DA Capstone 2-Predicting Loan Defaulters

October 7, 2021

[1]:
```
# This Python 3 environment comes with many helpful analytics libraries␣
↪installed
# It is defined by the kaggle/python docker image: https://github.com/kaggle/
↪docker-python
# For example, here's several helpful packages to load in

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the "../input/" directory.
# For example, running this (by clicking run or pressing Shift+Enter) will list␣
↪the files in the input directory

import matplotlib.pyplot as plt
import seaborn as sns
import xlrd
# Any results you write to the current directory are saved as output.
```

[2]:
```
# Perform preliminary data inspection and report the findings as the structure␣
↪of the data, missing values, duplicates, etc.
train = pd.read_excel (r'data.xlsx')
```

[3]:
```
train.head()
```

[3]:
```
   UniqueID  disbursed_amount  asset_cost    ltv  branch_id  supplier_id  \
0    420825             50578       58400  89.55         67        22807
1    417566             53278       61360  89.63         67        22807
2    539055             52378       60300  88.39         67        22807
3    529269             46349       61500  76.42         67        22807
4    563215             43594       78256  57.50         67        22744

   manufacturer_id  Current_pincode_ID Date.of.Birth Employment.Type  …  \
0               45                1441    1984-01-01        Salaried  …
1               45                1497    1985-08-24   Self employed  …
2               45                1495    1977-12-09   Self employed  …
3               45                1502    1988-06-01        Salaried  …
4               86                1499    1994-07-14   Self employed  …
```

```
   SEC.SANCTIONED.AMOUNT  SEC.DISBURSED.AMOUNT  PRIMARY.INSTAL.AMT  \
0                      0                     0                   0
1                      0                     0                   0
2                      0                     0                   0
3                      0                     0                   0
4                      0                     0                   0

   SEC.INSTAL.AMT  NEW.ACCTS.IN.LAST.SIX.MONTHS  \
0               0                             0
1               0                             0
2               0                             0
3               0                             0
4               0                             0

   DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS  AVERAGE.ACCT.AGE  \
0                                    0         0yrs 0mon
1                                    0         0yrs 0mon
2                                    0         0yrs 0mon
3                                    0         0yrs 0mon
4                                    0         0yrs 0mon

   CREDIT.HISTORY.LENGTH  NO.OF_INQUIRIES  loan_default
0              0yrs 0mon                0             0
1              0yrs 0mon                0             0
2              0yrs 0mon                1             1
3              0yrs 0mon                0             0
4              0yrs 0mon                0             0

[5 rows x 41 columns]
```

[4]:
```python
# getting the shapes of the datasets
print("Shape of Train :", train.shape)
```

```
Shape of Train : (233154, 41)
```

[5]:
```python
# get the info of train
train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 233154 entries, 0 to 233153
Data columns (total 41 columns):
 #   Column                               Non-Null Count   Dtype
---  ------                               --------------   -----
 0   UniqueID                             233154 non-null  int64
 1   disbursed_amount                     233154 non-null  int64
 2   asset_cost                           233154 non-null  int64
 3   ltv                                  233154 non-null  float64
```

```
 4   branch_id                            233154 non-null  int64
 5   supplier_id                          233154 non-null  int64
 6   manufacturer_id                      233154 non-null  int64
 7   Current_pincode_ID                   233154 non-null  int64
 8   Date.of.Birth                        233154 non-null  datetime64[ns]
 9   Employment.Type                      225493 non-null  object
 10  DisbursalDate                        233154 non-null  datetime64[ns]
 11  State_ID                             233154 non-null  int64
 12  Employee_code_ID                     233154 non-null  int64
 13  MobileNo_Avl_Flag                    233154 non-null  int64
 14  Aadhar_flag                          233154 non-null  int64
 15  PAN_flag                             233154 non-null  int64
 16  VoterID_flag                         233154 non-null  int64
 17  Driving_flag                         233154 non-null  int64
 18  Passport_flag                        233154 non-null  int64
 19  PERFORM_CNS.SCORE                    233154 non-null  int64
 20  PERFORM_CNS.SCORE.DESCRIPTION        233154 non-null  object
 21  PRI.NO.OF.ACCTS                      233154 non-null  int64
 22  PRI.ACTIVE.ACCTS                     233154 non-null  int64
 23  PRI.OVERDUE.ACCTS                    233154 non-null  int64
 24  PRI.CURRENT.BALANCE                  233154 non-null  int64
 25  PRI.SANCTIONED.AMOUNT                233154 non-null  int64
 26  PRI.DISBURSED.AMOUNT                 233154 non-null  int64
 27  SEC.NO.OF.ACCTS                      233154 non-null  int64
 28  SEC.ACTIVE.ACCTS                     233154 non-null  int64
 29  SEC.OVERDUE.ACCTS                    233154 non-null  int64
 30  SEC.CURRENT.BALANCE                  233154 non-null  int64
 31  SEC.SANCTIONED.AMOUNT                233154 non-null  int64
 32  SEC.DISBURSED.AMOUNT                 233154 non-null  int64
 33  PRIMARY.INSTAL.AMT                   233154 non-null  int64
 34  SEC.INSTAL.AMT                       233154 non-null  int64
 35  NEW.ACCTS.IN.LAST.SIX.MONTHS         233154 non-null  int64
 36  DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS  233154 non-null  int64
 37  AVERAGE.ACCT.AGE                     233154 non-null  object
 38  CREDIT.HISTORY.LENGTH                233154 non-null  object
 39  NO.OF_INQUIRIES                      233154 non-null  int64
 40  loan_default                         233154 non-null  int64
dtypes: datetime64[ns](2), float64(1), int64(34), object(4)
memory usage: 72.9+ MB
```

[6]:
```python
# getting the data types of train
train.dtypes
```

[6]:
```
UniqueID                                 int64
disbursed_amount                         int64
asset_cost                               int64
ltv                                    float64
```

```
branch_id                                          int64
supplier_id                                        int64
manufacturer_id                                    int64
Current_pincode_ID                                 int64
Date.of.Birth                             datetime64[ns]
Employment.Type                                   object
DisbursalDate                             datetime64[ns]
State_ID                                           int64
Employee_code_ID                                   int64
MobileNo_Avl_Flag                                  int64
Aadhar_flag                                        int64
PAN_flag                                           int64
VoterID_flag                                       int64
Driving_flag                                       int64
Passport_flag                                      int64
PERFORM_CNS.SCORE                                  int64
PERFORM_CNS.SCORE.DESCRIPTION                     object
PRI.NO.OF.ACCTS                                    int64
PRI.ACTIVE.ACCTS                                   int64
PRI.OVERDUE.ACCTS                                  int64
PRI.CURRENT.BALANCE                                int64
PRI.SANCTIONED.AMOUNT                              int64
PRI.DISBURSED.AMOUNT                               int64
SEC.NO.OF.ACCTS                                    int64
SEC.ACTIVE.ACCTS                                   int64
SEC.OVERDUE.ACCTS                                  int64
SEC.CURRENT.BALANCE                                int64
SEC.SANCTIONED.AMOUNT                              int64
SEC.DISBURSED.AMOUNT                               int64
PRIMARY.INSTAL.AMT                                 int64
SEC.INSTAL.AMT                                     int64
NEW.ACCTS.IN.LAST.SIX.MONTHS                       int64
DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS                int64
AVERAGE.ACCT.AGE                                  object
CREDIT.HISTORY.LENGTH                             object
NO.OF_INQUIRIES                                    int64
loan_default                                       int64
dtype: object
```

[7]:
```
# The presented data might also contain some missing values therefore,
↪exploration will also lead to devising strategies to fill in the missing
↪values while exploring the data
# checking if there exists any NULL values in the train set
train.isnull().sum()
```

[7]:
```
UniqueID                               0
disbursed_amount                       0
```

```
asset_cost                              0
ltv                                     0
branch_id                               0
supplier_id                             0
manufacturer_id                         0
Current_pincode_ID                      0
Date.of.Birth                           0
Employment.Type                      7661
DisbursalDate                           0
State_ID                                0
Employee_code_ID                        0
MobileNo_Avl_Flag                       0
Aadhar_flag                             0
PAN_flag                                0
VoterID_flag                            0
Driving_flag                            0
Passport_flag                           0
PERFORM_CNS.SCORE                       0
PERFORM_CNS.SCORE.DESCRIPTION           0
PRI.NO.OF.ACCTS                         0
PRI.ACTIVE.ACCTS                        0
PRI.OVERDUE.ACCTS                       0
PRI.CURRENT.BALANCE                     0
PRI.SANCTIONED.AMOUNT                   0
PRI.DISBURSED.AMOUNT                    0
SEC.NO.OF.ACCTS                         0
SEC.ACTIVE.ACCTS                        0
SEC.OVERDUE.ACCTS                       0
SEC.CURRENT.BALANCE                     0
SEC.SANCTIONED.AMOUNT                   0
SEC.DISBURSED.AMOUNT                    0
PRIMARY.INSTAL.AMT                      0
SEC.INSTAL.AMT                          0
NEW.ACCTS.IN.LAST.SIX.MONTHS            0
DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS     0
AVERAGE.ACCT.AGE                        0
CREDIT.HISTORY.LENGTH                   0
NO.OF_INQUIRIES                         0
loan_default                            0
dtype: int64
```

[8]:
```
# checking the values present in the Employement.Type attribute in the train␣
↪and test sets
train['Employment.Type'].value_counts()
```

[8]:
```
Self employed    127635
Salaried          97858
```

```
Name: Employment.Type, dtype: int64
```

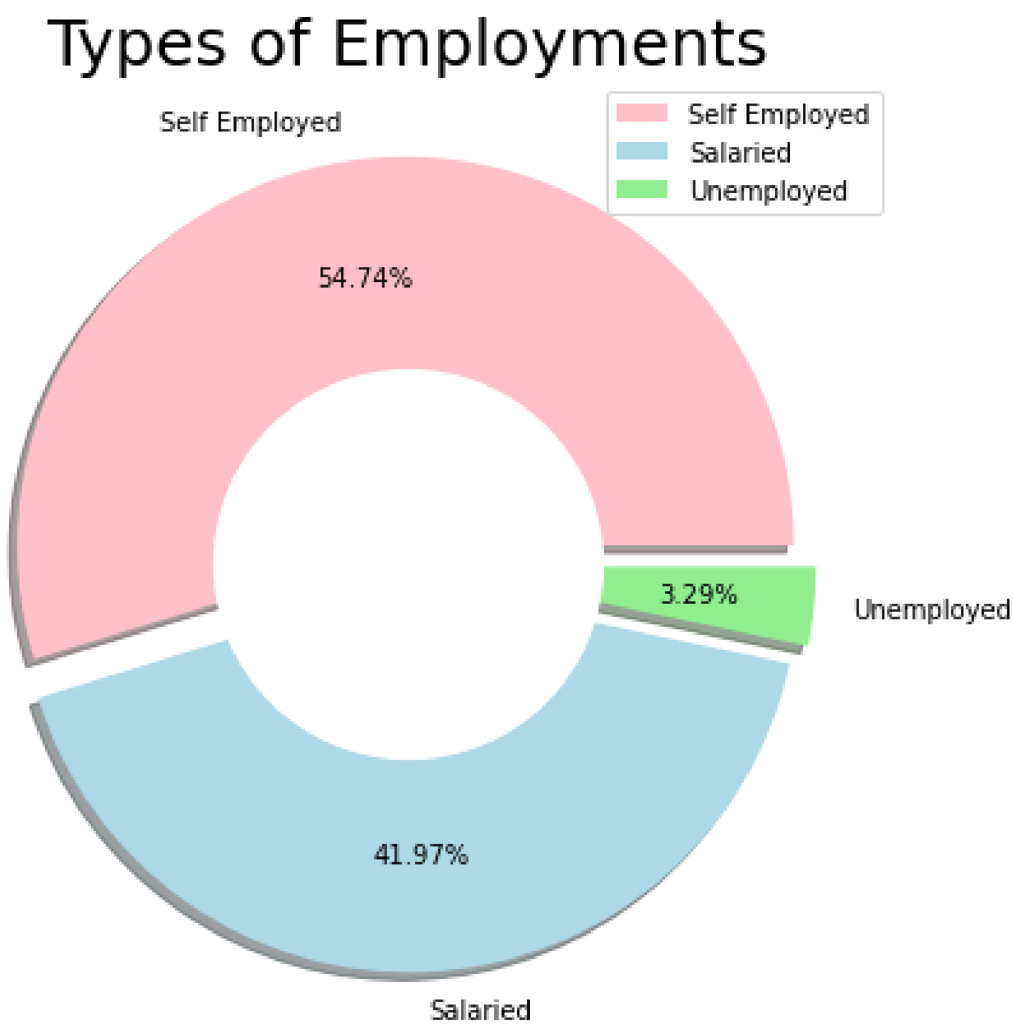
[9]:
```
# filling the missing values in the Employment.Type attribute of train and test␣
↪sets

# Employement Type has two types of Employment i.e., self employed and salaried
# but the empty values must be the people who don't  work at all that's why it␣
↪is empty
# let's fill unemployed in the place of Null values

train['Employment.Type'].fillna('Unemployed', inplace = True)

# let's check if there is any null values still left or not
print("Null values left in the train set:", train.isnull().sum().sum())
```

```
Null values left in the train set: 0
```

[10]:
```
train['Employment.Type'].value_counts()
```

[10]:
```
Self employed    127635
Salaried          97858
Unemployed         7661
Name: Employment.Type, dtype: int64
```

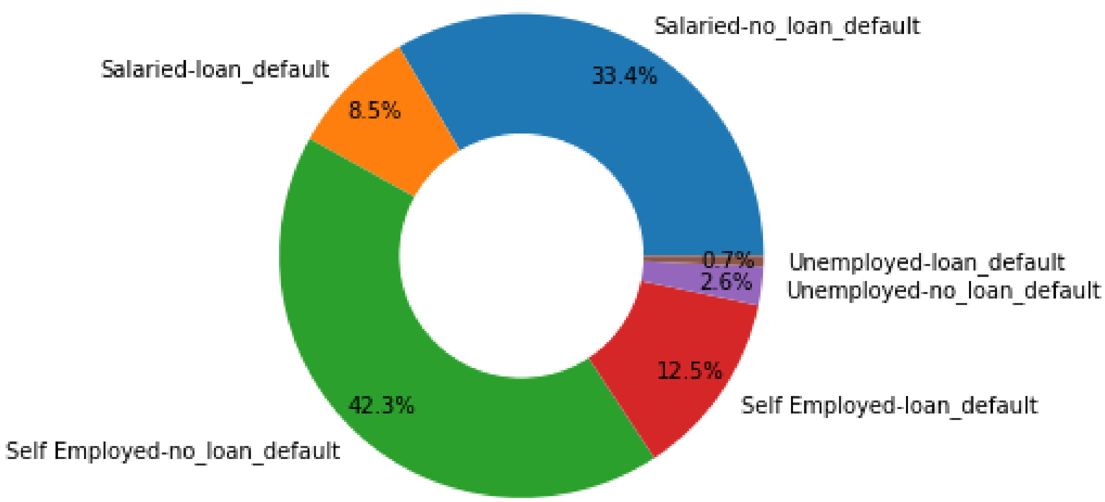
[11]:
```
# plotting a donut chart

size = [127635, 97858, 7661]
colors = ['pink', 'lightblue', 'lightgreen']
labels = "Self Employed", "Salaried", "Unemployed"
explode = [0.05, 0.05, 0.05]

circle = plt.Circle((0, 0), 0.5, color = 'white')

plt.rcParams['figure.figsize'] = (7, 7)
plt.pie(size, colors = colors, labels = labels, explode = explode, shadow =␣
↪True, pctdistance = 0.7, autopct = '%.2f%%')
plt.title('Types of Employments', fontsize = 25)
plt.axis('off')
p = plt.gcf()
p.gca().add_artist(circle)
plt.legend()
plt.show()
```



[12]:
```
employ_loandef=train.groupby(['Employment.Type', 'loan_default']).size()
emp_loan_prct=employ_loandef.apply(lambda x: 100*x/employ_loandef.sum())
emp_loan_prct
```

[12]:
```
Employment.Type  loan_default
Salaried         0               33.431981
                 1                8.539420
Self employed    0               42.280210
                 1               12.462578
Unemployed       0                2.580698
                 1                0.705113
dtype: float64
```

[13]:
```
mylabel= "Salaried-no_loan_default","Salaried-loan_default","Self␣
↪Employed-no_loan_default","Self Employed-loan_default",␣
↪"Unemployed-no_loan_default","Unemployed-loan_default"
circle = plt.Circle((0, 0), 0.5, color = 'white')
plt.rcParams['figure.figsize'] = (5, 15)
plt.pie(employ_loandef,labels=mylabel, autopct='%1.1f%%', pctdistance=0.85)
p = plt.gcf()
p.gca().add_artist(circle)
plt.show()
```



[14]:
```
# encodings for type of employments

train['Employment.Type'] = train['Employment.Type'].replace(('Self employed',␣
↪'Salaried', 'Unemployed'), (2, 1, 0))

# checking the values  of employement type
train['Employment.Type'].value_counts()
```

[14]:
```
2    127635
1     97858
0      7661
Name: Employment.Type, dtype: int64
```

[15]:
```
train.columns
```

[15]: Index(['UniqueID', 'disbursed_amount', 'asset_cost', 'ltv', 'branch_id',
       'supplier_id', 'manufacturer_id', 'Current_pincode_ID', 'Date.of.Birth',
       'Employment.Type', 'DisbursalDate', 'State_ID', 'Employee_code_ID',
       'MobileNo_Avl_Flag', 'Aadhar_flag', 'PAN_flag', 'VoterID_flag',
       'Driving_flag', 'Passport_flag', 'PERFORM_CNS.SCORE',
       'PERFORM_CNS.SCORE.DESCRIPTION', 'PRI.NO.OF.ACCTS', 'PRI.ACTIVE.ACCTS',
       'PRI.OVERDUE.ACCTS', 'PRI.CURRENT.BALANCE', 'PRI.SANCTIONED.AMOUNT',
       'PRI.DISBURSED.AMOUNT', 'SEC.NO.OF.ACCTS', 'SEC.ACTIVE.ACCTS',
       'SEC.OVERDUE.ACCTS', 'SEC.CURRENT.BALANCE', 'SEC.SANCTIONED.AMOUNT',
       'SEC.DISBURSED.AMOUNT', 'PRIMARY.INSTAL.AMT', 'SEC.INSTAL.AMT',
       'NEW.ACCTS.IN.LAST.SIX.MONTHS', 'DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS',
       'AVERAGE.ACCT.AGE', 'CREDIT.HISTORY.LENGTH', 'NO.OF_INQUIRIES',
       'loan_default'],
      dtype='object')

[16]:
```python
#let's check the unique values of ids in different branchs

print("Total no. of Unique Ids :", train['UniqueID'].nunique())
print("Total no. of Unique Branches :", train['branch_id'].nunique())
print("Total no. of Unique Suppliers :", train['supplier_id'].nunique())
print("Total no. of Unique Manufactures :", train['manufacturer_id'].nunique())
print("Total no. of Unique Current pincode Ids :", train['Current_pincode_ID'].
 ↪nunique())
print("Total no. of Unique State IDs :",train['State_ID'].nunique())
print("Total no. of Unique Employee code IDs :", train['Employee_code_ID'].
 ↪nunique())
```

```
Total no. of Unique Ids : 233154
Total no. of Unique Branches : 82
Total no. of Unique Suppliers : 2953
Total no. of Unique Manufactures : 11
Total no. of Unique Current pincode Ids : 6698
Total no. of Unique State IDs : 22
Total no. of Unique Employee code IDs : 3270
```

[17]:
```python
# Provide the statistical description of the quantitative data variables
# let's decsribe the train set
train.describe()
```

[17]:

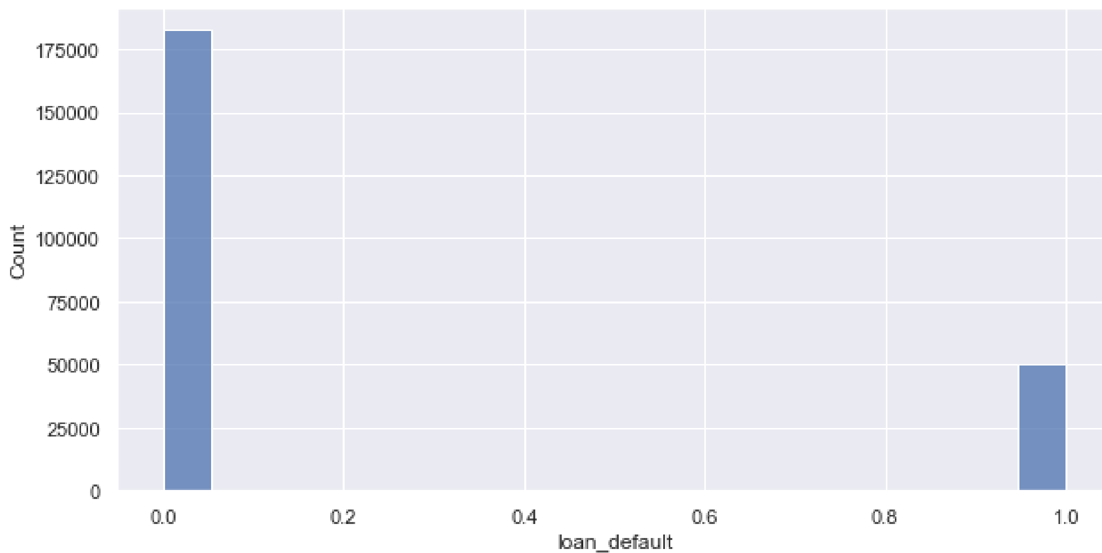
	UniqueID	disbursed_amount	asset_cost	ltv \
count	233154.000000	233154.000000	2.331540e+05	233154.000000
mean	535917.573376	54356.993528	7.586507e+04	74.746530
std	68315.693711	12971.314171	1.894478e+04	11.456636
min	417428.000000	13320.000000	3.700000e+04	10.030000
25%	476786.250000	47145.000000	6.571700e+04	68.880000
50%	535978.500000	53803.000000	7.094600e+04	76.800000
75%	595039.750000	60413.000000	7.920175e+04	83.670000

```
max    671084.000000     990572.000000  1.628992e+06      95.000000

           branch_id    supplier_id  manufacturer_id  Current_pincode_ID  \
count  233154.000000  233154.000000    233154.000000       233154.000000
mean       72.936094   19638.635035        69.028054         3396.880247
std        69.834995    3491.949566        22.141304         2238.147502
min         1.000000   10524.000000        45.000000            1.000000
25%        14.000000   16535.000000        48.000000         1511.000000
50%        61.000000   20333.000000        86.000000         2970.000000
75%       130.000000   23000.000000        86.000000         5677.000000
max       261.000000   24803.000000       156.000000         7345.000000

       Employment.Type       State_ID  …  SEC.OVERDUE.ACCTS  \
count    233154.000000  233154.000000  …      233154.000000
mean          1.514570       7.262243  …           0.007244
std           0.561699       4.482230  …           0.111079
min           0.000000       1.000000  …           0.000000
25%           1.000000       4.000000  …           0.000000
50%           2.000000       6.000000  …           0.000000
75%           2.000000      10.000000  …           0.000000
max           2.000000      22.000000  …           8.000000

       SEC.CURRENT.BALANCE  SEC.SANCTIONED.AMOUNT  SEC.DISBURSED.AMOUNT  \
count         2.331540e+05           2.331540e+05          2.331540e+05
mean          5.427793e+03           7.295923e+03          7.179998e+03
std           1.702370e+05           1.831560e+05          1.825925e+05
min          -5.746470e+05           0.000000e+00          0.000000e+00
25%           0.000000e+00           0.000000e+00          0.000000e+00
50%           0.000000e+00           0.000000e+00          0.000000e+00
75%           0.000000e+00           0.000000e+00          0.000000e+00
max           3.603285e+07           3.000000e+07          3.000000e+07

       PRIMARY.INSTAL.AMT  SEC.INSTAL.AMT  NEW.ACCTS.IN.LAST.SIX.MONTHS  \
count        2.331540e+05    2.331540e+05                 233154.000000
mean         1.310548e+04    3.232684e+02                      0.381833
std          1.513679e+05    1.555369e+04                      0.955107
min          0.000000e+00    0.000000e+00                      0.000000
25%          0.000000e+00    0.000000e+00                      0.000000
50%          0.000000e+00    0.000000e+00                      0.000000
75%          1.999000e+03    0.000000e+00                      0.000000
max          2.564281e+07    4.170901e+06                     35.000000

       DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS  NO.OF_INQUIRIES   loan_default
count                        233154.000000    233154.000000  233154.000000
mean                              0.097481         0.206615       0.217071
std                               0.384439         0.706498       0.412252
min                               0.000000         0.000000       0.000000
```

```
25%                                   0.000000         0.000000        0.000000
50%                                   0.000000         0.000000        0.000000
75%                                   0.000000         0.000000        0.000000
max                                  20.000000        36.000000        1.000000

[8 rows x 36 columns]
```

[21]:
```
# Explain how is the target variable distributed overall
# Histogram for loan defaulter
#train.plot.bar(y='loan_default', rot=0)
#plt.bar('loan_default')
sns.set(rc={'figure.figsize':(10,5)})
sns.histplot(x='loan_default',data=train)
```

[21]:
```
<AxesSubplot:xlabel='loan_default', ylabel='Count'>
```

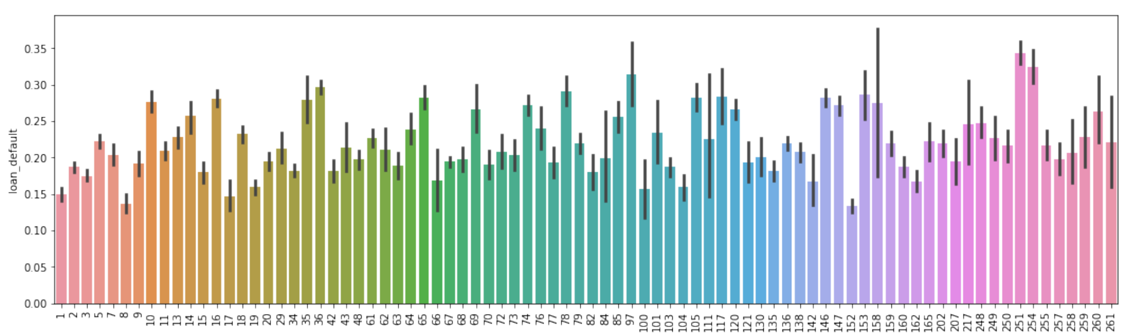


[19]:
```
fig, ax = plt.subplots(figsize=(18,5))
sns.barplot(y='loan_default',x='branch_id',data=train,ax=ax)
plt.xticks(rotation=90)
#plt.bar('loan_default','branch_id',data=train)
```

[19]:
```
(array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 16,
        17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33,
        34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50,
        51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67,
        68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81]),
 [Text(0, 0, '1'),
  Text(1, 0, '2'),
```

```
Text(2, 0, '3'),
Text(3, 0, '5'),
Text(4, 0, '7'),
Text(5, 0, '8'),
Text(6, 0, '9'),
Text(7, 0, '10'),
Text(8, 0, '11'),
Text(9, 0, '13'),
Text(10, 0, '14'),
Text(11, 0, '15'),
Text(12, 0, '16'),
Text(13, 0, '17'),
Text(14, 0, '18'),
Text(15, 0, '19'),
Text(16, 0, '20'),
Text(17, 0, '29'),
Text(18, 0, '34'),
Text(19, 0, '35'),
Text(20, 0, '36'),
Text(21, 0, '42'),
Text(22, 0, '43'),
Text(23, 0, '48'),
Text(24, 0, '61'),
Text(25, 0, '62'),
Text(26, 0, '63'),
Text(27, 0, '64'),
Text(28, 0, '65'),
Text(29, 0, '66'),
Text(30, 0, '67'),
Text(31, 0, '68'),
Text(32, 0, '69'),
Text(33, 0, '70'),
Text(34, 0, '72'),
Text(35, 0, '73'),
Text(36, 0, '74'),
Text(37, 0, '76'),
Text(38, 0, '77'),
Text(39, 0, '78'),
Text(40, 0, '79'),
Text(41, 0, '82'),
Text(42, 0, '84'),
Text(43, 0, '85'),
Text(44, 0, '97'),
Text(45, 0, '100'),
Text(46, 0, '101'),
Text(47, 0, '103'),
Text(48, 0, '104'),
```

```
Text(49, 0, '105'),
Text(50, 0, '111'),
Text(51, 0, '117'),
Text(52, 0, '120'),
Text(53, 0, '121'),
Text(54, 0, '130'),
Text(55, 0, '135'),
Text(56, 0, '136'),
Text(57, 0, '138'),
Text(58, 0, '142'),
Text(59, 0, '146'),
Text(60, 0, '147'),
Text(61, 0, '152'),
Text(62, 0, '153'),
Text(63, 0, '158'),
Text(64, 0, '159'),
Text(65, 0, '160'),
Text(66, 0, '162'),
Text(67, 0, '165'),
Text(68, 0, '202'),
Text(69, 0, '207'),
Text(70, 0, '217'),
Text(71, 0, '248'),
Text(72, 0, '249'),
Text(73, 0, '250'),
Text(74, 0, '251'),
Text(75, 0, '254'),
Text(76, 0, '255'),
Text(77, 0, '257'),
Text(78, 0, '258'),
Text(79, 0, '259'),
Text(80, 0, '260'),
Text(81, 0, '261')])
```



[23]:
```
train.corr()
```

[23]:
```
                                     UniqueID  disbursed_amount  asset_cost  \
UniqueID                             1.000000          0.145575    0.144671
disbursed_amount                     0.145575          1.000000    0.752668
asset_cost                           0.144671          0.752668    1.000000
ltv                                 -0.030537          0.376747   -0.301188
branch_id                           -0.004509          0.018328    0.018586
supplier_id                          0.025747          0.078596    0.150002
manufacturer_id                     -0.021514          0.095203   -0.005256
Current_pincode_ID                   0.030500          0.068540    0.275878
Employment.Type                      0.013869         -0.041106    0.015449
State_ID                            -0.057734         -0.023956   -0.026348
Employee_code_ID                     0.075983          0.010120    0.008711
MobileNo_Avl_Flag                         NaN               NaN         NaN
Aadhar_flag                         -0.001372         -0.015589   -0.093716
PAN_flag                            -0.014291          0.014490    0.048075
VoterID_flag                         0.008201          0.011665    0.083899
Driving_flag                        -0.026072          0.001381    0.020484
Passport_flag                       -0.001382          0.008291    0.005542
PERFORM_CNS.SCORE                   -0.014016          0.011409   -0.047398
PRI.NO.OF.ACCTS                     -0.013584          0.034461   -0.026798
PRI.ACTIVE.ACCTS                    -0.020862          0.039542   -0.021692
PRI.OVERDUE.ACCTS                   -0.007915          0.019422   -0.014782
PRI.CURRENT.BALANCE                 -0.007259          0.015713   -0.004356
PRI.SANCTIONED.AMOUNT               -0.002910          0.006294   -0.001180
PRI.DISBURSED.AMOUNT                -0.002563          0.006412   -0.001221
SEC.NO.OF.ACCTS                     -0.044180         -0.014916   -0.022600
SEC.ACTIVE.ACCTS                    -0.039888         -0.014275   -0.020359
SEC.OVERDUE.ACCTS                   -0.027486         -0.007305   -0.013455
SEC.CURRENT.BALANCE                 -0.013428         -0.004227   -0.008697
SEC.SANCTIONED.AMOUNT               -0.017114         -0.005271   -0.010776
SEC.DISBURSED.AMOUNT                -0.016857         -0.005119   -0.010594
PRIMARY.INSTAL.AMT                  -0.009321          0.002350   -0.004356
SEC.INSTAL.AMT                      -0.013518         -0.005525   -0.005738
NEW.ACCTS.IN.LAST.SIX.MONTHS        -0.003842          0.035880   -0.021968
DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS -0.009186          0.020198   -0.007932
NO.OF_INQUIRIES                      0.016537          0.038513   -0.017301
loan_default                         0.033848          0.077675    0.014261

                                          ltv  branch_id  supplier_id  \
UniqueID                            -0.030537  -0.004509     0.025747
disbursed_amount                     0.376747   0.018328     0.078596
asset_cost                          -0.301188   0.018586     0.150002
ltv                                  1.000000   0.006382    -0.103732
branch_id                            0.006382   1.000000     0.225744
supplier_id                         -0.103732   0.225744     1.000000
```

manufacturer_id	0.121003	-0.049957	-0.063916
Current_pincode_ID	-0.300112	0.048192	0.187401
Employment.Type	-0.077431	-0.010865	0.080914
State_ID	0.004882	0.187782	0.105502
Employee_code_ID	0.004940	0.088764	0.095792
MobileNo_Avl_Flag	NaN	NaN	NaN
Aadhar_flag	0.106953	-0.036384	-0.081161
PAN_flag	-0.037786	0.030206	0.036804
VoterID_flag	-0.098393	0.021387	0.079355
Driving_flag	-0.024060	-0.015503	-0.015682
Passport_flag	0.004467	-0.008537	-0.008773
PERFORM_CNS.SCORE	0.084993	-0.018049	-0.077029
PRI.NO.OF.ACCTS	0.086571	-0.010384	-0.070536
PRI.ACTIVE.ACCTS	0.087929	-0.006403	-0.066930
PRI.OVERDUE.ACCTS	0.051770	-0.005546	-0.042692
PRI.CURRENT.BALANCE	0.026714	-0.004133	-0.022582
PRI.SANCTIONED.AMOUNT	0.009955	-0.003370	-0.012959
PRI.DISBURSED.AMOUNT	0.010177	-0.003412	-0.012950
SEC.NO.OF.ACCTS	0.013019	-0.003194	-0.025809
SEC.ACTIVE.ACCTS	0.010565	-0.001957	-0.019885
SEC.OVERDUE.ACCTS	0.010256	-0.007104	-0.015665
SEC.CURRENT.BALANCE	0.007195	0.001418	-0.008729
SEC.SANCTIONED.AMOUNT	0.008800	0.002404	-0.010175
SEC.DISBURSED.AMOUNT	0.008733	0.002457	-0.010031
PRIMARY.INSTAL.AMT	0.007358	0.004955	-0.011953
SEC.INSTAL.AMT	0.000481	0.004265	-0.005826
NEW.ACCTS.IN.LAST.SIX.MONTHS	0.083824	-0.008354	-0.053264
DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS	0.041921	-0.001730	-0.024622
NO.OF_INQUIRIES	0.085725	-0.001813	-0.030988
loan_default	0.098208	0.030193	0.027357

	manufacturer_id	Current_pincode_ID \
UniqueID	-0.021514	0.030500
disbursed_amount	0.095203	0.068540
asset_cost	-0.005256	0.275878
ltv	0.121003	-0.300112
branch_id	-0.049957	0.048192
supplier_id	-0.063916	0.187401
manufacturer_id	1.000000	-0.104349
Current_pincode_ID	-0.104349	1.000000
Employment.Type	-0.026248	0.150706
State_ID	0.026412	0.047903
Employee_code_ID	-0.045009	-0.042609
MobileNo_Avl_Flag	NaN	NaN
Aadhar_flag	0.041251	-0.309285
PAN_flag	-0.029272	0.041118
VoterID_flag	-0.046489	0.281934

Driving_flag	0.004777	0.039667
Passport_flag	0.016338	-0.008913
PERFORM_CNS.SCORE	0.073012	-0.161664
PRI.NO.OF.ACCTS	0.068371	-0.084687
PRI.ACTIVE.ACCTS	0.065589	-0.117882
PRI.OVERDUE.ACCTS	0.025992	-0.062177
PRI.CURRENT.BALANCE	0.031722	-0.055948
PRI.SANCTIONED.AMOUNT	0.017242	-0.025843
PRI.DISBURSED.AMOUNT	0.017080	-0.025697
SEC.NO.OF.ACCTS	0.008392	-0.051766
SEC.ACTIVE.ACCTS	0.005898	-0.049852
SEC.OVERDUE.ACCTS	0.008152	-0.036228
SEC.CURRENT.BALANCE	0.003106	-0.021048
SEC.SANCTIONED.AMOUNT	0.003206	-0.026059
SEC.DISBURSED.AMOUNT	0.003301	-0.025630
PRIMARY.INSTAL.AMT	0.014116	0.002275
SEC.INSTAL.AMT	0.003020	-0.006305
NEW.ACCTS.IN.LAST.SIX.MONTHS	0.050079	-0.100080
DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS	0.019438	-0.057419
NO.OF_INQUIRIES	0.010012	-0.058821
loan_default	-0.025039	0.028419

	Employment.Type	State_ID	…	\
UniqueID	0.013869	-0.057734	…	
disbursed_amount	-0.041106	-0.023956	…	
asset_cost	0.015449	-0.026348	…	
ltv	-0.077431	0.004882	…	
branch_id	-0.010865	0.187782	…	
supplier_id	0.080914	0.105502	…	
manufacturer_id	-0.026248	0.026412	…	
Current_pincode_ID	0.150706	0.047903	…	
Employment.Type	1.000000	-0.054752	…	
State_ID	-0.054752	1.000000	…	
Employee_code_ID	0.037087	0.117309	…	
MobileNo_Avl_Flag	NaN	NaN	…	
Aadhar_flag	-0.094274	-0.169882	…	
PAN_flag	0.021053	0.113939	…	
VoterID_flag	0.093215	0.151839	…	
Driving_flag	0.010113	0.027886	…	
Passport_flag	-0.003849	-0.000915	…	
PERFORM_CNS.SCORE	-0.009866	-0.042323	…	
PRI.NO.OF.ACCTS	0.016616	-0.008704	…	
PRI.ACTIVE.ACCTS	0.011993	-0.021061	…	
PRI.OVERDUE.ACCTS	0.031922	-0.012381	…	
PRI.CURRENT.BALANCE	0.025183	-0.006359	…	
PRI.SANCTIONED.AMOUNT	0.014675	-0.004304	…	
PRI.DISBURSED.AMOUNT	0.014857	-0.004129	…	

SEC.NO.OF.ACCTS	0.000948	0.031245	…
SEC.ACTIVE.ACCTS	0.000275	0.032142	…
SEC.OVERDUE.ACCTS	0.002742	0.015679	…
SEC.CURRENT.BALANCE	0.001027	0.013231	…
SEC.SANCTIONED.AMOUNT	0.001225	0.017842	…
SEC.DISBURSED.AMOUNT	0.001179	0.017651	…
PRIMARY.INSTAL.AMT	0.009282	0.015843	…
SEC.INSTAL.AMT	-0.004238	0.009850	…
NEW.ACCTS.IN.LAST.SIX.MONTHS	0.005380	-0.014942	…
DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS	0.014894	-0.000655	…
NO.OF_INQUIRIES	0.010080	-0.016129	…
loan_default	0.025377	0.048075	…

	SEC.OVERDUE.ACCTS	SEC.CURRENT.BALANCE \
UniqueID	-0.027486	-0.013428
disbursed_amount	-0.007305	-0.004227
asset_cost	-0.013455	-0.008697
ltv	0.010256	0.007195
branch_id	-0.007104	0.001418
supplier_id	-0.015665	-0.008729
manufacturer_id	0.008152	0.003106
Current_pincode_ID	-0.036228	-0.021048
Employment.Type	0.002742	0.001027
State_ID	0.015679	0.013231
Employee_code_ID	-0.006537	-0.003543
MobileNo_Avl_Flag	NaN	NaN
Aadhar_flag	0.007979	0.006420
PAN_flag	0.020503	0.011181
VoterID_flag	-0.004257	-0.005591
Driving_flag	0.002497	0.001667
Passport_flag	-0.000497	0.003205
PERFORM_CNS.SCORE	0.035100	0.020128
PRI.NO.OF.ACCTS	0.030364	0.020945
PRI.ACTIVE.ACCTS	0.037422	0.022018
PRI.OVERDUE.ACCTS	0.041624	0.012623
PRI.CURRENT.BALANCE	0.015225	0.019378
PRI.SANCTIONED.AMOUNT	0.007967	0.010750
PRI.DISBURSED.AMOUNT	0.007872	0.010682
SEC.NO.OF.ACCTS	0.510394	0.281865
SEC.ACTIVE.ACCTS	0.526209	0.361325
SEC.OVERDUE.ACCTS	1.000000	0.187383
SEC.CURRENT.BALANCE	0.187383	1.000000
SEC.SANCTIONED.AMOUNT	0.234422	0.929196
SEC.DISBURSED.AMOUNT	0.230982	0.929995
PRIMARY.INSTAL.AMT	0.002066	0.010393
SEC.INSTAL.AMT	0.077731	0.096351
NEW.ACCTS.IN.LAST.SIX.MONTHS	0.057046	0.046465

```
DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS             0.108810             0.044987
NO.OF_INQUIRIES                                 0.004975             0.002715
loan_default                                   -0.001371            -0.005531

                                     SEC.SANCTIONED.AMOUNT  \
UniqueID                                         -0.017114
disbursed_amount                                 -0.005271
asset_cost                                       -0.010776
ltv                                               0.008800
branch_id                                         0.002404
supplier_id                                      -0.010175
manufacturer_id                                   0.003206
Current_pincode_ID                               -0.026059
Employment.Type                                   0.001225
State_ID                                          0.017842
Employee_code_ID                                 -0.002259
MobileNo_Avl_Flag                                      NaN
Aadhar_flag                                       0.007591
PAN_flag                                          0.015279
VoterID_flag                                     -0.006619
Driving_flag                                      0.002600
Passport_flag                                     0.003539
PERFORM_CNS.SCORE                                 0.024365
PRI.NO.OF.ACCTS                                   0.024877
PRI.ACTIVE.ACCTS                                  0.026805
PRI.OVERDUE.ACCTS                                 0.017194
PRI.CURRENT.BALANCE                               0.021179
PRI.SANCTIONED.AMOUNT                             0.011859
PRI.DISBURSED.AMOUNT                              0.011771
SEC.NO.OF.ACCTS                                   0.359193
SEC.ACTIVE.ACCTS                                  0.452693
SEC.OVERDUE.ACCTS                                 0.234422
SEC.CURRENT.BALANCE                               0.929196
SEC.SANCTIONED.AMOUNT                             1.000000
SEC.DISBURSED.AMOUNT                              0.999646
PRIMARY.INSTAL.AMT                                0.011448
SEC.INSTAL.AMT                                    0.113395
NEW.ACCTS.IN.LAST.SIX.MONTHS                      0.055801
DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS               0.055198
NO.OF_INQUIRIES                                   0.002675
loan_default                                     -0.006354

                                     SEC.DISBURSED.AMOUNT  PRIMARY.INSTAL.AMT  \
UniqueID                                        -0.016857           -0.009321
disbursed_amount                                -0.005119            0.002350
asset_cost                                      -0.010594           -0.004356
ltv                                              0.008733            0.007358
```

```
branch_id                             0.002457            0.004955
supplier_id                          -0.010031           -0.011953
manufacturer_id                       0.003301            0.014116
Current_pincode_ID                   -0.025630            0.002275
Employment.Type                       0.001179            0.009282
State_ID                              0.017651            0.015843
Employee_code_ID                     -0.002224            0.013470
MobileNo_Avl_Flag                          NaN                 NaN
Aadhar_flag                           0.007540            0.014832
PAN_flag                              0.014907           -0.010851
VoterID_flag                         -0.006580           -0.018820
Driving_flag                          0.002583            0.006219
Passport_flag                         0.003506            0.012914
PERFORM_CNS.SCORE                     0.023953            0.071309
PRI.NO.OF.ACCTS                       0.024551            0.269546
PRI.ACTIVE.ACCTS                      0.026325            0.192397
PRI.OVERDUE.ACCTS                     0.016855            0.103095
PRI.CURRENT.BALANCE                   0.021152            0.152499
PRI.SANCTIONED.AMOUNT                 0.011843            0.072591
PRI.DISBURSED.AMOUNT                  0.011757            0.073058
SEC.NO.OF.ACCTS                       0.353330            0.009986
SEC.ACTIVE.ACCTS                      0.445950            0.007063
SEC.OVERDUE.ACCTS                     0.230982            0.002066
SEC.CURRENT.BALANCE                   0.929995            0.010393
SEC.SANCTIONED.AMOUNT                 0.999646            0.011448
SEC.DISBURSED.AMOUNT                  1.000000            0.011469
PRIMARY.INSTAL.AMT                    0.011469            1.000000
SEC.INSTAL.AMT                        0.112880            0.010849
NEW.ACCTS.IN.LAST.SIX.MONTHS          0.054886            0.127796
DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS   0.054374            0.090364
NO.OF_INQUIRIES                       0.002581            0.008967
loan_default                         -0.006248           -0.010616

                                    SEC.INSTAL.AMT  \
UniqueID                                 -0.013518
disbursed_amount                         -0.005525
asset_cost                               -0.005738
ltv                                       0.000481
branch_id                                 0.004265
supplier_id                              -0.005826
manufacturer_id                           0.003020
Current_pincode_ID                       -0.006305
Employment.Type                          -0.004238
State_ID                                  0.009850
Employee_code_ID                          0.001894
MobileNo_Avl_Flag                              NaN
Aadhar_flag                               0.006273
```

PAN_flag	-0.000219
VoterID_flag	-0.006019
Driving_flag	-0.000843
Passport_flag	0.000524
PERFORM_CNS.SCORE	0.015655
PRI.NO.OF.ACCTS	0.023421
PRI.ACTIVE.ACCTS	0.021210
PRI.OVERDUE.ACCTS	0.009324
PRI.CURRENT.BALANCE	0.006799
PRI.SANCTIONED.AMOUNT	0.003883
PRI.DISBURSED.AMOUNT	0.003806
SEC.NO.OF.ACCTS	0.235277
SEC.ACTIVE.ACCTS	0.148895
SEC.OVERDUE.ACCTS	0.077731
SEC.CURRENT.BALANCE	0.096351
SEC.SANCTIONED.AMOUNT	0.113395
SEC.DISBURSED.AMOUNT	0.112880
PRIMARY.INSTAL.AMT	0.010849
SEC.INSTAL.AMT	1.000000
NEW.ACCTS.IN.LAST.SIX.MONTHS	0.028160
DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS	0.018613
NO.OF_INQUIRIES	0.002736
loan_default	-0.001548

	NEW.ACCTS.IN.LAST.SIX.MONTHS \
UniqueID	-0.003842
disbursed_amount	0.035880
asset_cost	-0.021968
ltv	0.083824
branch_id	-0.008354
supplier_id	-0.053264
manufacturer_id	0.050079
Current_pincode_ID	-0.100080
Employment.Type	0.005380
State_ID	-0.014942
Employee_code_ID	0.005473
MobileNo_Avl_Flag	NaN
Aadhar_flag	0.059269
PAN_flag	0.007046
VoterID_flag	-0.059591
Driving_flag	-0.000929
Passport_flag	0.008052
PERFORM_CNS.SCORE	0.346025
PRI.NO.OF.ACCTS	0.537622
PRI.ACTIVE.ACCTS	0.702943
PRI.OVERDUE.ACCTS	0.109896
PRI.CURRENT.BALANCE	0.198566

```
PRI.SANCTIONED.AMOUNT                                  0.095074
PRI.DISBURSED.AMOUNT                                   0.094681
SEC.NO.OF.ACCTS                                        0.119791
SEC.ACTIVE.ACCTS                                       0.134577
SEC.OVERDUE.ACCTS                                      0.057046
SEC.CURRENT.BALANCE                                    0.046465
SEC.SANCTIONED.AMOUNT                                  0.055801
SEC.DISBURSED.AMOUNT                                   0.054886
PRIMARY.INSTAL.AMT                                     0.127796
SEC.INSTAL.AMT                                         0.028160
NEW.ACCTS.IN.LAST.SIX.MONTHS                           1.000000
DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS                    0.182769
NO.OF_INQUIRIES                                        0.264709
loan_default                                          -0.029400

                                     DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS  \
UniqueID                                                       -0.009186
disbursed_amount                                                0.020198
asset_cost                                                     -0.007932
ltv                                                             0.041921
branch_id                                                      -0.001730
supplier_id                                                    -0.024622
manufacturer_id                                                 0.019438
Current_pincode_ID                                             -0.057419
Employment.Type                                                 0.014894
State_ID                                                       -0.000655
Employee_code_ID                                                0.003624
MobileNo_Avl_Flag                                                    NaN
Aadhar_flag                                                     0.023823
PAN_flag                                                       -0.002562
VoterID_flag                                                   -0.027453
Driving_flag                                                    0.011311
Passport_flag                                                   0.006695
PERFORM_CNS.SCORE                                               0.164745
PRI.NO.OF.ACCTS                                                 0.314148
PRI.ACTIVE.ACCTS                                                0.382584
PRI.OVERDUE.ACCTS                                               0.471564
PRI.CURRENT.BALANCE                                             0.244243
PRI.SANCTIONED.AMOUNT                                           0.122028
PRI.DISBURSED.AMOUNT                                            0.122650
SEC.NO.OF.ACCTS                                                 0.100321
SEC.ACTIVE.ACCTS                                                0.109194
SEC.OVERDUE.ACCTS                                               0.108810
SEC.CURRENT.BALANCE                                             0.044987
SEC.SANCTIONED.AMOUNT                                           0.055198
SEC.DISBURSED.AMOUNT                                            0.054374
PRIMARY.INSTAL.AMT                                              0.090364
```

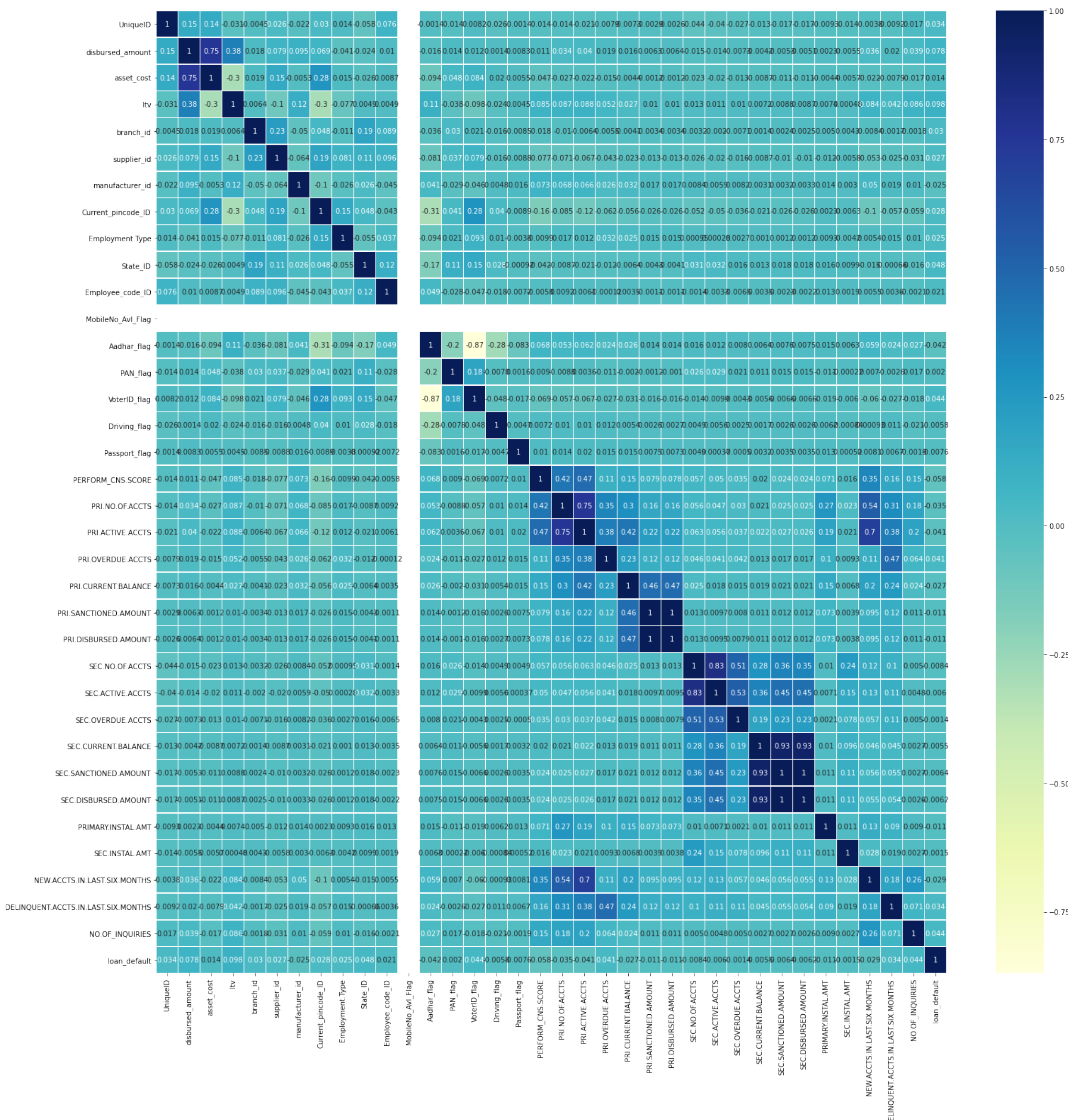
SEC.INSTAL.AMT	0.018613
NEW.ACCTS.IN.LAST.SIX.MONTHS	0.182769
DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS	1.000000
NO.OF_INQUIRIES	0.070715
loan_default	0.034462

	NO.OF_INQUIRIES	loan_default
UniqueID	0.016537	0.033848
disbursed_amount	0.038513	0.077675
asset_cost	-0.017301	0.014261
ltv	0.085725	0.098208
branch_id	-0.001813	0.030193
supplier_id	-0.030988	0.027357
manufacturer_id	0.010012	-0.025039
Current_pincode_ID	-0.058821	0.028419
Employment.Type	0.010080	0.025377
State_ID	-0.016129	0.048075
Employee_code_ID	-0.002108	0.020657
MobileNo_Avl_Flag	NaN	NaN
Aadhar_flag	0.027317	-0.041593
PAN_flag	0.016657	0.002046
VoterID_flag	-0.018285	0.043747
Driving_flag	-0.021179	-0.005821
Passport_flag	-0.001908	-0.007602
PERFORM_CNS.SCORE	0.152838	-0.057929
PRI.NO.OF.ACCTS	0.180744	-0.035456
PRI.ACTIVE.ACCTS	0.195484	-0.041451
PRI.OVERDUE.ACCTS	0.063780	0.040872
PRI.CURRENT.BALANCE	0.023861	-0.027386
PRI.SANCTIONED.AMOUNT	0.010743	-0.011304
PRI.DISBURSED.AMOUNT	0.010935	-0.011155
SEC.NO.OF.ACCTS	0.005035	-0.008385
SEC.ACTIVE.ACCTS	0.004830	-0.005993
SEC.OVERDUE.ACCTS	0.004975	-0.001371
SEC.CURRENT.BALANCE	0.002715	-0.005531
SEC.SANCTIONED.AMOUNT	0.002675	-0.006354
SEC.DISBURSED.AMOUNT	0.002581	-0.006248
PRIMARY.INSTAL.AMT	0.008967	-0.010616
SEC.INSTAL.AMT	0.002736	-0.001548
NEW.ACCTS.IN.LAST.SIX.MONTHS	0.264709	-0.029400
DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS	0.070715	0.034462
NO.OF_INQUIRIES	1.000000	0.043678
loan_default	0.043678	1.000000

[36 rows x 36 columns]

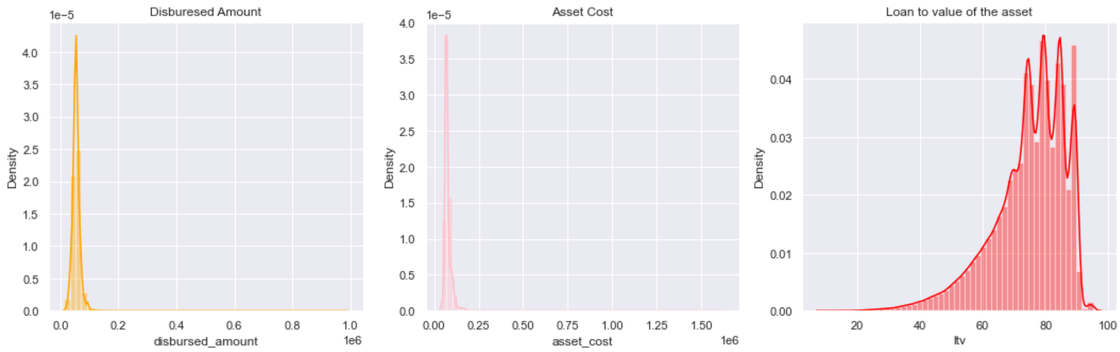
[30]:

```python
fig, ax = plt.subplots(figsize=(25,25))
dataplot = sns.heatmap(train.corr(), cmap="YlGnBu", annot=True,linewidth=0.
↪4,ax=ax)

# asset_cost & disbursed amount: 0.75
# primary no. of account & pri. active account: 0.75
# primary disbursed amount & primary sanctioned amount: 1
# sec. no. of account & sec. active account: 0.83
# Sec. overdue account & sec. active account: 0.53
# Sec. overdue account & sec. no. of account: 0.51
# sec. sanctioned amount & Sec. current balance: 0.93
# sec. sanctioned amount & sec. disbursed amount: 0.93
# new account in six months & pri. active account: 0.7
# new account in six months & primary no. of account: 0.54
```



[24]:
```python
# check the distribution of disbursed amount

plt.rcParams['figure.figsize'] = (18, 5)

plt.subplot(1, 3, 1)
sns.distplot(train['disbursed_amount'],  color = 'orange')
plt.title('Disburesed Amount')

plt.subplot(1, 3, 2)
sns.distplot(train['asset_cost'], color = 'pink')
plt.title('Asset Cost')

plt.subplot(1, 3, 3)
sns.distplot(train['ltv'], color = 'red')
plt.title('Loan to value of the asset')

plt.show()
```

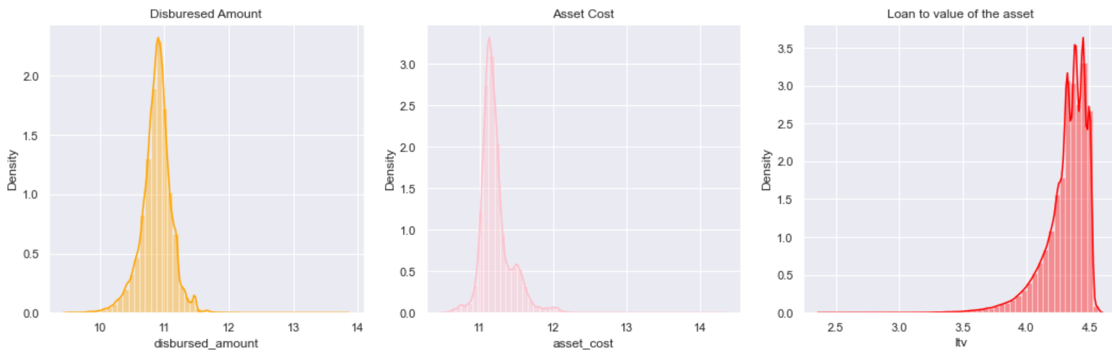
```
C:\Users\HP\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `histplot` (an axes-level function for
histograms).
  warnings.warn(msg, FutureWarning)
C:\Users\HP\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `histplot` (an axes-level function for
histograms).
  warnings.warn(msg, FutureWarning)
C:\Users\HP\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `histplot` (an axes-level function for
histograms).
  warnings.warn(msg, FutureWarning)
```




```
[25]: #performing log transformations on disbursed amount, ltv, and asset cost

train['disbursed_amount'] = np.log1p(train['disbursed_amount'])
train['ltv'] = np.log1p(train['ltv'])
train['asset_cost'] = np.log1p(train['asset_cost'])


plt.rcParams['figure.figsize'] = (18, 5)

plt.subplot(1, 3, 1)
sns.distplot(train['disbursed_amount'],  color = 'orange')
plt.title('Disburesed Amount')

plt.subplot(1, 3, 2)
sns.distplot(train['asset_cost'], color = 'pink')
plt.title('Asset Cost')

plt.subplot(1, 3, 3)
sns.distplot(train['ltv'], color = 'red')
plt.title('Loan to value of the asset')

plt.show()
```

```
C:\Users\HP\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `histplot` (an axes-level function for
histograms).
  warnings.warn(msg, FutureWarning)
C:\Users\HP\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `histplot` (an axes-level function for
histograms).
  warnings.warn(msg, FutureWarning)
```

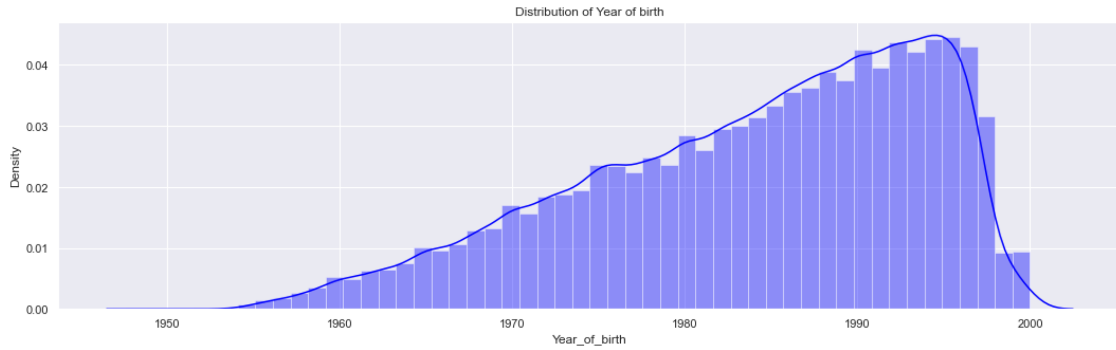
```
C:\Users\HP\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `histplot` (an axes-level function for
histograms).
  warnings.warn(msg, FutureWarning)
```



[26]:
```
# date of birth is an useless attribute
#  the only thing we can extract the is the year of birth
# let's first convert the date into date-time format
plt.rcParams['figure.figsize'] = (18, 5)

train['Date.of.Birth'] = pd.to_datetime(train['Date.of.Birth'], errors =␣
 ↪'coerce')

# extracting the year of birth of the customers
train['Year_of_birth'] = train['Date.of.Birth'].dt.year

# checking the values inside date of year
sns.distplot(train['Year_of_birth'], color = 'blue')
plt.title('Distribution of Year of birth')
```

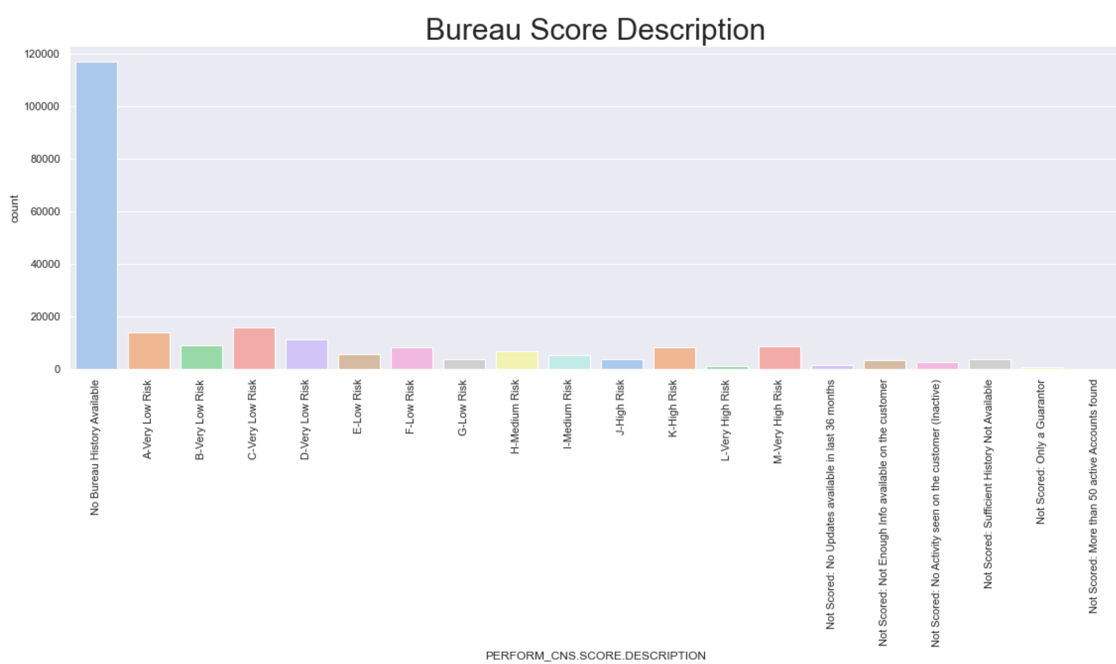
```
C:\Users\HP\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `histplot` (an axes-level function for
histograms).
  warnings.warn(msg, FutureWarning)
```

[26]: Text(0.5, 1.0, 'Distribution of Year of birth')




```
[27]: # checking the bureau score description

plt.rcParams['figure.figsize'] = (19, 6)
sns.countplot(train['PERFORM_CNS.SCORE.DESCRIPTION'], palette = 'pastel')
plt.title('Bureau Score Description', fontsize = 30)
plt.xticks(rotation = 90)
plt.show()
```

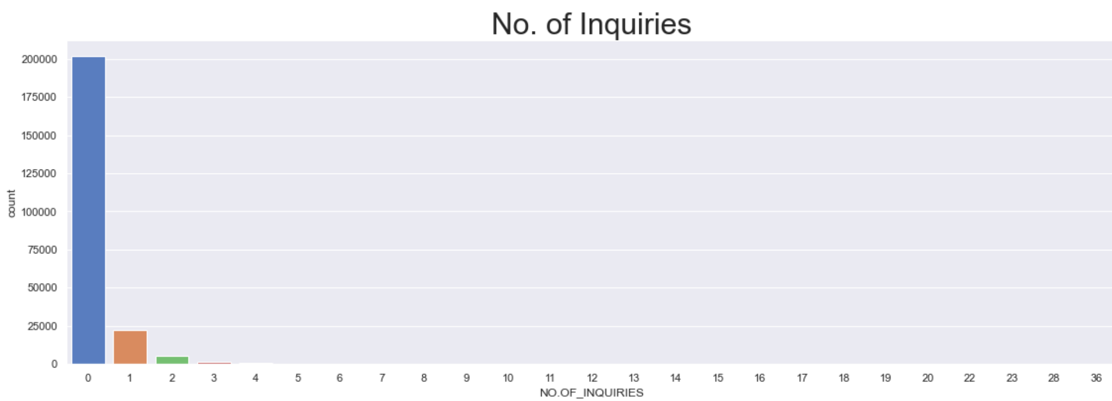
```
C:\Users\HP\anaconda3\lib\site-packages\seaborn\_decorators.py:36:
FutureWarning: Pass the following variable as a keyword arg: x. From version
0.12, the only valid positional argument will be `data`, and passing other
arguments without an explicit keyword will result in an error or
misinterpretation.
  warnings.warn(
```



[28]:
```
# plotting a countplot

sns.countplot(train['NO.OF_INQUIRIES'], palette = 'muted')
plt.title('No. of Inquiries',  fontsize = 30)
```

```
C:\Users\HP\anaconda3\lib\site-packages\seaborn\_decorators.py:36:
FutureWarning: Pass the following variable as a keyword arg: x. From version
0.12, the only valid positional argument will be `data`, and passing other
arguments without an explicit keyword will result in an error or
misinterpretation.
  warnings.warn(
```

[28]:
```
Text(0.5, 1.0, 'No. of Inquiries')
```



[159]:
```
# removing unnecassary columns

train1 = train.drop(['UniqueID', 'supplier_id', 'Current_pincode_ID', 'Date.of.
↪Birth', 'DisbursalDate', 'Employee_code_ID','Year_of_birth','AVERAGE.ACCT.
↪AGE','CREDIT.HISTORY.LENGTH','PERFORM_CNS.SCORE.DESCRIPTION'], axis = 1)

# checking the new columns of data
train1.columns
```

[159]:
```
Index(['disbursed_amount', 'asset_cost', 'ltv', 'branch_id', 'manufacturer_id',
       'Employment.Type', 'State_ID', 'MobileNo_Avl_Flag', 'Aadhar_flag',
       'PAN_flag', 'VoterID_flag', 'Driving_flag', 'Passport_flag',
       'PERFORM_CNS.SCORE', 'PRI.NO.OF.ACCTS', 'PRI.ACTIVE.ACCTS',
       'PRI.OVERDUE.ACCTS', 'PRI.CURRENT.BALANCE', 'PRI.SANCTIONED.AMOUNT',
       'PRI.DISBURSED.AMOUNT', 'SEC.NO.OF.ACCTS', 'SEC.ACTIVE.ACCTS',
       'SEC.OVERDUE.ACCTS', 'SEC.CURRENT.BALANCE', 'SEC.SANCTIONED.AMOUNT',
       'SEC.DISBURSED.AMOUNT', 'PRIMARY.INSTAL.AMT', 'SEC.INSTAL.AMT',
       'NEW.ACCTS.IN.LAST.SIX.MONTHS', 'DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS',
```

```
       'NO.OF_INQUIRIES', 'loan_default'],
      dtype='object')
```

[160]:
```python
train1.sample(5)
```

[160]:
```
        disbursed_amount  asset_cost       ltv  branch_id  manufacturer_id  \
162017         10.758881   11.155250  4.243339         70               51
211979         10.255200   11.272318  3.664331          1               86
167352         10.654786   11.087069  4.224788         79               45
174877         10.899162   11.580341  3.978934        160               51
126029         11.022131   11.478241  4.192983        104               51

        Employment.Type  State_ID  MobileNo_Avl_Flag  Aadhar_flag  PAN_flag  \
162017                1         4                  1            1         0
211979                1         3                  1            0         0
167352                1        18                  1            1         0
174877                1         1                  1            1         0
126029                1        10                  1            1         0

        …  SEC.OVERDUE.ACCTS  SEC.CURRENT.BALANCE  SEC.SANCTIONED.AMOUNT  \
162017  …                  0                    0                      0
211979  …                  0                    0                      0
167352  …                  0                    0                      0
174877  …                  0                    0                      0
126029  …                  0                    0                      0

        SEC.DISBURSED.AMOUNT  PRIMARY.INSTAL.AMT  SEC.INSTAL.AMT  \
162017                     0                3072               0
211979                     0                   0               0
167352                     0               21079               0
174877                     0                   0               0
126029                     0                2089               0

        NEW.ACCTS.IN.LAST.SIX.MONTHS  DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS  \
162017                             0                                    0
211979                             0                                    0
167352                             0                                    0
174877                             1                                    0
126029                             0                                    0

        NO.OF_INQUIRIES  loan_default
162017                0             1
211979                0             0
167352                0             0
174877                0             0
126029                0             0
```

```
[5 rows x 32 columns]
```

[161]:
```
# some attributes are categorical but they are in integer so let's convert them␣
↪into category

train1['branch_id'] = train1['branch_id'].astype('category')
train1['manufacturer_id'] = train1['manufacturer_id'].astype('category')
train1['State_ID'] = train1['State_ID'].astype('category')

from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()
train1['branch_id'] = le.fit_transform(train1['branch_id'])
train1['manufacturer_id'] = le.fit_transform(train1['manufacturer_id'])
train1['State_ID'] = le.fit_transform(train1['State_ID'])

# checking the values in these attributes
#data['branch_id'].value_counts()
#data['manufacturer_id'].value_counts()
#data['State_ID'].value_counts()
```

[162]:
```
# checking the target variable : loan defaulter
y_train = train1.iloc[:, -1]
```

[163]:
```
y_train.value_counts()
```

[163]:
```
0    182543
1     50611
Name: loan_default, dtype: int64
```

[164]:
```
train1.head()
```

[164]:
```
   disbursed_amount  asset_cost       ltv  branch_id  manufacturer_id  \
0         10.831292   10.975088  4.505902         30                0
1         10.883298   11.024530  4.506785         30                0
2         10.866261   11.007104  4.493009         30                0
3         10.743977   11.026809  4.349245         30                0
4         10.682698   11.267754  4.069027         30                5

   Employment.Type  State_ID  MobileNo_Avl_Flag  Aadhar_flag  PAN_flag  …  \
0                1         5                  1            1         0  …
1                2         5                  1            1         0  …
2                2         5                  1            1         0  …
3                1         5                  1            1         0  …
4                2         5                  1            1         0  …

   SEC.OVERDUE.ACCTS  SEC.CURRENT.BALANCE  SEC.SANCTIONED.AMOUNT  \
```

```
0                   0                     0                        0
1                   0                     0                        0
2                   0                     0                        0
3                   0                     0                        0
4                   0                     0                        0

   SEC.DISBURSED.AMOUNT  PRIMARY.INSTAL.AMT  SEC.INSTAL.AMT  \
0                     0                   0               0
1                     0                   0               0
2                     0                   0               0
3                     0                   0               0
4                     0                   0               0

   NEW.ACCTS.IN.LAST.SIX.MONTHS  DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS  \
0                             0                                    0
1                             0                                    0
2                             0                                    0
3                             0                                    0
4                             0                                    0

   NO.OF_INQUIRIES  loan_default
0                0             0
1                0             0
2                1             1
3                0             0
4                0             0

[5 rows x 32 columns]
```

[168]:
```
X = train1.iloc[:233154,:31]
```

[169]:
```
X = X.astype({"disbursed_amount":'int', "asset_cost":'int',"ltv":'int'})
```

[170]:
```
X.head()
```

[170]:
```
   disbursed_amount  asset_cost  ltv  branch_id  manufacturer_id  \
0                10          10    4         30                0
1                10          11    4         30                0
2                10          11    4         30                0
3                10          11    4         30                0
4                10          11    4         30                5

   Employment.Type  State_ID  MobileNo_Avl_Flag  Aadhar_flag  PAN_flag  …  \
0                1         5                  1            1         0  …
1                2         5                  1            1         0  …
2                2         5                  1            1         0  …
3                1         5                  1            1         0  …
```

```
4                2        5                  1            1         0  …

   SEC.ACTIVE.ACCTS  SEC.OVERDUE.ACCTS  SEC.CURRENT.BALANCE  \
0                 0                  0                    0
1                 0                  0                    0
2                 0                  0                    0
3                 0                  0                    0
4                 0                  0                    0

   SEC.SANCTIONED.AMOUNT  SEC.DISBURSED.AMOUNT  PRIMARY.INSTAL.AMT  \
0                      0                     0                   0
1                      0                     0                   0
2                      0                     0                   0
3                      0                     0                   0
4                      0                     0                   0

   SEC.INSTAL.AMT  NEW.ACCTS.IN.LAST.SIX.MONTHS  \
0               0                             0
1               0                             0
2               0                             0
3               0                             0
4               0                             0

   DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS  NO.OF_INQUIRIES
0                                    0                0
1                                    0                0
2                                    0                1
3                                    0                0
4                                    0                0

[5 rows x 31 columns]
```

[171]:
```
X.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 233154 entries, 0 to 233153
Data columns (total 31 columns):
 #   Column                               Non-Null Count   Dtype
---  ------                               --------------   -----
 0   disbursed_amount                     233154 non-null  int32
 1   asset_cost                           233154 non-null  int32
 2   ltv                                  233154 non-null  int32
 3   branch_id                            233154 non-null  int64
 4   manufacturer_id                      233154 non-null  int64
 5   Employment.Type                      233154 non-null  int64
 6   State_ID                             233154 non-null  int64
 7   MobileNo_Avl_Flag                    233154 non-null  int64
```

```
 8   Aadhar_flag                          233154 non-null  int64
 9   PAN_flag                             233154 non-null  int64
 10  VoterID_flag                         233154 non-null  int64
 11  Driving_flag                         233154 non-null  int64
 12  Passport_flag                        233154 non-null  int64
 13  PERFORM_CNS.SCORE                    233154 non-null  int64
 14  PRI.NO.OF.ACCTS                      233154 non-null  int64
 15  PRI.ACTIVE.ACCTS                     233154 non-null  int64
 16  PRI.OVERDUE.ACCTS                    233154 non-null  int64
 17  PRI.CURRENT.BALANCE                  233154 non-null  int64
 18  PRI.SANCTIONED.AMOUNT                233154 non-null  int64
 19  PRI.DISBURSED.AMOUNT                 233154 non-null  int64
 20  SEC.NO.OF.ACCTS                      233154 non-null  int64
 21  SEC.ACTIVE.ACCTS                     233154 non-null  int64
 22  SEC.OVERDUE.ACCTS                    233154 non-null  int64
 23  SEC.CURRENT.BALANCE                  233154 non-null  int64
 24  SEC.SANCTIONED.AMOUNT                233154 non-null  int64
 25  SEC.DISBURSED.AMOUNT                 233154 non-null  int64
 26  PRIMARY.INSTAL.AMT                   233154 non-null  int64
 27  SEC.INSTAL.AMT                       233154 non-null  int64
 28  NEW.ACCTS.IN.LAST.SIX.MONTHS         233154 non-null  int64
 29  DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS  233154 non-null  int64
 30  NO.OF_INQUIRIES                      233154 non-null  int64
dtypes: int32(3), int64(28)
memory usage: 52.5 MB
```

[172]:
```
y = train1["loan_default"]
```

[173]:
```
y.head()
```

[173]:
```
0    0
1    0
2    1
3    0
4    0
Name: loan_default, dtype: int64
```

[174]:
```
 from sklearn.model_selection import train_test_split
```

[175]:
```
x_train, x_test, y_train, y_test = train_test_split(X, y, test_size = 0.25,␣
↪random_state = 0)
```

[176]:
```
from sklearn.preprocessing import StandardScaler
```

[177]:
```
from sklearn.linear_model import LogisticRegression
```

[ ]:
```

```

[178]:

```python
# all parameters not specified are set to their defaults
logisticRegr = LogisticRegression()
```

[179]:

```python
logisticRegr.fit(x_train,y_train)
```

```
C:\Users\HP\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:763:
ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression
  n_iter_i = _check_optimize_result(
```

[179]:

```
LogisticRegression()
```

[181]:

```python
y_pred=logisticRegr.predict(x_test)
```

[182]:

```python
from sklearn import metrics
cnf_matrix = metrics.confusion_matrix(y_test, y_pred)
cnf_matrix
```

[182]:

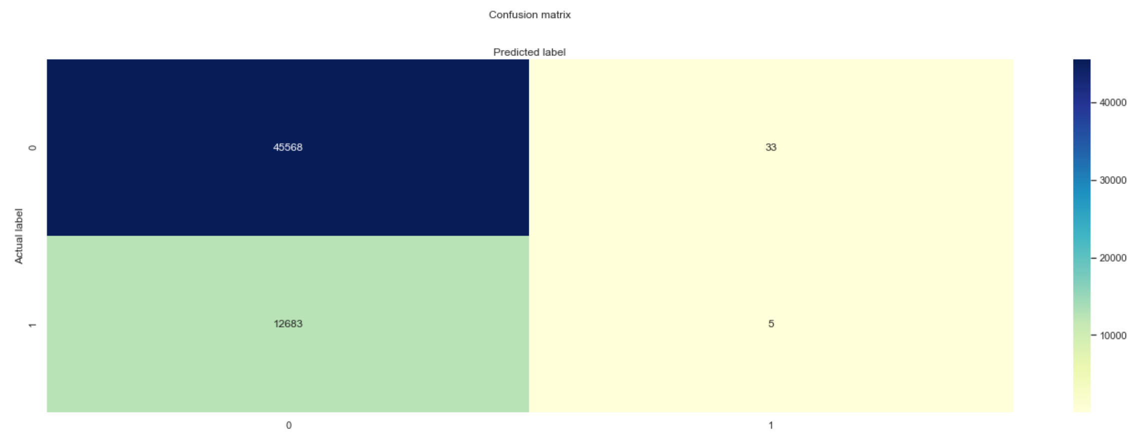
```
array([[45568,    33],
       [12683,     5]], dtype=int64)
```

[196]:

```python
# cm = confusion_matrix(y_valid, y_pred)
import matplotlib.pyplot as plt
fig, ax = plt.subplots()
sns.heatmap(cnf_matrix, annot = True,cmap="YlGnBu" ,fmt='g')
ax.xaxis.set_label_position("top")
plt.tight_layout()
plt.title('Confusion matrix', y=1.1)
plt.ylabel('Actual label')
plt.xlabel('Predicted label')
```

[196]:

```
Text(0.5, 384.16, 'Predicted label')
```



[185]:
```
print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
print("Precision:",metrics.precision_score(y_test, y_pred))
print("Recall:",metrics.recall_score(y_test, y_pred))
```

```
Accuracy: 0.7818456312511795
Precision: 0.13157894736842105
Recall: 0.0003940731399747793
```