



Data Glacier

Your Deep Learning Partner

CROSS SELLING RECOMMENDATION

Group Name

LISP01-Data-Analysts

Name, Email, Country, College/Company, Specialization

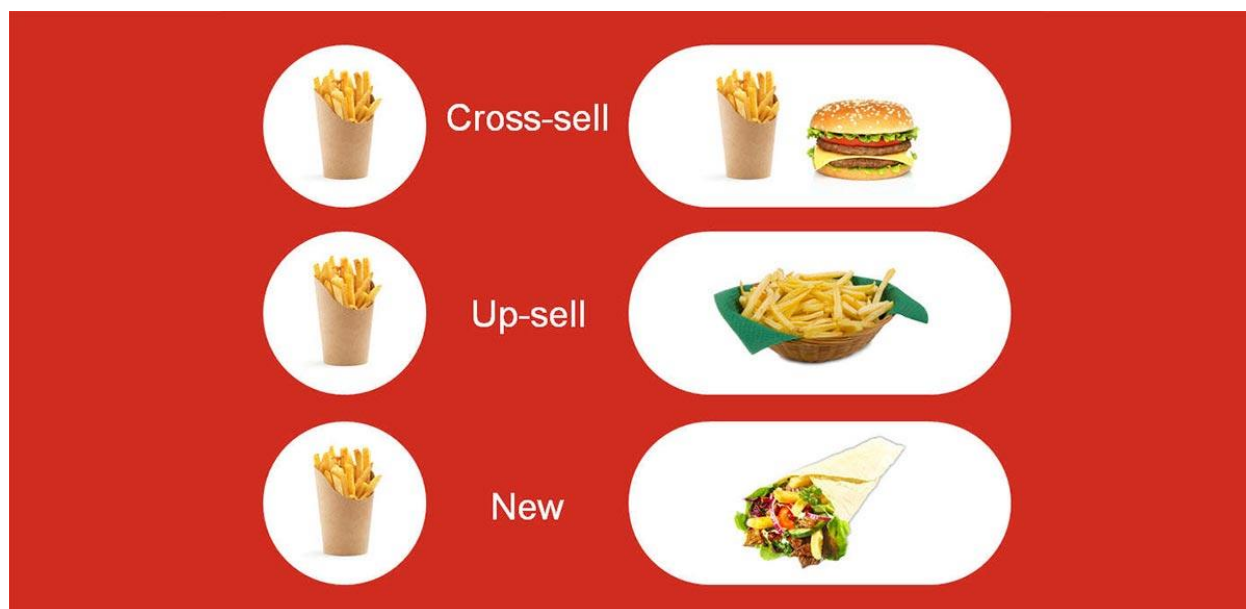
Lauro Ribeiro, lauroc.r.volei@hotmail.com, Ireland, Data
Glacier, Data Analyst

Lasisi Salmah, lasisisalmah52@gmail.com, Nigeria, Data
Glacier, Data Analyst

Buse Gungor, busegungor2303@gmail.com, Turkey, Data
Glacier, Data Analyst

Problem Description

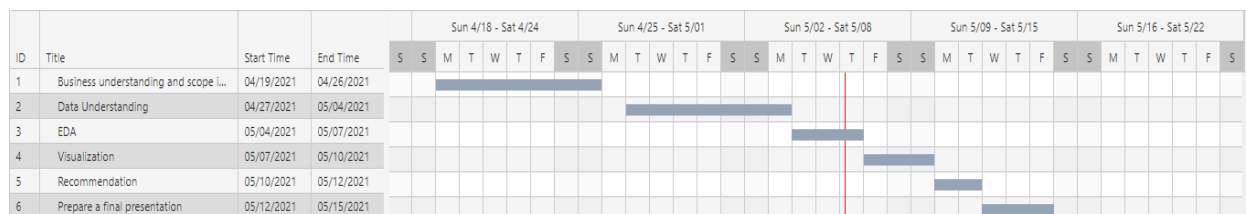
XYZ credit union in Latin America is performing very well in selling the Banking products (eg: Credit card, deposit account, retirement account, safe deposit box etc.) but their existing customer is not buying more than 1 product which means bank is not performing good in cross selling (Bank is not able to sell their other offerings to existing customer). As ABC analytics firm, need to inspect the data and suggest what action bank can take to increase cross selling without using ML.



Business Understanding

Many Bank's customers receive recommendations while others rarely see any, resulting in an uneven customer experience. Our task is to predict which products their existing customers will use next month based on their past behavior and similar customers.

Project Lifecycle



Data Intake Report

Name: Cross Selling Recommendation

Report date: 26.04.2021

Internship Batch: LISP01

Version: 1.0

Data intake by: Buse Gungor

Data intake reviewer: <intern who reviewed the report>

Data storage location: <https://drive.google.com/file/d/16-nzZR91-ijrfjUcl2PniTpOgrvFAYkA/view>

Tabular data details:

Dataset name	train
Total number of observations	13647309
Total number of files	<Number of files received>
Total number of features	48
Base format of the file	csv
Size of the data	4.9+ GB

Dataset name	test
Total number of observations	929615
Total number of files	<Number of files received>
Total number of features	24
Base format of the file	csv
Size of the data	170.2+ MB

- Contains the details of the raw data given in these tables.
- Although the train data is 4.5GB + in size, 2.5GB + part is used in order to be able to easily process on it and to easily observe and analyze the data.
- In the train dataset, 8 of the 48 variables are continuous numerical variables, 23 of them are discrete numerical variables. In categorical variables, 12 of them are nominal categorical variables, and 5 of them are ordinal variables. (raw data)
- In the test dataset, 2 of the 24 variables are continuous numerical variables, 7 of them are discrete numerical variables. In categorical variables, 13 of them are nominal categorical variables, and 2 of them are ordinal variables. (raw data)
- Columns with missing values in train dataset (raw data): ind_empleado, pais_residencia, sexo, age, fecha_alta, ind_nuevo, antiguedad, indrel, ult_fec_cli_1t, indrel_1mes, tiprel_1mes, index, indext, conyuemp, canal_entrada, indfall, tipodom, cod_prov, nomprov, ind_actividad_cliente, renta, segmento, ind_nomina_ult1, ind_nom_pens_ult1

- No missing values were observed in the test data.
- Columns are "ult_fec_cli_1t", "conyuemp" that are null percentage %99 in all train dataset. Drop these columns.
- Changing column names to make sense of the effect of variables on the problem.
- Converting columns containing dates to "DateTime".
- It has been observed that some columns that should be numeric variables are objects. These must be converted to the variable type.

Github Repo link: <https://github.com/LISP-01-DataAnalysts/DataGlacier-Internship-GroupProject>