Clustering-for-Customer-Segmentation-Understanding

Submitted by:

Yogesh.C.Mudliar

About this project

- Not all customers are the same. To know which group is your customer and their Preferences are a big part of success in your business.
- Unsupervised machine learning can help marketers know their audience globally and engage them with their products accordingly.
- Here, we can classify millions of people's interests through their social media activity and also through other surveys, online and offline, and cluster them into a specific group of their interest.

Objective

- A case requires to develop a customer segmentation to give recommendations like saving plans, loans, wealth management, etc. on target customers groups

About Dataset

- The sample Dataset summarizes the usage behavior of about 9000 active credit card holders during the last 6 months. The file is at a customer level with 18 behavioral variables

```
1  df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8950 entries, 0 to 8949
Data columns (total 18 columns):
 #   Column                            Non-Null Count  Dtype
---  ------                            --------------  -----
 0   CUST_ID                           8950 non-null   object
 1   BALANCE                           8950 non-null   float64
 2   BALANCE_FREQUENCY                 8950 non-null   float64
 3   PURCHASES                         8950 non-null   float64
 4   ONEOFF_PURCHASES                  8950 non-null   float64
 5   INSTALLMENTS_PURCHASES            8950 non-null   float64
 6   CASH_ADVANCE                      8950 non-null   float64
 7   PURCHASES_FREQUENCY               8950 non-null   float64
 8   ONEOFF_PURCHASES_FREQUENCY        8950 non-null   float64
 9   PURCHASES_INSTALLMENTS_FREQUENCY  8950 non-null   float64
 10  CASH_ADVANCE_FREQUENCY            8950 non-null   float64
 11  CASH_ADVANCE_TRX                  8950 non-null   int64
 12  PURCHASES_TRX                     8950 non-null   int64
 13  CREDIT_LIMIT                      8949 non-null   float64
 14  PAYMENTS                          8950 non-null   float64
 15  MINIMUM_PAYMENTS                  8637 non-null   float64
 16  PRC_FULL_PAYMENT                  8950 non-null   float64
 17  TENURE                            8950 non-null   int64
dtypes: float64(14), int64(3), object(1)
```

Tools used

- Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, Jupiter notebook, Git-Github, VScode, Stremlit are used to build the whole model.

PROJECT PROCESS

- **Research and business understanding:**

The first thing you have to do before you solve a problem is to define exactly what it is. You need to be able to translate data questions into something actionable.

- **Data pre-processing:**

Data preprocessing can refer to manipulation or dropping of data before it is used in order to ensure or enhance performance, and is an important step in the data mining process.

- **Exploratory Data analysis:**

Exploratory data analysis is an approach of analyzing data sets to summarize their main characteristics, often using statistical graphics and other data visualization methods.

CONTINUE…

- **Model Building:**

Model building process where different machine learning algorithms are used to make different machine learning models for various applications.

- **Model Deployment:**

Model Deployment is the process where various ML algorithms are deployed on various platforms like flask, streamlit, various open source platforms, etc. Here we have used Streamlit to deploy our ML project.

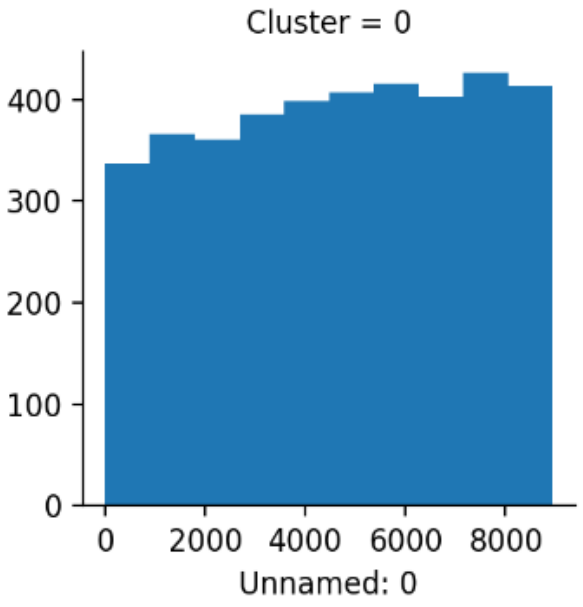
INSIGHTS GENERATED

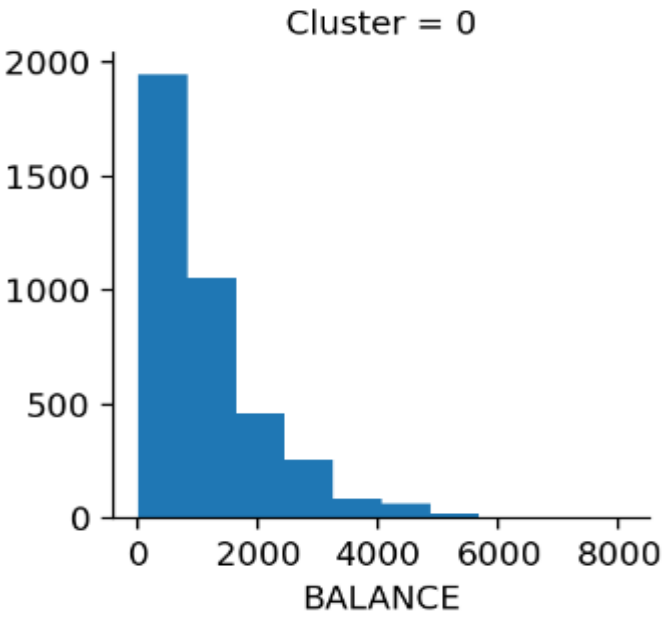
- OUTPUT SCREEN

INPUTS

- Balance – 40.900749
- Balance Frequency – 0.818182
- Purchases – 95.40
- One-off purchases – 0.00
- Instalment Purchases – 95.40
- Cash Advance – 0.000
- Purchases Frequency – 0.166667
- One-off Purchases Frequency – 0.00
- Purchases Instalments Frequency – 0.083333
- Cash advance frequency – 0.000
- Cash Advance TRX – 0
- Purchases TRX – 2
- Credit Limit – 1000
- Payments – 201.802084
- Minimum Payments – 139.509787
- PRC Full Payments – 0
- Tenure – 12

PREDICTION

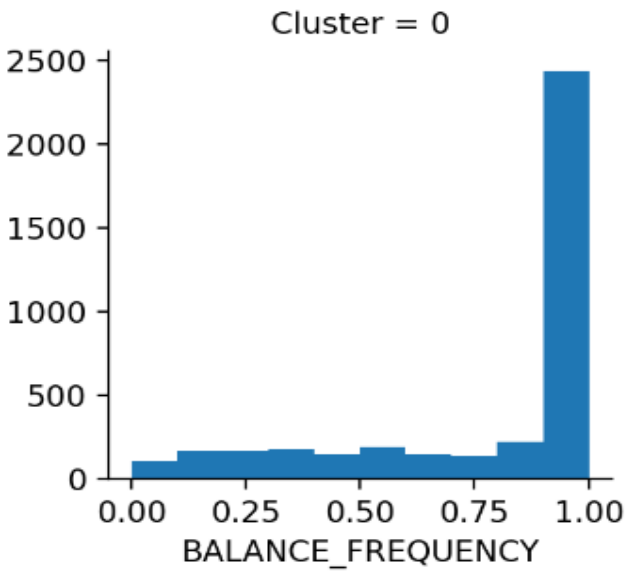


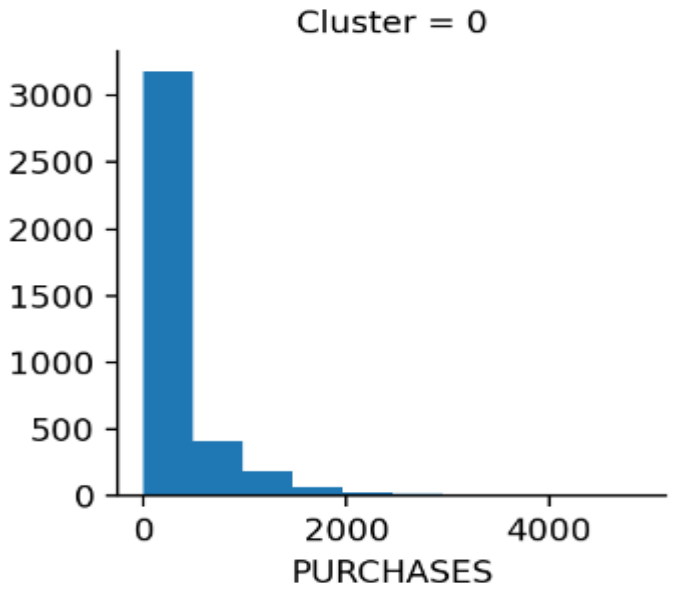


The visual shows the overall analysis and it belongs to Cluster 0 and the highest frequency is for 7000 and lowest for 0.

The visual shows that it belongs to Cluster 0 and the highest balance is 2000.

CONTINUE…

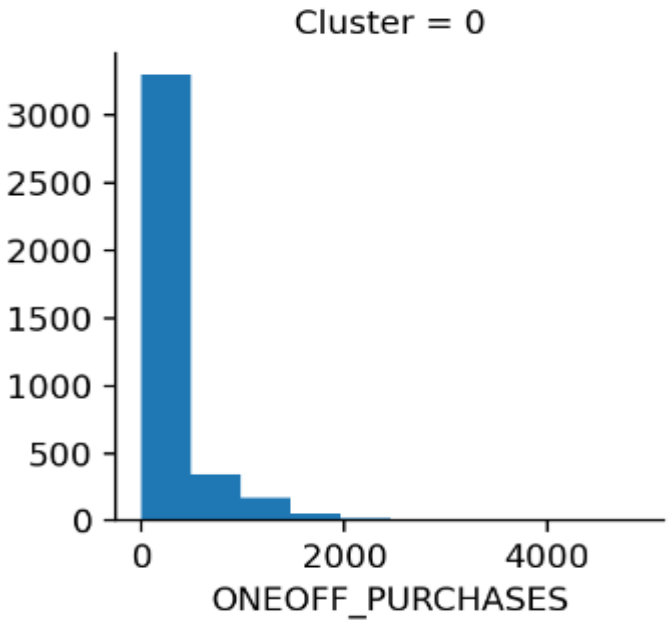




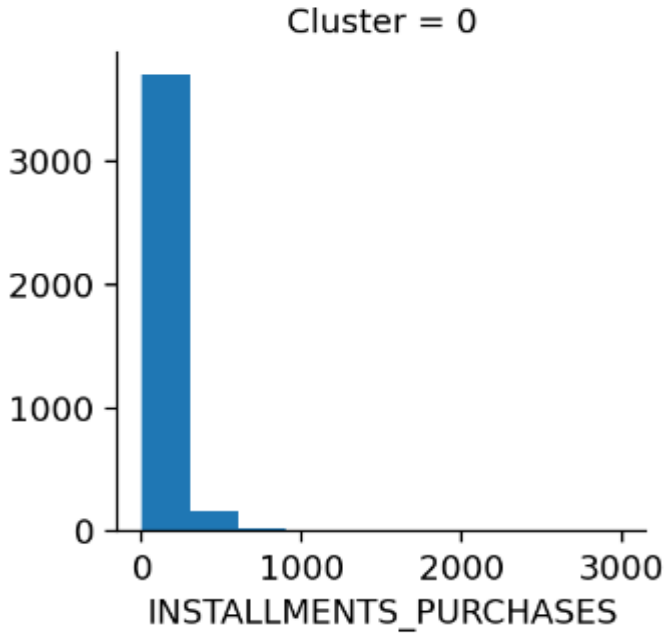
The visual shows the Balance Frequency which has highest frequency for 1.00 and lowest for 0.75.

The visual shows Purchases that it belongs to Cluster 0 and has highest frequency of 3000.

CONTINUE…

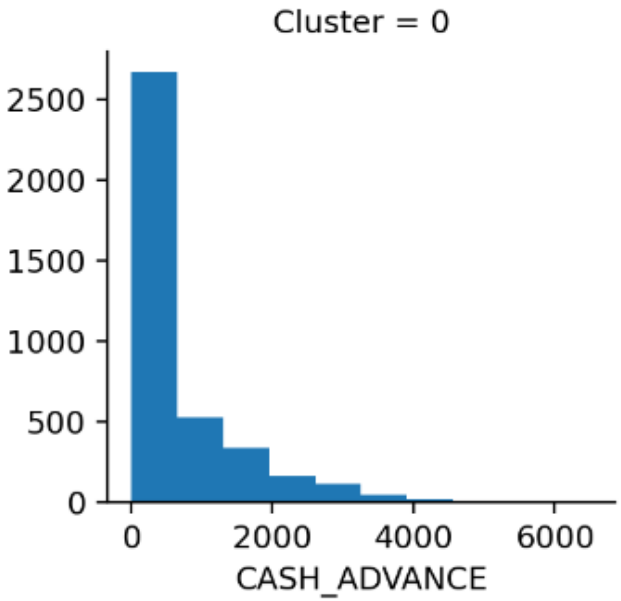


The visual shows the One-off Purchases which has highest frequency of 3000.

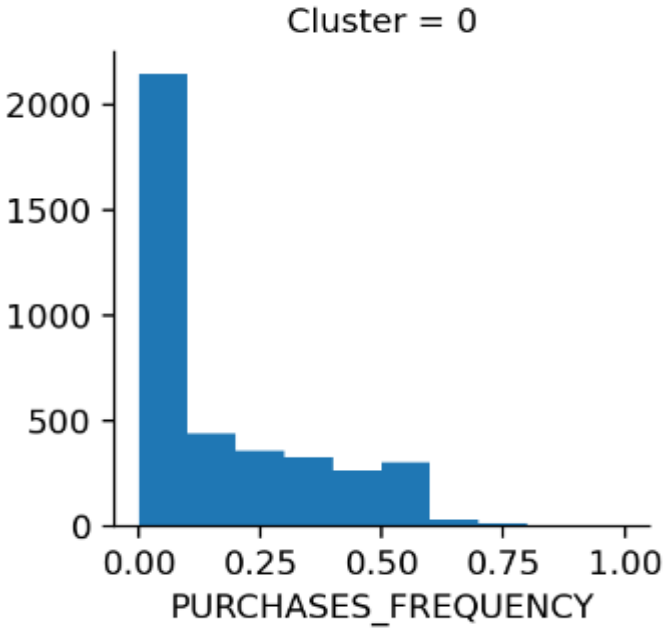


The visual shows Purchases that it belongs to Cluster 0 and has highest frequency of 3000.

CONTINUE…

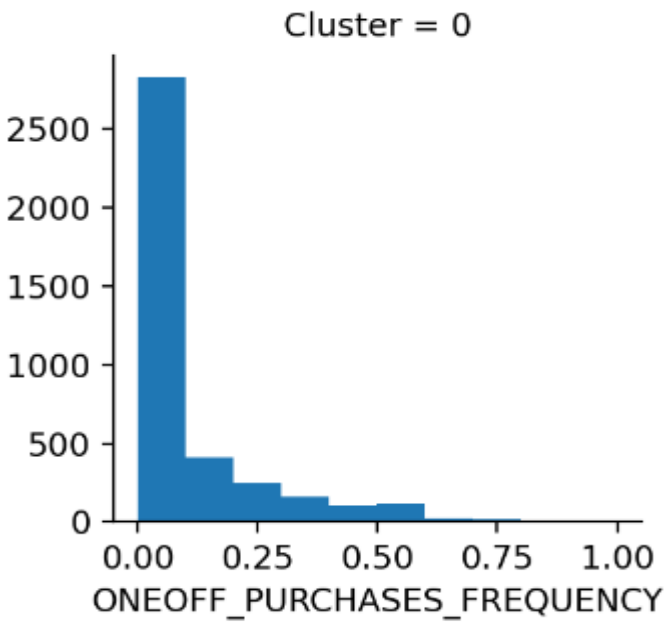


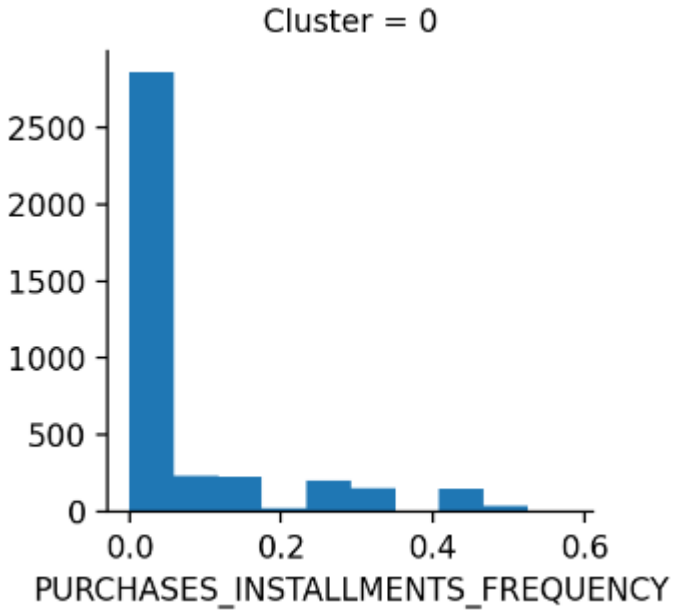
The visual shows the Cash Advance that belongs to Cluster 0 and has highest frequency of 2500.



The visual shows Purchases Frequency that it belongs to Cluster 0 and has highest frequency of 2000.

CONTINUE…

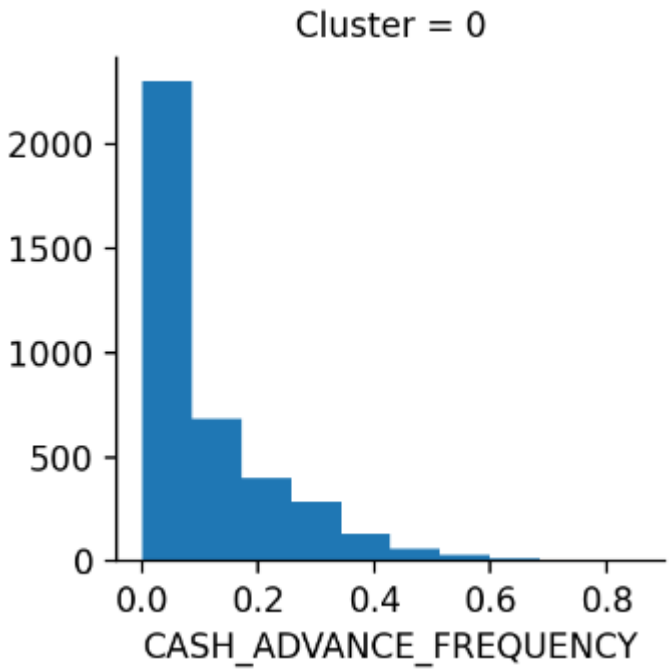


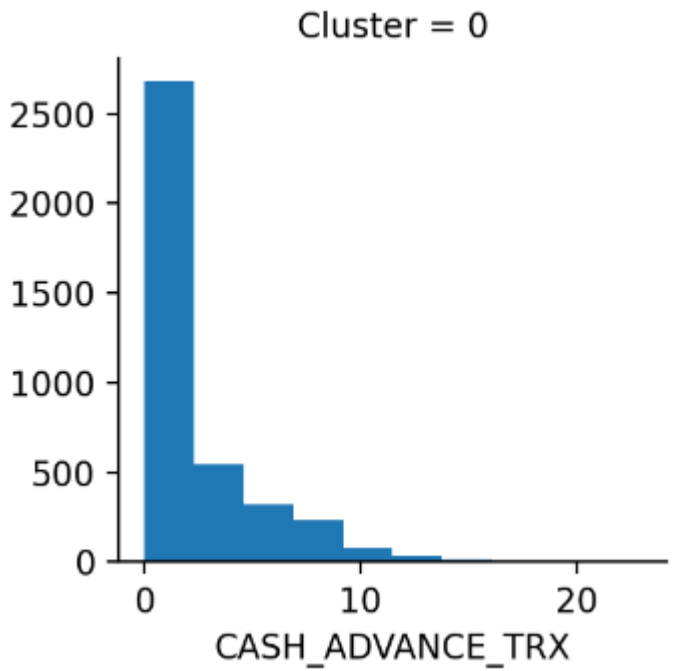


The visual shows the One-off Purchase Frequency which has highest frequency of 3000.

The visual shows Purchases Installments Frequency that it belongs to Cluster 0 and has highest frequency of 3000.

CONTINUE…

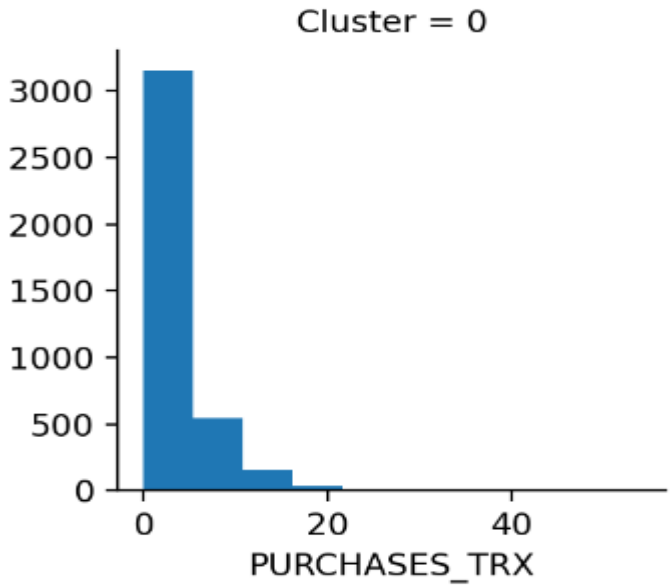


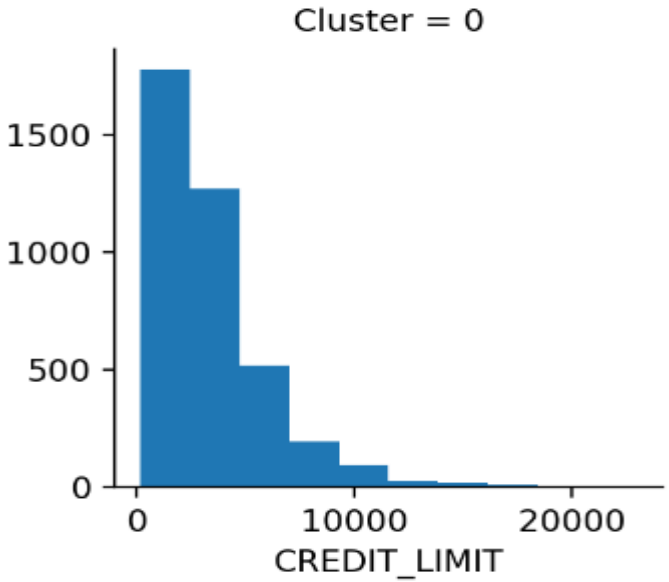


The visual shows the Cash Advance Frequency which belong to Cluster 0 which has highest frequency of 2500.

The visual shows Cash Advance TRX that it belongs to Cluster 0 and has highest frequency of 2500.

CONTINUE…

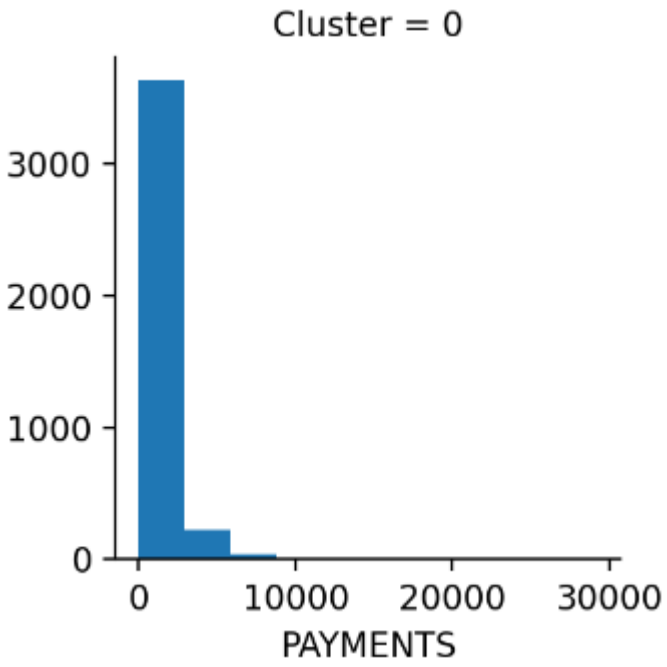




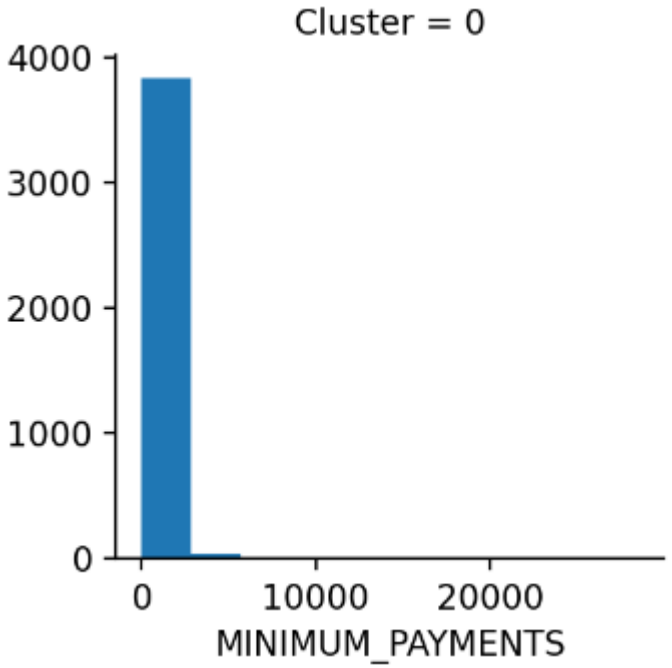
The visual shows the Purchases TRX which has highest frequency of 3000.

The visual shows Credit Limit that it belongs to Cluster 0 and has highest frequency of 1500.

CONTINUE…

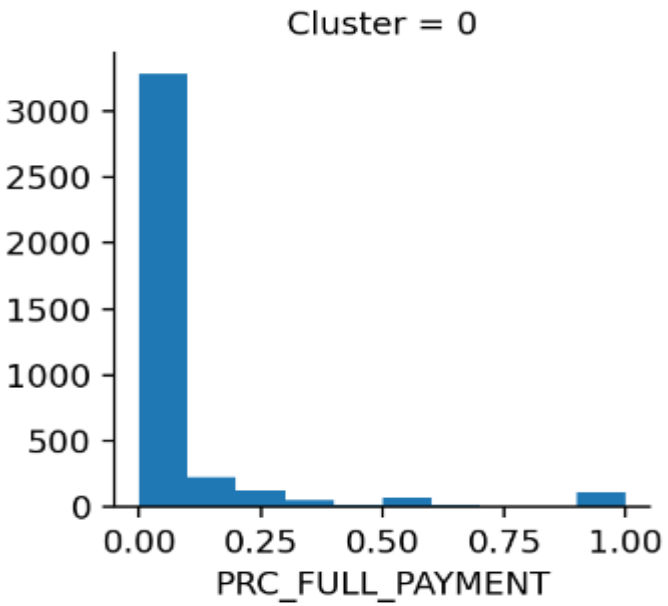


The visual shows the Payments for Cluster 0 which has highest frequency of 3000.

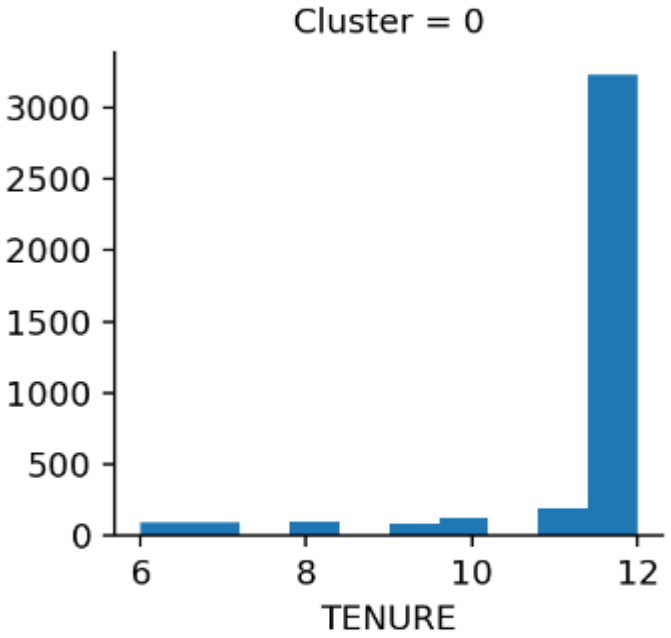


The visual shows Minimum Payments that it belongs to Cluster 0 and has highest frequency of 4000.

CONTINUE…



The visual shows the PRC Full Payment which has highest frequency of 3000.



The visual shows the Tenure for cluster 0 which has highest frequency of 3000.

Thank
you