rnjsqaoqn

October 10, 2025

Credit Card Fraud Detection

Goal:

Predict the probability of an online credit card transaction being fraudulent, based on different properties of the transactions.

1 1. Setup Environment

[1]:
```
# Data Manipulation
import numpy as np
import pandas as pd

# Data Visualization
import seaborn as sns
import matplotlib.pyplot as plt
import matplotlib.lines as mlines

# Time
import time
import datetime

# Machine Learning
from   sklearn.preprocessing import LabelEncoder, minmax_scale
from   sklearn.ensemble import RandomForestClassifier
from   sklearn.decomposition import PCA
from   sklearn.model_selection import train_test_split, GridSearchCV
from   sklearn.metrics import (confusion_matrix , classification_report,␣
 ↪accuracy_score, roc_auc_score,
                               plot_roc_curve, precision_recall_curve,␣
 ↪plot_precision_recall_curve)
from   sklearn.calibration import calibration_curve
from   sklearn.calibration import CalibratedClassifierCV

from   xgboost import XGBClassifier
from   lightgbm import LGBMClassifier

from   imblearn.over_sampling import RandomOverSampler
```

```
from   scipy.stats import chi2_contingency,  f_oneway

import gc
import warnings
from   tqdm import tqdm


# Set Options
pd.set_option('display.max_rows', 100)
pd.set_option('display.max_columns', 500)
%matplotlib inline
warnings.filterwarnings("ignore")
```

2 2. Data Overview

[2]:
```
%%time
df_id   = pd.read_csv(r'C:\Users\admin\Desktop\PYTHON␣
 ↪LEARNING\Data_Projects_01\MLp Projects\CREDIT CARD FRAUD␣
 ↪DETECTION\Datasets\train_identity.csv')
df_tran = pd.read_csv(r'C:\Users\admin\Desktop\PYTHON␣
 ↪LEARNING\Data_Projects_01\MLp Projects\CREDIT CARD FRAUD␣
 ↪DETECTION\Datasets\train_transaction.csv')
```

```
Wall time: 26.5 s
```

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

[3]:

	TransactionID	id_01	id_02	id_03	id_04	id_05	id_06	id_07	id_08	\
0	2987004	0.0	70787.0	NaN	NaN	NaN	NaN	NaN	NaN	
1	2987008	-5.0	98945.0	NaN	NaN	0.0	-5.0	NaN	NaN	
2	2987010	-5.0	191631.0	0.0	0.0	0.0	0.0	NaN	NaN	
3	2987011	-5.0	221832.0	NaN	NaN	0.0	-6.0	NaN	NaN	
4	2987016	0.0	7460.0	0.0	0.0	1.0	0.0	NaN	NaN	

	id_09	id_10	id_11	id_12	id_13	id_14	id_15	id_16	id_17	id_18	\
0	NaN	NaN	100.0	NotFound	NaN	-480.0	New	NotFound	166.0	NaN	
1	NaN	NaN	100.0	NotFound	49.0	-300.0	New	NotFound	166.0	NaN	
2	0.0	0.0	100.0	NotFound	52.0	NaN	Found	Found	121.0	NaN	
3	NaN	NaN	100.0	NotFound	52.0	NaN	New	NotFound	225.0	NaN	
4	0.0	0.0	100.0	NotFound	NaN	-300.0	Found	Found	166.0	15.0	

	id_19	id_20	id_21	id_22	id_23	id_24	id_25	id_26	id_27	id_28	\
0	542.0	144.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	New	
1	621.0	500.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	New	
2	410.0	142.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Found	
3	176.0	507.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	New	

```
4  529.0  575.0    NaN    NaN   NaN    NaN    NaN    NaN   NaN  Found

      id_29             id_30               id_31  id_32      id_33  \
0  NotFound       Android 7.0  samsung browser 6.2   32.0  2220x1080
1  NotFound        iOS 11.1.2   mobile safari 11.0   32.0   1334x750
2     Found               NaN          chrome 62.0    NaN        NaN
3  NotFound               NaN          chrome 62.0    NaN        NaN
4     Found  Mac OS X 10_11_6          chrome 62.0   24.0   1280x800

            id_34 id_35 id_36 id_37 id_38 DeviceType  \
0  match_status:2     T     F     T     T     mobile
1  match_status:1     T     F     F     T     mobile
2             NaN     F     F     T     T    desktop
3             NaN     F     F     T     T    desktop
4  match_status:2     T     F     T     T    desktop

                      DeviceInfo
0  SAMSUNG SM-G892A Build/NRD90M
1                     iOS Device
2                        Windows
3                            NaN
4                          MacOS
```

2.0.1 Identity Data Description

Variables in this table are identity information – network connection information (IP, ISP, Proxy, etc) and digital signature (UA/browser/os/version, etc) associated with transactions. They're collected by Vesta's fraud protection system and digital security partners. (The field names are masked and pairwise dictionary will not be provided for privacy protection and contract agreement)

Categorical Features: - DeviceType - DeviceInfo - id_12 - id_38

[4]:
```
df_tran.head()
```

[4]:
```
   TransactionID  isFraud  TransactionDT  TransactionAmt ProductCD  card1  \
0        2987000        0          86400            68.5         W  13926
1        2987001        0          86401            29.0         W   2755
2        2987002        0          86469            59.0         W   4663
3        2987003        0          86499            50.0         W  18132
4        2987004        0          86506            50.0         H   4497

   card2  card3       card4  card5   card6  addr1  addr2  dist1  dist2  \
0    NaN  150.0    discover  142.0  credit  315.0   87.0   19.0    NaN
1  404.0  150.0  mastercard  102.0  credit  325.0   87.0    NaN    NaN
2  490.0  150.0        visa  166.0   debit  330.0   87.0  287.0    NaN
3  567.0  150.0  mastercard  117.0   debit  476.0   87.0    NaN    NaN
4  514.0  150.0  mastercard  102.0  credit  420.0   87.0    NaN    NaN
```

```
  P_emaildomain R_emaildomain   C1   C2   C3   C4   C5   C6   C7   C8   C9  \
0           NaN           NaN  1.0  1.0  0.0  0.0  0.0  1.0  0.0  0.0  1.0
1     gmail.com           NaN  1.0  1.0  0.0  0.0  0.0  1.0  0.0  0.0  0.0
2   outlook.com           NaN  1.0  1.0  0.0  0.0  0.0  1.0  0.0  0.0  1.0
3     yahoo.com           NaN  2.0  5.0  0.0  0.0  0.0  4.0  0.0  0.0  1.0
4     gmail.com           NaN  1.0  1.0  0.0  0.0  0.0  1.0  0.0  1.0  0.0

   C10  C11  C12   C13  C14     D1     D2    D3    D4   D5  D6  D7  D8  D9  \
0  0.0  2.0  0.0   1.0  1.0   14.0    NaN  13.0   NaN  NaN NaN NaN NaN NaN
1  0.0  1.0  0.0   1.0  1.0    0.0    NaN   NaN   0.0  NaN NaN NaN NaN NaN
2  0.0  1.0  0.0   1.0  1.0    0.0    NaN   NaN   0.0  NaN NaN NaN NaN NaN
3  0.0  1.0  0.0  25.0  1.0  112.0  112.0   0.0  94.0  0.0 NaN NaN NaN NaN
4  1.0  1.0  0.0   1.0  1.0    0.0    NaN   NaN   NaN  NaN NaN NaN NaN NaN

    D10    D11  D12  D13  D14    D15   M1   M2   M3   M4   M5   M6   M7   M8  \
0  13.0   13.0  NaN  NaN  NaN    0.0    T    T    T   M2    F    T  NaN  NaN
1   0.0    NaN  NaN  NaN  NaN    0.0  NaN  NaN  NaN   M0    T    T  NaN  NaN
2   0.0  315.0  NaN  NaN  NaN  315.0    T    T    T   M0    F    F    F    F
3  84.0    NaN  NaN  NaN  NaN  111.0  NaN  NaN  NaN   M0    T    F  NaN  NaN
4   NaN    NaN  NaN  NaN  NaN    NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

    M9   V1   V2   V3   V4   V5   V6   V7   V8   V9  V10  V11  V12  V13  V14  \
0  NaN  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0
1  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  0.0  0.0  1.0
2    F  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0
3  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  1.0  1.0  1.0
4  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

   V15  V16  V17  V18  V19  V20  V21  V22  V23  V24  V25  V26  V27  V28  V29  \
0  0.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  0.0
1  0.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  0.0
2  0.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  0.0
3  0.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  0.0
4  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

   V30  V31  V32  V33  V34  V35  V36  V37  V38  V39  V40  V41  V42  V43  V44  \
0  0.0  0.0  0.0  0.0  0.0  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN
1  0.0  0.0  0.0  0.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  1.0  0.0  0.0  1.0
2  0.0  0.0  0.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  1.0  0.0  0.0  1.0
3  0.0  0.0  0.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  1.0  0.0  0.0  1.0
4  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

   V45  V46  V47  V48  V49  V50  V51  V52  V53  V54  V55  V56  V57  V58  V59  \
0  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  1.0  1.0  1.0  1.0  0.0  0.0  0.0
1  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  0.0
2  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  0.0
3  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  0.0
```

```
4  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

   V60  V61  V62  V63  V64  V65  V66  V67  V68  V69  V70  V71  V72  V73  V74  \
0  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
1  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
2  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
3  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
4  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

   V75  V76  V77  V78  V79  V80  V81  V82  V83  V84  V85  V86  V87  V88  V89  \
0  1.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  1.0  1.0  1.0  0.0
1  0.0  0.0  1.0  1.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  0.0
2  1.0  1.0  1.0  1.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  0.0
3  1.0  1.0  1.0  1.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  0.0
4  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

   V90  V91  V92  V93  V94  V95   V96   V97  V98   V99  V100  V101  V102  \
0  0.0  0.0  0.0  0.0  0.0  0.0   1.0   0.0  0.0   0.0   0.0   0.0   1.0
1  0.0  0.0  0.0  0.0  0.0  0.0   0.0   0.0  0.0   0.0   0.0   0.0   0.0
2  0.0  0.0  0.0  0.0  0.0  0.0   0.0   0.0  0.0   0.0   0.0   0.0   0.0
3  0.0  0.0  0.0  0.0  0.0  1.0  48.0  28.0  0.0  10.0   4.0   1.0  38.0
4  NaN  NaN  NaN  NaN  NaN  0.0   0.0   0.0  0.0   0.0   0.0   0.0   0.0

   V103  V104  V105  V106  V107  V108  V109  V110  V111  V112  V113  V114  \
0   0.0   0.0   0.0   0.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0
1   0.0   0.0   0.0   0.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0
2   0.0   0.0   0.0   0.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0
3  24.0   0.0   0.0   0.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0
4   0.0   0.0   0.0   0.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0

   V115  V116  V117  V118  V119  V120  V121  V122  V123  V124  V125  V126  \
0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   0.0
1   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   0.0
2   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   0.0
3   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0  50.0
4   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   0.0

     V127   V128  V129   V130   V131  V132    V133   V134  V135  V136  V137  \
0   117.0    0.0   0.0    0.0    0.0   0.0   117.0    0.0   0.0   0.0   0.0
1     0.0    0.0   0.0    0.0    0.0   0.0     0.0    0.0   0.0   0.0   0.0
2     0.0    0.0   0.0    0.0    0.0   0.0     0.0    0.0   0.0   0.0   0.0
3  1758.0  925.0   0.0  354.0  135.0  50.0  1404.0  790.0   0.0   0.0   0.0
4     0.0    0.0   0.0    0.0    0.0   0.0     0.0    0.0   0.0   0.0   0.0

   V138  V139  V140  V141  V142  V143  V144   V145  V146  V147  V148  V149  \
0   NaN   NaN   NaN   NaN   NaN   NaN   NaN    NaN   NaN   NaN   NaN   NaN
1   NaN   NaN   NaN   NaN   NaN   NaN   NaN    NaN   NaN   NaN   NaN   NaN
```

```
2   NaN   NaN   NaN   NaN   NaN   NaN   NaN    NaN   NaN   NaN   NaN   NaN
3   NaN   NaN   NaN   NaN   NaN   NaN   NaN    NaN   NaN   NaN   NaN   NaN
4   0.0   0.0   0.0   0.0   0.0   6.0  18.0  140.0   0.0   0.0   0.0   0.0

     V150  V151  V152  V153  V154  V155  V156  V157  V158          V159  \
0     NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN           NaN
1     NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN           NaN
2     NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN           NaN
3     NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN           NaN
4  1803.0  49.0  64.0   0.0   0.0   0.0   0.0   0.0   0.0  15557.990234

             V160  V161  V162  V163   V164    V165    V166  V167  V168  V169  \
0             NaN   NaN   NaN   NaN    NaN     NaN     NaN   NaN   NaN   NaN
1             NaN   NaN   NaN   NaN    NaN     NaN     NaN   NaN   NaN   NaN
2             NaN   NaN   NaN   NaN    NaN     NaN     NaN   NaN   NaN   NaN
3             NaN   NaN   NaN   NaN    NaN     NaN     NaN   NaN   NaN   NaN
4  169690.796875   0.0   0.0   0.0  515.0  5155.0  2840.0   0.0   0.0   0.0

   V170  V171  V172  V173  V174  V175  V176  V177  V178  V179  V180  V181  \
0   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
1   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
2   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
3   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
4   1.0   1.0   0.0   0.0   0.0   0.0   1.0   0.0   0.0   0.0   0.0   0.0

   V182  V183  V184  V185  V186  V187  V188  V189  V190  V191  V192  V193  \
0   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
1   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
2   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
3   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
4   0.0   0.0   0.0   0.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0

   V194  V195  V196  V197  V198  V199  V200  V201  V202  V203  V204  V205  \
0   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
1   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
2   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
3   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
4   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   0.0   0.0   0.0   0.0

   V206  V207  V208  V209  V210  V211  V212  V213  V214  V215  V216  V217  \
0   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
1   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
2   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
3   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
4   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0

   V218  V219  V220  V221  V222  V223  V224  V225  V226  V227  V228  V229  \
```

```
0   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
1   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
2   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
3   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
4   0.0   0.0   0.0   1.0   1.0   0.0   0.0   0.0   0.0   0.0   1.0   1.0

   V230  V231  V232  V233  V234  V235  V236  V237  V238  V239  V240  V241  \
0   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
1   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
2   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
3   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
4   1.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   1.0   1.0

   V242  V243  V244  V245  V246  V247  V248  V249  V250  V251  V252  V253  \
0   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
1   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
2   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
3   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
4   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0

   V254  V255  V256  V257  V258  V259  V260  V261  V262  V263  V264  V265  \
0   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
1   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
2   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
3   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
4   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   0.0   0.0   0.0

   V266  V267  V268  V269  V270  V271  V272  V273  V274  V275  V276  V277  \
0   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
1   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
2   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
3   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
4   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0

   V278  V279  V280  V281  V282  V283  V284  V285  V286  V287  V288  V289  \
0   NaN   0.0   0.0   0.0   1.0   1.0   0.0   0.0   0.0   0.0   0.0   0.0
1   NaN   0.0   0.0   0.0   1.0   1.0   0.0   0.0   0.0   0.0   0.0   0.0
2   NaN   0.0   0.0   0.0   1.0   1.0   0.0   0.0   0.0   0.0   0.0   0.0
3   NaN   1.0  28.0   0.0   0.0   0.0   0.0  10.0   0.0   4.0   0.0   0.0
4   0.0   0.0   0.0   0.0   1.0   1.0   0.0   0.0   0.0   0.0   0.0   0.0

   V290  V291  V292  V293  V294  V295  V296  V297  V298  V299  V300  V301  \
0   1.0   1.0   1.0   0.0   1.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
1   1.0   1.0   1.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
2   1.0   1.0   1.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
3   1.0   1.0   1.0   1.0  38.0  24.0   0.0   0.0   0.0   0.0   0.0   0.0
4   1.0   1.0   1.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
```

	V302	V303	V304	V305	V306	V307	V308	V309	V310	V311	V312	\
0	0.0	0.0	0.0	1.0	0.0	117.0	0.0	0.0	0.0	0.0	0.0	
1	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
2	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
3	0.0	0.0	0.0	1.0	50.0	1758.0	925.0	0.0	354.0	0.0	135.0	
4	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

	V313	V314	V315	V316	V317	V318	V319	V320	V321	V322	V323	V324	\
0	0.0	0.0	0.0	0.0	117.0	0.0	0.0	0.0	0.0	NaN	NaN	NaN	
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN	NaN	NaN	
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN	NaN	NaN	
3	0.0	0.0	0.0	50.0	1404.0	790.0	0.0	0.0	0.0	NaN	NaN	NaN	
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

	V325	V326	V327	V328	V329	V330	V331	V332	V333	V334	V335	V336	\
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

	V337	V338	V339
0	NaN	NaN	NaN
1	NaN	NaN	NaN
2	NaN	NaN	NaN
3	NaN	NaN	NaN
4	0.0	0.0	0.0

2.0.2 Transaction Data Description

- **TransactionDT**: timedelta from a given reference datetime (not an actual timestamp)
- **TransactionAMT**: transaction payment amount in USD
- **ProductCD**: product code, the product for each transaction
- **card1 - card6**: payment card information, such as card type, card category, issue bank, country, etc.
- **addr**: address
- **dist**: distance
- **P_ and (R___) emaildomain**: purchaser and recipient email domain
- **C1-C14**: counting, such as how many addresses are found to be associated with the payment card, etc. The actual meaning is masked.
- **D1-D15**: timedelta, such as days between previous transaction, etc.
- **M1-M9**: match, such as names on card and address, etc.
- **Vxxx**: Vesta engineered rich features, including ranking, counting, and other entity relations.

3 3. Optimize Memory Used by Data

Memory occupied by the dataframe (in mb)

[5]:
```
df_id.memory_usage(deep=True).sum() / 1024**2
```

[5]: 157.63398933410645

[6]:
```
df_tran.memory_usage(deep=True).sum() / 1024**2
```

[6]: 2100.701406478882

[7]:
```
print('int64 min: ', np.iinfo(np.int64).min)
print('int64 max: ', np.iinfo(np.int64).max)
```

```
int64 min:  -9223372036854775808
int64 max:  9223372036854775807
```

[8]:
```
print('int8 min: ', np.iinfo(np.int8).min)
print('int8 max: ', np.iinfo(np.int8).max)
```

```
int8 min:  -128
int8 max:  127
```

[9]:
```
# Reduce memory usage
def reduce_mem_usage(df, verbose=True):
    numerics = ['int16', 'int32', 'int64', 'float16', 'float32', 'float64']
    start_mem = df.memory_usage(deep=True).sum() / 1024**2
    for col in df.columns:
        col_type = df[col].dtypes
        if col_type in numerics:
            c_min = df[col].min()
            c_max = df[col].max()
            if str(col_type)[:3] == 'int':
                if c_min > np.iinfo(np.int8).min and c_max < np.iinfo(np.int8).
 ↪max:
                    df[col] = df[col].astype(np.int8)
                elif c_min > np.iinfo(np.int16).min and c_max < np.iinfo(np.
 ↪int16).max:
                    df[col] = df[col].astype(np.int16)
                elif c_min > np.iinfo(np.int32).min and c_max < np.iinfo(np.
 ↪int32).max:
                    df[col] = df[col].astype(np.int32)
                elif c_min > np.iinfo(np.int64).min and c_max < np.iinfo(np.
 ↪int64).max:
                    df[col] = df[col].astype(np.int64)
            else:
                if c_min > np.finfo(np.float16).min and c_max < np.finfo(np.
 ↪float16).max:
                    df[col] = df[col].astype(np.float16)
```

```
                elif c_min > np.finfo(np.float32).min and c_max < np.finfo(np.
↪float32).max:
                    df[col] = df[col].astype(np.float32)
                else:
                    df[col] = df[col].astype(np.float64)
    end_mem = df.memory_usage(deep=True).sum() / 1024**2
    if verbose: print('Mem. usage decreased to {:5.2f} Mb ({:.1f}% reduction)'.
↪format(end_mem, 100 * (start_mem - end_mem) / start_mem))
    return df
```

[10]:
```
# Reduce the memory size of the dataframe
df_id   = reduce_mem_usage(df_id)
df_tran = reduce_mem_usage(df_tran)
```

```
Mem. usage decreased to 138.38 Mb (12.2% reduction)
Mem. usage decreased to 867.89 Mb (58.7% reduction)
```

4 4. Basic Data Stats

Shape of dataframe

[11]:
```
df_id.shape
```

[11]: (144233, 41)

[12]:
```
df_tran.shape
```

[12]: (590540, 394)

Check how many transactions has ID info

[13]:
```
df_tran.TransactionID.isin(df_id.TransactionID).sum()
```

[13]: 144233

Summary of dataframe

[14]:
```
from pandas_summary import DataFrameSummary
df_id_summary = DataFrameSummary(df_id)
df_id_summary.summary()
```

[14]:

	TransactionID	id_01	id_02	id_03	id_04	\
count	144233.0	144233.0	140872.0	66324.0	66324.0	
mean	3236329.311288	NaN	174716.59375	0.0	-0.0	
std	178849.571186	0.0	159651.8125	0.0	0.0	
min	2987004.0	-100.0	1.0	-13.0	-28.0	
25%	3077142.0	-10.0	67992.0	0.0	0.0	
50%	3198818.0	-5.0	125800.5	0.0	0.0	

75%	3392923.0	-5.0	228749.0	0.0	0.0
max	3577534.0	0.0	999595.0	10.0	0.0
counts	144233	144233	140872	66324	66324
uniques	144233	77	115655	24	15
missing	0	0	3361	77909	77909
missing_perc	0%	0%	2.33%	54.02%	54.02%
types	numeric	numeric	numeric	numeric	numeric

	id_05	id_06	id_07	id_08	id_09	id_10
count	136865.0	136865.0	5155.0	5155.0	74926.0	74926.0
mean	NaN	NaN	inf	-inf	0.0	-0.0
std	0.0	0.0	11.382812	26.078125	0.0	0.0
min	-72.0	-100.0	-46.0	-100.0	-36.0	-100.0
25%	0.0	-6.0	5.0	-48.0	0.0	0.0
50%	0.0	0.0	14.0	-34.0	0.0	0.0
75%	1.0	0.0	22.0	-23.0	0.0	0.0
max	52.0	0.0	61.0	0.0	25.0	0.0
counts	136865	136865	5155	5155	74926	74926
uniques	93	101	84	94	46	62
missing	7368	7368	139078	139078	69307	69307
missing_perc	5.11%	5.11%	96.43%	96.43%	48.05%	48.05%
types	numeric	numeric	numeric	numeric	numeric	numeric

	id_11	id_12	id_13	id_14	id_15	id_16
count	140978.0	NaN	127320.0	80044.0	NaN	NaN
mean	NaN	NaN	NaN	NaN	NaN	NaN
std	0.0	NaN	0.0	NaN	NaN	NaN
min	90.0	NaN	10.0	-660.0	NaN	NaN
25%	100.0	NaN	49.0	-360.0	NaN	NaN
50%	100.0	NaN	52.0	-300.0	NaN	NaN
75%	100.0	NaN	52.0	-300.0	NaN	NaN
max	100.0	NaN	64.0	720.0	NaN	NaN
counts	140978	144233	127320	80044	140985	129340
uniques	146	2	54	25	3	2
missing	3255	0	16913	64189	3248	14893
missing_perc	2.26%	0%	11.73%	44.50%	2.25%	10.33%
types	numeric	bool	numeric	numeric	categorical	bool

	id_17	id_18	id_19	id_20	id_21	id_22
count	139369.0	45113.0	139318.0	139261.0	5159.0	5169.0
mean	NaN	inf	NaN	NaN	inf	inf
std	0.0	1.561523	NaN	NaN	inf	6.898438
min	100.0	10.0	100.0	100.0	100.0	10.0
25%	166.0	13.0	266.0	256.0	252.0	14.0
50%	166.0	15.0	341.0	472.0	252.0	14.0
75%	225.0	15.0	427.0	533.0	486.5	14.0
max	229.0	29.0	671.0	661.0	854.0	44.0

counts	139369	45113	139318	139261	5159	5169
uniques	104	18	522	394	490	25
missing	4864	99120	4915	4972	139074	139064
missing_perc	3.37%	68.72%	3.41%	3.45%	96.42%	96.42%
types	numeric	numeric	numeric	numeric	numeric	numeric

	id_23	id_24	id_25	id_26	id_27	id_28
count	NaN	4747.0	5132.0	5163.0	NaN	NaN
mean	NaN	12.789062	inf	inf	NaN	NaN
std	NaN	2.371094	97.4375	32.09375	NaN	NaN
min	NaN	11.0	100.0	100.0	NaN	NaN
25%	NaN	11.0	321.0	119.0	NaN	NaN
50%	NaN	11.0	321.0	149.0	NaN	NaN
75%	NaN	15.0	371.0	169.0	NaN	NaN
max	NaN	26.0	548.0	216.0	NaN	NaN
counts	5169	4747	5132	5163	5169	140978
uniques	3	12	341	95	2	2
missing	139064	139486	139101	139070	139064	3255
missing_perc	96.42%	96.71%	96.44%	96.42%	96.42%	2.26%
types	categorical	numeric	numeric	numeric	bool	bool

	id_29	id_30	id_31	id_32	id_33
count	NaN	NaN	NaN	77586.0	NaN
mean	NaN	NaN	NaN	NaN	NaN
std	NaN	NaN	NaN	0.0	NaN
min	NaN	NaN	NaN	0.0	NaN
25%	NaN	NaN	NaN	24.0	NaN
50%	NaN	NaN	NaN	24.0	NaN
75%	NaN	NaN	NaN	32.0	NaN
max	NaN	NaN	NaN	32.0	NaN
counts	140978	77565	140282	77586	73289
uniques	2	75	130	4	260
missing	3255	66668	3951	66647	70944
missing_perc	2.26%	46.22%	2.74%	46.21%	49.19%
types	bool	categorical	categorical	numeric	categorical

	id_34	id_35	id_36	id_37	id_38	DeviceType
count	NaN	NaN	NaN	NaN	NaN	NaN
mean	NaN	NaN	NaN	NaN	NaN	NaN
std	NaN	NaN	NaN	NaN	NaN	NaN
min	NaN	NaN	NaN	NaN	NaN	NaN
25%	NaN	NaN	NaN	NaN	NaN	NaN
50%	NaN	NaN	NaN	NaN	NaN	NaN
75%	NaN	NaN	NaN	NaN	NaN	NaN
max	NaN	NaN	NaN	NaN	NaN	NaN
counts	77805	140985	140985	140985	140985	140810
uniques	4	2	2	2	2	2

```
missing             66428    3248    3248    3248    3248       3423
missing_perc       46.06%   2.25%   2.25%   2.25%   2.25%      2.37%
types         categorical    bool    bool    bool    bool       bool

               DeviceInfo
count                 NaN
mean                  NaN
std                   NaN
min                   NaN
25%                   NaN
50%                   NaN
75%                   NaN
max                   NaN
counts             118666
uniques              1786
missing             25567
missing_perc       17.73%
types         categorical
```

[15]:
```
df_tran.describe()
```

[15]:
```
       TransactionID        isFraud  TransactionDT  TransactionAmt  \
count   5.905400e+05  590540.000000   5.905400e+05   590540.000000
mean    3.282270e+06       0.034990   7.372311e+06             NaN
std     1.704744e+05       0.183755   4.617224e+06             NaN
min     2.987000e+06       0.000000   8.640000e+04        0.250977
25%     3.134635e+06       0.000000   3.027058e+06       43.312500
50%     3.282270e+06       0.000000   7.306528e+06       68.750000
75%     3.429904e+06       0.000000   1.124662e+07      125.000000
max     3.577539e+06       1.000000   1.581113e+07    31936.000000

                card1      card2      card3      card5      addr1      addr2  \
count  590540.000000  581607.0  588975.0  586281.0  524834.0  524834.0
mean     9898.734658       NaN       NaN       NaN       NaN       NaN
std      4901.170153       NaN       0.0       0.0       NaN       0.0
min      1000.000000     100.0     100.0     100.0     100.0      10.0
25%      6019.000000     214.0     150.0     166.0     204.0      87.0
50%      9678.000000     361.0     150.0     226.0     299.0      87.0
75%     14184.000000     512.0     150.0     226.0     330.0      87.0
max     18396.000000     600.0     231.0     237.0     540.0     102.0

          dist1    dist2        C1        C2        C3        C4        C5  \
count  238269.0  37627.0  590540.0  590540.0  590540.0  590540.0  590540.0
mean        NaN      inf       NaN       NaN       0.0       NaN       NaN
std         NaN      inf       NaN       NaN       0.0       NaN       NaN
min         0.0      0.0       0.0       0.0       0.0       0.0       0.0
25%         3.0      7.0       1.0       1.0       0.0       0.0       0.0
```

```
50%         8.0     37.0       1.0       1.0       0.0       0.0       0.0
75%        24.0    206.0       3.0       3.0       0.0       0.0       1.0
max     10288.0  11624.0    4684.0    5692.0      26.0    2252.0     349.0

              C6        C7        C8        C9       C10       C11       C12  \
count   590540.0  590540.0  590540.0  590540.0  590540.0  590540.0  590540.0
mean         NaN       NaN       NaN       NaN       NaN       NaN       NaN
std          NaN       NaN       NaN       0.0       NaN       NaN       NaN
min          0.0       0.0       0.0       0.0       0.0       0.0       0.0
25%          1.0       0.0       0.0       0.0       0.0       1.0       0.0
50%          1.0       0.0       0.0       1.0       0.0       1.0       0.0
75%          2.0       0.0       0.0       2.0       0.0       2.0       0.0
max       2252.0    2256.0    3332.0     210.0    3256.0    3188.0    3188.0

              C13       C14        D1        D2        D3        D4        D5  \
count   590540.0  590540.0  589271.0  309743.0  327662.0  421618.0  280699.0
mean         NaN       NaN       NaN       NaN       NaN       NaN       NaN
std          NaN       NaN       NaN       NaN       NaN       NaN       NaN
min          0.0       0.0       0.0       0.0       0.0    -122.0       0.0
25%          1.0       1.0       0.0      26.0       1.0       0.0       1.0
50%          3.0       1.0       3.0      97.0       8.0      26.0      10.0
75%         12.0       2.0     122.0     276.0      27.0     253.0      32.0
max       2918.0    1429.0     640.0     640.0     819.0     869.0     819.0

             D6        D7            D8            D9        D10       D11  \
count   73187.0   38917.0  74926.000000  74926.000000  514518.0  311253.0
mean        NaN       inf           NaN      0.000000       NaN       NaN
std         NaN       inf           NaN      0.000000       NaN       NaN
min       -83.0       0.0      0.000000      0.000000       0.0     -53.0
25%         0.0       0.0      0.958496      0.208374       0.0       0.0
50%         0.0       0.0     37.875000      0.666504      15.0      43.0
75%        40.0      17.0    188.000000      0.833496     197.0     274.0
max       873.0     843.0   1708.000000      0.958496     876.0     670.0

             D12       D13       D14       D15        V1        V2        V3  \
count   64717.0   61952.0   62187.0  501427.0  311253.0  311253.0  311253.0
mean        inf       inf       inf       NaN       NaN       NaN       NaN
std         inf       inf       inf       NaN       0.0       0.0       0.0
min       -83.0       0.0    -193.0     -83.0       0.0       0.0       0.0
25%         0.0       0.0       0.0       0.0       1.0       1.0       1.0
50%         0.0       0.0       0.0      52.0       1.0       1.0       1.0
75%        13.0       0.0       2.0     314.0       1.0       1.0       1.0
max       648.0     847.0     878.0     879.0       1.0       8.0       9.0

              V4        V5        V6        V7        V8        V9       V10  \
count   311253.0  311253.0  311253.0  311253.0  311253.0  311253.0  311253.0
mean         NaN       NaN       NaN       NaN       NaN       NaN       NaN
```

std	0.0	0.0	0.0	0.0	0.0	0.0	0.0
min	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25%	1.0	1.0	1.0	1.0	1.0	1.0	0.0
50%	1.0	1.0	1.0	1.0	1.0	1.0	0.0
75%	1.0	1.0	1.0	1.0	1.0	1.0	1.0
max	6.0	6.0	9.0	9.0	8.0	8.0	4.0

	V11	V12	V13	V14	V15	V16	V17 \
count	311253.0	514467.0	514467.0	514467.0	514467.0	514467.0	514467.0
mean	NaN	NaN	NaN	NaN	0.0	0.0	NaN
std	0.0	0.0	0.0	0.0	0.0	0.0	0.0
min	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25%	0.0	0.0	0.0	1.0	0.0	0.0	0.0
50%	0.0	1.0	1.0	1.0	0.0	0.0	0.0
75%	1.0	1.0	1.0	1.0	0.0	0.0	0.0
max	5.0	3.0	6.0	1.0	7.0	15.0	15.0

	V18	V19	V20	V21	V22	V23	V24 \
count	514467.0	514467.0	514467.0	514467.0	514467.0	514467.0	514467.0
mean	NaN	NaN	NaN	NaN	NaN	NaN	NaN
std	0.0	0.0	0.0	0.0	0.0	0.0	0.0
min	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25%	0.0	1.0	1.0	0.0	0.0	1.0	1.0
50%	0.0	1.0	1.0	0.0	0.0	1.0	1.0
75%	0.0	1.0	1.0	0.0	0.0	1.0	1.0
max	15.0	7.0	15.0	5.0	8.0	13.0	13.0

	V25	V26	V27	V28	V29	V30	V31 \
count	514467.0	514467.0	514467.0	514467.0	514467.0	514467.0	514467.0
mean	NaN	NaN	0.0	0.0	NaN	NaN	NaN
std	0.0	0.0	0.0	0.0	0.0	0.0	0.0
min	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25%	1.0	1.0	0.0	0.0	0.0	0.0	0.0
50%	1.0	1.0	0.0	0.0	0.0	0.0	0.0
75%	1.0	1.0	0.0	0.0	1.0	1.0	0.0
max	7.0	13.0	4.0	4.0	5.0	9.0	7.0

	V32	V33	V34	V35	V36	V37	V38 \
count	514467.0	514467.0	514467.0	421571.0	421571.0	421571.0	421571.0
mean	NaN	NaN	NaN	NaN	NaN	NaN	NaN
std	0.0	0.0	0.0	0.0	0.0	0.0	0.0
min	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25%	0.0	0.0	0.0	0.0	0.0	1.0	1.0
50%	0.0	0.0	0.0	1.0	1.0	1.0	1.0
75%	0.0	0.0	0.0	1.0	1.0	1.0	1.0
max	15.0	7.0	13.0	3.0	5.0	54.0	54.0

```
            V39       V40       V41       V42       V43       V44       V45  \
count  421571.0  421571.0  421571.0  421571.0  421571.0  421571.0  421571.0
mean        NaN       NaN       NaN       NaN       NaN       NaN       NaN
std         0.0       0.0       0.0       0.0       0.0       0.0       0.0
min         0.0       0.0       0.0       0.0       0.0       0.0       0.0
25%         0.0       0.0       1.0       0.0       0.0       1.0       1.0
50%         0.0       0.0       1.0       0.0       0.0       1.0       1.0
75%         0.0       0.0       1.0       0.0       0.0       1.0       1.0
max        15.0      24.0       1.0       8.0       8.0      48.0      48.0

            V46       V47       V48       V49       V50       V51       V52  \
count  421571.0  421571.0  421571.0  421571.0  421571.0  421571.0  421571.0
mean        NaN       NaN       NaN       NaN       NaN       NaN       NaN
std         0.0       0.0       0.0       0.0       0.0       0.0       0.0
min         0.0       0.0       0.0       0.0       0.0       0.0       0.0
25%         1.0       1.0       0.0       0.0       0.0       0.0       0.0
50%         1.0       1.0       0.0       0.0       0.0       0.0       0.0
75%         1.0       1.0       1.0       1.0       0.0       0.0       0.0
max         6.0      12.0       5.0       5.0       5.0       6.0      12.0

            V53       V54       V55       V56       V57       V58       V59  \
count  513444.0  513444.0  513444.0  513444.0  513444.0  513444.0  513444.0
mean        NaN       NaN       NaN       NaN       NaN       NaN       NaN
std         0.0       0.0       0.0       0.0       0.0       0.0       0.0
min         0.0       0.0       0.0       0.0       0.0       0.0       0.0
25%         0.0       0.0       1.0       1.0       0.0       0.0       0.0
50%         1.0       1.0       1.0       1.0       0.0       0.0       0.0
75%         1.0       1.0       1.0       1.0       0.0       0.0       0.0
max         5.0       6.0      17.0      51.0       6.0      10.0      16.0

            V60       V61       V62       V63       V64       V65       V66  \
count  513444.0  513444.0  513444.0  513444.0  513444.0  513444.0  513444.0
mean        NaN       NaN       NaN       NaN       NaN       NaN       NaN
std         0.0       0.0       0.0       0.0       0.0       0.0       0.0
min         0.0       0.0       0.0       0.0       0.0       0.0       0.0
25%         0.0       1.0       1.0       0.0       0.0       1.0       1.0
50%         0.0       1.0       1.0       0.0       0.0       1.0       1.0
75%         0.0       1.0       1.0       0.0       0.0       1.0       1.0
max        16.0       6.0      10.0       7.0       7.0       1.0       7.0

            V67       V68       V69       V70       V71       V72       V73  \
count  513444.0  513444.0  513444.0  513444.0  513444.0  513444.0  513444.0
mean        NaN       0.0       NaN       NaN       NaN       NaN       NaN
std         0.0       0.0       0.0       0.0       0.0       0.0       0.0
min         0.0       0.0       0.0       0.0       0.0       0.0       0.0
25%         1.0       0.0       0.0       0.0       0.0       0.0       0.0
50%         1.0       0.0       0.0       0.0       0.0       0.0       0.0
```

```
75%         1.0       0.0       1.0       1.0       0.0       0.0       0.0
max         8.0       2.0       5.0       6.0       6.0      10.0       7.0

            V74       V75       V76       V77       V78       V79       V80  \
count  513444.0  501376.0  501376.0  501376.0  501376.0  501376.0  501376.0
mean        NaN       NaN       NaN       NaN       NaN       NaN       NaN
std         0.0       0.0       0.0       0.0       0.0       0.0       0.0
min         0.0       0.0       0.0       0.0       0.0       0.0       0.0
25%         0.0       0.0       0.0       1.0       1.0       0.0       0.0
50%         0.0       1.0       1.0       1.0       1.0       0.0       0.0
75%         0.0       1.0       1.0       1.0       1.0       0.0       0.0
max         8.0       4.0       6.0      30.0      31.0       7.0      19.0

            V81       V82       V83       V84       V85       V86       V87  \
count  501376.0  501376.0  501376.0  501376.0  501376.0  501376.0  501376.0
mean        NaN       NaN       NaN       NaN       NaN       NaN       NaN
std         0.0       0.0       0.0       0.0       0.0       0.0       0.0
min         0.0       0.0       0.0       0.0       0.0       0.0       0.0
25%         0.0       1.0       1.0       0.0       0.0       1.0       1.0
50%         0.0       1.0       1.0       0.0       0.0       1.0       1.0
75%         0.0       1.0       1.0       0.0       0.0       1.0       1.0
max        19.0       7.0       7.0       7.0       7.0      30.0      30.0

            V88       V89       V90       V91       V92       V93       V94  \
count  501376.0  501376.0  501376.0  501376.0  501376.0  501376.0  501376.0
mean        NaN       0.0       NaN       NaN       NaN       NaN       NaN
std         0.0       0.0       0.0       0.0       0.0       0.0       0.0
min         0.0       0.0       0.0       0.0       0.0       0.0       0.0
25%         1.0       0.0       0.0       0.0       0.0       0.0       0.0
50%         1.0       0.0       0.0       0.0       0.0       0.0       0.0
75%         1.0       0.0       1.0       1.0       0.0       0.0       0.0
max         1.0       2.0       5.0       6.0       7.0       7.0       2.0

            V95       V96       V97       V98       V99      V100      V101  \
count  590226.0  590226.0  590226.0  590226.0  590226.0  590226.0  590226.0
mean        NaN       NaN       NaN       0.0       NaN       NaN       NaN
std         NaN       NaN       NaN       0.0       0.0       0.0       NaN
min         0.0       0.0       0.0       0.0       0.0       0.0       0.0
25%         0.0       0.0       0.0       0.0       0.0       0.0       0.0
50%         0.0       0.0       0.0       0.0       0.0       0.0       0.0
75%         0.0       1.0       0.0       0.0       1.0       0.0       0.0
max       880.0    1410.0     976.0      12.0      88.0      28.0     869.0

           V102      V103      V104      V105      V106      V107      V108  \
count  590226.0  590226.0  590226.0  590226.0  590226.0  590226.0  590226.0
mean        NaN       NaN       0.0       NaN       NaN       NaN       NaN
std         NaN       NaN       0.0       0.0       0.0       0.0       0.0
```

```
min         0.0       0.0       0.0       0.0       0.0       0.0       0.0
25%         0.0       0.0       0.0       0.0       0.0       1.0       1.0
50%         0.0       0.0       0.0       0.0       0.0       1.0       1.0
75%         0.0       0.0       0.0       0.0       0.0       1.0       1.0
max      1285.0     928.0      15.0      99.0      55.0       1.0       7.0

           V109      V110      V111      V112      V113      V114      V115  \
count  590226.0  590226.0  590226.0  590226.0  590226.0  590226.0  590226.0
mean        NaN       NaN       NaN       NaN       NaN       NaN       NaN
std         0.0       0.0       0.0       0.0       0.0       0.0       0.0
min         0.0       0.0       0.0       0.0       0.0       0.0       0.0
25%         1.0       1.0       1.0       1.0       1.0       1.0       1.0
50%         1.0       1.0       1.0       1.0       1.0       1.0       1.0
75%         1.0       1.0       1.0       1.0       1.0       1.0       1.0
max         7.0       7.0       9.0       9.0       9.0       6.0       6.0

           V116      V117      V118      V119      V120      V121      V122  \
count  590226.0  590226.0  590226.0  590226.0  590226.0  590226.0  590226.0
mean        NaN       NaN       NaN       NaN       NaN       NaN       NaN
std         0.0       0.0       0.0       0.0       0.0       0.0       0.0
min         0.0       0.0       0.0       0.0       0.0       0.0       0.0
25%         1.0       1.0       1.0       1.0       1.0       1.0       1.0
50%         1.0       1.0       1.0       1.0       1.0       1.0       1.0
75%         1.0       1.0       1.0       1.0       1.0       1.0       1.0
max         6.0       3.0       3.0       3.0       3.0       3.0       3.0

           V123      V124      V125           V126           V127  \
count  590226.0  590226.0  590226.0  590226.000000  590226.000000
mean        NaN       NaN       NaN     129.979431     336.611603
std         0.0       0.0       0.0    2346.951660    4238.666992
min         0.0       0.0       0.0       0.000000       0.000000
25%         1.0       1.0       1.0       0.000000       0.000000
50%         1.0       1.0       1.0       0.000000       0.000000
75%         1.0       1.0       1.0       0.000000     107.949997
max        13.0      13.0      13.0  160000.000000  160000.000000

              V128      V129      V130      V131           V132  \
count  590226.000000  590226.0  590226.0  590226.0  590226.000000
mean      204.094025       NaN       NaN       NaN     103.513191
std      3010.258789       NaN       NaN       NaN    2266.106201
min         0.000000       0.0       0.0       0.0       0.000000
25%         0.000000       0.0       0.0       0.0       0.000000
50%         0.000000       0.0       0.0       0.0       0.000000
75%         0.000000       0.0      59.0       0.0       0.000000
max    160000.000000   55136.0   55136.0   55136.0   93736.000000

              V133           V134           V135           V136  \
```

```
count  590226.00000  590226.000000  590226.000000  590226.000000
mean      204.88916     145.972336      17.250130      38.821198
std      3796.31665    2772.986816     293.847565     451.808411
min         0.00000       0.000000       0.000000       0.000000
25%         0.00000       0.000000       0.000000       0.000000
50%         0.00000       0.000000       0.000000       0.000000
75%         0.00000       0.000000       0.000000       0.000000
max    133915.00000   98476.000000   90750.000000   90750.000000

                V137     V138     V139     V140     V141     V142     V143  \
count  590226.000000  81945.0  81945.0  81945.0  81945.0  81945.0  81951.0
mean       26.365089      0.0      NaN      NaN      0.0      0.0      NaN
std       348.332703      0.0      0.0      0.0      0.0      0.0      NaN
min         0.000000      0.0      0.0      0.0      0.0      0.0      0.0
25%         0.000000      0.0      0.0      0.0      0.0      0.0      0.0
50%         0.000000      0.0      1.0      1.0      0.0      0.0      0.0
75%         0.000000      0.0      1.0      1.0      0.0      0.0      0.0
max     90750.000000     22.0     33.0     33.0      5.0      9.0    869.0

          V144     V145     V146     V147     V148     V149     V150     V151  \
count  81951.0  81951.0  81945.0  81945.0  81945.0  81945.0  81951.0  81951.0
mean       NaN      NaN      0.0      0.0      0.0      0.0      NaN      NaN
std        0.0      NaN      0.0      0.0      0.0      0.0      NaN      0.0
min        0.0      0.0      0.0      0.0      0.0      0.0      1.0      1.0
25%        0.0      0.0      0.0      0.0      0.0      0.0      1.0      1.0
50%        0.0      0.0      0.0      0.0      1.0      1.0      1.0      1.0
75%        0.0      1.0      0.0      0.0      1.0      1.0      1.0      1.0
max       62.0    297.0     24.0     26.0     20.0     20.0   3388.0     57.0

          V152     V153     V154     V155     V156     V157     V158     V159  \
count  81951.0  81945.0  81945.0  81945.0  81945.0  81945.0  81945.0  81951.0
mean       NaN      0.0      0.0      0.0      0.0      NaN      NaN      NaN
std        0.0      0.0      0.0      0.0      0.0      0.0      0.0      NaN
min        1.0      0.0      0.0      0.0      0.0      0.0      0.0      0.0
25%        1.0      0.0      0.0      0.0      0.0      0.0      0.0      0.0
50%        1.0      1.0      1.0      1.0      1.0      1.0      1.0      0.0
75%        1.0      1.0      1.0      1.0      1.0      1.0      1.0      0.0
max       69.0     18.0     18.0     24.0     24.0     24.0     24.0  55136.0

                V160     V161     V162     V163          V164          V165  \
count   81951.000000  81945.0  81945.0  81945.0  81951.000000  81951.000000
mean    47453.191406      NaN      NaN      NaN    877.888794   2239.911865
std    142076.062500      NaN      NaN      NaN   6049.166016   8223.258789
min         0.000000      0.0      0.0      0.0      0.000000      0.000000
25%         0.000000      0.0      0.0      0.0      0.000000      0.000000
50%         0.000000      0.0      0.0      0.0      0.000000      0.000000
75%         0.000000      0.0      0.0      0.0      0.000000      0.000000
```

```
max    641511.437500   3300.0   3300.0   3300.0  93736.000000  98476.000000

                V166      V167      V168      V169      V170      V171  \
count   81951.000000  139631.0  139631.0  139819.0  139819.0  139819.0
mean      359.469391       NaN       NaN       0.0       NaN       NaN
std      1244.463257       NaN       NaN       0.0       0.0       0.0
min         0.000000       0.0       0.0       0.0       0.0       0.0
25%         0.000000       0.0       0.0       0.0       1.0       1.0
50%         0.000000       0.0       0.0       0.0       1.0       1.0
75%         0.000000       0.0       1.0       0.0       1.0       1.0
max    104060.000000     872.0     964.0      19.0      48.0      61.0

           V172      V173      V174      V175      V176      V177      V178  \
count  139631.0  139631.0  139819.0  139819.0  139631.0  139631.0  139631.0
mean        0.0       0.0       0.0       0.0       NaN       NaN       NaN
std         0.0       0.0       0.0       0.0       0.0       NaN       NaN
min         0.0       0.0       0.0       0.0       0.0       0.0       0.0
25%         0.0       0.0       0.0       0.0       1.0       0.0       0.0
50%         0.0       0.0       0.0       0.0       1.0       0.0       0.0
75%         0.0       0.0       0.0       0.0       1.0       0.0       0.0
max        31.0       7.0       8.0      14.0      48.0     861.0    1235.0

           V179      V180      V181      V182      V183      V184      V185  \
count  139631.0  139819.0  139631.0  139631.0  139631.0  139819.0  139819.0
mean        NaN       NaN       0.0       NaN       NaN       0.0       0.0
std         NaN       0.0       0.0       0.0       0.0       0.0       0.0
min         0.0       0.0       0.0       0.0       0.0       0.0       0.0
25%         0.0       0.0       0.0       0.0       0.0       0.0       0.0
50%         0.0       0.0       0.0       0.0       0.0       0.0       0.0
75%         0.0       0.0       0.0       0.0       0.0       0.0       0.0
max       920.0      83.0      24.0      83.0      41.0      16.0      31.0

           V186      V187      V188      V189      V190      V191      V192  \
count  139631.0  139631.0  139819.0  139819.0  139631.0  139631.0  139631.0
mean        NaN       NaN       NaN       NaN       NaN       NaN       NaN
std         0.0       0.0       0.0       0.0       0.0       0.0       0.0
min         0.0       0.0       0.0       0.0       0.0       0.0       0.0
25%         1.0       1.0       1.0       1.0       1.0       1.0       1.0
50%         1.0       1.0       1.0       1.0       1.0       1.0       1.0
75%         1.0       1.0       1.0       1.0       1.0       1.0       1.0
max        38.0     218.0      30.0      30.0      42.0      21.0      44.0

           V193      V194      V195      V196      V197      V198      V199  \
count  139631.0  139819.0  139819.0  139631.0  139819.0  139819.0  139631.0
mean        NaN       NaN       NaN       NaN       NaN       NaN       NaN
std         0.0       0.0       0.0       0.0       0.0       0.0       0.0
min         0.0       0.0       0.0       0.0       0.0       0.0       0.0
```

```
25%        1.0       1.0       1.0       1.0       1.0       1.0       1.0
50%        1.0       1.0       1.0       1.0       1.0       1.0       1.0
75%        1.0       1.0       1.0       1.0       1.0       1.0       1.0
max       37.0       7.0      16.0      38.0      14.0      21.0      45.0

           V200      V201           V202           V203           V204  \
count  139819.0  139819.0  139631.000000  139631.000000  139631.000000
mean        NaN       NaN     444.147125    1078.327515     686.956909
std         0.0       0.0    4683.828125    9105.608398    6048.980957
min         0.0       0.0       0.000000       0.000000       0.000000
25%         1.0       1.0       0.000000       0.000000       0.000000
50%         1.0       1.0       0.000000       0.000000       0.000000
75%         1.0       1.0       0.000000      30.924400      20.000000
max        45.0      55.0  104060.000000  139777.000000  104060.000000

           V205      V206      V207      V208      V209      V210  \
count  139631.0  139631.0  139631.0  139819.0  139819.0  139819.0
mean        NaN       NaN       NaN       NaN       NaN       NaN
std         NaN       NaN       NaN       NaN       NaN       NaN
min         0.0       0.0       0.0       0.0       0.0       0.0
25%         0.0       0.0       0.0       0.0       0.0       0.0
50%         0.0       0.0       0.0       0.0       0.0       0.0
75%         0.0       0.0       0.0       0.0       0.0       0.0
max     55136.0   55136.0   55136.0    3300.0    8048.0    3300.0

                V211           V212           V213          V214  \
count  139631.000000  139631.000000  139631.000000  139631.00000
mean      385.137024     765.988464     536.302734      38.43755
std      4541.837891    7496.120605    5471.664551     571.83429
min         0.000000       0.000000       0.000000       0.00000
25%         0.000000       0.000000       0.000000       0.00000
50%         0.000000       0.000000       0.000000       0.00000
75%         0.000000       0.000000       0.000000       0.00000
max     92888.000000  129006.000000   97628.000000  104060.00000

                V215           V216      V217      V218      V219      V220  \
count  139631.000000  139631.000000  130430.0  130430.0  130430.0  141416.0
mean      133.208221      71.107147       NaN       NaN       NaN       0.0
std      1040.453735     680.267639       NaN       NaN       NaN       0.0
min         0.000000       0.000000       0.0       0.0       0.0       0.0
25%         0.000000       0.000000       0.0       0.0       0.0       0.0
50%         0.000000       0.000000       0.0       0.0       0.0       0.0
75%         0.000000       0.000000       0.0       1.0       1.0       0.0
max    104060.000000  104060.000000     303.0     400.0     378.0      25.0

           V221      V222      V223      V224      V225      V226      V227  \
count  141416.0  141416.0  130430.0  130430.0  130430.0  130430.0  141416.0
```

mean	NaN	NaN	0.0	0.0	0.0	0.0	0.0
std	NaN	NaN	0.0	0.0	0.0	0.0	NaN
min	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25%	1.0	1.0	0.0	0.0	0.0	0.0	0.0
50%	1.0	1.0	0.0	0.0	0.0	0.0	0.0
75%	1.0	1.0	0.0	0.0	0.0	0.0	0.0
max	384.0	384.0	16.0	144.0	51.0	242.0	360.0

	V228	V229	V230	V231	V232	V233	V234 \
count	130430.0	130430.0	130430.0	130430.0	130430.0	130430.0	141416.0
mean	NaN	NaN	NaN	NaN	NaN	NaN	NaN
std	0.0	0.0	0.0	NaN	NaN	NaN	0.0
min	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25%	1.0	1.0	1.0	0.0	0.0	0.0	0.0
50%	1.0	1.0	1.0	0.0	0.0	0.0	0.0
75%	1.0	1.0	1.0	0.0	0.0	0.0	0.0
max	54.0	176.0	65.0	293.0	337.0	332.0	121.0

	V235	V236	V237	V238	V239	V240	V241 \
count	130430.0	130430.0	130430.0	141416.0	141416.0	130430.0	130430.0
mean	0.0	0.0	0.0	0.0	0.0	NaN	NaN
std	0.0	0.0	0.0	0.0	0.0	0.0	0.0
min	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25%	0.0	0.0	0.0	0.0	0.0	1.0	1.0
50%	0.0	0.0	0.0	0.0	0.0	1.0	1.0
75%	0.0	0.0	0.0	0.0	0.0	1.0	1.0
max	23.0	45.0	39.0	23.0	23.0	7.0	5.0

	V242	V243	V244	V245	V246	V247	V248 \
count	130430.0	130430.0	130430.0	141416.0	130430.0	130430.0	130430.0
mean	NaN	NaN	NaN	NaN	NaN	NaN	NaN
std	0.0	0.0	0.0	NaN	0.0	0.0	0.0
min	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25%	1.0	1.0	1.0	1.0	1.0	1.0	1.0
50%	1.0	1.0	1.0	1.0	1.0	1.0	1.0
75%	1.0	1.0	1.0	1.0	1.0	1.0	1.0
max	20.0	57.0	22.0	262.0	45.0	18.0	36.0

	V249	V250	V251	V252	V253	V254	V255 \
count	130430.0	141416.0	141416.0	130430.0	130430.0	130430.0	141416.0
mean	NaN	NaN	NaN	NaN	NaN	NaN	NaN
std	0.0	0.0	0.0	0.0	0.0	0.0	0.0
min	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25%	1.0	1.0	1.0	1.0	1.0	1.0	1.0
50%	1.0	1.0	1.0	1.0	1.0	1.0	1.0
75%	1.0	1.0	1.0	1.0	1.0	1.0	1.0
max	22.0	18.0	18.0	24.0	163.0	60.0	87.0

```
              V256        V257        V258        V259        V260        V261        V262  \
count   141416.0    130430.0    130430.0    141416.0    130430.0    130430.0    130430.0
mean         NaN         NaN         NaN         NaN         NaN         NaN         NaN
std          0.0         0.0         0.0         NaN         0.0         0.0         0.0
min          0.0         0.0         0.0         0.0         0.0         0.0         0.0
25%          1.0         1.0         1.0         1.0         1.0         1.0         1.0
50%          1.0         1.0         1.0         1.0         1.0         1.0         1.0
75%          1.0         1.0         1.0         1.0         1.0         1.0         1.0
max         87.0        48.0        66.0        285.0         8.0        49.0        20.0

                 V263           V264           V265        V266        V267  \
count  130430.000000  130430.000000  130430.000000  130430.0  130430.0
mean      117.390648     201.657608     153.520554         NaN         NaN
std      1294.851562    2284.827637    1605.512329         NaN         NaN
min         0.000000       0.000000       0.000000         0.0         0.0
25%         0.000000       0.000000       0.000000         0.0         0.0
50%         0.000000       0.000000       0.000000         0.0         0.0
75%         0.000000      33.593498      20.897526         0.0         0.0
max    153600.000000  153600.000000  153600.000000   55136.0   55136.0

              V268        V269        V270        V271        V272        V273  \
count   130430.0    130430.0    141416.0    141416.0    141416.0    130430.0
mean         NaN         NaN         NaN         NaN         NaN         NaN
std          NaN         NaN         NaN         NaN         NaN         NaN
min          0.0         0.0         0.0         0.0         0.0         0.0
25%          0.0         0.0         0.0         0.0         0.0         0.0
50%          0.0         0.0         0.0         0.0         0.0         0.0
75%          0.0         0.0         0.0         0.0         0.0         0.0
max      55136.0     55136.0      4000.0      4000.0      4000.0     51200.0

                 V274        V275           V276           V277           V278  \
count  130430.000000   130430.0  130430.000000  130430.000000  130430.000000
mean      107.151642         NaN      31.797276      51.956646      42.328224
std      1258.734131         NaN     615.659729     732.145386     660.611816
min         0.000000         0.0       0.000000       0.000000       0.000000
25%         0.000000         0.0       0.000000       0.000000       0.000000
50%         0.000000         0.0       0.000000       0.000000       0.000000
75%         0.000000         0.0       0.000000       0.000000       0.000000
max     66000.000000     51200.0  104060.000000  104060.000000  104060.000000

              V279        V280        V281        V282        V283        V284        V285  \
count   590528.0    590528.0    589271.0    589271.0    589271.0    590528.0    590528.0
mean         NaN         NaN         0.0         NaN         NaN         0.0         NaN
std          NaN         NaN         0.0         0.0         0.0         0.0         0.0
min          0.0         0.0         0.0         0.0         0.0         0.0         0.0
25%          0.0         0.0         0.0         0.0         0.0         0.0         0.0
```

```
50%         0.0       0.0       0.0       1.0       1.0       0.0       0.0
75%         0.0       1.0       0.0       1.0       1.0       0.0       1.0
max       880.0     975.0      22.0      32.0      68.0      12.0      95.0

            V286      V287      V288      V289      V290      V291      V292  \
count   590528.0  590528.0  589271.0  589271.0  590528.0  590528.0  590528.0
mean         0.0       NaN       NaN       NaN       NaN       NaN       NaN
std          0.0       0.0       0.0       0.0       0.0       NaN       NaN
min          0.0       0.0       0.0       0.0       1.0       1.0       1.0
25%          0.0       0.0       0.0       0.0       1.0       1.0       1.0
50%          0.0       0.0       0.0       0.0       1.0       1.0       1.0
75%          0.0       0.0       0.0       0.0       1.0       1.0       1.0
max          8.0      31.0      10.0      12.0      67.0    1055.0     323.0

            V293      V294      V295      V296      V297      V298      V299  \
count   590528.0  590528.0  590528.0  589271.0  590528.0  590528.0  590528.0
mean         NaN       NaN       NaN       NaN       0.0       NaN       NaN
std          NaN       NaN       NaN       0.0       0.0       0.0       0.0
min          0.0       0.0       0.0       0.0       0.0       0.0       0.0
25%          0.0       0.0       0.0       0.0       0.0       0.0       0.0
50%          0.0       0.0       0.0       0.0       0.0       0.0       0.0
75%          0.0       0.0       0.0       0.0       0.0       0.0       0.0
max        869.0    1286.0     928.0      93.0      12.0      93.0      49.0

            V300      V301      V302      V303      V304      V305  \
count   589271.0  589271.0  590528.0  590528.0  590528.0  590528.0
mean         0.0       0.0       NaN       NaN       NaN       NaN
std          0.0       0.0       0.0       0.0       0.0       0.0
min          0.0       0.0       0.0       0.0       0.0       1.0
25%          0.0       0.0       0.0       0.0       0.0       1.0
50%          0.0       0.0       0.0       0.0       0.0       1.0
75%          0.0       0.0       0.0       0.0       0.0       1.0
max         11.0      13.0      16.0      20.0      16.0       2.0

                V306           V307           V308      V309          V310  \
count   590528.000000  590528.000000  590528.000000  590528.0  590528.0000
mean       139.748718     408.682404     230.413162       NaN           NaN
std       2348.849609    4391.993164    3021.924316       NaN           NaN
min          0.000000       0.000000       0.000000       0.0       0.0000
25%          0.000000       0.000000       0.000000       0.0       0.0000
50%          0.000000       0.000000       0.000000       0.0       0.0000
75%          0.000000     151.380676      35.970001       0.0     107.9375
max     108800.000000  145765.000000  108800.000000   55136.0   55136.0000

            V311      V312      V313      V314      V315           V316  \
count   590528.0  590528.0  589271.0  589271.0  589271.0  590528.000000
mean         NaN       NaN       NaN       NaN       NaN     109.818542
```

```
std         NaN       NaN       NaN       NaN       NaN    2270.033203
min         0.0       0.0       0.0       0.0       0.0       0.000000
25%         0.0       0.0       0.0       0.0       0.0       0.000000
50%         0.0       0.0       0.0       0.0       0.0       0.000000
75%         0.0       0.0       0.0       0.0       0.0       0.000000
max     55136.0   55136.0    4816.0    7520.0    4816.0   93736.000000

                V317           V318           V319           V320  \
count   590528.000000  590528.000000  590528.000000  590528.000000
mean       247.606750     162.153412      18.372477      42.073135
std       3980.042725    2793.343506     332.304840     473.499298
min          0.000000       0.000000       0.000000       0.000000
25%          0.000000       0.000000       0.000000       0.000000
50%          0.000000       0.000000       0.000000       0.000000
75%          0.000000       0.000000       0.000000       0.000000
max     134021.000000   98476.000000  104060.000000  104060.000000

                V321     V322     V323     V324     V325     V326     V327  \
count   590528.000000  82351.0  82351.0  82351.0  82351.0  82351.0  82351.0
mean        28.326582      NaN      NaN      NaN      0.0      NaN      0.0
std        382.053162      NaN      NaN      NaN      0.0      0.0      0.0
min          0.000000      0.0      0.0      0.0      0.0      0.0      0.0
25%          0.000000      0.0      0.0      0.0      0.0      0.0      0.0
50%          0.000000      0.0      0.0      0.0      0.0      0.0      0.0
75%          0.000000      0.0      1.0      0.0      0.0      0.0      0.0
max     104060.000000    880.0   1411.0    976.0     12.0     44.0     18.0

          V328     V329     V330           V331           V332           V333  \
count  82351.0  82351.0  82351.0   82351.000000   82351.000000   82351.000000
mean       0.0      NaN      0.0     721.741882    1375.783691    1014.622803
std        0.0      0.0      0.0    6217.223633   11169.276367    7955.735352
min        0.0      0.0      0.0       0.000000       0.000000       0.000000
25%        0.0      0.0      0.0       0.000000       0.000000       0.000000
50%        0.0      0.0      0.0       0.000000       0.000000       0.000000
75%        0.0      0.0      0.0       0.000000      25.000000       0.000000
max       15.0     99.0     55.0  160000.000000  160000.000000  160000.000000

          V334     V335     V336           V337           V338           V339
count  82351.0  82351.0  82351.0   82351.000000   82351.000000   82351.000000
mean       NaN      NaN      NaN      55.352417     151.160538     100.700874
std        NaN      NaN      NaN     668.486816    1095.034302     814.946716
min        0.0      0.0      0.0       0.000000       0.000000       0.000000
25%        0.0      0.0      0.0       0.000000       0.000000       0.000000
50%        0.0      0.0      0.0       0.000000       0.000000       0.000000
75%        0.0      0.0      0.0       0.000000       0.000000       0.000000
max    55136.0  55136.0  55136.0  104060.000000  104060.000000  104060.000000
```

Check class imbalance

[16]:
```
df_tran.loc[:, 'isFraud'].value_counts()
```

[16]:
```
0    569877
1     20663
Name: isFraud, dtype: int64
```

[17]:
```
df_tran.loc[:, 'isFraud'].value_counts(normalize=True)*100
```

[17]:
```
0    96.500999
1     3.499001
Name: isFraud, dtype: float64
```

Lot of interesting things can be observed here:

- Rows in identity dataset are less than transaction dataset, that means only a subset of transactions in transactions dataset has identity data
- Both datasets have the common and unique key as TransactionID, both can be joined at this unique key
- id_24, id_25, dist2, D7 and many more columns have 90%+ missing values, which means that these columns are probably useless so need to drop it for now
- Columns from V1 to V339 in transaction dataset are numeric whereas columns from id_01 to id_39 are of mixed datatype
- TransactionDT column is a timedelta from a given reference datetime (not an actual timestamp). But reference datetime is not known, so need to assume it and convert it to date format
- Target class is imbalanced. So no need to drop the columns where one category contains the majority of rows

5 5. Data Preprocessing for EDA

Merge the datasets

[18]:
```
df = df_tran.merge(df_id, how='left', on='TransactionID')

del df_tran, df_id

gc.collect()
```

[18]:
```
0
```

[19]:
```
df.shape
```

[19]:
```
(590540, 434)
```

Add missing flag

[20]:
```
for col in df.columns:
    df[col+"_missing_flag"] = df[col].isnull()

df.head()
```

[20]:

	TransactionID	isFraud	TransactionDT	TransactionAmt	ProductCD	card1	\
0	2987000	0	86400	68.5	W	13926	
1	2987001	0	86401	29.0	W	2755	
2	2987002	0	86469	59.0	W	4663	
3	2987003	0	86499	50.0	W	18132	
4	2987004	0	86506	50.0	H	4497	

	card2	card3	card4	card5	card6	addr1	addr2	dist1	dist2	\
0	NaN	150.0	discover	142.0	credit	315.0	87.0	19.0	NaN	
1	404.0	150.0	mastercard	102.0	credit	325.0	87.0	NaN	NaN	
2	490.0	150.0	visa	166.0	debit	330.0	87.0	287.0	NaN	
3	567.0	150.0	mastercard	117.0	debit	476.0	87.0	NaN	NaN	
4	514.0	150.0	mastercard	102.0	credit	420.0	87.0	NaN	NaN	

	P_emaildomain	R_emaildomain	C1	C2	C3	C4	C5	C6	C7	C8	C9	\
0	NaN	NaN	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	
1	gmail.com	NaN	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	
2	outlook.com	NaN	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	
3	yahoo.com	NaN	2.0	5.0	0.0	0.0	0.0	4.0	0.0	0.0	1.0	
4	gmail.com	NaN	1.0	1.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	

	C10	C11	C12	C13	C14	D1	D2	D3	D4	D5	D6	D7	D8	D9	\
0	0.0	2.0	0.0	1.0	1.0	14.0	NaN	13.0	NaN	NaN	NaN	NaN	NaN	NaN	
1	0.0	1.0	0.0	1.0	1.0	0.0	NaN	NaN	0.0	NaN	NaN	NaN	NaN	NaN	
2	0.0	1.0	0.0	1.0	1.0	0.0	NaN	NaN	0.0	NaN	NaN	NaN	NaN	NaN	
3	0.0	1.0	0.0	25.0	1.0	112.0	112.0	0.0	94.0	0.0	NaN	NaN	NaN	NaN	
4	1.0	1.0	0.0	1.0	1.0	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

	D10	D11	D12	D13	D14	D15	M1	M2	M3	M4	M5	M6	M7	M8	\
0	13.0	13.0	NaN	NaN	NaN	0.0	T	T	T	M2	F	T	NaN	NaN	
1	0.0	NaN	NaN	NaN	NaN	0.0	NaN	NaN	NaN	M0	T	T	NaN	NaN	
2	0.0	315.0	NaN	NaN	NaN	315.0	T	T	T	M0	F	F	F	F	
3	84.0	NaN	NaN	NaN	NaN	111.0	NaN	NaN	NaN	M0	T	F	NaN	NaN	
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

	M9	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	\
0	NaN	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	1.0	1.0	1.0	
1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.0	0.0	1.0	
2	F	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	1.0	1.0	1.0	
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	1.0	1.0	
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

```
   V15  V16  V17  V18  V19  V20  V21  V22  V23  V24  V25  V26  V27  V28  V29  \
0  0.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  0.0
1  0.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  0.0
2  0.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  0.0
3  0.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  0.0
4  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

   V30  V31  V32  V33  V34  V35  V36  V37  V38  V39  V40  V41  V42  V43  V44  \
0  0.0  0.0  0.0  0.0  0.0  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN
1  0.0  0.0  0.0  0.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  1.0  0.0  0.0  1.0
2  0.0  0.0  0.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  1.0  0.0  0.0  1.0
3  0.0  0.0  0.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  1.0  0.0  0.0  1.0
4  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

   V45  V46  V47  V48  V49  V50  V51  V52  V53  V54  V55  V56  V57  V58  V59  \
0  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  1.0  1.0  1.0  1.0  0.0  0.0  0.0
1  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  0.0
2  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  0.0
3  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  0.0
4  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

   V60  V61  V62  V63  V64  V65  V66  V67  V68  V69  V70  V71  V72  V73  V74  \
0  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
1  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
2  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
3  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
4  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

   V75  V76  V77  V78  V79  V80  V81  V82  V83  V84  V85  V86  V87  V88  V89  \
0  1.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  1.0  1.0  1.0  0.0
1  0.0  0.0  1.0  1.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  0.0
2  1.0  1.0  1.0  1.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  0.0
3  1.0  1.0  1.0  1.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  0.0
4  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

   V90  V91  V92  V93  V94  V95   V96   V97  V98   V99  V100  V101  V102  \
0  0.0  0.0  0.0  0.0  0.0  0.0   1.0   0.0  0.0   0.0   0.0   0.0   1.0
1  0.0  0.0  0.0  0.0  0.0  0.0   0.0   0.0  0.0   0.0   0.0   0.0   0.0
2  0.0  0.0  0.0  0.0  0.0  0.0   0.0   0.0  0.0   0.0   0.0   0.0   0.0
3  0.0  0.0  0.0  0.0  0.0  1.0  48.0  28.0  0.0  10.0   4.0   1.0  38.0
4  NaN  NaN  NaN  NaN  NaN  0.0   0.0   0.0  0.0   0.0   0.0   0.0   0.0

   V103  V104  V105  V106  V107  V108  V109  V110  V111  V112  V113  V114  \
0   0.0   0.0   0.0   0.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0
1   0.0   0.0   0.0   0.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0
2   0.0   0.0   0.0   0.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0
3  24.0   0.0   0.0   0.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0
```

```
4   0.0   0.0   0.0   0.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0

    V115  V116  V117  V118  V119  V120  V121  V122  V123  V124  V125  V126  \
0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   0.0
1   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   0.0
2   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   0.0
3   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0  50.0
4   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   0.0

     V127   V128  V129   V130   V131  V132    V133   V134  V135  V136  V137  \
0   117.0    0.0   0.0    0.0    0.0   0.0   117.0    0.0   0.0   0.0   0.0
1     0.0    0.0   0.0    0.0    0.0   0.0     0.0    0.0   0.0   0.0   0.0
2     0.0    0.0   0.0    0.0    0.0   0.0     0.0    0.0   0.0   0.0   0.0
3  1758.0  925.0   0.0  354.0  135.0  50.0  1404.0  790.0   0.0   0.0   0.0
4     0.0    0.0   0.0    0.0    0.0   0.0     0.0    0.0   0.0   0.0   0.0

    V138  V139  V140  V141  V142  V143  V144   V145  V146  V147  V148  V149  \
0   NaN   NaN   NaN   NaN   NaN   NaN   NaN    NaN   NaN   NaN   NaN   NaN
1   NaN   NaN   NaN   NaN   NaN   NaN   NaN    NaN   NaN   NaN   NaN   NaN
2   NaN   NaN   NaN   NaN   NaN   NaN   NaN    NaN   NaN   NaN   NaN   NaN
3   NaN   NaN   NaN   NaN   NaN   NaN   NaN    NaN   NaN   NaN   NaN   NaN
4   0.0   0.0   0.0   0.0   0.0   6.0  18.0  140.0   0.0   0.0   0.0   0.0

     V150  V151  V152  V153  V154  V155  V156  V157  V158     V159  \
0     NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN      NaN
1     NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN      NaN
2     NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN      NaN
3     NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN      NaN
4  1803.0  49.0  64.0   0.0   0.0   0.0   0.0   0.0   0.0  15560.0

            V160  V161  V162  V163   V164    V165    V166  V167  V168  V169  \
0            NaN   NaN   NaN   NaN    NaN     NaN     NaN   NaN   NaN   NaN
1            NaN   NaN   NaN   NaN    NaN     NaN     NaN   NaN   NaN   NaN
2            NaN   NaN   NaN   NaN    NaN     NaN     NaN   NaN   NaN   NaN
3            NaN   NaN   NaN   NaN    NaN     NaN     NaN   NaN   NaN   NaN
4  169690.796875   0.0   0.0   0.0  515.0  5155.0  2840.0   0.0   0.0   0.0

    V170  V171  V172  V173  V174  V175  V176  V177  V178  V179  V180  V181  \
0   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
1   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
2   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
3   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
4   1.0   1.0   0.0   0.0   0.0   0.0   1.0   0.0   0.0   0.0   0.0   0.0

    V182  V183  V184  V185  V186  V187  V188  V189  V190  V191  V192  V193  \
0   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
1   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
```

```
2   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
3   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
4   0.0   0.0   0.0   0.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0

   V194  V195  …  V130_missing_flag  V131_missing_flag  V132_missing_flag  \
0   NaN   NaN  …              False              False              False
1   NaN   NaN  …              False              False              False
2   NaN   NaN  …              False              False              False
3   NaN   NaN  …              False              False              False
4   1.0   1.0  …              False              False              False

   V133_missing_flag  V134_missing_flag  V135_missing_flag  V136_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V137_missing_flag  V138_missing_flag  V139_missing_flag  V140_missing_flag  \
0              False               True               True               True
1              False               True               True               True
2              False               True               True               True
3              False               True               True               True
4              False              False              False              False

   V141_missing_flag  V142_missing_flag  V143_missing_flag  V144_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V145_missing_flag  V146_missing_flag  V147_missing_flag  V148_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V149_missing_flag  V150_missing_flag  V151_missing_flag  V152_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V153_missing_flag  V154_missing_flag  V155_missing_flag  V156_missing_flag  \
```

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

   V157_missing_flag  V158_missing_flag  V159_missing_flag  V160_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V161_missing_flag  V162_missing_flag  V163_missing_flag  V164_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V165_missing_flag  V166_missing_flag  V167_missing_flag  V168_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V169_missing_flag  V170_missing_flag  V171_missing_flag  V172_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V173_missing_flag  V174_missing_flag  V175_missing_flag  V176_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V177_missing_flag  V178_missing_flag  V179_missing_flag  V180_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False
```

```
   V181_missing_flag  V182_missing_flag  V183_missing_flag  V184_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V185_missing_flag  V186_missing_flag  V187_missing_flag  V188_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V189_missing_flag  V190_missing_flag  V191_missing_flag  V192_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V193_missing_flag  V194_missing_flag  V195_missing_flag  V196_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V197_missing_flag  V198_missing_flag  V199_missing_flag  V200_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V201_missing_flag  V202_missing_flag  V203_missing_flag  V204_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V205_missing_flag  V206_missing_flag  V207_missing_flag  V208_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
```

```
3              True              True              True              True
4             False             False             False             False

   V209_missing_flag  V210_missing_flag  V211_missing_flag  V212_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V213_missing_flag  V214_missing_flag  V215_missing_flag  V216_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V217_missing_flag  V218_missing_flag  V219_missing_flag  V220_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V221_missing_flag  V222_missing_flag  V223_missing_flag  V224_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V225_missing_flag  V226_missing_flag  V227_missing_flag  V228_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V229_missing_flag  V230_missing_flag  V231_missing_flag  V232_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V233_missing_flag  V234_missing_flag  V235_missing_flag  V236_missing_flag  \
0               True               True               True               True
```

```
1                True                True                True                True
2                True                True                True                True
3                True                True                True                True
4               False               False               False               False

   V237_missing_flag  V238_missing_flag  V239_missing_flag  V240_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V241_missing_flag  V242_missing_flag  V243_missing_flag  V244_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V245_missing_flag  V246_missing_flag  V247_missing_flag  V248_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V249_missing_flag  V250_missing_flag  V251_missing_flag  V252_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V253_missing_flag  V254_missing_flag  V255_missing_flag  V256_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V257_missing_flag  V258_missing_flag  V259_missing_flag  V260_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False
```

```
   V261_missing_flag  V262_missing_flag  V263_missing_flag  V264_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V265_missing_flag  V266_missing_flag  V267_missing_flag  V268_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V269_missing_flag  V270_missing_flag  V271_missing_flag  V272_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V273_missing_flag  V274_missing_flag  V275_missing_flag  V276_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V277_missing_flag  V278_missing_flag  V279_missing_flag  V280_missing_flag  \
0               True               True              False              False
1               True               True              False              False
2               True               True              False              False
3               True               True              False              False
4              False              False              False              False

   V281_missing_flag  V282_missing_flag  V283_missing_flag  V284_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

  V285_missing_flag  V286_missing_flag  V287_missing_flag V288_missing_flag  \
0             False              False              False             False
1             False              False              False             False
2             False              False              False             False
3             False              False              False             False
```

```
4             False              False              False             False

  V289_missing_flag  V290_missing_flag  V291_missing_flag  V292_missing_flag  \
0             False              False              False              False
1             False              False              False              False
2             False              False              False              False
3             False              False              False              False
4             False              False              False              False

   V293_missing_flag  V294_missing_flag  V295_missing_flag V296_missing_flag  \
0              False              False              False             False
1              False              False              False             False
2              False              False              False             False
3              False              False              False             False
4              False              False              False             False

   V297_missing_flag  V298_missing_flag  V299_missing_flag V300_missing_flag  \
0              False              False              False             False
1              False              False              False             False
2              False              False              False             False
3              False              False              False             False
4              False              False              False             False

  V301_missing_flag V302_missing_flag V303_missing_flag V304_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4             False             False             False             False

   V305_missing_flag V306_missing_flag V307_missing_flag V308_missing_flag  \
0              False             False              False             False
1              False             False              False             False
2              False             False              False             False
3              False             False              False             False
4              False             False              False             False

  V309_missing_flag V310_missing_flag V311_missing_flag V312_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4             False             False             False             False

  V313_missing_flag  V314_missing_flag  V315_missing_flag  V316_missing_flag  \
0             False              False              False              False
1             False              False              False              False
```

```
2            False             False             False             False
3            False             False             False             False
4            False             False             False             False

   V317_missing_flag  V318_missing_flag  V319_missing_flag  V320_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V321_missing_flag  V322_missing_flag  V323_missing_flag  V324_missing_flag  \
0              False               True               True               True
1              False               True               True               True
2              False               True               True               True
3              False               True               True               True
4              False              False              False              False

   V325_missing_flag  V326_missing_flag  V327_missing_flag  V328_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V329_missing_flag  V330_missing_flag  V331_missing_flag  V332_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V333_missing_flag  V334_missing_flag  V335_missing_flag  V336_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V337_missing_flag  V338_missing_flag  V339_missing_flag  \
0               True               True               True
1               True               True               True
2               True               True               True
3               True               True               True
4              False              False              False

   id_01_missing_flag  id_02_missing_flag  id_03_missing_flag  \
```

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

   id_04_missing_flag  id_05_missing_flag  id_06_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True                True                True

   id_07_missing_flag  id_08_missing_flag  id_09_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True                True                True

   id_10_missing_flag  id_11_missing_flag  id_12_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True               False               False

   id_13_missing_flag  id_14_missing_flag  id_15_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True               False               False

   id_16_missing_flag  id_17_missing_flag  id_18_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4               False               False                True

   id_19_missing_flag  id_20_missing_flag  id_21_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4               False               False                True
```

```
   id_22_missing_flag  id_23_missing_flag  id_24_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True                True                True

   id_25_missing_flag  id_26_missing_flag  id_27_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True                True                True

   id_28_missing_flag  id_29_missing_flag  id_30_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4               False               False               False

   id_31_missing_flag  id_32_missing_flag  id_33_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4               False               False               False

   id_34_missing_flag  id_35_missing_flag  id_36_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4               False               False               False

   id_37_missing_flag  id_38_missing_flag  DeviceType_missing_flag  \
0                True                True                     True
1                True                True                     True
2                True                True                     True
3                True                True                     True
4               False               False                    False

   DeviceInfo_missing_flag
0                     True
1                     True
2                     True
```

```
3                      True
4                     False

[5 rows x 868 columns]
```

Clean Data Let's drop the columns which may not be useful for our analysis

Create a missing value flag column for the columns we are dropping which have more than 90% missing values, there might be some specific pattern associated with missing values and transaction being fraud

[21]:
```
drop_cols = []

for col in df.columns:
    missing_share = df[col].isnull().sum()/df.shape[0]
    if missing_share > 0.9:
        drop_cols.append(col)
        print(col)
        # df[col + "_missing_flag"] = df[col].isnull()

good_cols = [col for col in df.columns if col not in drop_cols]
```

```
dist2
D7
id_07
id_08
id_18
id_21
id_22
id_23
id_24
id_25
id_26
id_27
```

[22]:
```
drop_cols = []
for col in good_cols:
    unique_value = df[col].nunique()
    if unique_value == 1:
        drop_cols.append(col)
        print(col)
good_cols = [col for col in good_cols if col not in drop_cols]
```

```
TransactionID_missing_flag
isFraud_missing_flag
TransactionDT_missing_flag
TransactionAmt_missing_flag
ProductCD_missing_flag
```

```
card1_missing_flag
C1_missing_flag
C2_missing_flag
C3_missing_flag
C4_missing_flag
C5_missing_flag
C6_missing_flag
C7_missing_flag
C8_missing_flag
C9_missing_flag
C10_missing_flag
C11_missing_flag
C12_missing_flag
C13_missing_flag
C14_missing_flag
```

[23]:
```
# Filter the data for relevant columns only
df = df[good_cols]
```

[24]:
```
df.shape
```

[24]:
```
(590540, 836)
```

Create date features

[25]:
```
START_DATE          = '2017-12-01'
startdate           = datetime.datetime.strptime(START_DATE, "%Y-%m-%d")
df["Date"]          = df['TransactionDT'].apply(lambda x: (startdate + datetime.
 ↪timedelta(seconds=x)))

df['_Weekdays']     = df['Date'].dt.dayofweek
df['_Hours']        = df['Date'].dt.hour
df['_Days']         = df['Date'].dt.day
```

[26]:
```
df = reduce_mem_usage(df)
```

```
Mem. usage decreased to 1453.88 Mb (0.8% reduction)
```

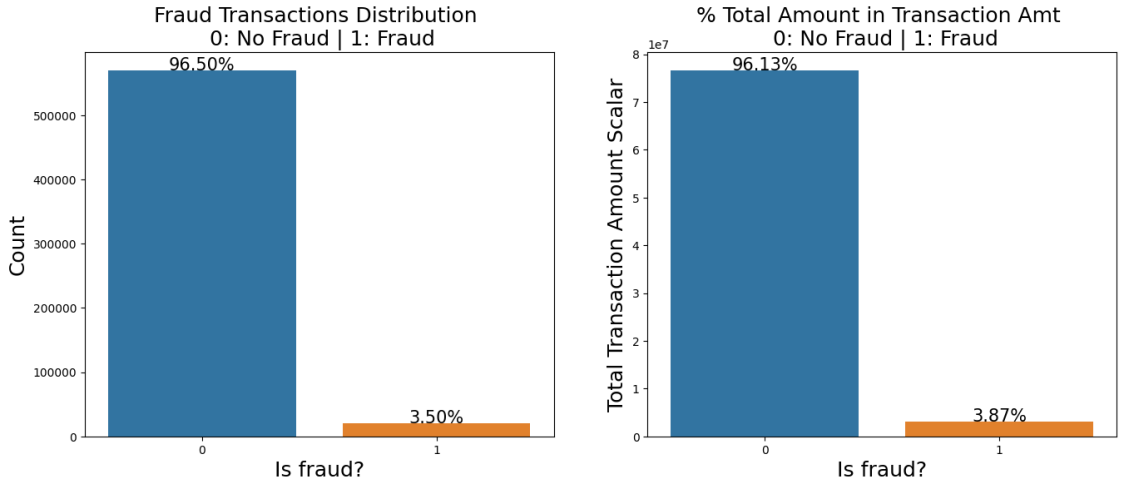
6 6. Exploratory Data Analysis

6.0.1 Check distribution of target variable

[27]:
```
df['isFraud'].value_counts()
```

[27]:
```
0    569877
1     20663
Name: isFraud, dtype: int64
```

[28]:
```
df['TransactionAmt'] = df['TransactionAmt'].astype(float)
total = len(df)
total_amt = df.groupby(['isFraud'])['TransactionAmt'].sum().sum()
plt.figure(figsize=(16,6))

plt.subplot(121)
g = sns.countplot(x='isFraud', data=df )
g.set_title("Fraud Transactions Distribution \n 0: No Fraud | 1: Fraud",␣
 ↪fontsize=18)
g.set_xlabel("Is fraud?", fontsize=18)
g.set_ylabel('Count', fontsize=18)
for p in g.patches:
    height = p.get_height()
    g.text(p.get_x()+p.get_width()/2.,
            height + 3,
            '{:1.2f}%'.format(height/total*100),
            ha="center", fontsize=15)

perc_amt = (df.groupby(['isFraud'])['TransactionAmt'].sum())
perc_amt = perc_amt.reset_index()

plt.subplot(122)
g1 = sns.barplot(x='isFraud', y='TransactionAmt',  dodge=True, data=perc_amt)
g1.set_title("% Total Amount in Transaction Amt \n 0: No Fraud | 1: Fraud",␣
 ↪fontsize=18)
g1.set_xlabel("Is fraud?", fontsize=18)
g1.set_ylabel('Total Transaction Amount Scalar', fontsize=18)
for p in g1.patches:
    height = p.get_height()
    g1.text(p.get_x()+p.get_width()/2.,
            height + 3,
            '{:1.2f}%'.format(height/total_amt * 100),
            ha="center", fontsize=15)

plt.show()
```



[29]:
```
df.groupby('isFraud')['TransactionAmt'].mean()
```

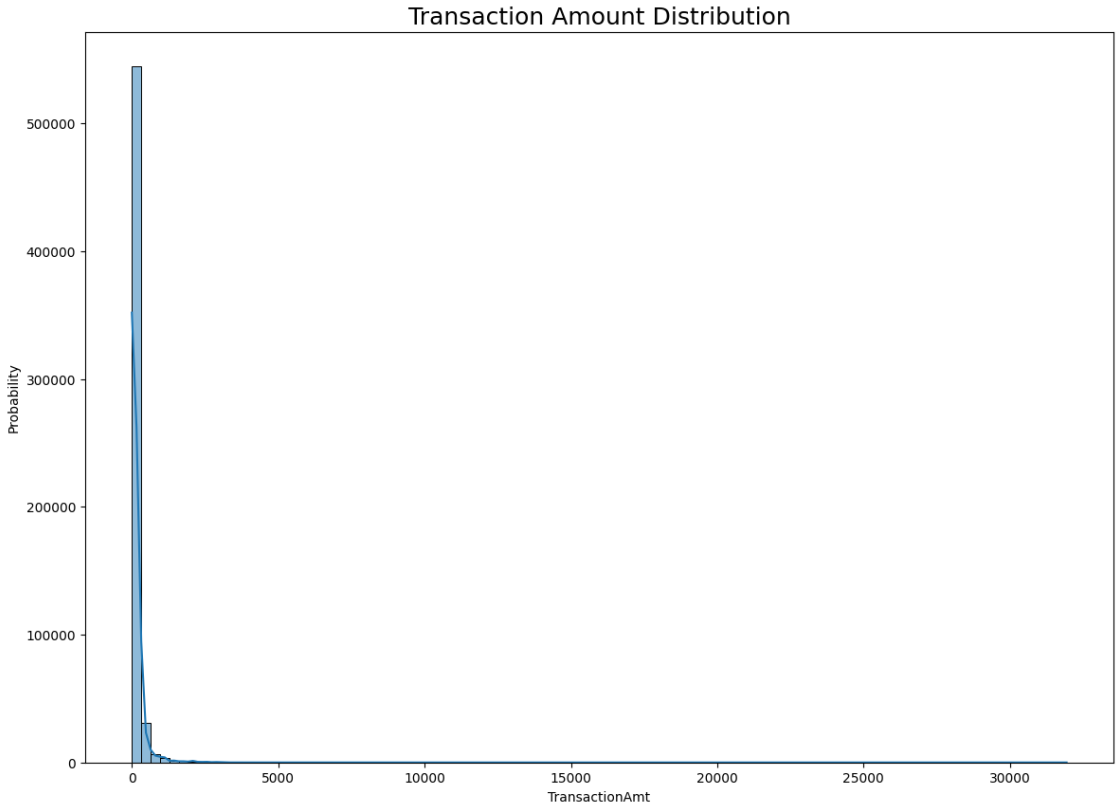
[29]:
```
isFraud
0    134.511857
1    149.244353
Name: TransactionAmt, dtype: float64
```

- The target variable is **imbalanced**. 3.5% transactions are Fraud
- Around same % of transaction amounts are fraud

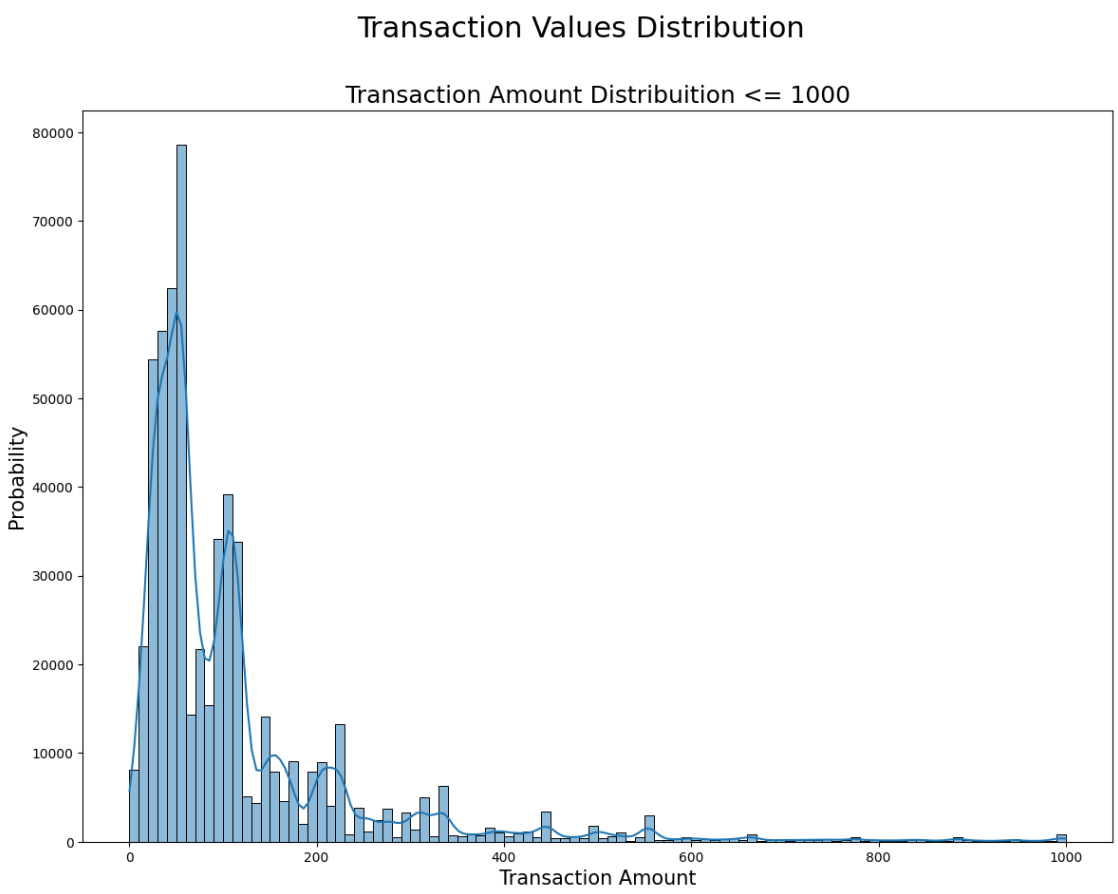
Let's explore the Transaction amount further

6.0.2 Check distribution of Transaction Amount

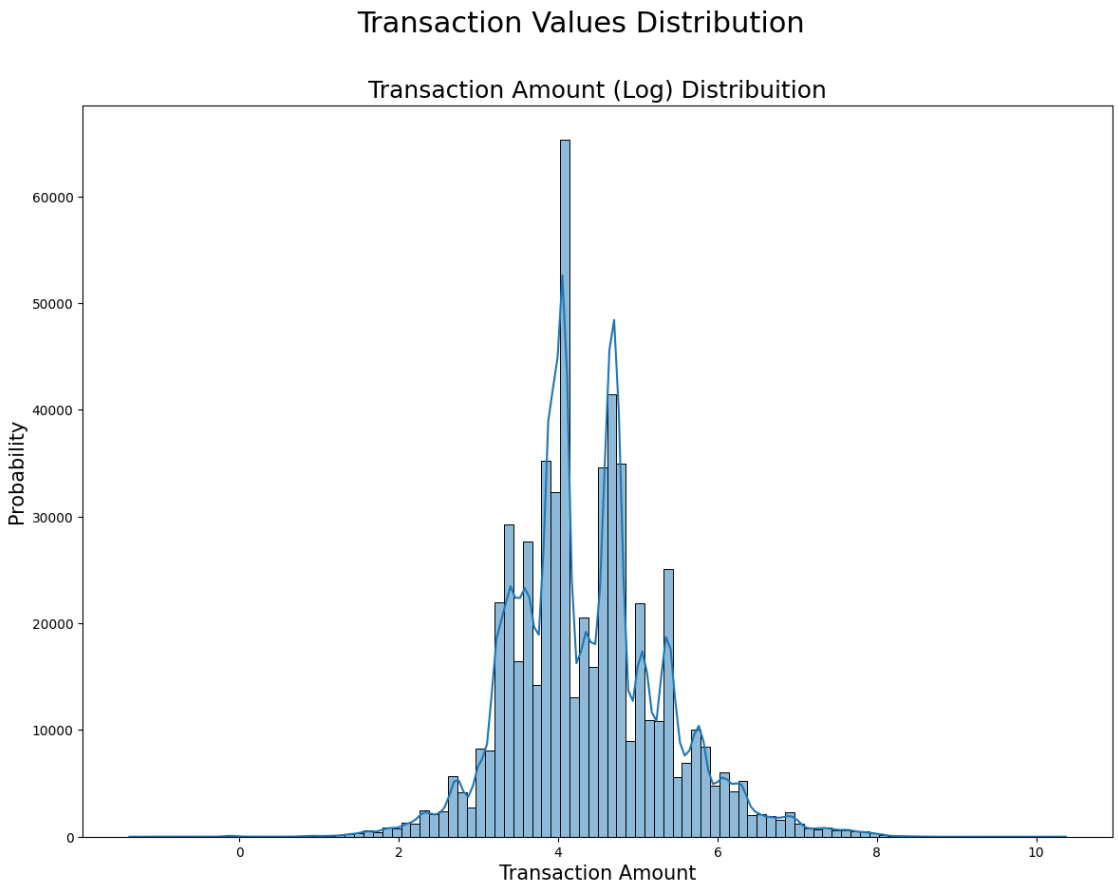
[30]:
```
plt.figure(figsize=(14,10))
sns.histplot(df['TransactionAmt'],bins=100,kde=True)
plt.title("Transaction Amount Distribution",fontsize=18)
plt.ylabel("Probability")
plt.show()
```



[31]:
```
# Distribution plot of Transaction Amount less than 1000
plt.figure(figsize=(14,10))
plt.suptitle('Transaction Values Distribution', fontsize=22)
sns.histplot(df[df['TransactionAmt'] <=␣
 ↪1000]['TransactionAmt'],bins=100,kde=True)
plt.title("Transaction Amount Distribuition <= 1000", fontsize=18)
plt.xlabel("Transaction Amount", fontsize=15)
plt.ylabel("Probability", fontsize=15)
plt.show()
```



[32]:
```
# Distribution plot of Transaction Amount less than 1000
plt.figure(figsize=(14,10))
plt.suptitle('Transaction Values Distribution', fontsize=22)
sns.histplot(np.log(df['TransactionAmt']),kde=True,bins=100)
plt.title("Transaction Amount (Log) Distribuition", fontsize=18)
plt.xlabel("Transaction Amount", fontsize=15)
plt.ylabel("Probability", fontsize=15)
plt.show()
```



- Transaction Amount is right skewed.
- Log of transaction amount is almost normally distributed, so use log of transaction amount while building the model

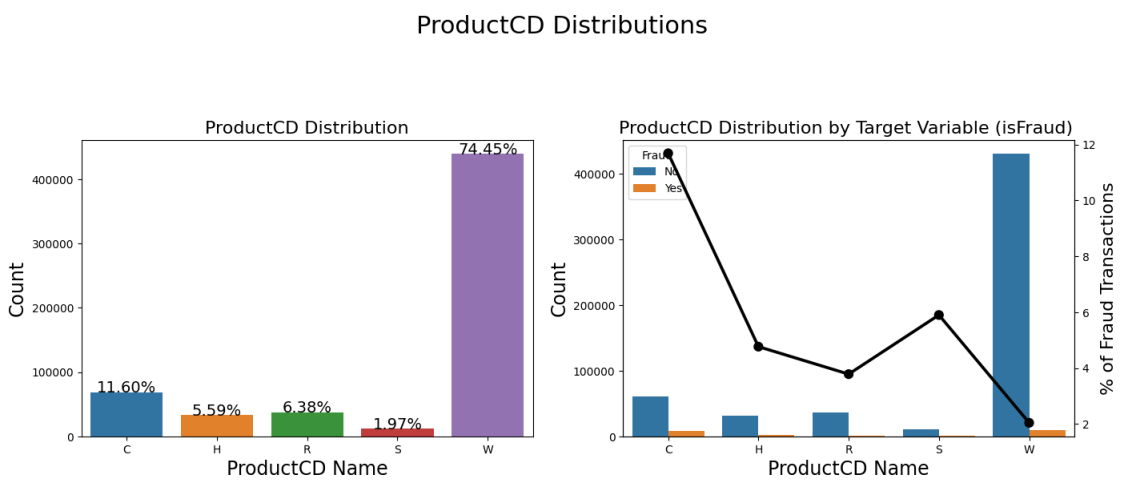
6.0.3 Product Features

- Distribution of ProductCD
- Distribution of Frauds by Product

[33]:
```
def plot_cat_feat_dist(df, col):
    tmp = pd.crosstab(df[col], df['isFraud'], normalize='index') * 100
    tmp = tmp.reset_index()
    tmp.rename(columns={0:'NoFraud', 1:'Fraud'}, inplace=True)

    plt.figure(figsize=(16,12))
    plt.suptitle(f'{col} Distributions', fontsize=22)

    plt.subplot(221)
    g = sns.countplot(x=col, data=df, order=tmp[col].values)
```

```
    g.set_title(f"{col} Distribution", fontsize=16)
    g.set_xlabel(f"{col} Name", fontsize=17)
    g.set_ylabel("Count", fontsize=17)
    for p in g.patches:
        height = p.get_height()
        g.text(p.get_x()+p.get_width()/2.,
                height + 3,
                '{:1.2f}%'.format(height/total*100),
                ha="center", fontsize=14)

    plt.subplot(222)
    g1 = sns.countplot(x=col, hue='isFraud', data=df, order=tmp[col].values)
    plt.legend(title='Fraud', loc='best', labels=['No', 'Yes'])
    gt = g1.twinx()
    gt = sns.pointplot(x=col, y='Fraud', data=tmp, color='black',␣
↪order=tmp[col].values, legend=False)
    gt.set_ylabel("% of Fraud Transactions", fontsize=16)

    g1.set_title(f"{col} Distribution by Target Variable (isFraud) ",␣
↪fontsize=16)
    g1.set_xlabel(f"{col} Name", fontsize=17)
    g1.set_ylabel("Count", fontsize=17)

    plt.subplots_adjust(hspace = 0.4, top = 0.85)
    plt.show()
```

[34]:
```
plot_cat_feat_dist(df, "ProductCD")
```



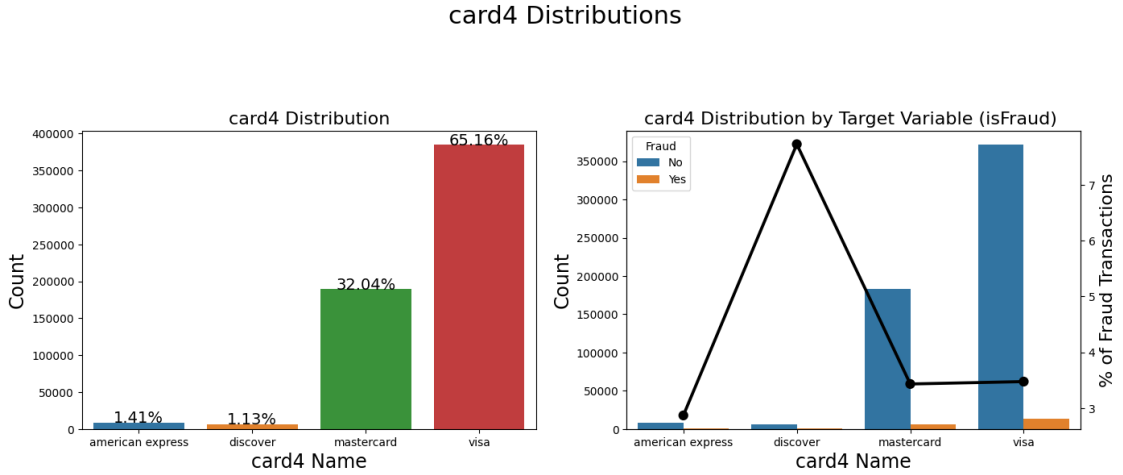
[35]:
```
df.groupby('ProductCD')['isFraud'].mean()
```

[35]:
```
ProductCD
C    0.116873
H    0.047662
R    0.037826
S    0.058996
W    0.020399
Name: isFraud, dtype: float64
```

- 75% of the transactions are for Product Catergory W
- 11.6% of the transactions are for Product Category C
- Fraud Transaction rate is maximum for Product Category C and minimum for Product Category W

6.0.4 Card Features

[36]:
```
# Card 4
plot_cat_feat_dist(df, "card4")
```

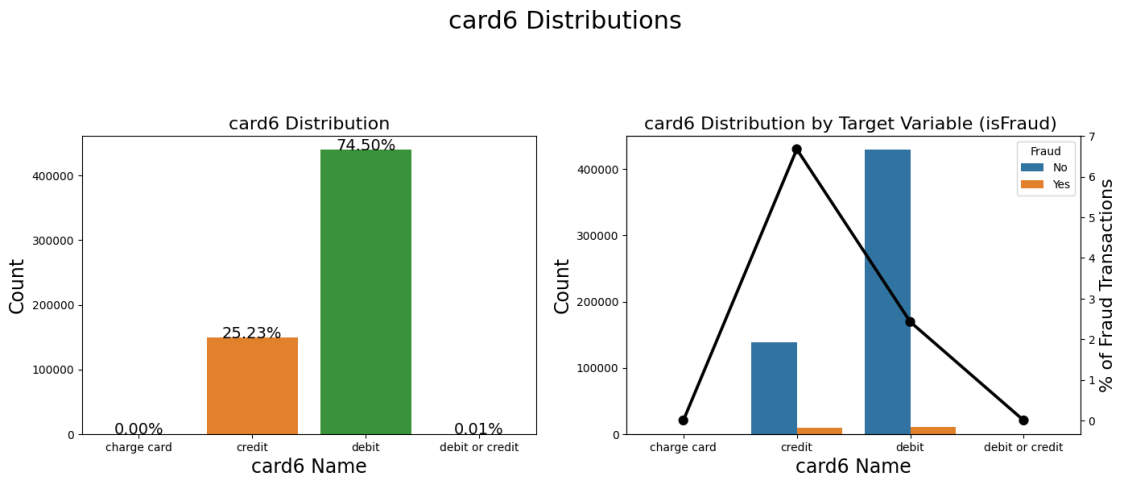


[37]:
```
df.groupby('card4')['isFraud'].mean()
```

[37]:
```
card4
american express    0.028698
discover            0.077282
mastercard          0.034331
visa                0.034756
Name: isFraud, dtype: float64
```

- 97% of transactions are from Mastercard(32%) and Visa(65%
- Fraud transaction rate is highest for discover cards(~8%) against ~3.5% of Mastercard and Visa and 2.87% in American Express

[38]:
```
# Card 6
plot_cat_feat_dist(df, "card6")
```



[39]:
```
df.groupby('card6')['isFraud'].mean()
```

[39]:
```
card6
charge card        0.000000
credit             0.066785
debit              0.024263
debit or credit    0.000000
Name: isFraud, dtype: float64
```

- Almost all the transactions are from Credit and Debit cards.
- Debit card transactions are almost 3 times as compared to credit card transactions.
- Fraud transaction rate is high for Credit cards as compared to Debit cards.

6.0.5 P emaildomain

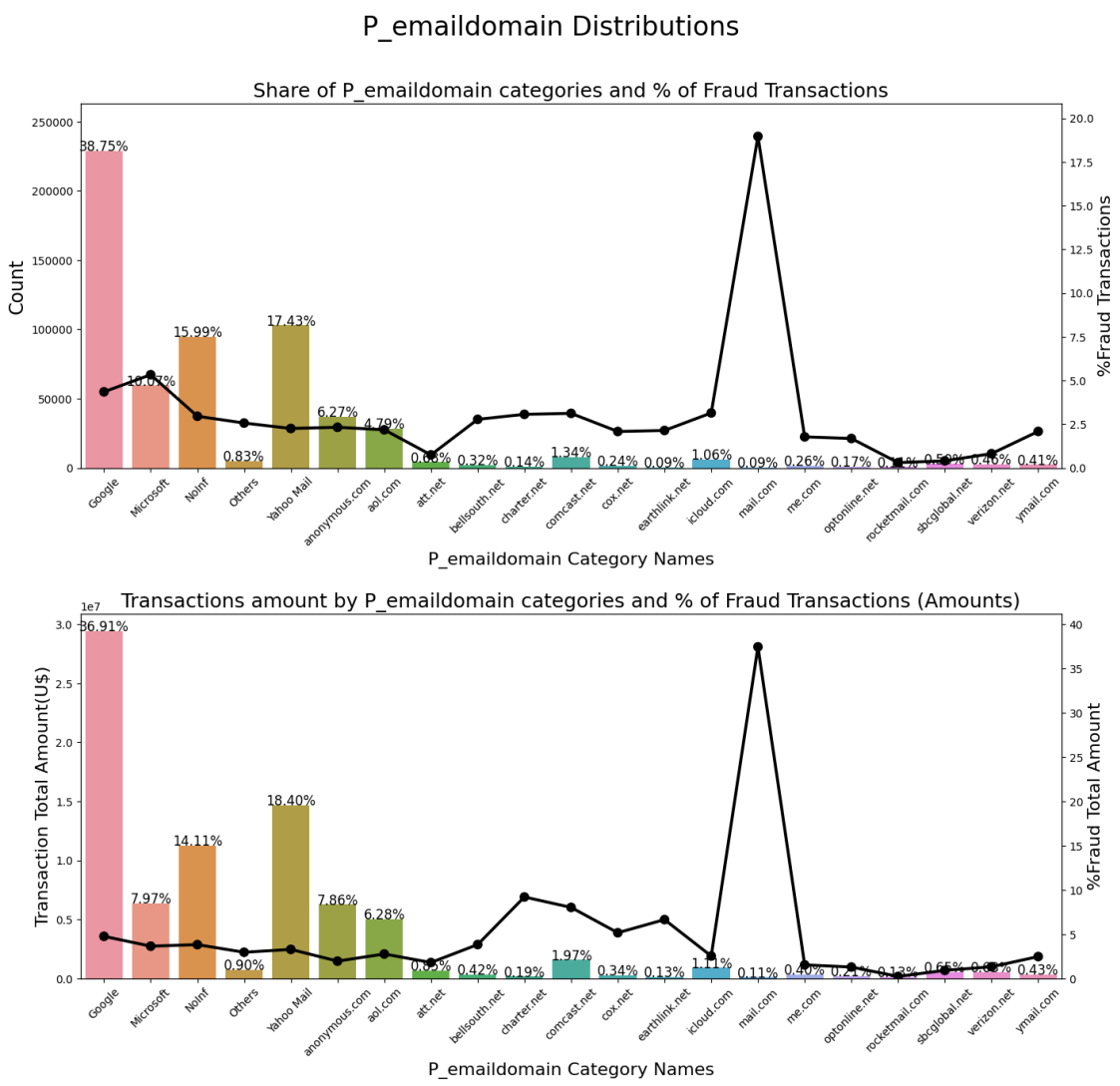
[40]:
```
df.loc[df['P_emaildomain'].isin(['gmail.com', 'gmail']),'P_emaildomain'] =␣
 ↪'Google'

df.loc[df['P_emaildomain'].isin(['yahoo.com', 'yahoo.com.mx',  'yahoo.co.uk',
                                 'yahoo.co.jp', 'yahoo.de', 'yahoo.fr',
                                 'yahoo.es']), 'P_emaildomain'] =␣
 ↪'Yahoo Mail'
df.loc[df['P_emaildomain'].isin(['hotmail.com','outlook.com','msn.com', 'live.
 ↪com.mx',
                                 'hotmail.es','hotmail.co.uk', 'hotmail.
 ↪de',
                                 'outlook.es', 'live.com', 'live.fr',
```

```
                                            'hotmail.fr']), 'P_emaildomain'] =␣
 ↪'Microsoft'
df.loc[df.P_emaildomain.isin(df.P_emaildomain\
                                            .value_counts()[df.P_emaildomain.
 ↪value_counts() <= 500 ]\
                                            .index), 'P_emaildomain'] = "Others"
df.P_emaildomain.fillna("NoInf", inplace=True)
```

[41]:
```
def plot_cat_with_amt(df, col, lim=2000):
    tmp = pd.crosstab(df[col], df['isFraud'], normalize='index') * 100
    tmp = tmp.reset_index()
    tmp.rename(columns={0:'NoFraud', 1:'Fraud'}, inplace=True)

    plt.figure(figsize=(16,14))
    plt.suptitle(f'{col} Distributions ', fontsize=24)

    plt.subplot(211)
    g = sns.countplot( x=col,  data=df, order=list(tmp[col].values))
    gt = g.twinx()
    gt = sns.pointplot(x=col, y='Fraud', data=tmp, order=list(tmp[col].values),
                       color='black', legend=False, )
    gt.set_ylim(0,tmp['Fraud'].max()*1.1)
    gt.set_ylabel("%Fraud Transactions", fontsize=16)
    g.set_title(f"Share of {col} categories and % of Fraud Transactions",␣
 ↪fontsize=18)
    g.set_xlabel(f"{col} Category Names", fontsize=16)
    g.set_ylabel("Count", fontsize=17)
    g.set_xticklabels(g.get_xticklabels(),rotation=45)
    sizes = []
    for p in g.patches:
        height = p.get_height()
        sizes.append(height)
        g.text(p.get_x()+p.get_width()/2.,
                height + 3,
                '{:1.2f}%'.format(height/total*100),
                ha="center",fontsize=12)

    g.set_ylim(0,max(sizes)*1.15)

    ########################################################################
    perc_amt = (df.groupby(['isFraud',col])['TransactionAmt'].sum() \
                / df.groupby([col])['TransactionAmt'].sum() * 100).
 ↪unstack('isFraud')
    perc_amt = perc_amt.reset_index()
    perc_amt.rename(columns={0:'NoFraud', 1:'Fraud'}, inplace=True)
    amt = df.groupby([col])['TransactionAmt'].sum().reset_index()
    perc_amt = perc_amt.fillna(0)
```

```
    plt.subplot(212)
    g1 = sns.barplot(x=col, y='TransactionAmt',
                     data=amt,
                     order=list(tmp[col].values))
    g1t = g1.twinx()
    g1t = sns.pointplot(x=col, y='Fraud', data=perc_amt,
                        order=list(tmp[col].values),
                     color='black', legend=False, )
    g1t.set_ylim(0,perc_amt['Fraud'].max()*1.1)
    g1t.set_ylabel("%Fraud Total Amount", fontsize=16)
    g.set_xticklabels(g.get_xticklabels(),rotation=45)
    g1.set_title(f"Transactions amount by {col} categories and % of Fraud␣
↪Transactions (Amounts)", fontsize=18)
    g1.set_xlabel(f"{col} Category Names", fontsize=16)
    g1.set_ylabel("Transaction Total Amount(U$)", fontsize=16)
    g1.set_xticklabels(g.get_xticklabels(),rotation=45)

    for p in g1.patches:
        height = p.get_height()
        g1.text(p.get_x()+p.get_width()/2.,
                height + 3,
                '{:1.2f}%'.format(height/total_amt*100),
                ha="center",fontsize=12)

    plt.subplots_adjust(hspace=.4, top = 0.9)
    plt.show()
```

[42]:
```
plot_cat_with_amt(df, 'P_emaildomain')
```



[43]:
```
df.groupby('P_emaildomain')['isFraud'].mean()
```

[43]:
```
P_emaildomain
Google           0.043496
Microsoft        0.053298
NoInf            0.029538
Others           0.025646
Yahoo Mail       0.022544
anonymous.com    0.023217
aol.com          0.021811
att.net          0.007439
bellsouth.net    0.027763
charter.net      0.030637
comcast.net      0.031187
```

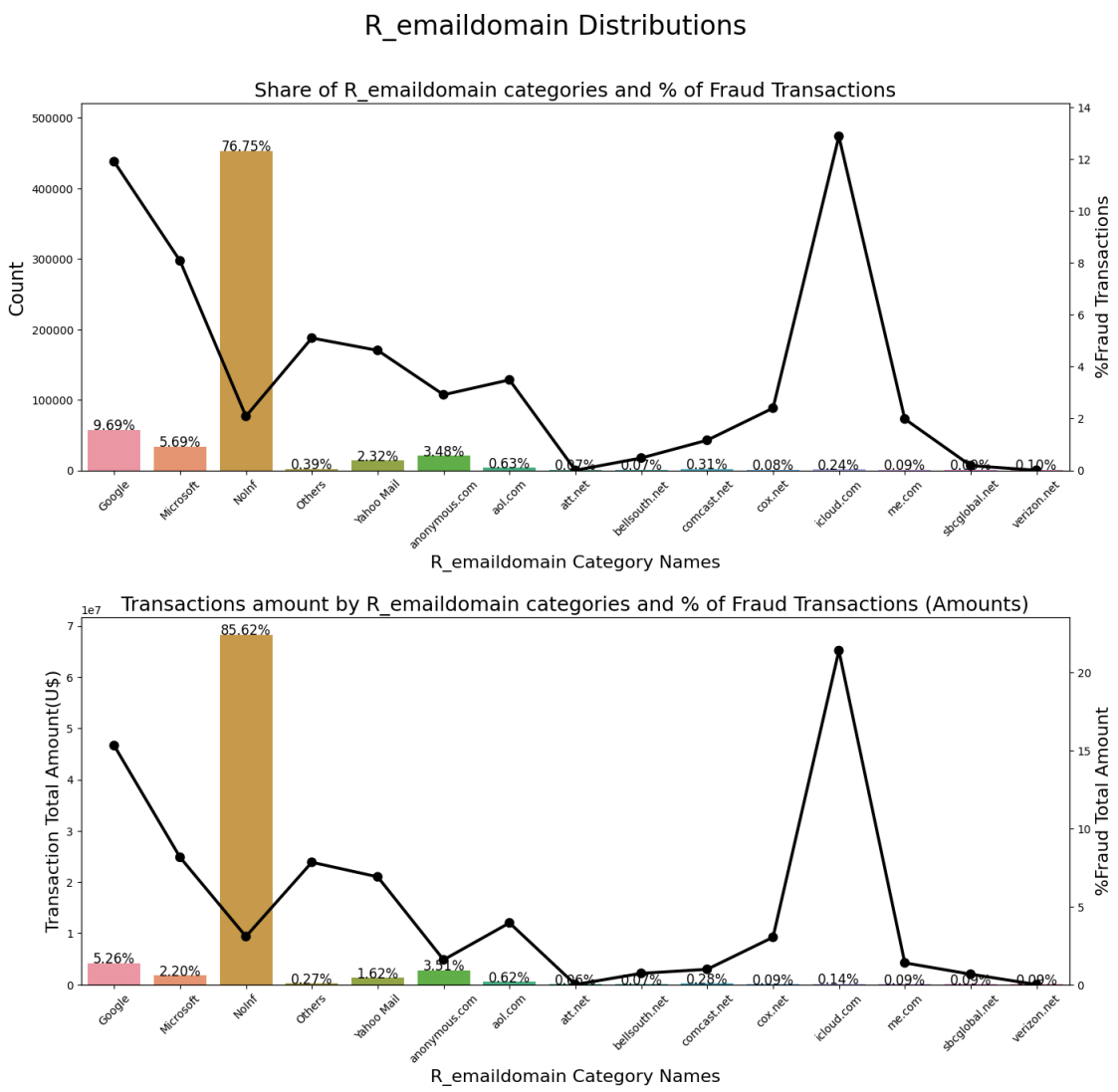
```
cox.net            0.020818
earthlink.net      0.021401
icloud.com         0.031434
mail.com           0.189624
me.com             0.017740
optonline.net      0.016815
rocketmail.com     0.003012
sbcglobal.net      0.004040
verizon.net        0.008133
ymail.com          0.020868
Name: isFraud, dtype: float64
```

- Majority of transactions are with P_emaildomain as Google, Microsoft and Yahoo Mail
- There isn't any information about P_emaildomain of around 16% transactions in terms of count and 14.11% in terms of amount
- Fraud transaction rate for Microsoft is high as compared to Google and Yahoo mail
- Fraud transaction rate (amount) for Google is high as comapred to Microsoft and Yahoo mail

6.0.6 R-Email Domain

[44]:
```
df.loc[df['R_emaildomain'].isin(['gmail.com', 'gmail']),'R_emaildomain'] =␣
 ↪'Google'

df.loc[df['R_emaildomain'].isin(['yahoo.com', 'yahoo.com.mx',  'yahoo.co.uk',
                                 'yahoo.co.jp', 'yahoo.de', 'yahoo.
 ↪fr',
                                 'yahoo.es']), 'R_emaildomain'] =␣
 ↪'Yahoo Mail'
df.loc[df['R_emaildomain'].isin(['hotmail.com','outlook.com','msn.com', 'live.
 ↪com.mx',
                                 'hotmail.es','hotmail.co.uk',␣
 ↪'hotmail.de',
                                 'outlook.es', 'live.com', 'live.
 ↪fr',
                                 'hotmail.fr']), 'R_emaildomain'] =␣
 ↪'Microsoft'
df.loc[df.R_emaildomain.isin(df.R_emaildomain\
                                 .value_counts()[df.R_emaildomain.
 ↪value_counts() <= 300 ]\
                                 .index), 'R_emaildomain'] = "Others"
df.R_emaildomain.fillna("NoInf", inplace=True)
```

[45]:
```
plot_cat_with_amt(df, 'R_emaildomain')
```



[46]:
```
df.groupby('R_emaildomain')['isFraud'].mean()
```

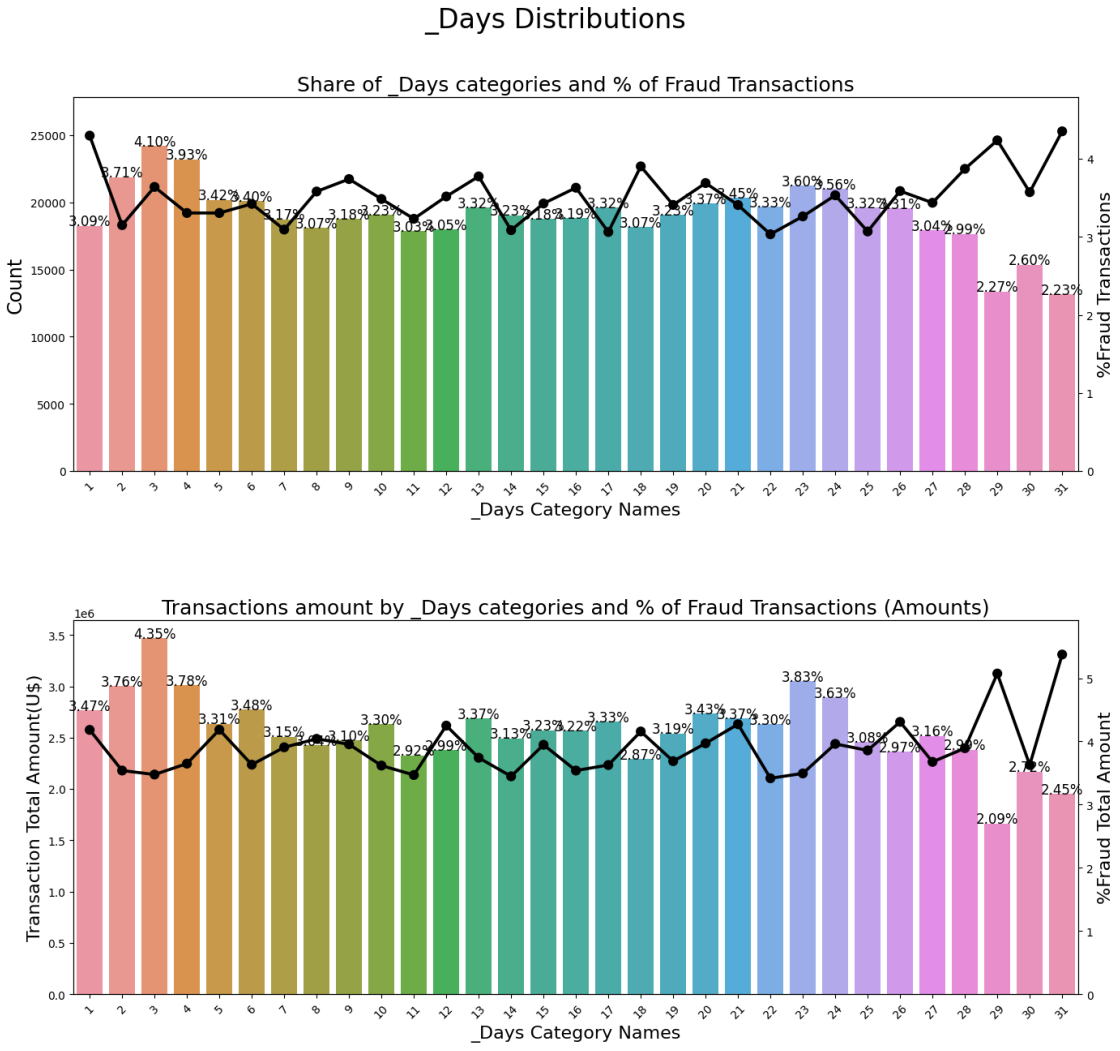
[46]:
```
R_emaildomain
Google           0.118986
Microsoft        0.080764
NoInf            0.020819
Others           0.050989
Yahoo Mail       0.046235
anonymous.com    0.029130
aol.com          0.034855
att.net          0.000000
bellsouth.net    0.004739
comcast.net      0.011589
cox.net          0.023965
```

```
icloud.com        0.128755
me.com            0.019784
sbcglobal.net     0.001812
verizon.net       0.000000
Name: isFraud, dtype: float64
```

- There isn't any information about R_emaildomain for Majority of transactions (76.75% count , 85.62% amount)
- Fraud transaction rate for Google is high as compared to Yahoo, anaonymous.com and Microsoft

6.0.7 Days of the Month

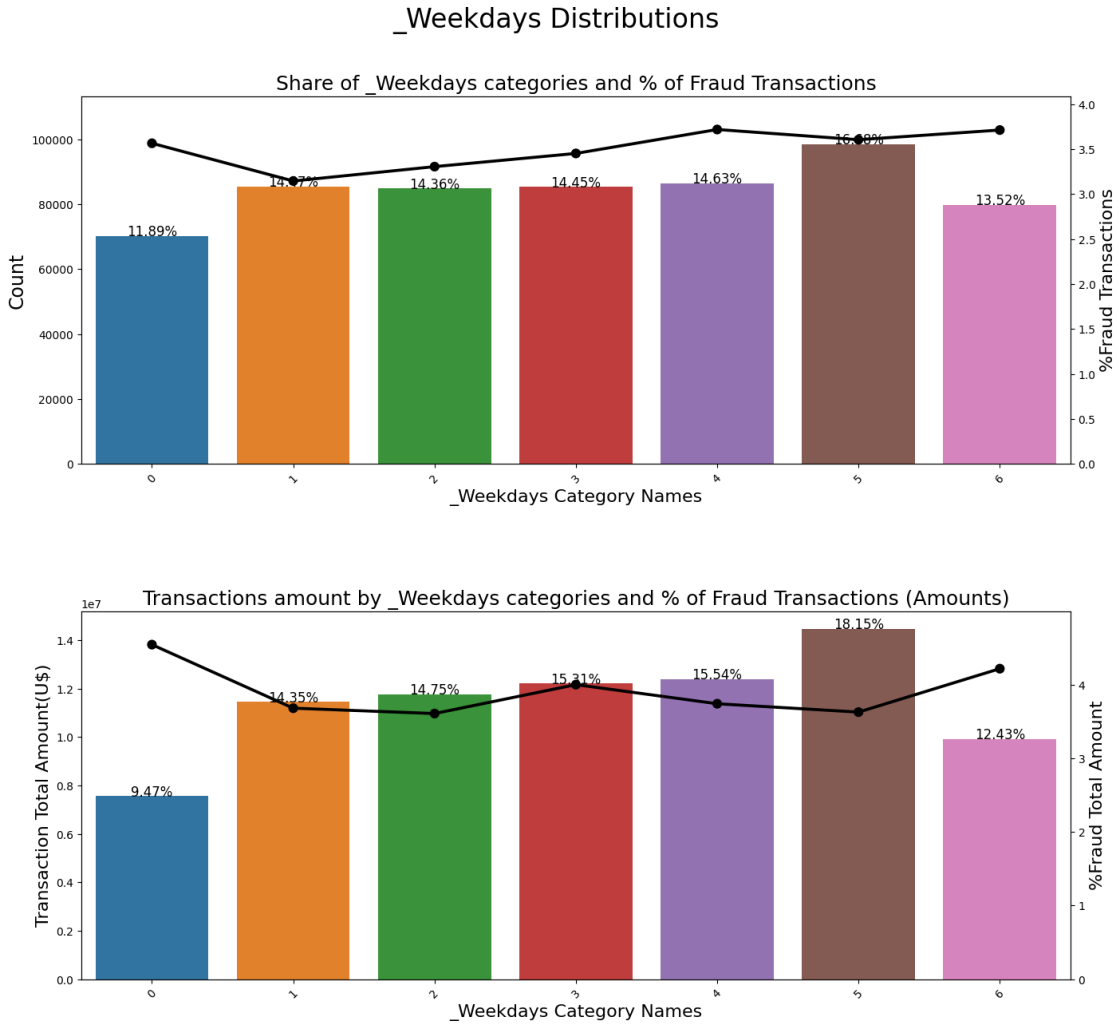
[47]:
```
plot_cat_with_amt(df, '_Days')
```



- The perc of fraud transactions is highest towards the beginning and the end of the month. Might be accelerated at the time of receiving pay-checks.
- Incidentally, fraud transaction rate is high on the days when number of transactions are less
- Day 29,30 and 31 are having less transactions, looks like people are cautious with spending in those times.

6.0.8 Days of the week

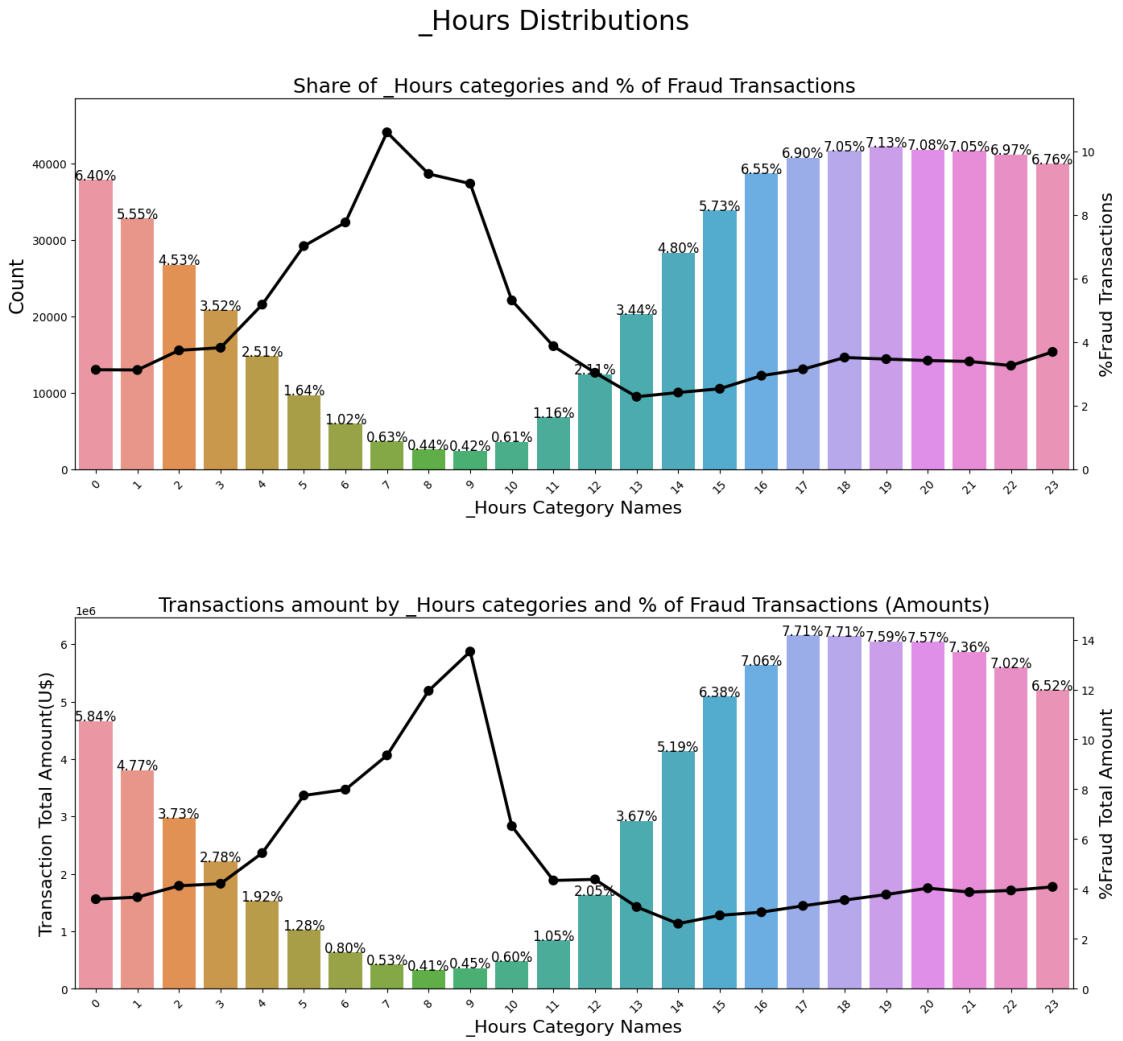
[48]:
```
plot_cat_with_amt(df, '_Weekdays')
```



- Surprisingly fraud transaction rate is high on the days when number of transactions and transaction amounts are less. Day 0 and 6
- Day 0 and 6 have less transactions, these might be weekend days

6.0.9 Hour of the Day

[49]:
```
plot_cat_with_amt(df, '_Hours')
```

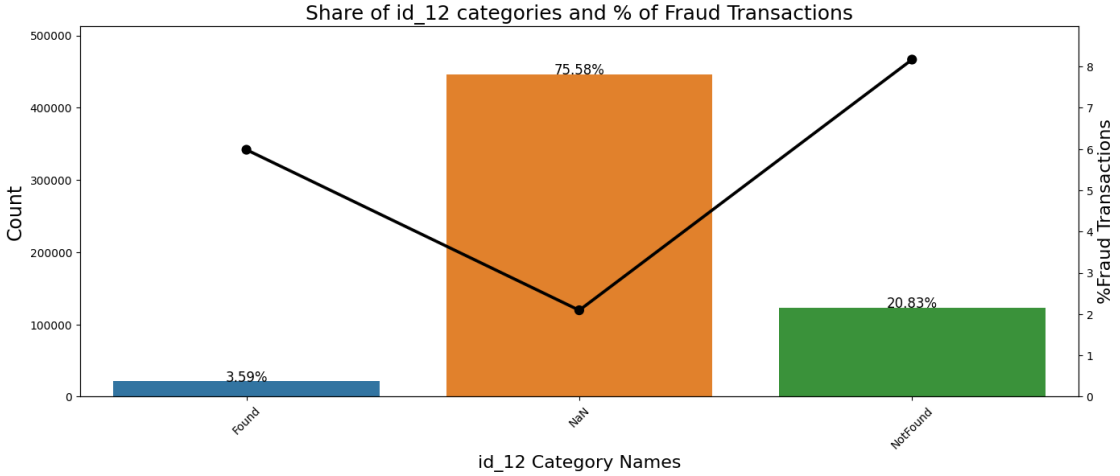


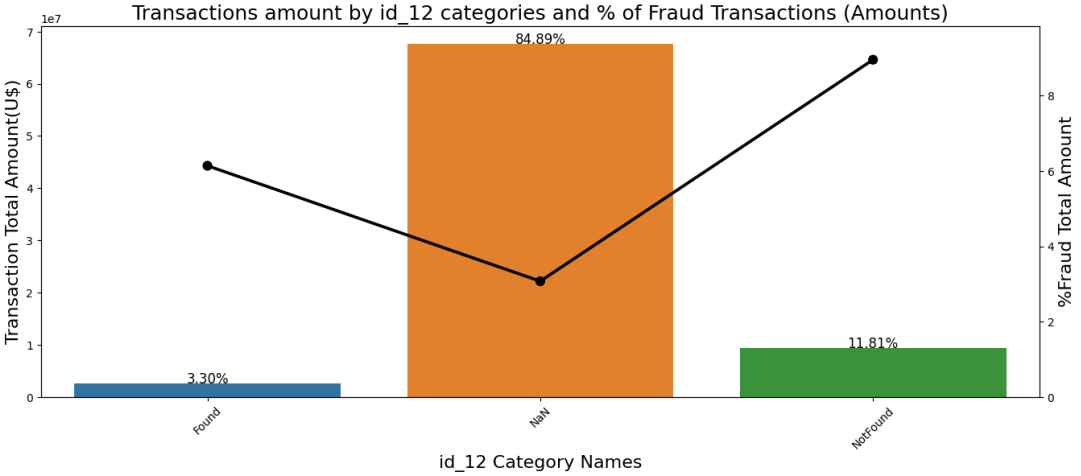
- Transactions start decreasing mid night but the fraud rate starts increasing
- Transactions from 3 AM to 12 PM needs to monitored very closely

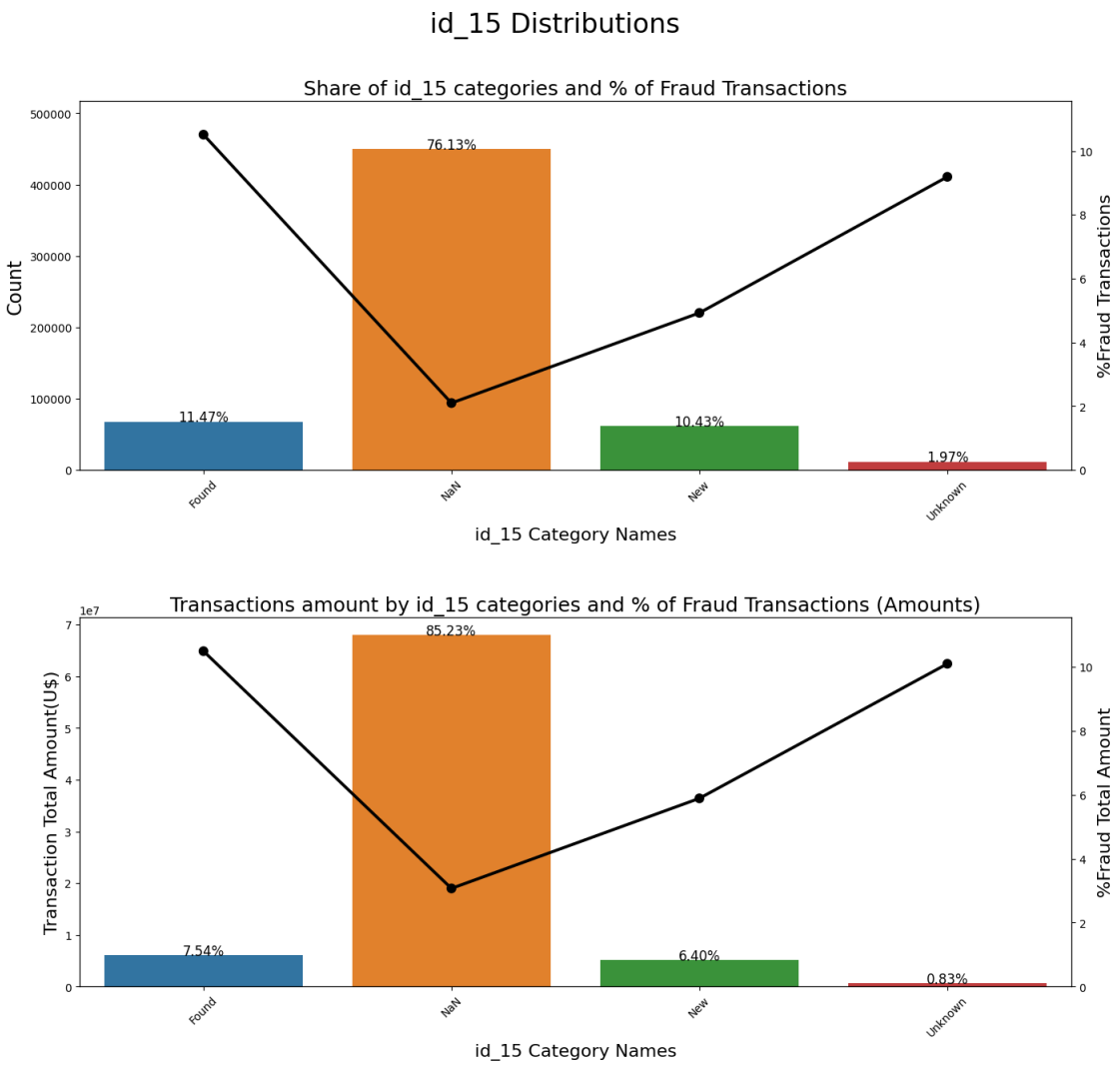
6.0.10 Columns from identity data

[50]:
```
for col in ['id_12', 'id_15', 'id_16', 'id_28', 'id_29']:
    df[col] = df[col].fillna('NaN')
    plot_cat_with_amt(df, col)
```

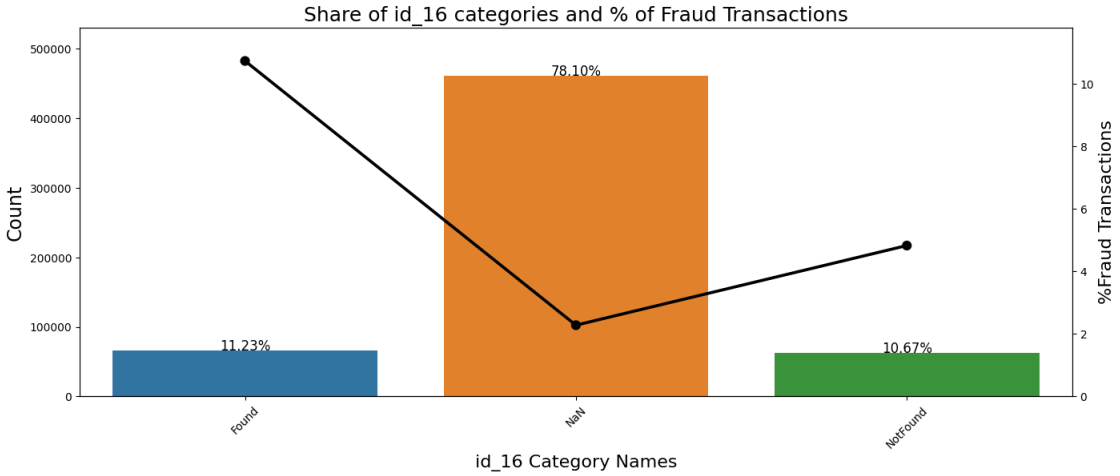
id_12 Distributions

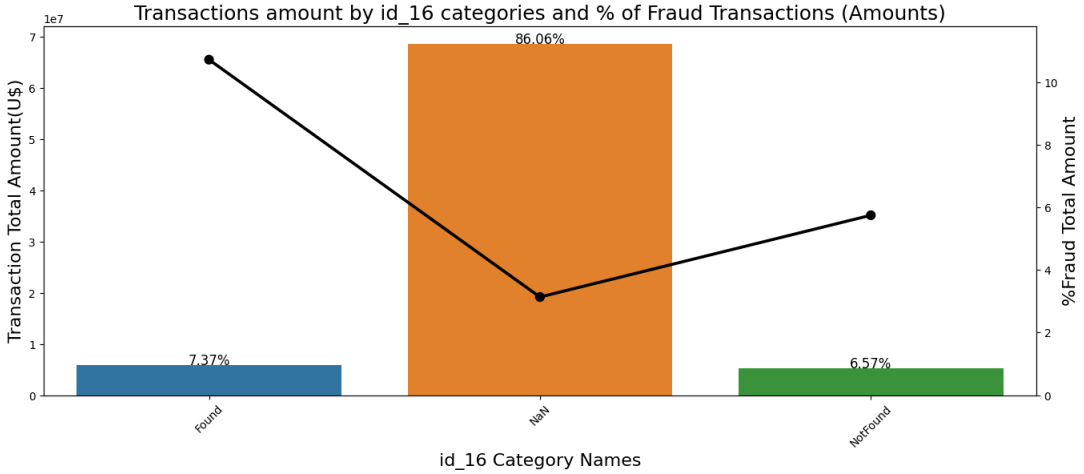


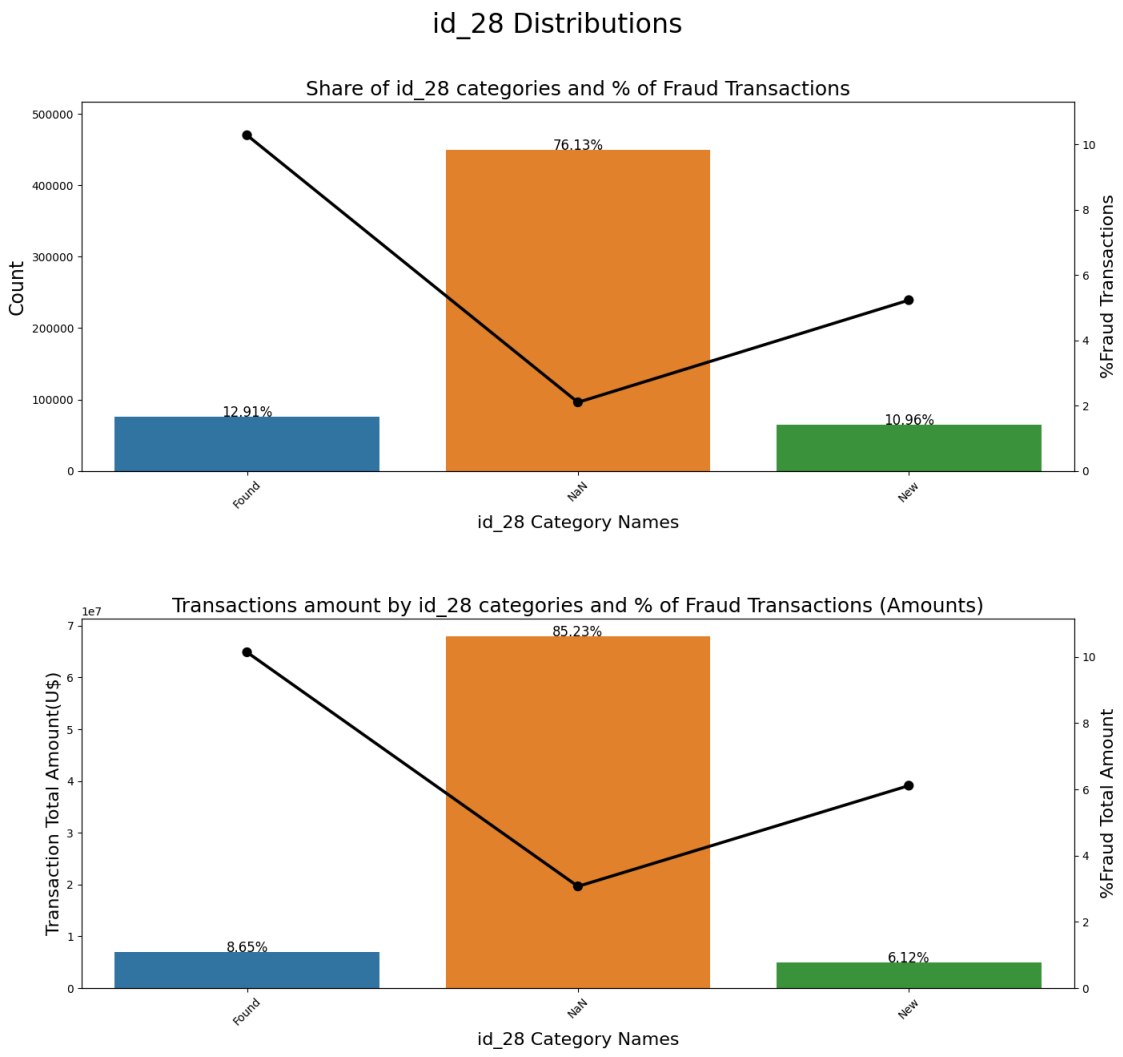


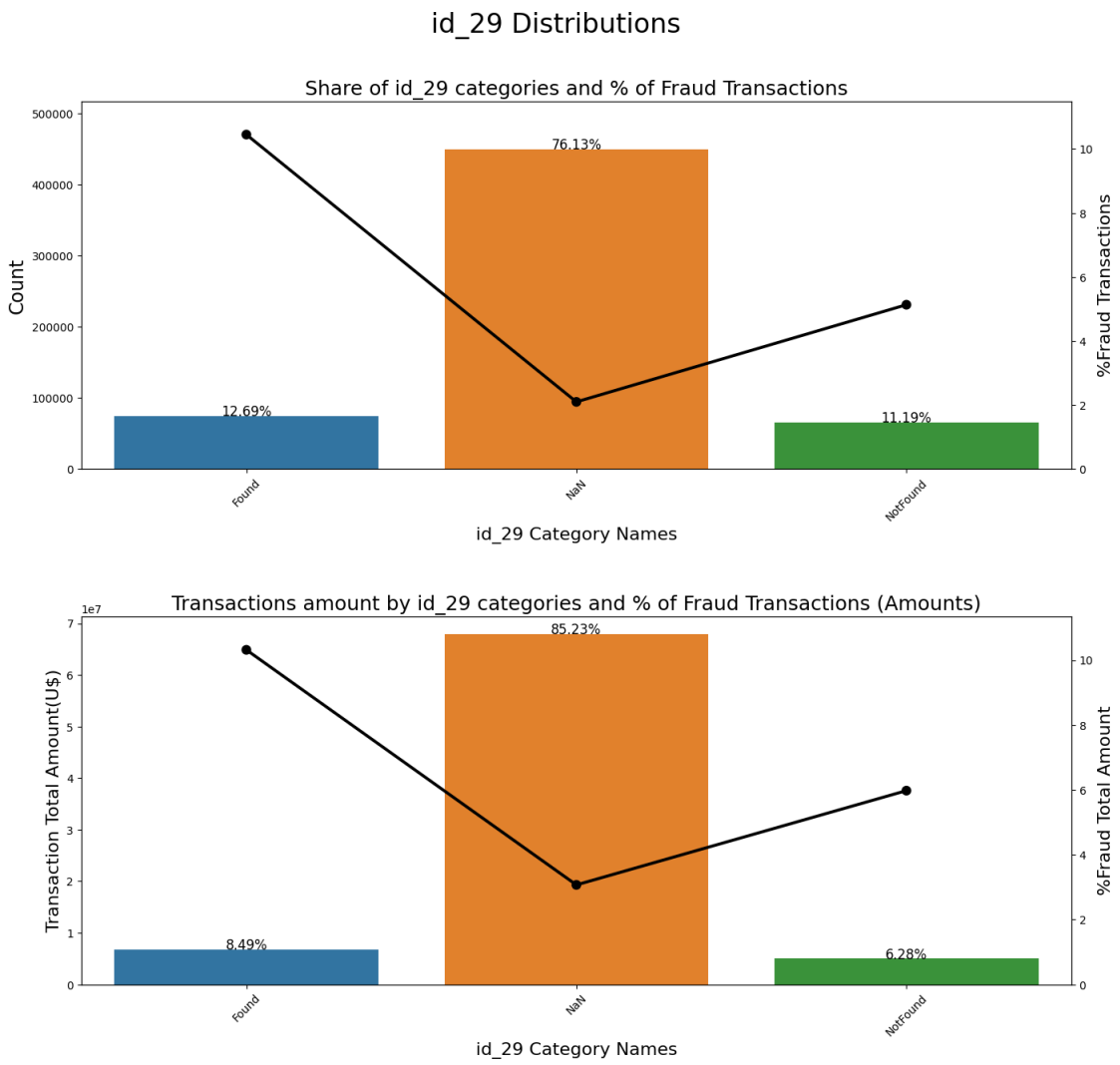

id_15 Distributions
Share of id_15 categories and % of Fraud Transactions
Count
%Fraud Transactions
76.13%
11.47%
10.43%
1.97%
Found
NaN
New
Unknown
id_15 Category Names
Transactions amount by id_15 categories and % of Fraud Transactions (Amounts)
1e7
Transaction Total Amount(U$)
%Fraud Total Amount
85.23%
7.54%
6.40%
0.83%
Found
NaN
New
Unknown
id_15 Category Names

id_16 Distributions

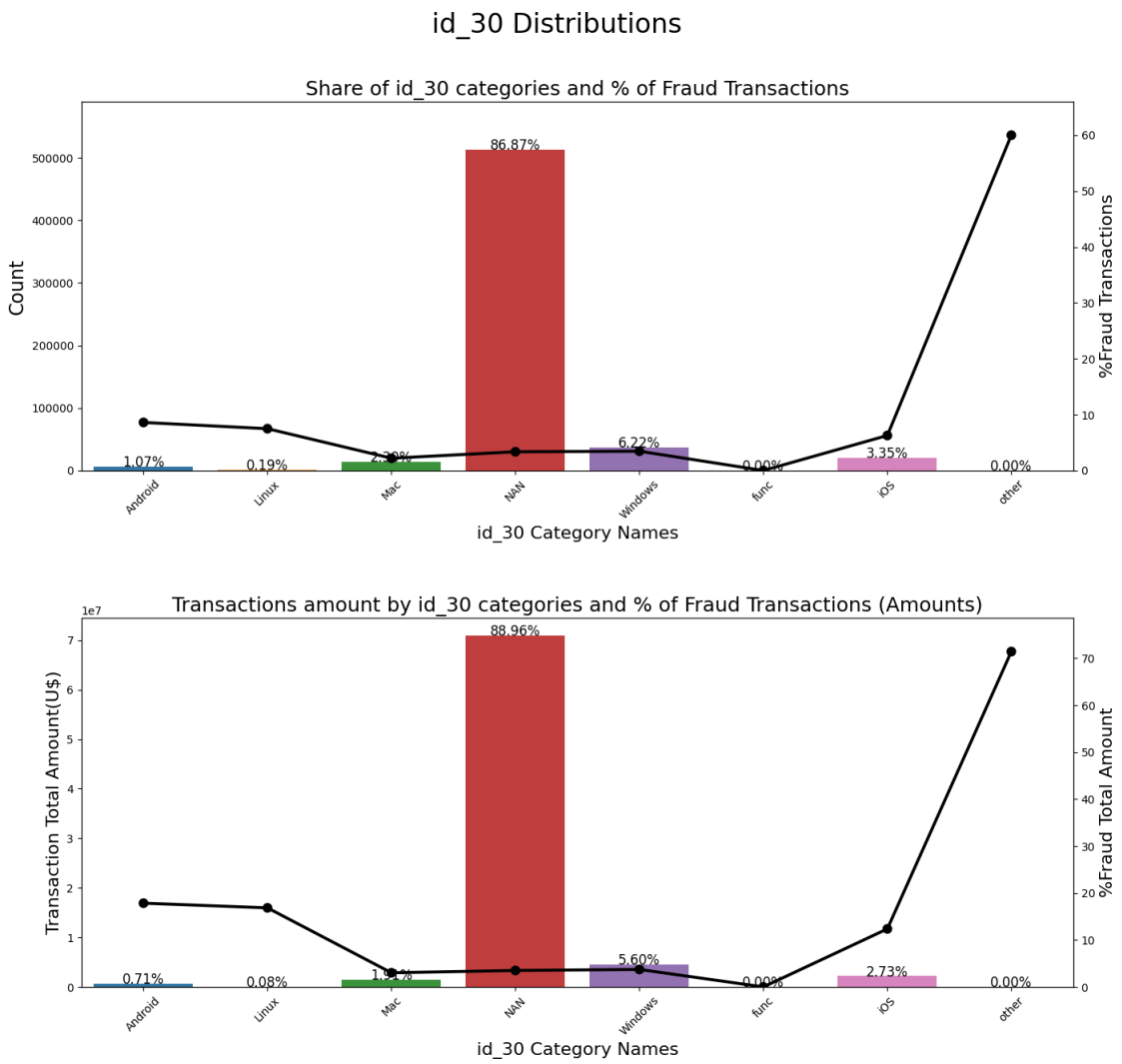





id_28 Distributions
Share of id_28 categories and % of Fraud Transactions
Count
%Fraud Transactions
76.13%
12.91%
10.96%
Found
NaN
New
id_28 Category Names
Transactions amount by id_28 categories and % of Fraud Transactions (Amounts)
1e7
Transaction Total Amount(U$)
%Fraud Total Amount
85.23%
8.65%
6.12%
Found
NaN
New
id_28 Category Names



[51]:
```
df.loc[df['id_30'].str.contains('Windows', na=False), 'id_30'] = 'Windows'
df.loc[df['id_30'].str.contains('iOS', na=False), 'id_30'] = 'iOS'
df.loc[df['id_30'].str.contains('Mac OS', na=False), 'id_30'] = 'Mac'
df.loc[df['id_30'].str.contains('Android', na=False), 'id_30'] = 'Android'
df['id_30'].fillna("NAN", inplace=True)

plot_cat_with_amt(df, "id_30")
```



[52]:
```
cat_columns = df.select_dtypes(include=['object']).columns
len(cat_columns)
```

[52]: 29

[53]:
```
binary_columns = [col for col in df.columns if df[col].nunique() == 2]
len(binary_columns)
```

[53]: 435

[54]:
```
num_columns = [col for col in df.columns if (col not in cat_columns) & (col not␣
 ↪in binary_columns)]
len(num_columns)
```

[54]: 389

[55]:
```
cat_columns = cat_columns.to_list() + binary_columns
```

7 7. Feature Engineering

[56]:
```
df.head()
```

[56]:
```
   TransactionID  isFraud  TransactionDT  TransactionAmt ProductCD  card1  \
0        2987000        0          86400            68.5         W  13926
1        2987001        0          86401            29.0         W   2755
2        2987002        0          86469            59.0         W   4663
3        2987003        0          86499            50.0         W  18132
4        2987004        0          86506            50.0         H   4497

   card2  card3       card4  card5   card6  addr1  addr2  dist1 P_emaildomain  \
0    NaN  150.0    discover  142.0  credit  315.0   87.0   19.0         NoInf
1  404.0  150.0  mastercard  102.0  credit  325.0   87.0    NaN        Google
2  490.0  150.0        visa  166.0   debit  330.0   87.0  287.0     Microsoft
3  567.0  150.0  mastercard  117.0   debit  476.0   87.0    NaN    Yahoo Mail
4  514.0  150.0  mastercard  102.0  credit  420.0   87.0    NaN        Google

  R_emaildomain   C1   C2   C3   C4   C5   C6   C7   C8   C9  C10  C11  C12  \
0         NoInf  1.0  1.0  0.0  0.0  0.0  1.0  0.0  0.0  1.0  0.0  2.0  0.0
1         NoInf  1.0  1.0  0.0  0.0  0.0  1.0  0.0  0.0  0.0  0.0  1.0  0.0
2         NoInf  1.0  1.0  0.0  0.0  0.0  1.0  0.0  0.0  1.0  0.0  1.0  0.0
3         NoInf  2.0  5.0  0.0  0.0  0.0  4.0  0.0  0.0  1.0  0.0  1.0  0.0
4         NoInf  1.0  1.0  0.0  0.0  0.0  1.0  0.0  1.0  0.0  1.0  1.0  0.0

    C13  C14     D1     D2    D3    D4   D5  D6  D8  D9   D10    D11  D12  \
0   1.0  1.0   14.0    NaN  13.0   NaN  NaN NaN NaN NaN  13.0   13.0  NaN
1   1.0  1.0    0.0    NaN   NaN   0.0  NaN NaN NaN NaN   0.0    NaN  NaN
2   1.0  1.0    0.0    NaN   NaN   0.0  NaN NaN NaN NaN   0.0  315.0  NaN
3  25.0  1.0  112.0  112.0   0.0  94.0  0.0 NaN NaN NaN  84.0    NaN  NaN
4   1.0  1.0    0.0    NaN   NaN   NaN  NaN NaN NaN NaN   NaN    NaN  NaN

   D13  D14    D15   M1   M2   M3   M4   M5   M6   M7   M8   M9   V1   V2  \
0  NaN  NaN    0.0    T    T    T   M2    F    T  NaN  NaN  NaN  1.0  1.0
1  NaN  NaN    0.0  NaN  NaN  NaN   M0    T    T  NaN  NaN  NaN  NaN  NaN
2  NaN  NaN  315.0    T    T    T   M0    F    F    F    F    F  1.0  1.0
3  NaN  NaN  111.0  NaN  NaN  NaN   M0    T    F  NaN  NaN  NaN  NaN  NaN
4  NaN  NaN    NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

    V3   V4   V5   V6   V7   V8   V9  V10  V11  V12  V13  V14  V15  V16  V17  \
0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0
1  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  0.0  0.0  1.0  0.0  0.0  0.0
```

```
2  1.0  1.0  1.0  1.0  1.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0
3  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  1.0  1.0  1.0  0.0  0.0  0.0
4  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

   V18  V19  V20  V21  V22  V23  V24  V25  V26  V27  V28  V29  V30  V31  V32  \
0  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0
1  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0
2  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0
3  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0
4  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

   V33  V34  V35  V36  V37  V38  V39  V40  V41  V42  V43  V44  V45  V46  V47  \
0  0.0  0.0  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN
1  0.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  1.0  0.0  0.0  1.0  1.0  1.0  1.0
2  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  1.0  0.0  0.0  1.0  1.0  1.0  1.0
3  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  1.0  0.0  0.0  1.0  1.0  1.0  1.0
4  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

   V48  V49  V50  V51  V52  V53  V54  V55  V56  V57  V58  V59  V60  V61  V62  \
0  NaN  NaN  NaN  NaN  NaN  1.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  1.0  1.0
1  0.0  0.0  0.0  0.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  0.0  0.0  1.0  1.0
2  0.0  0.0  0.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  1.0  1.0
3  0.0  0.0  0.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  1.0  1.0
4  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

   V63  V64  V65  V66  V67  V68  V69  V70  V71  V72  V73  V74  V75  V76  V77  \
0  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  1.0  1.0  1.0
1  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  1.0
2  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  1.0  1.0  1.0
3  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  1.0  1.0  1.0
4  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

   V78  V79  V80  V81  V82  V83  V84  V85  V86  V87  V88  V89  V90  V91  V92  \
0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0
1  1.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0
2  1.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0
3  1.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0
4  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

   V93  V94  V95   V96   V97  V98   V99  V100  V101  V102  V103  V104  V105  \
0  0.0  0.0  0.0   1.0   0.0  0.0   0.0   0.0   0.0   1.0   0.0   0.0   0.0
1  0.0  0.0  0.0   0.0   0.0  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
2  0.0  0.0  0.0   0.0   0.0  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
3  0.0  0.0  1.0  48.0  28.0  0.0  10.0   4.0   1.0  38.0  24.0   0.0   0.0
4  NaN  NaN  0.0   0.0   0.0  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0

   V106  V107  V108  V109  V110  V111  V112  V113  V114  V115  V116  V117  \
```

```
0   0.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0
1   0.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0
2   0.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0
3   0.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0
4   0.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0

   V118  V119  V120  V121  V122  V123  V124  V125  V126    V127   V128  V129  \
0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   0.0   117.0    0.0   0.0
1   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   0.0     0.0    0.0   0.0
2   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   0.0     0.0    0.0   0.0
3   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0  50.0  1758.0  925.0   0.0
4   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   0.0     0.0    0.0   0.0

    V130   V131  V132    V133   V134  V135  V136  V137  V138  V139  V140  \
0    0.0    0.0   0.0   117.0    0.0   0.0   0.0   0.0   NaN   NaN   NaN
1    0.0    0.0   0.0     0.0    0.0   0.0   0.0   0.0   NaN   NaN   NaN
2    0.0    0.0   0.0     0.0    0.0   0.0   0.0   0.0   NaN   NaN   NaN
3  354.0  135.0  50.0  1404.0  790.0   0.0   0.0   0.0   NaN   NaN   NaN
4    0.0    0.0   0.0     0.0    0.0   0.0   0.0   0.0   0.0   0.0   0.0

   V141  V142  V143  V144   V145  V146  V147  V148  V149    V150  V151  V152  \
0   NaN   NaN   NaN   NaN    NaN   NaN   NaN   NaN   NaN     NaN   NaN   NaN
1   NaN   NaN   NaN   NaN    NaN   NaN   NaN   NaN   NaN     NaN   NaN   NaN
2   NaN   NaN   NaN   NaN    NaN   NaN   NaN   NaN   NaN     NaN   NaN   NaN
3   NaN   NaN   NaN   NaN    NaN   NaN   NaN   NaN   NaN     NaN   NaN   NaN
4   0.0   0.0   6.0  18.0  140.0   0.0   0.0   0.0   0.0  1803.0  49.0  64.0

   V153  V154  V155  V156  V157  V158     V159          V160  V161  V162  \
0   NaN   NaN   NaN   NaN   NaN   NaN      NaN           NaN   NaN   NaN
1   NaN   NaN   NaN   NaN   NaN   NaN      NaN           NaN   NaN   NaN
2   NaN   NaN   NaN   NaN   NaN   NaN      NaN           NaN   NaN   NaN
3   NaN   NaN   NaN   NaN   NaN   NaN      NaN           NaN   NaN   NaN
4   0.0   0.0   0.0   0.0   0.0   0.0  15560.0  169690.796875   0.0   0.0

   V163   V164    V165    V166  V167  V168  V169  V170  V171  V172  V173  \
0   NaN    NaN     NaN     NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
1   NaN    NaN     NaN     NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
2   NaN    NaN     NaN     NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
3   NaN    NaN     NaN     NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
4   0.0  515.0  5155.0  2840.0   0.0   0.0   0.0   1.0   1.0   0.0   0.0

   V174  V175  V176  V177  V178  V179  V180  V181  V182  V183  V184  V185  \
0   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
1   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
2   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
3   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
4   0.0   0.0   1.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
```

```
   V186  V187  V188  V189  V190  V191  V192  V193  V194  V195  V196  V197  \
0   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
1   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
2   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
3   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
4   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0

   …  V134_missing_flag  V135_missing_flag  V136_missing_flag  \
0  …              False              False              False
1  …              False              False              False
2  …              False              False              False
3  …              False              False              False
4  …              False              False              False

   V137_missing_flag  V138_missing_flag  V139_missing_flag  V140_missing_flag  \
0              False               True               True               True
1              False               True               True               True
2              False               True               True               True
3              False               True               True               True
4              False              False              False              False

   V141_missing_flag  V142_missing_flag  V143_missing_flag  V144_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V145_missing_flag  V146_missing_flag  V147_missing_flag  V148_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V149_missing_flag  V150_missing_flag  V151_missing_flag  V152_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V153_missing_flag  V154_missing_flag  V155_missing_flag  V156_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
```

```
3              True              True              True              True
4             False             False             False             False

   V157_missing_flag  V158_missing_flag  V159_missing_flag  V160_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V161_missing_flag  V162_missing_flag  V163_missing_flag  V164_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V165_missing_flag  V166_missing_flag  V167_missing_flag  V168_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V169_missing_flag  V170_missing_flag  V171_missing_flag  V172_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V173_missing_flag  V174_missing_flag  V175_missing_flag  V176_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V177_missing_flag  V178_missing_flag  V179_missing_flag  V180_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V181_missing_flag  V182_missing_flag  V183_missing_flag  V184_missing_flag  \
0               True               True               True               True
```

```
1              True              True              True              True
2              True              True              True              True
3              True              True              True              True
4             False             False             False             False

   V185_missing_flag  V186_missing_flag  V187_missing_flag  V188_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V189_missing_flag  V190_missing_flag  V191_missing_flag  V192_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V193_missing_flag  V194_missing_flag  V195_missing_flag  V196_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V197_missing_flag  V198_missing_flag  V199_missing_flag  V200_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V201_missing_flag  V202_missing_flag  V203_missing_flag  V204_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V205_missing_flag  V206_missing_flag  V207_missing_flag  V208_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False
```

```
   V209_missing_flag  V210_missing_flag  V211_missing_flag  V212_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V213_missing_flag  V214_missing_flag  V215_missing_flag  V216_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V217_missing_flag  V218_missing_flag  V219_missing_flag  V220_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V221_missing_flag  V222_missing_flag  V223_missing_flag  V224_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V225_missing_flag  V226_missing_flag  V227_missing_flag  V228_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V229_missing_flag  V230_missing_flag  V231_missing_flag  V232_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V233_missing_flag  V234_missing_flag  V235_missing_flag  V236_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
```

```
4             False             False             False             False

   V237_missing_flag  V238_missing_flag  V239_missing_flag  V240_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V241_missing_flag  V242_missing_flag  V243_missing_flag  V244_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V245_missing_flag  V246_missing_flag  V247_missing_flag  V248_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V249_missing_flag  V250_missing_flag  V251_missing_flag  V252_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V253_missing_flag  V254_missing_flag  V255_missing_flag  V256_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V257_missing_flag  V258_missing_flag  V259_missing_flag  V260_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V261_missing_flag  V262_missing_flag  V263_missing_flag  V264_missing_flag  \
0               True               True               True               True
1               True               True               True               True
```

```
2              True               True               True               True
3              True               True               True               True
4             False              False              False              False

   V265_missing_flag  V266_missing_flag  V267_missing_flag  V268_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V269_missing_flag  V270_missing_flag  V271_missing_flag  V272_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V273_missing_flag  V274_missing_flag  V275_missing_flag  V276_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V277_missing_flag  V278_missing_flag  V279_missing_flag  V280_missing_flag  \
0               True               True              False              False
1               True               True              False              False
2               True               True              False              False
3               True               True              False              False
4              False              False              False              False

   V281_missing_flag  V282_missing_flag  V283_missing_flag  V284_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

  V285_missing_flag  V286_missing_flag  V287_missing_flag V288_missing_flag  \
0             False              False              False             False
1             False              False              False             False
2             False              False              False             False
3             False              False              False             False
4             False              False              False             False

  V289_missing_flag  V290_missing_flag  V291_missing_flag  V292_missing_flag  \
```

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

  V293_missing_flag V294_missing_flag V295_missing_flag V296_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4             False             False             False             False

   V297_missing_flag V298_missing_flag V299_missing_flag V300_missing_flag  \
0              False             False             False             False
1              False             False             False             False
2              False             False             False             False
3              False             False             False             False
4              False             False             False             False

  V301_missing_flag V302_missing_flag V303_missing_flag V304_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4             False             False             False             False

  V305_missing_flag  V306_missing_flag  V307_missing_flag  V308_missing_flag  \
0             False              False              False              False
1             False              False              False              False
2             False              False              False              False
3             False              False              False              False
4             False              False              False              False

   V309_missing_flag  V310_missing_flag  V311_missing_flag  V312_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V313_missing_flag  V314_missing_flag  V315_missing_flag  V316_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False
```

```
   V317_missing_flag  V318_missing_flag  V319_missing_flag  V320_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V321_missing_flag  V322_missing_flag  V323_missing_flag  V324_missing_flag  \
0              False               True               True               True
1              False               True               True               True
2              False               True               True               True
3              False               True               True               True
4              False              False              False              False

   V325_missing_flag  V326_missing_flag  V327_missing_flag  V328_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V329_missing_flag  V330_missing_flag  V331_missing_flag  V332_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V333_missing_flag  V334_missing_flag  V335_missing_flag  V336_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V337_missing_flag  V338_missing_flag  V339_missing_flag  \
0               True               True               True
1               True               True               True
2               True               True               True
3               True               True               True
4              False              False              False

   id_01_missing_flag  id_02_missing_flag  id_03_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
```

```
3                True                True                True
4               False               False                True

   id_04_missing_flag  id_05_missing_flag  id_06_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True                True                True

   id_07_missing_flag  id_08_missing_flag  id_09_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True                True                True

   id_10_missing_flag  id_11_missing_flag  id_12_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True               False               False

   id_13_missing_flag  id_14_missing_flag  id_15_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True               False               False

   id_16_missing_flag  id_17_missing_flag  id_18_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4               False               False                True

   id_19_missing_flag  id_20_missing_flag  id_21_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4               False               False                True

   id_22_missing_flag  id_23_missing_flag  id_24_missing_flag  \
0                True                True                True
```

```
1                True                True                True
2                True                True                True
3                True                True                True
4                True                True                True

   id_25_missing_flag  id_26_missing_flag  id_27_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True                True                True

   id_28_missing_flag  id_29_missing_flag  id_30_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4               False               False               False

   id_31_missing_flag  id_32_missing_flag  id_33_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4               False               False               False

   id_34_missing_flag  id_35_missing_flag  id_36_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4               False               False               False

   id_37_missing_flag  id_38_missing_flag  DeviceType_missing_flag  \
0                True                True                     True
1                True                True                     True
2                True                True                     True
3                True                True                     True
4               False               False                    False

   DeviceInfo_missing_flag                Date  _Weekdays  _Hours  _Days
0                     True 2017-12-02 00:00:00          5       0      2
1                     True 2017-12-02 00:00:01          5       0      2
2                     True 2017-12-02 00:01:09          5       0      2
3                     True 2017-12-02 00:01:39          5       0      2
4                    False 2017-12-02 00:01:46          5       0      2
```

```
[5 rows x 840 columns]
```

7.0.1 Domain Specific Features

[57]:
```
df['Trans_min_mean'] = df['TransactionAmt'] - np.
 ↪nanmean(df['TransactionAmt'],dtype="float64")
df['Trans_min_std']  = df['Trans_min_mean'] / np.nanstd(df['TransactionAmt'].
 ↪astype("float64"),dtype="float64")
```

[58]:
```
df['TransactionAmt_to_mean_card1'] = df['TransactionAmt'] / df.
 ↪groupby(['card1'])['TransactionAmt'].transform('mean')
df['TransactionAmt_to_mean_card4'] = df['TransactionAmt'] / df.
 ↪groupby(['card4'])['TransactionAmt'].transform('mean')
df['TransactionAmt_to_std_card1']  = df['TransactionAmt'] / df.
 ↪groupby(['card1'])['TransactionAmt'].transform('std')
df['TransactionAmt_to_std_card4']  = df['TransactionAmt'] / df.
 ↪groupby(['card4'])['TransactionAmt'].transform('std')
```

[59]:
```
df['TransactionAmt'] = np.log(df['TransactionAmt'])
```

[60]:
```
df.head()
```

[60]:

	TransactionID	isFraud	TransactionDT	TransactionAmt	ProductCD	card1	\
0	2987000	0	86400	4.226834	W	13926	
1	2987001	0	86401	3.367296	W	2755	
2	2987002	0	86469	4.077537	W	4663	
3	2987003	0	86499	3.912023	W	18132	
4	2987004	0	86506	3.912023	H	4497	

	card2	card3	card4	card5	card6	addr1	addr2	dist1	P_emaildomain	\
0	NaN	150.0	discover	142.0	credit	315.0	87.0	19.0	NoInf	
1	404.0	150.0	mastercard	102.0	credit	325.0	87.0	NaN	Google	
2	490.0	150.0	visa	166.0	debit	330.0	87.0	287.0	Microsoft	
3	567.0	150.0	mastercard	117.0	debit	476.0	87.0	NaN	Yahoo Mail	
4	514.0	150.0	mastercard	102.0	credit	420.0	87.0	NaN	Google	

	R_emaildomain	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	\
0	NoInf	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	2.0	0.0	
1	NoInf	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	
2	NoInf	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0	
3	NoInf	2.0	5.0	0.0	0.0	0.0	4.0	0.0	0.0	1.0	0.0	1.0	0.0	
4	NoInf	1.0	1.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0	1.0	0.0	

	C13	C14	D1	D2	D3	D4	D5	D6	D8	D9	D10	D11	D12	\
0	1.0	1.0	14.0	NaN	13.0	NaN	NaN	NaN	NaN	NaN	13.0	13.0	NaN	
1	1.0	1.0	0.0	NaN	NaN	0.0	NaN	NaN	NaN	NaN	0.0	NaN	NaN	

```
2   1.0  1.0    0.0    NaN   NaN   0.0  NaN NaN NaN NaN   0.0  315.0  NaN
3  25.0  1.0  112.0  112.0   0.0  94.0  0.0 NaN NaN NaN  84.0    NaN  NaN
4   1.0  1.0    0.0    NaN   NaN   NaN  NaN NaN NaN NaN   NaN    NaN  NaN

   D13  D14    D15   M1   M2   M3   M4   M5   M6   M7   M8   M9   V1   V2  \
0  NaN  NaN    0.0    T    T    T   M2    F    T  NaN  NaN  NaN  1.0  1.0
1  NaN  NaN    0.0  NaN  NaN  NaN   M0    T    T  NaN  NaN  NaN  NaN  NaN
2  NaN  NaN  315.0    T    T    T   M0    F    F    F    F    F  1.0  1.0
3  NaN  NaN  111.0  NaN  NaN  NaN   M0    T    F  NaN  NaN  NaN  NaN  NaN
4  NaN  NaN    NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

    V3   V4   V5   V6   V7   V8   V9  V10  V11  V12  V13  V14  V15  V16  V17  \
0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0
1  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  0.0  0.0  1.0  0.0  0.0  0.0
2  1.0  1.0  1.0  1.0  1.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0
3  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  1.0  1.0  1.0  0.0  0.0  0.0
4  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

   V18  V19  V20  V21  V22  V23  V24  V25  V26  V27  V28  V29  V30  V31  V32  \
0  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0
1  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0
2  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0
3  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0
4  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

   V33  V34  V35  V36  V37  V38  V39  V40  V41  V42  V43  V44  V45  V46  V47  \
0  0.0  0.0  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN
1  0.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  1.0  0.0  0.0  1.0  1.0  1.0  1.0
2  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  1.0  0.0  0.0  1.0  1.0  1.0  1.0
3  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  1.0  0.0  0.0  1.0  1.0  1.0  1.0
4  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

   V48  V49  V50  V51  V52  V53  V54  V55  V56  V57  V58  V59  V60  V61  V62  \
0  NaN  NaN  NaN  NaN  NaN  1.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  1.0  1.0
1  0.0  0.0  0.0  0.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  0.0  0.0  1.0  1.0
2  0.0  0.0  0.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  1.0  1.0
3  0.0  0.0  0.0  0.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  1.0  1.0
4  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

   V63  V64  V65  V66  V67  V68  V69  V70  V71  V72  V73  V74  V75  V76  V77  \
0  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  1.0  1.0  1.0
1  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  1.0
2  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  1.0  1.0  1.0
3  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  1.0  1.0  1.0
4  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

   V78  V79  V80  V81  V82  V83  V84  V85  V86  V87  V88  V89  V90  V91  V92  \
```

```
0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0
1  1.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0
2  1.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0
3  1.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0  1.0  1.0  1.0  0.0  0.0  0.0  0.0
4  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN

   V93  V94  V95   V96   V97  V98   V99  V100  V101  V102  V103  V104  V105  \
0  0.0  0.0  0.0   1.0   0.0  0.0   0.0   0.0   0.0   1.0   0.0   0.0   0.0
1  0.0  0.0  0.0   0.0   0.0  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
2  0.0  0.0  0.0   0.0   0.0  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
3  0.0  0.0  1.0  48.0  28.0  0.0  10.0   4.0   1.0  38.0  24.0   0.0   0.0
4  NaN  NaN  0.0   0.0   0.0  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0

   V106  V107  V108  V109  V110  V111  V112  V113  V114  V115  V116  V117  \
0   0.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0
1   0.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0
2   0.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0
3   0.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0
4   0.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0

   V118  V119  V120  V121  V122  V123  V124  V125  V126    V127   V128  V129  \
0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   0.0   117.0    0.0   0.0
1   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   0.0     0.0    0.0   0.0
2   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   0.0     0.0    0.0   0.0
3   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0  50.0  1758.0  925.0   0.0
4   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   0.0     0.0    0.0   0.0

    V130   V131  V132    V133   V134  V135  V136  V137  V138  V139  V140  \
0    0.0    0.0   0.0   117.0    0.0   0.0   0.0   0.0   NaN   NaN   NaN
1    0.0    0.0   0.0     0.0    0.0   0.0   0.0   0.0   NaN   NaN   NaN
2    0.0    0.0   0.0     0.0    0.0   0.0   0.0   0.0   NaN   NaN   NaN
3  354.0  135.0  50.0  1404.0  790.0   0.0   0.0   0.0   NaN   NaN   NaN
4    0.0    0.0   0.0     0.0    0.0   0.0   0.0   0.0   0.0   0.0   0.0

   V141  V142  V143  V144   V145  V146  V147  V148  V149    V150  V151  V152  \
0   NaN   NaN   NaN   NaN    NaN   NaN   NaN   NaN   NaN     NaN   NaN   NaN
1   NaN   NaN   NaN   NaN    NaN   NaN   NaN   NaN   NaN     NaN   NaN   NaN
2   NaN   NaN   NaN   NaN    NaN   NaN   NaN   NaN   NaN     NaN   NaN   NaN
3   NaN   NaN   NaN   NaN    NaN   NaN   NaN   NaN   NaN     NaN   NaN   NaN
4   0.0   0.0   6.0  18.0  140.0   0.0   0.0   0.0   0.0  1803.0  49.0  64.0

   V153  V154  V155  V156  V157  V158     V159          V160  V161  V162  \
0   NaN   NaN   NaN   NaN   NaN   NaN      NaN           NaN   NaN   NaN
1   NaN   NaN   NaN   NaN   NaN   NaN      NaN           NaN   NaN   NaN
2   NaN   NaN   NaN   NaN   NaN   NaN      NaN           NaN   NaN   NaN
3   NaN   NaN   NaN   NaN   NaN   NaN      NaN           NaN   NaN   NaN
4   0.0   0.0   0.0   0.0   0.0   0.0  15560.0  169690.796875   0.0   0.0
```

```
   V163   V164    V165    V166  V167  V168  V169  V170  V171  V172  V173  \
0   NaN    NaN     NaN     NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
1   NaN    NaN     NaN     NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
2   NaN    NaN     NaN     NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
3   NaN    NaN     NaN     NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
4   0.0  515.0  5155.0  2840.0   0.0   0.0   0.0   1.0   1.0   0.0   0.0

   V174  V175  V176  V177  V178  V179  V180  V181  V182  V183  V184  V185  \
0   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
1   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
2   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
3   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
4   0.0   0.0   1.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0

   V186  V187  V188  V189  V190  V191  V192  V193  V194  V195  V196  V197  \
0   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
1   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
2   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
3   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
4   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0

   …  V140_missing_flag  V141_missing_flag  V142_missing_flag  \
0  …               True               True               True
1  …               True               True               True
2  …               True               True               True
3  …               True               True               True
4  …              False              False              False

   V143_missing_flag  V144_missing_flag  V145_missing_flag  V146_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V147_missing_flag  V148_missing_flag  V149_missing_flag  V150_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V151_missing_flag  V152_missing_flag  V153_missing_flag  V154_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
```

```
3              True              True              True              True
4             False             False             False             False

   V155_missing_flag  V156_missing_flag  V157_missing_flag  V158_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V159_missing_flag  V160_missing_flag  V161_missing_flag  V162_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V163_missing_flag  V164_missing_flag  V165_missing_flag  V166_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V167_missing_flag  V168_missing_flag  V169_missing_flag  V170_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V171_missing_flag  V172_missing_flag  V173_missing_flag  V174_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V175_missing_flag  V176_missing_flag  V177_missing_flag  V178_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V179_missing_flag  V180_missing_flag  V181_missing_flag  V182_missing_flag  \
0               True               True               True               True
```

```
1              True              True              True              True
2              True              True              True              True
3              True              True              True              True
4             False             False             False             False

   V183_missing_flag  V184_missing_flag  V185_missing_flag  V186_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V187_missing_flag  V188_missing_flag  V189_missing_flag  V190_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V191_missing_flag  V192_missing_flag  V193_missing_flag  V194_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V195_missing_flag  V196_missing_flag  V197_missing_flag  V198_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V199_missing_flag  V200_missing_flag  V201_missing_flag  V202_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V203_missing_flag  V204_missing_flag  V205_missing_flag  V206_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False
```

```
   V207_missing_flag  V208_missing_flag  V209_missing_flag  V210_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V211_missing_flag  V212_missing_flag  V213_missing_flag  V214_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V215_missing_flag  V216_missing_flag  V217_missing_flag  V218_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V219_missing_flag  V220_missing_flag  V221_missing_flag  V222_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V223_missing_flag  V224_missing_flag  V225_missing_flag  V226_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V227_missing_flag  V228_missing_flag  V229_missing_flag  V230_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V231_missing_flag  V232_missing_flag  V233_missing_flag  V234_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
```

```
4             False             False             False             False

   V235_missing_flag  V236_missing_flag  V237_missing_flag  V238_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V239_missing_flag  V240_missing_flag  V241_missing_flag  V242_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V243_missing_flag  V244_missing_flag  V245_missing_flag  V246_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V247_missing_flag  V248_missing_flag  V249_missing_flag  V250_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V251_missing_flag  V252_missing_flag  V253_missing_flag  V254_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V255_missing_flag  V256_missing_flag  V257_missing_flag  V258_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V259_missing_flag  V260_missing_flag  V261_missing_flag  V262_missing_flag  \
0               True               True               True               True
1               True               True               True               True
```

```
2              True              True              True              True
3              True              True              True              True
4             False             False             False             False

   V263_missing_flag  V264_missing_flag  V265_missing_flag  V266_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V267_missing_flag  V268_missing_flag  V269_missing_flag  V270_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V271_missing_flag  V272_missing_flag  V273_missing_flag  V274_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V275_missing_flag  V276_missing_flag  V277_missing_flag  V278_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V279_missing_flag  V280_missing_flag  V281_missing_flag  V282_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V283_missing_flag  V284_missing_flag  V285_missing_flag  V286_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V287_missing_flag  V288_missing_flag  V289_missing_flag  V290_missing_flag  \
```

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

  V291_missing_flag  V292_missing_flag  V293_missing_flag V294_missing_flag  \
0             False              False              False             False
1             False              False              False             False
2             False              False              False             False
3             False              False              False             False
4             False              False              False             False

  V295_missing_flag  V296_missing_flag  V297_missing_flag  V298_missing_flag  \
0             False              False              False              False
1             False              False              False              False
2             False              False              False              False
3             False              False              False              False
4             False              False              False              False

  V299_missing_flag V300_missing_flag V301_missing_flag V302_missing_flag  \
0             False             False              False             False
1             False             False              False             False
2             False             False              False             False
3             False             False              False             False
4             False             False              False             False

   V303_missing_flag V304_missing_flag V305_missing_flag V306_missing_flag  \
0              False             False              False             False
1              False             False              False             False
2              False             False              False             False
3              False             False              False             False
4              False             False              False             False

  V307_missing_flag V308_missing_flag V309_missing_flag V310_missing_flag  \
0             False             False              False             False
1             False             False              False             False
2             False             False              False             False
3             False             False              False             False
4             False             False              False             False

  V311_missing_flag  V312_missing_flag  V313_missing_flag  V314_missing_flag  \
0             False              False              False              False
1             False              False              False              False
2             False              False              False              False
3             False              False              False              False
4             False              False              False              False
```

```
   V315_missing_flag  V316_missing_flag  V317_missing_flag  V318_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V319_missing_flag  V320_missing_flag  V321_missing_flag  V322_missing_flag  \
0              False              False              False               True
1              False              False              False               True
2              False              False              False               True
3              False              False              False               True
4              False              False              False              False

   V323_missing_flag  V324_missing_flag  V325_missing_flag  V326_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V327_missing_flag  V328_missing_flag  V329_missing_flag  V330_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V331_missing_flag  V332_missing_flag  V333_missing_flag  V334_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V335_missing_flag  V336_missing_flag  V337_missing_flag  V338_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V339_missing_flag  id_01_missing_flag  id_02_missing_flag  \
0               True                True                True
1               True                True                True
2               True                True                True
```

```
3               True                True                True
4              False               False               False

   id_03_missing_flag  id_04_missing_flag  id_05_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True                True                True

   id_06_missing_flag  id_07_missing_flag  id_08_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True                True                True

   id_09_missing_flag  id_10_missing_flag  id_11_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True                True               False

   id_12_missing_flag  id_13_missing_flag  id_14_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4               False                True               False

   id_15_missing_flag  id_16_missing_flag  id_17_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4               False               False               False

   id_18_missing_flag  id_19_missing_flag  id_20_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True               False               False

   id_21_missing_flag  id_22_missing_flag  id_23_missing_flag  \
0                True                True                True
```

```
1                True                True                True
2                True                True                True
3                True                True                True
4                True                True                True

   id_24_missing_flag  id_25_missing_flag  id_26_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True                True                True

   id_27_missing_flag  id_28_missing_flag  id_29_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True               False               False

   id_30_missing_flag  id_31_missing_flag  id_32_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4               False               False               False

   id_33_missing_flag  id_34_missing_flag  id_35_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4               False               False               False

   id_36_missing_flag  id_37_missing_flag  id_38_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4               False               False               False

   DeviceType_missing_flag  DeviceInfo_missing_flag                Date  \
0                     True                     True 2017-12-02 00:00:00
1                     True                     True 2017-12-02 00:00:01
2                     True                     True 2017-12-02 00:01:09
3                     True                     True 2017-12-02 00:01:39
4                    False                    False 2017-12-02 00:01:46
```

```
   _Weekdays  _Hours  _Days  Trans_min_mean  Trans_min_std  \
0          5       0      2      -66.527347      -0.278174
1          5       0      2     -106.027347      -0.443337
2          5       0      2      -76.027347      -0.317897
3          5       0      2      -85.027347      -0.355529
4          5       0      2      -85.027347      -0.355529

   TransactionAmt_to_mean_card1  TransactionAmt_to_mean_card4  \
0                      0.194638                      0.257761
1                      0.123780                      0.219053
2                      0.608151                      0.443070
3                      0.405136                      0.377678
4                      0.515616                      0.377678

   TransactionAmt_to_std_card1  TransactionAmt_to_std_card4
0                     0.184560                     0.170241
1                     0.063004                     0.114214
2                     0.589226                     0.258550
3                     0.259460                     0.196921
4                     0.882898                     0.196921

[5 rows x 846 columns]
```

8 8. Dimensionality Reduction - PCA

[61]:
```
def perform_PCA(df, cols, n_components, prefix='PCA_', rand_seed=4):
    pca = PCA(n_components=n_components, random_state=rand_seed)
    principalComponents = pca.fit_transform(df[cols])
    principalDf = pd.DataFrame(principalComponents)
    df.drop(cols, axis=1, inplace=True)

    principalDf.rename(columns=lambda x: str(prefix)+str(x), inplace=True)
    df = pd.concat([df, principalDf], axis=1)
    return df
```

[62]:
```
# Columns starting from V1 to V339
filter_col = df.columns[53:392]
```

Impute missing values in the mas_v columns, later use minmax_scale function to scale the values in these columns

[63]:
```
from   sklearn.preprocessing import minmax_scale

# Fill na values and scale V columns
for col in filter_col:
    df[col] = df[col].fillna((df[col].min() - 2))
```

```
    df[col] = (minmax_scale(df[col], feature_range=(0,1)))

# Perform PCA
df = perform_PCA(df, filter_col, prefix='PCA_V_', n_components=30)
```

[64]:
```
df = reduce_mem_usage(df)
```

```
Mem. usage decreased to 1170.85 Mb (4.4% reduction)
```

[65]:
```
df.head()
```

[65]:

	TransactionID	isFraud	TransactionDT	TransactionAmt	ProductCD	card1
0	2987000	0	86400	4.226562	W	13926
1	2987001	0	86401	3.367188	W	2755
2	2987002	0	86469	4.078125	W	4663
3	2987003	0	86499	3.912109	W	18132
4	2987004	0	86506	3.912109	H	4497

	card2	card3	card4	card5	card6	addr1	addr2	dist1	P_emaildomain
0	NaN	150.0	discover	142.0	credit	315.0	87.0	19.0	NoInf
1	404.0	150.0	mastercard	102.0	credit	325.0	87.0	NaN	Google
2	490.0	150.0	visa	166.0	debit	330.0	87.0	287.0	Microsoft
3	567.0	150.0	mastercard	117.0	debit	476.0	87.0	NaN	Yahoo Mail
4	514.0	150.0	mastercard	102.0	credit	420.0	87.0	NaN	Google

	R_emaildomain	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
0	NoInf	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	2.0	0.0
1	NoInf	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0
2	NoInf	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0
3	NoInf	2.0	5.0	0.0	0.0	0.0	4.0	0.0	0.0	1.0	0.0	1.0	0.0
4	NoInf	1.0	1.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0	1.0	0.0

	C13	C14	D1	D2	D3	D4	D5	D6	D8	D9	D10	D11	D12
0	1.0	1.0	14.0	NaN	13.0	NaN	NaN	NaN	NaN	NaN	13.0	13.0	NaN
1	1.0	1.0	0.0	NaN	NaN	0.0	NaN	NaN	NaN	NaN	0.0	NaN	NaN
2	1.0	1.0	0.0	NaN	NaN	0.0	NaN	NaN	NaN	NaN	0.0	315.0	NaN
3	25.0	1.0	112.0	112.0	0.0	94.0	0.0	NaN	NaN	NaN	84.0	NaN	NaN
4	1.0	1.0	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

	D13	D14	D15	M1	M2	M3	M4	M5	M6	M7	M8	M9	id_01
0	NaN	NaN	0.0	T	T	T	M2	F	T	NaN	NaN	NaN	NaN
1	NaN	NaN	0.0	NaN	NaN	NaN	M0	T	T	NaN	NaN	NaN	NaN
2	NaN	NaN	315.0	T	T	T	M0	F	F	F	F	F	NaN
3	NaN	NaN	111.0	NaN	NaN	NaN	M0	T	F	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.0

	id_02	id_03	id_04	id_05	id_06	id_09	id_10	id_11	id_12	id_13

```
0      NaN    NaN    NaN    NaN    NaN    NaN    NaN    NaN       NaN    NaN
1      NaN    NaN    NaN    NaN    NaN    NaN    NaN    NaN       NaN    NaN
2      NaN    NaN    NaN    NaN    NaN    NaN    NaN    NaN       NaN    NaN
3      NaN    NaN    NaN    NaN    NaN    NaN    NaN    NaN       NaN    NaN
4  70787.0    NaN    NaN    NaN    NaN    NaN    NaN  100.0  NotFound    NaN

   id_14 id_15     id_16  id_17  id_19  id_20 id_28     id_29    id_30  \
0    NaN   NaN       NaN    NaN    NaN    NaN   NaN       NaN      NAN
1    NaN   NaN       NaN    NaN    NaN    NaN   NaN       NaN      NAN
2    NaN   NaN       NaN    NaN    NaN    NaN   NaN       NaN      NAN
3    NaN   NaN       NaN    NaN    NaN    NaN   NaN       NaN      NAN
4 -480.0   New  NotFound  166.0  542.0  144.0   New  NotFound  Android

                id_31  id_32      id_33           id_34 id_35 id_36 id_37  \
0                 NaN    NaN        NaN             NaN   NaN   NaN   NaN
1                 NaN    NaN        NaN             NaN   NaN   NaN   NaN
2                 NaN    NaN        NaN             NaN   NaN   NaN   NaN
3                 NaN    NaN        NaN             NaN   NaN   NaN   NaN
4  samsung browser 6.2   32.0  2220x1080  match_status:2     T     F     T

  id_38 DeviceType                    DeviceInfo  card2_missing_flag  \
0   NaN        NaN                           NaN                True
1   NaN        NaN                           NaN               False
2   NaN        NaN                           NaN               False
3   NaN        NaN                           NaN               False
4     T     mobile  SAMSUNG SM-G892A Build/NRD90M               False

   card3_missing_flag  card4_missing_flag  card5_missing_flag  \
0               False               False               False
1               False               False               False
2               False               False               False
3               False               False               False
4               False               False               False

   card6_missing_flag  addr1_missing_flag  addr2_missing_flag  \
0               False               False               False
1               False               False               False
2               False               False               False
3               False               False               False
4               False               False               False

   dist1_missing_flag  dist2_missing_flag  P_emaildomain_missing_flag  \
0               False                True                        True
1                True                True                       False
2               False                True                       False
3                True                True                       False
4                True                True                       False
```

```
   R_emaildomain_missing_flag  D1_missing_flag  D2_missing_flag  \
0                        True            False             True
1                        True            False             True
2                        True            False             True
3                        True            False            False
4                        True            False             True

   D3_missing_flag  D4_missing_flag  D5_missing_flag  D6_missing_flag  \
0            False             True             True             True
1             True            False             True             True
2             True            False             True             True
3            False            False            False             True
4             True             True             True             True

   D7_missing_flag  D8_missing_flag  D9_missing_flag  D10_missing_flag  \
0             True             True             True             False
1             True             True             True             False
2             True             True             True             False
3             True             True             True             False
4             True             True             True              True

   D11_missing_flag  D12_missing_flag  D13_missing_flag  D14_missing_flag  \
0             False              True              True              True
1              True              True              True              True
2             False              True              True              True
3              True              True              True              True
4              True              True              True              True

   D15_missing_flag  M1_missing_flag  M2_missing_flag  M3_missing_flag  \
0             False            False            False            False
1             False             True             True             True
2             False            False            False            False
3             False             True             True             True
4              True             True             True             True

   M4_missing_flag  M5_missing_flag  M6_missing_flag  M7_missing_flag  \
0            False            False            False             True
1            False            False            False             True
2            False            False            False            False
3            False            False            False             True
4             True             True             True             True

   M8_missing_flag  M9_missing_flag  V1_missing_flag  V2_missing_flag  \
0             True             True            False            False
1             True             True             True             True
2            False            False            False            False
```

```
3             True             True             True             True
4             True             True             True             True

   V3_missing_flag  V4_missing_flag  V5_missing_flag  V6_missing_flag  \
0            False            False            False            False
1             True             True             True             True
2            False            False            False            False
3             True             True             True             True
4             True             True             True             True

   V7_missing_flag  V8_missing_flag  V9_missing_flag  V10_missing_flag  \
0            False            False            False             False
1             True             True             True              True
2            False            False            False             False
3             True             True             True              True
4             True             True             True              True

   V11_missing_flag  V12_missing_flag  V13_missing_flag  V14_missing_flag  \
0             False             False             False             False
1              True             False             False             False
2             False             False             False             False
3              True             False             False             False
4              True              True              True              True

   V15_missing_flag  V16_missing_flag  V17_missing_flag  V18_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V19_missing_flag  V20_missing_flag  V21_missing_flag  V22_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V23_missing_flag  V24_missing_flag  V25_missing_flag  V26_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V27_missing_flag  V28_missing_flag  V29_missing_flag  V30_missing_flag  \
0             False             False             False             False
```

```
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V31_missing_flag  V32_missing_flag  V33_missing_flag  V34_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V35_missing_flag  V36_missing_flag  V37_missing_flag  V38_missing_flag  \
0              True              True              True              True
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V39_missing_flag  V40_missing_flag  V41_missing_flag  V42_missing_flag  \
0              True              True              True              True
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V43_missing_flag  V44_missing_flag  V45_missing_flag  V46_missing_flag  \
0              True              True              True              True
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V47_missing_flag  V48_missing_flag  V49_missing_flag  V50_missing_flag  \
0              True              True              True              True
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V51_missing_flag  V52_missing_flag  V53_missing_flag  V54_missing_flag  \
0              True              True             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True
```

```
   V55_missing_flag  V56_missing_flag  V57_missing_flag  V58_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V59_missing_flag  V60_missing_flag  V61_missing_flag  V62_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V63_missing_flag  V64_missing_flag  V65_missing_flag  V66_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V67_missing_flag  V68_missing_flag  V69_missing_flag  V70_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V71_missing_flag  V72_missing_flag  V73_missing_flag  V74_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V75_missing_flag  V76_missing_flag  V77_missing_flag  V78_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V79_missing_flag  V80_missing_flag  V81_missing_flag  V82_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
```

```
4             True             True             True             True

   V83_missing_flag  V84_missing_flag  V85_missing_flag  V86_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V87_missing_flag  V88_missing_flag  V89_missing_flag  V90_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V91_missing_flag  V92_missing_flag  V93_missing_flag  V94_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V95_missing_flag  V96_missing_flag  V97_missing_flag  V98_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4             False             False             False             False

   V99_missing_flag  V100_missing_flag  V101_missing_flag  V102_missing_flag  \
0             False              False              False              False
1             False              False              False              False
2             False              False              False              False
3             False              False              False              False
4             False              False              False              False

   V103_missing_flag  V104_missing_flag  V105_missing_flag  V106_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V107_missing_flag  V108_missing_flag  V109_missing_flag  V110_missing_flag  \
0              False              False              False              False
1              False              False              False              False
```

```
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V111_missing_flag  V112_missing_flag  V113_missing_flag  V114_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V115_missing_flag  V116_missing_flag  V117_missing_flag  V118_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V119_missing_flag  V120_missing_flag  V121_missing_flag  V122_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V123_missing_flag  V124_missing_flag  V125_missing_flag  V126_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V127_missing_flag  V128_missing_flag  V129_missing_flag  V130_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V131_missing_flag  V132_missing_flag  …  V170_missing_flag  \
0              False              False  …               True
1              False              False  …               True
2              False              False  …               True
3              False              False  …               True
4              False              False  …              False

   V171_missing_flag  V172_missing_flag  V173_missing_flag  V174_missing_flag  \
```

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

   V175_missing_flag  V176_missing_flag  V177_missing_flag  V178_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V179_missing_flag  V180_missing_flag  V181_missing_flag  V182_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V183_missing_flag  V184_missing_flag  V185_missing_flag  V186_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V187_missing_flag  V188_missing_flag  V189_missing_flag  V190_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V191_missing_flag  V192_missing_flag  V193_missing_flag  V194_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V195_missing_flag  V196_missing_flag  V197_missing_flag  V198_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False
```

```
   V199_missing_flag  V200_missing_flag  V201_missing_flag  V202_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V203_missing_flag  V204_missing_flag  V205_missing_flag  V206_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V207_missing_flag  V208_missing_flag  V209_missing_flag  V210_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V211_missing_flag  V212_missing_flag  V213_missing_flag  V214_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V215_missing_flag  V216_missing_flag  V217_missing_flag  V218_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V219_missing_flag  V220_missing_flag  V221_missing_flag  V222_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V223_missing_flag  V224_missing_flag  V225_missing_flag  V226_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
```

```
3              True              True              True              True
4             False             False             False             False

   V227_missing_flag  V228_missing_flag  V229_missing_flag  V230_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V231_missing_flag  V232_missing_flag  V233_missing_flag  V234_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V235_missing_flag  V236_missing_flag  V237_missing_flag  V238_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V239_missing_flag  V240_missing_flag  V241_missing_flag  V242_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V243_missing_flag  V244_missing_flag  V245_missing_flag  V246_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V247_missing_flag  V248_missing_flag  V249_missing_flag  V250_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V251_missing_flag  V252_missing_flag  V253_missing_flag  V254_missing_flag  \
0               True               True               True               True
```

```
1                True               True               True               True
2                True               True               True               True
3                True               True               True               True
4               False              False              False              False

   V255_missing_flag  V256_missing_flag  V257_missing_flag  V258_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V259_missing_flag  V260_missing_flag  V261_missing_flag  V262_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V263_missing_flag  V264_missing_flag  V265_missing_flag  V266_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V267_missing_flag  V268_missing_flag  V269_missing_flag  V270_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V271_missing_flag  V272_missing_flag  V273_missing_flag  V274_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V275_missing_flag  V276_missing_flag  V277_missing_flag  V278_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False
```

```
   V279_missing_flag  V280_missing_flag  V281_missing_flag  V282_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V283_missing_flag  V284_missing_flag  V285_missing_flag  V286_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V287_missing_flag  V288_missing_flag  V289_missing_flag  V290_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V291_missing_flag  V292_missing_flag  V293_missing_flag  V294_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V295_missing_flag  V296_missing_flag  V297_missing_flag  V298_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V299_missing_flag  V300_missing_flag  V301_missing_flag  V302_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V303_missing_flag  V304_missing_flag  V305_missing_flag  V306_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
```

```
4             False             False             False             False

   V307_missing_flag  V308_missing_flag  V309_missing_flag  V310_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V311_missing_flag  V312_missing_flag  V313_missing_flag  V314_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V315_missing_flag  V316_missing_flag  V317_missing_flag  V318_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V319_missing_flag  V320_missing_flag  V321_missing_flag  V322_missing_flag  \
0              False              False              False               True
1              False              False              False               True
2              False              False              False               True
3              False              False              False               True
4              False              False              False              False

   V323_missing_flag  V324_missing_flag  V325_missing_flag  V326_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V327_missing_flag  V328_missing_flag  V329_missing_flag  V330_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V331_missing_flag  V332_missing_flag  V333_missing_flag  V334_missing_flag  \
0               True               True               True               True
1               True               True               True               True
```

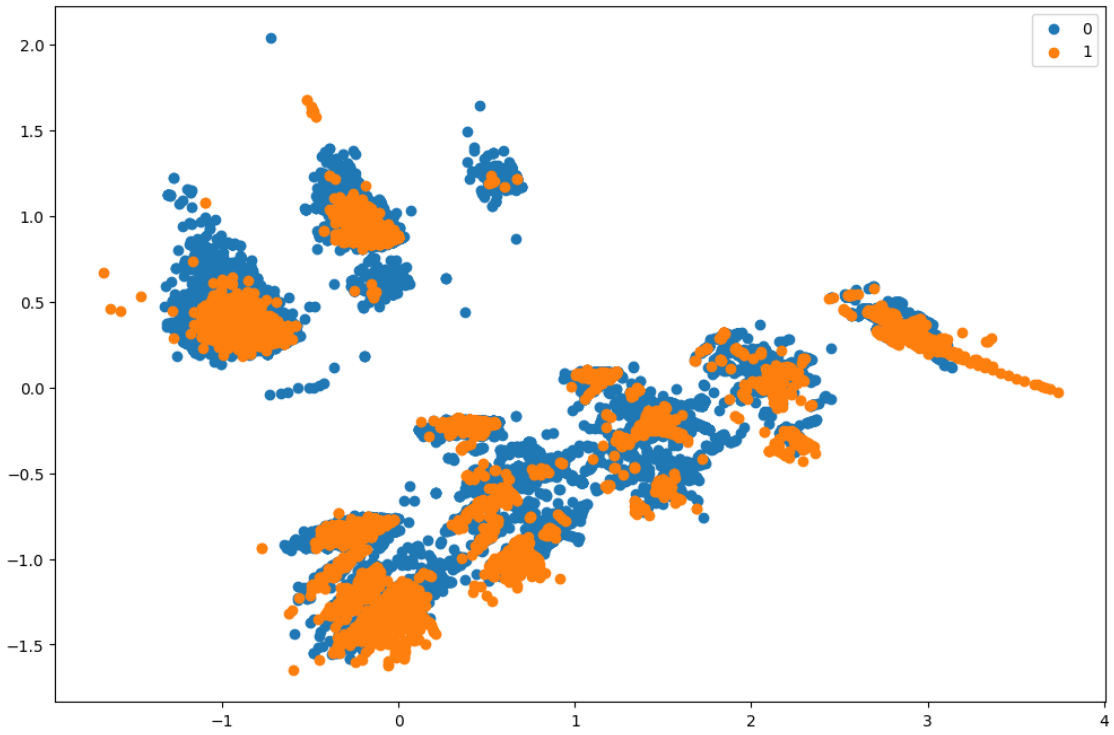
```
2              True              True              True              True
3              True              True              True              True
4             False             False             False             False

   V335_missing_flag  V336_missing_flag  V337_missing_flag  V338_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V339_missing_flag  id_01_missing_flag  id_02_missing_flag  \
0               True                True                True
1               True                True                True
2               True                True                True
3               True                True                True
4              False               False               False

   id_03_missing_flag  id_04_missing_flag  id_05_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True                True                True

   id_06_missing_flag  id_07_missing_flag  id_08_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True                True                True

   id_09_missing_flag  id_10_missing_flag  id_11_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True                True               False

   id_12_missing_flag  id_13_missing_flag  id_14_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4               False                True               False

   id_15_missing_flag  id_16_missing_flag  id_17_missing_flag  \
```

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

   id_18_missing_flag  id_19_missing_flag  id_20_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True               False               False

   id_21_missing_flag  id_22_missing_flag  id_23_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True                True                True

   id_24_missing_flag  id_25_missing_flag  id_26_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True                True                True

   id_27_missing_flag  id_28_missing_flag  id_29_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True               False               False

   id_30_missing_flag  id_31_missing_flag  id_32_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4               False               False               False

   id_33_missing_flag  id_34_missing_flag  id_35_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4               False               False               False
```

```
   id_36_missing_flag  id_37_missing_flag  id_38_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4               False               False               False

   DeviceType_missing_flag  DeviceInfo_missing_flag                Date  \
0                     True                     True 2017-12-02 00:00:00
1                     True                     True 2017-12-02 00:00:01
2                     True                     True 2017-12-02 00:01:09
3                     True                     True 2017-12-02 00:01:39
4                    False                    False 2017-12-02 00:01:46

   _Weekdays  _Hours  _Days  Trans_min_mean  Trans_min_std  \
0          5       0      2           -66.5      -0.278076
1          5       0      2          -106.0      -0.443359
2          5       0      2           -76.0      -0.317871
3          5       0      2           -85.0      -0.355469
4          5       0      2           -85.0      -0.355469

   TransactionAmt_to_mean_card1  TransactionAmt_to_mean_card4  \
0                      0.194580                      0.257812
1                      0.123779                      0.218994
2                      0.607910                      0.443115
3                      0.405029                      0.377686
4                      0.515625                      0.377686

   TransactionAmt_to_std_card1  TransactionAmt_to_std_card4   PCA_V_0  \
0                     0.184560                     0.170288 -0.157349
1                     0.063004                     0.114197 -0.086426
2                     0.589226                     0.258545 -0.800781
3                     0.259460                     0.196899 -0.237305
4                     0.882898                     0.196899  2.904297

    PCA_V_1   PCA_V_2   PCA_V_3   PCA_V_4   PCA_V_5   PCA_V_6   PCA_V_7  \
0  0.918945 -0.843750  0.307861 -0.089417  0.003073 -0.020004 -0.187622
1 -0.800293 -0.152222 -0.363281 -0.101929 -0.002268  0.032410 -0.068970
2  0.316895  0.273193 -0.026337  0.043182 -0.008049 -0.039246 -0.217041
3 -0.811523 -0.123718 -0.423584 -0.067444  0.025040  0.110596 -0.253906
4  0.380127  0.480469 -0.009628 -0.171631  1.170898 -0.178345  0.004597

    PCA_V_8   PCA_V_9  PCA_V_10  PCA_V_11  PCA_V_12  PCA_V_13  PCA_V_14  \
0  0.038269  0.002508 -0.010406  0.034119 -0.044281 -0.089722  0.045013
1  0.040070 -0.180054 -0.059326  0.002270  0.018936 -0.029465  0.016541
2  0.017593  0.033508 -0.000374 -0.015335  0.020706 -0.046051 -0.006680
```

```
3  0.004715  0.170288 -0.012352 -0.014534  0.005116  0.031982 -0.013611
4  0.043732 -0.001608  0.015762 -0.017914 -0.020660 -0.005581 -0.010246

   PCA_V_15  PCA_V_16  PCA_V_17  PCA_V_18  PCA_V_19  PCA_V_20  PCA_V_21  \
0  0.001526 -0.003292  0.018463 -0.018341  0.009972 -0.026993 -0.021591
1 -0.006153 -0.004082  0.010201 -0.001767 -0.022842  0.006413 -0.021011
2  0.004841  0.001108  0.007488 -0.007820 -0.006615 -0.008324  0.009743
3 -0.017181 -0.001876 -0.084717  0.050446  0.140747  0.059052 -0.020187
4 -0.003710 -0.011406 -0.008102  0.013367  0.017807 -0.013626  0.000090

   PCA_V_22  PCA_V_23  PCA_V_24  PCA_V_25  PCA_V_26  PCA_V_27  PCA_V_28  \
0 -0.054565  0.025299  0.018677 -0.006100  0.004215 -0.042999  0.008232
1  0.054779 -0.042572 -0.026566  0.003536  0.001620  0.001561 -0.003544
2 -0.007683  0.003206  0.002855  0.001932  0.003839  0.002985 -0.019638
3  0.066467 -0.010803 -0.017715  0.025543  0.003864  0.003452  0.026138
4 -0.005260 -0.141479  0.203003  0.014557  0.012848  0.002218  0.014000

   PCA_V_29
0 -0.007919
1 -0.003748
2 -0.003443
3 -0.041901
4 -0.001691

[5 rows x 537 columns]
```

[66]:
```
# Plot first 2 PCA features and colour by target variable
plt.figure(figsize=(12, 8));
groups = df.groupby("isFraud")
for name, group in groups:
    plt.scatter(group["PCA_V_0"], group["PCA_V_1"], label=name)
plt.legend()
plt.show()
```



9 9. Feature Encoding

[67]:
```
cat_cols = df.select_dtypes(include=['object','category']).columns.tolist()
```

[68]:
```
binary_cols = [col for col in df.columns if df[col].dropna().nunique() ==2 and␣
↪col not in cat_cols]
```

[69]:
```
num_cols = [col for col in df.
↪select_dtypes(include=['int64','int32','int16','int8','float64','float32','float16']).
↪columns if col not in binary_cols]
```

[70]:
```
len(cat_cols)
```

[70]: 29

[71]:
```
len(binary_cols)
```

[71]: 415

[72]:
```
len(num_cols)
```

[72]: 92

[73]:
```
cat_cols = cat_cols + binary_cols
len(cat_cols)
```

[73]:
```
444
```

[ ]:
```

```

[74]:
```
# Frequecny encoding variables
frequency_encoded_variables = []
for col in cat_cols:
    if df[col].nunique() > 30:
        print(col, df[col].nunique())
        frequency_encoded_variables.append(col)
```

```
id_31 130
id_33 260
DeviceInfo 1786
```

[75]:
```
for variable in tqdm(frequency_encoded_variables):
    # group by frequency
    fq = df.groupby(variable).size()/len(df)
    # mapping values to dataframe
    df.loc[:, "{}".format(variable)] = df[variable].map(fq)
    cat_columns.remove(variable)
```

```
100%|
    | 3/3 [00:00<00:00,  4.14it/s]
```

[76]:
```
df.head()
```

[76]:

	TransactionID	isFraud	TransactionDT	TransactionAmt	ProductCD	card1	\
0	2987000	0	86400	4.226562	W	13926	
1	2987001	0	86401	3.367188	W	2755	
2	2987002	0	86469	4.078125	W	4663	
3	2987003	0	86499	3.912109	W	18132	
4	2987004	0	86506	3.912109	H	4497	

	card2	card3	card4	card5	card6	addr1	addr2	dist1	P_emaildomain	\
0	NaN	150.0	discover	142.0	credit	315.0	87.0	19.0	NoInf	
1	404.0	150.0	mastercard	102.0	credit	325.0	87.0	NaN	Google	
2	490.0	150.0	visa	166.0	debit	330.0	87.0	287.0	Microsoft	
3	567.0	150.0	mastercard	117.0	debit	476.0	87.0	NaN	Yahoo Mail	
4	514.0	150.0	mastercard	102.0	credit	420.0	87.0	NaN	Google	

	R_emaildomain	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	\
0	NoInf	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	2.0	0.0	
1	NoInf	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	

```
2         NoInf  1.0  1.0  0.0  0.0  0.0  1.0  0.0  0.0  1.0  0.0  1.0  0.0
3         NoInf  2.0  5.0  0.0  0.0  0.0  4.0  0.0  0.0  1.0  0.0  1.0  0.0
4         NoInf  1.0  1.0  0.0  0.0  0.0  1.0  0.0  1.0  0.0  1.0  1.0  0.0

    C13  C14     D1     D2    D3    D4   D5  D6  D8  D9   D10    D11  D12  \
0   1.0  1.0   14.0    NaN  13.0   NaN  NaN NaN NaN NaN  13.0   13.0  NaN
1   1.0  1.0    0.0    NaN   NaN   0.0  NaN NaN NaN NaN   0.0    NaN  NaN
2   1.0  1.0    0.0    NaN   NaN   0.0  NaN NaN NaN NaN   0.0  315.0  NaN
3  25.0  1.0  112.0  112.0   0.0  94.0  0.0 NaN NaN NaN  84.0    NaN  NaN
4   1.0  1.0    0.0    NaN   NaN   NaN  NaN NaN NaN NaN   NaN    NaN  NaN

   D13  D14    D15   M1   M2   M3   M4   M5   M6   M7   M8   M9  id_01  \
0  NaN  NaN    0.0    T    T    T   M2    F    T  NaN  NaN  NaN    NaN
1  NaN  NaN    0.0  NaN  NaN  NaN   M0    T    T  NaN  NaN  NaN    NaN
2  NaN  NaN  315.0    T    T    T   M0    F    F    F    F    F    NaN
3  NaN  NaN  111.0  NaN  NaN  NaN   M0    T    F  NaN  NaN  NaN    NaN
4  NaN  NaN    NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN    0.0

     id_02  id_03  id_04  id_05  id_06  id_09  id_10  id_11     id_12  id_13  \
0      NaN    NaN    NaN    NaN    NaN    NaN    NaN    NaN       NaN    NaN
1      NaN    NaN    NaN    NaN    NaN    NaN    NaN    NaN       NaN    NaN
2      NaN    NaN    NaN    NaN    NaN    NaN    NaN    NaN       NaN    NaN
3      NaN    NaN    NaN    NaN    NaN    NaN    NaN    NaN       NaN    NaN
4  70787.0    NaN    NaN    NaN    NaN    NaN    NaN  100.0  NotFound    NaN

   id_14 id_15     id_16  id_17  id_19  id_20 id_28     id_29    id_30  \
0    NaN   NaN       NaN    NaN    NaN    NaN   NaN       NaN      NAN
1    NaN   NaN       NaN    NaN    NaN    NaN   NaN       NaN      NAN
2    NaN   NaN       NaN    NaN    NaN    NaN   NaN       NaN      NAN
3    NaN   NaN       NaN    NaN    NaN    NaN   NaN       NaN      NAN
4 -480.0   New  NotFound  166.0  542.0  144.0   New  NotFound  Android

      id_31  id_32     id_33           id_34 id_35 id_36 id_37 id_38  \
0       NaN    NaN       NaN             NaN   NaN   NaN   NaN   NaN
1       NaN    NaN       NaN             NaN   NaN   NaN   NaN   NaN
2       NaN    NaN       NaN             NaN   NaN   NaN   NaN   NaN
3       NaN    NaN       NaN             NaN   NaN   NaN   NaN   NaN
4  0.001797   32.0  0.000921  match_status:2     T     F     T     T

  DeviceType  DeviceInfo  card2_missing_flag  card3_missing_flag  \
0        NaN         NaN                True               False
1        NaN         NaN               False               False
2        NaN         NaN               False               False
3        NaN         NaN               False               False
4     mobile    0.000015               False               False

   card4_missing_flag  card5_missing_flag  card6_missing_flag  \
```

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

   addr1_missing_flag  addr2_missing_flag  dist1_missing_flag  \
0               False               False               False
1               False               False                True
2               False               False               False
3               False               False                True
4               False               False                True

   dist2_missing_flag  P_emaildomain_missing_flag  R_emaildomain_missing_flag  \
0                True                        True                        True
1                True                       False                        True
2                True                       False                        True
3                True                       False                        True
4                True                       False                        True

   D1_missing_flag  D2_missing_flag  D3_missing_flag  D4_missing_flag  \
0            False             True            False             True
1            False             True             True            False
2            False             True             True            False
3            False            False            False            False
4            False             True             True             True

   D5_missing_flag  D6_missing_flag  D7_missing_flag  D8_missing_flag  \
0             True             True             True             True
1             True             True             True             True
2             True             True             True             True
3            False             True             True             True
4             True             True             True             True

   D9_missing_flag  D10_missing_flag  D11_missing_flag  D12_missing_flag  \
0             True              False              False              True
1             True              False               True              True
2             True              False              False              True
3             True              False               True              True
4             True               True               True              True

   D13_missing_flag  D14_missing_flag  D15_missing_flag  M1_missing_flag  \
0              True              True             False            False
1              True              True             False             True
2              True              True             False            False
3              True              True             False             True
4              True              True              True             True
```

```
   M2_missing_flag  M3_missing_flag  M4_missing_flag  M5_missing_flag  \
0            False            False            False            False
1             True             True            False            False
2            False            False            False            False
3             True             True            False            False
4             True             True             True             True

   M6_missing_flag  M7_missing_flag  M8_missing_flag  M9_missing_flag  \
0            False             True             True             True
1            False             True             True             True
2            False            False            False            False
3            False             True             True             True
4             True             True             True             True

   V1_missing_flag  V2_missing_flag  V3_missing_flag  V4_missing_flag  \
0            False            False            False            False
1             True             True             True             True
2            False            False            False            False
3             True             True             True             True
4             True             True             True             True

   V5_missing_flag  V6_missing_flag  V7_missing_flag  V8_missing_flag  \
0            False            False            False            False
1             True             True             True             True
2            False            False            False            False
3             True             True             True             True
4             True             True             True             True

   V9_missing_flag  V10_missing_flag  V11_missing_flag  V12_missing_flag  \
0            False             False             False             False
1             True              True              True             False
2            False             False             False             False
3             True              True              True             False
4             True              True              True              True

   V13_missing_flag  V14_missing_flag  V15_missing_flag  V16_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V17_missing_flag  V18_missing_flag  V19_missing_flag  V20_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
```

```
3             False             False             False             False
4              True              True              True              True

   V21_missing_flag  V22_missing_flag  V23_missing_flag  V24_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V25_missing_flag  V26_missing_flag  V27_missing_flag  V28_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V29_missing_flag  V30_missing_flag  V31_missing_flag  V32_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V33_missing_flag  V34_missing_flag  V35_missing_flag  V36_missing_flag  \
0             False             False              True              True
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V37_missing_flag  V38_missing_flag  V39_missing_flag  V40_missing_flag  \
0              True              True              True              True
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V41_missing_flag  V42_missing_flag  V43_missing_flag  V44_missing_flag  \
0              True              True              True              True
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V45_missing_flag  V46_missing_flag  V47_missing_flag  V48_missing_flag  \
0              True              True              True              True
```

```
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V49_missing_flag  V50_missing_flag  V51_missing_flag  V52_missing_flag  \
0              True              True              True              True
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V53_missing_flag  V54_missing_flag  V55_missing_flag  V56_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V57_missing_flag  V58_missing_flag  V59_missing_flag  V60_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V61_missing_flag  V62_missing_flag  V63_missing_flag  V64_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V65_missing_flag  V66_missing_flag  V67_missing_flag  V68_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V69_missing_flag  V70_missing_flag  V71_missing_flag  V72_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True
```

```
   V73_missing_flag  V74_missing_flag  V75_missing_flag  V76_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V77_missing_flag  V78_missing_flag  V79_missing_flag  V80_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V81_missing_flag  V82_missing_flag  V83_missing_flag  V84_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V85_missing_flag  V86_missing_flag  V87_missing_flag  V88_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V89_missing_flag  V90_missing_flag  V91_missing_flag  V92_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True              True              True

   V93_missing_flag  V94_missing_flag  V95_missing_flag  V96_missing_flag  \
0             False             False             False             False
1             False             False             False             False
2             False             False             False             False
3             False             False             False             False
4              True              True             False             False

   V97_missing_flag  V98_missing_flag  V99_missing_flag  V100_missing_flag  \
0             False             False             False              False
1             False             False             False              False
2             False             False             False              False
3             False             False             False              False
```

```
4            False            False            False            False

   V101_missing_flag  V102_missing_flag  V103_missing_flag  V104_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V105_missing_flag  V106_missing_flag  V107_missing_flag  V108_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V109_missing_flag  V110_missing_flag  V111_missing_flag  V112_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V113_missing_flag  V114_missing_flag  V115_missing_flag  V116_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V117_missing_flag  V118_missing_flag  V119_missing_flag  V120_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V121_missing_flag  V122_missing_flag  V123_missing_flag  V124_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V125_missing_flag  V126_missing_flag  V127_missing_flag  V128_missing_flag  \
0              False              False              False              False
1              False              False              False              False
```

```
2             False             False             False             False
3             False             False             False             False
4             False             False             False             False

   V129_missing_flag  V130_missing_flag  V131_missing_flag  V132_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   …  V170_missing_flag  V171_missing_flag  V172_missing_flag  \
0  …               True               True               True
1  …               True               True               True
2  …               True               True               True
3  …               True               True               True
4  …              False              False              False

   V173_missing_flag  V174_missing_flag  V175_missing_flag  V176_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V177_missing_flag  V178_missing_flag  V179_missing_flag  V180_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V181_missing_flag  V182_missing_flag  V183_missing_flag  V184_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V185_missing_flag  V186_missing_flag  V187_missing_flag  V188_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V189_missing_flag  V190_missing_flag  V191_missing_flag  V192_missing_flag  \
```

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

   V193_missing_flag  V194_missing_flag  V195_missing_flag  V196_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V197_missing_flag  V198_missing_flag  V199_missing_flag  V200_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V201_missing_flag  V202_missing_flag  V203_missing_flag  V204_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V205_missing_flag  V206_missing_flag  V207_missing_flag  V208_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V209_missing_flag  V210_missing_flag  V211_missing_flag  V212_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V213_missing_flag  V214_missing_flag  V215_missing_flag  V216_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False
```

```
   V217_missing_flag  V218_missing_flag  V219_missing_flag  V220_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V221_missing_flag  V222_missing_flag  V223_missing_flag  V224_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V225_missing_flag  V226_missing_flag  V227_missing_flag  V228_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V229_missing_flag  V230_missing_flag  V231_missing_flag  V232_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V233_missing_flag  V234_missing_flag  V235_missing_flag  V236_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V237_missing_flag  V238_missing_flag  V239_missing_flag  V240_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V241_missing_flag  V242_missing_flag  V243_missing_flag  V244_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
```

```
3              True              True              True              True
4             False             False             False             False

   V245_missing_flag  V246_missing_flag  V247_missing_flag  V248_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V249_missing_flag  V250_missing_flag  V251_missing_flag  V252_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V253_missing_flag  V254_missing_flag  V255_missing_flag  V256_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V257_missing_flag  V258_missing_flag  V259_missing_flag  V260_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V261_missing_flag  V262_missing_flag  V263_missing_flag  V264_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V265_missing_flag  V266_missing_flag  V267_missing_flag  V268_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V269_missing_flag  V270_missing_flag  V271_missing_flag  V272_missing_flag  \
0               True               True               True               True
```

```
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V273_missing_flag  V274_missing_flag  V275_missing_flag  V276_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V277_missing_flag  V278_missing_flag  V279_missing_flag  V280_missing_flag  \
0               True               True              False              False
1               True               True              False              False
2               True               True              False              False
3               True               True              False              False
4              False              False              False              False

   V281_missing_flag  V282_missing_flag  V283_missing_flag  V284_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V285_missing_flag  V286_missing_flag  V287_missing_flag  V288_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V289_missing_flag  V290_missing_flag  V291_missing_flag  V292_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V293_missing_flag  V294_missing_flag  V295_missing_flag  V296_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False
```

```
   V297_missing_flag  V298_missing_flag  V299_missing_flag  V300_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V301_missing_flag  V302_missing_flag  V303_missing_flag  V304_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V305_missing_flag  V306_missing_flag  V307_missing_flag  V308_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V309_missing_flag  V310_missing_flag  V311_missing_flag  V312_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V313_missing_flag  V314_missing_flag  V315_missing_flag  V316_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V317_missing_flag  V318_missing_flag  V319_missing_flag  V320_missing_flag  \
0              False              False              False              False
1              False              False              False              False
2              False              False              False              False
3              False              False              False              False
4              False              False              False              False

   V321_missing_flag  V322_missing_flag  V323_missing_flag  V324_missing_flag  \
0              False               True               True               True
1              False               True               True               True
2              False               True               True               True
3              False               True               True               True
```

```
4             False             False             False             False

   V325_missing_flag  V326_missing_flag  V327_missing_flag  V328_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V329_missing_flag  V330_missing_flag  V331_missing_flag  V332_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V333_missing_flag  V334_missing_flag  V335_missing_flag  V336_missing_flag  \
0               True               True               True               True
1               True               True               True               True
2               True               True               True               True
3               True               True               True               True
4              False              False              False              False

   V337_missing_flag  V338_missing_flag  V339_missing_flag  \
0               True               True               True
1               True               True               True
2               True               True               True
3               True               True               True
4              False              False              False

   id_01_missing_flag  id_02_missing_flag  id_03_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4               False               False                True

   id_04_missing_flag  id_05_missing_flag  id_06_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True                True                True

   id_07_missing_flag  id_08_missing_flag  id_09_missing_flag  \
0                True                True                True
1                True                True                True
```

```
2                True                True                True
3                True                True                True
4                True                True                True

   id_10_missing_flag  id_11_missing_flag  id_12_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True               False               False

   id_13_missing_flag  id_14_missing_flag  id_15_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True               False               False

   id_16_missing_flag  id_17_missing_flag  id_18_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4               False               False                True

   id_19_missing_flag  id_20_missing_flag  id_21_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4               False               False                True

   id_22_missing_flag  id_23_missing_flag  id_24_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True                True                True

   id_25_missing_flag  id_26_missing_flag  id_27_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4                True                True                True

   id_28_missing_flag  id_29_missing_flag  id_30_missing_flag  \
```

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

   id_31_missing_flag  id_32_missing_flag  id_33_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4               False               False               False

   id_34_missing_flag  id_35_missing_flag  id_36_missing_flag  \
0                True                True                True
1                True                True                True
2                True                True                True
3                True                True                True
4               False               False               False

   id_37_missing_flag  id_38_missing_flag  DeviceType_missing_flag  \
0                True                True                     True
1                True                True                     True
2                True                True                     True
3                True                True                     True
4               False               False                    False

   DeviceInfo_missing_flag                Date  _Weekdays  _Hours  _Days  \
0                     True 2017-12-02 00:00:00          5       0      2
1                     True 2017-12-02 00:00:01          5       0      2
2                     True 2017-12-02 00:01:09          5       0      2
3                     True 2017-12-02 00:01:39          5       0      2
4                    False 2017-12-02 00:01:46          5       0      2

   Trans_min_mean  Trans_min_std  TransactionAmt_to_mean_card1  \
0           -66.5      -0.278076                      0.194580
1          -106.0      -0.443359                      0.123779
2           -76.0      -0.317871                      0.607910
3           -85.0      -0.355469                      0.405029
4           -85.0      -0.355469                      0.515625

   TransactionAmt_to_mean_card4  TransactionAmt_to_std_card1  \
0                      0.257812                     0.184560
1                      0.218994                     0.063004
2                      0.443115                     0.589226
3                      0.377686                     0.259460
4                      0.377686                     0.882898
```

```
   TransactionAmt_to_std_card4   PCA_V_0   PCA_V_1   PCA_V_2   PCA_V_3  \
0                     0.170288 -0.157349  0.918945 -0.843750  0.307861
1                     0.114197 -0.086426 -0.800293 -0.152222 -0.363281
2                     0.258545 -0.800781  0.316895  0.273193 -0.026337
3                     0.196899 -0.237305 -0.811523 -0.123718 -0.423584
4                     0.196899  2.904297  0.380127  0.480469 -0.009628

    PCA_V_4   PCA_V_5   PCA_V_6   PCA_V_7   PCA_V_8   PCA_V_9  PCA_V_10  \
0 -0.089417  0.003073 -0.020004 -0.187622  0.038269  0.002508 -0.010406
1 -0.101929 -0.002268  0.032410 -0.068970  0.040070 -0.180054 -0.059326
2  0.043182 -0.008049 -0.039246 -0.217041  0.017593  0.033508 -0.000374
3 -0.067444  0.025040  0.110596 -0.253906  0.004715  0.170288 -0.012352
4 -0.171631  1.170898 -0.178345  0.004597  0.043732 -0.001608  0.015762

   PCA_V_11  PCA_V_12  PCA_V_13  PCA_V_14  PCA_V_15  PCA_V_16  PCA_V_17  \
0  0.034119 -0.044281 -0.089722  0.045013  0.001526 -0.003292  0.018463
1  0.002270  0.018936 -0.029465  0.016541 -0.006153 -0.004082  0.010201
2 -0.015335  0.020706 -0.046051 -0.006680  0.004841  0.001108  0.007488
3 -0.014534  0.005116  0.031982 -0.013611 -0.017181 -0.001876 -0.084717
4 -0.017914 -0.020660 -0.005581 -0.010246 -0.003710 -0.011406 -0.008102

   PCA_V_18  PCA_V_19  PCA_V_20  PCA_V_21  PCA_V_22  PCA_V_23  PCA_V_24  \
0 -0.018341  0.009972 -0.026993 -0.021591 -0.054565  0.025299  0.018677
1 -0.001767 -0.022842  0.006413 -0.021011  0.054779 -0.042572 -0.026566
2 -0.007820 -0.006615 -0.008324  0.009743 -0.007683  0.003206  0.002855
3  0.050446  0.140747  0.059052 -0.020187  0.066467 -0.010803 -0.017715
4  0.013367  0.017807 -0.013626  0.000090 -0.005260 -0.141479  0.203003

   PCA_V_25  PCA_V_26  PCA_V_27  PCA_V_28  PCA_V_29
0 -0.006100  0.004215 -0.042999  0.008232 -0.007919
1  0.003536  0.001620  0.001561 -0.003544 -0.003748
2  0.001932  0.003839  0.002985 -0.019638 -0.003443
3  0.025543  0.003864  0.003452  0.026138 -0.041901
4  0.014557  0.012848  0.002218  0.014000 -0.001691

[5 rows x 537 columns]
```

[77]:
```
# Label encode the variables
for col in cat_cols:
    lbl        = LabelEncoder()
    lbl.fit(list(df[col].values))
    df[col] = lbl.transform(list(df[col].values))
```

[78]:
```
df = reduce_mem_usage(df)
```

```
Mem. usage decreased to 384.45 Mb (81.5% reduction)
```

10 10. Data Preprocessing for Model Building

[79]:
```
df.loc[:, 'isFraud'].value_counts()
```

[79]:
```
0    569877
1     20663
Name: isFraud, dtype: int64
```

[80]:
```
df = df.drop(['TransactionID','TransactionDT','Date'], axis=1)
```

[81]:
```
# Split the y variable series and x variables dataset
X = df.drop(['isFraud'],axis=1)
y = df.isFraud.astype(bool)

# Delete train df
del df

# Collect garbage
gc.collect()
```

[81]:
```
0
```

[82]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.3,␣
↪random_state = 0,stratify=y)
```

[83]:
```
X_train.head()
```

[83]:

	TransactionAmt	ProductCD	card1	card2	card3	card4	card5	card6	\
448539	4.679688	4	6598	111.0	150.0	2	195.0	2	
321311	4.355469	4	12839	321.0	150.0	4	226.0	2	
497320	3.892578	4	14649	548.0	150.0	4	226.0	2	
350951	4.058594	4	6489	295.0	150.0	4	226.0	2	
98132	5.296875	2	5714	170.0	150.0	4	195.0	1	

	addr1	addr2	dist1	P_emaildomain	R_emaildomain	C1	C2	C3	\
448539	264.0	87.0	6.0	0	2	2.0	4.0	0.0	
321311	264.0	87.0	0.0	0	2	1.0	1.0	0.0	
497320	441.0	87.0	86.0	4	2	2.0	4.0	0.0	
350951	184.0	87.0	NaN	0	2	154.0	148.0	0.0	
98132	498.0	87.0	NaN	5	6	1.0	1.0	0.0	

	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	\
448539	0.0	1.0	1.0	0.0	0.0	1.0	0.0	2.0	0.0	7.0	2.0	
321311	0.0	2.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0	12.0	1.0	
497320	0.0	0.0	2.0	0.0	0.0	3.0	0.0	4.0	0.0	45.0	2.0	
350951	0.0	108.0	101.0	0.0	0.0	92.0	0.0	107.0	0.0	504.0	129.0	
98132	1.0	0.0	1.0	0.0	1.0	0.0	1.0	1.0	0.0	1.0	1.0	

```
          D1     D2    D3     D4    D5  D6        D8        D9    D10   D11  \
448539   64.0   64.0  18.0   63.0  10.0 NaN       NaN       NaN   63.0  35.0
321311  114.0  114.0  14.0  115.0  14.0 NaN       NaN       NaN  115.0  45.0
497320   85.0   85.0   3.0  329.0   3.0 NaN       NaN       NaN  329.0  85.0
350951  577.0  577.0  54.0   54.0  54.0 NaN       NaN       NaN  577.0   NaN
98132     0.0    NaN   NaN    NaN   NaN NaN  7.917969  0.916504    NaN   NaN

        D12  D13  D14    D15  M1  M2  M3  M4  M5  M6  M7  M8  M9  id_01  \
448539  NaN  NaN  NaN   63.0   1   1   1   3   2   0   0   1   1    NaN
321311  NaN  NaN  NaN  115.0   1   1   1   3   2   1   1   1   1    NaN
497320  NaN  NaN  NaN   85.0   1   1   1   0   0   0   0   0   0    NaN
350951  NaN  NaN  NaN  110.0   2   2   2   3   2   0   2   2   2    NaN
98132   NaN  NaN  NaN    NaN   2   2   2   3   2   2   2   2   2   -5.0

           id_02  id_03  id_04  id_05  id_06  id_09  id_10  id_11  id_12  \
448539       NaN    NaN    NaN    NaN    NaN    NaN    NaN    NaN      1
321311       NaN    NaN    NaN    NaN    NaN    NaN    NaN    NaN      1
497320       NaN    NaN    NaN    NaN    NaN    NaN    NaN    NaN      1
350951       NaN    NaN    NaN    NaN    NaN    NaN    NaN    NaN      1
98132   105681.0    0.0    0.0    2.0    0.0    0.0    0.0  100.0      2

        id_13  id_14  id_15  id_16  id_17  id_19  id_20  id_28  id_29  id_30  \
448539    NaN    NaN      1      1    NaN    NaN    NaN      1      1      3
321311    NaN    NaN      1      1    NaN    NaN    NaN      1      1      3
497320    NaN    NaN      1      1    NaN    NaN    NaN      1      1      3
350951    NaN    NaN      1      1    NaN    NaN    NaN      1      1      3
98132    52.0 -360.0      0      0  166.0  621.0  272.0      0      0      4

        id_31  id_32  id_33  id_34  id_35  id_36  id_37  id_38  DeviceType  \
448539    103    NaN     74      4      2      2      2      2           2
321311    103    NaN     74      4      2      2      2      2           2
497320    103    NaN     74      4      2      2      2      2           2
350951    103    NaN     74      4      2      2      2      2           2
98132      84   24.0     72      3      1      0      1      1           0

        DeviceInfo  card2_missing_flag  card3_missing_flag  \
448539         141                   0                   0
321311         141                   0                   0
497320         141                   0                   0
350951         141                   0                   0
98132          140                   0                   0

        card4_missing_flag  card5_missing_flag  card6_missing_flag  \
448539                   0                   0                   0
321311                   0                   0                   0
497320                   0                   0                   0
```

```
350951                   0                   0                   0
98132                    0                   0                   0

        addr1_missing_flag  addr2_missing_flag  dist1_missing_flag  \
448539                   0                   0                   0
321311                   0                   0                   0
497320                   0                   0                   0
350951                   0                   0                   1
98132                    0                   0                   1

        dist2_missing_flag  P_emaildomain_missing_flag  \
448539                   1                           0
321311                   1                           0
497320                   1                           0
350951                   1                           0
98132                    1                           0

        R_emaildomain_missing_flag  D1_missing_flag  D2_missing_flag  \
448539                           1                0                0
321311                           1                0                0
497320                           1                0                0
350951                           1                0                0
98132                            0                0                1

        D3_missing_flag  D4_missing_flag  D5_missing_flag  D6_missing_flag  \
448539                0                0                0                1
321311                0                0                0                1
497320                0                0                0                1
350951                0                0                0                1
98132                 1                1                1                1

        D7_missing_flag  D8_missing_flag  D9_missing_flag  D10_missing_flag  \
448539                1                1                1                 0
321311                1                1                1                 0
497320                1                1                1                 0
350951                1                1                1                 0
98132                 1                0                0                 1

        D11_missing_flag  D12_missing_flag  D13_missing_flag  \
448539                 0                 1                 1
321311                 0                 1                 1
497320                 0                 1                 1
350951                 1                 1                 1
98132                  1                 1                 1

        D14_missing_flag  D15_missing_flag  M1_missing_flag  M2_missing_flag  \
448539                 1                 0                0                0
```

```
321311                1                 0                0                0
497320                1                 0                0                0
350951                1                 0                1                1
98132                 1                 1                1                1

        M3_missing_flag  M4_missing_flag  M5_missing_flag  M6_missing_flag  \
448539                0                1                1                0
321311                0                1                1                0
497320                0                0                0                0
350951                1                1                1                0
98132                 1                1                1                1

        M7_missing_flag  M8_missing_flag  M9_missing_flag  V1_missing_flag  \
448539                0                0                0                0
321311                0                0                0                0
497320                0                0                0                0
350951                1                1                1                1
98132                 1                1                1                1

        V2_missing_flag  V3_missing_flag  V4_missing_flag  V5_missing_flag  \
448539                0                0                0                0
321311                0                0                0                0
497320                0                0                0                0
350951                1                1                1                1
98132                 1                1                1                1

        V6_missing_flag  V7_missing_flag  V8_missing_flag  V9_missing_flag  \
448539                0                0                0                0
321311                0                0                0                0
497320                0                0                0                0
350951                1                1                1                1
98132                 1                1                1                1

        V10_missing_flag  V11_missing_flag  V12_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 1                 1                 0
98132                  1                 1                 1

        V13_missing_flag  V14_missing_flag  V15_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1
```

```
        V16_missing_flag  V17_missing_flag  V18_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V19_missing_flag  V20_missing_flag  V21_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V22_missing_flag  V23_missing_flag  V24_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V25_missing_flag  V26_missing_flag  V27_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V28_missing_flag  V29_missing_flag  V30_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V31_missing_flag  V32_missing_flag  V33_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V34_missing_flag  V35_missing_flag  V36_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
```

```
98132                 1                 1                 1

        V37_missing_flag  V38_missing_flag  V39_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V40_missing_flag  V41_missing_flag  V42_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V43_missing_flag  V44_missing_flag  V45_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V46_missing_flag  V47_missing_flag  V48_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V49_missing_flag  V50_missing_flag  V51_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V52_missing_flag  V53_missing_flag  V54_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V55_missing_flag  V56_missing_flag  V57_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
```

```
497320                0                 0                 0
350951                0                 0                 0
98132                 1                 1                 1

        V58_missing_flag  V59_missing_flag  V60_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V61_missing_flag  V62_missing_flag  V63_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V64_missing_flag  V65_missing_flag  V66_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V67_missing_flag  V68_missing_flag  V69_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V70_missing_flag  V71_missing_flag  V72_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V73_missing_flag  V74_missing_flag  V75_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V76_missing_flag  V77_missing_flag  V78_missing_flag  \
```

```
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V79_missing_flag  V80_missing_flag  V81_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V82_missing_flag  V83_missing_flag  V84_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V85_missing_flag  V86_missing_flag  V87_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V88_missing_flag  V89_missing_flag  V90_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V91_missing_flag  V92_missing_flag  V93_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V94_missing_flag  V95_missing_flag  V96_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 0                 0
```

```
        V97_missing_flag  V98_missing_flag  V99_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  0                 0                 0

        V100_missing_flag  V101_missing_flag  V102_missing_flag  \
448539                  0                  0                  0
321311                  0                  0                  0
497320                  0                  0                  0
350951                  0                  0                  0
98132                   0                  0                  0

        V103_missing_flag  V104_missing_flag  V105_missing_flag  \
448539                  0                  0                  0
321311                  0                  0                  0
497320                  0                  0                  0
350951                  0                  0                  0
98132                   0                  0                  0

        V106_missing_flag  V107_missing_flag  V108_missing_flag  \
448539                  0                  0                  0
321311                  0                  0                  0
497320                  0                  0                  0
350951                  0                  0                  0
98132                   0                  0                  0

        V109_missing_flag  V110_missing_flag  V111_missing_flag  \
448539                  0                  0                  0
321311                  0                  0                  0
497320                  0                  0                  0
350951                  0                  0                  0
98132                   0                  0                  0

        V112_missing_flag  V113_missing_flag  V114_missing_flag  \
448539                  0                  0                  0
321311                  0                  0                  0
497320                  0                  0                  0
350951                  0                  0                  0
98132                   0                  0                  0

        V115_missing_flag  V116_missing_flag  V117_missing_flag  \
448539                  0                  0                  0
321311                  0                  0                  0
497320                  0                  0                  0
```

```
350951                 0                  0                  0
98132                  0                  0                  0

        V118_missing_flag  V119_missing_flag  V120_missing_flag  \
448539                  0                  0                  0
321311                  0                  0                  0
497320                  0                  0                  0
350951                  0                  0                  0
98132                   0                  0                  0

        V121_missing_flag  V122_missing_flag  V123_missing_flag  \
448539                  0                  0                  0
321311                  0                  0                  0
497320                  0                  0                  0
350951                  0                  0                  0
98132                   0                  0                  0

        V124_missing_flag  V125_missing_flag  V126_missing_flag  \
448539                  0                  0                  0
321311                  0                  0                  0
497320                  0                  0                  0
350951                  0                  0                  0
98132                   0                  0                  0

        V127_missing_flag  V128_missing_flag  V129_missing_flag  \
448539                  0                  0                  0
321311                  0                  0                  0
497320                  0                  0                  0
350951                  0                  0                  0
98132                   0                  0                  0

        V130_missing_flag  V131_missing_flag  V132_missing_flag  \
448539                  0                  0                  0
321311                  0                  0                  0
497320                  0                  0                  0
350951                  0                  0                  0
98132                   0                  0                  0

        V133_missing_flag  V134_missing_flag  V135_missing_flag  …  \
448539                  0                  0                  0  …
321311                  0                  0                  0  …
497320                  0                  0                  0  …
350951                  0                  0                  0  …
98132                   0                  0                  0  …

        V169_missing_flag  V170_missing_flag  V171_missing_flag  \
448539                  1                  1                  1
```

```
321311                 1                 1                 1
497320                 1                 1                 1
350951                 1                 1                 1
98132                  0                 0                 0

        V172_missing_flag  V173_missing_flag  V174_missing_flag  \
448539                  1                 1                  1
321311                  1                 1                  1
497320                  1                 1                  1
350951                  1                 1                  1
98132                   0                 0                  0

        V175_missing_flag  V176_missing_flag  V177_missing_flag  \
448539                  1                 1                  1
321311                  1                 1                  1
497320                  1                 1                  1
350951                  1                 1                  1
98132                   0                 0                  0

        V178_missing_flag  V179_missing_flag  V180_missing_flag  \
448539                  1                 1                  1
321311                  1                 1                  1
497320                  1                 1                  1
350951                  1                 1                  1
98132                   0                 0                  0

        V181_missing_flag  V182_missing_flag  V183_missing_flag  \
448539                  1                 1                  1
321311                  1                 1                  1
497320                  1                 1                  1
350951                  1                 1                  1
98132                   0                 0                  0

        V184_missing_flag  V185_missing_flag  V186_missing_flag  \
448539                  1                 1                  1
321311                  1                 1                  1
497320                  1                 1                  1
350951                  1                 1                  1
98132                   0                 0                  0

        V187_missing_flag  V188_missing_flag  V189_missing_flag  \
448539                  1                 1                  1
321311                  1                 1                  1
497320                  1                 1                  1
350951                  1                 1                  1
98132                   0                 0                  0
```

	V190_missing_flag	V191_missing_flag	V192_missing_flag	\
448539	1	1	1	
321311	1	1	1	
497320	1	1	1	
350951	1	1	1	
98132	0	0	0	

	V193_missing_flag	V194_missing_flag	V195_missing_flag	\
448539	1	1	1	
321311	1	1	1	
497320	1	1	1	
350951	1	1	1	
98132	0	0	0	

	V196_missing_flag	V197_missing_flag	V198_missing_flag	\
448539	1	1	1	
321311	1	1	1	
497320	1	1	1	
350951	1	1	1	
98132	0	0	0	

	V199_missing_flag	V200_missing_flag	V201_missing_flag	\
448539	1	1	1	
321311	1	1	1	
497320	1	1	1	
350951	1	1	1	
98132	0	0	0	

	V202_missing_flag	V203_missing_flag	V204_missing_flag	\
448539	1	1	1	
321311	1	1	1	
497320	1	1	1	
350951	1	1	1	
98132	0	0	0	

	V205_missing_flag	V206_missing_flag	V207_missing_flag	\
448539	1	1	1	
321311	1	1	1	
497320	1	1	1	
350951	1	1	1	
98132	0	0	0	

	V208_missing_flag	V209_missing_flag	V210_missing_flag	\
448539	1	1	1	
321311	1	1	1	
497320	1	1	1	
350951	1	1	1	

```
98132                  0                  0                  0

        V211_missing_flag  V212_missing_flag  V213_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V214_missing_flag  V215_missing_flag  V216_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V217_missing_flag  V218_missing_flag  V219_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V220_missing_flag  V221_missing_flag  V222_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V223_missing_flag  V224_missing_flag  V225_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V226_missing_flag  V227_missing_flag  V228_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V229_missing_flag  V230_missing_flag  V231_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
```

```
497320                 1                  1                  1
350951                 1                  1                  1
98132                  0                  0                  0

        V232_missing_flag  V233_missing_flag  V234_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V235_missing_flag  V236_missing_flag  V237_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V238_missing_flag  V239_missing_flag  V240_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V241_missing_flag  V242_missing_flag  V243_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V244_missing_flag  V245_missing_flag  V246_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V247_missing_flag  V248_missing_flag  V249_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V250_missing_flag  V251_missing_flag  V252_missing_flag  \
```

```
448539                 1                  1                  1
321311                 1                  1                  1
497320                 1                  1                  1
350951                 1                  1                  1
98132                  0                  0                  0

        V253_missing_flag  V254_missing_flag  V255_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V256_missing_flag  V257_missing_flag  V258_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V259_missing_flag  V260_missing_flag  V261_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V262_missing_flag  V263_missing_flag  V264_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V265_missing_flag  V266_missing_flag  V267_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V268_missing_flag  V269_missing_flag  V270_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0
```

```
        V271_missing_flag  V272_missing_flag  V273_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V274_missing_flag  V275_missing_flag  V276_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V277_missing_flag  V278_missing_flag  V279_missing_flag  \
448539                  1                  1                  0
321311                  1                  1                  0
497320                  1                  1                  0
350951                  1                  1                  0
98132                   0                  0                  0

        V280_missing_flag  V281_missing_flag  V282_missing_flag  \
448539                  0                  0                  0
321311                  0                  0                  0
497320                  0                  0                  0
350951                  0                  0                  0
98132                   0                  0                  0

        V283_missing_flag  V284_missing_flag  V285_missing_flag  \
448539                  0                  0                  0
321311                  0                  0                  0
497320                  0                  0                  0
350951                  0                  0                  0
98132                   0                  0                  0

        V286_missing_flag  V287_missing_flag  V288_missing_flag  \
448539                  0                  0                  0
321311                  0                  0                  0
497320                  0                  0                  0
350951                  0                  0                  0
98132                   0                  0                  0

        V289_missing_flag  V290_missing_flag  V291_missing_flag  \
448539                  0                  0                  0
321311                  0                  0                  0
497320                  0                  0                  0
```

```
350951                 0                 0                 0
98132                  0                 0                 0

        V292_missing_flag  V293_missing_flag  V294_missing_flag  \
448539                  0                  0                  0
321311                  0                  0                  0
497320                  0                  0                  0
350951                  0                  0                  0
98132                   0                  0                  0

        V295_missing_flag  V296_missing_flag  V297_missing_flag  \
448539                  0                  0                  0
321311                  0                  0                  0
497320                  0                  0                  0
350951                  0                  0                  0
98132                   0                  0                  0

        V298_missing_flag  V299_missing_flag  V300_missing_flag  \
448539                  0                  0                  0
321311                  0                  0                  0
497320                  0                  0                  0
350951                  0                  0                  0
98132                   0                  0                  0

        V301_missing_flag  V302_missing_flag  V303_missing_flag  \
448539                  0                  0                  0
321311                  0                  0                  0
497320                  0                  0                  0
350951                  0                  0                  0
98132                   0                  0                  0

        V304_missing_flag  V305_missing_flag  V306_missing_flag  \
448539                  0                  0                  0
321311                  0                  0                  0
497320                  0                  0                  0
350951                  0                  0                  0
98132                   0                  0                  0

        V307_missing_flag  V308_missing_flag  V309_missing_flag  \
448539                  0                  0                  0
321311                  0                  0                  0
497320                  0                  0                  0
350951                  0                  0                  0
98132                   0                  0                  0

        V310_missing_flag  V311_missing_flag  V312_missing_flag  \
448539                  0                  0                  0
```

```
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  0                 0                 0

        V313_missing_flag  V314_missing_flag  V315_missing_flag  \
448539                  0                 0                  0
321311                  0                 0                  0
497320                  0                 0                  0
350951                  0                 0                  0
98132                   0                 0                  0

        V316_missing_flag  V317_missing_flag  V318_missing_flag  \
448539                  0                 0                  0
321311                  0                 0                  0
497320                  0                 0                  0
350951                  0                 0                  0
98132                   0                 0                  0

        V319_missing_flag  V320_missing_flag  V321_missing_flag  \
448539                  0                 0                  0
321311                  0                 0                  0
497320                  0                 0                  0
350951                  0                 0                  0
98132                   0                 0                  0

        V322_missing_flag  V323_missing_flag  V324_missing_flag  \
448539                  1                 1                  1
321311                  1                 1                  1
497320                  1                 1                  1
350951                  1                 1                  1
98132                   0                 0                  0

        V325_missing_flag  V326_missing_flag  V327_missing_flag  \
448539                  1                 1                  1
321311                  1                 1                  1
497320                  1                 1                  1
350951                  1                 1                  1
98132                   0                 0                  0

        V328_missing_flag  V329_missing_flag  V330_missing_flag  \
448539                  1                 1                  1
321311                  1                 1                  1
497320                  1                 1                  1
350951                  1                 1                  1
98132                   0                 0                  0
```

```
        V331_missing_flag  V332_missing_flag  V333_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V334_missing_flag  V335_missing_flag  V336_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V337_missing_flag  V338_missing_flag  V339_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        id_01_missing_flag  id_02_missing_flag  id_03_missing_flag  \
448539                   1                   1                   1
321311                   1                   1                   1
497320                   1                   1                   1
350951                   1                   1                   1
98132                    0                   0                   0

        id_04_missing_flag  id_05_missing_flag  id_06_missing_flag  \
448539                   1                   1                   1
321311                   1                   1                   1
497320                   1                   1                   1
350951                   1                   1                   1
98132                    0                   0                   0

        id_07_missing_flag  id_08_missing_flag  id_09_missing_flag  \
448539                   1                   1                   1
321311                   1                   1                   1
497320                   1                   1                   1
350951                   1                   1                   1
98132                    1                   1                   0

        id_10_missing_flag  id_11_missing_flag  id_12_missing_flag  \
448539                   1                   1                   1
321311                   1                   1                   1
497320                   1                   1                   1
350951                   1                   1                   1
```

```
98132                    0                    0                    0

        id_13_missing_flag  id_14_missing_flag  id_15_missing_flag  \
448539                   1                   1                   1
321311                   1                   1                   1
497320                   1                   1                   1
350951                   1                   1                   1
98132                    0                   0                   0

        id_16_missing_flag  id_17_missing_flag  id_18_missing_flag  \
448539                   1                   1                   1
321311                   1                   1                   1
497320                   1                   1                   1
350951                   1                   1                   1
98132                    0                   0                   1

        id_19_missing_flag  id_20_missing_flag  id_21_missing_flag  \
448539                   1                   1                   1
321311                   1                   1                   1
497320                   1                   1                   1
350951                   1                   1                   1
98132                    0                   0                   1

        id_22_missing_flag  id_23_missing_flag  id_24_missing_flag  \
448539                   1                   1                   1
321311                   1                   1                   1
497320                   1                   1                   1
350951                   1                   1                   1
98132                    1                   1                   1

        id_25_missing_flag  id_26_missing_flag  id_27_missing_flag  \
448539                   1                   1                   1
321311                   1                   1                   1
497320                   1                   1                   1
350951                   1                   1                   1
98132                    1                   1                   1

        id_28_missing_flag  id_29_missing_flag  id_30_missing_flag  \
448539                   1                   1                   1
321311                   1                   1                   1
497320                   1                   1                   1
350951                   1                   1                   1
98132                    0                   0                   0

        id_31_missing_flag  id_32_missing_flag  id_33_missing_flag  \
448539                   1                   1                   1
321311                   1                   1                   1
```

```
497320                   1                   1                   1
350951                   1                   1                   1
98132                    0                   0                   0

        id_34_missing_flag  id_35_missing_flag  id_36_missing_flag  \
448539                   1                   1                   1
321311                   1                   1                   1
497320                   1                   1                   1
350951                   1                   1                   1
98132                    0                   0                   0

        id_37_missing_flag  id_38_missing_flag  DeviceType_missing_flag  \
448539                   1                   1                        1
321311                   1                   1                        1
497320                   1                   1                        1
350951                   1                   1                        1
98132                    0                   0                        0

        DeviceInfo_missing_flag  _Weekdays  _Hours  _Days  Trans_min_mean  \
448539                        1          3       3     12       -27.09375
321311                        1          5      14      3       -57.09375
497320                        1          0      19     30       -86.00000
350951                        1          6       1     11       -77.06250
98132                         0          5      22     23        65.00000

        Trans_min_std  TransactionAmt_to_mean_card1  \
448539      -0.113281                      0.461914
321311      -0.238770                      0.645508
497320      -0.359619                      0.470703
350951      -0.322266                      0.555176
98132        0.271729                      1.193359

        TransactionAmt_to_mean_card4  TransactionAmt_to_std_card1  \
448539                      0.815430                     0.313331
321311                      0.585449                     0.401900
497320                      0.367920                     0.326992
350951                      0.435059                     0.570935
98132                       1.501953                     0.718662

        TransactionAmt_to_std_card4   PCA_V_0   PCA_V_1   PCA_V_2   PCA_V_3  \
448539                     0.425049 -0.782227  0.368896  0.263428  0.007843
321311                     0.341553 -0.782227  0.368896  0.263428  0.007904
497320                     0.214722 -0.801758  0.317871  0.270020 -0.027130
350951                     0.253906 -0.304443 -0.825684 -0.107727 -0.468994
98132                      0.876465  2.806641  0.345703  0.377197 -0.017654

         PCA_V_4   PCA_V_5   PCA_V_6   PCA_V_7   PCA_V_8   PCA_V_9  PCA_V_10  \
```

```
448539  0.013977 -0.007053  0.002455  0.334717 -0.010490 -0.220459 -0.042755
321311  0.013786 -0.007095  0.002304  0.334961 -0.010521 -0.220825 -0.042938
497320  0.058411 -0.003780 -0.014931 -0.225342  0.021576  0.033905 -0.024246
350951 -0.162598  0.029373  0.138550  0.065613 -0.067749  0.260010  0.052063
98132  -0.042450 -0.244263  0.023148  0.011810  0.000937  0.023010 -0.009560

        PCA_V_11  PCA_V_12  PCA_V_13  PCA_V_14  PCA_V_15  PCA_V_16  PCA_V_17  \
448539  0.021561 -0.005577 -0.020416  0.005253 -0.000285  0.004505 -0.006889
321311  0.020325 -0.003403 -0.022324  0.005749  0.000531  0.004292 -0.009888
497320  0.010124  0.020370  0.017105 -0.014145 -0.013794 -0.005302 -0.077637
350951 -0.091125  0.068176  0.014664  0.010857 -0.011421 -0.020767  0.009926
98132  -0.022293 -0.021683 -0.013680 -0.005581 -0.009674  0.001465 -0.008263

        PCA_V_18  PCA_V_19  PCA_V_20  PCA_V_21  PCA_V_22  PCA_V_23  PCA_V_24  \
448539  0.008362 -0.011055 -0.006924 -0.033722  0.011093 -0.011307 -0.007355
321311  0.009216 -0.005642 -0.005966 -0.034119  0.012009 -0.009644 -0.006516
497320  0.056580  0.115479  0.013458 -0.012100  0.003590  0.023392  0.003922
350951  0.008850  0.005054 -0.041290  0.041382 -0.051819  0.042450  0.026703
98132  -0.008881  0.005920 -0.004478 -0.000713 -0.003057  0.008438 -0.017868

        PCA_V_25  PCA_V_26  PCA_V_27  PCA_V_28  PCA_V_29
448539 -0.004627  0.002234 -0.005222  0.004154 -0.001096
321311 -0.002327  0.000332 -0.006203  0.002287 -0.005352
497320  0.032867  0.010475  0.007317  0.008362 -0.037079
350951 -0.046417 -0.029648  0.022217  0.004562 -0.008606
98132   0.000072  0.013344 -0.005264 -0.000665  0.007454

[5 rows x 533 columns]
```

11 11. Model Building

12 XGBoost Classifier

[ ]:

[84]:
```
X_train = X_train.replace([np.inf, -np.inf], np.nan)
X_test = X_test.replace([np.inf, -np.inf], np.nan)

median_values = X_train.median()
X_train = X_train.fillna(median_values)
X_test = X_test.fillna(median_values)
```

[85]:
```
%%time
xgb = XGBClassifier(nthread=-1, random_state=0)
xgb.fit(X_train, y_train)
xgb
```

```
[13:34:41] WARNING: C:/Users/Administrator/workspace/xgboost-
win64_release_1.4.0/src/learner.cc:1095: Starting in XGBoost 1.3.0, the default
evaluation metric used with the objective 'binary:logistic' was changed from
'error' to 'logloss'. Explicitly set eval_metric if you'd like to restore the
old behavior.
Wall time: 7min 4s
```

[85]:
```
XGBClassifier(base_score=0.5, booster='gbtree', colsample_bylevel=1,
              colsample_bynode=1, colsample_bytree=1, gamma=0, gpu_id=-1,
              importance_type='gain', interaction_constraints='',
              learning_rate=0.300000012, max_delta_step=0, max_depth=6,
              min_child_weight=1, missing=nan, monotone_constraints='()',
              n_estimators=100, n_jobs=8, nthread=-1, num_parallel_tree=1,
              random_state=0, reg_alpha=0, reg_lambda=1, scale_pos_weight=1,
              subsample=1, tree_method='exact', validate_parameters=1,
              verbosity=None)
```

[86]:
```
y_pred_xgb = xgb.predict(X_test)
y_prob_pred_xgb = xgb.predict_proba(X_test)
y_prob_pred_xgb = [x[1] for x in y_prob_pred_xgb]
print("Y predicted : ",y_pred_xgb)
print("Y probability predicted : ",y_prob_pred_xgb[:5])
```

```
Y predicted :  [False False False … False False False]
Y probability predicted :  [0.0032196122, 0.0040678284, 0.007803707,
0.023133403, 0.0024363862]
```

13 12. Evaluation Metrics

- Accuracy Score
- Confusion Matrix
- Classification Report
- AUC Score
- Concordance Index
- ROC curve
- PR curve

[87]:
```
from bisect import bisect_left, bisect_right

def concordance(actuals, preds):
    ones_preds  = [p for a,p in zip(actuals, preds) if a == 1]
    zeros_preds = [p for a,p in zip(actuals, preds) if a == 0]
    n_ones      = len([x for x in actuals if x == 1])
    n_total_pairs =  float(n_ones) * float(len(actuals) - n_ones)
    # print("Total Pairs: ", n_total_pairs)

    zeros_sorted = sorted(zeros_preds)
```

```
    conc = 0; disc = 0; ties = 0;
    for i, one_pred in enumerate(ones_preds):
        cur_conc = bisect_left(zeros_sorted, one_pred)
        cur_ties = bisect_right(zeros_sorted, one_pred) - cur_conc
        conc += cur_conc
        ties += cur_ties
        disc += float(len(zeros_sorted)) - cur_ties - cur_conc

    concordance = conc/n_total_pairs
    discordance = disc/n_total_pairs
    ties_perc = ties/n_total_pairs
    return concordance
```

[88]:
```
def compute_evaluation_metric(model, x_test, y_actual, y_predicted,␣
↪y_predicted_prob):
    print("\nAccuracy Score : ", accuracy_score(y_actual, y_predicted))
    print("\nAUC Score : ", roc_auc_score(y_actual, y_predicted_prob))
    print("\nConfusion Matrix : \n", confusion_matrix(y_actual, y_predicted))
    print("\nClassification Report : \n", classification_report(y_actual,␣
↪y_predicted))

    # Concordance index if function exists
    try:
        print("\nConcordance Index : ", concordance(y_actual, y_predicted_prob))
    except:
        print("\nConcordance Index : Function 'concordance' not defined")

    # ROC curve
    print("\nROC curve : \n")
    plot_roc_curve(model, x_test, y_actual)
    plt.show()

    # PR curve
    print("\nPR curve : \n")
    plot_precision_recall_curve(model, x_test, y_actual)
    plt.show()

    # --- Compute TPR, FPR, TNR, FNR ---
    tn, fp, fn, tp = confusion_matrix(y_actual, y_predicted).ravel()

    tpr = tp / (tp + fn)  # True Positive Rate / Recall
    fpr = fp / (fp + tn)  # False Positive Rate
    tnr = tn / (tn + fp)  # True Negative Rate / Specificity
    fnr = fn / (fn + tp)  # False Negative Rate

    print("\nAdditional Metrics:")
    print(f"TPR (Recall) : {tpr:.4f}")
```

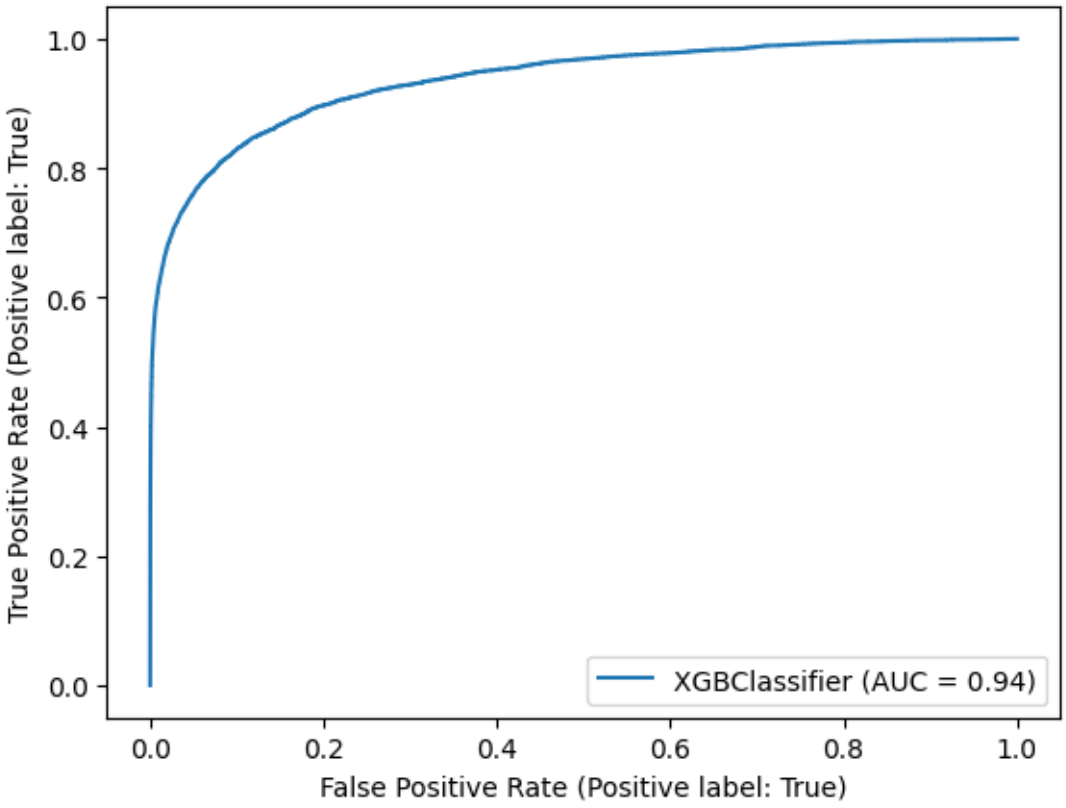
```
    print(f"FPR : {fpr:.4f}")
    print(f"TNR (Specificity) : {tnr:.4f}")
    print(f"FNR : {fnr:.4f}")
```

[89]:
```
concordance(y_test.values, y_prob_pred_xgb)
```

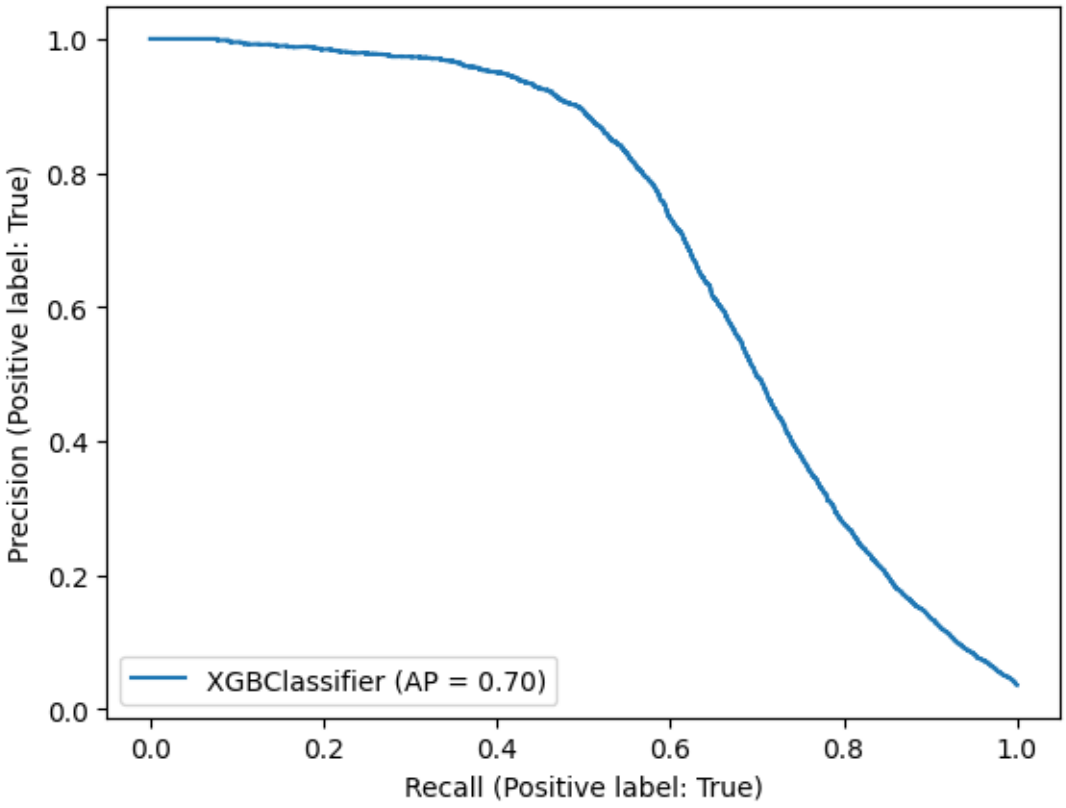
[89]:
```
0.9363753273299149
```

[90]:
```
compute_evaluation_metric(xgb, X_test, y_test, y_pred_xgb, y_prob_pred_xgb)
```

```
Accuracy Score :  0.9801086011672933

AUC Score :  0.9363753419553209

Confusion Matrix :
 [[170646    317]
 [  3207   2992]]

Classification Report :
              precision    recall  f1-score   support

       False       0.98      1.00      0.99    170963
        True       0.90      0.48      0.63      6199

    accuracy                           0.98    177162
   macro avg       0.94      0.74      0.81    177162
weighted avg       0.98      0.98      0.98    177162


Concordance Index :  0.9363753273299149

ROC curve :
```


1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate (Positive label: True)
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate (Positive label: True)
XGBClassifier (AUC = 0.94)

PR curve :




```
Additional Metrics:
TPR (Recall) : 0.4827
FPR : 0.0019
TNR (Specificity) : 0.9981
FNR : 0.5173
```

14 13. Capture Rates and Calibration Curve

Divide the data in 10 equal bins as per predicted probability scores. Then, compute the percentage of the total target class 1 captured in every bin.

Ideally the proportion should be decreasing as we go down ever bin. Let's check it out

[91]:
```
# Create Validation set
validation_df = {'y_test' : y_test, 'y_pred' : y_pred_xgb, 'y_pred_prob' :␣
 ↪y_prob_pred_xgb}
validation_df = pd.DataFrame(data = validation_df)

# Add binning column to the dataframe
validation_df['bin_y_pred_prob'] = pd.qcut(validation_df['y_pred_prob'], q=10)
```

```
validation_df.head()
```

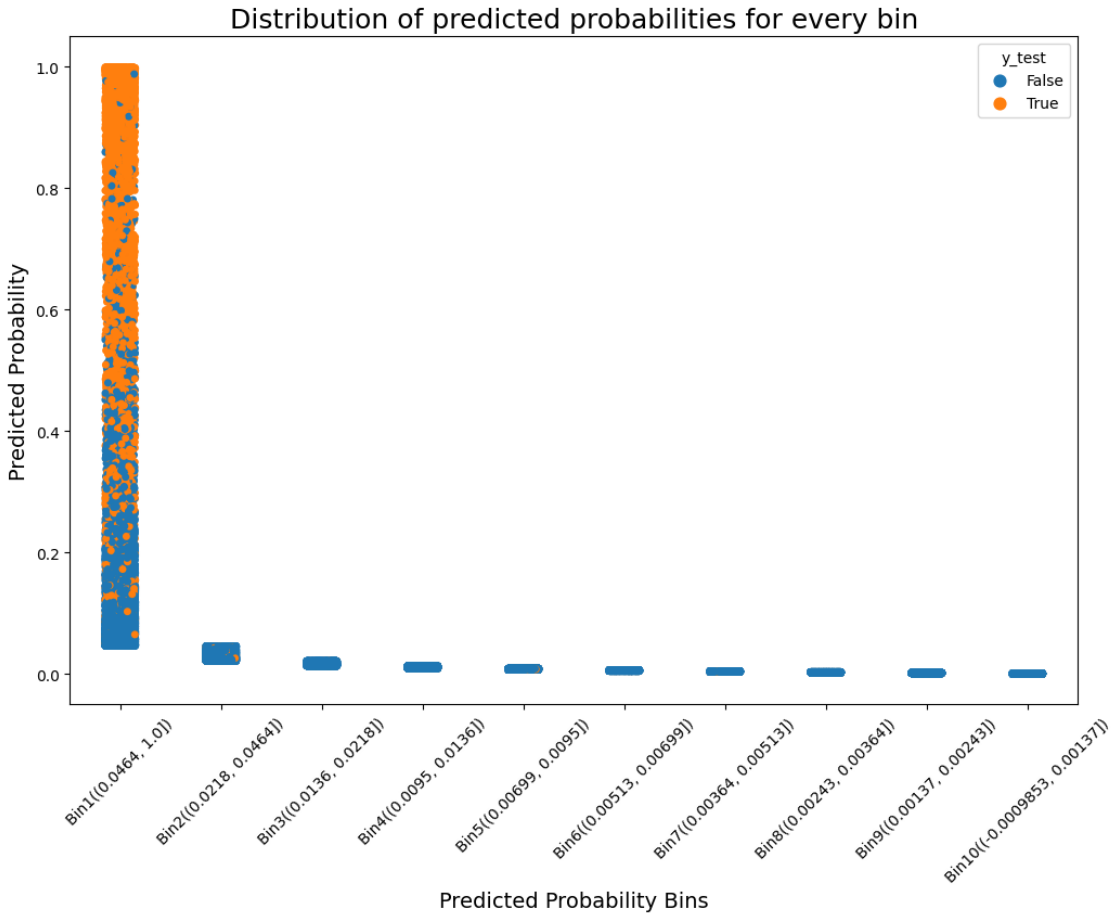
[91]:

	y_test	y_pred	y_pred_prob	bin_y_pred_prob
274635	False	False	0.003220	(0.00243, 0.00364]
115501	False	False	0.004068	(0.00364, 0.00513]
387765	False	False	0.007804	(0.00699, 0.0095]
123591	False	False	0.023133	(0.0218, 0.0464]
462833	False	False	0.002436	(0.00243, 0.00364]

[92]:
```
x_label = []
for i in range(len(validation_df['bin_y_pred_prob'].cat.categories[::-1].
 ↪astype('str'))):
    x_label.append("Bin" + str(i + 1)+ "(" + validation_df['bin_y_pred_prob'].
 ↪cat.categories[::-1].astype('str')[i] + ")")
```

Capture Rates Plot

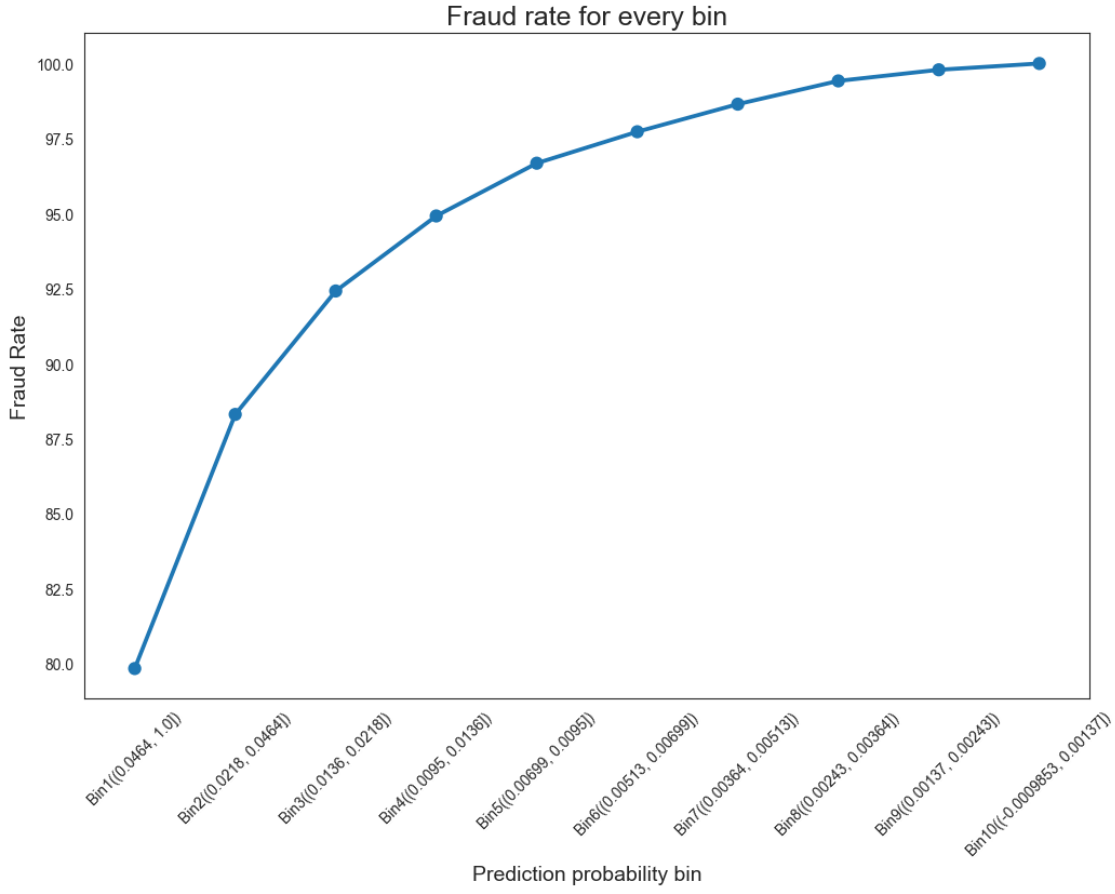
[93]:
```
# Plot Distribution of predicted probabilities for every bin
plt.figure(figsize=(12, 8));
sns.stripplot(validation_df.bin_y_pred_prob, validation_df.y_pred_prob, jitter␣
 ↪= 0.15, hue = validation_df.y_test, order = validation_df['bin_y_pred_prob'].
 ↪cat.categories[::-1])
plt.title("Distribution of predicted probabilities for every bin", fontsize=18)
plt.xlabel("Predicted Probability Bins", fontsize=14);
plt.ylabel("Predicted Probability", fontsize=14);
plt.xticks(np.arange(10), x_label, rotation=45);
plt.show()
```



Gains Table

[94]:
```
# Aggregate the data
gains_df              = validation_df.groupby(["bin_y_pred_prob","y_test"]).
 ↪agg({'y_test': ['count']})
gains_df.columns      = gains_df.columns.map(''.join)
gains_df['prob_bin'] = gains_df.index.get_level_values(0)
gains_df['y_test']    = gains_df.index.get_level_values(1)
gains_df.reset_index(drop = True, inplace = True)
gains_df

# Get infection rate and percentage infections
gains_table = gains_df.pivot(index='prob_bin', columns='y_test',␣
 ↪values='y_testcount')
gains_table['prob_bin'] = gains_table.index
gains_table = gains_table.iloc[::-1]
gains_table['prob_bin'] = x_label
gains_table.reset_index(drop = True, inplace = True)
gains_table = gains_table[['prob_bin', 0, 1]]
```

```
gains_table.columns = ['prob_bin', "not_fraud", "fraud"]
gains_table['perc_fraud'] = gains_table['fraud']/gains_table['fraud'].sum()
gains_table['perc_not_fraud'] = gains_table['not_fraud']/
 ↪gains_table['not_fraud'].sum()
gains_table['cum_perc_fraud'] = 100*(gains_table.fraud.cumsum() / gains_table.
 ↪fraud.sum())
gains_table['cum_perc_not_fraud'] = 100*(gains_table.not_fraud.cumsum() /␣
 ↪gains_table.not_fraud.sum())
gains_table


# Plot
plt.figure(figsize=(12, 8));
sns.set_style("white")
sns.pointplot(x = "prob_bin", y = "cum_perc_fraud", data = gains_table, legend␣
 ↪= False, order=gains_table.prob_bin)
plt.xticks(rotation=45);
plt.ylabel("Fraud Rate", fontsize=14)
plt.xlabel("Prediction probability bin", fontsize=14)
plt.title("Fraud rate for every bin", fontsize=18)
plt.show()
```

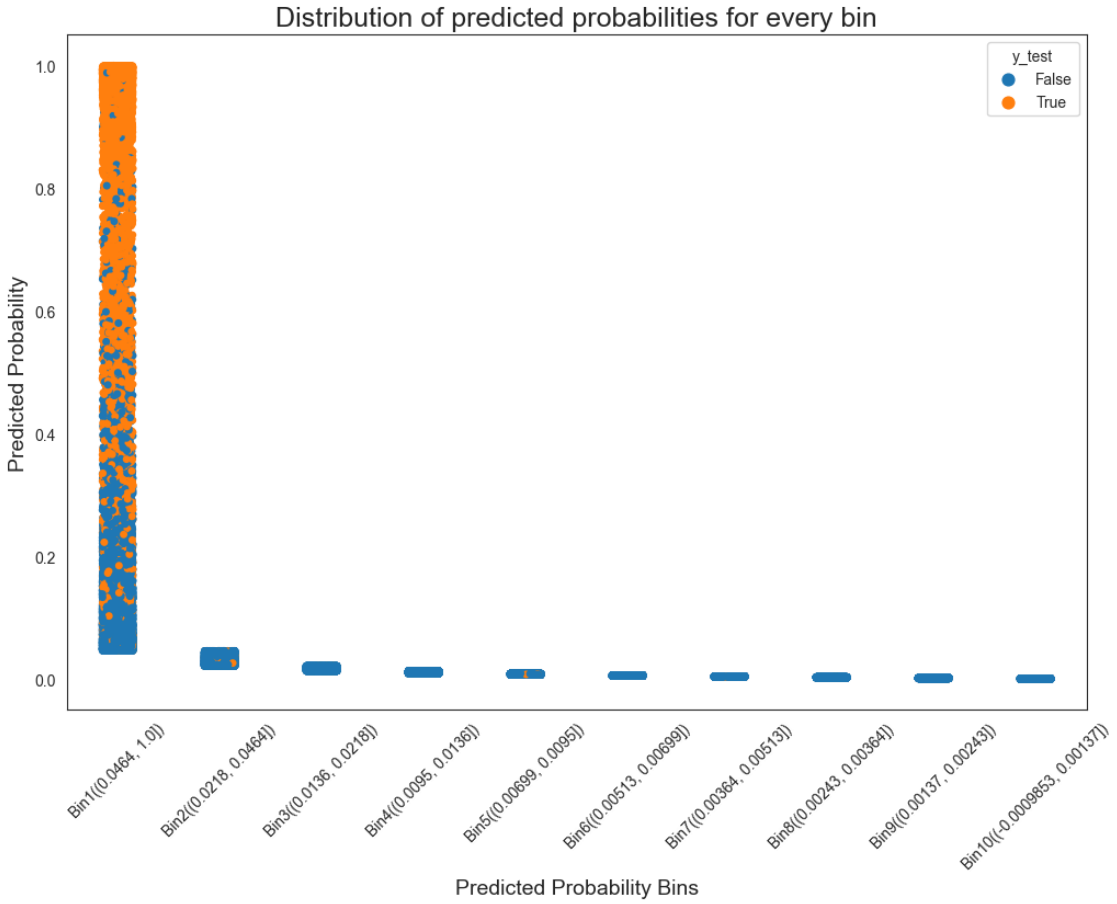


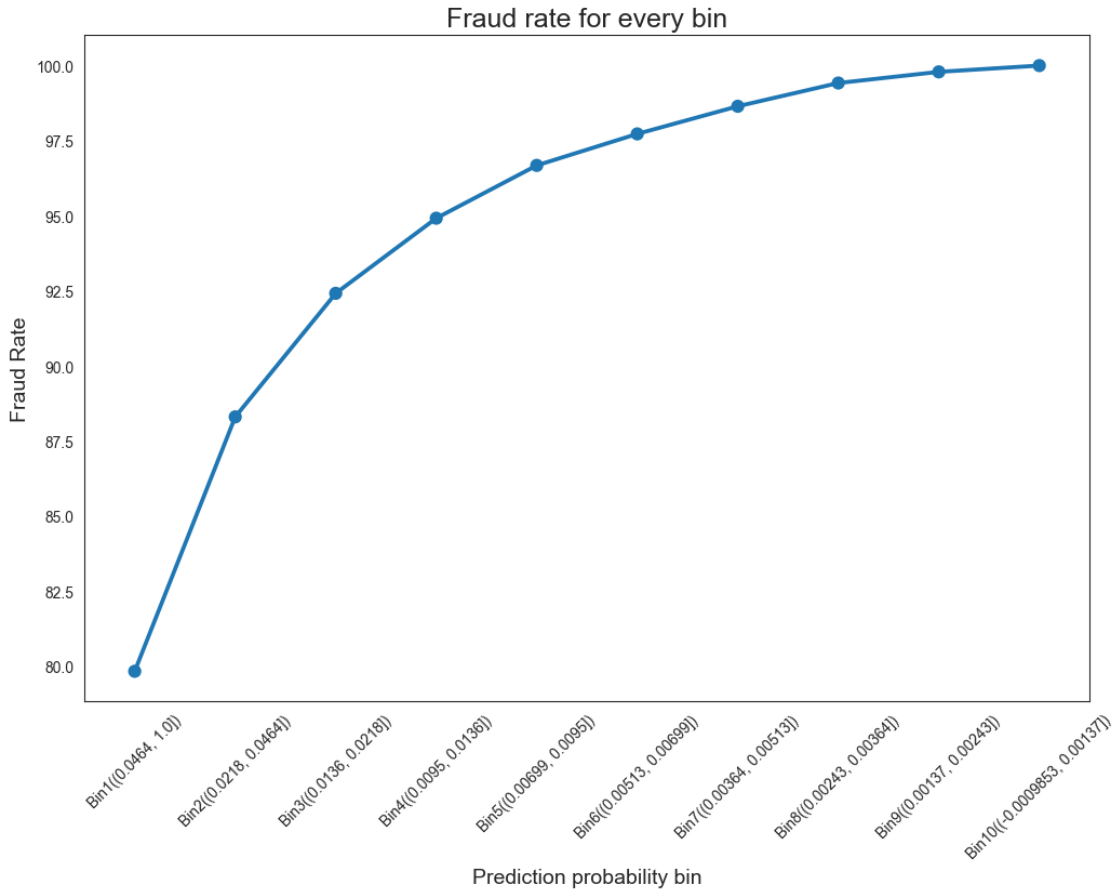
Ideally the slope should be high initially and should decrease as we move further to the right. This is not really a good model.

[95]:
```
def captures(y_test, y_pred, y_pred_prob):
    # Create Validation set
    validation_df = {'y_test' : y_test, 'y_pred' : y_pred, 'y_pred_prob' :␣
 ↪y_pred_prob}
    validation_df = pd.DataFrame(data = validation_df)

    # Add binning column to the dataframe
    try:
        validation_df['bin_y_pred_prob'] = pd.
 ↪qcut(validation_df['y_pred_prob'], q=10)
    except:
        validation_df['bin_y_pred_prob'] = pd.
 ↪qcut(validation_df['y_pred_prob'], q=10, duplicates='drop')

    # Change x label and column names
    x_label = []
    for i in range(len(validation_df['bin_y_pred_prob'].cat.categories[::-1].
 ↪astype('str'))):
        x_label.append("Bin" + str(i + 1)+ "(" +␣
 ↪validation_df['bin_y_pred_prob'].cat.categories[::-1].astype('str')[i] + ")")

    # Plot Distribution of predicted probabilities for every bin
    plt.figure(figsize=(12, 8));
    sns.stripplot(validation_df.bin_y_pred_prob, validation_df.y_pred_prob,␣
 ↪jitter = 0.15, hue = validation_df.y_test, order =␣
 ↪validation_df['bin_y_pred_prob'].cat.categories[::-1])
    plt.title("Distribution of predicted probabilities for every bin",␣
 ↪fontsize=18)
    plt.xlabel("Predicted Probability Bins", fontsize=14);
    plt.ylabel("Predicted Probability", fontsize=14);
    try:
        plt.xticks(np.arange(10), x_label, rotation=45);
    except:
        pass
    plt.show()

    # Aggregate the data
    gains_df              = validation_df.groupby(["bin_y_pred_prob","y_test"]).
 ↪agg({'y_test': ['count']})
    gains_df.columns      = gains_df.columns.map(''.join)
    gains_df['prob_bin'] = gains_df.index.get_level_values(0)
```

```
    gains_df['y_test']   = gains_df.index.get_level_values(1)
    gains_df.reset_index(drop = True, inplace = True)
    gains_df

    # Get infection rate and percentage infections
    gains_table = gains_df.pivot(index='prob_bin', columns='y_test',␣
↪values='y_testcount')
    gains_table['prob_bin'] = gains_table.index
    gains_table = gains_table.iloc[::-1]
    gains_table['prob_bin'] = x_label
    gains_table.reset_index(drop = True, inplace = True)
    gains_table = gains_table[['prob_bin', 0, 1]]
    gains_table.columns = ['prob_bin', "not_fraud", "fraud"]
    gains_table['perc_fraud'] = gains_table['fraud']/gains_table['fraud'].sum()
    gains_table['perc_not_fraud'] = gains_table['not_fraud']/
↪gains_table['not_fraud'].sum()
    gains_table['cum_perc_fraud'] = 100*(gains_table.fraud.cumsum() /␣
↪gains_table.fraud.sum())
    gains_table['cum_perc_not_fraud'] = 100*(gains_table.not_fraud.cumsum() /␣
↪gains_table.not_fraud.sum())
    gains_table


    # Plot
    plt.figure(figsize=(12, 8));
    sns.set_style("white")
    sns.pointplot(x = "prob_bin", y = "cum_perc_fraud", data = gains_table,␣
↪legend = False, order=gains_table.prob_bin)
    plt.xticks(rotation=45);
    plt.ylabel("Fraud Rate", fontsize=14)
    plt.xlabel("Prediction probability bin", fontsize=14)
    plt.title("Fraud rate for every bin", fontsize=18)
    plt.show()
    return gains_table
```

[96]:
```
captures(y_test, y_pred_xgb, y_prob_pred_xgb)
```


Distribution of predicted probabilities for every bin
y_test
False
True
Predicted Probability
1.0
0.8
0.6
0.4
0.2
0.0
Bin1((0.0464, 1.0])
Bin2((0.0218, 0.0464])
Bin3((0.0136, 0.0218])
Bin4((0.0095, 0.0136])
Bin5((0.00699, 0.0095])
Bin6((0.00513, 0.00699])
Bin7((0.00364, 0.00513])
Bin8((0.00243, 0.00364])
Bin9((0.00137, 0.00243])
Bin10((-0.0009853, 0.00137])
Predicted Probability Bins



[96]:

	prob_bin	not_fraud	fraud	perc_fraud	perc_not_fraud	\
0	Bin1((0.0464, 1.0])	12768	4949	0.798355	0.074683	
1	Bin2((0.0218, 0.0464])	17191	525	0.084691	0.100554	
2	Bin3((0.0136, 0.0218])	17461	255	0.041136	0.102133	
3	Bin4((0.0095, 0.0136])	17561	155	0.025004	0.102718	
4	Bin5((0.00699, 0.0095])	17607	109	0.017583	0.102987	
5	Bin6((0.00513, 0.00699])	17651	65	0.010486	0.103245	
6	Bin7((0.00364, 0.00513])	17659	57	0.009195	0.103291	
7	Bin8((0.00243, 0.00364])	17668	48	0.007743	0.103344	
8	Bin9((0.00137, 0.00243])	17693	23	0.003710	0.103490	
9	Bin10((-0.0009853, 0.00137])	17704	13	0.002097	0.103555	

	cum_perc_fraud	cum_perc_not_fraud
0	79.835457	7.468283
1	88.304565	17.523675
2	92.418132	27.736996
3	94.918535	38.008809
4	96.676883	48.307529
5	97.725440	58.631985

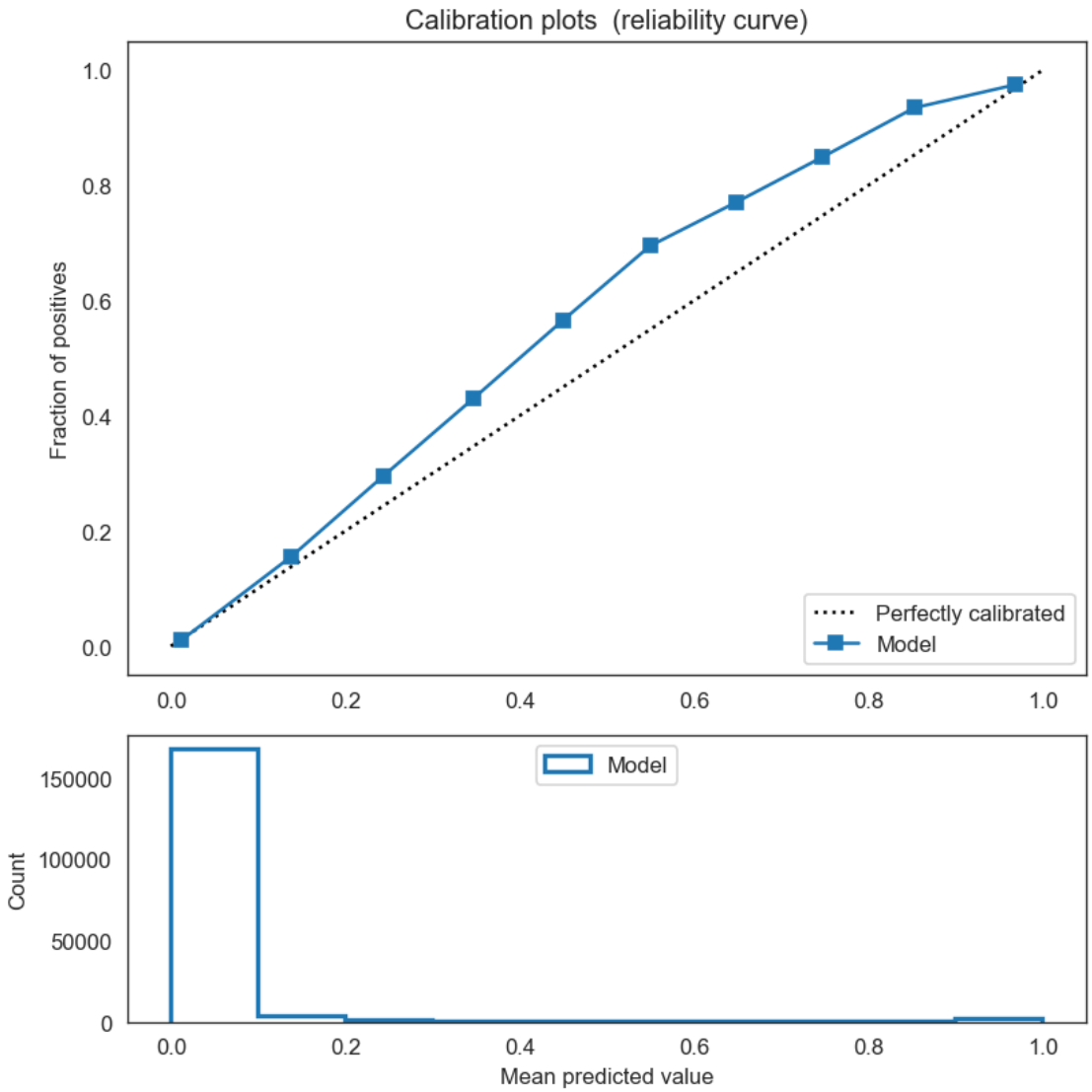
```
6       98.644943          68.961120
7       99.419261          79.295520
8       99.790289          89.644543
9      100.000000         100.000000
```

14.1 Calibration Curve

[97]:
```
from sklearn.calibration import calibration_curve
import matplotlib.pyplot as plt
```

[98]:
```
def draw_calibration_curve(y_test, y_prob, n_bins=10):
    plt.figure(figsize=(7, 7), dpi=120)
    ax1 = plt.subplot2grid((3, 1), (0, 0), rowspan=2)
    ax2 = plt.subplot2grid((3, 1), (2, 0))
    ax1.plot([0, 1], [0, 1], "k:", label="Perfectly calibrated")


    fraction_of_positives, mean_predicted_value = calibration_curve(y_test,␣
 ↪y_prob, n_bins=10)

    ax1.plot(mean_predicted_value, fraction_of_positives, "s-", label="%s" %␣
 ↪("Model", ))
    ax2.hist(y_prob, range=(0, 1), bins=10, label="Model", histtype="step",␣
 ↪lw=2)

    # Labels
    ax1.set_ylabel("Fraction of positives")
    ax1.set_ylim([-0.05, 1.05])
    ax1.legend(loc="lower right")
    ax1.set_title('Calibration plots  (reliability curve)')

    ax2.set_xlabel("Mean predicted value")
    ax2.set_ylabel("Count")
    ax2.legend(loc="upper center", ncol=2)

    plt.tight_layout()
    plt.show()
```

[99]:
```
draw_calibration_curve(y_test, y_prob_pred_xgb, n_bins=10)
```



15 14. Model Comparison and Evaluation

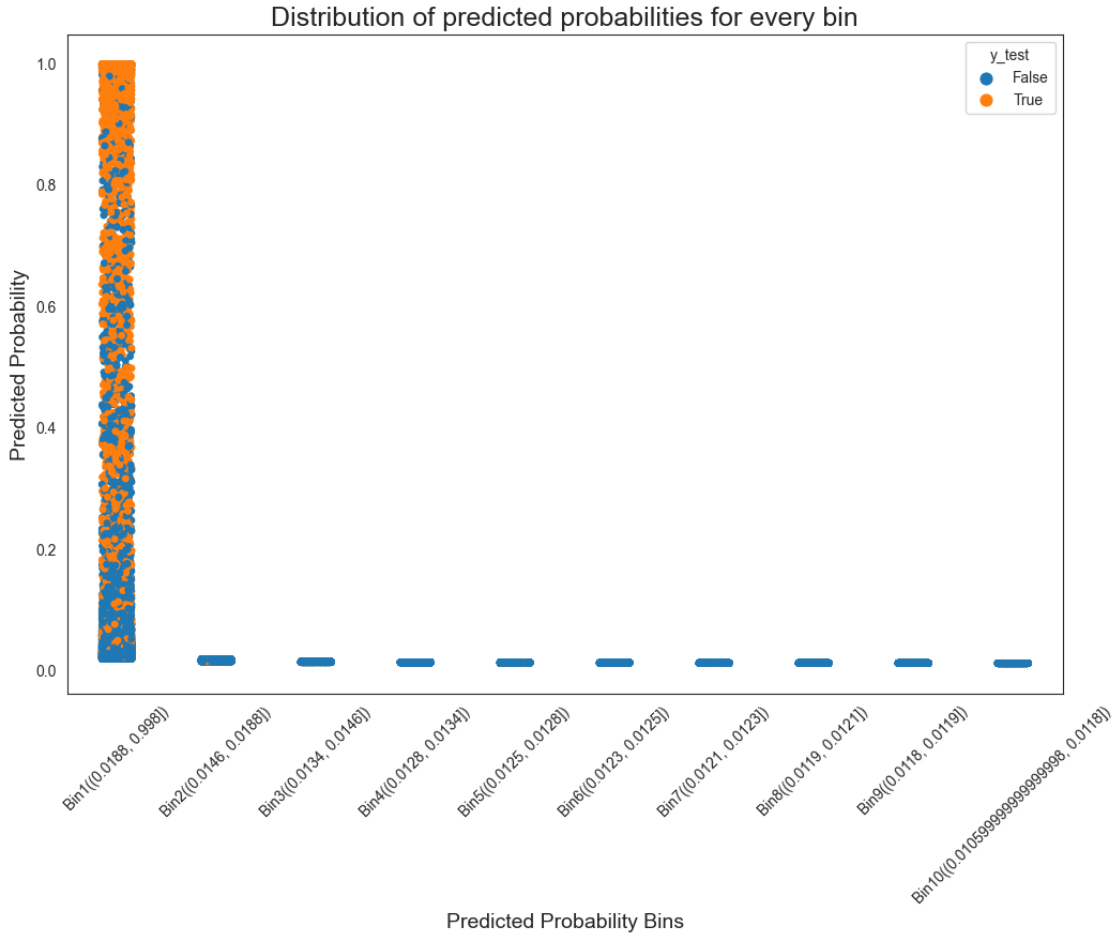
15.1 A) Logistic regression

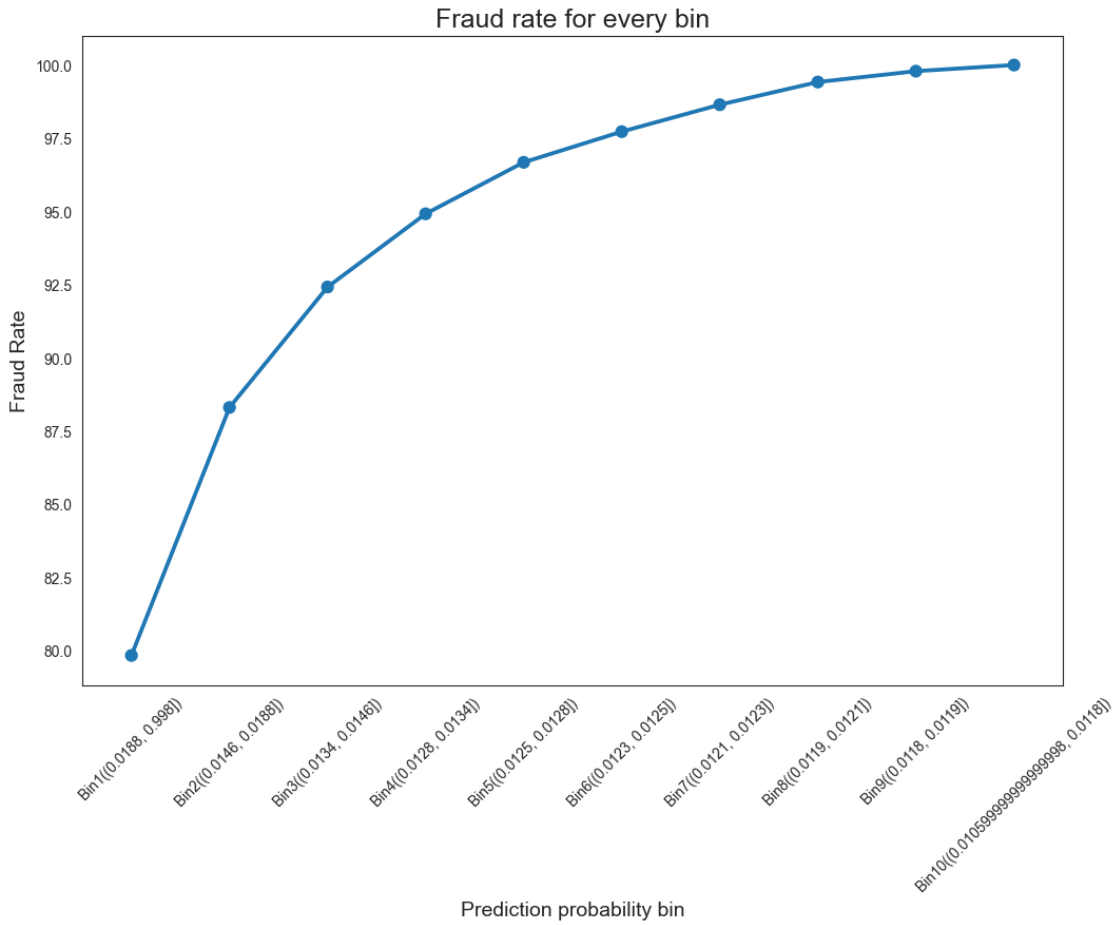
[100]:
```
# Prediction
y_pred_xgb_test = xgb.predict(X_test)
y_prob_pred_xgb_test = xgb.predict_proba(X_test)[:, 1]
```

[101]:
```
from sklearn.linear_model import LogisticRegression
X = np.array(y_prob_pred_xgb_test)
clf = LogisticRegression(random_state=0).fit(X.reshape(-1, 1), y_test)
```

[102]:
```
y_prob_pred_calib = clf.predict_proba(X.reshape(-1, 1))[:, 1]
y_pred_calib      = clf.predict(X.reshape(-1, 1))
```

[103]:
```
captures(y_test, y_pred_calib, y_prob_pred_calib)
```





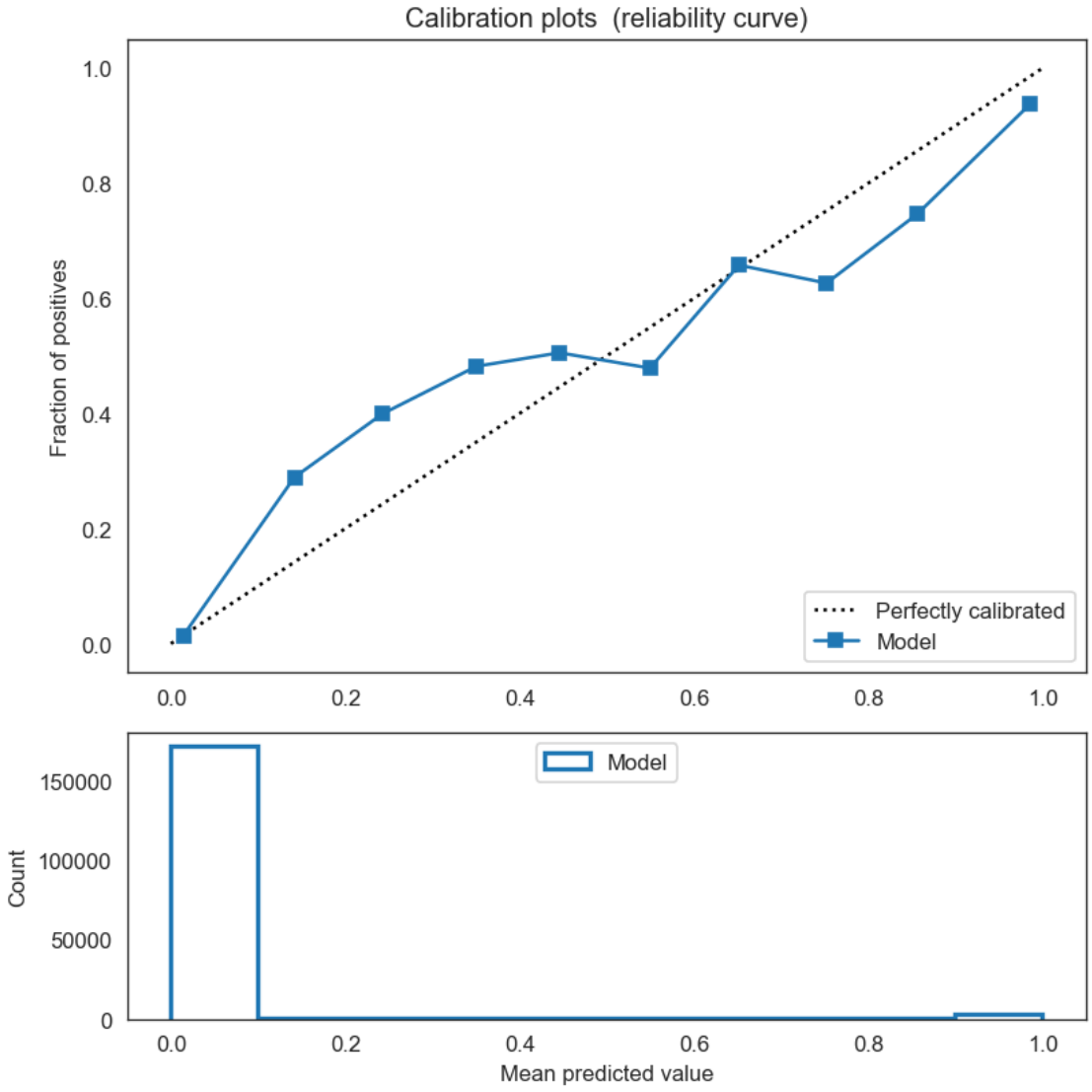
[103]:

	prob_bin	not_fraud	fraud	perc_fraud
0	Bin1((0.0188, 0.998])	12768	4949	0.798355
1	Bin2((0.0146, 0.0188])	17191	525	0.084691
2	Bin3((0.0134, 0.0146])	17461	255	0.041136
3	Bin4((0.0128, 0.0134])	17561	155	0.025004
4	Bin5((0.0125, 0.0128])	17607	109	0.017583
5	Bin6((0.0123, 0.0125])	17651	65	0.010486
6	Bin7((0.0121, 0.0123])	17659	57	0.009195
7	Bin8((0.0119, 0.0121])	17668	48	0.007743
8	Bin9((0.0118, 0.0119])	17693	23	0.003710
9	Bin10((0.010599999999999998, 0.0118])	17704	13	0.002097

	perc_not_fraud	cum_perc_fraud	cum_perc_not_fraud
0	0.074683	79.835457	7.468283
1	0.100554	88.304565	17.523675
2	0.102133	92.418132	27.736996
3	0.102718	94.918535	38.008809
4	0.102987	96.676883	48.307529

```
5       0.103245       97.725440          58.631985
6       0.103291       98.644943          68.961120
7       0.103344       99.419261          79.295520
8       0.103490       99.790289          89.644543
9       0.103555      100.000000         100.000000
```

[104]:
```
draw_calibration_curve(y_test, y_prob_pred_calib, n_bins=10)
```



15.2 B) LightGBM

[107]:
```
from lightgbm import LGBMClassifier
```

[108]:
```
%%time
lgbc = LGBMClassifier(random_state=0, n_jobs = -1)
lgbc.fit(X_train,y_train)
lgbc
```

```
Wall time: 29.5 s
```

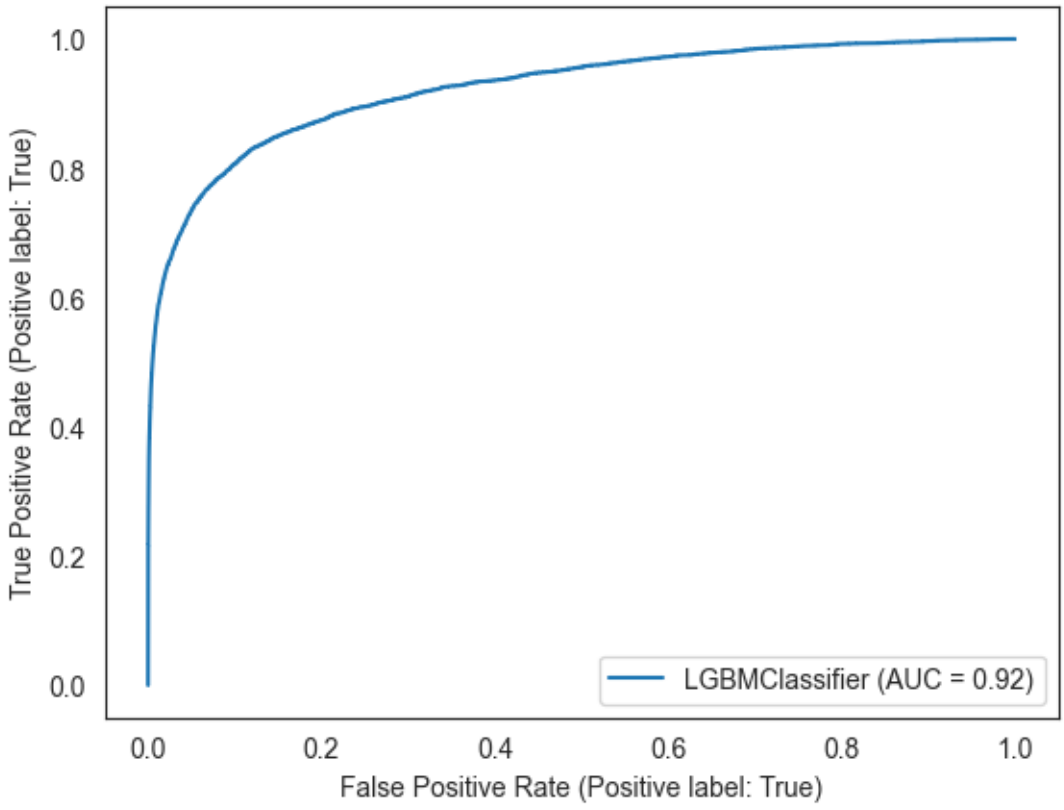
[108]:
```
LGBMClassifier(random_state=0)
```

[109]:
```
y_pred_lgbc = lgbc.predict(X_test)
y_prob_pred_lgbc = lgbc.predict_proba(X_test)
y_prob_pred_lgbc = [x[1] for x in y_prob_pred_lgbc]
print("Y predicted : ",y_pred_lgbc)
print("Y probability predicted : ",y_prob_pred_lgbc[:5])
```

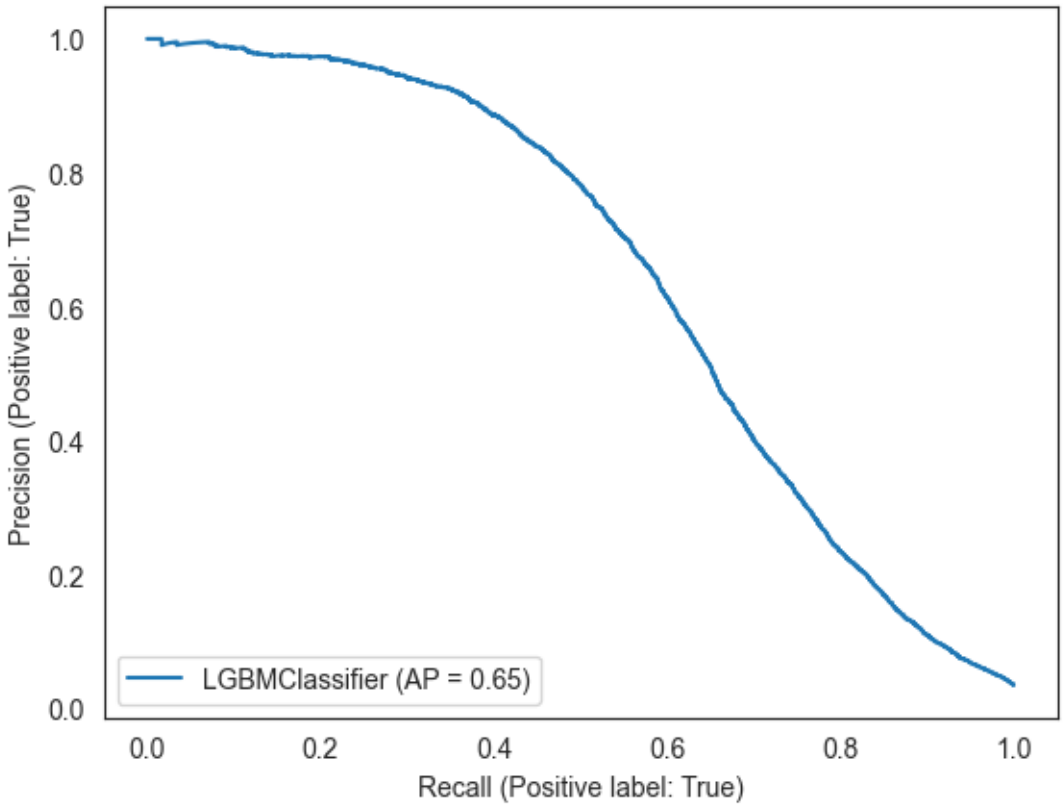
```
Y predicted :  [False False False … False False False]
Y probability predicted :  [0.016160027443867835, 0.0080773782284121,
0.00825509576482752, 0.06708271917846045, 0.007287115458060899]
```

[110]:
```
compute_evaluation_metric(lgbc, X_test, y_test, y_pred_lgbc, y_prob_pred_lgbc)
```

```
Accuracy Score :  0.9774613066007383

AUC Score :  0.9239690737882372

Confusion Matrix :
 [[170585    378]
 [  3615   2584]]

Classification Report :
              precision    recall  f1-score   support

       False       0.98      1.00      0.99    170963
        True       0.87      0.42      0.56      6199

    accuracy                           0.98    177162
   macro avg       0.93      0.71      0.78    177162
weighted avg       0.98      0.98      0.97    177162


Concordance Index :  0.9239690501988727

ROC curve :
```

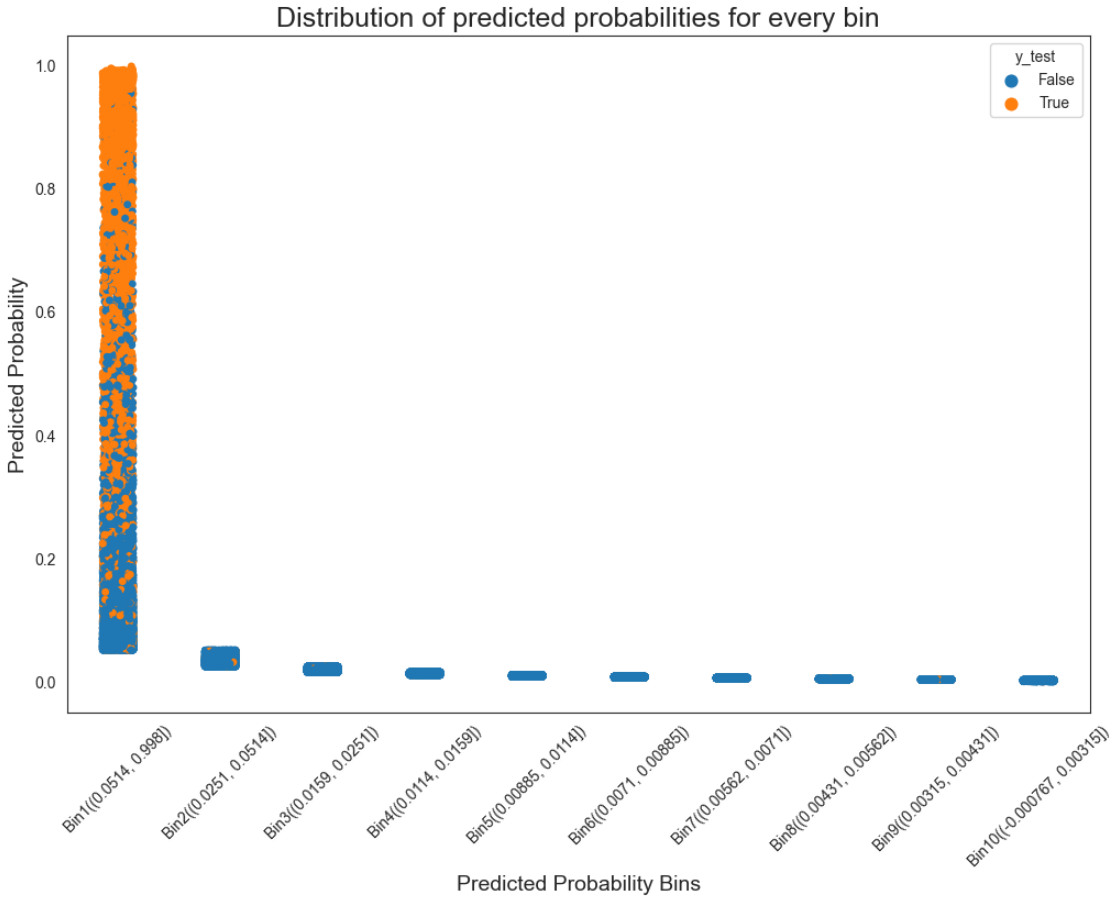

1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate (Positive label: True)
LGBMClassifier (AUC = 0.92)
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate (Positive label: True)

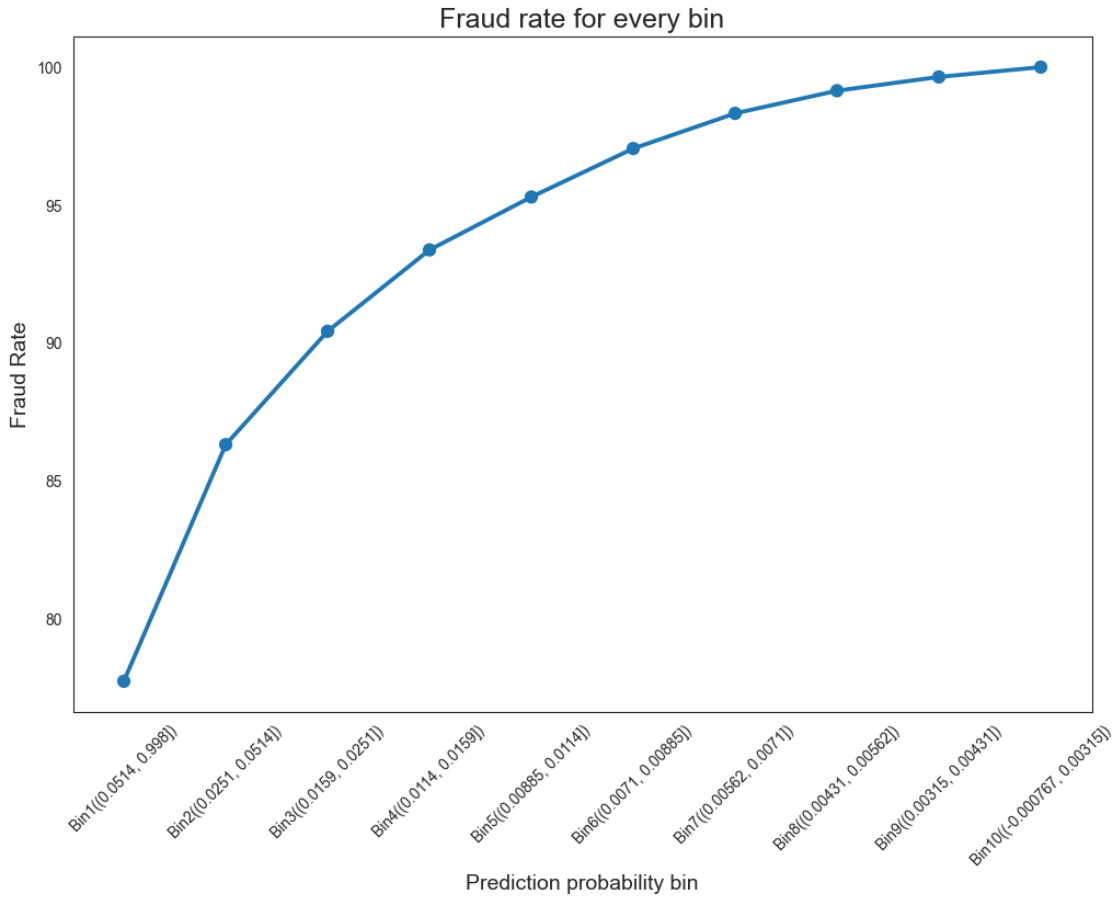
PR curve :




```
Additional Metrics:
TPR (Recall) : 0.4168
FPR : 0.0022
TNR (Specificity) : 0.9978
FNR : 0.5832
```

[111]:
```
captures(y_test, y_pred_lgbc, y_prob_pred_lgbc)
```


Distribution of predicted probabilities for every bin
y_test
False
True
Predicted Probability
1.0
0.8
0.6
0.4
0.2
0.0
Bin1((0.0514, 0.998])
Bin2((0.0251, 0.0514])
Bin3((0.0159, 0.0251])
Bin4((0.0114, 0.0159])
Bin5((0.00885, 0.0114])
Bin6((0.0071, 0.00885])
Bin7((0.00562, 0.0071])
Bin8((0.00431, 0.00562])
Bin9((0.00315, 0.00431])
Bin10((-0.000767, 0.00315])
Predicted Probability Bins



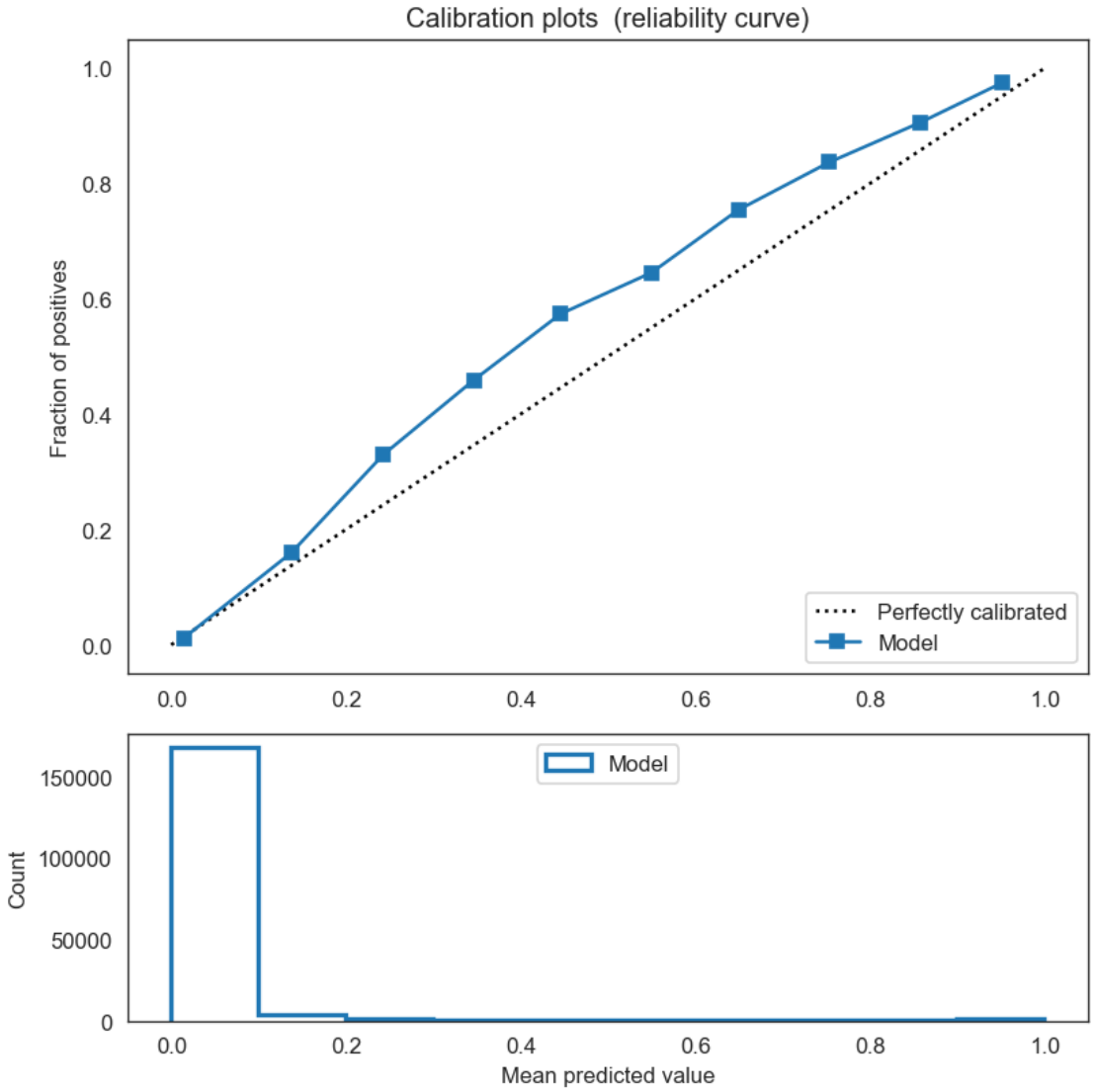
[111]:

	prob_bin	not_fraud	fraud	perc_fraud	perc_not_fraud	\
0	Bin1((0.0514, 0.998])	12899	4818	0.777222	0.075449	
1	Bin2((0.0251, 0.0514])	17184	532	0.085820	0.100513	
2	Bin3((0.0159, 0.0251])	17461	255	0.041136	0.102133	
3	Bin4((0.0114, 0.0159])	17533	183	0.029521	0.102554	
4	Bin5((0.00885, 0.0114])	17597	119	0.019197	0.102929	
5	Bin6((0.0071, 0.00885])	17607	109	0.017583	0.102987	
6	Bin7((0.00562, 0.0071])	17637	79	0.012744	0.103163	
7	Bin8((0.00431, 0.00562])	17665	51	0.008227	0.103326	
8	Bin9((0.00315, 0.00431])	17685	31	0.005001	0.103443	
9	Bin10((-0.000767, 0.00315])	17695	22	0.003549	0.103502	

	cum_perc_fraud	cum_perc_not_fraud
0	77.722213	7.544907
1	86.304243	17.596205
2	90.417809	27.809526
3	93.369898	38.064961
4	95.289563	48.357832
5	97.047911	58.656551

6	98.322310	68.972819
7	99.145023	79.305464
8	99.645104	89.649807
9	100.000000	100.000000

[112]:
```
draw_calibration_curve(y_test, y_prob_pred_lgbc, n_bins=10)
```



- With LGBM, Accuracy score is 97.7%. It's almost similar to XGBoost model
- AUC score has imporved to 92.6 from 88.5
- Recall and f-1 score have also improved, but it's still not upto the mark

15.3 C) Random Forest Classifier

[113]:
```
from sklearn.ensemble import RandomForestClassifier, ExtraTreesClassifier,␣
↪GradientBoostingClassifier
```

[114]:
```
X_train.head()
```

[114]:

	TransactionAmt	ProductCD	card1	card2	card3	card4	card5	card6	\
448539	4.679688	4	6598	111.0	150.0	2	195.0	2	
321311	4.355469	4	12839	321.0	150.0	4	226.0	2	
497320	3.892578	4	14649	548.0	150.0	4	226.0	2	
350951	4.058594	4	6489	295.0	150.0	4	226.0	2	
98132	5.296875	2	5714	170.0	150.0	4	195.0	1	

	addr1	addr2	dist1	P_emaildomain	R_emaildomain	C1	C2	C3	\
448539	264.0	87.0	6.0	0	2	2.0	4.0	0.0	
321311	264.0	87.0	0.0	0	2	1.0	1.0	0.0	
497320	441.0	87.0	86.0	4	2	2.0	4.0	0.0	
350951	184.0	87.0	8.0	0	2	154.0	148.0	0.0	
98132	498.0	87.0	8.0	5	6	1.0	1.0	0.0	

	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	\
448539	0.0	1.0	1.0	0.0	0.0	1.0	0.0	2.0	0.0	7.0	2.0	
321311	0.0	2.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0	12.0	1.0	
497320	0.0	0.0	2.0	0.0	0.0	3.0	0.0	4.0	0.0	45.0	2.0	
350951	0.0	108.0	101.0	0.0	0.0	92.0	0.0	107.0	0.0	504.0	129.0	
98132	1.0	0.0	1.0	0.0	1.0	0.0	1.0	1.0	0.0	1.0	1.0	

	D1	D2	D3	D4	D5	D6	D8	D9	D10	\
448539	64.0	64.0	18.0	63.0	10.0	0.0	37.843750	0.666504	63.0	
321311	114.0	114.0	14.0	115.0	14.0	0.0	37.843750	0.666504	115.0	
497320	85.0	85.0	3.0	329.0	3.0	0.0	37.843750	0.666504	329.0	
350951	577.0	577.0	54.0	54.0	54.0	0.0	37.843750	0.666504	577.0	
98132	0.0	97.0	8.0	26.0	10.0	0.0	7.917969	0.916504	15.0	

	D11	D12	D13	D14	D15	M1	M2	M3	M4	M5	M6	M7	M8	M9	id_01	\
448539	35.0	0.0	0.0	0.0	63.0	1	1	1	3	2	0	0	1	1	-5.0	
321311	45.0	0.0	0.0	0.0	115.0	1	1	1	3	2	1	1	1	1	-5.0	
497320	85.0	0.0	0.0	0.0	85.0	1	1	1	0	0	0	0	0	0	-5.0	
350951	43.0	0.0	0.0	0.0	110.0	2	2	2	3	2	0	2	2	2	-5.0	
98132	43.0	0.0	0.0	0.0	52.0	2	2	2	3	2	2	2	2	2	-5.0	

	id_02	id_03	id_04	id_05	id_06	id_09	id_10	id_11	id_12	\
448539	125673.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	1	
321311	125673.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	1	
497320	125673.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	1	
350951	125673.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	1	

```
98132   105681.0    0.0    0.0    2.0    0.0    0.0    0.0  100.0      2

        id_13  id_14  id_15  id_16  id_17  id_19  id_20  id_28  id_29  id_30  \
448539   52.0 -300.0      1      1  166.0  341.0  472.0      1      1      3
321311   52.0 -300.0      1      1  166.0  341.0  472.0      1      1      3
497320   52.0 -300.0      1      1  166.0  341.0  472.0      1      1      3
350951   52.0 -300.0      1      1  166.0  341.0  472.0      1      1      3
98132    52.0 -360.0      0      0  166.0  621.0  272.0      0      0      4

        id_31  id_32  id_33  id_34  id_35  id_36  id_37  id_38  DeviceType  \
448539    103   24.0     74      4      2      2      2      2           2
321311    103   24.0     74      4      2      2      2      2           2
497320    103   24.0     74      4      2      2      2      2           2
350951    103   24.0     74      4      2      2      2      2           2
98132      84   24.0     72      3      1      0      1      1           0

        DeviceInfo  card2_missing_flag  card3_missing_flag  \
448539         141                   0                   0
321311         141                   0                   0
497320         141                   0                   0
350951         141                   0                   0
98132          140                   0                   0

        card4_missing_flag  card5_missing_flag  card6_missing_flag  \
448539                   0                   0                   0
321311                   0                   0                   0
497320                   0                   0                   0
350951                   0                   0                   0
98132                    0                   0                   0

        addr1_missing_flag  addr2_missing_flag  dist1_missing_flag  \
448539                   0                   0                   0
321311                   0                   0                   0
497320                   0                   0                   0
350951                   0                   0                   1
98132                    0                   0                   1

        dist2_missing_flag  P_emaildomain_missing_flag  \
448539                   1                           0
321311                   1                           0
497320                   1                           0
350951                   1                           0
98132                    1                           0

        R_emaildomain_missing_flag  D1_missing_flag  D2_missing_flag  \
448539                           1                0                0
321311                           1                0                0
```

```
497320                        1               0               0
350951                        1               0               0
98132                         0               0               1

        D3_missing_flag  D4_missing_flag  D5_missing_flag  D6_missing_flag  \
448539                0                0                0                1
321311                0                0                0                1
497320                0                0                0                1
350951                0                0                0                1
98132                 1                1                1                1

        D7_missing_flag  D8_missing_flag  D9_missing_flag  D10_missing_flag  \
448539                1                1                1                 0
321311                1                1                1                 0
497320                1                1                1                 0
350951                1                1                1                 0
98132                 1                0                0                 1

        D11_missing_flag  D12_missing_flag  D13_missing_flag  \
448539                 0                 1                 1
321311                 0                 1                 1
497320                 0                 1                 1
350951                 1                 1                 1
98132                  1                 1                 1

        D14_missing_flag  D15_missing_flag  M1_missing_flag  M2_missing_flag  \
448539                 1                 0                0                0
321311                 1                 0                0                0
497320                 1                 0                0                0
350951                 1                 0                1                1
98132                  1                 1                1                1

        M3_missing_flag  M4_missing_flag  M5_missing_flag  M6_missing_flag  \
448539                0                1                1                0
321311                0                1                1                0
497320                0                0                0                0
350951                1                1                1                0
98132                 1                1                1                1

        M7_missing_flag  M8_missing_flag  M9_missing_flag  V1_missing_flag  \
448539                0                0                0                0
321311                0                0                0                0
497320                0                0                0                0
350951                1                1                1                1
98132                 1                1                1                1

        V2_missing_flag  V3_missing_flag  V4_missing_flag  V5_missing_flag  \
```

```
448539                0                0                0                0
321311                0                0                0                0
497320                0                0                0                0
350951                1                1                1                1
98132                 1                1                1                1

        V6_missing_flag  V7_missing_flag  V8_missing_flag  V9_missing_flag  \
448539                0                0                0                0
321311                0                0                0                0
497320                0                0                0                0
350951                1                1                1                1
98132                 1                1                1                1

        V10_missing_flag  V11_missing_flag  V12_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 1                 1                 0
98132                  1                 1                 1

        V13_missing_flag  V14_missing_flag  V15_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V16_missing_flag  V17_missing_flag  V18_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V19_missing_flag  V20_missing_flag  V21_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V22_missing_flag  V23_missing_flag  V24_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1
```

```
        V25_missing_flag  V26_missing_flag  V27_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V28_missing_flag  V29_missing_flag  V30_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V31_missing_flag  V32_missing_flag  V33_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V34_missing_flag  V35_missing_flag  V36_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V37_missing_flag  V38_missing_flag  V39_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V40_missing_flag  V41_missing_flag  V42_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V43_missing_flag  V44_missing_flag  V45_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
```

```
350951                0                 0                 0
98132                 1                 1                 1

        V46_missing_flag  V47_missing_flag  V48_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V49_missing_flag  V50_missing_flag  V51_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V52_missing_flag  V53_missing_flag  V54_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V55_missing_flag  V56_missing_flag  V57_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V58_missing_flag  V59_missing_flag  V60_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V61_missing_flag  V62_missing_flag  V63_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V64_missing_flag  V65_missing_flag  V66_missing_flag  \
448539                 0                 0                 0
```

```
321311                0                 0                 0
497320                0                 0                 0
350951                0                 0                 0
98132                 1                 1                 1

        V67_missing_flag  V68_missing_flag  V69_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V70_missing_flag  V71_missing_flag  V72_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V73_missing_flag  V74_missing_flag  V75_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V76_missing_flag  V77_missing_flag  V78_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V79_missing_flag  V80_missing_flag  V81_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V82_missing_flag  V83_missing_flag  V84_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1
```

```
        V85_missing_flag  V86_missing_flag  V87_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V88_missing_flag  V89_missing_flag  V90_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V91_missing_flag  V92_missing_flag  V93_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 1                 1

        V94_missing_flag  V95_missing_flag  V96_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  1                 0                 0

        V97_missing_flag  V98_missing_flag  V99_missing_flag  \
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  0                 0                 0

        V100_missing_flag  V101_missing_flag  V102_missing_flag  \
448539                  0                  0                  0
321311                  0                  0                  0
497320                  0                  0                  0
350951                  0                  0                  0
98132                   0                  0                  0

        V103_missing_flag  V104_missing_flag  V105_missing_flag  \
448539                  0                  0                  0
321311                  0                  0                  0
497320                  0                  0                  0
350951                  0                  0                  0
```

```
98132                   0                   0                   0

        V106_missing_flag  V107_missing_flag  V108_missing_flag  \
448539                  0                   0                   0
321311                  0                   0                   0
497320                  0                   0                   0
350951                  0                   0                   0
98132                   0                   0                   0

        V109_missing_flag  V110_missing_flag  V111_missing_flag  \
448539                  0                   0                   0
321311                  0                   0                   0
497320                  0                   0                   0
350951                  0                   0                   0
98132                   0                   0                   0

        V112_missing_flag  V113_missing_flag  V114_missing_flag  \
448539                  0                   0                   0
321311                  0                   0                   0
497320                  0                   0                   0
350951                  0                   0                   0
98132                   0                   0                   0

        V115_missing_flag  V116_missing_flag  V117_missing_flag  \
448539                  0                   0                   0
321311                  0                   0                   0
497320                  0                   0                   0
350951                  0                   0                   0
98132                   0                   0                   0

        V118_missing_flag  V119_missing_flag  V120_missing_flag  \
448539                  0                   0                   0
321311                  0                   0                   0
497320                  0                   0                   0
350951                  0                   0                   0
98132                   0                   0                   0

        V121_missing_flag  V122_missing_flag  V123_missing_flag  \
448539                  0                   0                   0
321311                  0                   0                   0
497320                  0                   0                   0
350951                  0                   0                   0
98132                   0                   0                   0

        V124_missing_flag  V125_missing_flag  V126_missing_flag  \
448539                  0                   0                   0
321311                  0                   0                   0
```

```
497320                 0                  0                  0
350951                 0                  0                  0
98132                  0                  0                  0

        V127_missing_flag  V128_missing_flag  V129_missing_flag  \
448539                  0                  0                  0
321311                  0                  0                  0
497320                  0                  0                  0
350951                  0                  0                  0
98132                   0                  0                  0

        V130_missing_flag  V131_missing_flag  V132_missing_flag  \
448539                  0                  0                  0
321311                  0                  0                  0
497320                  0                  0                  0
350951                  0                  0                  0
98132                   0                  0                  0

        V133_missing_flag  V134_missing_flag  V135_missing_flag  …  \
448539                  0                  0                  0  …
321311                  0                  0                  0  …
497320                  0                  0                  0  …
350951                  0                  0                  0  …
98132                   0                  0                  0  …

        V169_missing_flag  V170_missing_flag  V171_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V172_missing_flag  V173_missing_flag  V174_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V175_missing_flag  V176_missing_flag  V177_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V178_missing_flag  V179_missing_flag  V180_missing_flag  \
```

```
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V181_missing_flag  V182_missing_flag  V183_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V184_missing_flag  V185_missing_flag  V186_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V187_missing_flag  V188_missing_flag  V189_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V190_missing_flag  V191_missing_flag  V192_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V193_missing_flag  V194_missing_flag  V195_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V196_missing_flag  V197_missing_flag  V198_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0
```

```
        V199_missing_flag  V200_missing_flag  V201_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V202_missing_flag  V203_missing_flag  V204_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V205_missing_flag  V206_missing_flag  V207_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V208_missing_flag  V209_missing_flag  V210_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V211_missing_flag  V212_missing_flag  V213_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V214_missing_flag  V215_missing_flag  V216_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V217_missing_flag  V218_missing_flag  V219_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
```

```
350951                 1                  1                  1
98132                  0                  0                  0

        V220_missing_flag  V221_missing_flag  V222_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V223_missing_flag  V224_missing_flag  V225_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V226_missing_flag  V227_missing_flag  V228_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V229_missing_flag  V230_missing_flag  V231_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V232_missing_flag  V233_missing_flag  V234_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V235_missing_flag  V236_missing_flag  V237_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V238_missing_flag  V239_missing_flag  V240_missing_flag  \
448539                  1                  1                  1
```

```
321311                 1                 1                 1
497320                 1                 1                 1
350951                 1                 1                 1
98132                  0                 0                 0

        V241_missing_flag  V242_missing_flag  V243_missing_flag  \
448539                  1                 1                  1
321311                  1                 1                  1
497320                  1                 1                  1
350951                  1                 1                  1
98132                   0                 0                  0

        V244_missing_flag  V245_missing_flag  V246_missing_flag  \
448539                  1                 1                  1
321311                  1                 1                  1
497320                  1                 1                  1
350951                  1                 1                  1
98132                   0                 0                  0

        V247_missing_flag  V248_missing_flag  V249_missing_flag  \
448539                  1                 1                  1
321311                  1                 1                  1
497320                  1                 1                  1
350951                  1                 1                  1
98132                   0                 0                  0

        V250_missing_flag  V251_missing_flag  V252_missing_flag  \
448539                  1                 1                  1
321311                  1                 1                  1
497320                  1                 1                  1
350951                  1                 1                  1
98132                   0                 0                  0

        V253_missing_flag  V254_missing_flag  V255_missing_flag  \
448539                  1                 1                  1
321311                  1                 1                  1
497320                  1                 1                  1
350951                  1                 1                  1
98132                   0                 0                  0

        V256_missing_flag  V257_missing_flag  V258_missing_flag  \
448539                  1                 1                  1
321311                  1                 1                  1
497320                  1                 1                  1
350951                  1                 1                  1
98132                   0                 0                  0
```

```
        V259_missing_flag  V260_missing_flag  V261_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V262_missing_flag  V263_missing_flag  V264_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V265_missing_flag  V266_missing_flag  V267_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V268_missing_flag  V269_missing_flag  V270_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V271_missing_flag  V272_missing_flag  V273_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V274_missing_flag  V275_missing_flag  V276_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V277_missing_flag  V278_missing_flag  V279_missing_flag  \
448539                  1                  1                  0
321311                  1                  1                  0
497320                  1                  1                  0
350951                  1                  1                  0
```

```
98132                   0                   0                   0

        V280_missing_flag  V281_missing_flag  V282_missing_flag  \
448539                  0                   0                   0
321311                  0                   0                   0
497320                  0                   0                   0
350951                  0                   0                   0
98132                   0                   0                   0

        V283_missing_flag  V284_missing_flag  V285_missing_flag  \
448539                  0                   0                   0
321311                  0                   0                   0
497320                  0                   0                   0
350951                  0                   0                   0
98132                   0                   0                   0

        V286_missing_flag  V287_missing_flag  V288_missing_flag  \
448539                  0                   0                   0
321311                  0                   0                   0
497320                  0                   0                   0
350951                  0                   0                   0
98132                   0                   0                   0

        V289_missing_flag  V290_missing_flag  V291_missing_flag  \
448539                  0                   0                   0
321311                  0                   0                   0
497320                  0                   0                   0
350951                  0                   0                   0
98132                   0                   0                   0

        V292_missing_flag  V293_missing_flag  V294_missing_flag  \
448539                  0                   0                   0
321311                  0                   0                   0
497320                  0                   0                   0
350951                  0                   0                   0
98132                   0                   0                   0

        V295_missing_flag  V296_missing_flag  V297_missing_flag  \
448539                  0                   0                   0
321311                  0                   0                   0
497320                  0                   0                   0
350951                  0                   0                   0
98132                   0                   0                   0

        V298_missing_flag  V299_missing_flag  V300_missing_flag  \
448539                  0                   0                   0
321311                  0                   0                   0
```

```
497320                 0                 0                 0
350951                 0                 0                 0
98132                  0                 0                 0

        V301_missing_flag  V302_missing_flag  V303_missing_flag  \
448539                  0                 0                 0
321311                  0                 0                 0
497320                  0                 0                 0
350951                  0                 0                 0
98132                   0                 0                 0

        V304_missing_flag  V305_missing_flag  V306_missing_flag  \
448539                  0                 0                 0
321311                  0                 0                 0
497320                  0                 0                 0
350951                  0                 0                 0
98132                   0                 0                 0

        V307_missing_flag  V308_missing_flag  V309_missing_flag  \
448539                  0                 0                 0
321311                  0                 0                 0
497320                  0                 0                 0
350951                  0                 0                 0
98132                   0                 0                 0

        V310_missing_flag  V311_missing_flag  V312_missing_flag  \
448539                  0                 0                 0
321311                  0                 0                 0
497320                  0                 0                 0
350951                  0                 0                 0
98132                   0                 0                 0

        V313_missing_flag  V314_missing_flag  V315_missing_flag  \
448539                  0                 0                 0
321311                  0                 0                 0
497320                  0                 0                 0
350951                  0                 0                 0
98132                   0                 0                 0

        V316_missing_flag  V317_missing_flag  V318_missing_flag  \
448539                  0                 0                 0
321311                  0                 0                 0
497320                  0                 0                 0
350951                  0                 0                 0
98132                   0                 0                 0

        V319_missing_flag  V320_missing_flag  V321_missing_flag  \
```

```
448539                 0                 0                 0
321311                 0                 0                 0
497320                 0                 0                 0
350951                 0                 0                 0
98132                  0                 0                 0

        V322_missing_flag  V323_missing_flag  V324_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V325_missing_flag  V326_missing_flag  V327_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V328_missing_flag  V329_missing_flag  V330_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V331_missing_flag  V332_missing_flag  V333_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V334_missing_flag  V335_missing_flag  V336_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0

        V337_missing_flag  V338_missing_flag  V339_missing_flag  \
448539                  1                  1                  1
321311                  1                  1                  1
497320                  1                  1                  1
350951                  1                  1                  1
98132                   0                  0                  0
```

```
        id_01_missing_flag  id_02_missing_flag  id_03_missing_flag  \
448539                   1                   1                   1
321311                   1                   1                   1
497320                   1                   1                   1
350951                   1                   1                   1
98132                    0                   0                   0

        id_04_missing_flag  id_05_missing_flag  id_06_missing_flag  \
448539                   1                   1                   1
321311                   1                   1                   1
497320                   1                   1                   1
350951                   1                   1                   1
98132                    0                   0                   0

        id_07_missing_flag  id_08_missing_flag  id_09_missing_flag  \
448539                   1                   1                   1
321311                   1                   1                   1
497320                   1                   1                   1
350951                   1                   1                   1
98132                    1                   1                   0

        id_10_missing_flag  id_11_missing_flag  id_12_missing_flag  \
448539                   1                   1                   1
321311                   1                   1                   1
497320                   1                   1                   1
350951                   1                   1                   1
98132                    0                   0                   0

        id_13_missing_flag  id_14_missing_flag  id_15_missing_flag  \
448539                   1                   1                   1
321311                   1                   1                   1
497320                   1                   1                   1
350951                   1                   1                   1
98132                    0                   0                   0

        id_16_missing_flag  id_17_missing_flag  id_18_missing_flag  \
448539                   1                   1                   1
321311                   1                   1                   1
497320                   1                   1                   1
350951                   1                   1                   1
98132                    0                   0                   1

        id_19_missing_flag  id_20_missing_flag  id_21_missing_flag  \
448539                   1                   1                   1
321311                   1                   1                   1
497320                   1                   1                   1
```

```
350951                   1                   1                   1
98132                    0                   0                   1

        id_22_missing_flag  id_23_missing_flag  id_24_missing_flag  \
448539                   1                   1                   1
321311                   1                   1                   1
497320                   1                   1                   1
350951                   1                   1                   1
98132                    1                   1                   1

        id_25_missing_flag  id_26_missing_flag  id_27_missing_flag  \
448539                   1                   1                   1
321311                   1                   1                   1
497320                   1                   1                   1
350951                   1                   1                   1
98132                    1                   1                   1

        id_28_missing_flag  id_29_missing_flag  id_30_missing_flag  \
448539                   1                   1                   1
321311                   1                   1                   1
497320                   1                   1                   1
350951                   1                   1                   1
98132                    0                   0                   0

        id_31_missing_flag  id_32_missing_flag  id_33_missing_flag  \
448539                   1                   1                   1
321311                   1                   1                   1
497320                   1                   1                   1
350951                   1                   1                   1
98132                    0                   0                   0

        id_34_missing_flag  id_35_missing_flag  id_36_missing_flag  \
448539                   1                   1                   1
321311                   1                   1                   1
497320                   1                   1                   1
350951                   1                   1                   1
98132                    0                   0                   0

        id_37_missing_flag  id_38_missing_flag  DeviceType_missing_flag  \
448539                   1                   1                        1
321311                   1                   1                        1
497320                   1                   1                        1
350951                   1                   1                        1
98132                    0                   0                        0

        DeviceInfo_missing_flag  _Weekdays  _Hours  _Days  Trans_min_mean  \
448539                        1          3       3     12       -27.09375
```

```
321311                        1          5      14      3       -57.09375
497320                        1          0      19     30       -86.00000
350951                        1          6       1     11       -77.06250
98132                         0          5      22     23        65.00000

        Trans_min_std  TransactionAmt_to_mean_card1  \
448539      -0.113281                      0.461914
321311      -0.238770                      0.645508
497320      -0.359619                      0.470703
350951      -0.322266                      0.555176
98132        0.271729                      1.193359

        TransactionAmt_to_mean_card4  TransactionAmt_to_std_card1  \
448539                      0.815430                     0.313331
321311                      0.585449                     0.401900
497320                      0.367920                     0.326992
350951                      0.435059                     0.570935
98132                       1.501953                     0.718662

        TransactionAmt_to_std_card4   PCA_V_0   PCA_V_1   PCA_V_2   PCA_V_3  \
448539                     0.425049 -0.782227  0.368896  0.263428  0.007843
321311                     0.341553 -0.782227  0.368896  0.263428  0.007904
497320                     0.214722 -0.801758  0.317871  0.270020 -0.027130
350951                     0.253906 -0.304443 -0.825684 -0.107727 -0.468994
98132                      0.876465  2.806641  0.345703  0.377197 -0.017654

         PCA_V_4   PCA_V_5   PCA_V_6   PCA_V_7   PCA_V_8   PCA_V_9  PCA_V_10  \
448539  0.013977 -0.007053  0.002455  0.334717 -0.010490 -0.220459 -0.042755
321311  0.013786 -0.007095  0.002304  0.334961 -0.010521 -0.220825 -0.042938
497320  0.058411 -0.003780 -0.014931 -0.225342  0.021576  0.033905 -0.024246
350951 -0.162598  0.029373  0.138550  0.065613 -0.067749  0.260010  0.052063
98132  -0.042450 -0.244263  0.023148  0.011810  0.000937  0.023010 -0.009560

        PCA_V_11  PCA_V_12  PCA_V_13  PCA_V_14  PCA_V_15  PCA_V_16  PCA_V_17  \
448539  0.021561 -0.005577 -0.020416  0.005253 -0.000285  0.004505 -0.006889
321311  0.020325 -0.003403 -0.022324  0.005749  0.000531  0.004292 -0.009888
497320  0.010124  0.020370  0.017105 -0.014145 -0.013794 -0.005302 -0.077637
350951 -0.091125  0.068176  0.014664  0.010857 -0.011421 -0.020767  0.009926
98132  -0.022293 -0.021683 -0.013680 -0.005581 -0.009674  0.001465 -0.008263

        PCA_V_18  PCA_V_19  PCA_V_20  PCA_V_21  PCA_V_22  PCA_V_23  PCA_V_24  \
448539  0.008362 -0.011055 -0.006924 -0.033722  0.011093 -0.011307 -0.007355
321311  0.009216 -0.005642 -0.005966 -0.034119  0.012009 -0.009644 -0.006516
497320  0.056580  0.115479  0.013458 -0.012100  0.003590  0.023392  0.003922
350951  0.008850  0.005054 -0.041290  0.041382 -0.051819  0.042450  0.026703
98132  -0.008881  0.005920 -0.004478 -0.000713 -0.003057  0.008438 -0.017868
```

```
        PCA_V_25  PCA_V_26  PCA_V_27  PCA_V_28  PCA_V_29
448539 -0.004627  0.002234 -0.005222  0.004154 -0.001096
321311 -0.002327  0.000332 -0.006203  0.002287 -0.005352
497320  0.032867  0.010475  0.007317  0.008362 -0.037079
350951 -0.046417 -0.029648  0.022217  0.004562 -0.008606
98132   0.000072  0.013344 -0.005264 -0.000665  0.007454

[5 rows x 533 columns]
```

[118]:
```
X_train.isnull().sum()
```

[118]:
```
TransactionAmt    0
ProductCD         0
card1             0
card2             0
card3             0
                 ..
PCA_V_25          0
PCA_V_26          0
PCA_V_27          0
PCA_V_28          0
PCA_V_29          0
Length: 533, dtype: int64
```

[120]:
```
%%time
rfc = RandomForestClassifier(random_state=0, n_jobs = -1)
rfc.fit(X_train, y_train)
rfc
```

```
Wall time: 2min 45s
```

[120]:
```
RandomForestClassifier(n_jobs=-1, random_state=0)
```

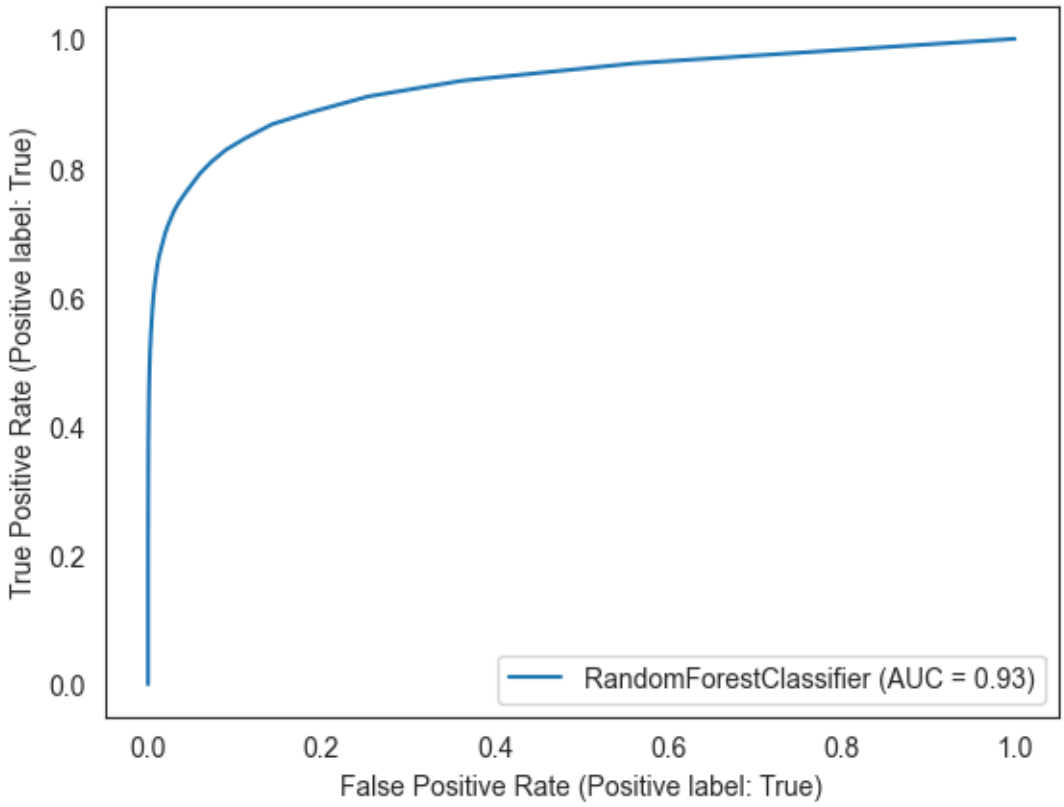
[121]:
```
y_pred_rfc = rfc.predict(X_test)
y_prob_pred_rfc = rfc.predict_proba(X_test)[:, 1]

print("Y predicted : ",y_pred_rfc)
print("Y probability predicted : ",y_prob_pred_rfc[:5])
```

```
Y predicted :  [False False False … False False False]
Y probability predicted :  [0.02 0.   0.   0.06 0.02]
```

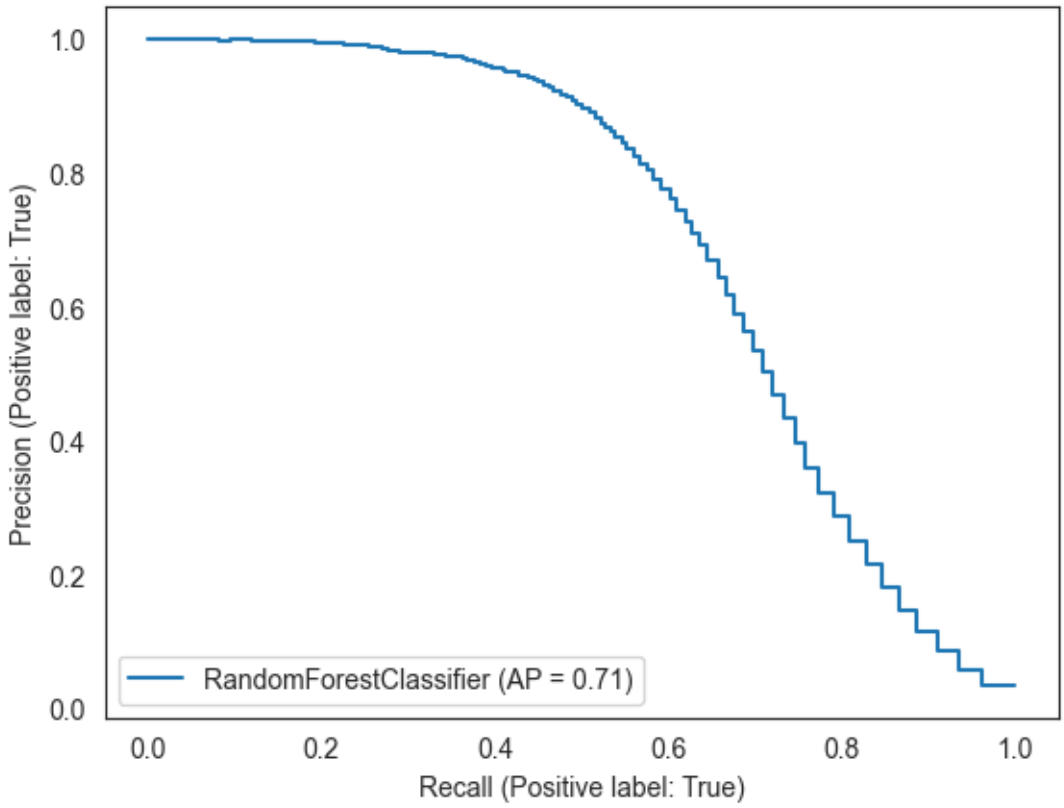
[123]:
```
compute_evaluation_metric(rfc, X_test, y_test, y_pred_rfc, y_prob_pred_rfc)
```

```
Accuracy Score :  0.9792901412266739

AUC Score :  0.9295181538168427
```

```
Confusion Matrix :
 [[170811    152]
 [  3517   2682]]

Classification Report :
              precision    recall  f1-score   support

       False       0.98      1.00      0.99    170963
        True       0.95      0.43      0.59      6199

    accuracy                           0.98    177162
   macro avg       0.96      0.72      0.79    177162
weighted avg       0.98      0.98      0.98    177162


Concordance Index :  0.914790096309497

ROC curve :
```



```
PR curve :
```

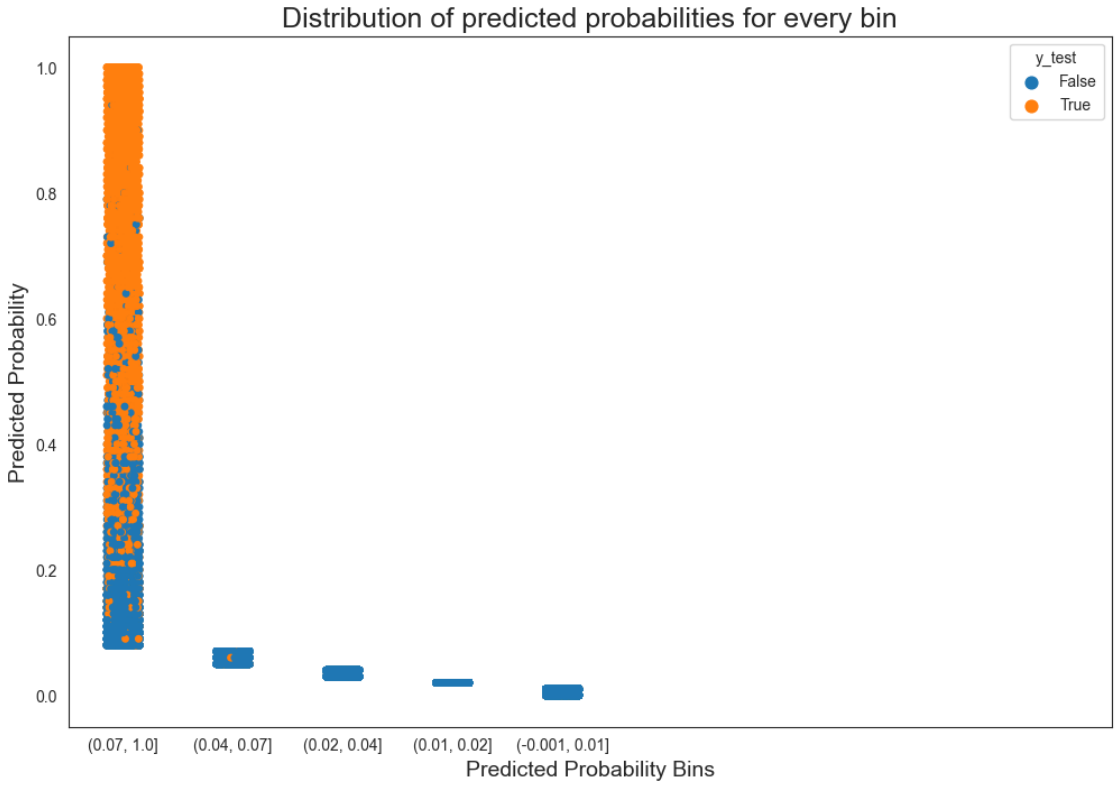


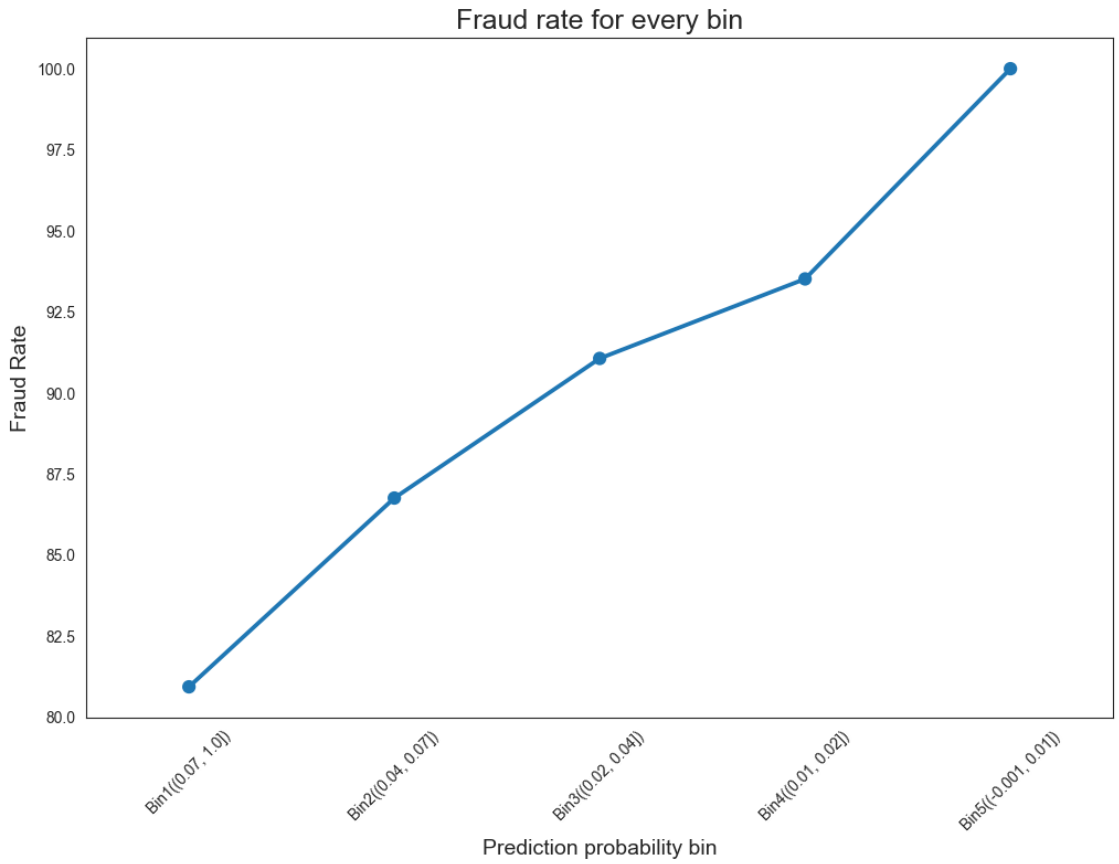

```
Additional Metrics:
TPR (Recall) : 0.4327
FPR : 0.0009
TNR (Specificity) : 0.9991
FNR : 0.5673
```

[124]:
```
concordance(y_test.values, y_prob_pred_rfc)
```

[124]:
```
0.914790096309497
```

[125]:
```
captures(y_test, y_pred_rfc, y_prob_pred_rfc)
```


Distribution of predicted probabilities for every bin
y_test
False
True
Predicted Probability
1.0
0.8
0.6
0.4
0.2
0.0
(0.07, 1.0]
(0.04, 0.07]
(0.02, 0.04]
(0.01, 0.02]
(-0.001, 0.01]
Predicted Probability Bins

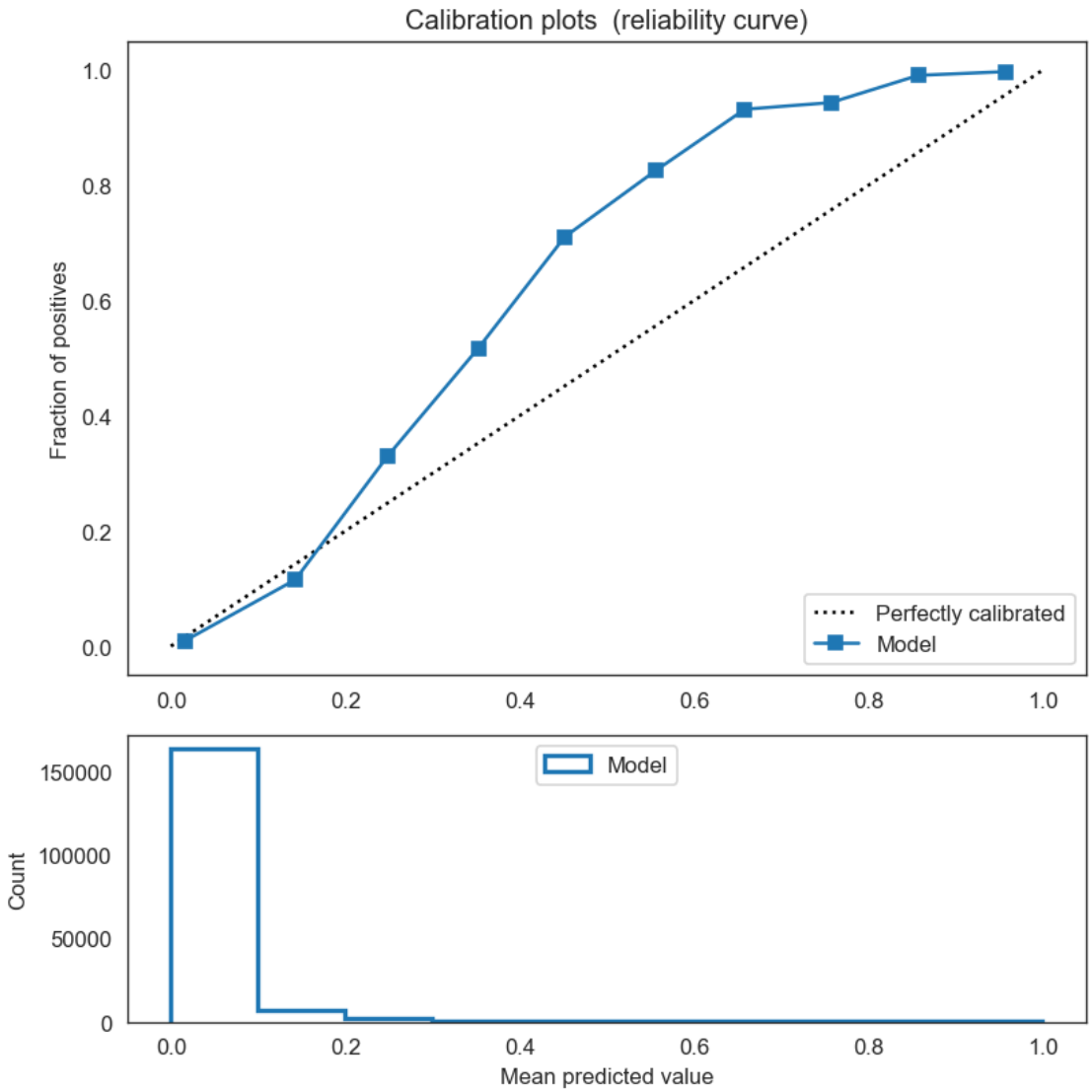


[125]:

	prob_bin	not_fraud	fraud	perc_fraud	perc_not_fraud	\
0	Bin1((0.07, 1.0])	12482	5017	0.809324	0.073010	
1	Bin2((0.04, 0.07])	11939	361	0.058235	0.069834	
2	Bin3((0.02, 0.04])	18984	267	0.043071	0.111042	
3	Bin4((0.01, 0.02])	18595	152	0.024520	0.108766	
4	Bin5((-0.001, 0.01])	108963	402	0.064849	0.637348	

	cum_perc_fraud	cum_perc_not_fraud
0	80.932408	7.300995
1	86.755928	14.284377
2	91.063075	25.388534
3	93.515083	36.265157
4	100.000000	100.000000

[126]:
```
draw_calibration_curve(y_test, y_prob_pred_rfc, n_bins=10)
```



16 15. Handling Class Imbalance

16.0.1 Handle Class Imbalance with Random Oversampler

[130]:
```
from   imblearn.over_sampling import RandomOverSampler
ros = RandomOverSampler()
X_train_ros, y_train_ros = ros.fit_resample(X_train, y_train)
y_train_ros.value_counts()
```

[130]:
```
False    398914
True     398914
Name: isFraud, dtype: int64
```

[131]:
```
%%time
lgbc_ros = LGBMClassifier(random_state=0)
lgbc_ros.fit(X_train_ros,y_train_ros)
lgbc_ros
```

```
Wall time: 56.9 s
```

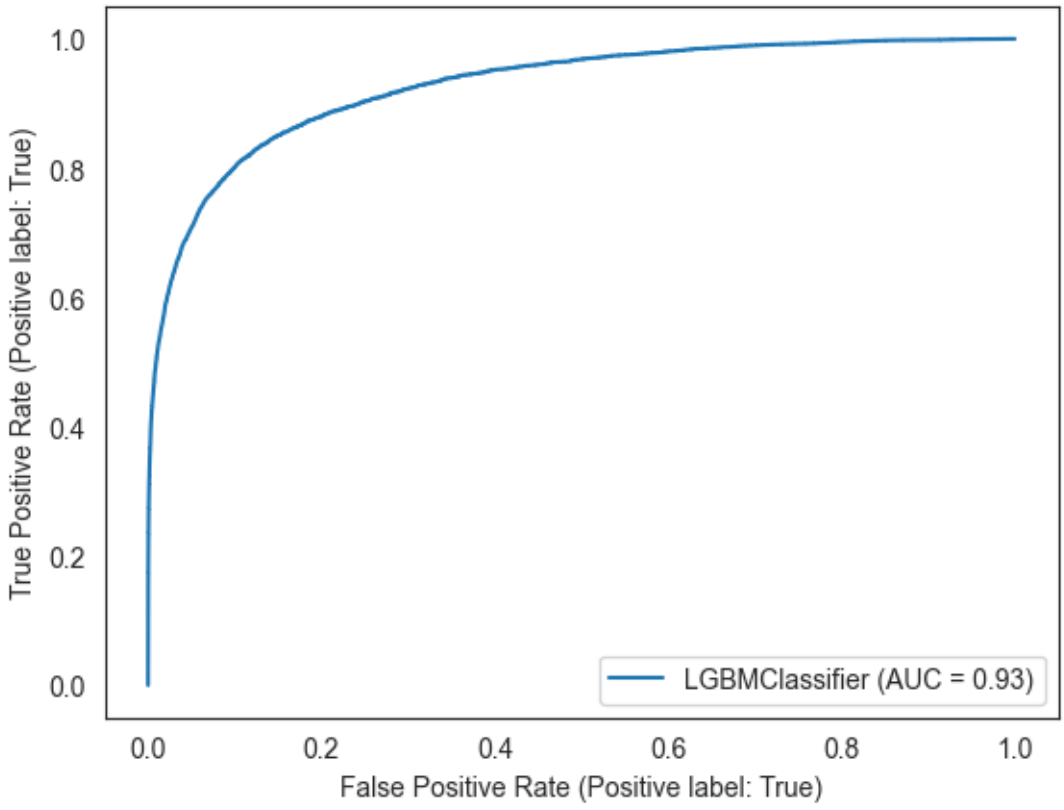
[131]:
```
LGBMClassifier(random_state=0)
```

[133]:
```
y_pred_lgbcros = lgbc_ros.predict(X_test)
y_prob_pred_lgbcros = lgbc_ros.predict_proba(X_test)[:, 1]

print("Y predicted : ",y_pred_lgbcros)
print("Y probability predicted : ",y_prob_pred_lgbcros[:5])
```

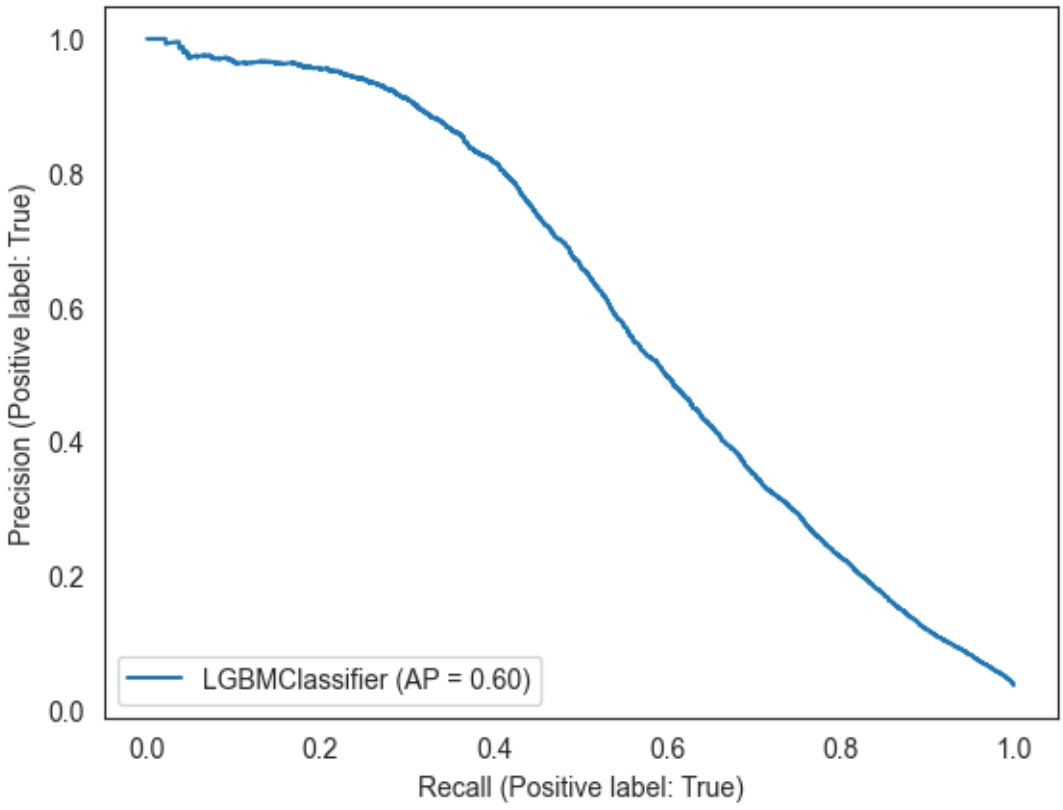
```
Y predicted :  [False False False … False False False]
Y probability predicted :  [0.23093612 0.12247041 0.13222139 0.56481875
0.10460942]
```

[135]:
```
compute_evaluation_metric(lgbc_ros, X_test, y_test, y_pred_lgbcros,␣
↪y_prob_pred_lgbcros)
```

```
Accuracy Score :  0.886696921461713

AUC Score :  0.9271778128567145

Confusion Matrix :
 [[152037  18926]
 [  1147   5052]]

Classification Report :
              precision    recall  f1-score   support

       False       0.99      0.89      0.94    170963
        True       0.21      0.81      0.33      6199

    accuracy                           0.89    177162
   macro avg       0.60      0.85      0.64    177162
weighted avg       0.97      0.89      0.92    177162


Concordance Index :  0.9271778029491814

ROC curve :
```

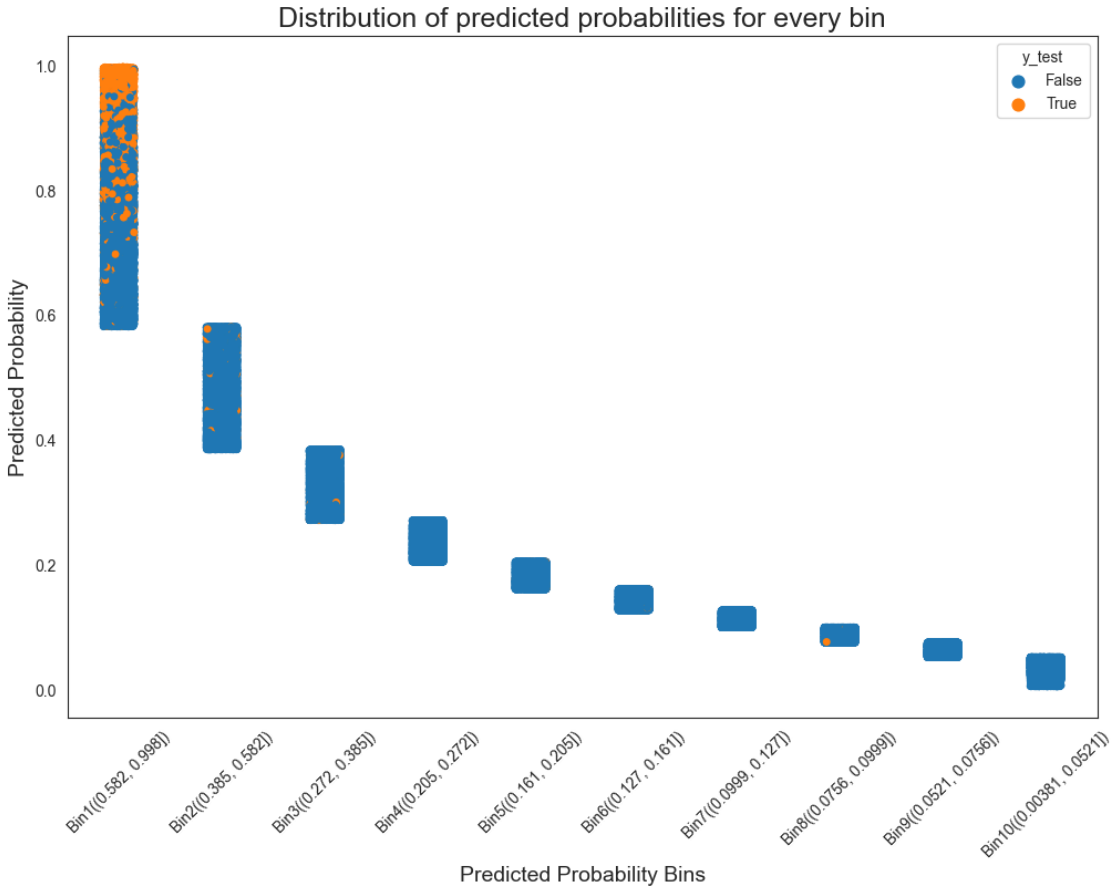

1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate (Positive label: True)
LGBMClassifier (AUC = 0.93)
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate (Positive label: True)

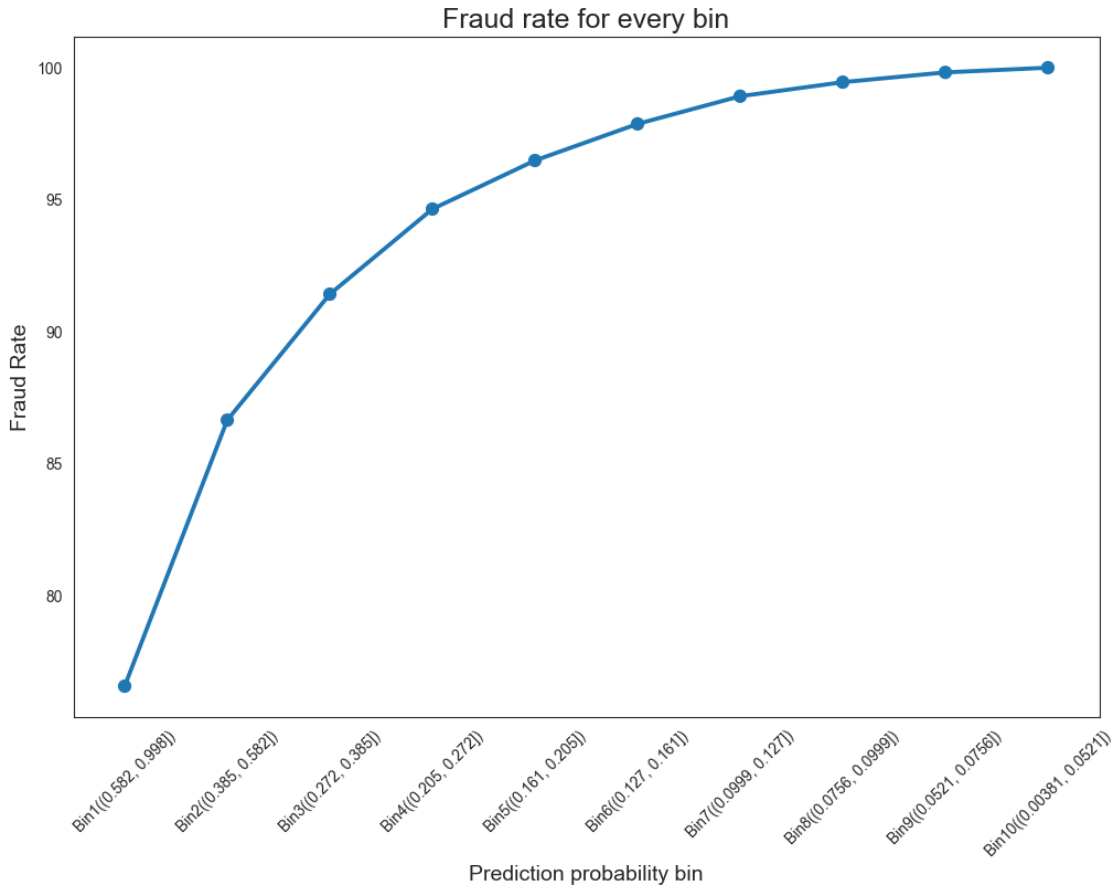
PR curve :




```
Additional Metrics:
TPR (Recall) : 0.8150
FPR : 0.1107
TNR (Specificity) : 0.8893
FNR : 0.1850
```

[136]:
```
captures(y_test, y_pred_lgbcros, y_prob_pred_lgbcros)
```


Distribution of predicted probabilities for every bin
y_test
False
True
Predicted Probability
1.0
0.8
0.6
0.4
0.2
0.0
Bin1((0.582, 0.998])
Bin2((0.385, 0.582])
Bin3((0.272, 0.385])
Bin4((0.205, 0.272])
Bin5((0.161, 0.205])
Bin6((0.127, 0.161])
Bin7((0.0999, 0.127])
Bin8((0.0756, 0.0999])
Bin9((0.0521, 0.0756])
Bin10((0.00381, 0.0521])
Predicted Probability Bins



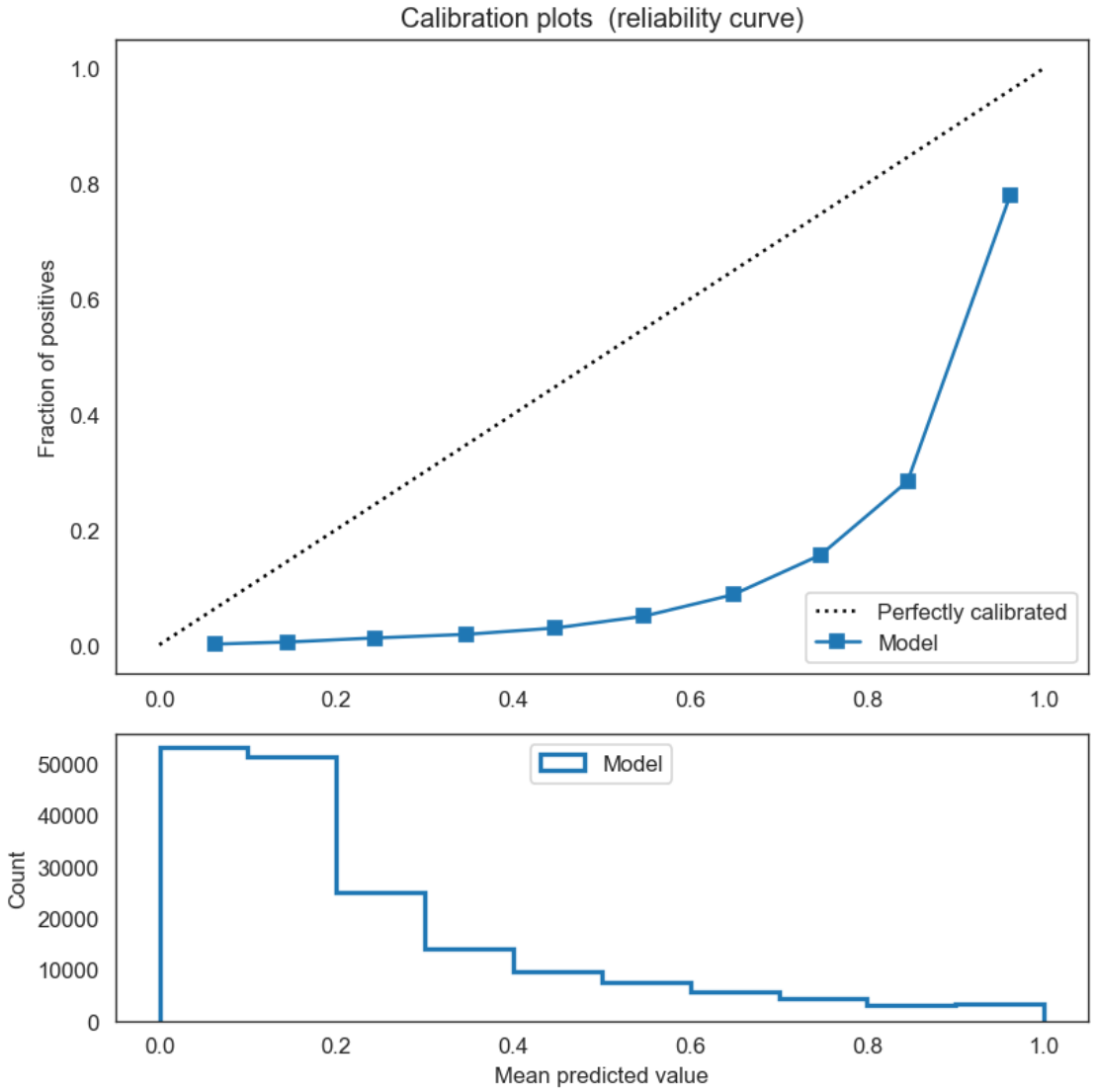
[136]:

	prob_bin	not_fraud	fraud	perc_fraud	perc_not_fraud	\
0	Bin1((0.582, 0.998])	12971	4746	0.765607	0.075870	
1	Bin2((0.385, 0.582])	17091	625	0.100823	0.099969	
2	Bin3((0.272, 0.385])	17420	296	0.047750	0.101893	
3	Bin4((0.205, 0.272])	17516	200	0.032263	0.102455	
4	Bin5((0.161, 0.205])	17602	114	0.018390	0.102958	
5	Bin6((0.127, 0.161])	17630	86	0.013873	0.103122	
6	Bin7((0.0999, 0.127])	17651	65	0.010486	0.103245	
7	Bin8((0.0756, 0.0999])	17683	33	0.005323	0.103432	
8	Bin9((0.0521, 0.0756])	17693	23	0.003710	0.103490	
9	Bin10((0.00381, 0.0521])	17706	11	0.001774	0.103566	

	cum_perc_fraud	cum_perc_not_fraud
0	76.560736	7.587022
1	86.643007	17.583922
2	91.417971	27.773261
3	94.644297	38.018753
4	96.483304	48.314548
5	97.870624	58.626720

6	98.919181	68.951177
7	99.451524	79.294350
8	99.822552	89.643373
9	100.000000	100.000000

[137]:
```
draw_calibration_curve(y_test, y_prob_pred_lgbcros, n_bins=10)
```



- After balancing the class, accuracy score is 0.88 and AUC score is 92.5%
- Accuracy has decreased as compared to the previos model, but AUC has improved
- Additionally the recall has improved significantly at the cost of precision.

17 16. Cost Sensitive Learning with Class weights

[139]:
```
%%time
lgbc_bal = LGBMClassifier(random_state=0, class_weight='balanced')
lgbc_bal.fit(X_train, y_train)
lgbc_bal
```

```
Wall time: 32 s
```

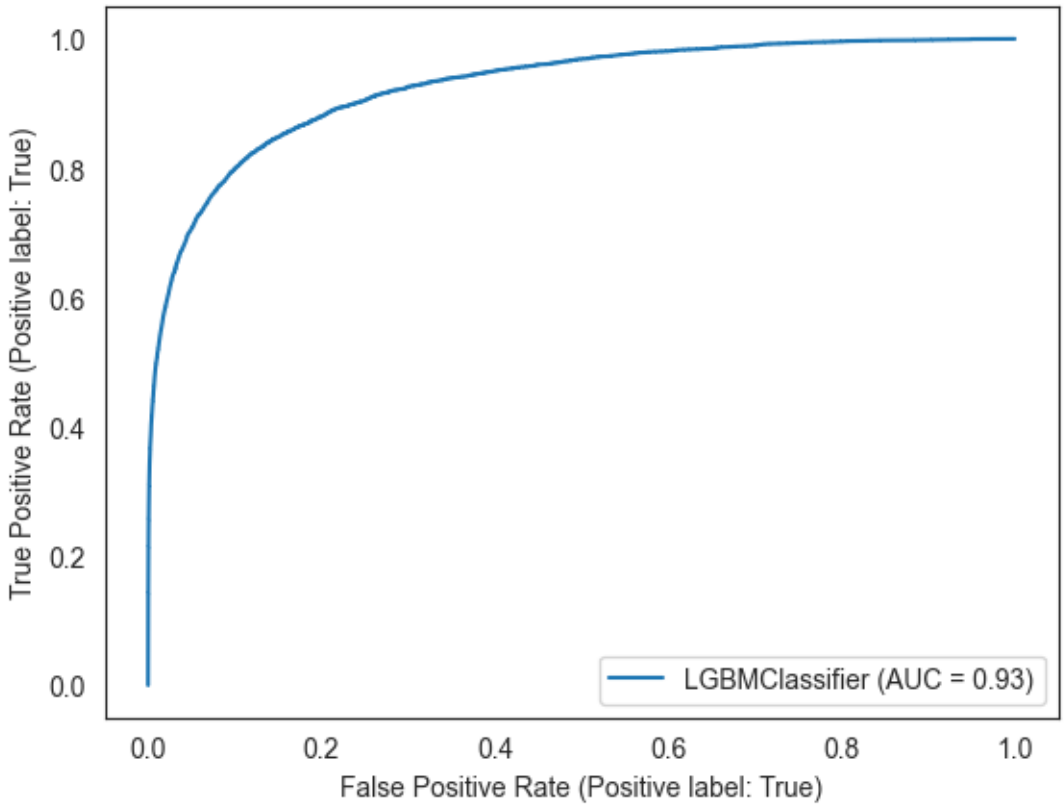
[139]:
```
LGBMClassifier(class_weight='balanced', random_state=0)
```

[140]:
```
y_pred_lgbcbal = lgbc_bal.predict(X_test)
y_prob_pred_lgbcbal = lgbc_bal.predict_proba(X_test)[:, 1]

print("Y predicted : ",y_pred_lgbcbal)
print("Y probability predicted : ",y_prob_pred_lgbcbal[:5])
```

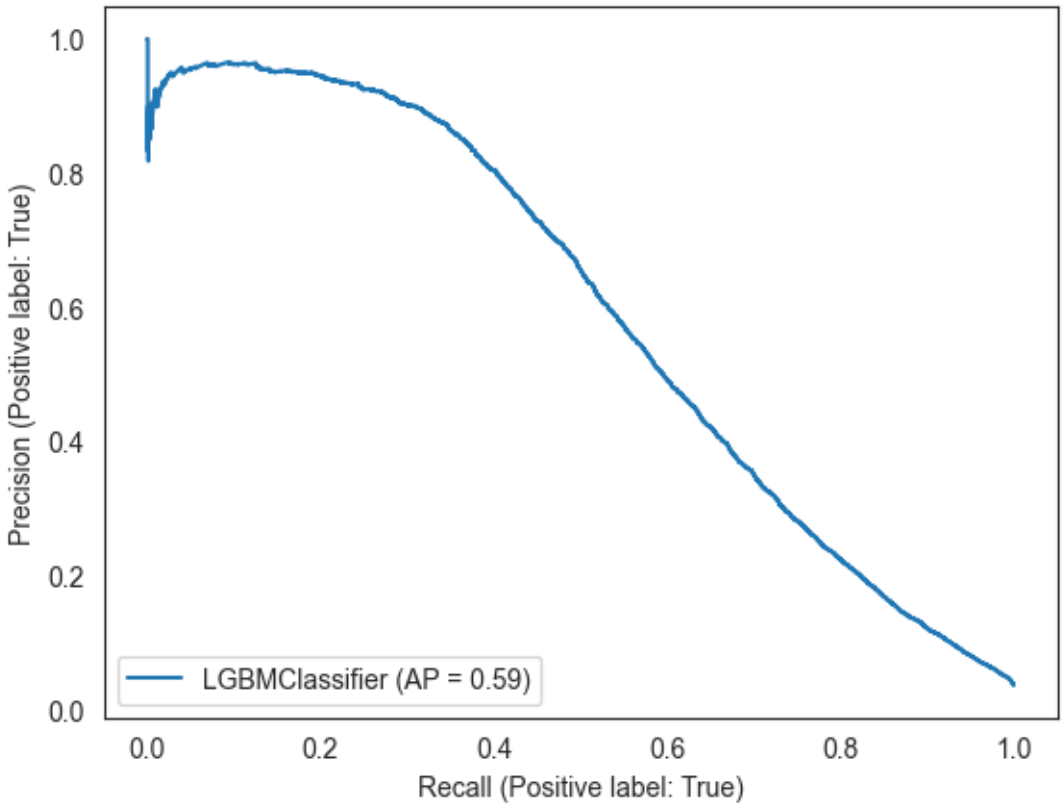
```
Y predicted :  [False False False … False False False]
Y probability predicted :  [0.22700362 0.15390235 0.13087413 0.52349433
0.14032882]
```

[141]:
```
compute_evaluation_metric(lgbc_bal, X_test, y_test, y_pred_lgbcbal,␣
 ↪y_prob_pred_lgbcbal)
```

```
Accuracy Score :  0.8862284237025999

AUC Score :  0.927124006459723

Confusion Matrix :
 [[151969  18994]
 [  1162   5037]]

Classification Report :
              precision    recall  f1-score   support

       False       0.99      0.89      0.94    170963
        True       0.21      0.81      0.33      6199

    accuracy                           0.89    177162
   macro avg       0.60      0.85      0.64    177162
weighted avg       0.97      0.89      0.92    177162


Concordance Index :  0.9271239946650406

ROC curve :
```

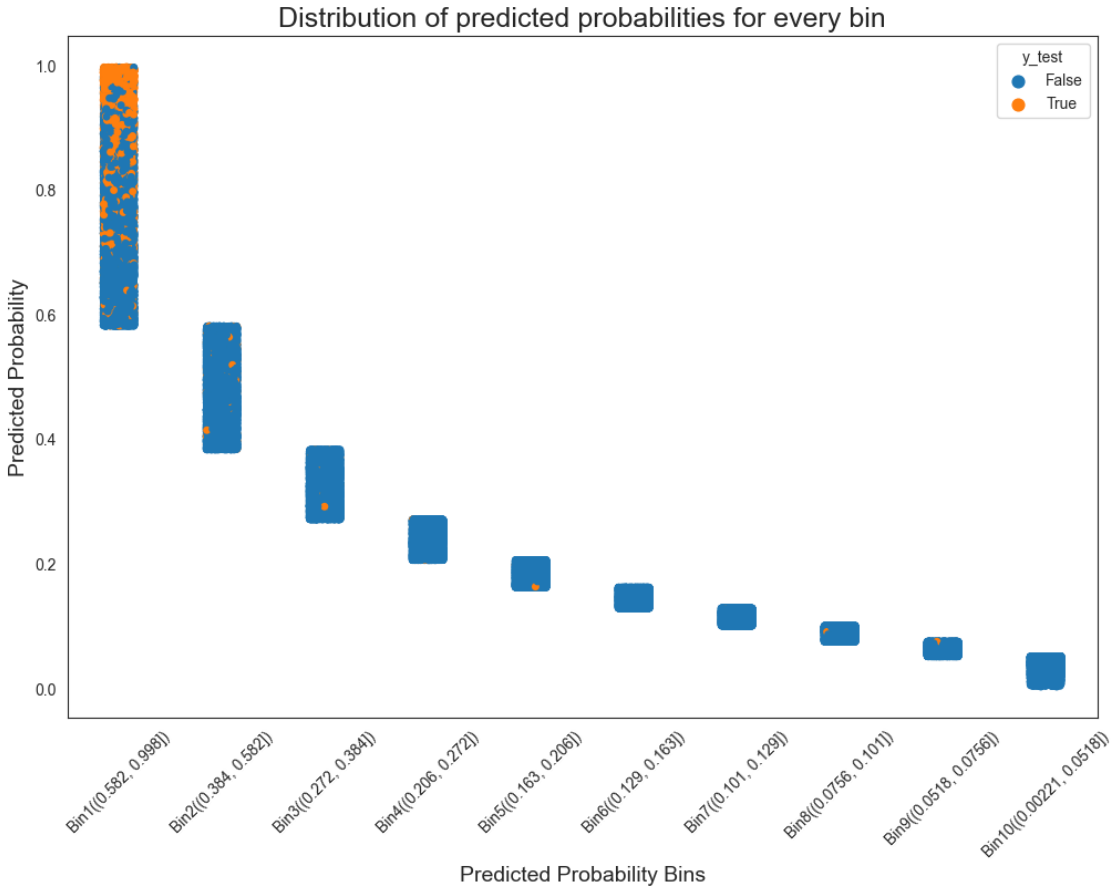

1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate (Positive label: True)
LGBMClassifier (AUC = 0.93)
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate (Positive label: True)

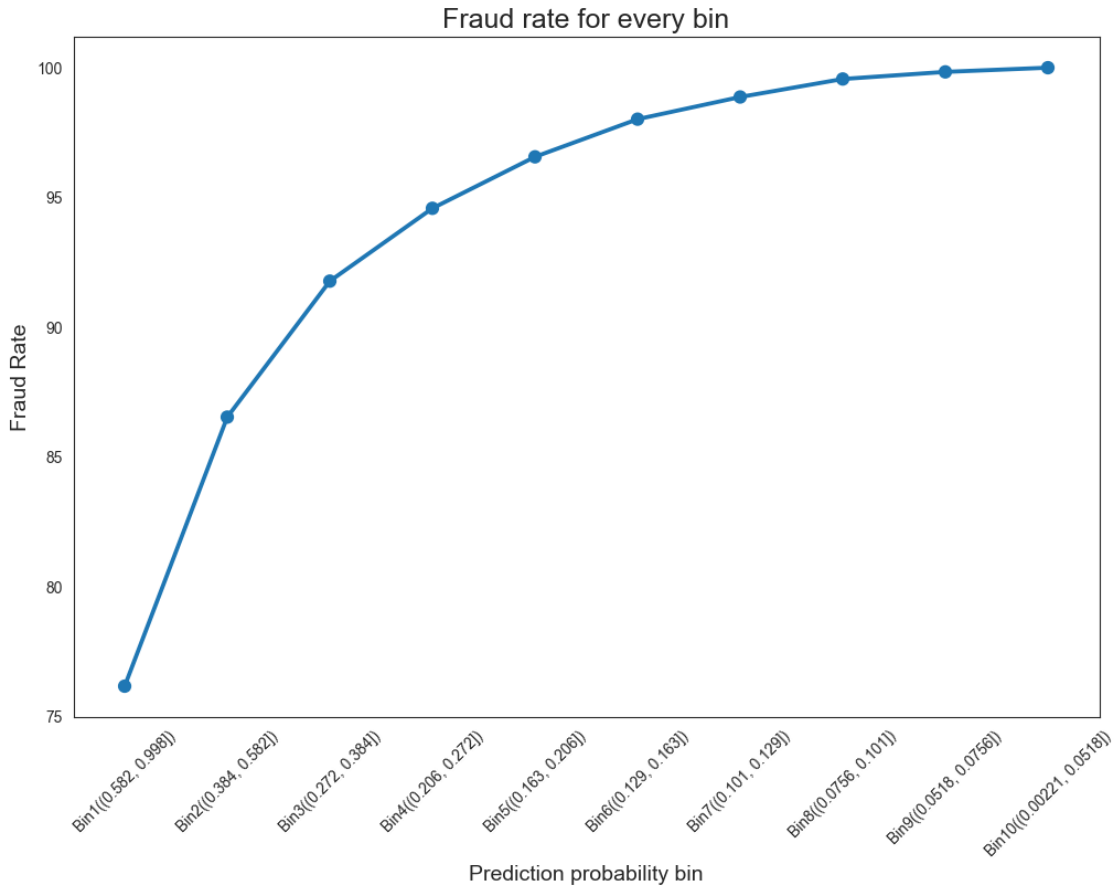
PR curve :




```
Additional Metrics:
TPR (Recall) : 0.8126
FPR : 0.1111
TNR (Specificity) : 0.8889
FNR : 0.1874
```

[142]:
```
captures(y_test, y_pred_lgbcbal, y_prob_pred_lgbcbal)
```


Distribution of predicted probabilities for every bin
y_test
False
True
Predicted Probability
1.0
0.8
0.6
0.4
0.2
0.0
Bin1((0.582, 0.998])
Bin2((0.384, 0.582])
Bin3((0.272, 0.384])
Bin4((0.206, 0.272])
Bin5((0.163, 0.206])
Bin6((0.129, 0.163])
Bin7((0.101, 0.129])
Bin8((0.0756, 0.101])
Bin9((0.0518, 0.0756])
Bin10((0.00221, 0.0518])
Predicted Probability Bins



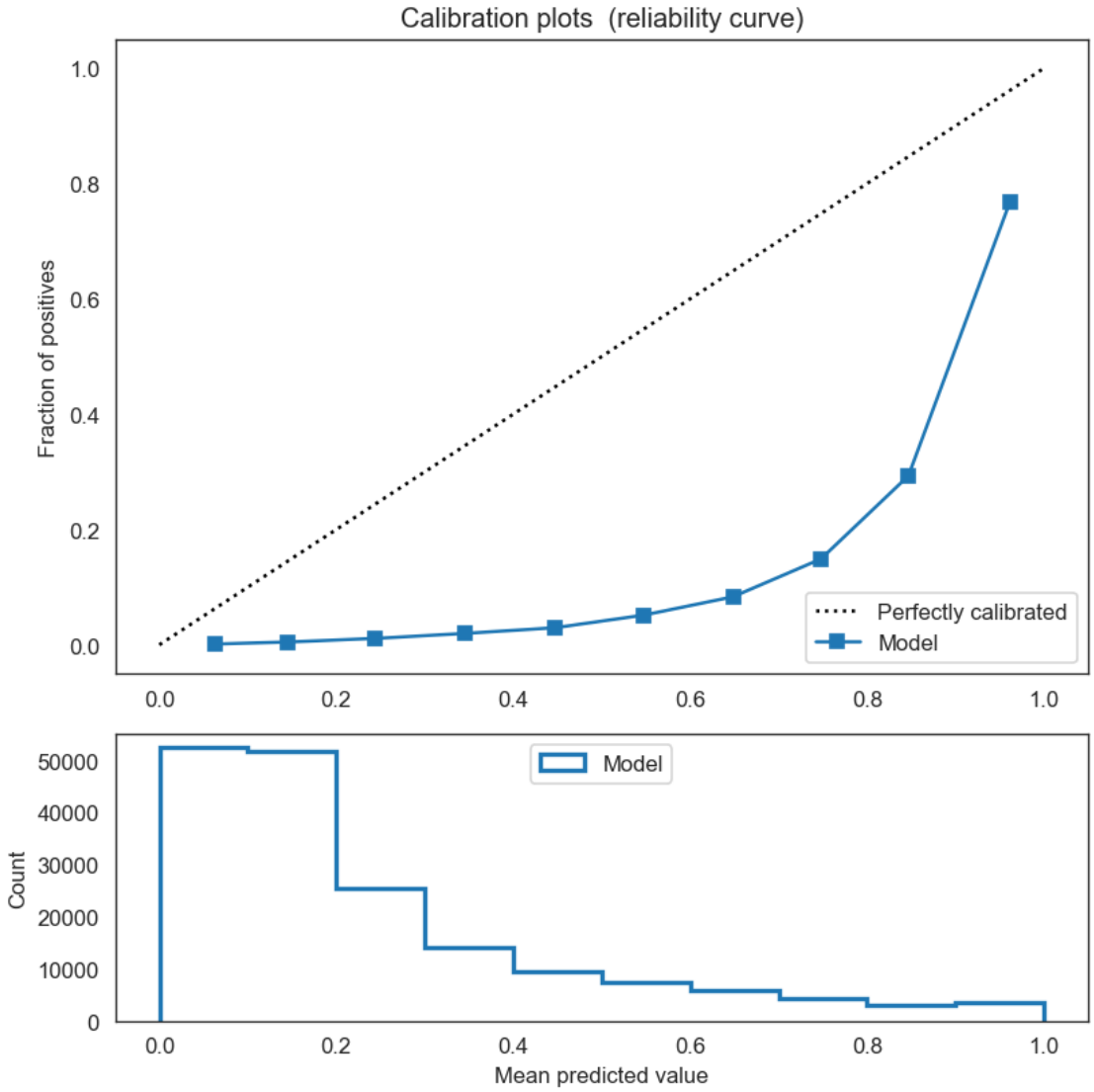
[142]:

	prob_bin	not_fraud	fraud	perc_fraud	perc_not_fraud	\
0	Bin1((0.582, 0.998])	12996	4721	0.761574	0.076016	
1	Bin2((0.384, 0.582])	17073	643	0.103726	0.099864	
2	Bin3((0.272, 0.384])	17391	325	0.052428	0.101724	
3	Bin4((0.206, 0.272])	17542	174	0.028069	0.102607	
4	Bin5((0.163, 0.206])	17593	123	0.019842	0.102905	
5	Bin6((0.129, 0.163])	17626	90	0.014518	0.103098	
6	Bin7((0.101, 0.129])	17663	53	0.008550	0.103315	
7	Bin8((0.0756, 0.101])	17673	43	0.006937	0.103373	
8	Bin9((0.0518, 0.0756])	17699	17	0.002742	0.103525	
9	Bin10((0.00221, 0.0518])	17707	10	0.001613	0.103572	

	cum_perc_fraud	cum_perc_not_fraud
0	76.157445	7.601645
1	86.530085	17.588016
2	91.772867	27.760393
3	94.579771	38.021092
4	96.563962	48.311623
5	98.015809	58.621456

6	98.870786	68.952931
7	99.564446	79.290256
8	99.838684	89.642788
9	100.000000	100.000000

[143]:
```
draw_calibration_curve(y_test, y_prob_pred_lgbcbal, n_bins=10)
```



18 17. Model Calibration

[144]:
```
from sklearn.calibration import CalibratedClassifierCV
```

[145]:
```
lgbc_bal = LGBMClassifier(random_state=0)
calibrated_clf = CalibratedClassifierCV(base_estimator=lgbc_bal, cv=3,␣
 ↪method='sigmoid')
calibrated_clf.fit(X_train, y_train)
```

[145]:
```
CalibratedClassifierCV(base_estimator=LGBMClassifier(random_state=0), cv=3)
```

[146]:
```
y_pred_calib = calibrated_clf.predict(X_test)
y_prob_pred_calib = calibrated_clf.predict_proba(X_test)[:, 1]
```

[147]:
```
len(calibrated_clf.calibrated_classifiers_)
```

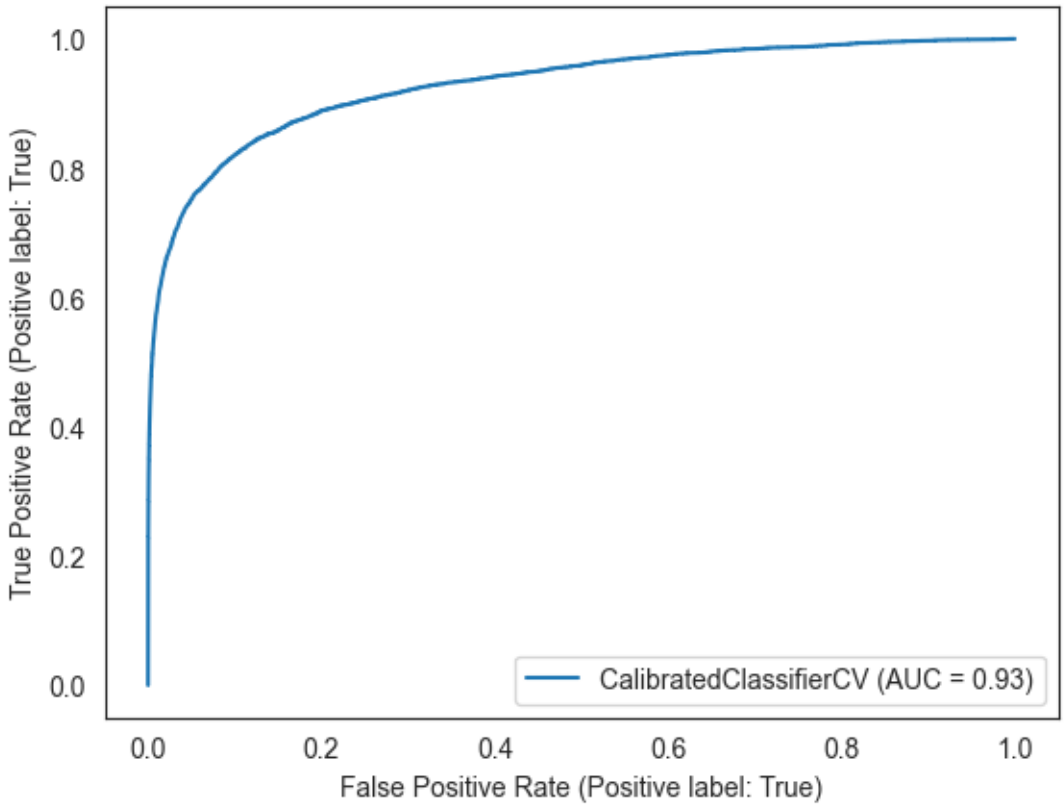
[147]:
```
3
```

[148]:
```
print("Y predicted : ", y_pred_calib)
print("Y probability predicted : ", y_prob_pred_calib[:5])
```

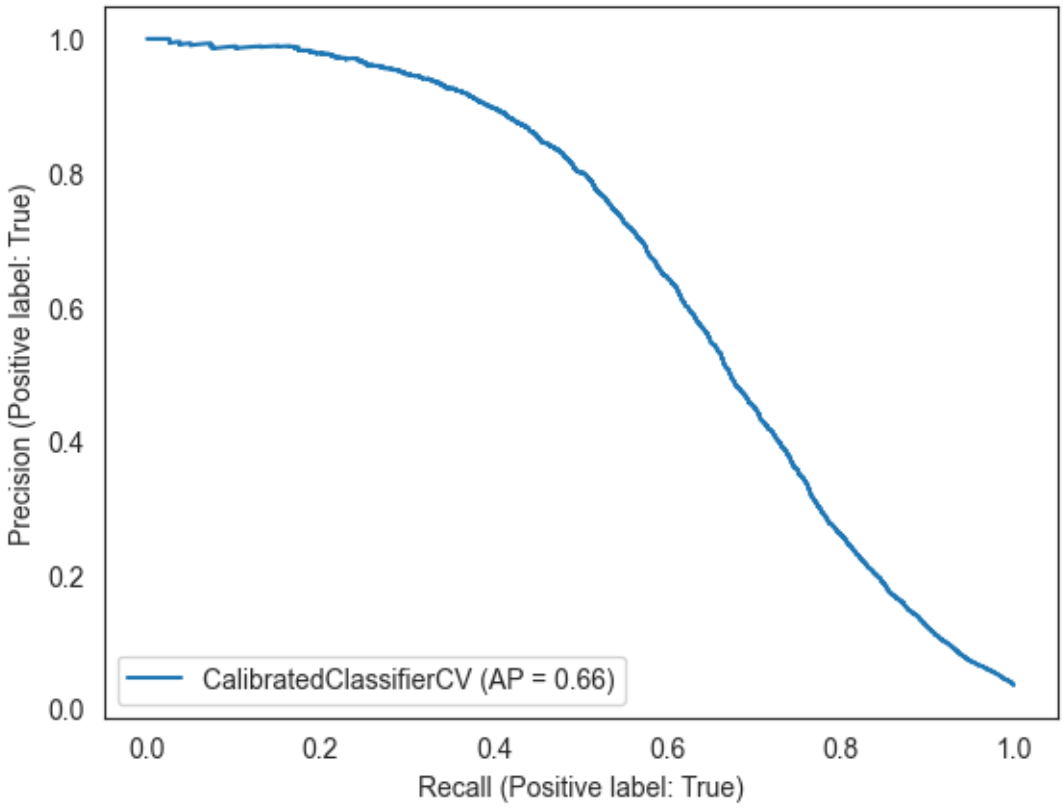
```
Y predicted :  [False False False … False False False]
Y probability predicted :  [0.01510527 0.013539   0.01360709 0.0235388
0.01342325]
```

[150]:
```
compute_evaluation_metric(calibrated_clf, X_test, y_test, y_pred_calib,␣
 ↪y_prob_pred_calib)
```

```
Accuracy Score :  0.9781160745532338

AUC Score :  0.9292560410642979

Confusion Matrix :
 [[170502    461]
 [  3416   2783]]

Classification Report :
              precision    recall  f1-score   support

       False       0.98      1.00      0.99    170963
        True       0.86      0.45      0.59      6199

    accuracy                           0.98    177162
   macro avg       0.92      0.72      0.79    177162
weighted avg       0.98      0.98      0.97    177162


Concordance Index :  0.9292560391771487

ROC curve :
```

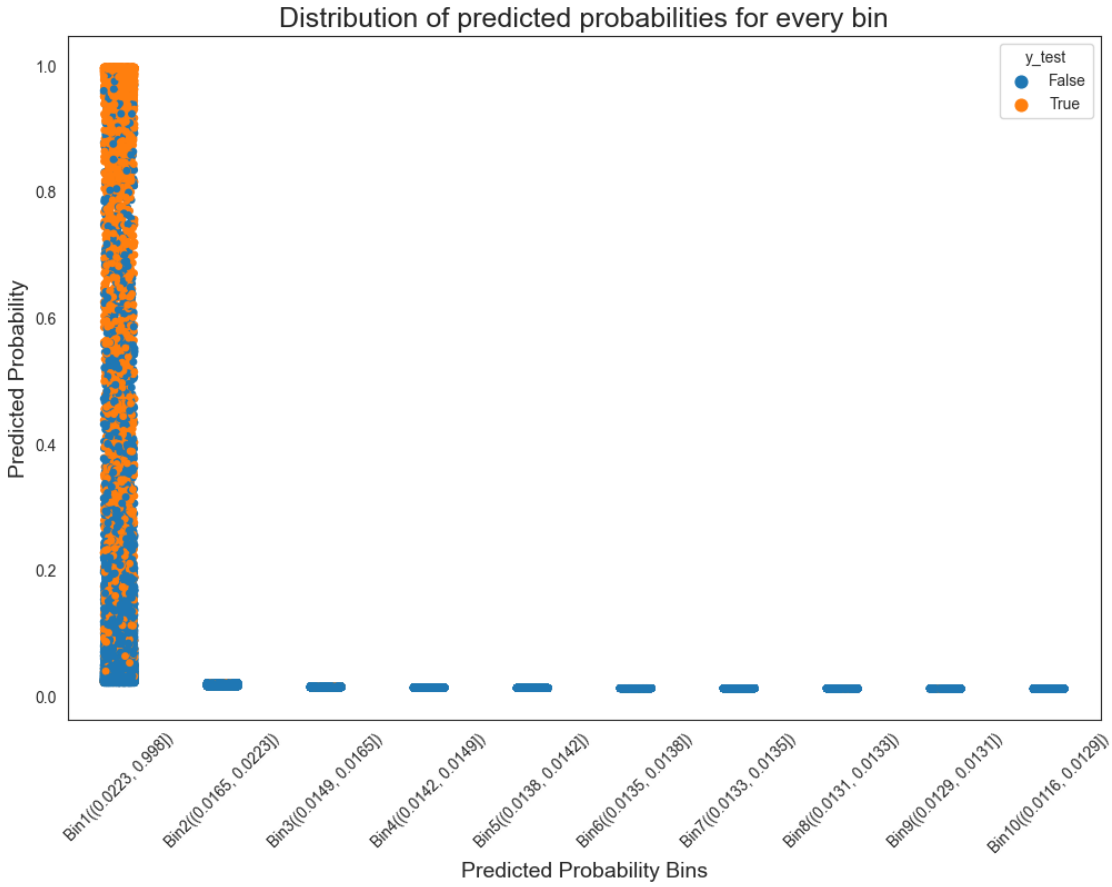

1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate (Positive label: True)
CalibratedClassifierCV (AUC = 0.93)
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate (Positive label: True)

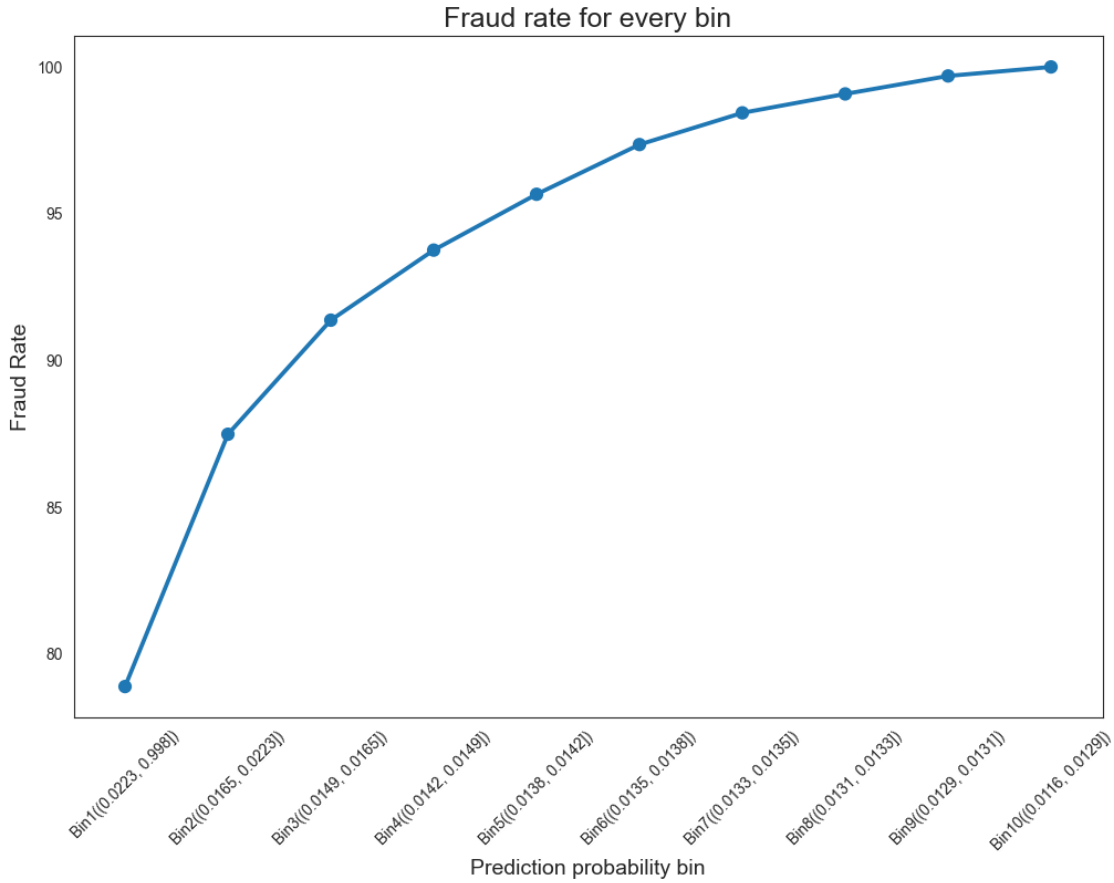
PR curve :




```
Additional Metrics:
TPR (Recall) : 0.4489
FPR : 0.0027
TNR (Specificity) : 0.9973
FNR : 0.5511
```

[151]:
```
captures(y_test, y_pred_calib, y_prob_pred_calib)
```


Distribution of predicted probabilities for every bin
y_test
False
True
Predicted Probability
1.0
0.8
0.6
0.4
0.2
0.0
Bin1((0.0223, 0.998])
Bin2((0.0165, 0.0223])
Bin3((0.0149, 0.0165])
Bin4((0.0142, 0.0149])
Bin5((0.0138, 0.0142])
Bin6((0.0135, 0.0138])
Bin7((0.0133, 0.0135])
Bin8((0.0131, 0.0133])
Bin9((0.0129, 0.0131])
Bin10((0.0116, 0.0129])
Predicted Probability Bins



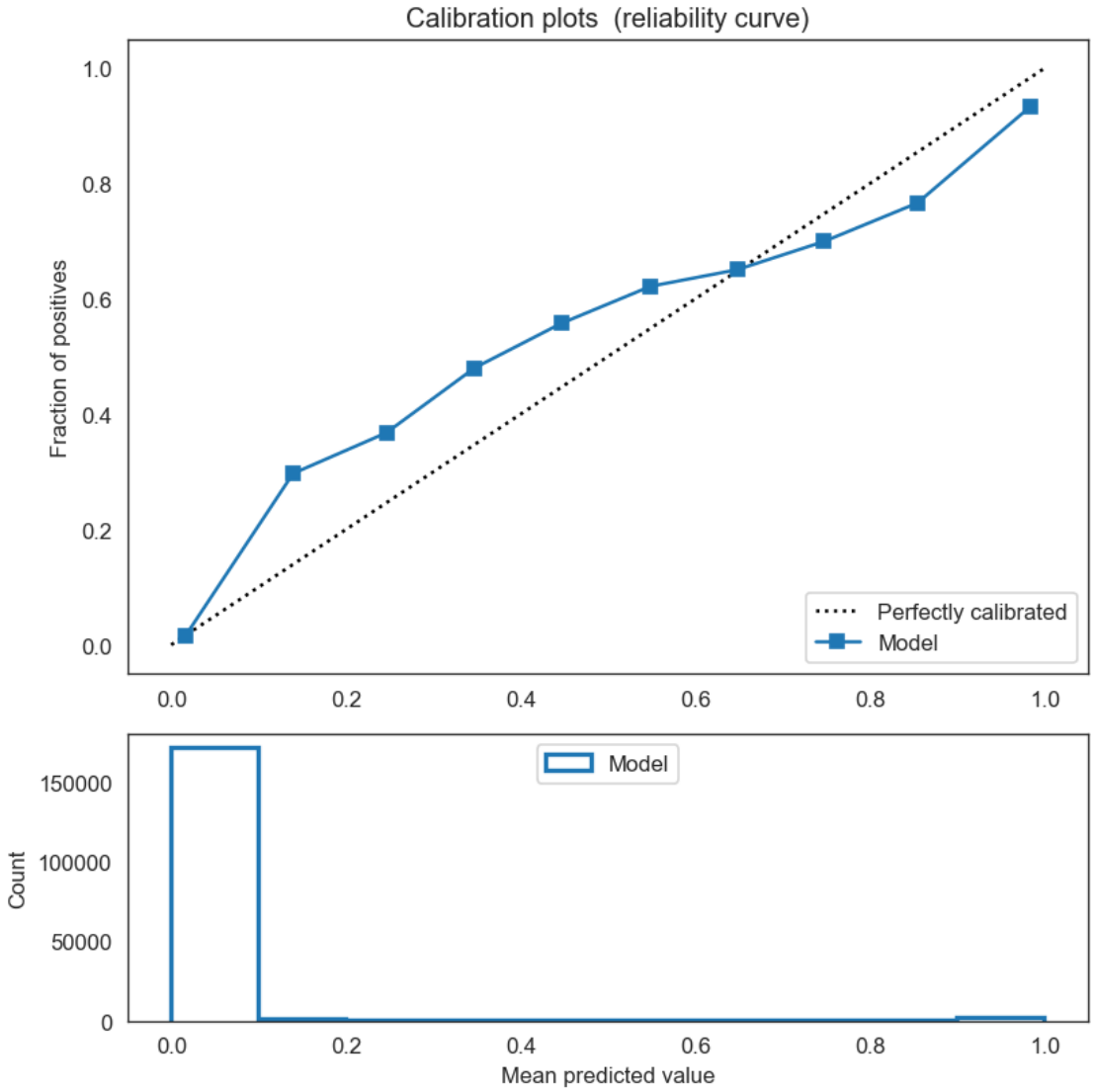
[151]:

	prob_bin	not_fraud	fraud	perc_fraud	perc_not_fraud	\
0	Bin1((0.0223, 0.998])	12827	4890	0.788837	0.075028	
1	Bin2((0.0165, 0.0223])	17183	533	0.085982	0.100507	
2	Bin3((0.0149, 0.0165])	17475	241	0.038877	0.102215	
3	Bin4((0.0142, 0.0149])	17568	148	0.023875	0.102759	
4	Bin5((0.0138, 0.0142])	17598	118	0.019035	0.102935	
5	Bin6((0.0135, 0.0138])	17611	105	0.016938	0.103011	
6	Bin7((0.0133, 0.0135])	17649	67	0.010808	0.103233	
7	Bin8((0.0131, 0.0133])	17676	40	0.006453	0.103391	
8	Bin9((0.0129, 0.0131])	17678	38	0.006130	0.103402	
9	Bin10((0.0116, 0.0129])	17698	19	0.003065	0.103519	

	cum_perc_fraud	cum_perc_not_fraud
0	78.883691	7.502793
1	87.481852	17.553506
2	91.369576	27.775016
3	93.757058	38.050923
4	95.660590	48.344379
5	97.354412	58.645438

6	98.435231	68.968724
7	99.080497	79.307803
8	99.693499	89.648053
9	100.000000	100.000000

[152]:
```
draw_calibration_curve(y_test, y_prob_pred_calib, n_bins=10)
```



19 18. Model Tuning

[153]:
```
%%time

lgbmclassifier = LGBMClassifier(
    class_weight=None,
    learning_rate=0.1,
    max_depth=8,
    n_estimators=100,
    num_leaves=256,
    reg_alpha=0.5,
    random_state=0
)

lgbmclassifier.fit(X_train, y_train)
```

```
Wall time: 34.7 s
```

[153]:
```
LGBMClassifier(max_depth=8, num_leaves=256, random_state=0, reg_alpha=0.5)
```

[154]:
```
y_g_pred = lgbmclassifier.predict(X_test)
y_prob_g_pred = lgbmclassifier.predict_proba(X_test)[:, 1]
print("Y predicted : ",y_g_pred)
print("Y probability predicted : ",y_prob_g_pred[:5])
```

```
Y predicted :  [False False False … False False False]
Y probability predicted :  [0.00549227 0.00615294 0.007173   0.09576845
0.00734595]
```

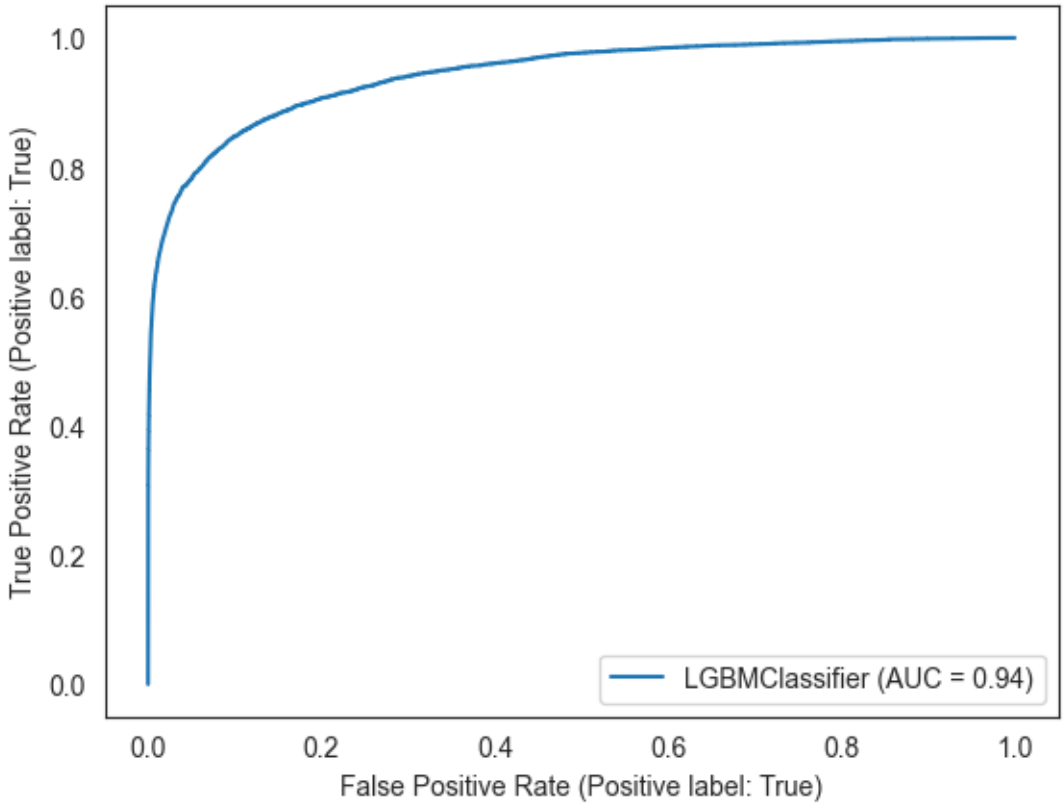
[155]:
```
compute_evaluation_metric(lgbmclassifier, X_test, y_test, y_g_pred,␣
 ↪y_prob_g_pred)
```

```
Accuracy Score :  0.9796909043700116

AUC Score :  0.9439534102237099

Confusion Matrix :
 [[170676    287]
 [  3311   2888]]

Classification Report :
              precision    recall  f1-score   support

       False       0.98      1.00      0.99    170963
        True       0.91      0.47      0.62      6199

    accuracy                           0.98    177162
```

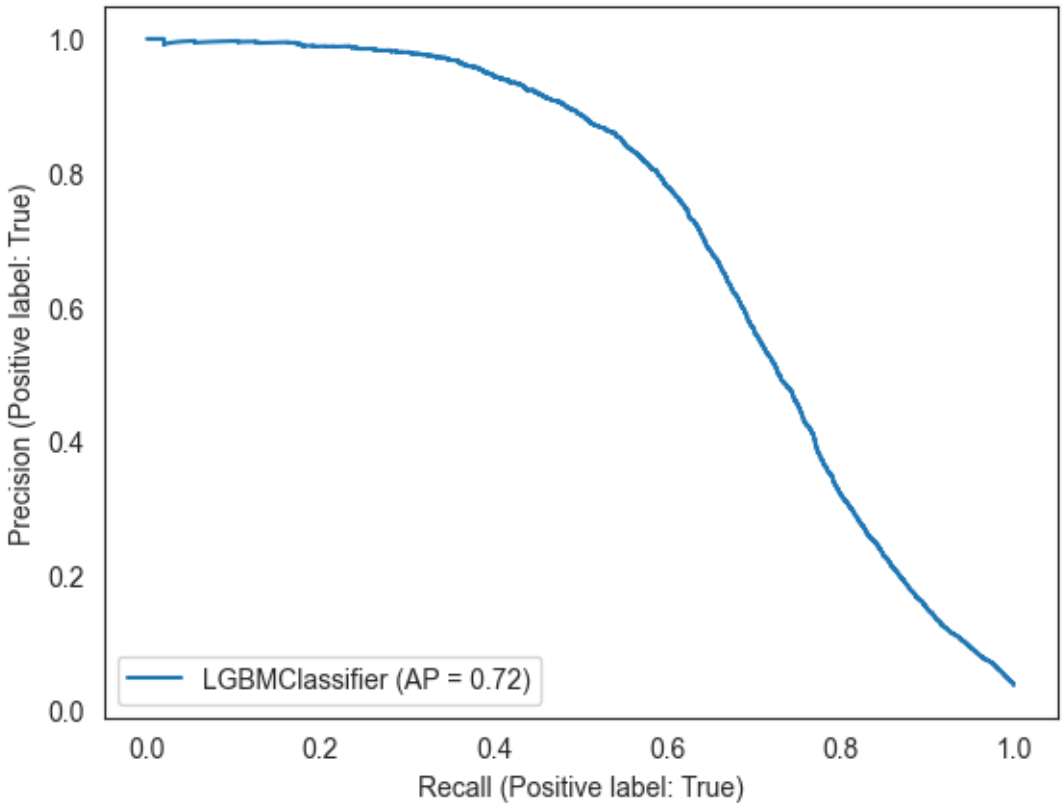
```
   macro avg       0.95      0.73      0.80    177162
weighted avg       0.98      0.98      0.98    177162
```

```
Concordance Index :  0.9439534040904752
```

```
ROC curve :
```

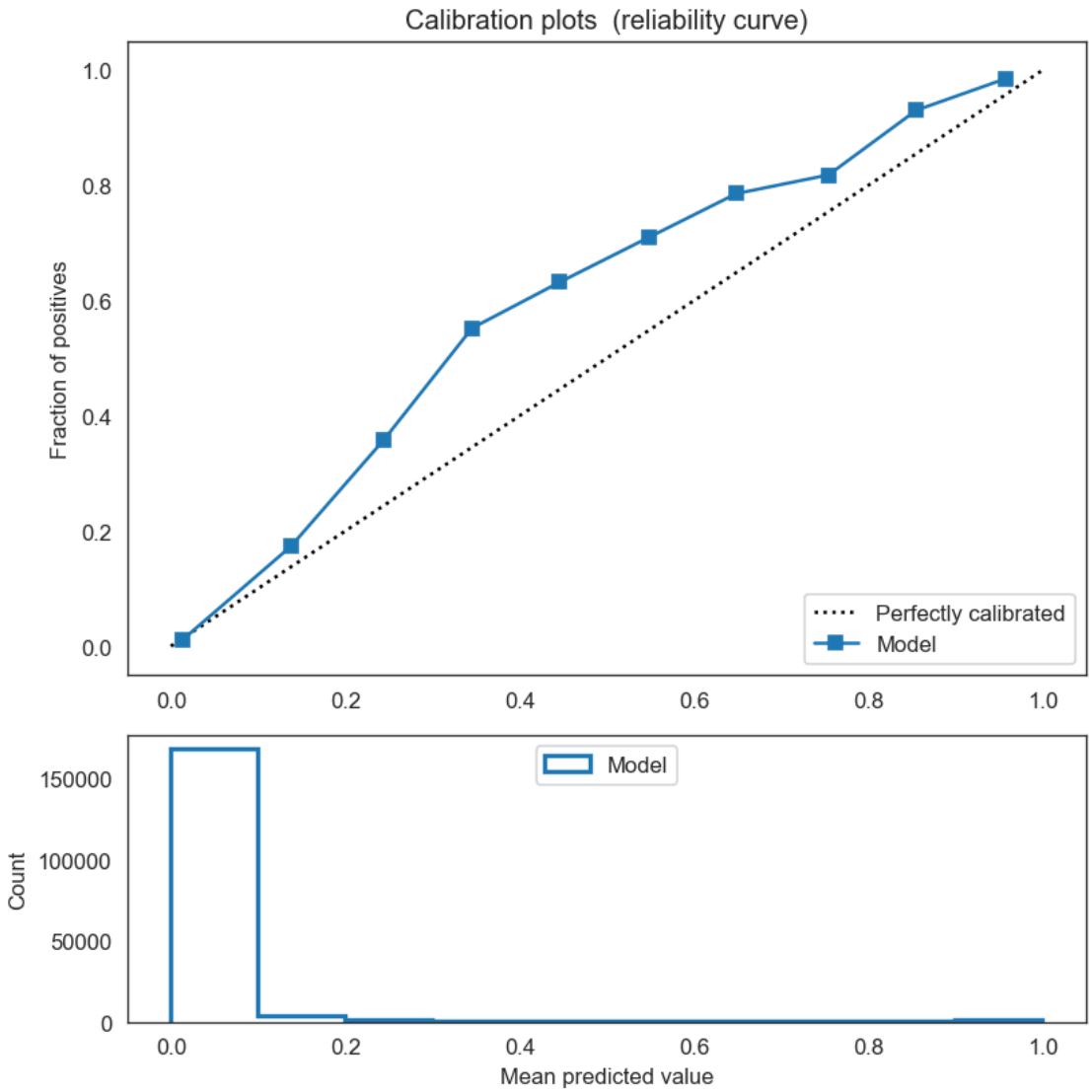



```
PR curve :
```




```
Additional Metrics:
TPR (Recall) : 0.4659
FPR : 0.0017
TNR (Specificity) : 0.9983
FNR : 0.5341
```

[156]:
```
draw_calibration_curve(y_test, y_prob_g_pred, n_bins=10)
```

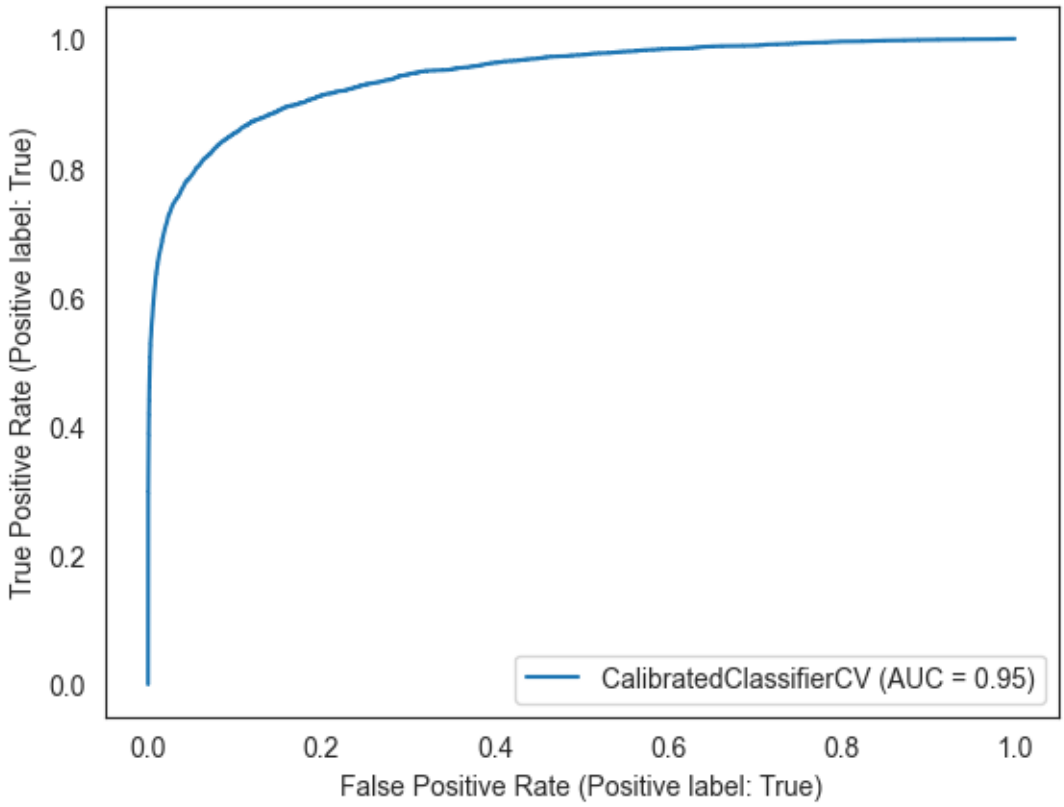


[157]:
```
# Calibrate
calibrated_clf = CalibratedClassifierCV(base_estimator=lgbmclassifier, cv=3)
calibrated_clf.fit(X_train, y_train)
y_pred_calib = calibrated_clf.predict(X_test)
y_prob_pred_calib = calibrated_clf.predict_proba(X_test)[:, 1]
```

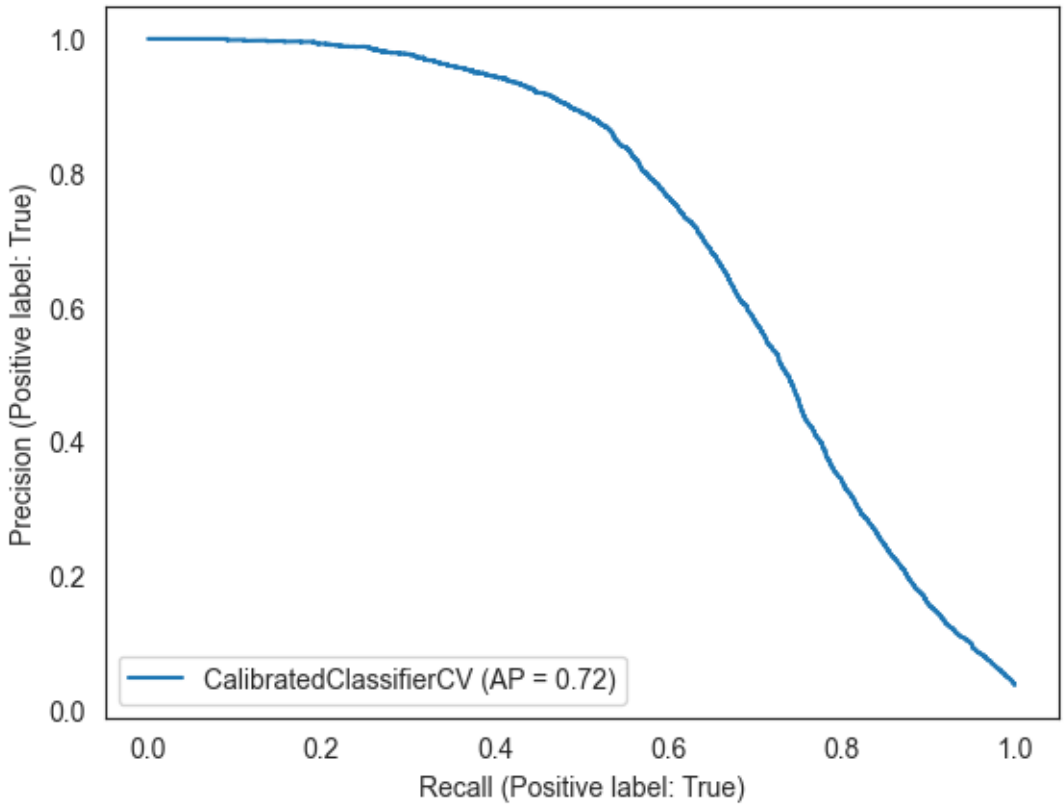
[158]:
```
compute_evaluation_metric(calibrated_clf, X_test, y_test, y_pred_calib,␣
↪y_prob_pred_calib)
```

```
Accuracy Score :  0.9804416296948556

AUC Score :  0.9456479526988175
```

```
Confusion Matrix :
 [[170572    391]
 [  3074   3125]]

Classification Report :
              precision    recall  f1-score   support

       False       0.98      1.00      0.99    170963
        True       0.89      0.50      0.64      6199

    accuracy                           0.98    177162
   macro avg       0.94      0.75      0.82    177162
weighted avg       0.98      0.98      0.98    177162


Concordance Index :  0.9456479526988175

ROC curve :
```

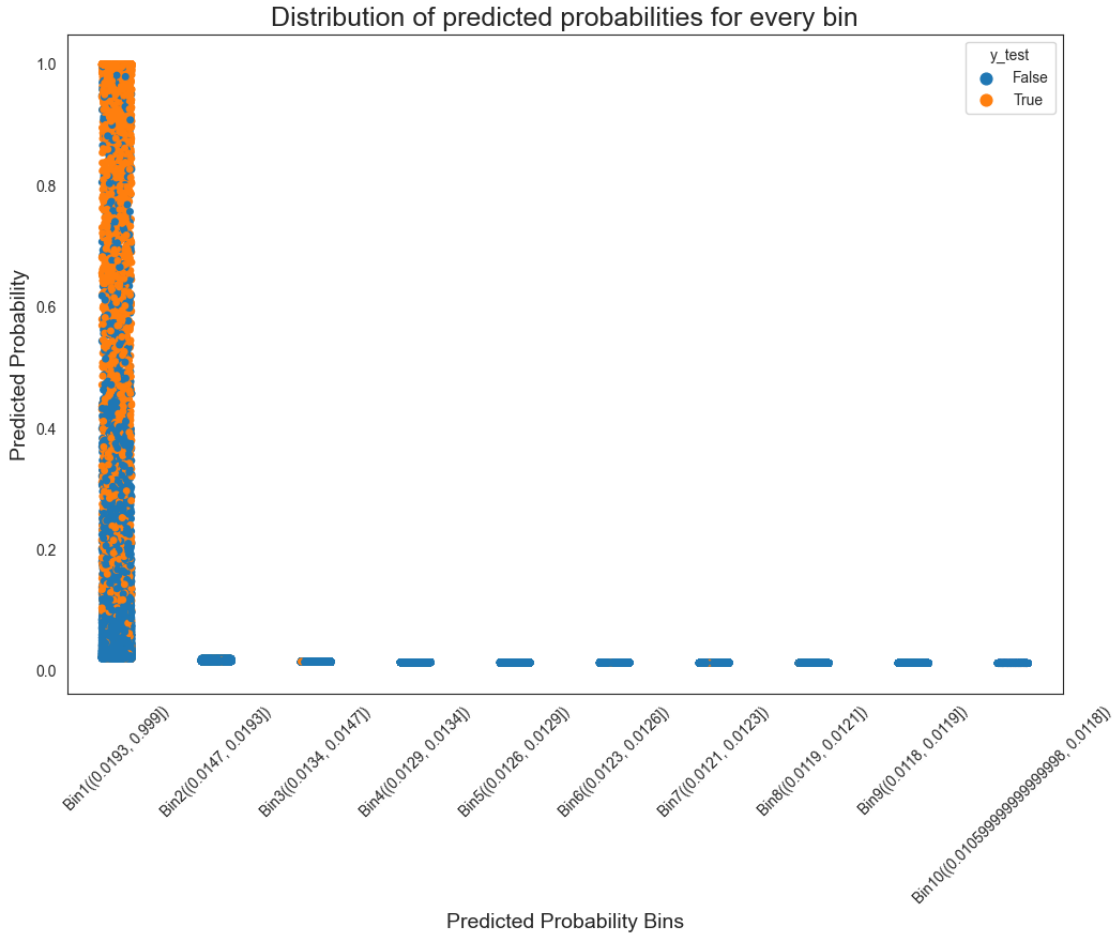


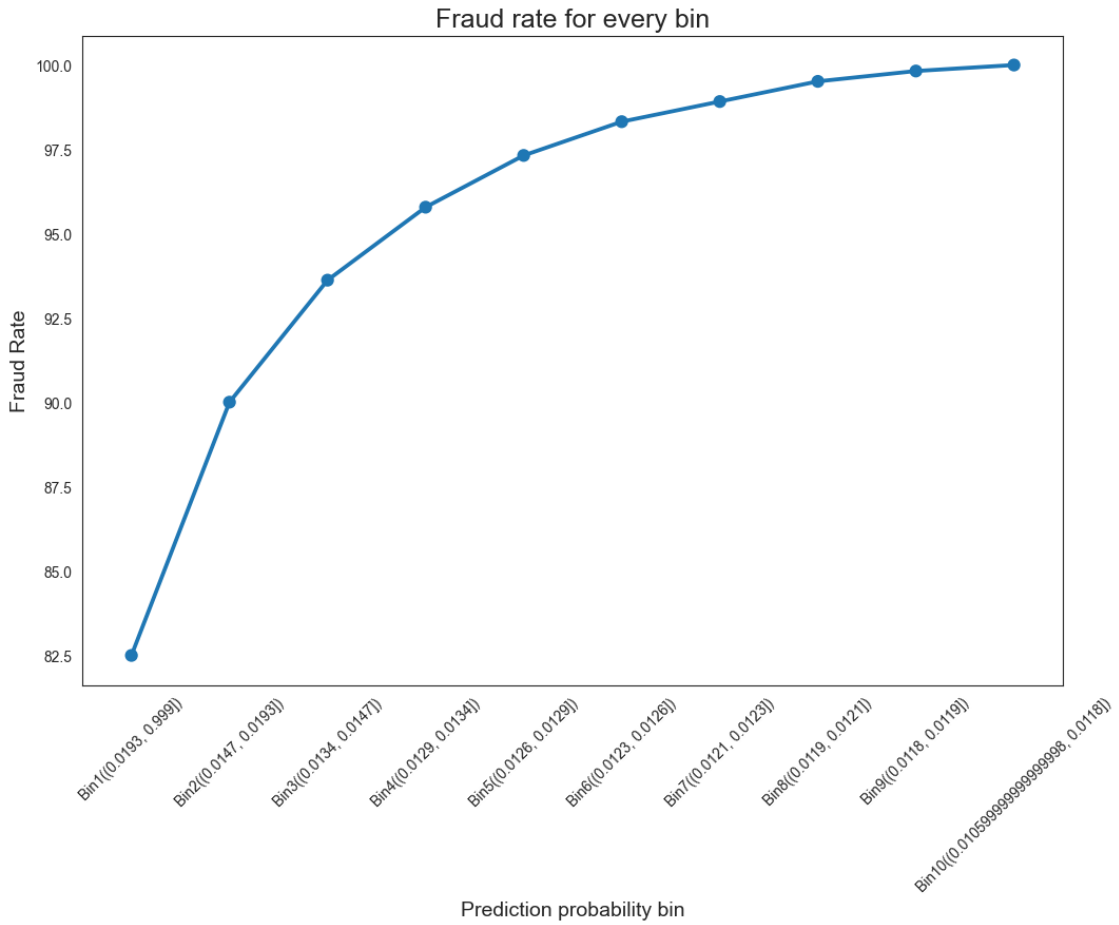
```
PR curve :
```




```
Additional Metrics:
TPR (Recall) : 0.5041
FPR : 0.0023
TNR (Specificity) : 0.9977
FNR : 0.4959
```

[160]:
```
captures(y_test, y_pred_calib, y_prob_pred_calib)
```


Distribution of predicted probabilities for every bin
y_test
False
True
Predicted Probability
1.0
0.8
0.6
0.4
0.2
0.0
Bin1((0.0193, 0.999])
Bin2((0.0147, 0.0193])
Bin3((0.0134, 0.0147])
Bin4((0.0129, 0.0134])
Bin5((0.0126, 0.0129])
Bin6((0.0123, 0.0126])
Bin7((0.0121, 0.0123])
Bin8((0.0119, 0.0121])
Bin9((0.0118, 0.0119])
Bin10((0.0105999999999998, 0.0118])
Predicted Probability Bins

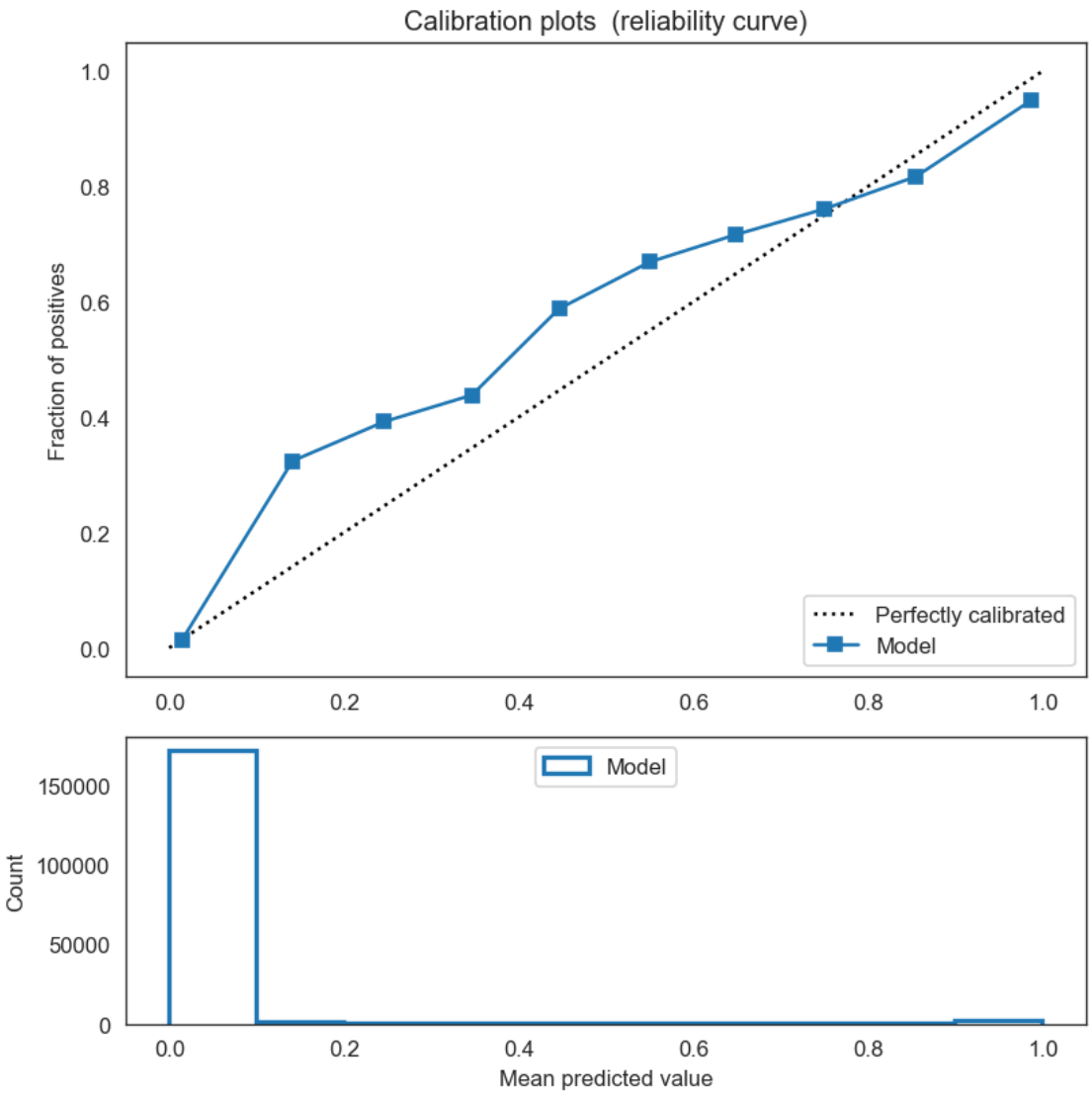


[160]:

	prob_bin	not_fraud	fraud	perc_fraud	perc_not_fraud	cum_perc_fraud	cum_perc_not_fraud
0	Bin1((0.0193, 0.999])	12602	5115	0.825133	0.073712	82.513309	7.371186
1	Bin2((0.0147, 0.0193])	17251	465	0.075012	0.100905	90.014518	17.461673
2	Bin3((0.0134, 0.0147])	17492	224	0.036135	0.102315	93.628005	27.693127
3	Bin4((0.0129, 0.0134])	17582	134	0.021616	0.102841	95.789643	37.977223
4	Bin5((0.0126, 0.0129])	17621	95	0.015325	0.103069	97.322149	48.284132
5	Bin6((0.0123, 0.0126])	17654	62	0.010002			
6	Bin7((0.0121, 0.0123])	17679	37	0.005969			
7	Bin8((0.0119, 0.0121])	17679	37	0.005969			
8	Bin9((0.0118, 0.0119])	17697	19	0.003065			
9	Bin10((0.010599999999999998, 0.0118])	17706	11	0.001774			

```
5       0.103262       98.322310       58.610343
6       0.103408       98.919181       68.951177
7       0.103408       99.516051       79.292011
8       0.103514       99.822552       89.643373
9       0.103566      100.000000      100.000000
```

[159]:
```
draw_calibration_curve(y_test, y_prob_pred_calib, n_bins=10)
```



- Accuracy score is 0.98
- AUC score and Concordance index are 0.95, which are the best so far
- Classfication report is also balanced between both the classes
- ROC curve and PR are the best so far

20 19. Feature Importance

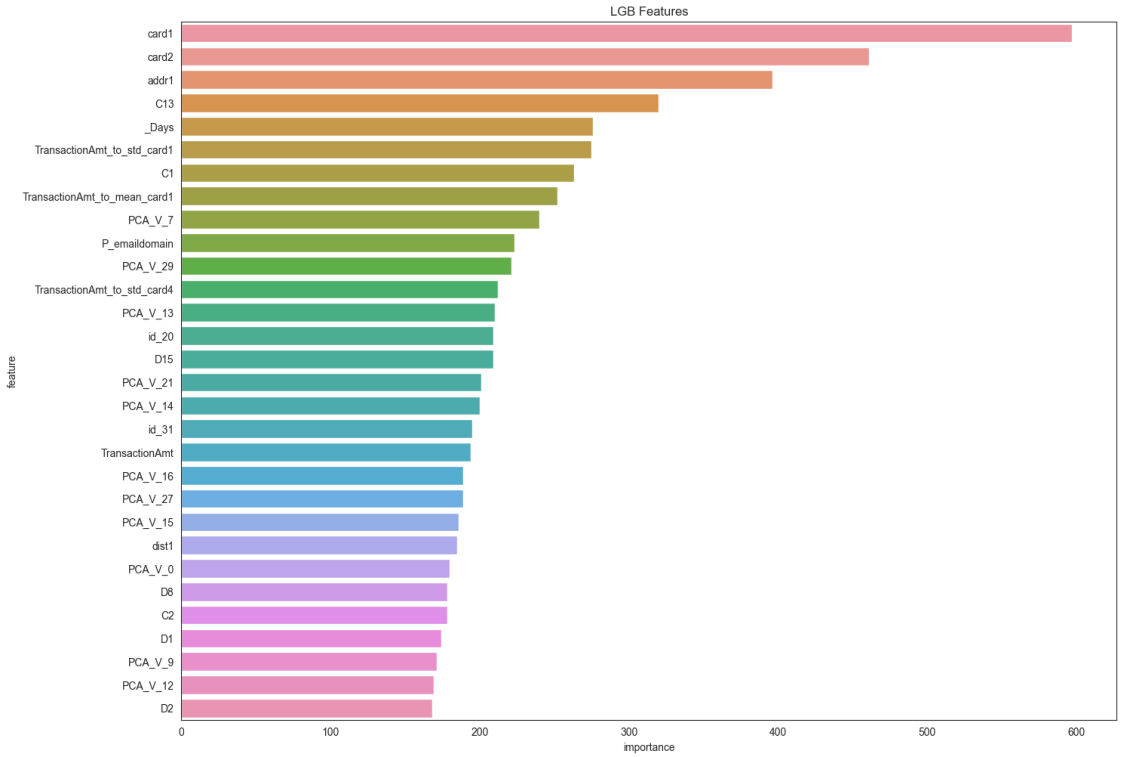
[161]:
```
feature_importance_df = pd.DataFrame({'feature' : X_train.columns, 'importance'␣
↪: lgbmclassifier.feature_importances_ })
```

[162]:
```
feature_importance_df = feature_importance_df.sort_values(by="importance",␣
↪ascending=False)
feature_importance_df = feature_importance_df.iloc[:30,:]
feature_importance_df
```

[162]:

	feature	importance
2	card1	597
3	card2	461
8	addr1	396
25	C13	320
496	_Days	276
501	TransactionAmt_to_std_card1	275
13	C1	263
499	TransactionAmt_to_mean_card1	252
510	PCA_V_7	240
11	P_emaildomain	223
532	PCA_V_29	221
502	TransactionAmt_to_std_card4	212
516	PCA_V_13	210
66	id_20	209
40	D15	209
524	PCA_V_21	201
517	PCA_V_14	200
70	id_31	195
0	TransactionAmt	194
519	PCA_V_16	189
530	PCA_V_27	189
518	PCA_V_15	186
10	dist1	185
503	PCA_V_0	180
33	D8	178
14	C2	178
27	D1	174
512	PCA_V_9	171
515	PCA_V_12	169
28	D2	168

[163]:
```
plt.figure(figsize=(16, 12));
sns.barplot(x="importance", y="feature", data=feature_importance_df.
↪sort_values(by="importance", ascending=False));
plt.title('LGB Features');
```

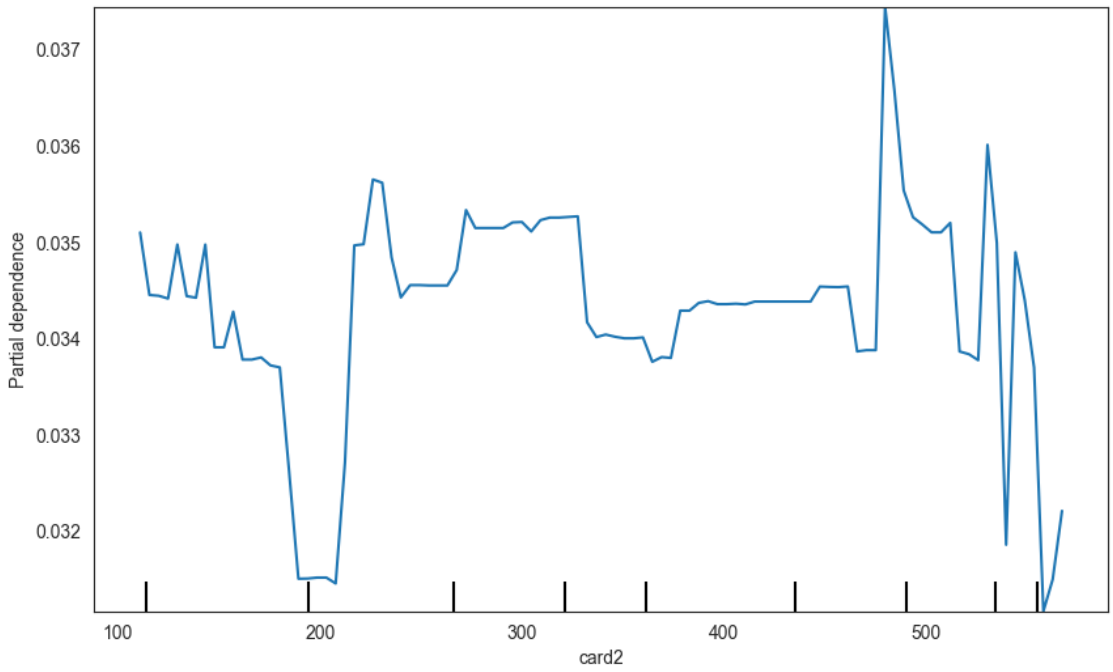


- card1 is contributing the most in predicting if a transaction is fraud or not
- card2, addr1, C13, P_emaildomain, C1,_days etc are some of the most important features in predicting the fraud
- Certain card types, addresses and emails are at high risk of fraud, so there is a need to monitor these carefully

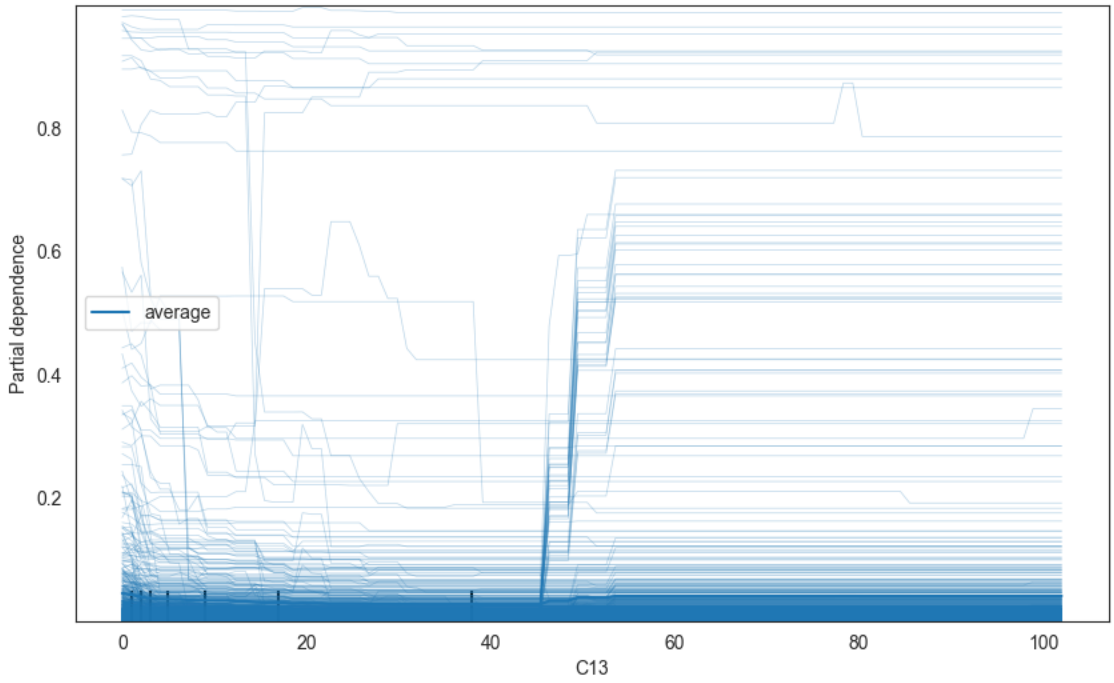
21 20. Partial Dependence and Individual Conditional Expectations (ICE)

Plot Partial Dependence

[168]:
```
from sklearn.inspection import PartialDependenceDisplay
import matplotlib.pyplot as plt

fig, ax = plt.subplots(figsize=(10, 6))
PartialDependenceDisplay.from_estimator(
    lgbmclassifier,        # trained LightGBM model
    X_train,               # your training data
    features=['card2'],    # feature name as a list
    ax=ax
)
plt.show()
```



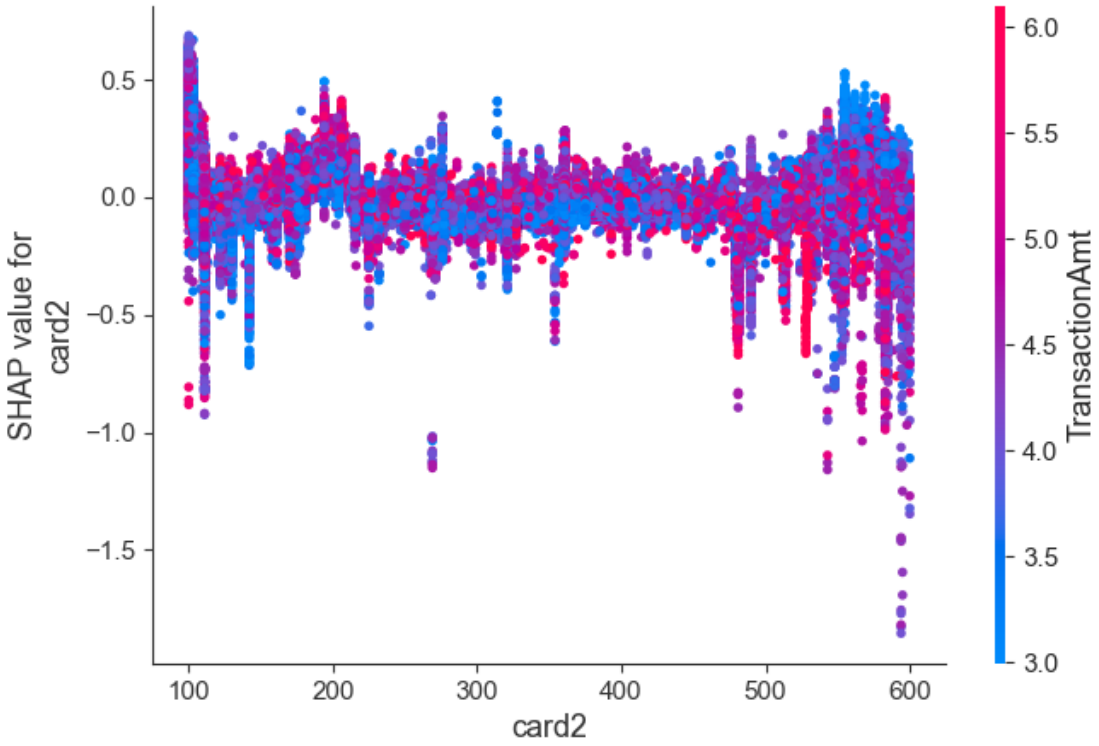
[170]:
```
fig, ax = plt.subplots(figsize=(10, 6))

PartialDependenceDisplay.from_estimator(
    lgbmclassifier,
    X_train,
    features=['C13'],
    kind='both',
    ax=ax
)

plt.show()
```



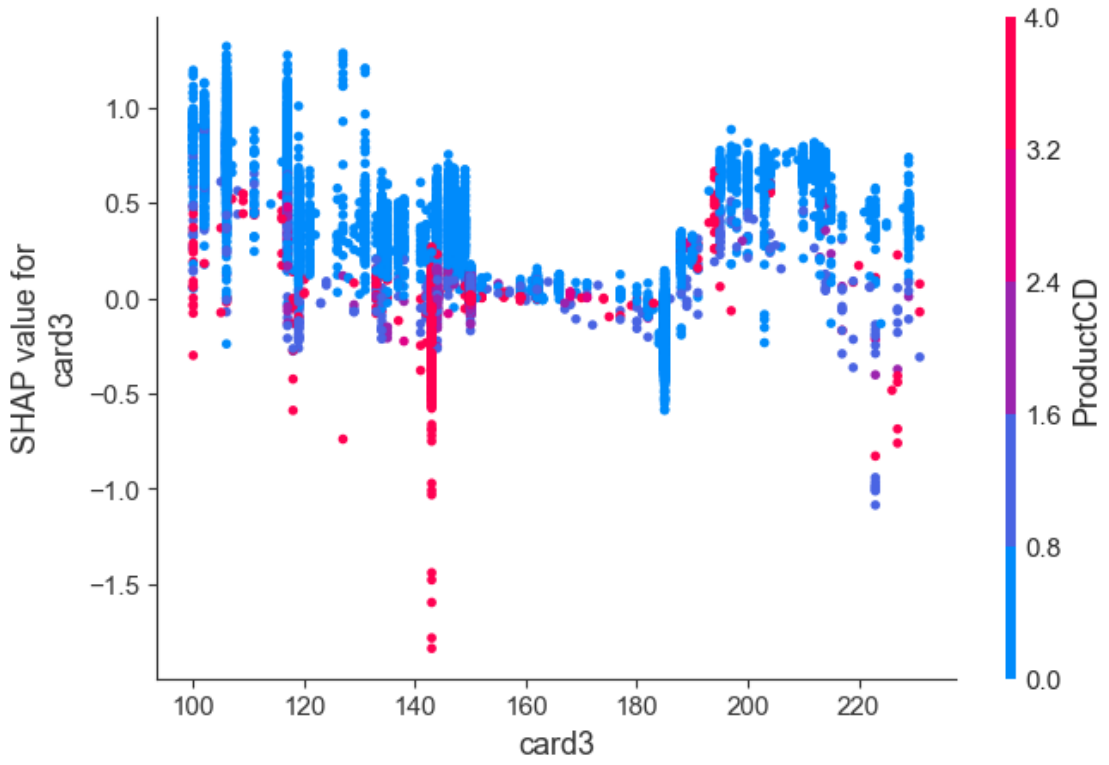
22 21. SHAP Values

[172]:
```
import shap
shap_model = shap.TreeExplainer(lgbmclassifier)
shap_values = shap_model.shap_values(X_train)
```

[173]:
```
shap.dependence_plot("card2", shap_values[0], X_train)
```



[174]:
```
shap.dependence_plot("card3", shap_values[0], X_train)
```



[175]:
```
shap.initjs()
shap.force_plot(shap_model.expected_value[1], shap_values[1][14], X_train.
 ↪iloc[14, :])
```

```
<IPython.core.display.HTML object>
```

[175]: <shap.plots._force.AdditiveForceVisualizer at 0x1e026cad048>

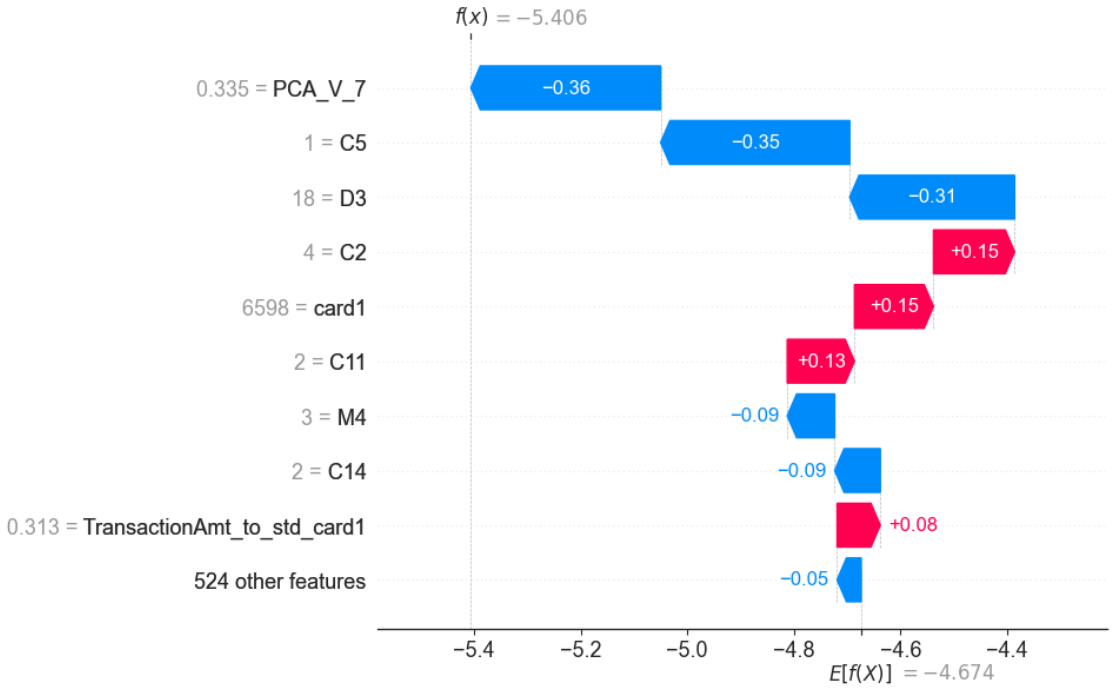
[176]:
```
shap.initjs()  # needed to show viz
shap.force_plot(shap_model.expected_value[1], shap_values[1][14], X_train.
 ↪iloc[14, :], link='logit')
```

```
<IPython.core.display.HTML object>
```

[176]: <shap.plots._force.AdditiveForceVisualizer at 0x1e02525f898>

[183]:
```
index = 0

shap.plots.waterfall(
    shap.Explanation(
        values=shap_values[1][index],
        base_values=shap_model.expected_value[1],
        data=X_train.iloc[index],
```

```
        feature_names=X_train.columns
    )
)
```



[ ]:

22.1 Author

MIT TRIVEDI

22.2 Project

Credit Card Fraud Detection

22.3 Date

October 2025

22.4 Contact

LinkedIn · GitHub · Email

22.5 Summary

This project presents an **end-to-end credit card fraud detection pipeline** featuring:

- **Models:** LightGBM and XGBoost
- **Feature Engineering:** Handling missing values, encoding, new features, etc.
- **Model Evaluation:** AUC, confusion matrix, precision, recall
- **Explainability:** SHAP visualizations for model interpretability

All **code and notebooks are available in this repository** for reproducibility and experimentation.

[ ]: