

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)

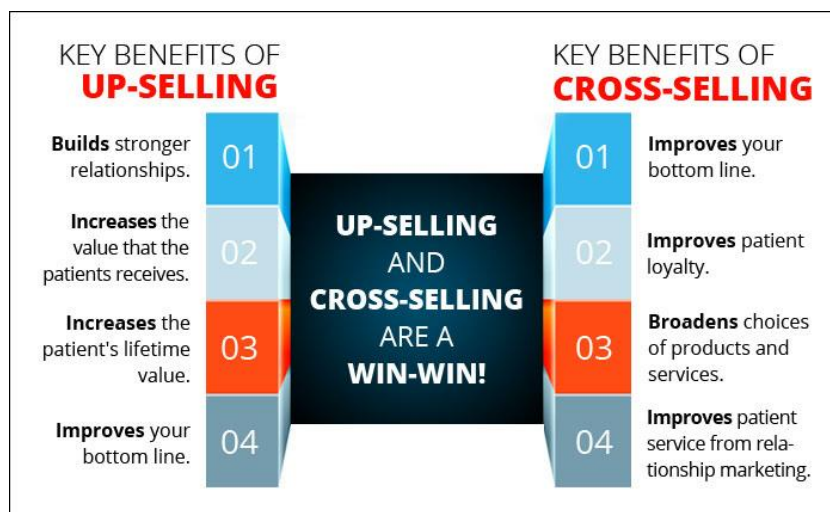
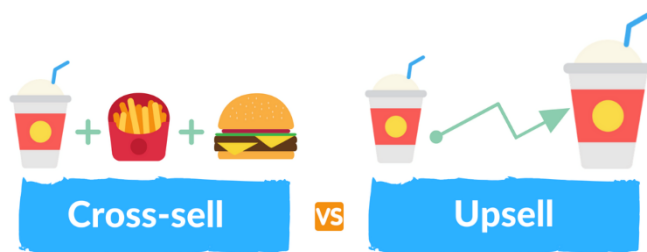
Country: India, South Africa, USA

College/Company: Data Glacier

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)



## Categorical Features:

- *sexo* : gender
- *ind\_nuevo* : New customer Index. 1 if the customer registered in the last 6 months.
- *segmento* : segmentation: 01 - VIP, 02 - Individuals 03 - college graduated
- *indext* : Foreigner index (S (Yes) or N (No) if the customer's birth country is different than the bank country)
- *indresi* : Residence index (S (Yes) or N (No) if the residence country is the same than the bank country)
- *indrel* : 1 (First/Primary), 99 (Primary customer during the month but not 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)
- *indfall* : Deceased index. N/S.
- *tiprel\_1mes* : Customer relation type at the beginning of the month, A (active), I (inactive), P (former customer),R (Potential)
- *ind\_actividad\_cliente* : Activity index (1, active customer; 0, inactive customer)
- *tipodom* : Address type. 1, primary address
- *pais\_residencia* : Customer's Country residence
- *canal\_entrada* : channel used by the customer to join

## Numerical Features:

- *age* : Age
- *antiguedad* : Customer seniority (in months)
- *renta* : Gross income of the household

## DATA CLEANING AND TRANSFORMATION

- 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'.
- To resolve outlier, we will ignore age groups less than 18 and age groups greater than 81
- The following columns were changed to numeric data types: age, antigüedad, indrel\_1mes
- N/A values in ind\_nomina\_ult1 and ind\_nom\_pens\_ult1 were replaced with zero.
- Delete columns that are almost entirely empty (ult\_fec\_cli\_1t, conyuemp)
- Fix the Unicode character in nomprov ( Coruña ,A was changed to CORUNA, A )
- tipodom doesn't seem to be useful, and the province code is not needed because the name of the province exists in nomprov.
- fecha\_dato and fecha\_alta datatypes were changed to datetime format.
- Replace NA values for ind\_actividad\_cliente with the median
- NA values in segmento and ind\_empleado were replaced with the mode.
- NA values in the rest of the columns were replaced with "UNKNOWN"
- The following products are not purchased so they are deleted: ind\_ahor\_fin\_ult1, ind\_av  
al\_fin\_ult1, ind\_cder\_fin\_ult1, ind\_ctju\_fin\_ult1, ind\_deco\_fin\_ult1, ind\_deme\_fin\_ult1, in  
d\_pres\_fin\_ult1, ind\_viv\_fin\_ult1

## REVIEW COMMENT

Based on the points discussed above, the team have unanimously agreed to follow this and proceed with further data analysis and gain some insights from it.