



Project: Customer Segmentation

Week 7: Deliverables

Team Member Details:

- Brennan Clinch, bclinch98@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/Business Understanding:

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 task 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.

Project Lifecycle (Deadlines)

Task	Beginning Date	End Date
Problem understanding and data investigation (week 7)	9/13/2022	9/19/2022
Data understanding and beginning of data cleaning (week 8)	9/19/2022	9/26/2022
Finish data cleaning (week 9)	9/27/2022	10/2/2022
EDA (week 10)	10/3/2022	10/9/2022
Finish EDA and have proposed modeling technique (week 11)	10/10/2022	10/16/2022
Model selection and building (week 12)	10/17/2022	10/23/2022
Final Presentation (week 13)	10/24/2022	10/30/2022

Data Intake Report

Name: Customer Segmentation

Report date: 18.09.2022

Internship Batch: LISUM12

Version: 1.0

Data intake by: Kutay Selçuk

Data intake reviewer: Brennan Clinch, Rohit Sunku and S.Z.

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

Tabular data details:

Total number of observations	1,000,000
Total number of files	1
Total number of features	48
Base format of the file	.csv
Size of the data	366.2+ MB

Proposed Approach:

- There is no duplication in file.
- There are null values in file as they are mentioned below.

Column Name	Number of NaN
ind_empleado	10782
pais_residencia	10782
sexo	10786
fecha_alta	10782
ind_nuevo	10782
indrel	10782
ult_fec_cli_1t	998899
indrel_1mes	10782

tiprel_1mes	10782
indresi	10782
indext	10782
conyuemp	999822
canal_entrada	10861
indfall	10782
tipodom	10782
cod_prov	17734
nomprov	17734

ind_actividad_cliente	10782
renta	175183
ind_nomina_ult1	5402
ind_nom_pens_ult1	5402