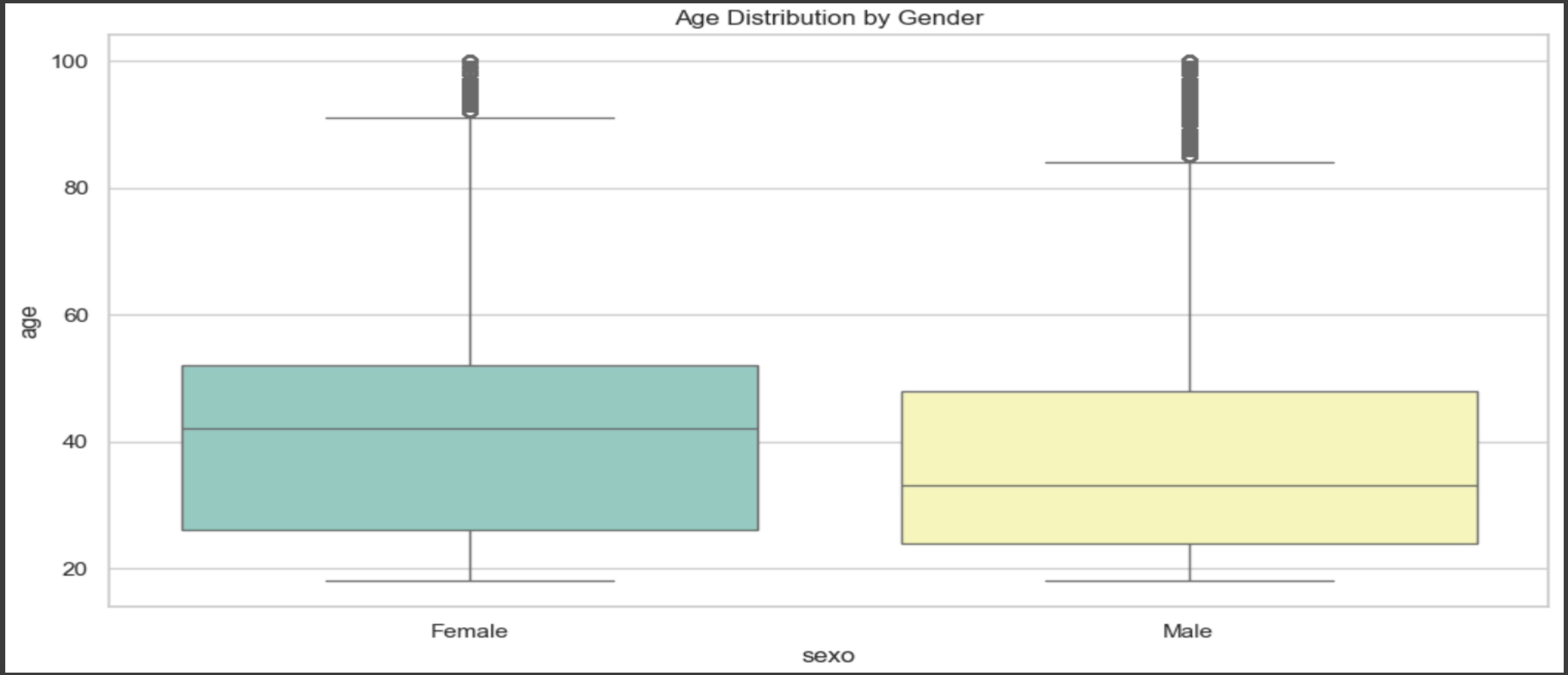
3. Top 10 Provinces by Customer Count

- Certain provinces (like **Madrid**, **Barcelona**, etc.) dominate the customer base.
- Indicates geographic concentration — great for region-specific marketing strategies.

1. Demographics Analysis

Data Glacier

Your Deep Learning Partner



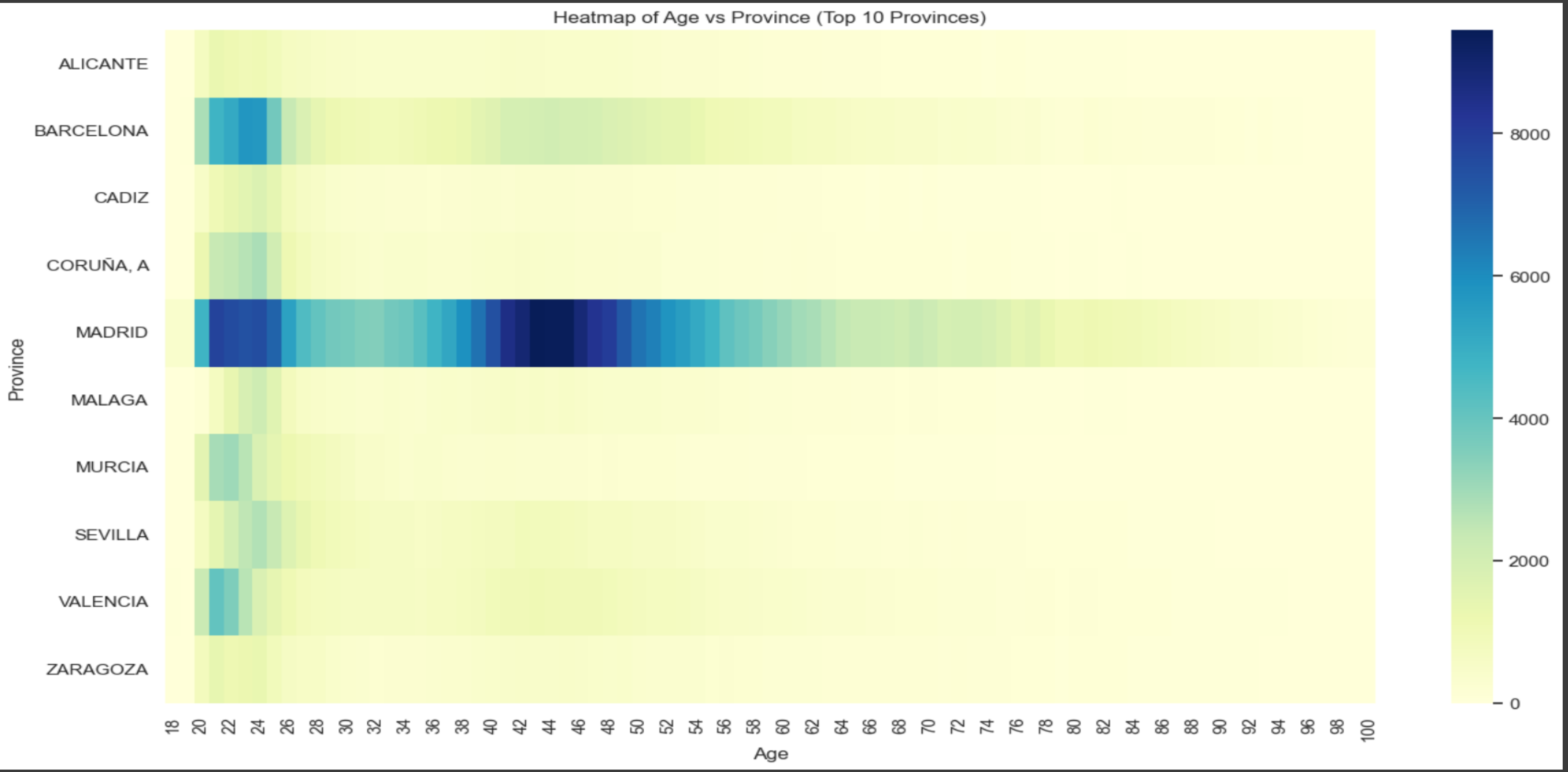
4. Age by Gender Boxplot

- Female customers tend to be **slightly older** on average.
- The age spread is similar for both genders but has a few more high-age outliers.

1. Demographics Analysis

Data Glacier

Your Deep Learning Partner

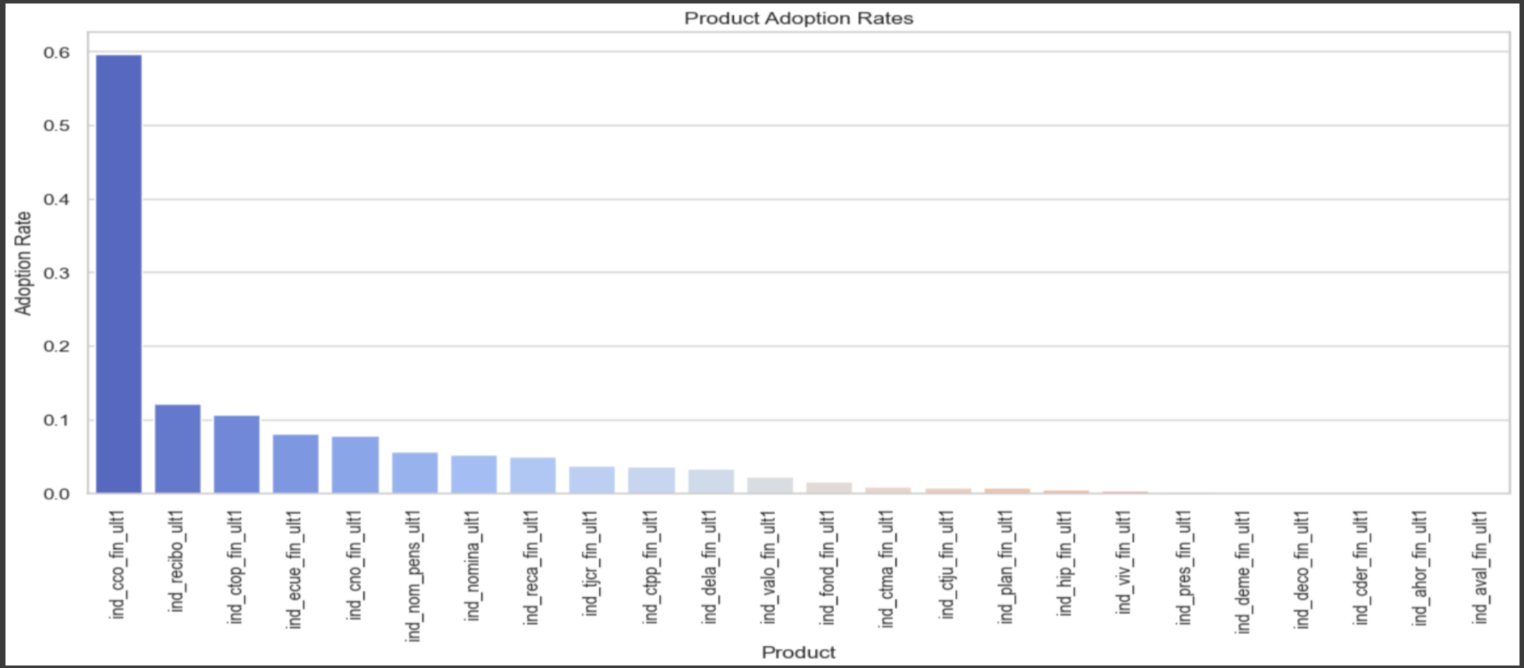


5. Heatmap of Age vs. Province

- Each province shows a **strong concentration around ages 40-55**.
- Useful to identify which regions have a younger vs older customer base.

2. Product Adoption Rates:

Data Glacier
Your Deep Learning Partner

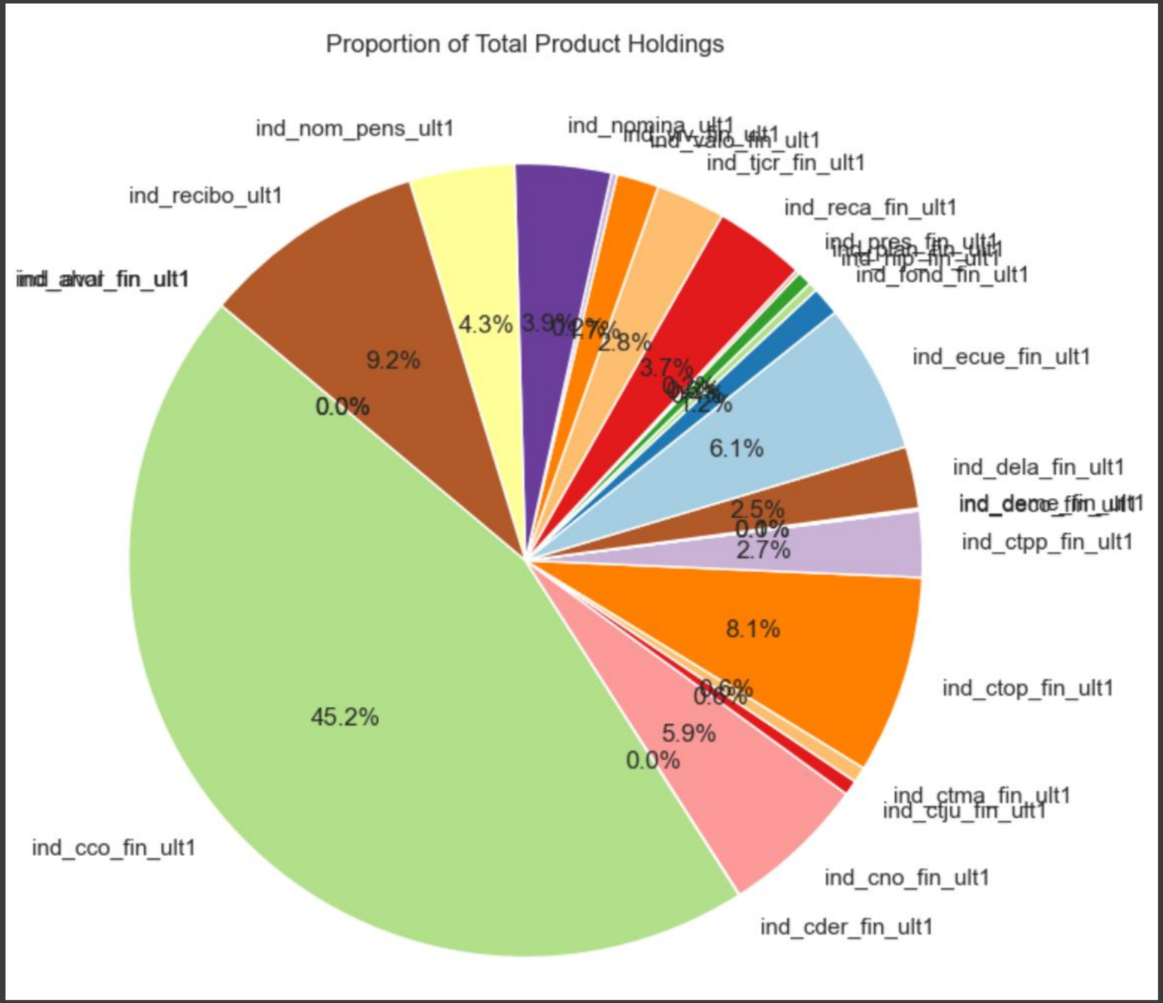


- Most commonly held products:
  - ind_recibo_ult1 (**12.1%**): Utility Bill Payments
  - ind_nom_pens_ult1 (**5.5%**): Pension Deposits
  - ind_nomina_ult1 (**5.1%**): Salary Deposits
  - ind_tjcr_fin_ult1 (**3.8%**): Credit Cards
- Least commonly held products:
  - ind_plan_fin_ult1, ind_pres_fin_ult1, ind_viv_fin_ult1: All below **1%**

2. Product Adoption Rates:

Data Glacier

Your Deep Learning Partner



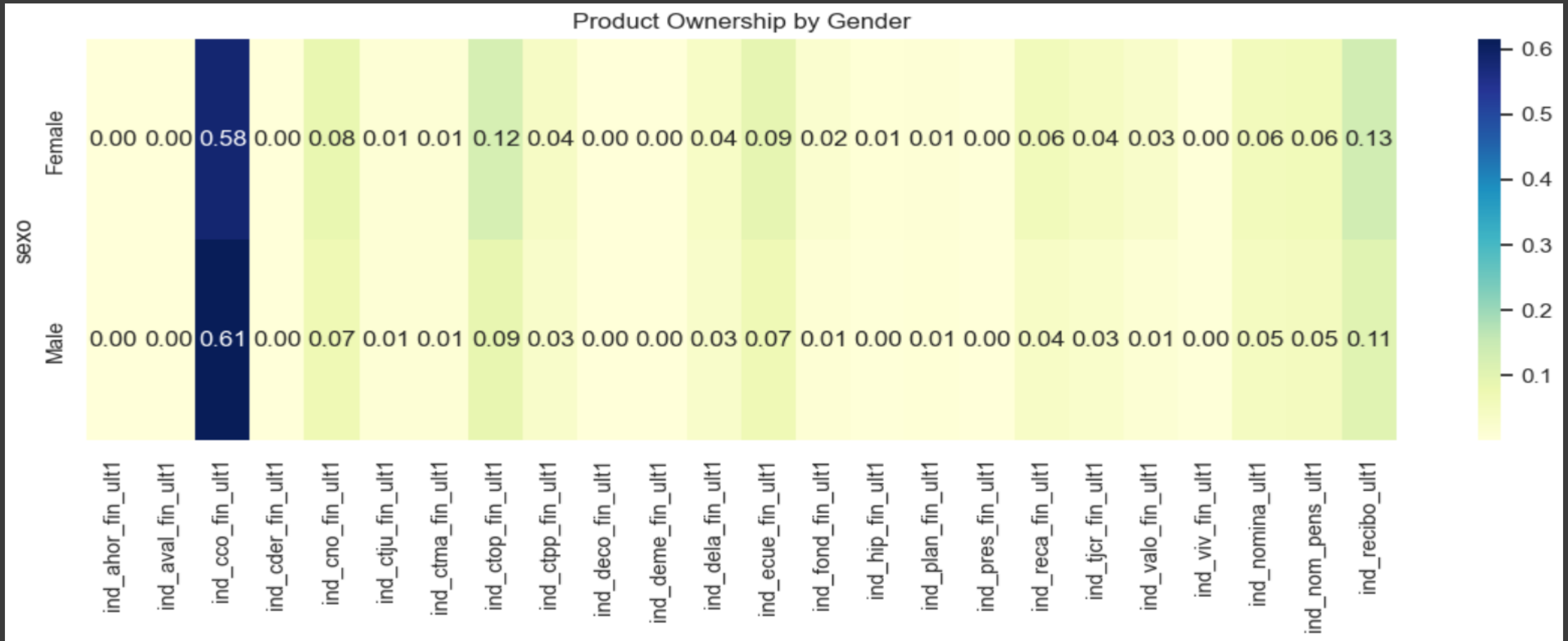
Pie Chart: Total Product Holdings

- A few product types dominate the customer portfolio space.
- Products like **salary/pension deposits** and **recurring payments** represent the bulk of ownership.

3. Product Ownership by Segment:

Data Glacier

Your Deep Learning Partner



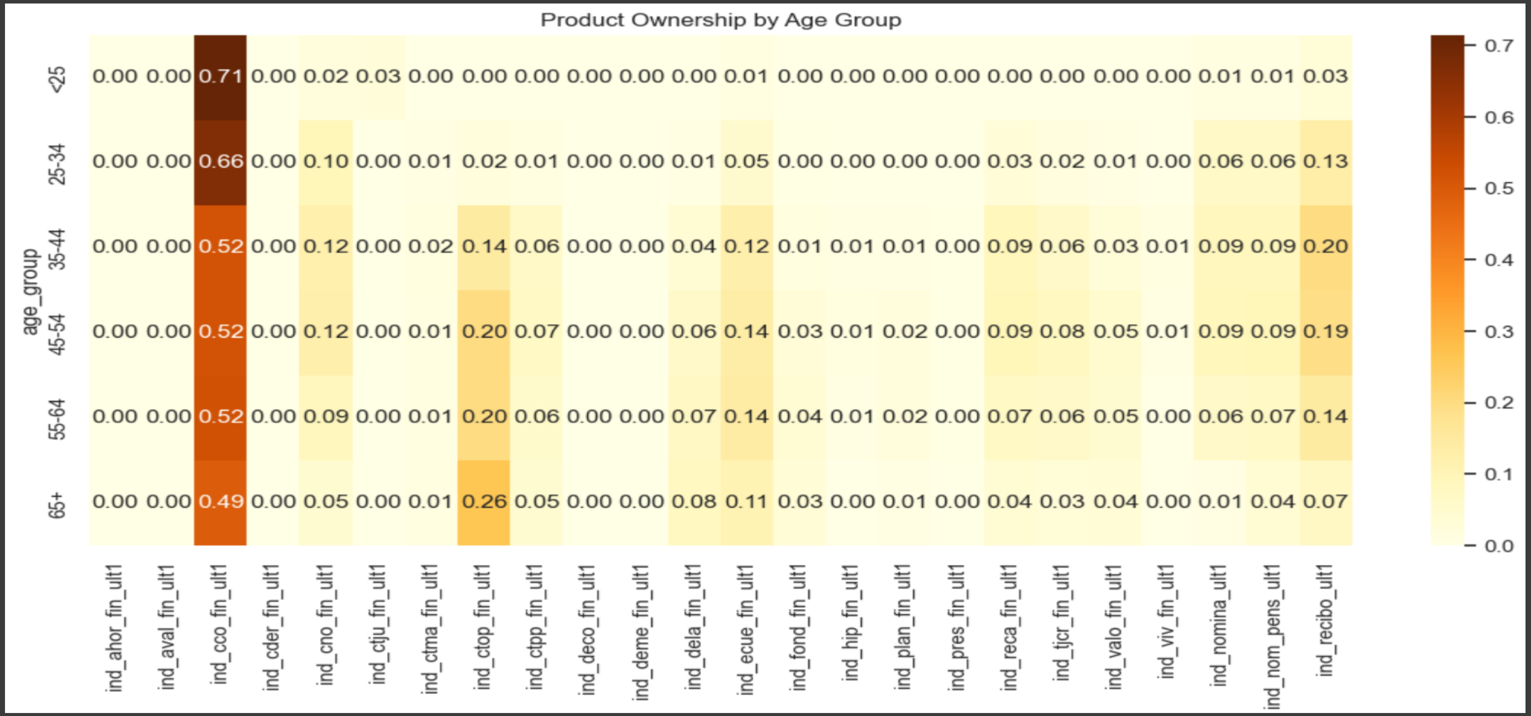
1. Gender-Based Ownership

- **Females** slightly lead in products like:
  - ind_nomina_ult1 (salary deposits)
  - ind_recibo_ult1 (recurring bill payments)

- **Males** show marginally higher ownership in:
  - Investment-related products (ind_valo_fin_ult1, ind_fond_fin_ult1)

3. Product Ownership by Segment:

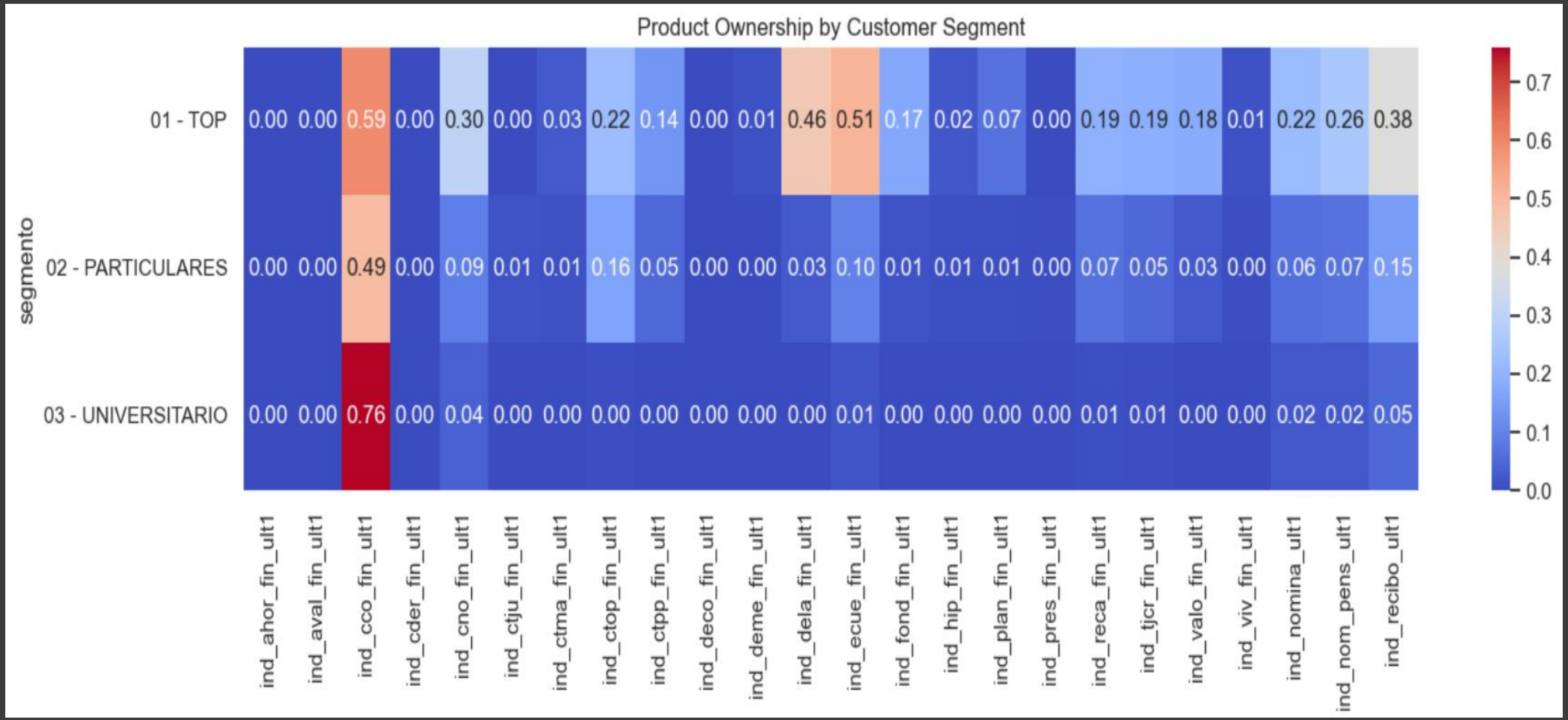
Data Glacier
Your Deep Learning Partner



2. Age Group Trends

- **Young adults (<25)** rarely hold any product.
- **Ages 35-54** dominate across almost every product — especially:
    - Salary deposits (ind_nomina_ult1)
    - Credit cards (ind_tjcr_fin_ult1)
- **Older groups (65+)** tend to show less ownership of credit or investment products.

3. Product Ownership by Segment:

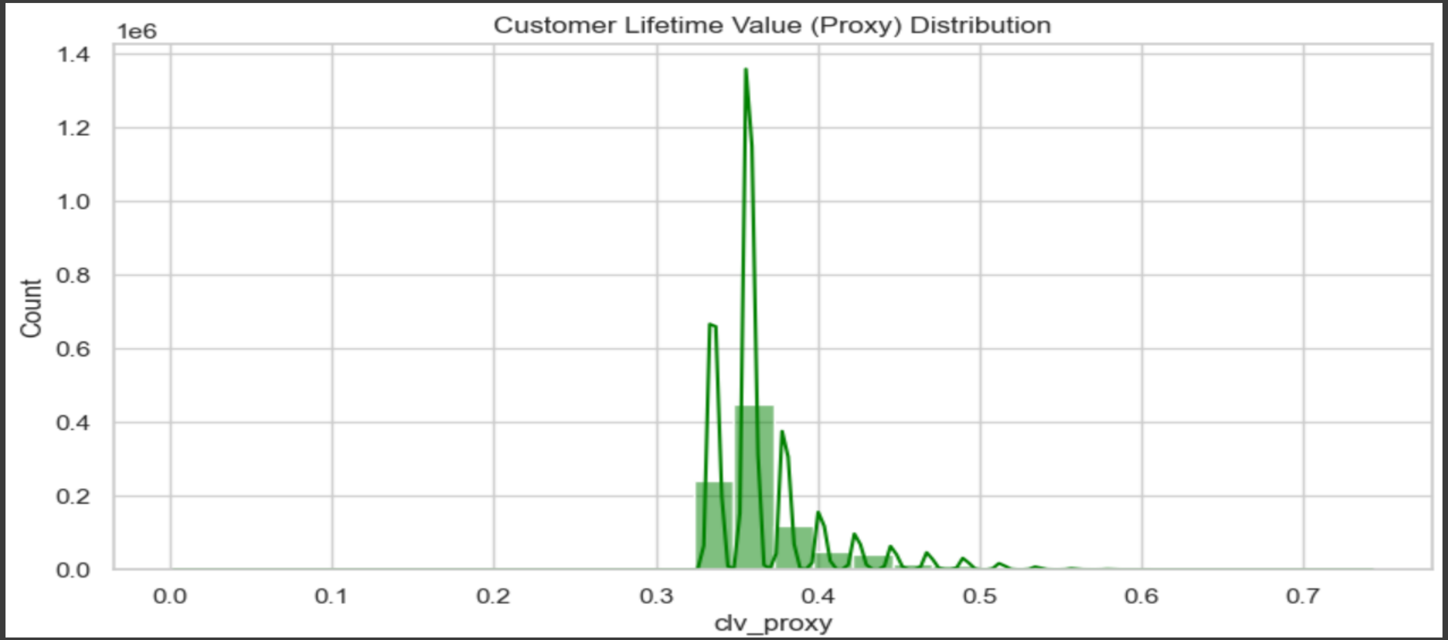


3. Segment-Based Trends

- segmento indicates customer types like "01 - VIP", "02 - Individuals", "03 - College students":
  - **VIPs** show highest ownership in multiple financial products, especially credit and investment tools.
  - **College students** exhibit minimal product ownership — mostly basic accounts.

4. Customer Lifetime Value Proxies

Data Glacier
Your Deep Learning Partner



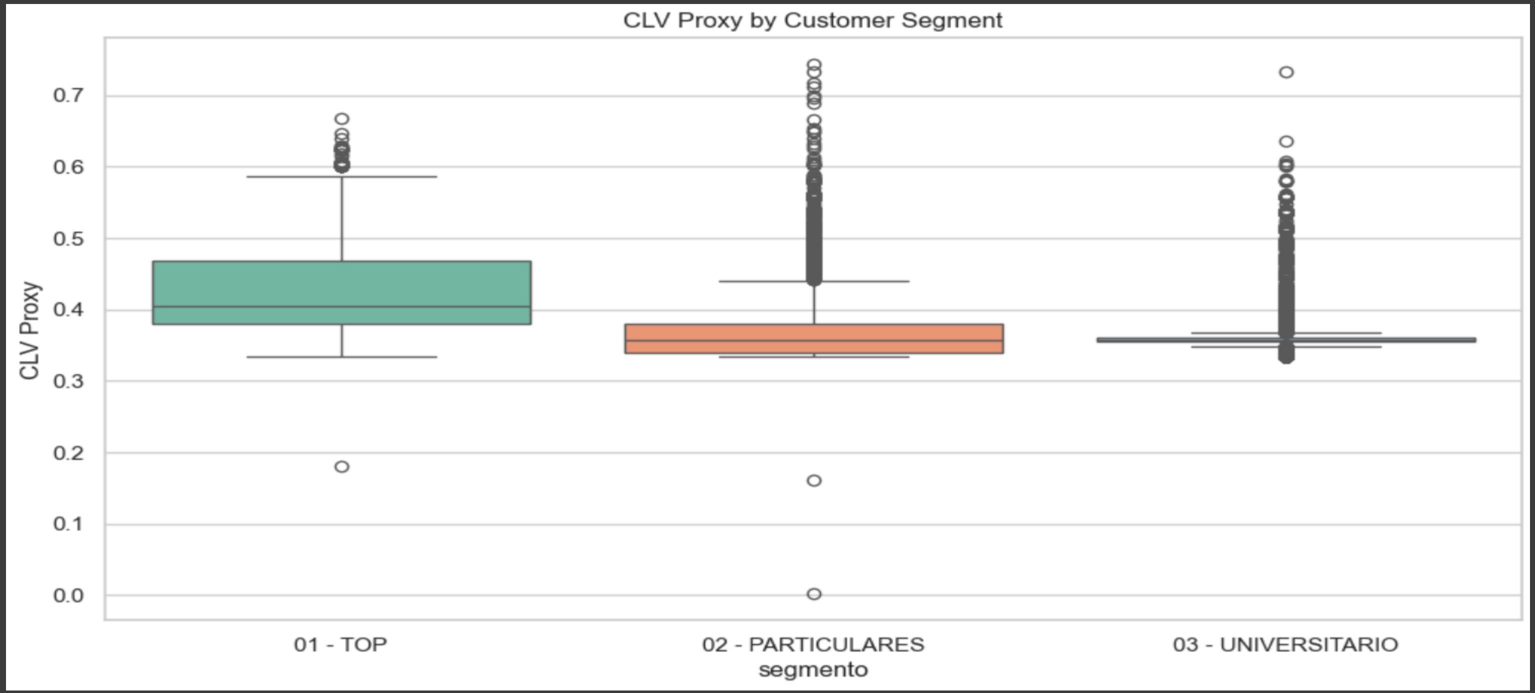
1. Distribution of CLV Proxy

- Most customers fall within **mid** CLV proxy scores.
- Right-skewed: a smaller segment shows high-value potential (top 10–15%).

4. Customer Lifetime Value Proxies

Data Glacier

Your Deep Learning Partner

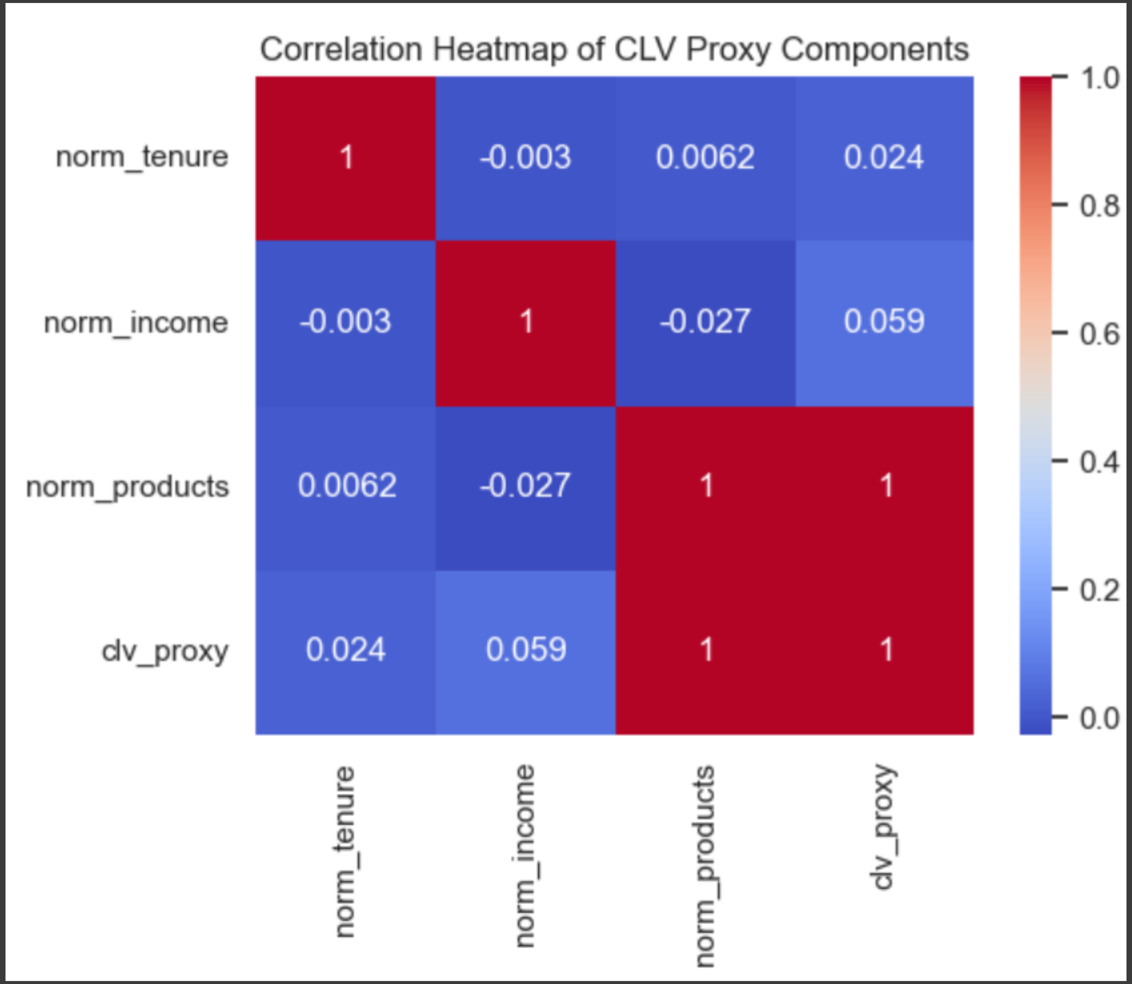


2. Boxplot by Segment

- **VIPs** (01 - VIP) have the highest median and spread of CLV.
- **College students** (03 - UNIVERSITARIO) score lowest in CLV — expected due to low tenure, product count, and income.
- **Mass Market Individuals** (02 - PARTICULARES) span the full spectrum, indicating a diverse customer base.

4. Customer Lifetime Value Proxies

Data Glacier
Your Deep Learning Partner

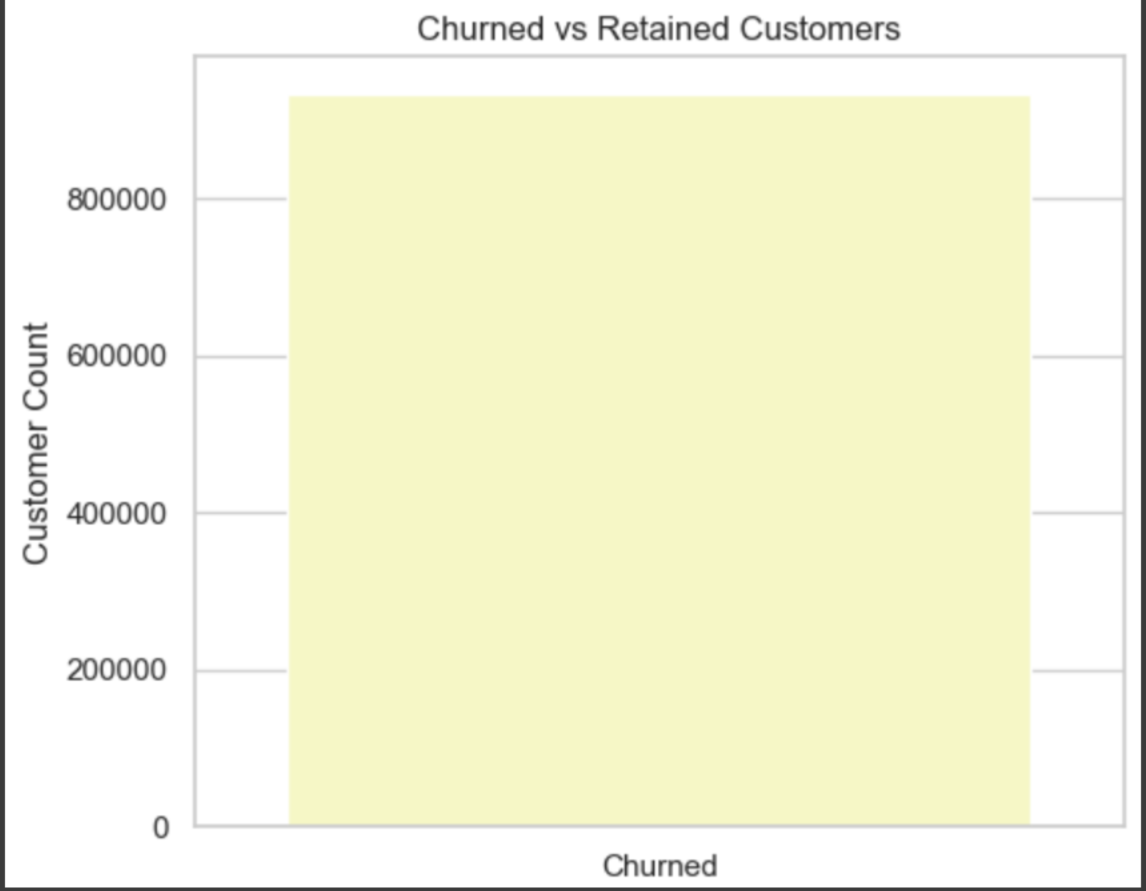


3. Correlation Heatmap

- **Product count** (norm_products) is most strongly correlated with overall CLV score.
- **Income and tenure** also contribute but less dominantly.

5. Churn Indicators

Data Glacier
Your Deep Learning Partner

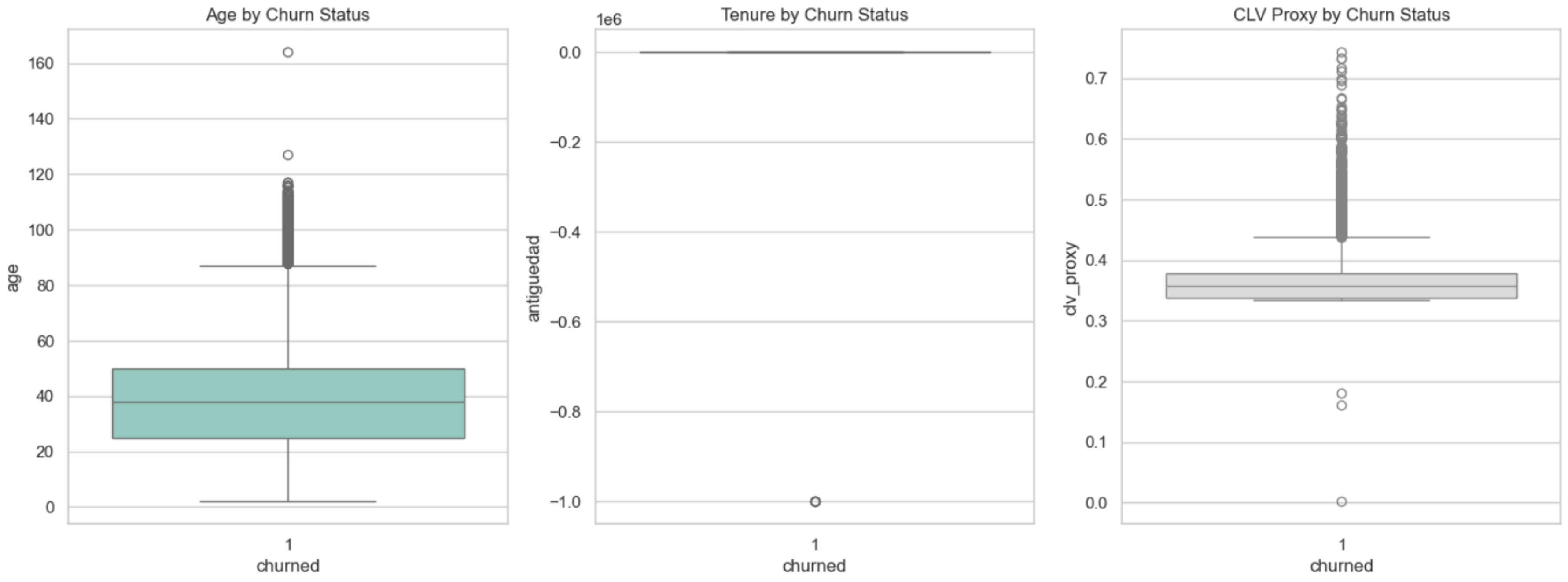


1. Churned vs Retained Customers

- A **very small fraction** of the customers is marked as "churned".
- Indicates the dataset mostly includes **active customers**.

5. Churn Indicators

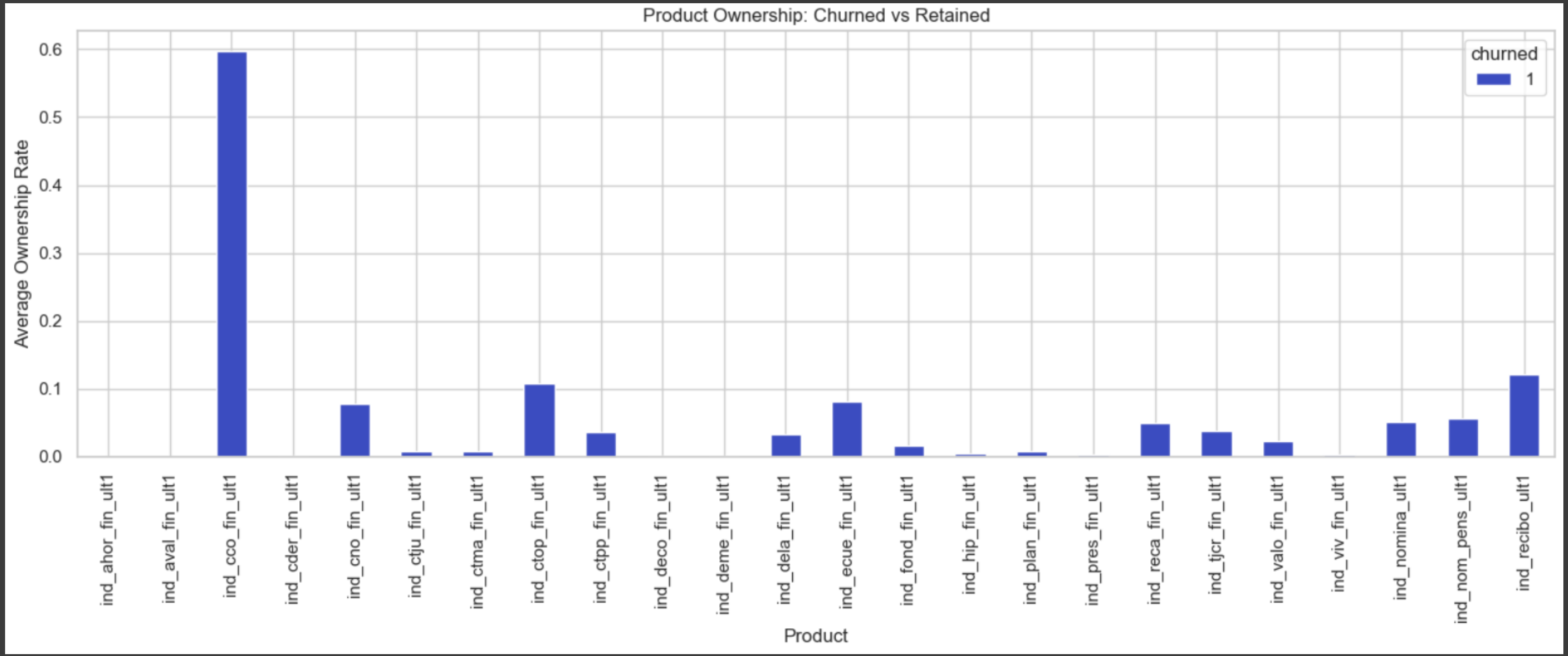
Data Glacier
Your Deep Learning Partner



2. Age, Tenure, and CLV Comparisons

- **Churned customers** tend to:
  - Be **older on average**
  - Have **shorter tenure** (surprising — might reflect new users abandoning)
  - Have **lower CLV scores** overall
- Retained users dominate in higher tenure and product count distributions.

5. Churn Indicators



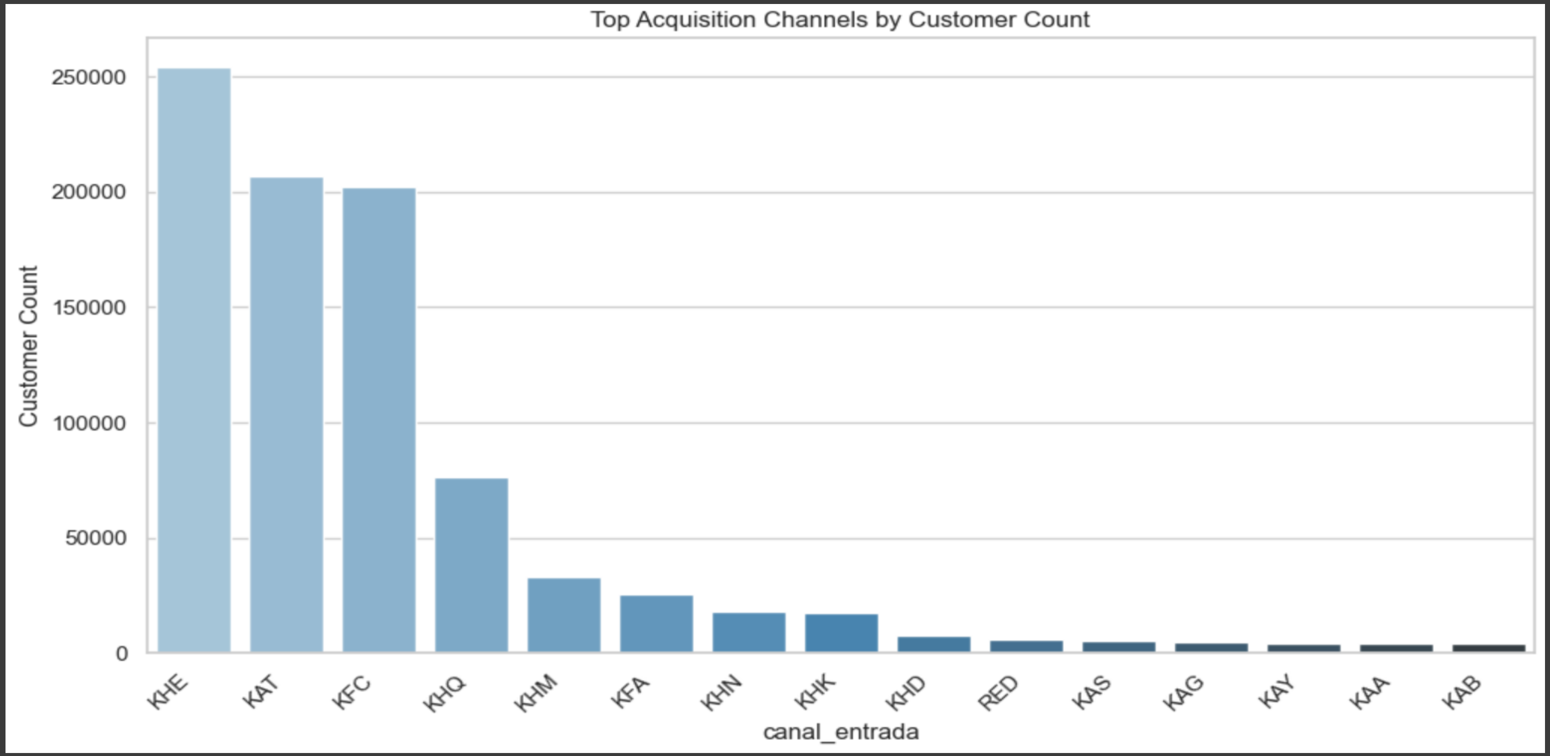
3. Product Ownership Drop-Off

- Churned users have **lower ownership rates across nearly all products**.
- Largest relative drop in products like:
  - Credit Cards (ind_tjcr_fin_ult1)
  - Salary/Pension Accounts

6. Channel Effectiveness

Data Glacier

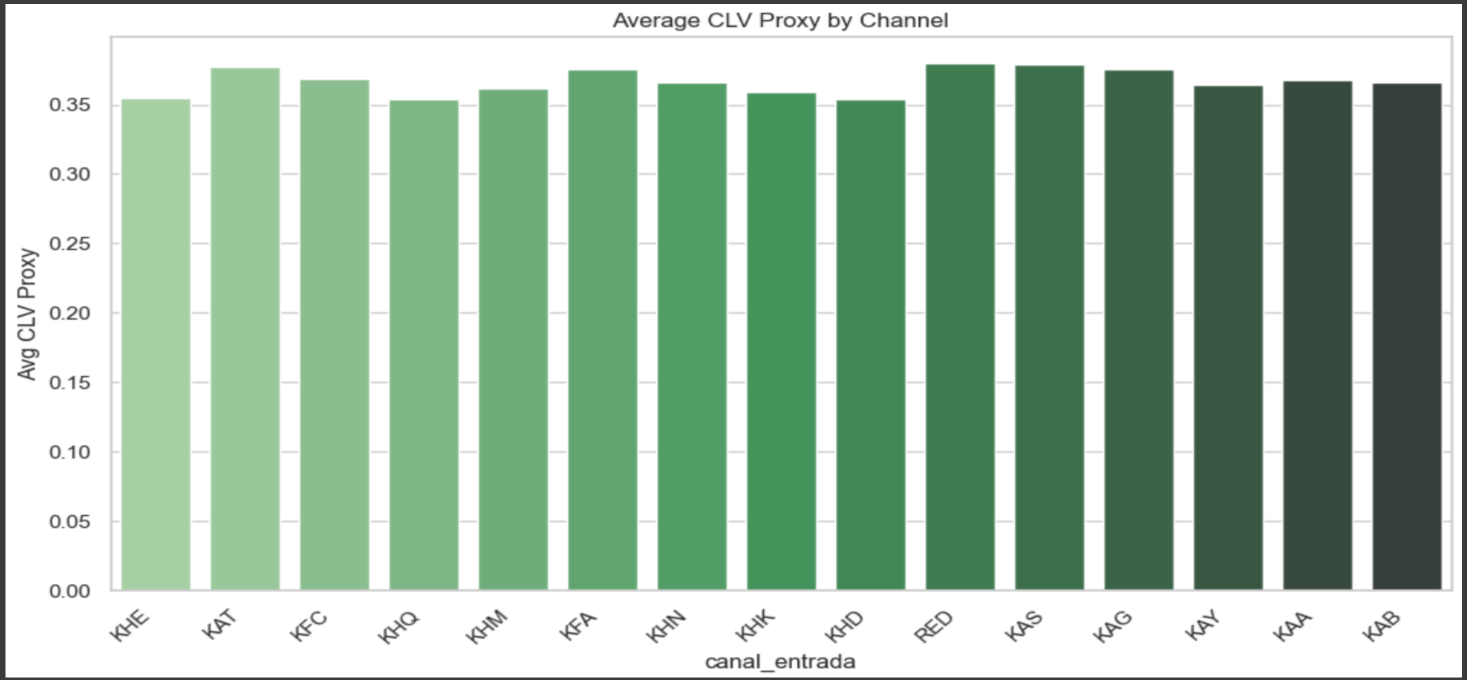
Your Deep Learning Partner



1. Top Acquisition Channels by Customer Count

- A few channels dominate customer acquisition:
    - KHE, KAT, and KFC are among the most used.
- These likely represent physical or digital acquisition pathways.

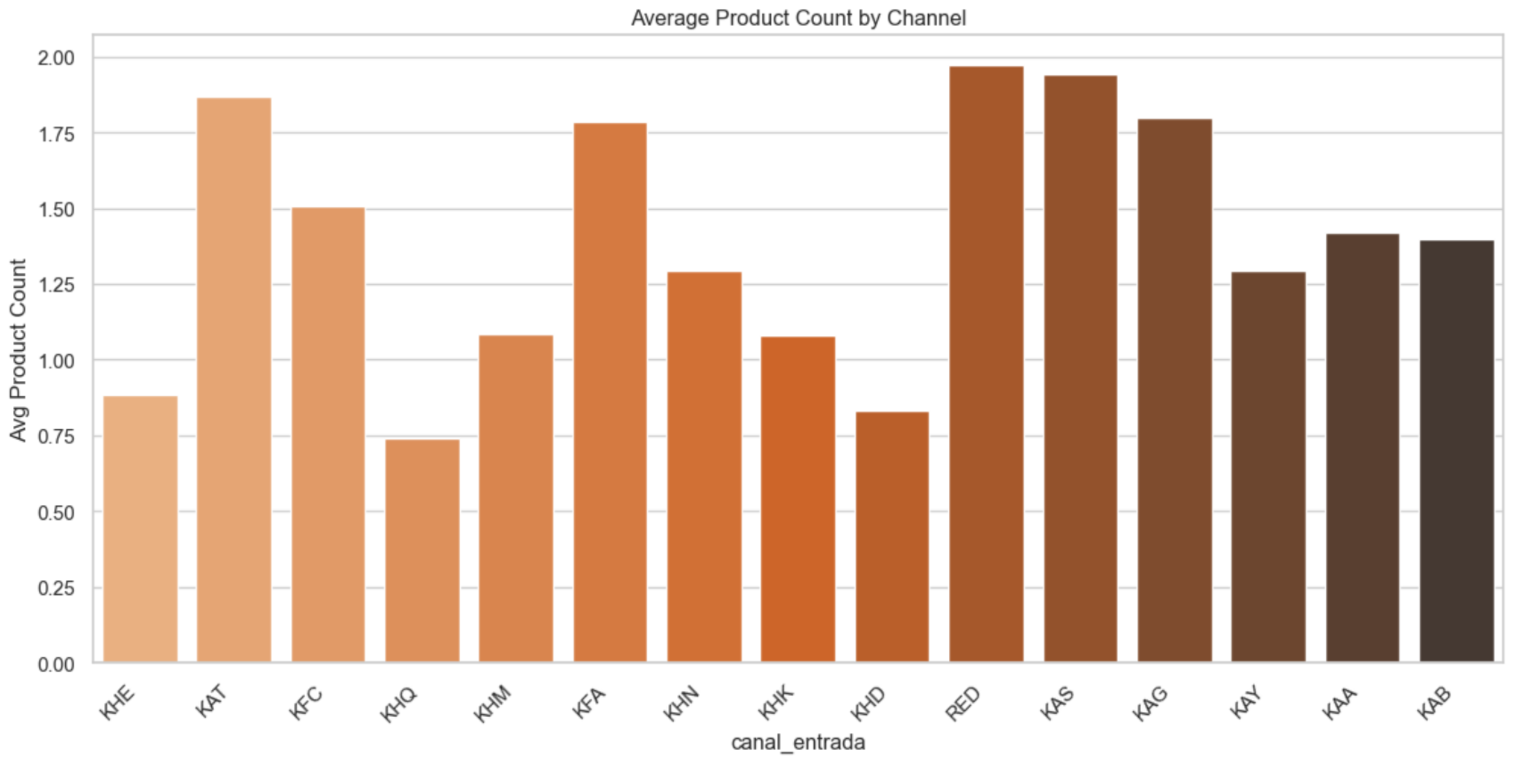
6. Channel Effectiveness



2. Average CLV Proxy by Channel

- High customer volume **does not always mean high value**:
    - Channels like KAT and KFA show **higher CLV**, despite smaller customer bases.
- Mass channels may acquire many users, but **niche or referral-based ones attract higher value clients**.

6. Channel Effectiveness

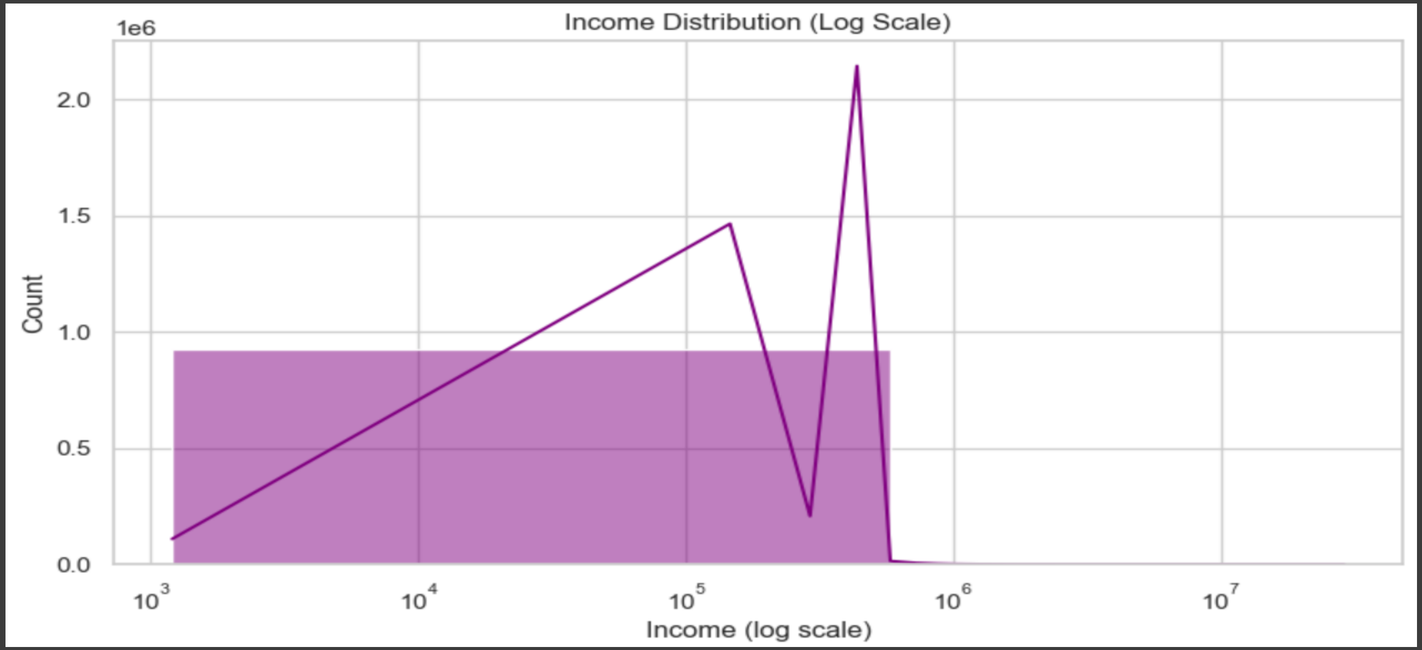


3. Average Product Count by Channel

- Channels with higher CLV also typically yield **more product engagement**.
- KFA and KAT again appear as effective **quality acquisition routes**.

7. Income Analysis

Data Glacier
Your Deep Learning Partner



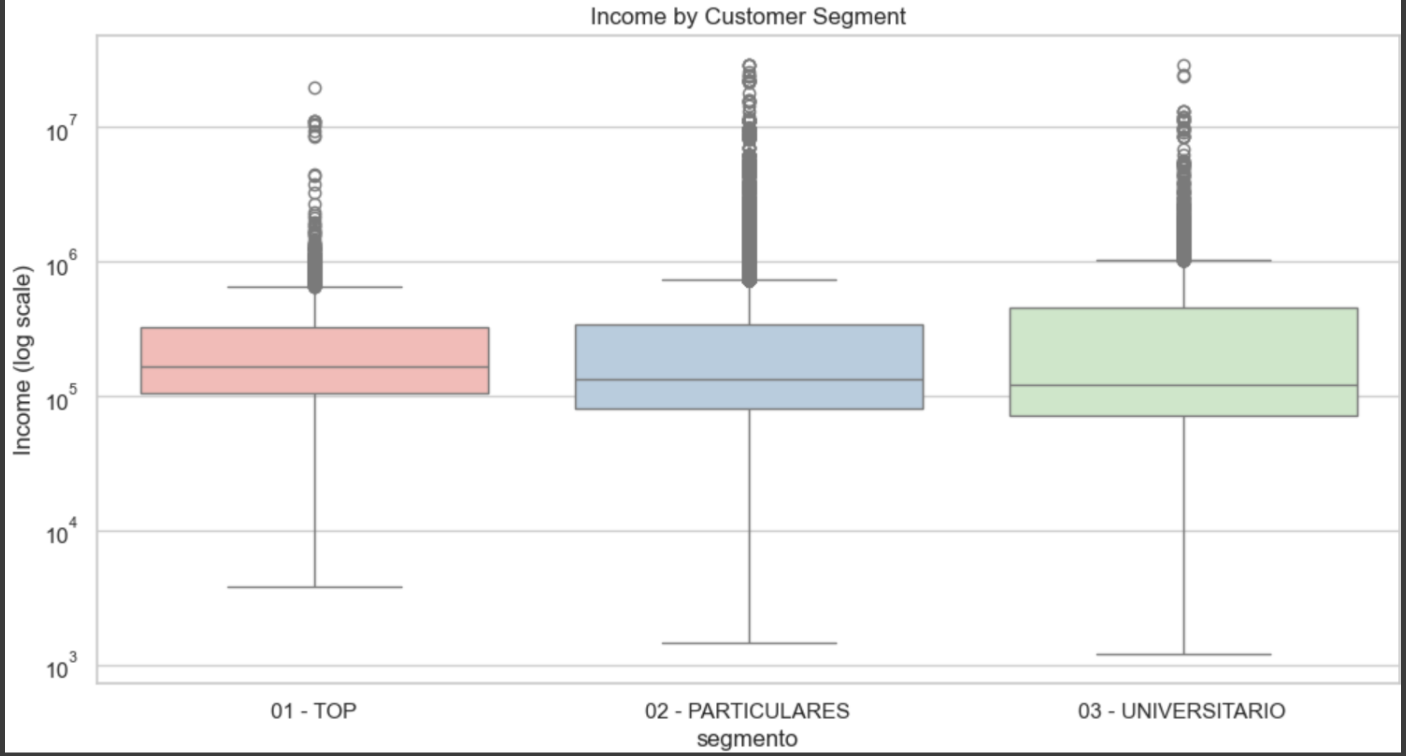
1. Income Distribution (Log Scale)

- Highly **right-skewed**: majority of incomes lie below ~60,000.
- A few customers report **extremely high income** (>100,000), suggesting income outliers.

7. Income Analysis

Data Glacier

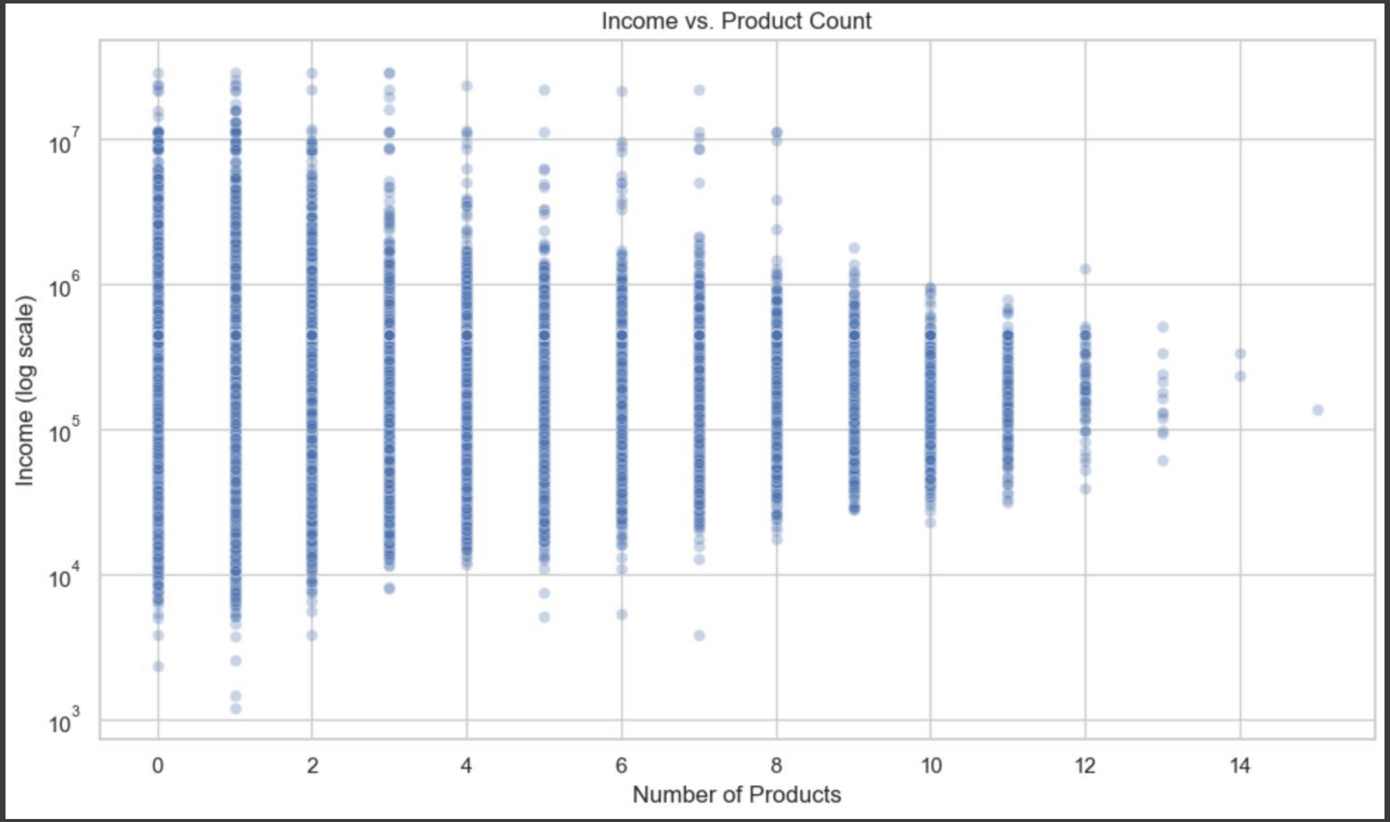
Your Deep Learning Partner



2. Income by Customer Segment

- **VIPs** (01 - VIP) predictably show the **highest income range**.
- **College students** (03 - UNIVERSITARIO) have the lowest and most compact income distribution.
- Segments are well-separated, validating the segmentation strategy by income

7. Income Analysis



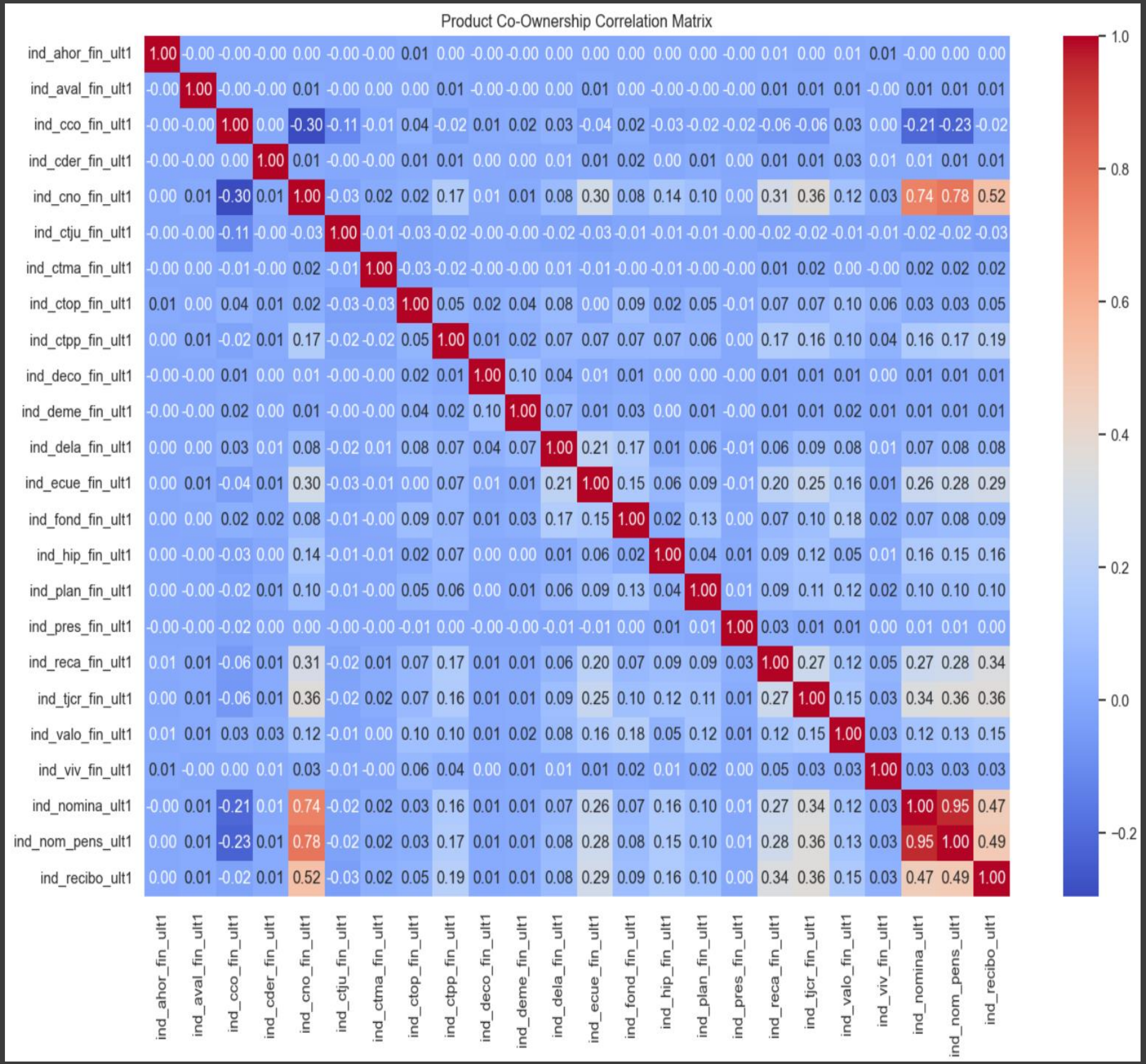
3. Income vs. Product Count

- Positive trend: **higher product count tends to correlate with higher income**.
- However, some **high-income customers own few products**, indicating potential for upselling.

8. Product Co-Ownership Patterns

Data Glacier

Your Deep Learning Partner



Key Product Correlations

- **High Positive Correlation Pairs**:
  - ind_nomina_ult1 (salary) ↔ ind_recibo_ult1 (bill payments): 0.66
    ➤ Customers receiving salaries tend to set up recurring payments.
  - ind_nom_pens_ult1 (pension) ↔ ind_recibo_ult1: 0.45
    ➤ Similar trend with pension-based income.
- **Investment-related Products** like ind_fond_fin_ult1, ind_valo_fin_ult1, ind_deco_fin_ult1 are often held together:
  - Moderate correlations (~0.3–0.5), indicating bundled behaviours.
- **Minimal or Near-Zero Correlations**:
  - Savings accounts (ind_ahor_fin_ult1) and insurance (ind_plan_fin_ult1) don't strongly align with others.

Data Glacier

Your Deep Learning Partner

Final Recommendations

Final Recommendation:

Data Glacier
Your Deep Learning Partner

Product Adoption Patterns

- High adoption in salary deposits, bill payments, and credit cards.
- Very low adoption of investment and insurance products.
- ***Cross-sell opportunity***: Promote underutilized products like investment funds to salary account holders.

Demographic & Segment Analysis

- Females slightly lead in essential banking services; males show more interest in investments.
- Ages 35-54 dominate product holdings.
- VIPs hold diverse product portfolios; students are underpenetrated.
- ***Cross-sell opportunity***: Bundle investment and savings products for VIPs and the 35-54 age group.

CLV-Based Targeting

- Higher CLV is linked with tenure, income, and product count.
- CLV proxy helps identify top-tier clients for premium services.
- ***Cross-sell opportunity***: Upsell additional products to high CLV clients with moderate product usage.

Final Recommendation:

Data Glacier
Your Deep Learning Partner

Channel Performance

- Mass acquisition channels deliver volume but not value.
- Channels like KAT and KFA yield higher CLV and product engagement.
- ***Cross-sell opportunity***: Invest in quality acquisition channels for better LTV.

Product Pairing Insights

- Salary accounts strongly correlate with bill payments.
- Investment products are frequently co-owned.
- ***Cross-sell opportunity***: Automate bundled recommendations (e.g., salary + bills + investment starter).

Behavioral & Anomaly Insights

- Users with zero product ownership and high income are prime cross-sell targets.
- Anomalous age and income entries should be reviewed.
- ***Cross-sell opportunity***: Trigger onboarding flows for dormant accounts.

Final Recommendation:

Strategic Recommendations

- Prioritize high-CLV customers for personalized upselling.
- Develop segment-specific product bundles.
- Promote low-adoption products through targeted campaigns.
- Use acquisition channel metrics to guide marketing investments.
- Launch proactive outreach to zero-product customers with entry-level offers.

Thank You

Data Glacier

Your Deep Learning Partner