Market Segmentation in Insurance using Unsupervised Machine learning - K Means and Agglomerative Clustering

Market segmentation is the process of dividing a broad consumer or business market into smaller groups, or segments, of consumers or businesses that have common needs, interests, or characteristics. The goal of market segmentation is to tailor marketing efforts to specific segments more effectively, ensuring that each segment's unique needs and preferences are met.

Key Concepts in Market Segmentation:

1. **Segmentation Criteria**:
   - **Demographic Segmentation**: Dividing the market based on variables like age, gender, income, education, occupation, family size, etc.
   - **Geographic Segmentation**: Segmenting the market based on geographic locations such as countries, regions, cities, or even neighborhoods.

- **Psychographic Segmentation**: Dividing the market based on lifestyle, social class, personality traits, values, interests, and attitudes.
- **Behavioral Segmentation**: Segmenting the market based on consumer behavior, including purchase patterns, brand loyalty, usage rate, benefits sought, and readiness to buy.
- **Firmographic Segmentation** (for B2B markets): Dividing businesses based on attributes like industry, company size, location, and buying behavior.

2. **Benefits of Market Segmentation**:
   - **Targeted Marketing**: Enables companies to focus their marketing strategies on specific groups, increasing the effectiveness of campaigns.
   - **Product Development**: Helps in designing and developing products that meet the specific needs of different segments.
   - **Competitive Advantage**: By addressing the unique needs of a segment, companies can differentiate themselves from competitors.
   - **Resource Optimization**: Allows businesses to allocate resources more efficiently by focusing on the most profitable segments.
3. **Examples of Market Segmentation**:
   - A cosmetics company may segment its market by age, creating separate product lines for teens, young adults, and older adults.
   - A travel agency might segment its customers based on travel behavior, offering packages tailored to adventure seekers, luxury travelers, or budget-conscious families.
4. **Steps in Market Segmentation**:
   - **Identify the Market**: Define the broad market to be segmented.
   - **Segment the Market**: Use relevant criteria to divide the market into distinct groups.
   - **Evaluate Segments**: Assess the potential of each segment based on factors like size, growth potential, and competition.
   - **Select Target Segments**: Choose the segments that align best with the company's objectives and resources.
   - **Develop Positioning**: Create a marketing strategy that effectively communicates the value proposition to the chosen segments.

Application in Business:

Market segmentation is widely used in various industries, from consumer goods to services, to ensure that marketing efforts resonate with the right audience, thus driving sales, enhancing customer satisfaction, and building brand loyalty.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

import warnings
warnings.filterwarnings('ignore')
```

```
df = pd.read_csv('market_customer_data.csv')

df.head()
```

```
  CUST_ID      BALANCE  BALANCE_FREQUENCY  PURCHASES  ONEOFF_PURCHASES
\
0  C10001    40.900749           0.818182      95.40              0.00

1  C10002  3202.467416           0.909091       0.00              0.00

2  C10003  2495.148862           1.000000     773.17            773.17

3  C10004  1666.670542           0.636364    1499.00           1499.00

4  C10005   817.714335           1.000000      16.00             16.00


   INSTALLMENTS_PURCHASES  CASH_ADVANCE  PURCHASES_FREQUENCY  \
0                    95.4      0.000000             0.166667
1                     0.0   6442.945483             0.000000
2                     0.0      0.000000             1.000000
3                     0.0    205.788017             0.083333
4                     0.0      0.000000             0.083333

   ONEOFF_PURCHASES_FREQUENCY  PURCHASES_INSTALLMENTS_FREQUENCY  \
0                    0.000000                          0.083333
1                    0.000000                          0.000000
2                    1.000000                          0.000000
3                    0.083333                          0.000000
4                    0.083333                          0.000000

   CASH_ADVANCE_FREQUENCY  CASH_ADVANCE_TRX  PURCHASES_TRX
CREDIT_LIMIT  \
0                0.000000                 0              2
1000.0
1                0.250000                 4              0
7000.0
2                0.000000                 0             12
7500.0
3                0.083333                 1              1
7500.0
4                0.000000                 0              1
1200.0

      PAYMENTS  MINIMUM_PAYMENTS  PRC_FULL_PAYMENT  TENURE
0   201.802084        139.509787          0.000000      12
1  4103.032597       1072.340217          0.222222      12
2   622.066742        627.284787          0.000000      12
3     0.000000               NaN          0.000000      12
4   678.334763        244.791237          0.000000      12
```

```
df.tail()
```

```
     CUST_ID     BALANCE  BALANCE_FREQUENCY  PURCHASES
ONEOFF_PURCHASES  \
8945  C19186   28.493517           1.000000     291.12
0.00
8946  C19187   19.183215           1.000000     300.00
0.00
8947  C19188   23.398673           0.833333     144.40
0.00
8948  C19189   13.457564           0.833333       0.00
0.00
8949  C19190  372.708075           0.666667    1093.25
1093.25

      INSTALLMENTS_PURCHASES  CASH_ADVANCE  PURCHASES_FREQUENCY  \
8945                  291.12      0.000000             1.000000
8946                  300.00      0.000000             1.000000
8947                  144.40      0.000000             0.833333
8948                    0.00     36.558778             0.000000
8949                    0.00    127.040008             0.666667

      ONEOFF_PURCHASES_FREQUENCY  PURCHASES_INSTALLMENTS_FREQUENCY  \
8945                    0.000000                          0.833333
8946                    0.000000                          0.833333
8947                    0.000000                          0.666667
8948                    0.000000                          0.000000
8949                    0.666667                          0.000000

      CASH_ADVANCE_FREQUENCY  CASH_ADVANCE_TRX  PURCHASES_TRX
CREDIT_LIMIT  \
8945                0.000000                 0              6
1000.0
8946                0.000000                 0              6
1000.0
8947                0.000000                 0              5
1000.0
8948                0.166667                 2              0
500.0
8949                0.333333                 2             23
1200.0

        PAYMENTS  MINIMUM_PAYMENTS  PRC_FULL_PAYMENT  TENURE
8945  325.594462         48.886365              0.50       6
8946  275.861322               NaN              0.00       6
8947   81.270775         82.418369              0.25       6
8948   52.549959         55.755628              0.25       6
8949   63.165404         88.288956              0.00       6
```

```
df.shape
```

```
(8950, 18)

df.columns

Index(['CUST_ID', 'BALANCE', 'BALANCE_FREQUENCY', 'PURCHASES',
       'ONEOFF_PURCHASES', 'INSTALLMENTS_PURCHASES', 'CASH_ADVANCE',
       'PURCHASES_FREQUENCY', 'ONEOFF_PURCHASES_FREQUENCY',
       'PURCHASES_INSTALLMENTS_FREQUENCY', 'CASH_ADVANCE_FREQUENCY',
       'CASH_ADVANCE_TRX', 'PURCHASES_TRX', 'CREDIT_LIMIT',
'PAYMENTS',
       'MINIMUM_PAYMENTS', 'PRC_FULL_PAYMENT', 'TENURE'],
      dtype='object')
```

The data provided appears to be related to customer credit card usage and financial behavior. Here's a breakdown of each feature:

1. **`CUST_ID`**:
   - **Type**: Categorical
   - **Description**: A unique identifier for each customer. This is likely a string or integer that uniquely identifies each record in the dataset.
2. **`BALANCE`**:
   - **Type**: Numerical
   - **Description**: The outstanding balance on the customer's credit card. This represents the amount of money owed by the customer at a given time.
3. **`BALANCE_FREQUENCY`**:
   - **Type**: Numerical (float, typically between 0 and 1)
   - **Description**: This represents how frequently the balance is updated or the frequency of balance usage by the customer. It could be an indicator of how often the customer checks or updates their balance.
4. **`PURCHASES`**:
   - **Type**: Numerical
   - **Description**: The total amount of money spent by the customer on purchases using the credit card.
5. **`ONEOFF_PURCHASES`**:
   - **Type**: Numerical
   - **Description**: The amount spent on one-time purchases. These are typically larger, non-recurring purchases.
6. **`INSTALLMENTS_PURCHASES`**:
   - **Type**: Numerical
   - **Description**: The amount spent on purchases made in installments. This could refer to payments spread over multiple billing cycles.
7. **`CASH_ADVANCE`**:
   - **Type**: Numerical
   - **Description**: The total amount of cash advances taken by the customer. A cash advance is a service provided by credit card issuers allowing cardholders to withdraw cash, usually at a high interest rate.

8. **`PURCHASES_FREQUENCY`**:
   - **Type**: Numerical (float, typically between 0 and 1)
   - **Description**: The frequency of purchases made by the customer. A value closer to 1 indicates that the customer makes purchases frequently.
9. **`ONEOFF_PURCHASES_FREQUENCY`**:
   - **Type**: Numerical (float, typically between 0 and 1)
   - **Description**: The frequency of one-time purchases. A higher value indicates more frequent one-off purchases.
10. **`PURCHASES_INSTALLMENTS_FREQUENCY`**:
    - **Type**: Numerical (float, typically between 0 and 1)
    - **Description**: The frequency of purchases made in installments. A higher value suggests more frequent use of installment purchases.
11. **`CASH_ADVANCE_FREQUENCY`**:
    - **Type**: Numerical (float, typically between 0 and 1)
    - **Description**: The frequency of cash advances taken by the customer. A value closer to 1 indicates frequent cash advances.
12. **`CASH_ADVANCE_TRX`**:
    - **Type**: Numerical (integer)
    - **Description**: The number of transactions where the customer took a cash advance.
13. **`PURCHASES_TRX`**:
    - **Type**: Numerical (integer)
    - **Description**: The number of purchase transactions made by the customer.
14. **`CREDIT_LIMIT`**:
    - **Type**: Numerical
    - **Description**: The credit limit assigned to the customer. This is the maximum amount that the customer is allowed to borrow using the credit card.
15. **`PAYMENTS`**:
    - **Type**: Numerical
    - **Description**: The total amount of payments made by the customer towards their credit card balance.
16. **`MINIMUM_PAYMENTS`**:
    - **Type**: Numerical
    - **Description**: The minimum amount that the customer is required to pay on their credit card balance to avoid penalties.
17. **`PRC_FULL_PAYMENT`**:
    - **Type**: Numerical (float, typically between 0 and 1)
    - **Description**: The percentage of times the customer has fully paid the credit card balance. A value closer to 1 suggests that the customer often pays off their balance in full.
18. **`TENURE`**:
    - **Type**: Numerical (integer)
    - **Description**: The number of months the customer has been with the credit card issuer. This indicates the duration of the customer's relationship with the issuer.

This dataset is useful for analyzing customer credit card behavior, which can be leveraged for customer segmentation, credit risk analysis, and targeted marketing strategies.

```
df.duplicated().sum()
```

```
0
```

```
df.isnull().sum()
```

```
CUST_ID                             0
BALANCE                             0
BALANCE_FREQUENCY                   0
PURCHASES                           0
ONEOFF_PURCHASES                    0
INSTALLMENTS_PURCHASES              0
CASH_ADVANCE                        0
PURCHASES_FREQUENCY                 0
ONEOFF_PURCHASES_FREQUENCY          0
PURCHASES_INSTALLMENTS_FREQUENCY    0
CASH_ADVANCE_FREQUENCY              0
CASH_ADVANCE_TRX                    0
PURCHASES_TRX                       0
CREDIT_LIMIT                        1
PAYMENTS                            0
MINIMUM_PAYMENTS                  313
PRC_FULL_PAYMENT                    0
TENURE                              0
dtype: int64
```

```
df["MINIMUM_PAYMENTS"] =
df["MINIMUM_PAYMENTS"].fillna(df["MINIMUM_PAYMENTS"].mean())
df["CREDIT_LIMIT"] =
df["CREDIT_LIMIT"].fillna(df["CREDIT_LIMIT"].mean())
```

```
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                      8950 non-null   float64
 14  PAYMENTS                          8950 non-null   float64
 15  MINIMUM_PAYMENTS                  8950 non-null   float64
 16  PRC_FULL_PAYMENT                  8950 non-null   float64
 17  TENURE                            8950 non-null   int64
dtypes: float64(14), int64(3), object(1)
memory usage: 1.2+ MB

df.describe()

            BALANCE  BALANCE_FREQUENCY     PURCHASES  ONEOFF_PURCHASES
\
count   8950.000000        8950.000000   8950.000000       8950.000000

mean    1564.474828           0.877271   1003.204834        592.437371

std     2081.531879           0.236904   2136.634782       1659.887917

min        0.000000           0.000000      0.000000          0.000000

25%      128.281915           0.888889     39.635000          0.000000

50%      873.385231           1.000000    361.280000         38.000000

75%     2054.140036           1.000000   1110.130000        577.405000

max    19043.138560           1.000000  49039.570000      40761.250000


       INSTALLMENTS_PURCHASES  CASH_ADVANCE  PURCHASES_FREQUENCY  \
count             8950.000000   8950.000000          8950.000000
mean               411.067645    978.871112             0.490351
std                904.338115   2097.163877             0.401371
min                  0.000000      0.000000             0.000000
25%                  0.000000      0.000000             0.083333
50%                 89.000000      0.000000             0.500000
75%                468.637500   1113.821139             0.916667
max              22500.000000  47137.211760             1.000000

       ONEOFF_PURCHASES_FREQUENCY  PURCHASES_INSTALLMENTS_FREQUENCY  \
count                 8950.000000                       8950.000000
mean                     0.202458                          0.364437
std                      0.298336                          0.397448
min                      0.000000                          0.000000
25%                      0.000000                          0.000000
50%                      0.083333                          0.166667
75%                      0.300000                          0.750000
max                      1.000000                          1.000000
```

```
       CASH_ADVANCE_FREQUENCY  CASH_ADVANCE_TRX  PURCHASES_TRX
CREDIT_LIMIT  \
count             8950.000000       8950.000000    8950.000000
8950.000000
mean                 0.135144          3.248827      14.709832
4494.449450
std                  0.200121          6.824647      24.857649
3638.612411
min                  0.000000          0.000000       0.000000
50.000000
25%                  0.000000          0.000000       1.000000
1600.000000
50%                  0.000000          0.000000       7.000000
3000.000000
75%                  0.222222          4.000000      17.000000
6500.000000
max                  1.500000        123.000000     358.000000
30000.000000

            PAYMENTS  MINIMUM_PAYMENTS  PRC_FULL_PAYMENT       TENURE
count   8950.000000       8950.000000       8950.000000  8950.000000
mean    1733.143852        864.206542          0.153715    11.517318
std     2895.063757       2330.588021          0.292499     1.338331
min        0.000000          0.019163          0.000000     6.000000
25%      383.276166        170.857654          0.000000    12.000000
50%      856.901546        335.628312          0.000000    12.000000
75%     1901.134317        864.206542          0.142857    12.000000
max    50721.483360      76406.207520          1.000000    12.000000
```

```
df.drop(columns=["CUST_ID"],axis=1,inplace=True)
```

```
df.duplicated().sum()
```

```
0
```

```
object_columns = df.select_dtypes(include=['object']).columns
print("Object type columns:")
print(object_columns)

numerical_columns = df.select_dtypes(include=['int64',
'float64']).columns
print("\nNumerical type columns:")
print(numerical_columns)
```

```
Object type columns:
Index([], dtype='object')

Numerical type columns:
Index(['BALANCE', 'BALANCE_FREQUENCY', 'PURCHASES',
'ONEOFF_PURCHASES',
```

```
       'INSTALLMENTS_PURCHASES', 'CASH_ADVANCE',
'PURCHASES_FREQUENCY',
       'ONEOFF_PURCHASES_FREQUENCY',
'PURCHASES_INSTALLMENTS_FREQUENCY',
       'CASH_ADVANCE_FREQUENCY', 'CASH_ADVANCE_TRX', 'PURCHASES_TRX',
       'CREDIT_LIMIT', 'PAYMENTS', 'MINIMUM_PAYMENTS',
'PRC_FULL_PAYMENT',
       'TENURE'],
      dtype='object')
```

```
def classify_features(df):
    categorical_features = []
    non_categorical_features = []
    discrete_features = []
    continuous_features = []

    for column in df.columns:
        if df[column].dtype == 'object':
            if df[column].nunique() < 10:
                categorical_features.append(column)
            else:
                non_categorical_features.append(column)
        elif df[column].dtype in ['int64', 'float64']:
            if df[column].nunique() < 10:
                discrete_features.append(column)
            else:
                continuous_features.append(column)

    return categorical_features, non_categorical_features,
discrete_features, continuous_features

categorical, non_categorical, discrete, continuous =
classify_features(df)

print("Categorical Features:", categorical)
print("Non-Categorical Features:", non_categorical)
print("Discrete Features:", discrete)
print("Continuous Features:", continuous)
```

```
Categorical Features: []
Non-Categorical Features: []
Discrete Features: ['TENURE']
Continuous Features: ['BALANCE', 'BALANCE_FREQUENCY', 'PURCHASES',
'ONEOFF_PURCHASES', 'INSTALLMENTS_PURCHASES', 'CASH_ADVANCE',
'PURCHASES_FREQUENCY', 'ONEOFF_PURCHASES_FREQUENCY',
'PURCHASES_INSTALLMENTS_FREQUENCY', 'CASH_ADVANCE_FREQUENCY',
'CASH_ADVANCE_TRX', 'PURCHASES_TRX', 'CREDIT_LIMIT', 'PAYMENTS',
'MINIMUM_PAYMENTS', 'PRC_FULL_PAYMENT']
```

```
for i in discrete:
    print(i)
```

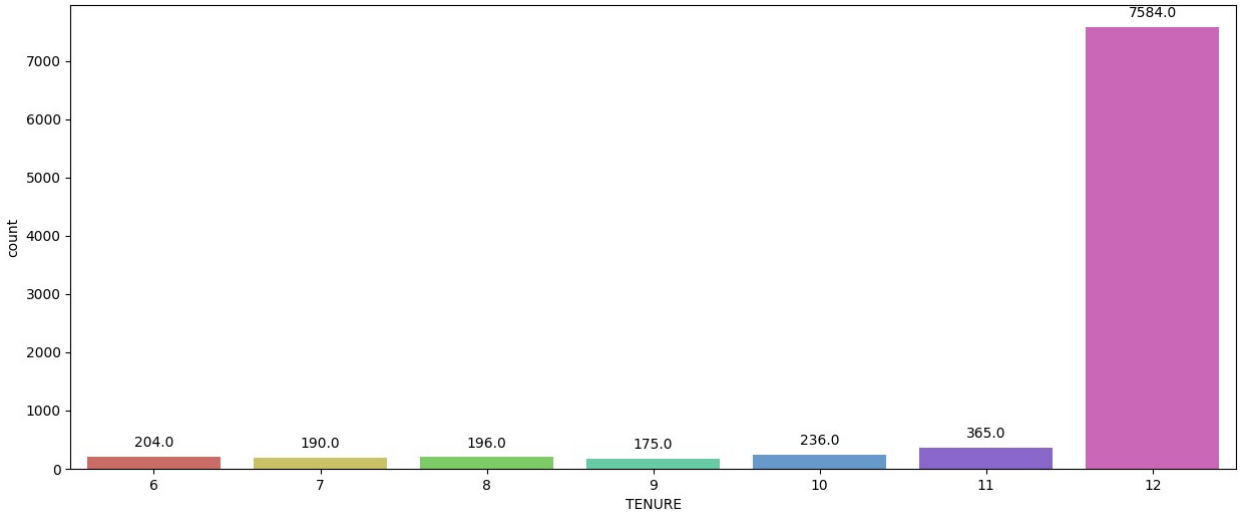
```
    print(df[i].unique())
    print()
```

```
TENURE
[12  8 11  9 10  7  6]
```

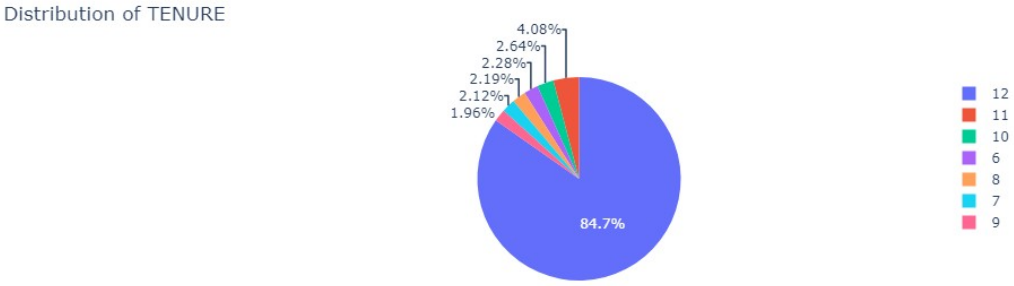
```
for i in discrete:
    print(df[i].value_counts())
    print()
```

```
TENURE
12    7584
11     365
10     236
6      204
8      196
7      190
9      175
Name: count, dtype: int64
```

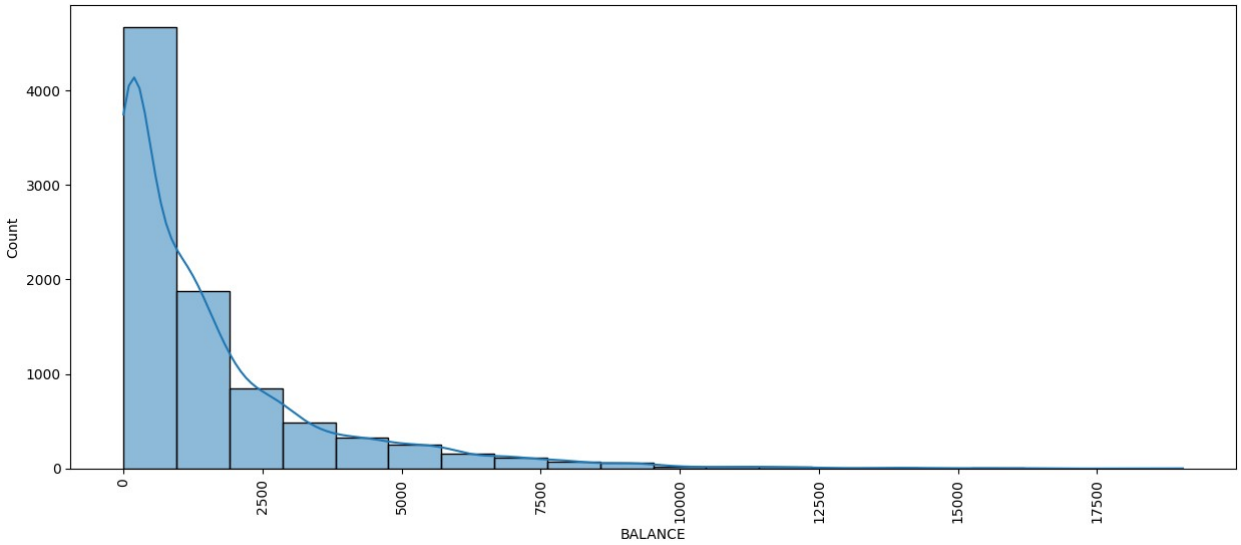
```
for i in discrete:
    plt.figure(figsize=(15, 6))
    ax = sns.countplot(x=i, data=df, palette='hls')

    for p in ax.patches:
        height = p.get_height()
        ax.annotate(f'{height}',
                    xy=(p.get_x() + p.get_width() / 2., height),
                    xytext=(0, 10),
                    textcoords='offset points',
                    ha='center', va='center')

    plt.show()
```

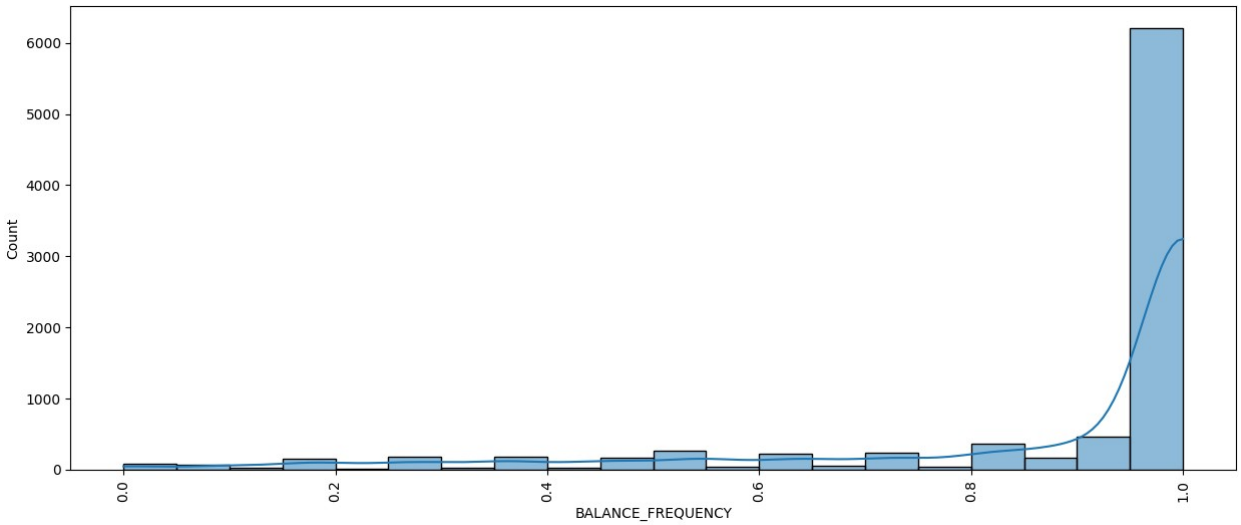


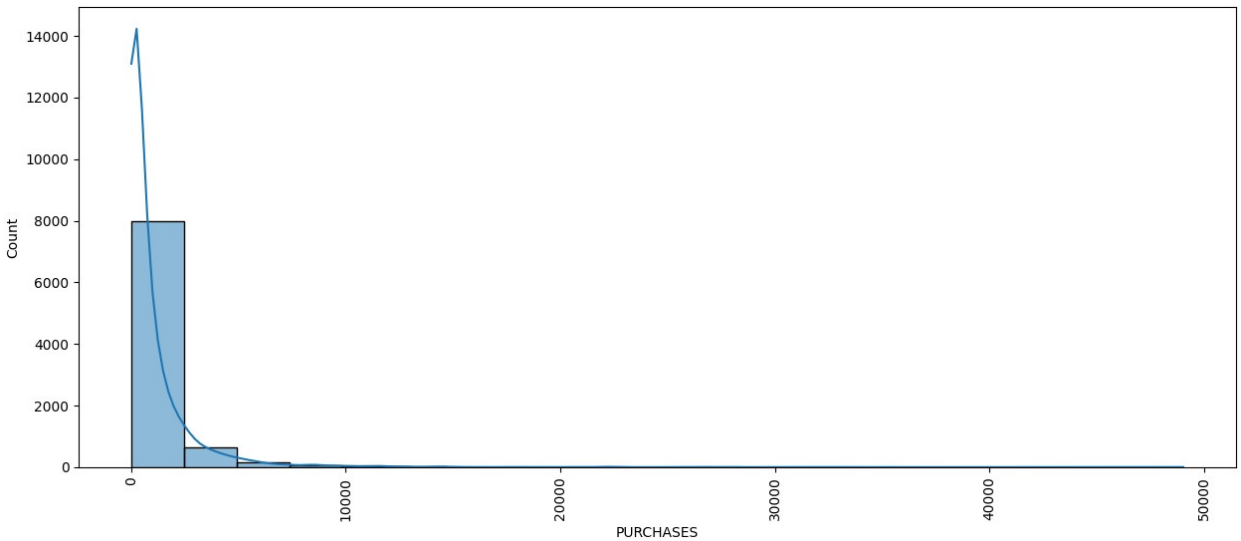

```
import plotly.express as px

for i in discrete:
    counts = df[i].value_counts()
    fig = px.pie(counts, values=counts.values, names=counts.index,
title=f'Distribution of {i}')
    fig.show()
```

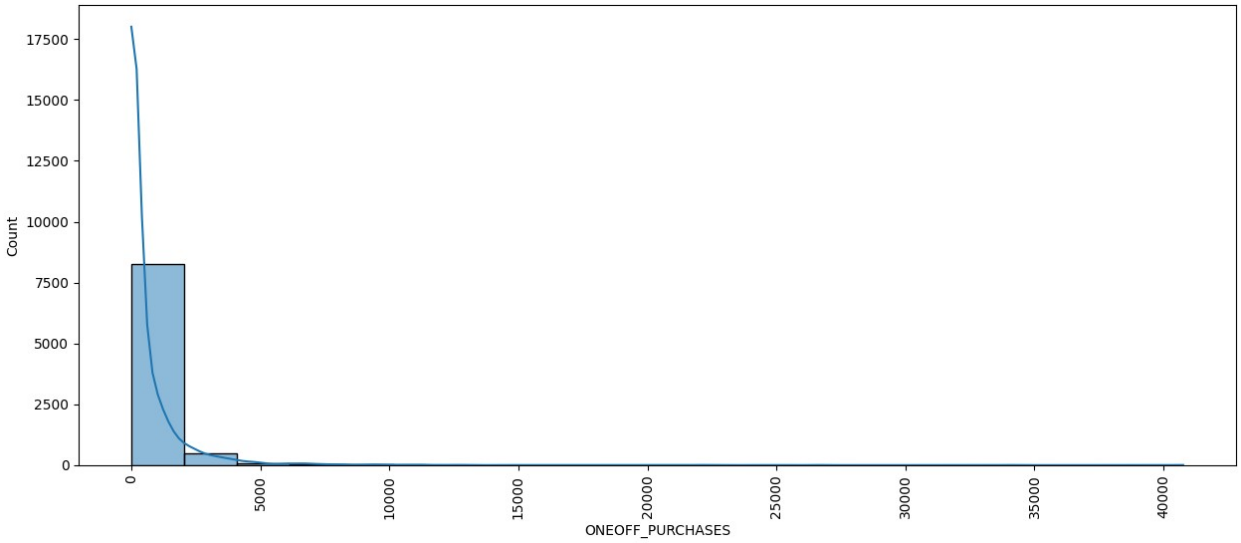


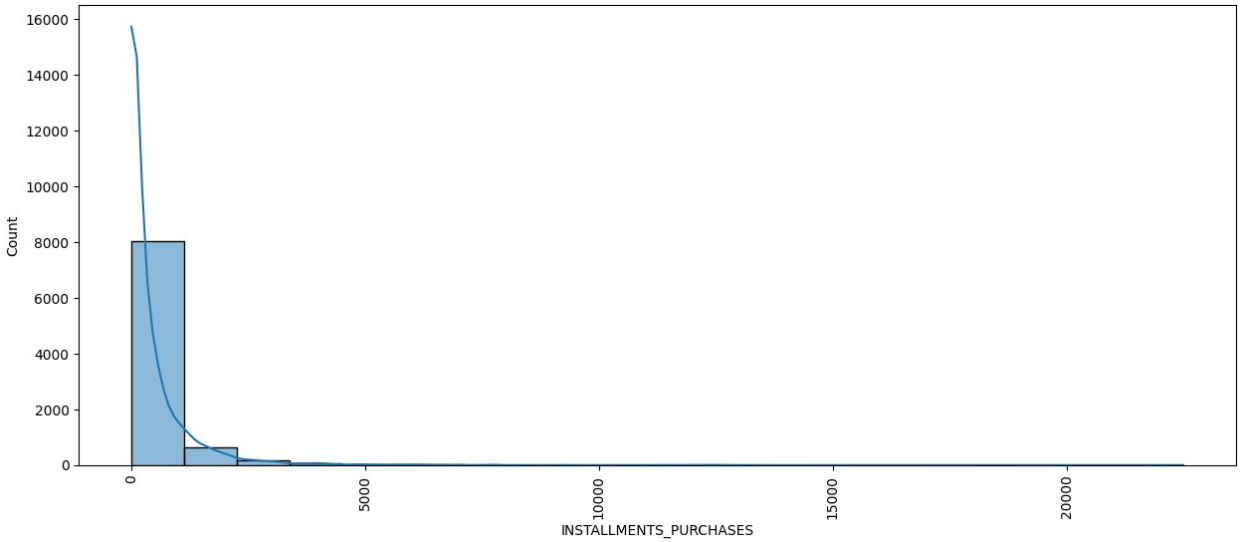

```
for i in continuous:
    plt.figure(figsize=(15,6))
    sns.histplot(df[i], bins = 20, kde = True, palette='hls')
    plt.xticks(rotation = 90)
    plt.show()
```

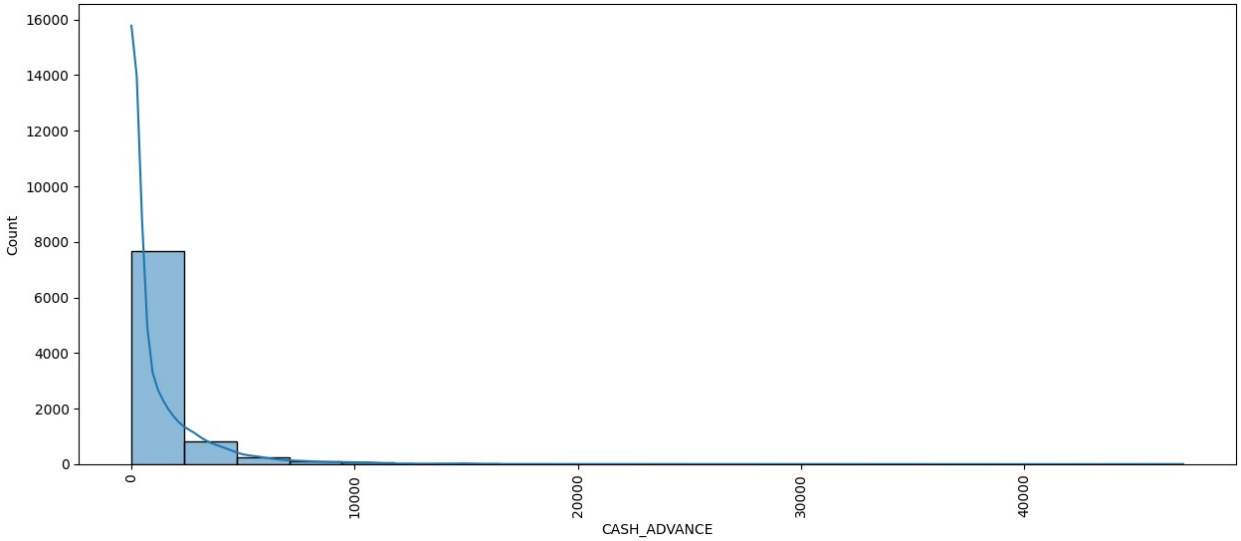

Count
4000
3000
2000
1000
0
0
2500
5000
7500
10000
12500
15000
17500
BALANCE

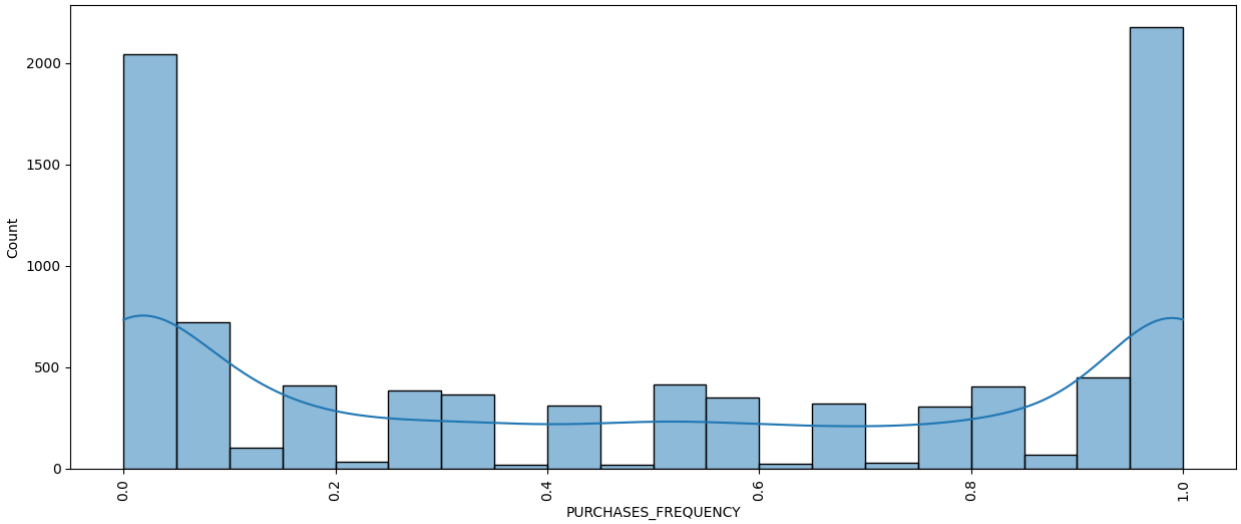

Count
6000
5000
4000
3000
2000
1000
0
0.0
0.2
0.4
0.6
0.8
1.0
BALANCE_FREQUENCY

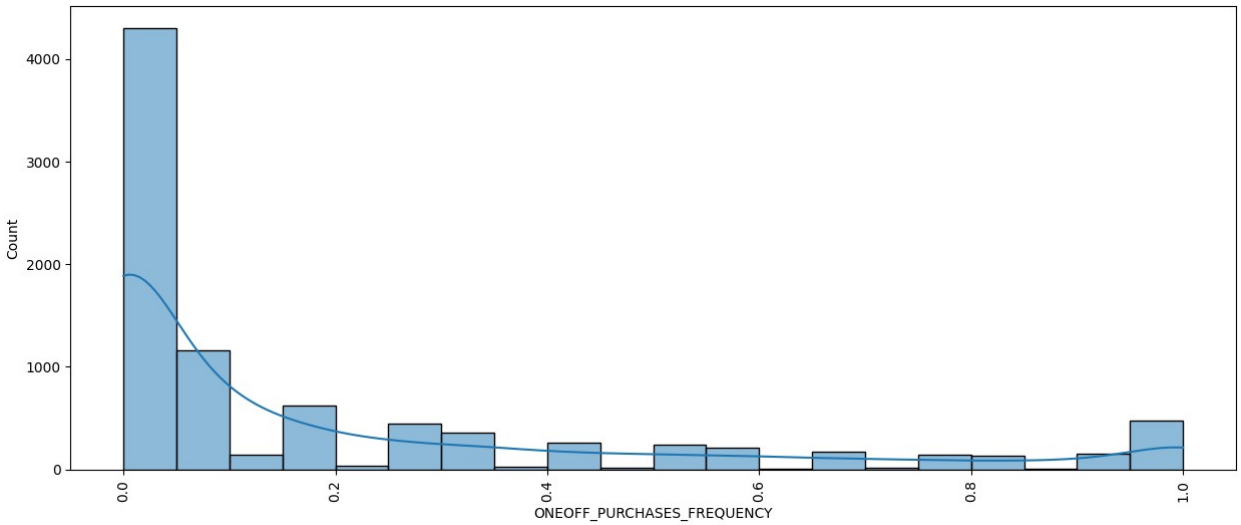

Count
14000
12000
10000
8000
6000
4000
2000
0
0
10000
20000
30000
40000
50000
PURCHASES

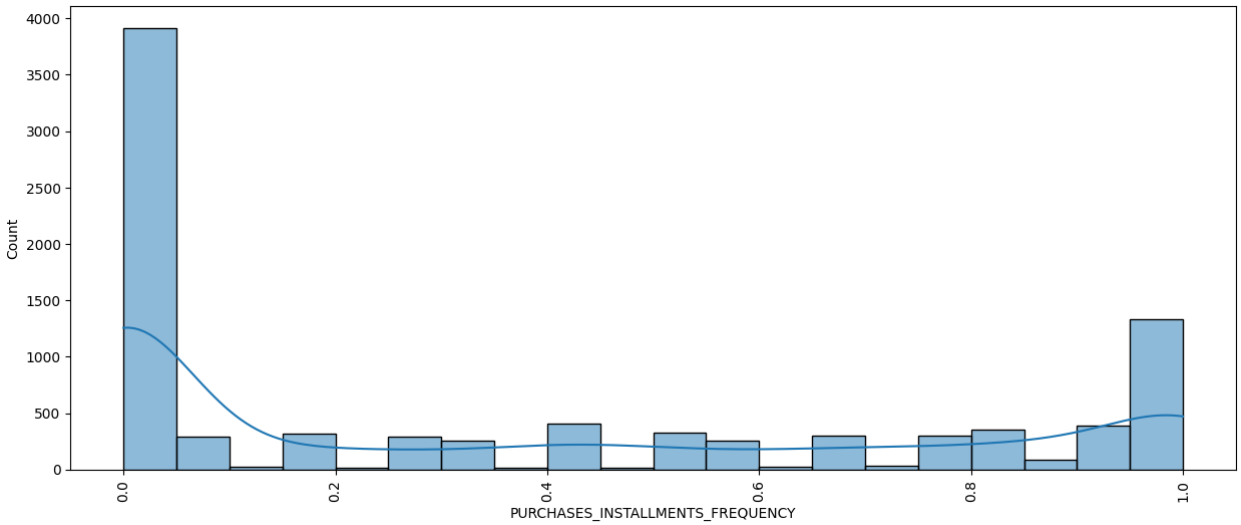

Count
17500
15000
12500
10000
7500
5000
2500
0
0
5000
10000
15000
20000
25000
30000
35000
40000
ONEOFF_PURCHASES

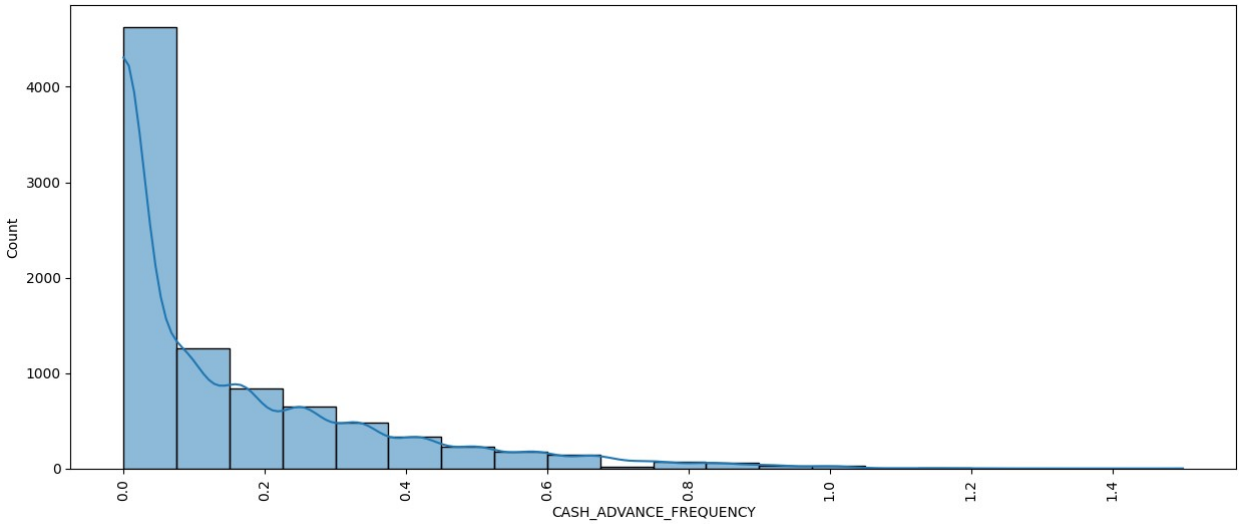

Count
16000
14000
12000
10000
8000
6000
4000
2000
0
0
5000
10000
15000
20000
INSTALLMENTS_PURCHASES

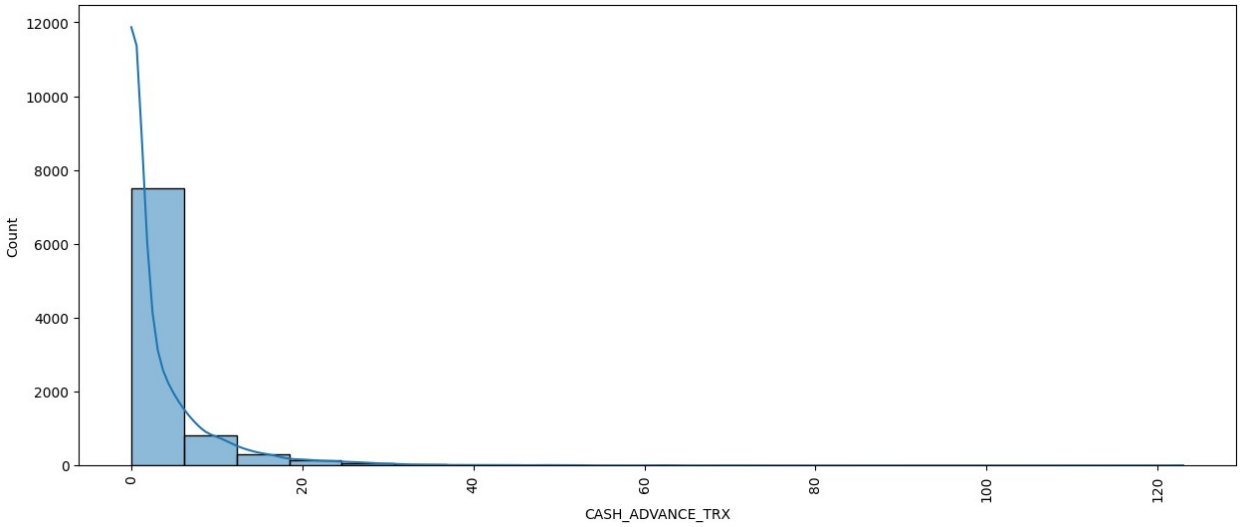

Count
16000
14000
12000
10000
8000
6000
4000
2000
0
0
10000
20000
30000
40000
CASH_ADVANCE

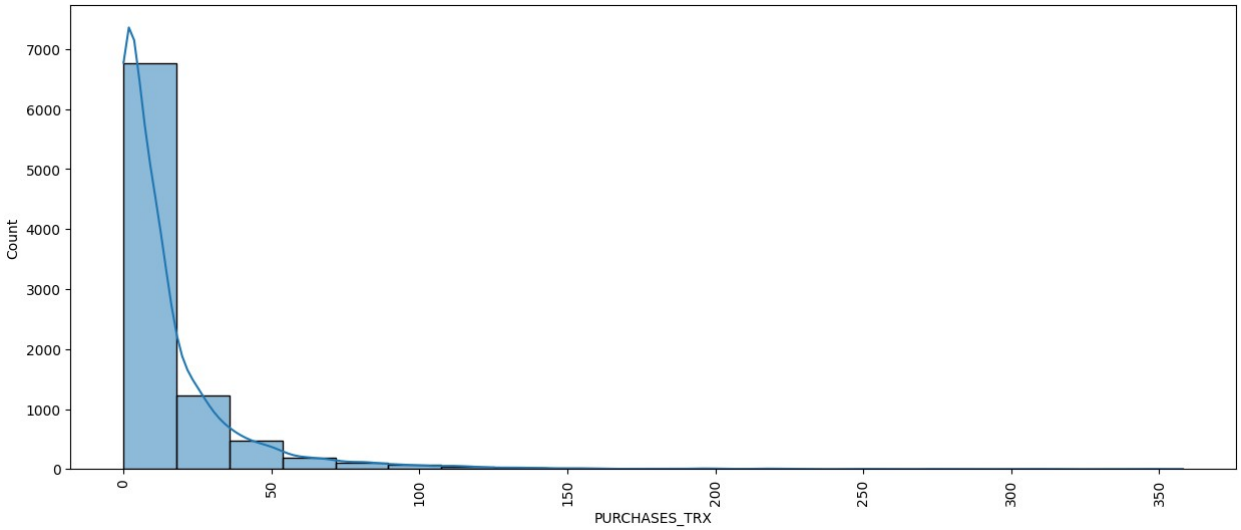

Count
2000
1500
1000
500
0
0.0
0.2
0.4
0.6
0.8
1.0
PURCHASES_FREQUENCY

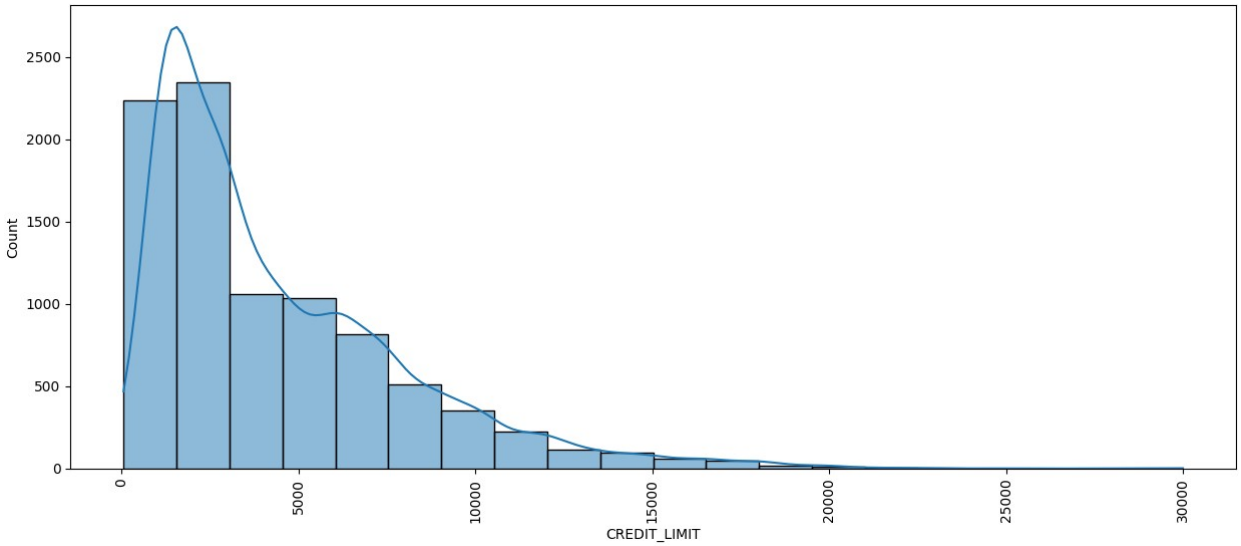

Count
4000
3000
2000
1000
0
0.0
0.2
0.4
0.6
0.8
1.0
ONEOFF_PURCHASES_FREQUENCY

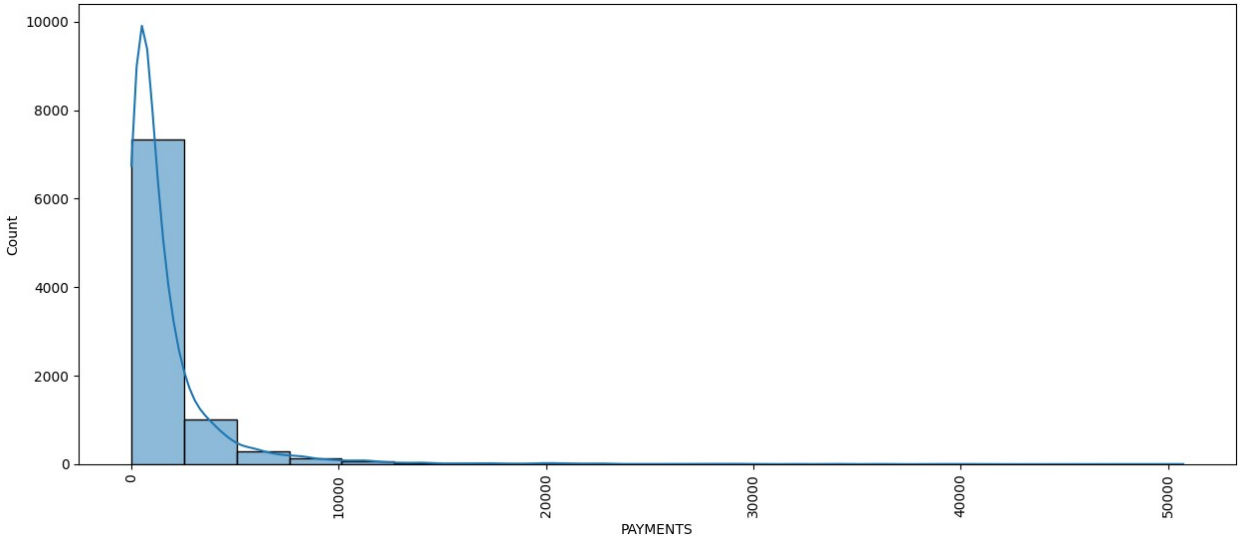

Count
4000
3500
3000
2500
2000
1500
1000
500
0
0.0
0.2
0.4
0.6
0.8
1.0
PURCHASES_INSTALLMENTS_FREQUENCY

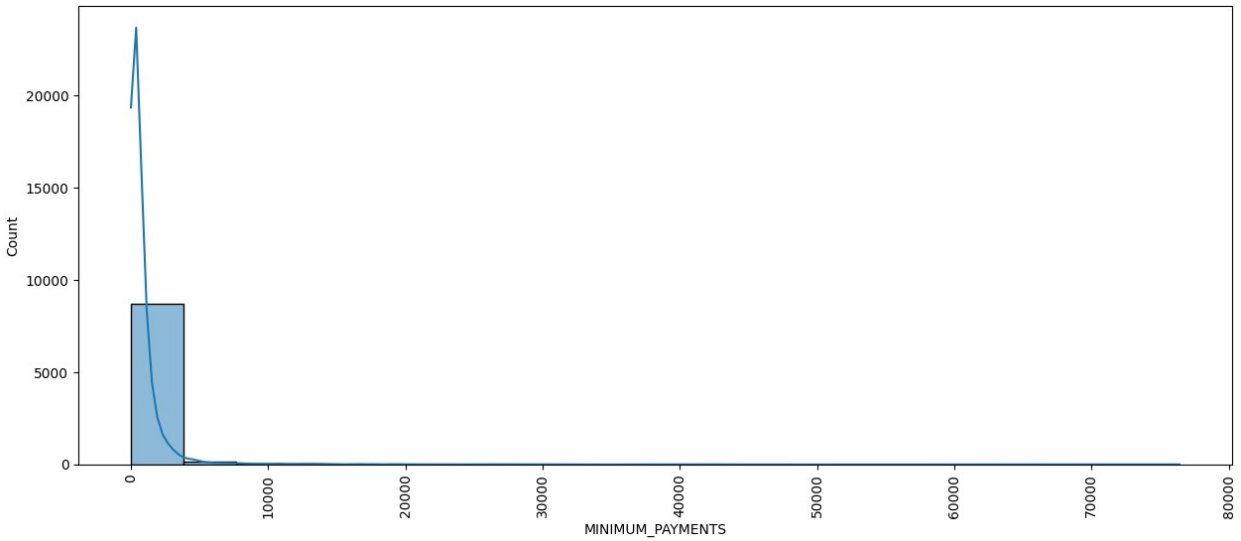

Count
4000
3000
2000
1000
0
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
CASH_ADVANCE_FREQUENCY

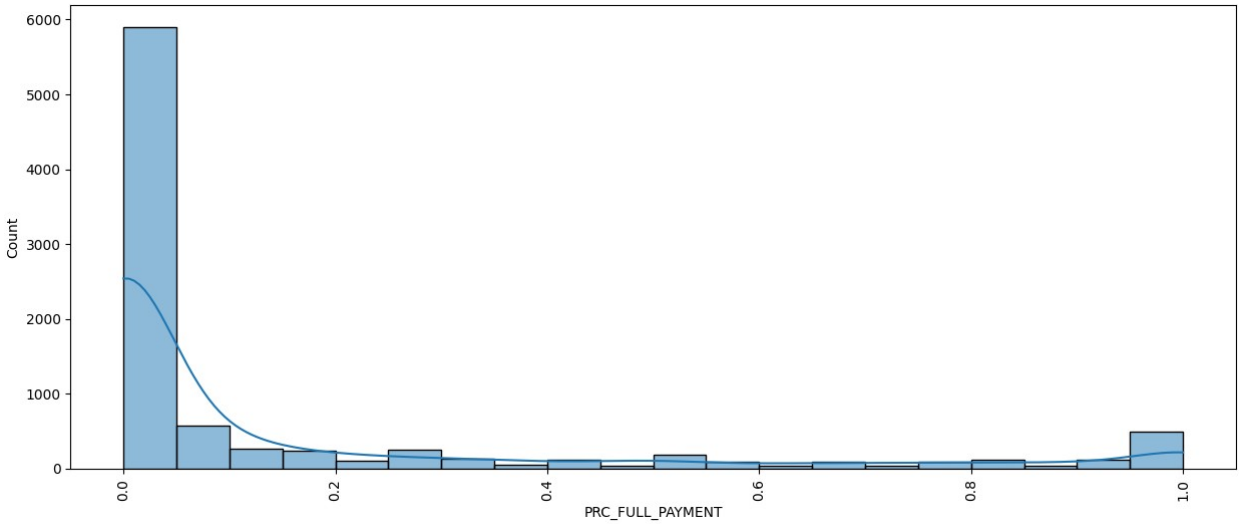

Count
12000
10000
8000
6000
4000
2000
0
0
20
40
60
80
100
120
CASH_ADVANCE_TRX


Count
7000
6000
5000
4000
3000
2000
1000
0
0
50
100
150
200
250
300
350
PURCHASES_TRX

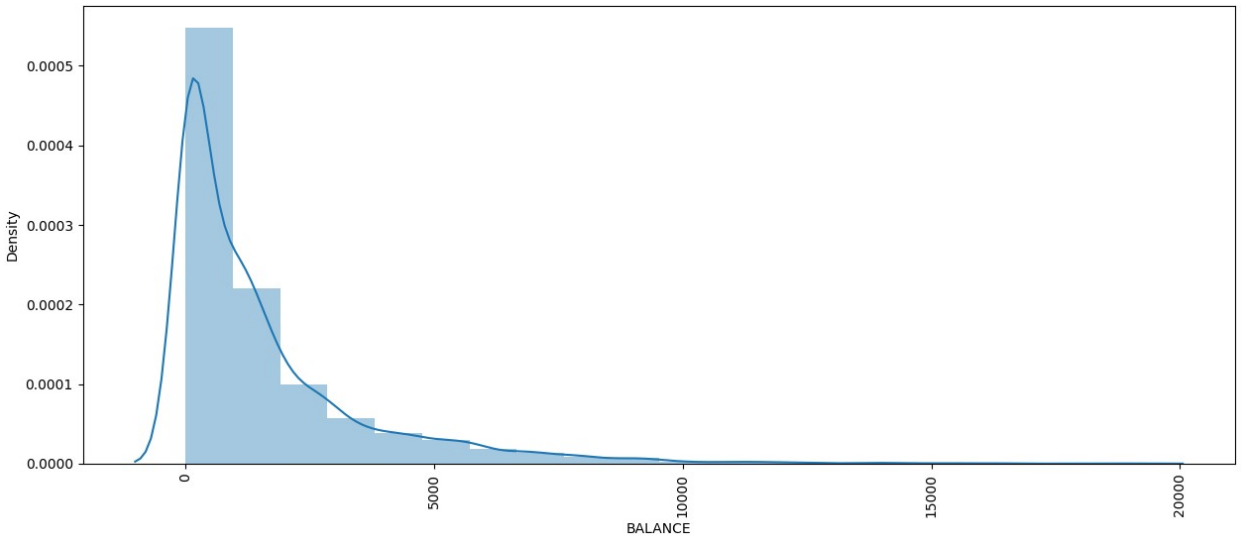

Count
2500
2000
1500
1000
500
0
0
5000
10000
15000
20000
25000
30000
CREDIT_LIMIT

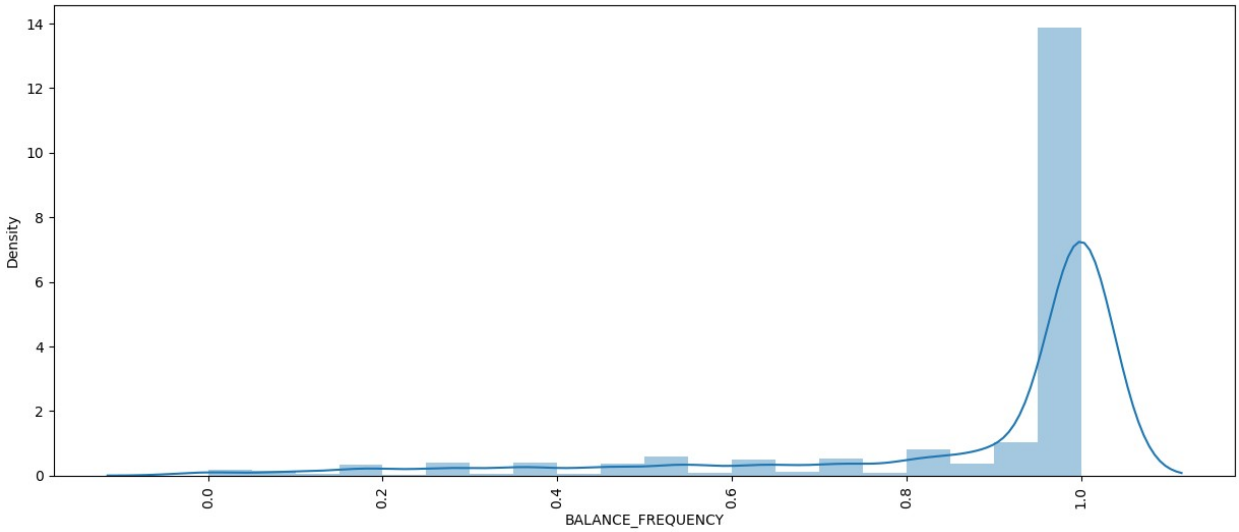

Count
10000
8000
6000
4000
2000
0
0
10000
20000
30000
40000
50000
PAYMENTS

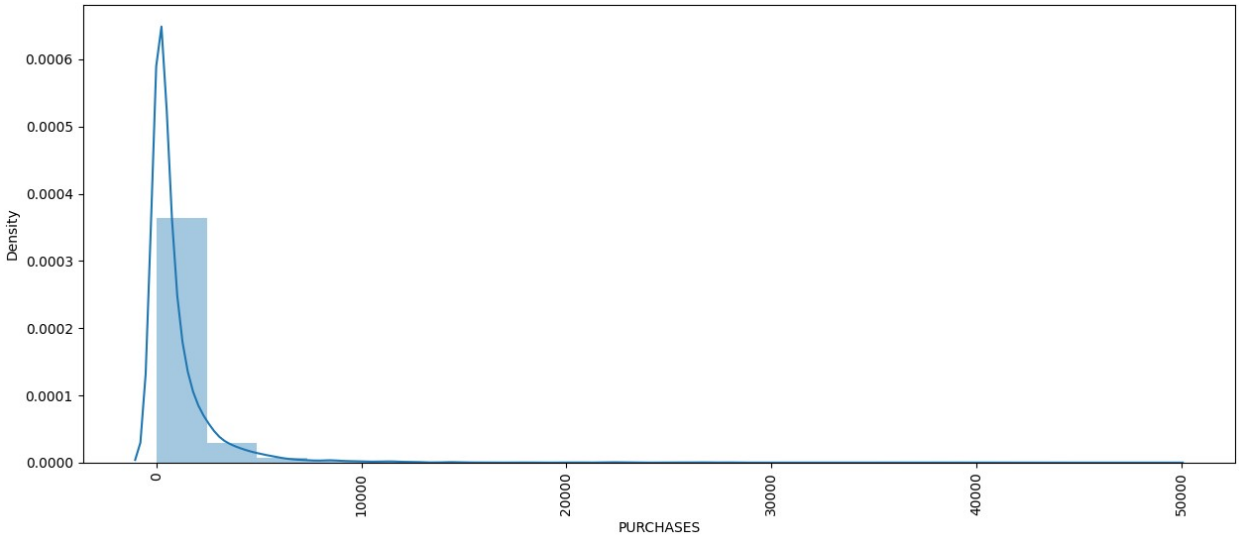

Count
20000
15000
10000
5000
0
0
10000
20000
30000
40000
50000
60000
70000
80000
MINIMUM_PAYMENTS

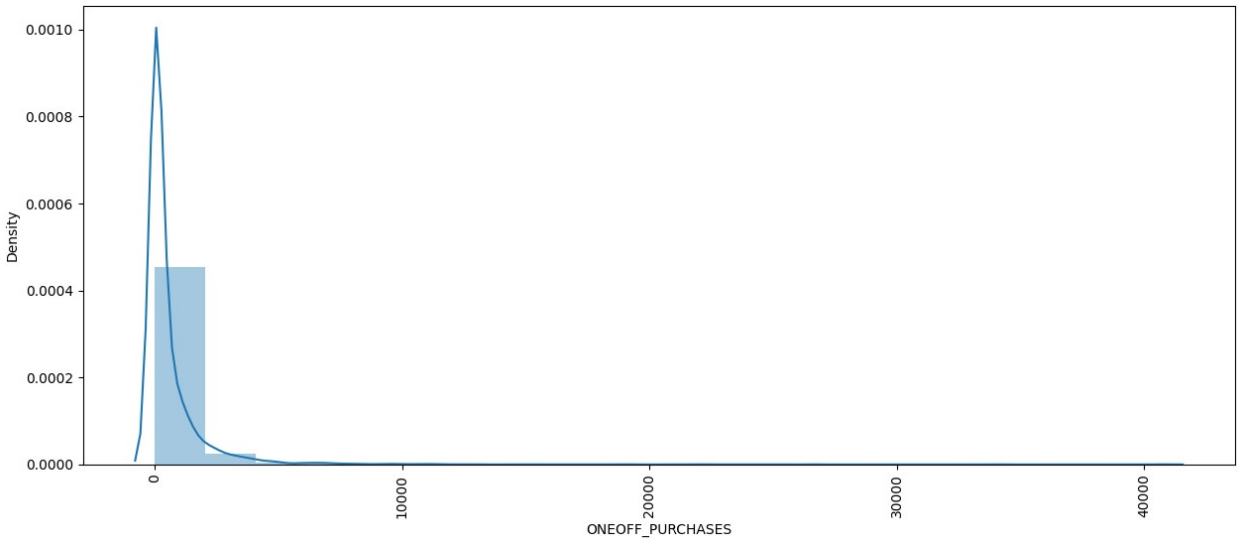


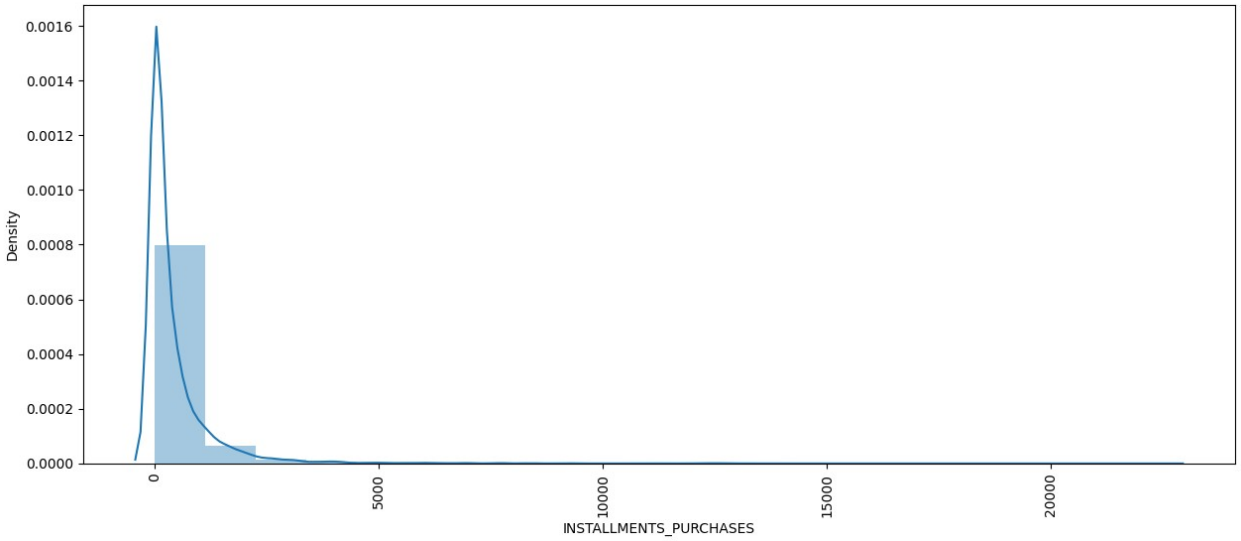

```
for i in continuous:
    plt.figure(figsize=(15,6))
    sns.distplot(df[i], bins = 20, kde = True)
    plt.xticks(rotation = 90)
    plt.show()
```

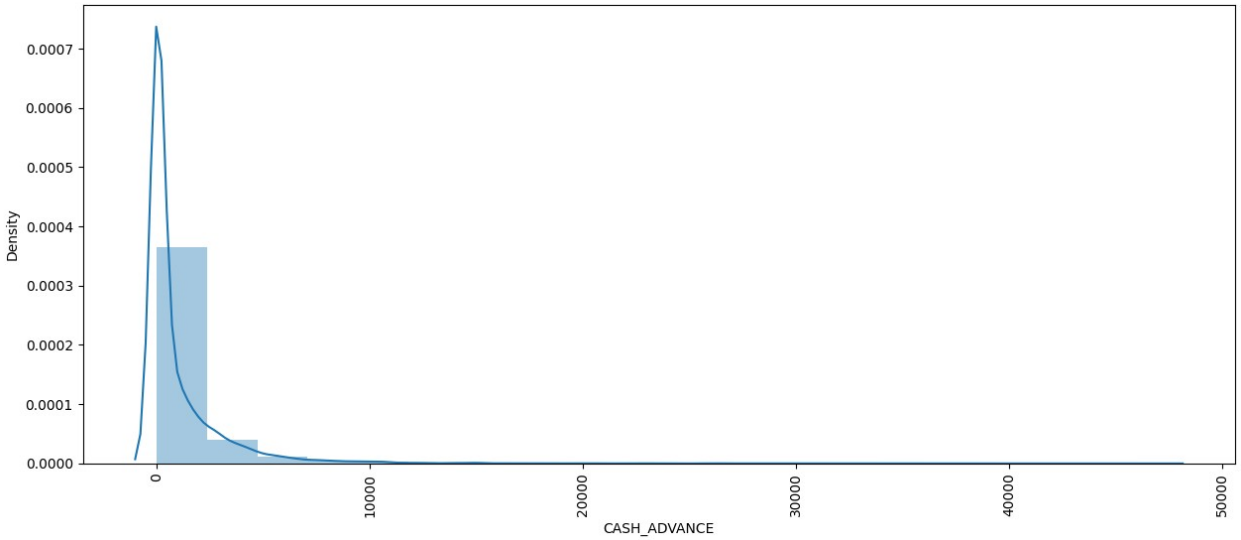


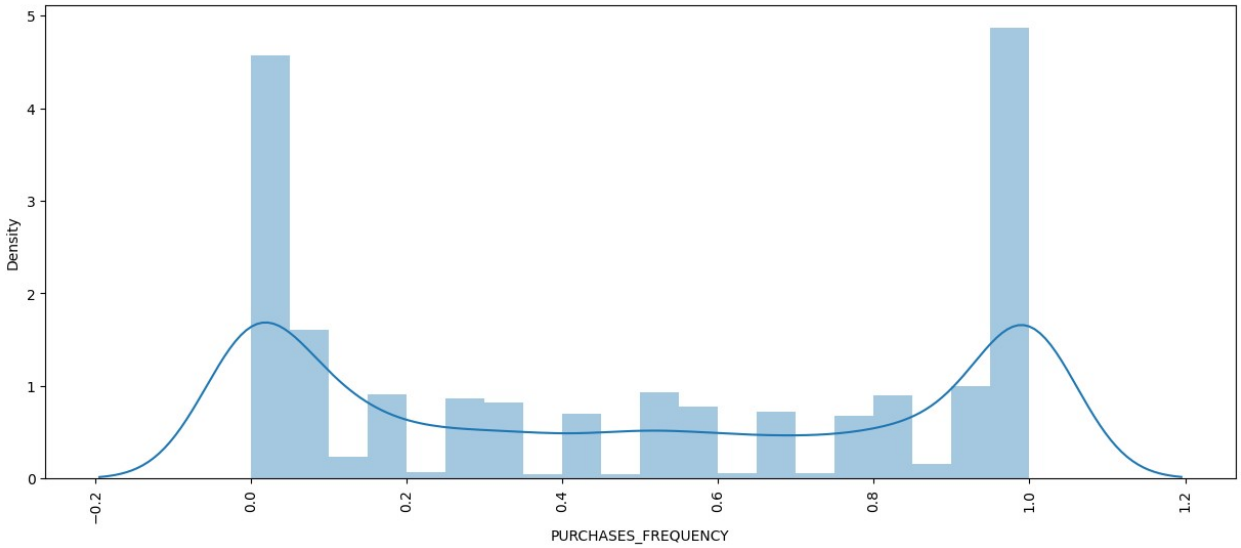

Density
BALANCE_FREQUENCY

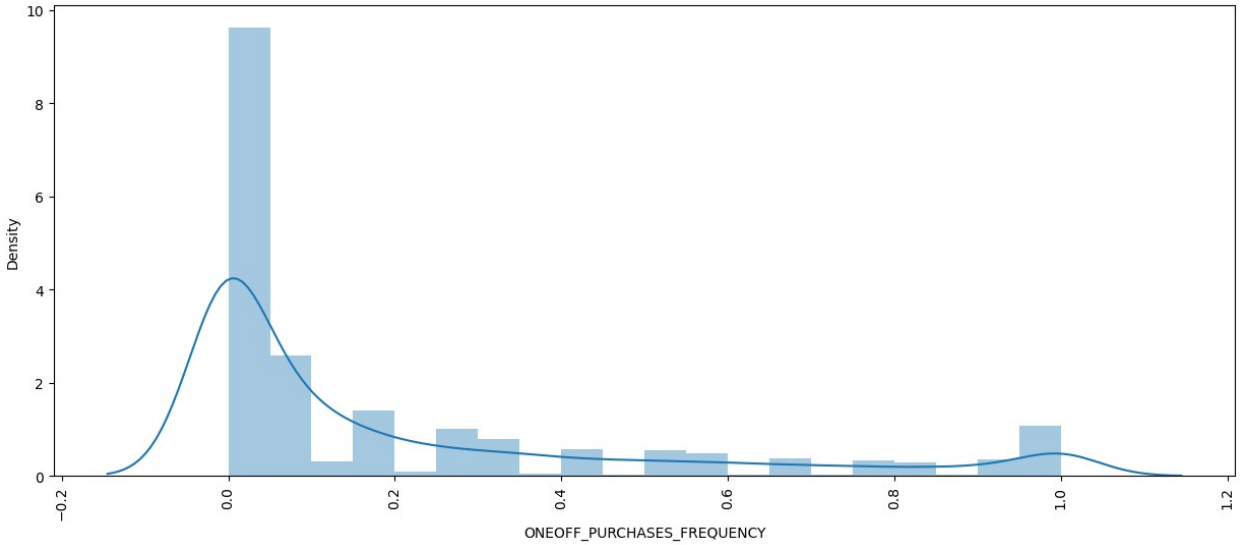

Density
PURCHASES

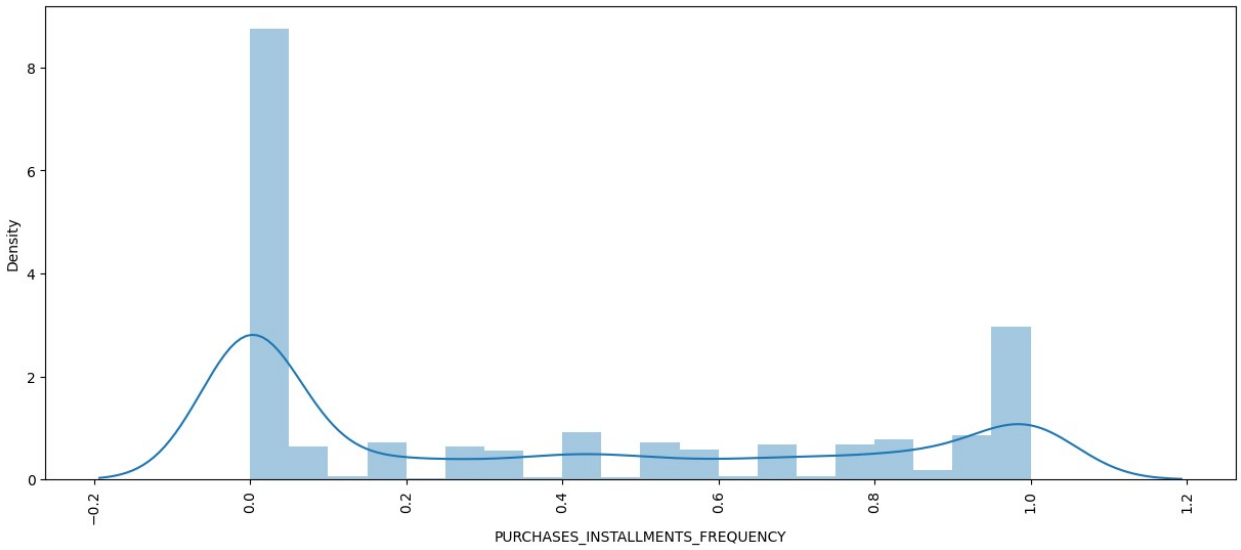

Density
ONEOFF_PURCHASES

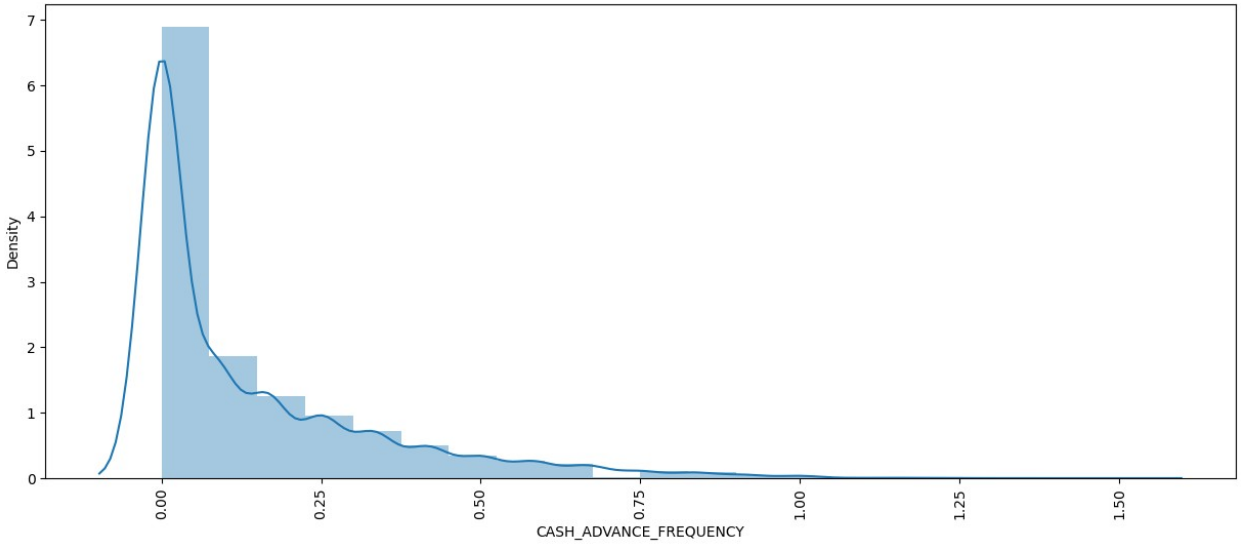

Density
0.0016
0.0014
0.0012
0.0010
0.0008
0.0006
0.0004
0.0002
0.0000
0
5000
10000
15000
20000
INSTALLMENTS_PURCHASES

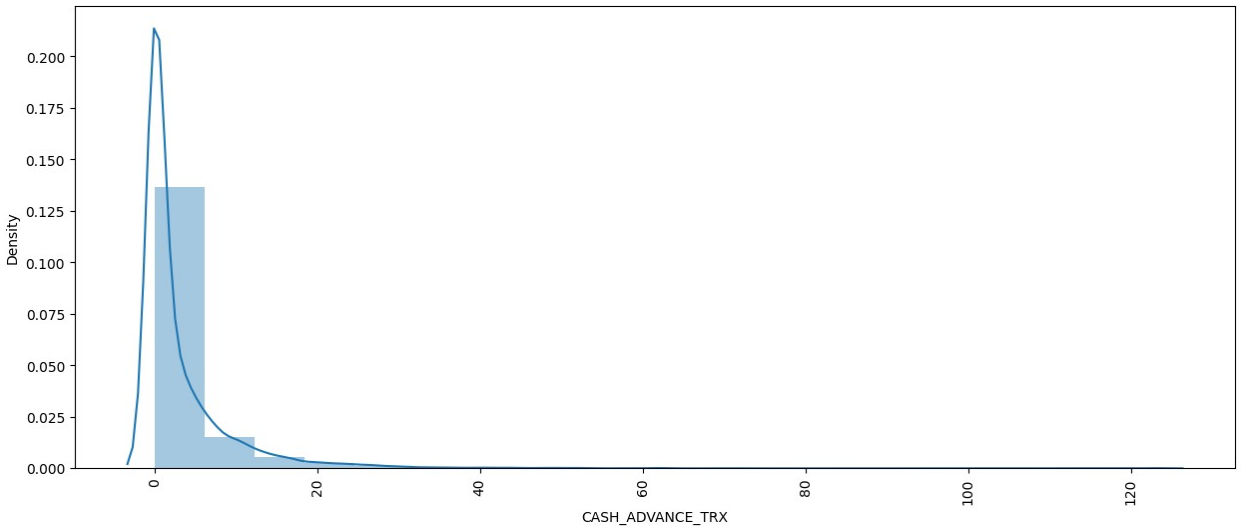

Density
0.0007
0.0006
0.0005
0.0004
0.0003
0.0002
0.0001
0.0000
0
10000
20000
30000
40000
50000
CASH_ADVANCE

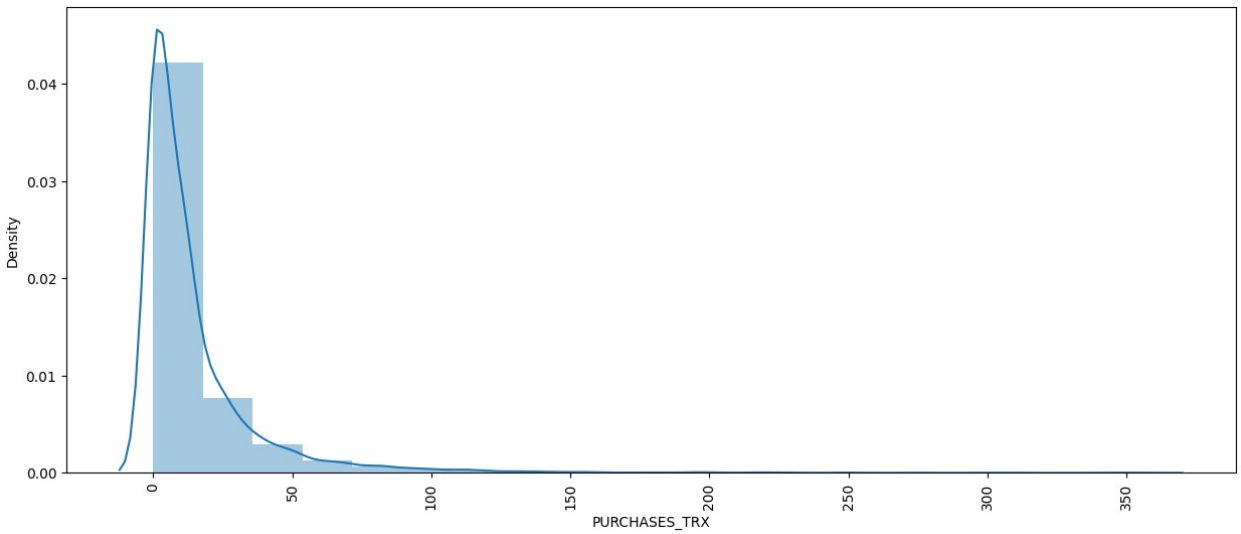

Density
5
4
3
2
1
0
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
PURCHASES_FREQUENCY

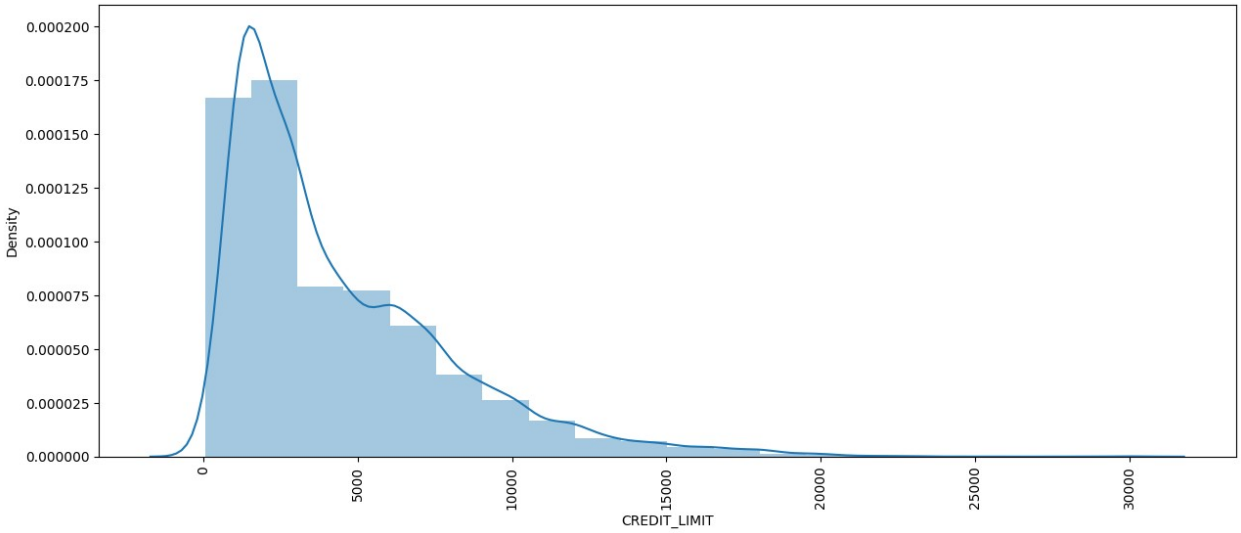

Density
ONEOFF_PURCHASES_FREQUENCY

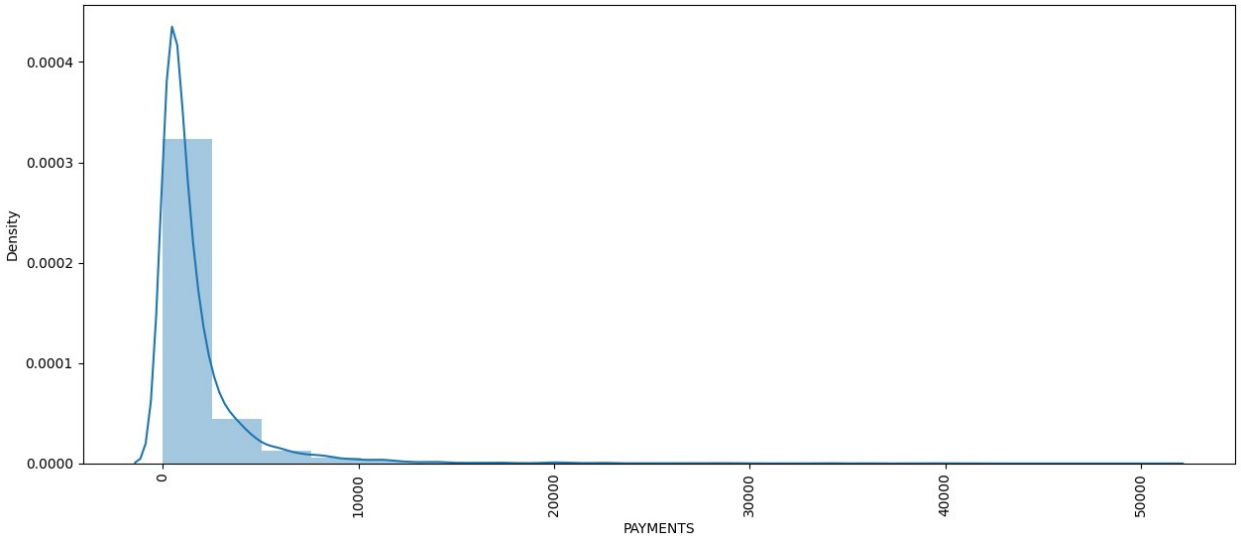

Density
PURCHASES_INSTALLMENTS_FREQUENCY

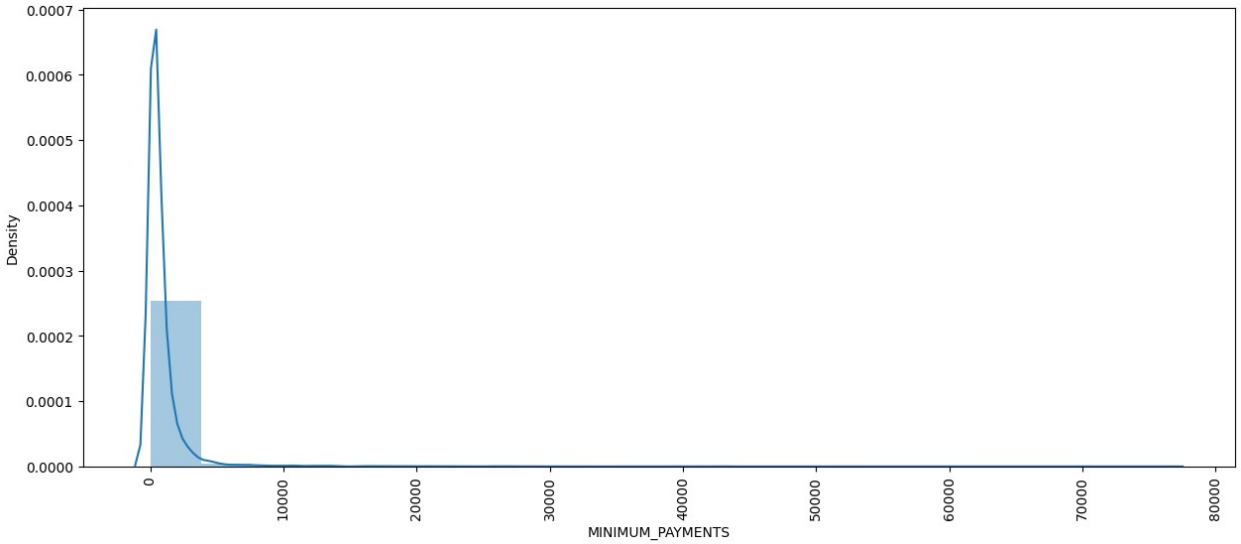

Density
CASH_ADVANCE_FREQUENCY

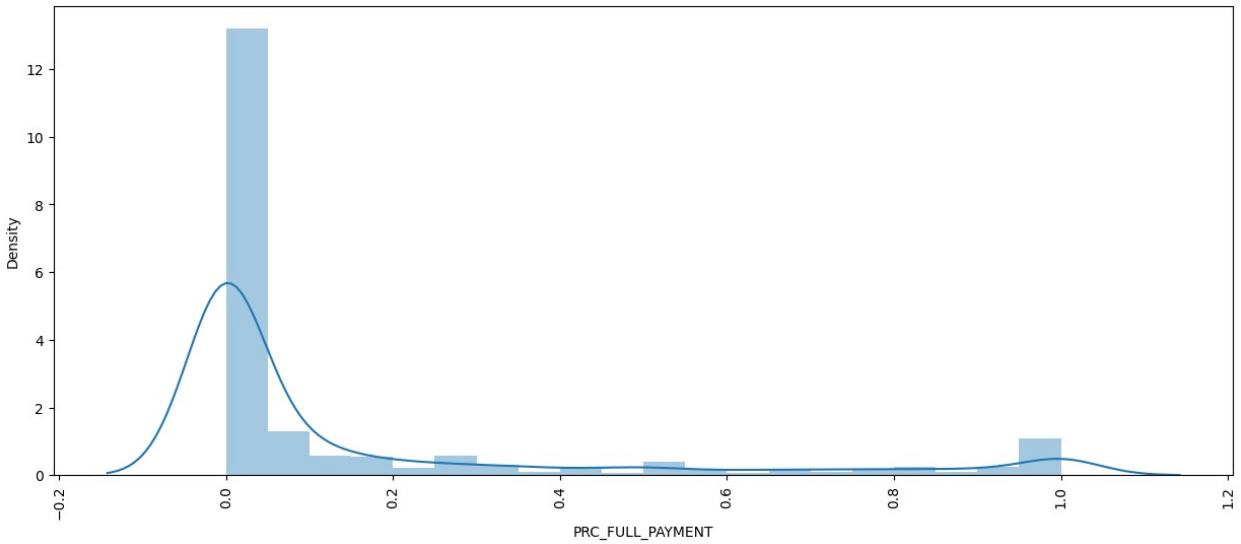

Density
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
0
20
40
60
80
100
120
CASH_ADVANCE_TRX


Density
0.04
0.03
0.02
0.01
0.00
0
50
100
150
200
250
300
350
PURCHASES_TRX

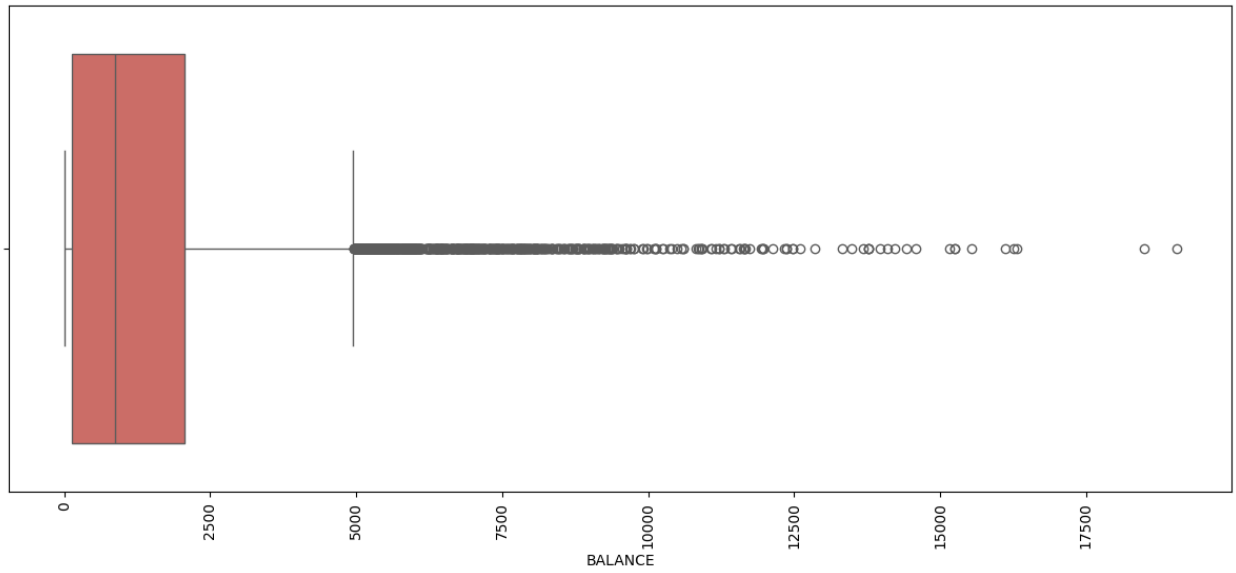

Density
0.000200
0.000175
0.000150
0.000125
0.000100
0.000075
0.000050
0.000025
0.000000
0
5000
10000
15000
20000
25000
30000
CREDIT_LIMIT

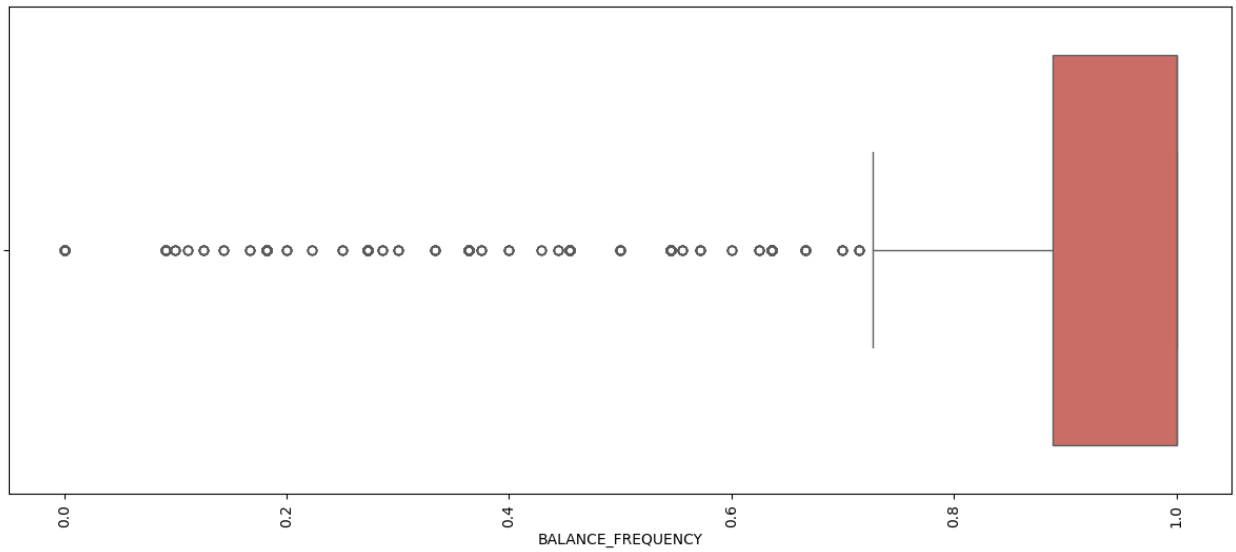

Density
0.0004
0.0003
0.0002
0.0001
0.0000
0
10000
20000
30000
40000
50000
PAYMENTS

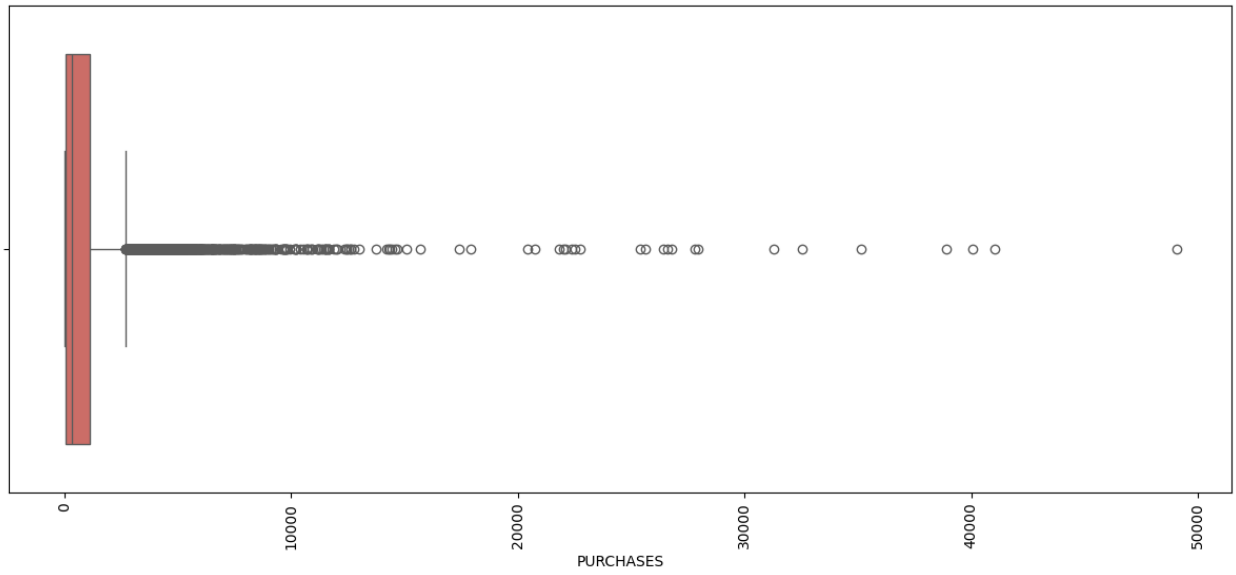

Density
0.0007
0.0006
0.0005
0.0004
0.0003
0.0002
0.0001
0.0000
0
10000
20000
30000
40000
50000
60000
70000
80000
MINIMUM_PAYMENTS

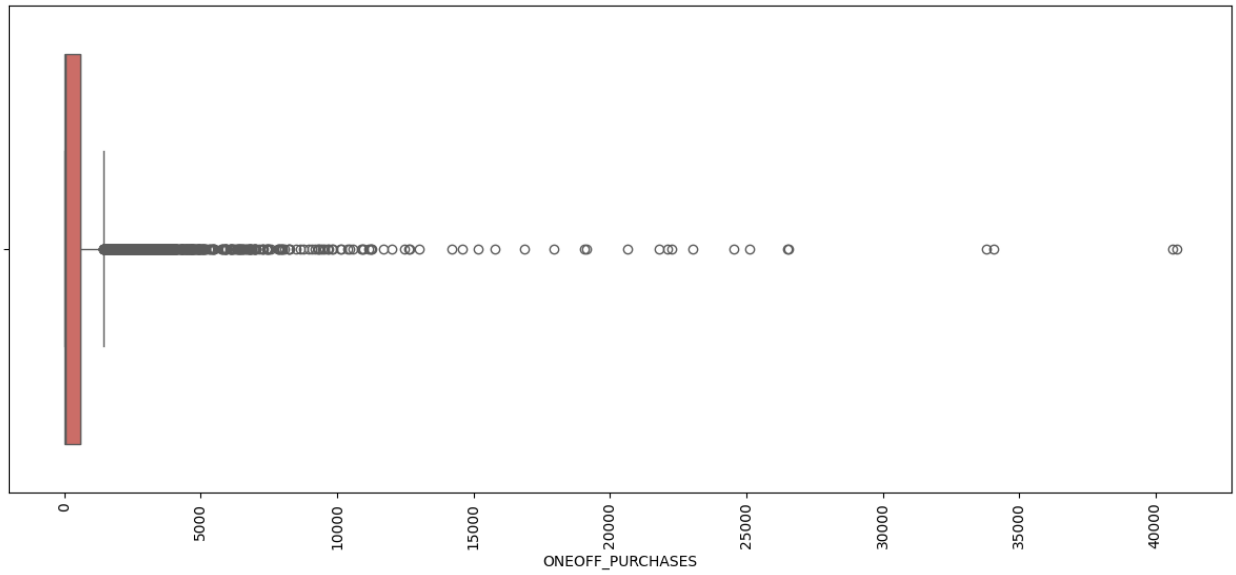

Density
12
10
8
6
4
2
0
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
PRC_FULL_PAYMENT

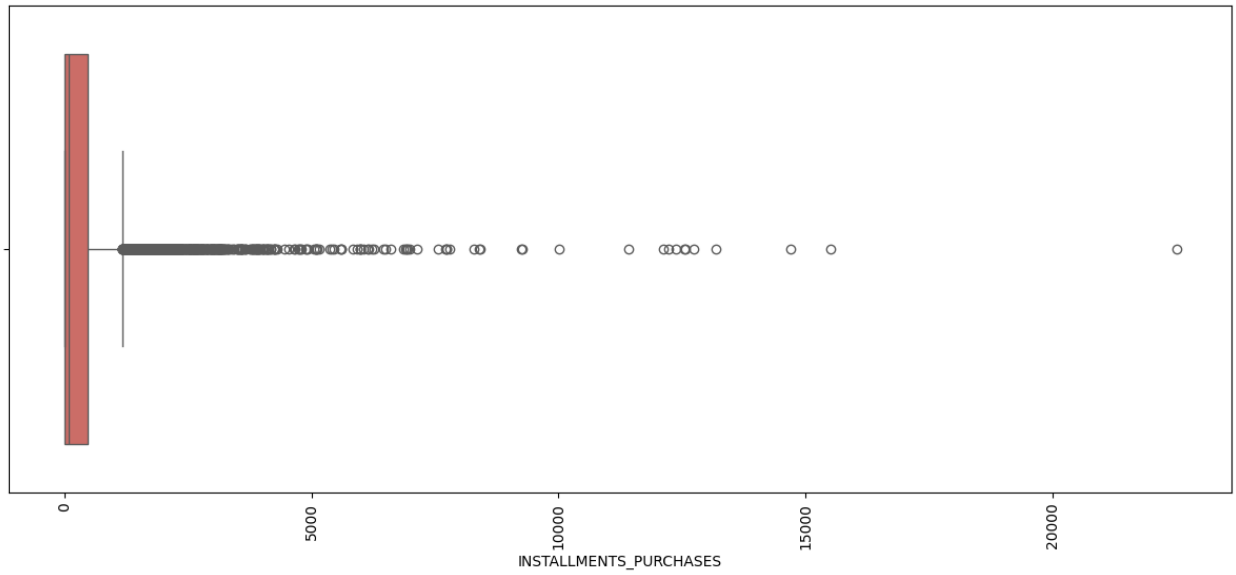
```
for i in continuous:
    plt.figure(figsize=(15, 6))
    sns.boxplot(x=i, data=df, palette='hls')
    plt.xticks(rotation=90)
    plt.show()
```

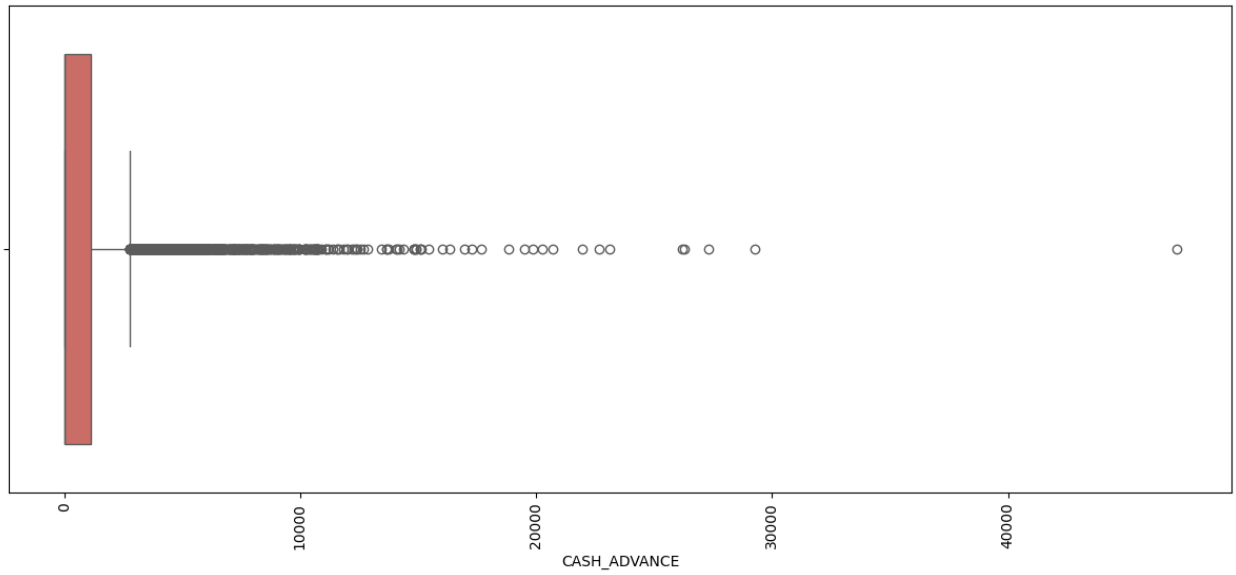


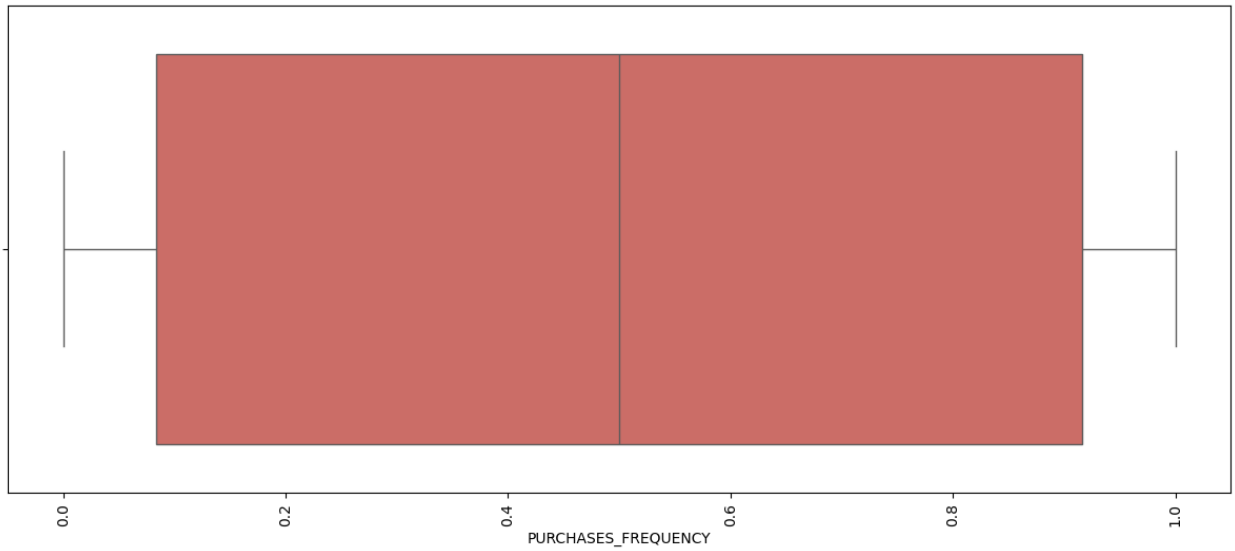


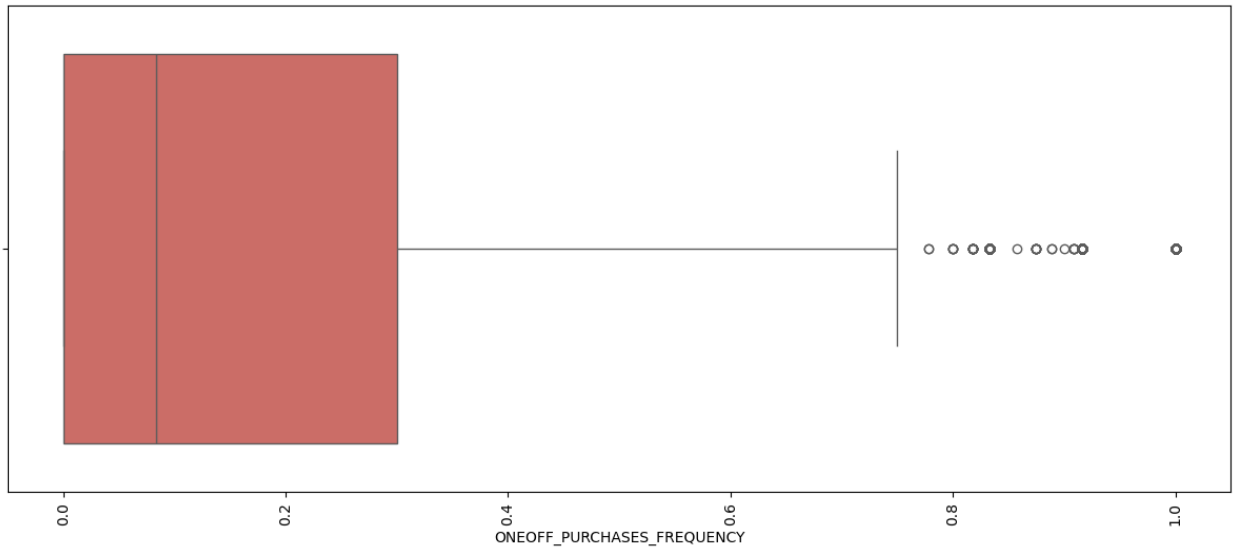

0
10000
20000
30000
40000
50000
PURCHASES

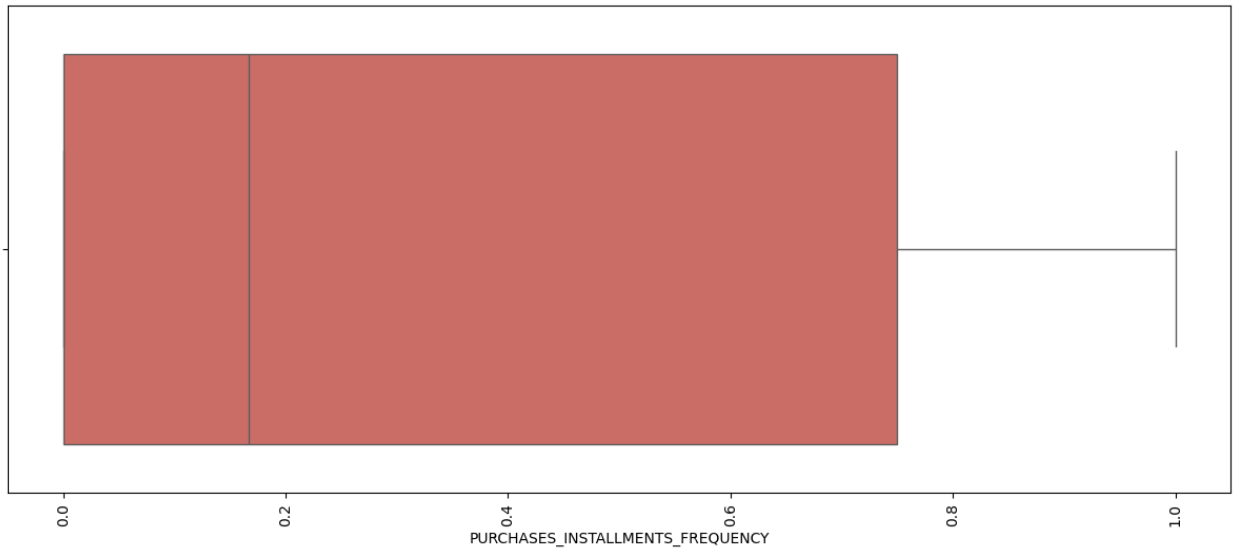

0
5000
10000
15000
20000
25000
30000
35000
40000
ONEOFF_PURCHASES

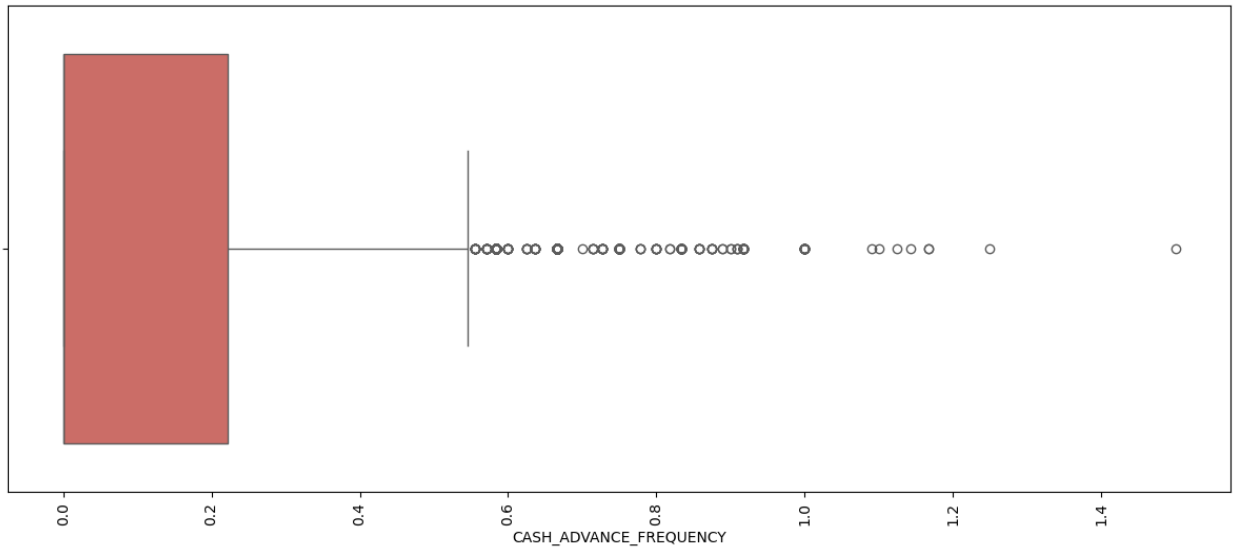

0
5000
10000
15000
20000
INSTALLMENTS_PURCHASES

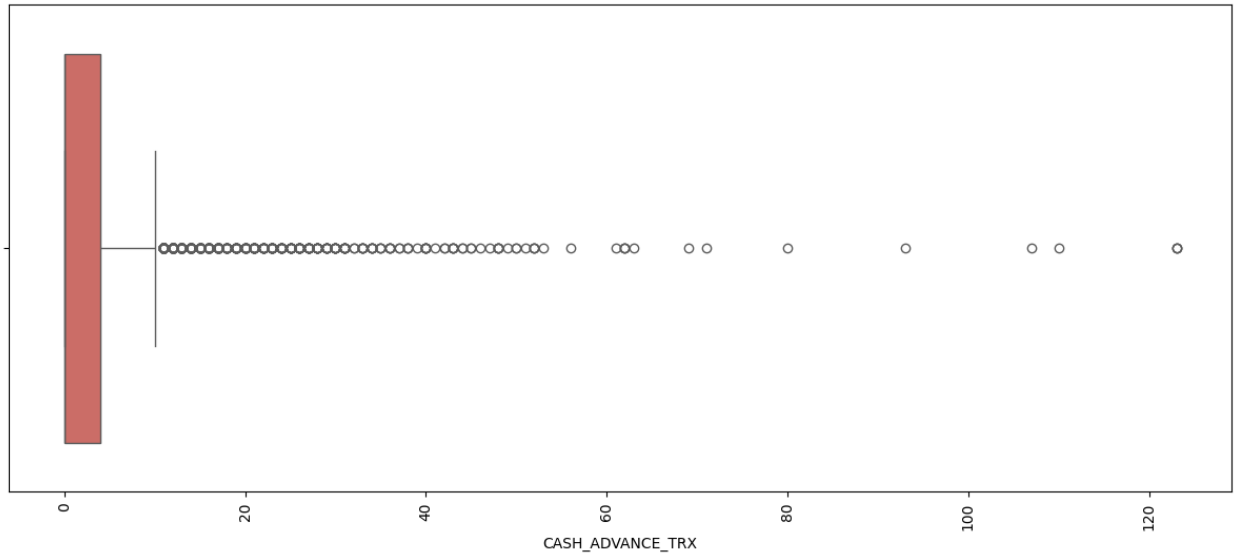

0
10000
20000
30000
40000
CASH_ADVANCE

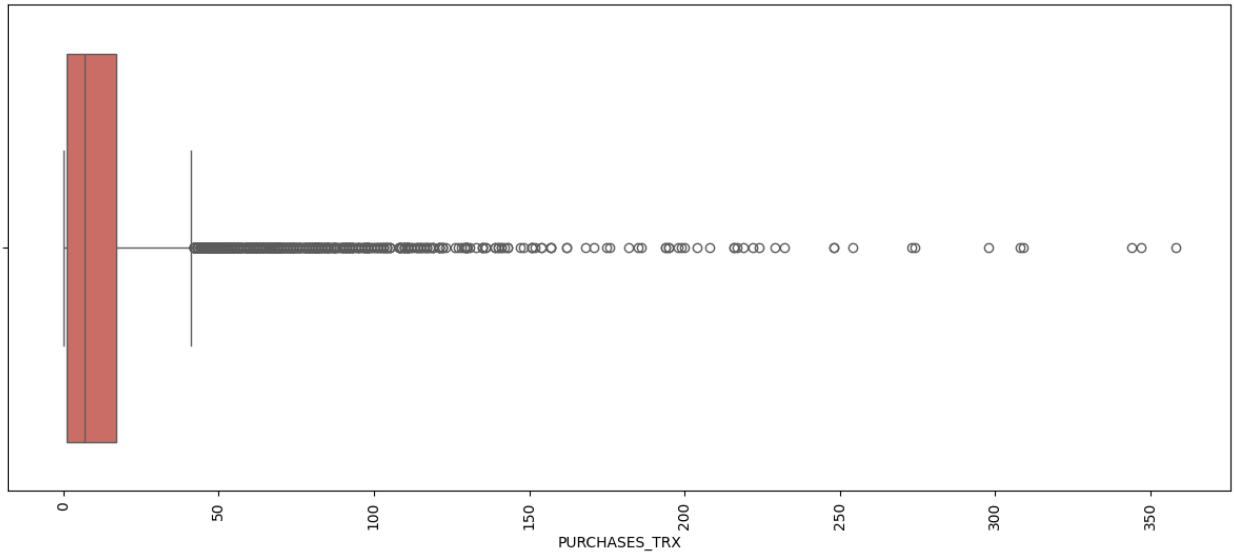

0.0
0.2
0.4
0.6
0.8
1.0
PURCHASES_FREQUENCY

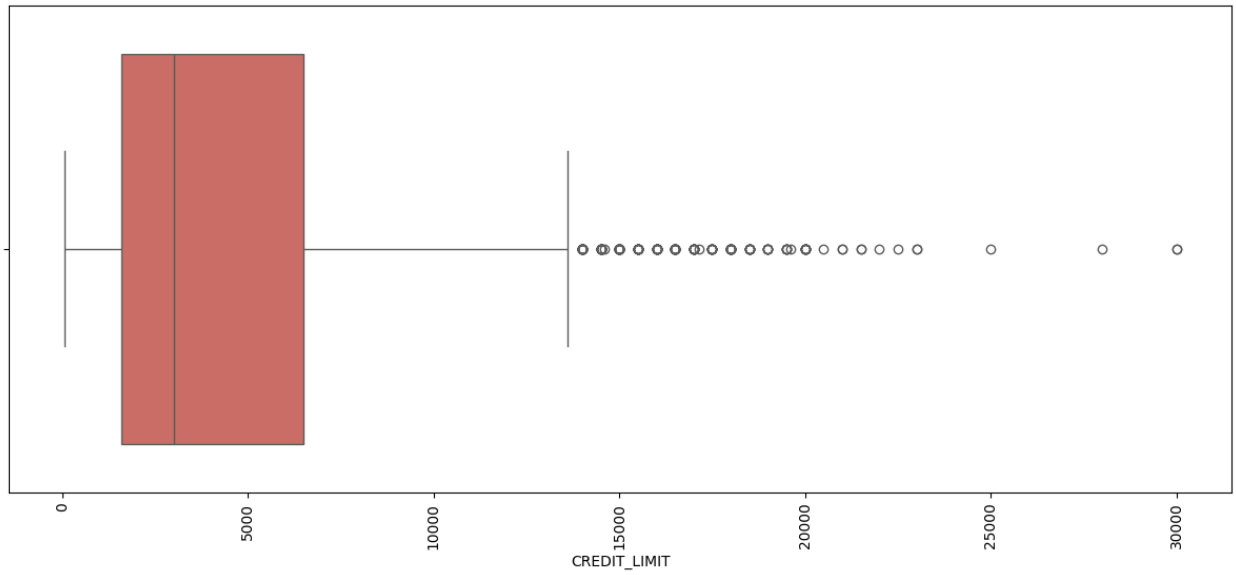

0.0
0.2
0.4
0.6
0.8
1.0
ONEOFF_PURCHASES_FREQUENCY

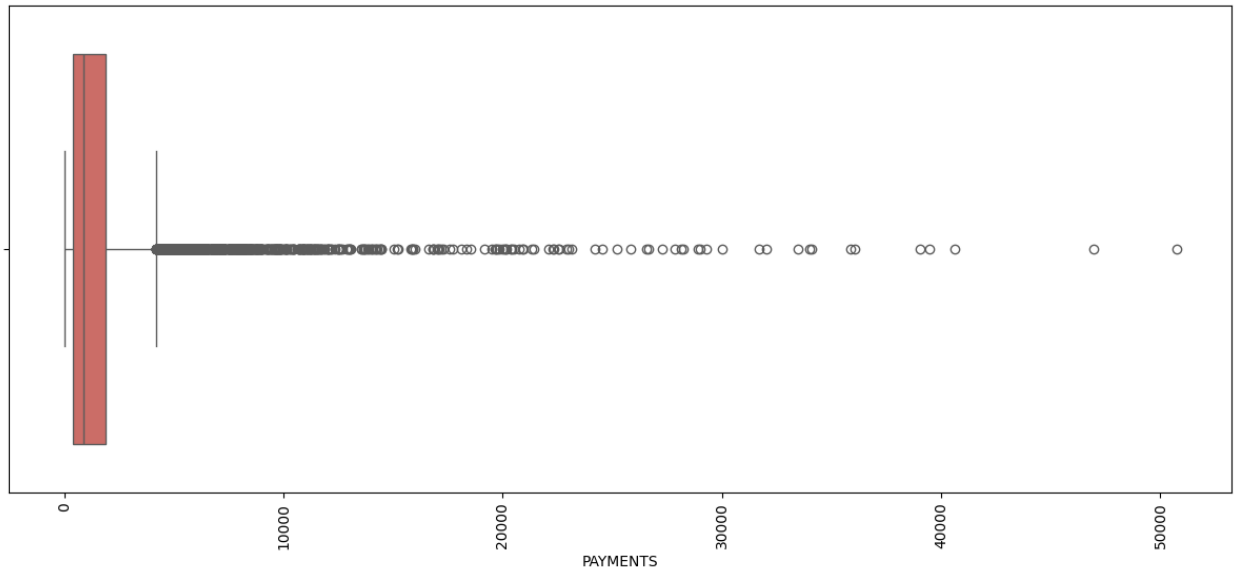

0.0
0.2
0.4
0.6
0.8
1.0
PURCHASES_INSTALLMENTS_FREQUENCY

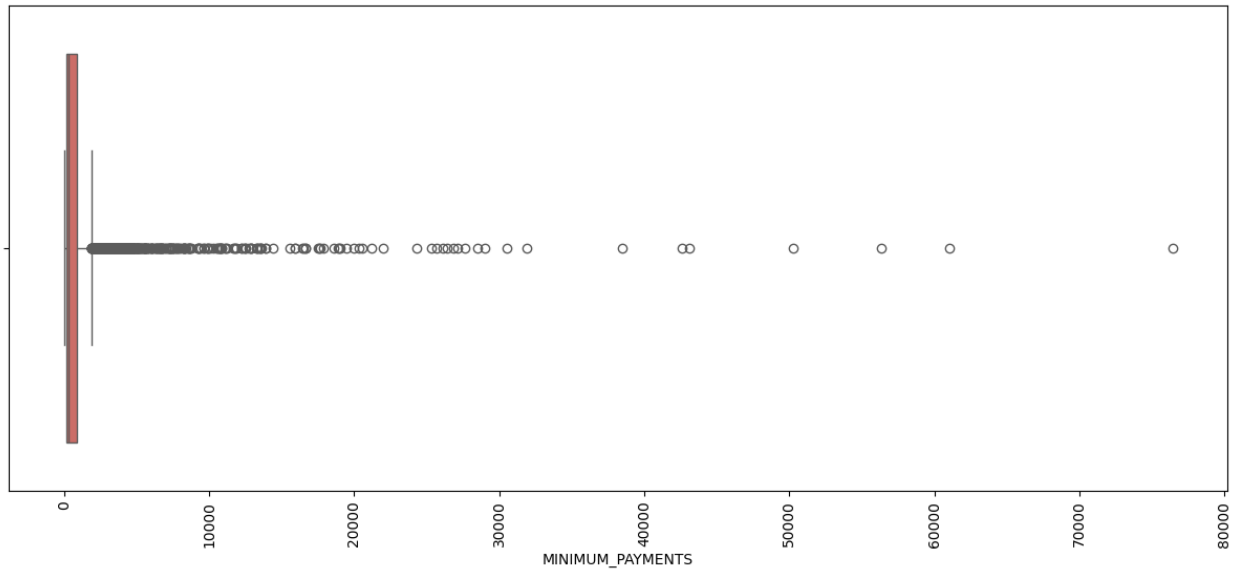

0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
CASH_ADVANCE_FREQUENCY

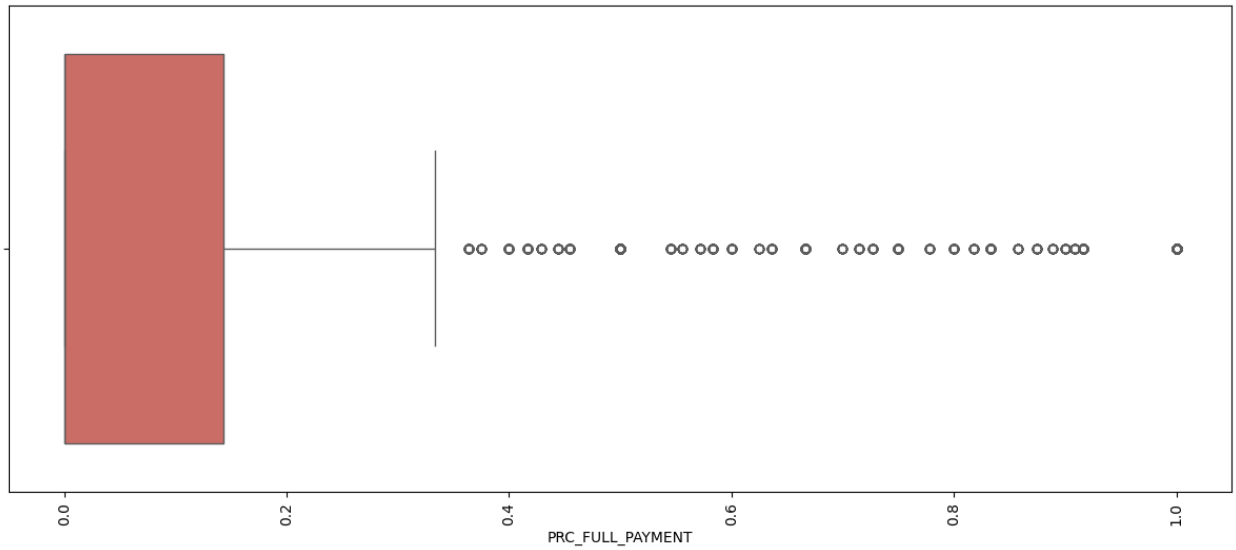

0
20
40
60
80
100
120
CASH_ADVANCE_TRX


0
50
100
150
200
250
300
350
PURCHASES_TRX

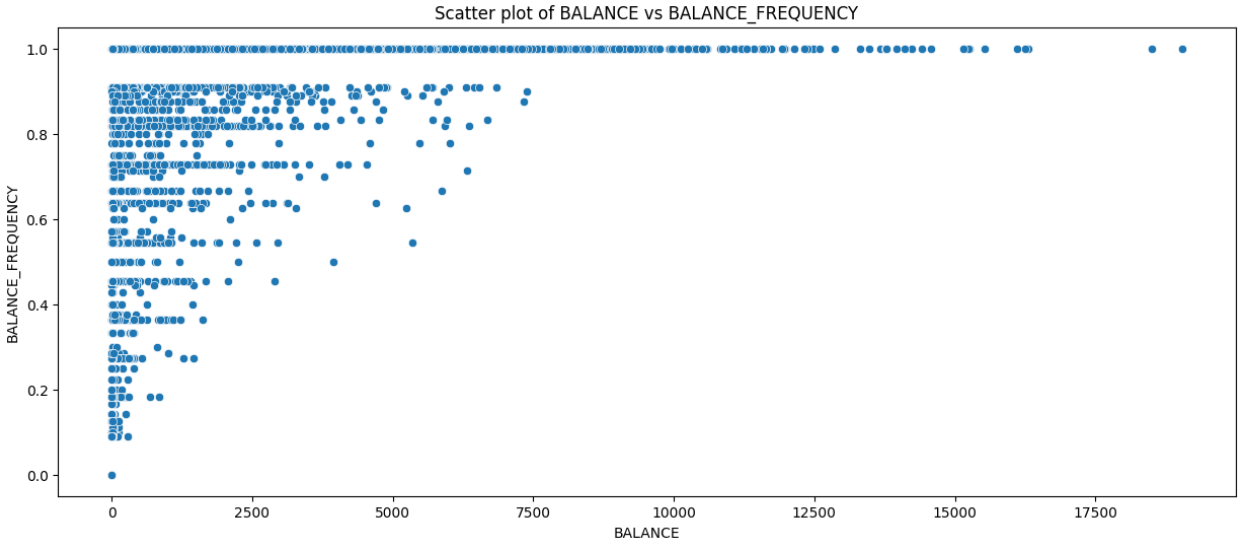

0
5000
10000
15000
20000
25000
30000
CREDIT_LIMIT

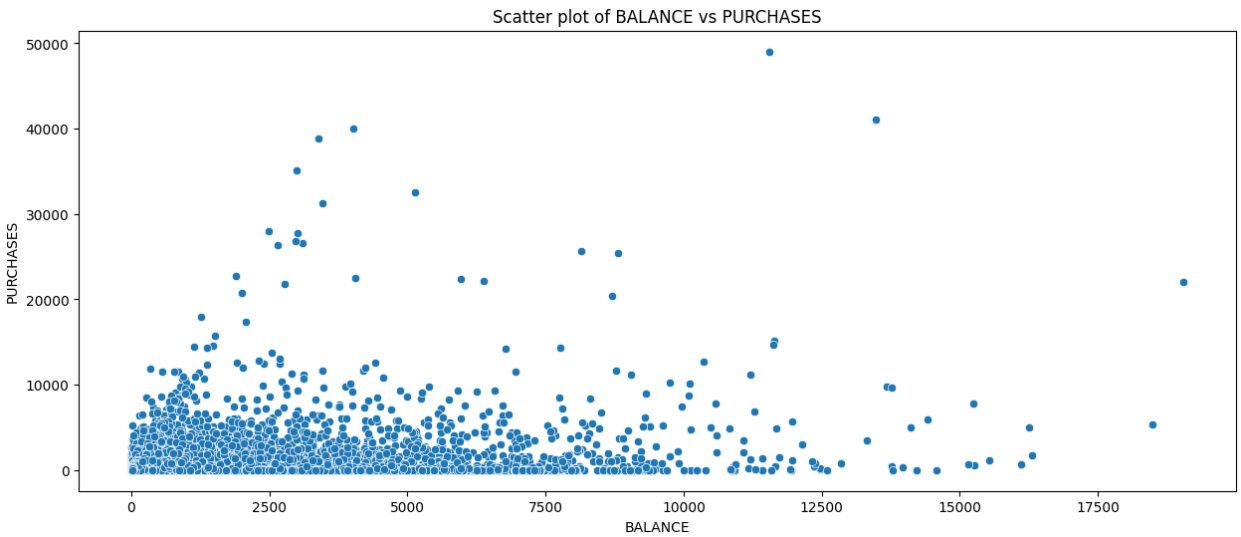

0
10000
20000
30000
40000
50000
PAYMENTS

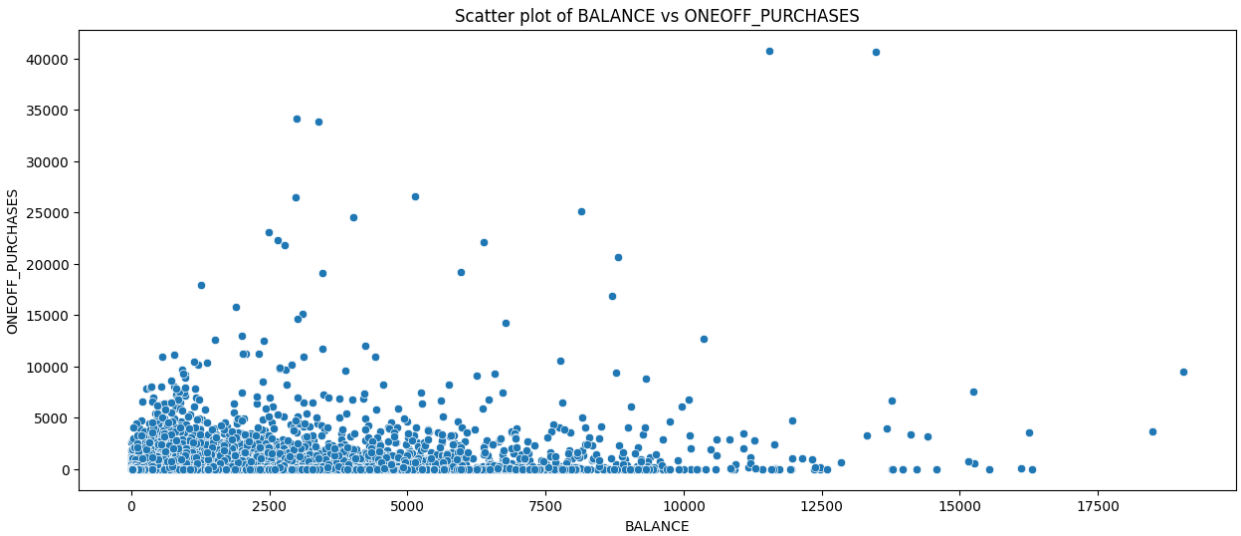


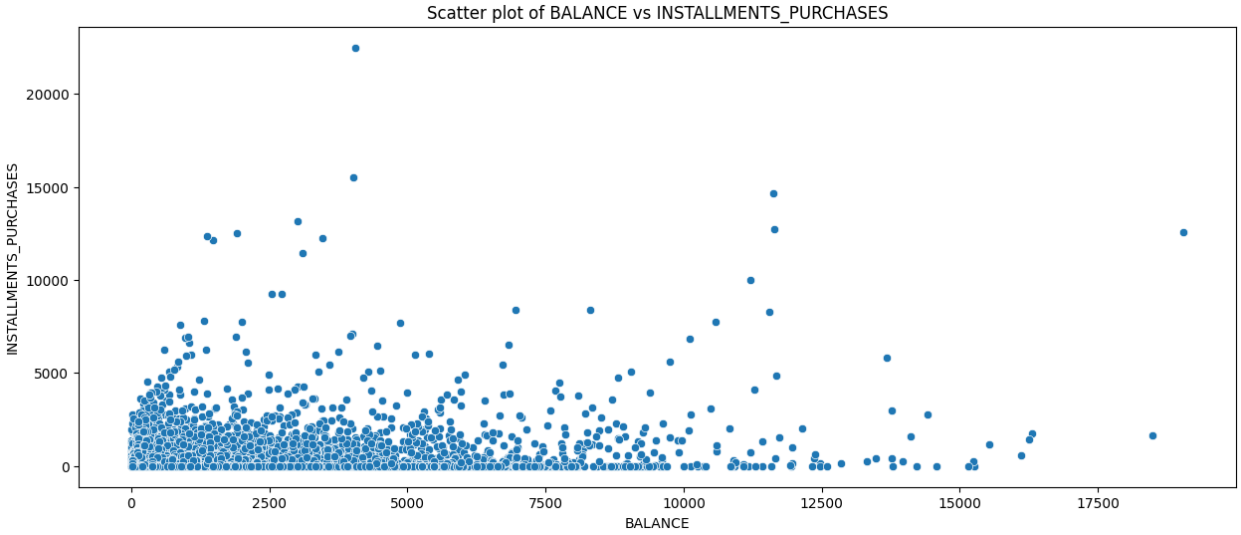


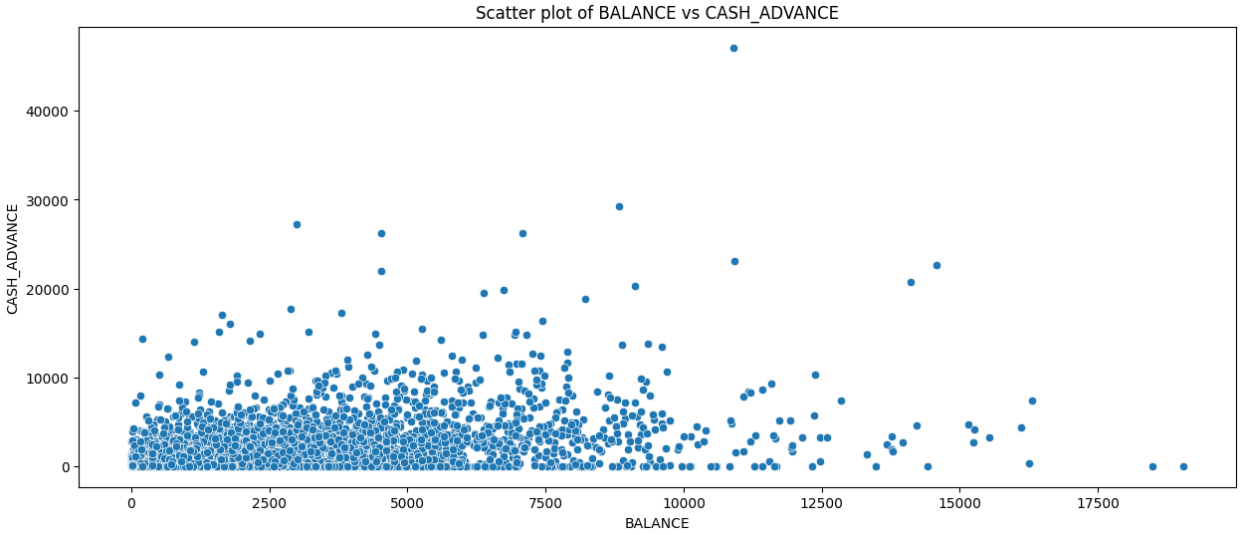

```
for i in range(len(continuous)):
    for j in range(i + 1, len(continuous)):
        plt.figure(figsize=(15, 6))
        sns.scatterplot(x=continuous[i], y=continuous[j], data=df,
palette='hls')
        plt.title(f'Scatter plot of {continuous[i]} vs
{continuous[j]}')
        plt.show()
```

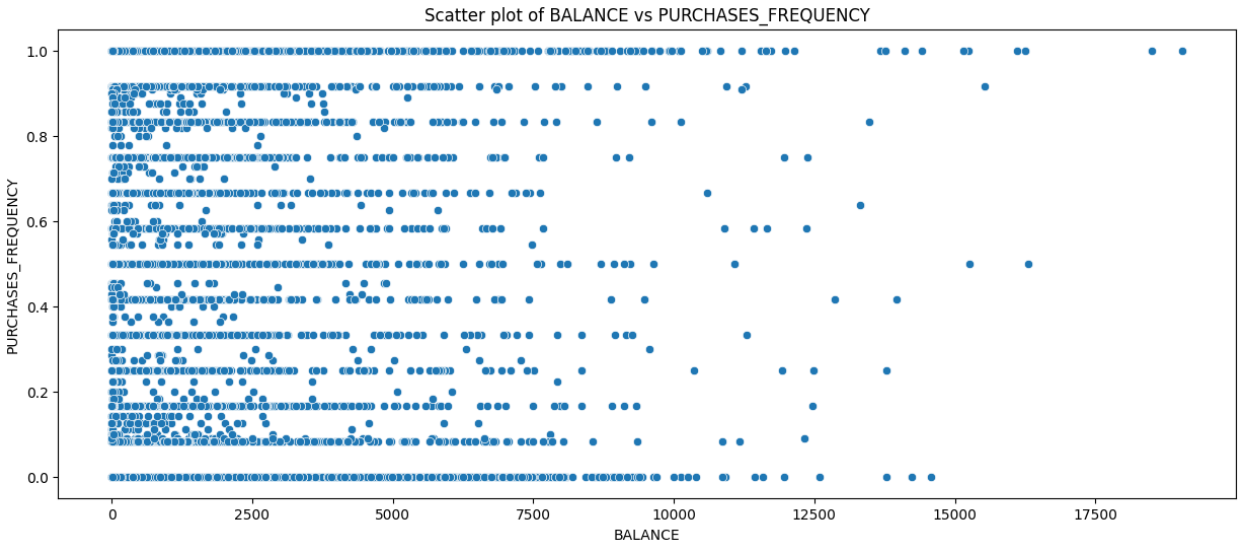

Scatter plot of BALANCE vs BALANCE_FREQUENCY
BALANCE_FREQUENCY
BALANCE
0.0
0.2
0.4
0.6
0.8
1.0
0
2500
5000
7500
10000
12500
15000
17500

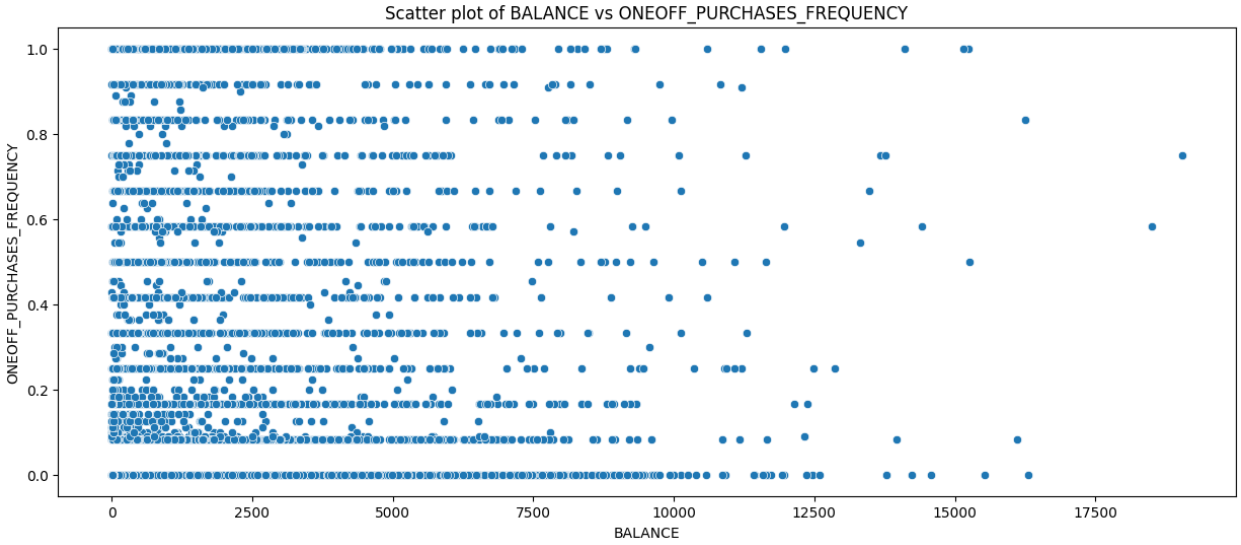

Scatter plot of BALANCE vs PURCHASES
PURCHASES
BALANCE
0
10000
20000
30000
40000
50000
0
2500
5000
7500
10000
12500
15000
17500

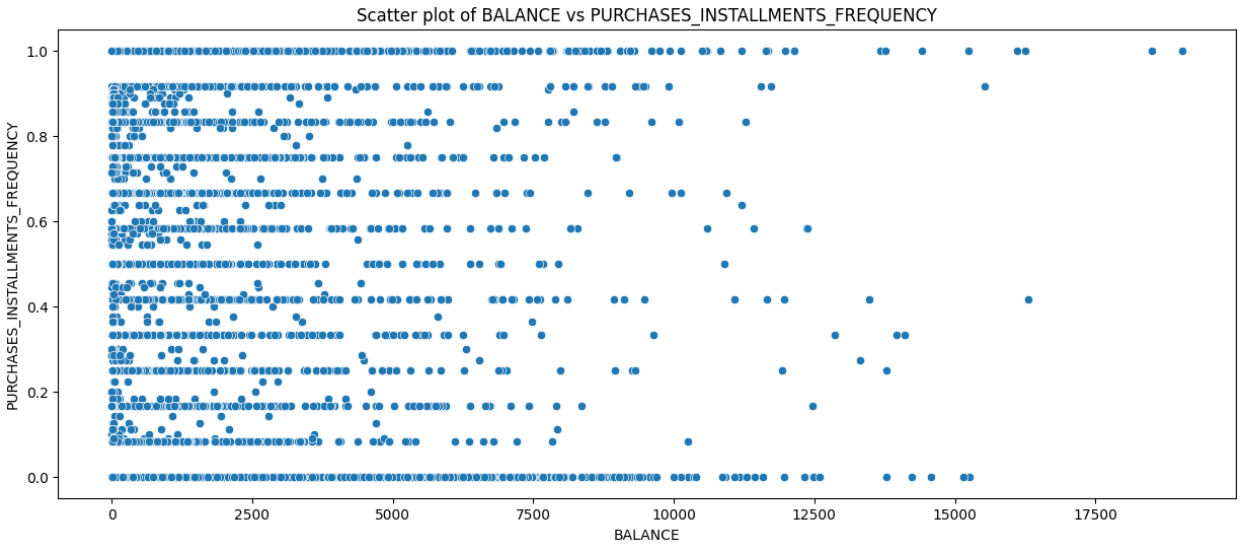

Scatter plot of BALANCE vs ONEOFF_PURCHASES
ONEOFF_PURCHASES
BALANCE
0
5000
10000
15000
20000
25000
30000
35000
40000
0
2500
5000
7500
10000
12500
15000
17500

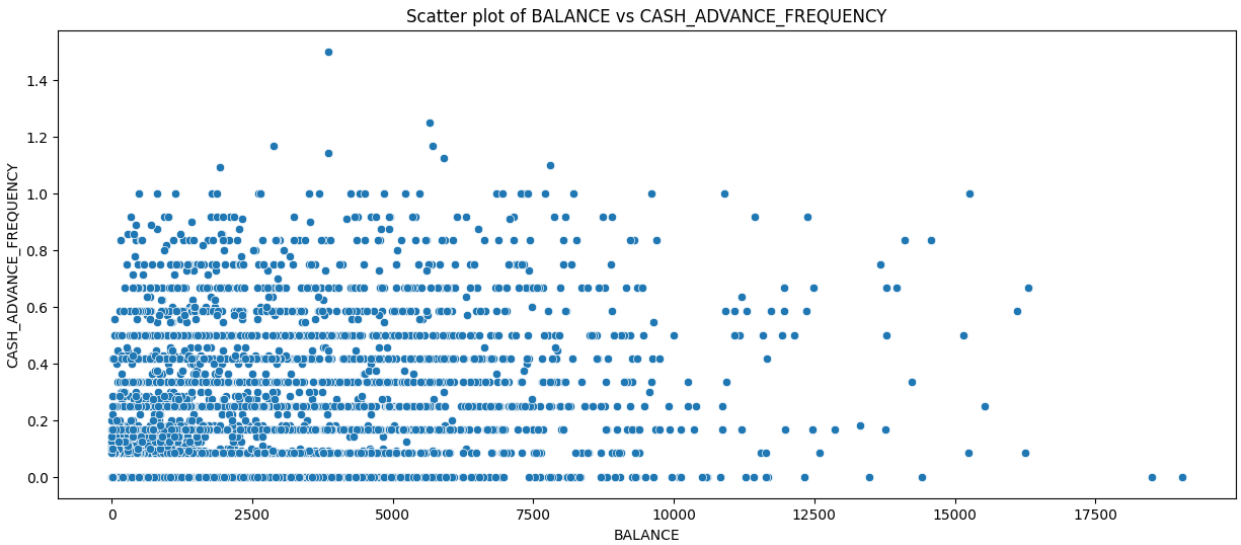

Scatter plot of BALANCE vs INSTALLMENTS_PURCHASES
INSTALLMENTS_PURCHASES
BALANCE
0
5000
10000
15000
20000
0
2500
5000
7500
10000
12500
15000
17500

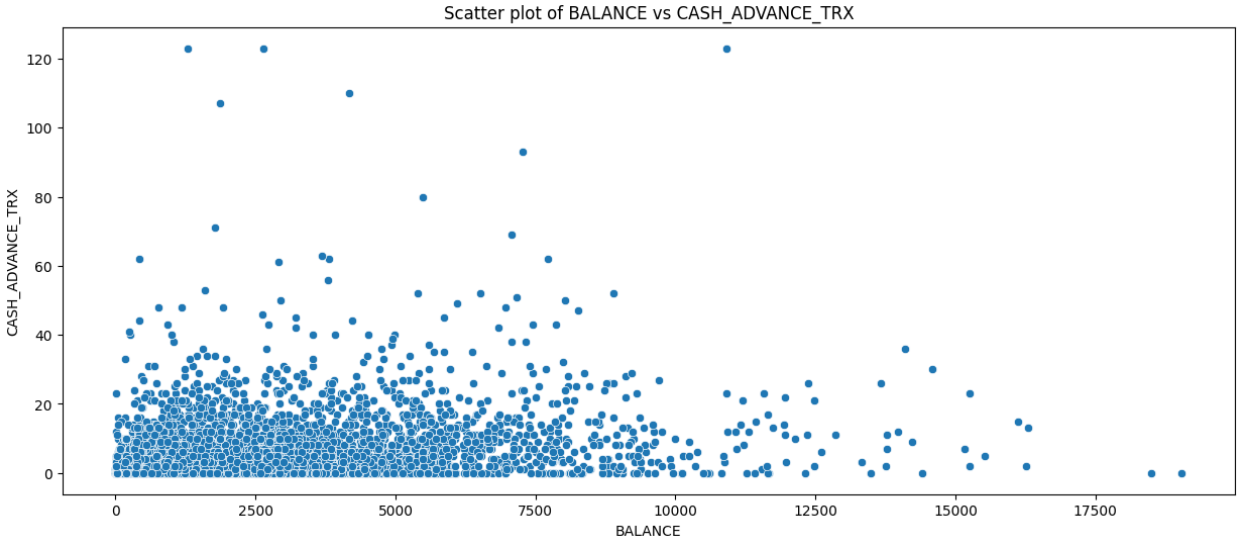

Scatter plot of BALANCE vs CASH_ADVANCE
CASH_ADVANCE
BALANCE
0
10000
20000
30000
40000
0
2500
5000
7500
10000
12500
15000
17500

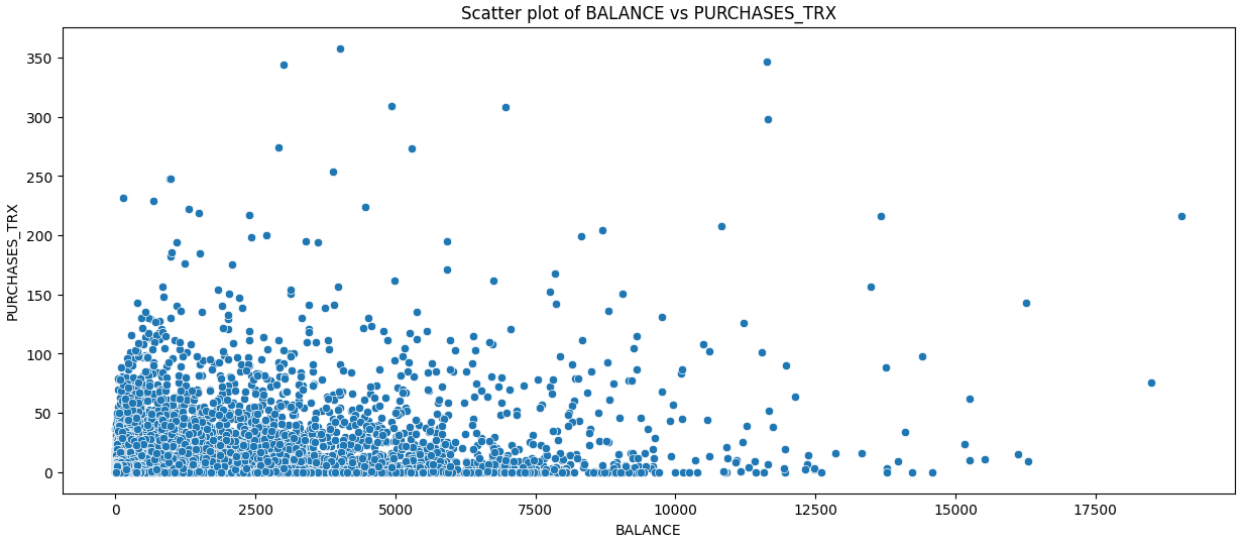

Scatter plot of BALANCE vs PURCHASES_FREQUENCY
PURCHASES_FREQUENCY
BALANCE
0.0
0.2
0.4
0.6
0.8
1.0
0
2500
5000
7500
10000
12500
15000
17500

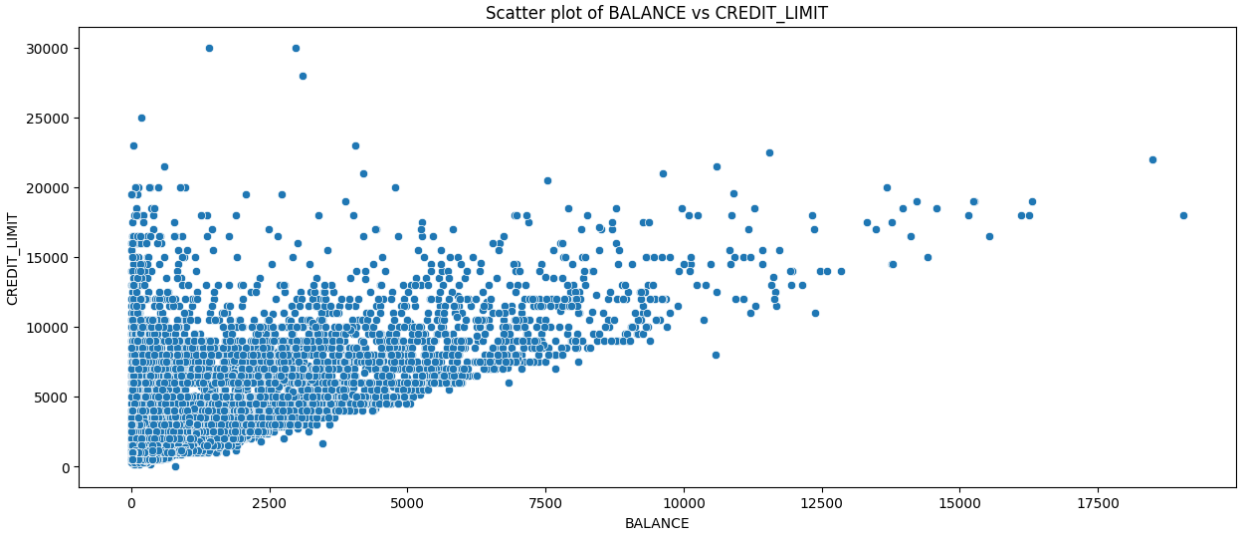

Scatter plot of BALANCE vs ONEOFF_PURCHASES_FREQUENCY
ONEOFF_PURCHASES_FREQUENCY
BALANCE
0.0
0.2
0.4
0.6
0.8
1.0
0
2500
5000
7500
10000
12500
15000
17500

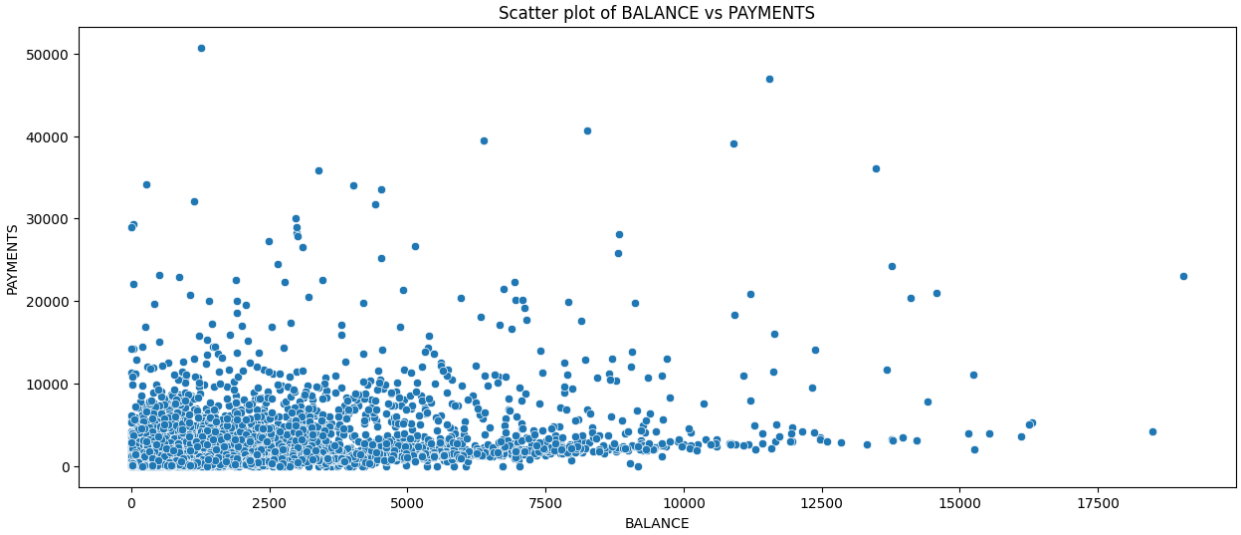

Scatter plot of BALANCE vs PURCHASES_INSTALLMENTS_FREQUENCY
PURCHASES_INSTALLMENTS_FREQUENCY
BALANCE
0.0
0.2
0.4
0.6
0.8
1.0
0
2500
5000
7500
10000
12500
15000
17500

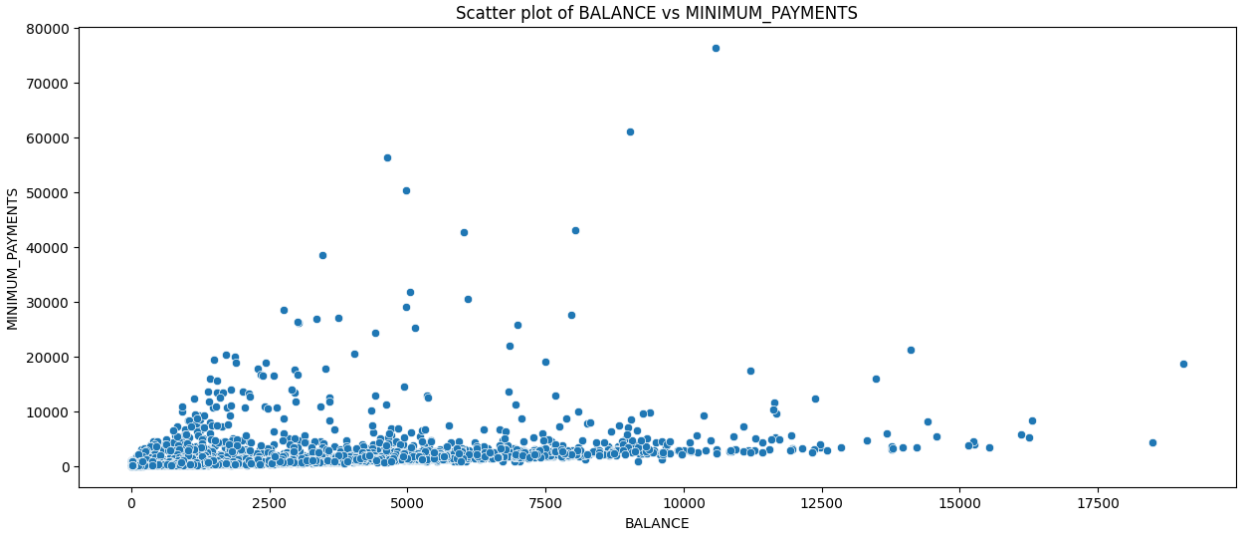

Scatter plot of BALANCE vs CASH_ADVANCE_FREQUENCY
CASH_ADVANCE_FREQUENCY
BALANCE
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0
2500
5000
7500
10000
12500
15000
17500

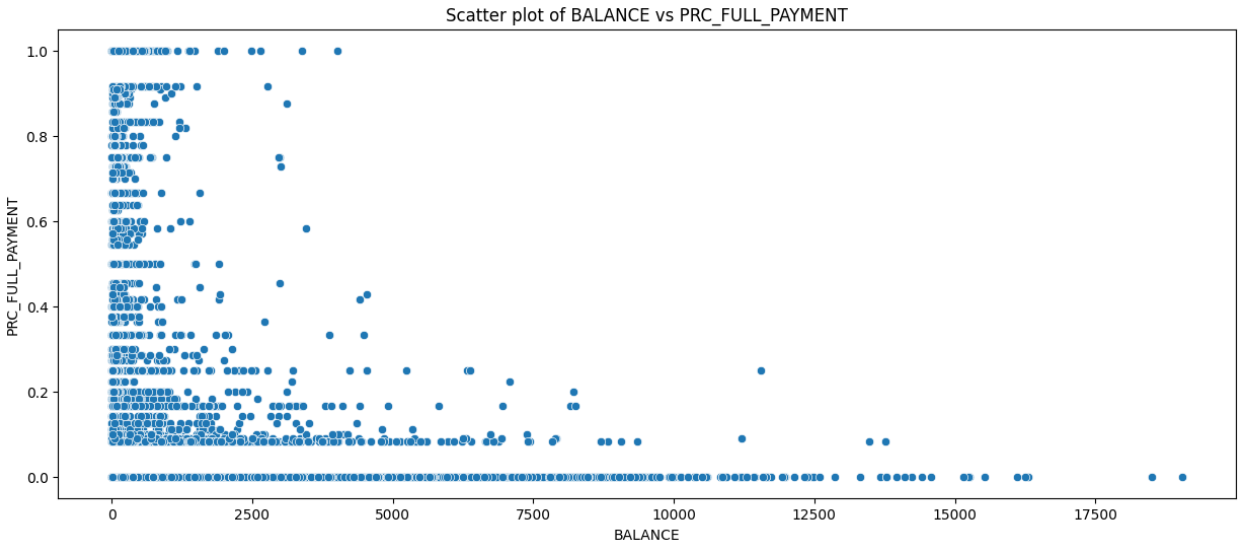

Scatter plot of BALANCE vs CASH_ADVANCE_TRX
CASH_ADVANCE_TRX
BALANCE
0
20
40
60
80
100
120
0
2500
5000
7500
10000
12500
15000
17500

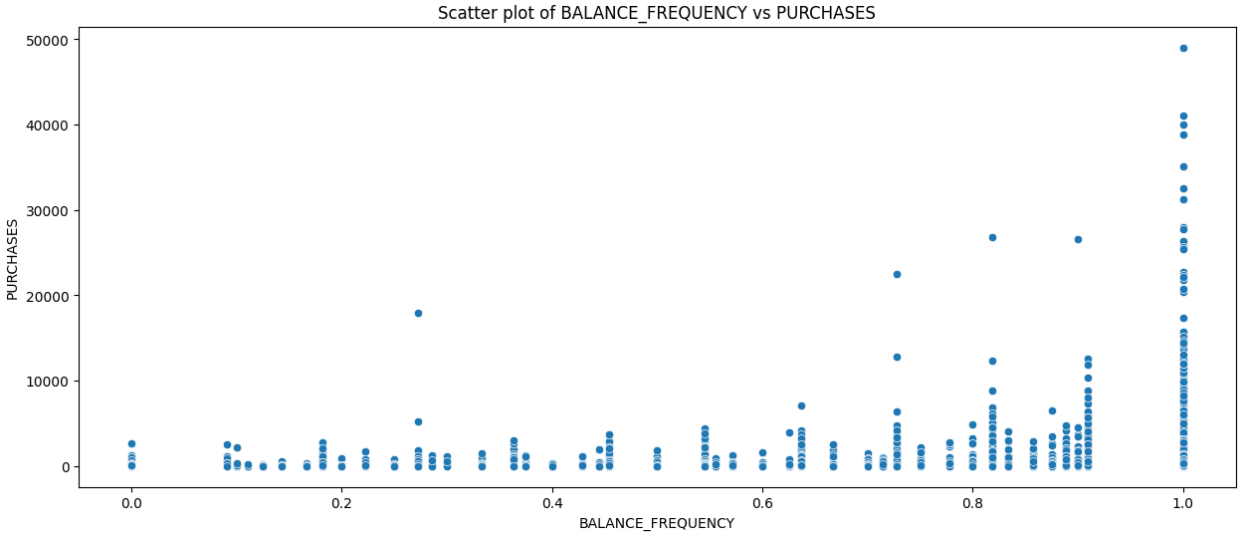

Scatter plot of BALANCE vs PURCHASES_TRX
PURCHASES_TRX
BALANCE
0
50
100
150
200
250
300
350
0
2500
5000
7500
10000
12500
15000
17500

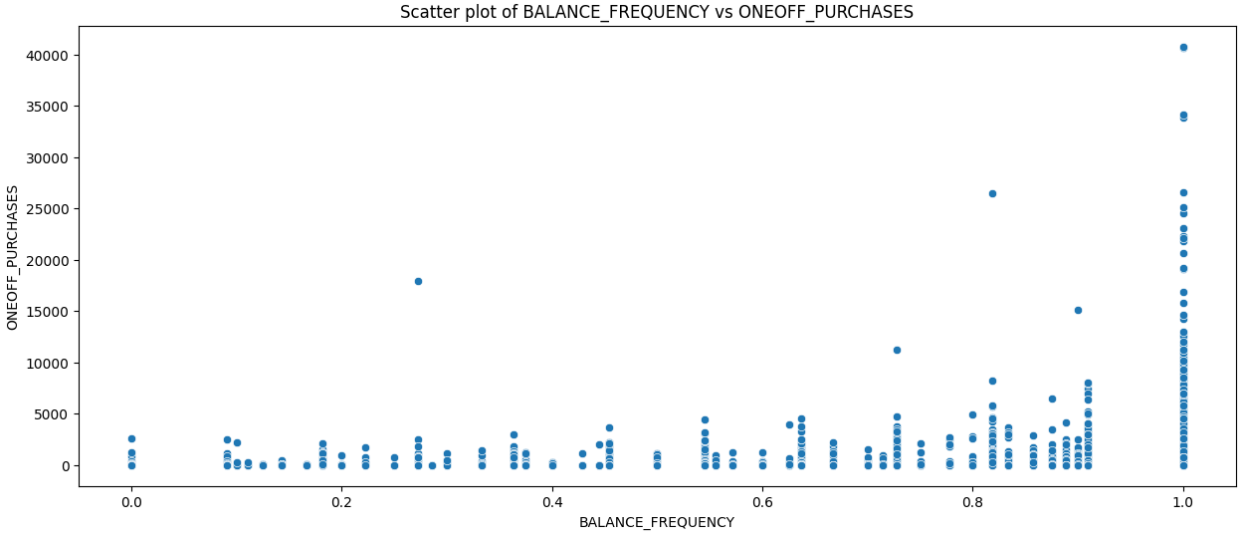

Scatter plot of BALANCE vs CREDIT_LIMIT
CREDIT_LIMIT
BALANCE
0
5000
10000
15000
20000
25000
30000
0
2500
5000
7500
10000
12500
15000
17500

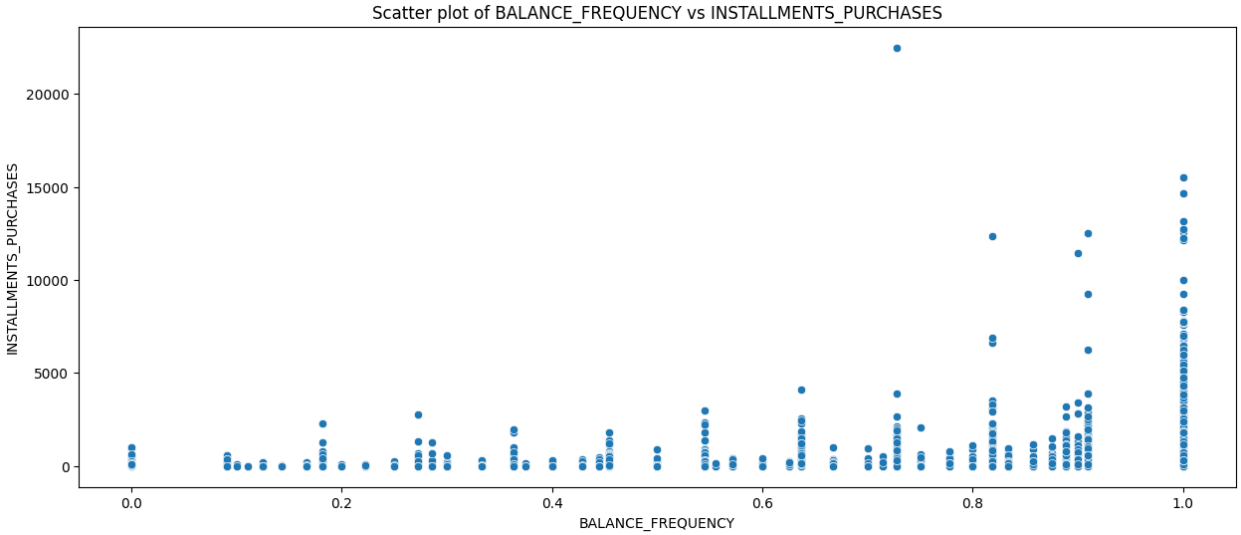

Scatter plot of BALANCE vs PAYMENTS
PAYMENTS
BALANCE
0
10000
20000
30000
40000
50000
0
2500
5000
7500
10000
12500
15000
17500

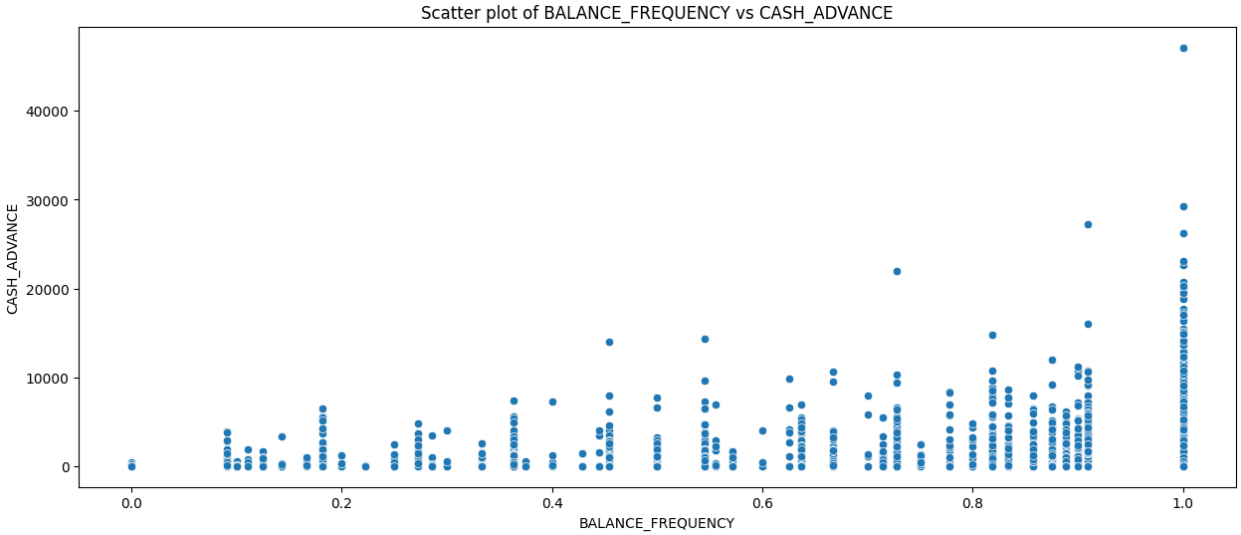

Scatter plot of BALANCE vs MINIMUM_PAYMENTS
MINIMUM_PAYMENTS
BALANCE
0
10000
20000
30000
40000
50000
60000
70000
80000
0
2500
5000
7500
10000
12500
15000
17500

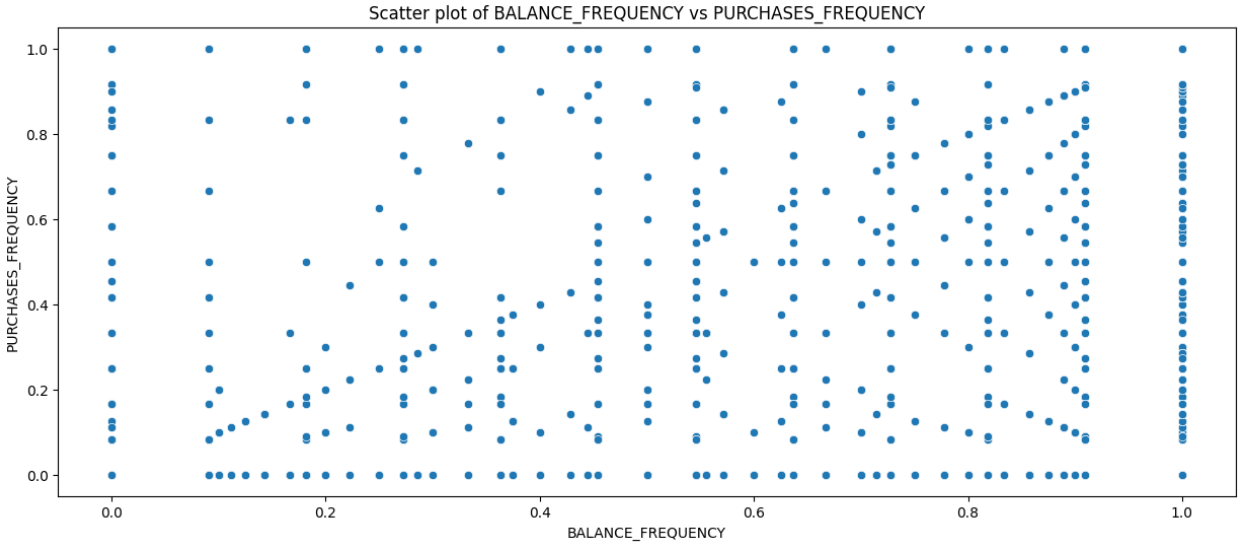

Scatter plot of BALANCE vs PRC_FULL_PAYMENT
PRC_FULL_PAYMENT
BALANCE
0.0
0.2
0.4
0.6
0.8
1.0
0
2500
5000
7500
10000
12500
15000
17500

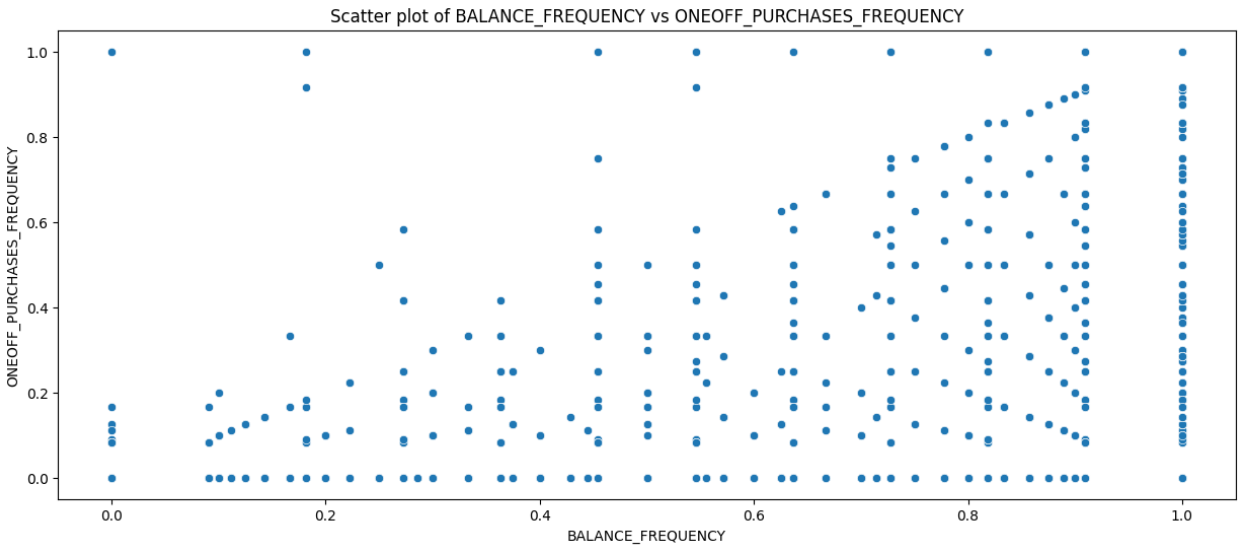

Scatter plot of BALANCE_FREQUENCY vs PURCHASES
PURCHASES
BALANCE_FREQUENCY
0
10000
20000
30000
40000
50000
0.0
0.2
0.4
0.6
0.8
1.0

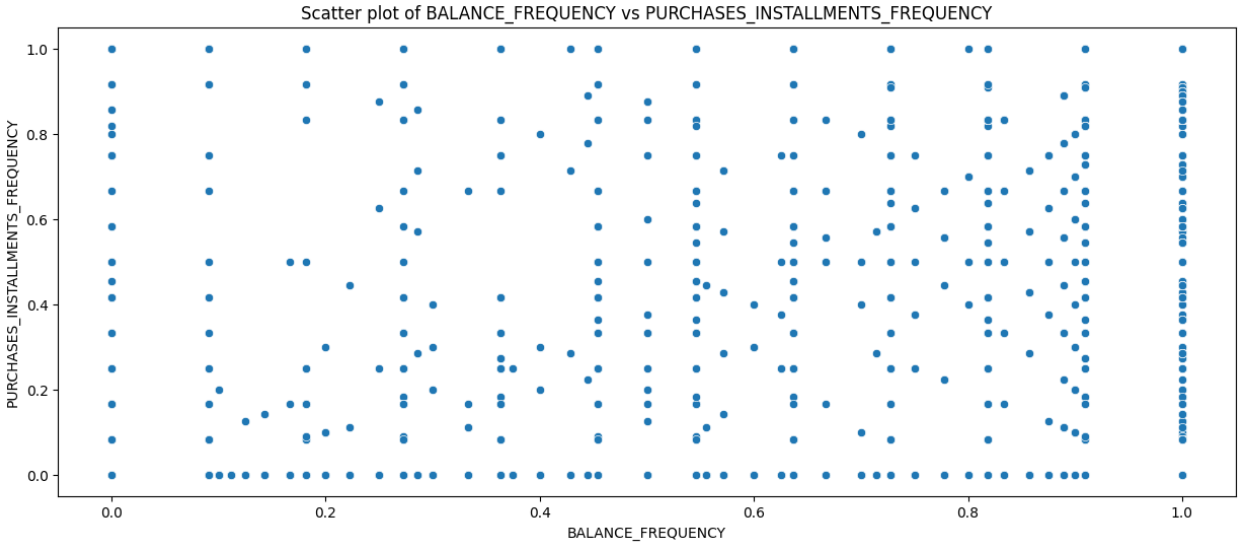

Scatter plot of BALANCE_FREQUENCY vs ONEOFF_PURCHASES
ONEOFF_PURCHASES
BALANCE_FREQUENCY
0
5000
10000
15000
20000
25000
30000
35000
40000
0.0
0.2
0.4
0.6
0.8
1.0

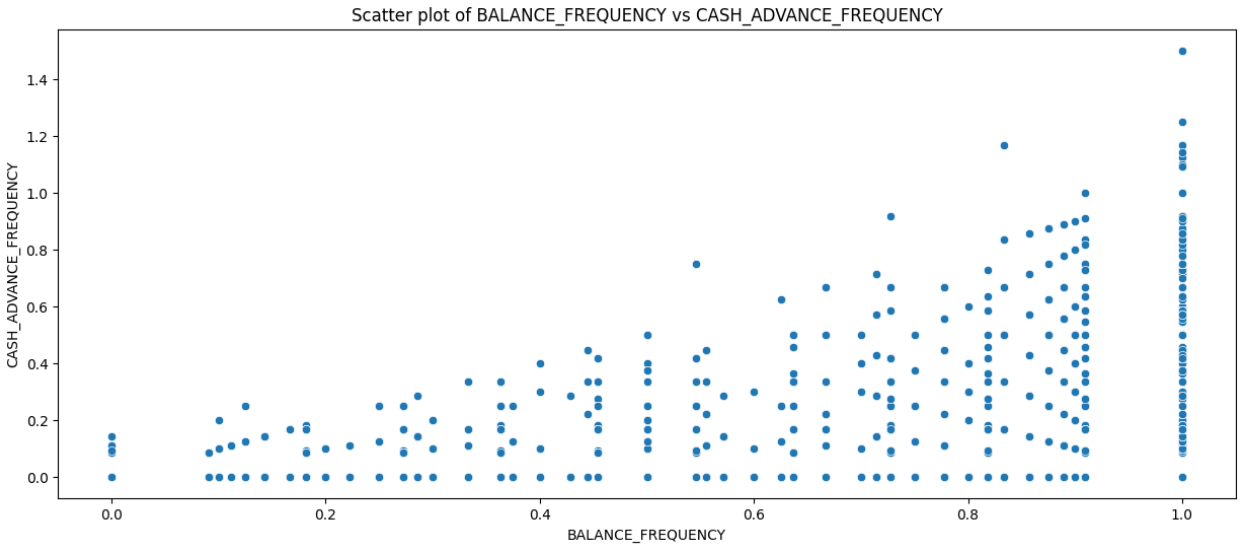

Scatter plot of BALANCE_FREQUENCY vs INSTALLMENTS_PURCHASES
INSTALLMENTS_PURCHASES
BALANCE_FREQUENCY
0
5000
10000
15000
20000
0.0
0.2
0.4
0.6
0.8
1.0

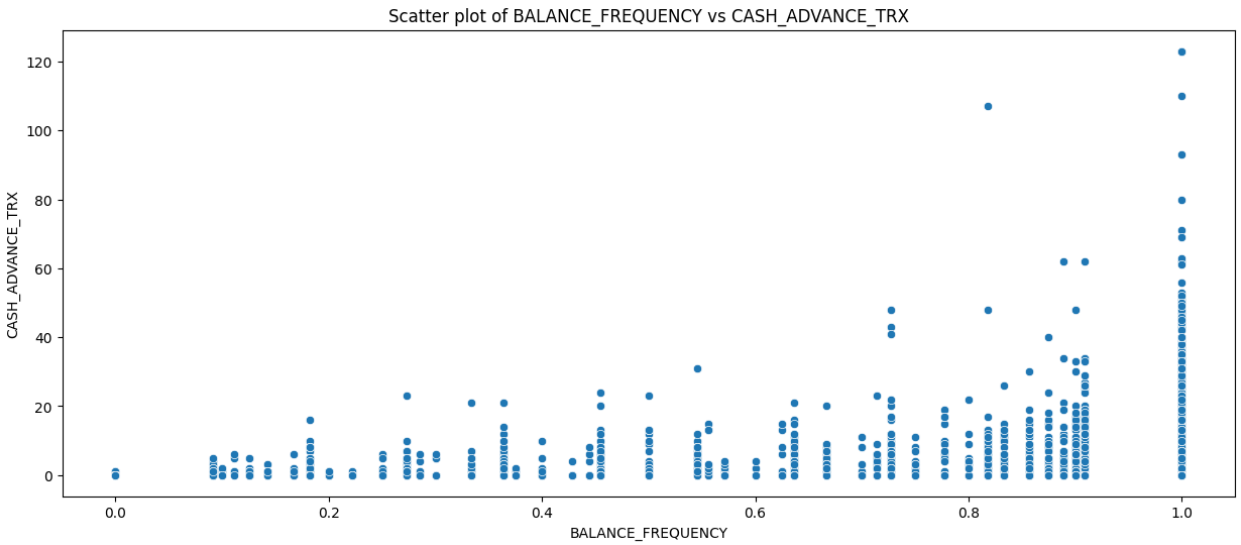

Scatter plot of BALANCE_FREQUENCY vs CASH_ADVANCE
CASH_ADVANCE
BALANCE_FREQUENCY
0
10000
20000
30000
40000
0.0
0.2
0.4
0.6
0.8
1.0

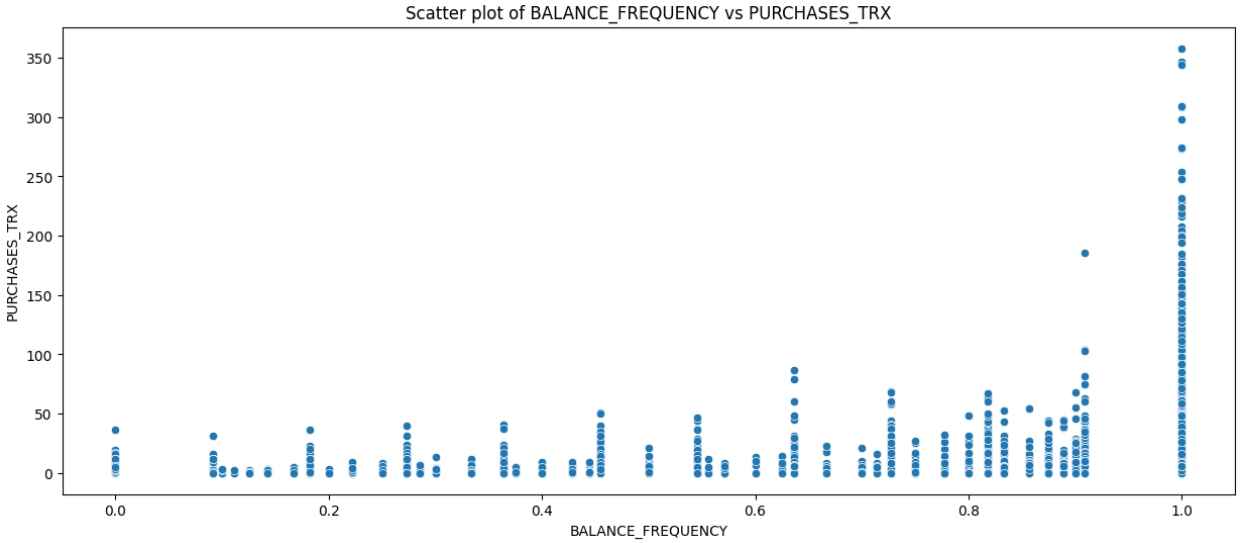

Scatter plot of BALANCE_FREQUENCY vs PURCHASES_FREQUENCY
PURCHASES_FREQUENCY
BALANCE_FREQUENCY
0.0
0.2
0.4
0.6
0.8
1.0

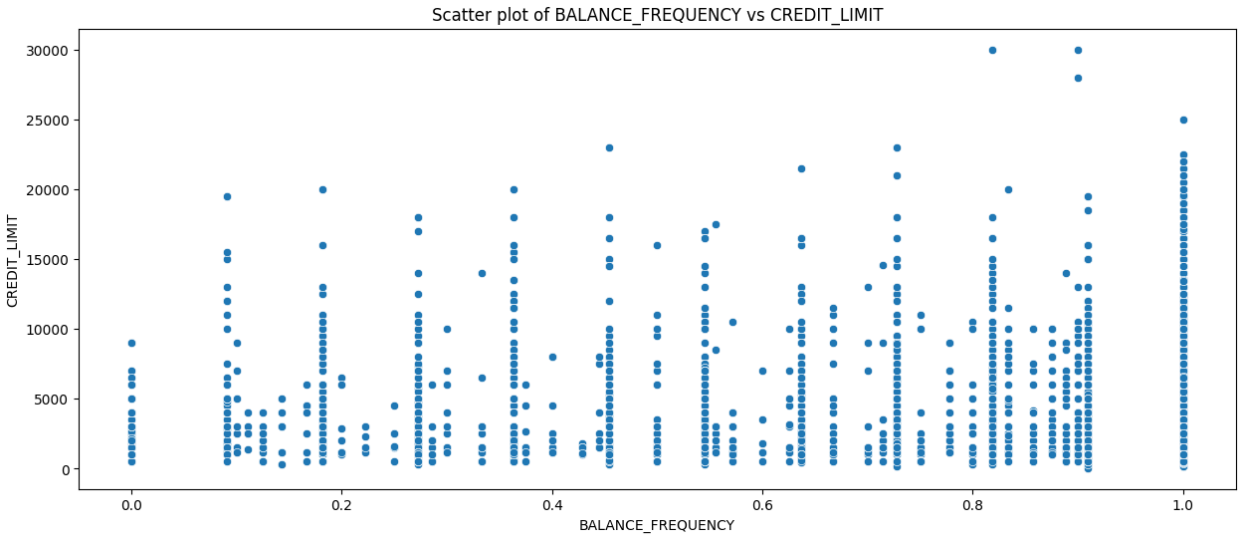

Scatter plot of BALANCE_FREQUENCY vs ONEOFF_PURCHASES_FREQUENCY
ONEOFF_PURCHASES_FREQUENCY
BALANCE_FREQUENCY
0.0
0.2
0.4
0.6
0.8
1.0

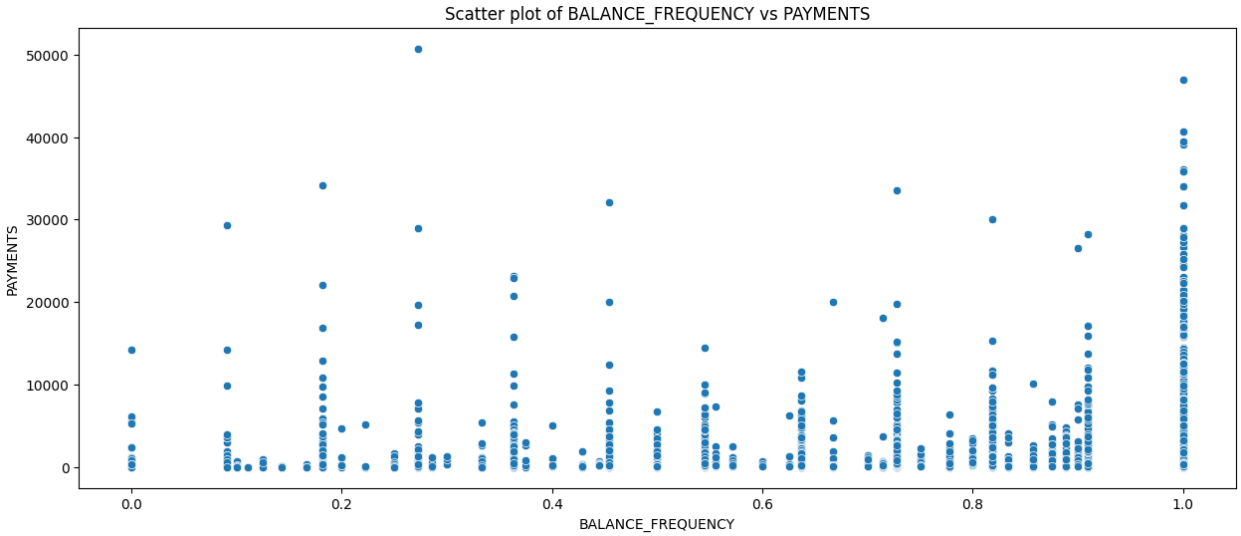

Scatter plot of BALANCE_FREQUENCY vs PURCHASES_INSTALLMENTS_FREQUENCY
PURCHASES_INSTALLMENTS_FREQUENCY
BALANCE_FREQUENCY
0.0
0.2
0.4
0.6
0.8
1.0

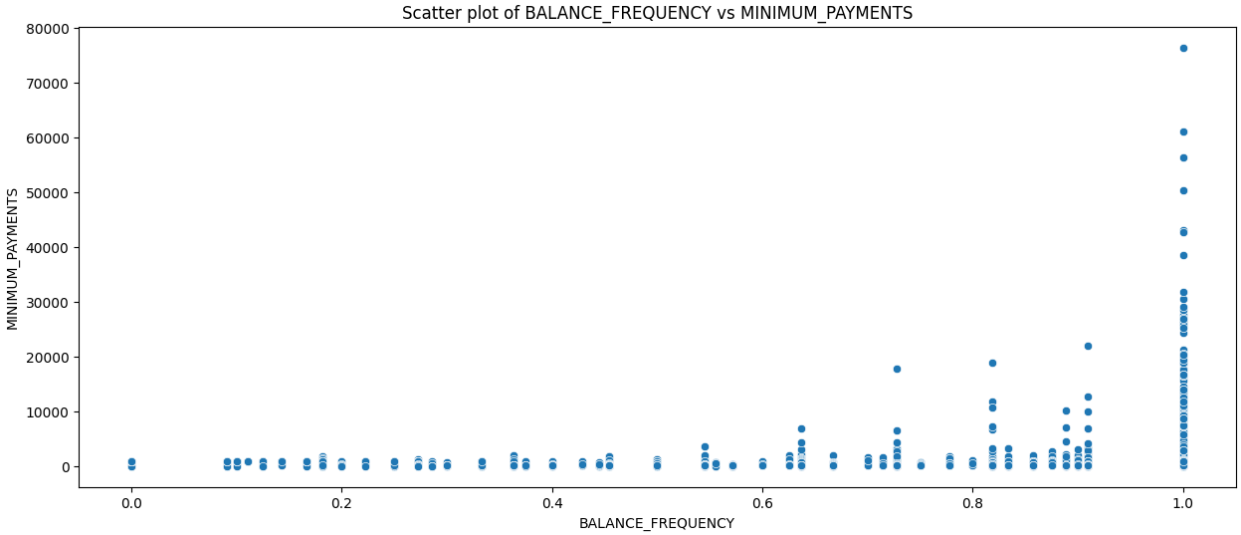

Scatter plot of BALANCE_FREQUENCY vs CASH_ADVANCE_FREQUENCY
CASH_ADVANCE_FREQUENCY
BALANCE_FREQUENCY
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4

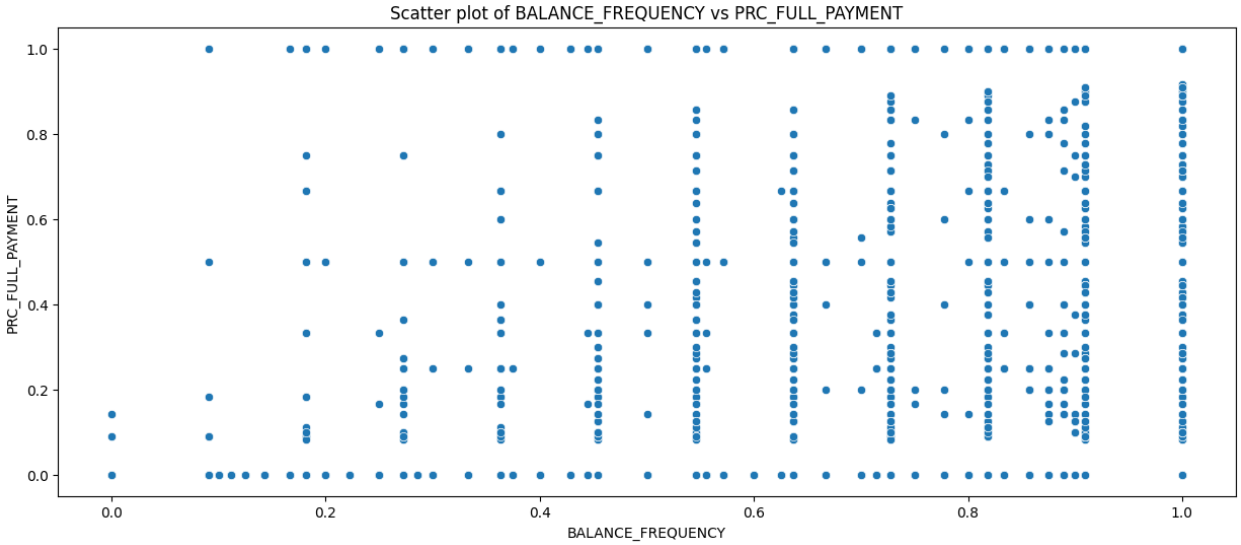

Scatter plot of BALANCE_FREQUENCY vs CASH_ADVANCE_TRX
CASH_ADVANCE_TRX
BALANCE_FREQUENCY
0
20
40
60
80
100
120

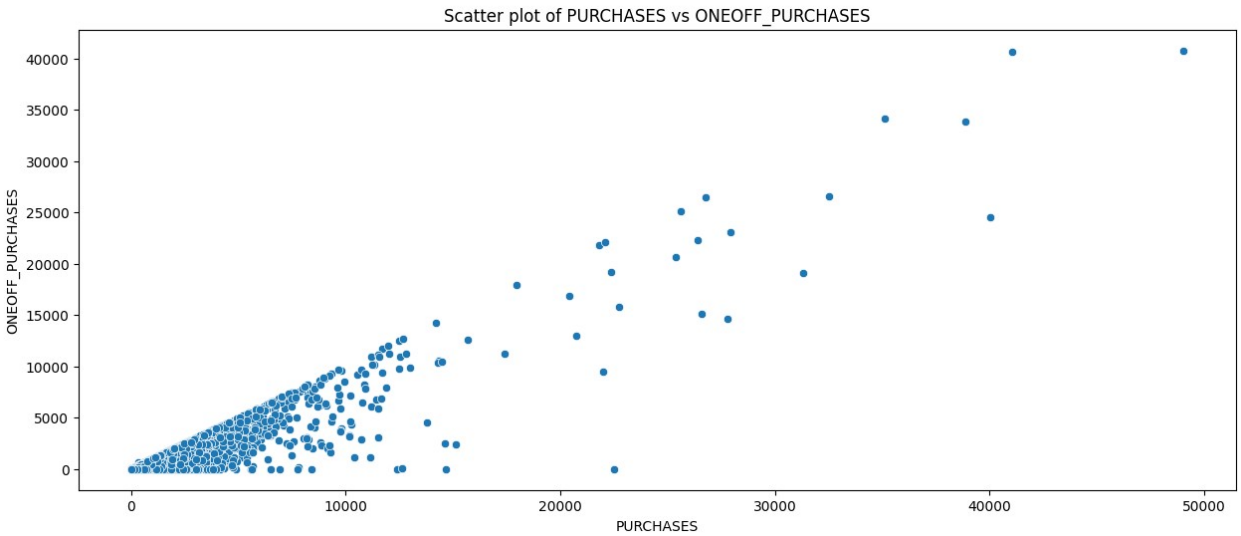

Scatter plot of BALANCE_FREQUENCY vs PURCHASES_TRX
PURCHASES_TRX
BALANCE_FREQUENCY
0
50
100
150
200
250
300
350
0.0
0.2
0.4
0.6
0.8
1.0

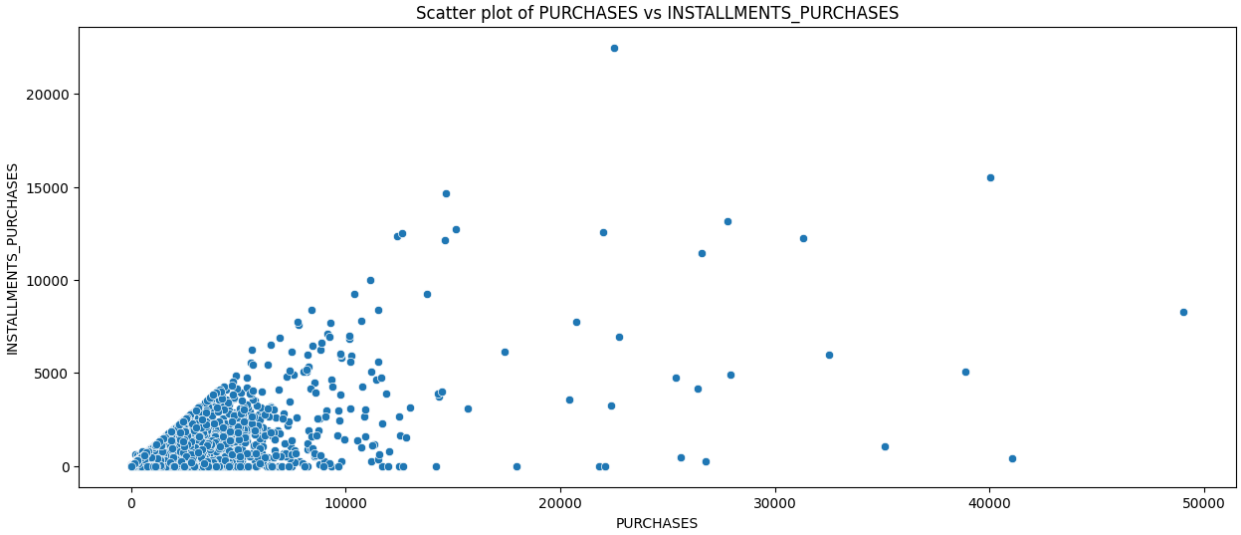

Scatter plot of BALANCE_FREQUENCY vs CREDIT_LIMIT
CREDIT_LIMIT
BALANCE_FREQUENCY
0
5000
10000
15000
20000
25000
30000
0.0
0.2
0.4
0.6
0.8
1.0

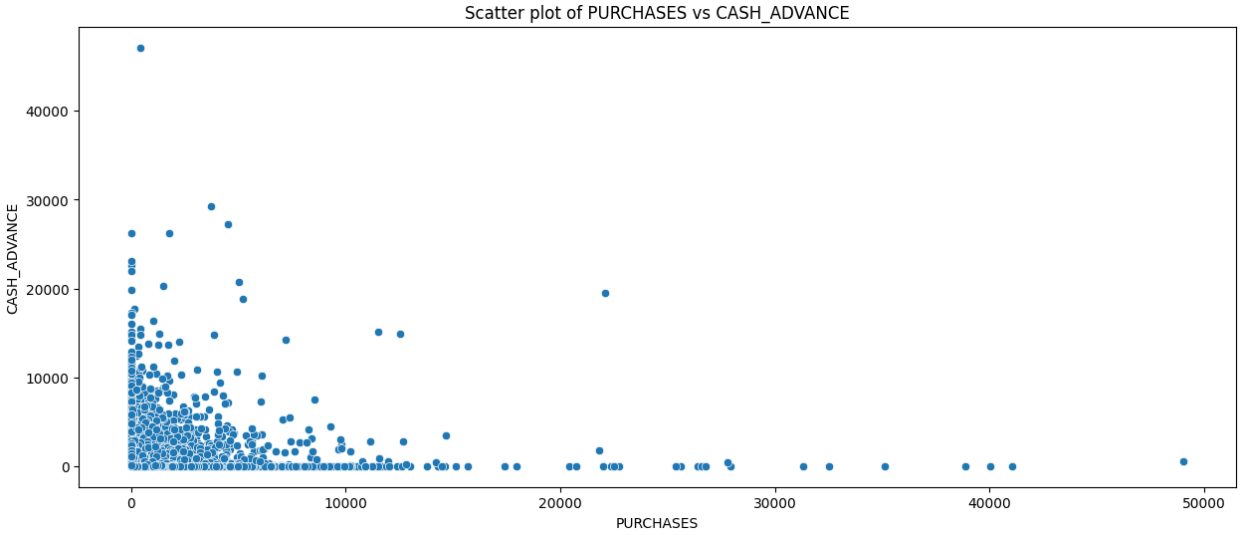

Scatter plot of BALANCE_FREQUENCY vs PAYMENTS
PAYMENTS
BALANCE_FREQUENCY
0
10000
20000
30000
40000
50000
0.0
0.2
0.4
0.6
0.8
1.0

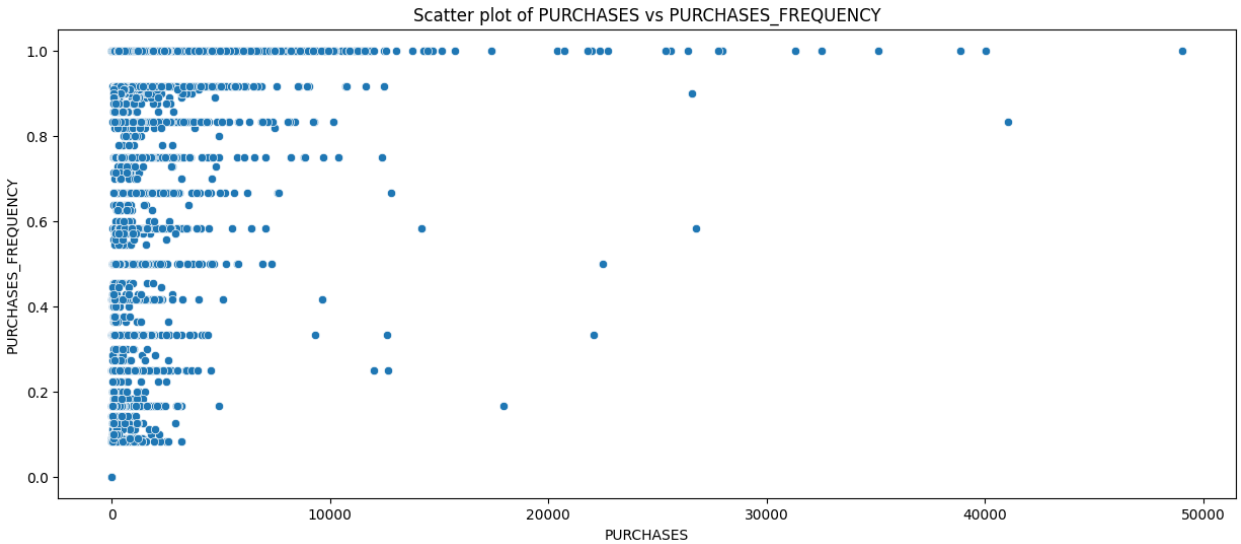

Scatter plot of BALANCE_FREQUENCY vs MINIMUM_PAYMENTS
MINIMUM_PAYMENTS
BALANCE_FREQUENCY
0
10000
20000
30000
40000
50000
60000
70000
80000
0.0
0.2
0.4
0.6
0.8
1.0

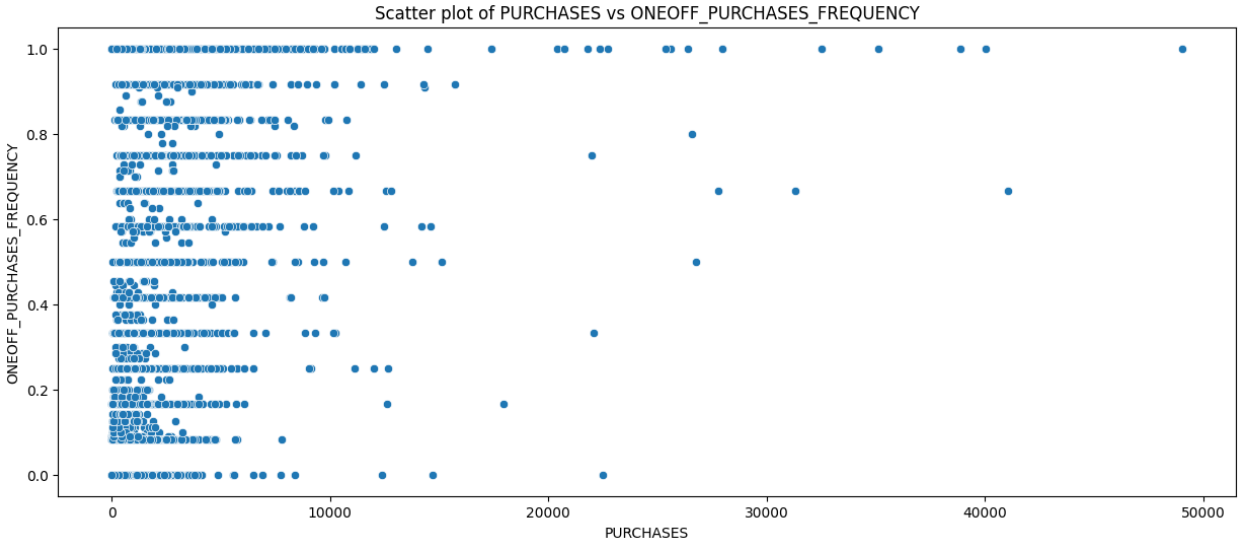

Scatter plot of BALANCE_FREQUENCY vs PRC_FULL_PAYMENT
PRC_FULL_PAYMENT
BALANCE_FREQUENCY
0.0
0.2
0.4
0.6
0.8
1.0

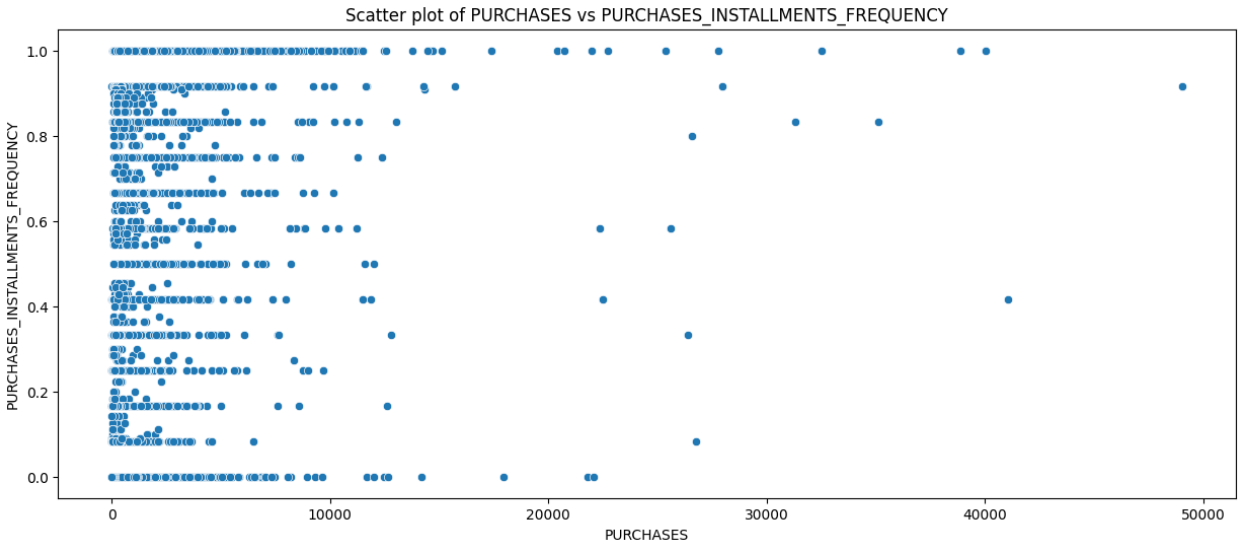

Scatter plot of PURCHASES vs ONEOFF_PURCHASES
ONEOFF_PURCHASES
PURCHASES
0
5000
10000
15000
20000
25000
30000
35000
40000
0
10000
20000
30000
40000
50000

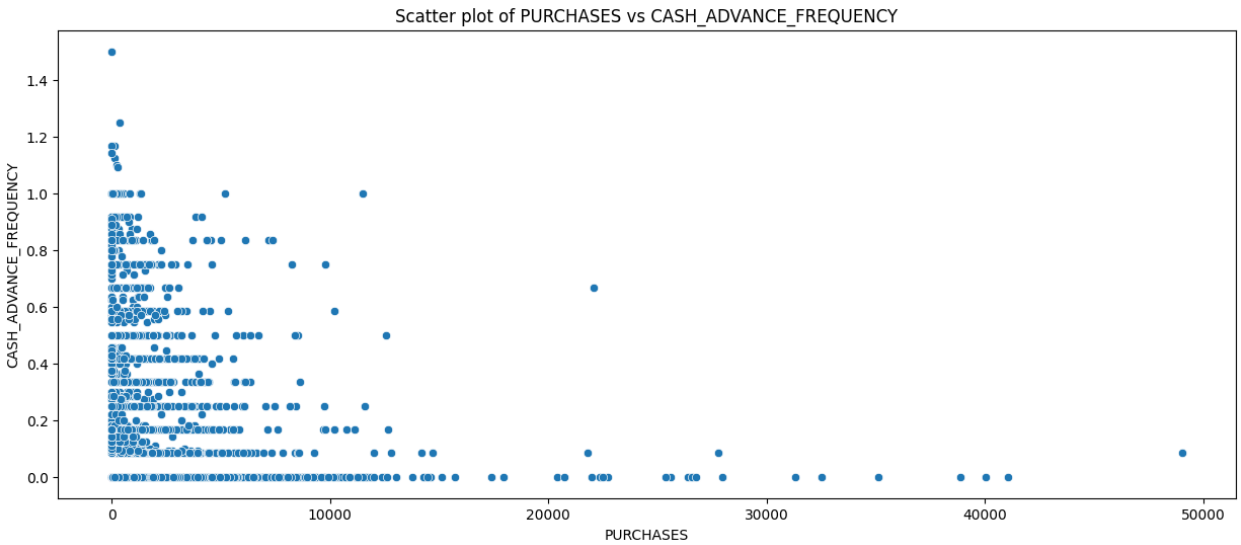

Scatter plot of PURCHASES vs INSTALLMENTS_PURCHASES
INSTALLMENTS_PURCHASES
PURCHASES
0
5000
10000
15000
20000
0
10000
20000
30000
40000
50000

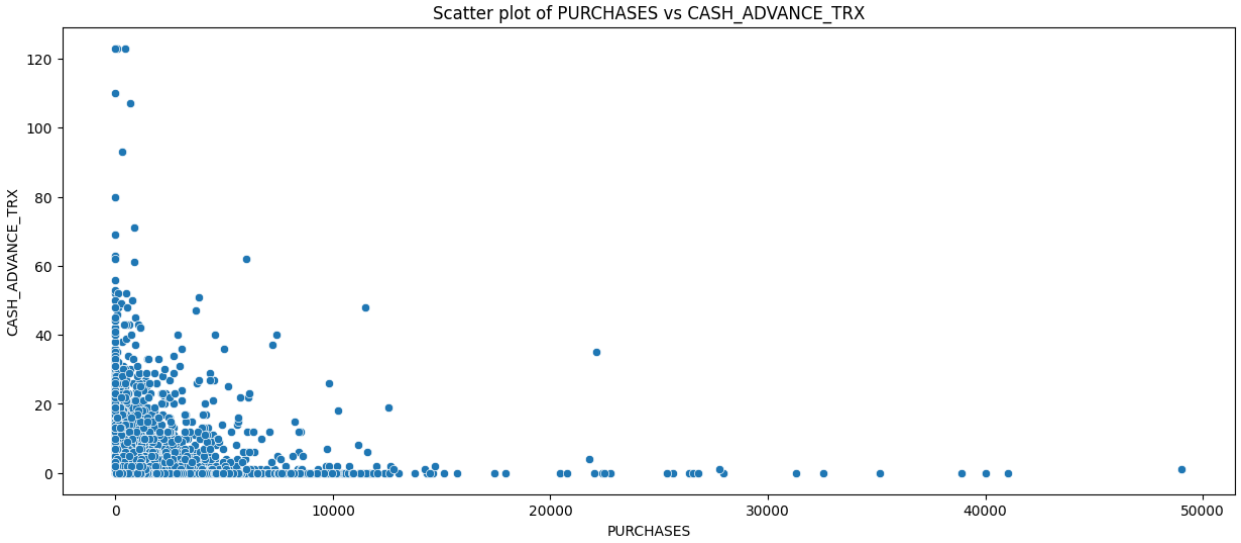

Scatter plot of PURCHASES vs CASH_ADVANCE
CASH_ADVANCE
PURCHASES
0
10000
20000
30000
40000
0
10000
20000
30000
40000
50000

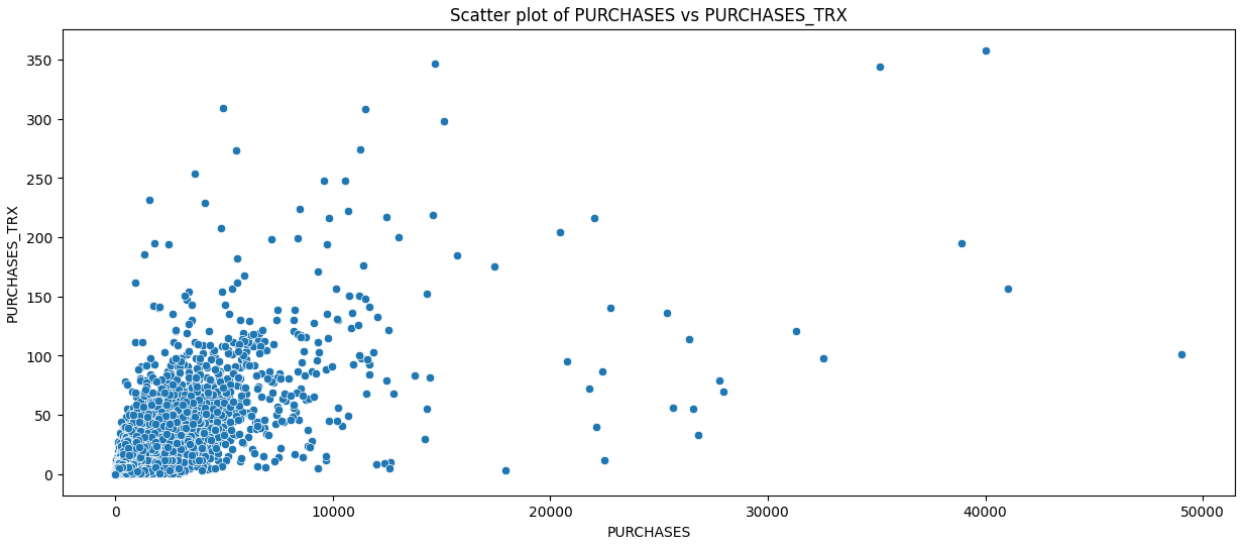

Scatter plot of PURCHASES vs PURCHASES_FREQUENCY
PURCHASES_FREQUENCY
PURCHASES
0.0
0.2
0.4
0.6
0.8
1.0
0
10000
20000
30000
40000
50000

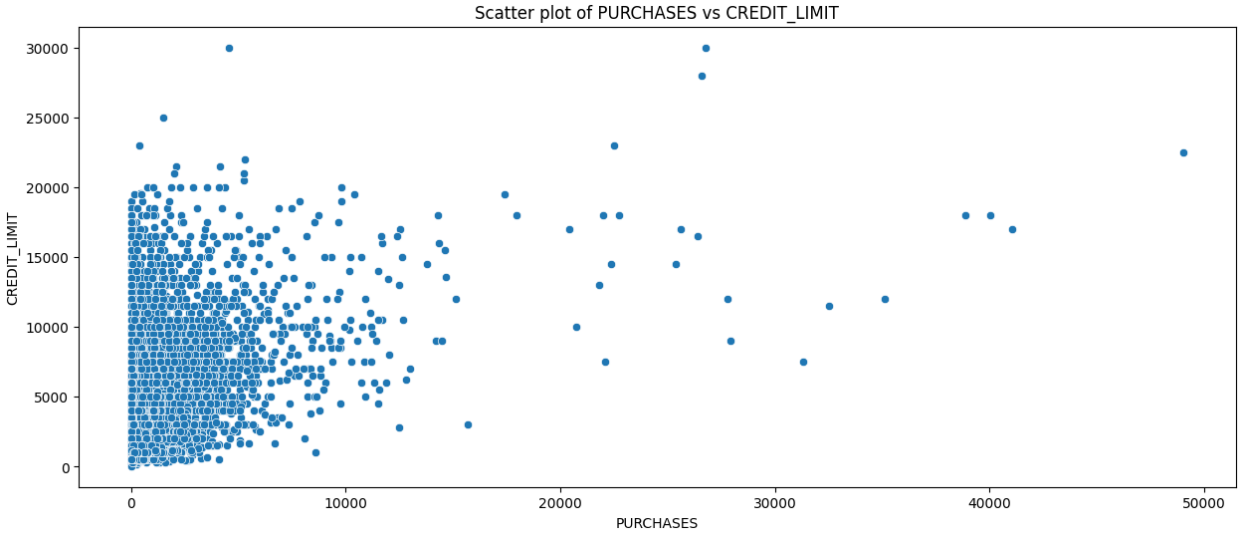

Scatter plot of PURCHASES vs ONEOFF_PURCHASES_FREQUENCY
ONEOFF_PURCHASES_FREQUENCY
PURCHASES
0.0
0.2
0.4
0.6
0.8
1.0
0
10000
20000
30000
40000
50000

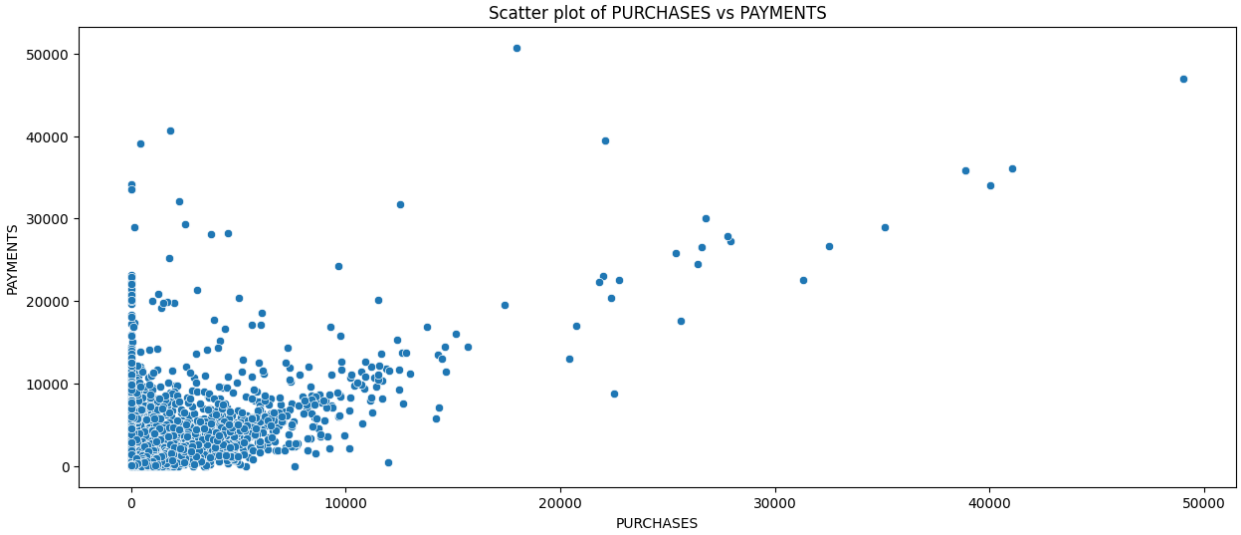

Scatter plot of PURCHASES vs PURCHASES_INSTALLMENTS_FREQUENCY
PURCHASES_INSTALLMENTS_FREQUENCY
PURCHASES
0.0
0.2
0.4
0.6
0.8
1.0
0
10000
20000
30000
40000
50000

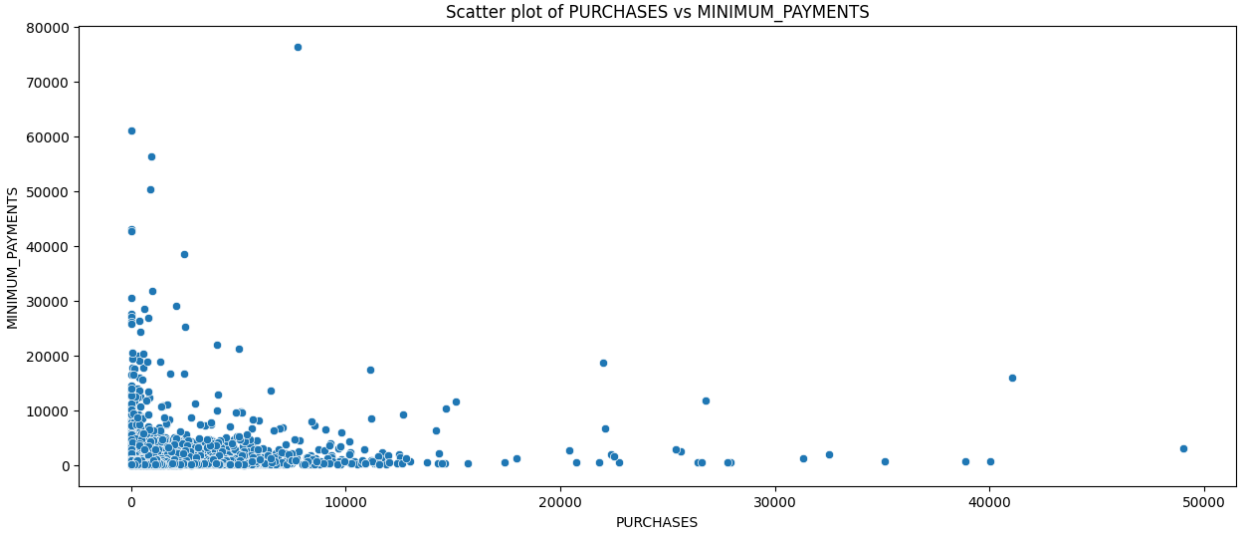

Scatter plot of PURCHASES vs CASH_ADVANCE_FREQUENCY
CASH_ADVANCE_FREQUENCY
PURCHASES
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0
10000
20000
30000
40000
50000

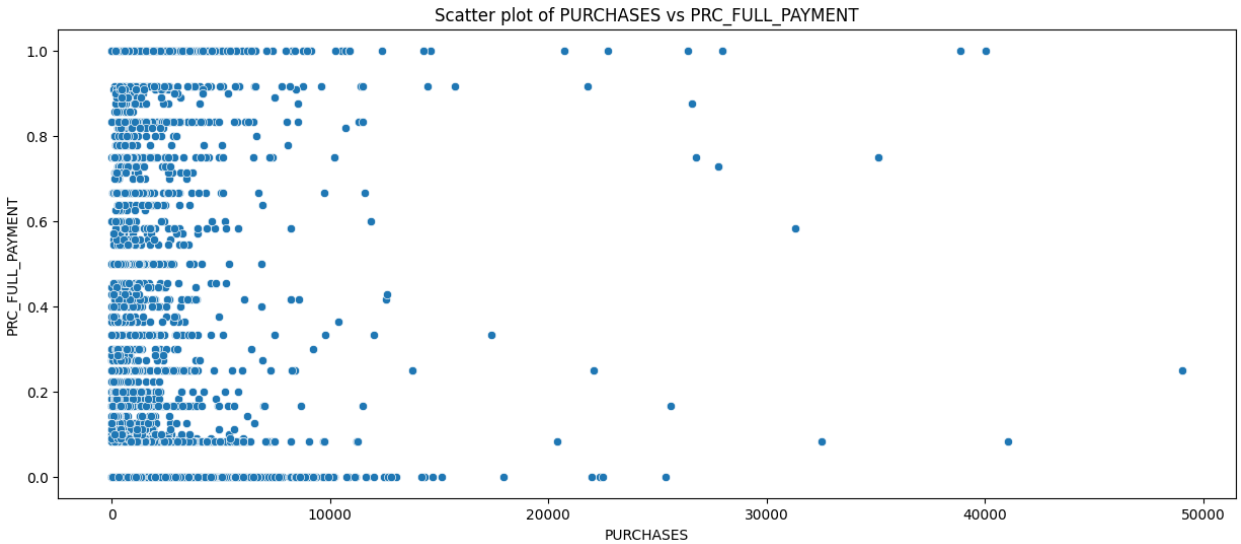

Scatter plot of PURCHASES vs CASH_ADVANCE_TRX
CASH_ADVANCE_TRX
PURCHASES
0
20
40
60
80
100
120
0
10000
20000
30000
40000
50000

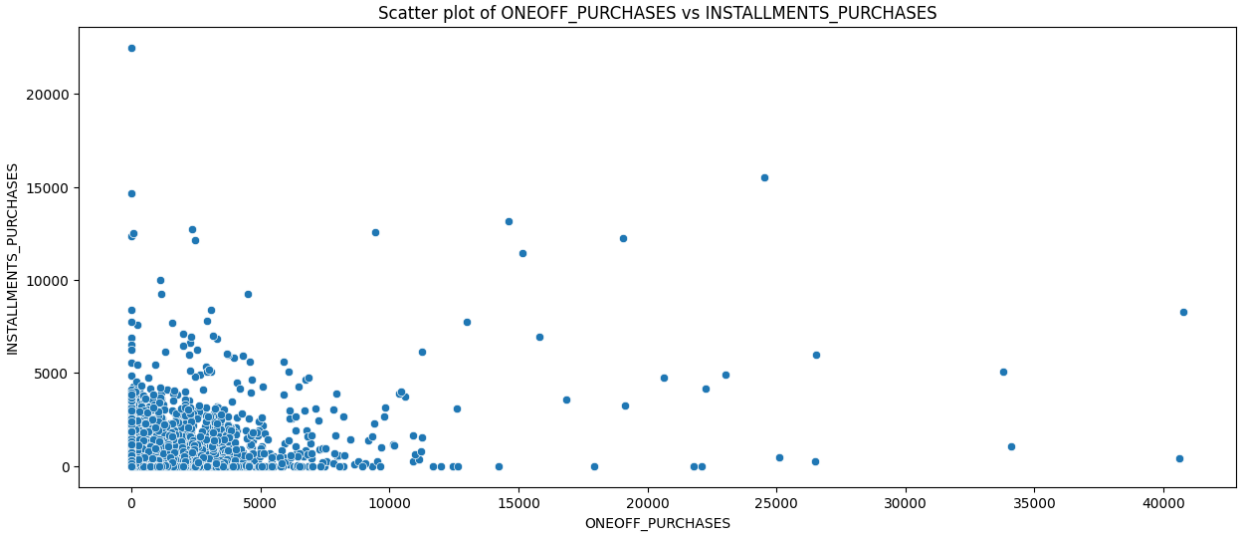

Scatter plot of PURCHASES vs PURCHASES_TRX
PURCHASES_TRX
PURCHASES
0
50
100
150
200
250
300
350
0
10000
20000
30000
40000
50000

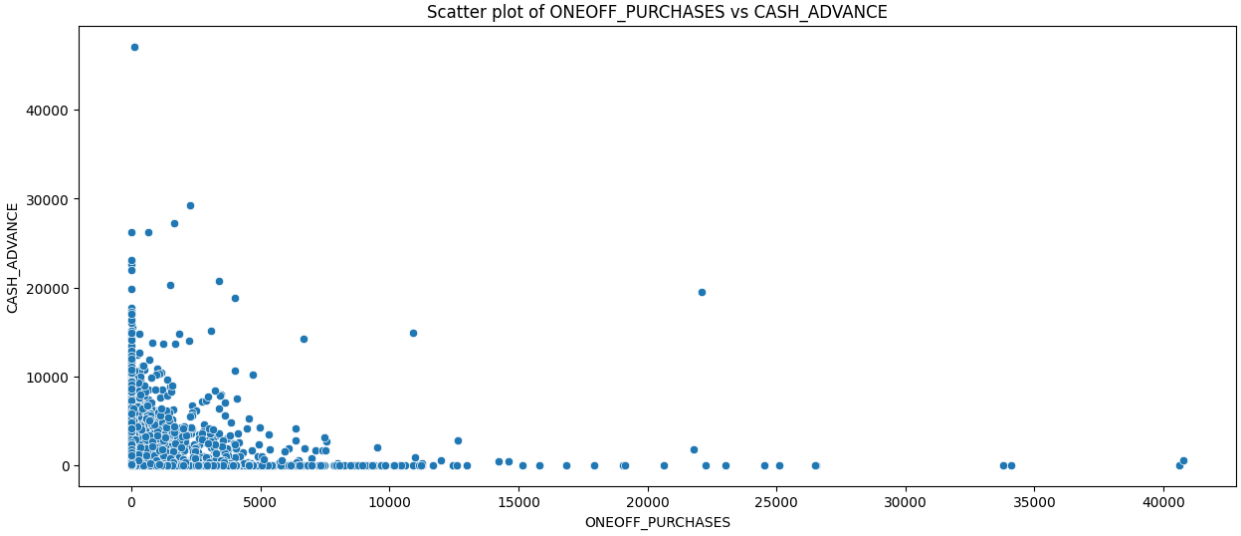

Scatter plot of PURCHASES vs CREDIT_LIMIT
CREDIT_LIMIT
PURCHASES
0
5000
10000
15000
20000
25000
30000
0
10000
20000
30000
40000
50000

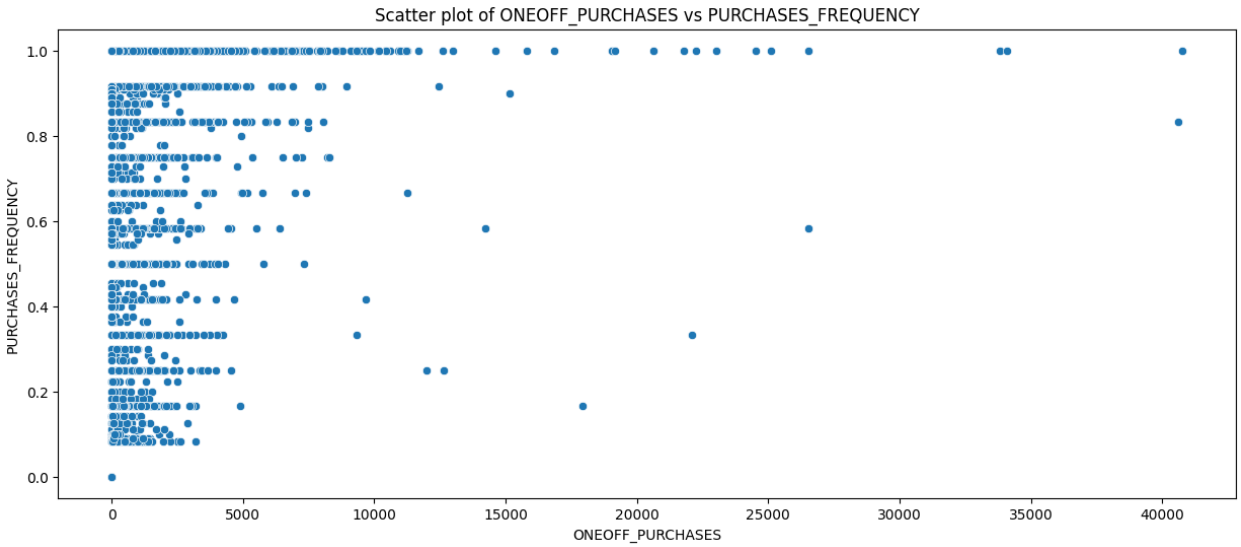

Scatter plot of PURCHASES vs PAYMENTS
PAYMENTS
PURCHASES
0
10000
20000
30000
40000
50000

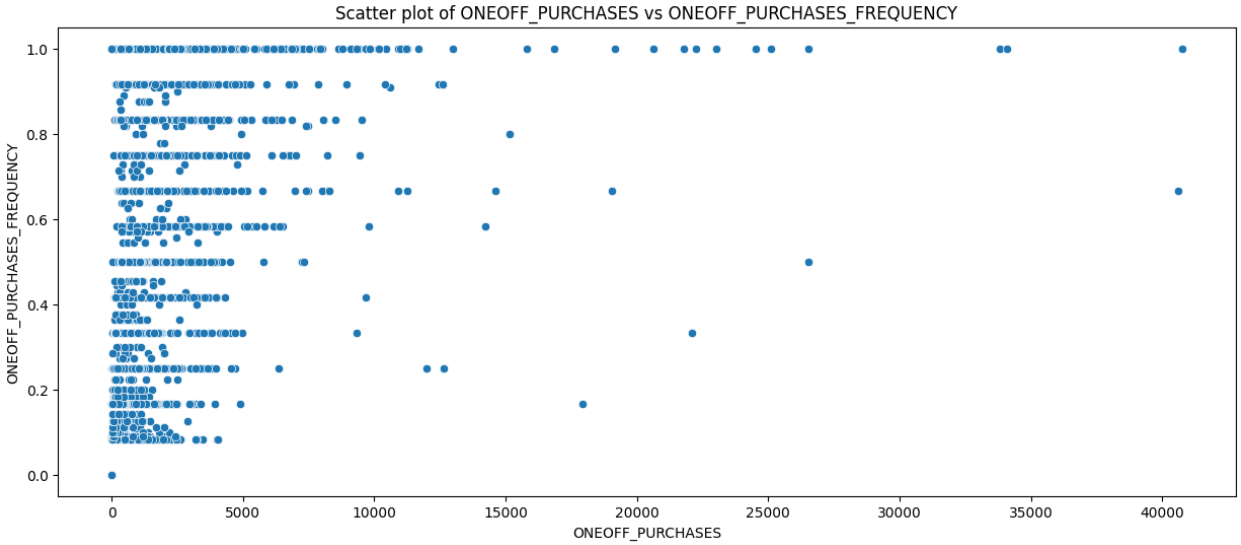

Scatter plot of PURCHASES vs MINIMUM_PAYMENTS
MINIMUM_PAYMENTS
PURCHASES
0
10000
20000
30000
40000
50000
60000
70000
80000

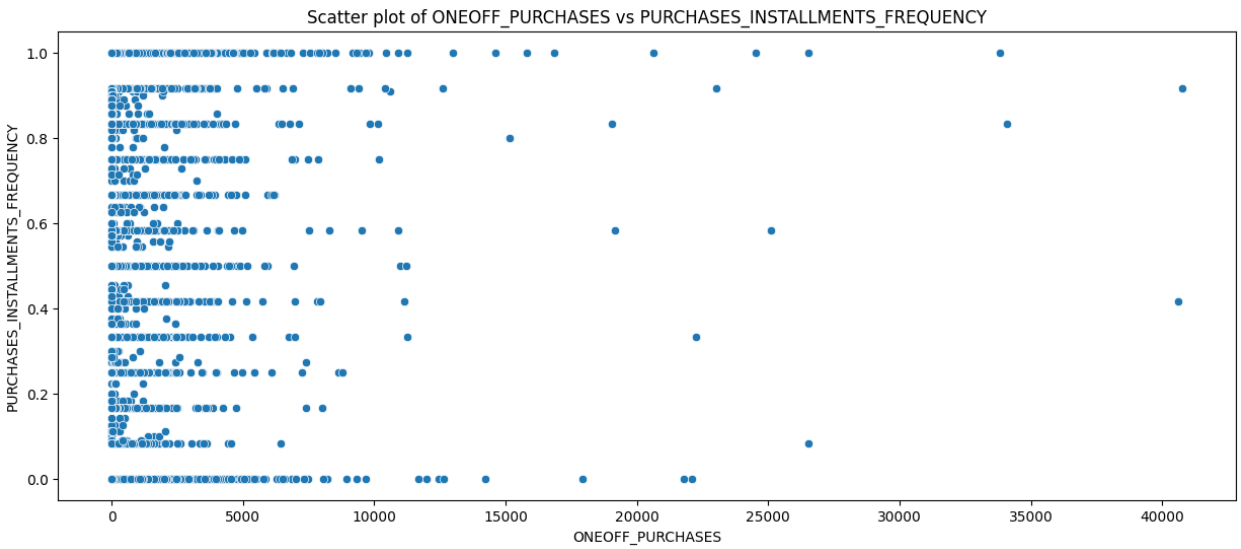

Scatter plot of PURCHASES vs PRC_FULL_PAYMENT
PRC_FULL_PAYMENT
PURCHASES
0.0
0.2
0.4
0.6
0.8
1.0
0
10000
20000
30000
40000
50000

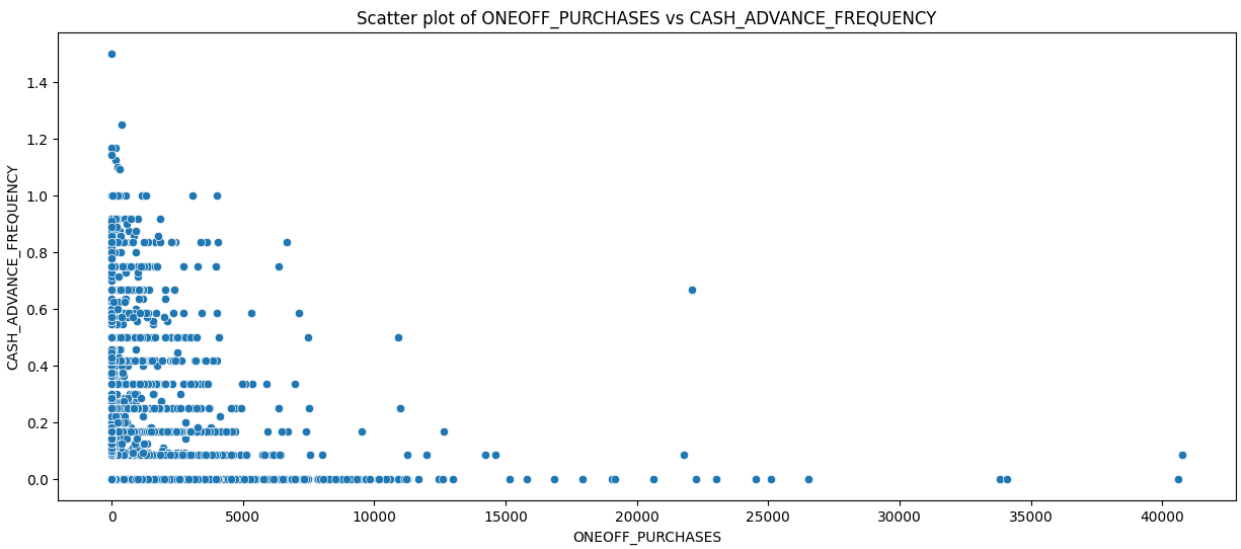

Scatter plot of ONEOFF_PURCHASES vs INSTALLMENTS_PURCHASES
INSTALLMENTS_PURCHASES
ONEOFF_PURCHASES
0
5000
10000
15000
20000
25000
30000
35000
40000

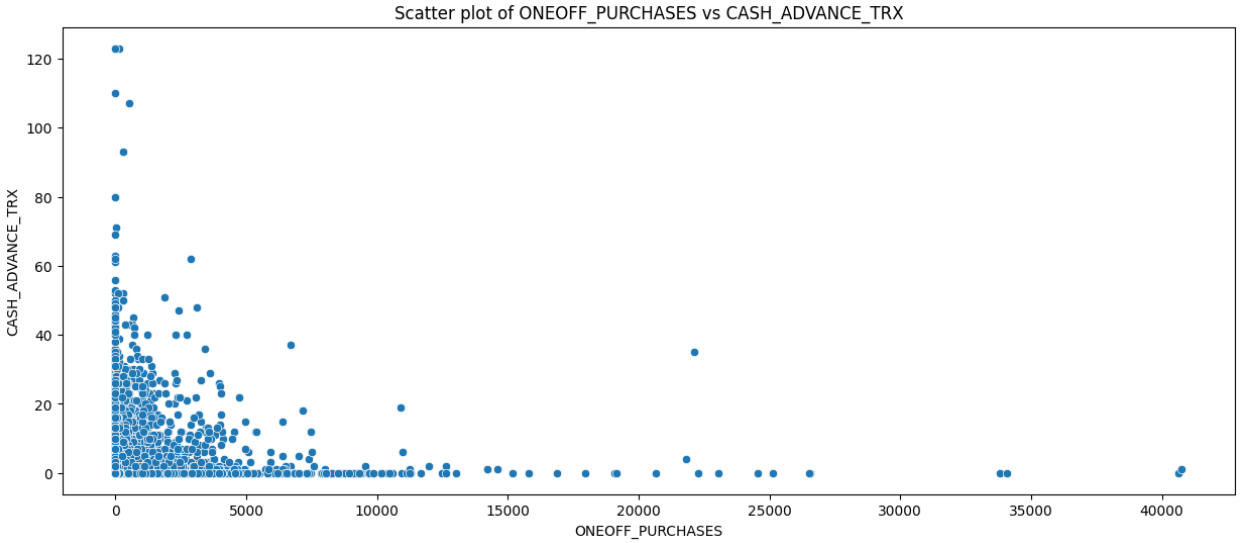

Scatter plot of ONEOFF_PURCHASES vs CASH_ADVANCE
CASH_ADVANCE
ONEOFF_PURCHASES
0
10000
20000
30000
40000
5000
15000
25000
35000

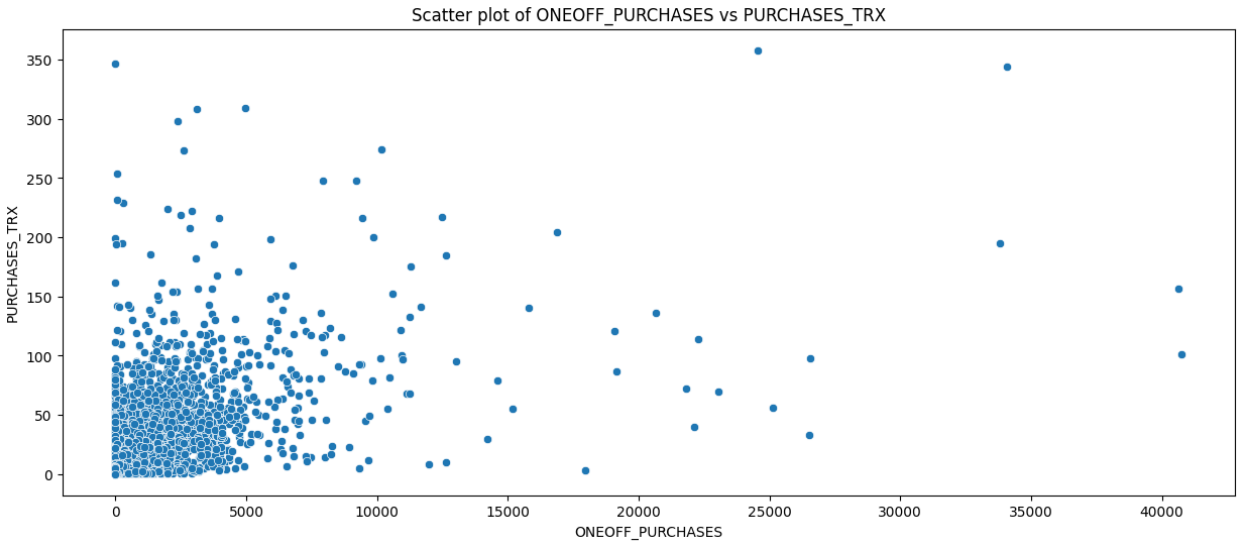

Scatter plot of ONEOFF_PURCHASES vs PURCHASES_FREQUENCY
PURCHASES_FREQUENCY
ONEOFF_PURCHASES
0.0
0.2
0.4
0.6
0.8
1.0
0
5000
10000
15000
20000
25000
30000
35000
40000

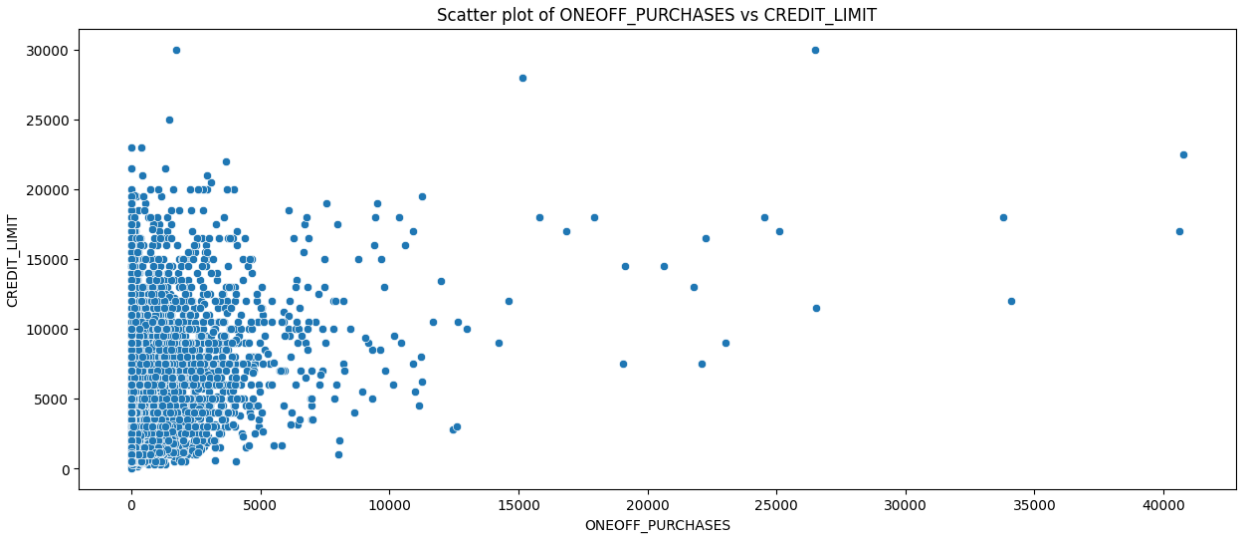

Scatter plot of ONEOFF_PURCHASES vs ONEOFF_PURCHASES_FREQUENCY
ONEOFF_PURCHASES_FREQUENCY
ONEOFF_PURCHASES
0.0
0.2
0.4
0.6
0.8
1.0
0
5000
10000
15000
20000
25000
30000
35000
40000

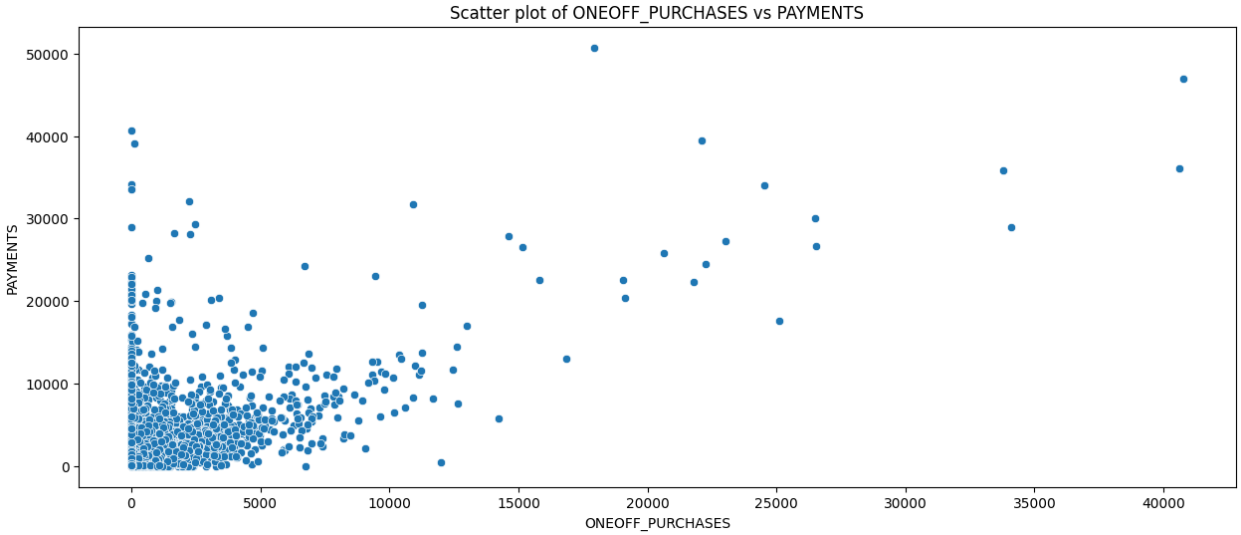

Scatter plot of ONEOFF_PURCHASES vs PURCHASES_INSTALLMENTS_FREQUENCY
PURCHASES_INSTALLMENTS_FREQUENCY
ONEOFF_PURCHASES
0.0
0.2
0.4
0.6
0.8
1.0
0
5000
10000
15000
20000
25000
30000
35000
40000

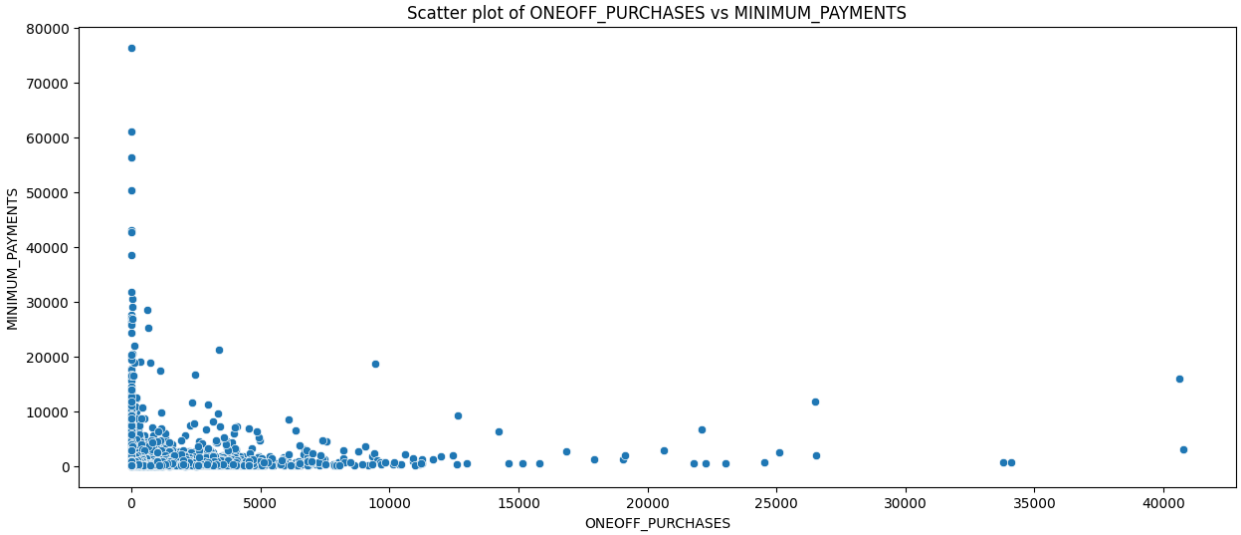

Scatter plot of ONEOFF_PURCHASES vs CASH_ADVANCE_FREQUENCY
CASH_ADVANCE_FREQUENCY
ONEOFF_PURCHASES
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0
5000
10000
15000
20000
25000
30000
35000
40000

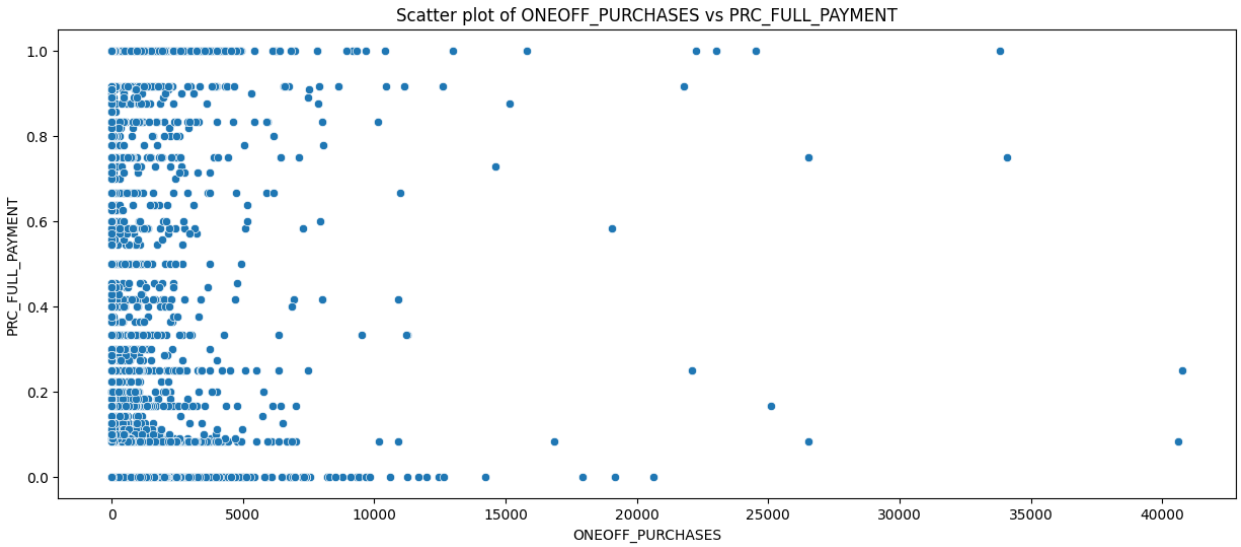

Scatter plot of ONEOFF_PURCHASES vs CASH_ADVANCE_TRX
CASH_ADVANCE_TRX
ONEOFF_PURCHASES
0
20
40
60
80
100
120
0
5000
10000
15000
20000
25000
30000
35000
40000

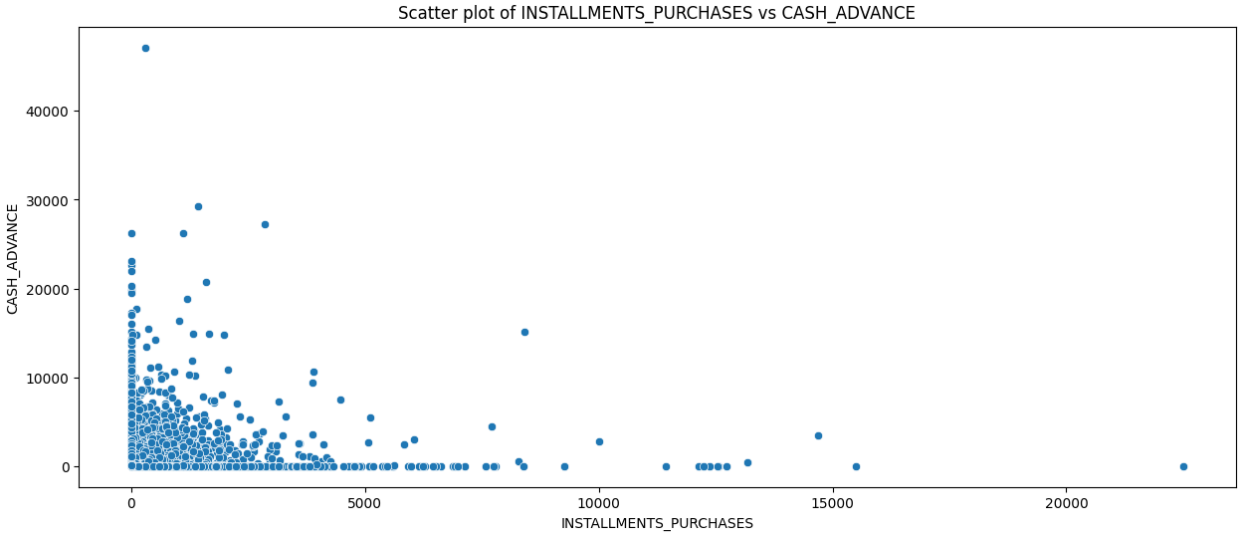

Scatter plot of ONEOFF_PURCHASES vs PURCHASES_TRX
PURCHASES_TRX
ONEOFF_PURCHASES
0
50
100
150
200
250
300
350
0
5000
10000
15000
20000
25000
30000
35000
40000

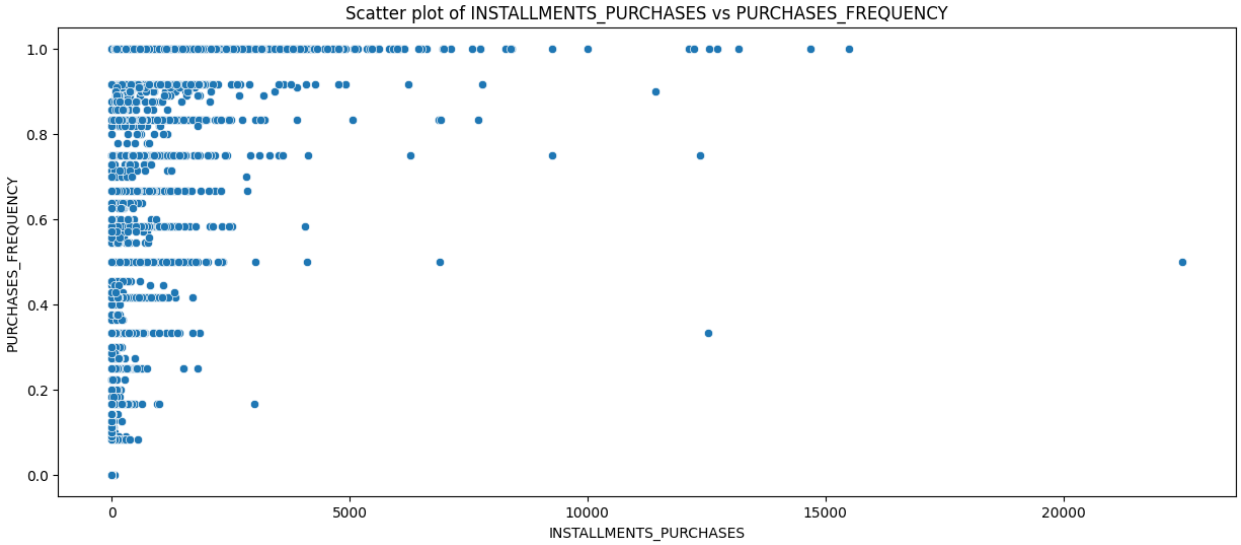

Scatter plot of ONEOFF_PURCHASES vs CREDIT_LIMIT
CREDIT_LIMIT
ONEOFF_PURCHASES
0
5000
10000
15000
20000
25000
30000
0
5000
10000
15000
20000
25000
30000
35000
40000

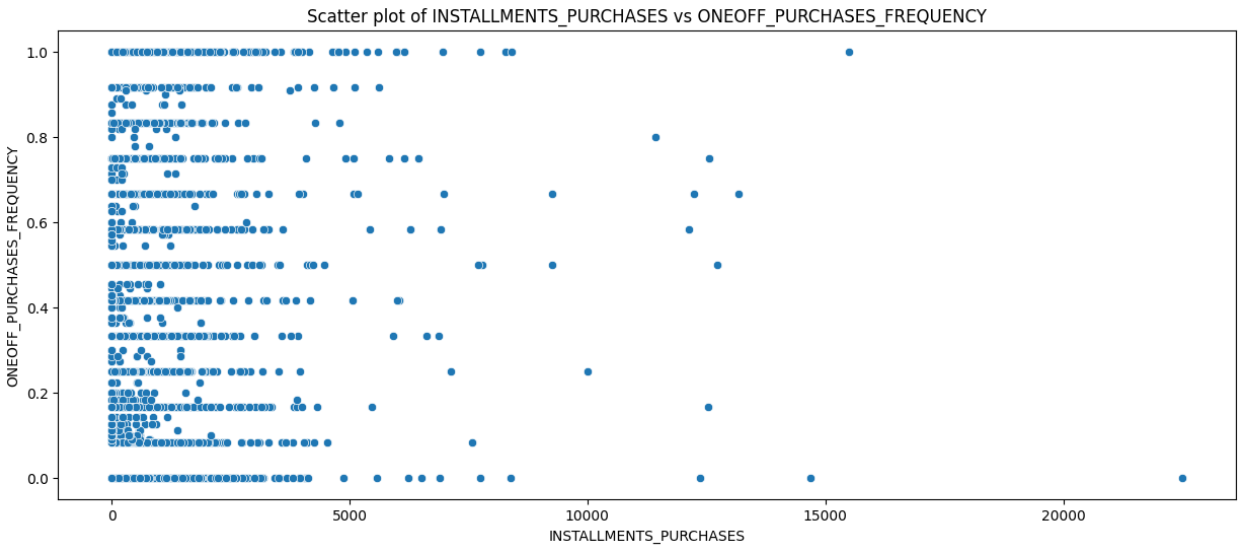

Scatter plot of ONEOFF_PURCHASES vs PAYMENTS
PAYMENTS
ONEOFF_PURCHASES
0
10000
20000
30000
40000
50000
0
5000
10000
15000
20000
25000
30000
35000
40000

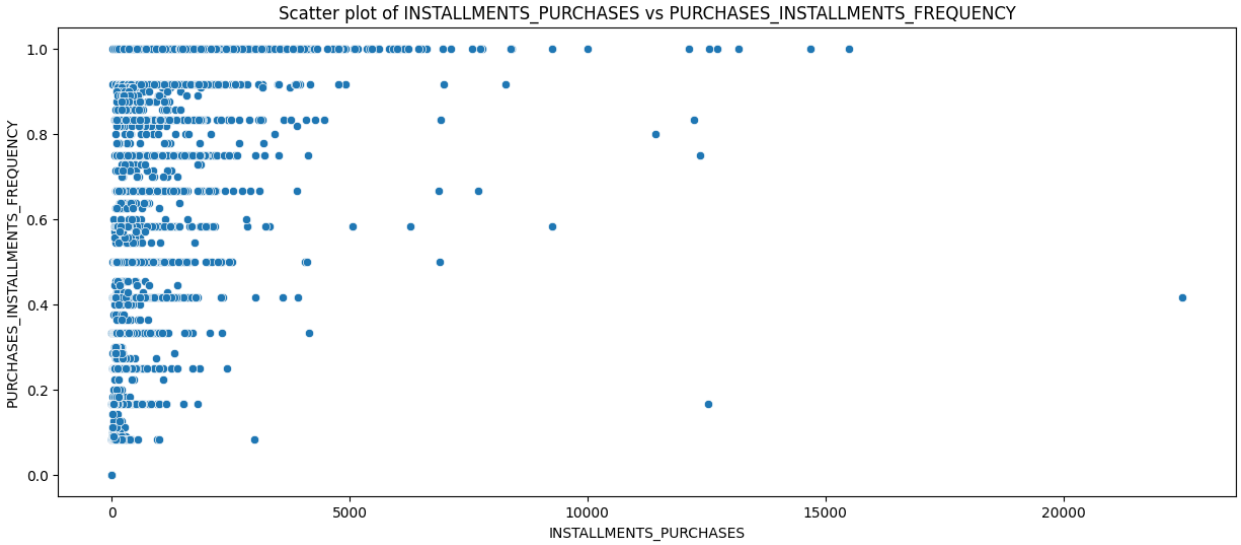

Scatter plot of ONEOFF_PURCHASES vs MINIMUM_PAYMENTS
MINIMUM_PAYMENTS
ONEOFF_PURCHASES
0
10000
20000
30000
40000
50000
60000
70000
80000
0
5000
10000
15000
20000
25000
30000
35000
40000

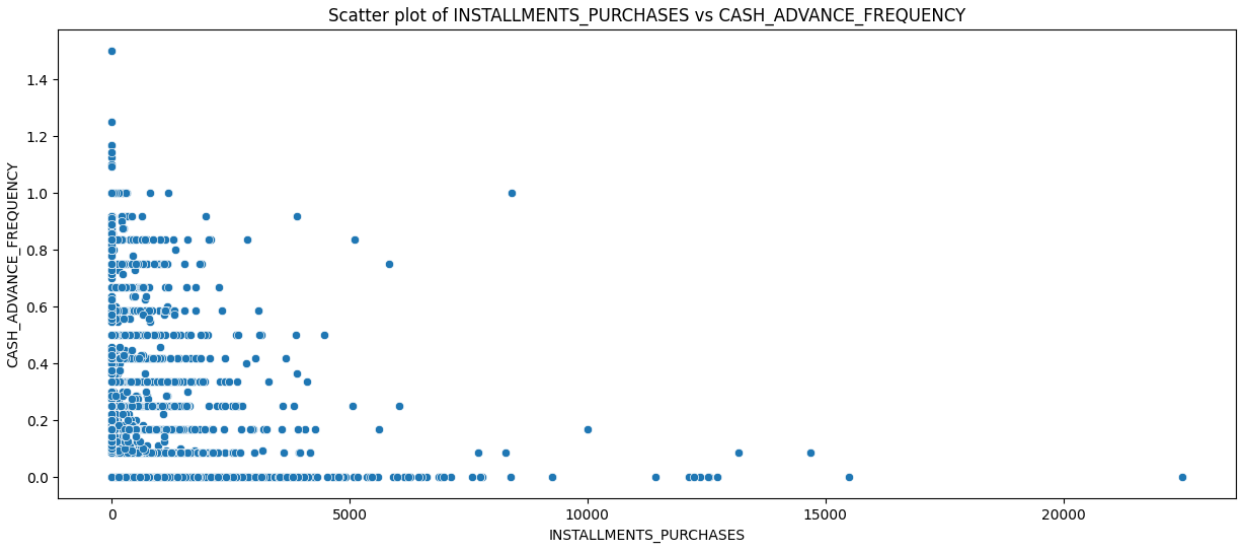

Scatter plot of ONEOFF_PURCHASES vs PRC_FULL_PAYMENT
PRC_FULL_PAYMENT
ONEOFF_PURCHASES
0.0
0.2
0.4
0.6
0.8
1.0
0
5000
10000
15000
20000
25000
30000
35000
40000

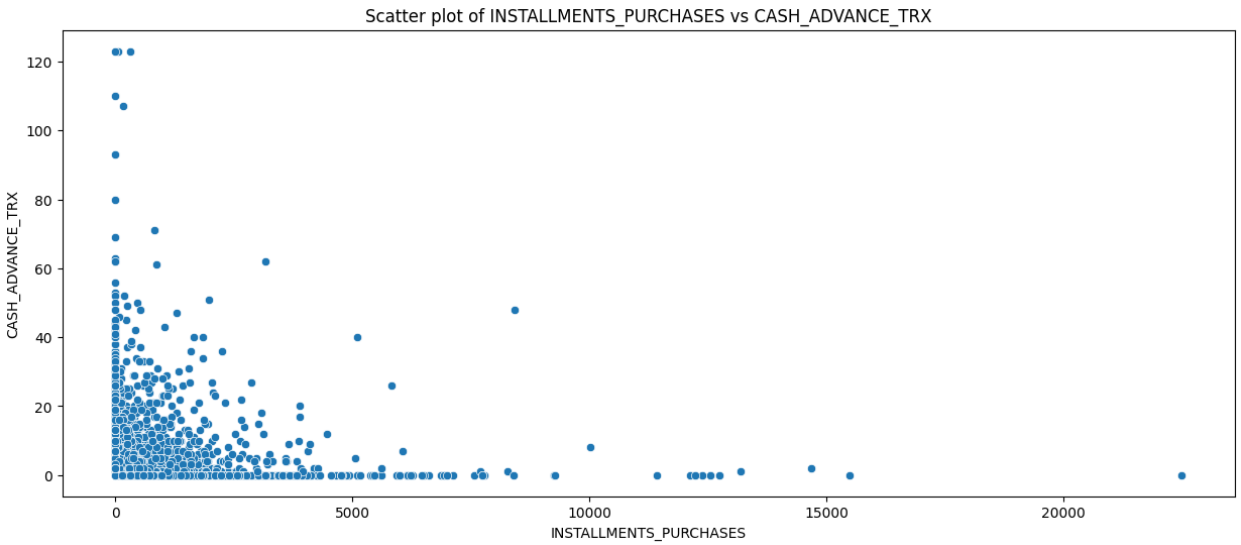

Scatter plot of INSTALLMENTS_PURCHASES vs CASH_ADVANCE
CASH_ADVANCE
INSTALLMENTS_PURCHASES
0
10000
20000
30000
40000
0
5000
10000
15000
20000

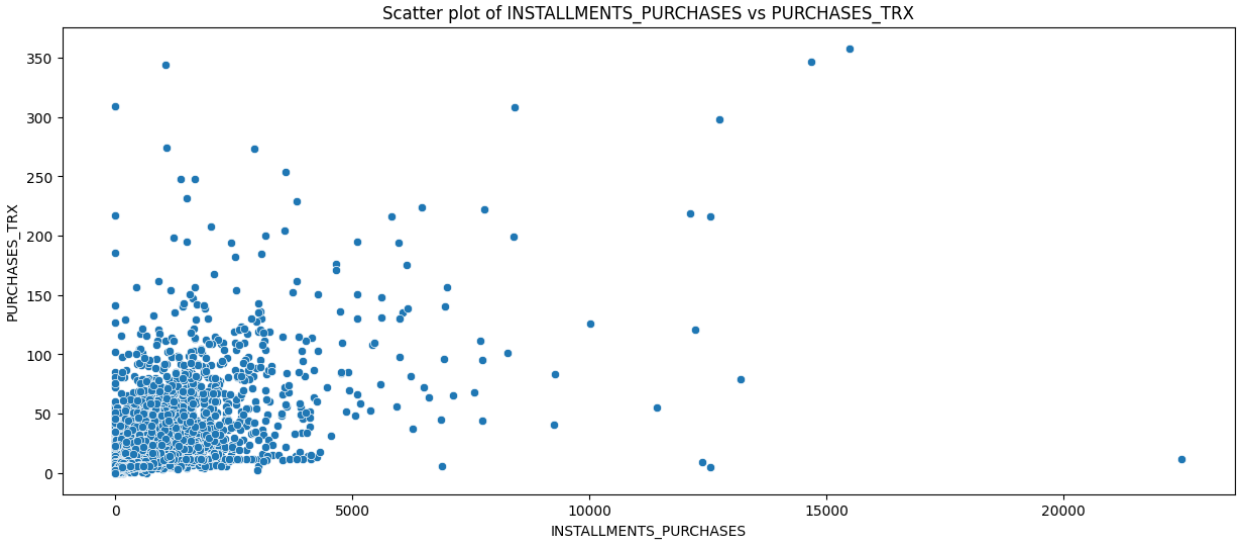

Scatter plot of INSTALLMENTS_PURCHASES vs PURCHASES_FREQUENCY
PURCHASES_FREQUENCY
INSTALLMENTS_PURCHASES
0.0
0.2
0.4
0.6
0.8
1.0
0
5000
10000
15000
20000

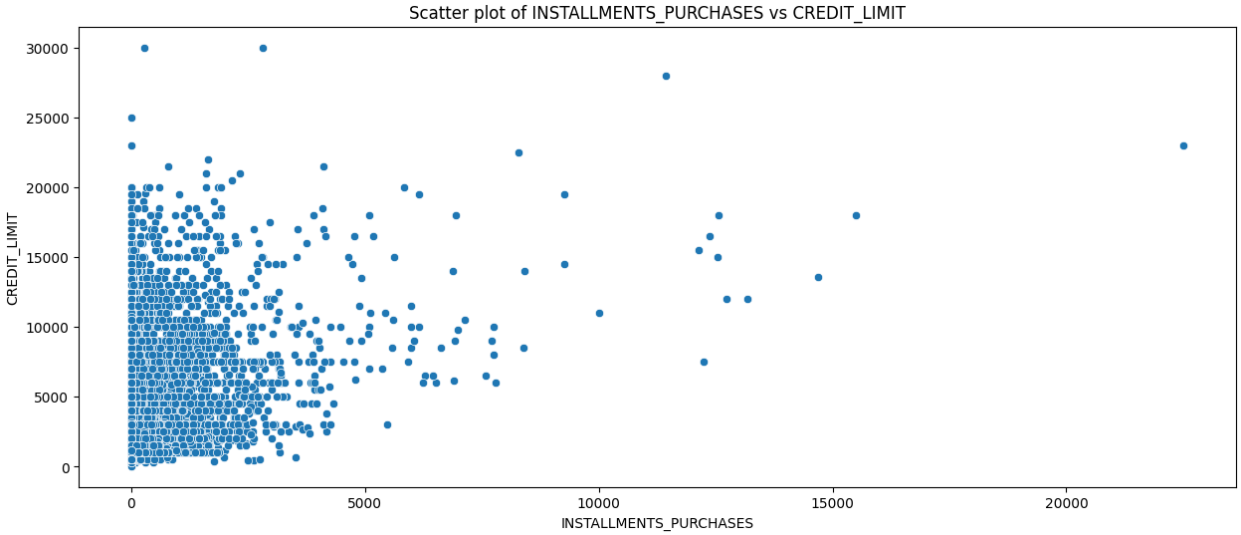

Scatter plot of INSTALLMENTS_PURCHASES vs ONEOFF_PURCHASES_FREQUENCY
ONEOFF_PURCHASES_FREQUENCY
INSTALLMENTS_PURCHASES
0.0
0.2
0.4
0.6
0.8
1.0
0
5000
10000
15000
20000

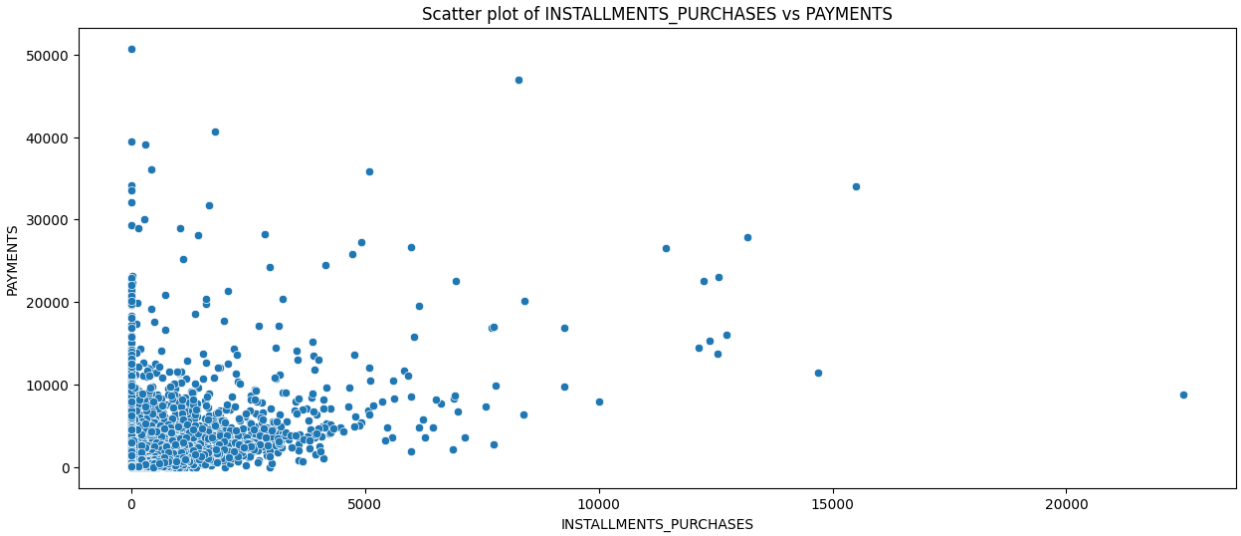

Scatter plot of INSTALLMENTS_PURCHASES vs PURCHASES_INSTALLMENTS_FREQUENCY
PURCHASES_INSTALLMENTS_FREQUENCY
INSTALLMENTS_PURCHASES
0.0
0.2
0.4
0.6
0.8
1.0
0
5000
10000
15000
20000

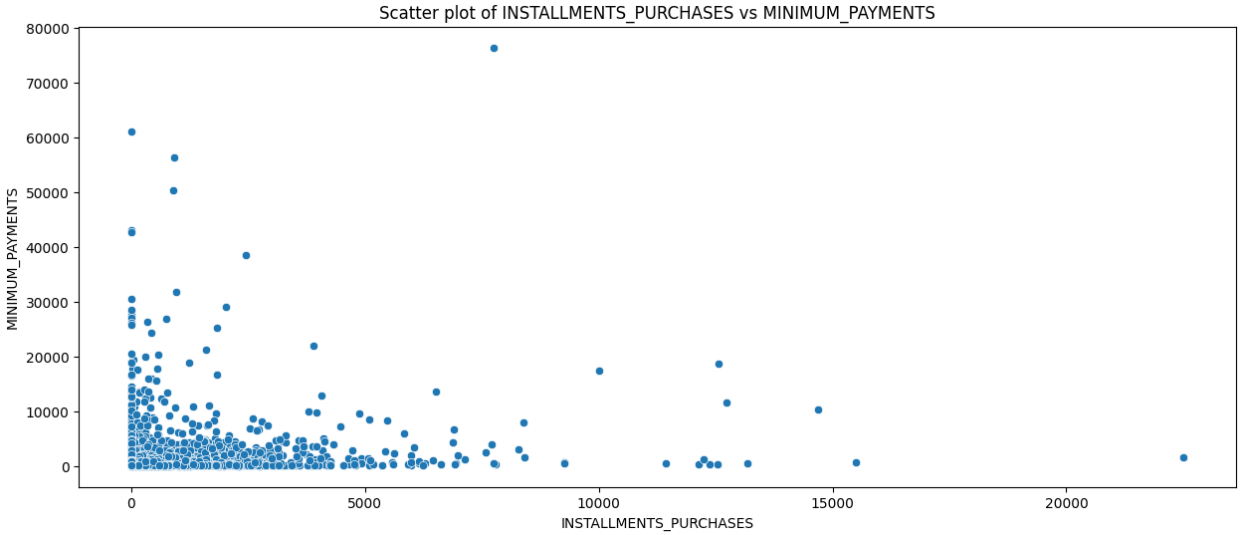

Scatter plot of INSTALLMENTS_PURCHASES vs CASH_ADVANCE_FREQUENCY
CASH_ADVANCE_FREQUENCY
INSTALLMENTS_PURCHASES
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0
5000
10000
15000
20000

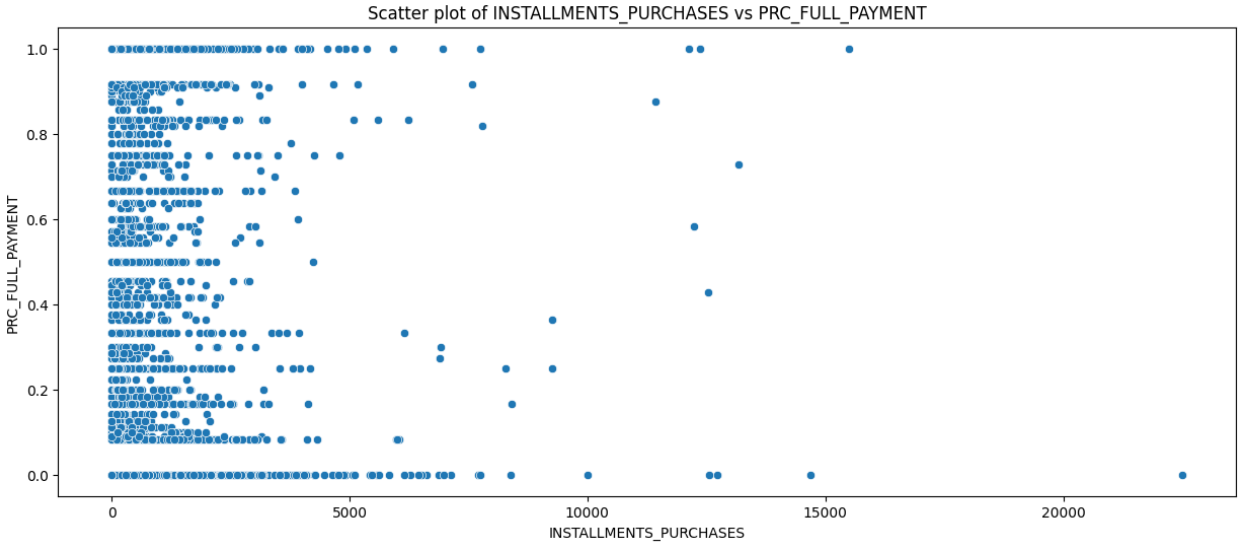

Scatter plot of INSTALLMENTS_PURCHASES vs CASH_ADVANCE_TRX
CASH_ADVANCE_TRX
INSTALLMENTS_PURCHASES
0
20
40
60
80
100
120
0
5000
10000
15000
20000

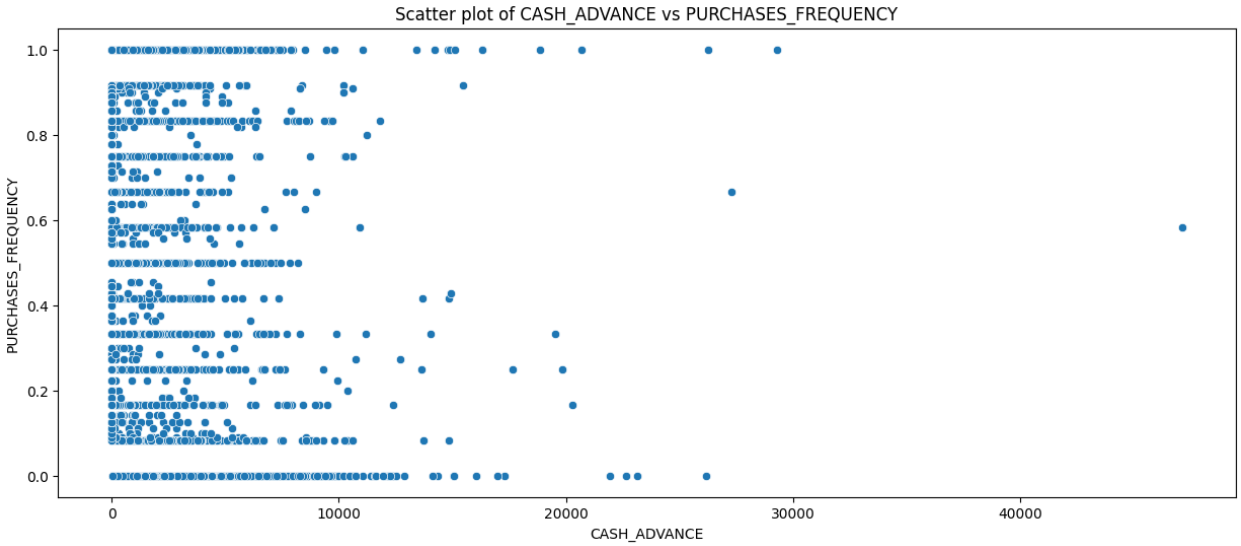

Scatter plot of INSTALLMENTS_PURCHASES vs PURCHASES_TRX
PURCHASES_TRX
INSTALLMENTS_PURCHASES
0
50
100
150
200
250
300
350
0
5000
10000
15000
20000

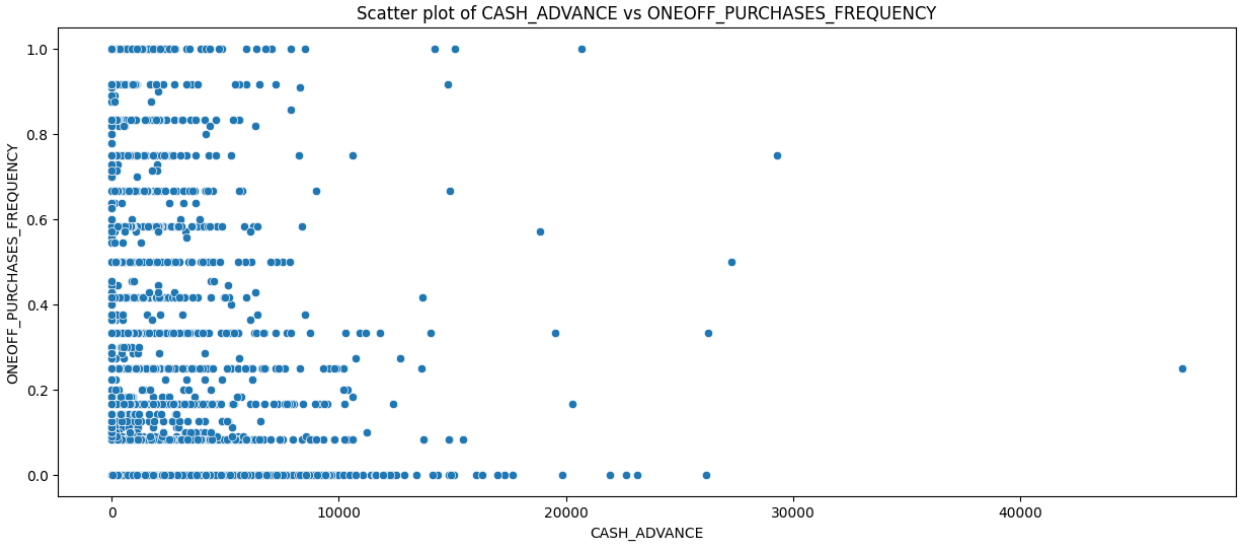

Scatter plot of INSTALLMENTS_PURCHASES vs CREDIT_LIMIT
CREDIT_LIMIT
INSTALLMENTS_PURCHASES
0
5000
10000
15000
20000
25000
30000
0
5000
10000
15000
20000

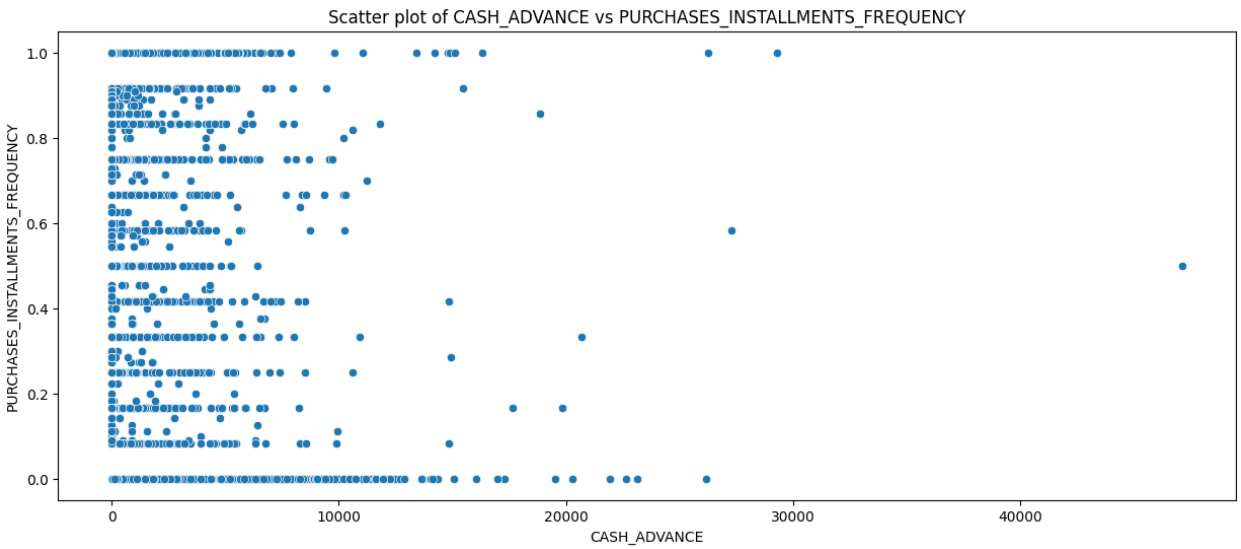

Scatter plot of INSTALLMENTS_PURCHASES vs PAYMENTS
PAYMENTS
INSTALLMENTS_PURCHASES
0
10000
20000
30000
40000
50000
0
5000
10000
15000
20000

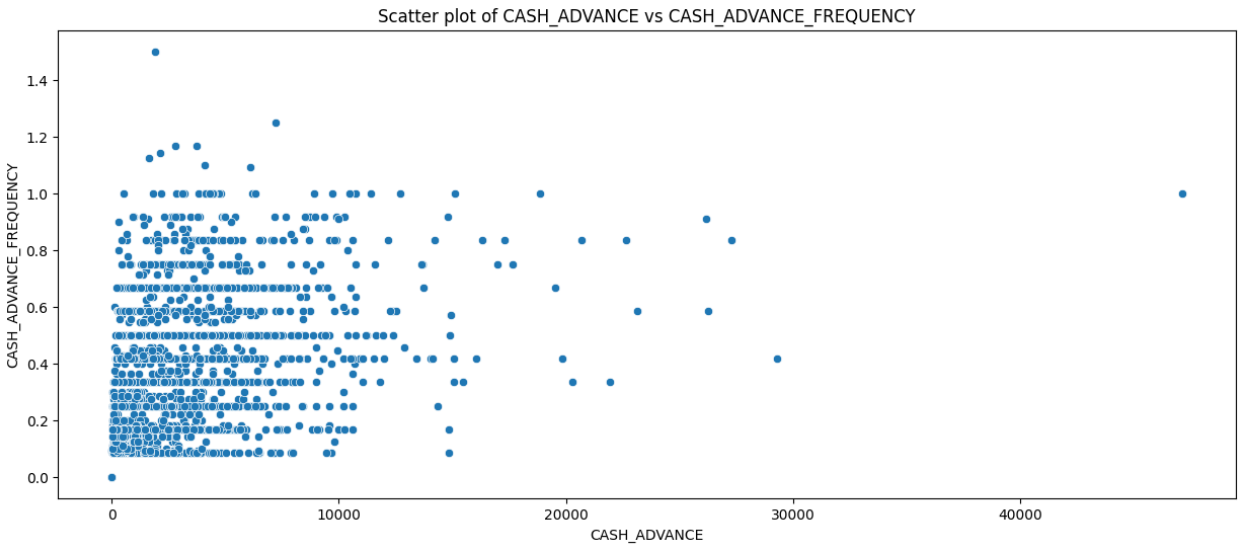

Scatter plot of INSTALLMENTS_PURCHASES vs MINIMUM_PAYMENTS
MINIMUM_PAYMENTS
INSTALLMENTS_PURCHASES
0
10000
20000
30000
40000
50000
60000
70000
80000
0
5000
10000
15000
20000

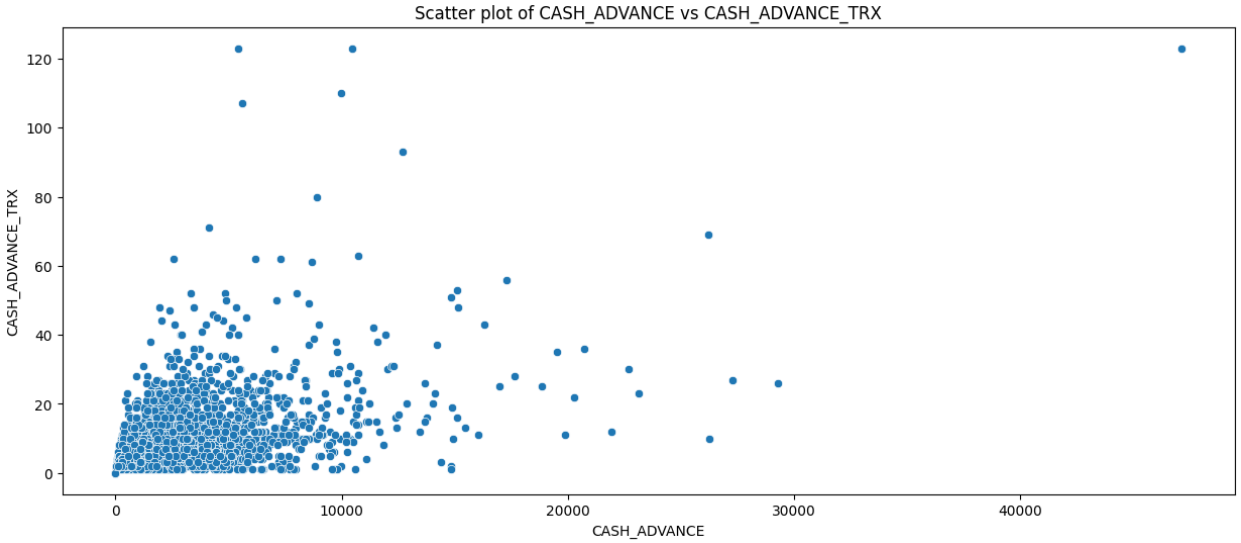

Scatter plot of INSTALLMENTS_PURCHASES vs PRC_FULL_PAYMENT
PRC_FULL_PAYMENT
INSTALLMENTS_PURCHASES
0.0
0.2
0.4
0.6
0.8
1.0
0
5000
10000
15000
20000

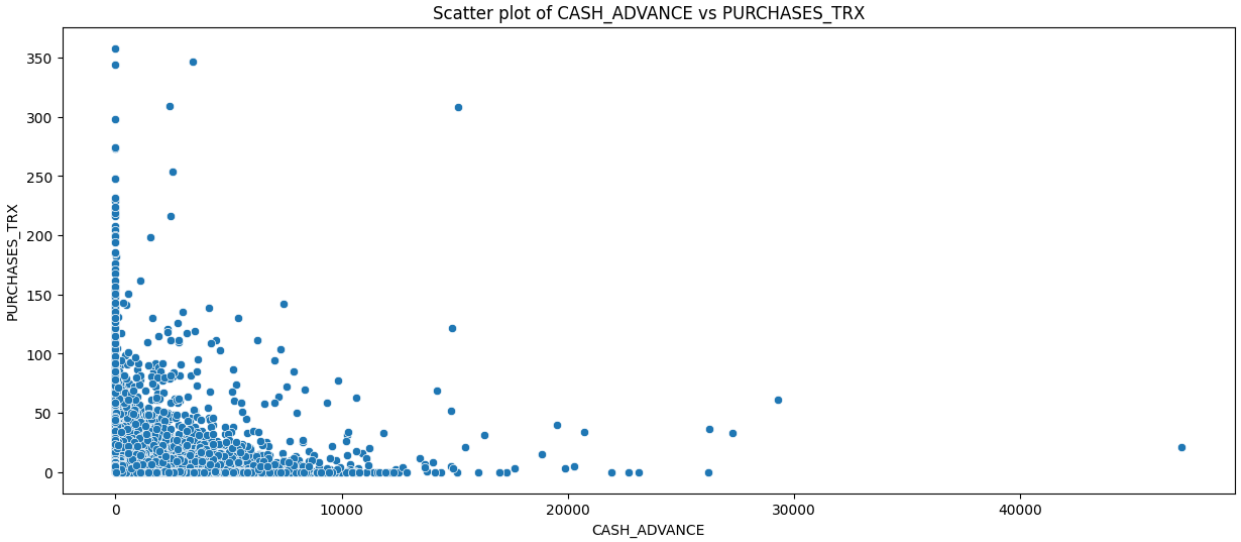

Scatter plot of CASH_ADVANCE vs PURCHASES_FREQUENCY
PURCHASES_FREQUENCY
CASH_ADVANCE
0.0
0.2
0.4
0.6
0.8
1.0
0
10000
20000
30000
40000

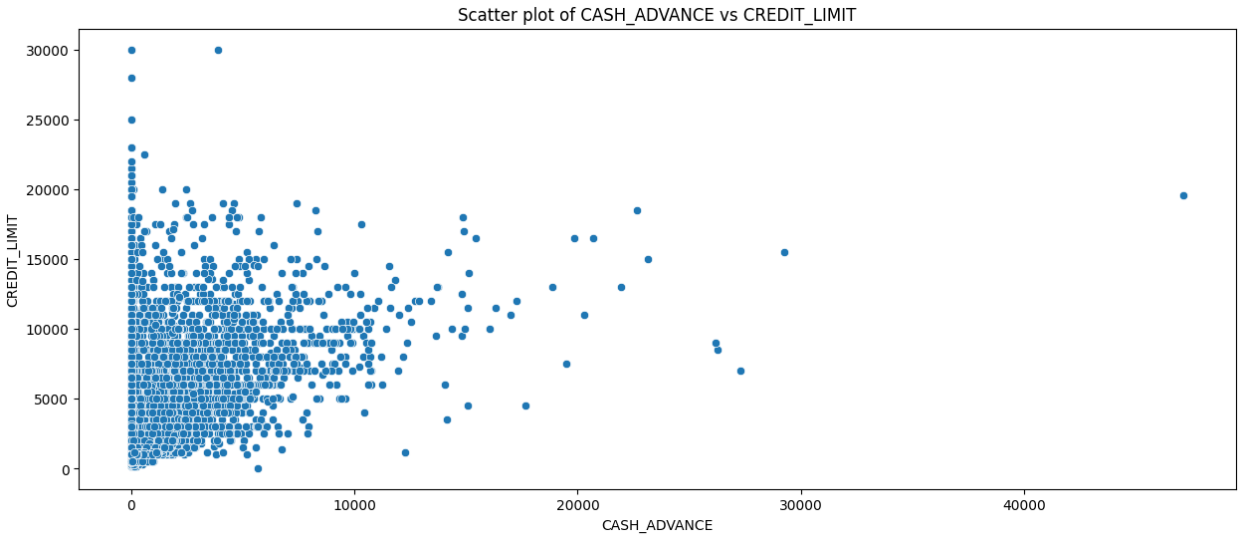

Scatter plot of CASH_ADVANCE vs ONEOFF_PURCHASES_FREQUENCY
ONEOFF_PURCHASES_FREQUENCY
CASH_ADVANCE
0.0
0.2
0.4
0.6
0.8
1.0
0
10000
20000
30000
40000

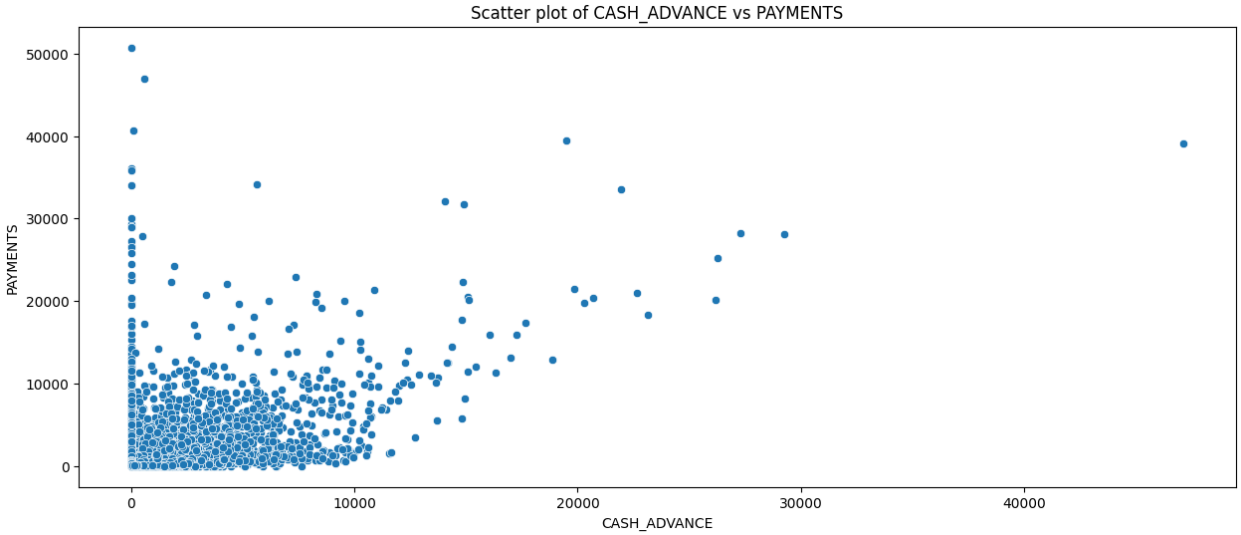

Scatter plot of CASH_ADVANCE vs PURCHASES_INSTALLMENTS_FREQUENCY
PURCHASES_INSTALLMENTS_FREQUENCY
CASH_ADVANCE
0.0
0.2
0.4
0.6
0.8
1.0
0
10000
20000
30000
40000

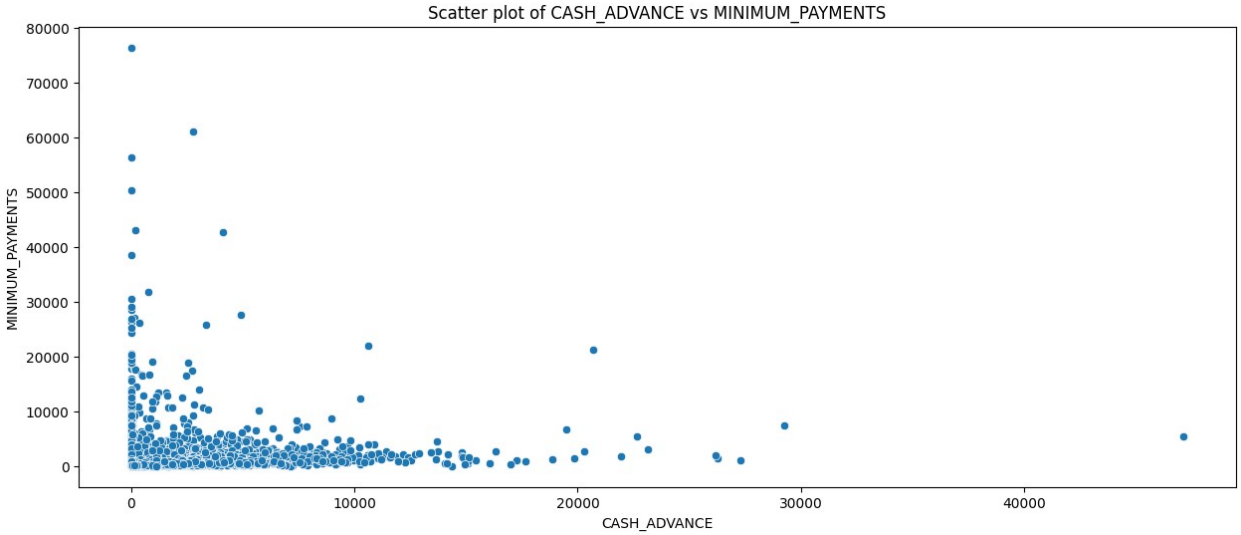

Scatter plot of CASH_ADVANCE vs CASH_ADVANCE_FREQUENCY
CASH_ADVANCE_FREQUENCY
CASH_ADVANCE
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0
10000
20000
30000
40000

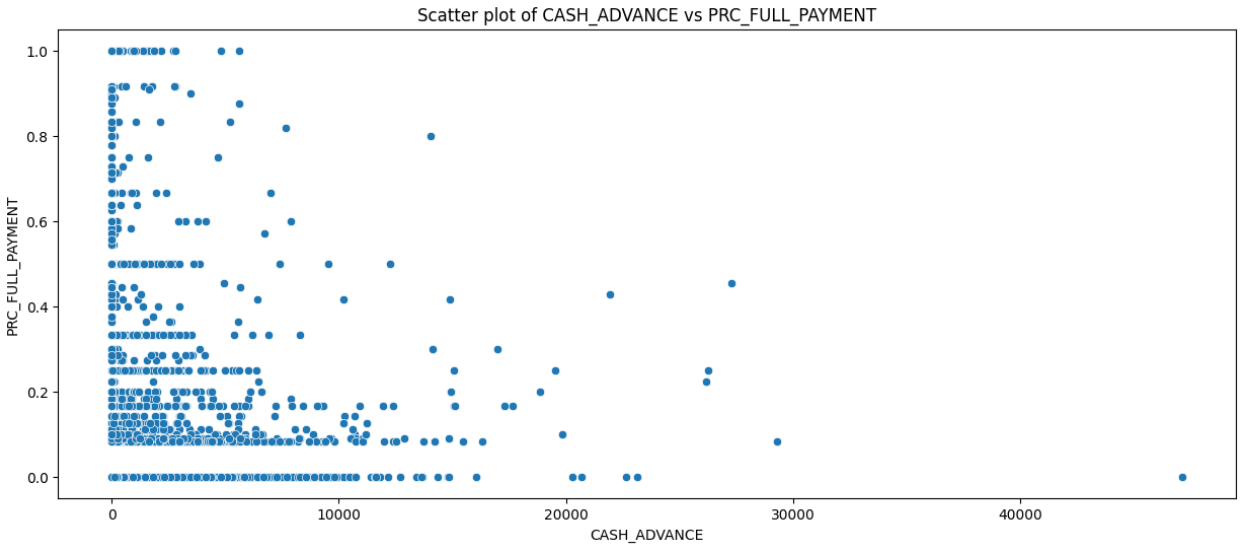

Scatter plot of CASH_ADVANCE vs CASH_ADVANCE_TRX
CASH_ADVANCE_TRX
CASH_ADVANCE
0
20
40
60
80
100
120
0
10000
20000
30000
40000

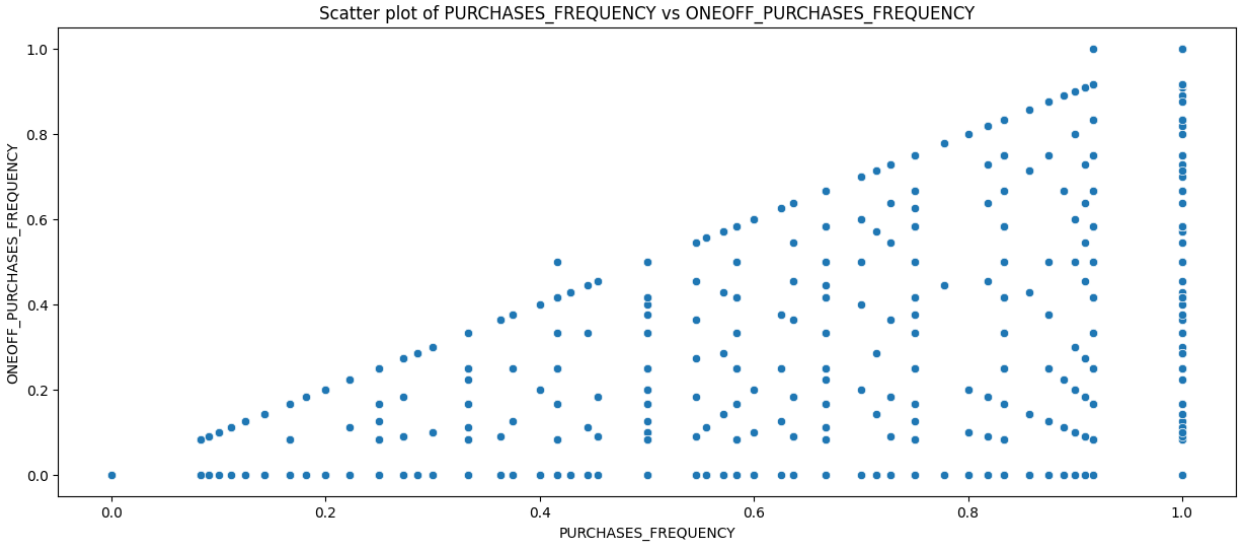

Scatter plot of CASH_ADVANCE vs PURCHASES_TRX
PURCHASES_TRX
CASH_ADVANCE
0
50
100
150
200
250
300
350
0
10000
20000
30000
40000

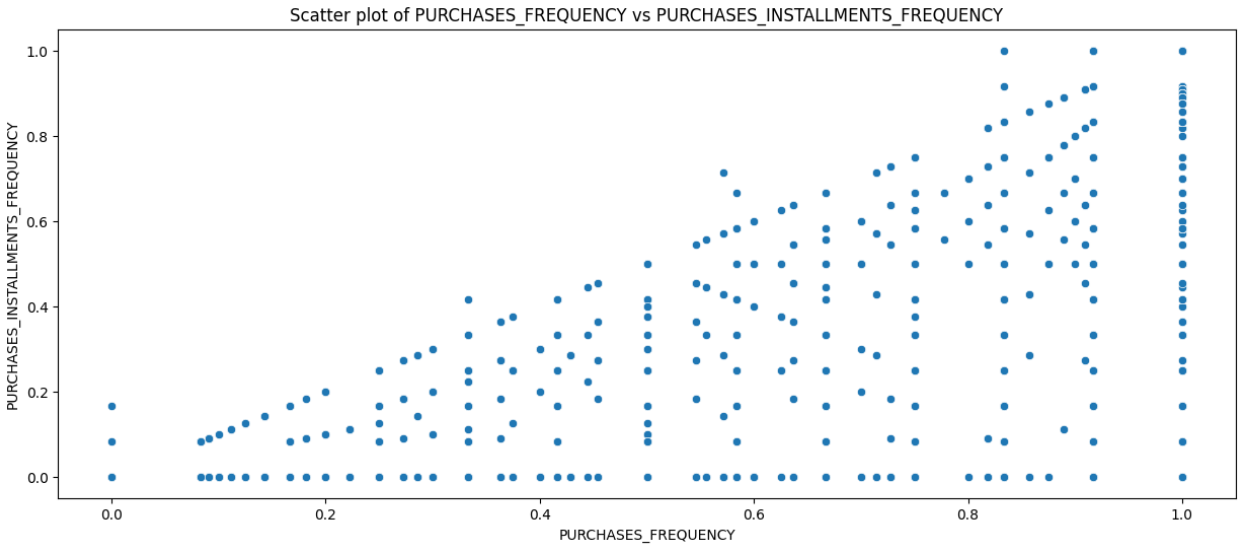

Scatter plot of CASH_ADVANCE vs CREDIT_LIMIT
CREDIT_LIMIT
CASH_ADVANCE
0
5000
10000
15000
20000
25000
30000
0
10000
20000
30000
40000

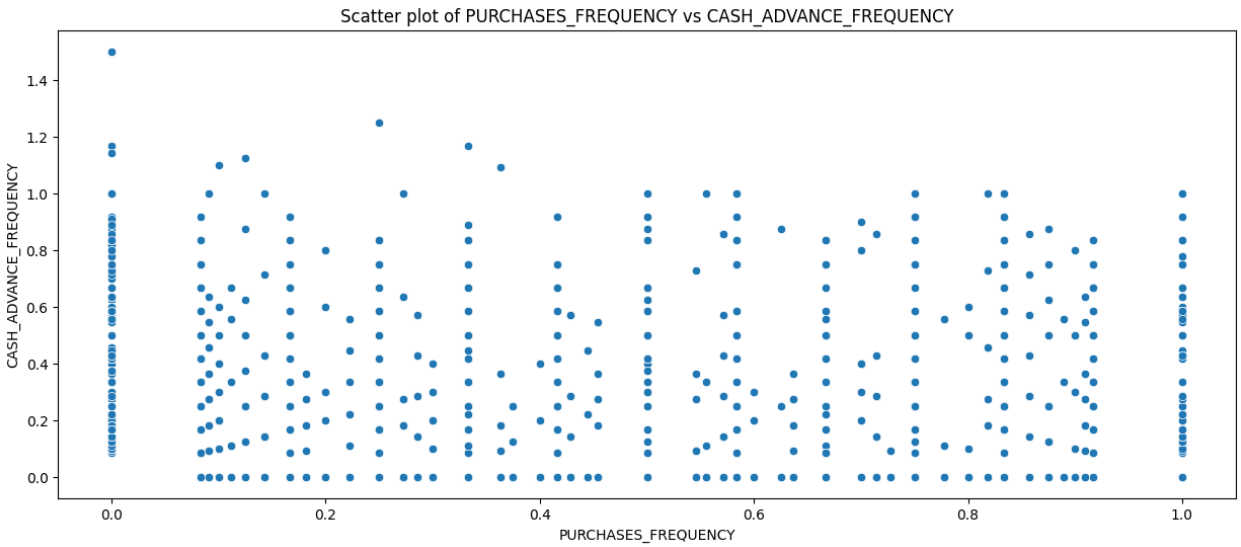

Scatter plot of CASH_ADVANCE vs PAYMENTS
PAYMENTS
CASH_ADVANCE
0
10000
20000
30000
40000
50000

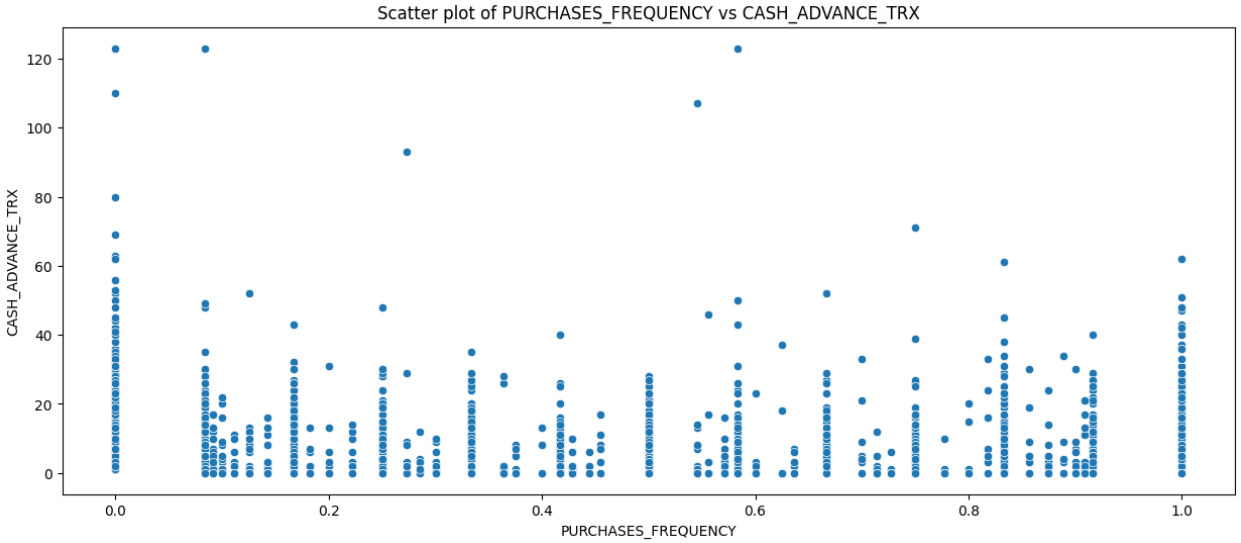

Scatter plot of CASH_ADVANCE vs MINIMUM_PAYMENTS
MINIMUM_PAYMENTS
CASH_ADVANCE
0
10000
20000
30000
40000
50000
60000
70000
80000

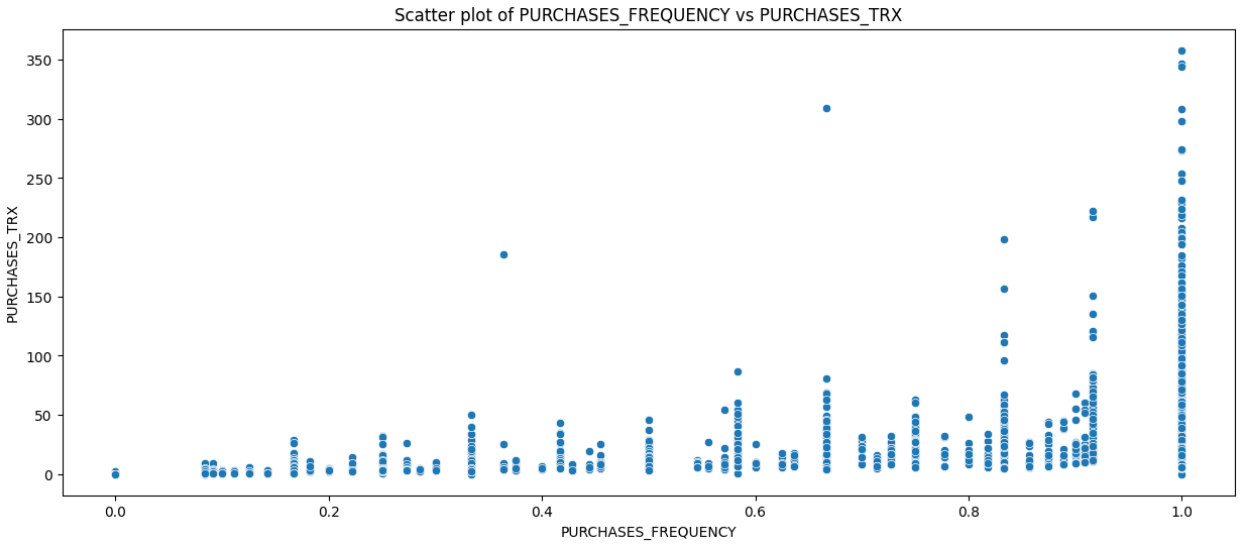

Scatter plot of CASH_ADVANCE vs PRC_FULL_PAYMENT
PRC_FULL_PAYMENT
CASH_ADVANCE
0.0
0.2
0.4
0.6
0.8
1.0
0
10000
20000
30000
40000

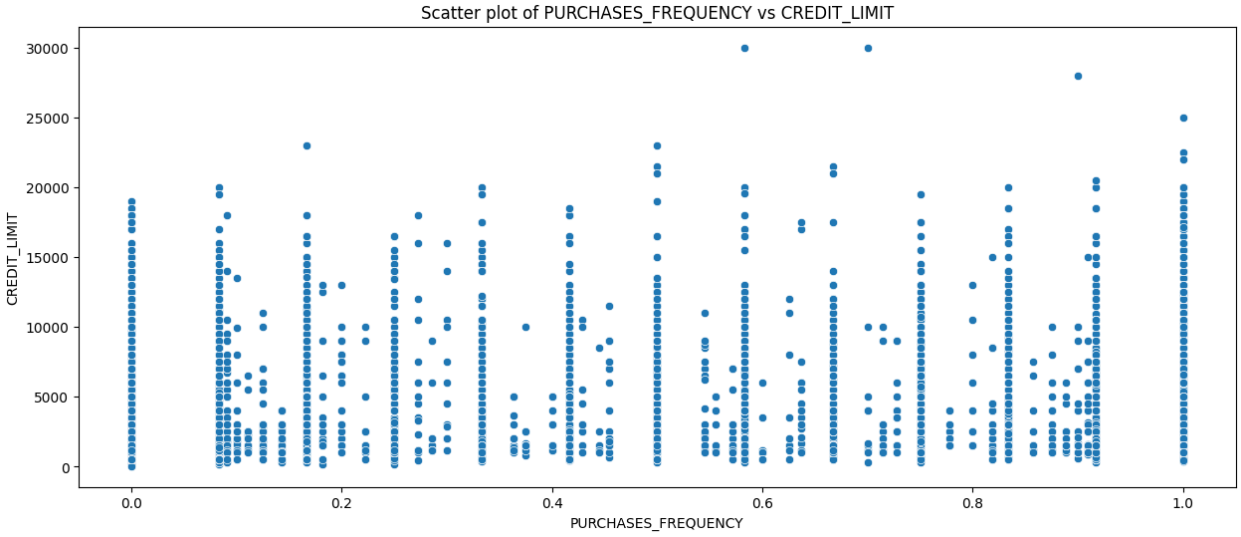

Scatter plot of PURCHASES_FREQUENCY vs ONEOFF_PURCHASES_FREQUENCY
ONEOFF_PURCHASES_FREQUENCY
PURCHASES_FREQUENCY
0.0
0.2
0.4
0.6
0.8
1.0

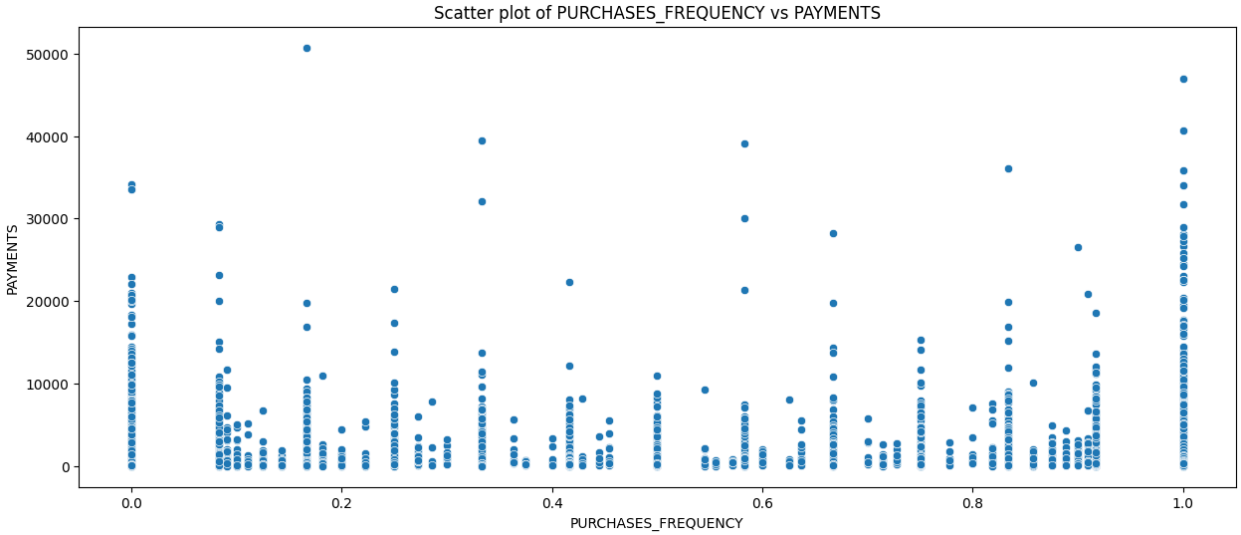

Scatter plot of PURCHASES_FREQUENCY vs PURCHASES_INSTALLMENTS_FREQUENCY
PURCHASES_INSTALLMENTS_FREQUENCY
PURCHASES_FREQUENCY
0.0
0.2
0.4
0.6
0.8
1.0

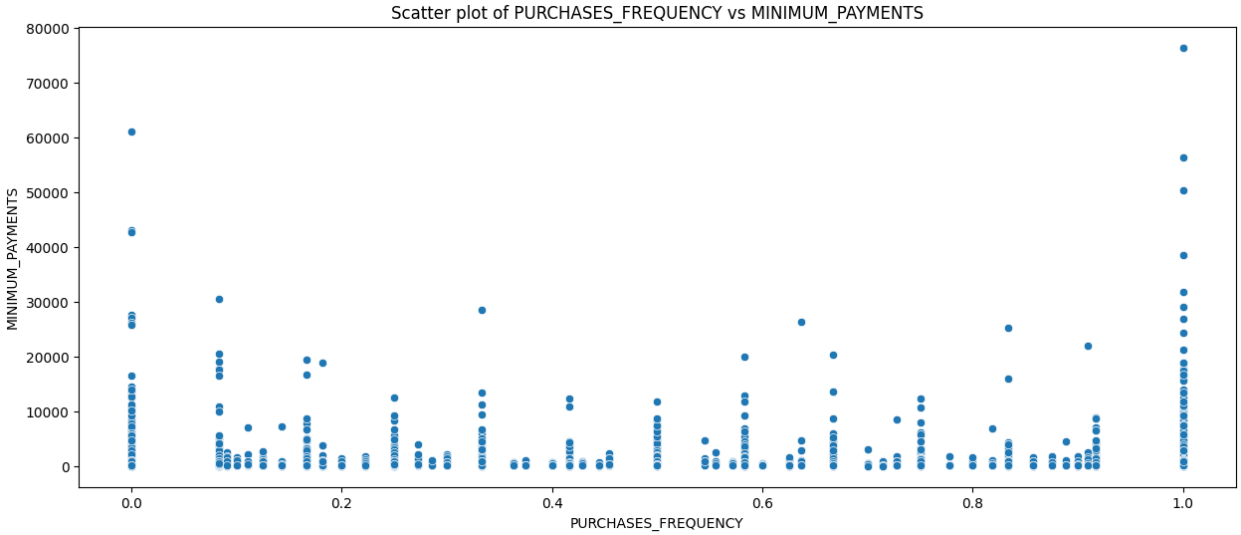

Scatter plot of PURCHASES_FREQUENCY vs CASH_ADVANCE_FREQUENCY
CASH_ADVANCE_FREQUENCY
PURCHASES_FREQUENCY
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4

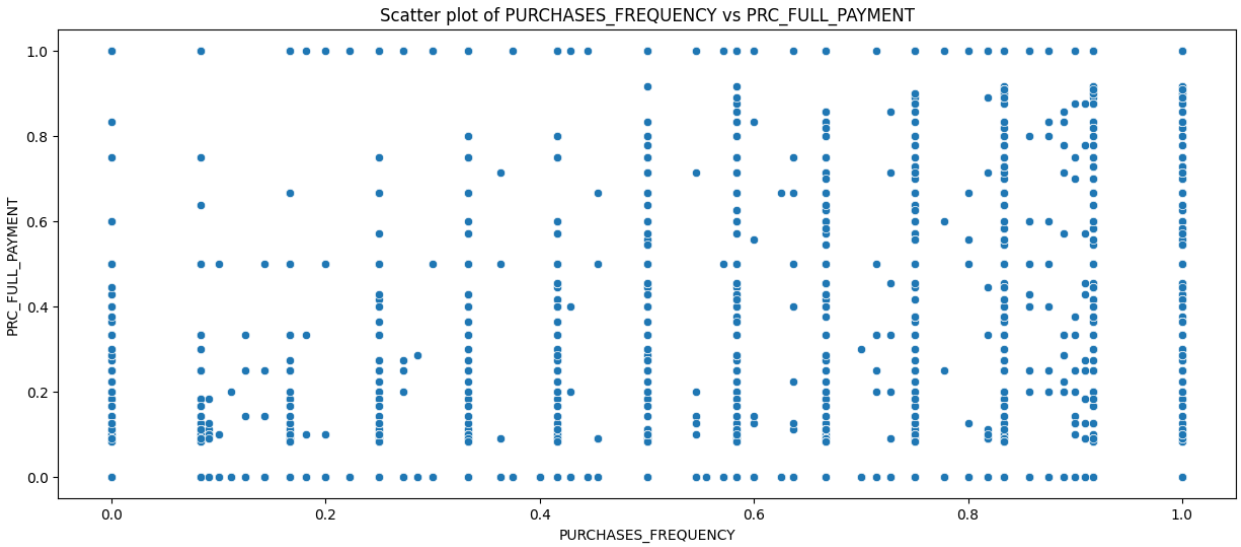

Scatter plot of PURCHASES_FREQUENCY vs CASH_ADVANCE_TRX
CASH_ADVANCE_TRX
PURCHASES_FREQUENCY
0
20
40
60
80
100
120
0.0
0.2
0.4
0.6
0.8
1.0

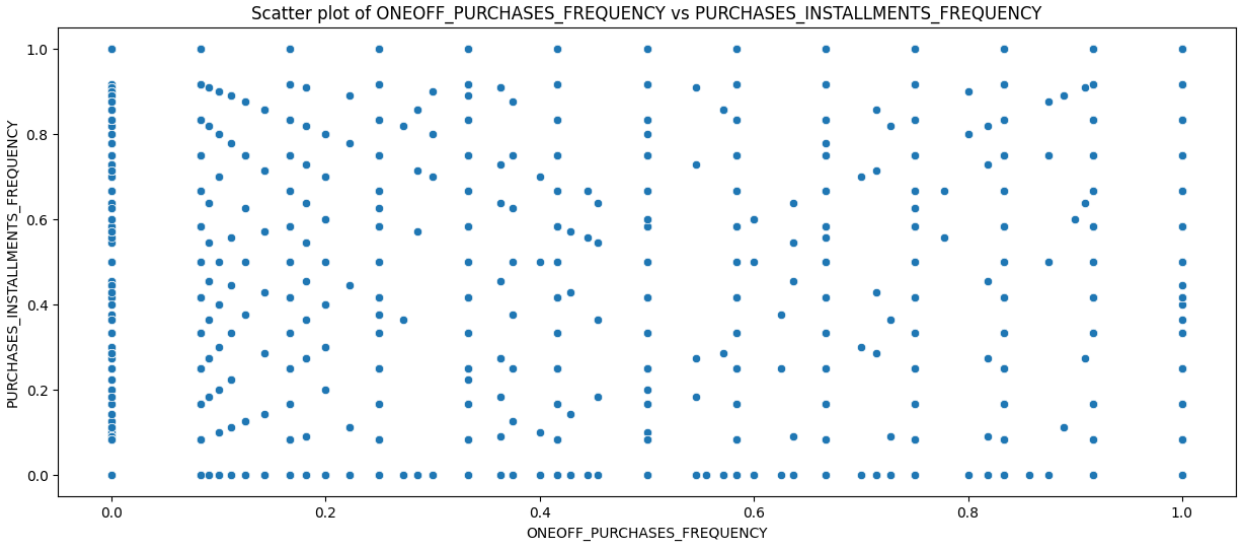

Scatter plot of PURCHASES_FREQUENCY vs PURCHASES_TRX
PURCHASES_TRX
PURCHASES_FREQUENCY
0
50
100
150
200
250
300
350
0.0
0.2
0.4
0.6
0.8
1.0

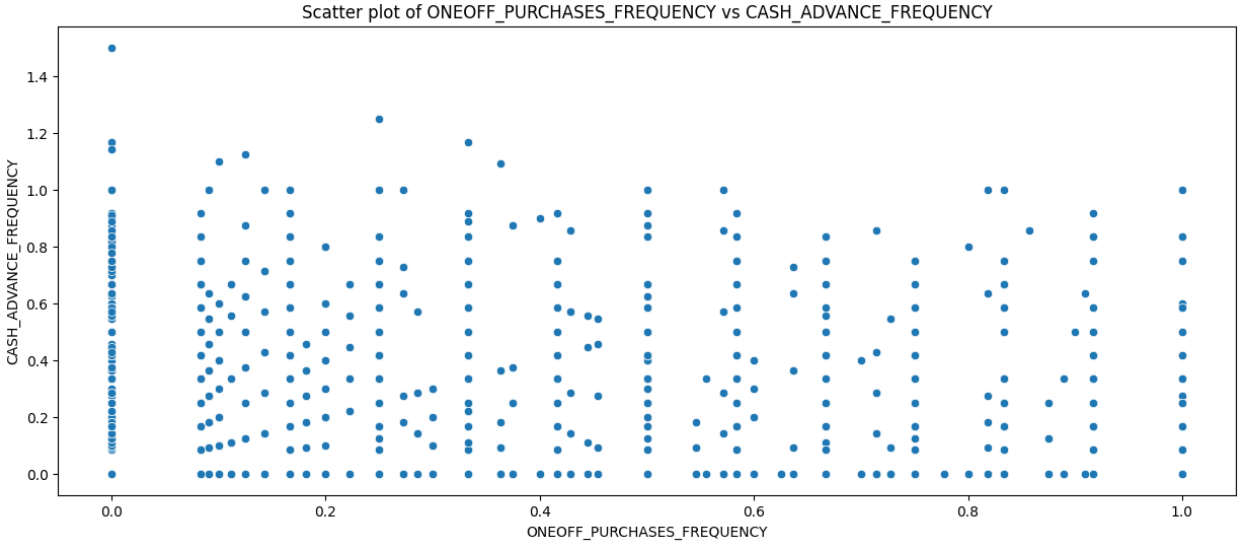

Scatter plot of PURCHASES_FREQUENCY vs CREDIT_LIMIT
CREDIT_LIMIT
PURCHASES_FREQUENCY
0
5000
10000
15000
20000
25000
30000
0.0
0.2
0.4
0.6
0.8
1.0

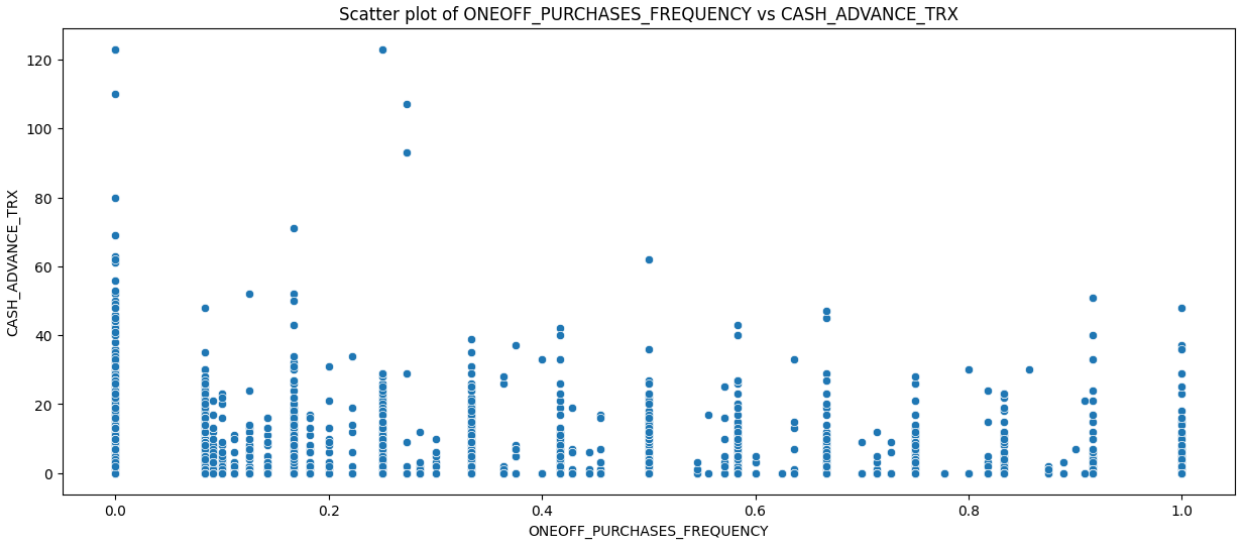

Scatter plot of PURCHASES_FREQUENCY vs PAYMENTS
PAYMENTS
PURCHASES_FREQUENCY
0
10000
20000
30000
40000
50000
0.0
0.2
0.4
0.6
0.8
1.0

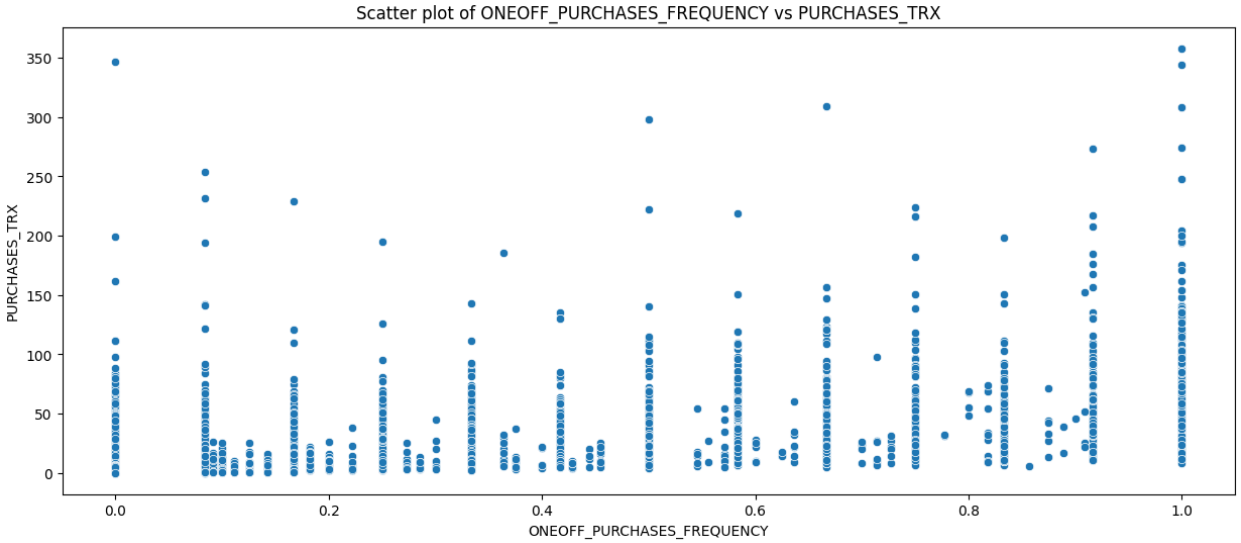

Scatter plot of PURCHASES_FREQUENCY vs MINIMUM_PAYMENTS
MINIMUM_PAYMENTS
PURCHASES_FREQUENCY
0
10000
20000
30000
40000
50000
60000
70000
80000
0.0
0.2
0.4
0.6
0.8
1.0

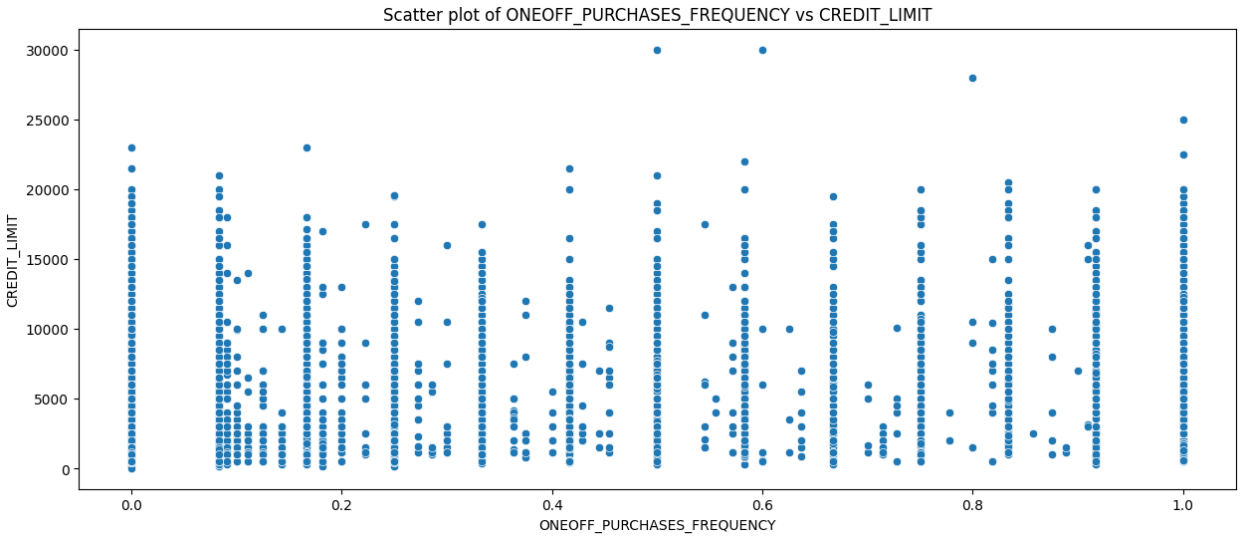

Scatter plot of PURCHASES_FREQUENCY vs PRC_FULL_PAYMENT
PRC_FULL_PAYMENT
PURCHASES_FREQUENCY
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0

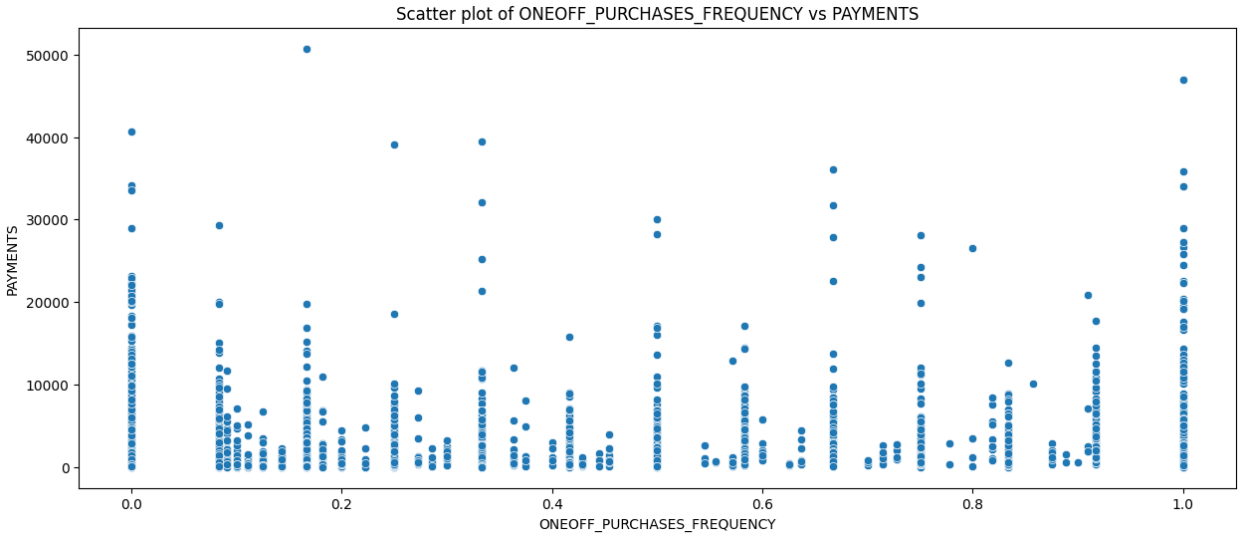

Scatter plot of ONEOFF_PURCHASES_FREQUENCY vs PURCHASES_INSTALLMENTS_FREQUENCY
PURCHASES_INSTALLMENTS_FREQUENCY
ONEOFF_PURCHASES_FREQUENCY
0.0
0.2
0.4
0.6
0.8
1.0

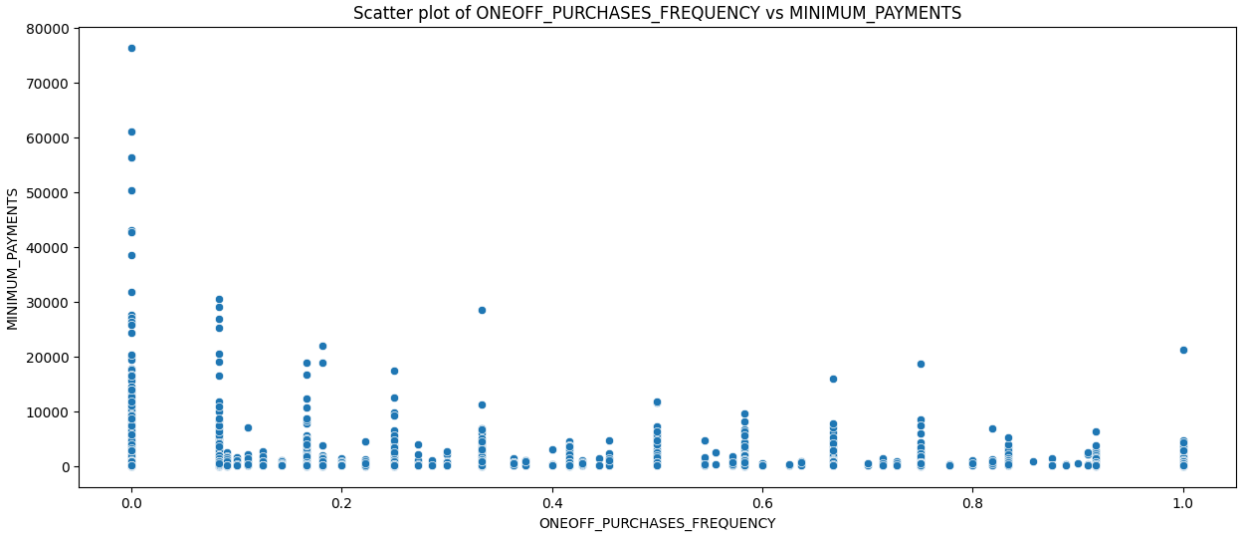

Scatter plot of ONEOFF_PURCHASES_FREQUENCY vs CASH_ADVANCE_FREQUENCY
CASH_ADVANCE_FREQUENCY
ONEOFF_PURCHASES_FREQUENCY
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4

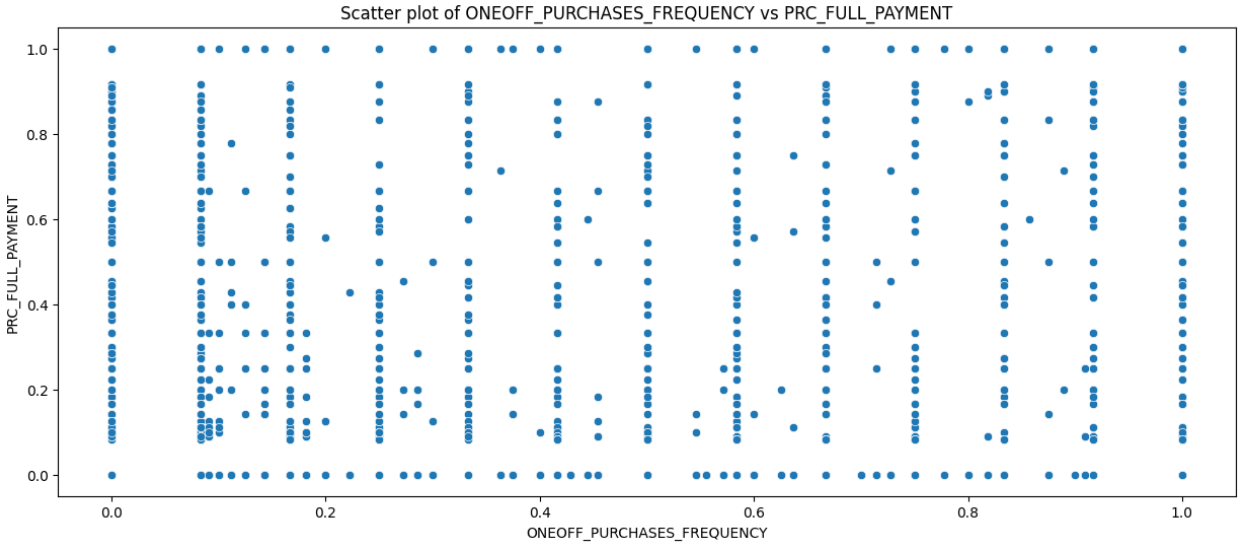

Scatter plot of ONEOFF_PURCHASES_FREQUENCY vs CASH_ADVANCE_TRX
CASH_ADVANCE_TRX
ONEOFF_PURCHASES_FREQUENCY
0
20
40
60
80
100
120
0.0
0.2
0.4
0.6
0.8
1.0

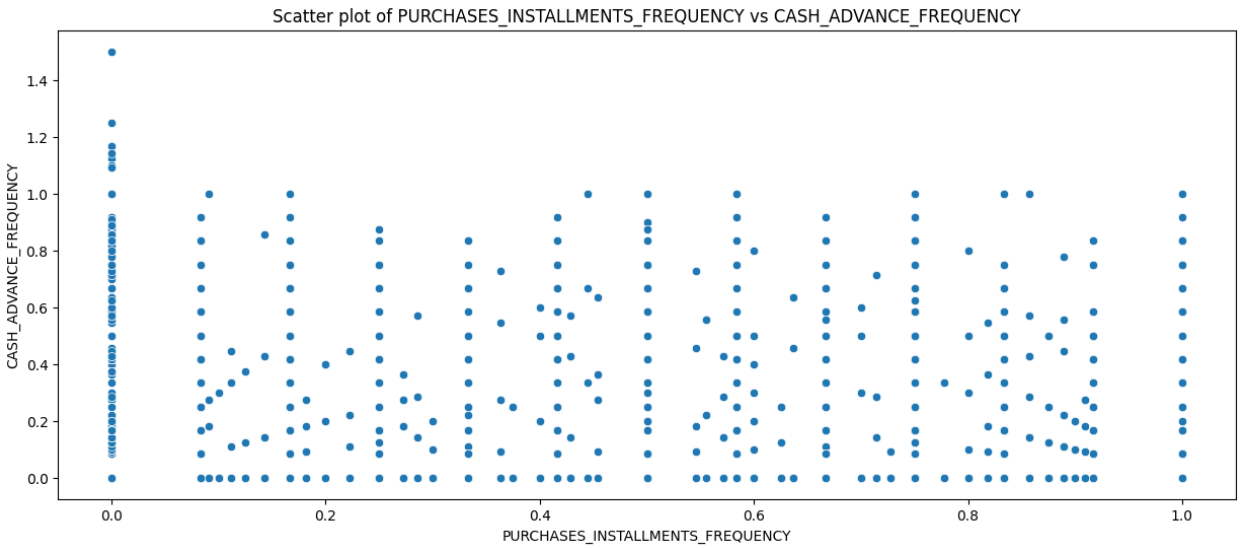

Scatter plot of ONEOFF_PURCHASES_FREQUENCY vs PURCHASES_TRX
PURCHASES_TRX
ONEOFF_PURCHASES_FREQUENCY
0
50
100
150
200
250
300
350
0.0
0.2
0.4
0.6
0.8
1.0

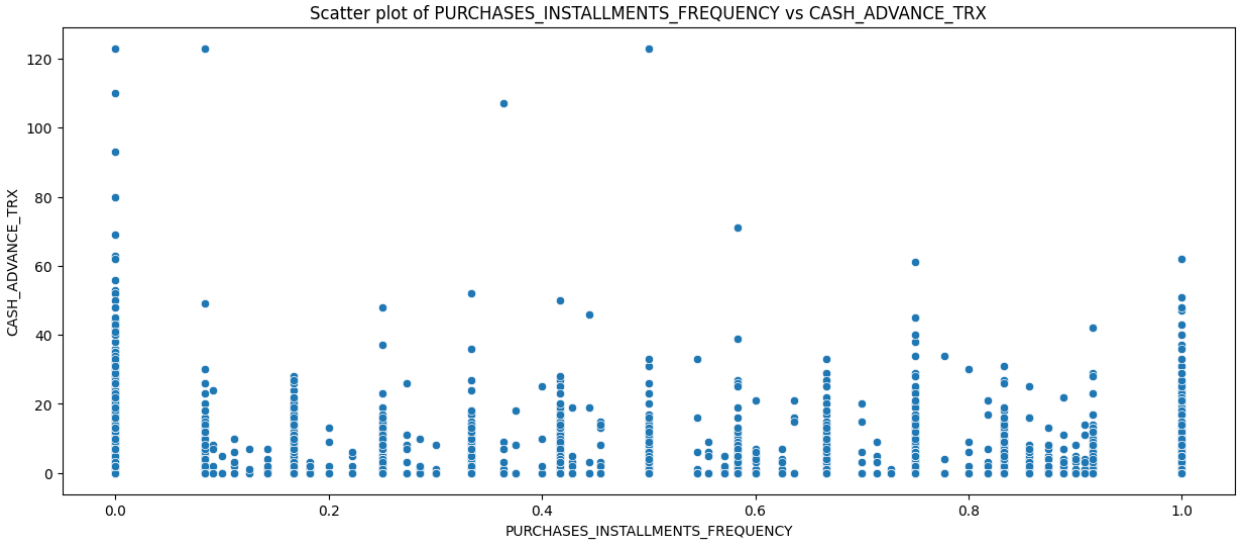

Scatter plot of ONEOFF_PURCHASES_FREQUENCY vs CREDIT_LIMIT
CREDIT_LIMIT
ONEOFF_PURCHASES_FREQUENCY
0
5000
10000
15000
20000
25000
30000
0.0
0.2
0.4
0.6
0.8
1.0

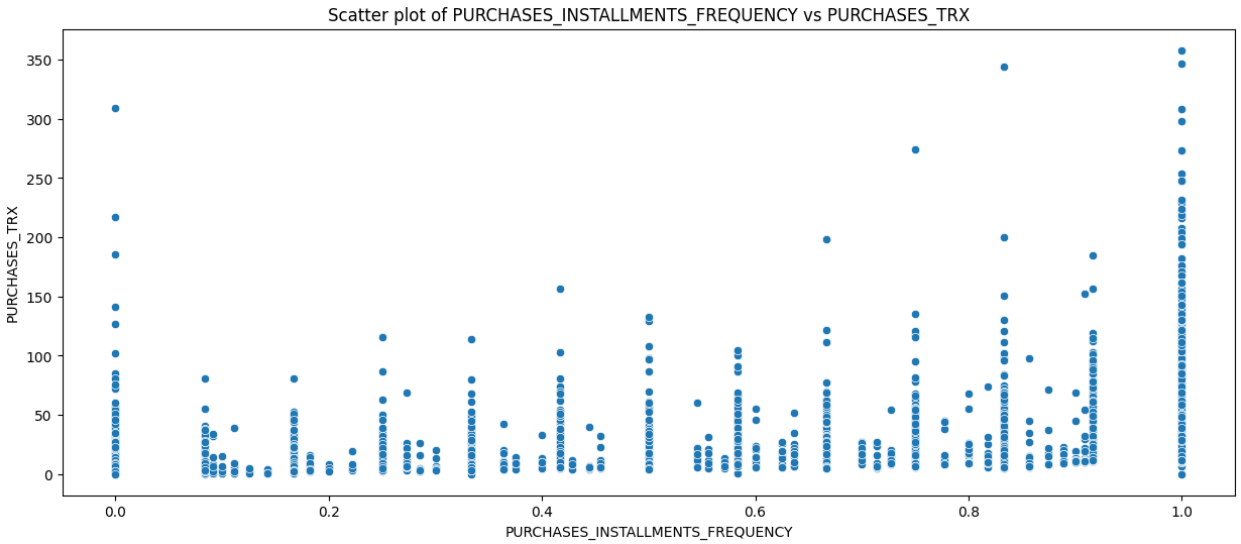

Scatter plot of ONEOFF_PURCHASES_FREQUENCY vs PAYMENTS
PAYMENTS
ONEOFF_PURCHASES_FREQUENCY
0
10000
20000
30000
40000
50000
0.0
0.2
0.4
0.6
0.8
1.0

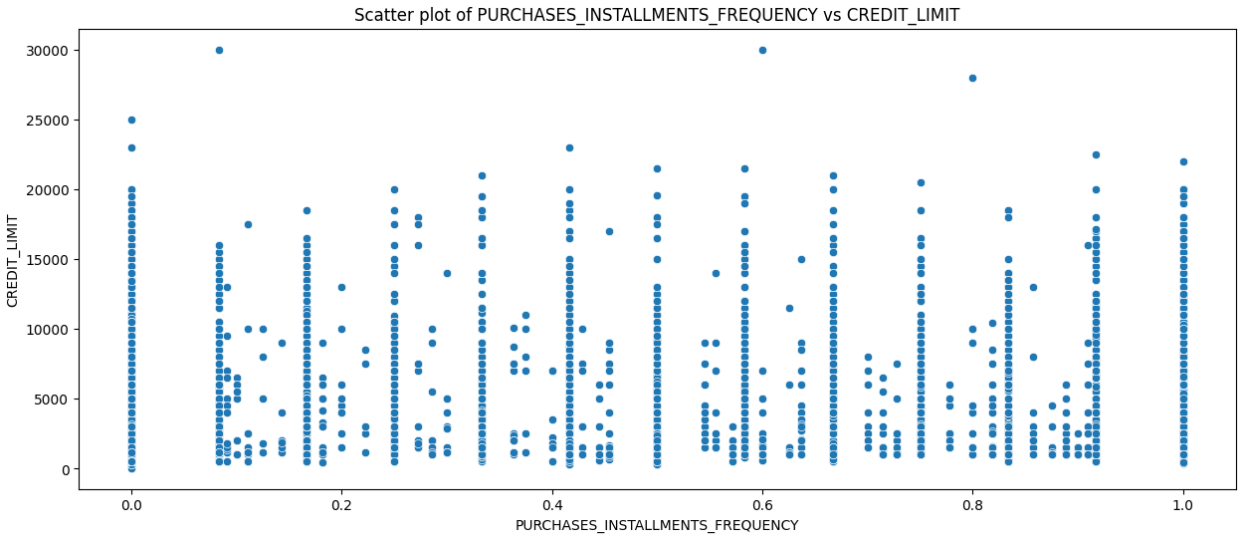

Scatter plot of ONEOFF_PURCHASES_FREQUENCY vs MINIMUM_PAYMENTS
MINIMUM_PAYMENTS
ONEOFF_PURCHASES_FREQUENCY
0
10000
20000
30000
40000
50000
60000
70000
80000
0.0
0.2
0.4
0.6
0.8
1.0

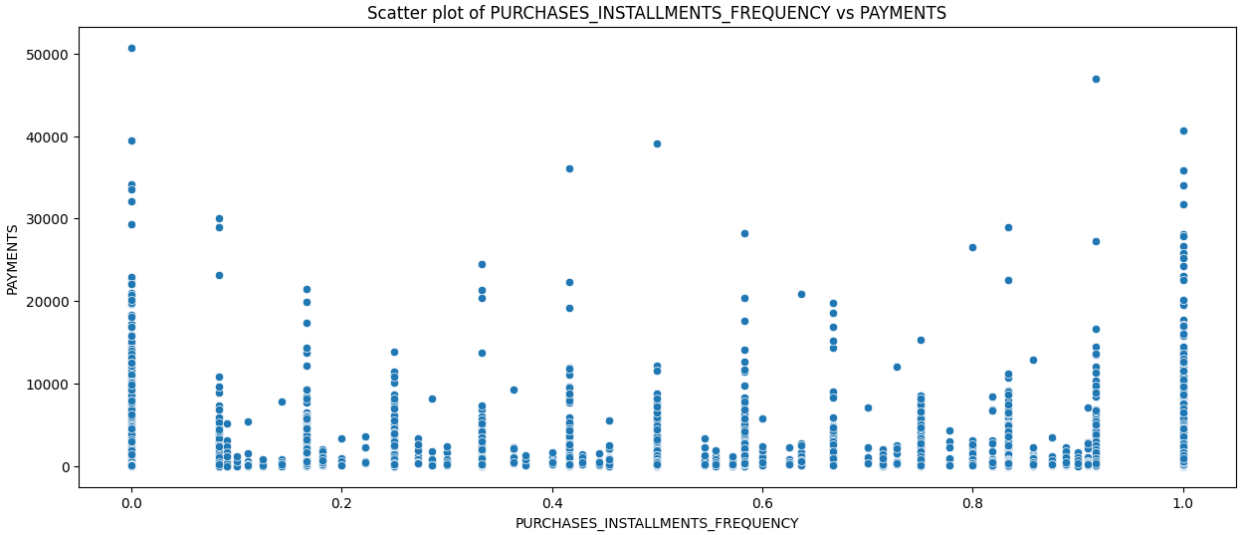

Scatter plot of ONEOFF_PURCHASES_FREQUENCY vs PRC_FULL_PAYMENT
PRC_FULL_PAYMENT
ONEOFF_PURCHASES_FREQUENCY
0.0
0.2
0.4
0.6
0.8
1.0

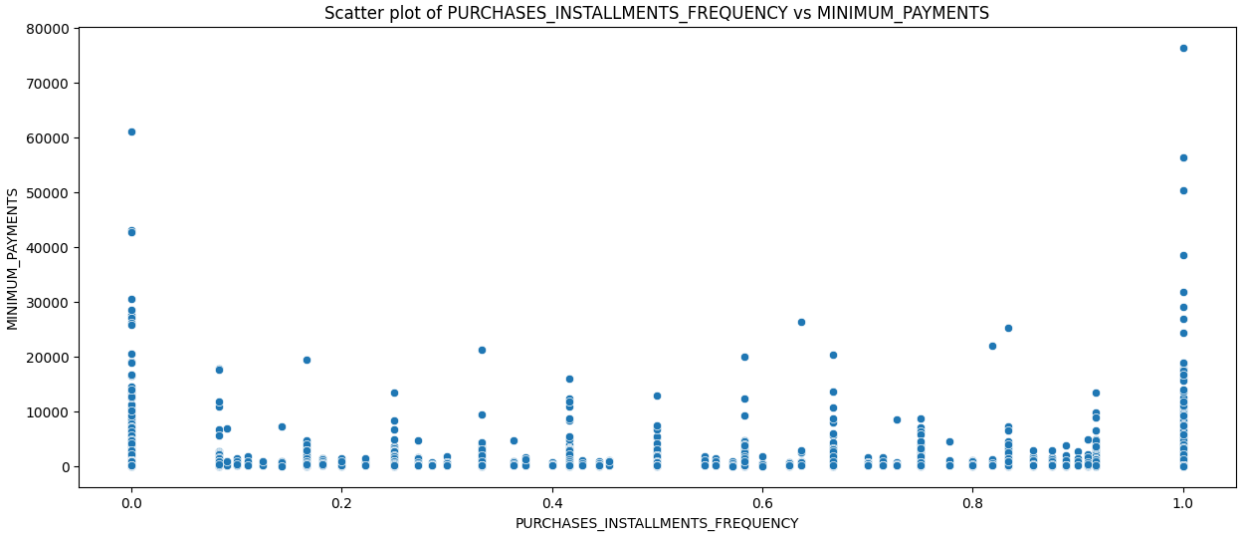

Scatter plot of PURCHASES_INSTALLMENTS_FREQUENCY vs CASH_ADVANCE_FREQUENCY
CASH_ADVANCE_FREQUENCY
PURCHASES_INSTALLMENTS_FREQUENCY
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4

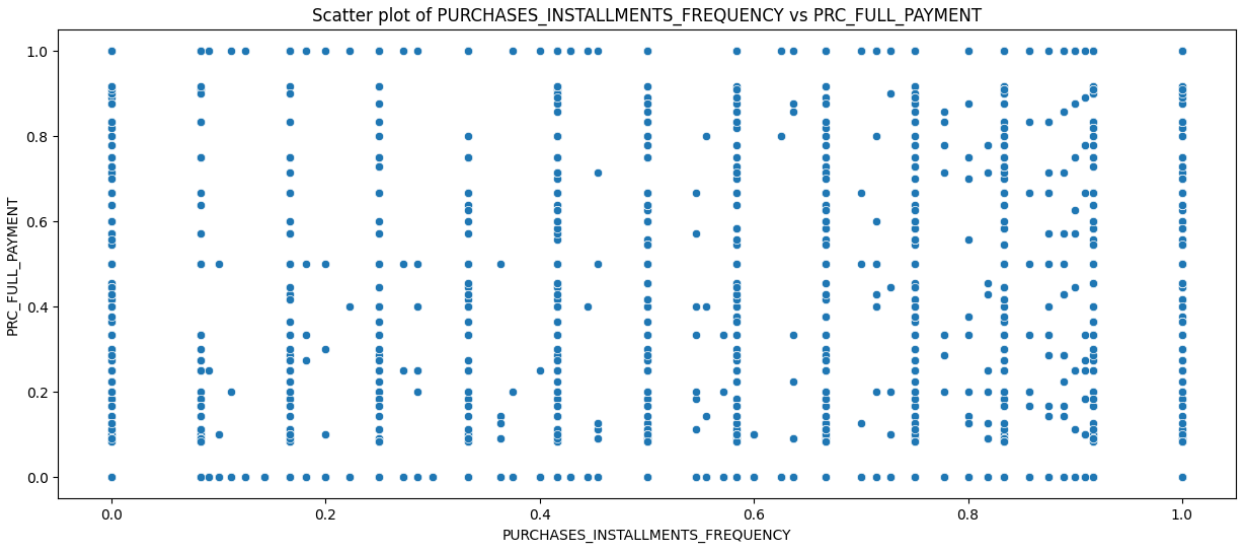

Scatter plot of PURCHASES_INSTALLMENTS_FREQUENCY vs CASH_ADVANCE_TRX
CASH_ADVANCE_TRX
PURCHASES_INSTALLMENTS_FREQUENCY
0
20
40
60
80
100
120
0.0
0.2
0.4
0.6
0.8
1.0

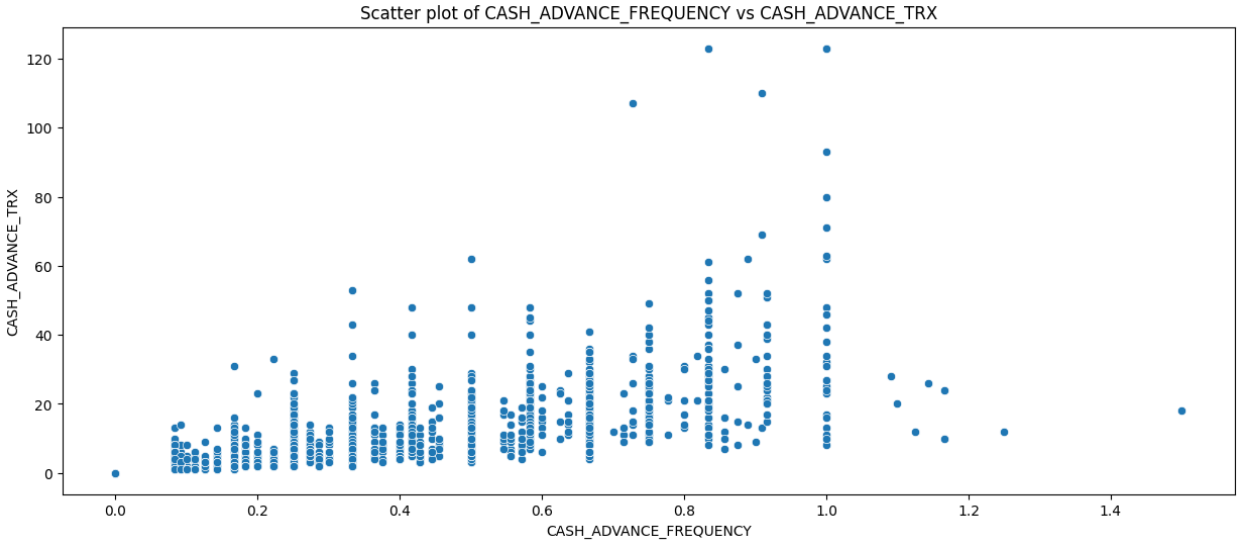

Scatter plot of PURCHASES_INSTALLMENTS_FREQUENCY vs PURCHASES_TRX
PURCHASES_TRX
PURCHASES_INSTALLMENTS_FREQUENCY
0
50
100
150
200
250
300
350
0.0
0.2
0.4
0.6
0.8
1.0

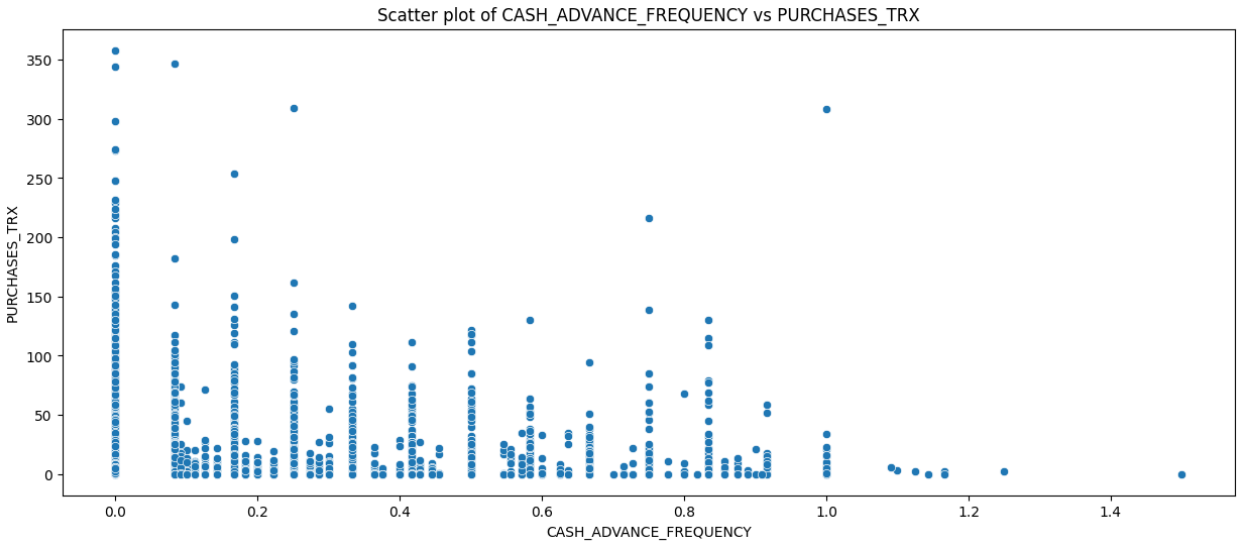

Scatter plot of PURCHASES_INSTALLMENTS_FREQUENCY vs CREDIT_LIMIT
CREDIT_LIMIT
PURCHASES_INSTALLMENTS_FREQUENCY
0
5000
10000
15000
20000
25000
30000
0.0
0.2
0.4
0.6
0.8
1.0

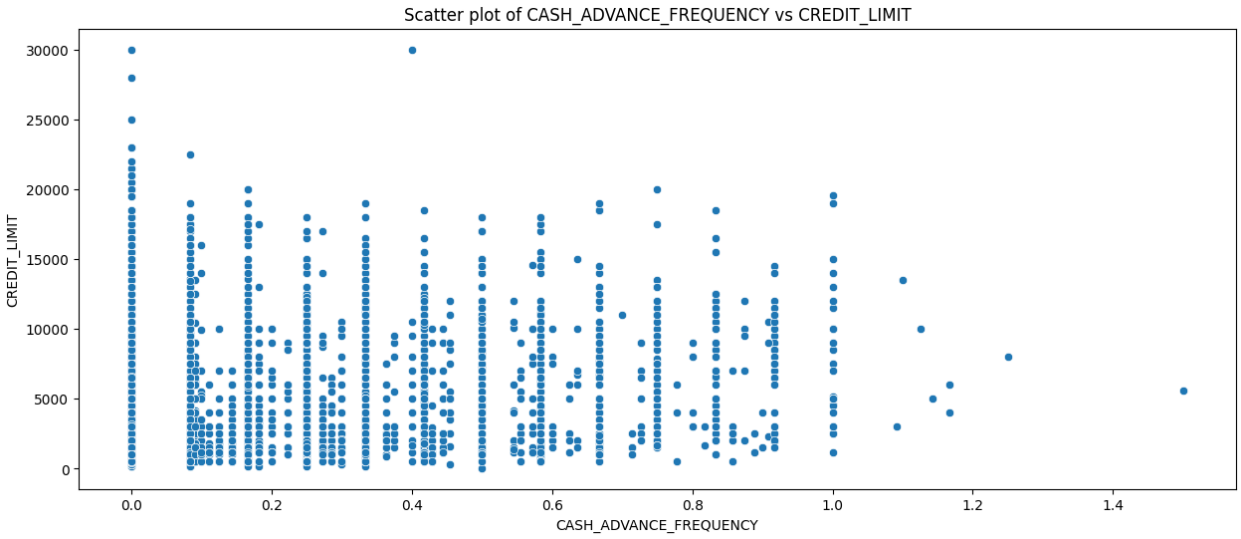

Scatter plot of PURCHASES_INSTALLMENTS_FREQUENCY vs PAYMENTS
PAYMENTS
PURCHASES_INSTALLMENTS_FREQUENCY
0
10000
20000
30000
40000
50000
0.0
0.2
0.4
0.6
0.8
1.0

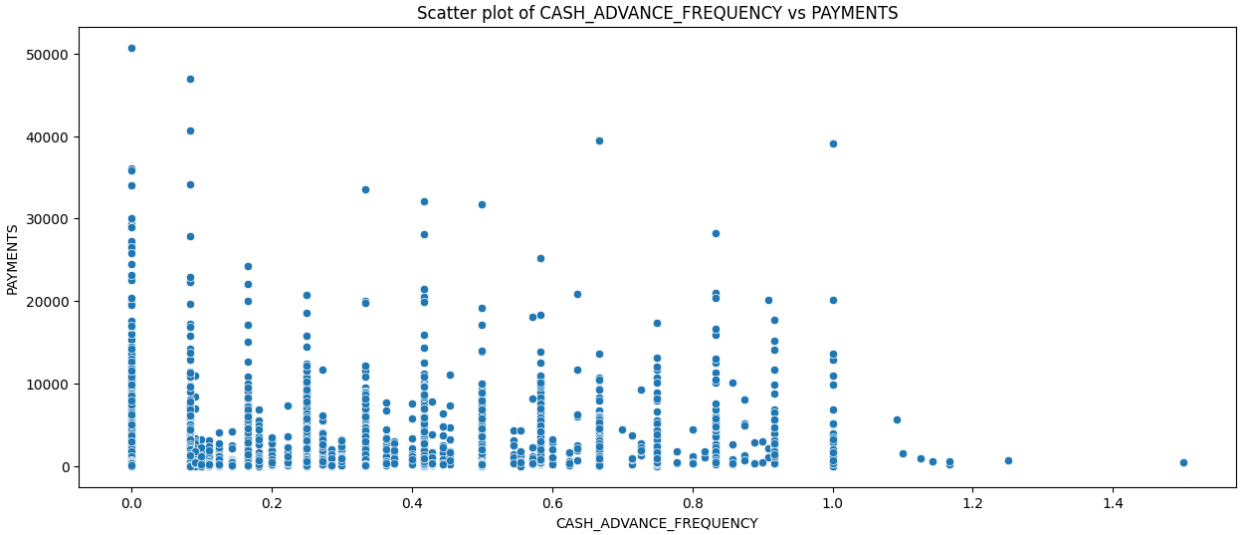

Scatter plot of PURCHASES_INSTALLMENTS_FREQUENCY vs MINIMUM_PAYMENTS
MINIMUM_PAYMENTS
PURCHASES_INSTALLMENTS_FREQUENCY
0
10000
20000
30000
40000
50000
60000
70000
80000
0.0
0.2
0.4
0.6
0.8
1.0

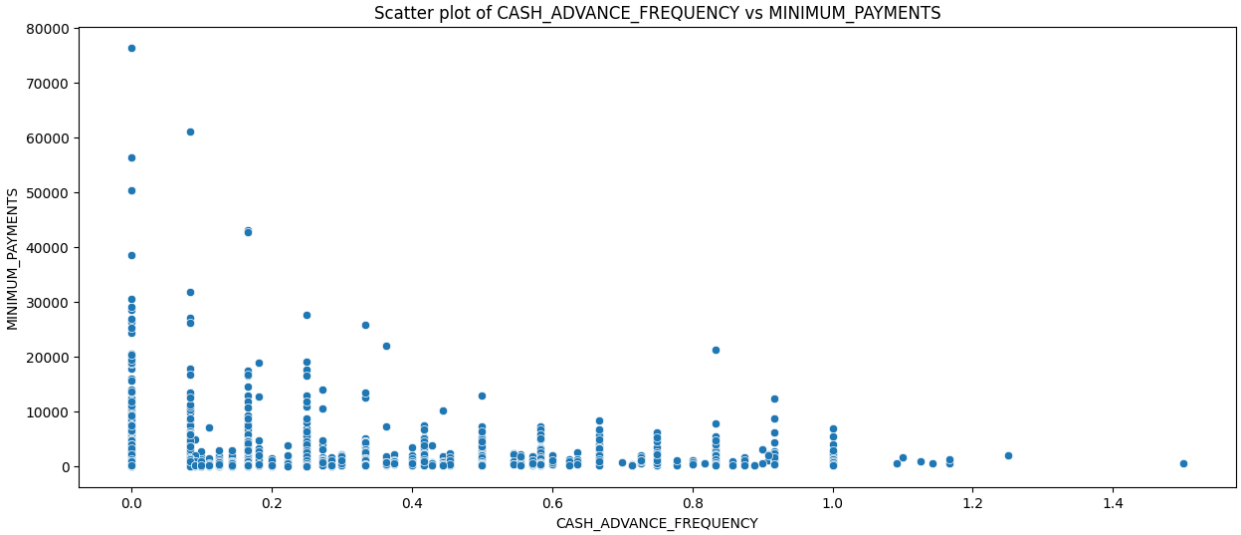

Scatter plot of PURCHASES_INSTALLMENTS_FREQUENCY vs PRC_FULL_PAYMENT
PRC_FULL_PAYMENT
PURCHASES_INSTALLMENTS_FREQUENCY
0.0
0.2
0.4
0.6
0.8
1.0

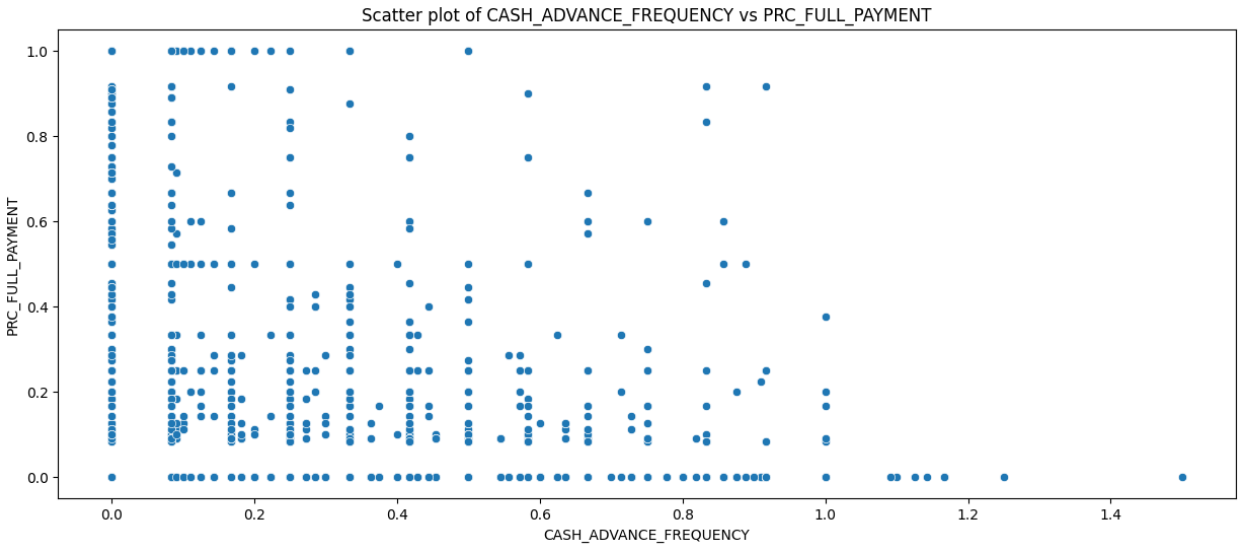

Scatter plot of CASH_ADVANCE_FREQUENCY vs CASH_ADVANCE_TRX
CASH_ADVANCE_TRX
CASH_ADVANCE_FREQUENCY
0
20
40
60
80
100
120
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4

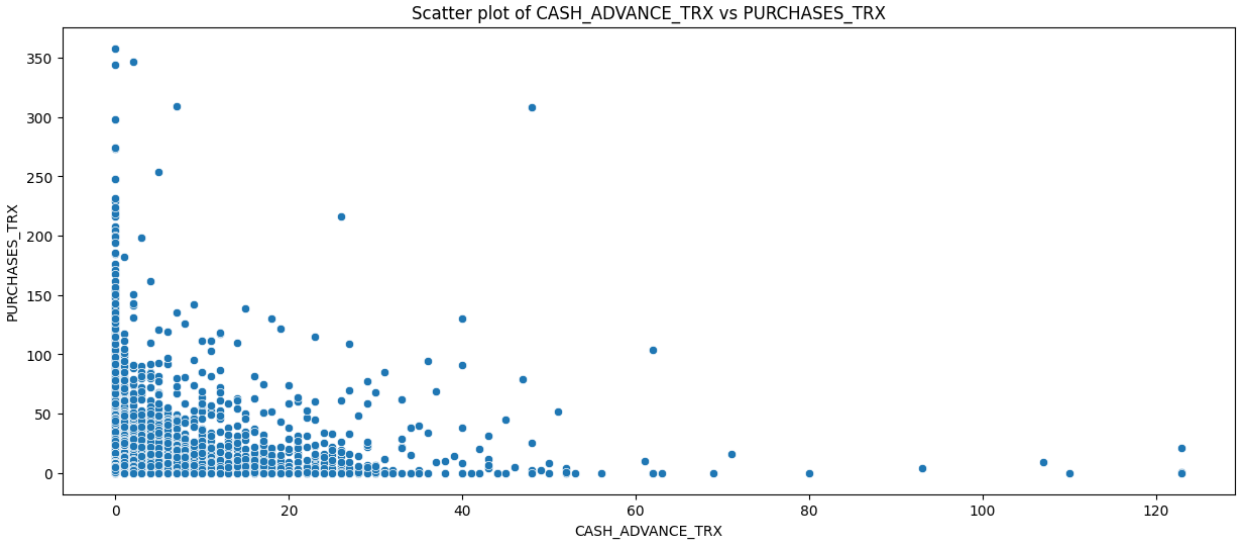

Scatter plot of CASH_ADVANCE_FREQUENCY vs PURCHASES_TRX
PURCHASES_TRX
CASH_ADVANCE_FREQUENCY
0
50
100
150
200
250
300
350
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4

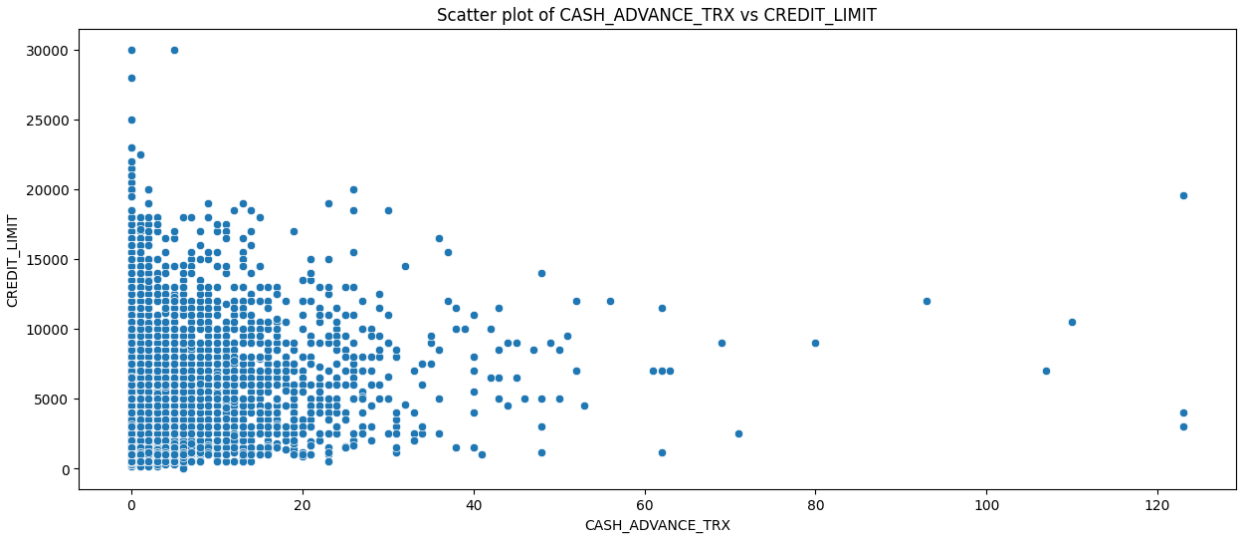

Scatter plot of CASH_ADVANCE_FREQUENCY vs CREDIT_LIMIT
CREDIT_LIMIT
CASH_ADVANCE_FREQUENCY
0
5000
10000
15000
20000
25000
30000
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4

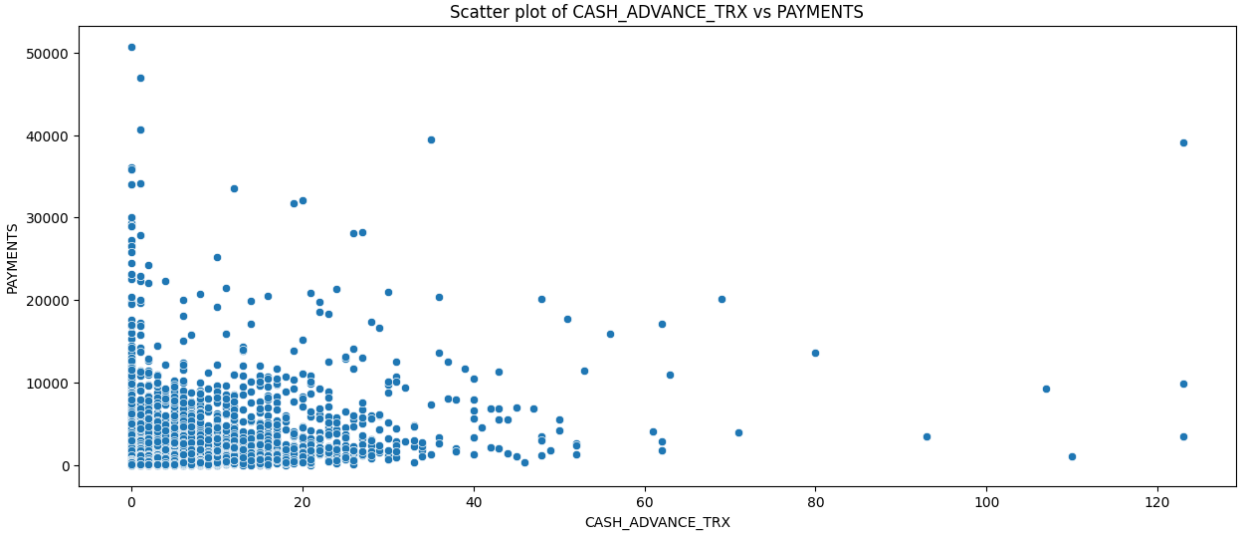

Scatter plot of CASH_ADVANCE_FREQUENCY vs PAYMENTS
PAYMENTS
CASH_ADVANCE_FREQUENCY
0
10000
20000
30000
40000
50000
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4

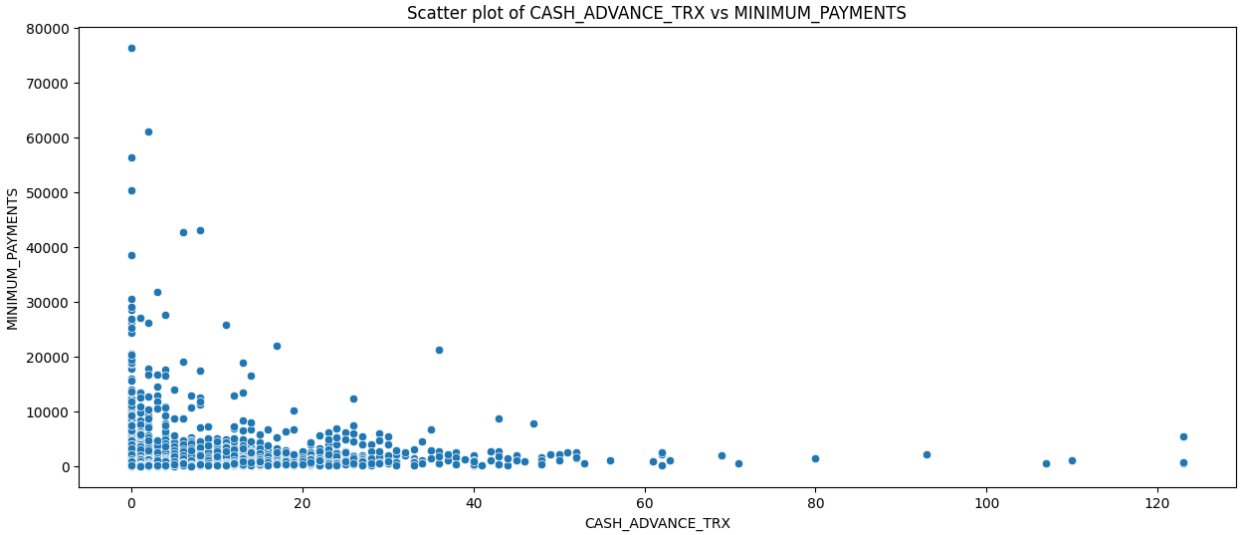

Scatter plot of CASH_ADVANCE_FREQUENCY vs MINIMUM_PAYMENTS
MINIMUM_PAYMENTS
CASH_ADVANCE_FREQUENCY
0
10000
20000
30000
40000
50000
60000
70000
80000
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4

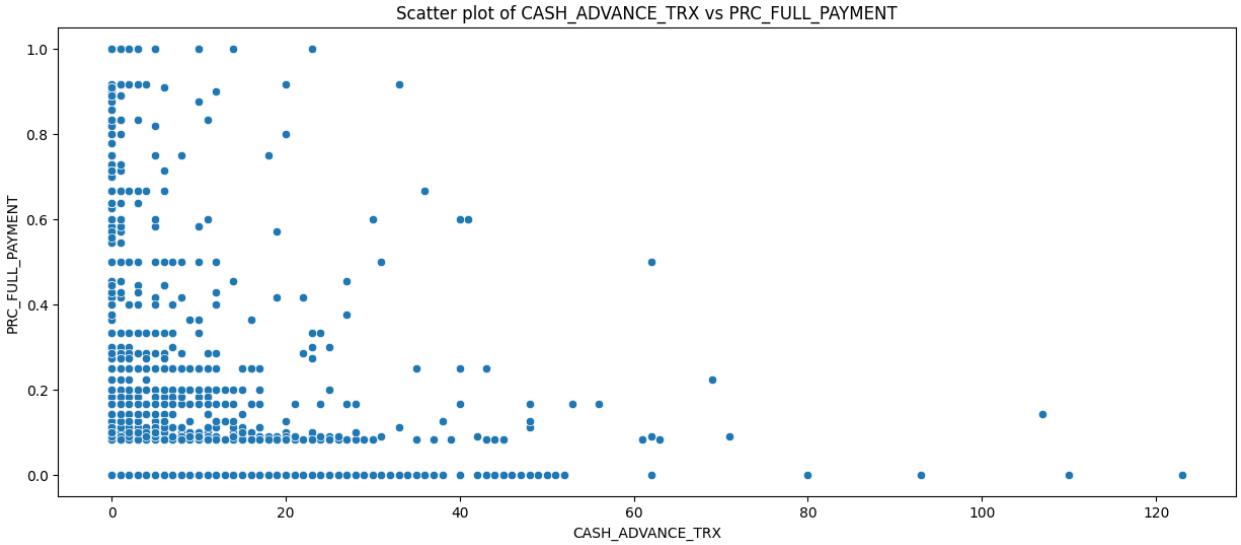

Scatter plot of CASH_ADVANCE_FREQUENCY vs PRC_FULL_PAYMENT
PRC_FULL_PAYMENT
CASH_ADVANCE_FREQUENCY
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4

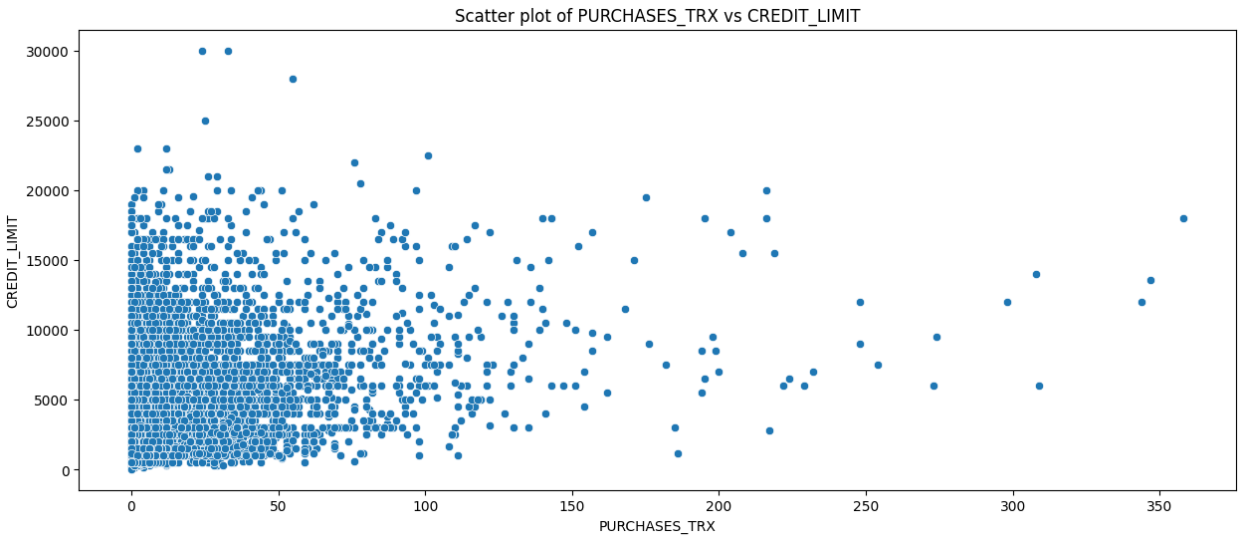

Scatter plot of CASH_ADVANCE_TRX vs PURCHASES_TRX
PURCHASES_TRX
CASH_ADVANCE_TRX
0
50
100
150
200
250
300
350
0
20
40
60
80
100
120

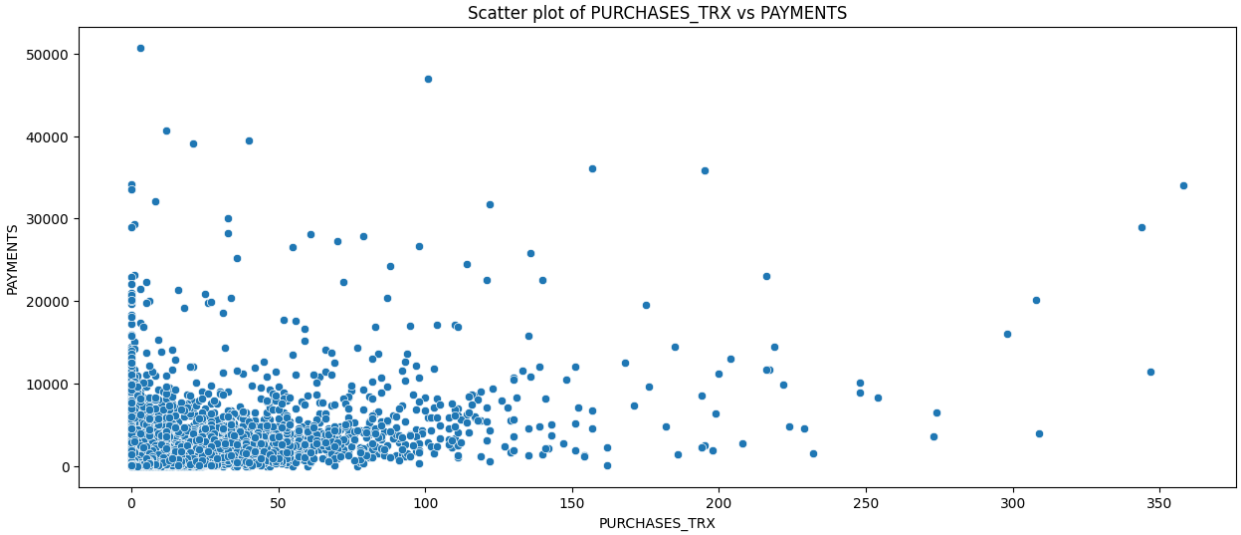

Scatter plot of CASH_ADVANCE_TRX vs CREDIT_LIMIT
CREDIT_LIMIT
CASH_ADVANCE_TRX
0
5000
10000
15000
20000
25000
30000
0
20
40
60
80
100
120

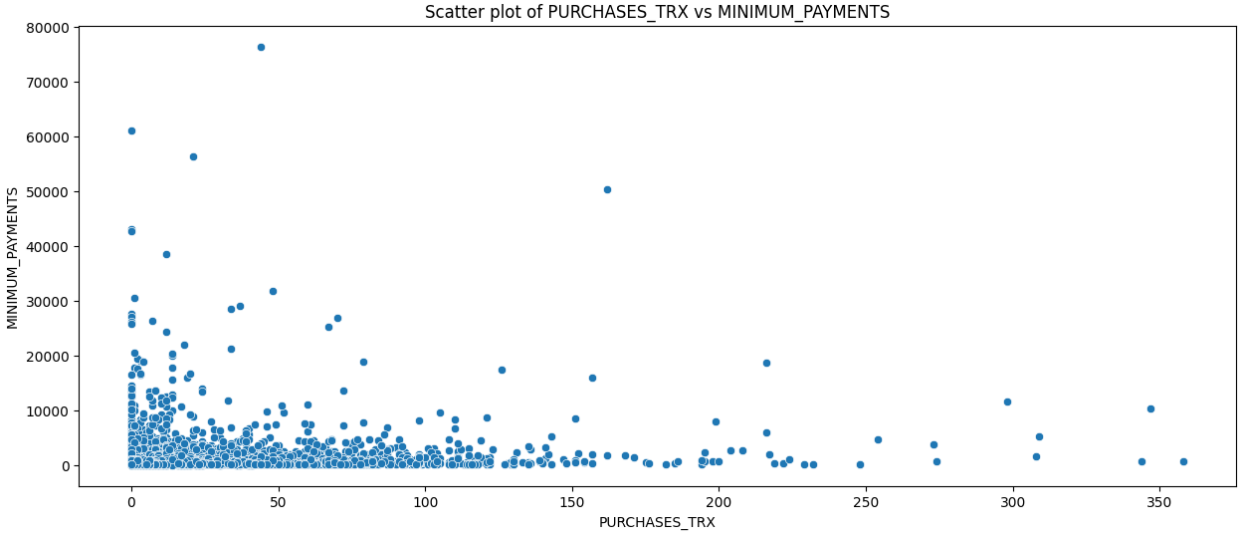

Scatter plot of CASH_ADVANCE_TRX vs PAYMENTS
PAYMENTS
CASH_ADVANCE_TRX
0
10000
20000
30000
40000
50000
0
20
40
60
80
100
120

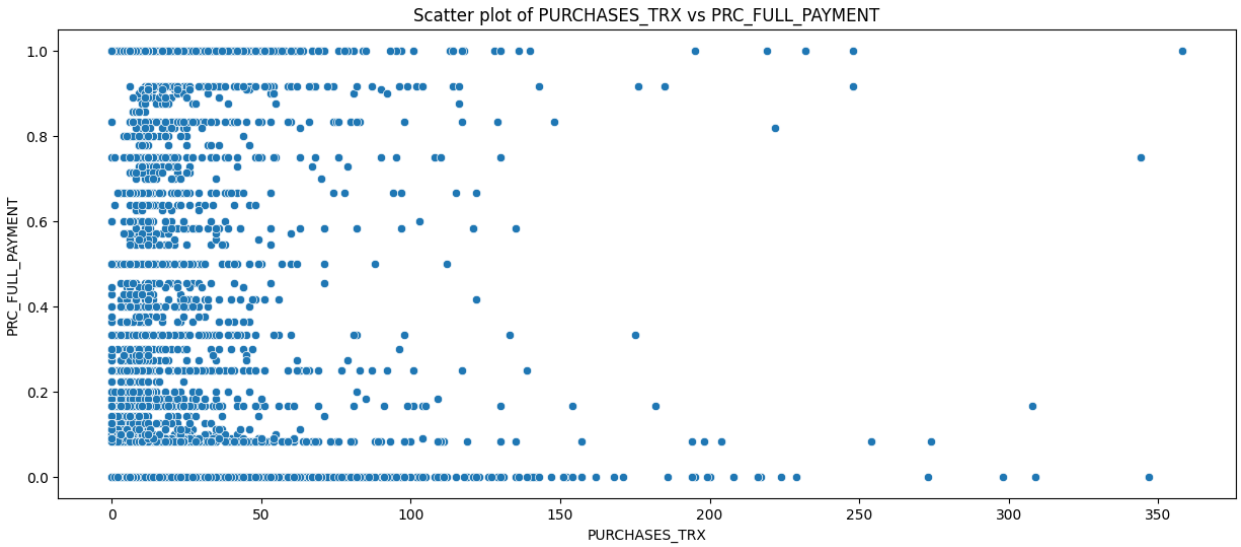

Scatter plot of CASH_ADVANCE_TRX vs MINIMUM_PAYMENTS
MINIMUM_PAYMENTS
CASH_ADVANCE_TRX
0
10000
20000
30000
40000
50000
60000
70000
80000
0
20
40
60
80
100
120

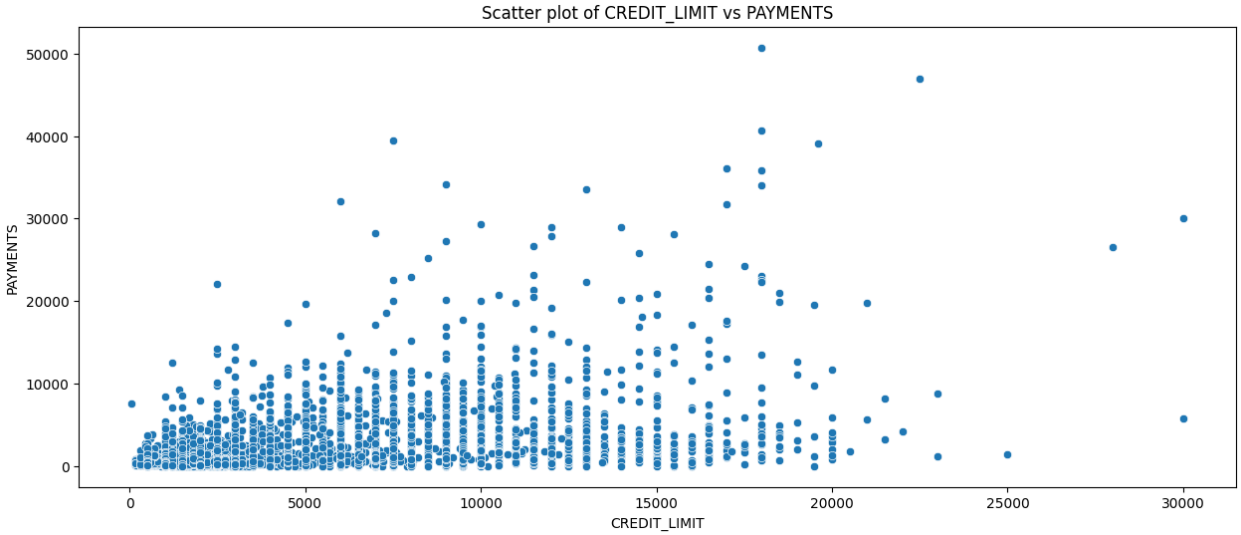

Scatter plot of CASH_ADVANCE_TRX vs PRC_FULL_PAYMENT
PRC_FULL_PAYMENT
CASH_ADVANCE_TRX
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
120

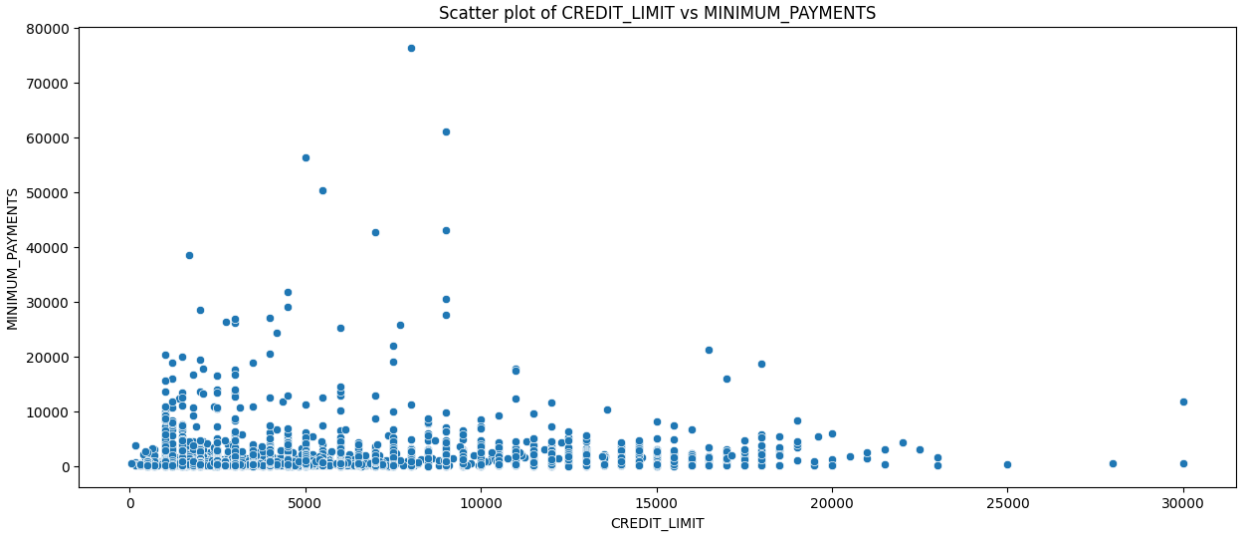

Scatter plot of PURCHASES_TRX vs CREDIT_LIMIT
CREDIT_LIMIT
PURCHASES_TRX
0
5000
10000
15000
20000
25000
30000
0
50
100
150
200
250
300
350

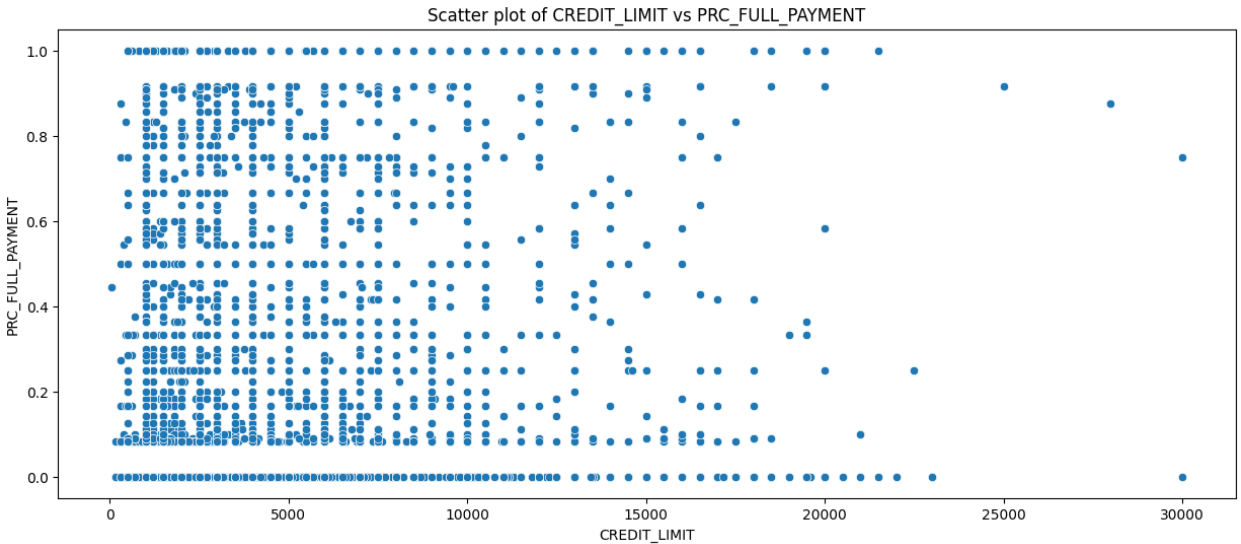

Scatter plot of PURCHASES_TRX vs PAYMENTS
PAYMENTS
PURCHASES_TRX
0
10000
20000
30000
40000
50000
0
50
100
150
200
250
300
350

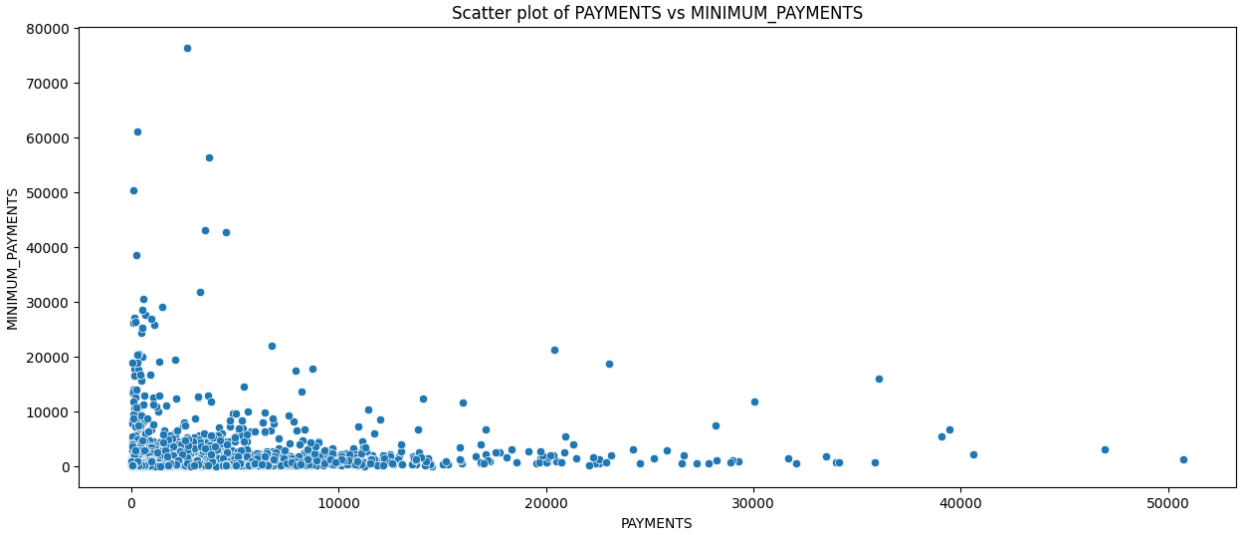

Scatter plot of PURCHASES_TRX vs MINIMUM_PAYMENTS
MINIMUM_PAYMENTS
PURCHASES_TRX
0
10000
20000
30000
40000
50000
60000
70000
80000
0
50
100
150
200
250
300
350

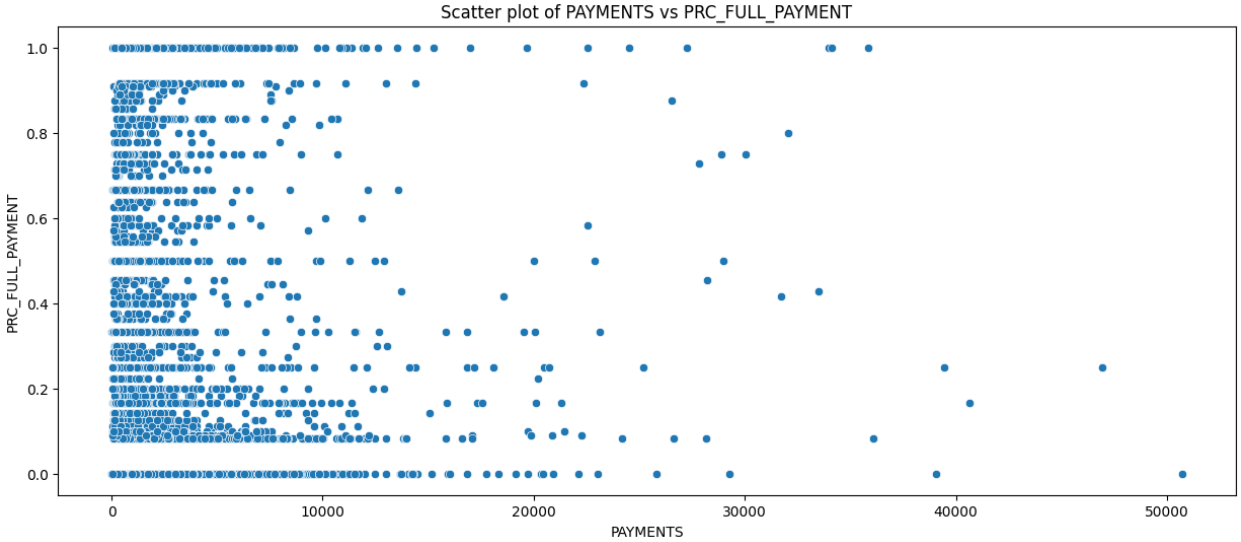

Scatter plot of PURCHASES_TRX vs PRC_FULL_PAYMENT
PRC_FULL_PAYMENT
PURCHASES_TRX
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200
250
300
350

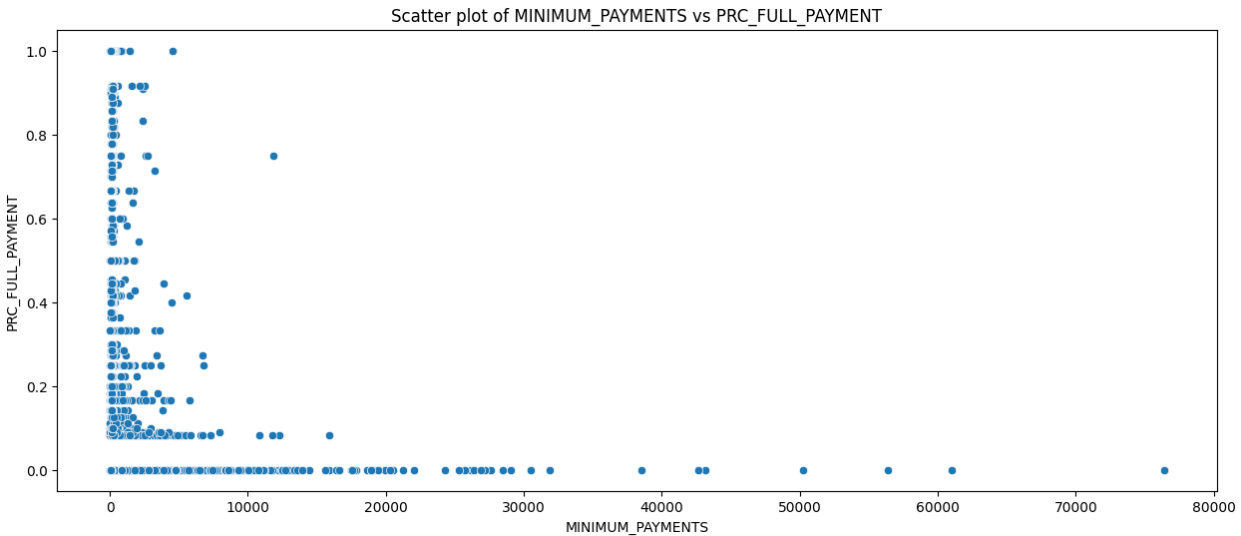

Scatter plot of CREDIT_LIMIT vs PAYMENTS
PAYMENTS
CREDIT_LIMIT
0
10000
20000
30000
40000
50000
0
5000
10000
15000
20000
25000
30000


Scatter plot of CREDIT_LIMIT vs MINIMUM_PAYMENTS
MINIMUM_PAYMENTS
CREDIT_LIMIT
0
10000
20000
30000
40000
50000
60000
70000
80000
0
5000
10000
15000
20000
25000
30000

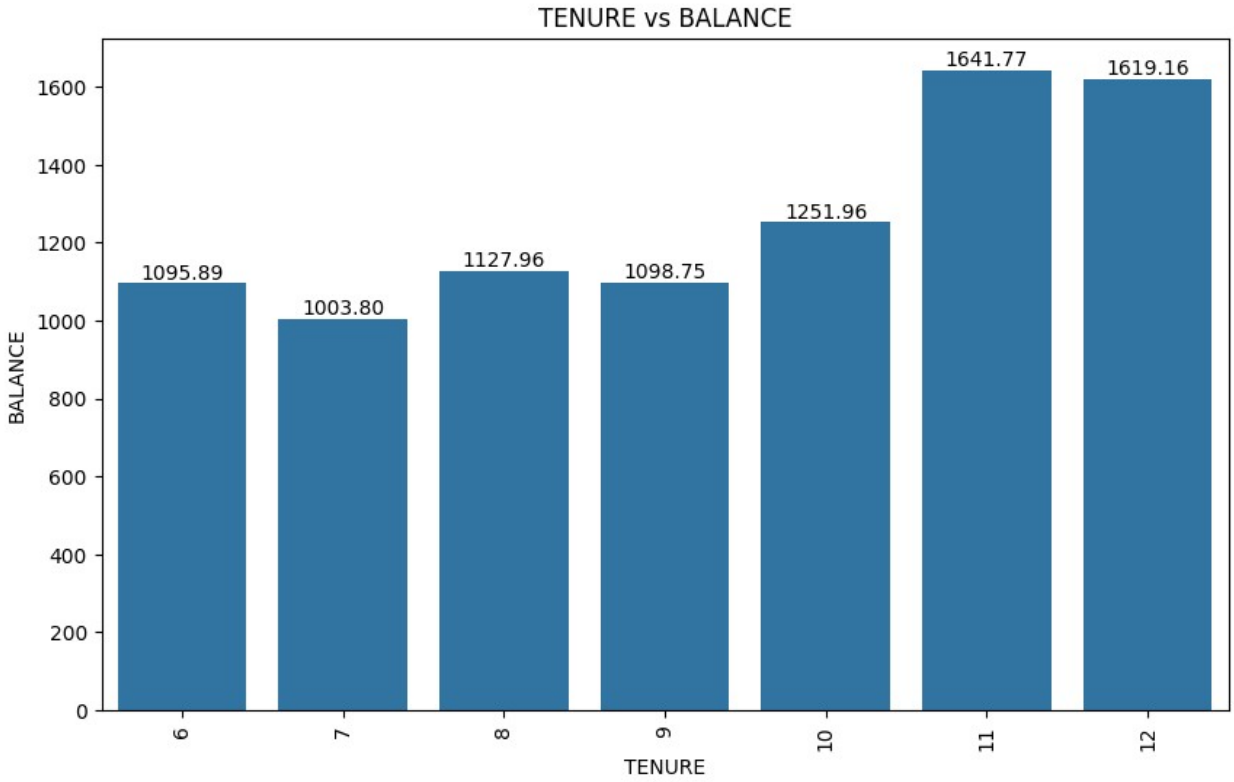

Scatter plot of CREDIT_LIMIT vs PRC_FULL_PAYMENT
PRC_FULL_PAYMENT
CREDIT_LIMIT
0.0
0.2
0.4
0.6
0.8
1.0
0
5000
10000
15000
20000
25000
30000

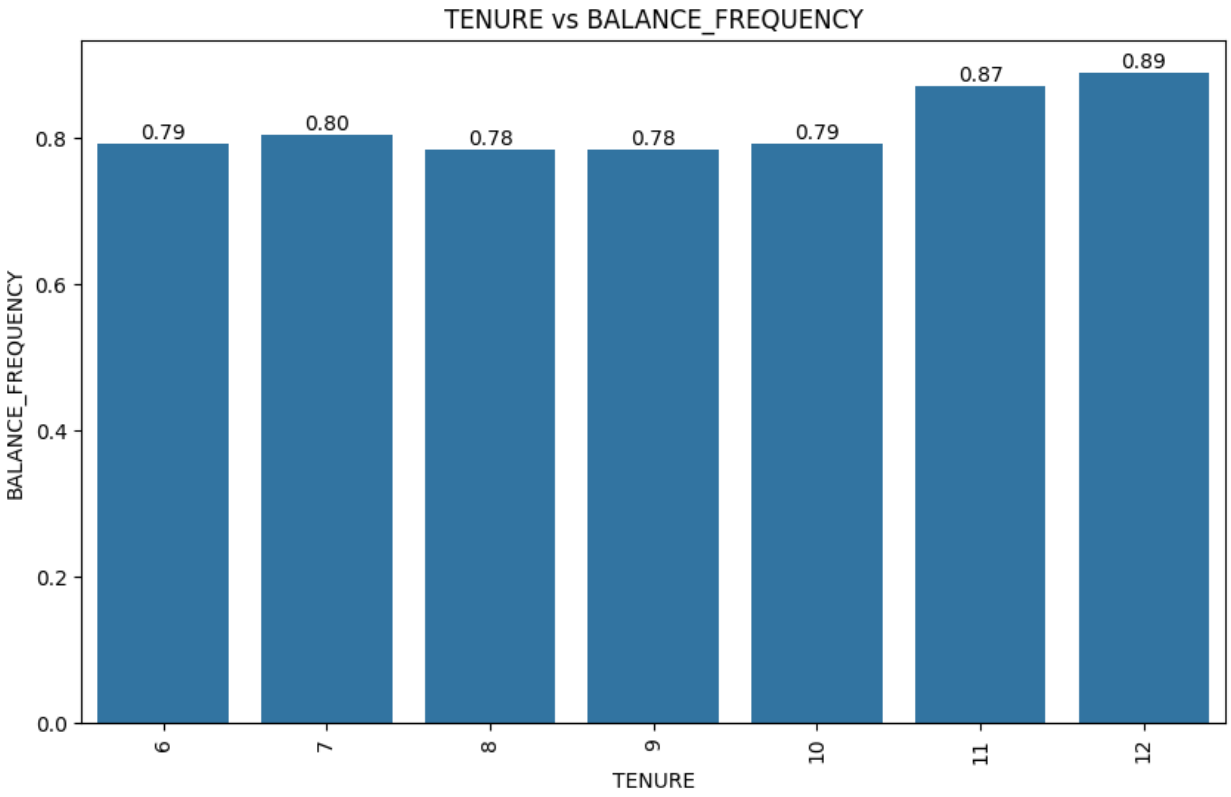

Scatter plot of PAYMENTS vs MINIMUM_PAYMENTS
MINIMUM_PAYMENTS
PAYMENTS
0
10000
20000
30000
40000
50000
60000
70000
80000

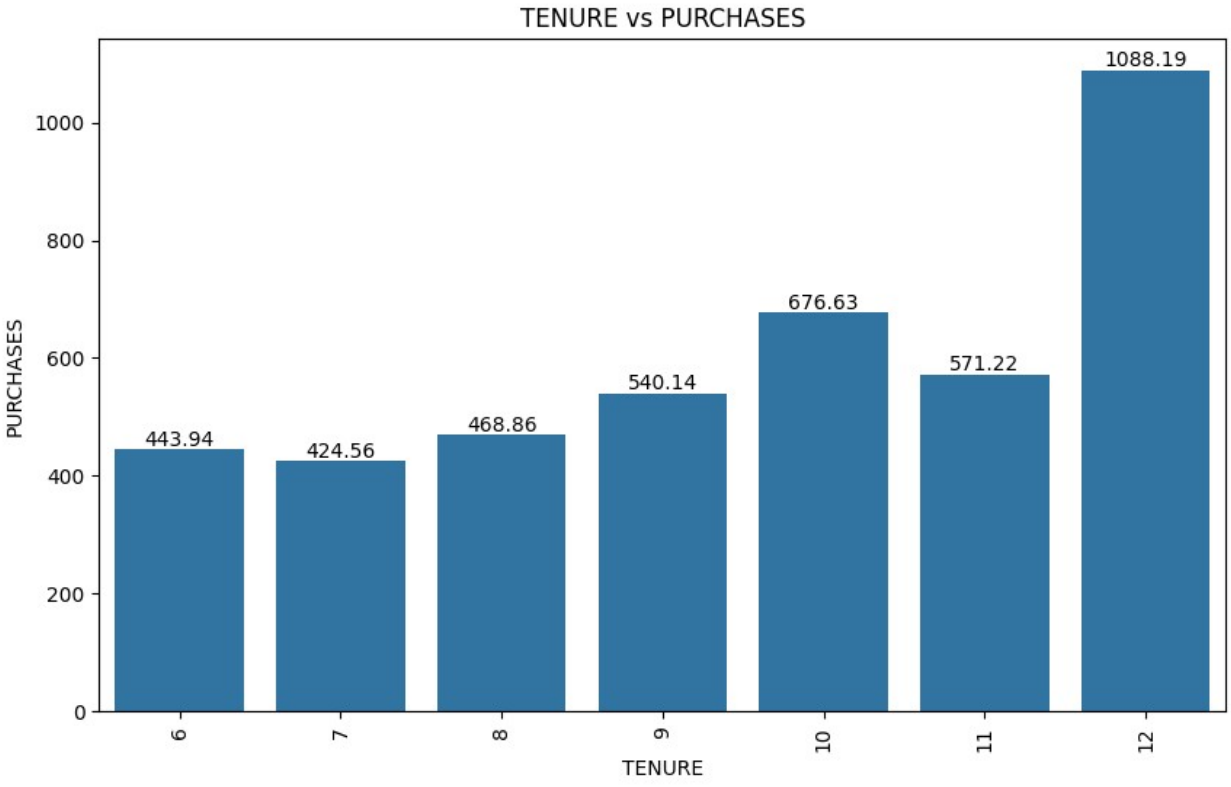

Scatter plot of PAYMENTS vs PRC_FULL_PAYMENT
PRC_FULL_PAYMENT
PAYMENTS
0.0
0.2
0.4
0.6
0.8
1.0
0
10000
20000
30000
40000
50000

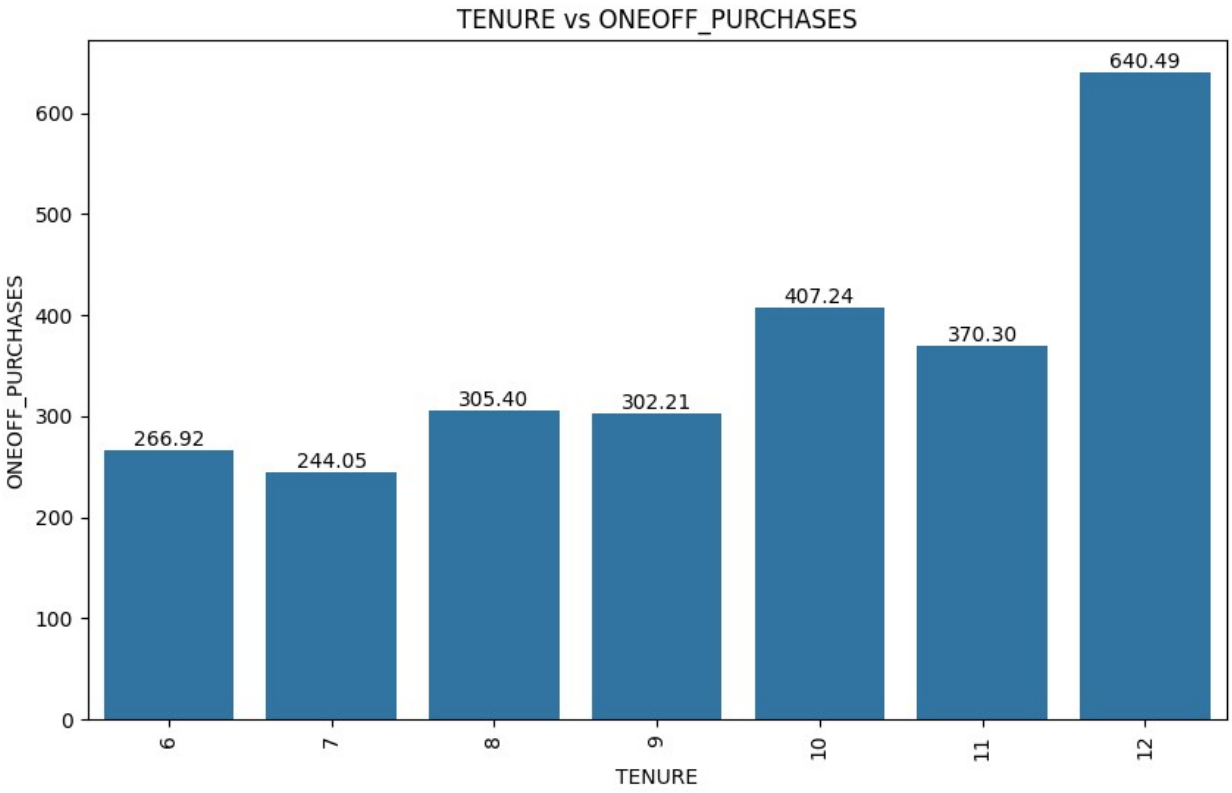

Scatter plot of MINIMUM_PAYMENTS vs PRC_FULL_PAYMENT
PRC_FULL_PAYMENT
MINIMUM_PAYMENTS
0.0
0.2
0.4
0.6
0.8
1.0
0
10000
20000
30000
40000
50000
60000
70000
80000

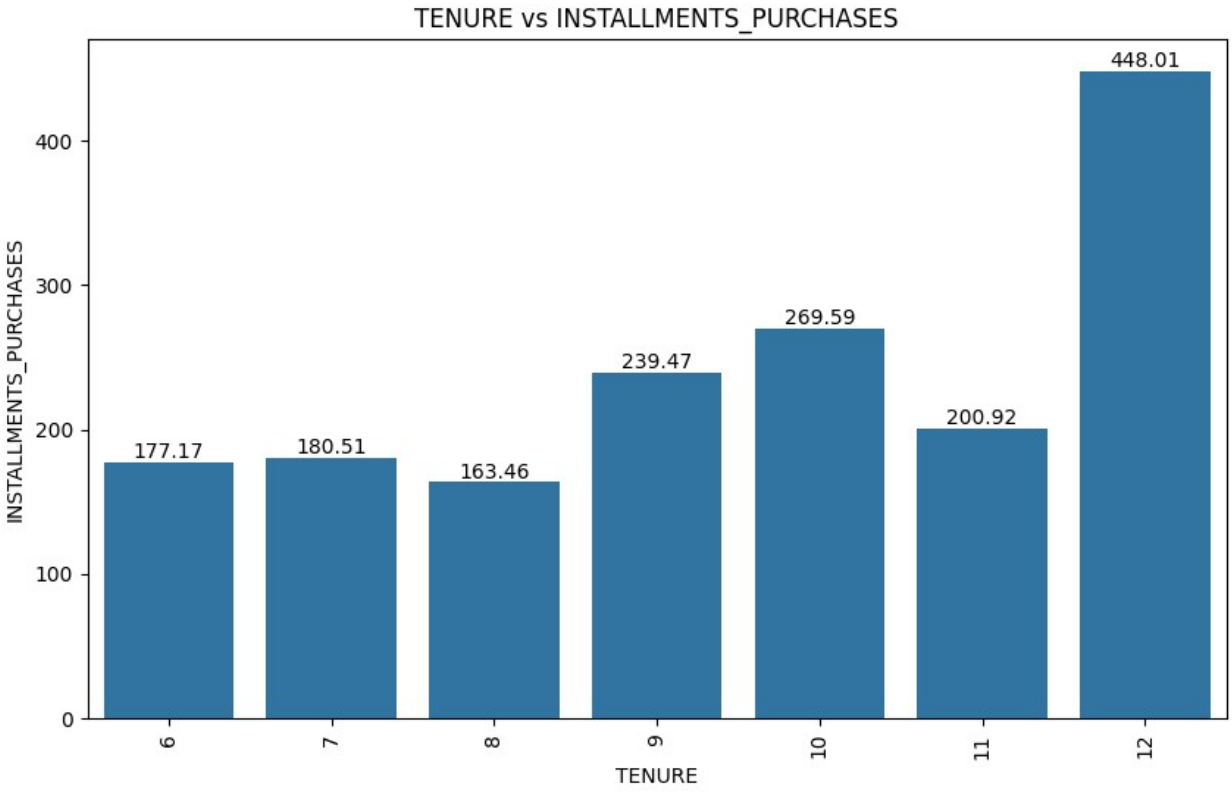
```
for dis in discrete:
    for cont in continuous:
        plt.figure(figsize=(10, 6))
        ax = sns.barplot(data=df, x=dis, y=cont, ci=None)
        plt.title(f'{dis} vs {cont}')

        for p in ax.patches:
            height = p.get_height()
            ax.annotate(f'{height:.2f}', (p.get_x() + p.get_width() /
2., height),
                        ha='center', va='bottom', fontsize=10,
color='black', rotation=0)
        plt.xticks(rotation = 90)
        plt.show()
```

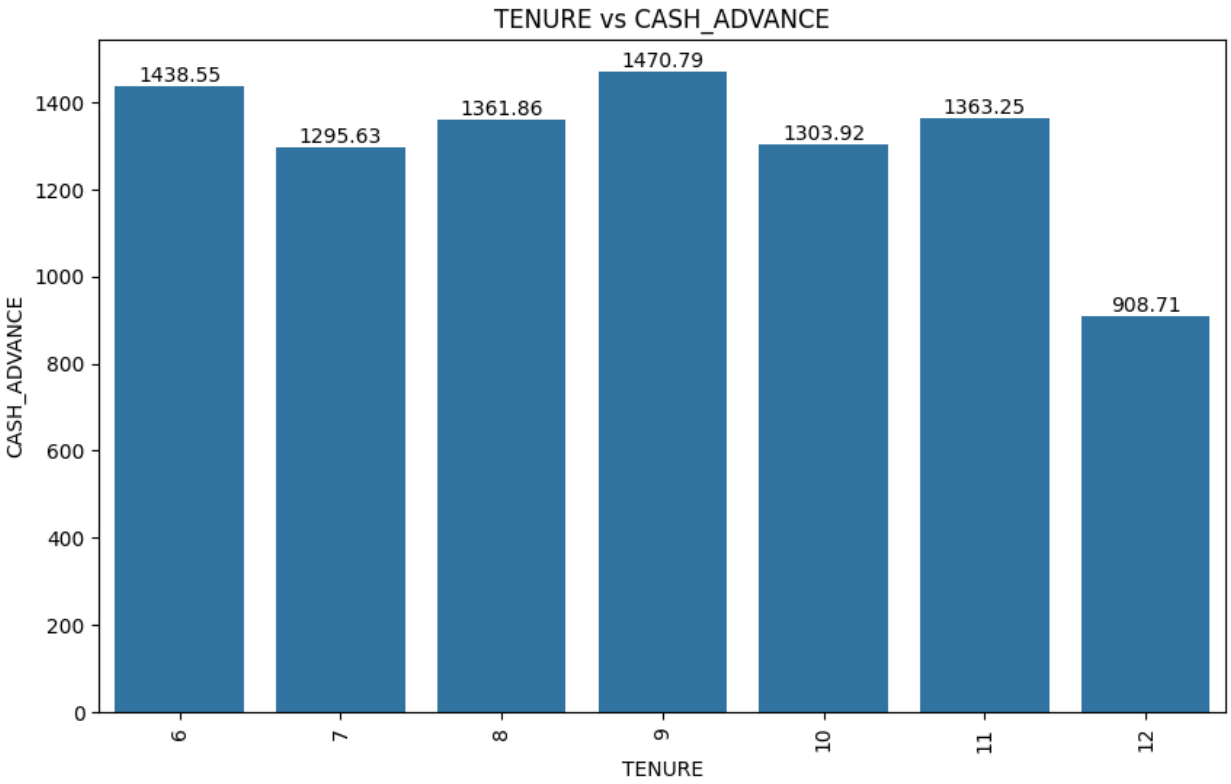


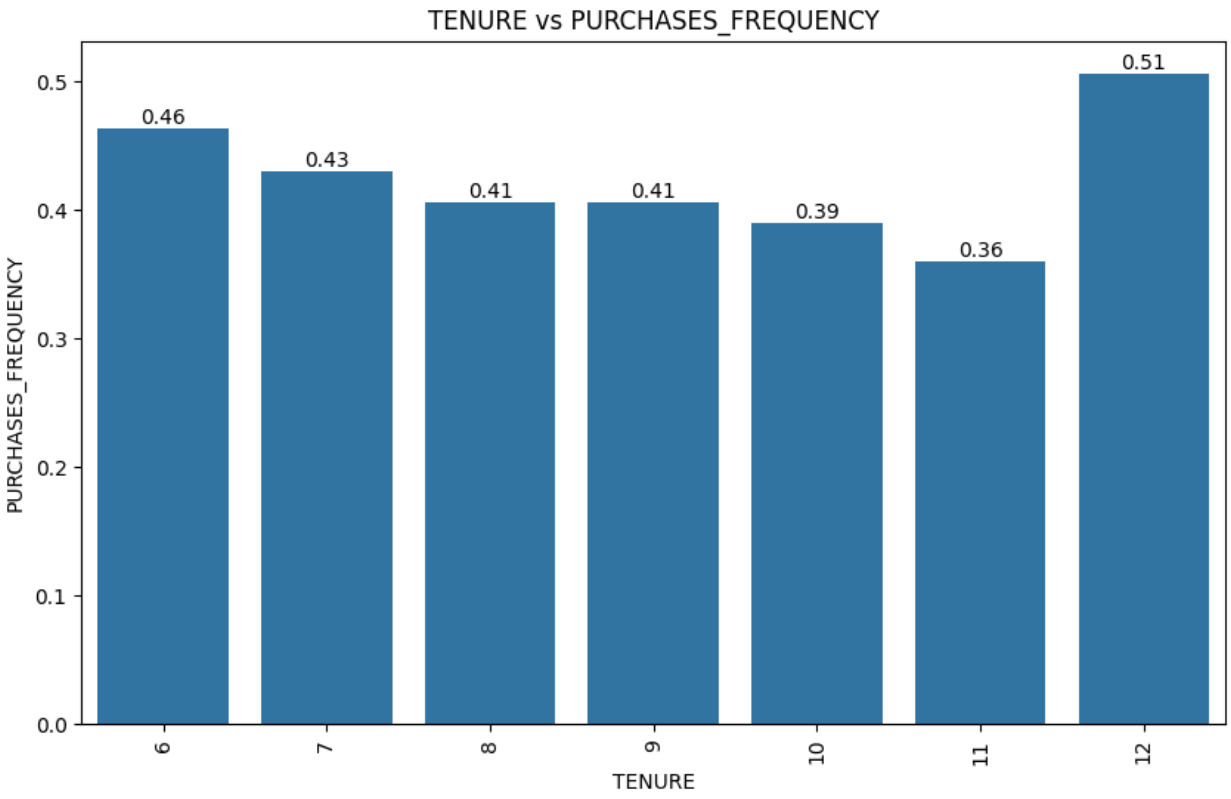

TENURE vs BALANCE_FREQUENCY
BALANCE_FREQUENCY
0.0
0.2
0.4
0.6
0.8
0.79
0.80
0.78
0.78
0.79
0.87
0.89
6
7
8
9
10
11
12
TENURE

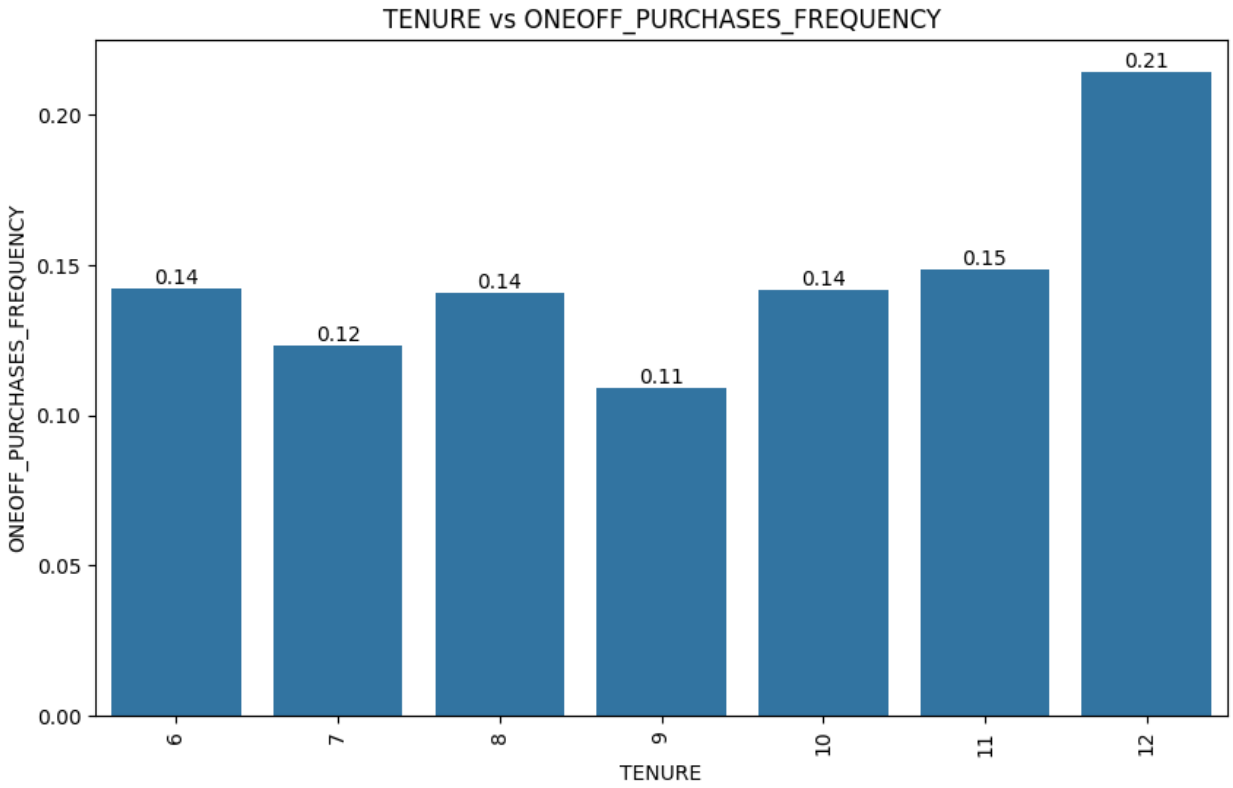

TENURE vs PURCHASES
PURCHASES
0
200
400
600
800
1000
443.94
424.56
468.86
540.14
676.63
571.22
1088.19
6
7
8
9
10
11
12
TENURE

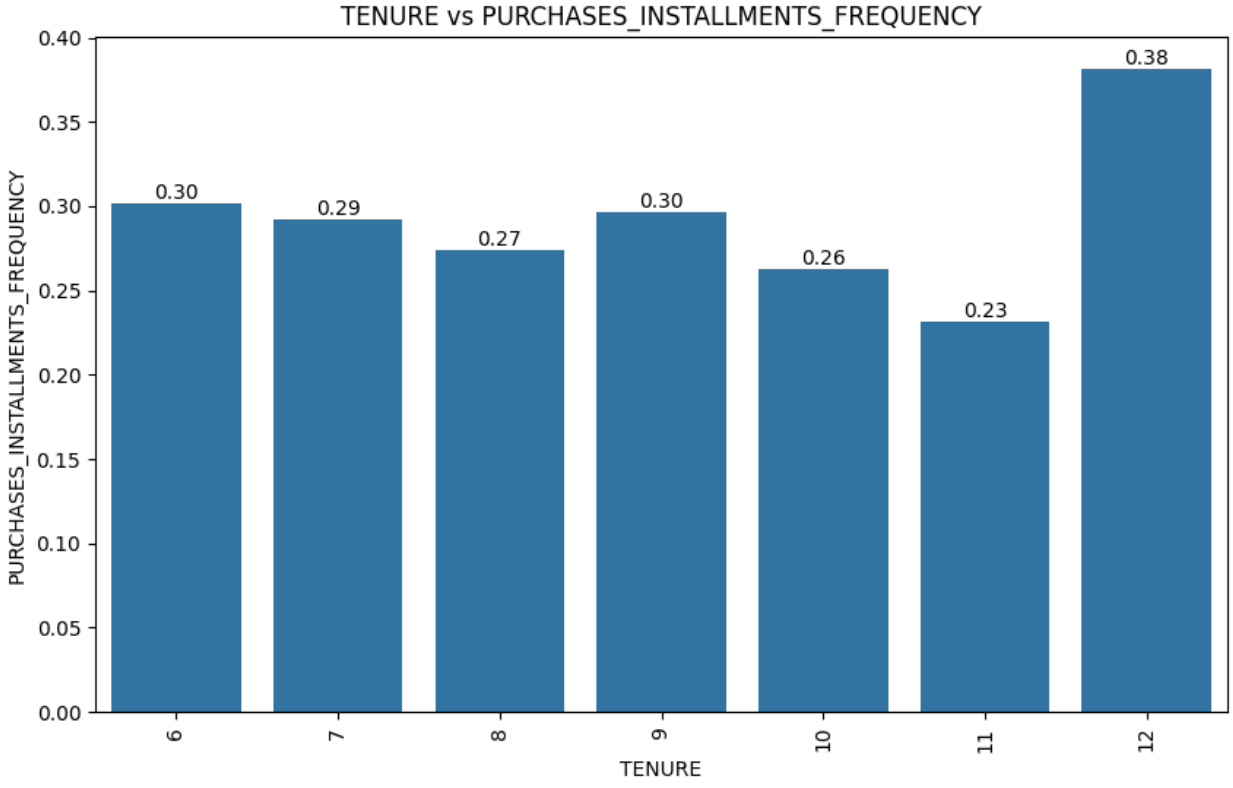

TENURE vs ONEOFF_PURCHASES
ONEOFF_PURCHASES
TENURE
0
100
200
300
400
500
600
6
7
8
9
10
11
12
266.92
244.05
305.40
302.21
407.24
370.30
640.49

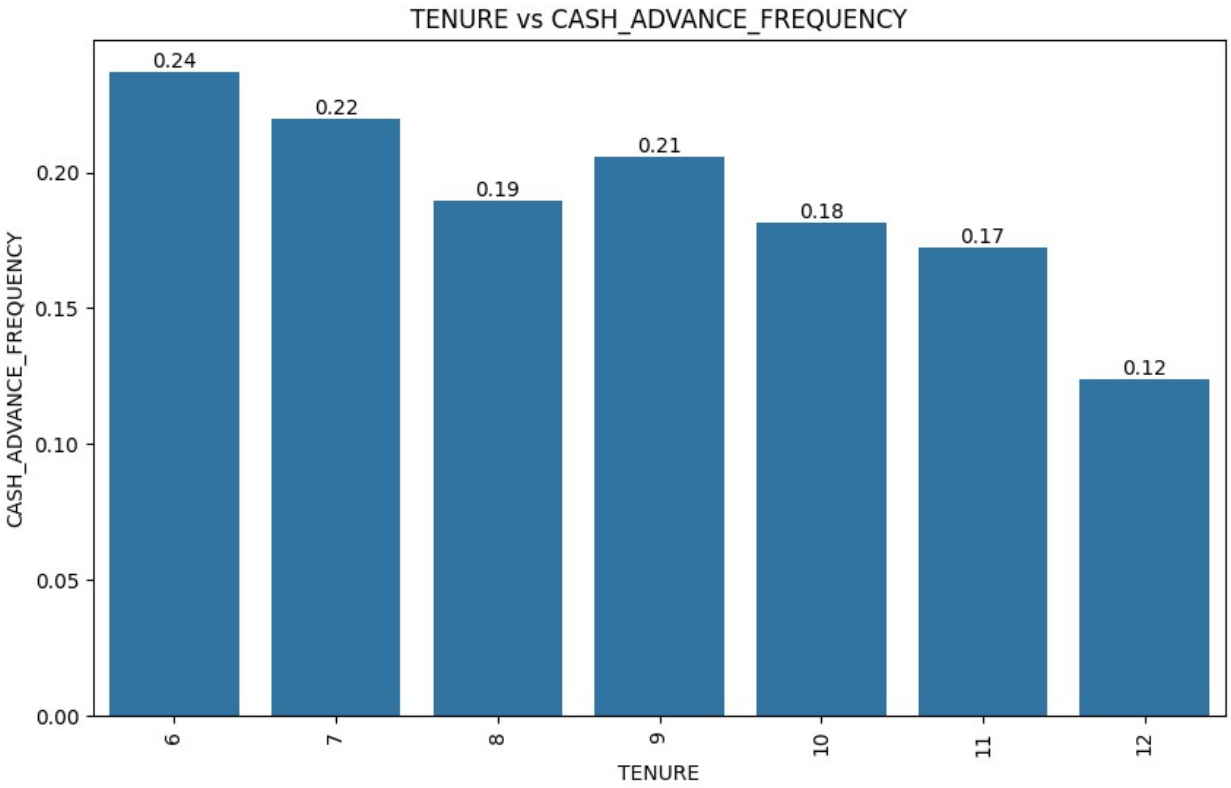

TENURE vs INSTALLMENTS_PURCHASES
INSTALLMENTS_PURCHASES
TENURE
0
100
200
300
400
6
7
8
9
10
11
12
177.17
180.51
163.46
239.47
269.59
200.92
448.01

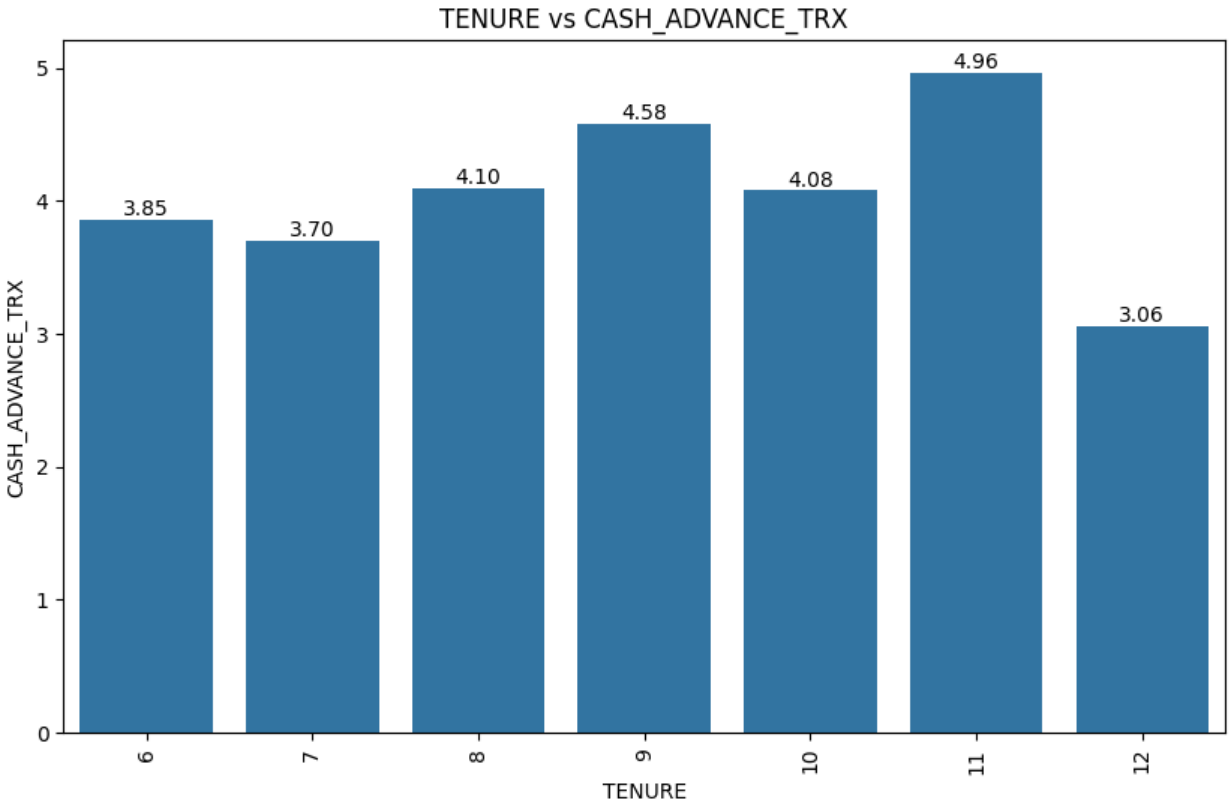

TENURE vs CASH_ADVANCE
CASH_ADVANCE
TENURE
1438.55
1295.63
1361.86
1470.79
1303.92
1363.25
908.71
0
200
400
600
800
1000
1200
1400
6
7
8
9
10
11
12

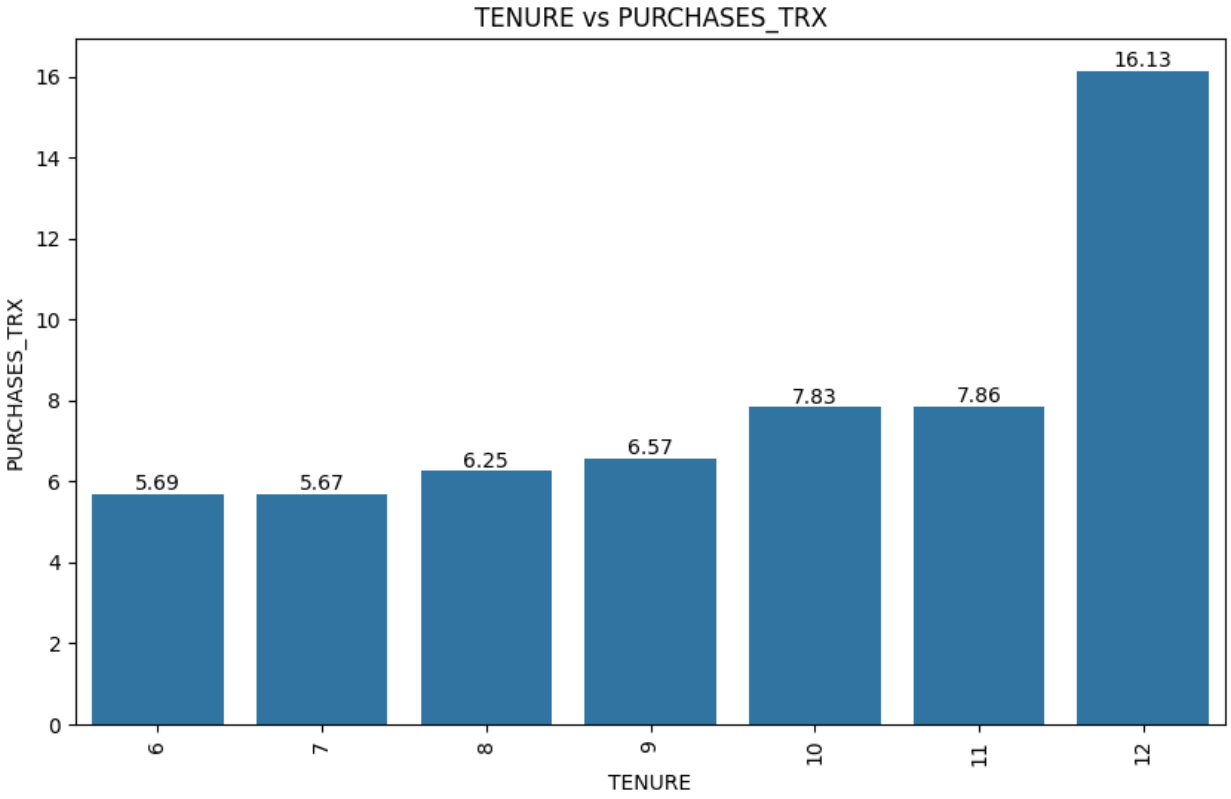

TENURE vs PURCHASES_FREQUENCY
PURCHASES_FREQUENCY
TENURE
0.46
0.43
0.41
0.41
0.39
0.36
0.51
0.0
0.1
0.2
0.3
0.4
0.5
6
7
8
9
10
11
12

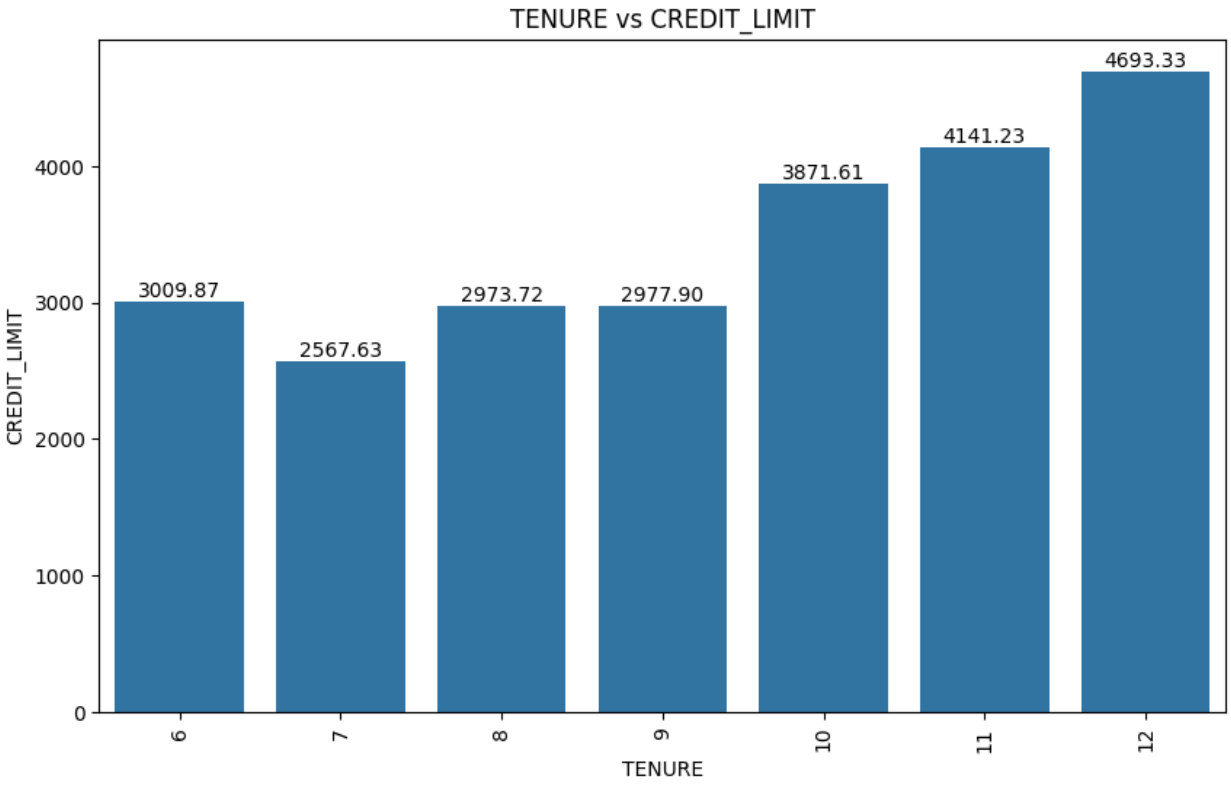

TENURE vs ONEOFF_PURCHASES_FREQUENCY
ONEOFF_PURCHASES_FREQUENCY
0.00
0.05
0.10
0.15
0.20
0.14
0.12
0.14
0.11
0.14
0.15
0.21
6
7
8
9
10
11
12
TENURE

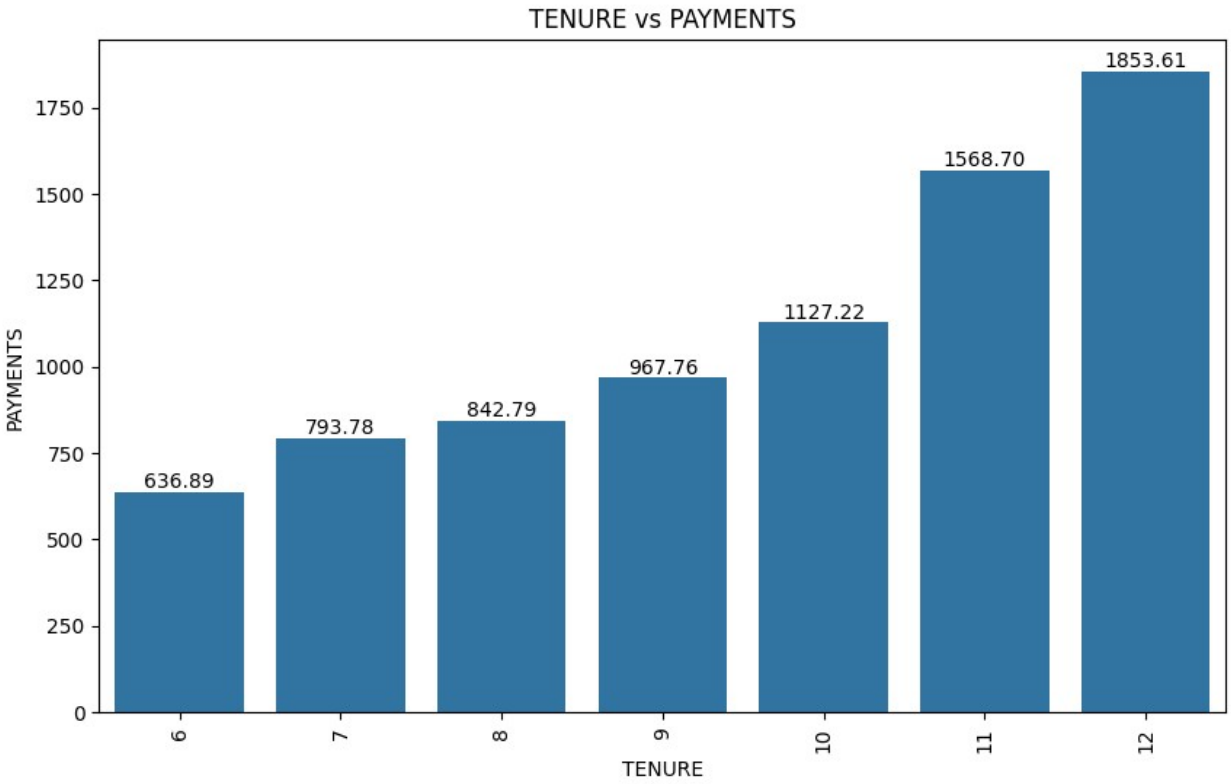

TENURE vs PURCHASES_INSTALLMENTS_FREQUENCY
PURCHASES_INSTALLMENTS_FREQUENCY
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.30
0.29
0.27
0.30
0.26
0.23
0.38
6
7
8
9
10
11
12
TENURE

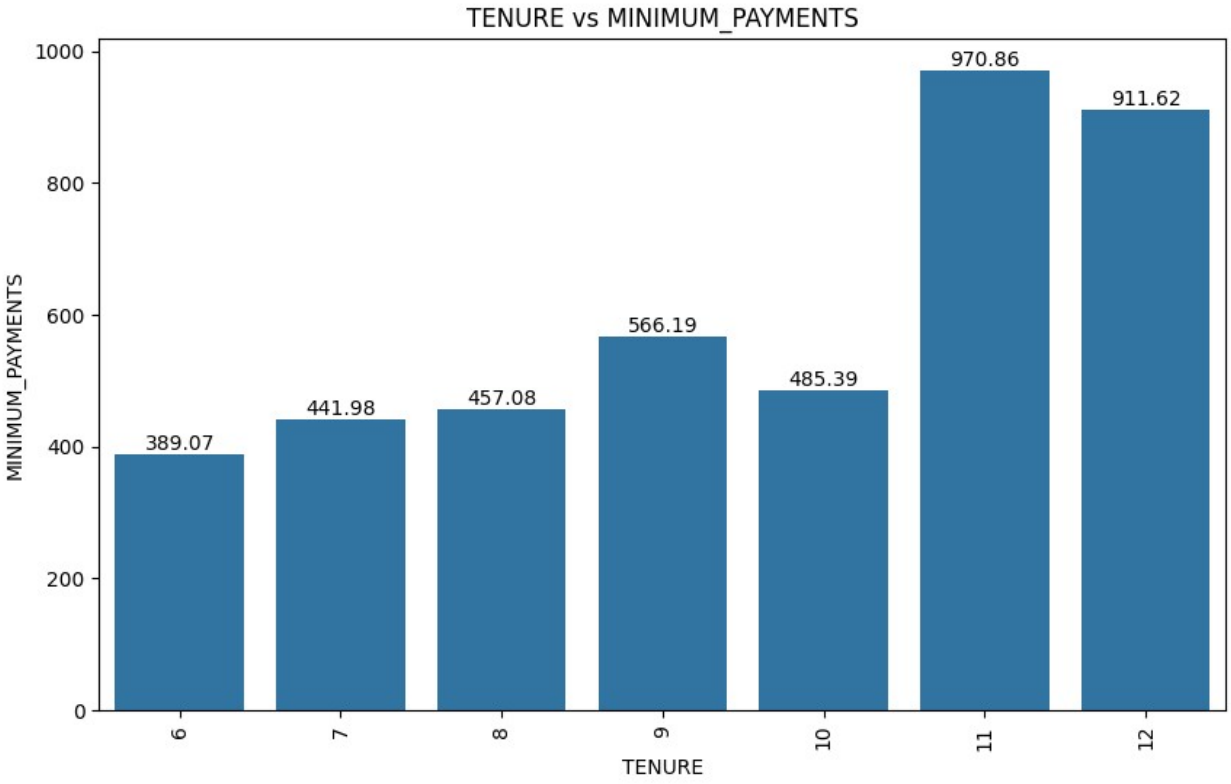

TENURE vs CASH_ADVANCE_FREQUENCY
CASH_ADVANCE_FREQUENCY
0.00
0.05
0.10
0.15
0.20
0.24
0.22
0.19
0.21
0.18
0.17
0.12
6
7
8
9
10
11
12
TENURE

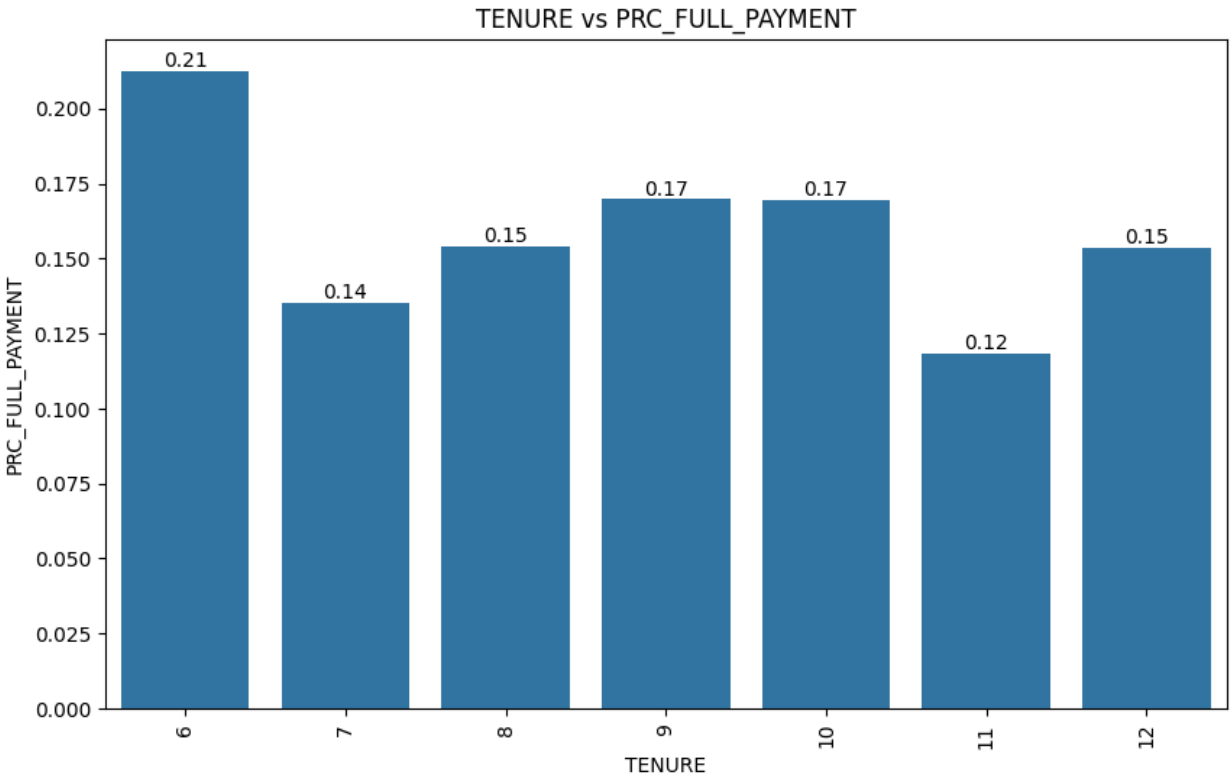

TENURE vs CASH_ADVANCE_TRX
CASH_ADVANCE_TRX
0
1
2
3
4
5
3.85
3.70
4.10
4.58
4.08
4.96
3.06
6
7
8
9
10
11
12
TENURE


TENURE vs PURCHASES_TRX
PURCHASES_TRX
TENURE
0
2
4
6
8
10
12
14
16
5.69
5.67
6.25
6.57
7.83
7.86
16.13
6
7
8
9
10
11
12

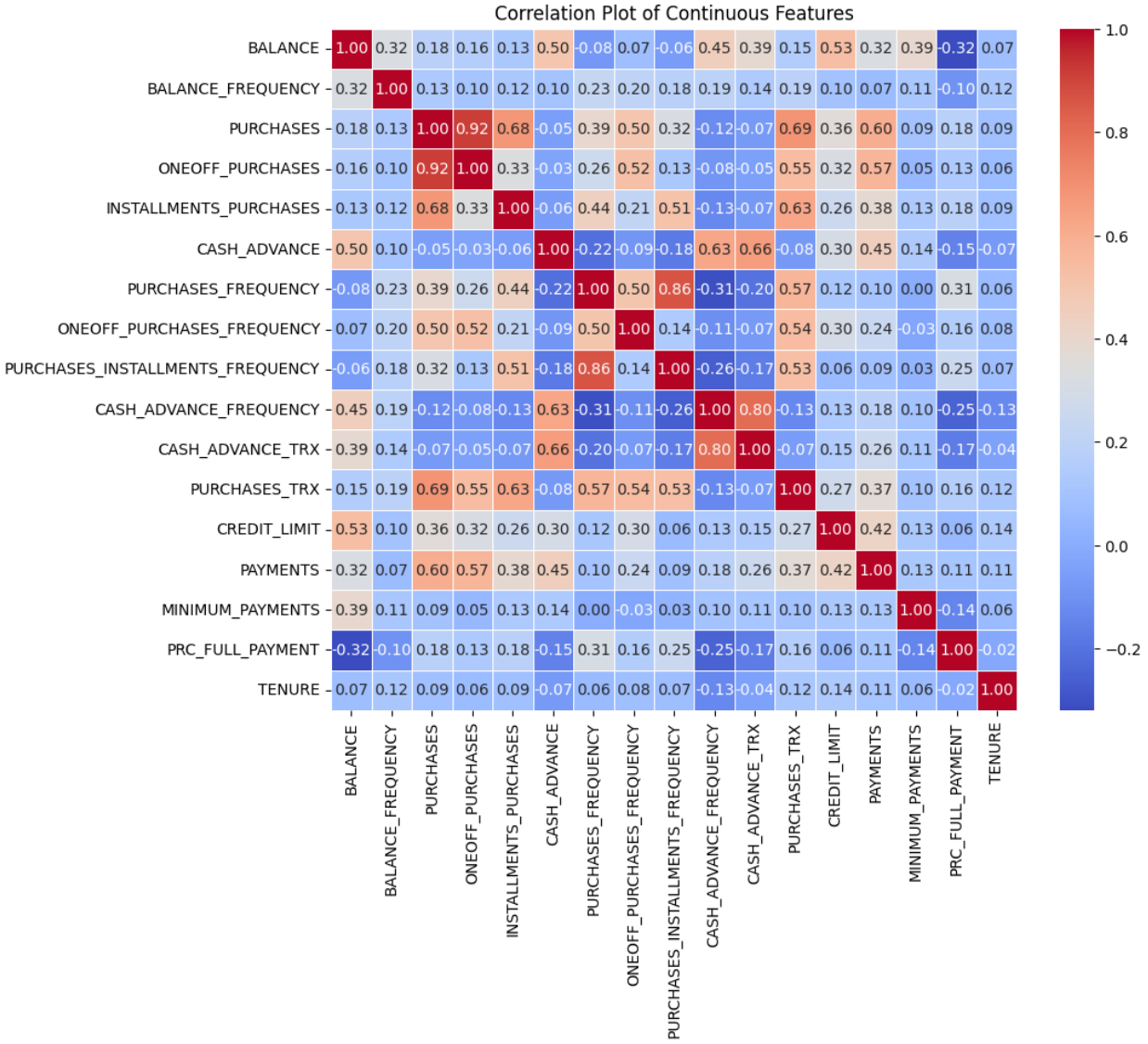

TENURE vs CREDIT_LIMIT
CREDIT_LIMIT
TENURE
0
1000
2000
3000
4000
3009.87
2567.63
2973.72
2977.90
3871.61
4141.23
4693.33
6
7
8
9
10
11
12


TENURE vs PAYMENTS
PAYMENTS
TENURE
636.89
793.78
842.79
967.76
1127.22
1568.70
1853.61
0
250
500
750
1000
1250
1500
1750
6
7
8
9
10
11
12


TENURE vs MINIMUM_PAYMENTS
MINIMUM_PAYMENTS
TENURE
389.07
441.98
457.08
566.19
485.39
970.86
911.62
0
200
400
600
800
1000
6
7
8
9
10
11
12




```
corr_matrix = df.corr()

plt.figure(figsize=(10, 8))
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', fmt=".2f",
linewidths=0.5)
plt.title('Correlation Plot of Continuous Features')
plt.show()
```




```
corr_matrix

                                  BALANCE  BALANCE_FREQUENCY
PURCHASES  \
BALANCE                          1.000000           0.322412
0.181261
BALANCE_FREQUENCY                0.322412           1.000000
0.133674
PURCHASES                        0.181261           0.133674
1.000000
ONEOFF_PURCHASES                 0.164350           0.104323
0.916845
INSTALLMENTS_PURCHASES           0.126469           0.124292
0.679896
CASH_ADVANCE                     0.496692           0.099388  -
0.051474
```

```
PURCHASES_FREQUENCY                -0.077944            0.229715
0.393017
ONEOFF_PURCHASES_FREQUENCY          0.073166            0.202415
0.498430
PURCHASES_INSTALLMENTS_FREQUENCY -0.063186              0.176079
0.315567
CASH_ADVANCE_FREQUENCY              0.449218            0.191873  -
0.120143
CASH_ADVANCE_TRX                    0.385152            0.141555  -
0.067175
PURCHASES_TRX                       0.154338            0.189626
0.689561
CREDIT_LIMIT                        0.531267            0.095795
0.356959
PAYMENTS                            0.322802            0.065008
0.603264
MINIMUM_PAYMENTS                    0.394282            0.114249
0.093515
PRC_FULL_PAYMENT                   -0.318959           -0.095082
0.180379
TENURE                              0.072692            0.119776
0.086288

                                   ONEOFF_PURCHASES
INSTALLMENTS_PURCHASES  \
BALANCE                                 0.164350
0.126469
BALANCE_FREQUENCY                       0.104323
0.124292
PURCHASES                               0.916845
0.679896
ONEOFF_PURCHASES                        1.000000
0.330622
INSTALLMENTS_PURCHASES                  0.330622
1.000000
CASH_ADVANCE                           -0.031326                -
0.064244
PURCHASES_FREQUENCY                     0.264937
0.442418
ONEOFF_PURCHASES_FREQUENCY              0.524891
0.214042
PURCHASES_INSTALLMENTS_FREQUENCY        0.127729
0.511351
CASH_ADVANCE_FREQUENCY                 -0.082628                -
0.132318
CASH_ADVANCE_TRX                       -0.046212                -
0.073999
PURCHASES_TRX                           0.545523
0.628108
```

```
CREDIT_LIMIT                                0.319721
0.256496
PAYMENTS                                    0.567292
0.384084
MINIMUM_PAYMENTS                            0.048597
0.131687
PRC_FULL_PAYMENT                            0.132763
0.182569
TENURE                                      0.064150
0.086143

                                  CASH_ADVANCE  PURCHASES_FREQUENCY  \
BALANCE                               0.496692            -0.077944
BALANCE_FREQUENCY                     0.099388             0.229715
PURCHASES                            -0.051474             0.393017
ONEOFF_PURCHASES                     -0.031326             0.264937
INSTALLMENTS_PURCHASES               -0.064244             0.442418
CASH_ADVANCE                          1.000000            -0.215507
PURCHASES_FREQUENCY                  -0.215507             1.000000
ONEOFF_PURCHASES_FREQUENCY           -0.086754             0.501343
PURCHASES_INSTALLMENTS_FREQUENCY     -0.177070             0.862934
CASH_ADVANCE_FREQUENCY                0.628522            -0.308478
CASH_ADVANCE_TRX                      0.656498            -0.203478
PURCHASES_TRX                        -0.075850             0.568430
CREDIT_LIMIT                          0.303983             0.119778
PAYMENTS                              0.453238             0.103464
MINIMUM_PAYMENTS                      0.139223             0.002976
PRC_FULL_PAYMENT                     -0.152935             0.305802
TENURE                               -0.068312             0.061506

                                  ONEOFF_PURCHASES_FREQUENCY  \
BALANCE                                             0.073166
BALANCE_FREQUENCY                                   0.202415
PURCHASES                                           0.498430
ONEOFF_PURCHASES                                    0.524891
INSTALLMENTS_PURCHASES                              0.214042
CASH_ADVANCE                                       -0.086754
PURCHASES_FREQUENCY                                 0.501343
ONEOFF_PURCHASES_FREQUENCY                          1.000000
PURCHASES_INSTALLMENTS_FREQUENCY                    0.142329
CASH_ADVANCE_FREQUENCY                             -0.111716
CASH_ADVANCE_TRX                                   -0.069088
PURCHASES_TRX                                       0.544869
CREDIT_LIMIT                                        0.295030
PAYMENTS                                            0.243537
MINIMUM_PAYMENTS                                   -0.029963
PRC_FULL_PAYMENT                                    0.157531
TENURE                                              0.082466

                                  PURCHASES_INSTALLMENTS_FREQUENCY  \
```

```
BALANCE                                                -0.063186
BALANCE_FREQUENCY                                       0.176079
PURCHASES                                               0.315567
ONEOFF_PURCHASES                                        0.127729
INSTALLMENTS_PURCHASES                                  0.511351
CASH_ADVANCE                                           -0.177070
PURCHASES_FREQUENCY                                     0.862934
ONEOFF_PURCHASES_FREQUENCY                              0.142329
PURCHASES_INSTALLMENTS_FREQUENCY                        1.000000
CASH_ADVANCE_FREQUENCY                                 -0.262958
CASH_ADVANCE_TRX                                       -0.169207
PURCHASES_TRX                                           0.529975
CREDIT_LIMIT                                            0.060752
PAYMENTS                                                0.085551
MINIMUM_PAYMENTS                                        0.029590
PRC_FULL_PAYMENT                                        0.250087
TENURE                                                  0.073275

                                    CASH_ADVANCE_FREQUENCY
CASH_ADVANCE_TRX  \
BALANCE                                           0.449218
0.385152
BALANCE_FREQUENCY                                 0.191873
0.141555
PURCHASES                                        -0.120143         -
0.067175
ONEOFF_PURCHASES                                 -0.082628         -
0.046212
INSTALLMENTS_PURCHASES                           -0.132318         -
0.073999
CASH_ADVANCE                                      0.628522
0.656498
PURCHASES_FREQUENCY                              -0.308478         -
0.203478
ONEOFF_PURCHASES_FREQUENCY                       -0.111716         -
0.069088
PURCHASES_INSTALLMENTS_FREQUENCY                 -0.262958         -
0.169207
CASH_ADVANCE_FREQUENCY                            1.000000
0.799561
CASH_ADVANCE_TRX                                  0.799561
1.000000
PURCHASES_TRX                                    -0.131168         -
0.066157
CREDIT_LIMIT                                      0.132616
0.149699
PAYMENTS                                          0.183192
0.255278
MINIMUM_PAYMENTS                                  0.097898
```

```
0.109185
PRC_FULL_PAYMENT                                -0.249773          -
0.169784
TENURE                                          -0.133372          -
0.043421

                                  PURCHASES_TRX  CREDIT_LIMIT
PAYMENTS  \
BALANCE                                0.154338      0.531267
0.322802
BALANCE_FREQUENCY                      0.189626      0.095795
0.065008
PURCHASES                              0.689561      0.356959
0.603264
ONEOFF_PURCHASES                       0.545523      0.319721
0.567292
INSTALLMENTS_PURCHASES                 0.628108      0.256496
0.384084
CASH_ADVANCE                          -0.075850      0.303983
0.453238
PURCHASES_FREQUENCY                    0.568430      0.119778
0.103464
ONEOFF_PURCHASES_FREQUENCY             0.544869      0.295030
0.243537
PURCHASES_INSTALLMENTS_FREQUENCY       0.529975      0.060752
0.085551
CASH_ADVANCE_FREQUENCY                -0.131168      0.132616
0.183192
CASH_ADVANCE_TRX                      -0.066157      0.149699
0.255278
PURCHASES_TRX                          1.000000      0.272877
0.370832
CREDIT_LIMIT                           0.272877      1.000000
0.421852
PAYMENTS                               0.370832      0.421852
1.000000
MINIMUM_PAYMENTS                       0.095858      0.125134
0.125046
PRC_FULL_PAYMENT                       0.162066      0.055671
0.112138
TENURE                                 0.121874      0.139034
0.106136

                                  MINIMUM_PAYMENTS  PRC_FULL_PAYMENT
TENURE
BALANCE                                   0.394282         -0.318959
0.072692
BALANCE_FREQUENCY                         0.114249         -0.095082
0.119776
```

```
PURCHASES                                0.093515          0.180379
0.086288
ONEOFF_PURCHASES                         0.048597          0.132763
0.064150
INSTALLMENTS_PURCHASES                   0.131687          0.182569
0.086143
CASH_ADVANCE                             0.139223         -0.152935 -
0.068312
PURCHASES_FREQUENCY                      0.002976          0.305802
0.061506
ONEOFF_PURCHASES_FREQUENCY              -0.029963          0.157531
0.082466
PURCHASES_INSTALLMENTS_FREQUENCY         0.029590          0.250087
0.073275
CASH_ADVANCE_FREQUENCY                   0.097898         -0.249773 -
0.133372
CASH_ADVANCE_TRX                         0.109185         -0.169784 -
0.043421
PURCHASES_TRX                            0.095858          0.162066
0.121874
CREDIT_LIMIT                             0.125134          0.055671
0.139034
PAYMENTS                                 0.125046          0.112138
0.106136
MINIMUM_PAYMENTS                         1.000000         -0.139674
0.057257
PRC_FULL_PAYMENT                        -0.139674          1.000000 -
0.016486
TENURE                                   0.057257         -0.016486
1.000000
```

```
threshold = 0.7
features_to_drop = []
for i in range(len(corr_matrix.columns)):
    for j in range(i):
        if abs(corr_matrix.iloc[i, j]) > threshold:
            colname = corr_matrix.columns[i]
            if colname not in features_to_drop:
                features_to_drop.append(colname)

features_to_drop
```

```
['ONEOFF_PURCHASES', 'PURCHASES_INSTALLMENTS_FREQUENCY',
'CASH_ADVANCE_TRX']
```

```
df = df.drop(columns=features_to_drop)

Q1 = df.quantile(0.25)
Q3 = df.quantile(0.75)
IQR = Q3 - Q1
```

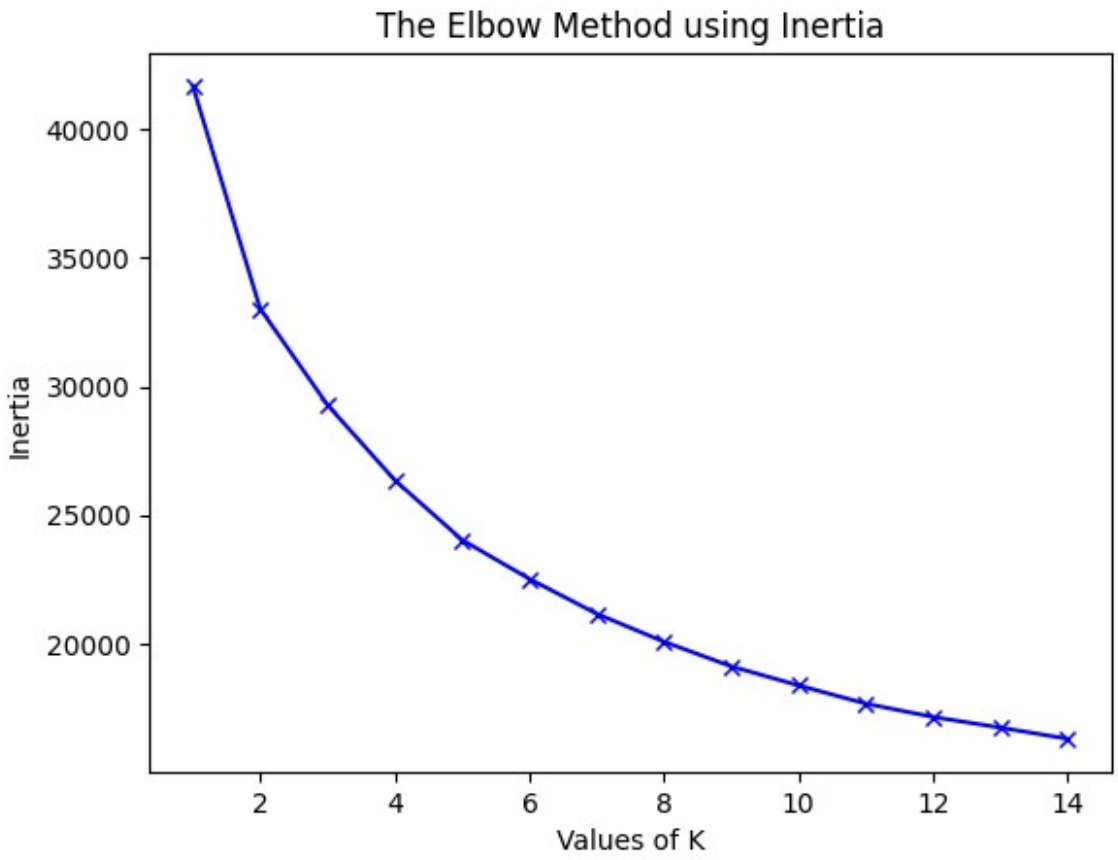
```
outliers = ((df < (Q1 - 1.5 * IQR)) | (df > (Q3 + 1.5 *
IQR))).any(axis=1)

outliers

0       False
1        True
2        True
3        True
4       False
        ...
8945     True
8946     True
8947     True
8948     True
8949     True
Length: 8950, dtype: bool

df_cleaned = df[~outliers]

df_cleaned

          BALANCE  BALANCE_FREQUENCY  PURCHASES
INSTALLMENTS_PURCHASES  \
0       40.900749           0.818182      95.40
95.40
4      817.714335           1.000000      16.00
0.00
7     1823.652743           1.000000     436.20
436.20
8     1014.926473           1.000000     861.49
200.00
11     630.794744           0.818182    1492.18
0.00
...           ...                ...        ...                       ..
.
8738   981.286008           1.000000    1370.00
0.00
8742    87.026009           1.000000     605.52
605.52
8747    16.428326           0.909091     441.50
316.80
8759    67.377243           1.000000     295.00
295.00
8760   307.127754           1.000000     909.30
500.00

      CASH_ADVANCE  PURCHASES_FREQUENCY  ONEOFF_PURCHASES_FREQUENCY  \
0         0.000000             0.166667                    0.000000
4         0.000000             0.083333                    0.083333
```

```
7         0.000000             1.000000                    0.000000
8         0.000000             0.333333                    0.083333
11        0.000000             0.250000                    0.250000
...            ...                  ...                         ...
8738      0.000000             0.083333                    0.083333
8742      0.000000             1.000000                    0.000000
8747      0.000000             1.000000                    0.166667
8759      0.000000             0.500000                    0.000000
8760    237.378894             0.583333                    0.166667

      CASH_ADVANCE_FREQUENCY  PURCHASES_TRX  CREDIT_LIMIT     PAYMENTS
\
0                   0.000000              2        1000.0  201.802084

4                   0.000000              1        1200.0  678.334763

7                   0.000000             12        2300.0  679.065082

8                   0.000000              5        7000.0  688.278568

11                  0.000000              6        2000.0  705.618627

...                      ...            ...           ...         ...

8738                0.000000              1        1400.0  596.685481

8742                0.000000             12        1500.0  511.637312

8747                0.000000             14        1000.0  482.547848

8759                0.000000              6        1000.0  245.689379

8760                0.166667             12        1000.0  943.278170


      MINIMUM_PAYMENTS  PRC_FULL_PAYMENT  TENURE
0           139.509787          0.000000      12
4           244.791237          0.000000      12
7           532.033990          0.000000      12
8           311.963409          0.000000      12
11          155.549069          0.000000      12
...                ...               ...     ...
8738        451.584847          0.000000      12
8742        175.012705          0.000000      12
8747         91.328536          0.333333      12
8759        167.126034          0.300000      12
8760        179.258575          0.000000      12

[3207 rows x 14 columns]
```

```
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.cluster import
KMeans,AgglomerativeClustering,DBSCAN,SpectralClustering
from sklearn.mixture import GaussianMixture
from sklearn.metrics import silhouette_samples, silhouette_score
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report
from sklearn import tree
from sklearn import metrics

scalar=StandardScaler()

scaled_df = scalar.fit_transform(df_cleaned)

pca = PCA(n_components=2)
principal_components = pca.fit_transform(scaled_df)
pca_df =
pd.DataFrame(data=principal_components ,columns=["PCA1","PCA2"])
pca_df
```

```
          PCA1      PCA2
0    -0.543394 -2.871282
1    -0.984140 -1.505669
2     1.282342  0.081815
3     0.545613 -0.165682
4     1.026556 -1.381102
...        ...       ...
3202 -0.043462 -0.619531
3203  2.001030 -1.251574
3204  2.568172 -2.711847
3205  0.999517 -2.719791
3206  1.389334 -0.410095

[3207 rows x 2 columns]
```

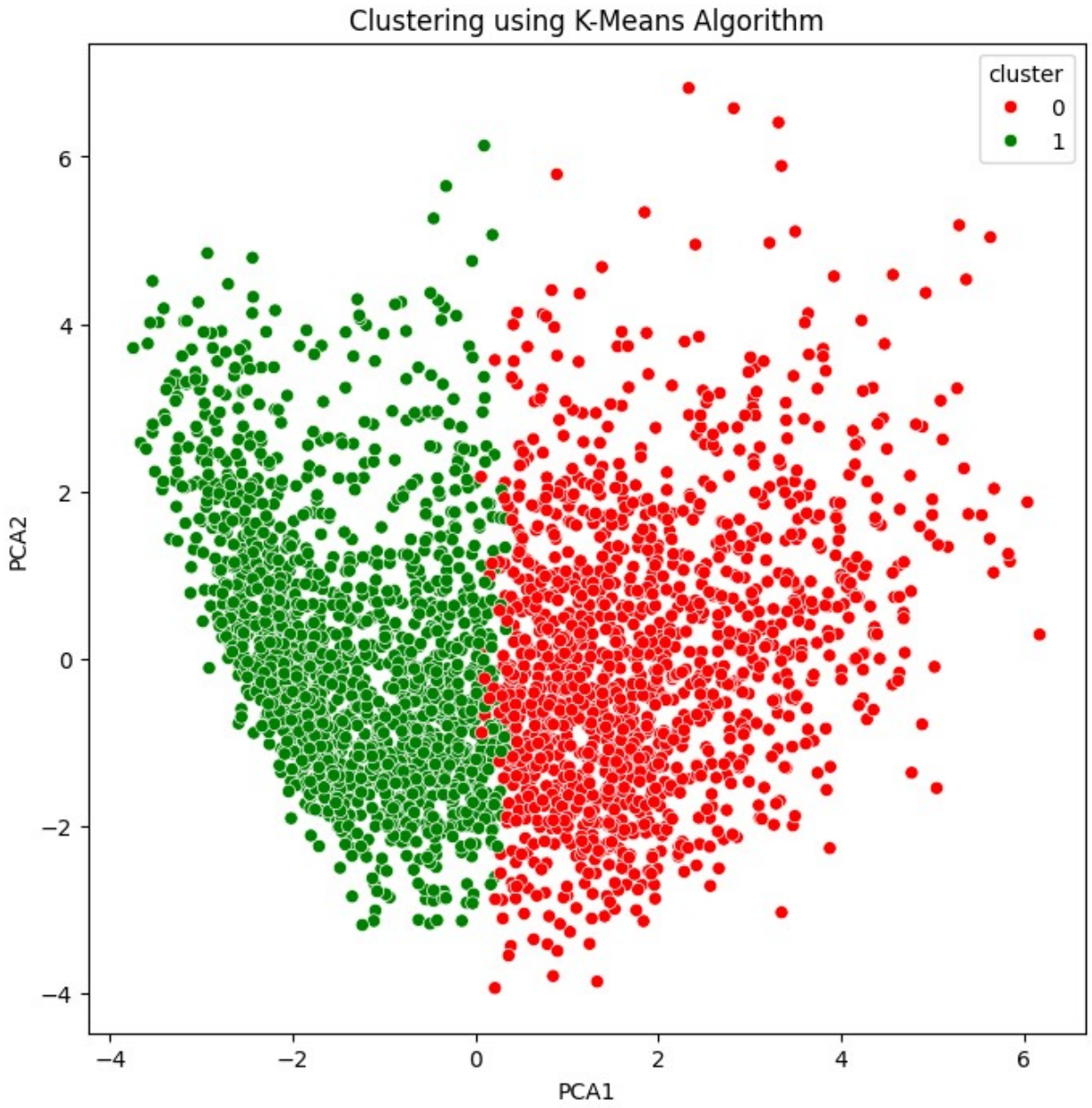
```
inertia = []
range_val = range(1,15)
for i in range_val:
    kmean = KMeans(n_clusters=i)
    kmean.fit_predict(pd.DataFrame(scaled_df))
    inertia.append(kmean.inertia_)
plt.plot(range_val,inertia,'bx-')
plt.xlabel('Values of K')
plt.ylabel('Inertia')
plt.title('The Elbow Method using Inertia')
plt.show()
```




```
kmeans_model=KMeans(2)
kmeans_model.fit_predict(scaled_df)
pca_df_kmeans=
pd.concat([pca_df,pd.DataFrame({'cluster':kmeans_model.labels_})],axis
=1)

pca_df_kmeans
```

```
          PCA1      PCA2  cluster
0    -0.543394 -2.871282        1
1    -0.984140 -1.505669        1
2     1.282342  0.081815        0
3     0.545613 -0.165682        0
4     1.026556 -1.381102        0
...        ...       ...      ...
3202 -0.043462 -0.619531        1
3203  2.001030 -1.251574        0
3204  2.568172 -2.711847        0
3205  0.999517 -2.719791        0
3206  1.389334 -0.410095        0

[3207 rows x 3 columns]
```

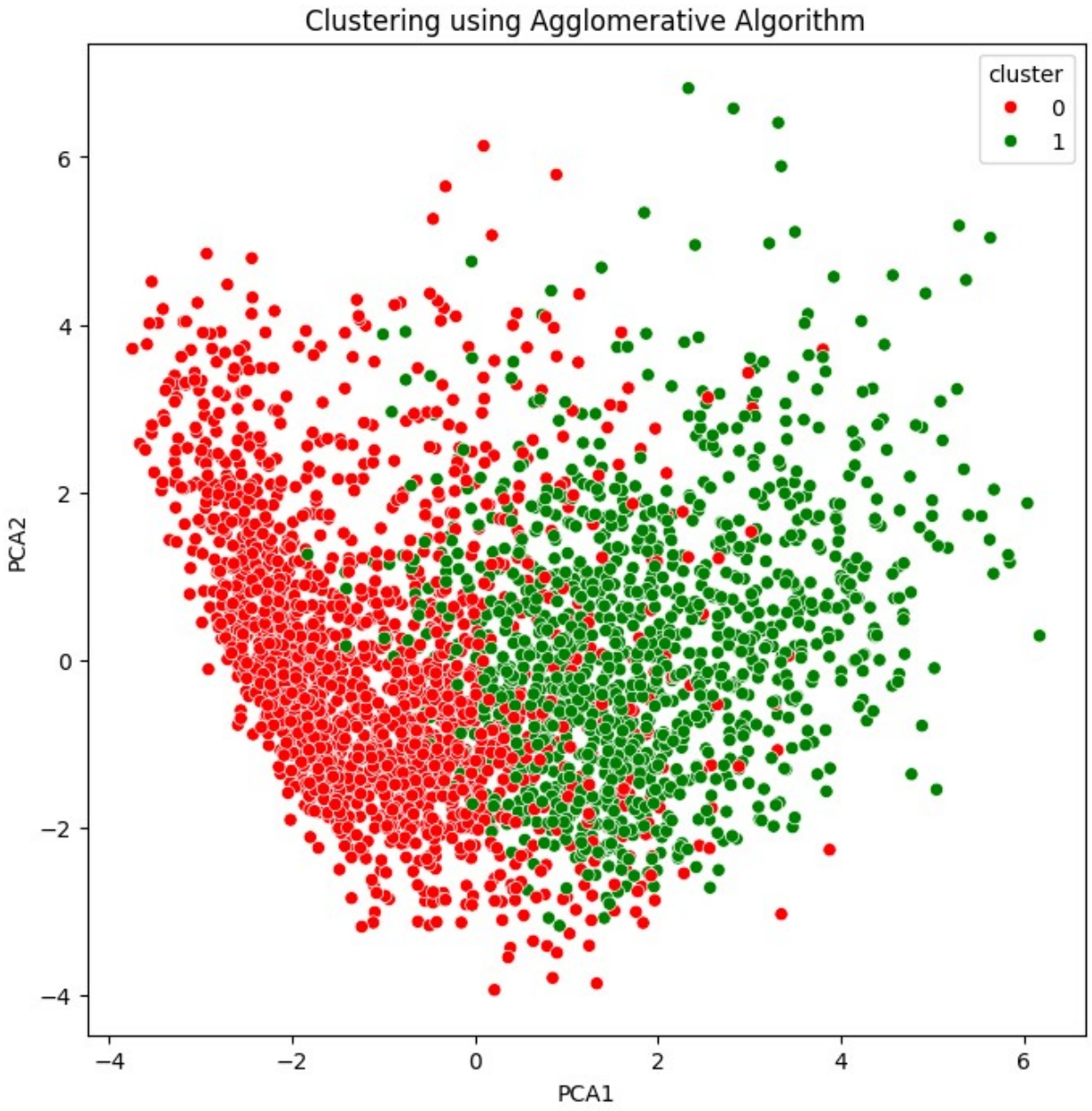
```
plt.figure(figsize=(8,8))
ax=sns.scatterplot(x="PCA1",y="PCA2",hue="cluster",data=pca_df_kmeans,
palette=['red','green'])
plt.title("Clustering using K-Means Algorithm")
plt.show()
```




```
from sklearn.metrics import silhouette_score

sil_coeff = silhouette_score(pca_df_kmeans.drop("cluster", axis=1),
pca_df_kmeans["cluster"])
print("Silhouette Coefficient:", round(sil_coeff, 3))
```

```
Silhouette Coefficient: 0.404
```

```
from sklearn.cluster import AgglomerativeClustering

cluster = AgglomerativeClustering(n_clusters=2, affinity='euclidean',
linkage='ward')
labels = cluster.fit_predict(scaled_df)

pca_df_agg =
pd.concat([pca_df,pd.DataFrame({'cluster':cluster.labels_})],axis=1)

plt.figure(figsize=(8,8))
ax=sns.scatterplot(x="PCA1",y="PCA2",hue="cluster",data=pca_df_agg,pal
ette=['red','green'])
plt.title("Clustering using Agglomerative Algorithm")
plt.show()
```




```
sil_coeff = silhouette_score(pca_df_agg.drop("cluster", axis=1),
pca_df_agg["cluster"])
print("Silhouette Coefficient:", round(sil_coeff, 3))

Silhouette Coefficient: 0.324
```

Thanks !!!