Data Glacier

Your Deep Learning Partner

CROSS SELLING RECOMMENDATION PROJECT

Virtual Internship

16-May-2023

Team Member's Details

Group Name: Individual

Name: Ian Kihara Wangui

Email: eandavid6@gmail.com

Company: DataGlacier

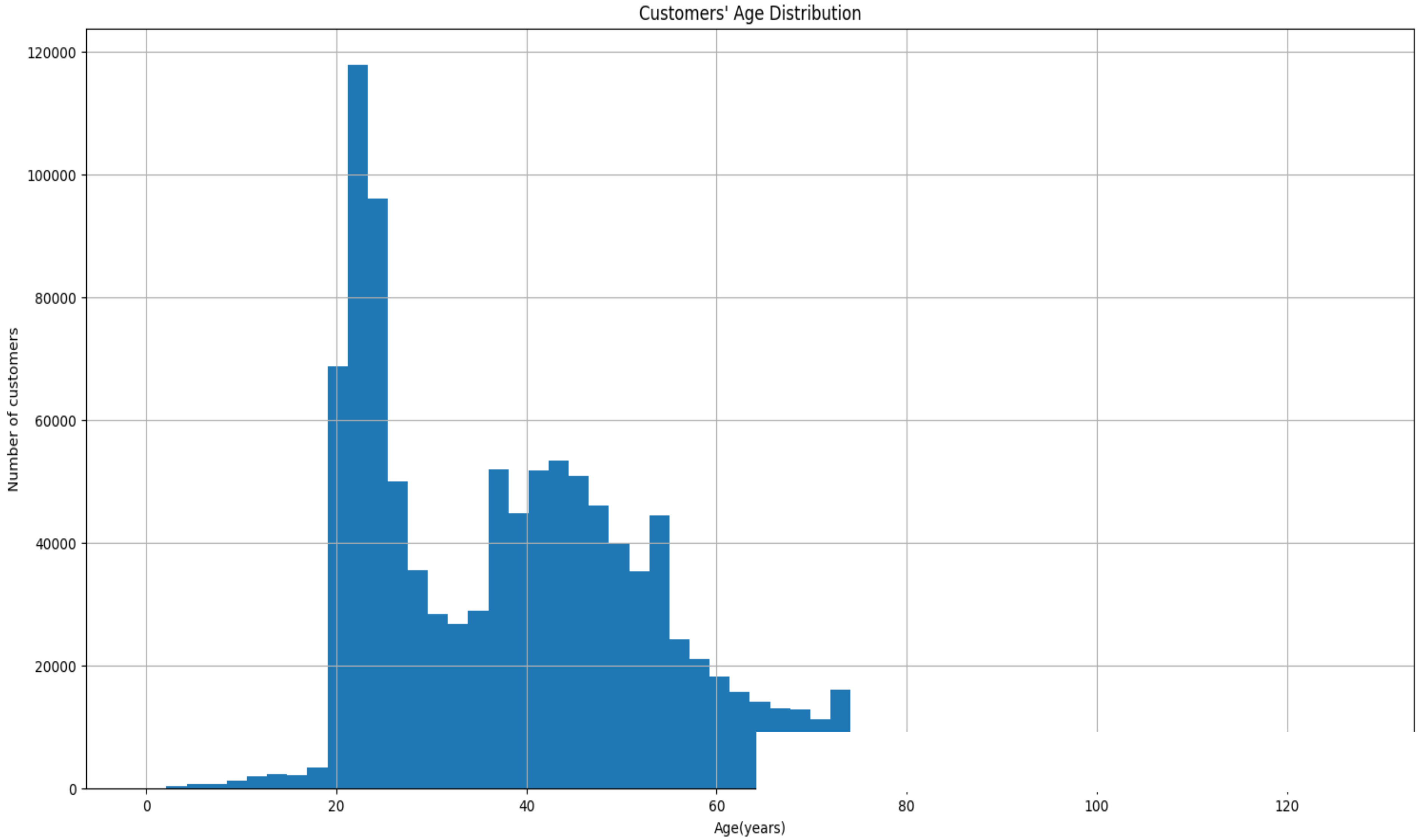
Specialization: Data Analyst

Github Repo Link: https://github.com/Iandavidk/DataGlacier/tree/main/Week-11-Cross_Selling_Presentation

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 not buying more than 1 product.
- This means XYZ Credit Union 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.
- **Objective : As a data analyst at ABC analytics you need to inspect the data and suggest what action XYZ Credit Union can take to increase cross selling (without using Machine Learning models)**

1. Age Distribution of customers



- It's interesting that the distribution is bimodal.
- There are a large number of university aged students, and then another peak around middle-age.

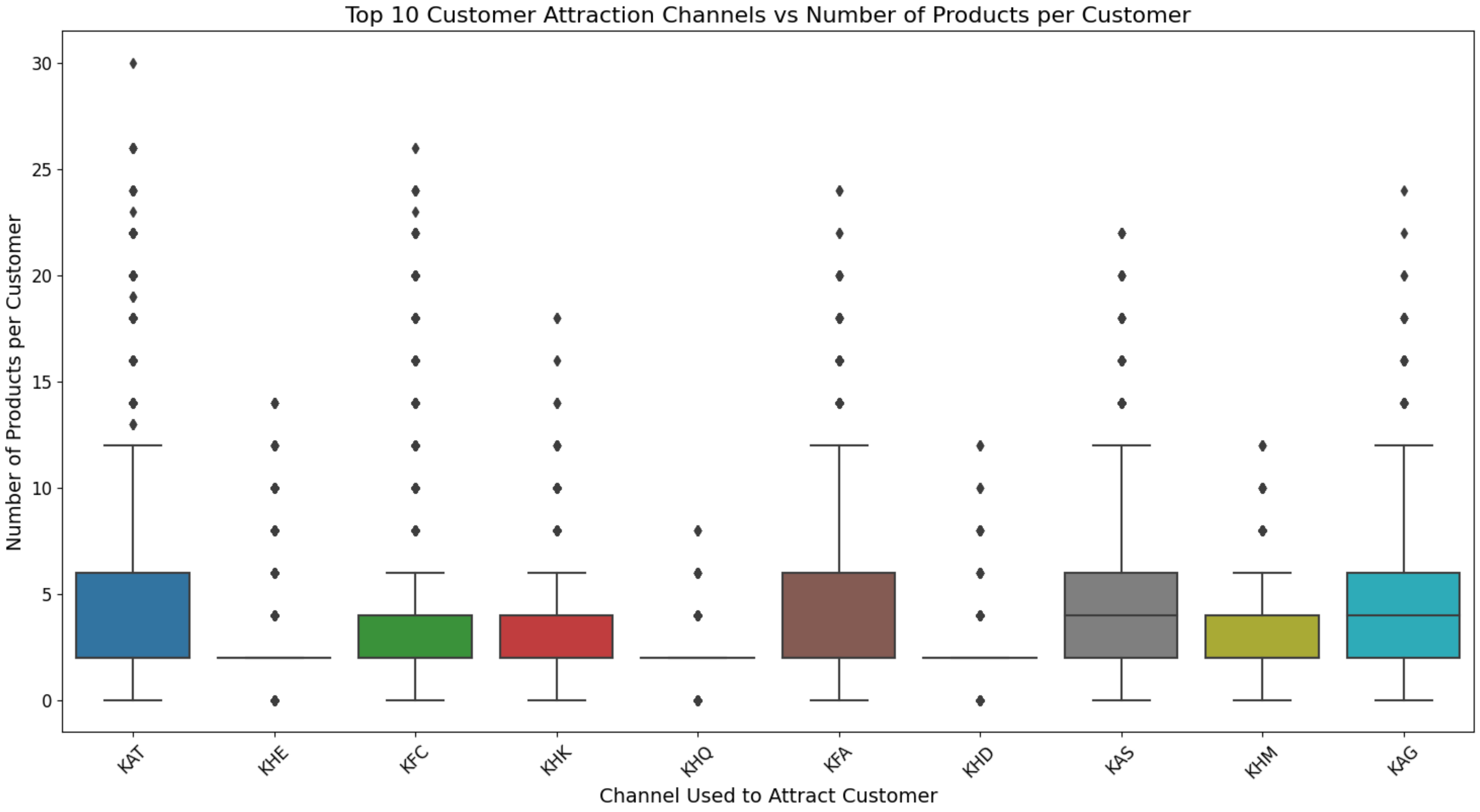
2. Customer attraction by channel

```
# Customer attraction by channel        Show code        [18]

KHE    0.306298
KAT    0.255559
KFC    0.239331
KFA    0.033062
KHQ    0.022468
KHK    0.017727
KHD    0.008926
KAS    0.006966
KHM    0.006582
KAG    0.006272
RED    0.005378
KAA    0.005255
KHN    0.005210
KAY    0.005118
KAB    0.005019
Name: channel_used_join, dtype: float64
```

- We can see that the most popular customer attraction channels are : 'KHE', 'KAT' and 'KFC'.

3. Customer attraction by channel and products



- It appears that customers who were attracted through channels such as 'KHE', 'KAT', and 'KFC' tend to have a higher number of products per customer compared to those who were attracted through other channels.
- However, it's important to note that there are outliers in each channel group, indicating that there are customers who have a high number of products regardless of the channel used to attract them.
- Overall, this visualization suggests that the customer attraction channel may have some influence on the number of products per customer, but it's not the only factor at play.

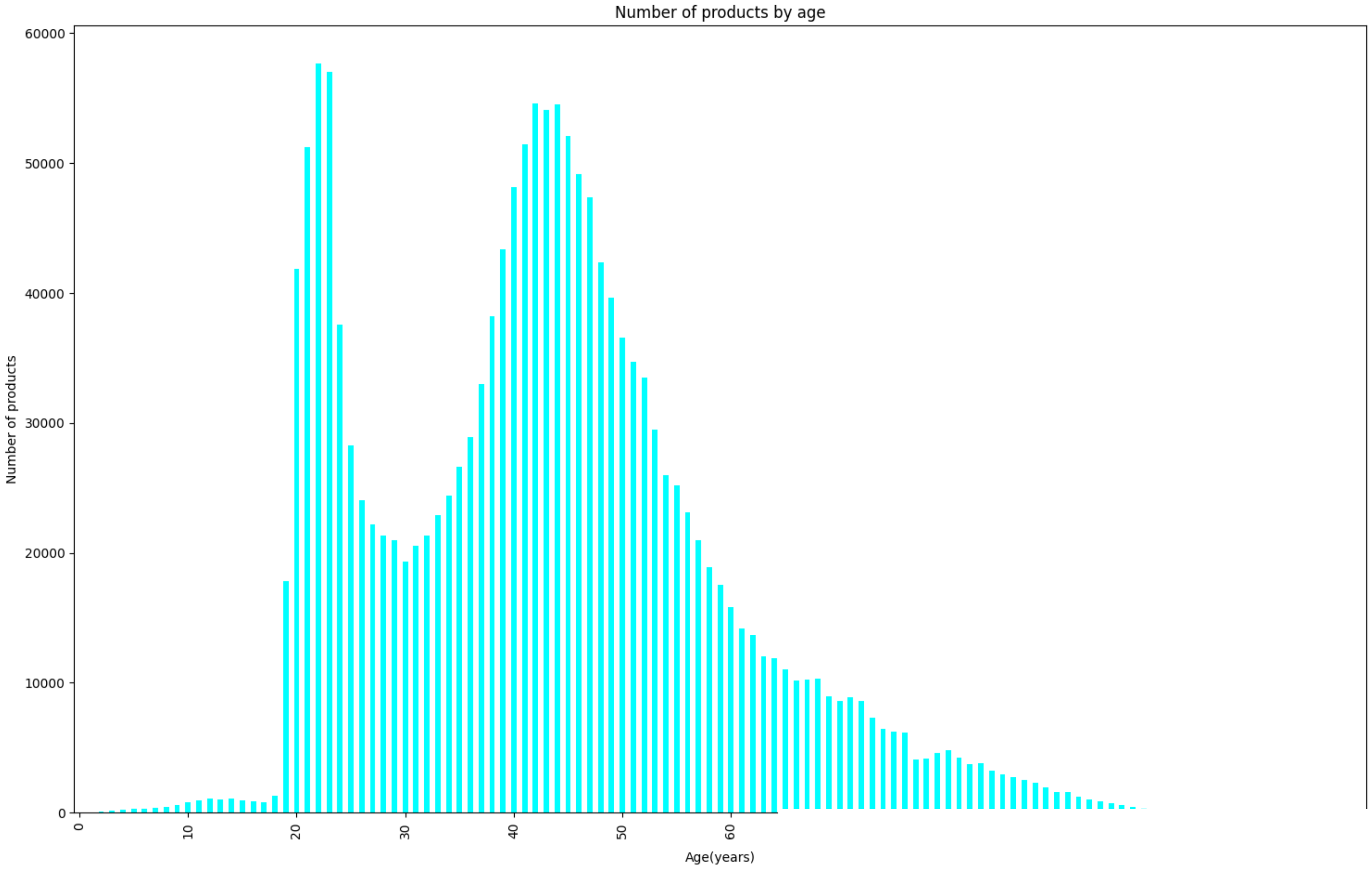
4. Total number of products per customer

```
# Total number of products per customer Show code                [28]

1.0     556985
0.0     190776
2.0     159808
3.0      65472
4.0      38710
5.0      25443
6.0      18613
7.0      12570
8.0       6835
9.0       3041
10.0      1201
11.0       402
12.0       105
13.0        24
15.0         1
Name: total_products, dtype: int64
```

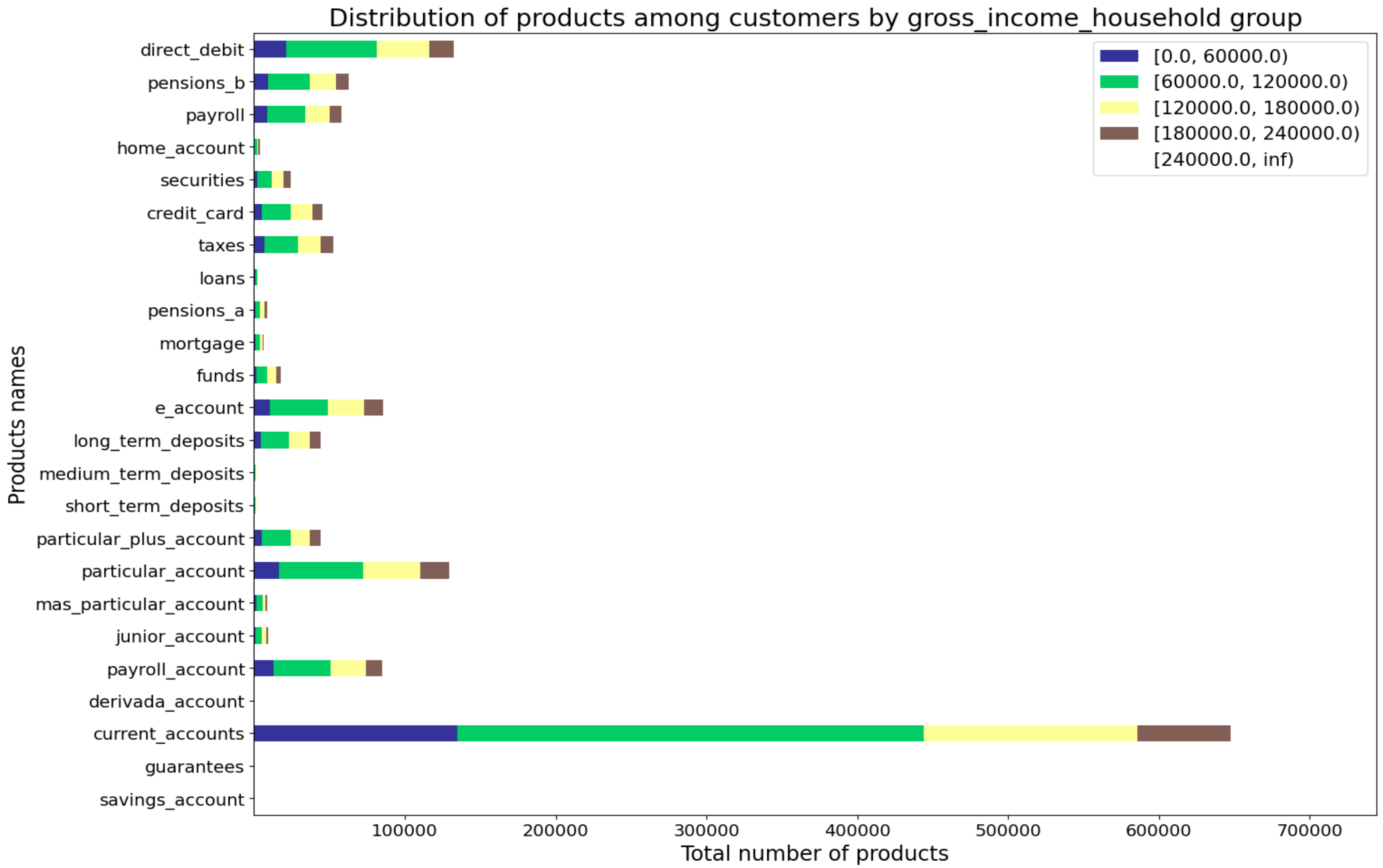
- Most of the customers used one or two products and rarely use more than five products.

5. Total number of products by age



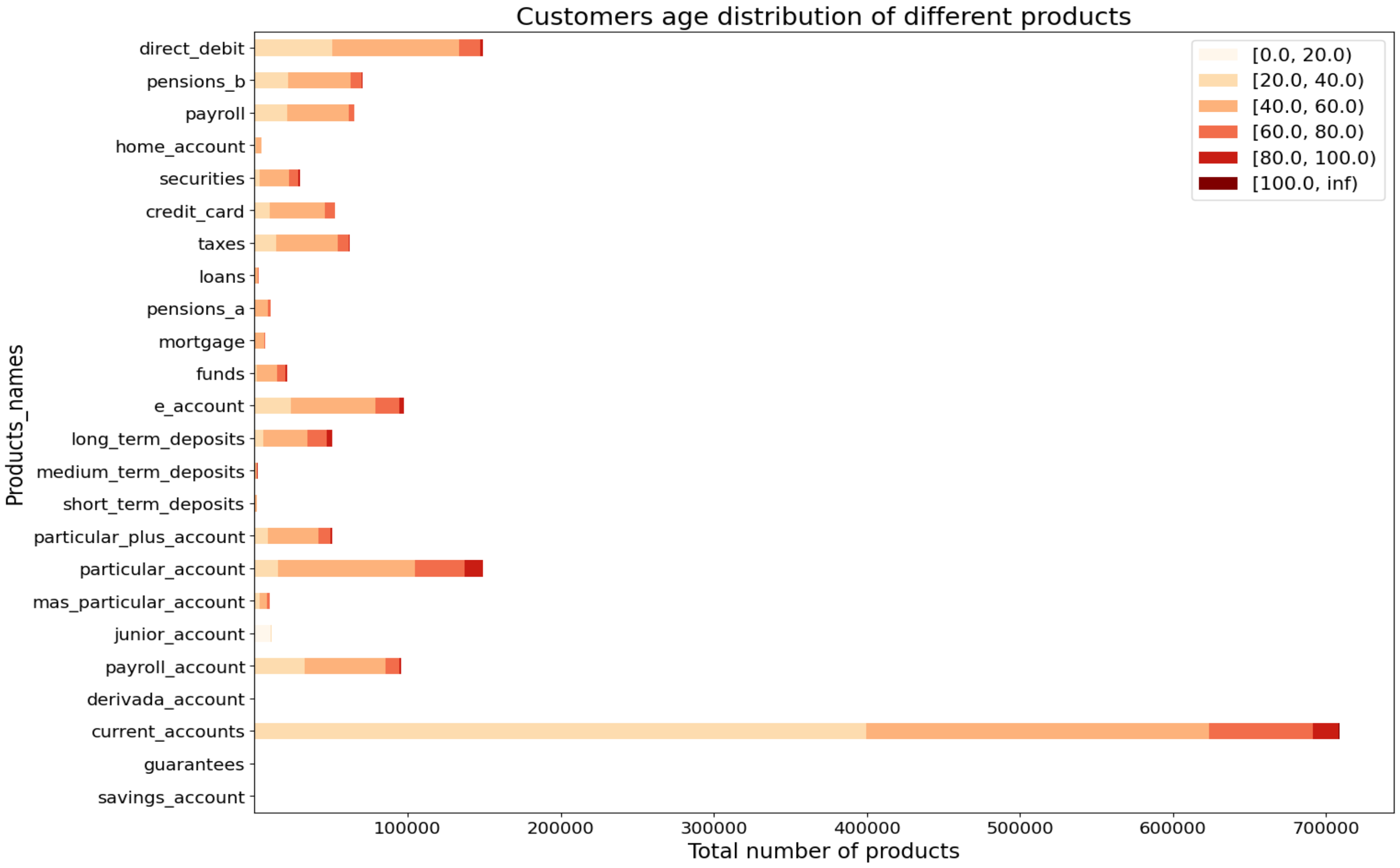
- As we see bimodal distribution with most of the products used by middle aged customers between 35 and 55 years old, followed by young customers in thier twenties.

6. Total number of products by income



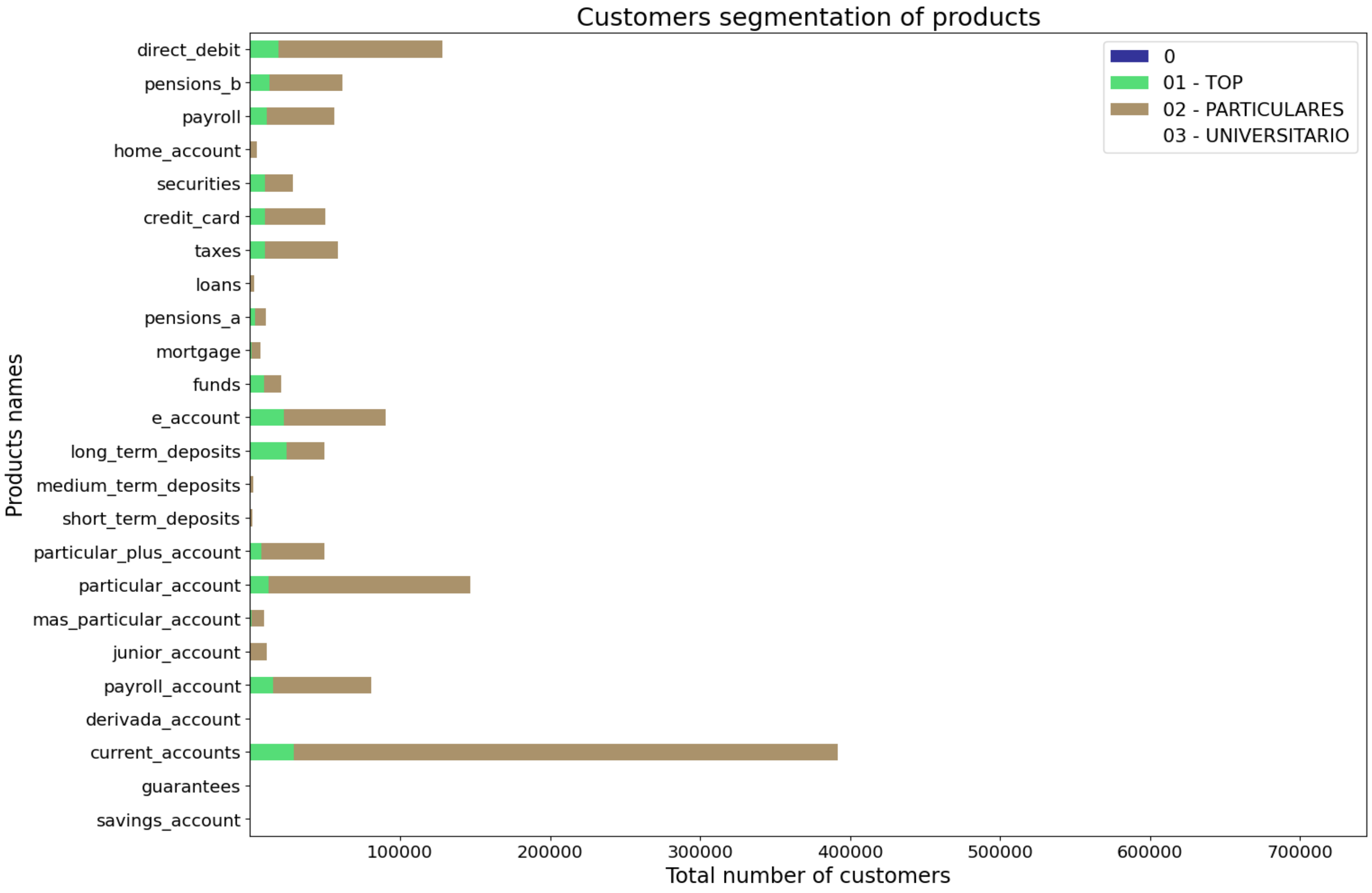
- From the visualization, we can infer that customers with higher gross income tend to have more products with the bank. The (60,000-120,000) group has the highest number of products, while the group with the lowest gross income (above 240,000) has the lowest number of products.
- The most popular products among all income groups are current accounts, followed by particular accounts and direct debit accounts.
- The least popular products are loans and mortgages, which are only held by a small percentage of customers across all income groups.

7. Total number of products by age group



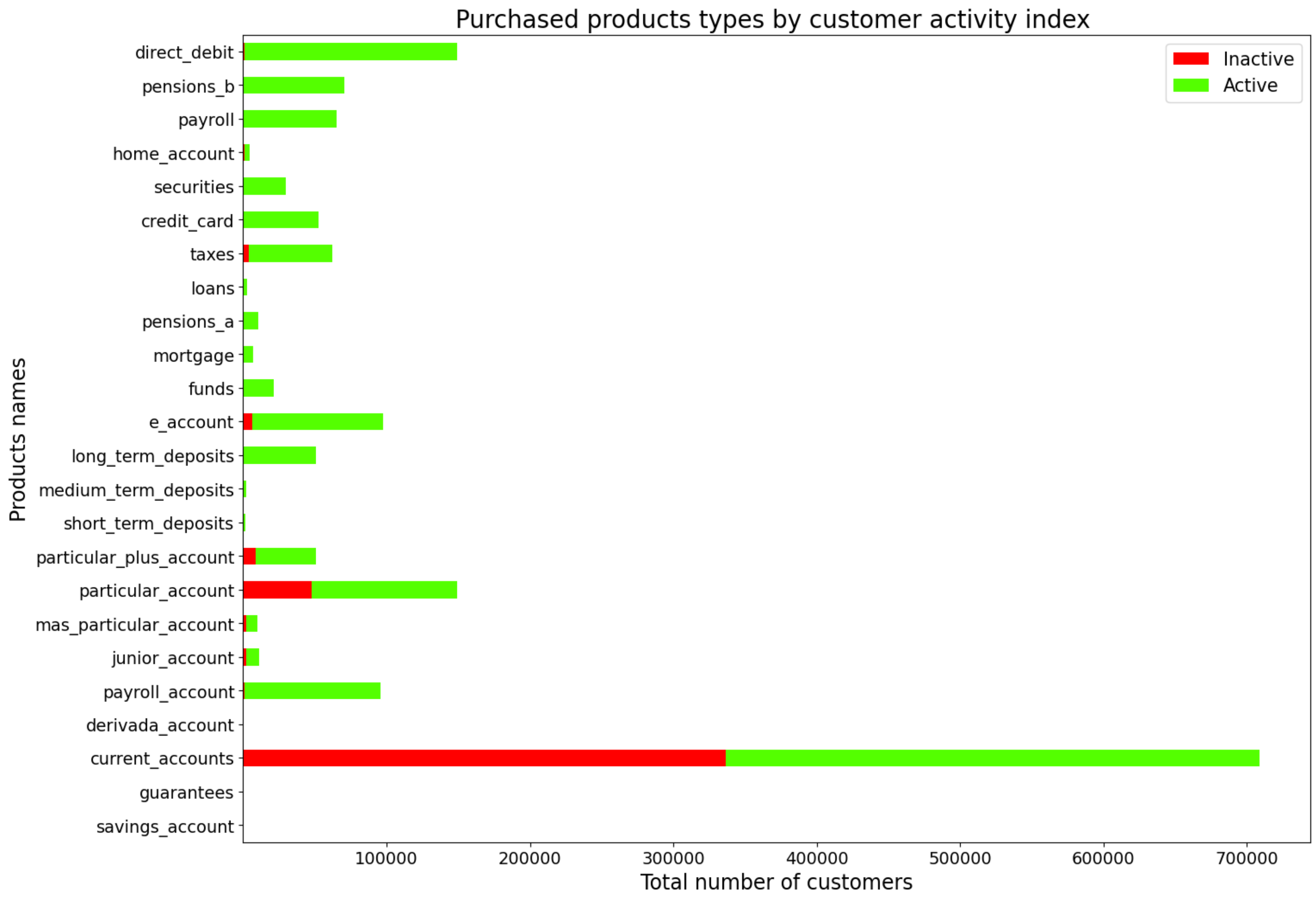
- The majority of customers who have a mortgage are between 40 and 60 years old.
- Customers who have a payroll account, pensions, and direct debit are mostly between 20 and 60 years old.
- Customers who have short-term deposits, medium-term deposits, and long-term deposits are mostly between 40 and 80 years old.
- Customers who have securities are mostly between 60 and 80 years old.
- Customers who have e-account, funds, and credit card are mostly between 20 and 40 years old.
- Customers who have home account, taxes, and loans are mostly between 20 and 60 years old.

8. Distribution of products by segment



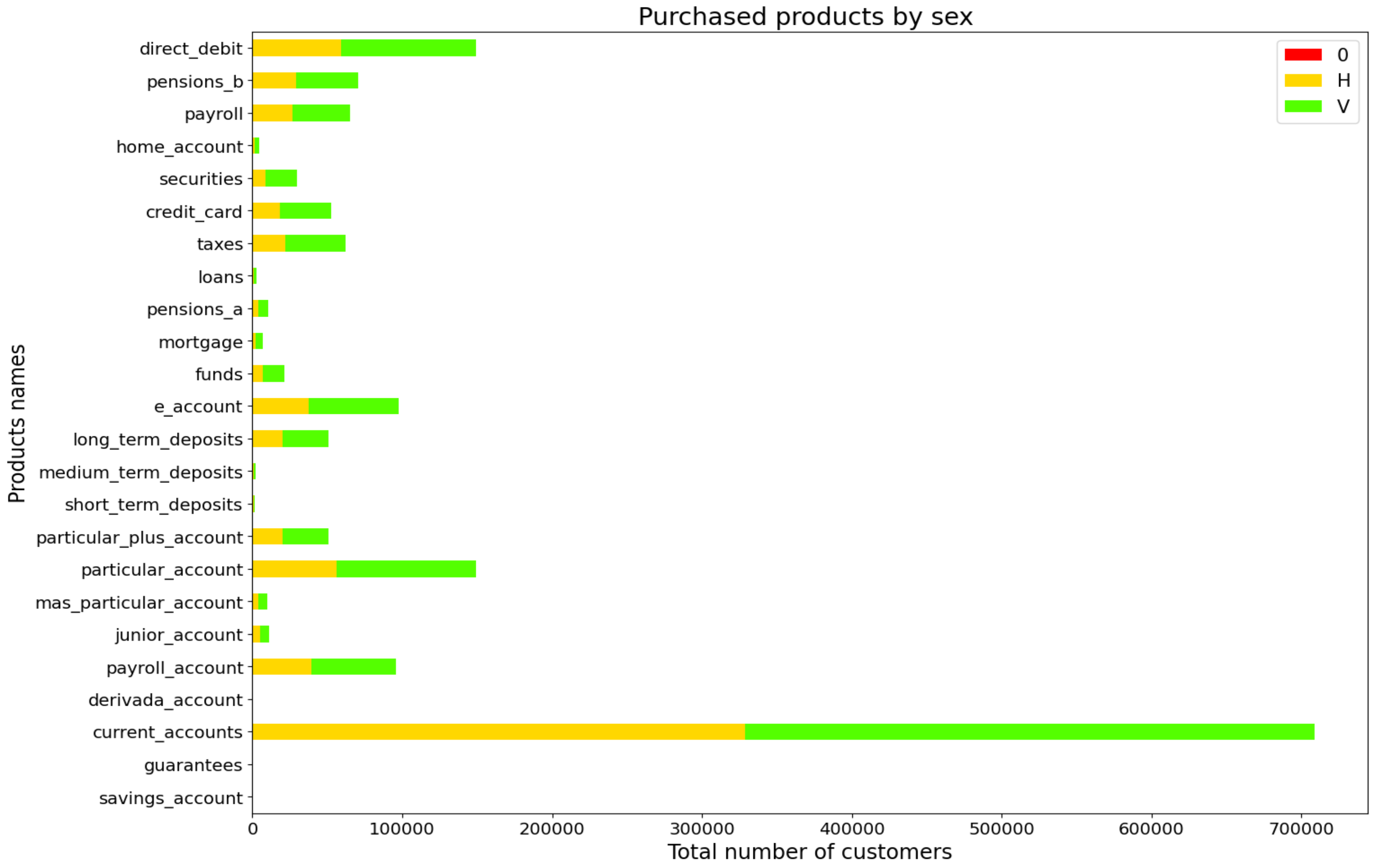
- The majority of customers for each product belong to the "02 - PARTICULARES" segment, followed by the " segment.
- The "03 - UNIVERSITARIO" segment has the lowest number of customers for each product.
- The distribution of customers across segments varies for each product, with some products having a more even distribution across segments while others are dominated by a single segment.
- Overall, the "02 - PARTICULARES" segment has the highest number of customers across all products, followed by the "01 - TOP" segment.

9. Distribution of customers by activity index



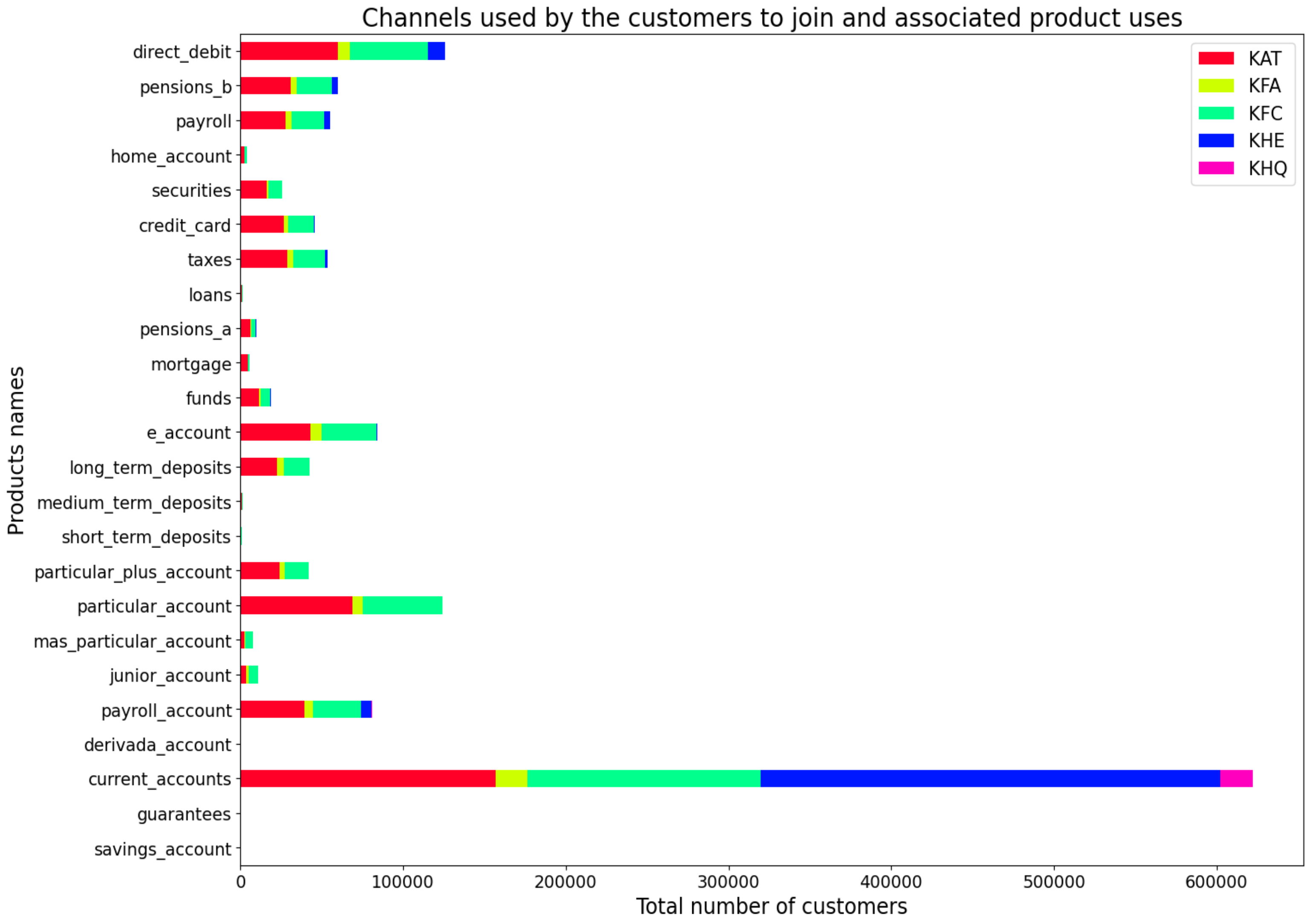
- The "Active" customers have a higher number of purchased products across all product types compared to the "Inactive" customers.
- The "Active" customers have a higher number of "long_term_deposits", "funds", "mortgage", "pensions_a", "loans", "credit_card", "securities", "home_account", "payroll", "pensions_b", and "direct_debit" products compared to the "Inactive" customers.
- On the other hand, the "Inactive" customers have a higher number of "savings_account", "guarantees", "current_accounts", "derivada_account", "payroll_account", "junior_account", "mas_particular_account", "particular_account", "particular_plus_account", "short_term_deposits", "medium_term_deposits", and "e_account" products compared to the "Active" customers.

10. Distribution of products by Sex



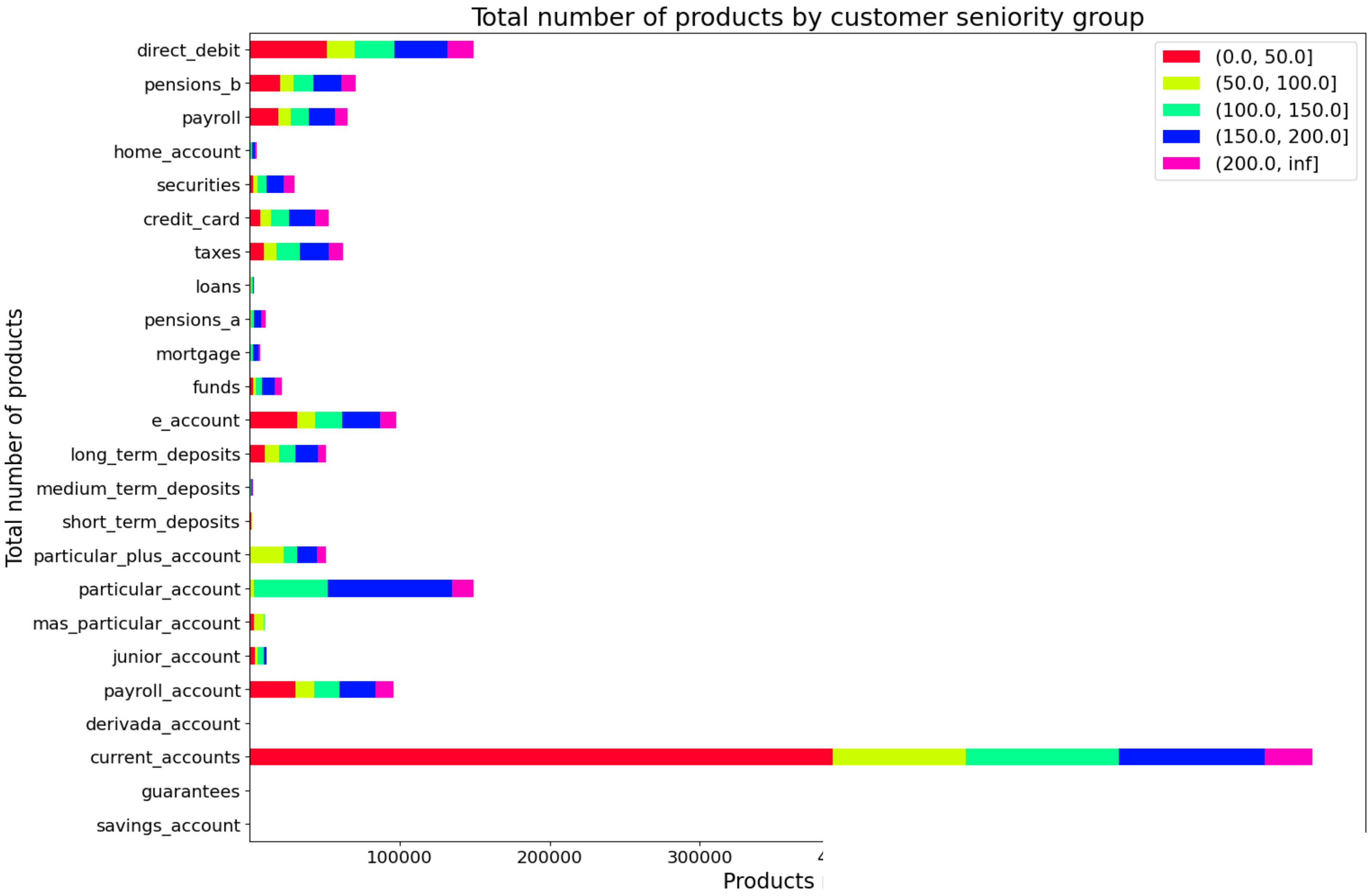
- We can infer that the distribution of purchased products types varies by customer sex.
- The number of purchased products is higher for female customers compared to male customers across all product types.
- Overall, the most purchased products by both male and female customers are "particular_account", "current_accounts", "e_account", and "direct_debit" products.

11. Distribution of products by channel join



- we can infer that the channels used by customers to join the bank are associated with the types of products they purchase.
- Customers who joined through channels KHE, KAT, KFC, KFA, and KHQ tend to purchase more "particular_account", "current_accounts", "e_account", and "direct_debit" products compared to other products.
- This information can be useful for the bank to target specific products to customers based on the channels they used to join.

12. Distribution of products by seniority group



- From the visualization above, we can infer that the total number of products purchased by customers is positively correlated with their seniority group.
- Customers who have been with the bank for a longer period tend to purchase more products compared to those who have been with the bank for a shorter period.
- This information can be useful for the bank to target specific products to customers based on their seniority group.

Final Recommendations

Based on the exploratory data analysis section above, the following recommendations can be made to increase cross-selling of the banking products:

1. **Target specific products to customers based on their seniority group.** Customers who have been with the bank for a longer period tend to purchase more products compared to those who have been with the bank for a shorter period. Therefore, the bank can target lower seniority groups with specific products that are likely to be of interest to them.
2. **Offer incentives or discounts to customers who purchase multiple products.** This can encourage customers to purchase more products and increase cross-selling. Based on analysis above customers who have purchased one product are more likely to purchase another product. Therefore, offering incentives or discounts to customers who purchase multiple products can increase the likelihood of cross-selling.
3. **Analyze the products that are frequently purchased together and create bundled packages to offer to customers.** This can increase the likelihood of customers purchasing multiple products. For example, based on the analysis above, customers who have a savings account are more likely to have a mortgage and a pension plan. Therefore, the bank can create bundled packages that include these products to encourage customers to purchase multiple products.
4. **Provide personalized recommendations to customers based on their transaction history and purchase behavior.** This can increase the relevance of the recommendations and encourage customers to purchase more products. Based on the analysis above, customers who have a salary account are more likely to have a credit card and a direct debit. Therefore, the bank can provide personalized recommendations to customers who have a salary account based on their transaction history and purchase behavior to encourage them to purchase a credit card and a direct debit.
5. **Conduct targeted marketing campaigns to promote specific products to customers who have not yet purchased them.** This can increase awareness of the products and encourage customers to try them out. The EDA analysis shows that there are some products that have low purchase rates, such as securities and funds. Therefore, the bank can conduct targeted marketing campaigns to promote these products to customers who have not yet purchased them.

ML Model Recommendations

Based on the dataset above, the following machine-learning models are recommended to increase cross-selling of banking products:

1. **Logistic Regression** - This model is useful for predicting binary outcomes, such as whether a customer will purchase a banking product or not. It can also provide insights into which variables are most important in predicting the outcome.

2. **Random Forest** - This model is useful for predicting multiple outcomes and can handle both categorical and continuous variables. It can also provide insights into which variables are most important in predicting the outcome.

3. **Gradient Boosting** - This model is useful for predicting multiple outcomes and can handle both categorical and continuous variables. It can also provide insights into which variables are most important in predicting the outcome.

The basis for selecting these models is their ability to handle both categorical and continuous variables, provide insights into variable importance, and handle multiple outcomes. Additionally, these models have been shown to perform well in similar banking industry applications.

Thank You

Data Glacier
Your Deep Learning Partner