

Micro-Credit Defaulter Model

Submitted by:

Deepak Tripathi

**ACKNOWLEDGMENT**

We are using this opportunity to express my gratitude to everyone who supported us throughout the course of this group project.

We are thankful for their aspiring guidance, invaluably constructive criticism and friendly advice during the project work.

I am sincerely grateful to them for sharing their truthful and illuminating views on a number of issues related to the project.

Further, we were fortunate to have **Khushboo Garg** as our mentor.

He has readily shared his immense knowledge in data analytics and guide us in a manner that the outcome resulted in enhancing our data skills.

**INTRODUCTION**

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 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.

Business Problem Framing

This is a classic Business problem which helps Micro Financing Institutions and other Lending companies reduce Credit risks by recognizing potential Defaulters.

Review of Literature

The Microfinance revolution in India as a powerful tool for poverty alleviation and women

empowerment. He stated that the viability of microfinance institutions are under question mark and the

outreach is too small and there is a need for an all round effort to help develop the fledgling

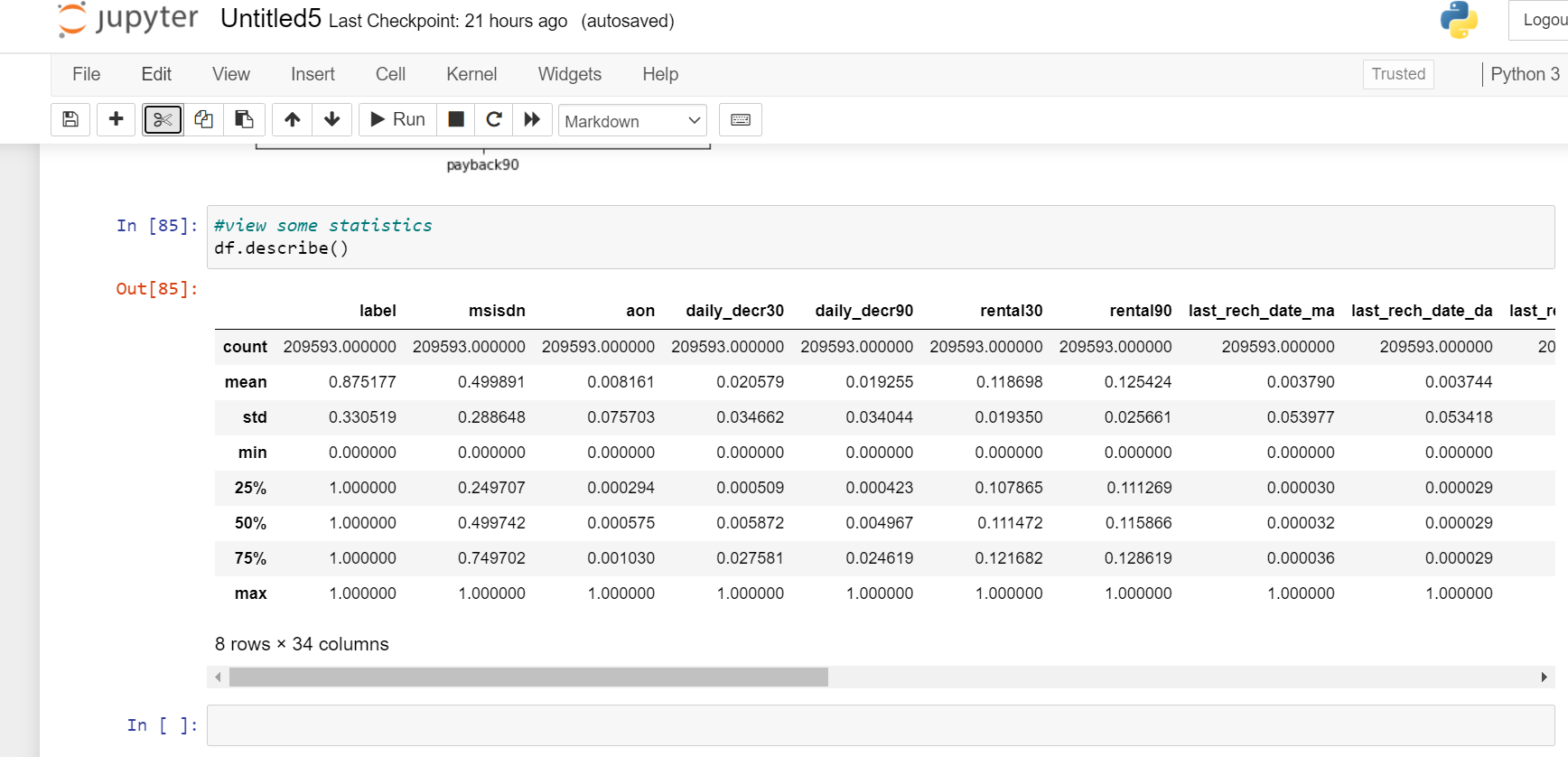
Microfinance Industry while tackling the trade off between outreach and sustainability.

**Analytical Problem Framing**

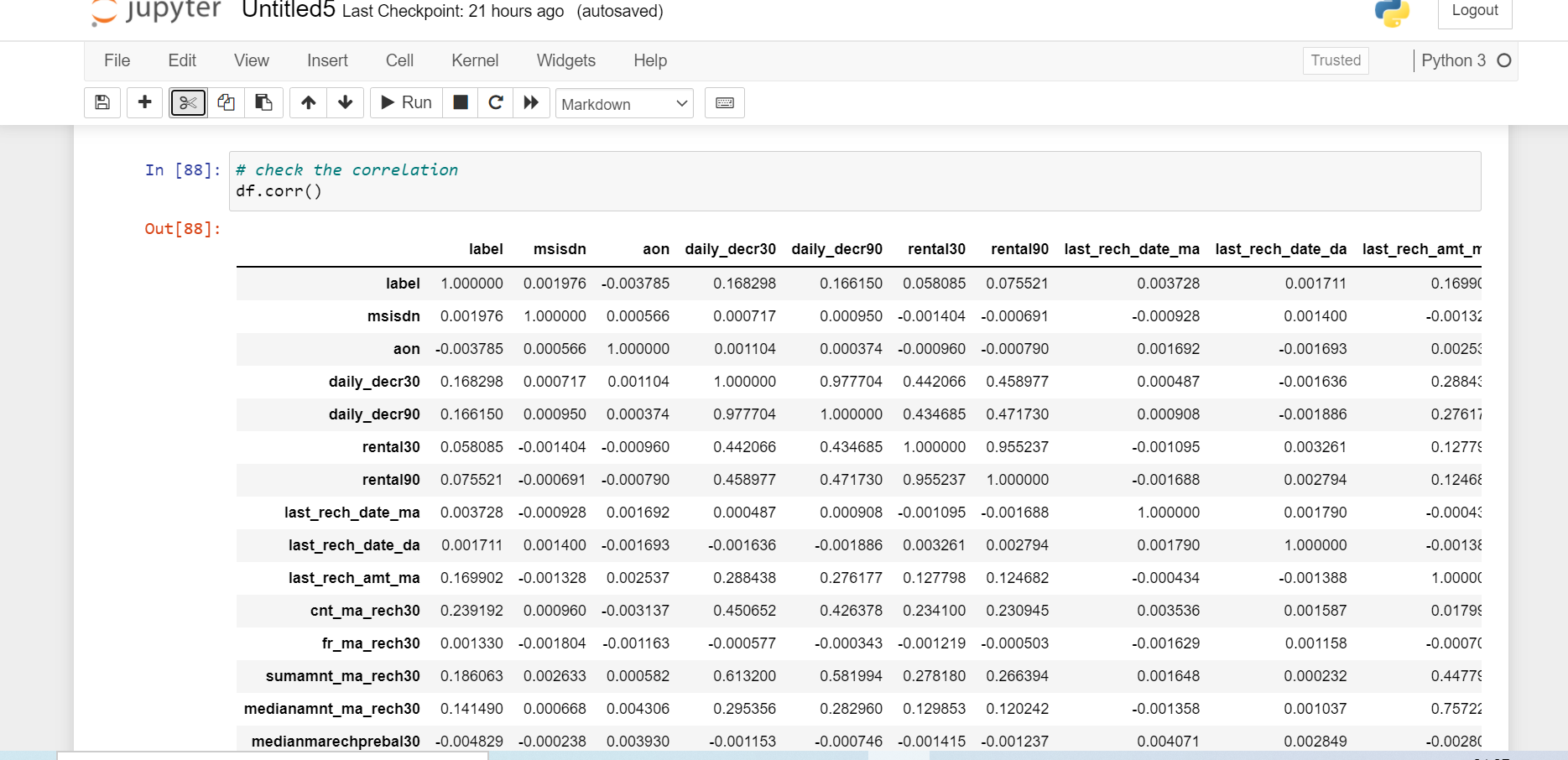
Mathematical/ Analytical Modelling of the Problem

**Statistical**: - By using describe method we are able to find out the mean, median, count, std, min, 25% ,50% ,75% ,max.

Through which we are able to compare the data. Easy to find out the data spread.



Correlation:- **Correlational to** **a non-experimental research method which studies the relationship between two variables with the help of statistical analysis.**



* Data Sources and their formats

Data Sources :-Data Source provided by the flip robo.

* Data Pre-processing Done

We need to import some libraries for the read, for the visualization, for converting string into numeric form.

import pandas as pd

import numpy as np

import matplotlib as mpl

import matplotlib.pyplot as plt

import matplotlib.figure as fig

import seaborn as sns

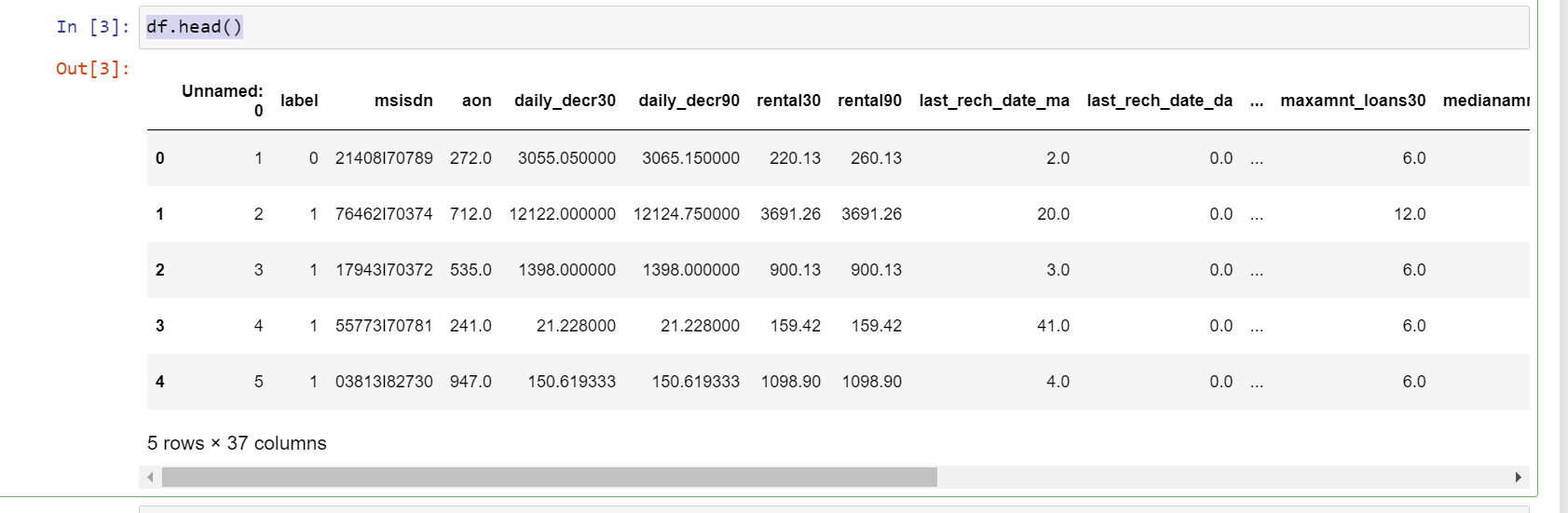
import sklearn.preprocessing.LabelEncoder

import warnings

import pandas as pd

df = pd.read\_excel('micro.xlsx')

df.head()

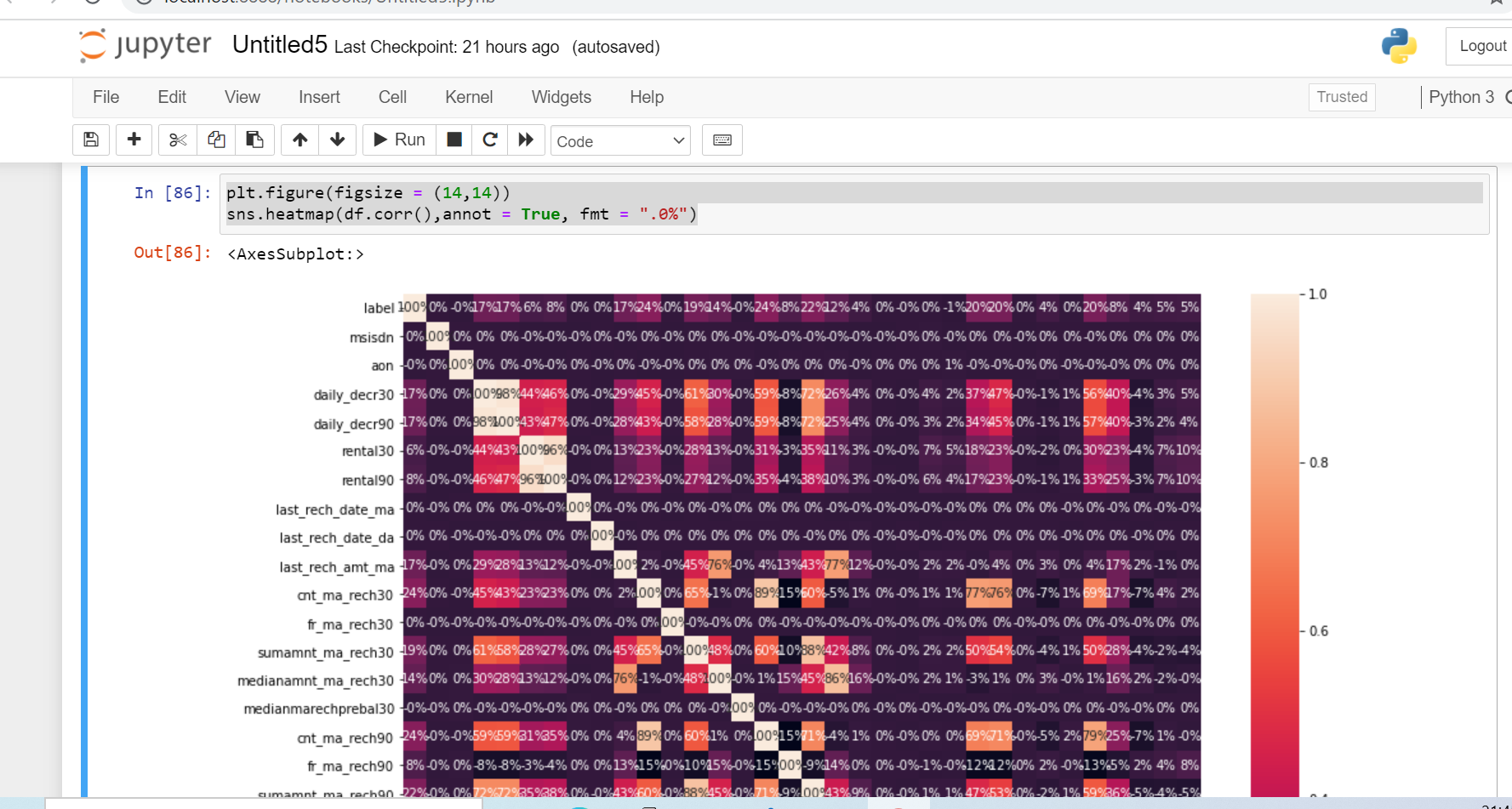


import matplotlib.pyplot as plt

Heatmap :-

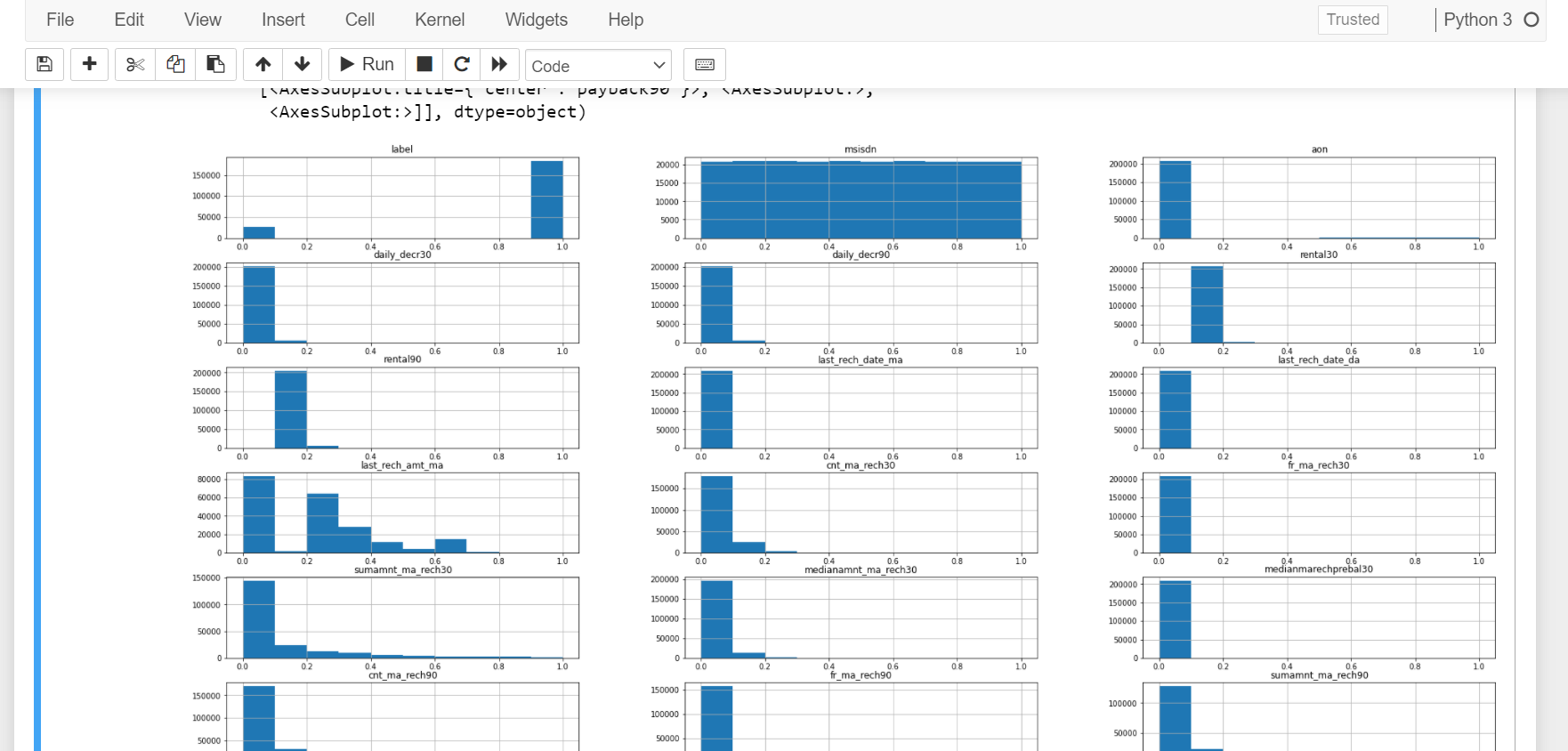
plt.figure(figsize = (14,14))

sns.heatmap(df.corr(),annot = True, fmt = ".0%")

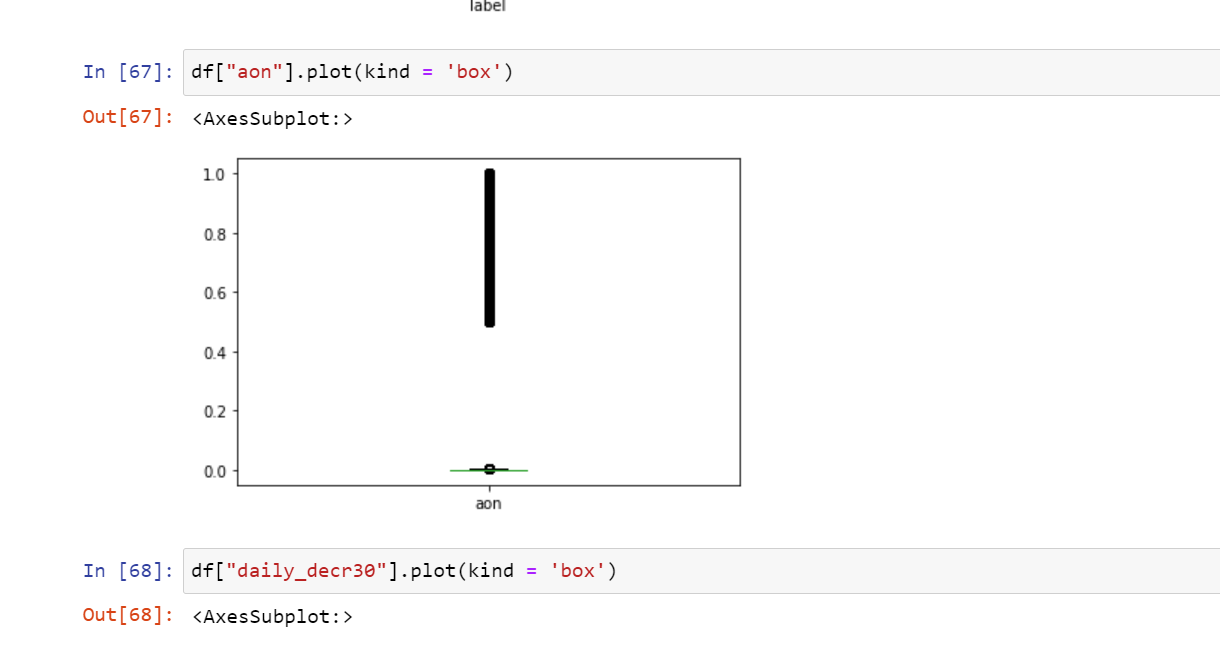
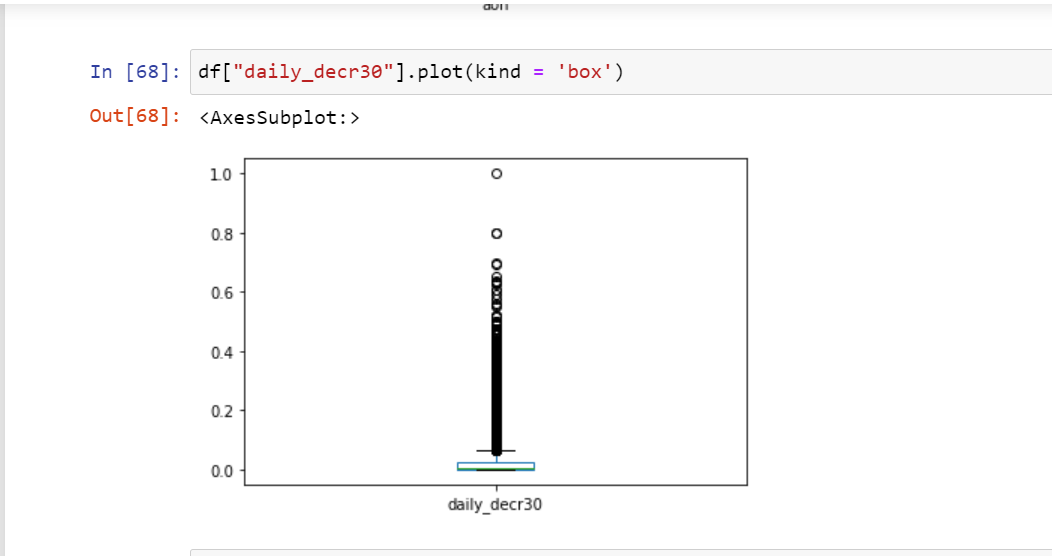
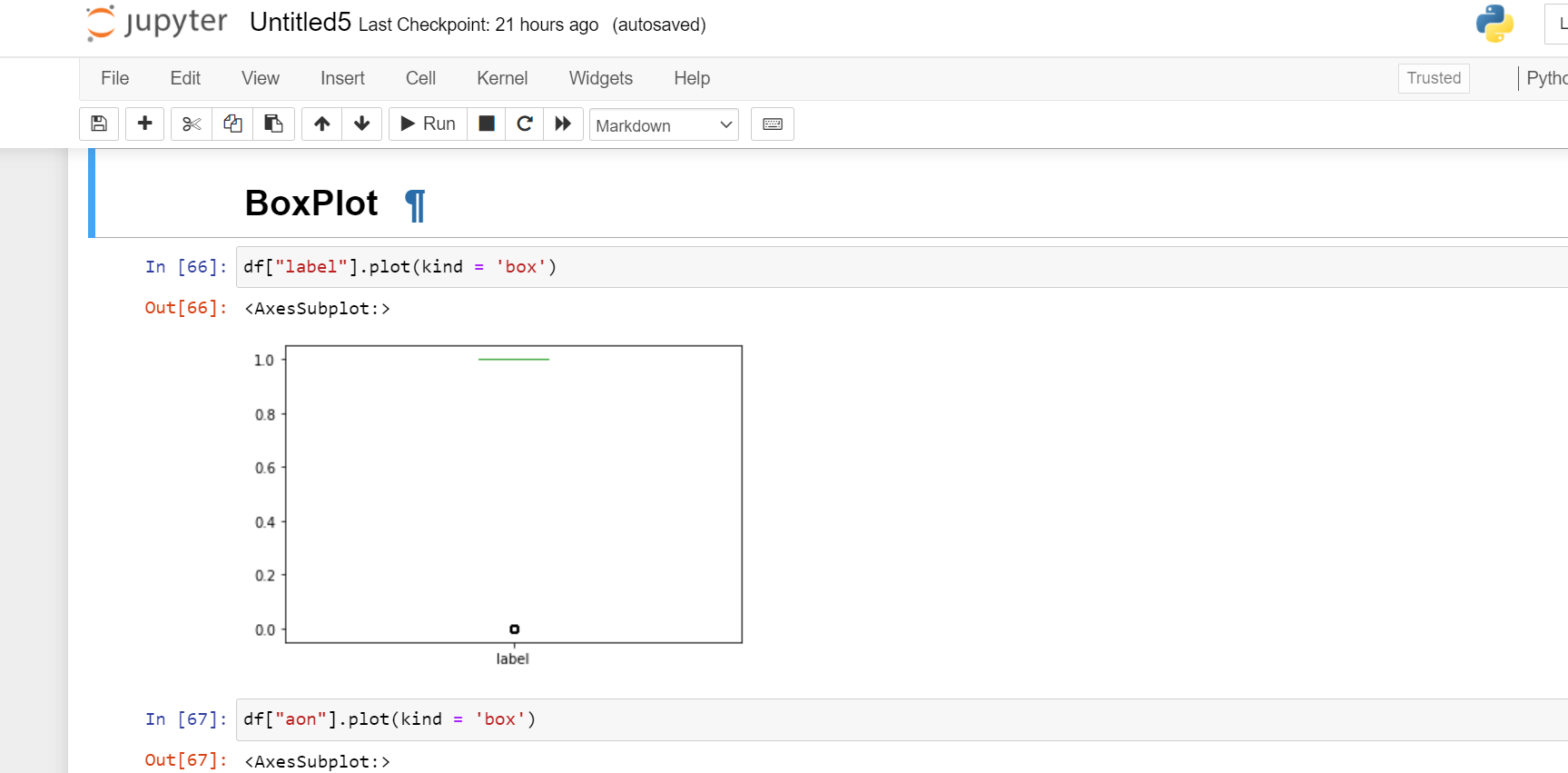


**Histogram**

df.hist(figsize=(30,30),layout=(12,3),sharex=False)



# BoxPlot



# Label Encoder

# Label encoder convert the string into numerical data.

# from sklearn.preprocessing import LabelEncoder

# le = LabelEncoder()

# for column in df.columns:

# if df[column].dtype == np.number:

# continue

# df[column] = LabelEncoder().fit\_transform(df[column])

# MinMaxScaler

ransform features by scaling each feature to a given range.

This estimator scales and translates each feature individually such that it is in the given range on the training set, e.g. between zero and one.

from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler()

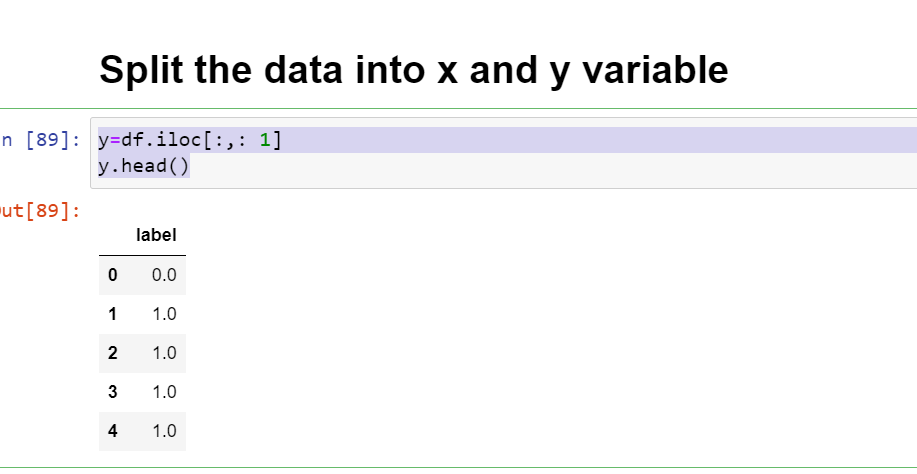
scaled\_values = scaler.fit\_transform(df)

df.loc[:,:] = scaled\_values

.

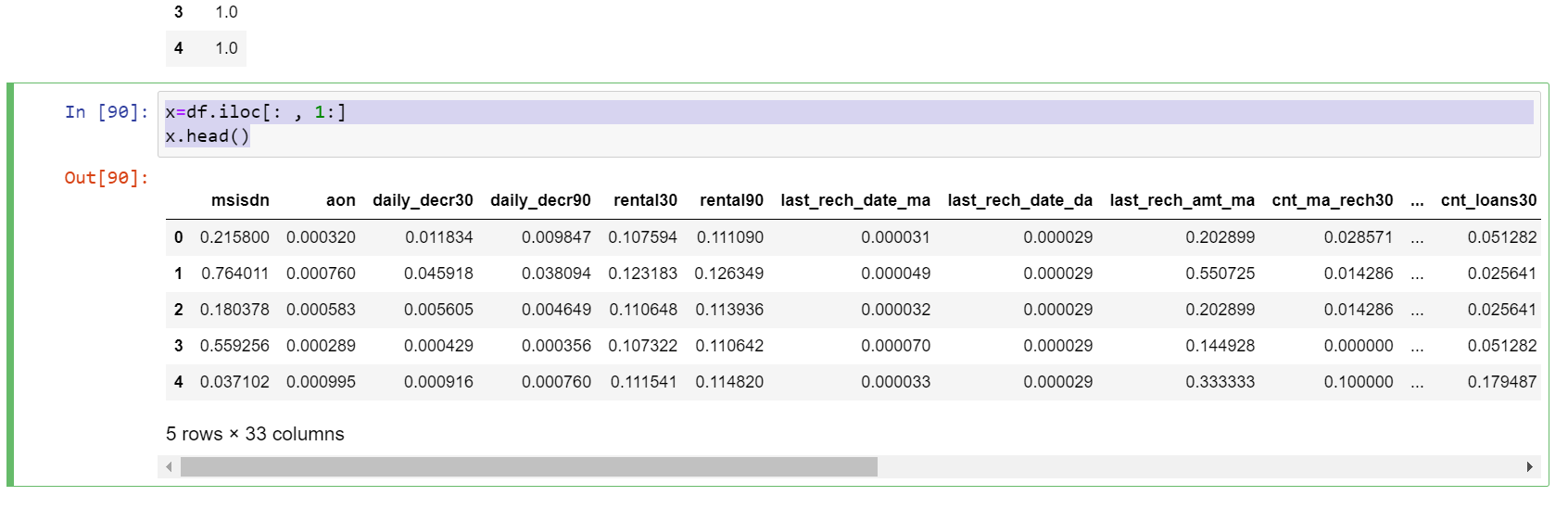
* **Data Inputs- Logic- Output Relationships**

# Split the data into x and y variable



# x=df.iloc[: , 1:]

# x.head()



**Model/s Development and Evaluation**

# Split the data into Training and testing.

# from sklearn.model\_selection import train\_test\_split

# x\_train, x\_test, y\_train, y\_test = train\_test\_split(x,y,test\_size=30, random\_state= 0)

# LogisticRegression

# from sklearn.linear\_model import LogisticRegression

# lg = LogisticRegression()

# lg.fit(x\_train,y\_train)

# lg.score(x\_train,y\_train)

# pred=lg.predict(x\_test)

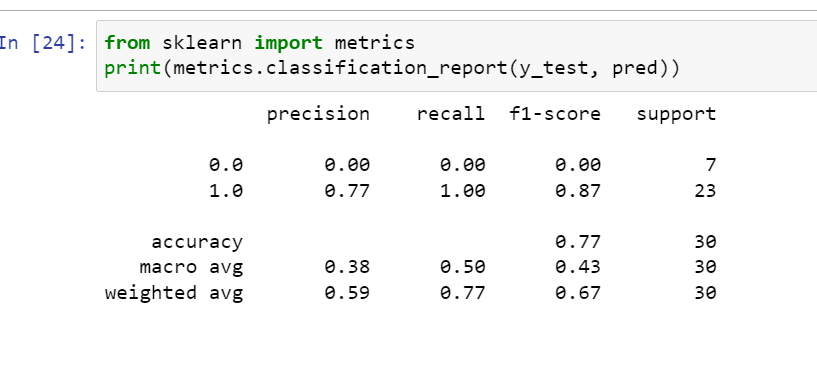
# from sklearn.metrics import accuracy\_score

# print(round(accuracy\_score(y\_test,pred),2))

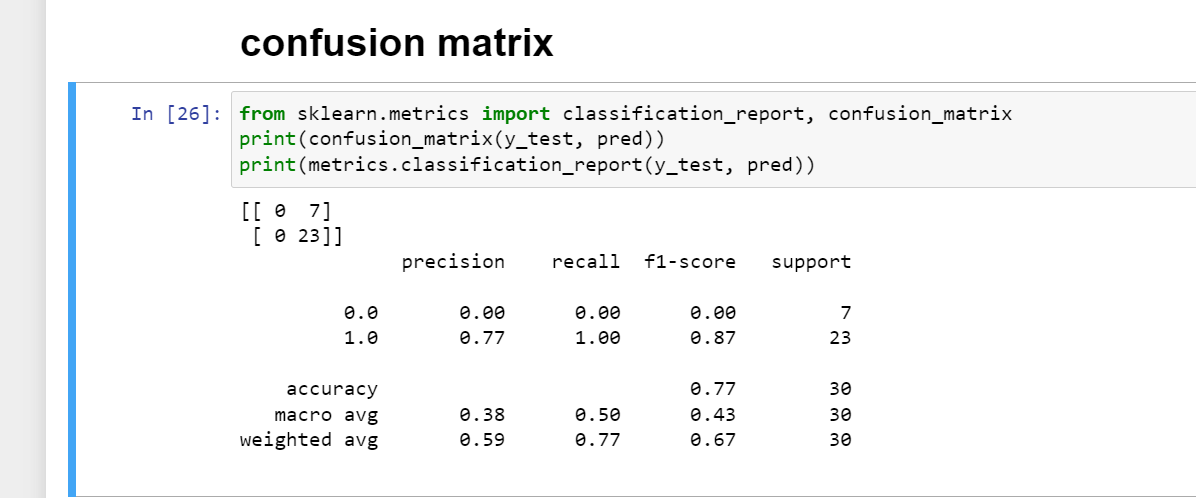
# Score of logistic regression is 77 percent.



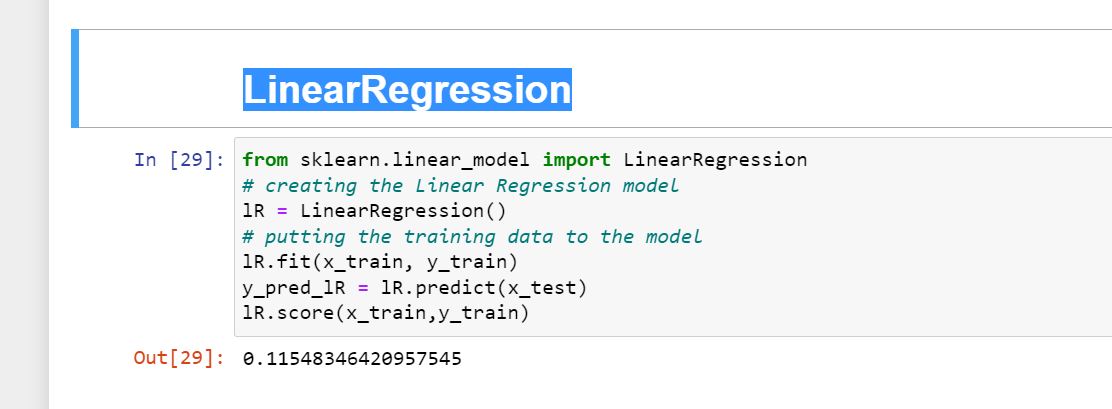
# Metrics



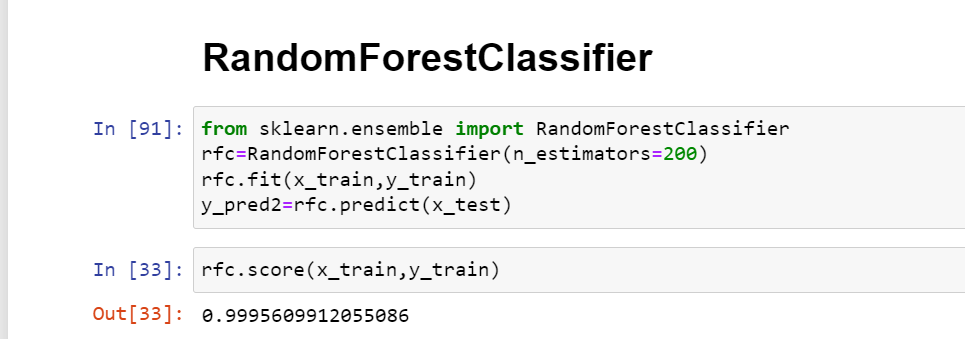
# confusion matrix



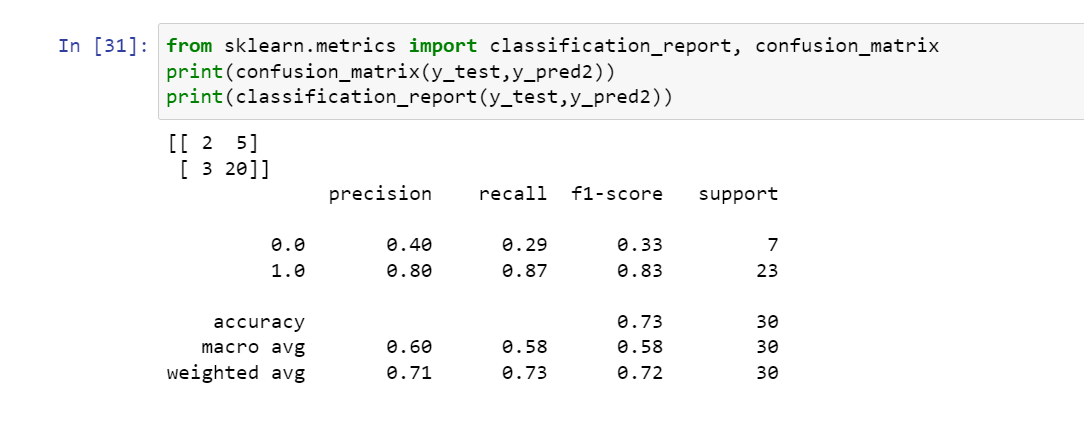
# LinearRegression



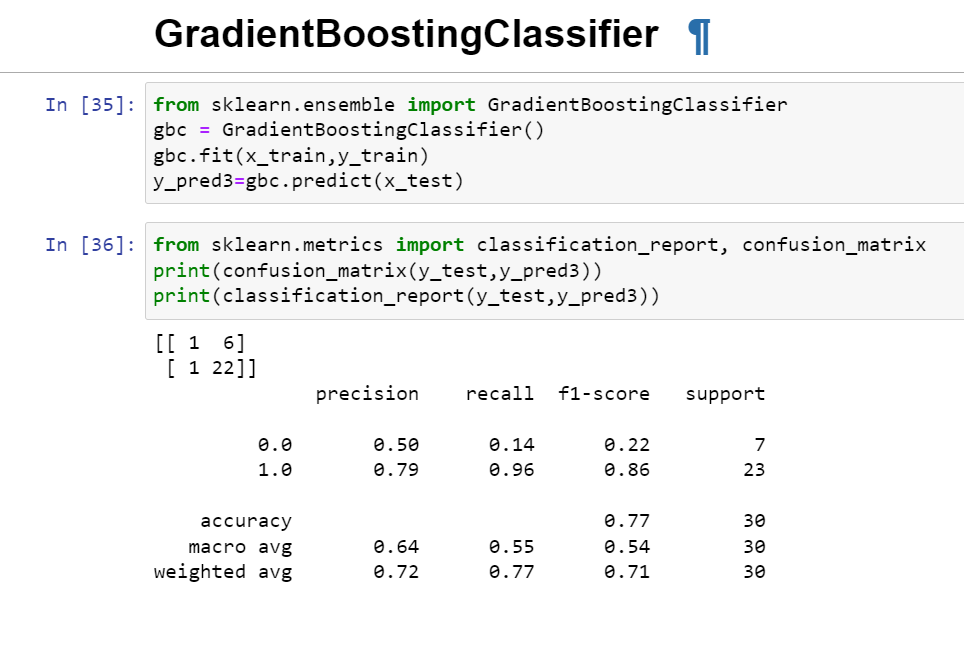
# RandomForestClassifier



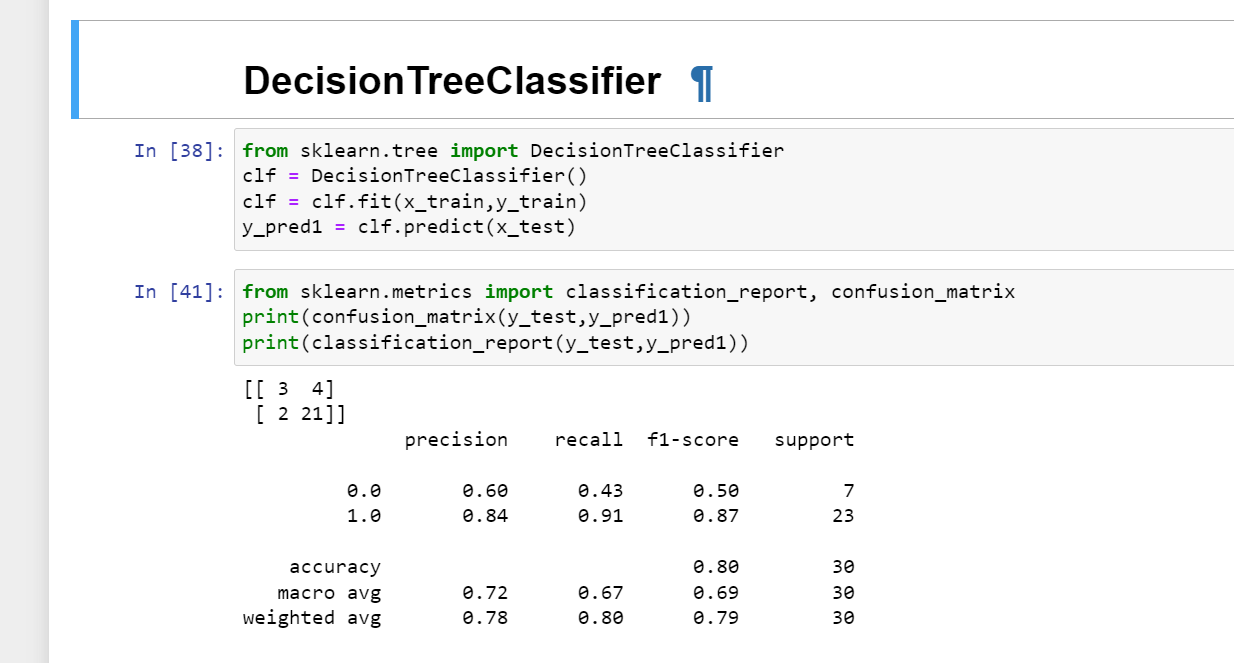
# Cunfusion metrices



# GradientBoostingClassifier



# DecisionTreeClassifier



Run and Evaluate selected models

According to the performance metrics, Random Forrest scores highest in accuracy. Hence, Random Forrest looks like the best fit for this data.