MICRO CREDIT DEFAULTER PROJECT

PROBLEM STATEMENT:

A Microfinance Institution (MFI) is an organization that offers financial services to low income populations. MFS becomes very useful when targeting especially the unbanked poor families living in remote areas with not much sources of income. The Microfinance services (MFS) provided by MFI are Group Loans, Agricultural Loans, Individual Business Loans and so on. Many microfinance institutions (MFI), experts and donors are supporting the idea of using mobile financial services (MFS) which they feel are more convenient and efficient, and cost saving, than the traditional high-touch model used since long for the purpose of delivering microfinance services. Though, the MFI industry is primarily focusing on low income families and are very useful in such areas, the implementation of MFS has been uneven with both significant challenges and successes. Today, microfinance is widely accepted as a poverty-reduction tool, representing $70 billion in outstanding loans and a global outreach of 200 million clients.

We are working with one such client that is in Telecom Industry. They are a fixed wireless telecommunications network provider. They have launched various products and have developed its business and organization based on the budget operator model, offering better products at Lower Prices to all value conscious customers through a strategy of disruptive innovation that focuses on the subscriber.

They understand the importance of communication and how it affects a person's life, thus, focusing on providing their services and products to low income families and poor customers that can help them in the need of hour.

They are collaborating with an MFI to provide micro-credit on mobile balances to be paid back in 5 days. The Consumer is believed to be defaulter if he deviates from the path of paying back the loaned amount within the time duration of 5 days.

For the loan amount of 5 (in Indonesian Rupiah), payback amount should be 6 (in Indonesian Rupiah), while, for the loan amount of 10 (in Indonesian Rupiah), the payback amount should be 12 (in Indonesian Rupiah).

The sample data is provided to us from our client database. It is hereby given to you for this exercise. In order to improve the selection of customers for the credit, the client wants some predictions that could help them in further investment and improvement in selection of customers.

UNDERSTANDING:

Nowadays, the institution provides small loans to poor clients who typically lack collateral, steady employment, verifiable credit history to improve borrowers from the selection of customers for the credit and the client wants some predictions that could help them in further investment and improvement in selection of customers. The Microfinance Institution provided Group Loans, Agricultural Loans and Individual Business Loans which helps to small community. Micro Credit is the extension of very small loans. It is designed to support entrepreneurship and alleviate poverty. Today, microfinance is widely accepted as a poverty-reduction tool, representing $70 billion in outstanding loans and a global outreach of 200 million clients.

EDA STEPS AND VISUALIZATIONS:

1. Data Collection

```
1 df=pd.read_csv(r'C:/Users/user/Desktop/FRTech Internship Project/Micro Credit Project/Data file.csv')
2 df.head()
```

|	Unnamed: 0	label	msisdn	aon	daily_decr30	daily_decr90	rental30	rental90	last_rech_date_ma	last_rech_date_da	...
0	1	0	21408I70789	272.0	3055.050000	3065.150000	220.13	260.13	2.0	0.0	...
1	2	1	76462I70374	712.0	12122.000000	12124.750000	3691.26	3691.26	20.0	0.0	...
2	3	1	17943I70372	535.0	1398.000000	1398.000000	900.13	900.13	3.0	0.0	...
3	4	1	55773I70781	241.0	21.228000	21.228000	159.42	159.42	41.0	0.0	...
4	5	1	03813I82730	947.0	150.619333	150.619333	1098.90	1098.90	4.0	0.0	...

5 rows × 37 columns

2. Data Cleaning

```
1  total=oversampled.isnull().sum().sort_values(ascending=False)
2  percent = (oversampled.isnull().sum()/oversampled.isnull().count()).sort_values(ascending=False)
3  missing = pd.concat([total, percent], axis=1, keys=['Total' , 'Percent'])
4  missing.head()
```

	Total	Percent
label	0	0.0
rental30	0	0.0
pcircle	0	0.0
payback90	0	0.0
payback30	0	0.0

3. Univariate Analysis

```
1  sns.distplot(df['label']);
```
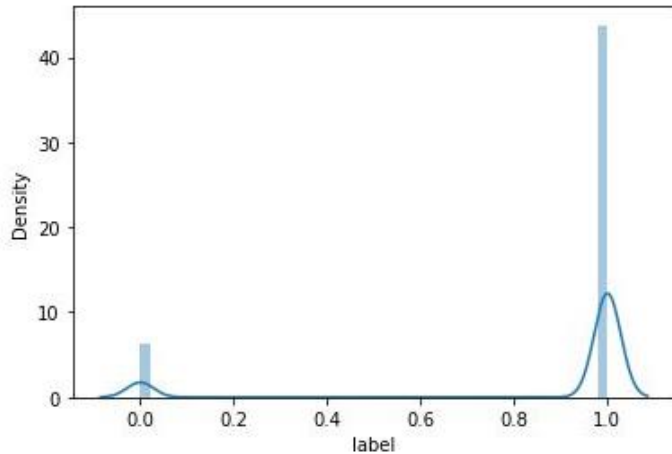


```
1  from sklearn.preprocessing import StandardScaler
2  # Data Scaling Formula Z=(x-mean)/std
3  scaler = StandardScaler()
4  X_scaled=scaler.fit_transform(X)
5  X_scaled
```

```
array([[-0.23621479, -0.31934721, -0.13811893, ..., -0.85648223,
         0.      ,  2.24155209],
       [-0.11036683, -0.21530584,  0.149985  , ...,  0.02496435,
         0.      , -0.44262019],
       [ 0.64217573,  0.40683817, -0.7143268 , ...,  0.15088528,
         0.      , -0.76657202],
       ...,
       [-0.71720003, -0.71698862,  0.149985  , ..., -0.85648223,
         0.      , -1.36819684],
       [-0.11651456, -0.22038831, -0.7143268 , ..., -0.47871941,
         0.      , -0.67401435],
       [ 0.28816543,  0.11416982,  0.149985  , ..., -0.47871941,
         0.      , -0.90540852]])
```

4. Bivariate Analysis

```
In [134]:  1  plt.scatter(y_test,y_pred)
           2  plt.xlabel('Actual Label')
           3  plt.ylabel('Predicted Label')
           4  plt.title('Actual vs model predicted')
           5  plt.show
```

Out[134]: <function matplotlib.pyplot.show(close=None, block=None)>
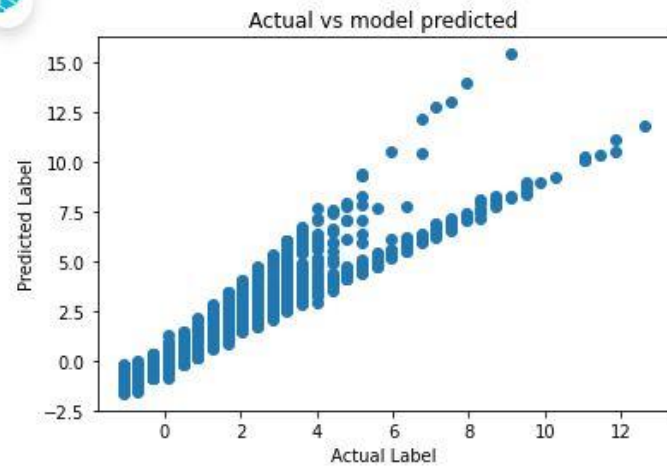

```
In [123]:  1  corr = MiceImputed.corr()
           2  mask = np.triu(np.ones_like(corr, dtype=np.bool))
           3  f, ax = plt.subplots(figsize=(20,20))
           4  cmap = sns.diverging_palette(250, 25, as_cmap=True)
           5  sns.heatmap(corr, mask=mask, cmap=cmap, vmax=None, center=0, square=True, annot=True, linewidths=.5, cbar_kws={"shrink": .9})
           6  plt.show()
```
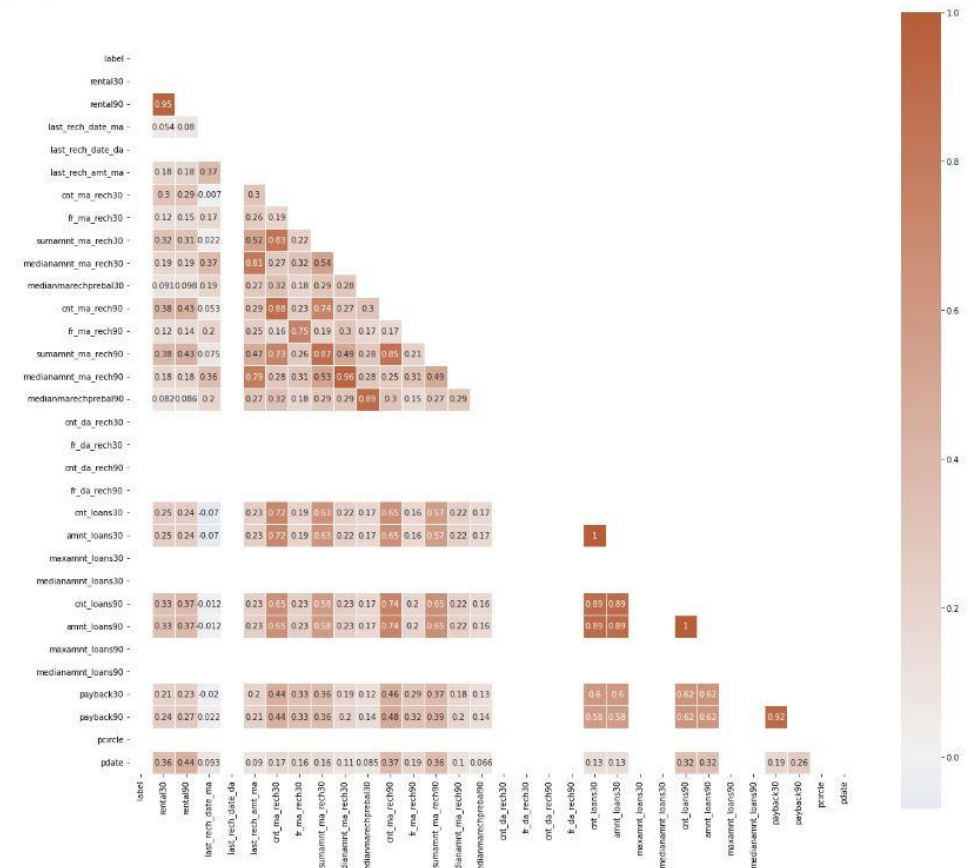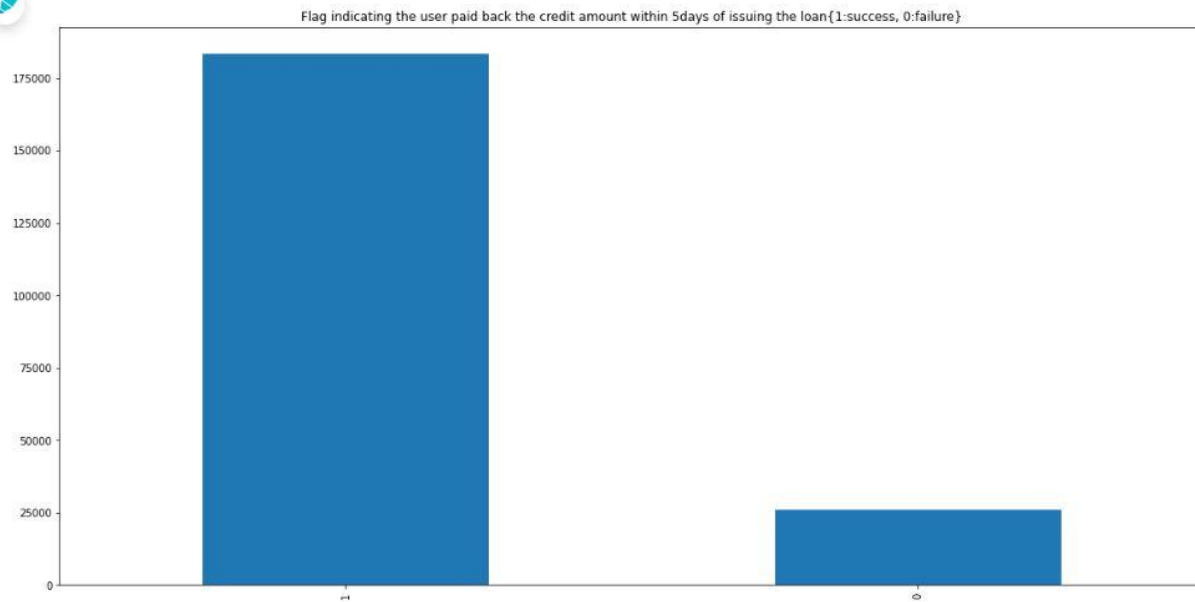
```
plt.subplot()
df['label'].value_counts().plot(kind='bar',title='Flag indicating the user paid back the credit amount within 5days of issuing the loan{1:success, 0:failu
```

`<AxesSubplot:title={'center':'Flag indicating the user paid back the credit amount within 5days of issuing the loan{1:success, 0:failure}'}>`


Flag indicating the user paid back the credit amount within 5days of issuing the loan{1:success, 0:failure}

```
plt.figure(figsize=(10,5))
plt.title('User paid back the credit amount within 5 days of issuing the loan')
ax=sns.countplot(x='label',data=df)
plt.xlabel('Label')
plt.ylabel('Issuing the loan')
for i in ax.patches:
    ax.annotate(int(i.get_height()), (i.get_x()+0.25, i.get_height()+1), va='bottom',color='black')
```


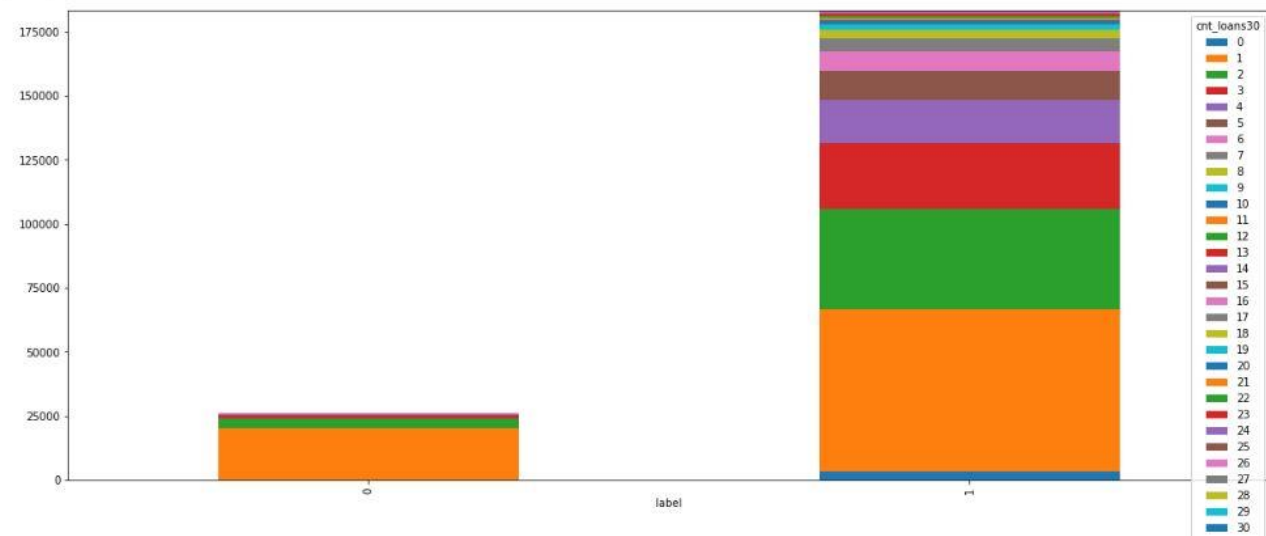User paid back the credit amount within 5 days of issuing the loan
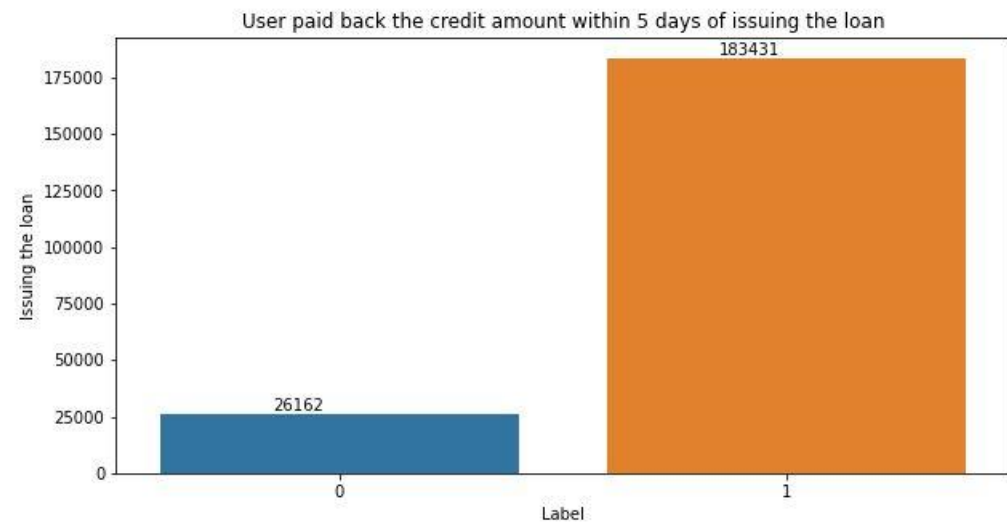
```
1  plt.figure(figsize=(20,10))
2  sns.countplot(df['label'])
3  plt.show()
```
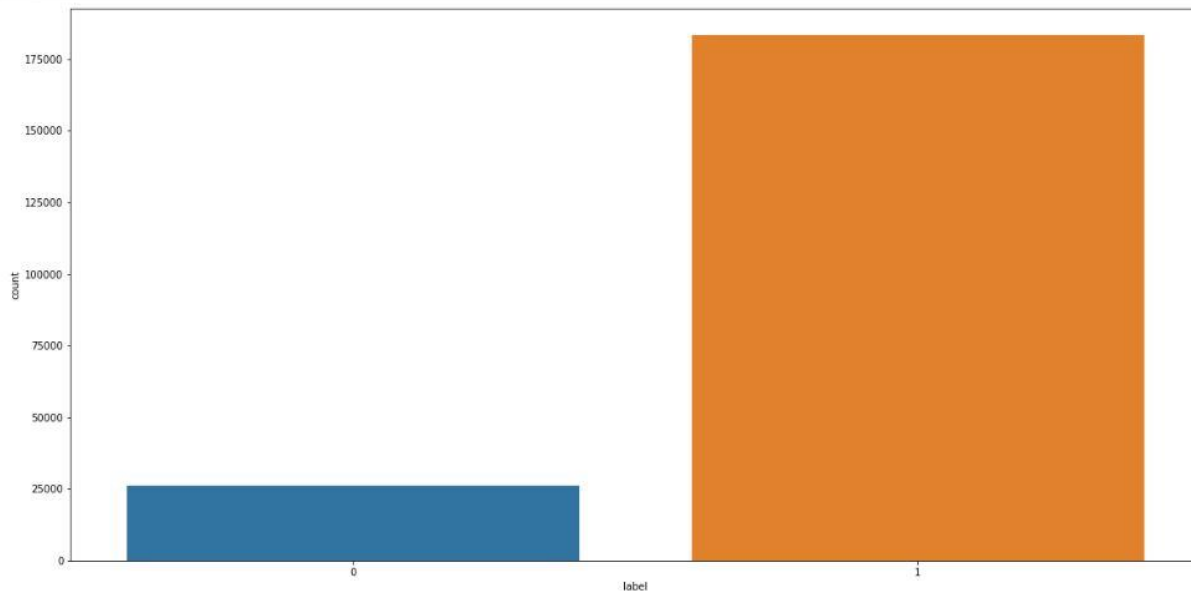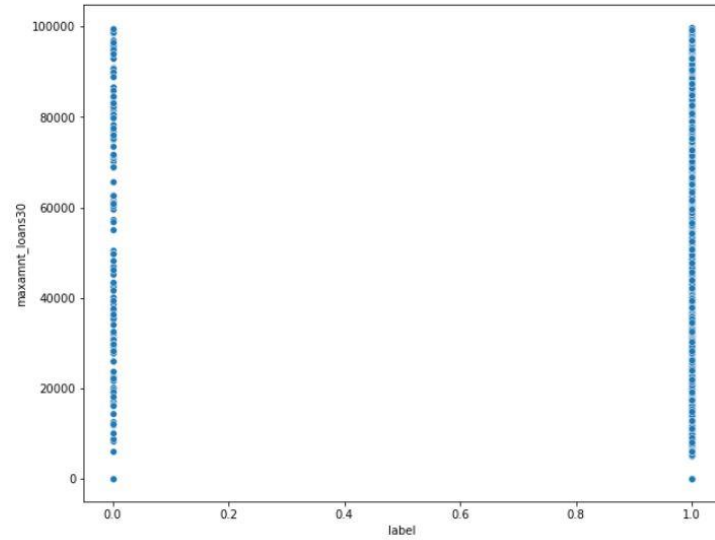


```
1  data=pd.crosstab(df['label'],df['cnt_loans30'])
2  data.plot.bar(stacked=True,figsize=(20,8))
3  plt.xticks(rotation=90)
4  plt.show()
```
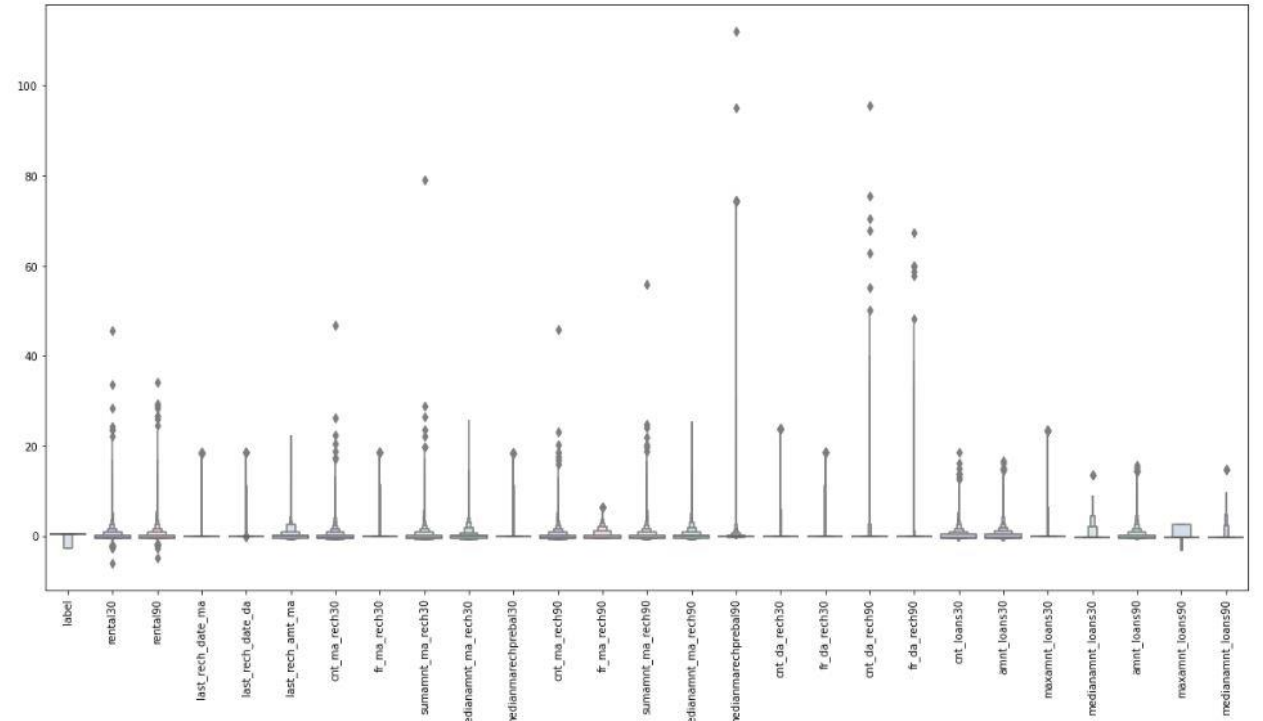
```
1  plt.figure(figsize=(10,8))
2  sns.scatterplot(df['label'],df['maxamnt_loans30'])
3  print(df[['label' , 'maxamnt_loans30']].corr())
4  plt.show()
```

```
                 label  maxamnt_loans30
label         1.000000         0.000248
maxamnt_loans30  0.000248         1.000000
```

```
1  color=["#D0DBEE", "#C2C4E2", "#EED4E5", "#D1E6DC", "#BDE2E2"]
2  plt.figure(figsize=(20,10))
3  sns.boxenplot(data=features,palette=color)
4  plt.xticks(rotation=90)
5  plt.show()
```
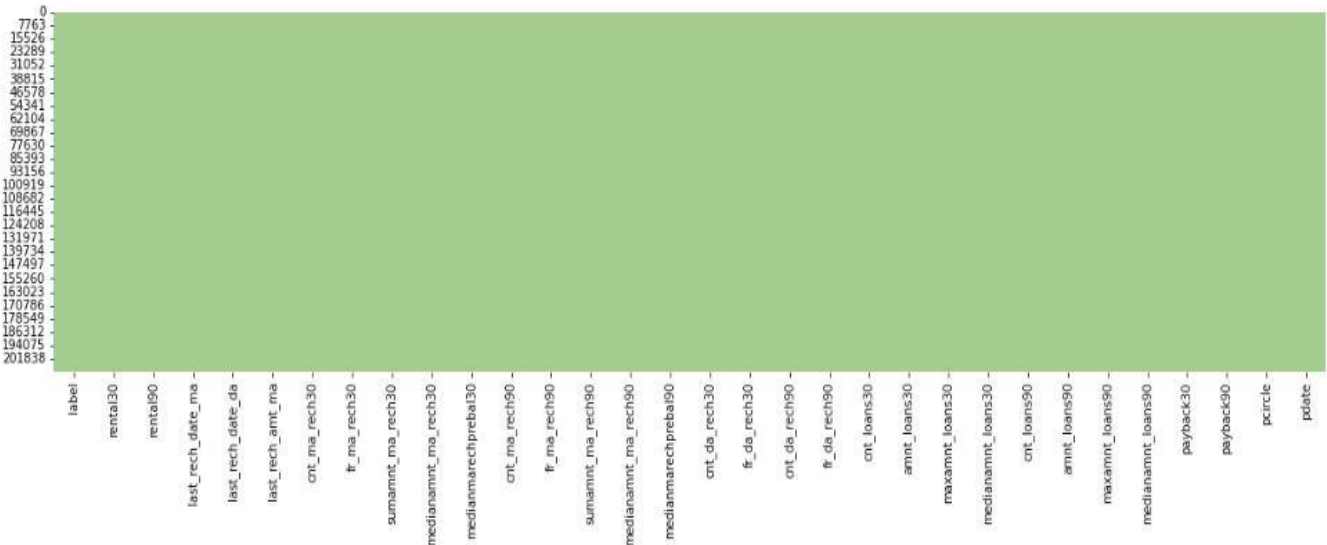
```python
# Visualizing relationship
plt.figure(figsize=(15,10), facecolor='pink')
plotnumber = 1

for column in X:
    if plotnumber<=8 :
        ax = plt.subplot(2,4,plotnumber)
        plt.scatter(X[column],y)
        plt.xlabel(column,fontsize=10)
        plt.ylabel('Label',fontsize=10)
    plotnumber+=1
plt.tight_layout()
```
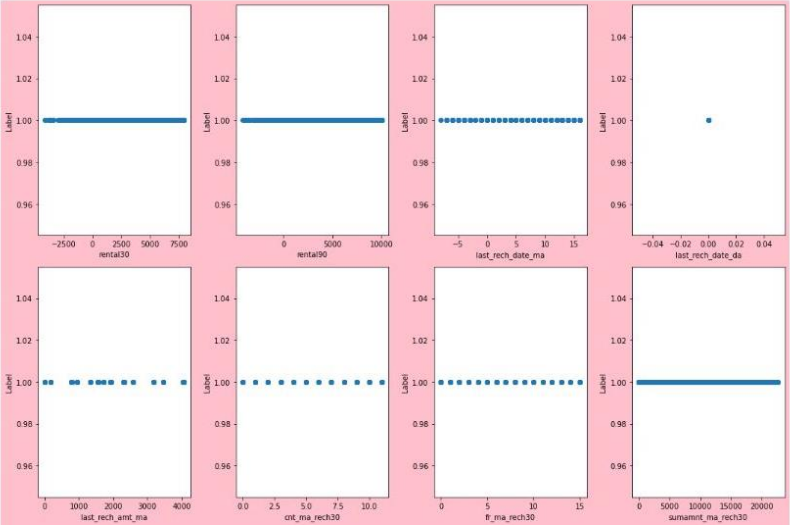


```python
plt.figure(figsize=(20,5))
sns.heatmap(df.isnull(),cbar=False,cmap='crest')
```

Out[114]: <AxesSubplot:>

STEPS AND ASSUMPTIONS TO COMPLETE THE PROJECT:

1. The model of the independent   variables and dependent variables are exactly vary with the variables.

2. It can accordingly manipulating the strategy of the areas that will yield high returns as it make easier for the   institution   in improving the client datasets in the micro-credit project.

3. It will take a good way for the management to understand the label of defaulter and non-defaulter of loan.

4. By doing visualization there are many things to be noted when it will according to work each other it means that from which aspect it is going to work when the defaulter and non-defaulter is the main target of the micro-credit   case because according to this it will predicted.

5. By preprocessing the data it means that from the micro-credit case label encoder helps the dataset column to transform to fit another column in to it.

6. Presumptions are by using regression label encoding, data scaling, precision that it means the relationship between the dependent and independent variables look fairly linear. Thus, our linearity assumption is satisfied.

MODEL DASHBOARD

Actual vs model predicted
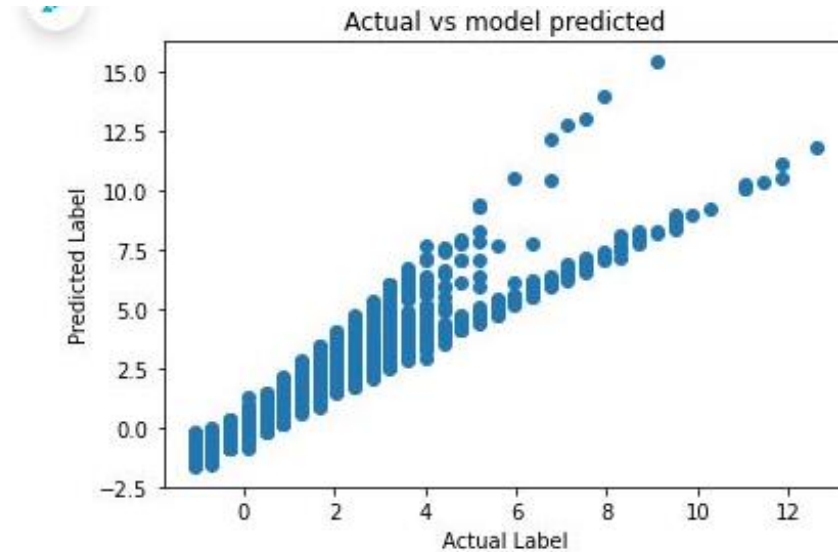
It shows the effect of the model and compares it against the null model.

FINALIZED MODEL:

the accuracy score was 87%

```
==================Train Result=====================
Accuracy Score: 87.55819105067648
_____
CLASSIFICATION REPORT :
              0         1   accuracy    macro avg   weighted avg
precision  1.000000  0.875576  0.875582    0.937788      0.891063
recall     0.000383  1.000000  0.875582    0.500192      0.875582
f1-score   0.000766  0.933661  0.875582    0.467214      0.817547
support    18261.000000 128454.000000 0.875582 146715.000000 146715.000000
_____
Cofusion Matrix:
[[    7  18254]
 [    0 128454]]


==================Test Result=====================
Accuracy: 87.43280638697159
_____
CLASSIFICATION REPORT :
              0         1   accuracy    macro avg   weighted avg
precision  0.0      0.874342  0.874328    0.437171      0.764476
recall     0.0      0.999982  0.874328    0.499991      0.874328
f1-score   0.0      0.932951  0.874328    0.466475      0.815720
support    7901.0  54977.000000 0.874328 62878.000000 62878.000000
_____
Confusion Matrix:
[[    0  7901]
 [    1 54976]]
```

CONCLUSION

I would like to conclude here that by doing the research on this project realized that the small creditors are at very large and its datasets are very lengthy the micro-finance institution has lot to do improve the condition and provide good facilities to the poor customers like Group Loans, Agricultural Loans, Individual Business Loans . By using data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data. There are many things to be noted when it will according to work each other. By preprocessing the data it means that from the micro-credit case label encoder helps the dataset column to transform to fit another column in to it. The model of the project are ready to analyse the independent and dependent variable.

THANK YOU