

Customer segmentation

Group Name: individual work

Name: Maha Haj Meftah

Email: hajmeftahmaha@gmail.com

Country: Tunisia

College: University of Carthage, National Engineering School Of Carthage

Specialization: Data analyst

Problem Description:

XYZ bank wants to roll out Christmas offers to their customers. They want to roll out personalized offer to particular set of customers. They don't want more than 5 group as this will be inefficient for their campaign.

Data understanding:

The data contains customer data for Latin American bank. Each customer has a customer id which uniquely identifies them. The data contains customers local customers and foreign customers. The demographic data of the customers is provided also. Additionally, the data contains information about how active each customer is, customer seniority and the number of bank products each customer use.

-At the end of this file you'll find the meaning of each column.

-The following table shows the columns with missing values and their numbers:

| Column | Number of NA |
|-----------------------|--------------|
| ind_empleado | 10782 |
| pais_residencia | 10782 |
| sexo | 10786 |
| fecha_alta | 10782 |
| ind_nuevo | 10782 |
| renta | 175183 |
| ind_actividad_cliente | 10782 |
| nomprov | 17734 |
| indfall | 10782 |
| canal_entrada | 10861 |

| | |
|-------------------|--------|
| conyuemp | 999822 |
| indext | 10782 |
| indresi | 10782 |
| | |
| | |
| indrel | 10782 |
| ult_fec_cli_1t | 998899 |
| indrel_1mes | 10782 |
| tiprel_1mes | 10782 |
| ind_nomina_ult1 | 5402 |
| ind_nom_pens_ult1 | 5402 |

Approaches

For handling the NAS:

- Drop the columns with missing values higher than 80%
- replace the missing values in categorical features with mode
- replace the missing values in numerical features with mean
- drop column because it has only one type of value such 1 because it has no significant information on the data

For handling the outliers :

- use the IQR-interquartile to identify the outliers(

Other approaches :

- dropping the columns like :ncodpers ,cod_prov,unnamed

Label encoding(for categorical features):

-indrel_1mes :

- 1 primary
- 2 co-owner
- 3 former primary

-tripel_1mes:

- A active
- I inactive
- P potential

-hot encoding for variables with 2 categories

-frequency encoding for multi-categorical variables

meaning of each column:

| | |
|-----------------|---|
| fecha_datos | The table is partitioned for this column |
| ncodpers | Customer code |
| ind_employeado | Employee index: A active, B ex employed, F filial, N not employee, P pasive |
| pais_residencia | Customer's Country residence |
| sexo | Customer's sex |
| age | Age |
| fecha_alta | The date in which the customer became as the first holder of a contract in the bank |
| ind_nuevo | New customer Index. 1 if the customer registered in the last 6 months. |
| antiguedad | Customer seniority (in months) |
| indrel | 1 (First/Primary), 99 (Primary customer during the month but not at the end of the month) |
| ult_fec_cliente | Last date as primary customer (if he isn't 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) |
| tiprel_1mes | Customer relation type at the beginning of the month, A (active), I (inactive), P (former customer),R (Potential) |
| indresi | Residence index (S (Yes) or N (No) if the residence country is the same than the bank country) |
| indext | Foreigner index (S (Yes) or N (No) if the customer's birth country is different than the bank country) |
| conyuemp | Spouse index. 1 if the customer is spouse of an employee |
| canal_entrada | channel used by the customer to join |
| indfall | Deceased index. N/S |
| tipodom | Addres type. 1, primary address |
| cod_prov | Province code (customer's address) |
| nomprov | Province name |

| | |
|-----------------------|---|
| ind_actividad_cliente | Activity index (1, active customer; 0, inactive customer) |
| renta | Gross income of the household |
| ind_ahor_fin_ult1 | Saving Account |
| ind_aval_fin_ult1 | Guarantees |
| ind_cco_fin_ult1 | Current Accounts |
| ind_cder_fin_ult1 | Derivada Account |
| ind_cno_fin_ult1 | Payroll Account |
| ind_ctju_fin_ult1 | Junior Account |
| ind_ctma_fin_ult1 | Más particular Account |
| ind_ctop_fin_ult1 | particular Account |
| ind_ctpp_fin_ult1 | particular Plus Account |
| ind_deco_fin_ult1 | Short-term deposits |
| ind_deme_fin_ult1 | Medium-term deposits |
| ind_dela_fin_ult1 | Long-term deposits |
| ind_ecue_fin_ult1 | e-account |
| ind_fond_fin_ult1 | Funds |
| ind_hip_fin_ult1 | Mortgage |
| ind_plan_fin_ult1 | Pensions |
| ind_pres_fin_ult1 | Loans |