**Mphasis HyperGraf KYC based Customer Segmentation**

Customer segmentation aggregates the data captured in the KYC details and creates customer groups. The KYC data comes in various categories such as Geographic, Demographic, Behavior and Financial variables.  This solution leverages machine learning to identify patterns in the data and segment the customers. These segments can be utilized for marketing analytics and campaigns.

**Max number of samples** = 100000

**Min number of samples** = 1000

* Type of clustering:

Default = **hdbscan**

1. hdbscan
2. kmeans
3. kmodes

* Type of segmentation:

Default = **Combined**

1. Geographic segmentation
2. Demographic segmentation
3. Behavior segmentation
4. Financial segmentation
5. Combined

* Recommended clustering parameters for different clustering:

Note:

1. **hdbscan:** (Integer)

* **min\_cluster\_size = (Default =** 1000**)**
* **min\_samples = (Default =** 80**)**

1. **kmeans:** (Integer)

* **n\_clusters = (Default =** 5**)**

1. **kmodes:** (Integer)

* **n\_clusters = (Default =** 5**)**

Sample format of Input data:

* **Sample ‘Input.csv’ :**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **customer\_code** | **age** | **customer\_seniority** | **income** | **gender** | **customer\_status** |
| 816978 | 55 | 80 | 67804.53 | V | 1 |
| 154073 | 50 | 187 | 109185.8 | V | 0 |
| 739359 | 48 | 91 | 100030.8 | H | 1 |

* **Sample ‘parameters.json’ :**

{

"Clustering algorithm":"kmodes",

"Segmentation type":"combined",

"Cluster parameters": [10]

}

**All the attributes must be one of below mentioned:**

Note: if other columns are added they will be not considered

* **'customer\_code'**: Unique ID for customers
* Columns allowed for Geographic segmentation columns:

(any two from the following needs to be present for Geographic segmentation)

* **'area\_type':** Type of area (Categorical variable)
* **'new\_territory'**: yes or no (Categorical variable)
* **'country':** Country code (US) (Categorical variable)
* **'city':** City code (Ex: NY) (Categorical variable)
* **'population\_density':** Density of population (Numeric variable)
* Columns allowed for Demographical segmentation columns:

(any two from the following needs to be present for Demographical segmentation)

* **'age':** Age in years/month (Numeric variable)
* **'gender':** M/F/Others/Unknown (Categorical variable)
* **'race':** Race (Categorical variable)
* **'religion':** Religion (Categorical variable)
* **'years\_of\_education'**: In years/months (Numeric variable)
* **'income':** Gross income/ Avg.income (Numeric variable)
* Columns allowed for Behavioral segmentation columns:

(any two from the following needs to be present for Behavioral segmentation)

* **'customer\_seniority'**: Seniority of customers in years/Months (Numeric variable)
* **'customer\_status':** Active customer (1)/ Inactive customer(0)/…. (Categorical variable)
* **'new\_customer':** New customer or not 0/1 (Categorical variable)
* **'employee':** Person is an existing employee or not ex: current\_emp (1)/ previous\_emp(0)/ not\_emp(-1)(Categorical variable)
* Columns allowed for Financial segmentation columns:

(any two from the following needs to be present for Financial segmentation)

* **'loan\_amount'**: Loan Amount (Numeric variable)
* **'credit\_rating':** Credit card Rating (Categorical variable)
* **'yearly\_spend'**: Amount spend in a (year/month) (Numeric variable variable)
* **'income':** Amount of money earned (Year/month) (Numeric variable)