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

df_original = pd.read_csv('CC_GENERAL.csv', index_col='CUST_ID')
df = df_original.copy()

df.shape

(8950, 17)

df.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')

df.sample(10)

              BALANCE  BALANCE_FREQUENCY  PURCHASES
ONEOFF_PURCHASES  \
CUST_ID

C10476   1715.459275           1.000000      96.75
34.65

C10562     68.039722           0.454545     734.11
734.11

C10115    156.216753           0.818182    1354.86
585.63

C18675   2373.686499           1.000000    1310.47
0.00

C14085   1066.884128           1.000000       0.00
0.00

C14892   1594.485356           1.000000    2650.91
2123.90

C14291   2224.065672           1.000000      81.84
81.84

C10456     37.753941           0.363636     243.36
243.36

C16125    252.717781           1.000000     463.34
0.00

C17261    274.604824           1.000000     341.37
0.00
```

CUST_ID	INSTALLMENTS_PURCHASES	CASH_ADVANCE	PURCHASES_FREQUENCY	\
C10476	62.10	19.141227	0.333333	
C10562	0.00	0.000000	0.333333	
C10115	769.23	0.000000	0.916667	
C18675	1310.47	3667.381175	1.000000	
C14085	0.00	215.632774	0.000000	
C14892	527.01	0.000000	1.000000	
C14291	0.00	1570.966236	0.083333	
C10456	0.00	0.000000	0.083333	
C16125	463.34	0.000000	0.833333	
C17261	341.37	0.000000	0.583333	

CUST_ID	ONEOFF_PURCHASES_FREQUENCY	PURCHASES_INSTALLMENTS_FREQUENCY	\
C10476	0.083333	0.250000	
C10562	0.333333	0.000000	
C10115	0.333333	0.833333	
C18675	0.000000	1.000000	
C14085	0.000000	0.000000	
C14892	0.250000	1.000000	
C14291	0.083333	0.000000	
C10456	0.083333	0.000000	
C16125	0.000000	0.833333	
C17261	0.000000	0.500000	

CUST_ID	CASH_ADVANCE_FREQUENCY	CASH_ADVANCE_TRX	PURCHASES_TRX	\
C10476	0.083333	1	7	
C10562	0.000000	0	6	
C10115	0.000000	0	34	
C18675	0.250000	5	31	
C14085	0.250000	6	0	
C14892	0.000000	0	33	
C14291	0.666667	12	5	
C10456	0.000000	0	2	
C16125	0.000000	0	10	
C17261	0.000000	0	7	

```
          CREDIT_LIMIT     PAYMENTS  MINIMUM_PAYMENTS  PRC_FULL_PAYMENT
TENURE
CUST_ID

C10476         1800.0   452.225512        411.612521              0.00
12
C10562         5000.0  1537.307484        102.554198              0.50
12
C10115         5500.0   858.976682        108.319927              0.00
12
C18675         3000.0  4613.536289        936.940012              0.00
12
C14085         1200.0   222.800881        263.186377              0.00
12
C14892         7000.0  1416.618009        427.147060              0.00
12
C14291         2500.0   487.001643        968.791469              0.00
12
C10456         1800.0   546.417272        134.985783              0.25
12
C16125         4000.0   162.949236        184.464721              0.00
12
C17261         2500.0   101.053841        171.581036              0.00
12

df.info()

<class 'pandas.core.frame.DataFrame'>
Index: 8950 entries, C10001 to C19190
Data columns (total 17 columns):
 #   Column                            Non-Null Count  Dtype
---  ------                            --------------  -----
 0   BALANCE                           8950 non-null   float64
 1   BALANCE_FREQUENCY                 8950 non-null   float64
 2   PURCHASES                         8950 non-null   float64
 3   ONEOFF_PURCHASES                  8950 non-null   float64
 4   INSTALLMENTS_PURCHASES            8950 non-null   float64
 5   CASH_ADVANCE                      8950 non-null   float64
 6   PURCHASES_FREQUENCY               8950 non-null   float64
 7   ONEOFF_PURCHASES_FREQUENCY        8950 non-null   float64
 8   PURCHASES_INSTALLMENTS_FREQUENCY  8950 non-null   float64
 9   CASH_ADVANCE_FREQUENCY            8950 non-null   float64
 10  CASH_ADVANCE_TRX                  8950 non-null   int64
 11  PURCHASES_TRX                     8950 non-null   int64
 12  CREDIT_LIMIT                      8949 non-null   float64
 13  PAYMENTS                          8950 non-null   float64
 14  MINIMUM_PAYMENTS                  8637 non-null   float64
 15  PRC_FULL_PAYMENT                  8950 non-null   float64
 16  TENURE                            8950 non-null   int64
```

```
dtypes: float64(14), int64(3)
memory usage: 1.2+ MB
```

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

```
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
```

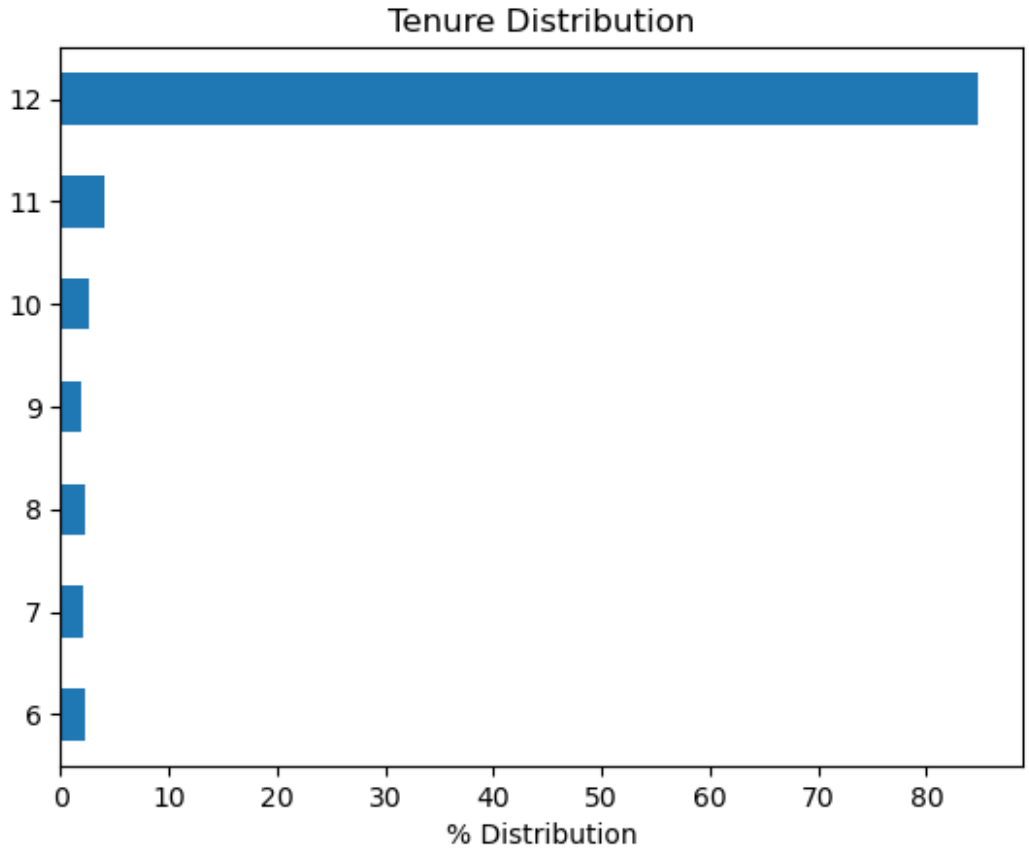
```
df.fillna(df.median(), inplace=True)
```

```
for col in df.columns:
    print('{:33} : {:6} : {:}'.format(col, df[col].nunique(),
df[col].dtype))
```

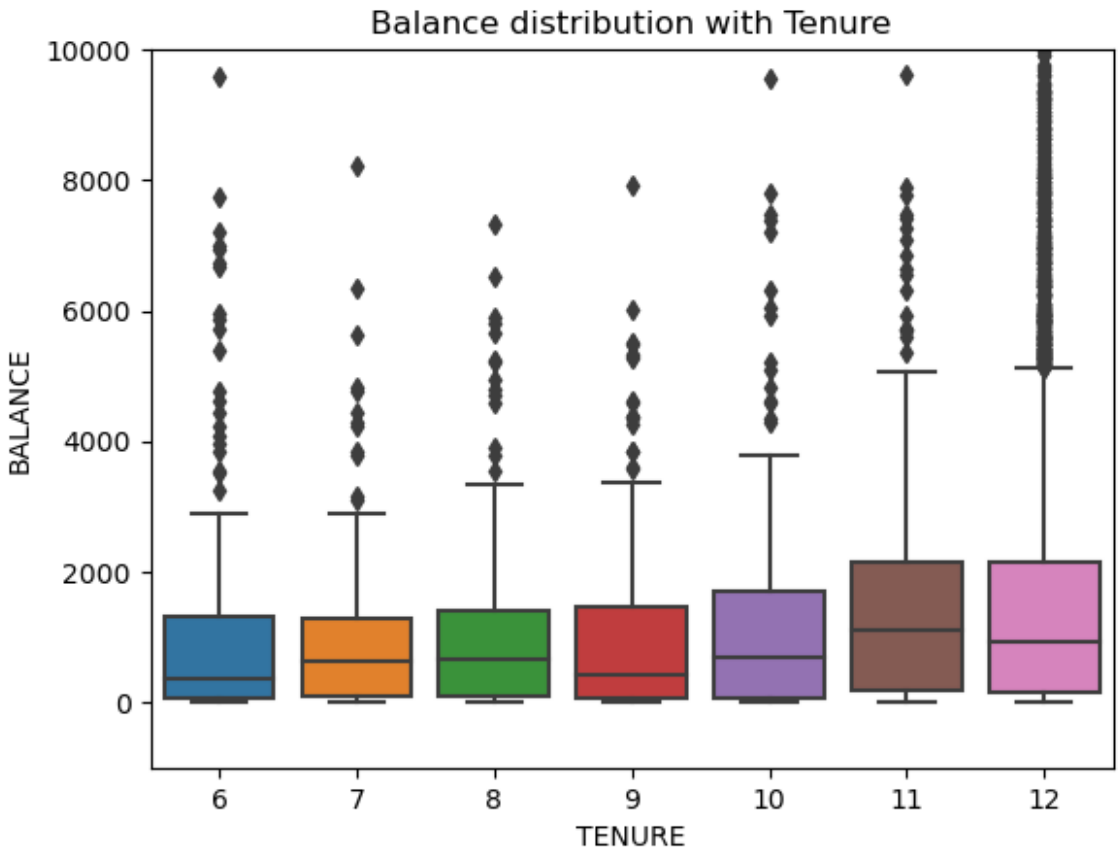
```
BALANCE                           :   8871 : float64
BALANCE_FREQUENCY                 :     43 : float64
PURCHASES                         :   6203 : float64
ONEOFF_PURCHASES                  :   4014 : float64
INSTALLMENTS_PURCHASES            :   4452 : float64
CASH_ADVANCE                      :   4323 : float64
PURCHASES_FREQUENCY               :     47 : float64
ONEOFF_PURCHASES_FREQUENCY        :     47 : float64
PURCHASES_INSTALLMENTS_FREQUENCY  :     47 : float64
CASH_ADVANCE_FREQUENCY            :     54 : float64
CASH_ADVANCE_TRX                  :     65 : int64
PURCHASES_TRX                     :    173 : int64
CREDIT_LIMIT                      :    205 : float64
PAYMENTS                          :   8711 : float64
MINIMUM_PAYMENTS                  :   8636 : float64
PRC_FULL_PAYMENT                  :     47 : float64
TENURE                            :      7 : int64
```

```
(1e2*df['TENURE'].value_counts().sort_index()/
len(df)).plot(kind='barh')
```

```
plt.title('Tenure Distribution')
plt.xlabel('% Distribution');
```




```
sns.boxplot(x="TENURE", y="BALANCE", data=df)
plt.ylim(-10**3, 10**4)
plt.title('Balance distribution with Tenure');
```




```
fig, axs = plt.subplots(nrows=4, ncols=4, figsize=(15, 15))
for i in range(4):
    for j in range(4):
        sns.distplot(df[df.columns[4 * i + j]], ax=axs[i,j])
plt.show()
```

```
/var/folders/nz/v4c3dsmn1293s89vq32wmzq00000gn/T/
ipykernel_27289/1523716901.py:4: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.

Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df[df.columns[4 * i + j]], ax=axs[i,j])
/var/folders/nz/v4c3dsmn1293s89vq32wmzq00000gn/T/ipykernel_27289/15237
16901.py:4: UserWarning:
```

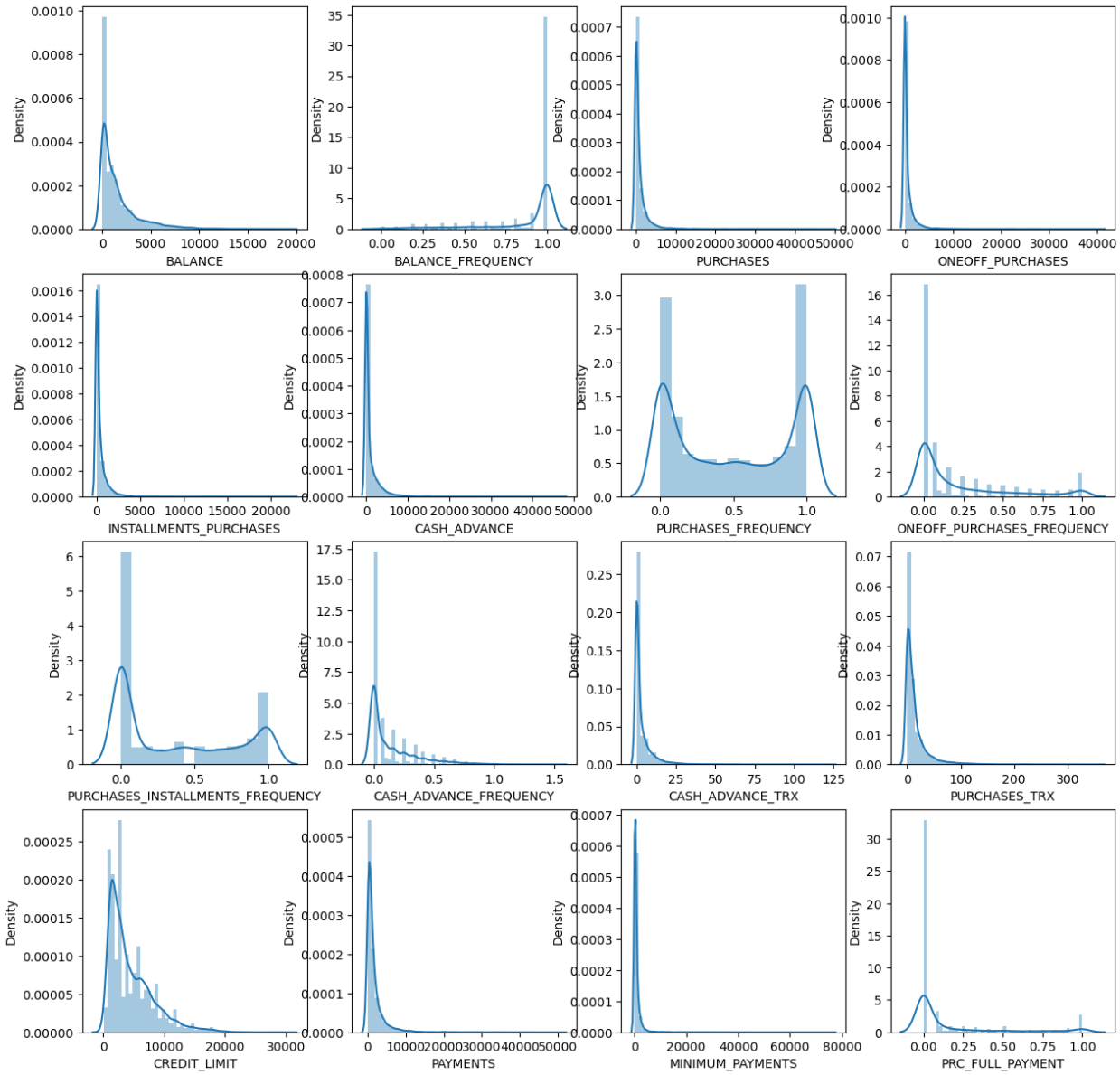
```
`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.

Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df[df.columns[4 * i + j]], ax=axs[i,j])
/var/folders/nz/v4c3dsmn1293s89vq32wmzq00000gn/T/ipykernel_27289/15237
16901.py:4: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.

Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df[df.columns[4 * i + j]], ax=axs[i,j])
/var/folders/nz/v4c3dsmn1293s89vq32wmzq00000gn/T/ipykernel_27289/15237
16901.py:4: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.

Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df[df.columns[4 * i + j]], ax=axs[i,j])
/var/folders/nz/v4c3dsmn1293s89vq32wmzq00000gn/T/ipykernel_27289/15237
16901.py:4: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.

Please adapt your code to use either `displot` (a figure-level
function with
```

```
similar flexibility) or `histplot` (an axes-level function for
histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df[df.columns[4 * i + j]], ax=axs[i,j])
/var/folders/nz/v4c3dsmn1293s89vq32wmzq00000gn/T/ipykernel_27289/15237
16901.py:4: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.

Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df[df.columns[4 * i + j]], ax=axs[i,j])
/var/folders/nz/v4c3dsmn1293s89vq32wmzq00000gn/T/ipykernel_27289/15237
16901.py:4: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.

Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df[df.columns[4 * i + j]], ax=axs[i,j])
/var/folders/nz/v4c3dsmn1293s89vq32wmzq00000gn/T/ipykernel_27289/15237
16901.py:4: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.

Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751
```

```
  sns.distplot(df[df.columns[4 * i + j]], ax=axs[i,j])
/var/folders/nz/v4c3dsmn1293s89vq32wmzq00000gn/T/ipykernel_27289/15237
16901.py:4: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.

Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df[df.columns[4 * i + j]], ax=axs[i,j])
/var/folders/nz/v4c3dsmn1293s89vq32wmzq00000gn/T/ipykernel_27289/15237
16901.py:4: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.

Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df[df.columns[4 * i + j]], ax=axs[i,j])
/var/folders/nz/v4c3dsmn1293s89vq32wmzq00000gn/T/ipykernel_27289/15237
16901.py:4: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.

Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df[df.columns[4 * i + j]], ax=axs[i,j])
/var/folders/nz/v4c3dsmn1293s89vq32wmzq00000gn/T/ipykernel_27289/15237
16901.py:4: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn
```

```
v0.14.0.

Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df[df.columns[4 * i + j]], ax=axs[i,j])
/var/folders/nz/v4c3dsmn1293s89vq32wmzq00000gn/T/ipykernel_27289/15237
16901.py:4: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.

Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df[df.columns[4 * i + j]], ax=axs[i,j])
/var/folders/nz/v4c3dsmn1293s89vq32wmzq00000gn/T/ipykernel_27289/15237
16901.py:4: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.

Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df[df.columns[4 * i + j]], ax=axs[i,j])
/var/folders/nz/v4c3dsmn1293s89vq32wmzq00000gn/T/ipykernel_27289/15237
16901.py:4: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.

Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
```

```
histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df[df.columns[4 * i + j]], ax=axs[i,j])
/var/folders/nz/v4c3dsmn1293s89vq32wmzq00000gn/T/ipykernel_27289/15237
16901.py:4: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.

Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df[df.columns[4 * i + j]], ax=axs[i,j])
```




```
df.shape
```

```
(8950, 17)
```

```
from sklearn.cluster import KMeans
k = 5
kmeans = KMeans(n_clusters=k, random_state=1)
df['k_5_label'] = kmeans.fit_predict(df)
```

```
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
```

```
kmeans.inertia_
```

```
181759406976.1244
```

```
profile = df.groupby('k_5_label').mean().T
```

```
round(profile)
```

k_5_label	0	1	2	3	4
BALANCE	2047.0	5685.0	5519.0	800.0	4346.0
BALANCE_FREQUENCY	1.0	1.0	1.0	1.0	1.0
PURCHASES	1341.0	2722.0	26257.0	524.0	1112.0
ONEOFF_PURCHASES	834.0	1700.0	21002.0	258.0	160.0
INSTALLMENTS_PURCHASES	508.0	1022.0	5255.0	266.0	952.0
CASH_ADVANCE	1078.0	4592.0	1492.0	487.0	1029.0
PURCHASES_FREQUENCY	1.0	1.0	1.0	0.0	0.0
ONEOFF_PURCHASES_FREQUENCY	0.0	0.0	1.0	0.0	0.0
PURCHASES_INSTALLMENTS_FREQUENCY	0.0	0.0	1.0	0.0	0.0
CASH_ADVANCE_FREQUENCY	0.0	0.0	0.0	0.0	0.0
CASH_ADVANCE_TRX	3.0	10.0	2.0	2.0	3.0
PURCHASES_TRX	19.0	33.0	123.0	10.0	22.0
CREDIT_LIMIT	7376.0	11381.0	16160.0	2271.0	4591.0
PAYMENTS	1934.0	6651.0	28434.0	920.0	1563.0
MINIMUM_PAYMENTS	771.0	2116.0	3222.0	528.0	26690.0
PRC_FULL_PAYMENT	0.0	0.0	0.0	0.0	0.0
TENURE	12.0	12.0	12.0	11.0	12.0

```
round(pd.DataFrame(kmeans.cluster_centers_.T))
```

	0	1	2	3	4
0	2042.0	5686.0	5519.0	800.0	4346.0
1	1.0	1.0	1.0	1.0	1.0
2	1339.0	2725.0	26257.0	524.0	1112.0

```
3    833.0   1699.0  21002.0   258.0    160.0
4    506.0   1026.0   5255.0   266.0    952.0
5   1076.0   4583.0   1492.0   487.0   1029.0
6      1.0      1.0      1.0     0.0      0.0
7      0.0      0.0      1.0     0.0      0.0
8      0.0      0.0      1.0     0.0      0.0
9      0.0      0.0      0.0     0.0      0.0
10     3.0     10.0      2.0     2.0      3.0
11    19.0     33.0    123.0    10.0     22.0
12  7372.0  11377.0  16160.0  2271.0   4591.0
13  1935.0   6628.0  28434.0   920.0   1563.0
14   770.0   2114.0   3222.0   528.0  26690.0
15     0.0      0.0      0.0     0.0      0.0
16    12.0     12.0     12.0    11.0     12.0
```

```
from sklearn.cluster import MiniBatchKMeans

minibatch_kmeans = MiniBatchKMeans(n_clusters=5, random_state=1)
df['k_5_batch'] = minibatch_kmeans.fit_predict(df)
```

```
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 3 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
```

```
pd.crosstab(df['k_5_label'], df['k_5_batch'])
```

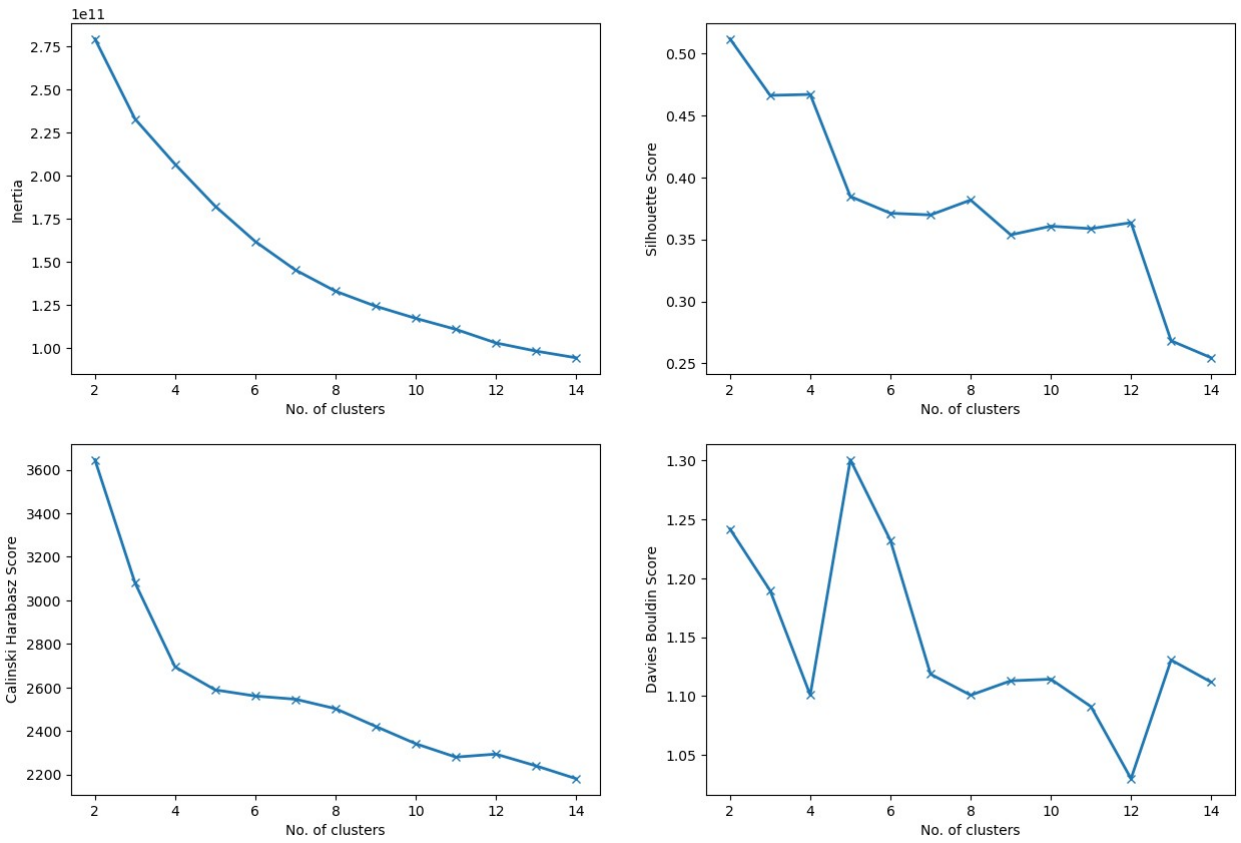
```
k_5_batch    0     1     2     3    4
k_5_label
0          471   902  1178     0   31
1          354     0    85     0  250
2            0     0     0     0   25
3            7  1982     0  3623    2
4           37     0     0     2    1
```

```
from sklearn.metrics import silhouette_score, calinski_harabasz_score,
davies_bouldin_score

def evaluate_metrics(df, min_clust=2, max_clust=10, rand_state=1):
    inertias = []
    silhouette = []
    ch_score = []
    db_score = []
    for n_clust in range(min_clust, max_clust):
        kmeans = KMeans(n_clusters=n_clust, random_state=rand_state)
        y_label = kmeans.fit_predict(df)
        inertias.append(kmeans.inertia_)
        silhouette.append(silhouette_score(df, y_label))
        ch_score.append(calinski_harabasz_score(df, y_label))
        db_score.append(davies_bouldin_score(df, y_label))
```

```
    fig, ax = plt.subplots(2, 2, figsize=(15, 10))
    ax[0][0].plot(range(min_clust, max_clust), inertias, '-x',
linewidth=2)
    ax[0][0].set_xlabel('No. of clusters')
    ax[0][0].set_ylabel('Inertia')

    ax[0][1].plot(range(min_clust, max_clust), silhouette, '-x',
linewidth=2)
    ax[0][1].set_xlabel('No. of clusters')
    ax[0][1].set_ylabel('Silhouette Score')

    ax[1][0].plot(range(min_clust, max_clust), ch_score, '-x',
linewidth=2)
    ax[1][0].set_xlabel('No. of clusters')
    ax[1][0].set_ylabel('Calinski Harabasz Score')

    ax[1][1].plot(range(min_clust, max_clust), db_score, '-x',
linewidth=2)
    ax[1][1].set_xlabel('No. of clusters')
    ax[1][1].set_ylabel('Davies Bouldin Score')
    fig.suptitle('Metrics to evaluate the number of clusters')
    plt.show()

evaluate_metrics(df.iloc[:, :-2], min_clust=2, max_clust=15,
rand_state=0)
```

```
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
```

```
to suppress the warning
  warnings.warn(
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
```




```
df = df_original.copy()
df.fillna(df.median(), inplace=True)

from sklearn.preprocessing import StandardScaler
df_scaled = StandardScaler().fit_transform(df)

evaluate_metrics(df_scaled, min_clust=2, max_clust=15, rand_state=0)
```

```
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
```

```
  warnings.warn(
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
/Users/sravya/anaconda3/lib/python3.10/site-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
```

```
to suppress the warning
  warnings.warn(
```