Project: Customer Segmentation

Week 10: Deliverables

Group Name: cust_seg

Specialization: Data Science

Submitted to: Data Glacier

Internship Batch: LISUM12

Team Member Details:

- Brennan Clinch, bclincher98@gmail.com, USA, North Carolina State University, Data Science
- Rohit Sunku, rgs8890@gmail.com, UK, Le Wagon, Data Science
- Kutay Selçuk, kutay.selcuk@ozu.edu.tr, Turkey, Ozyegin University, Data Science
- Zhan Shi, zhanshi@g.ucla.edu, USA, University of California Los Angeles, Data Science

Problem Description:

XYZ bank wants to roll out Christmas offers to their customers. But Bank does not want to roll out the same offer to all customers. Instead, they want to roll out personalized offers to a particular set of customers. If they manually start understanding the category of the customer then this will be not efficient and also they will not be able to uncover the hidden pattern in the data (pattern which group certain kind of customer in one category). Bank approached ABC analytics company to solve their problem. Bank also shared information with ABC analytics that they don't want **more than 5 groups** as this will be inefficient for their campaign.

ABC analytics assigned this talk to their analytics team and instructed their team to come up with the approach and feature which group similar behavior in one category and others in a different category.

Github Repo Link

https://github.com/Brennan-Clinch/Customer-Segmentation-Project

EDA Performed on Data

Since our goal is to divide our customers up into no more than 5 groups, we were looking to do exploratory data analysis on the customers and then use our findings to help with dividing them up into 5 different segments when we do our modeling later. The areas we looked at for the EDA are mentioned below along with their results.

1) What were the most used products by customers?

 The products that were used the most by customers were those who had a current account, followed by a Particular account, then Direct debit, along with e-account and Payroll account.

2) How many total customers are in the bank overall?

The total number of customers with a customer ID are 510,617.

3) Channel with highest number of customers

The KAT channel is the channel used by the most customers as when we did our cleaning we replaced the null values with the KAT channel as it was the most popular channel.

4) What age range has the most amount of customers? And if the age ranges have an effect on the type of products being bought.

The age range with the most amount of customers is those who are between 18-30. The age range does have a large effect on the type of products being bought as those between 18-30 had the highest percentage of Current accounts (60-80%) compared to those over 30 at around < 40% for those between 40-65 and over 65.

5) What products are most used based on the seniority of the customer from when they started?

Customer seniority has a large effect on the amount of products being used by the customers. Current accounts are most used by young people, or those with 0 and 50 months of seniority.

The amount of customers with a Particular account does increase as the seniority increases but eventually decreases when they are older than 200 months.

6) Does the city of the customer have an effect on the products being bought?

Yes, as Madrid is the most popular city with the most amount of customers.

7) What effect does gender have on the products being bought?

Males have lower product usage than females, so special offers could be made to males.

8) Income by city, age.

While Madrid had the most amount of customers therefore it easily took the place for most income groups, the city with the most customers with income less than 30,000 was Valencia. So this indicates that certain offers could be given to customers who make less than 30,000 and live in Valencia.

For age, since the most popular group is those aged 18-30, they got most incomes starting at less than 30,000 and then between 30,000 and 50,000. Customers that got the higher salaries were between 41 and 65.

Final Recommendations

In order to group our customers into 5 separate clusters for our modeling which is the next step we can use our recommendations to help.

- We can focus on ages that are older than 18-30 as that is where most of the customers are.
- Males have lower product usage than females, so giving away special offers to males could be a strategy.
- Since Spain has the most amount of customers overall (mainly just Spain for the data we have) along with Madrid being the most popular city, the bank can focus on other cities/areas in Spain.
- The bank can focus on individuals separately with regards to their seniority with certain offers given to people with middle and higher seniority and then to people with lower seniority.