In [1]:
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings(action="ignore")
```

In [2]:
```python
data=pd.read_csv('CC GENERAL.csv')
data
```

Out[2]:

	CUST_ID	BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES	CASH_ADVA
0	C10001	40.900749	0.818182	95.40	0.00	95.40	0.000
1	C10002	3202.467416	0.909091	0.00	0.00	0.00	6442.945
2	C10003	2495.148862	1.000000	773.17	773.17	0.00	0.000
3	C10004	1666.670542	0.636364	1499.00	1499.00	0.00	205.788
4	C10005	817.714335	1.000000	16.00	16.00	0.00	0.000
...	...	...	...	...	...	...	
8945	C19186	28.493517	1.000000	291.12	0.00	291.12	0.000
8946	C19187	19.183215	1.000000	300.00	0.00	300.00	0.000
8947	C19188	23.398673	0.833333	144.40	0.00	144.40	0.000
8948	C19189	13.457564	0.833333	0.00	0.00	0.00	36.558
8949	C19190	372.708075	0.666667	1093.25	1093.25	0.00	127.040

8950 rows × 18 columns

In [3]:
```python
data.describe().T
```

Out[3]:

	count	mean	std	min	25%	50%	75%
BALANCE	8950.0	1564.474828	2081.531879	0.000000	128.281915	873.385231	2054.140036
BALANCE_FREQUENCY	8950.0	0.877271	0.236904	0.000000	0.888889	1.000000	1.000000
PURCHASES	8950.0	1003.204834	2136.634782	0.000000	39.635000	361.280000	1110.130000
ONEOFF_PURCHASES	8950.0	592.437371	1659.887917	0.000000	0.000000	38.000000	577.405000
INSTALLMENTS_PURCHASES	8950.0	411.067645	904.338115	0.000000	0.000000	89.000000	468.637500
CASH_ADVANCE	8950.0	978.871112	2097.163877	0.000000	0.000000	0.000000	1113.821139
PURCHASES_FREQUENCY	8950.0	0.490351	0.401371	0.000000	0.083333	0.500000	0.916667
ONEOFF_PURCHASES_FREQUENCY	8950.0	0.202458	0.298336	0.000000	0.000000	0.083333	0.300000
PURCHASES_INSTALLMENTS_FREQUENCY	8950.0	0.364437	0.397448	0.000000	0.000000	0.166667	0.750000
CASH_ADVANCE_FREQUENCY	8950.0	0.135144	0.200121	0.000000	0.000000	0.000000	0.222222
CASH_ADVANCE_TRX	8950.0	3.248827	6.824647	0.000000	0.000000	0.000000	4.000000
PURCHASES_TRX	8950.0	14.709832	24.857649	0.000000	1.000000	7.000000	17.000000
CREDIT_LIMIT	8949.0	4494.449450	3638.815725	50.000000	1600.000000	3000.000000	6500.000000
PAYMENTS	8950.0	1733.143852	2895.063757	0.000000	383.276166	856.901546	1901.134317
MINIMUM_PAYMENTS	8637.0	864.206542	2372.446607	0.019163	169.123707	312.343947	825.485459
PRC_FULL_PAYMENT	8950.0	0.153715	0.292499	0.000000	0.000000	0.000000	0.142857
TENURE	8950.0	11.517318	1.338331	6.000000	12.000000	12.000000	12.000000

In [4]:
```
data.isnull().sum().sort_values(ascending=False)
```

Out[4]:   MINIMUM_PAYMENTS   313
           CREDIT_LIMIT   1
           CUST_ID   0
           BALANCE   0
           PRC_FULL_PAYMENT   0
           PAYMENTS   0
           PURCHASES_TRX   0
           CASH_ADVANCE_TRX   0
           CASH_ADVANCE_FREQUENCY   0
           PURCHASES_INSTALLMENTS_FREQUENCY   0
           ONEOFF_PURCHASES_FREQUENCY   0
           PURCHASES_FREQUENCY   0
           CASH_ADVANCE   0
           INSTALLMENTS_PURCHASES   0
           ONEOFF_PURCHASES   0
           PURCHASES   0
           BALANCE_FREQUENCY   0
           TENURE   0
           dtype: int64

In [5]:
```python
cat_col = data.select_dtypes(include=['object']).columns
num_col = data.select_dtypes(exclude=['object']).columns
```
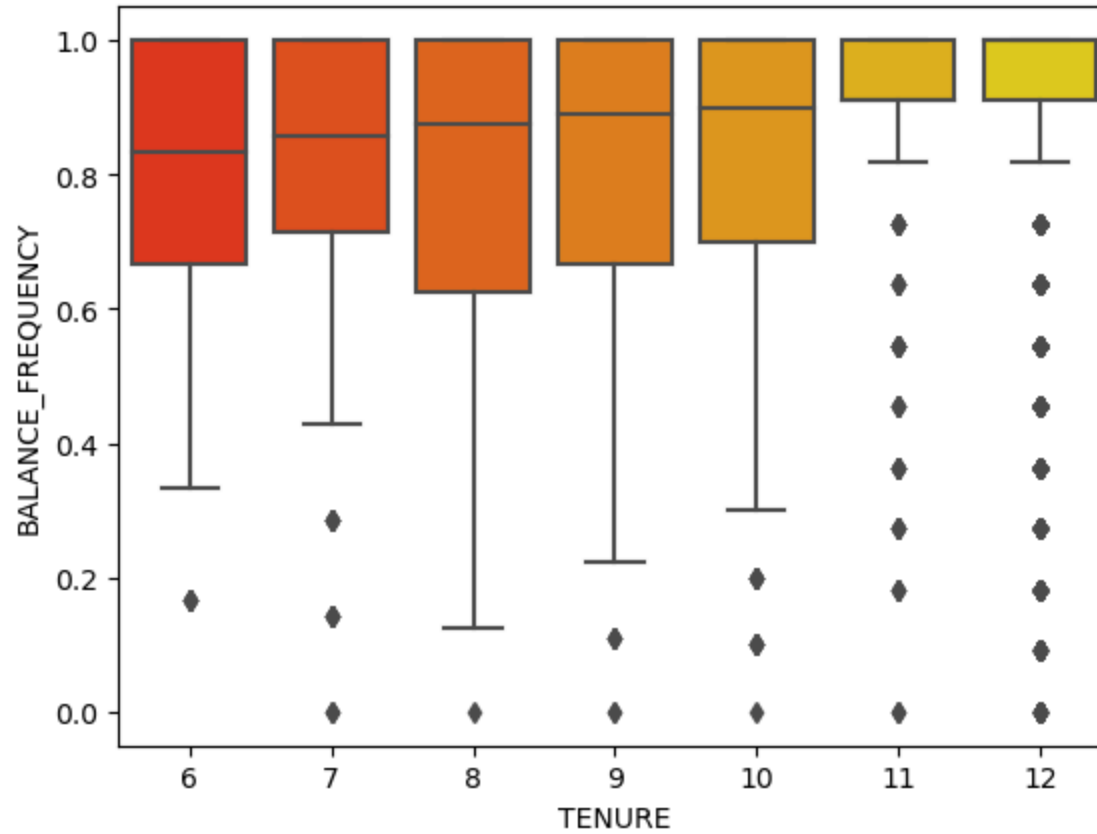
In [6]:
```python
df_v = data['TENURE'].value_counts().reset_index()
df_v.columns = ['TENURE', 'Counts']
plot = df_v.plot.pie(y='Counts', labels=df_v['TENURE'], figsize=(8, 8), autopct='%1.1f%%')
plt.ylabel('')  # Optional: remove the y-label
plt.title('TENURE')
plt.show()
```
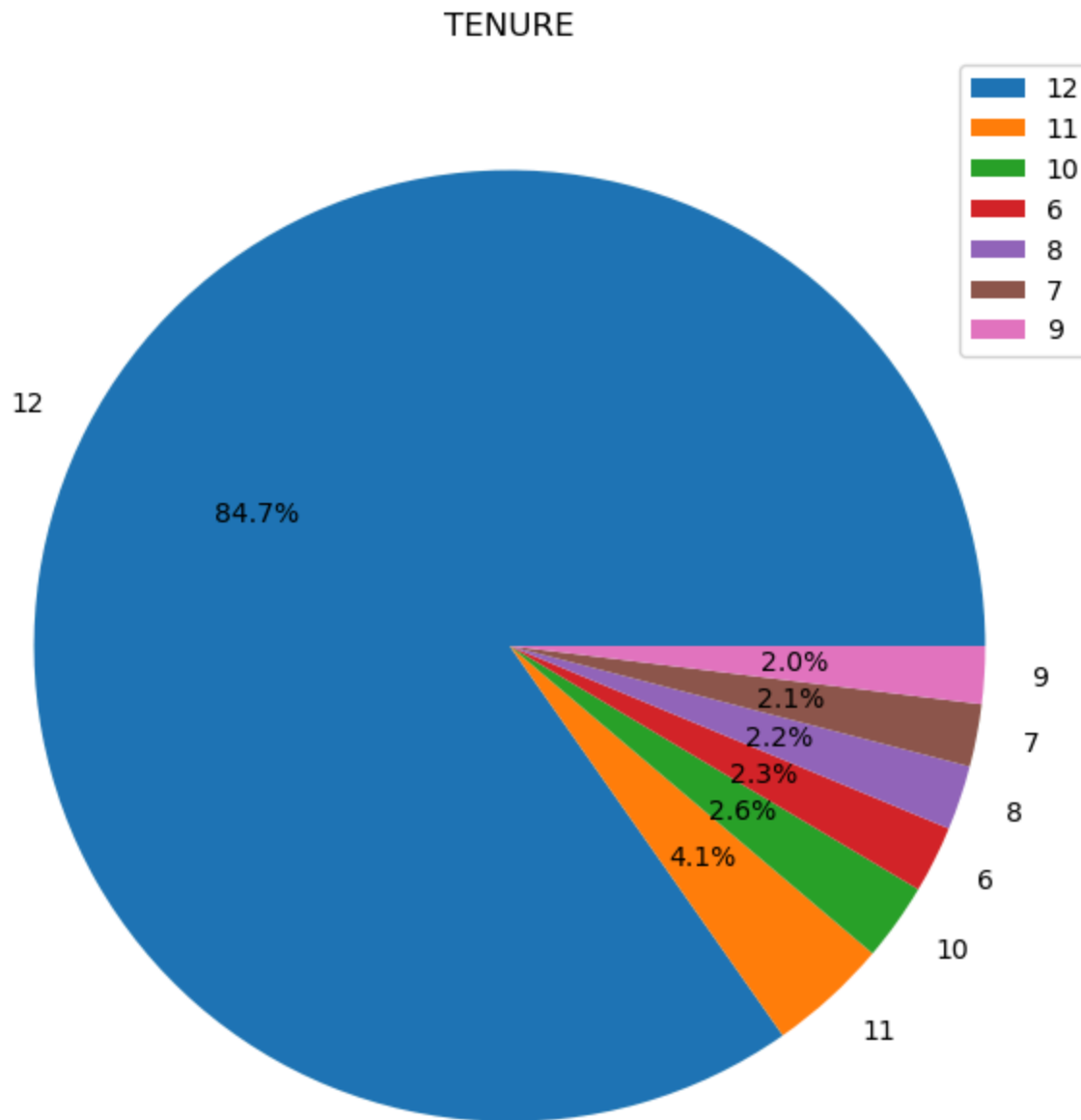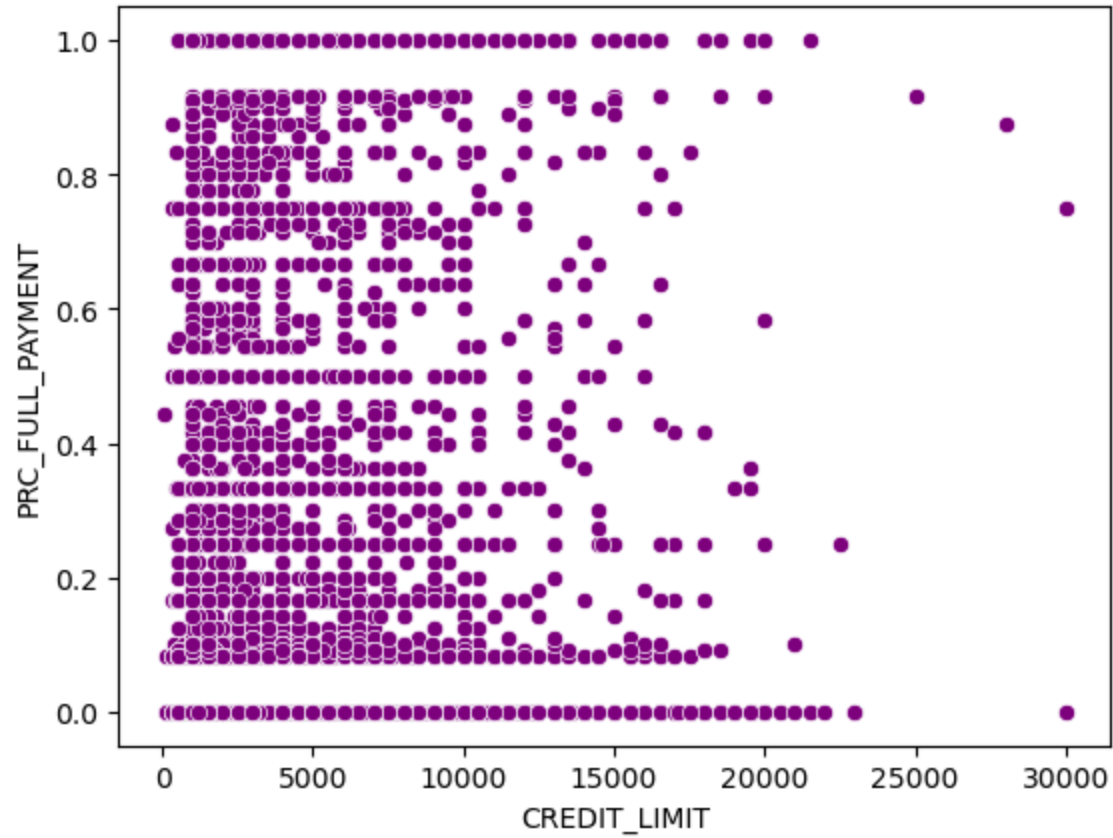
TENURE


```
In [7]: sns.boxplot(x='TENURE', y='CREDIT_LIMIT', data=data,palette='Pastel1');
```
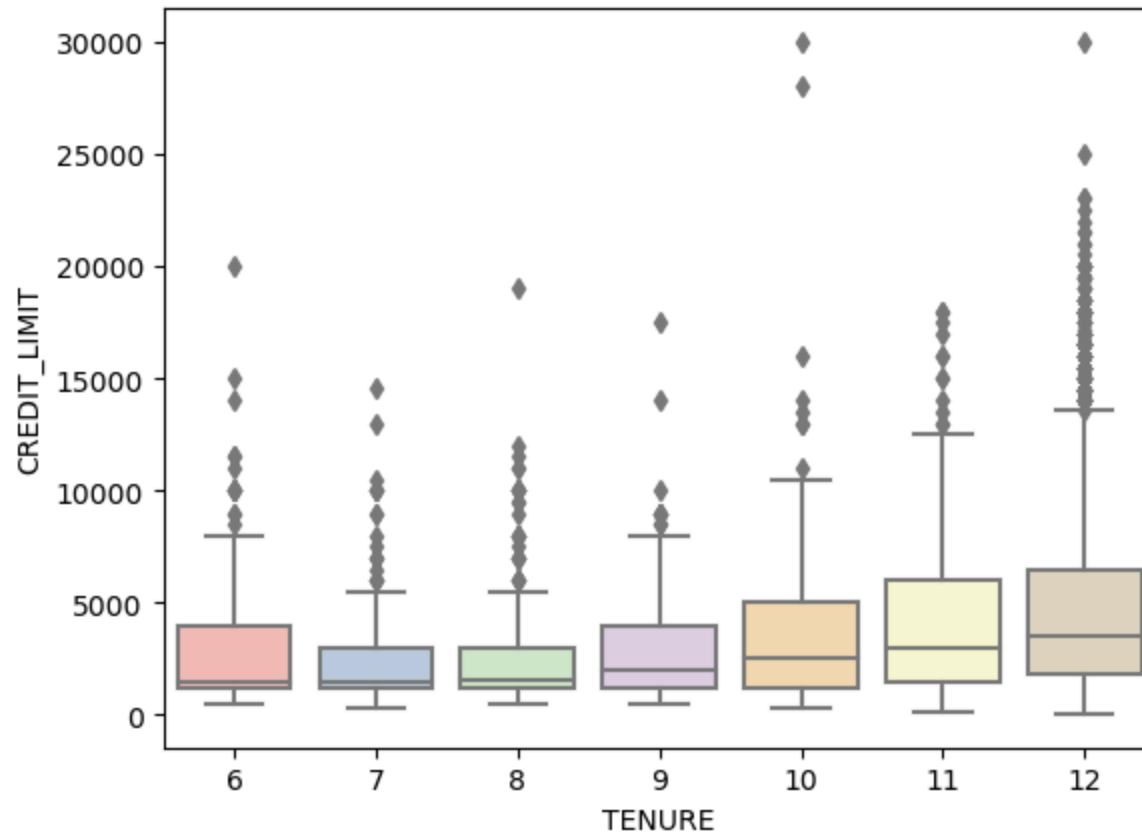
```
In [8]: sns.boxplot(x='TENURE', y='BALANCE_FREQUENCY', data=data,palette='autumn');
```
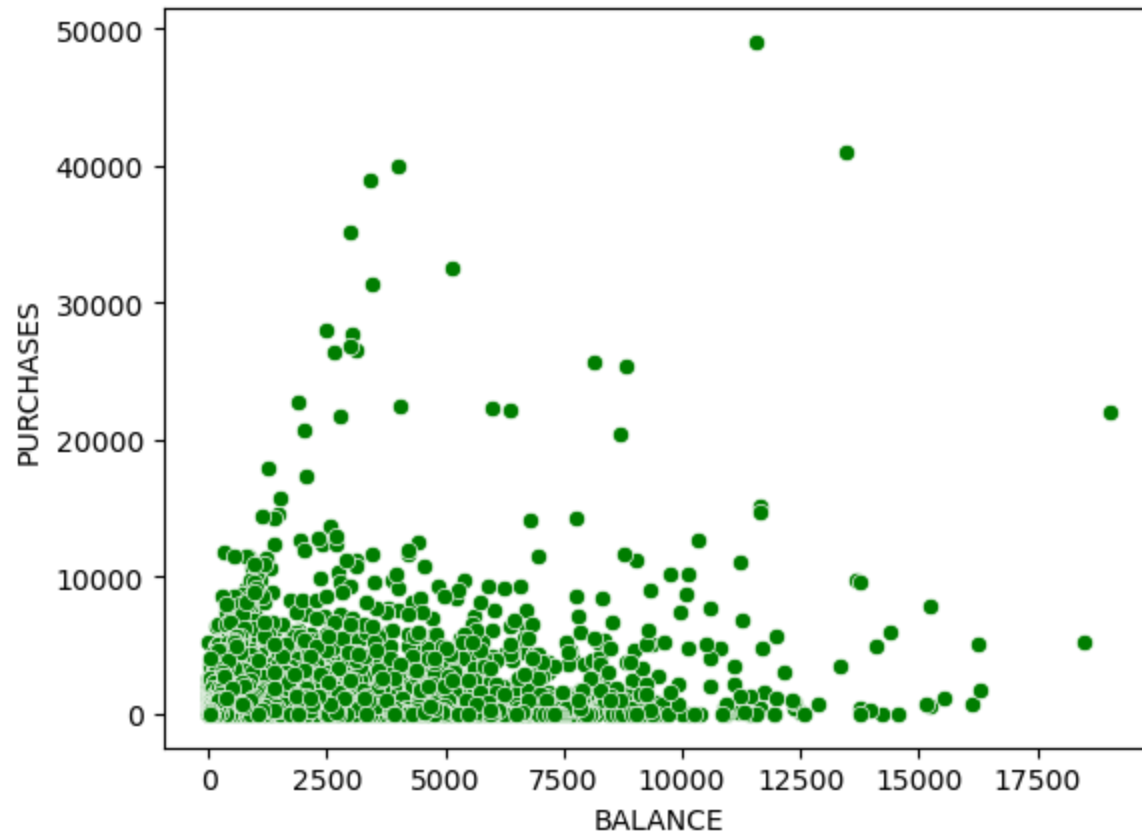
```
In [9]:   sns.scatterplot(x='CREDIT_LIMIT', y='PRC_FULL_PAYMENT', data=data, color='purple');
```
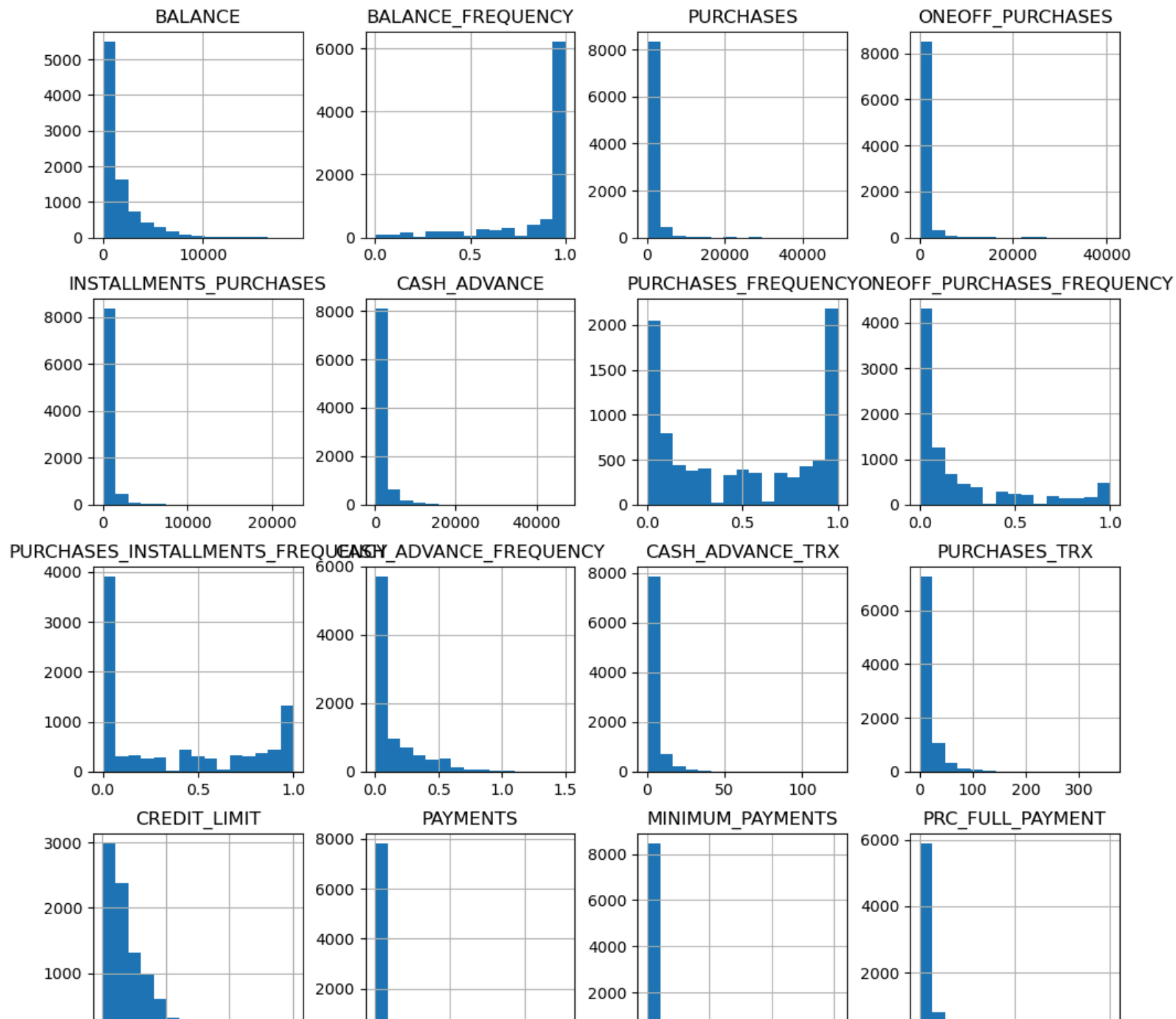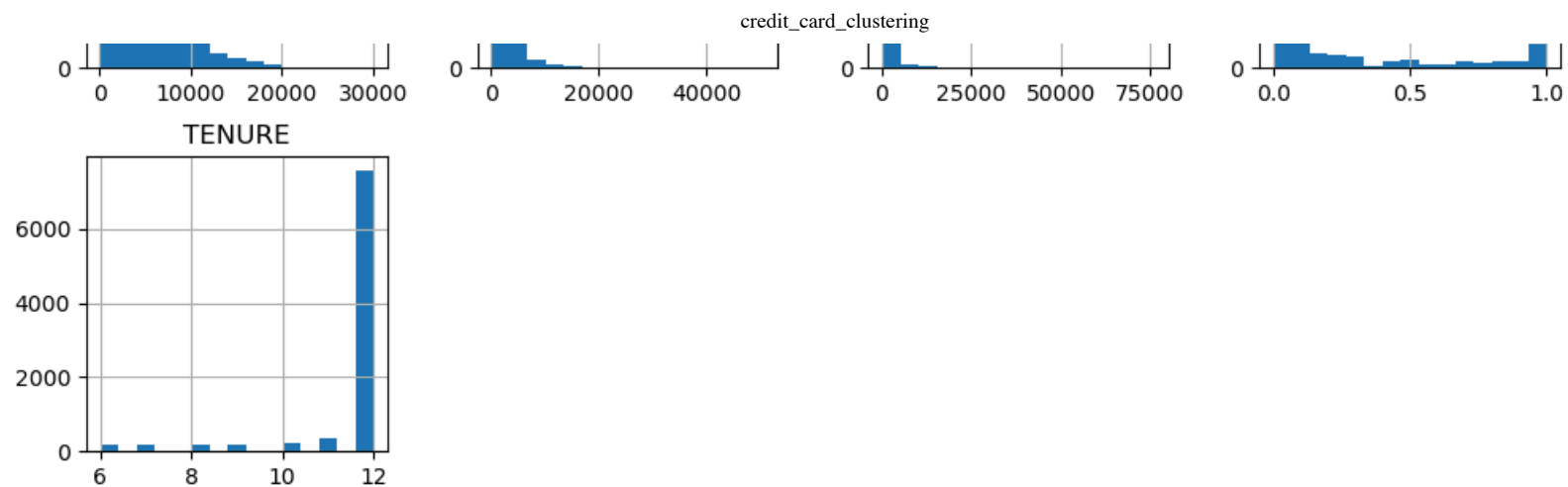
```
In [10]:  sns.scatterplot(x='BALANCE', y='PURCHASES', data=data, color='green')
```

```
Out[10]:  <Axes: xlabel='BALANCE', ylabel='PURCHASES'>
```

```
In [11]:   data[num_col].hist(bins=15, figsize=(12, 15), layout=(5, 4));
```

```
In [12]:  data[num_col].corr()
```

Out[12]:

	BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PUI
BALANCE	1.000000	0.322412	0.181261	0.164350	
BALANCE_FREQUENCY	0.322412	1.000000	0.133674	0.104323	
PURCHASES	0.181261	0.133674	1.000000	0.916845	
ONEOFF_PURCHASES	0.164350	0.104323	0.916845	1.000000	
INSTALLMENTS_PURCHASES	0.126469	0.124292	0.679896	0.330622	
CASH_ADVANCE	0.496692	0.099388	-0.051474	-0.031326	-
PURCHASES_FREQUENCY	-0.077944	0.229715	0.393017	0.264937	
ONEOFF_PURCHASES_FREQUENCY	0.073166	0.202415	0.498430	0.524891	
PURCHASES_INSTALLMENTS_FREQUENCY	-0.063186	0.176079	0.315567	0.127729	
CASH_ADVANCE_FREQUENCY	0.449218	0.191873	-0.120143	-0.082628	
CASH_ADVANCE_TRX	0.385152	0.141555	-0.067175	-0.046212	-
PURCHASES_TRX	0.154338	0.189626	0.689561	0.545523	
CREDIT_LIMIT	0.531283	0.095843	0.356963	0.319724	
PAYMENTS	0.322802	0.065008	0.603264	0.567292	
MINIMUM_PAYMENTS	0.398684	0.132569	0.093860	0.048755	
PRC_FULL_PAYMENT	-0.318959	-0.095082	0.180379	0.132763	
TENURE	0.072692	0.119776	0.086288	0.064150	

In [13]:
```python
plt.subplots(figsize=(20,15))
sns.heatmap(data[num_col].corr(),annot = True);
```
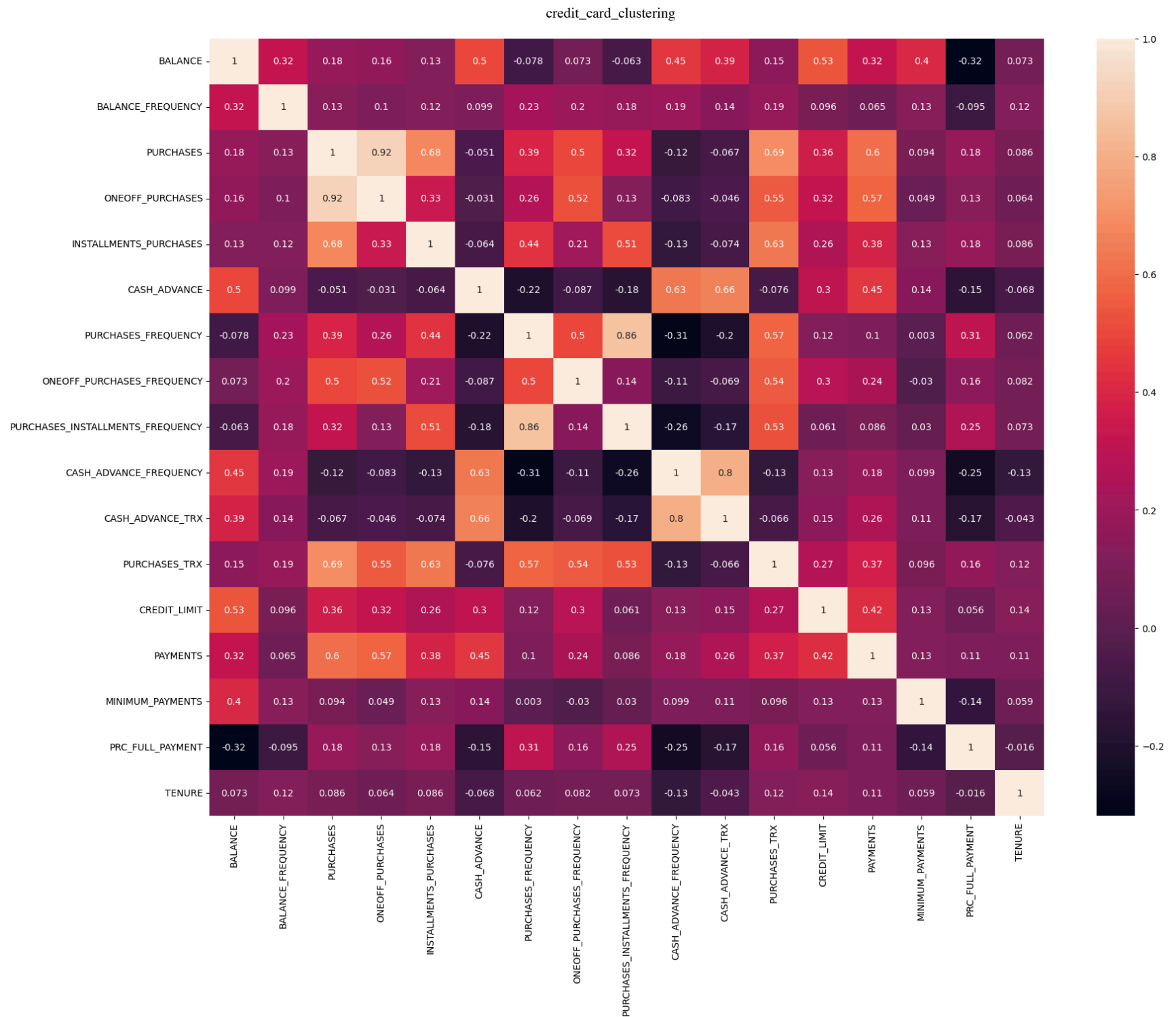
In [14]:
```python
from sklearn.impute import KNNImputer
imputer = KNNImputer()
imp_data = pd.DataFrame(imputer.fit_transform(data[num_col]), columns=data[num_col].columns)
imp_data.isna().sum()
```

Out[14]:
```
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                        0
PAYMENTS                            0
MINIMUM_PAYMENTS                    0
PRC_FULL_PAYMENT                    0
TENURE                              0
dtype: int64
```

In [15]:
```python
imp_data
```

Out[15]:

	BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES	CASH_ADVANCE	PUR(
0	40.900749	0.818182	95.40	0.00	95.40	0.000000	
1	3202.467416	0.909091	0.00	0.00	0.00	6442.945483	
2	2495.148862	1.000000	773.17	773.17	0.00	0.000000	
3	1666.670542	0.636364	1499.00	1499.00	0.00	205.788017	
4	817.714335	1.000000	16.00	16.00	0.00	0.000000	
...	...	...	...	...	...	...	
8945	28.493517	1.000000	291.12	0.00	291.12	0.000000	
8946	19.183215	1.000000	300.00	0.00	300.00	0.000000	
8947	23.398673	0.833333	144.40	0.00	144.40	0.000000	
8948	13.457564	0.833333	0.00	0.00	0.00	36.558778	
8949	372.708075	0.666667	1093.25	1093.25	0.00	127.040008	

8950 rows × 17 columns

In [16]:
```
pip install pycaret
```

```
Requirement already satisfied: pycaret in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages (3.3.
2)
Requirement already satisfied: ipython>=5.5.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-package
s (from pycaret) (8.15.0)
Requirement already satisfied: ipywidgets>=7.6.5 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-pack
ages (from pycaret) (8.0.4)
Requirement already satisfied: tqdm>=4.62.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages
(from pycaret) (4.65.0)
Requirement already satisfied: numpy<1.27,>=1.21 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-pack
ages (from pycaret) (1.24.3)
Requirement already satisfied: pandas<2.2.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages
(from pycaret) (2.0.3)
Requirement already satisfied: jinja2>=3 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages (fr
om pycaret) (3.1.2)
Requirement already satisfied: scipy<=1.11.4,>=1.6.1 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-
packages (from pycaret) (1.11.1)
Requirement already satisfied: joblib<1.4,>=1.2.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-pac
kages (from pycaret) (1.2.0)
Requirement already satisfied: scikit-learn>1.4.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-pac
kages (from pycaret) (1.4.2)
Requirement already satisfied: pyod>=1.1.3 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages
(from pycaret) (2.0.0)
Requirement already satisfied: imbalanced-learn>=0.12.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/si
te-packages (from pycaret) (0.12.3)
Requirement already satisfied: category-encoders>=2.4.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/si
te-packages (from pycaret) (2.6.3)
Requirement already satisfied: lightgbm>=3.0.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packag
es (from pycaret) (4.3.0)
Requirement already satisfied: numba>=0.55.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages
(from pycaret) (0.57.1)
Requirement already satisfied: requests>=2.27.1 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packa
ges (from pycaret) (2.31.0)
Requirement already satisfied: psutil>=5.9.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages
(from pycaret) (5.9.0)
Requirement already satisfied: markupsafe>=2.0.1 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-pack
ages (from pycaret) (2.1.1)
Requirement already satisfied: importlib-metadata>=4.12.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/
site-packages (from pycaret) (6.0.0)
Requirement already satisfied: nbformat>=4.2.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packag
es (from pycaret) (5.9.2)
Requirement already satisfied: cloudpickle in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages
(from pycaret) (2.2.1)
Requirement already satisfied: deprecation>=2.1.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-pac
kages (from pycaret) (2.1.0)
Requirement already satisfied: xxhash in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages (from
```

```
pycaret) (2.0.2)
Requirement already satisfied: matplotlib<3.8.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packa
ges (from pycaret) (3.7.2)
Requirement already satisfied: scikit-plot>=0.3.7 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-pac
kages (from pycaret) (0.3.7)
Requirement already satisfied: yellowbrick>=1.4 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packa
ges (from pycaret) (1.5)
Requirement already satisfied: plotly>=5.14.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-package
s (from pycaret) (5.22.0)
Requirement already satisfied: kaleido>=0.2.1 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-package
s (from pycaret) (0.2.1)
Requirement already satisfied: schemdraw==0.15 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packag
es (from pycaret) (0.15)
Requirement already satisfied: plotly-resampler>=0.8.3.1 in /Users/fabriciocamargo/anaconda3/lib/python3.11/s
ite-packages (from pycaret) (0.10.0)
Requirement already satisfied: statsmodels>=0.12.1 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-pa
ckages (from pycaret) (0.14.0)
Requirement already satisfied: sktime==0.26.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-package
s (from pycaret) (0.26.0)
Requirement already satisfied: tbats>=1.1.3 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages
(from pycaret) (1.1.3)
Requirement already satisfied: pmdarima>=2.0.4 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packag
es (from pycaret) (2.0.4)
Requirement already satisfied: wurlitzer in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages (fr
om pycaret) (3.0.2)
Requirement already satisfied: packaging in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages (fr
om sktime==0.26.0->pycaret) (23.1)
Requirement already satisfied: scikit-base<0.8.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-pack
ages (from sktime==0.26.0->pycaret) (0.7.8)
Requirement already satisfied: patsy>=0.5.1 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages
(from category-encoders>=2.4.0->pycaret) (0.5.3)
Requirement already satisfied: threadpoolctl>=2.0.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-p
ackages (from imbalanced-learn>=0.12.0->pycaret) (2.2.0)
Requirement already satisfied: zipp>=0.5 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages (fr
om importlib-metadata>=4.12.0->pycaret) (3.11.0)
Requirement already satisfied: backcall in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages (fro
m ipython>=5.5.0->pycaret) (0.2.0)
Requirement already satisfied: decorator in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages (fr
om ipython>=5.5.0->pycaret) (5.1.1)
Requirement already satisfied: jedi>=0.16 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages (f
rom ipython>=5.5.0->pycaret) (0.18.1)
Requirement already satisfied: matplotlib-inline in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-pack
ages (from ipython>=5.5.0->pycaret) (0.1.6)
Requirement already satisfied: pickleshare in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages
(from ipython>=5.5.0->pycaret) (0.7.5)
```

```
Requirement already satisfied: prompt-toolkit!=3.0.37,<3.1.0,>=3.0.30 in /Users/fabriciocamargo/anaconda3/li
b/python3.11/site-packages (from ipython>=5.5.0->pycaret) (3.0.36)
Requirement already satisfied: pygments>=2.4.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packag
es (from ipython>=5.5.0->pycaret) (2.15.1)
Requirement already satisfied: stack-data in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages (f
rom ipython>=5.5.0->pycaret) (0.2.0)
Requirement already satisfied: traitlets>=5 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages
(from ipython>=5.5.0->pycaret) (5.7.1)
Requirement already satisfied: pexpect>4.3 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages
(from ipython>=5.5.0->pycaret) (4.8.0)
Requirement already satisfied: appnope in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages (from
ipython>=5.5.0->pycaret) (0.1.2)
Requirement already satisfied: ipykernel>=4.5.1 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packa
ges (from ipywidgets>=7.6.5->pycaret) (6.25.0)
Requirement already satisfied: widgetsnbextension~=4.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/sit
e-packages (from ipywidgets>=7.6.5->pycaret) (4.0.5)
Requirement already satisfied: jupyterlab-widgets~=3.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/sit
e-packages (from ipywidgets>=7.6.5->pycaret) (3.0.5)
Requirement already satisfied: contourpy>=1.0.1 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packa
ges (from matplotlib<3.8.0->pycaret) (1.0.5)
Requirement already satisfied: cycler>=0.10 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages
(from matplotlib<3.8.0->pycaret) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-pack
ages (from matplotlib<3.8.0->pycaret) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-pack
ages (from matplotlib<3.8.0->pycaret) (1.4.4)
Requirement already satisfied: pillow>=6.2.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages
(from matplotlib<3.8.0->pycaret) (9.4.0)
Requirement already satisfied: pyparsing<3.1,>=2.3.1 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-
packages (from matplotlib<3.8.0->pycaret) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-p
ackages (from matplotlib<3.8.0->pycaret) (2.8.2)
Requirement already satisfied: fastjsonschema in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-package
s (from nbformat>=4.2.0->pycaret) (2.16.2)
Requirement already satisfied: jsonschema>=2.6 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packag
es (from nbformat>=4.2.0->pycaret) (4.17.3)
Requirement already satisfied: jupyter-core in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages
(from nbformat>=4.2.0->pycaret) (5.3.0)
Requirement already satisfied: llvmlite<0.41,>=0.40.0dev0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/
site-packages (from numba>=0.55.0->pycaret) (0.40.0)
Requirement already satisfied: pytz>=2020.1 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages
(from pandas<2.2.0->pycaret) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-package
s (from pandas<2.2.0->pycaret) (2023.3)
Requirement already satisfied: tenacity>=6.2.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packag
```

```
es (from plotly>=5.14.0->pycaret) (8.2.2)
Requirement already satisfied: dash>=2.9.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages
(from plotly-resampler>=0.8.3.1->pycaret) (2.17.0)
Requirement already satisfied: orjson<4.0.0,>=3.8.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-p
ackages (from plotly-resampler>=0.8.3.1->pycaret) (3.10.3)
Requirement already satisfied: tsdownsample>=0.1.3 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-pa
ckages (from plotly-resampler>=0.8.3.1->pycaret) (0.1.3)
Requirement already satisfied: Cython!=0.29.18,!=0.29.31,>=0.29 in /Users/fabriciocamargo/anaconda3/lib/pytho
n3.11/site-packages (from pmdarima>=2.0.4->pycaret) (3.0.10)
Requirement already satisfied: urllib3 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages (from
pmdarima>=2.0.4->pycaret) (1.26.16)
Requirement already satisfied: setuptools!=50.0.0,>=38.6.0 in /Users/fabriciocamargo/anaconda3/lib/python3.1
1/site-packages (from pmdarima>=2.0.4->pycaret) (68.0.0)
Requirement already satisfied: charset-normalizer<4,>=2 in /Users/fabriciocamargo/anaconda3/lib/python3.11/si
te-packages (from requests>=2.27.1->pycaret) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages
(from requests>=2.27.1->pycaret) (3.4)
Requirement already satisfied: certifi>=2017.4.17 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-pac
kages (from requests>=2.27.1->pycaret) (2023.7.22)
Requirement already satisfied: Flask<3.1,>=1.0.4 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-pack
ages (from dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (2.2.2)
Requirement already satisfied: Werkzeug<3.1 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages
(from dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (2.2.3)
Requirement already satisfied: dash-html-components==2.0.0 in /Users/fabriciocamargo/anaconda3/lib/python3.1
1/site-packages (from dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (2.0.0)
Requirement already satisfied: dash-core-components==2.0.0 in /Users/fabriciocamargo/anaconda3/lib/python3.1
1/site-packages (from dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (2.0.0)
Requirement already satisfied: dash-table==5.0.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-pack
ages (from dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (5.0.0)
Requirement already satisfied: typing-extensions>=4.1.1 in /Users/fabriciocamargo/anaconda3/lib/python3.11/si
te-packages (from dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (4.7.1)
Requirement already satisfied: retrying in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages (fro
m dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (1.3.4)
Requirement already satisfied: nest-asyncio in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages
(from dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (1.5.6)
Requirement already satisfied: comm>=0.1.1 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages
(from ipykernel>=4.5.1->ipywidgets>=7.6.5->pycaret) (0.1.2)
Requirement already satisfied: debugpy>=1.6.5 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-package
s (from ipykernel>=4.5.1->ipywidgets>=7.6.5->pycaret) (1.6.7)
Requirement already satisfied: jupyter-client>=6.1.12 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site
-packages (from ipykernel>=4.5.1->ipywidgets>=7.6.5->pycaret) (7.4.9)
Requirement already satisfied: pyzmq>=20 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages (fr
om ipykernel>=4.5.1->ipywidgets>=7.6.5->pycaret) (23.2.0)
Requirement already satisfied: tornado>=6.1 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages
(from ipykernel>=4.5.1->ipywidgets>=7.6.5->pycaret) (6.3.2)
```

```
Requirement already satisfied: parso<0.9.0,>=0.8.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-pa
ckages (from jedi>=0.16->ipython>=5.5.0->pycaret) (0.8.3)
Requirement already satisfied: attrs>=17.4.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages
(from jsonschema>=2.6->nbformat>=4.2.0->pycaret) (22.1.0)
Requirement already satisfied: pyrsistent!=0.17.0,!=0.17.1,!=0.17.2,>=0.14.0 in /Users/fabriciocamargo/anacon
da3/lib/python3.11/site-packages (from jsonschema>=2.6->nbformat>=4.2.0->pycaret) (0.18.0)
Requirement already satisfied: platformdirs>=2.5 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-pack
ages (from jupyter-core->nbformat>=4.2.0->pycaret) (3.10.0)
Requirement already satisfied: six in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages (from pat
sy>=0.5.1->category-encoders>=2.4.0->pycaret) (1.16.0)
Requirement already satisfied: ptyprocess>=0.5 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packag
es (from pexpect>4.3->ipython>=5.5.0->pycaret) (0.7.0)
Requirement already satisfied: wcwidth in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages (from
prompt-toolkit!=3.0.37,<3.1.0,>=3.0.30->ipython>=5.5.0->pycaret) (0.2.5)
Requirement already satisfied: executing in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages (fr
om stack-data->ipython>=5.5.0->pycaret) (0.8.3)
Requirement already satisfied: asttokens in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages (fr
om stack-data->ipython>=5.5.0->pycaret) (2.0.5)
Requirement already satisfied: pure-eval in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages (fr
om stack-data->ipython>=5.5.0->pycaret) (0.2.2)
Requirement already satisfied: itsdangerous>=2.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-pack
ages (from Flask<3.1,>=1.0.4->dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (2.0.1)
Requirement already satisfied: click>=8.0 in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages (f
rom Flask<3.1,>=1.0.4->dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (8.0.4)
Requirement already satisfied: entrypoints in /Users/fabriciocamargo/anaconda3/lib/python3.11/site-packages
(from jupyter-client>=6.1.12->ipykernel>=4.5.1->ipywidgets>=7.6.5->pycaret) (0.4)

[notice] A new release of pip is available: 23.3.1 -> 24.0
[notice] To update, run: pip install --upgrade pip
Note: you may need to restart the kernel to use updated packages.
```

In [17]:
```python
from pycaret.clustering import *
```

In [18]:
```python
clu = setup(data=imp_data, normalize=True,
            pca=True,
            remove_multicollinearity=True,
            session_id=123)
```

	Description	Value
0	Session id	123
1	Original data shape	(8950, 17)
2	Transformed data shape	(8950, 16)
3	Numeric features	17
4	Preprocess	True
5	Imputation type	simple
6	Numeric imputation	mean
7	Categorical imputation	mode
8	Remove multicollinearity	True
9	Multicollinearity threshold	0.900000
10	Normalize	True
11	Normalize method	zscore
12	PCA	True
13	PCA method	linear
14	PCA components	None
15	CPU Jobs	-1
16	Use GPU	False
17	Log Experiment	False
18	Experiment Name	cluster-default-name
19	USI	8c5f

In [19]:
```python
models()
```

Out[19]:

ID	Name	Reference
kmeans	K-Means Clustering	sklearn.cluster._kmeans.KMeans
ap	Affinity Propagation	sklearn.cluster._affinity_propagation.Affinity...
meanshift	Mean Shift Clustering	sklearn.cluster._mean_shift.MeanShift
sc	Spectral Clustering	sklearn.cluster._spectral.SpectralClustering
hclust	Agglomerative Clustering	sklearn.cluster._agglomerative.AgglomerativeCl...
dbscan	Density-Based Spatial Clustering	sklearn.cluster._dbscan.DBSCAN
optics	OPTICS Clustering	sklearn.cluster._optics.OPTICS
birch	Birch Clustering	sklearn.cluster._birch.Birch

```python
In [20]:   kmeans = create_model('kmeans')
```
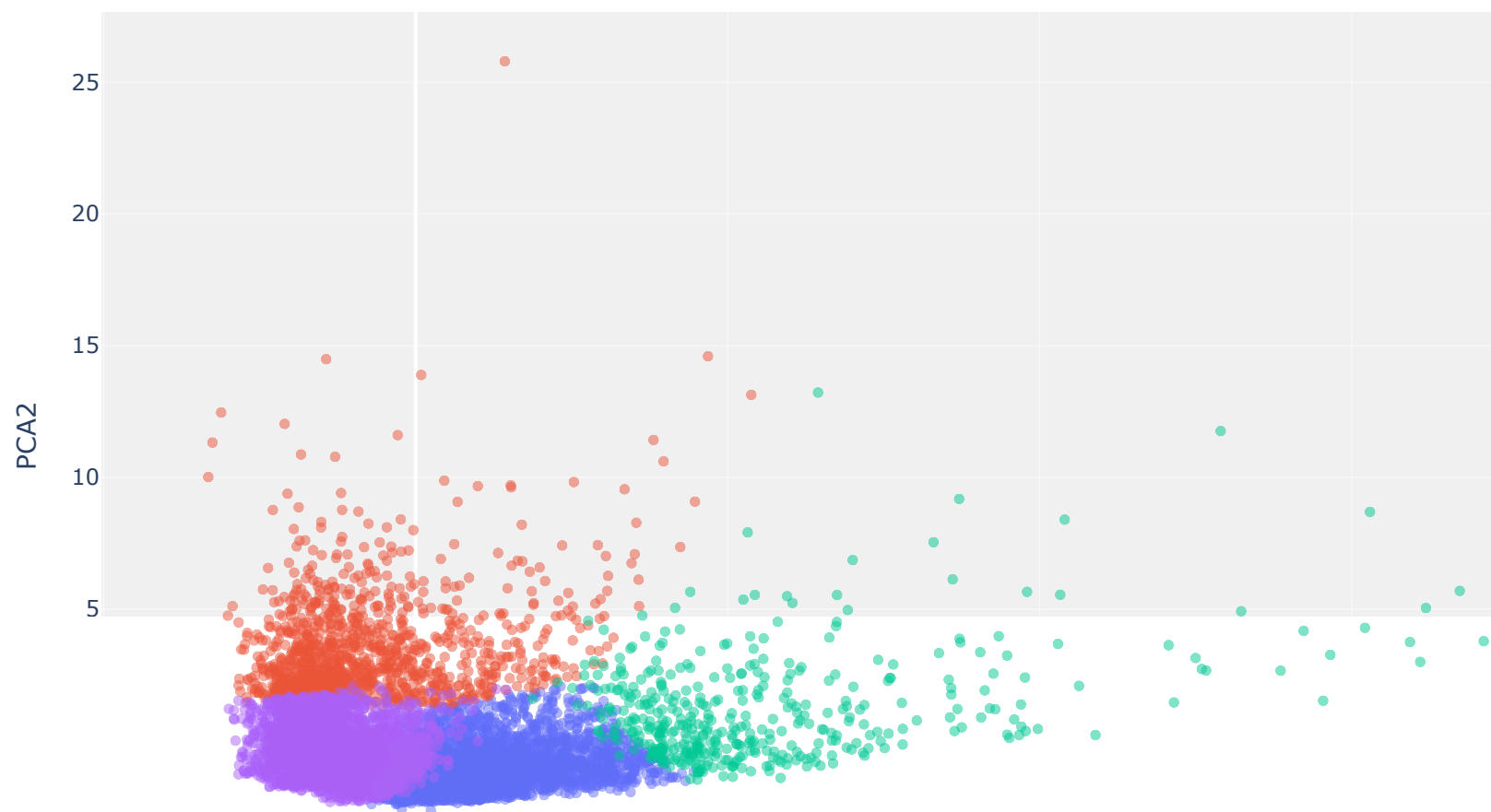
	Silhouette	Calinski-Harabasz	Davies-Bouldin	Homogeneity	Rand Index	Completeness
0	0.1987	1604.1436	1.5876	0	0	0

```python
In [21]:   kmean_results = assign_model(kmeans)
           kmean_results.head()
```

Out[21]:

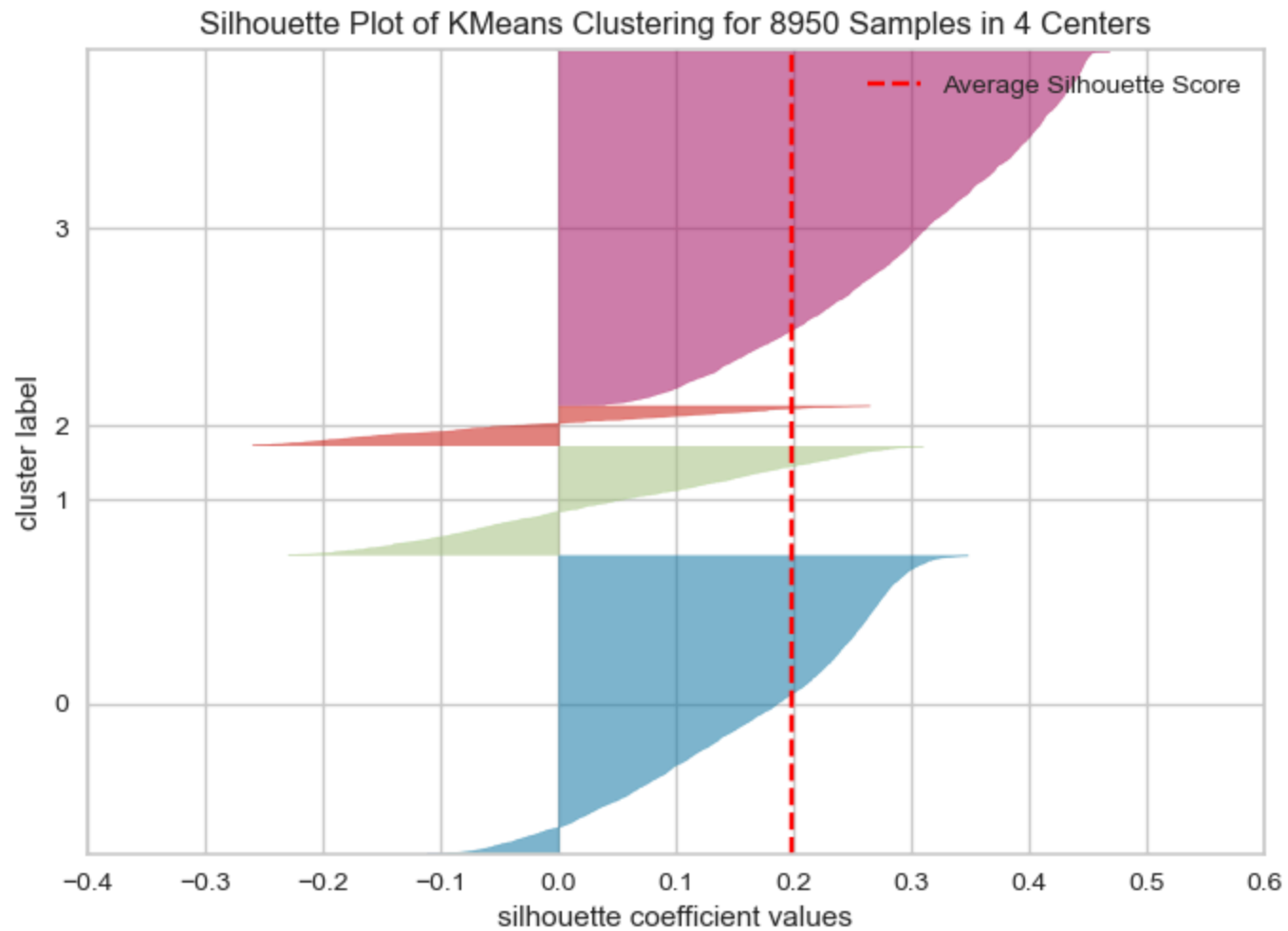|	BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES	CASH_ADVANCE	PURCHA
0	40.900749	0.818182	95.400002	0.000000	95.400002	0.000000	
1	3202.467529	0.909091	0.000000	0.000000	0.000000	6442.945312	
2	2495.148926	1.000000	773.169983	773.169983	0.000000	0.000000	
3	1666.670532	0.636364	1499.000000	1499.000000	0.000000	205.788010	
4	817.714355	1.000000	16.000000	16.000000	0.000000	0.000000	

In [22]:
```python
plot_model(kmeans)
```
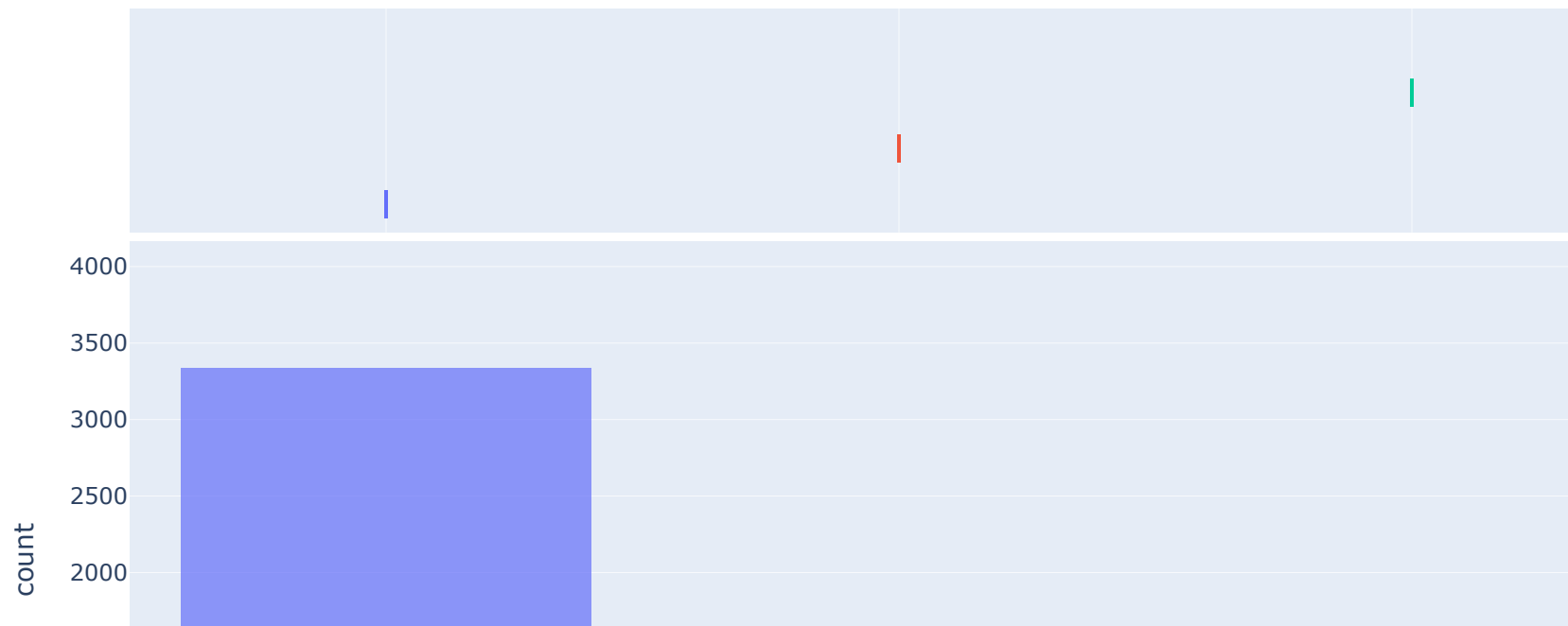
2D Cluster PCA Plot

In [23]:
```python
plot_model(kmeans, plot = 'elbow')
```

Distortion Score Elbow for KMeans Clustering
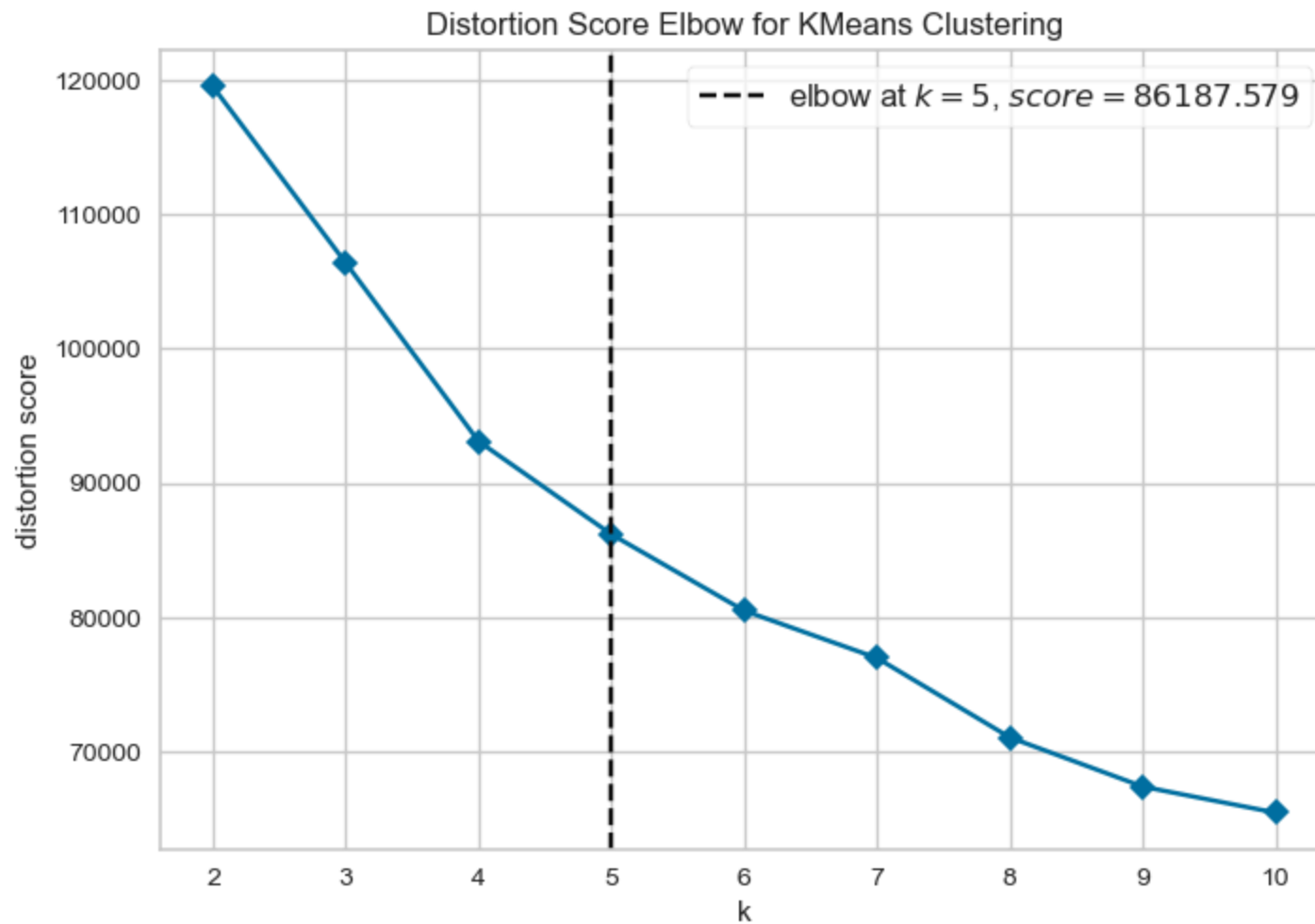
In [24]:
```
plot_model(kmeans, plot = 'silhouette')
```

Silhouette Plot of KMeans Clustering for 8950 Samples in 4 Centers

```
In [25]:  plot_model(kmeans, plot = 'distribution') #to see size of clusters
```

```
In [26]:  plot_model(kmeans, plot = 'distribution', feature = 'TENURE')
```
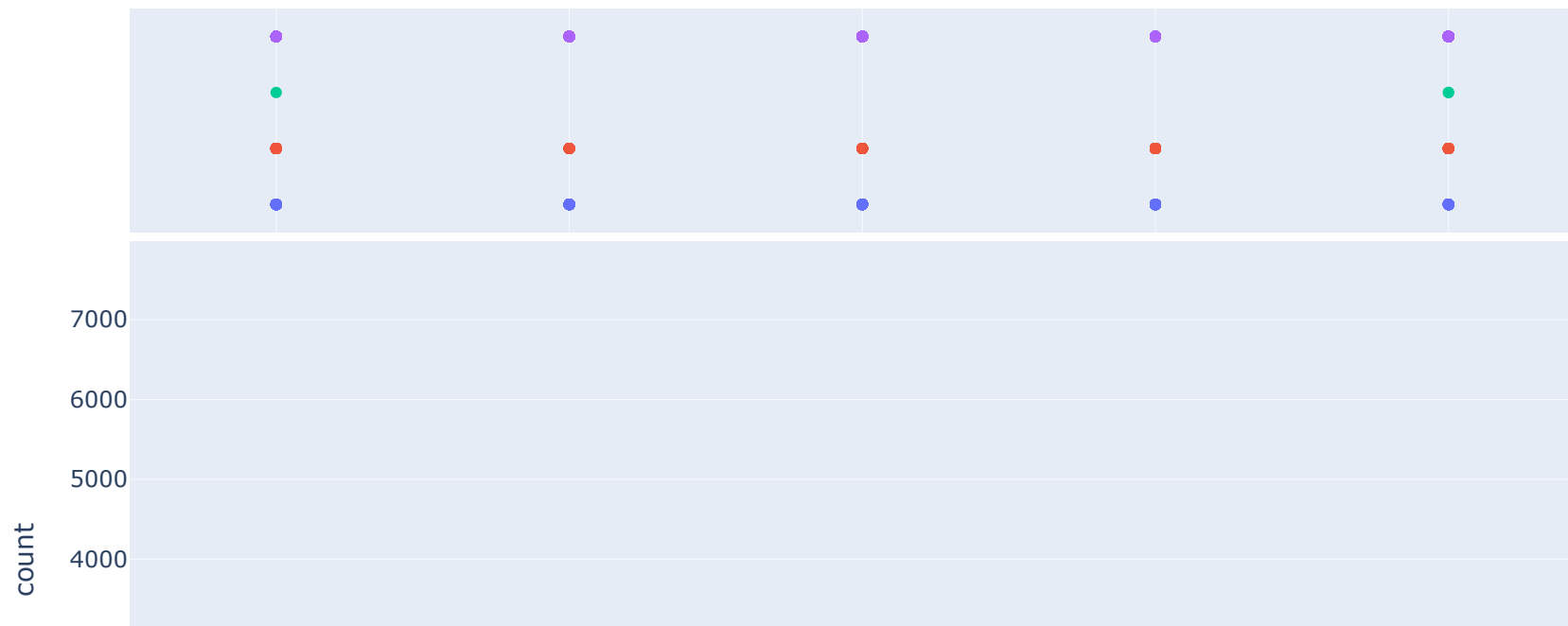
```
In [27]:  plot_model(kmeans, plot = 'distribution', feature = 'PAYMENTS')
```

```
In [28]: plot_model(kmeans, plot ='tsne')
```

3d TSNE Plot for Clusters