```
# install the full version of pycaret
!pip install pycaret[full]
```

```
  Stored in directory: /root/.cache/pip/wheels/43/07/ac/7c5a9d708d65247ac1f94066cf1d
  Building wheel for alembic (setup.py) ... done
  Created wheel for alembic: filename=alembic-1.4.1-py2.py3-none-any.whl size=158155
  Stored in directory: /root/.cache/pip/wheels/84/07/f7/12f7370ca47a66030c2edeedcc23
  Building wheel for prometheus-flask-exporter (setup.py) ... done
  Created wheel for prometheus-flask-exporter: filename=prometheus_flask_exporter-0.
  Stored in directory: /root/.cache/pip/wheels/c0/e2/9c/4f3ee23964802940f81a8b476d0b
  Building wheel for databricks-cli (setup.py) ... done
  Created wheel for databricks-cli: filename=databricks_cli-0.14.3-cp37-none-any.whl
  Stored in directory: /root/.cache/pip/wheels/5b/24/f3/34d8e3964dac4ba849d844273c49
  Building wheel for gpustat (setup.py) ... done
  Created wheel for gpustat: filename=gpustat-0.6.0-cp37-none-any.whl size=12621 sha
  Stored in directory: /root/.cache/pip/wheels/48/b4/d5/fb5b7f1d040f2ff20687e3bad686
  Building wheel for pyperclip (setup.py) ... done
  Created wheel for pyperclip: filename=pyperclip-1.8.2-cp37-none-any.whl size=11107
  Stored in directory: /root/.cache/pip/wheels/25/af/b8/3407109267803f4015e1ee2ff23b
Successfully built umap-learn pyod shap pynndescent phik htmlmin alembic prometheus-
ERROR: google-colab 1.0.0 has requirement requests~=2.23.0, but you'll have requests
ERROR: datascience 0.10.6 has requirement folium==0.2.1, but you'll have folium 0.8.
ERROR: phik 0.11.2 has requirement scipy>=1.5.2, but you'll have scipy 1.4.1 which i
ERROR: pyldavis 3.3.1 has requirement numpy>=1.20.0, but you'll have numpy 1.19.5 wh
ERROR: pyldavis 3.3.1 has requirement pandas>=1.2.0, but you'll have pandas 1.1.5 wh
ERROR: ray 1.3.0 has requirement protobuf>=3.15.3, but you'll have protobuf 3.12.4 w
ERROR: botocore 1.20.70 has requirement urllib3<1.27,>=1.25.4, but you'll have urlli
ERROR: awscli 1.19.70 has requirement colorama<0.4.4,>=0.2.5, but you'll have colora
Installing collected packages: threadpoolctl, scikit-learn, mlxtend, pynndescent, um
  Found existing installation: scikit-learn 0.22.2.post1
    Uninstalling scikit-learn-0.22.2.post1:
      Successfully uninstalled scikit-learn-0.22.2.post1
  Found existing installation: mlxtend 0.14.0

    Uninstalling mlxtend-0.14.0:
      Successfully uninstalled mlxtend-0.14.0
  Found existing installation: yellowbrick 0.9.1
    Uninstalling yellowbrick-0.9.1:
      Successfully uninstalled yellowbrick-0.9.1
  Found existing installation: requests 2.23.0
    Uninstalling requests-2.23.0:
      Successfully uninstalled requests-2.23.0
  Found existing installation: tqdm 4.41.1
    Uninstalling tqdm-4.41.1:
      Successfully uninstalled tqdm-4.41.1
  Found existing installation: pandas-profiling 1.4.1
    Uninstalling pandas-profiling-1.4.1:
      Successfully uninstalled pandas-profiling-1.4.1
  Found existing installation: lightgbm 2.2.3
    Uninstalling lightgbm-2.2.3:
      Successfully uninstalled lightgbm-2.2.3
  Found existing installation: imbalanced-learn 0.4.3
    Uninstalling imbalanced-learn-0.4.3:
      Successfully uninstalled imbalanced-learn-0.4.3
  Found existing installation: xgboost 0.90
    Uninstalling xgboost-0.90:
      Successfully uninstalled xgboost-0.90
```

```
    Successfully uninstalled xgboost-0.90
  Found existing installation: docutils 0.17
    Uninstalling docutils-0.17:
      Successfully uninstalled docutils-0.17
Successfully installed Boruta-0.3 Mako-1.1.4 aiohttp-3.7.4.post0 aiohttp-cors-0.7.0
```

This case requires to develop a customer segmentation to define marketing strategy. The sample Dataset summarizes the usage behavior of about 9000 active credit card holders during the last 6 months. The file is at a customer level with 18 behavioral variables.

Following is the Data Dictionary for Credit Card dataset :-

CUSTID : Identification of Credit Card holder (Categorical) BALANCE : Balance amount left in their account to make purchases (BALANCEFREQUENCY : How frequently the Balance is updated, score between 0 and 1 (1 = frequently updated, 0 = not frequently updated) PURCHASES : Amount of purchases made from account ONEOFFPURCHASES : Maximum purchase amount done in one-go INSTALLMENTSPURCHASES : Amount of purchase done in installment CASHADVANCE : Cash in advance given by the user PURCHASESFREQUENCY : How frequently the Purchases are being made, score between 0 and 1 (1 = frequently purchased, 0 = not frequently purchased) ONEOFFPURCHASESFREQUENCY : How frequently Purchases are happening in one-go (1 = frequently purchased, 0 = not frequently purchased) PURCHASESINSTALLMENTSFREQUENCY : How frequently purchases in installments are being done (1 = frequently done, 0 = not frequently done) CASHADVANCEFREQUENCY : How frequently the cash in advance being paid CASHADVANCETRX : Number of Transactions made with "Cash in Advanced" PURCHASESTRX : Numbe of purchase transactions made CREDITLIMIT : Limit of Credit Card for user PAYMENTS : Amount of Payment done by user MINIMUM_PAYMENTS : Minimum amount of payments made by user PRCFULLPAYMENT : Percent of full payment paid by user TENURE : Tenure of credit card service for user

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

```
df = pd.read_csv("/content/drive/MyDrive/Colab Notebooks/kaggle/csv/CC GENERAL.csv")
```

```
df.head()
```

	CUST_ID	BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PU
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	

```
!pip install dataprep
!pip install "dask[complete]"
!pip install "dask[delayed]"
```

```
Requirement already satisfied: webencodings in /usr/local/lib/python3.7/dist-package
Installing collected packages: ply, jsonpath-ng, regex, nltk, wordcloud, locket, par
  Found existing installation: regex 2019.12.20
    Uninstalling regex-2019.12.20:
      Successfully uninstalled regex-2019.12.20
  Found existing installation: nltk 3.2.5
    Uninstalling nltk-3.2.5:
      Successfully uninstalled nltk-3.2.5
  Found existing installation: wordcloud 1.5.0
    Uninstalling wordcloud-1.5.0:

      Successfully uninstalled wordcloud-1.5.0
  Found existing installation: dask 2.12.0
    Uninstalling dask-2.12.0:
      Successfully uninstalled dask-2.12.0
Successfully installed dask-2.30.0 dataprep-0.2.15 fsspec-2021.4.0 jsonpath-ng-1.5.2
Requirement already satisfied: dask[complete] in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: pyyaml in /usr/local/lib/python3.7/dist-packages (fro
Requirement already satisfied: bokeh!=2.0.0,>=1.0.0; extra == "complete" in /usr/loc
Requirement already satisfied: partd>=0.3.10; extra == "complete" in /usr/local/lib/
Requirement already satisfied: fsspec>=0.6.0; extra == "complete" in /usr/local/lib/
Requirement already satisfied: pandas>=0.23.0; extra == "complete" in /usr/local/lib
Requirement already satisfied: numpy>=1.13.0; extra == "complete" in /usr/local/lib/
Requirement already satisfied: cloudpickle>=0.2.2; extra == "complete" in /usr/local
Collecting distributed>=2.0; extra == "complete"
  Downloading https://files.pythonhosted.org/packages/63/f8/ac2c18adde6477bca3881c4d
     |████████████████████████████████| 706kB 10.3MB/s
Requirement already satisfied: toolz>=0.8.2; extra == "complete" in /usr/local/lib/p
Requirement already satisfied: Jinja2>=2.7 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: tornado>=5.1 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist
Requirement already satisfied: typing-extensions>=3.7.4 in /usr/local/lib/python3.7/
Requirement already satisfied: packaging>=16.8 in /usr/local/lib/python3.7/dist-pack
Requirement already satisfied: pillow>=7.1.0 in /usr/local/lib/python3.7/dist-packag
Requirement already satisfied: locket in /usr/local/lib/python3.7/dist-packages (fro
Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: msgpack>=0.6.0 in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: tblib>=1.6.0 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: psutil>=5.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: click>=6.6 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: zict>=0.1.3 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: sortedcontainers!=2.0.0,!=2.0.1 in /usr/local/lib/pyt
Requirement already satisfied: MarkupSafe>=0.23 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (f
Requirement already satisfied: pyparsing>=2.0.2 in /usr/local/lib/python3.7/dist-pac
```

```
Requirement already satisfied: heapdict in /usr/local/lib/python3.7/dist-packages (f
ERROR: distributed 2021.4.1 has requirement cloudpickle>=1.5.0, but you'll have clou
ERROR: distributed 2021.4.1 has requirement dask>=2021.03.0, but you'll have dask 2.
Installing collected packages: distributed
  Found existing installation: distributed 1.25.3
    Uninstalling distributed-1.25.3:
      Successfully uninstalled distributed-1.25.3
Successfully installed distributed-2021.4.1
Requirement already satisfied: dask[delayed] in /usr/local/lib/python3.7/dist-packag
Requirement already satisfied: pyyaml in /usr/local/lib/python3.7/dist-packages (fro
Requirement already satisfied: cloudpickle>=0.2.2; extra == "delayed" in /usr/local/
Requirement already satisfied: toolz>=0.8.2; extra == "delayed" in /usr/local/lib/py
```

```
from dataprep.eda import create_report,plot,plot_correlation,plot_missing
```

```
NumExpr defaulting to 2 threads.
```

```
create_report(df)
```

DataPrep Report | Overview | Variables ≡ | Interactions | Correlations

Missing Values

Number of Rows	8950
Missing Cells	314
Missing Cells (%)	0.2%
Duplicate Rows	0
Duplicate Rows (%)	0.0%
Total Size in Memory	1.7 MB
Average Row Size in Memory	199.0 B

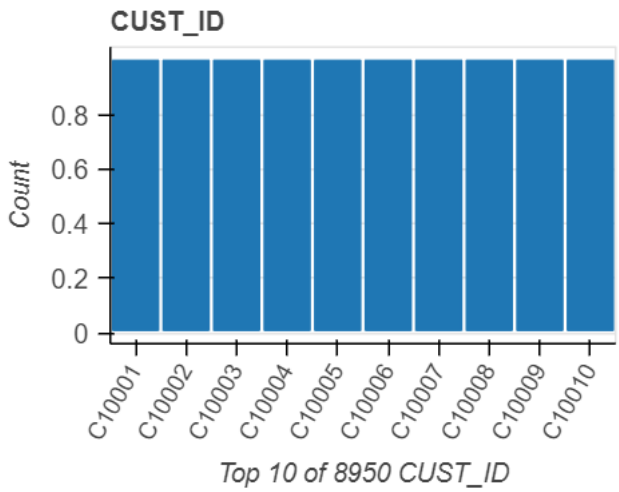
Numerical	17

Variables

CUST_ID
`categorical`

Show Details

Distinct Count	8950
Unique (%)	100.0%
Missing	0
Missing (%)	0.0%
Memory Size	620.6 KB

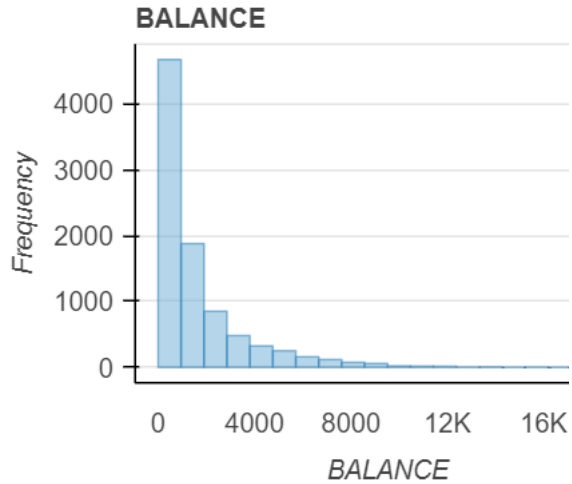


ANCE
.cal

ow Details

Distinct Count	8871
Unique (%)	99.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%

Memory Size	139.8 KB
Mean	1564.4748
Minimum	0
Maximum	19043.1386
Zeros	80
Zeros (%)	0.9%
Negatives	0
Negatives (%)	0.0%



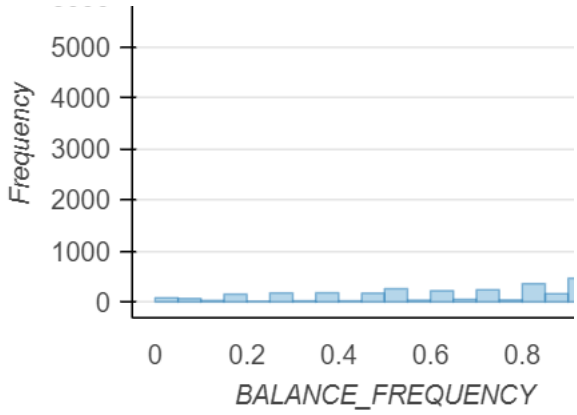
Distinct Count	43

Memory Size	139.8 KB

ANCE_FREQUE…
ical
ow Details

Unique (%)	0.5%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%

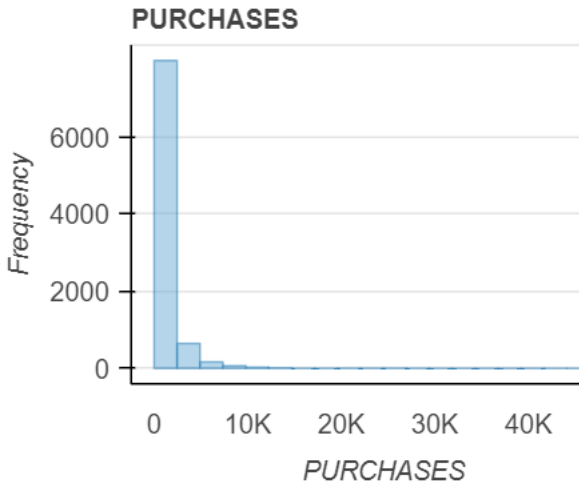
Mean	0.8773
Minimum	0
Maximum	1
Zeros	80
Zeros (%)	0.9%
Negatives	0
Negatives (%)	0.0%



CHASES
ical
ow Details

Distinct Count	6203
Unique (%)	69.3%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%

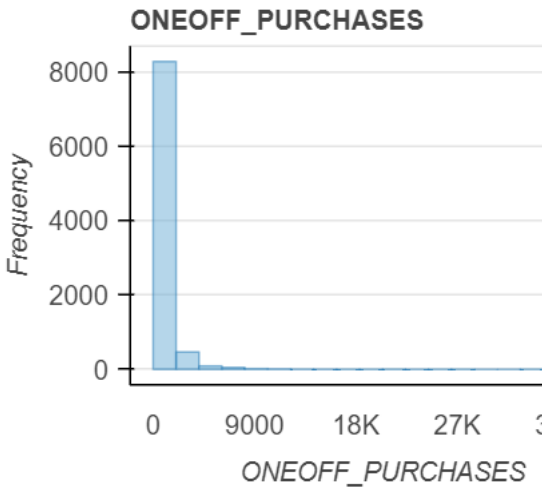
Memory Size	139.8 KB
Mean	1003.2048
Minimum	0
Maximum	49039.57
Zeros	2044
Zeros (%)	22.8%
Negatives	0
Negatives (%)	0.0%



FF_PURCHAS…
al
Details

Distinct Count	4014
Unique (%)	44.9%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%

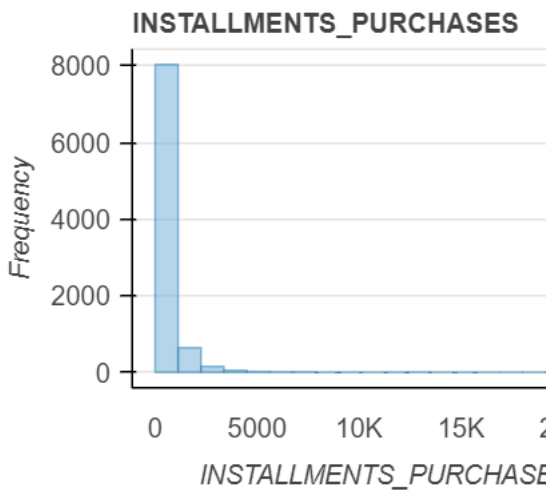
Memory Size	139.8 KB
Mean	592.4374
Minimum	0
Maximum	40761.25
Zeros	4302
Zeros (%)	48.1%
Negatives	0
Negatives (%)	0.0%



LLMENTS_PU…
al
Details

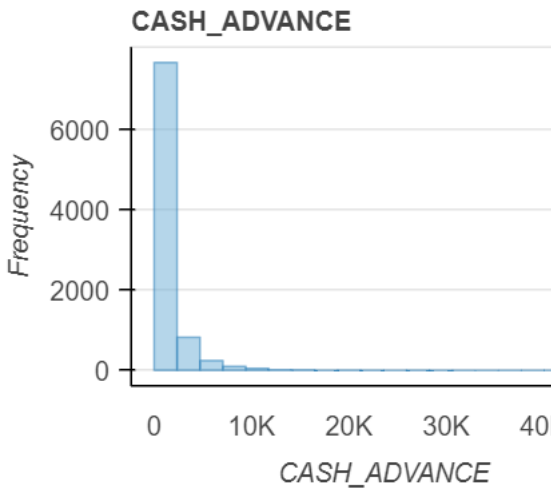
Distinct Count	4452
Unique (%)	49.7%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%

Memory Size	139.8 KB
Mean	411.0676
Minimum	0
Maximum	22500
Zeros	3916
Zeros (%)	43.8%
Negatives	0
Negatives (%)	0.0%



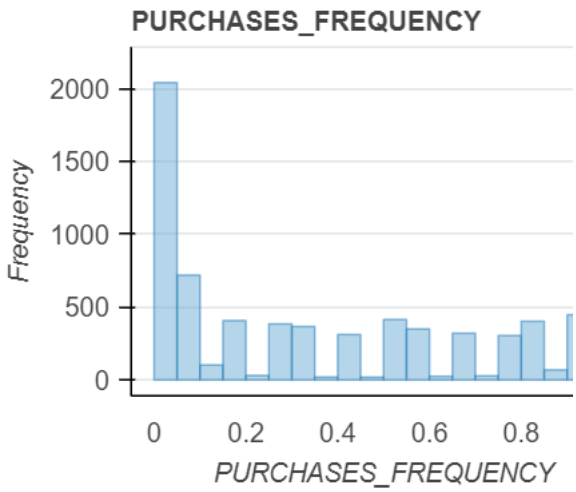
_ADVANCE
ical
Details

Distinct Count	4323	Memory Size	139.8 KB
Unique (%)	48.3%	Mean	978.8711
Missing	0	Minimum	0
Missing (%)	0.0%	Maximum	47137.2118
Infinite	0	Zeros	4628
Infinite (%)	0.0%	Zeros (%)	51.7%
		Negatives	0
		Negatives (%)	0.0%



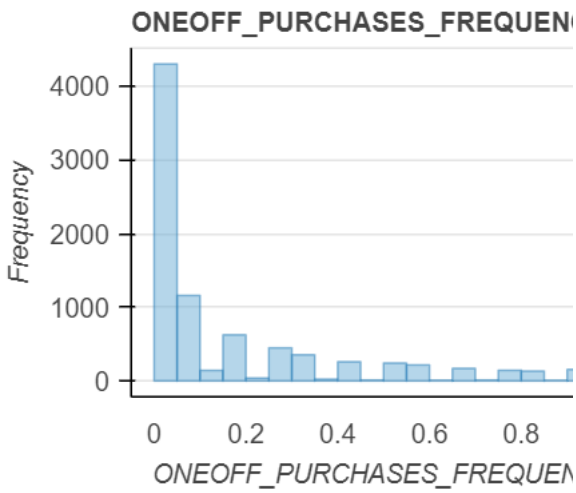
CHASES_FREQ...
ical
ow Details

Distinct Count	47	Memory Size	139.8 KB
Unique (%)	0.5%	Mean	0.4904
Missing	0	Minimum	0
Missing (%)	0.0%	Maximum	1
Infinite	0	Zeros	2043
Infinite (%)	0.0%	Zeros (%)	22.8%
		Negatives	0
		Negatives (%)	0.0%



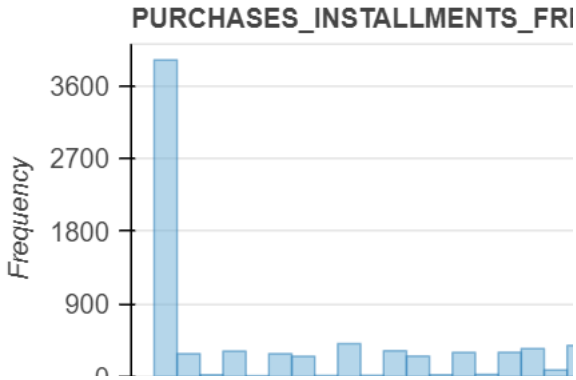
OFF_PURCHAS...
ical
ow Details

Distinct Count	47	Memory Size	139.8 KB
Unique (%)	0.5%	Mean	0.2025
Missing	0	Minimum	0
Missing (%)	0.0%	Maximum	1
Infinite	0	Zeros	4302
Infinite (%)	0.0%	Zeros (%)	48.1%
		Negatives	0
		Negatives (%)	0.0%

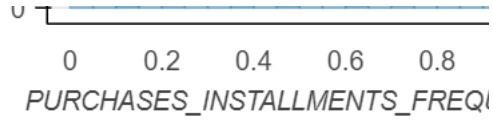


CHASES_INST...
ical
ow Details

Distinct Count	47	Memory Size	139.8 KB
Unique (%)	0.5%	Mean	0.3644
Missing	0	Minimum	0
Missing (%)	0.0%	Maximum	1
Infinite	0	Zeros	3915
		Zeros (%)	43.7%
		Negatives	0

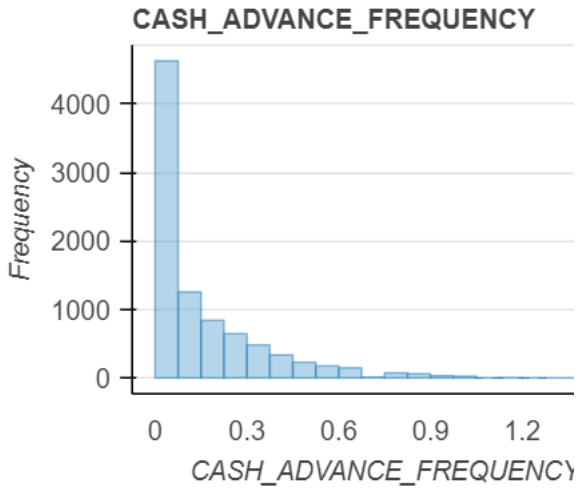


	Infinite (%)	0.0%	Negatives	0
			Negatives (%)	0.0%



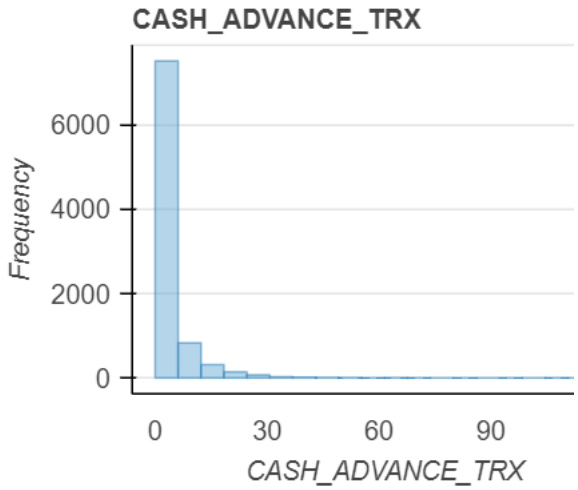
H_ADVANCE_F...
ical
ow Details

Distinct Count	54	Memory Size	139.8 KB
Unique (%)	0.6%	Mean	0.1351
Missing	0	Minimum	0
Missing (%)	0.0%	Maximum	1.5
Infinite	0	Zeros	4628
Infinite (%)	0.0%	Zeros (%)	51.7%
		Negatives	0
		Negatives (%)	0.0%



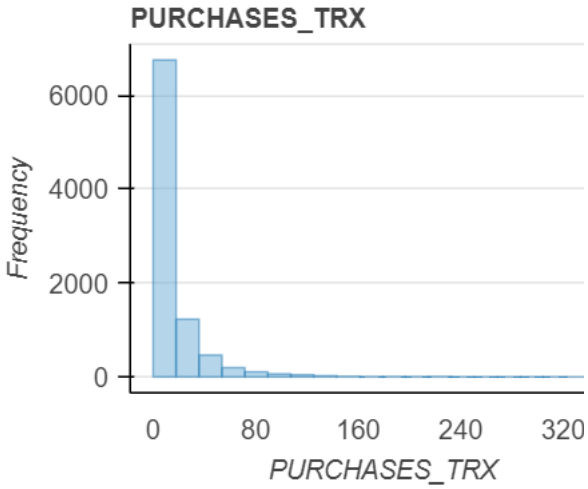
H_ADVANCE_T...
ical
ow Details

Distinct Count	65	Memory Size	139.8 KB
Unique (%)	0.7%	Mean	3.2488
Missing	0	Minimum	0
Missing (%)	0.0%	Maximum	123
Infinite	0	Zeros	4628
Infinite (%)	0.0%	Zeros (%)	51.7%
		Negatives	0
		Negatives (%)	0.0%



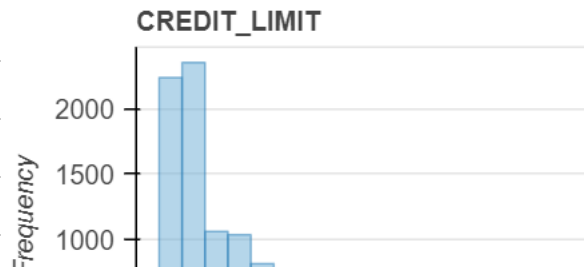
RCHASES_TRX
rical
how Details

Distinct Count	173	Memory Size	139.8 KB
Unique (%)	1.9%	Mean	14.7098
Missing	0	Minimum	0
Missing (%)	0.0%	Maximum	358
Infinite	0	Zeros	2044
Infinite (%)	0.0%	Zeros (%)	22.8%
		Negatives	0
		Negatives (%)	0.0%



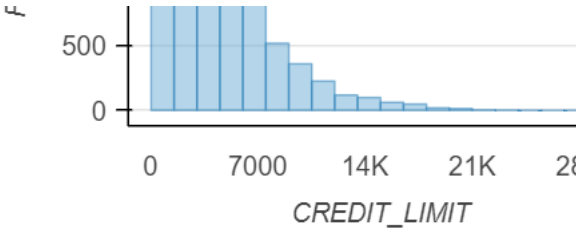
DIT_LIMIT
rical

Distinct Count	205	Memory Size	139.8 KB
Unique (%)	2.3%	Mean	4494.4495
Missing	1	Minimum	50
Missing	0.0%	Maximum	30000
		Zeros	0



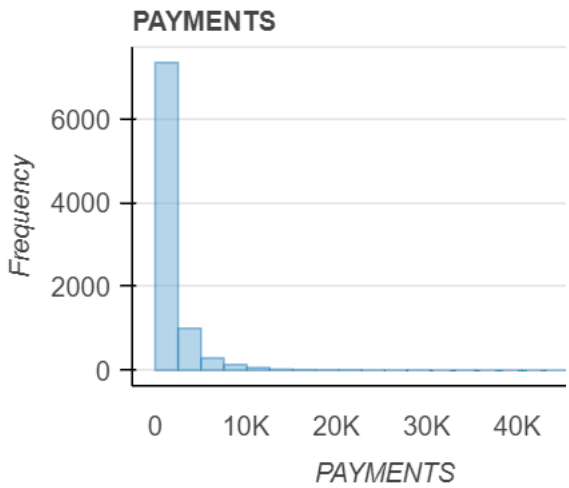
ow Details

(%)	0.0%
Infinite	0
Infinite (%)	0.0%
Zeros	0
Zeros (%)	0.0%
Negatives	0
Negatives (%)	0.0%



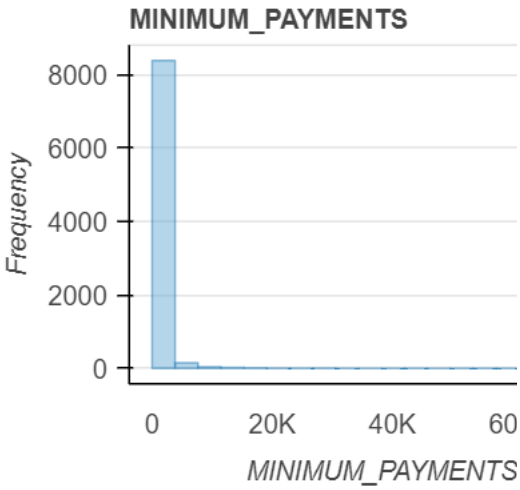
IENTS
.cal
w Details

Distinct Count	8711
Unique (%)	97.3%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Memory Size	139.8 KB
Mean	1733.1439
Minimum	0
Maximum	50721.4834
Zeros	240
Zeros (%)	2.7%
Negatives	0
Negatives (%)	0.0%



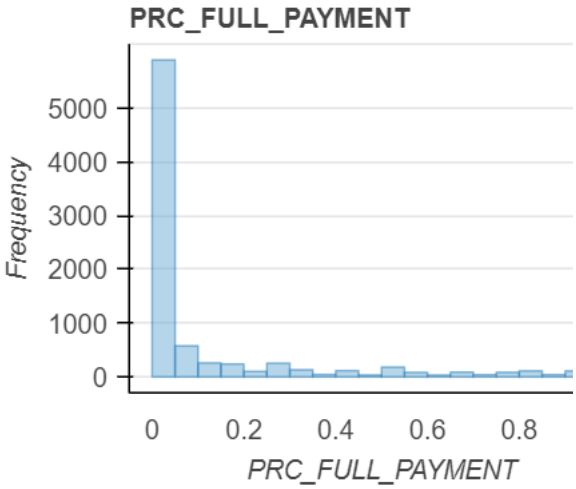
I_PAYMENTS
etails

Distinct Count	8636
Unique (%)	100.0%
Missing	313
Missing (%)	3.5%
Infinite	0
Infinite (%)	0.0%
Memory Size	135.0 KB
Mean	864.2065
Minimum	0.01916
Maximum	76406.2075
Zeros	0
Zeros (%)	0.0%
Negatives	0
Negatives (%)	0.0%

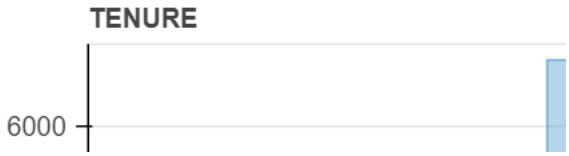


_FULL_PAYMENT
ical
ow Details

Distinct Count	47
Unique (%)	0.5%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Memory Size	139.8 KB
Mean	0.1537
Minimum	0
Maximum	1
Zeros	5903
Zeros (%)	66.0%
Negatives	0
Negatives (%)	0.0%



Distinct Count	7
Unique (%)	0.1%
Memory Size	139.8 KB
Mean	11.5173
Minimum	6

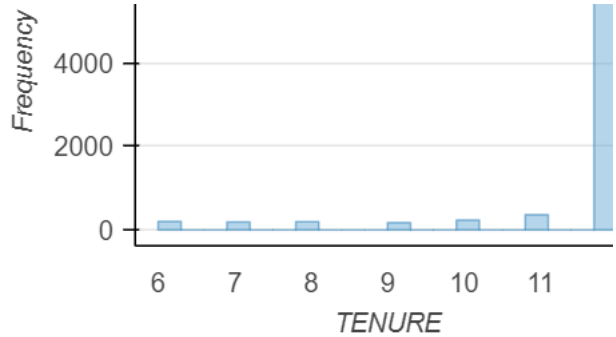


ENURE
merical

Show Details

Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%

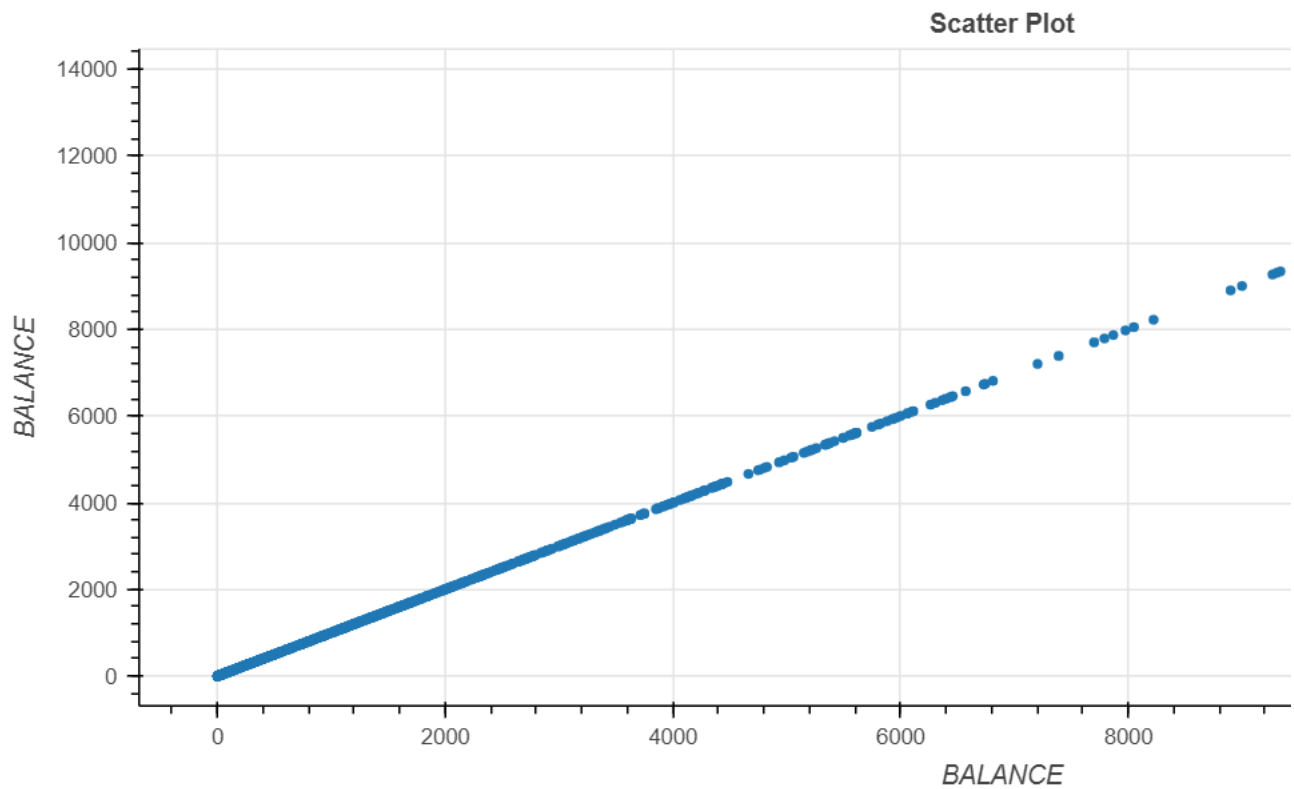
Maximum	12
Zeros	0
Zeros (%)	0.0%
Negatives	0
Negatives (%)	0.0%



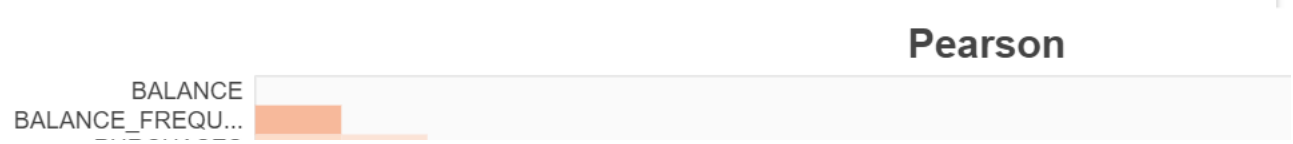
Interactions

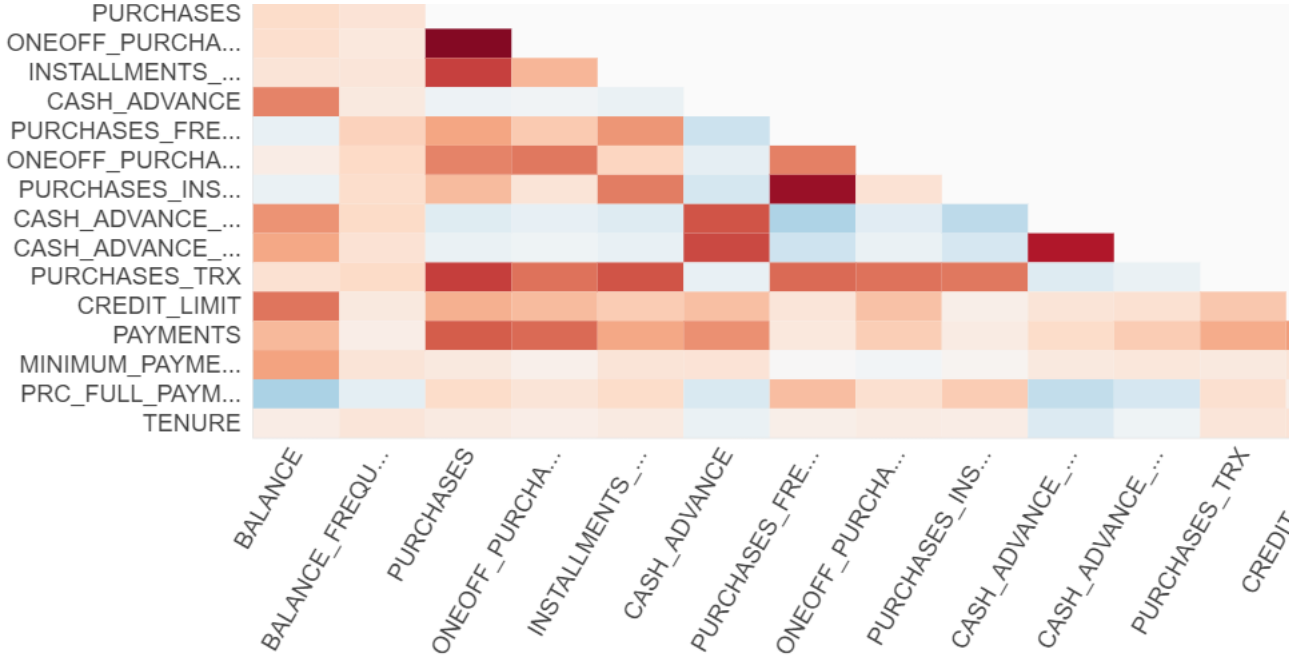
X-Axis: BALANCE

Y-Axis: BALANCE

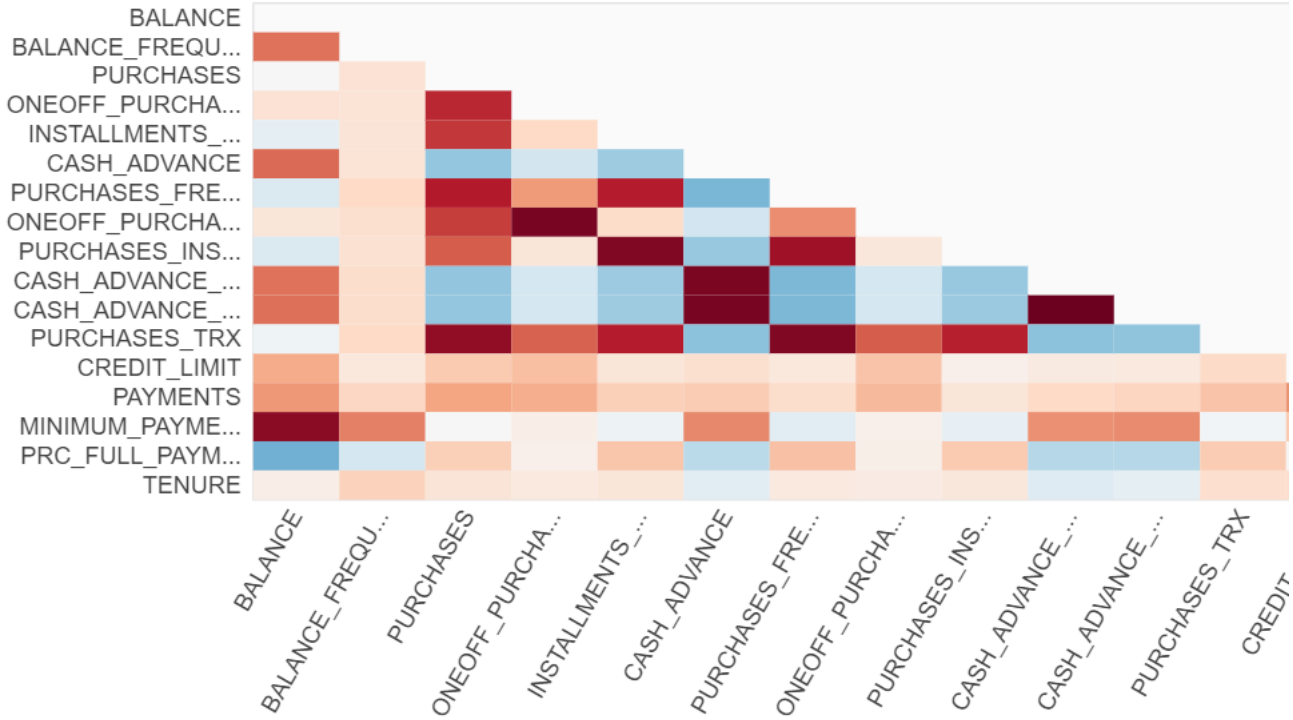


Correlations

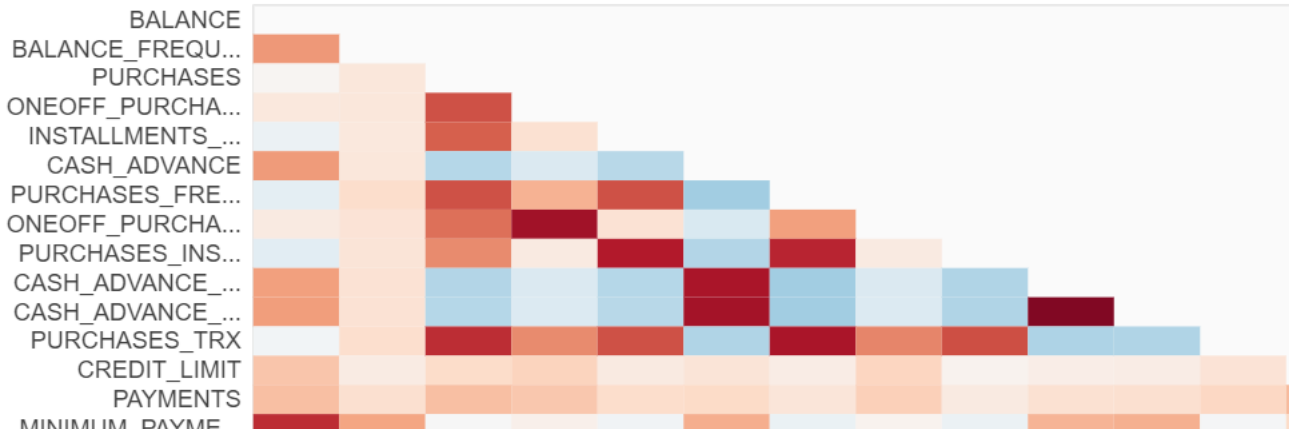




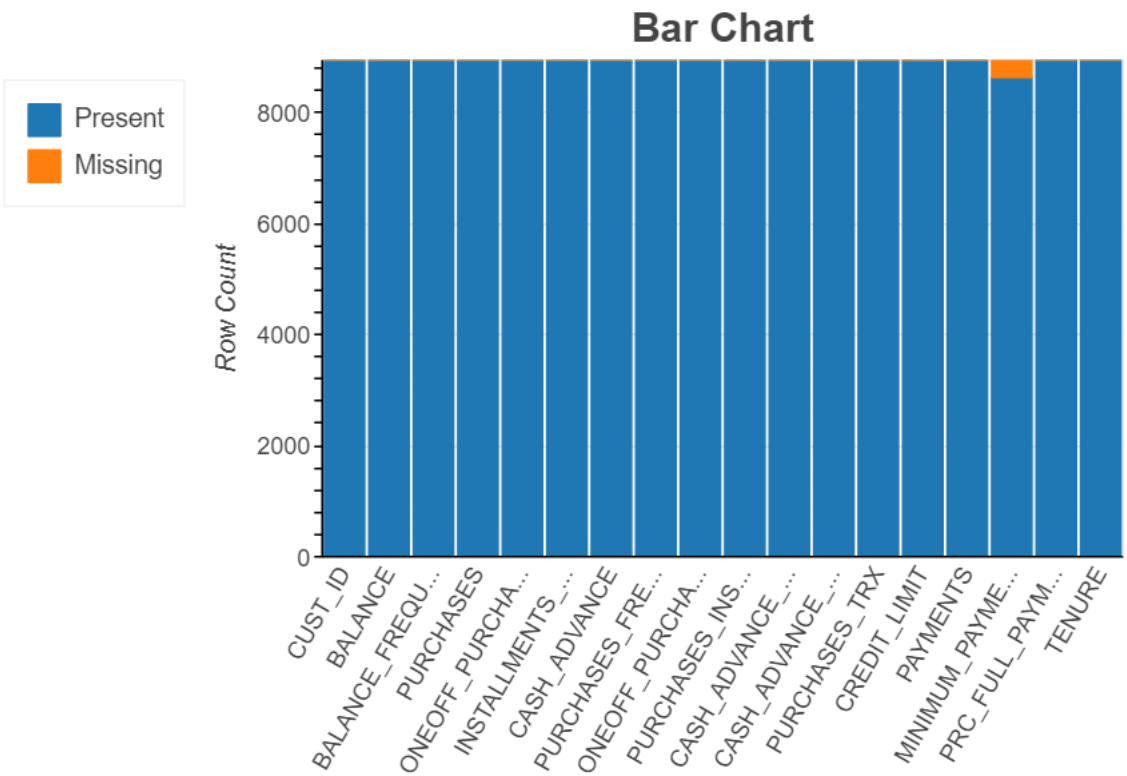
Spearman

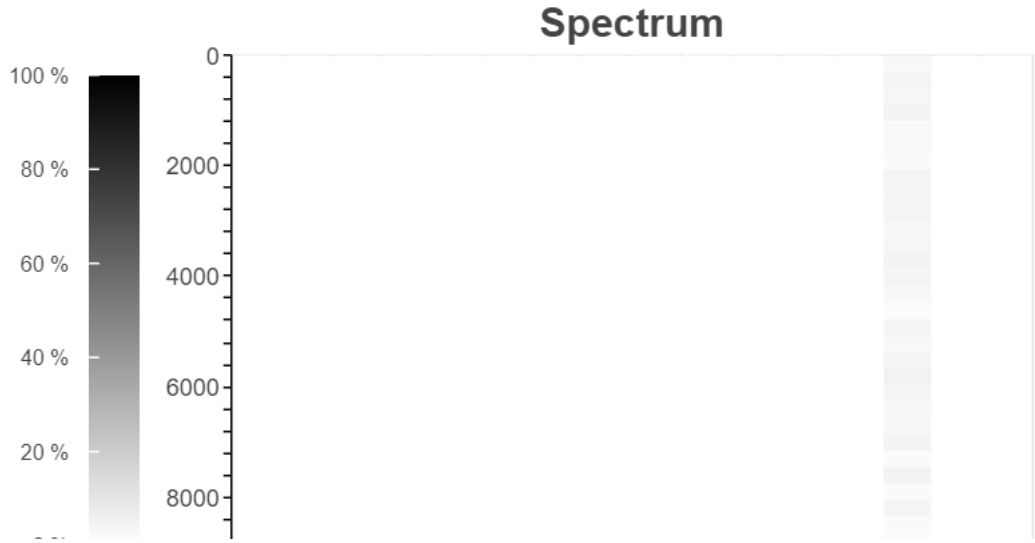


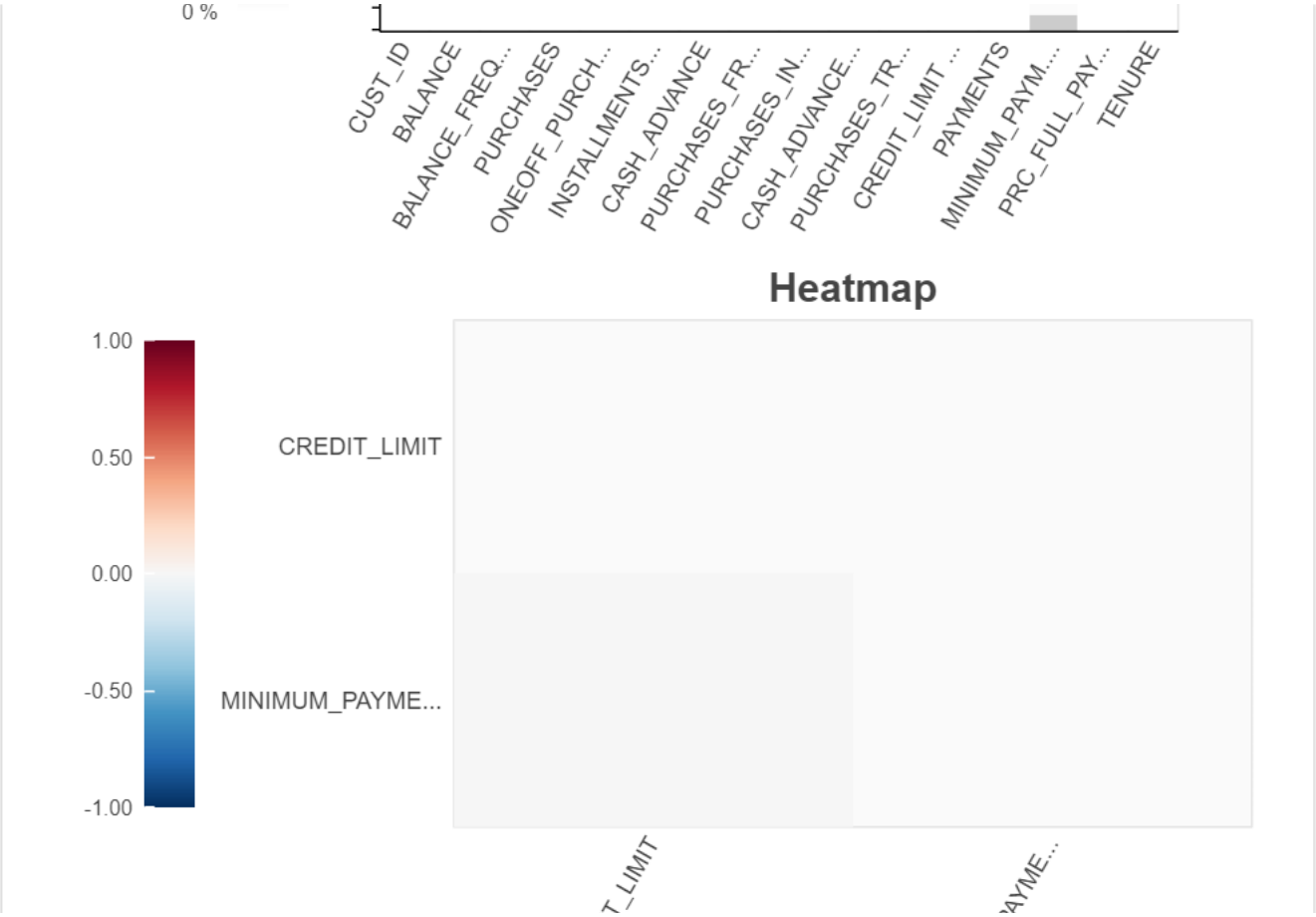
KendallTau



Missing Values








```
df.drop('CUST_ID', axis=1, inplace=True)
```

```
df.dropna(subset= ['CREDIT_LIMIT'], inplace=True)
```

```
df['MINIMUM_PAYMENTS'].fillna(df['MINIMUM_PAYMENTS'].median(), inplace=True)
```

Anomaly Detection Using Pycaret

PyCaret's Anomaly Detection Module is an unsupervised ML module that is used for identifying rare items,events or observation which raise suspicions by differing significantly from the majority of data. Typically ,the anomalous items will translate to some kind of problem such as bank fraud ,a structural defect,medical problems or errors .This module provide several pre-processing feature that prepares the data for modeling through setup function.This modules has over 12 ready-to-use algorithms and several plots to analyze the results of trained models.

```
# import anomaly detection modules
from pycaret.anomaly import *
```

```
# intialize the setup
exp_ano = setup(df)
```

	Description	Value
0	session_id	5368
1	Original Data	(8949, 17)
2	Missing Values	False
3	Numeric Features	16
4	Categorical Features	1
5	Ordinal Features	False
6	High Cardinality Features	False
7	High Cardinality Method	None
8	Transformed Data	(8949, 23)
9	CPU Jobs	-1
10	Use GPU	False
11	Log Experiment	False
12	Experiment Name	anomaly-default-name
13	USI	e8cb
14	Imputation Type	simple
15	Iterative Imputation Iteration	None
16	Numeric Imputer	mean
17	Iterative Imputation Numeric Model	None
18	Categorical Imputer	mode
19	Iterative Imputation Categorical Model	None
20	Unknown Categoricals Handling	least_frequent
21	Normalize	False
22	Normalize Method	None
23	Transformation	False
24	Transformation Method	None
25	PCA	False
26	PCA Method	None
27	PCA Components	None
28	Ignore Low Variance	False
29	Combine Rare Levels	False