

Week 9 – The challenges Document

Name: [Fatimah Asiri](#)

Username: alassirifatima@gmail.com

Batch code: [LISUM14 30 Sep – 30 Dec 2022](#)

Submission Date: [12/02/2022](#)

Submitted to: [Data Glacier](#).

This document contains the challenges on the
project

Data Science

Customer Segmentation

Group Project

Prepared by:
Fatimah Asiri

Problem Statement:

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 (a pattern that group certain kinds of customers 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.

ML Problem:

ABC analytics proposed a customer segmentation approach to Bank.

ABC analytics assigned this task to their analytics team and instructed their team to come up with an approach and feature that groups similar behavior customers in one category and others in a different category.

The challenges:

1- Change the columns names from Spanish to English.

fecha_dato	Date
ncodpers	Customer code
ind_empleado	Employee Index
pais_residencia	Country residence
sexo	Sex
fecha_alta	Register Date
ind_nuevo	New customer
antiguedad	Customer seniority
indrel_1mes	Customer type
tiprel_1mes	Customer relation type
indresi	Residence index
indext	Foreigner index
canal_entrada	channel
indfall	Deceased index.

tipodom	Addres type
cod_prov	Province code
nomprov	Province name
ind_actividad_cliente	Activity index
renta	Gross income
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	More 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	Nom Pensions
ind_recibo_ult1	Direct Debit

2- Handling missing data by mode for object and by the median for number.

3- Distribution for each column to get a better vision of data.

4- EDA Step contains distribution for same columns.