House Loan Data Analysis

In [1]:
```
!pip install plotly
```

```
Requirement already satisfied: plotly in c:\users\lenovo\downloads\anacon
da\lib\site-packages (5.6.0)
Requirement already satisfied: six in c:\users\lenovo\downloads\anaconda
\lib\site-packages (from plotly) (1.15.0)
Requirement already satisfied: tenacity>=6.2.0 in c:\users\lenovo\downloa
ds\anaconda\lib\site-packages (from plotly) (8.0.1)
```

In [2]:
```
!pip install cufflinks
```

```
Collecting cufflinks
  Downloading cufflinks-0.17.3.tar.gz (81 kB)
Requirement already satisfied: numpy>=1.9.2 in c:\users\lenovo\downloads
\anaconda\lib\site-packages (from cufflinks) (1.20.1)
Requirement already satisfied: pandas>=0.19.2 in c:\users\lenovo\download
s\anaconda\lib\site-packages (from cufflinks) (1.2.4)
Requirement already satisfied: plotly>=4.1.1 in c:\users\lenovo\downloads
\anaconda\lib\site-packages (from cufflinks) (5.6.0)
Requirement already satisfied: six>=1.9.0 in c:\users\lenovo\downloads\an
aconda\lib\site-packages (from cufflinks) (1.15.0)
Collecting colorlover>=0.2.1
  Downloading colorlover-0.3.0-py3-none-any.whl (8.9 kB)
Requirement already satisfied: setuptools>=34.4.1 in c:\users\lenovo\down
loads\anaconda\lib\site-packages (from cufflinks) (52.0.0.post20210125)
Requirement already satisfied: ipython>=5.3.0 in c:\users\lenovo\download
s\anaconda\lib\site-packages (from cufflinks) (7.22.0)
Requirement already satisfied: ipywidgets>=7.0.0 in c:\users\lenovo\downl
oads\anaconda\lib\site-packages (from cufflinks) (7.6.3)
Requirement already satisfied: decorator in c:\users\lenovo\downloads\ana
conda\lib\site-packages (from ipython>=5.3.0->cufflinks) (5.0.6)
Requirement already satisfied: jedi>=0.16 in c:\users\lenovo\downloads\an
aconda\lib\site-packages (from ipython>=5.3.0->cufflinks) (0.17.2)
Requirement already satisfied: colorama in c:\users\lenovo\downloads\anac
onda\lib\site-packages (from ipython>=5.3.0->cufflinks) (0.4.4)
Requirement already satisfied: traitlets>=4.2 in c:\users\lenovo\download
s\anaconda\lib\site-packages (from ipython>=5.3.0->cufflinks) (5.0.5)
Requirement already satisfied: backcall in c:\users\lenovo\downloads\anac
onda\lib\site-packages (from ipython>=5.3.0->cufflinks) (0.2.0)
Requirement already satisfied: prompt-toolkit!=3.0.0,!=3.0.1,<3.1.0,>=2.
0.0 in c:\users\lenovo\downloads\anaconda\lib\site-packages (from ipython
>=5.3.0->cufflinks) (3.0.17)
Requirement already satisfied: pygments in c:\users\lenovo\downloads\anac
onda\lib\site-packages (from ipython>=5.3.0->cufflinks) (2.8.1)
Requirement already satisfied: pickleshare in c:\users\lenovo\downloads\a
naconda\lib\site-packages (from ipython>=5.3.0->cufflinks) (0.7.5)
Requirement already satisfied: nbformat>=4.2.0 in c:\users\lenovo\downloa
ds\anaconda\lib\site-packages (from ipywidgets>=7.0.0->cufflinks) (5.1.3)
Requirement already satisfied: jupyterlab-widgets>=1.0.0 in c:\users\leno
vo\downloads\anaconda\lib\site-packages (from ipywidgets>=7.0.0->cufflink
s) (1.0.0)
Requirement already satisfied: widgetsnbextension~=3.5.0 in c:\users\leno
vo\downloads\anaconda\lib\site-packages (from ipywidgets>=7.0.0->cufflink
s) (3.5.1)
Requirement already satisfied: ipykernel>=4.5.1 in c:\users\lenovo\downlo
ads\anaconda\lib\site-packages (from ipywidgets>=7.0.0->cufflinks) (5.3.
4)
Requirement already satisfied: jupyter-client in c:\users\lenovo\download
s\anaconda\lib\site-packages (from ipykernel>=4.5.1->ipywidgets>=7.0.0->c
ufflinks) (6.1.12)
```

```
Requirement already satisfied: tornado>=4.2 in c:\users\lenovo\downloads
\anaconda\lib\site-packages (from ipykernel>=4.5.1->ipywidgets>=7.0.0->cu
fflinks) (6.1)
Requirement already satisfied: parso<0.8.0,>=0.7.0 in c:\users\lenovo\dow
nloads\anaconda\lib\site-packages (from jedi>=0.16->ipython>=5.3.0->cuffl
inks) (0.7.0)
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in c:\users\lenovo
\downloads\anaconda\lib\site-packages (from nbformat>=4.2.0->ipywidgets>=
7.0.0->cufflinks) (3.2.0)
Requirement already satisfied: ipython-genutils in c:\users\lenovo\downlo
ads\anaconda\lib\site-packages (from nbformat>=4.2.0->ipywidgets>=7.0.0->
cufflinks) (0.2.0)
Requirement already satisfied: jupyter-core in c:\users\lenovo\downloads
\anaconda\lib\site-packages (from nbformat>=4.2.0->ipywidgets>=7.0.0->cuf
flinks) (4.7.1)
Requirement already satisfied: attrs>=17.4.0 in c:\users\lenovo\downloads
\anaconda\lib\site-packages (from jsonschema!=2.5.0,>=2.4->nbformat>=4.2.
0->ipywidgets>=7.0.0->cufflinks) (20.3.0)
Requirement already satisfied: pyrsistent>=0.14.0 in c:\users\lenovo\down
loads\anaconda\lib\site-packages (from jsonschema!=2.5.0,>=2.4->nbformat>
=4.2.0->ipywidgets>=7.0.0->cufflinks) (0.17.3)
Requirement already satisfied: python-dateutil>=2.7.3 in c:\users\lenovo
\downloads\anaconda\lib\site-packages (from pandas>=0.19.2->cufflinks)
(2.8.1)
Requirement already satisfied: pytz>=2017.3 in c:\users\lenovo\downloads
\anaconda\lib\site-packages (from pandas>=0.19.2->cufflinks) (2021.1)
Requirement already satisfied: tenacity>=6.2.0 in c:\users\lenovo\downloa
ds\anaconda\lib\site-packages (from plotly>=4.1.1->cufflinks) (8.0.1)
Requirement already satisfied: wcwidth in c:\users\lenovo\downloads\anaco
nda\lib\site-packages (from prompt-toolkit!=3.0.0,!=3.0.1,<3.1.0,>=2.0.0-
>ipython>=5.3.0->cufflinks) (0.2.5)
Requirement already satisfied: notebook>=4.4.1 in c:\users\lenovo\downloa
ds\anaconda\lib\site-packages (from widgetsnbextension~=3.5.0->ipywidgets
>=7.0.0->cufflinks) (6.3.0)
Requirement already satisfied: terminado>=0.8.3 in c:\users\lenovo\downlo
ads\anaconda\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~
=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.9.4)
Requirement already satisfied: prometheus-client in c:\users\lenovo\downl
oads\anaconda\lib\site-packages (from notebook>=4.4.1->widgetsnbextension
~=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.10.1)
Requirement already satisfied: argon2-cffi in c:\users\lenovo\downloads\a
naconda\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.5.
0->ipywidgets>=7.0.0->cufflinks) (20.1.0)
Requirement already satisfied: Send2Trash>=1.5.0 in c:\users\lenovo\downl
oads\anaconda\lib\site-packages (from notebook>=4.4.1->widgetsnbextension
~=3.5.0->ipywidgets>=7.0.0->cufflinks) (1.5.0)
Requirement already satisfied: pyzmq>=17 in c:\users\lenovo\downloads\ana
conda\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.5.0-
>ipywidgets>=7.0.0->cufflinks) (20.0.0)
Requirement already satisfied: nbconvert in c:\users\lenovo\downloads\ana
conda\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.5.0-
>ipywidgets>=7.0.0->cufflinks) (6.0.7)
Requirement already satisfied: jinja2 in c:\users\lenovo\downloads\anacon
da\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.5.0->ip
ywidgets>=7.0.0->cufflinks) (2.11.3)
Requirement already satisfied: pywin32>=1.0 in c:\users\lenovo\downloads
\anaconda\lib\site-packages (from jupyter-core->nbformat>=4.2.0->ipywidge
ts>=7.0.0->cufflinks) (227)
Requirement already satisfied: pywinpty>=0.5 in c:\users\lenovo\downloads
\anaconda\lib\site-packages (from terminado>=0.8.3->notebook>=4.4.1->widg
etsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.5.7)
Requirement already satisfied: cffi>=1.0.0 in c:\users\lenovo\downloads\a
naconda\lib\site-packages (from argon2-cffi->notebook>=4.4.1->widgetsnbex
tension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (1.14.5)
Requirement already satisfied: pycparser in c:\users\lenovo\downloads\ana
conda\lib\site-packages (from cffi>=1.0.0->argon2-cffi->notebook>=4.4.1->
widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (2.20)
Requirement already satisfied: MarkupSafe>=0.23 in c:\users\lenovo\downlo
```

```
ads\anaconda\lib\site-packages (from jinja2->notebook>=4.4.1->widgetsnbex
tension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (1.1.1)
Requirement already satisfied: pandocfilters>=1.4.1 in c:\users\lenovo\do
wnloads\anaconda\lib\site-packages (from nbconvert->notebook>=4.4.1->widg
etsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (1.4.3)
Requirement already satisfied: nbclient<0.6.0,>=0.5.0 in c:\users\lenovo
\downloads\anaconda\lib\site-packages (from nbconvert->notebook>=4.4.1->w
idgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.5.3)
Requirement already satisfied: testpath in c:\users\lenovo\downloads\anac
onda\lib\site-packages (from nbconvert->notebook>=4.4.1->widgetsnbextensi
on~=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.4.4)
Requirement already satisfied: defusedxml in c:\users\lenovo\downloads\an
aconda\lib\site-packages (from nbconvert->notebook>=4.4.1->widgetsnbexten
sion~=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.7.1)
Requirement already satisfied: entrypoints>=0.2.2 in c:\users\lenovo\down
loads\anaconda\lib\site-packages (from nbconvert->notebook>=4.4.1->widget
snbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.3)
Requirement already satisfied: jupyterlab-pygments in c:\users\lenovo\dow
nloads\anaconda\lib\site-packages (from nbconvert->notebook>=4.4.1->widge
tsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.1.2)
Requirement already satisfied: bleach in c:\users\lenovo\downloads\anacon
da\lib\site-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension
~=3.5.0->ipywidgets>=7.0.0->cufflinks) (3.3.0)
Requirement already satisfied: mistune<2,>=0.8.1 in c:\users\lenovo\downl
oads\anaconda\lib\site-packages (from nbconvert->notebook>=4.4.1->widgets
nbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.8.4)
Requirement already satisfied: nest-asyncio in c:\users\lenovo\downloads
\anaconda\lib\site-packages (from nbclient<0.6.0,>=0.5.0->nbconvert->note
book>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (1.
5.1)
Requirement already satisfied: async-generator in c:\users\lenovo\downloa
ds\anaconda\lib\site-packages (from nbclient<0.6.0,>=0.5.0->nbconvert->no
tebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks)
(1.10)
Requirement already satisfied: packaging in c:\users\lenovo\downloads\ana
conda\lib\site-packages (from bleach->nbconvert->notebook>=4.4.1->widgets
nbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (20.9)
Requirement already satisfied: webencodings in c:\users\lenovo\downloads
\anaconda\lib\site-packages (from bleach->nbconvert->notebook>=4.4.1->wid
getsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.5.1)
Requirement already satisfied: pyparsing>=2.0.2 in c:\users\lenovo\downlo
ads\anaconda\lib\site-packages (from packaging->bleach->nbconvert->notebo
ok>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (2.4.
7)
Building wheels for collected packages: cufflinks
  Building wheel for cufflinks (setup.py): started
  Building wheel for cufflinks (setup.py): finished with status 'done'
  Created wheel for cufflinks: filename=cufflinks-0.17.3-py3-none-any.whl
size=68724 sha256=1fa21b9ca48d24f7d594694f4315231b470347d5982f763c195aac6
066395205
  Stored in directory: c:\users\lenovo\appdata\local\pip\cache\wheels\6b
\76\62\6da97734911ffcbdd559fd1a3f28526321f0ae699182a23866
Successfully built cufflinks
Installing collected packages: colorlover, cufflinks
Successfully installed colorlover-0.3.0 cufflinks-0.17.3
```

In [4]:

```
import pandas as pd
import sklearn
import numpy as np
import matplotlib.pyplot as plt
import os
import warnings
import seaborn as sns
from sklearn.preprocessing import OneHotEncoder
from sklearn.datasets import make_blobs
from sklearn.impute import SimpleImputer
from sklearn.pipeline import Pipeline
```

```
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import StandardScaler
from sklearn.svm import LinearSVC
from sklearn.metrics import roc_auc_score
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import roc_auc_score
from sklearn.calibration import CalibratedClassifierCV
from sklearn.metrics import confusion_matrix
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
from sklearn.linear_model import SGDClassifier
import plotly.offline as py
import plotly.graph_objs as go
from plotly.offline import init_notebook_mode, iplot
from sklearn.model_selection import train_test_split
init_notebook_mode(connected=True)
import cufflinks as cf
cf.go_offline()
import pickle
import gc

warnings.filterwarnings('ignore')
%matplotlib inline
```

In [5]:
```
!pip install lightgbm
```

```
Collecting lightgbm
  Downloading lightgbm-3.3.2-py3-none-win_amd64.whl (1.0 MB)
Requirement already satisfied: wheel in c:\users\lenovo\downloads\anaconda\lib\site-packages (from lightgbm) (0.36.2)
Requirement already satisfied: scikit-learn!=0.22.0 in c:\users\lenovo\downloads\anaconda\lib\site-packages (from lightgbm) (1.0.2)
Requirement already satisfied: scipy in c:\users\lenovo\downloads\anaconda\lib\site-packages (from lightgbm) (1.6.2)
Requirement already satisfied: numpy in c:\users\lenovo\downloads\anaconda\lib\site-packages (from lightgbm) (1.20.1)
Requirement already satisfied: joblib>=0.11 in c:\users\lenovo\downloads\anaconda\lib\site-packages (from scikit-learn!=0.22.0->lightgbm) (1.0.1)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\lenovo\downloads\anaconda\lib\site-packages (from scikit-learn!=0.22.0->lightgbm) (2.1.0)
Installing collected packages: lightgbm
Successfully installed lightgbm-3.3.2
```

In [6]:
```
import lightgbm as lgb
```

In [7]:
```
df=pd.read_csv('C:\\Users\\lenovo\\Downloads\\loan_data (1) (1).csv')
```

In [10]:
```
df.describe()
```

Out[10]:

	SK_ID_CURR	TARGET	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT
count	307511.000000	307511.000000	307511.000000	3.075110e+05	3.075110e+05
mean	278180.518577	0.080729	0.417052	1.687979e+05	5.990260e+05
std	102790.175348	0.272419	0.722121	2.371231e+05	4.024908e+05
min	100002.000000	0.000000	0.000000	2.565000e+04	4.500000e+04

25%	189145.500000	0.000000	0.000000	1.125000e+05	2.700000e+05
50%	278202.000000	0.000000	0.000000	1.471500e+05	5.135310e+05
75%	367142.500000	0.000000	1.000000	2.025000e+05	8.086500e+05
max	456255.000000	1.000000	19.000000	1.170000e+08	4.050000e+06

8 rows × 106 columns

In [11]:
```
df.columns
```

Out[11]:
```
Index(['SK_ID_CURR', 'TARGET', 'NAME_CONTRACT_TYPE', 'CODE_GENDER',
       'FLAG_OWN_CAR', 'FLAG_OWN_REALTY', 'CNT_CHILDREN', 'AMT_INCOME_TOT
AL',
       '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)
```

In [12]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 307511 entries, 0 to 307510
Columns: 122 entries, SK_ID_CURR to AMT_REQ_CREDIT_BUREAU_YEAR
dtypes: float64(65), int64(41), object(16)
memory usage: 286.2+ MB
```

In [13]:
```
df.isnull().sum()
```

Out[13]:
```
SK_ID_CURR                      0
TARGET                          0
NAME_CONTRACT_TYPE              0
CODE_GENDER                     0
FLAG_OWN_CAR                    0
                               ...
AMT_REQ_CREDIT_BUREAU_DAY     41519
AMT_REQ_CREDIT_BUREAU_WEEK    41519
AMT_REQ_CREDIT_BUREAU_MON     41519
AMT_REQ_CREDIT_BUREAU_QRT     41519
AMT_REQ_CREDIT_BUREAU_YEAR    41519
Length: 122, dtype: int64
```

In [14]:
```
df.head()
```

Out[14]:

	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLA
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	

5 rows × 122 columns

In [15]:
```
defaulters=(df.TARGET==1).sum()
payers=(df.TARGET==0).sum()
print((defaulters/payers)*100)
```

```
8.781828601345662
```

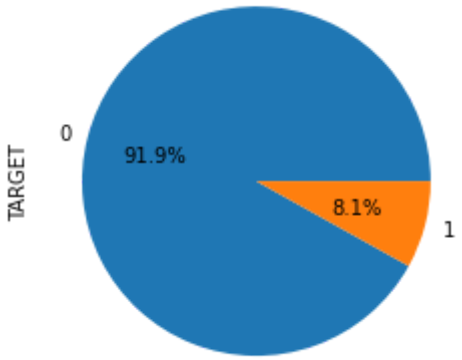
In [16]:
```
without_id=[column for column in df.columns if column!='SK_ID_CURR']
```

In [18]:
```
na=df[df.duplicated(subset=without_id,keep=False)]
print("Duplicates are: ",na.shape[0])
```

```
Duplicates are:  0
```

In [19]:
```
df.TARGET.value_counts().plot(kind='pie',autopct='%1.1f%%')
```

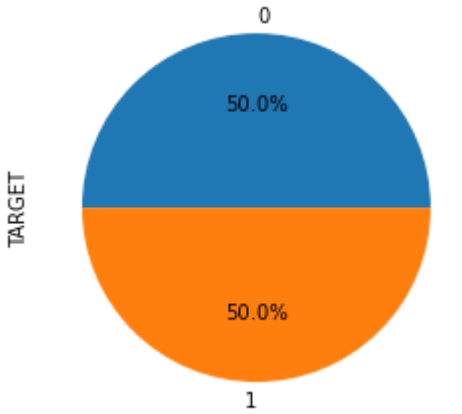
Out[19]:
```
<AxesSubplot:ylabel='TARGET'>
```



In [20]:
```
import matplotlib as plt
```

In [21]:
```
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=24
normalised_home_loan=pd.concat([unpaid_home_loan,paid_home_loan])
normalised_home_loan.TARGET.value_counts().plot(kind='pie',autopct="%1.1
```

Out[21]:
```
<AxesSubplot:ylabel='TARGET'>
```



In [22]:
```
import tensorflow as tf
```

In [23]:
```
normalised_home_loan.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 49650 entries, 207339 to 121862
Columns: 122 entries, SK_ID_CURR to AMT_REQ_CREDIT_BUREAU_YEAR
dtypes: float64(65), int64(41), object(16)
memory usage: 46.6+ MB
```

In [24]:
```
normalised_home_loan.head
```

Out[24]:
```
<bound method NDFrame.head of         SK_ID_CURR  TARGET NAME_CONTRACT_TY
PE CODE_GENDER FLAG_OWN_CAR  \
207339      340318       1         Cash loans           F            N
8756        110186       1         Cash loans           M            Y
230344      366811       1         Cash loans           F            N
178329      306645       1         Cash loans           M            Y
55586       164407       1         Cash loans           M            N
...            ...     ...                ...         ...          ...
130947      251878       0         Cash loans           F            Y
40467       146875       0         Cash loans           F            N
187004      316791       0         Cash loans           M            N
131755      252811       0         Cash loans           F            N
121862      241287       0         Cash loans           M            N

       FLAG_OWN_REALTY  CNT_CHILDREN  AMT_INCOME_TOTAL  AMT_CREDIT  \
207339               N             0          112500.0    405000.0
8756                 N             0          135000.0    544491.0
230344               Y             0          112500.0    225000.0
178329               Y             0          157500.0    595273.5
55586                N             0          157500.0    521451.0
...                ...           ...               ...         ...
130947               Y             0          135000.0    770913.0
40467                N             2          360000.0    260640.0
187004               Y             1          180000.0    688500.0
131755               Y             2          202500.0    312840.0
121862               N             0           58500.0    254700.0

        AMT_ANNUITY  ...  FLAG_DOCUMENT_18 FLAG_DOCUMENT_19 FLAG_DOCUMENT
_20  \
207339      21969.0  ...                 0                0
0
8756        17563.5  ...                 0                0
0
230344      17905.5  ...                 0                0
0
178329      29083.5  ...                 0                0
0
55586       35406.0  ...                 0                0
0
...             ...  ...               ...              ...
...
130947      24997.5  ...                 0                0
0
40467       29475.0  ...                 0                0
0
187004      22752.0  ...                 0                0
0
131755      18090.0  ...                 0                0
0
121862      13446.0  ...                 0                0
0

       FLAG_DOCUMENT_21 AMT_REQ_CREDIT_BUREAU_HOUR AMT_REQ_CREDIT_BUREAU_
```

```
DAY  \
207339                0                        0.0
0.0
8756                  0                        0.0
0.0
230344                0                        NaN
NaN
178329                0                        NaN
NaN
55586                 0                        0.0
0.0
...                 ...                        ...
...
130947                0                        0.0
0.0
40467                 0                        0.0
0.0
187004                0                        0.0
0.0
131755                0                        0.0
0.0
121862                0                        0.0
0.0

        AMT_REQ_CREDIT_BUREAU_WEEK  AMT_REQ_CREDIT_BUREAU_MON  \
207339                         0.0                        0.0
8756                           0.0                        0.0
230344                         NaN                        NaN
178329                         NaN                        NaN
55586                          0.0                        0.0
...                            ...                        ...
130947                         0.0                        1.0
40467                          0.0                        0.0
187004                         0.0                        0.0
131755                         0.0                        0.0
121862                         0.0                        0.0

        AMT_REQ_CREDIT_BUREAU_QRT  AMT_REQ_CREDIT_BUREAU_YEAR
207339                        0.0                        3.0
8756                          0.0                        0.0
230344                        NaN                        NaN
178329                        NaN                        NaN
55586                         0.0                        1.0
...                           ...                        ...
130947                        1.0                        1.0
40467                         0.0                        0.0
187004                        0.0                        0.0
131755                        1.0                        3.0
121862                        0.0                        0.0

[49650 rows x 122 columns]>
```

In [25]:
```
normalised_home_loan.dropna(axis=0)
normalised_home_loan.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 49650 entries, 207339 to 121862
Columns: 122 entries, SK_ID_CURR to AMT_REQ_CREDIT_BUREAU_YEAR
dtypes: float64(65), int64(41), object(16)
memory usage: 46.6+ MB
```

In [26]:
```
normalised_home_loan.isnull().sum()
```

Out[26]:
```
SK_ID_CURR                    0
TARGET                        0
NAME_CONTRACT_TYPE            0
```

```
CODE_GENDER                      0
FLAG_OWN_CAR                     0
                              ...
AMT_REQ_CREDIT_BUREAU_DAY     7648
AMT_REQ_CREDIT_BUREAU_WEEK    7648
AMT_REQ_CREDIT_BUREAU_MON     7648
AMT_REQ_CREDIT_BUREAU_QRT     7648
AMT_REQ_CREDIT_BUREAU_YEAR    7648
Length: 122, dtype: int64
```

In [27]:

```
print(pd.unique(normalised_home_loan.AMT_REQ_CREDIT_BUREAU_DAY))
print(pd.unique(normalised_home_loan.AMT_REQ_CREDIT_BUREAU_WEEK))
print(pd.unique(normalised_home_loan.AMT_REQ_CREDIT_BUREAU_MON))
print(pd.unique(normalised_home_loan.AMT_REQ_CREDIT_BUREAU_QRT))
print(pd.unique(normalised_home_loan.AMT_REQ_CREDIT_BUREAU_YEAR))
```

```
[ 0. nan  1.  2.  4.  3.  9.]
[ 0. nan  1.  2.  4.  3.  5.  6.]
[ 0. nan  1.  3.  5.  9.  2.  6.  8.  4. 11. 12.  7. 13. 10. 17. 15. 14.
 16. 18. 27.]
[ 0. nan  2.  3.  1.  4.  5.  6. 19.  7.]
[ 3.  0. nan  1.  5.  4.  2.  6.  7.  8.  9. 10. 14. 13. 12. 11. 22. 16.
 23. 17.]
```

In [28]:

```
normalised_home_loan.dropna(axis=0)
```

Out[28]:

	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR
279124	423360	1	Cash loans	M	Y
216116	350411	1	Cash loans	M	Y
133687	255050	1	Cash loans	M	Y
4159	104863	1	Cash loans	M	Y
208602	341779	1	Cash loans	F	Y
...	...	...	...	...	...
108677	226053	0	Cash loans	M	Y
258603	399273	0	Revolving loans	M	Y
51880	160079	0	Cash loans	M	Y
282820	427561	0	Cash loans	F	Y
207101	340051	0	Revolving loans	F	Y

1230 rows × 122 columns

In [29]:

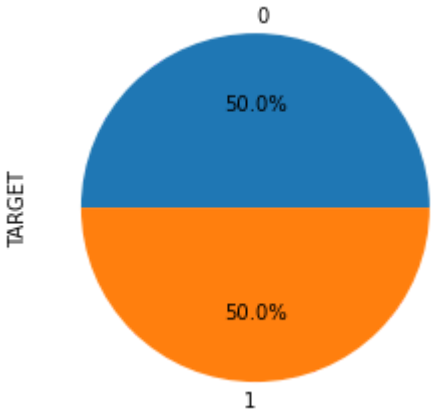
```
print(normalised_home_loan.info())
print(normalised_home_loan.isnull().sum())
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 49650 entries, 207339 to 121862
Columns: 122 entries, SK_ID_CURR to AMT_REQ_CREDIT_BUREAU_YEAR
dtypes: float64(65), int64(41), object(16)
memory usage: 46.6+ MB
None
SK_ID_CURR                      0
TARGET                          0
NAME_CONTRACT_TYPE              0
CODE_GENDER                     0
FLAG_OWN_CAR                    0
```

```
                              ...
AMT_REQ_CREDIT_BUREAU_DAY     7648
AMT_REQ_CREDIT_BUREAU_WEEK    7648
AMT_REQ_CREDIT_BUREAU_MON     7648
AMT_REQ_CREDIT_BUREAU_QRT     7648
AMT_REQ_CREDIT_BUREAU_YEAR    7648
Length: 122, dtype: int64
```

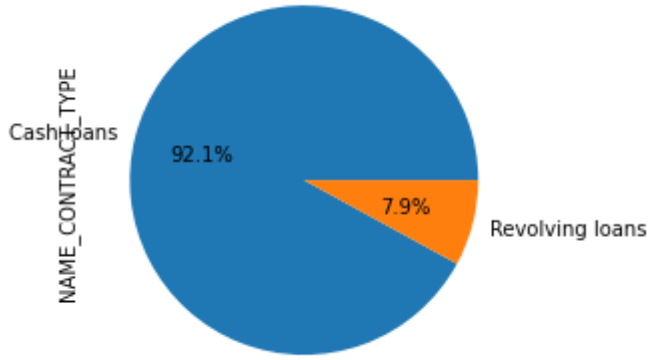
In [30]:
```
normalised_home_loan.TARGET.value_counts().plot(kind='pie',autopct="%1.1
```

Out[30]: <AxesSubplot:ylabel='TARGET'>



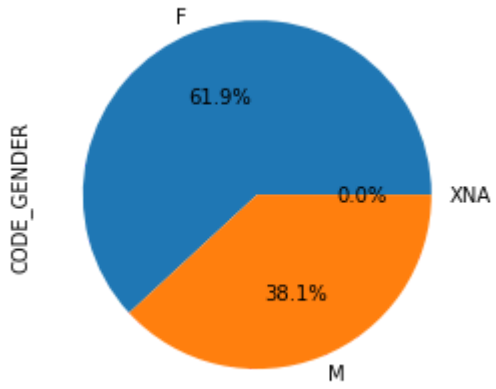
In [31]:
```
normalised_home_loan.NAME_CONTRACT_TYPE.value_counts().plot(kind='pie',a
```

Out[31]: <AxesSubplot:ylabel='NAME_CONTRACT_TYPE'>



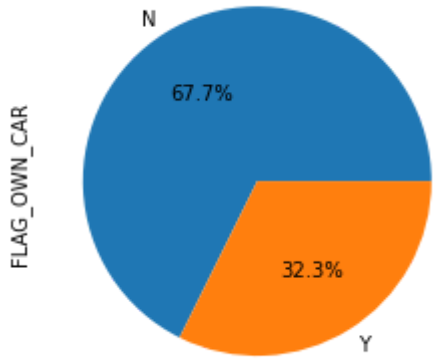
In [32]:
```
normalised_home_loan.CODE_GENDER.value_counts().plot(kind='pie',autopct=
```

Out[32]: <AxesSubplot:ylabel='CODE_GENDER'>



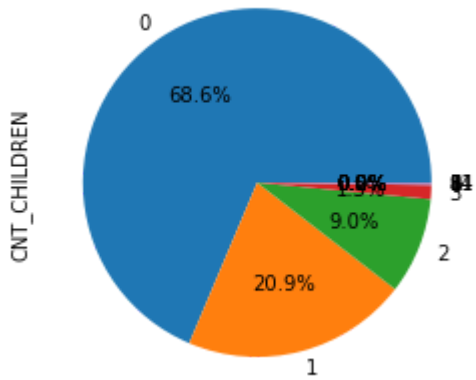
In [33]:
```
normalised_home_loan.FLAG_OWN_CAR.value_counts().plot(kind='pie',autopct
```

Out[33]: <AxesSubplot:ylabel='FLAG_OWN_CAR'>



In [34]:
```
normalised_home_loan.CNT_CHILDREN.value_counts().plot(kind='pie',autopct
```

Out[34]: <AxesSubplot:ylabel='CNT_CHILDREN'>



In [35]:
```
!pip install chart_studio
```

```
Collecting chart_studio
  Downloading chart_studio-1.1.0-py3-none-any.whl (64 kB)
Requirement already satisfied: plotly in c:\users\lenovo\downloads\anacon
da\lib\site-packages (from chart_studio) (5.6.0)
Requirement already satisfied: six in c:\users\lenovo\downloads\anaconda
```

```
\lib\site-packages (from chart_studio) (1.15.0)
Requirement already satisfied: requests in c:\users\lenovo\downloads\anac
onda\lib\site-packages (from chart_studio) (2.25.1)
Collecting retrying>=1.3.3
  Downloading retrying-1.3.3.tar.gz (10 kB)
Requirement already satisfied: tenacity>=6.2.0 in c:\users\lenovo\downloa
ds\anaconda\lib\site-packages (from plotly->chart_studio) (8.0.1)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\lenovo\down
loads\anaconda\lib\site-packages (from requests->chart_studio) (2020.12.
5)
Requirement already satisfied: idna<3,>=2.5 in c:\users\lenovo\downloads
\anaconda\lib\site-packages (from requests->chart_studio) (2.10)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\lenovo\d
ownloads\anaconda\lib\site-packages (from requests->chart_studio) (1.26.
4)
Requirement already satisfied: chardet<5,>=3.0.2 in c:\users\lenovo\downl
oads\anaconda\lib\site-packages (from requests->chart_studio) (4.0.0)
Building wheels for collected packages: retrying
  Building wheel for retrying (setup.py): started
  Building wheel for retrying (setup.py): finished with status 'done'
  Created wheel for retrying: filename=retrying-1.3.3-py3-none-any.whl si
ze=11429 sha256=5f1c444276e7eb7591bffea725580884d94b9a9f073162b24bc600e55
2577f04
  Stored in directory: c:\users\lenovo\appdata\local\pip\cache\wheels\c4
\a7\48\0a434133f6d56e878ca511c0e6c38326907c0792f67b476e56
Successfully built retrying
Installing collected packages: retrying, chart-studio
Successfully installed chart-studio-1.1.0 retrying-1.3.3
```

In [36]:

```
cf.set_config_file(theme='polar')

normalised_home_loan[normalised_home_loan['AMT_INCOME_TOTAL'] < 2000000]
   xTitle = 'Total Income', yTitle ='Count of applicants',
              title='Distribution of AMT_INCOME_TOTAL')
```

In [37]:
```
(normalised_home_loan[normalised_home_loan['AMT_INCOME_TOTAL']>1000000][
```

Out[37]:
```
0    64.864865
1    35.135135
Name: TARGET, dtype: float64
```

In [38]:
```
print((normalised_home_loan[normalised_home_loan['CNT_CHILDREN']>2]['TAR
print((normalised_home_loan[normalised_home_loan['CNT_CHILDREN']>5]['TAR
```

```
1    57.047872
0    42.952128
Name: TARGET, dtype: float64
1    81.818182
0    18.181818
Name: TARGET, dtype: float64
```

In [39]:
```
print((normalised_home_loan[normalised_home_loan['FLAG_OWN_CAR']=='N']['
print((normalised_home_loan[normalised_home_loan['FLAG_OWN_CAR']=='Y']['
```

```
1    51.350064
0    48.649936
Name: TARGET, dtype: float64
0    52.823962
1    47.176038
Name: TARGET, dtype: float64
```

In [40]:
```
print((normalised_home_loan[normalised_home_loan['CODE_GENDER']=='M']['T
print((normalised_home_loan[normalised_home_loan['CODE_GENDER']=='F']['T
```

```
1    56.280372
0    43.719628
Name: TARGET, dtype: float64
0    53.867691
1    46.132309
Name: TARGET, dtype: float64
```

In [41]:
```
print((normalised_home_loan[normalised_home_loan['NAME_CONTRACT_TYPE']==
print((normalised_home_loan[normalised_home_loan['NAME_CONTRACT_TYPE']==
```

```
1    50.802923
0    49.197077
Name: TARGET, dtype: float64
0    59.309995
1    40.690005
Name: TARGET, dtype: float64
```

In [42]:
```
normalised_home_loan=normalised_home_loan.sample(frac=1,random_state=5)
```

In [43]:
```
from sklearn.preprocessing import OrdinalEncoder

ordenc=OrdinalEncoder()
normalised_home_loan['NAME_CONTRACT_TYPE_CODE']=ordenc.fit_transform(nor
print(normalised_home_loan[['NAME_CONTRACT_TYPE','NAME_CONTRACT_TYPE_COD
print(normalised_home_loan['NAME_CONTRACT_TYPE_CODE'].value_counts())
```

```
       NAME_CONTRACT_TYPE  NAME_CONTRACT_TYPE_CODE
```

```
302218         Cash loans                      0.0
167526         Cash loans                      0.0
159305         Cash loans                      0.0
275427         Cash loans                      0.0
8837           Cash loans                      0.0
192094         Cash loans                      0.0
235115    Revolving loans                      1.0
79051          Cash loans                      0.0
123267    Revolving loans                      1.0
5517           Cash loans                      0.0
128624         Cash loans                      0.0
187583         Cash loans                      0.0
143193         Cash loans                      0.0
288269         Cash loans                      0.0
44320          Cash loans                      0.0
256898         Cash loans                      0.0
118237         Cash loans                      0.0
5980      Revolving loans                      1.0
96475          Cash loans                      0.0
249976         Cash loans                      0.0
0.0    45708
1.0     3942
Name: NAME_CONTRACT_TYPE_CODE, dtype: int64
```

In [44]:

```
normalised_home_loan['CODE_GENDER_CODE']=ordenc.fit_transform(normalised
print(normalised_home_loan[['CODE_GENDER','CODE_GENDER_CODE']].head(20))
print(normalised_home_loan['CODE_GENDER_CODE'].value_counts())
```

```
       CODE_GENDER  CODE_GENDER_CODE
302218           M               1.0
167526           F               0.0
159305           M               1.0
275427           F               0.0
8837             M               1.0
192094           M               1.0
235115           F               0.0
79051            F               0.0
123267           M               1.0
5517             F               0.0
128624           M               1.0
187583           F               0.0
143193           M               1.0
288269           F               0.0
44320            F               0.0
256898           F               0.0
118237           F               0.0
5980             M               1.0
96475            F               0.0
249976           F               0.0
0.0    30716
1.0    18932
2.0        2
Name: CODE_GENDER_CODE, dtype: int64
```

In [45]:

```
normalised_home_loan.loc[normalised_home_loan['CODE_GENDER_CODE']==2]
```

Out[45]:

	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR
83382	196708	0	Revolving loans	XNA	N
189640	319880	0	Revolving loans	XNA	Y

2 rows × 124 columns

In [46]:
```
normalised_home_loan['FLAG_OWN_CAR_CODE']=ordenc.fit_transform(normalise
print(normalised_home_loan[['FLAG_OWN_CAR','FLAG_OWN_CAR_CODE']].head(20
print(normalised_home_loan['FLAG_OWN_CAR_CODE'].value_counts())
```

```
       FLAG_OWN_CAR  FLAG_OWN_CAR_CODE
302218            N                0.0
167526            N                0.0
159305            N                0.0
275427            N                0.0
8837              N                0.0
192094            N                0.0
235115            N                0.0
79051             N                0.0
123267            N                0.0
5517              N                0.0
128624            N                0.0
187583            N                0.0
143193            N                0.0
288269            Y                1.0
44320             Y                1.0
256898            N                0.0
118237            N                0.0
5980              Y                1.0
96475             N                0.0
249976            N                0.0
0.0    33591
1.0    16059
Name: FLAG_OWN_CAR_CODE, dtype: int64
```

In [47]:
```
normalised_home_loan['CNT_CHILDREN_CODE']=ordenc.fit_transform(normalise
print(normalised_home_loan[['CNT_CHILDREN_CODE','CNT_CHILDREN']].head(20
print(normalised_home_loan['CNT_CHILDREN_CODE'].value_counts())
```

```
        CNT_CHILDREN_CODE  CNT_CHILDREN
302218                0.0             0
167526                0.0             0
159305                2.0             2
275427                0.0             0
8837                  0.0             0
192094                0.0             0
235115                0.0             0
79051                 0.0             0
123267                1.0             1
5517                  0.0             0
128624                0.0             0
187583                1.0             1
143193                0.0             0
288269                0.0             0
44320                 0.0             0
256898                0.0             0
118237                2.0             2
5980                  0.0             0
96475                 0.0             0
249976                0.0             0
0.0     34073
1.0     10381
2.0      4444
3.0       642
4.0        89
5.0        10
6.0         6
8.0         2
7.0         1
9.0         1
10.0        1
Name: CNT_CHILDREN_CODE, dtype: int64
```

In [48]:
```
normalised_home_loan=normalised_home_loan.sample(frac=1,random_state=45)
```

In [49]:
```
normalised_home_loan['TARGET'].value_counts()
```

Out[49]:
```
0    24825
1    24825
Name: TARGET, dtype: int64
```

In [50]:
```
y=normalised_home_loan.TARGET
```

In [51]:
```
y=y.sample(frac=1,random_state=45)
```

In [52]:
```
normalised_home_loan_features=['SK_ID_CURR','NAME_CONTRACT_TYPE_CODE','CI
```

In [53]:
```
from sklearn.model_selection import train_test_split
```

In [54]:
```
X=normalised_home_loan[normalised_home_loan_features]
```

In [55]:
```
X=X.sample(frac=1,random_state=45)
```

In [56]:
```
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
```

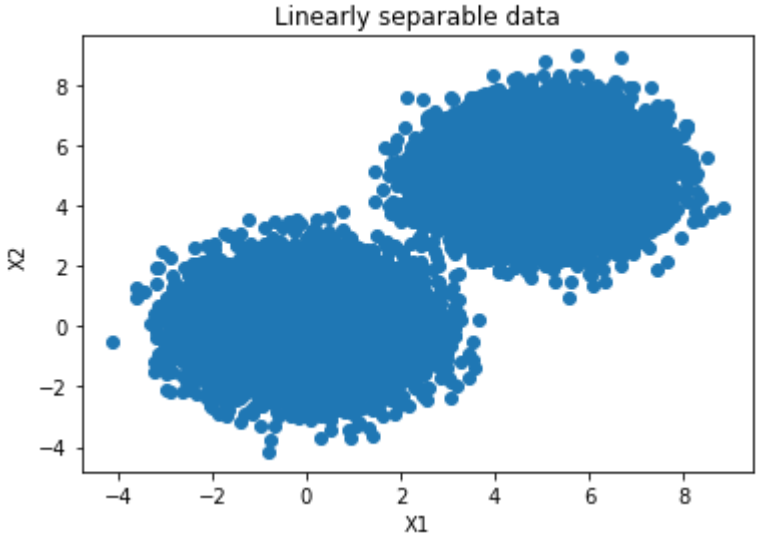
In [57]:
```
inputs, targets = make_blobs(n_samples = num_samples_total, centers = ce

X_train,X_test,y_train,y_test=train_test_split(inputs,targets,test_size=
```

In [58]:
```
print(X_train.shape, X_test.shape, y_train.shape, y_test.shape)
```

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

In [59]:
```
plt.pyplot.scatter(X_train[:,0], X_train[:,1])
plt.pyplot.title('Linearly separable data')
plt.pyplot.xlabel('X1')
plt.pyplot.ylabel('X2')
plt.pyplot.show()
```



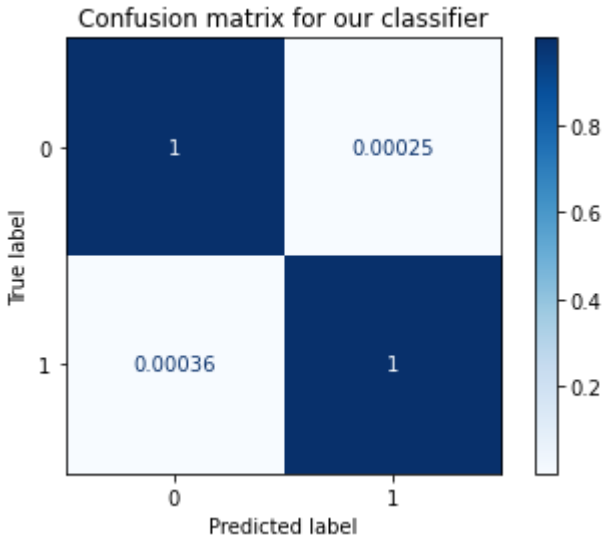
In [60]:
```
from sklearn import svm
from sklearn.metrics import plot_confusion_matrix
```

In [61]:
```
clf=svm.SVC(kernel='linear')
```

In [62]:
```
clf=clf.fit(X_train,y_train)
```

In [63]:
```
predictions = clf.predict(X_test)
```

In [64]:
```
matrix = plot_confusion_matrix(clf, X_test, y_test,
                               cmap=plt.cm.Blues,
                               normalize='true')
plt.pyplot.title('Confusion matrix for our classifier')
plt.pyplot.show(matrix)
plt.pyplot.show()
```



In [65]:
```
from sklearn.metrics import precision_score, recall_score,f1_score
```

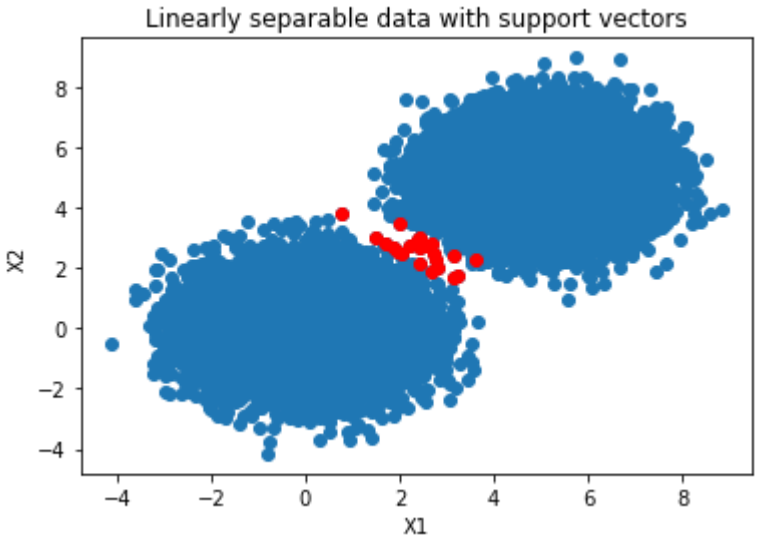
In [66]:
```
```

```
print(precision_score(y_test, predictions))
print(recall_score(y_test, predictions))
print(f1_score(y_test,predictions,average=None))
```

```
0.9997567797640764
0.9996352140077821
[0.99969368 0.99969599]
```

In [67]:

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



In [69]:

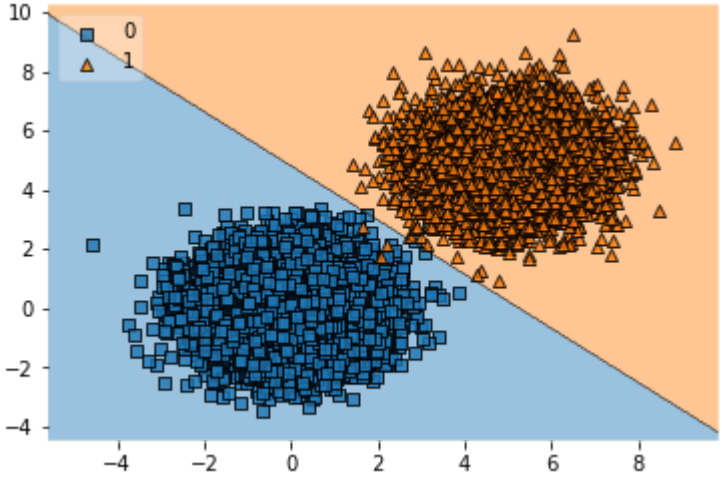
```
!pip install mlxtend
```

```
Collecting mlxtend
  Downloading mlxtend-0.19.0-py2.py3-none-any.whl (1.3 MB)
Requirement already satisfied: scikit-learn>=0.20.3 in c:\users\lenovo\do
wnloads\anaconda\lib\site-packages (from mlxtend) (1.0.2)
Requirement already satisfied: scipy>=1.2.1 in c:\users\lenovo\downloads
\anaconda\lib\site-packages (from mlxtend) (1.6.2)
Requirement already satisfied: setuptools in c:\users\lenovo\downloads\an
aconda\lib\site-packages (from mlxtend) (52.0.0.post20210125)
Requirement already satisfied: matplotlib>=3.0.0 in c:\users\lenovo\downl
oads\anaconda\lib\site-packages (from mlxtend) (3.3.4)
Requirement already satisfied: numpy>=1.16.2 in c:\users\lenovo\downloads
\anaconda\lib\site-packages (from mlxtend) (1.20.1)
Requirement already satisfied: pandas>=0.24.2 in c:\users\lenovo\download
s\anaconda\lib\site-packages (from mlxtend) (1.2.4)
Requirement already satisfied: joblib>=0.13.2 in c:\users\lenovo\download
s\anaconda\lib\site-packages (from mlxtend) (1.0.1)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.3 i
n c:\users\lenovo\downloads\anaconda\lib\site-packages (from matplotlib>=
3.0.0->mlxtend) (2.4.7)
Requirement already satisfied: cycler>=0.10 in c:\users\lenovo\downloads
\anaconda\lib\site-packages (from matplotlib>=3.0.0->mlxtend) (0.10.0)
Requirement already satisfied: pillow>=6.2.0 in c:\users\lenovo\downloads
\anaconda\lib\site-packages (from matplotlib>=3.0.0->mlxtend) (8.2.0)
Requirement already satisfied: python-dateutil>=2.1 in c:\users\lenovo\do
wnloads\anaconda\lib\site-packages (from matplotlib>=3.0.0->mlxtend) (2.
8.1)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\lenovo\downl
```

```
oads\anaconda\lib\site-packages (from matplotlib>=3.0.0->mlxtend) (1.3.1)
Requirement already satisfied: six in c:\users\lenovo\downloads\anaconda
\lib\site-packages (from cycler>=0.10->matplotlib>=3.0.0->mlxtend) (1.15.
0)
Requirement already satisfied: pytz>=2017.3 in c:\users\lenovo\downloads
\anaconda\lib\site-packages (from pandas>=0.24.2->mlxtend) (2021.1)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\lenovo\do
wnloads\anaconda\lib\site-packages (from scikit-learn>=0.20.3->mlxtend)
(2.1.0)
Installing collected packages: mlxtend
Successfully installed mlxtend-0.19.0
```

In [70]:
```
from mlxtend.plotting import plot_decision_regions
```

In [71]:
```
plot_decision_regions(X_test, y_test, clf=clf, legend=2)
plt.pyplot.show()
```



In [ ]:
```

```