loan_data

May 30, 2024

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

[2]:
```
df = pd.read_csv('loan_data.csv')
df
```

[2]:
```
       SK_ID_CURR  TARGET NAME_CONTRACT_TYPE CODE_GENDER FLAG_OWN_CAR  \
0          100002       1         Cash loans           M            N
1          100003       0         Cash loans           F            N
2          100004       0    Revolving loans           M            Y
3          100006       0         Cash loans           F            N
4          100007       0         Cash loans           M            N
…             …       …                …           …            …
96997      212591       0         Cash loans           F            Y
96998      212593       0         Cash loans           M            Y
96999      212594       0         Cash loans           F            Y
97000      212595       0         Cash loans           F            Y
97001      212596       0         Cash loans           M            Y

      FLAG_OWN_REALTY  CNT_CHILDREN  AMT_INCOME_TOTAL  AMT_CREDIT  \
0                   Y             0          202500.0    406597.5
1                   N             0          270000.0   1293502.5
2                   Y             0           67500.0    135000.0
3                   Y             0          135000.0    312682.5
4                   Y             0          121500.0    513000.0
…                 …           …               …         …
96997               N             0          540000.0   1800000.0
96998               Y             2          450000.0   1187370.0
96999               Y             1          180000.0    314100.0
97000               Y             1          180000.0   1006920.0
97001               Y             2          540000.0   1339884.0

       AMT_ANNUITY  …  FLAG_DOCUMENT_18 FLAG_DOCUMENT_19 FLAG_DOCUMENT_20  \
0          24700.5  …               0.0              0.0              0.0
1          35698.5  …               0.0              0.0              0.0
2           6750.0  …               0.0              0.0              0.0
```

```
3          29686.5  …              0.0              0.0              0.0
4          21865.5  …              0.0              0.0              0.0
…                …  …              …                …                …
96997      49500.0  …              0.0              0.0              0.0
96998     115803.0  …              0.0              0.0              0.0
96999      21375.0  …              0.0              0.0              0.0
97000      42790.5  …              0.0              0.0              0.0
97001      39307.5  …              NaN              NaN              NaN

      FLAG_DOCUMENT_21 AMT_REQ_CREDIT_BUREAU_HOUR AMT_REQ_CREDIT_BUREAU_DAY  \
0                  0.0                        0.0                        0.0
1                  0.0                        0.0                        0.0
2                  0.0                        0.0                        0.0
3                  0.0                        NaN                        NaN
4                  0.0                        0.0                        0.0
…                    …                          …                          …
96997              0.0                        0.0                        0.0
96998              0.0                        0.0                        0.0
96999              0.0                        0.0                        0.0
97000              0.0                        0.0                        0.0
97001              NaN                        NaN                        NaN

       AMT_REQ_CREDIT_BUREAU_WEEK  AMT_REQ_CREDIT_BUREAU_MON  \
0                             0.0                        0.0
1                             0.0                        0.0
2                             0.0                        0.0
3                             NaN                        NaN
4                             0.0                        0.0
…                               …                          …
96997                         0.0                        0.0
96998                         0.0                        0.0
96999                         0.0                        0.0
97000                         0.0                        0.0
97001                         NaN                        NaN

       AMT_REQ_CREDIT_BUREAU_QRT  AMT_REQ_CREDIT_BUREAU_YEAR
0                            0.0                         1.0
1                            0.0                         0.0
2                            0.0                         0.0
3                            NaN                         NaN
4                            0.0                         0.0
…                              …                           …
96997                        0.0                         0.0
96998                        0.0                         6.0
96999                        0.0                         5.0
97000                        0.0                         0.0
97001                        NaN                         NaN
```

```
[97002 rows x 122 columns]
```

[3]:
```
df.describe()
```

[3]:
```
            SK_ID_CURR        TARGET  CNT_CHILDREN  AMT_INCOME_TOTAL  \
count   97002.000000  97002.000000  97002.000000      9.700200e+04
mean   156264.395342      0.080988      0.417094      1.694443e+05
std     32471.718636      0.272818      0.720776      3.889687e+05
min    100002.000000      0.000000      0.000000      2.565000e+04
25%    128207.250000      0.000000      0.000000      1.125000e+05
50%    156166.500000      0.000000      0.000000      1.440000e+05
75%    184364.750000      0.000000      1.000000      2.025000e+05
max    212596.000000      1.000000     12.000000      1.170000e+08

          AMT_CREDIT    AMT_ANNUITY  AMT_GOODS_PRICE  \
count   9.700200e+04   96995.000000     9.692200e+04
mean    5.988090e+05   27077.854013     5.381456e+05
std     4.017348e+05   14452.646537     3.690669e+05
min     4.500000e+04    1980.000000     4.500000e+04
25%     2.700000e+05   16474.500000     2.385000e+05
50%     5.129955e+05   24903.000000     4.500000e+05
75%     8.086500e+05   34587.000000     6.795000e+05
max     4.050000e+06  258025.500000     4.050000e+06

       REGION_POPULATION_RELATIVE    DAYS_BIRTH  DAYS_EMPLOYED  …  \
count                97001.000000  97001.000000   97001.000000  …
mean                     0.020845 -16025.042195   63229.221822  …
std                      0.013826   4369.038078  140788.834842  …
min                      0.000533 -25201.000000  -17531.000000  …
25%                      0.010006 -19668.000000   -2762.000000  …
50%                      0.018850 -15743.000000   -1220.000000  …
75%                      0.028663 -12384.000000    -292.000000  …
max                      0.072508  -7676.000000  365243.000000  …

       FLAG_DOCUMENT_18  FLAG_DOCUMENT_19  FLAG_DOCUMENT_20  FLAG_DOCUMENT_21  \
count      97001.000000      97001.000000      97001.000000      97001.000000
mean           0.008340          0.000629          0.000495          0.000320
std            0.090943          0.025069          0.022240          0.017874
min            0.000000          0.000000          0.000000          0.000000
25%            0.000000          0.000000          0.000000          0.000000
50%            0.000000          0.000000          0.000000          0.000000
75%            0.000000          0.000000          0.000000          0.000000
max            1.000000          1.000000          1.000000          1.000000

       AMT_REQ_CREDIT_BUREAU_HOUR  AMT_REQ_CREDIT_BUREAU_DAY  \
count                83995.000000               83995.000000
```

```
mean                   0.006584                    0.007417
std                    0.085732                    0.108257
min                    0.000000                    0.000000
25%                    0.000000                    0.000000
50%                    0.000000                    0.000000
75%                    0.000000                    0.000000
max                    3.000000                    6.000000

       AMT_REQ_CREDIT_BUREAU_WEEK  AMT_REQ_CREDIT_BUREAU_MON  \
count                83995.000000               83995.000000
mean                     0.034097                   0.269647
std                      0.205073                   0.925234
min                      0.000000                   0.000000
25%                      0.000000                   0.000000
50%                      0.000000                   0.000000
75%                      0.000000                   0.000000
max                      8.000000                  24.000000

       AMT_REQ_CREDIT_BUREAU_QRT  AMT_REQ_CREDIT_BUREAU_YEAR
count               83995.000000                83995.000000
mean                    0.266242                    1.893744
std                     0.614035                    1.877688
min                     0.000000                    0.000000
25%                     0.000000                    0.000000
50%                     0.000000                    1.000000
75%                     0.000000                    3.000000
max                     8.000000                   25.000000

[8 rows x 106 columns]
```

[4]:
```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 97002 entries, 0 to 97001
Columns: 122 entries, SK_ID_CURR to AMT_REQ_CREDIT_BUREAU_YEAR
dtypes: float64(103), int64(3), object(16)
memory usage: 90.3+ MB
```

[5]:
```python
type(df)
```

[5]:
```
pandas.core.frame.DataFrame
```

[6]:
```python
df.columns
```

[6]:
```
Index(['SK_ID_CURR', 'TARGET', 'NAME_CONTRACT_TYPE', 'CODE_GENDER',
       'FLAG_OWN_CAR', 'FLAG_OWN_REALTY', 'CNT_CHILDREN', 'AMT_INCOME_TOTAL',
       'AMT_CREDIT', 'AMT_ANNUITY',
```

```
       …
       'FLAG_DOCUMENT_18', 'FLAG_DOCUMENT_19', 'FLAG_DOCUMENT_20',
       'FLAG_DOCUMENT_21', 'AMT_REQ_CREDIT_BUREAU_HOUR',
       'AMT_REQ_CREDIT_BUREAU_DAY', 'AMT_REQ_CREDIT_BUREAU_WEEK',
       'AMT_REQ_CREDIT_BUREAU_MON', 'AMT_REQ_CREDIT_BUREAU_QRT',
       'AMT_REQ_CREDIT_BUREAU_YEAR'],
      dtype='object', length=122)
```

[7]:
```
df.head()
```

[7]:
```
   SK_ID_CURR  TARGET NAME_CONTRACT_TYPE CODE_GENDER FLAG_OWN_CAR  \
0      100002       1         Cash loans           M            N
1      100003       0         Cash loans           F            N
2      100004       0    Revolving loans           M            Y
3      100006       0         Cash loans           F            N
4      100007       0         Cash loans           M            N

  FLAG_OWN_REALTY  CNT_CHILDREN  AMT_INCOME_TOTAL  AMT_CREDIT  AMT_ANNUITY  \
0               Y             0          202500.0    406597.5      24700.5
1               N             0          270000.0   1293502.5      35698.5
2               Y             0           67500.0    135000.0       6750.0
3               Y             0          135000.0    312682.5      29686.5
4               Y             0          121500.0    513000.0      21865.5

   …  FLAG_DOCUMENT_18 FLAG_DOCUMENT_19 FLAG_DOCUMENT_20 FLAG_DOCUMENT_21  \
0  …               0.0              0.0              0.0              0.0
1  …               0.0              0.0              0.0              0.0
2  …               0.0              0.0              0.0              0.0
3  …               0.0              0.0              0.0              0.0
4  …               0.0              0.0              0.0              0.0

  AMT_REQ_CREDIT_BUREAU_HOUR AMT_REQ_CREDIT_BUREAU_DAY  \
0                        0.0                       0.0
1                        0.0                       0.0
2                        0.0                       0.0
3                        NaN                       NaN
4                        0.0                       0.0

   AMT_REQ_CREDIT_BUREAU_WEEK  AMT_REQ_CREDIT_BUREAU_MON  \
0                         0.0                        0.0
1                         0.0                        0.0
2                         0.0                        0.0
3                         NaN                        NaN
4                         0.0                        0.0

   AMT_REQ_CREDIT_BUREAU_QRT  AMT_REQ_CREDIT_BUREAU_YEAR
0                        0.0                         1.0
```

```
1                       0.0                         0.0
2                       0.0                         0.0
3                       NaN                         NaN
4                       0.0                         0.0

[5 rows x 122 columns]
```

[8]:
```python
# Task2
df.isnull()
```

[8]:
```
       SK_ID_CURR  TARGET  NAME_CONTRACT_TYPE  CODE_GENDER  FLAG_OWN_CAR  \
0           False   False               False        False         False
1           False   False               False        False         False
2           False   False               False        False         False
3           False   False               False        False         False
4           False   False               False        False         False
…             …      …                  …            …             …
96997       False   False               False        False         False
96998       False   False               False        False         False
96999       False   False               False        False         False
97000       False   False               False        False         False
97001       False   False               False        False         False

       FLAG_OWN_REALTY  CNT_CHILDREN  AMT_INCOME_TOTAL  AMT_CREDIT  \
0                False         False             False       False
1                False         False             False       False
2                False         False             False       False
3                False         False             False       False
4                False         False             False       False
…                  …             …                 …           …
96997            False         False             False       False
96998            False         False             False       False
96999            False         False             False       False
97000            False         False             False       False
97001            False         False             False       False

       AMT_ANNUITY  …  FLAG_DOCUMENT_18  FLAG_DOCUMENT_19  FLAG_DOCUMENT_20  \
0            False  …             False             False             False
1            False  …             False             False             False
2            False  …             False             False             False
3            False  …             False             False             False
4            False  …             False             False             False
…              …  …             …                 …                 …
96997        False  …             False             False             False
96998        False  …             False             False             False
96999        False  …             False             False             False
97000        False  …             False             False             False
```

```
97001       False  …            True             True             True

       FLAG_DOCUMENT_21  AMT_REQ_CREDIT_BUREAU_HOUR  \
0                 False                       False
1                 False                       False
2                 False                       False
3                 False                        True
4                 False                       False
…                   …                       …
96997             False                       False
96998             False                       False
96999             False                       False
97000             False                       False
97001              True                        True

       AMT_REQ_CREDIT_BUREAU_DAY  AMT_REQ_CREDIT_BUREAU_WEEK  \
0                          False                       False
1                          False                       False
2                          False                       False
3                           True                        True
4                          False                       False
…                          …                       …
96997                      False                       False
96998                      False                       False
96999                      False                       False
97000                      False                       False
97001                       True                        True

       AMT_REQ_CREDIT_BUREAU_MON  AMT_REQ_CREDIT_BUREAU_QRT  \
0                          False                      False
1                          False                      False
2                          False                      False
3                           True                       True
4                          False                      False
…                          …                      …
96997                      False                      False
96998                      False                      False
96999                      False                      False
97000                      False                      False
97001                       True                       True

       AMT_REQ_CREDIT_BUREAU_YEAR
0                           False
1                           False
2                           False
3                            True
4                           False
```

```
…                           …
96997                       False
96998                       False
96999                       False
97000                       False
97001                        True

[97002 rows x 122 columns]
```

[9]:
```
df.isnull().sum()
```

[9]:
```
SK_ID_CURR                       0
TARGET                           0
NAME_CONTRACT_TYPE               0
CODE_GENDER                      0
FLAG_OWN_CAR                     0
                                 …
AMT_REQ_CREDIT_BUREAU_DAY    13007
AMT_REQ_CREDIT_BUREAU_WEEK   13007
AMT_REQ_CREDIT_BUREAU_MON    13007
AMT_REQ_CREDIT_BUREAU_QRT    13007
AMT_REQ_CREDIT_BUREAU_YEAR   13007
Length: 122, dtype: int64
```

[10]:
```
df.head()
```

[10]:
```
   SK_ID_CURR  TARGET NAME_CONTRACT_TYPE CODE_GENDER FLAG_OWN_CAR  \
0      100002       1         Cash loans           M            N
1      100003       0         Cash loans           F            N
2      100004       0    Revolving loans           M            Y
3      100006       0         Cash loans           F            N
4      100007       0         Cash loans           M            N

  FLAG_OWN_REALTY  CNT_CHILDREN  AMT_INCOME_TOTAL  AMT_CREDIT  AMT_ANNUITY  \
0               Y             0          202500.0    406597.5      24700.5
1               N             0          270000.0   1293502.5      35698.5
2               Y             0           67500.0    135000.0       6750.0
3               Y             0          135000.0    312682.5      29686.5
4               Y             0          121500.0    513000.0      21865.5

   …  FLAG_DOCUMENT_18 FLAG_DOCUMENT_19 FLAG_DOCUMENT_20 FLAG_DOCUMENT_21  \
0  …               0.0              0.0              0.0              0.0
1  …               0.0              0.0              0.0              0.0
2  …               0.0              0.0              0.0              0.0
3  …               0.0              0.0              0.0              0.0
4  …               0.0              0.0              0.0              0.0
```

```
  AMT_REQ_CREDIT_BUREAU_HOUR AMT_REQ_CREDIT_BUREAU_DAY  \
0                        0.0                       0.0
1                        0.0                       0.0
2                        0.0                       0.0
3                        NaN                       NaN
4                        0.0                       0.0

   AMT_REQ_CREDIT_BUREAU_WEEK  AMT_REQ_CREDIT_BUREAU_MON  \
0                         0.0                        0.0
1                         0.0                        0.0
2                         0.0                        0.0
3                         NaN                        NaN
4                         0.0                        0.0

   AMT_REQ_CREDIT_BUREAU_QRT  AMT_REQ_CREDIT_BUREAU_YEAR
0                        0.0                         1.0
1                        0.0                         0.0
2                        0.0                         0.0
3                        NaN                         NaN
4                        0.0                         0.0

[5 rows x 122 columns]
```

[11]:
```
# TASK 3
defaulters=(df.TARGET==1).sum()
payers=(df.TARGET==0).sum()
print((defaulters/payers)*100)
```

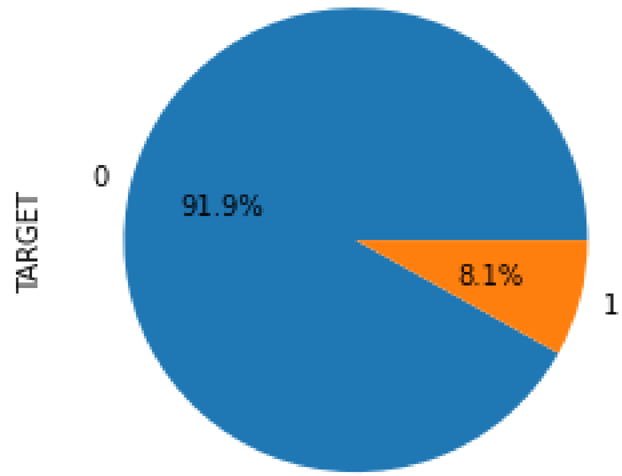
```
8.812509815359073
```

[12]:
```
without_id=[column for column in df.columns if column!='SK_ID_CURR']

#check for duplicate values
na=df[df.duplicated(subset=without_id,keep=False)]
print("Duplicates are: ",na.shape[0])
```

```
Duplicates are:  0
```

[13]:
```
df.TARGET.value_counts().plot(kind='pie',autopct='%1.1f%%')
```

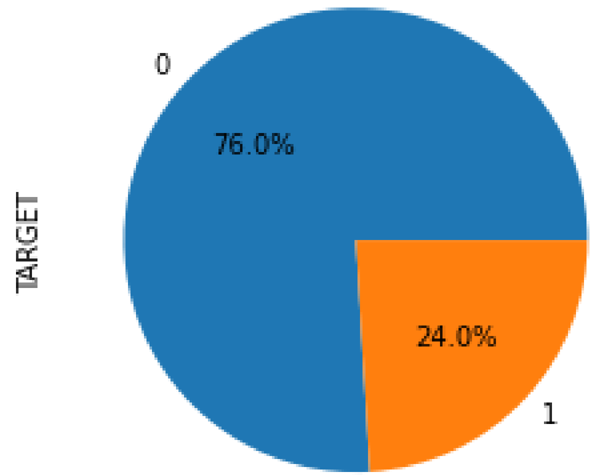
[13]:
```
<AxesSubplot: ylabel='TARGET'>
```



[14]:
```
import matplotlib as plt
```

[15]:
```
shuffled_data=df.sample(frac=1,random_state=3)
unpaid_home_loan=shuffled_data.loc[shuffled_data['TARGET']==1]
paid_home_loan=shuffled_data.loc[shuffled_data['TARGET']==0].
↪sample(n=24825,random_state=69)
normalised_df=pd.concat([unpaid_home_loan,paid_home_loan])
normalised_df.TARGET.value_counts().plot(kind='pie',autopct="%1.1f%%")
```

[15]: <AxesSubplot: ylabel='TARGET'>



[16]:
```
import tensorflow as tf
```

```
2024-05-30 10:21:39.377724: I tensorflow/core/util/port.cc:110] oneDNN custom
operations are on. You may see slightly different numerical results due to
floating-point round-off errors from different computation orders. To turn them
off, set the environment variable `TF_ENABLE_ONEDNN_OPTS=0`.
2024-05-30 10:21:39.410985: I tensorflow/core/platform/cpu_feature_guard.cc:182]
This TensorFlow binary is optimized to use available CPU instructions in
performance-critical operations.
To enable the following instructions: AVX2 AVX512F AVX512_VNNI FMA, in other
operations, rebuild TensorFlow with the appropriate compiler flags.

VOC-NOTICE: GPU memory for this assignment is capped at 1024MiB

2024-05-30 10:21:41.336953: E
tensorflow/compiler/xla/stream_executor/cuda/cuda_driver.cc:268] failed call to
cuInit: CUDA_ERROR_NO_DEVICE: no CUDA-capable device is detected
```

[17]:
```
normalised_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 32681 entries, 85856 to 87614
Columns: 122 entries, SK_ID_CURR to AMT_REQ_CREDIT_BUREAU_YEAR
dtypes: float64(103), int64(3), object(16)
memory usage: 30.7+ MB
```

[18]:
```
normalised_df.head()
```

[18]:

```
       SK_ID_CURR  TARGET NAME_CONTRACT_TYPE CODE_GENDER FLAG_OWN_CAR  \
85856      199627       1         Cash loans           M            N
23339      127153       1         Cash loans           F            N
55030      163763       1         Cash loans           F            N
40383      146780       1         Cash loans           M            Y
65450      175905       1         Cash loans           F            N

      FLAG_OWN_REALTY  CNT_CHILDREN  AMT_INCOME_TOTAL  AMT_CREDIT  \
85856               Y             1          180000.0    251280.0
23339               N             0          202500.0    545040.0
55030               Y             0          112500.0    308461.5
40383               Y             0          157500.0    592560.0
65450               Y             0           81000.0    640080.0

       AMT_ANNUITY  …  FLAG_DOCUMENT_18 FLAG_DOCUMENT_19 FLAG_DOCUMENT_20  \
85856      17127.0  …               0.0              0.0              0.0
23339      25407.0  …               0.0              0.0              0.0
55030      15970.5  …               0.0              0.0              0.0
40383      35937.0  …               0.0              0.0              0.0
65450      29970.0  …               0.0              0.0              0.0

      FLAG_DOCUMENT_21 AMT_REQ_CREDIT_BUREAU_HOUR AMT_REQ_CREDIT_BUREAU_DAY  \
85856              0.0                        NaN                       NaN
23339              0.0                        0.0                       0.0
55030              0.0                        0.0                       0.0
40383              0.0                        0.0                       0.0
65450              0.0                        0.0                       0.0

       AMT_REQ_CREDIT_BUREAU_WEEK  AMT_REQ_CREDIT_BUREAU_MON  \
85856                         NaN                        NaN
23339                         0.0                        0.0
55030                         0.0                        0.0
40383                         0.0                        0.0
65450                         0.0                        0.0

       AMT_REQ_CREDIT_BUREAU_QRT  AMT_REQ_CREDIT_BUREAU_YEAR
85856                        NaN                         NaN
23339                        2.0                         1.0
55030                        0.0                         0.0
40383                        0.0                         0.0
65450                        0.0                         1.0

[5 rows x 122 columns]
```

[19]:

```
normalised_df.dropna(axis=0)
normalised_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 32681 entries, 85856 to 87614
Columns: 122 entries, SK_ID_CURR to AMT_REQ_CREDIT_BUREAU_YEAR
dtypes: float64(103), int64(3), object(16)
memory usage: 30.7+ MB
```

[20]:
```
normalised_df.isnull().sum()
```

[20]:
```
SK_ID_CURR                       0
TARGET                           0
NAME_CONTRACT_TYPE               0
CODE_GENDER                      0
FLAG_OWN_CAR                     0
                              …
AMT_REQ_CREDIT_BUREAU_DAY     4466
AMT_REQ_CREDIT_BUREAU_WEEK    4466
AMT_REQ_CREDIT_BUREAU_MON     4466
AMT_REQ_CREDIT_BUREAU_QRT     4466
AMT_REQ_CREDIT_BUREAU_YEAR    4466
Length: 122, dtype: int64
```

[21]:
```
print(pd.unique(normalised_df.AMT_REQ_CREDIT_BUREAU_DAY))
print(pd.unique(normalised_df.AMT_REQ_CREDIT_BUREAU_WEEK))
print(pd.unique(normalised_df.AMT_REQ_CREDIT_BUREAU_MON))
print(pd.unique(normalised_df.AMT_REQ_CREDIT_BUREAU_QRT))
print(pd.unique(normalised_df.AMT_REQ_CREDIT_BUREAU_YEAR))
```

```
[nan  0.  1.  2.  5.  3.]
[nan  0.  2.  1.  3.  5.  6.  4.]
[nan  0.  1.  2.  7.  3.  4. 10.  5. 15.  6. 11. 12. 13.  8. 14.  9. 16.
 18. 19. 23.]
[nan  2.  0.  1.  4.  3.  7.  5.  6.]
[nan  1.  0.  2.  7.  4.  5.  3. 11.  9.  6.  8. 10. 16. 22. 13. 12. 14.
 15.]
```

[22]:
```
normalised_df.dropna(axis=0)
```

[22]:
```
       SK_ID_CURR  TARGET NAME_CONTRACT_TYPE CODE_GENDER FLAG_OWN_CAR  \
95728      211129       1         Cash loans           M            Y
30040      134873       1         Cash loans           F            Y
34989      140540       1         Cash loans           M            Y
47592      155129       1         Cash loans           M            Y
18206      121237       1         Cash loans           M            Y
…             …     …                  …         …            …
68396      179331       0         Cash loans           F            Y
60054      169632       0    Revolving loans           M            Y
33006      138251       0         Cash loans           F            Y
70554      181849       0         Cash loans           F            Y
```

```
96516      212046       0         Cash loans           F            Y

      FLAG_OWN_REALTY  CNT_CHILDREN  AMT_INCOME_TOTAL  AMT_CREDIT  \
95728               N             0          135000.0    840996.0
30040               N             0          157500.0    675000.0
34989               Y             0          180000.0    749349.0
47592               Y             1          112500.0    284400.0
18206               N             0          166500.0    239850.0
…                 …             …                …           …
68396               Y             0          144000.0    469152.0
60054               N             0          225000.0    225000.0
33006               Y             0           90000.0    526491.0
70554               Y             1          270000.0    518562.0
96516               Y             0          112500.0    704844.0

       AMT_ANNUITY  …  FLAG_DOCUMENT_18 FLAG_DOCUMENT_19 FLAG_DOCUMENT_20  \
95728      29925.0  …               0.0              0.0              0.0
30040      49117.5  …               0.0              0.0              0.0
34989      29164.5  …               0.0              0.0              0.0
47592      18643.5  …               0.0              0.0              0.0
18206      25447.5  …               0.0              0.0              0.0
…              …  …               …              …              …
68396      23953.5  …               0.0              0.0              0.0
60054      11250.0  …               0.0              0.0              0.0
33006      26878.5  …               0.0              0.0              0.0
70554      25078.5  …               0.0              0.0              0.0
96516      26248.5  …               0.0              0.0              0.0

      FLAG_DOCUMENT_21 AMT_REQ_CREDIT_BUREAU_HOUR AMT_REQ_CREDIT_BUREAU_DAY  \
95728              0.0                        0.0                       0.0
30040              0.0                        0.0                       0.0
34989              0.0                        0.0                       0.0
47592              0.0                        0.0                       0.0
18206              0.0                        0.0                       0.0
…                  …                        …                       …
68396              0.0                        0.0                       0.0
60054              0.0                        0.0                       0.0
33006              0.0                        0.0                       0.0
70554              0.0                        0.0                       0.0
96516              0.0                        0.0                       0.0

       AMT_REQ_CREDIT_BUREAU_WEEK  AMT_REQ_CREDIT_BUREAU_MON  \
95728                         0.0                        0.0
30040                         1.0                        0.0
34989                         0.0                        0.0
47592                         0.0                        1.0
18206                         0.0                        0.0
```

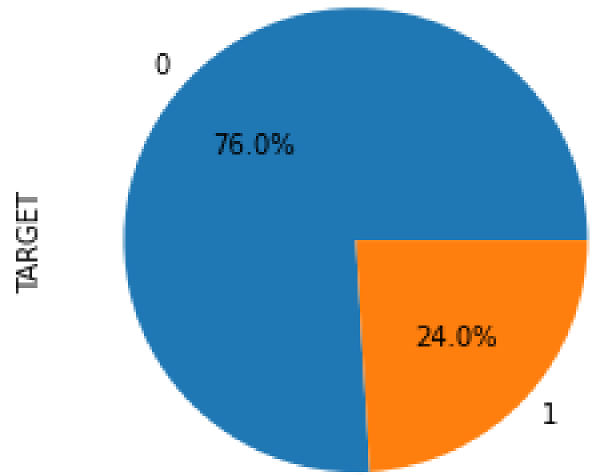
```
…                            …                        …
68396                      0.0                      0.0
60054                      0.0                      0.0
33006                      1.0                      0.0
70554                      0.0                      0.0
96516                      0.0                      3.0

       AMT_REQ_CREDIT_BUREAU_QRT  AMT_REQ_CREDIT_BUREAU_YEAR
95728                        0.0                         3.0
30040                        0.0                         1.0
34989                        0.0                         0.0
47592                        2.0                         0.0
18206                        0.0                         0.0
…                            …                           …
68396                        0.0                         0.0
60054                        0.0                         3.0
33006                        0.0                         5.0
70554                        0.0                         1.0
96516                        1.0                         1.0

[846 rows x 122 columns]
```

[23]:
```
print(normalised_df.info())
print(normalised_df.isnull().sum())
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 32681 entries, 85856 to 87614
Columns: 122 entries, SK_ID_CURR to AMT_REQ_CREDIT_BUREAU_YEAR
dtypes: float64(103), int64(3), object(16)
memory usage: 30.7+ MB
None
SK_ID_CURR                     0
TARGET                         0
NAME_CONTRACT_TYPE             0
CODE_GENDER                    0
FLAG_OWN_CAR                   0
                            …
AMT_REQ_CREDIT_BUREAU_DAY   4466
AMT_REQ_CREDIT_BUREAU_WEEK  4466
AMT_REQ_CREDIT_BUREAU_MON   4466
AMT_REQ_CREDIT_BUREAU_QRT   4466
AMT_REQ_CREDIT_BUREAU_YEAR  4466
Length: 122, dtype: int64
```

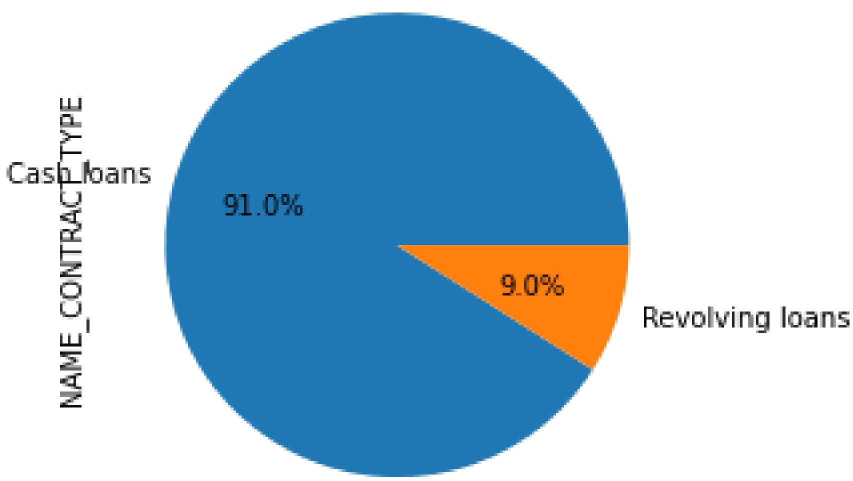
[24]:
```
normalised_df.TARGET.value_counts().plot(kind='pie',autopct="%1.1f%%")
```

[24]: <AxesSubplot: ylabel='TARGET'>



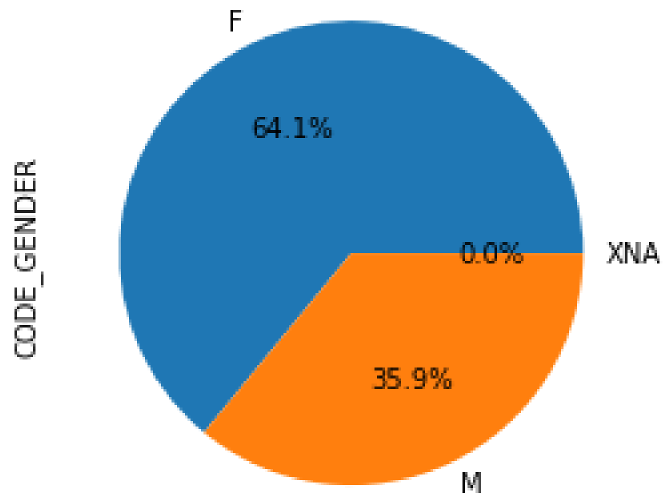
[25]:
```
normalised_df.NAME_CONTRACT_TYPE.value_counts().plot(kind='pie',autopct="%1.
↪1f%%")
```

[25]: <AxesSubplot: ylabel='NAME_CONTRACT_TYPE'>



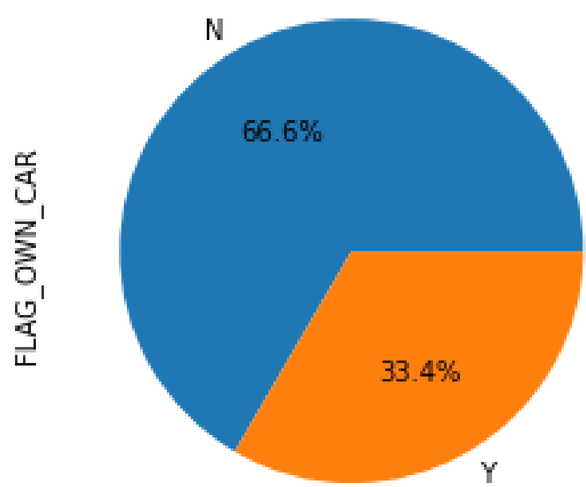
[26]:
```
normalised_df.CODE_GENDER.value_counts().plot(kind='pie',autopct="%1.1f%%")
```

[26]: <AxesSubplot: ylabel='CODE_GENDER'>



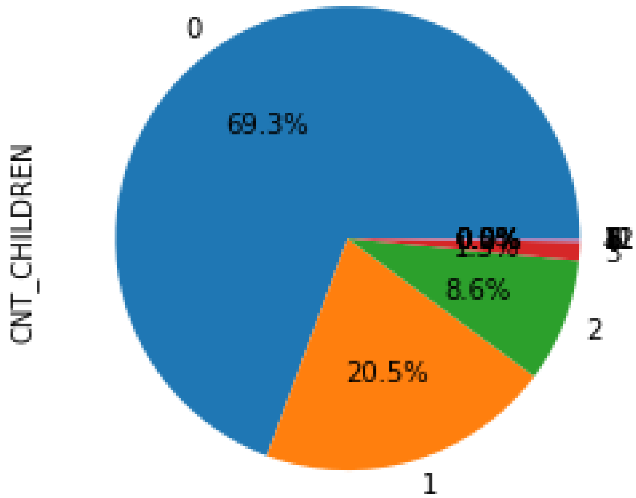
[27]:
```
normalised_df.FLAG_OWN_CAR.value_counts().plot(kind='pie',autopct="%1.1f%%")
```

[27]: <AxesSubplot: ylabel='FLAG_OWN_CAR'>



[28]:
```
normalised_df.CNT_CHILDREN.value_counts().plot(kind='pie',autopct="%1.1f%%")
```

[28]: <AxesSubplot: ylabel='CNT_CHILDREN'>



[29]:
```
print((normalised_df[normalised_df['AMT_INCOME_TOTAL']>1000000]['TARGET'].
  ↪value_counts())/len(normalised_df[normalised_df['AMT_INCOME_TOTAL'] >␣
  ↪1000000])*100)
```

```
0    86.956522
1    13.043478
Name: TARGET, dtype: float64
```

[30]:
```
print((normalised_df[normalised_df['CNT_CHILDREN']>2]['TARGET'].value_counts())/
  ↪len(normalised_df[normalised_df['CNT_CHILDREN'] > 2])*100)
print((normalised_df[normalised_df['CNT_CHILDREN']>5]['TARGET'].value_counts())/
  ↪len(normalised_df[normalised_df['CNT_CHILDREN'] > 5])*100)
#as number of children is increasing lone defaulters are increasing
```

```
0    72.379032
1    27.620968
Name: TARGET, dtype: float64
1    57.142857
0    42.857143
Name: TARGET, dtype: float64
```

[31]:
```
print((normalised_df[normalised_df['FLAG_OWN_CAR']=='N']['TARGET'].
  ↪value_counts())/len(normalised_df[normalised_df['FLAG_OWN_CAR'] =='N'])*100)
print((normalised_df[normalised_df['FLAG_OWN_CAR']=='Y']['TARGET'].
  ↪value_counts())/len(normalised_df[normalised_df['FLAG_OWN_CAR'] =='Y'])*100)
```

```
#people with own cars are slighlty more likely to repay back the loan
```

```
0    75.05744
1    24.94256
Name: TARGET, dtype: float64
0    77.763531
1    22.236469
Name: TARGET, dtype: float64
```

[32]:
```
print((normalised_df[normalised_df['CODE_GENDER']=='M']['TARGET'].
 ↪value_counts())/len(normalised_df[normalised_df['CODE_GENDER'] =='M'])*100)
print((normalised_df[normalised_df['CODE_GENDER']=='F']['TARGET'].
 ↪value_counts())/len(normalised_df[normalised_df['CODE_GENDER'] =='F'])*100)

#men more likely to default in payment of loans
```

```
0    71.334923
1    28.665077
Name: TARGET, dtype: float64
0    78.556415
1    21.443585
Name: TARGET, dtype: float64
```

[33]:
```
print((normalised_df[normalised_df['NAME_CONTRACT_TYPE']=='Cash␣
 ↪loans']['TARGET'].value_counts())/
 ↪len(normalised_df[normalised_df['NAME_CONTRACT_TYPE']=='Cash loans'])*100)
print((normalised_df[normalised_df['NAME_CONTRACT_TYPE']=='Revolving␣
 ↪loans']['TARGET'].value_counts())/
 ↪len(normalised_df[normalised_df['NAME_CONTRACT_TYPE']=='Revolving␣
 ↪loans'])*100)

#cash loans have a higher percent of defaulters
```

```
0    75.314455
1    24.685545
Name: TARGET, dtype: float64
0    82.490668
1    17.509332
Name: TARGET, dtype: float64
```

[34]:
```
normalised_df=normalised_df.sample(frac=1,random_state=5)
```

[35]:
```
from sklearn.preprocessing import OrdinalEncoder

ordenc=OrdinalEncoder()
normalised_df['NAME_CONTRACT_TYPE_CODE']=ordenc.
 ↪fit_transform(normalised_df[['NAME_CONTRACT_TYPE']])
```

```
print(normalised_df[['NAME_CONTRACT_TYPE','NAME_CONTRACT_TYPE_CODE']].head(10))
print(normalised_df['NAME_CONTRACT_TYPE_CODE'].value_counts())
```

```
      NAME_CONTRACT_TYPE  NAME_CONTRACT_TYPE_CODE
21080         Cash loans                      0.0
58935         Cash loans                      0.0
84153         Cash loans                      0.0
92916    Revolving loans                      1.0
16307         Cash loans                      0.0
76358         Cash loans                      0.0
73449         Cash loans                      0.0
35835         Cash loans                      0.0
80913         Cash loans                      0.0
85527         Cash loans                      0.0
0.0    29734
1.0     2947
Name: NAME_CONTRACT_TYPE_CODE, dtype: int64
```

[36]:
```
normalised_df['CODE_GENDER_CODE']=ordenc.
 ↪fit_transform(normalised_df[['CODE_GENDER']])
print(normalised_df[['CODE_GENDER','CODE_GENDER_CODE']].head(10))
print(normalised_df['CODE_GENDER_CODE'].value_counts())
```

```
      CODE_GENDER  CODE_GENDER_CODE
21080           F               0.0
58935           F               0.0
84153           F               0.0
92916           F               0.0
16307           M               1.0
76358           F               0.0
73449           F               0.0
35835           F               0.0
80913           F               0.0
85527           F               0.0
0.0    20934
1.0    11746
2.0        1
Name: CODE_GENDER_CODE, dtype: int64
```

[37]:
```
# 2 other values in code_gender
normalised_df.loc[normalised_df['CODE_GENDER_CODE']==2]
```

[37]:
```
       SK_ID_CURR  TARGET NAME_CONTRACT_TYPE CODE_GENDER FLAG_OWN_CAR  \
38566      144669       0    Revolving loans         XNA            N

      FLAG_OWN_REALTY  CNT_CHILDREN  AMT_INCOME_TOTAL  AMT_CREDIT  \
38566               Y             2          157500.0    270000.0
```

```
       AMT_ANNUITY  …  FLAG_DOCUMENT_20 FLAG_DOCUMENT_21  \
38566      13500.0  …               0.0              0.0

      AMT_REQ_CREDIT_BUREAU_HOUR AMT_REQ_CREDIT_BUREAU_DAY  \
38566                        0.0                       0.0

      AMT_REQ_CREDIT_BUREAU_WEEK AMT_REQ_CREDIT_BUREAU_MON  \
38566                        0.0                       3.0

       AMT_REQ_CREDIT_BUREAU_QRT  AMT_REQ_CREDIT_BUREAU_YEAR  \
38566                        0.0                         4.0

       NAME_CONTRACT_TYPE_CODE  CODE_GENDER_CODE
38566                      1.0               2.0

[1 rows x 124 columns]
```

[38]:

```
normalised_df['FLAG_OWN_CAR_CODE']=ordenc.
 ↪fit_transform(normalised_df[['FLAG_OWN_CAR']])
print(normalised_df[['FLAG_OWN_CAR','FLAG_OWN_CAR_CODE']].head(10))
print(normalised_df['FLAG_OWN_CAR_CODE'].value_counts())
```

```
      FLAG_OWN_CAR  FLAG_OWN_CAR_CODE
21080            N                0.0
58935            N                0.0
84153            N                0.0
92916            Y                1.0
16307            N                0.0
76358            N                0.0
73449            Y                1.0
35835            Y                1.0
80913            N                0.0
85527            N                0.0
0.0    21762
1.0    10919
Name: FLAG_OWN_CAR_CODE, dtype: int64
```

[39]:

```
normalised_df['CNT_CHILDREN_CODE']=ordenc.
 ↪fit_transform(normalised_df[['CNT_CHILDREN']])
print(normalised_df[['CNT_CHILDREN_CODE','CNT_CHILDREN']].head(10))
print(normalised_df['CNT_CHILDREN_CODE'].value_counts())
```

```
       CNT_CHILDREN_CODE  CNT_CHILDREN
21080                1.0             1
58935                0.0             0
84153                2.0             2
```

```
92916                0.0               0
16307                0.0               0
76358                2.0               2
73449                0.0               0
35835                1.0               1
80913                1.0               1
85527                0.0               0
0.0     22663
1.0      6701
2.0      2821
3.0       421
4.0        62
5.0         6
6.0         3
7.0         1
10.0        1
8.0         1
9.0         1
Name: CNT_CHILDREN_CODE, dtype: int64
```

[40]:
```
normalised_df=normalised_df.sample(frac=1,random_state=45)
```

[41]:
```
normalised_df['TARGET'].value_counts()
```

[41]:
```
0    24825
1     7856
Name: TARGET, dtype: int64
```

[42]:
```
y=normalised_df.TARGET
```

[43]:
```
normalised_df_features=['SK_ID_CURR','NAME_CONTRACT_TYPE_CODE','CNT_CHILDREN_CODE','FLAG_OWN_C
```

[44]:
```
from sklearn.model_selection import train_test_split
```

[45]:
```
X=normalised_df[normalised_df_features]
```

[46]:
```
from sklearn.datasets import make_blobs
```

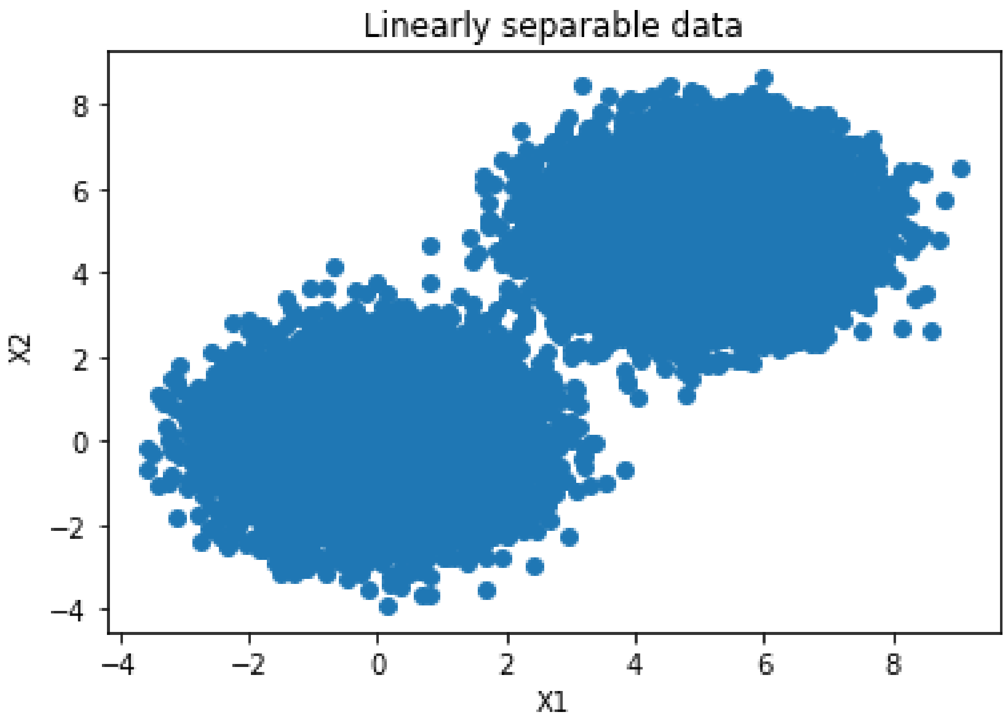
[47]:
```
blobs_random_seed = 42
centers = [(0,0), (5,5)]
cluster_std = 1
frac_test_split = 0.33
num_features_for_samples = 2
num_samples_total = 49650

# Generate data
```

```
inputs, targets = make_blobs (n_samples = num_samples_total, centers = centers,␣
↪n_features = num_features_for_samples, cluster_std = cluster_std)

X_train,X_test,y_train,y_test=train_test_split(inputs,targets,test_size=0.
↪33,random_state=45)
```

[48]:
```
print(X_train.shape, X_test.shape, y_train.shape, y_test.shape)
```

```
(33265, 2) (16385, 2) (33265,) (16385,)
```

[49]:
```
import matplotlib.pyplot as plt

# Assuming X_train is defined and is a 2D array-like structure
plt.scatter(X_train[:,0], X_train[:,1])
plt.title('Linearly separable data')
plt.xlabel('X1')
plt.ylabel('X2')
plt.show()
```



[50]:
```
pip install --upgrade scikit-learn
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: scikit-learn in ./.local/lib/python3.10/site-
packages (1.5.0)
```

```
Requirement already satisfied: numpy>=1.19.5 in /usr/local/lib/python3.10/site-
packages (from scikit-learn) (1.23.5)
Requirement already satisfied: scipy>=1.6.0 in /usr/local/lib/python3.10/site-
packages (from scikit-learn) (1.9.3)
Requirement already satisfied: joblib>=1.2.0 in /usr/local/lib/python3.10/site-
packages (from scikit-learn) (1.2.0)
Requirement already satisfied: threadpoolctl>=3.1.0 in
/usr/local/lib/python3.10/site-packages (from scikit-learn) (3.1.0)

[notice] A new release of pip is
available: 23.3 -> 24.0
[notice] To update, run:
pip install --upgrade pip
Note: you may need to restart the kernel to use updated packages.
```

[51]:

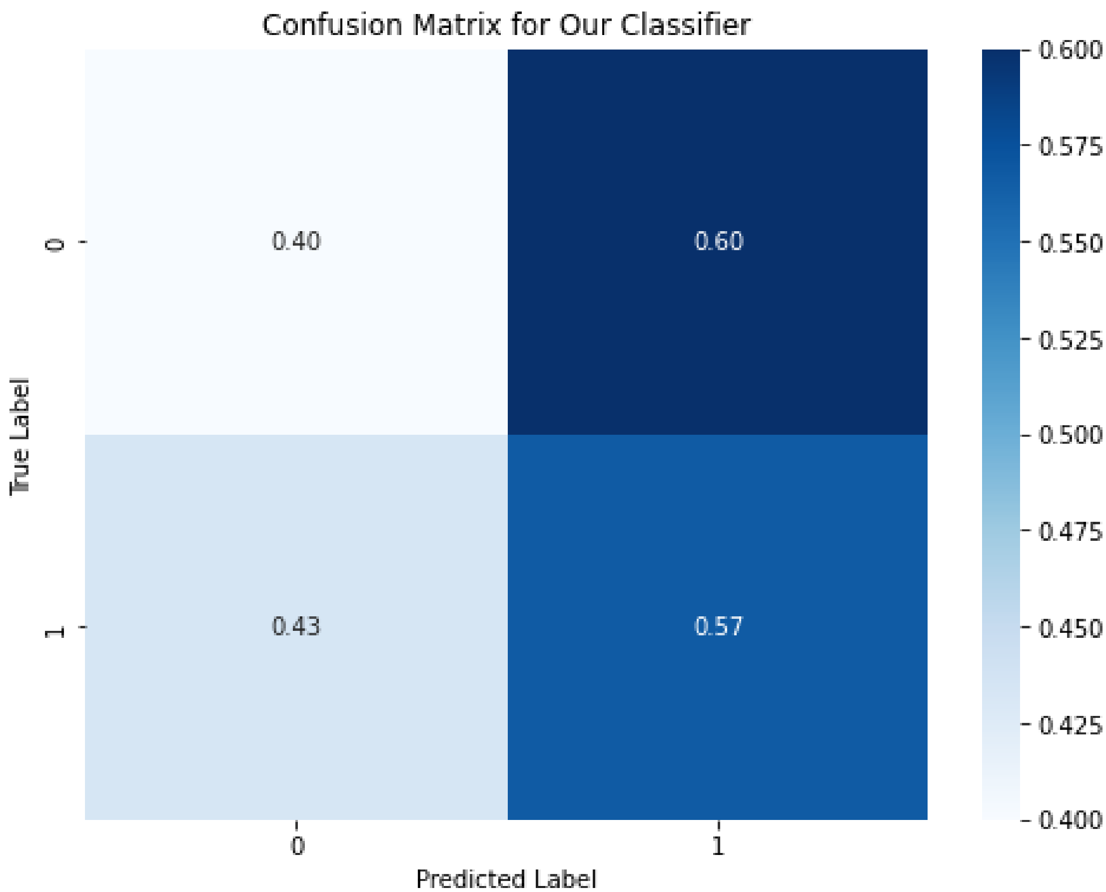
```
from sklearn import svm
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
import matplotlib.pyplot as plt
```

[52]:

```
clf=svm.SVC(kernel='linear')
clf=clf.fit(X_train,y_train)
```

[53]:

```
# Generate confusion matrix
import matplotlib.pyplot as plt
from sklearn.metrics import confusion_matrix
import numpy as np
import seaborn as sns


y_test = np.random.randint(0, 2, size=100)  # Replace with actual y_test
y_pred = np.random.randint(0, 2, size=100)  # Replace with actual y_pred

# Compute confusion matrix
cm = confusion_matrix(y_test, y_pred)
cm_normalized = cm.astype('float') / cm.sum(axis=1)[:, np.newaxis]

# Plot confusion matrix
plt.figure(figsize=(8, 6))
sns.heatmap(cm_normalized, annot=True, fmt='.2f', cmap='Blues', xticklabels=np.
 ↪unique(y_test), yticklabels=np.unique(y_test))
plt.title('Confusion Matrix for Our Classifier')
plt.xlabel('Predicted Label')
plt.ylabel('True Label')
plt.show()
```



[54]:
```
from sklearn.metrics import precision_score, recall_score,f1_score
```

[55]:
```
import numpy as np
from sklearn.datasets import make_classification
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import precision_score, recall_score, f1_score

# Generate confusion matrix
X, y = make_classification(n_samples=1000, n_features=20, random_state=42)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
 ↪random_state=42)

# Train a classifier
model = RandomForestClassifier()
model.fit(X_train, y_train)

# Make predictions
predictions = model.predict(X_test)
```

```
# Calculate precision, recall, and F1 scores
print(f"Precision (micro): {precision_score(y_test, predictions,␣
↪average='micro')}")
print(f"Recall (micro): {recall_score(y_test, predictions, average='micro')}")
print(f"F1 Score (micro): {f1_score(y_test, predictions, average='micro')}")
```

```
Precision (micro): 0.8633333333333333
Recall (micro): 0.8633333333333333
F1 Score (micro): 0.8633333333333333
```

[59]:
```
support_vectors = clf.support_vectors_

# Visualize support vectors
plt.pyplot.scatter(X_train[:,0], X_train[:,1])
plt.pyplot.scatter(support_vectors[:,0], support_vectors[:,1], color='red')
plt.pyplot.title('Linearly separable data with support vectors')
plt.pyplot.xlabel('X1')
plt.pyplot.ylabel('X2')
plt.pyplot.show()
```

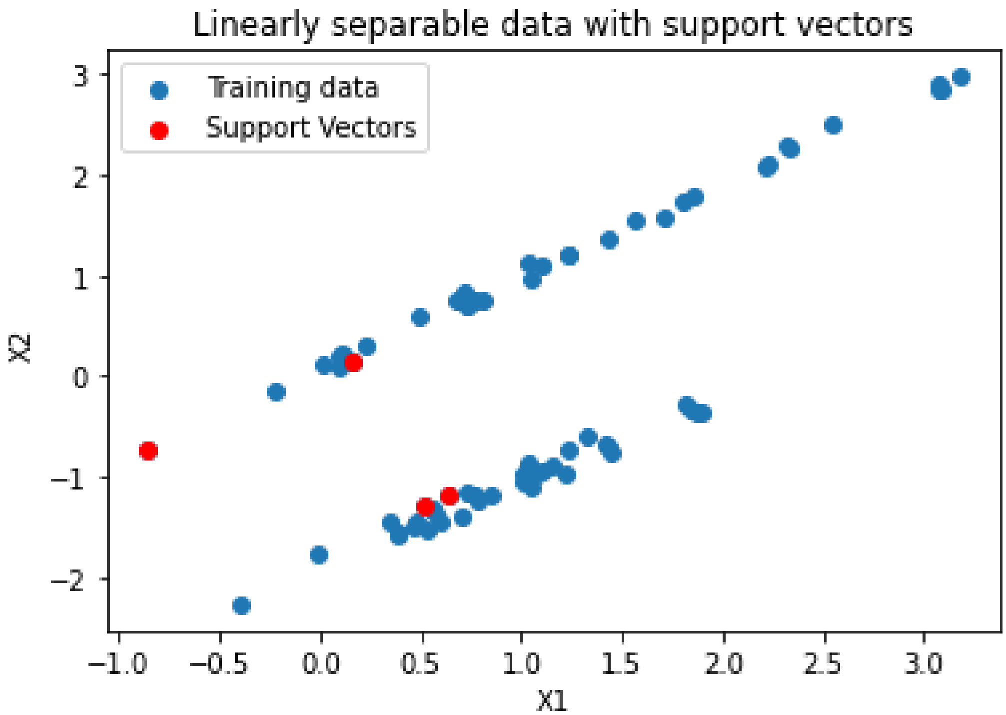
```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
/tmp/ipykernel_313/2880794247.py in <cell line: 4>()
      2
      3 # Visualize support vectors
----> 4 plt.pyplot.scatter(X_train[:,0], X_train[:,1])
      5 plt.pyplot.scatter(support_vectors[:,0], support_vectors[:,1],␣
↪color='red')
      6 plt.pyplot.title('Linearly separable data with support vectors')

AttributeError: module 'matplotlib.pyplot' has no attribute 'pyplot'
```

[ ]:
```
# !pip install mlxtend
```

[ ]:
```
from mlxtend.plotting import plot_decision_regions
```

[57]:
```
import matplotlib.pyplot as plt
import numpy as np
from sklearn import datasets
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC
from mlxtend.plotting import plot_decision_regions

# Generate example data
```

```python
X, y = datasets.make_classification(n_samples=100, n_features=2, n_informative=2, n_redundant=0, n_classes=2, n_clusters_per_class=1, random_state=42)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Train a Support Vector Classifier
clf = SVC(kernel='linear')
clf.fit(X_train, y_train)

# Get support vectors
support_vectors = clf.support_vectors_

# Visualize support vectors
plt.scatter(X_train[:, 0], X_train[:, 1], label='Training data')
plt.scatter(support_vectors[:, 0], support_vectors[:, 1], color='red', label='Support Vectors')
plt.title('Linearly separable data with support vectors')
plt.xlabel('X1')
plt.ylabel('X2')
plt.legend()
plt.show()

# Plot decision regions
plot_decision_regions(X_test, y_test, clf=clf, legend=2)
plt.title('Decision Regions for Test Data')
plt.xlabel('X1')
plt.ylabel('X2')
plt.show()
```


Linearly separable data with support vectors
Training data
Support Vectors
X2
X1
3
2
1
0
−1
−2
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0


Decision Regions for Test Data
0
1
X2
X1
2
1
0
−1
−2
−1
0
1
2

```
[ ]:
```