Specialization: Data Analyst

Group Name: Data Storytellers

Team member's details:

1. Laxmi Rathna Mrinalini Dutt Innamuri
   mrinalinidutt.ilr@gmail.com
   United Arab Emirates
   Repo link
   https://github.com/LaxmiInnamuri/DataAnalyst_CrossSellingRecommendation_GroupProject.git

2. Christy Abraham Jacob
   Chris.jacob7@gmail.com
   United Kingdom

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 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).

What action bank can take to increase cross selling?

Business Understanding

Customer centric view- Effective cross-sellers build a customer-centric view of opportunity and take a longer-term view of customer value rather than product-centric view of cross-selling.

Single view of the customer- Having into consideration product usage and transactional history, service history, etc. Identifying patterns of behavior across products is essential for understanding and anticipating customer needs to provide the most appropriate engagement, in the form of products they don't yet have, services they don't yet use.

Using customer data such as transaction behavior, existing accounts, lifestyle factors and demographics to customize cross-sell offers and provide personalized recommendations based on each customer's current and future needs.

Maximize segmentation- Customers face different needs at different times in their life. Selling bank products relevant to each new point in their life cycle simply makes sense. Segmenting your customers into groups based on most-likely cross-sell opportunities.

In order to properly prepare to cross sell your clients, you should know which items sell best for you and what items pair well with your best sellers.

Channels- Marketing is fundamentally about target audience, while the channel is the means of delivering a message to that audience. Thus, Channel used by customer also plays an important role in cross-selling.

Products & Services- In order to properly prepare to cross sell your clients, you should know which items sell best for you and what items pair well with your best sellers through studying past Sales data against the products & services you currently offer. When you're selling additional products and services to clients it's important to show them the additional value the products or service.

Cross Selling vs. Upselling-While cross selling and upselling in banking are both useful techniques for boosting sales, they are completely different. They can be used individually or together for maximum results.

Cross selling is when you sell another complementary product or service to an existing customer. The products or services you are cross selling should go well with what your customers are already purchasing.

Upselling happens when you sell your client more expensive products in addition to what they're already purchasing. In order to effectively upsell you must sell the value that comes with the upgraded version of a product or service. Comparison charts help show the extra value that comes with purchasing the next level up.

Data Understanding

Given data from test.csv & train.csv contains Customer Data.

Test.csv contains data for the date 28-06-2016

Train.csv contains data for the period from 28-01-2015 to 28-05-2016

Data Size

The size of training data set is about 4.9+ GB, which has 13,647,309 observations. The test data set has 929,615 observations.

Input Features

From column 1 to column 24 are the input features, which contain 21 categorical features and 3 continuous features. The input features contain customers' demographic and status with the bank information. On top of this, the observations are in the time series format. The data contains each customer's information from January 2015 to May 2016.

Data columns contains information Customer Code, Employee Index, Customer's Country residence, Sex, Age, First Contract Date, New Customer Index, Customer Seniority, Primary Customer Status & End Date, Customer Type, Customer Relation Type, Residence Index, Foreigner Index, Spouse Index, Channel, Deceased Index, Address Type, Province Code, Province Name, Active Index, Gross Income, Segmentation.

Output Features

From column 25 to column 48 are the output features, which contains the product purchased information according to each customer from January 2015 to May 2016. Each column stands for one product, and there are 24 products in total.

Id	Column Name	Data Type (given)	Description	Action
1	fecha_dato	object	The table is partitioned for this column	Convert to String
2	ncodpers	int64	Customer code	Identifying unquie ids
3	ind_empleado	object	Employee index: A active, B ex employed, F filial, N not employee, P pasive	Update missing values with N
4	pais_residencia	object	Customer's Country residence	Update missing value as 'UNKNOWN'
5	sexo	object	Customer's sex	Covert to string
6	age	object	Age	Convert to numeric Replace missing values with Mean
7	fecha_alta	object	The date in which the customer became as the first holder of a contract in the bank	Replace missing values with Median
8	ind_nuevo	float64	New customer Index. 1 if the customer registered in the last 6 months.	Covert to string. Update missing values to 1 as active months were below 6 months.
9	antiguedad	object	Customer seniority (in months)	Convert to numeric Update Missing values as 0
10	indrel	float64	1 (First/Primary), 99 (Primary customer during the month but not at the end of the month)	Convert to numeric. Missing number assumed as 1
11	ult_fec_cli_1t	object	Last date as primary customer (if he isn't at the end of the month)	Delete the column because of highest missing values
12	indrel_1mes	object	Customer type at the beginning of the month ,1 (First/Primary customer), 2 (co-owner),P (Potential),3 (former primary), 4(former co-owner)	Changing the astype to Category and where inplace=True as P and for null column as 'UNKNOWN'
13	tiprel_1mes	object	Customer relation type at the beginning of the month, A (active), I (inactive), P (former customer),R (Potential)	Update missing value as 'UNKNOWN'

14	indresi	object	Residence index (S (Yes) or N (No) if the residence country is the same than the bank country)	Update missing value as 'UNKNOWN'
15	indext	object	Foreigner index (S (Yes) or N (No) if the customer's birth country is different than the bank country)	Covert to string
16	conyuemp	object	Spouse index. 1 if the customer is spouse of an employee	Delete the column because of highest missing values
17	canal_entrada	float64	channel used by the customer to join	
18	indfall	float64	Deceased index. N/S	Changing the astype to Category and where inplace=True as A and for null column as 'UNKNOWN'
19	tipodom	float64	Addres type. 1, primary address	Delete the column as nomprov is sufficient for analysis
20	cod_prov	float64	Province code (customer's address)	Delete the column as nomprov is sufficient for analysis
21	nomprov	object	Province name	Update Missing values with UNKNOWN
22	ind_actividad_cliente	float64	Activity index (1, active customer; 0, inactive customer)	Replace missing values with Median
23	renta	float64	Gross income of the household	Replace missing values with Median
		object		Sort by unique values
24	segmento	int64	segmentation: 01 - VIP, 02 - Individuals 03 - college graduated	Assume missing values as 02 - Individuals hence update
25	ind_ahor_fin_ult1	int64	Saving Account	Consider as Target (Output features)
26	ind_aval_fin_ult1	int64	Guarantees	Consider as Target (Output features)
27	ind_cco_fin_ult1	int64	Current Accounts	Consider as Target (Output features)
28	ind_cder_fin_ult1	int64	Derivada Account	Consider as Target (Output features)

29	ind_cno_fin_ult1	int64	Payroll Account	Consider as Target (Output features)
30	ind_ctju_fin_ult1	int64	Junior Account	Consider as Target (Output features)
31	ind_ctma_fin_ult1	int64	Más particular Account	Consider as Target (Output features)
32	ind_ctop_fin_ult1	int64	particular Account	Consider as Target (Output features)
33	ind_ctpp_fin_ult1	int64	particular Plus Account	Consider as Target (Output features)
34	ind_deco_fin_ult1	int64	Short-term deposits	Consider as Target (Output features)
35	ind_deme_fin_ult 1	int64	Medium-term deposits	Consider as Target (Output features)
36	ind_dela_fin_ult1	int64	Long-term deposits	Consider as Target (Output features)
37	ind_ecue_fin_ult1	int64	e-account	Consider as Target (Output features)
38	ind_fond_fin_ult1	int64	Funds	Consider as Target (Output features)
39	ind_hip_fin_ult1	int64	Mortgage	Consider as Target (Output features)
40	ind_plan_fin_ult1	int64	Pensions	Consider as Target (Output features)
41	ind_pres_fin_ult1	int64	Loans	Consider as Target (Output features)
42	ind_reca_fin_ult1	int64	Taxes	Consider as Target (Output features)
43	ind_tjcr_fin_ult1	int64	Credit Card	Consider as Target (Output features)
44	ind_valo_fin_ult1	int64	Securities	Consider as Target (Output features)
45	ind_viv_fin_ult1	int64	Home Account	Consider as Target (Output features)
46	ind_nomina_ult1	float64	Payroll	Update missing values as 0. Consider as Target (Output features)
47	ind_nom_pens_ul t1	float64	Pensions	Update missing values as 0. Consider as Target (Output features)
48	ind_recibo_ult1	int64	Direct Debit	Consider as Target (Output features)

Project lifecycle along with deadline

Project Life Cycle	Deadline
1. Business understanding	19-Oct-22
2. Data Understanding	26-Oct-22
3. Data Cleansing and Transformation	2-Nov-22
4. Exploratory data analysis	9-Nov-22
5. EDA Recommendation (ppt)	16-Nov-22
6. Dashboard	23-Nov-22
7. Final Report	30-Nov-22