Customer Segmentation Project

Week 7

JULY 18, 2022

Contents

1. 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

2. 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).

3. Business understanding

- **Business Problem:**

XYZ bank wants to 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). Bank approached ABC analytics company to solve their problem. Bank also shared information with ABC analytics that they don't want **more than 5 group** as this will be inefficient for their campaign. The ABC analytics team's proposal to use Customer Segmentation, which is the process of dividing customers into groups based on common characteristics so companies can market to each group effectively and appropriately.

- **The Data:**

The existing data, which was provided by the bank, is the bank's customers data. However, the data contains many columns that will help the analytics team analyze the data and build a customer segmentation approach for the bank.

Since the data does not contain a dependent variable or (Target), We believe that machine learning (clustering) techniques would be appropriate to use for this type of data.

Size: 1000000 records, 48 columns.

- **Columns Description:**

Column Name	Description
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)
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 (Potential)
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
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
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

4. Project life cycle along with the deadline

The project's general view, along with the deadline, is described in the table below. The deadline for week 7 is 19 July 2022. The following week is added up accordingly.

Task Name	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13
Data understanding and exploration							
Feature engineering							
EDA							
EDA presentation							
Segmentation techniques							
Model selection and building							
Code and report submission							

Data Understanding and Exploration

1. Check for missing values
2. Check for duplicates
3. Check for outliers
4. Check skewness of data
5. Inspect data types
6. Explore individual columns, etc

Feature Engineering

1. Impute missing values
2. Create new columns as needed, etc

EDA and EDA presentation

Segmentation Techniques

1. RFM
2. Cohort Analysis
3. K Means Segmentation, etc

5. Data Intake Report

The dataset was downloaded from a Google Drive link provided by Data Glacier.

Name: Customer Segmentation

Report date: 18/07/2022

Internship Batch: LISUM10: 30

Version:1.0

Data intake by: M.A.S Group

Data intake reviewer: Data Glacier

Data storage location: https://drive.google.com/drive/folders/1bfCpJIKmp6IHxiLPWvOS2nU1dc24pViB

Tabular data details:

Total number of observations	1000000
Total number of files	1
Total number of features	48
Base format of the file	CSV
Size of the data	154 MB

6. GitHub Repo Link

The link for GitHub: https://github.com/kojomensahonums/Customer-segmentation-with-Data-Glacier