



Micro Credit Card Defaulter.

Submitted by:

Nomaan Sayed

**ACKNOWLEDGMENT**

This includes mentioning of all the references, research papers, data sources, professionals and other resources that helped you and guided you in completion of the project.

## **INTRODUCTION**

- **Business Problem Framing**

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.

- **Conceptual Background of the Domain Problem**

MFS is providing the loan or small credit to low income groups which should be paid within 5 days if exceeded that than the borrower is considered as a defaulter which is termed as label in this project.

- **Review of Literature**

It is very important to boost up the low income groups by helping them with some micro loans and getting their data of how many had repaid it.

- **Motivation for the Problem Undertaken**

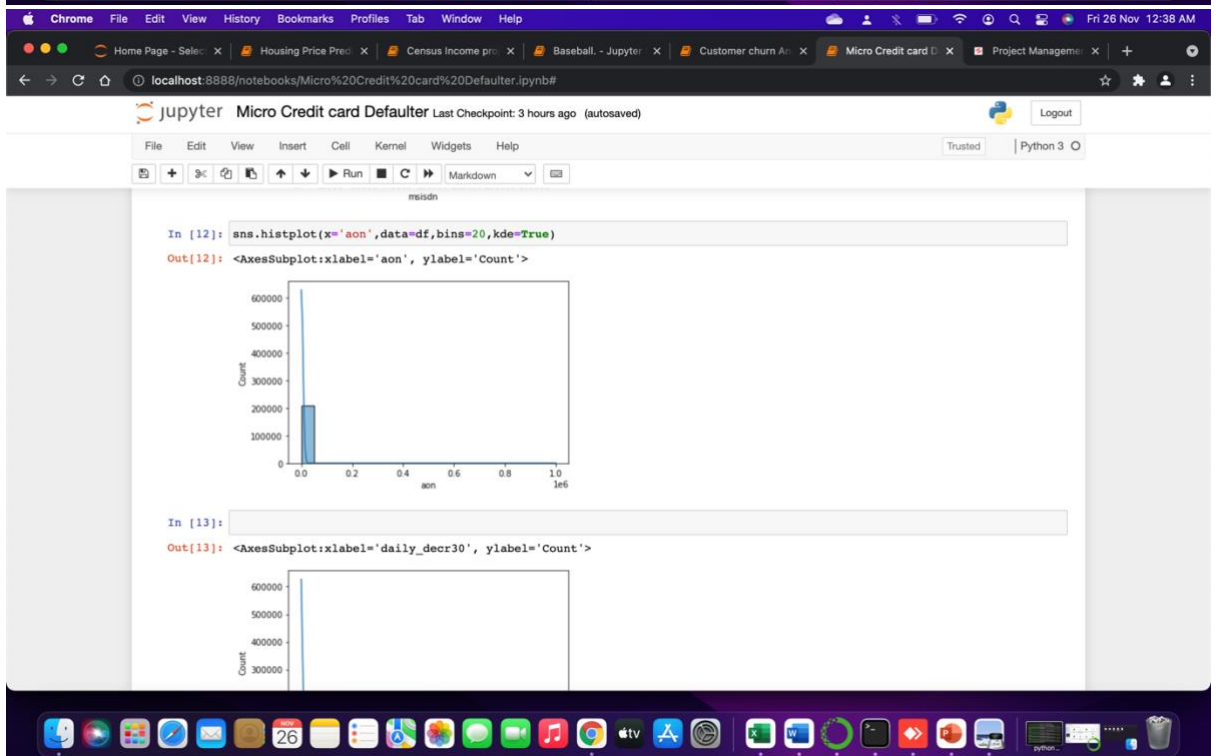
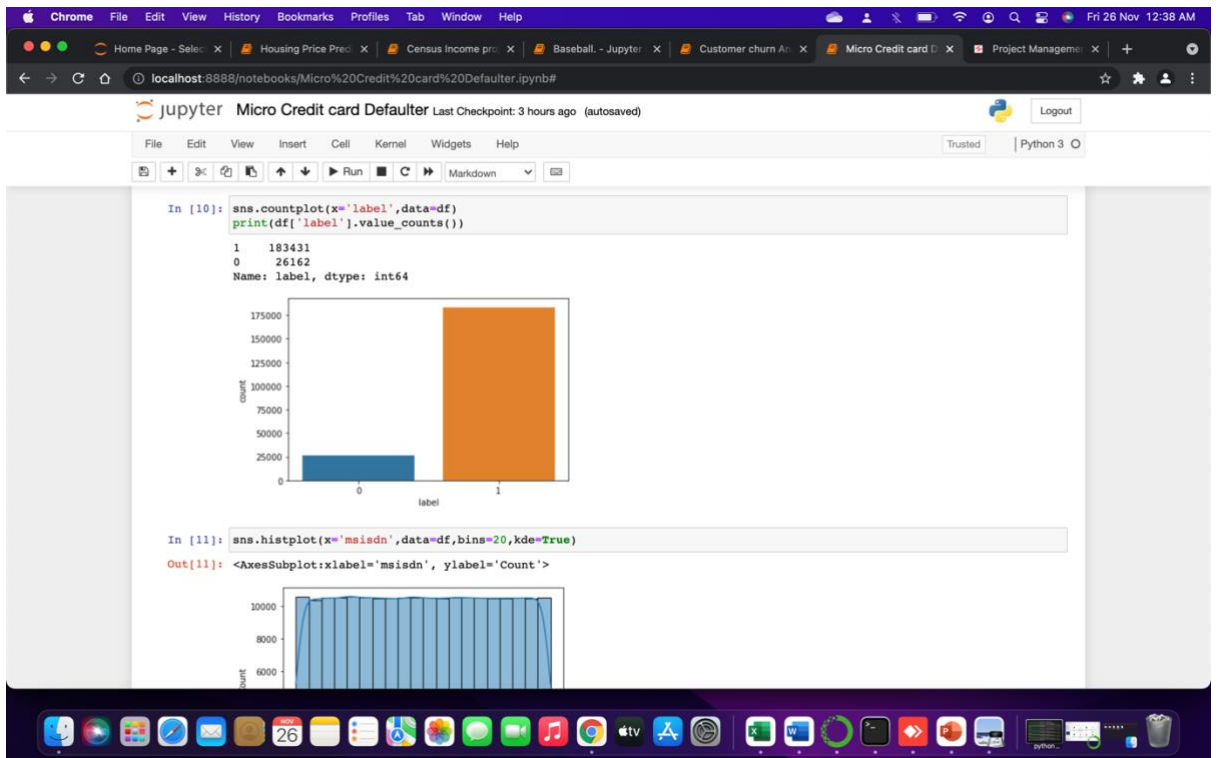
A good step to boost up the low income groups as no banks and NBFCs had ever gave them loan or credit, so their business and other things remains unchanged.

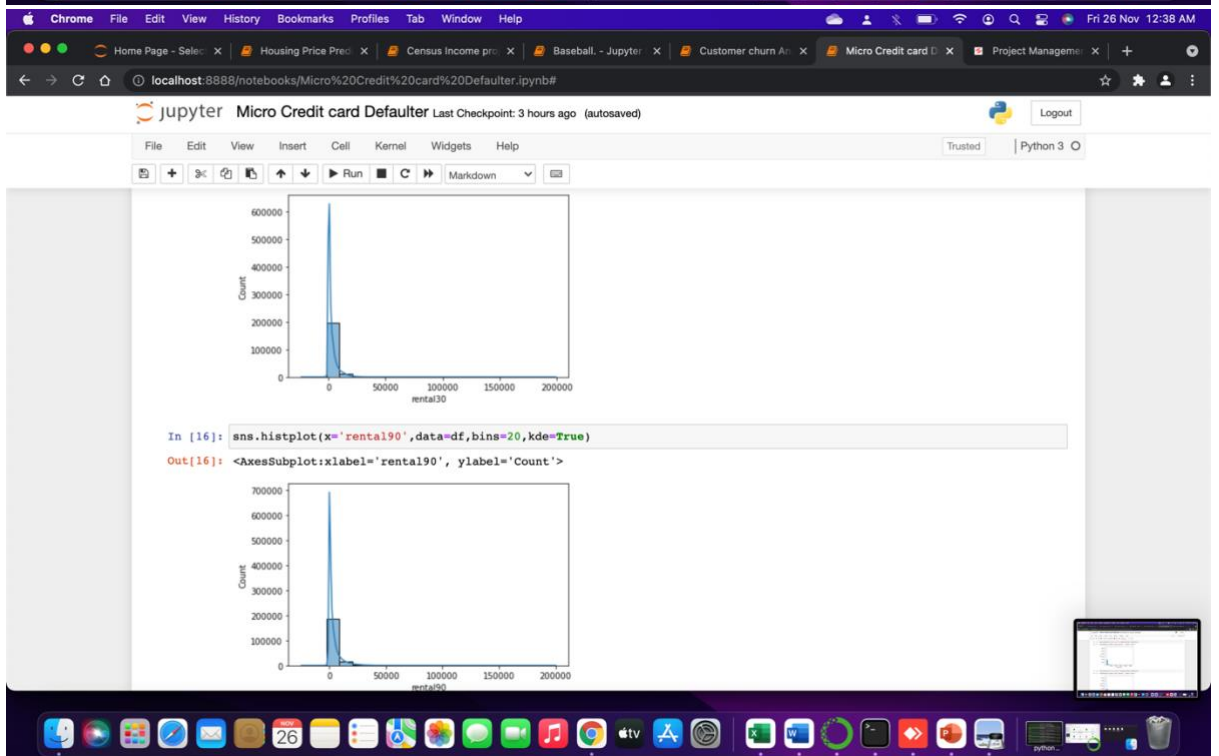
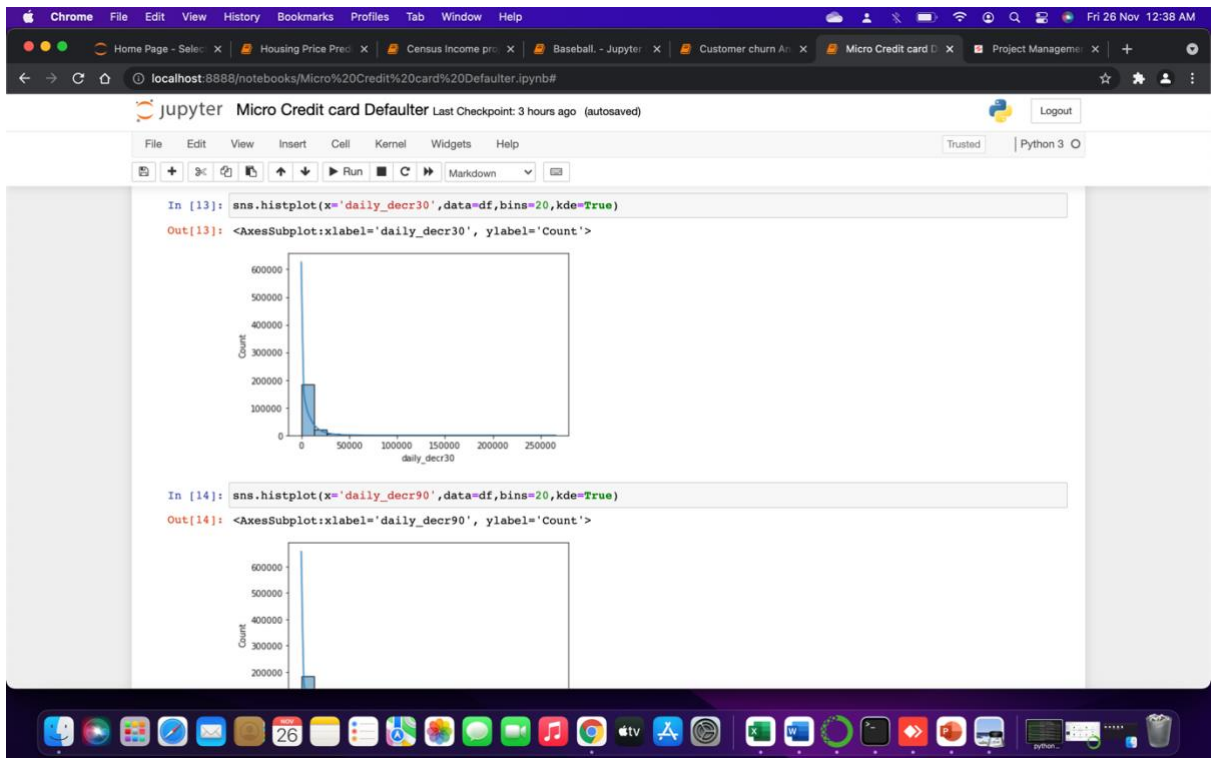
## **Analytical Problem Framing**

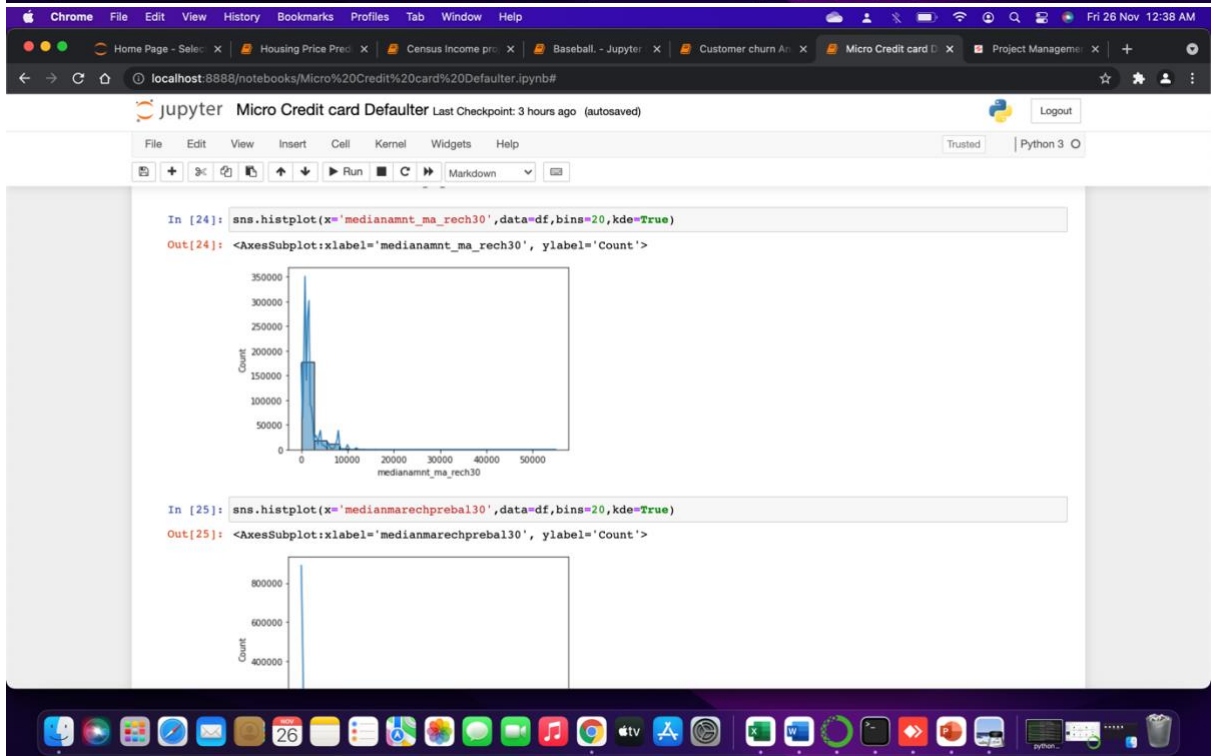
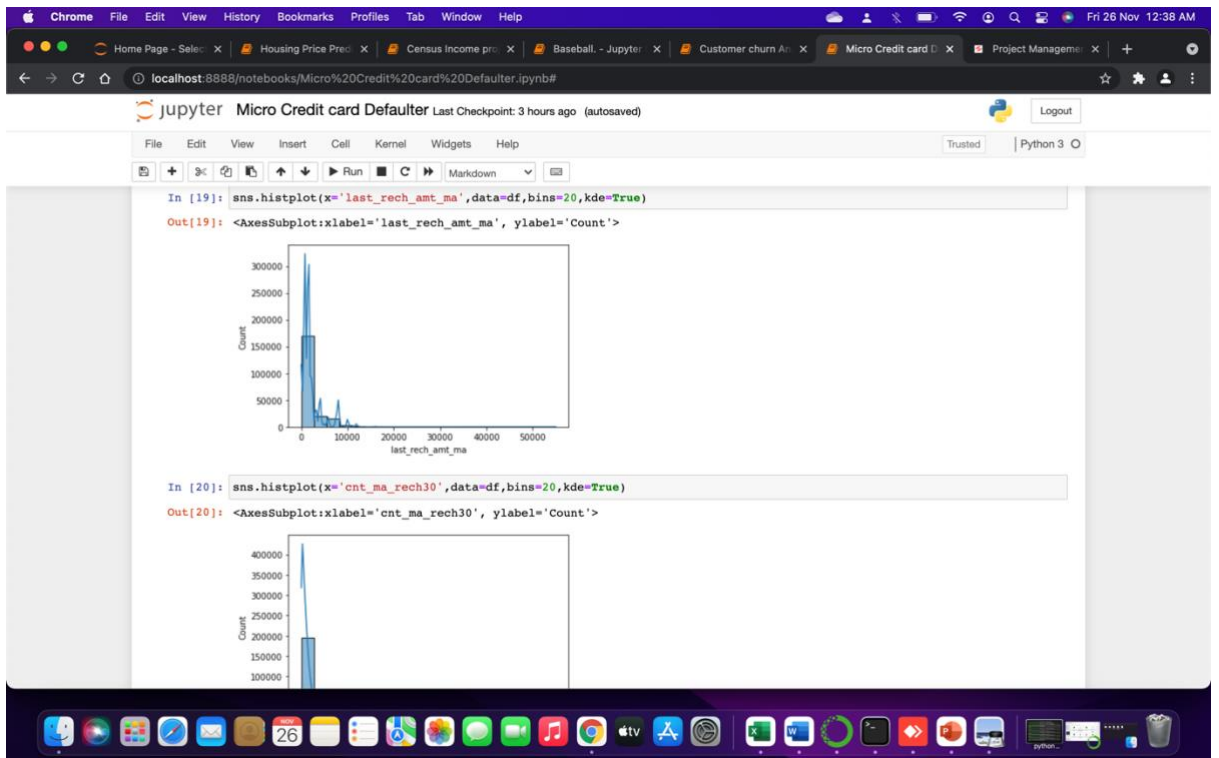
- Mathematical/ Analytical Modeling of the Problem

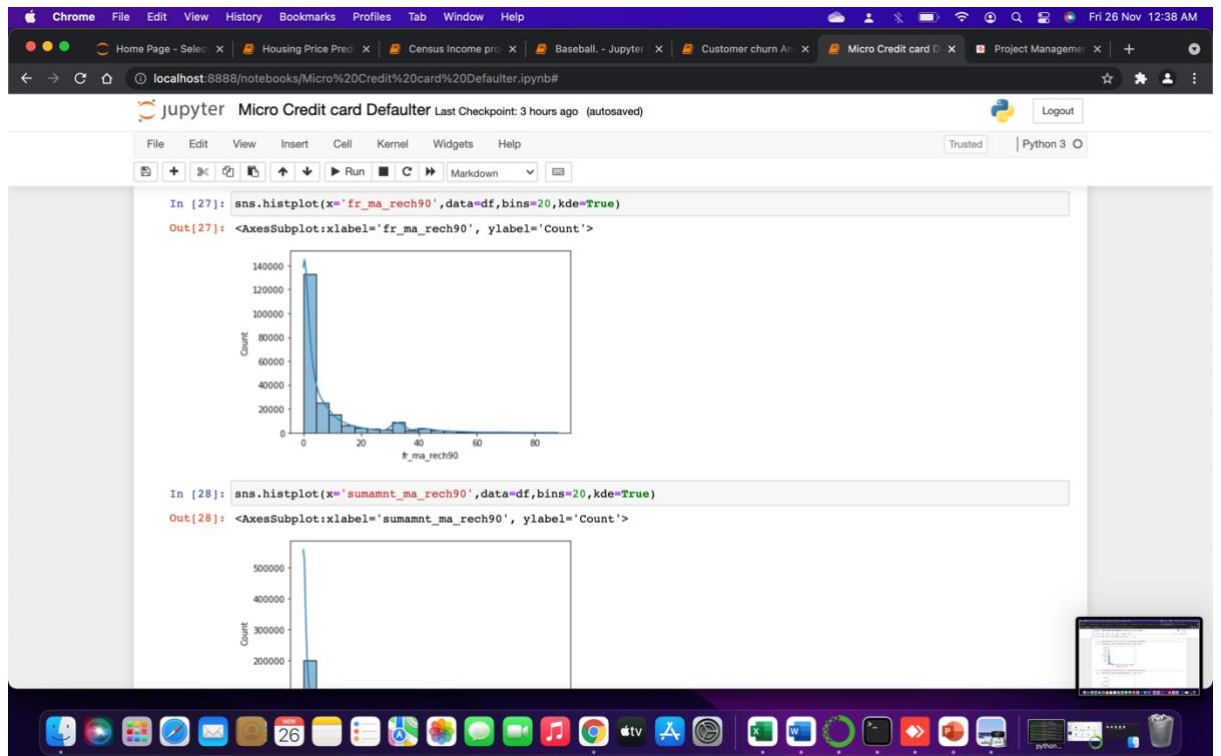
Histplot and Countplot is used in this problem to get the insights of data. Data was present in categorical and continuous type. Correlation is used to see the changes between the two data.

- Data Sources and their formats









- Data Preprocessing Done
  - 1- Used histogram and Countplot to get the insights of the data.
  - 2- Described the data and uses Correlation to get the result between two data
  - 3- Checked the outliers and skewness
  - 4- Removed the skewness using zscore
  - 5- Uses power transform to transform the data
  - 6- Scale the data as ML works much better in scaled data.



- Data Inputs- Logic- Output Relationships

In this project, label which shows how many can repay the credit or not.

- State the set of assumptions (if any) related to the problem under consideration

No such assumptions.

- Hardware and Software Requirements and Tools Used

Hardware :- Macbook Air- i5 9<sup>th</sup> gen, 8gb ram.

Software :- Jupyter Notebook and the libraries – Numpy, Pandas, Scikit, Scipy, power\_transform, Standard Scaler.

## **Model/s Development and Evaluation**

- Identification of possible problem-solving approaches (methods)

Firstly, we had seen the insights and cleaned the data using various techniques and finally used the logistic regression model as the label is categorical data.

Accuracy score was used to evaluate the performance of logistic regression model.

- Testing of Identified Approaches (Algorithms)

Logistic Regression

# Accuracy Score

- Run and Evaluate selected models

The image displays two screenshots of a Jupyter Notebook interface, likely running on a local host (localhost:8888). The notebook is titled "Micro Credit card Defaulter" and shows the process of model selection and evaluation.

**Model Selection**

In [99]:

```
from sklearn.linear_model import LogisticRegression
lr=LogisticRegression()
from sklearn.svm import SVC
svc=SVC(kernel='rbf')
from sklearn.tree import DecisionTreeClassifier
dtc=DecisionTreeClassifier()
from sklearn.naive_bayes import MultinomialNB
mnb=MultinomialNB()
from sklearn.neighbors import KNeighborsClassifier
knc=KNeighborsClassifier()
from sklearn.metrics import accuracy_score
import warnings
warnings.filterwarnings('ignore')
from sklearn.model_selection import train_test_split
```

In [100]:

```
for i in range(0,1000):
    x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=i)
    lr.fit(x_train,y_train)
    pred_train=lr.predict(x_train)
    pred_test=lr.predict(x_test)
    if round(accuracy_score(y_train,pred_train)*100,1)!=round(accuracy_score(y_test,pred_test)*100,1):
        print("At random state ",i,"The Model perform very well")
        print("At random state :-",i)
        print("Training accuracy_score is :-",accuracy_score(y_train,pred_train)*100)
        print("Testing accuracy_score is :-",accuracy_score(y_test,pred_test)*100)
```

At random state 4 The Model perform very well  
At random state :- 4  
Training accuracy\_score is :- 87.5979314402502  
Testing accuracy\_score is :- 87.57315826959403

**Cross Validation of Dataset**

In [101]:

```
from sklearn.metrics import classification_report
print(classification_report(y_test,pred_test))
```

	precision	recall	f1-score	support
0	0.14	0.06	0.08	4517
1	0.86	0.95	0.90	27776
accuracy			0.82	32293
macro avg	0.50	0.50	0.49	32293
weighted avg	0.76	0.82	0.79	32293

In [102]:

```
pred_lr=lr.predict(x_test)
from sklearn.model_selection import cross_val_score
lss=accuracy_score(y_test,pred_lr)
for j in range(2,10):
    lscscore=cross_val_score(lr,x,y,cv=j)
    lsc=lscscore.mean()
    print("At cv :- ",j)
    print("Cross validation score is :-",lsc*100)
    print("accuracy_score is :-",lss*100)
    print("\n")
```

At cv :- 2  
Cross validation score is :- 87.52547108932274  
accuracy\_score is :- 87.52670857461369

At cv :- 3  
Cross validation score is :- 87.57563623010599

Chrome File Edit View History Bookmarks Profiles Tab Window Help

Home Page - Sale... x Housing Price Pred... x Census Income pr... x Baseball. - Jupyter x Customer churn A... x Micro Credit card D... x Project Managem... x +

localhost:8888/notebooks/Micro%20Credit%20card%20Defaulter.ipynb#

jupyter Micro Credit card Defaulter Last Checkpoint: 3 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

### Cross Validation of Dataset

```
In [102]: pred_lr=lr.predict(x_test)
from sklearn.model_selection import cross_val_score
lss=accuracy_score(y_test,pred_lr)
for j in range(2,10):
    lsscore=cross_val_score(lr,x,y,cv=j)
    lsc=lsscore.mean()
    print("At cv i:- ",j)
    print("Cross validation score is :-",lsc*100)
    print("accuracy_score is :-",lss*100)
    print("\n")

At cv i:- 2
Cross validation score is :- 87.52547108932274
accuracy_score is :- 87.52670857461369

At cv i:- 3
Cross validation score is :- 87.57563623010599
accuracy_score is :- 87.52670857461369

At cv i:- 4
Cross validation score is :- 87.59483579742839
accuracy_score is :- 87.52670857461369

At cv i:- 5
Cross validation score is :- 87.596073452451
accuracy_score is :- 87.52670857461369

At cv i:- 6
```

Mac OS dock with various application icons.

Chrome File Edit View History Bookmarks Profiles Tab Window Help

Home Page - Sale... x Housing Price Pred... x Census Income pr... x Baseball. - Jupyter x Customer churn A... x Micro Credit card D... x Project Managem... x +

localhost:8888/notebooks/Micro%20Credit%20card%20Defaulter.ipynb#

jupyter Micro Credit card Defaulter Last Checkpoint: 3 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```
At cv i:- 5
Cross validation score is :- 87.596073452451
accuracy_score is :- 87.52670857461369

At cv i:- 6
Cross validation score is :- 87.57873351176075
accuracy_score is :- 87.52670857461369

At cv i:- 7
Cross validation score is :- 87.58430757151353
accuracy_score is :- 87.52670857461369

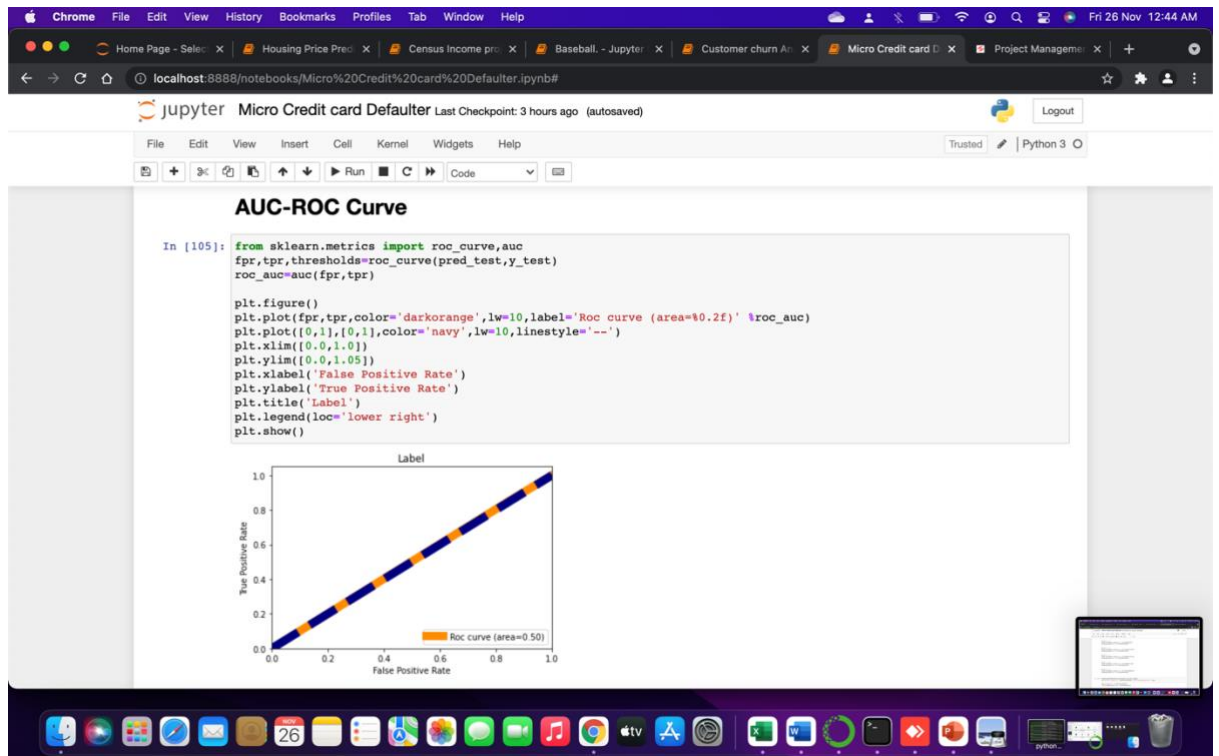
At cv i:- 8
Cross validation score is :- 87.5768756111465
accuracy_score is :- 87.52670857461369

At cv i:- 9
Cross validation score is :- 87.59793329489169
accuracy_score is :- 87.52670857461369

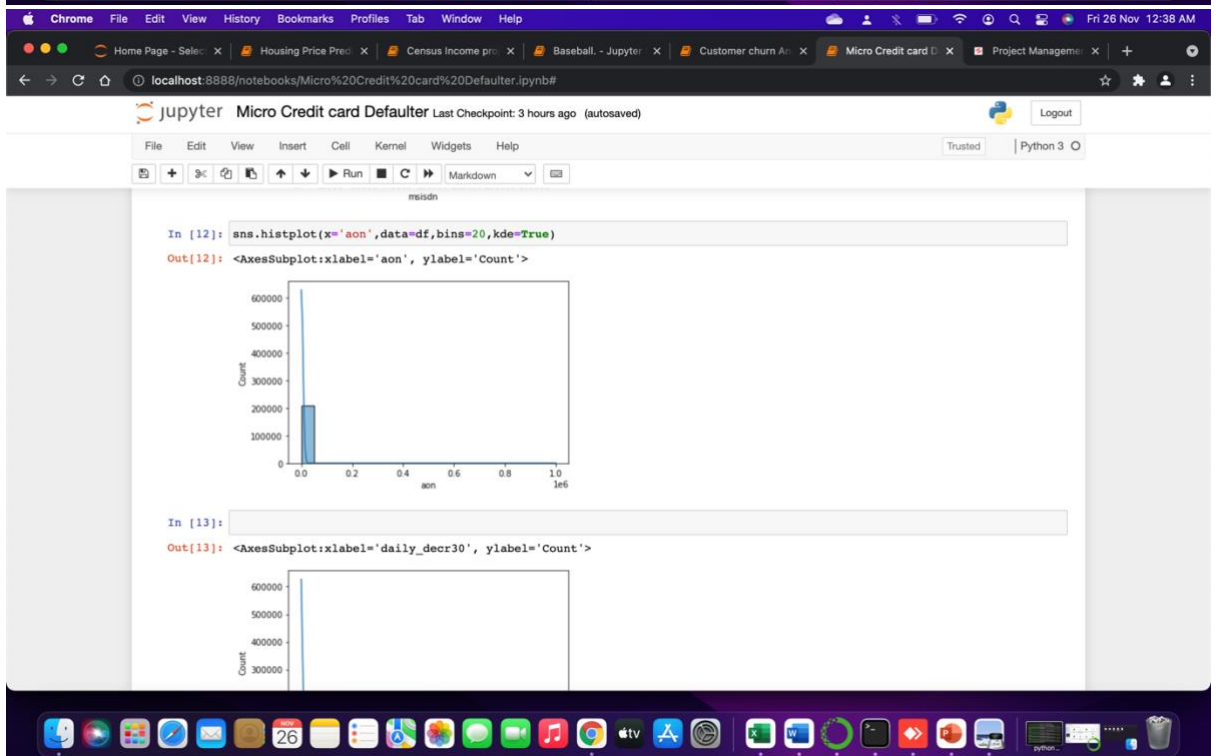
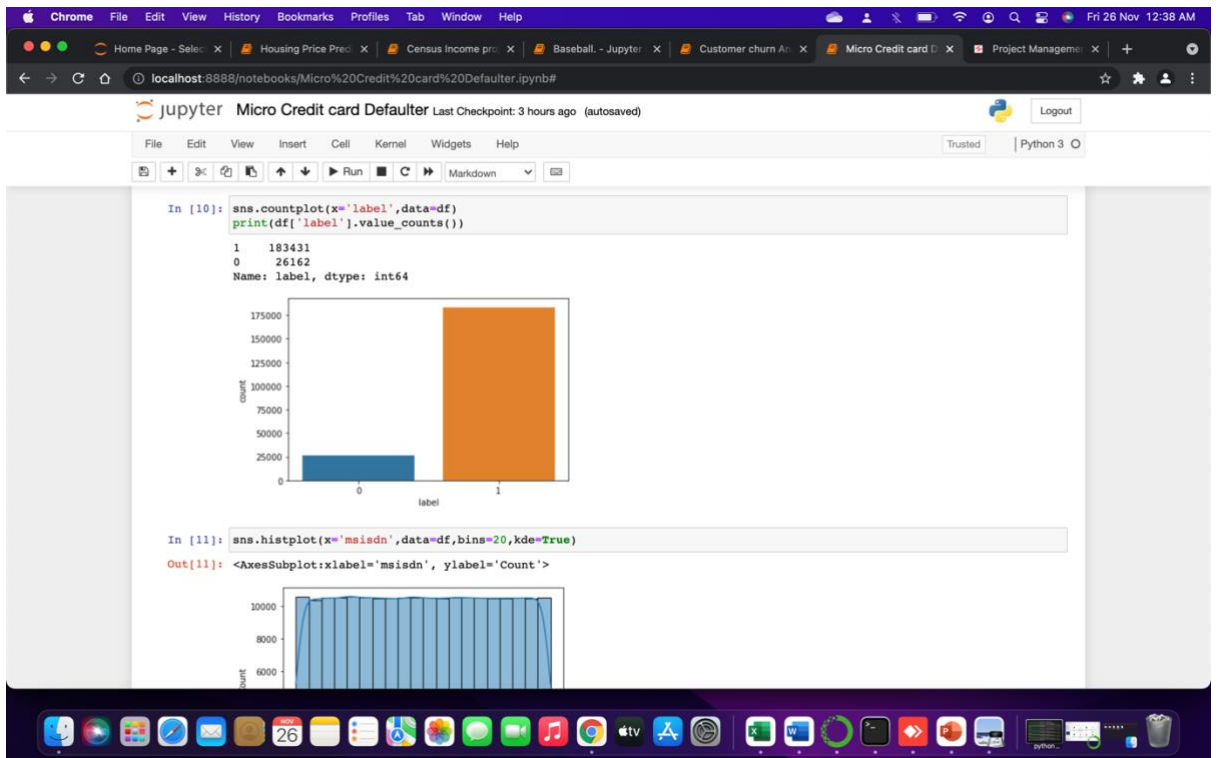
In [103]: lsscore_selected=cross_val_score(lr,x,y,cv=2).mean()
print("The cv score is: ",lsscore_selected,"\n The accuracy_score is: ",lss)

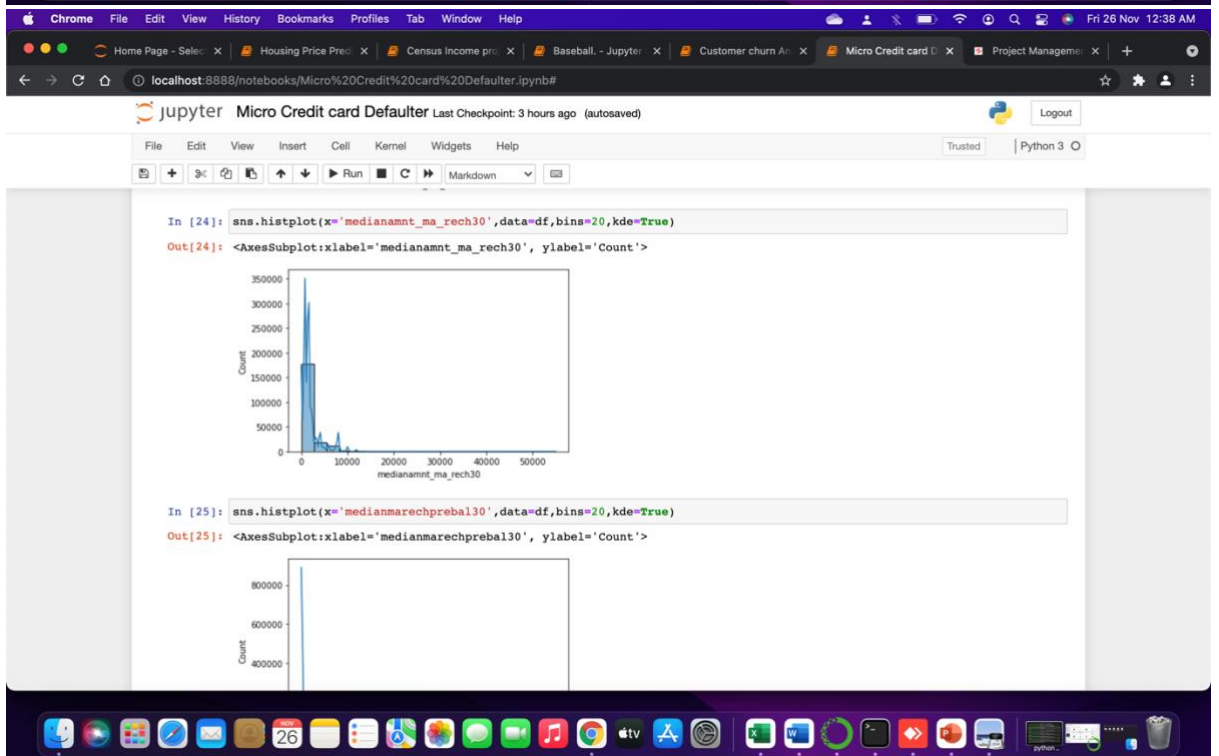
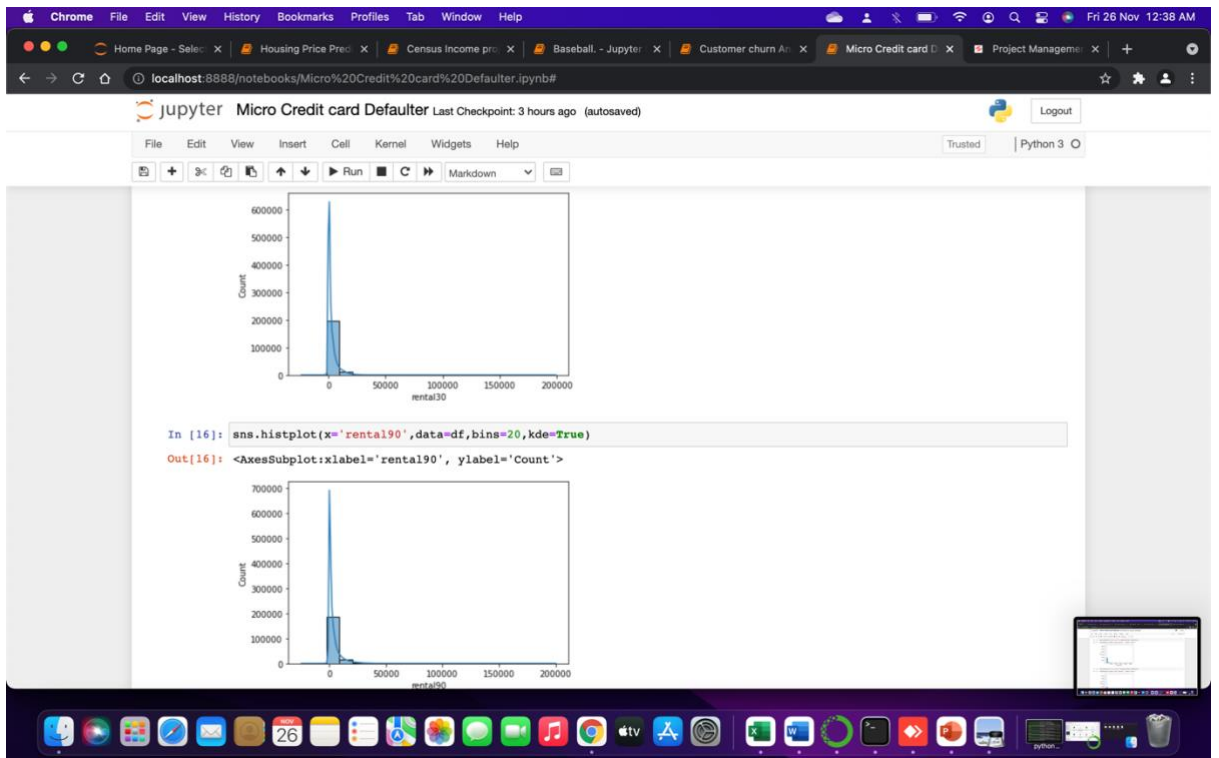
The cv score is: 0.8752547108932274
The accuracy_score is: 0.875267085746137
```

Mac OS dock with various application icons.

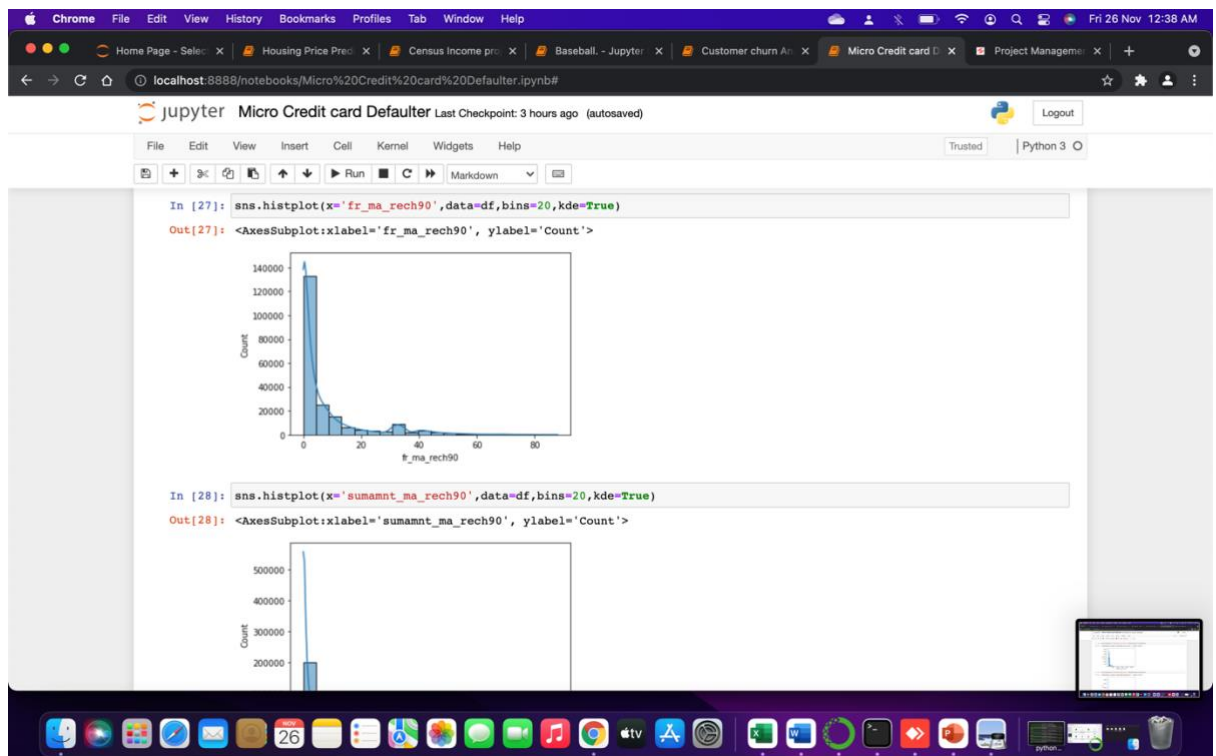


- Key Metrics for success in solving problem under consideration  
Accuracy score metrics is used.
- Visualizations









- Interpretation of the Results

From the visualizations we can make out the exact insights of the data and can decide on how to remove the outliers and clean up the data. From the model itself we can get the accuracy of the data.

## CONCLUSION

- Key Findings and Conclusions of the Study

Model is working good and giving the best results from our assumptions.

- Learning Outcomes of the Study in respect of Data Science

Before visualization it is very difficult to come to the conclusion just by seeing the data, as it is very huge. But

using visualization it shows the insights which we cannot assume and further there are many unwanted data which should be removed so using the data cleaning it makes it pure and gives the best results. It is like finding the needle in grass using a magnet. Data cleaning works as a magnet.

- Limitations of this work and Scope for Future Work

It gives the result for either paid the credit or not but it doesn't helps us how to get the solution of non-repayment.