Name: Deepthika Shiwani Muralikrishnan, Payal Upadhyay, Abida S Bhatti

Group Name: Team Data Analysts

Email: [deepthikamk@gmail.com](mailto:deepthikamk@gmail.com) , [payalupadhyay456@gmail.com](mailto:payalupadhyay456@gmail.com) , [sarwarabida123@gmail.com](mailto:sarwarabida123@gmail.com)

Country: India, South Africa, USA

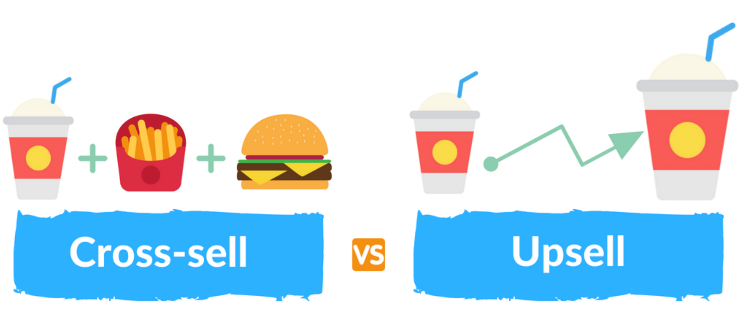
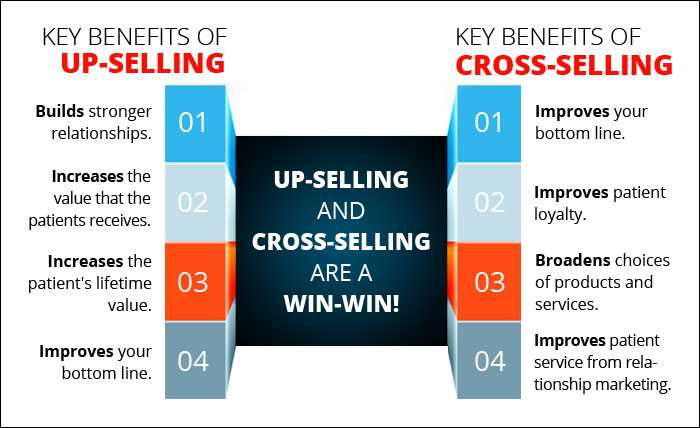
College/Company: -

Specialization: Data Analyst

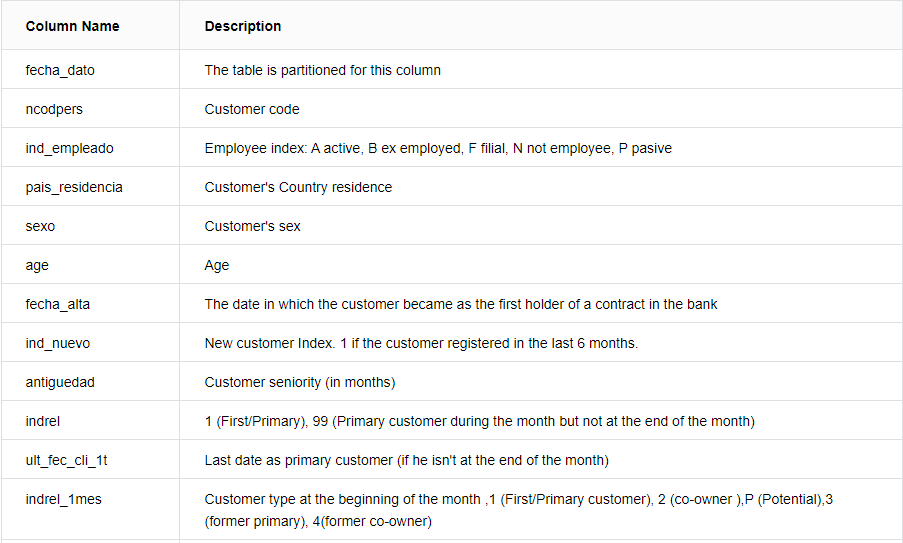
**PROBLEM DESCRIPTION**

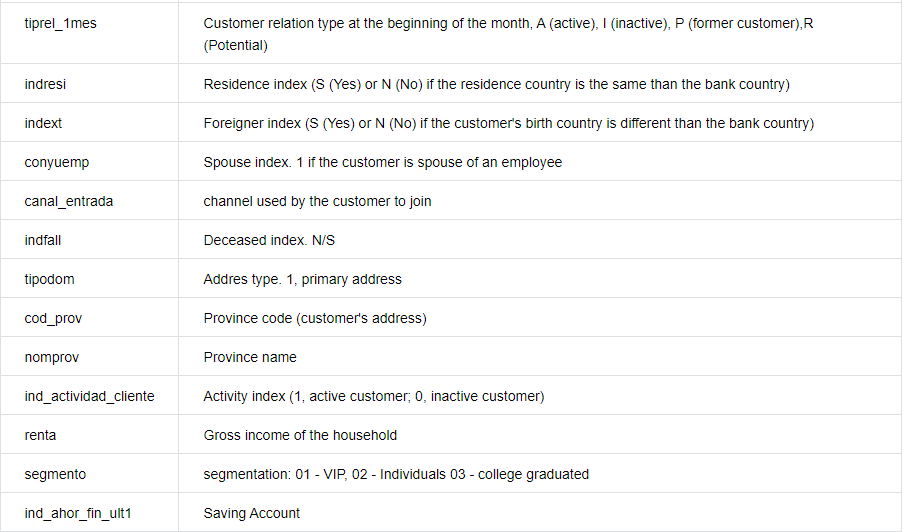
XYZ credit union in Latin America is performing very well in selling the Banking products (eg: Credit card, deposit account, retirement account, safe deposit box etc) but their existing customer is not buying more than 1 product which means bank is not performing good in cross selling (Bank is not able to sell their other offerings to existing customer). XYZ Credit Union decided to approach ABC analytics to solve their problem. ABC company came up with a framework which will be utilizing machine learning algorithm in the core to increase cross selling. But as a data analyst you need to inspect the data and suggest what action bank can take to increase cross selling (without using ML)

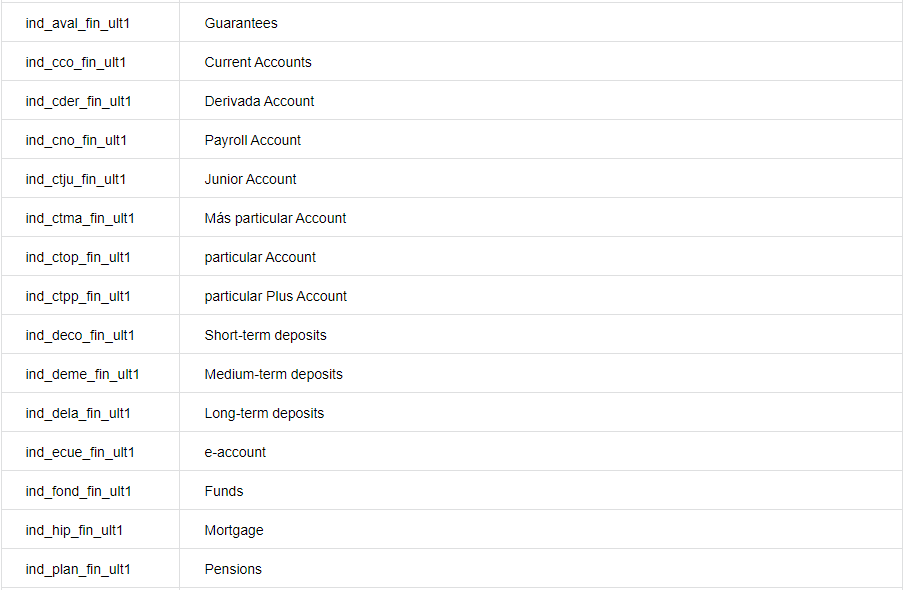


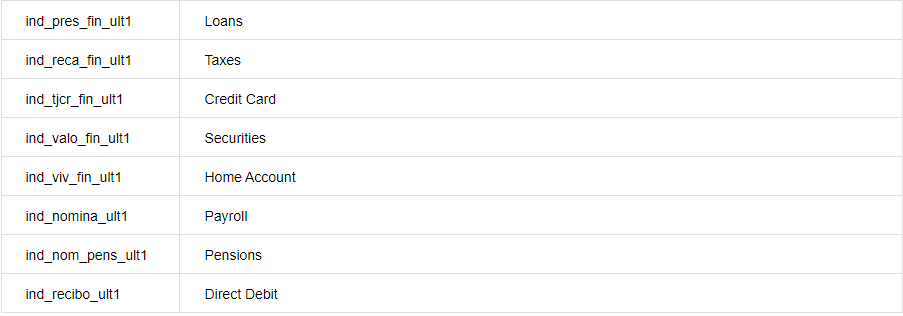


**DATA UNDERSTANDING**









There are 24 different types of products offered.

**PROBLEMS IN DATA**

**Train.csv**

Columns with NA values:

ind\_empleado 27734

pais\_residencia 27734

sexo 27804

fecha\_alta 27734

ind\_nuevo 27734

indrel 27734

ult\_fec\_cli\_1t 13622516

indrel\_1mes 149781

tiprel\_1mes 149781

indresi 27734

indext 27734

conyuemp 13645501

canal\_entrada 186126

indfall 27734

tipodom 27735

cod\_prov 93591

nomprov 93591

ind\_actividad\_cliente 27734

renta 2794375

segmento 189368

ind\_nomina\_ult1 16063

ind\_nom\_pens\_ult1 16063

20.47% of income values are NA

**Test.csv**

Columns with NA values:

ult\_fec\_cli\_1t 927932

indrel\_1mes 23

tiprel\_1mes 23

conyuemp 929511

canal\_entrada 2081

cod\_prov 3996

nomprov 3996

segmento 2248

Outliers:

* In train data set, the age group between 1-18 is only 0.8 % of the total data and have minimum impact in the overall analysis in regression.
* In train data set, age group between 80 -164 is only 2.7 % of the total data and have minimum impact in the overall analysis in regression.

**APPROACHES TO OVERCOME PROBLEMS**

In our data analysis we used following approaches to resolve NA values and outlier issues:

1.  To resolve NA issue in age column, create data set for average income of each group and apply it to each age row where income is ‘NA’.

Age Groups:

Group 1 (19-25 yr old customers):

E-account, saving account, short-term deposit, medium-term deposit, long-term deposit, loans, and credit cards.

Group 2 (26-45 yr old customers):

Mortgage, e-account, savings, credit cards, pension plan, loans, and securities.

Group 3 (46-65+ yr old customers):

E-account, loans, short-term deposit, savings, credit cards, securities, and pension plan.

2.  To resolve outlier, we will ignore age groups between 1-18 (.8% of dataset) and 85-164(2.7% of data set) as their impact is minimal on overall analysis.

3. Non-numeric missing values can be replaced with the mode (most occurring value) in those categories.

4. Delete products that nobody buys (only a few customers).

5. Delete rows where a lot/all values are missing.

Github repository link : <https://github.com/deepthikashiwani/Cross-selling-Recommendation-How-to-increase-cross-selling-of-Banking-Products>