



Data Glacier

Your Deep Learning Partner

Customer Segmentation Project

Virtual Internship

30-Aug-2022

Group Information

Group Name: M.A.S

Specialization: Data Science

Submitted to: Data Glacier canvas platform

Internship Batch: LISUM10: 30

Group Members	Three members		
Name	Email	Country	Collage/Company
Moath Mohammed Bin Musallam	moathmusallam@gmail.com	Saudi Arabia	University of East Anglia
Andrew Kojo Mensah-Onumah	kojoakmo@gmail.com	Ghana	Data Glacier
Shaimaa Saleh Obad Al-khawlani	s.khawlany@gmail.com	Yemen	Data Glacier

Problem description

Most banks around the world have variant large customer base with different income levels, ages, characteristics, values and lifestyles. XYZ bank wants to increase the production and the satisfactions of all customers categories by roll out Christmas offers to their customers. But Bank does not want to roll out same offer to all customers instead they want to roll out personalized offer to particular set of customers. If they manually start understanding the category of 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).

Objectives:

To increase the production and the satisfactions of all customers categories by roll out Christmas offers to their customers.

Data Analysis

Overview on Data

- Data provided by bank XYZ contained a million client user data.
- Data covers bank accounts created between 16/1/1995 and 27/2/2015
- Date for recording data was 28/1/2015 and 28/2/2015
- Data covers bank user details and services offered by bank. Grouping data this way is essential to solving the problem statement which is 'Target advertisement' during Christmas
- Out of the million customers, 993,000 of them were from Spain. 96% of the Spanish people were also based in Spain. Thus, market niche was focused on customers in Spain as they provided a good sample of the entire data
- XYZ bank offers 22 different products to its clients

Key Insights

Client Demographics

- Majority of customers are between 40 and 80 as that portion has dense concentration
- A very large proportion of customers have their income limits between 0 and 5,000,000
- Customers in the age group of 20, and 40-50 use the most products
- On average, customers use only 2 bank products; current account and particular account
- There are more female bank users than male, with females shading 57% of the customers
- Most bank users are found in the following provinces: Madrid, Barcelona, Valencia, Sevilla and Zaragoza
- Fewest bank users were found in the following provinces: Gipuzkoa, Bizkaia, Navarra, Ceuta and Melilla
- Top 3 channels used by customers to join are KAT, KFC and KHE
- The customer using the most bank products is Customer 797816
- 92.7% of customers do not pay taxes
- 99.9% of customers have no guarantees (or guarantors)

Key Insights

Product Demographics

- The most used products are Current account, Particular account, Direct debit, e-account and payroll account
- The least used products are Saving account, Derivada account, Short-term deposit, Medium-term deposit and Loans

Methodology

The technical recommendations will be to use the following methods. Since the data have no **dependent value**, **unsupervised learning** techniques should be applied.

The elbow method:

- The elbow method uses the sum of squared distance (SSE) to choose an ideal value of k based on the distance between the data points and their assigned clusters, we use this method to classify the number of clusters based on the data we have.

PCA:

- Principal component analysis (PCA) is a widely used statistical technique for unsupervised dimension reduction. It is believed that it improves the clustering results in practice (noise reduction)

Clustering:

- Clustering is the task of dividing data points into a number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups. And we have **5 clusters** based on the elbow method results.

Cluster Groups of customers

1st Cluster

Males only

Can be of any customer type at the beginning of the month

Have no guarantees

Use a maximum of 5 bank products

Older customers

2nd Cluster

Both sexes present

Largest income holder here is 30,000,000

Use a maximum of 11 bank products

Has the most customers

Most are customer relation type P at the beginning of the month

3rd Cluster

Both sexes present

Few teenagers and few older ones (above 100)

Largest income holder here is 25,000,000

Uses minimum of 2 and maximum of 15 bank products

Least represented group

Customers with least seniority in months

4th Cluster

Males only

Most are customer relation type A and I at the beginning of the month

Most are customer type 1 at the beginning of the month

Largest income holder here is 25,000,000

Use a maximum of 10 bank products

Fairly large representations in all 6 major channels

5th Cluster

Both sexes present

Largest income holder here is 20,000,000

Have no guarantees

Use a maximum of 5 bank products

Recommendations

1. According to the age insights it is highly recommended to focus on ages that is younger than and older than twenties ages.
2. Males has lower product usage than Females, so we recommend to give special offers for Males.
3. We can say that Spain is the only Country which uses bank's products, the bank has to focus on other countries.
4. As per the insights most of the customers are in Madrid then Barcelona, the other Cities are very low so special offers has to proposed to these cities to increase the customer numbers.
5. the bank has to give special offers for the products rather than current account since it is the most used product.
6. Special offer should be handed over to customers using over 12 products and especially the customer using 15 bank products
7. Least used products should be reconsidered

M.A.S

Thank You



Data Glacier

Your Deep Learning Partner