REPORT

DOCUMENTATION ON LOAN PREDICTION

MODEL NAME :- LOAN STATUS PREDICTION

AIM :-

THE AIM OF THIS PROJECT WAS TO COMPARE THE VARIOUS LOAN PREDICTION MODELS TO SHOW WHICH IS THE BEST ONE WITH LEAST ERROR AND COULD BE USED BY BANKS IN REAL WORLD TO PREDICT IF THE LOAN SHOULD BE APPROVED OR NOT TAKING THE RISK FACTOR IN MIND.AFTER COMPARING THE ANALYSING THE MODELS,IT WAS FOUND THAT THE PREDICTION MODEL BASED ON SVM AND LR PROVED TO BE THE MOST ACCURATE AND FITTING OF THEM ALL.

THIS CAN BE USEFUL IN REDUCING THE TIME AND MANPOWER REQUIRED TO APPROVE LOANS AND FILTER OUT THE PERFECT CANDIDATES FOR PROVIDING LOANS.

INPUT TO THE MODEL :-GENDER,MARRIED,DEPENDENTS,EDUCATION,SELF\_EMPLOYED,APPLICANT INMCONE,COAPPLICANT INCOME,LOAN AMOUNT,LOAN AMOUNT TERM,CREDICT HISTRORY,PROPERTY AREA

OUTPUT TO THE MODEL :-

LOAN STATUS

THIS MODEL HELPS TO WHICH DEPARTMENT PEOPLE?

BANK PEOPLE WILL MAKE CORRECT STATUS OF THEIR LOAN GIVING TO PEOPLE ON BASED UPON THEIR PERSONAL IMFORMATION AND SOME CRITIERIAS ABOUT THEIR LIVING AREA CONDITIONS AS OF CHECKING ALL THEIR PREVIOUS LOAN HISTORY AND ASKING ABOUT THEM, WILL THEY ARE ACTUALLY ABLE TO REPAY THEIR LOAN AMOUNT OR NOT?

PROBLEM STATEMENT :-

BANK PEOPLE HAVE A PRESENCE ACROSS ALL URBAN ,RURAL AND SEMIRURAL AREAS.THE CUSTOMER FIRST APPLIES FOR LOAN AND AFTER THAT THE COMPANY VALIDATES THE CUSTOMER ELIGIBILITY FOR THE LOAN.

AS FOR THE ELIGIBILTY ,IF THEY WANTS TO AUTOMATE THE LOAN ELIGIBILITY PROCESS(REAL-TIME)BASED ON CUSTOMER DETAIL PROVIDED WHILE FILLING OUT ONLINE APPLICATION FORMS.

THE DETAILS ARE MENTIONED ABOVE AS “INPUT TO THE MODEL”.

SO USING THE TRAINING DATASET WE WILL TRAIN OUR MODEL AND TRY TO PREDICT THE LOAN STATUS COLUMN ON TESTING DATASET.

PROCEDURE :-

STEP-1 : IMPORTING THE ESSENTIAL LIBRARAIES FOR THE PREDICTION.

STEP-2 : LOADING THE DATASET.

STEP-3 : HANDLING THE MISSING VALUES.

STEP-4 : CHECKING IS THERE ANY MISSING VALUES IF PRESENT REPLACE OR DROP THE MISSING VALUES.

STEP- 5: CLEANING THE DATSET , CONVERTING ALL THE CATEGORICAL VALUES INTO NUMERICAL VALUES FOR EASY PREDICTIONS.IF DATSET IS CATEGORICAL ERROR PREDICTIONS AND HANDLING THE DATA MAY CREATE PROBLEMS.

STEP- 6: DATA VISUALIZATION .IF IT IS MULTI LINEAR REGRESSION USE SEABORN REGRESSION TO SHOW THE RELATIONSHIP AMONG THE COLUMNS AS IT HAS INBUILT FEATURES LIKE COLOURS ,TITLES ETC.

STEP-7 : DIVIDING THE DATASET INTO TRAINING AND TESTING FOR TARINING OUR MACHINE WITH SOME DATA AND TESTING IT.

STEP-8 : APPLYTING THE BEST ALOGORITHM FOR BUILIDING THE MODEL.

STEP-9 : CHECKING THE ACCURACY SCORE.

STEP-10 : PREDICTING THE MODEL WITH OLD DATAPOINTS AND NEW DATAPOINTS.

DESCRIPTION ABOUT DATA COLUMNS :-

GENDER : MALE/FEMALE

MARRIED : APPILICANT MARRIED OR NOT

DEPENDENTS : NUMBER OF DEPENDENTS ON APPLICANT

EDUCATION : GRADUATE OR UNDER GRADUATE

SELF\_EMPLOYED : SELF EMPLOYED OR NOT

APPILICANT INCOME : SALARY

CO APPLICANT INCOME : SALARY OF SPOUSE

LOAN AMOUNT TERM : NUMBER OF MONTHS REQUIRED TO REPAY THE AMOUNT.

LOAN AMOUNT : AMOUNT IN THOUSANDS THEN THE PROBABILITY OF REPAYING WOULD BE LESSER

CREDIT CARD HISTORY :1 DENOTES THE CREDIT HISTORY IS GOOD AND 0 OTHERWISE.IT IS THE RECORD OF A BORROWERS RESPONSIBLE PAYMENT OF DEBTS

PROPERTY AREA :URBAN,SEMIURBAN,RURAL

LOAN STATUS : LOAN APPROVED OR NOT

LOAN ESSENTIAL PYTHON ML LIBRARIES :-

import pandas as pd

import matplotlib.pyplot as plt

import numpy as np

import seaborn as sns

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

from sklearn import svm

from sklearn.metrics import accuracy\_score

LOAD DATASET :-

df=pd.read\_csv("train\_u6lujuX\_CVtuZ9i (1).csv")

df

SIZE OF THE DATASET :-

df.shape

HANDLING THE MISSING VALUES IN DATASET :-

df.isnull().sum()

DROPPING MISSING VALUES :-

data=df.dropna()

CONVERTING THE CATEGORICAL VALUES INTO NUMERICAL VALUES :-

org\_data=data1.replace({'Married':{'No':0,'Yes':1},'Gender':{'Male':1,'Female':0},'Self\_Employed':{'No':0,'Yes':1},'Education':{'Graduate':1,'Not Graduate':0},'Property\_Area':{'Rural':0,'Semiurban':1,'Urban':2}})

TRAINING AND TESTING :-

x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.2,stratify=y,random\_state=2)

MODEL CLASSIFIER :-

model=LogisticRegression()

ACCURACY SCORE :-

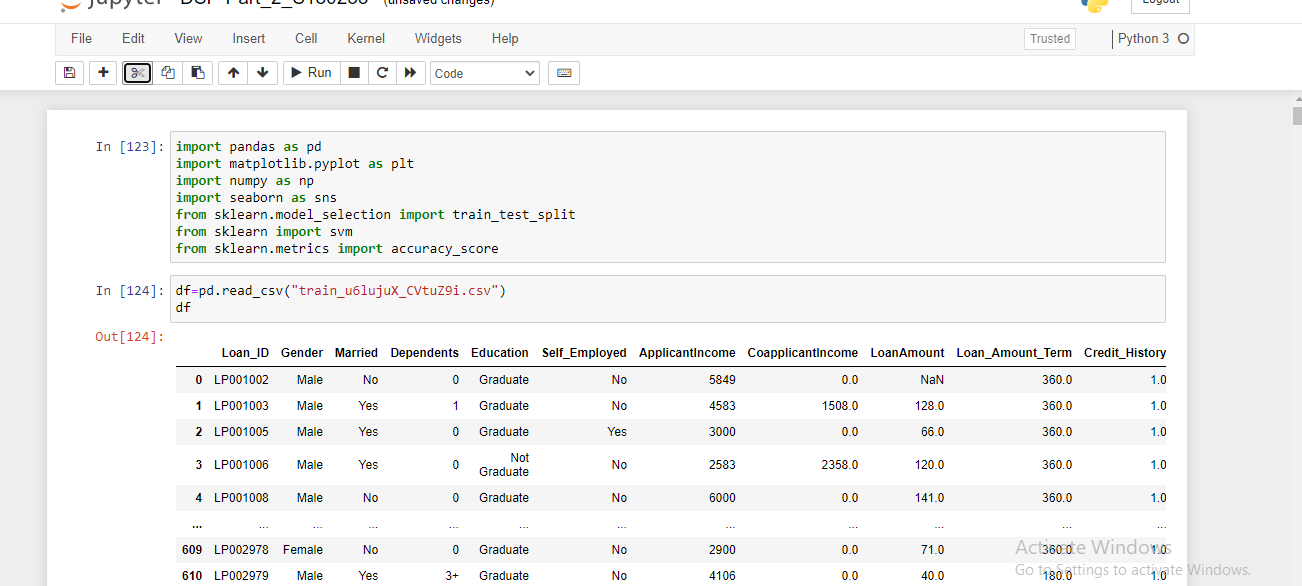
x\_train\_pred=model.predict(x\_train)

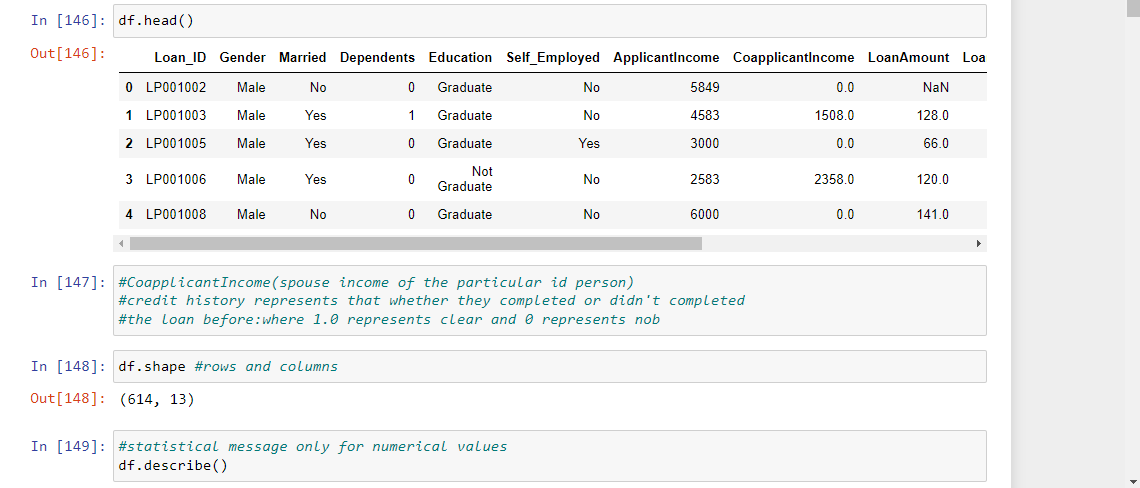
training\_data\_accuracy=accuracy\_score(x\_train\_pred,y\_train)

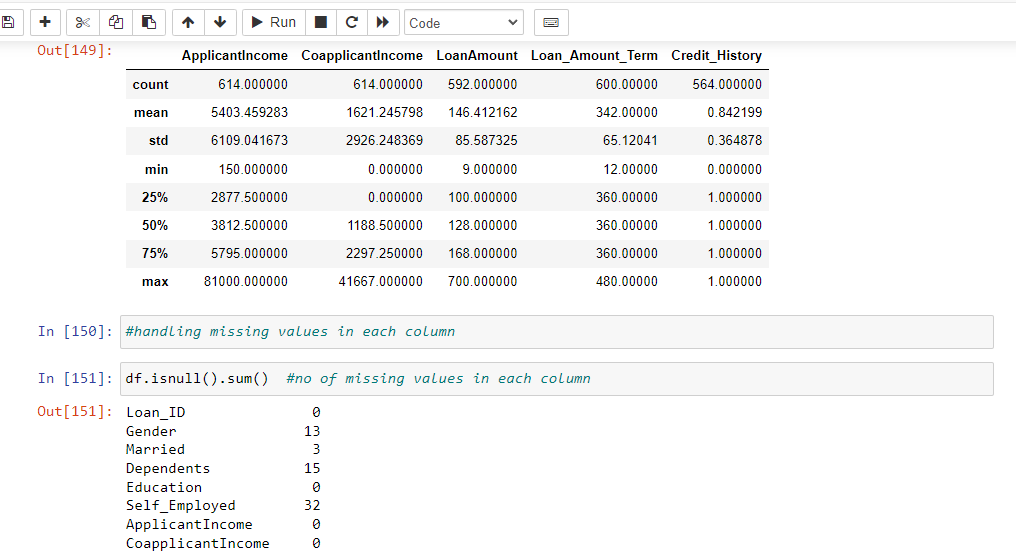
x\_test\_pred=model.predict(x\_test)

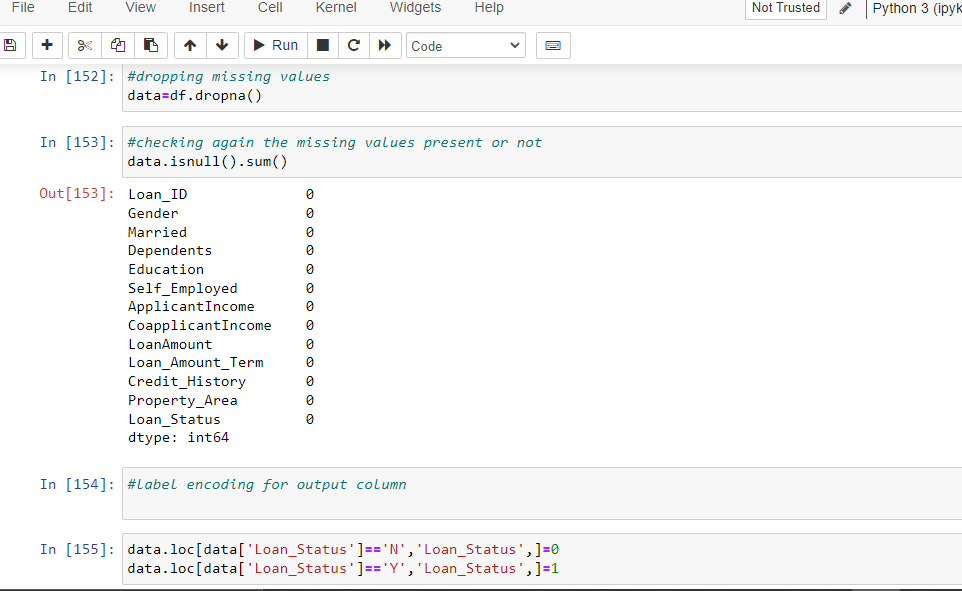
testing\_data\_accuracy=accuracy\_score(x\_test\_pred,y\_test)

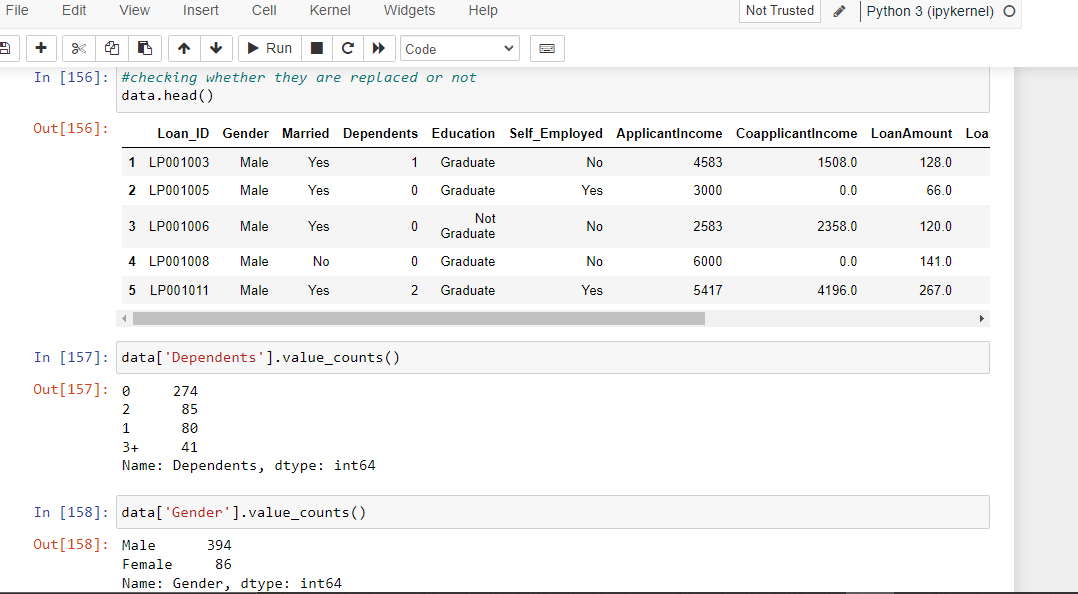
SOURCE PROGRAM :-

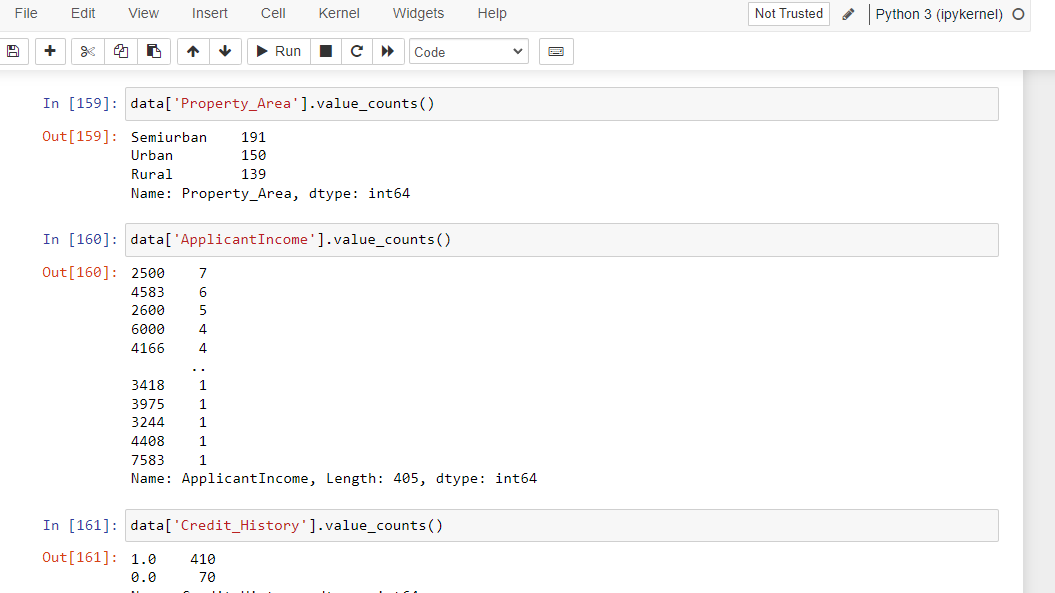


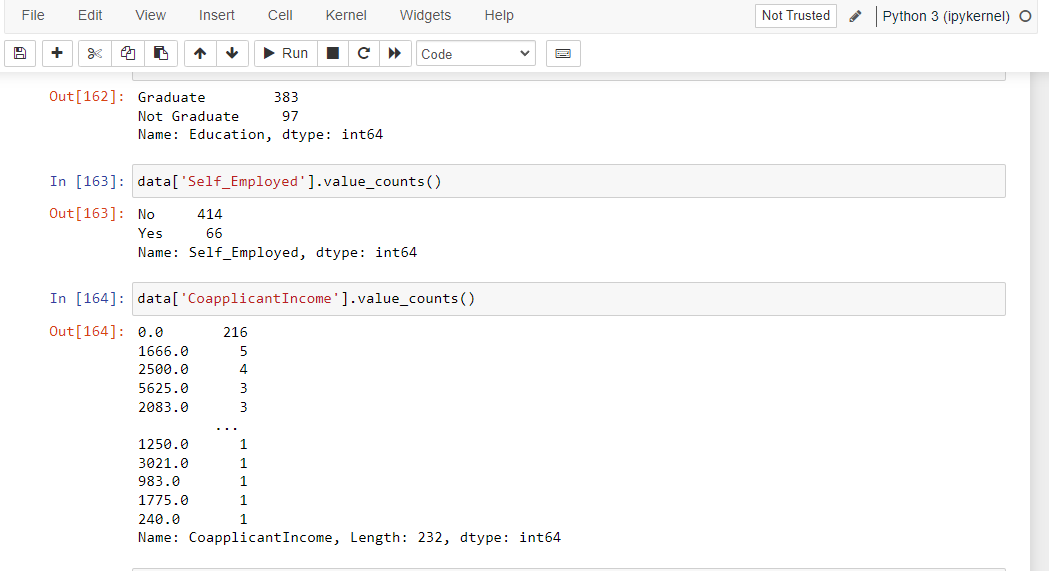


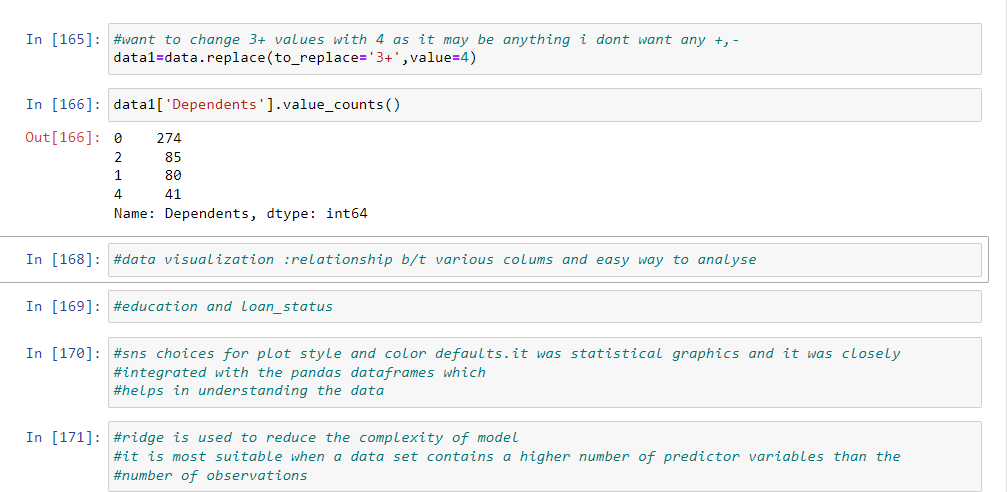


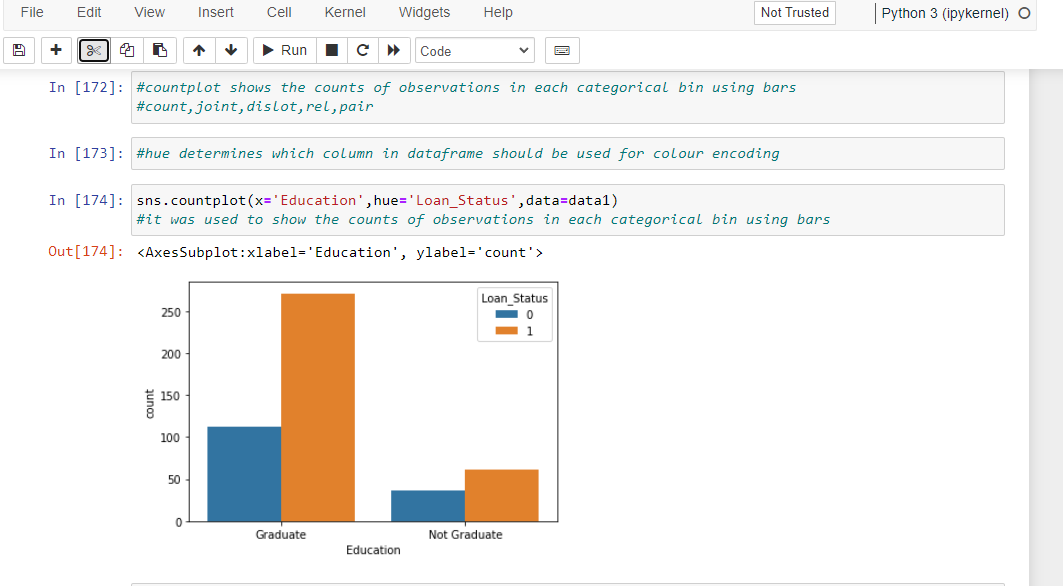


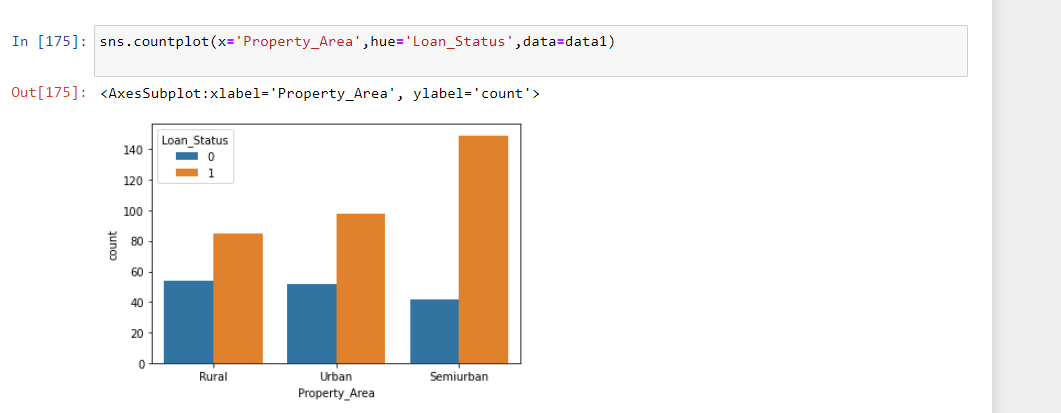


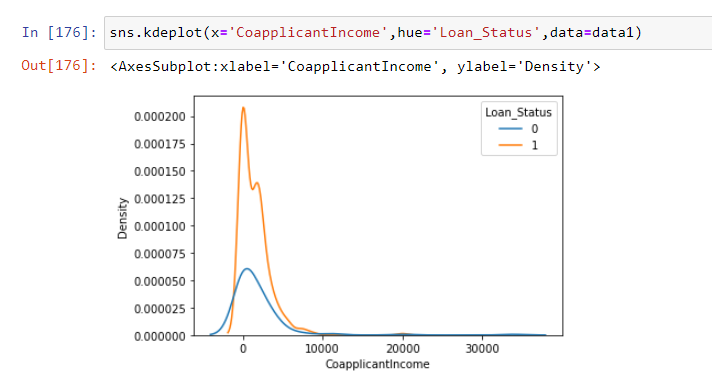


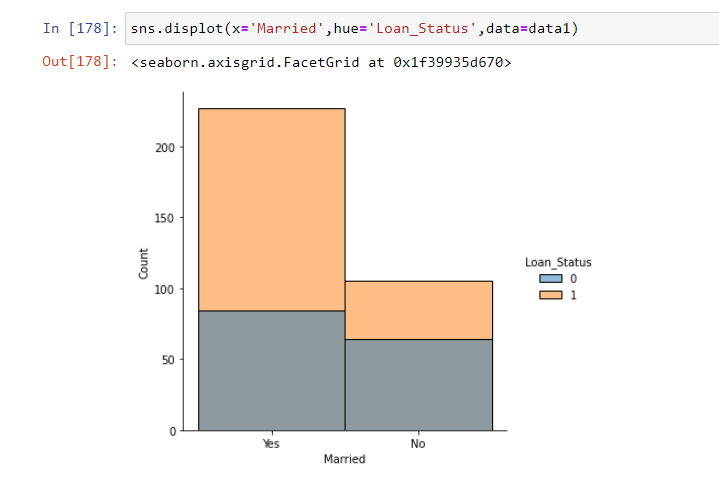


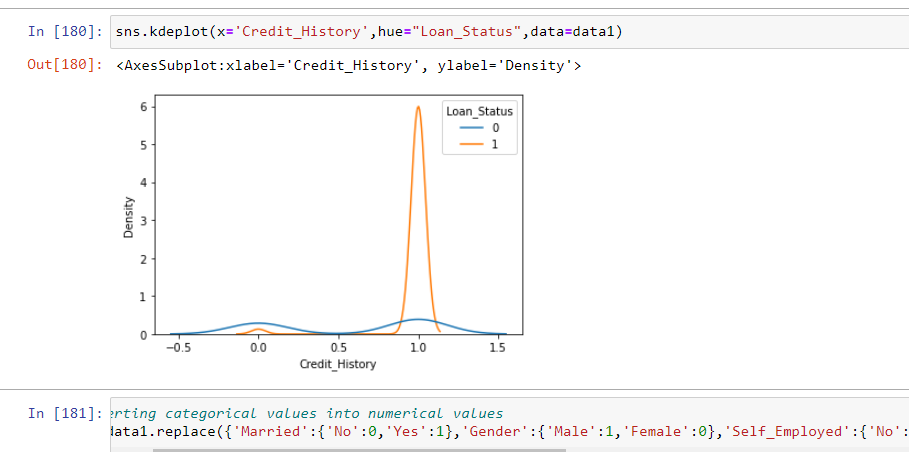


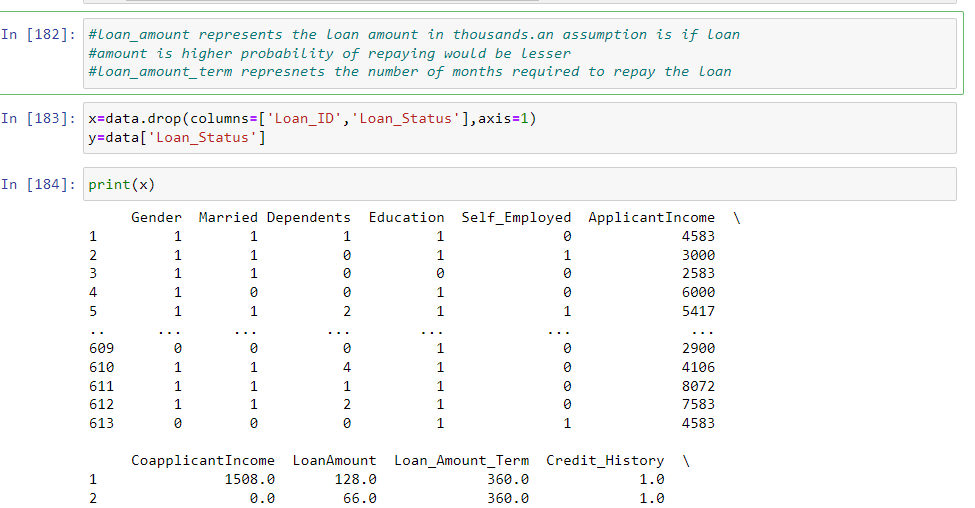


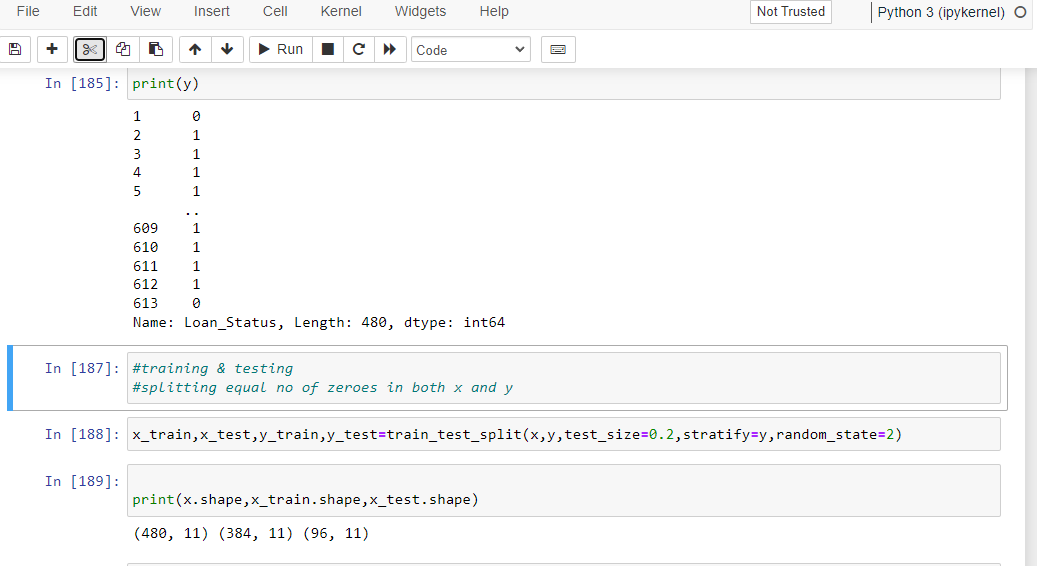


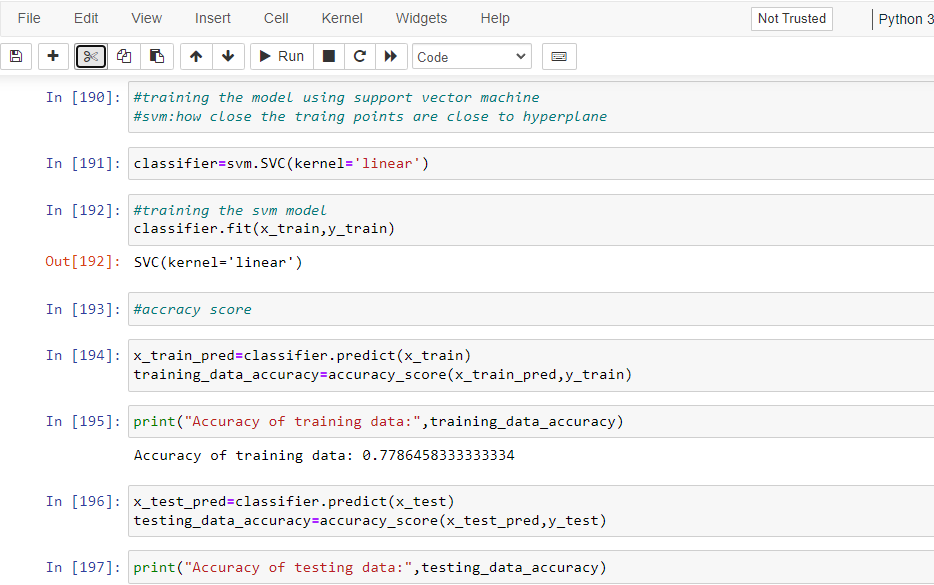


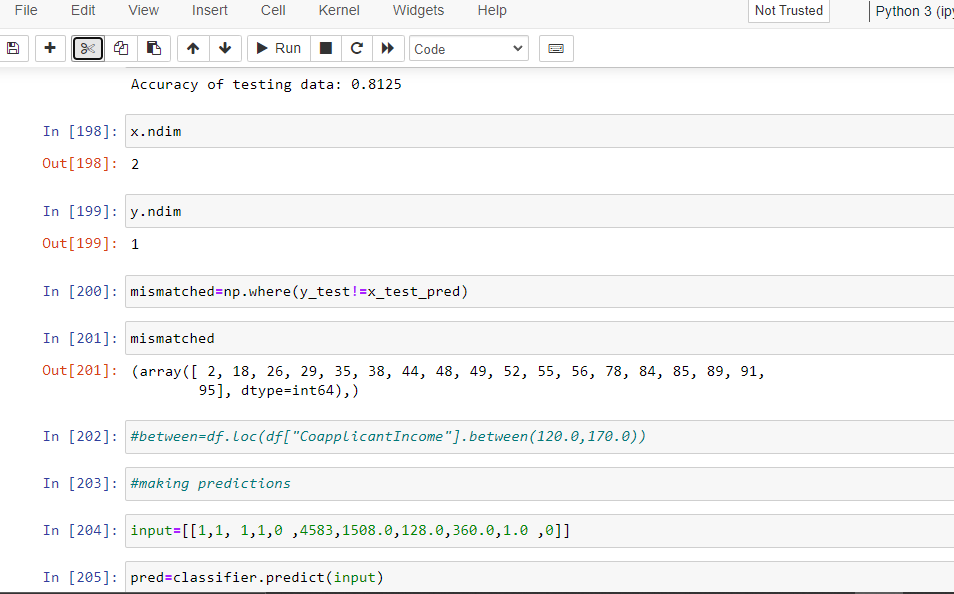


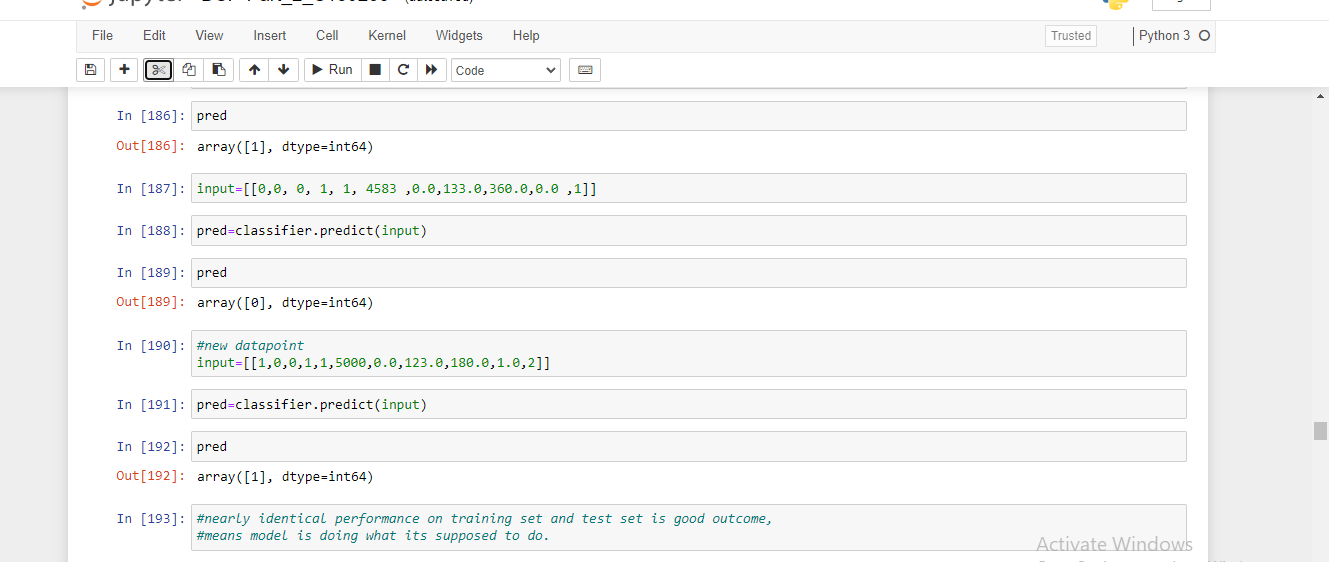


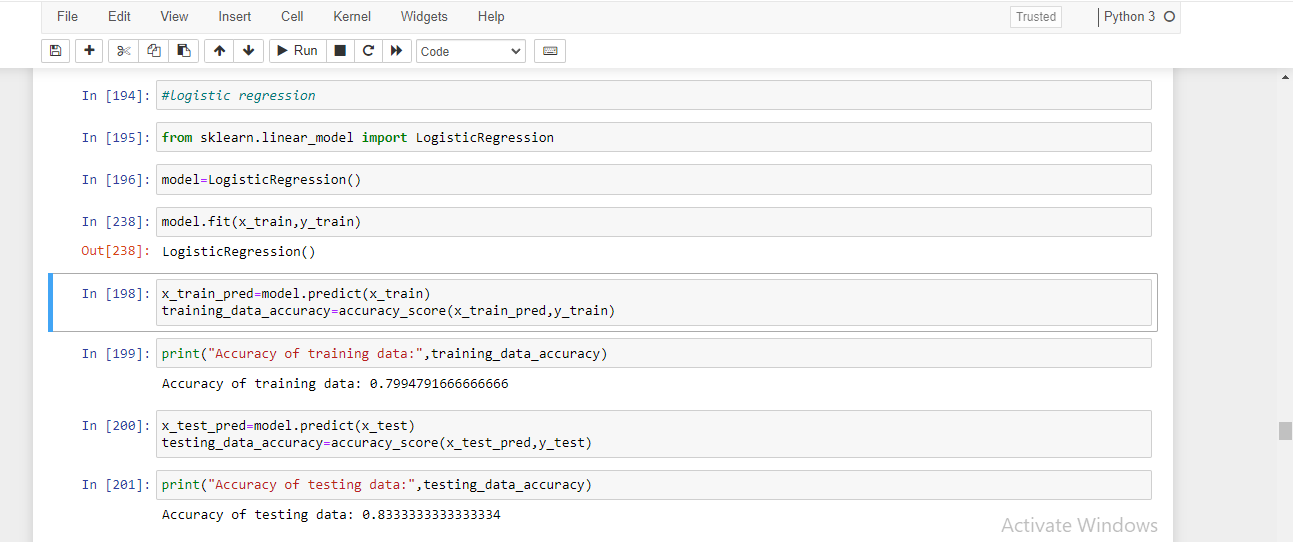


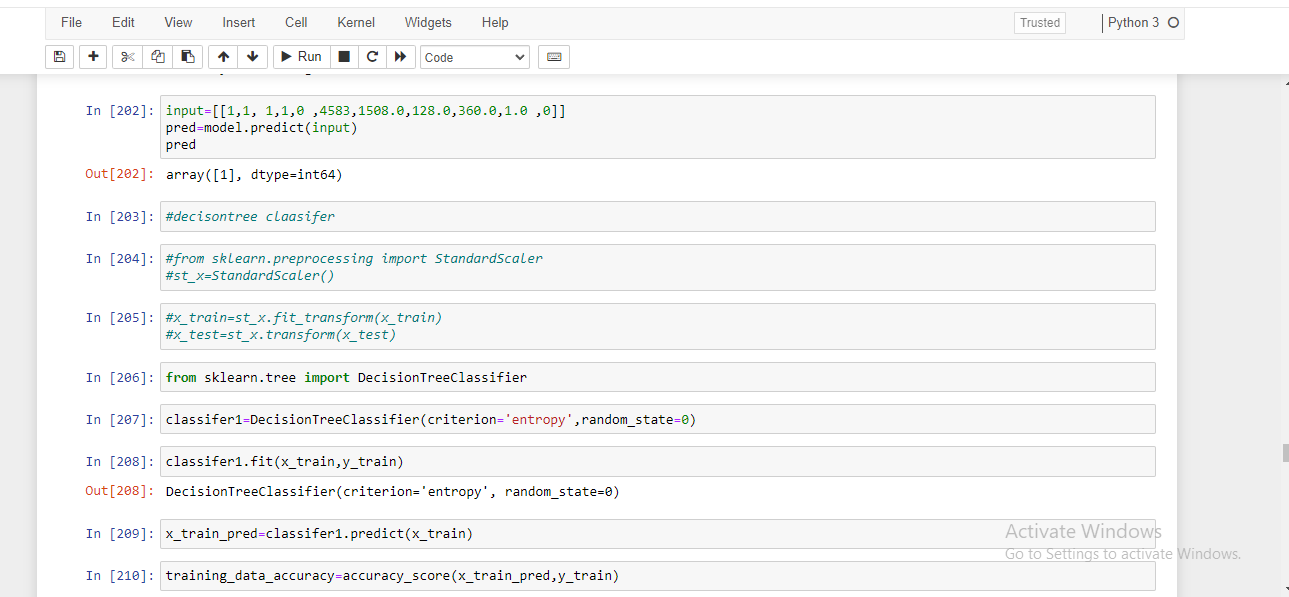


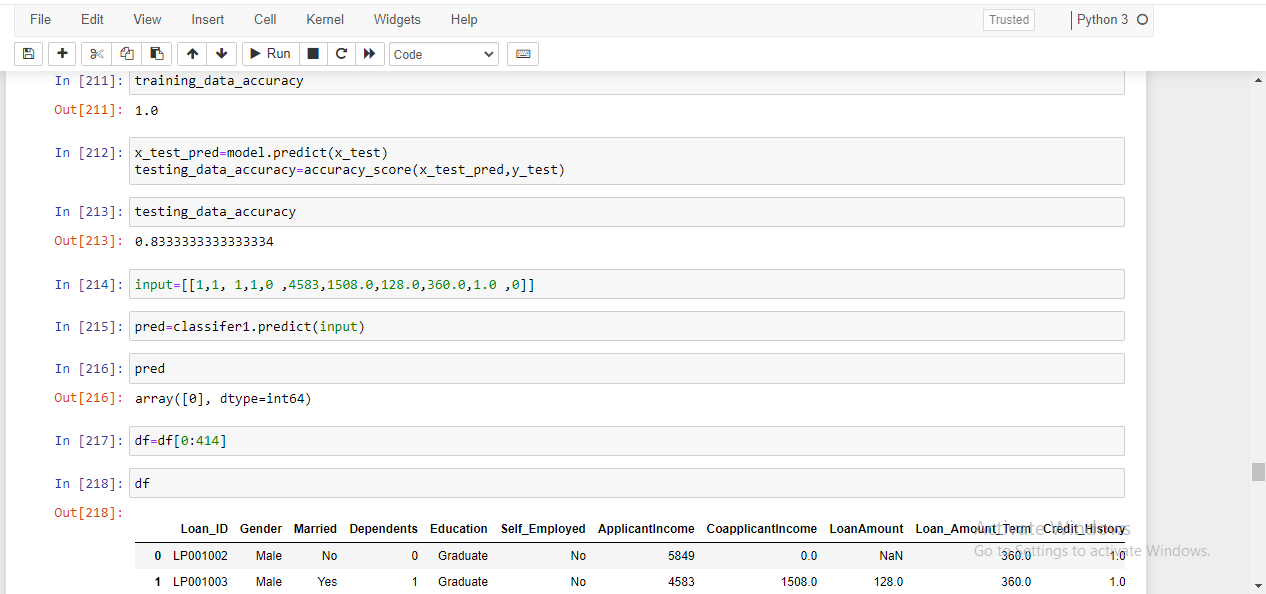


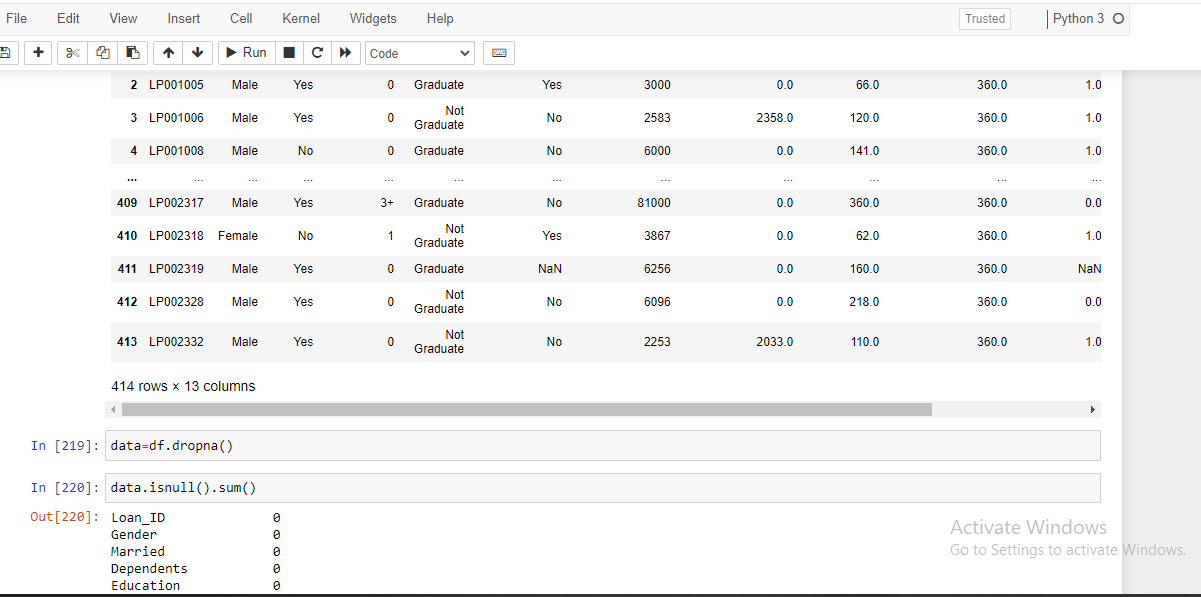


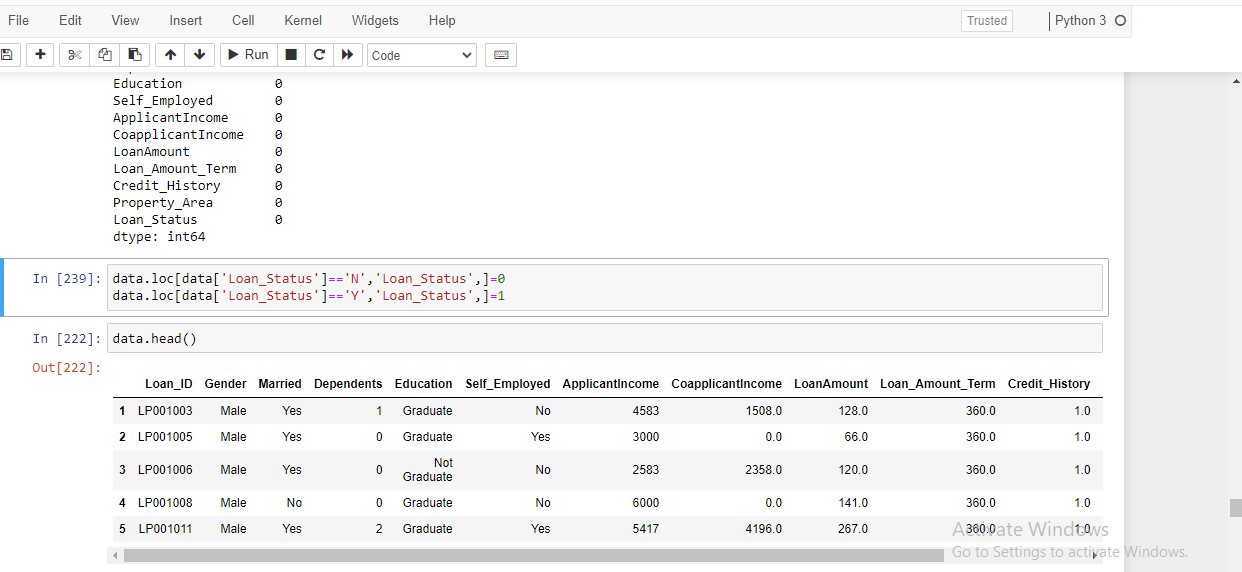


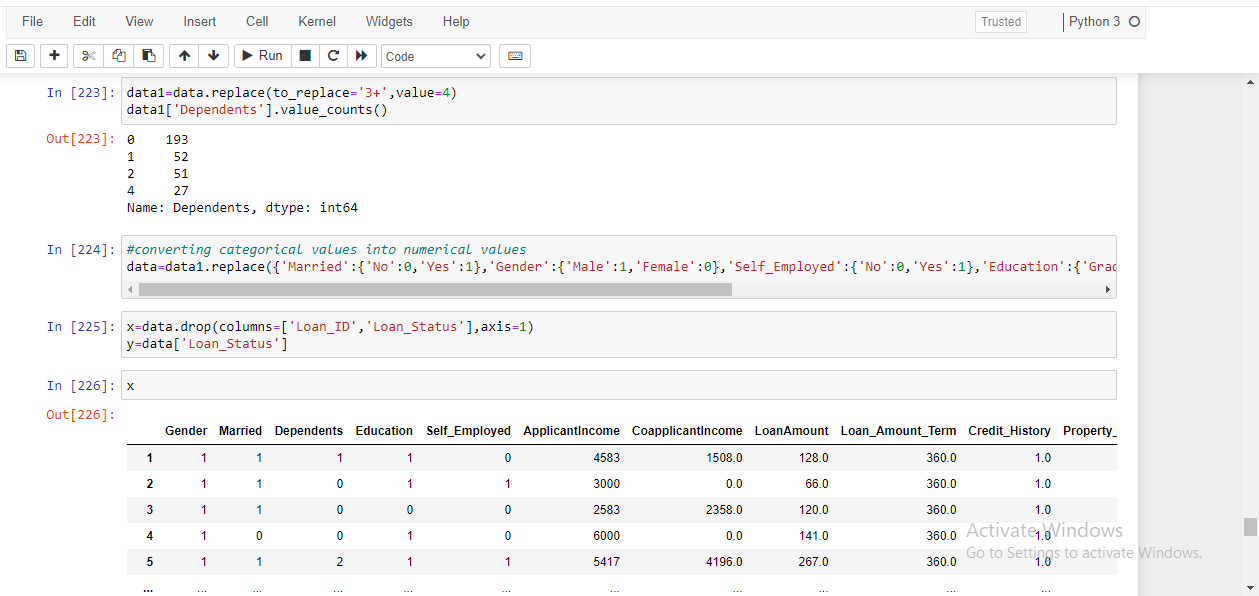


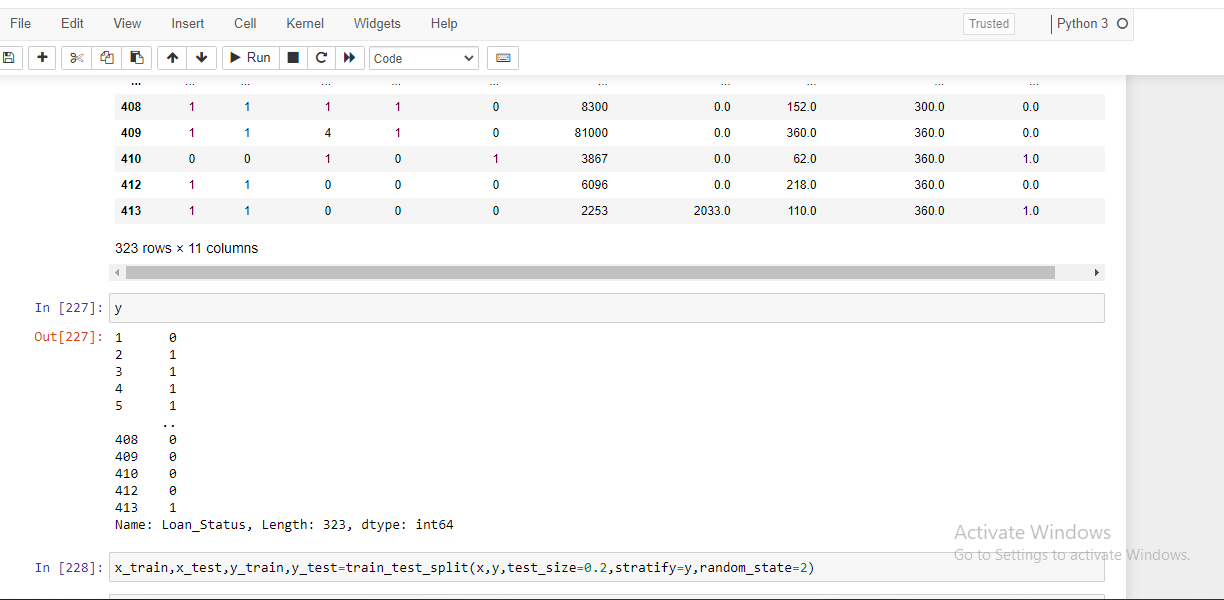


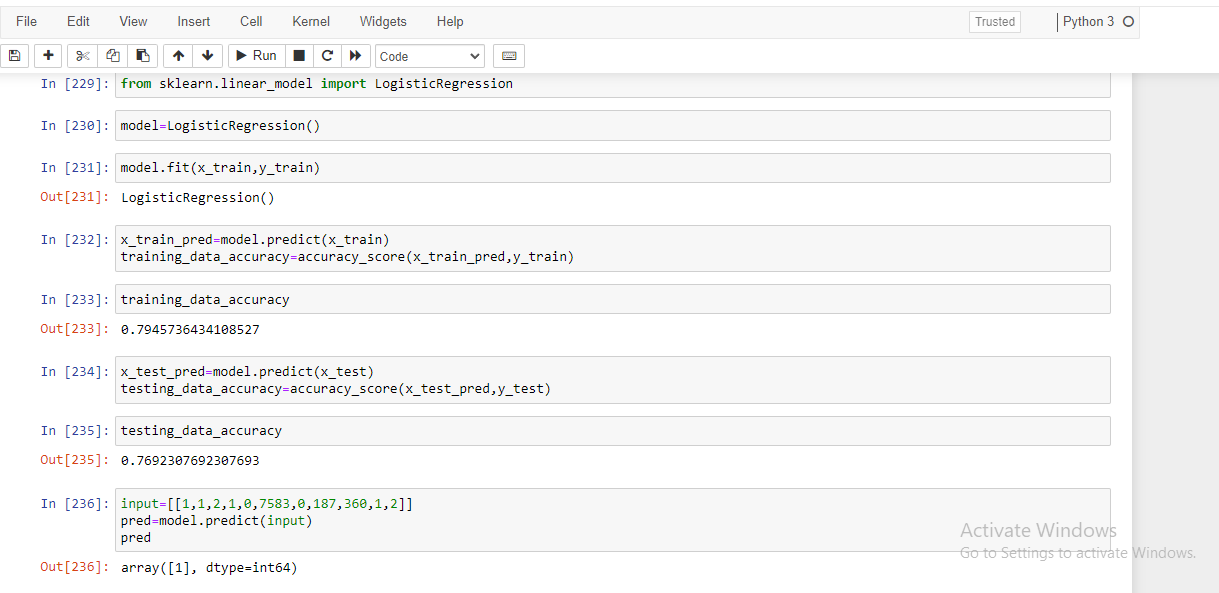


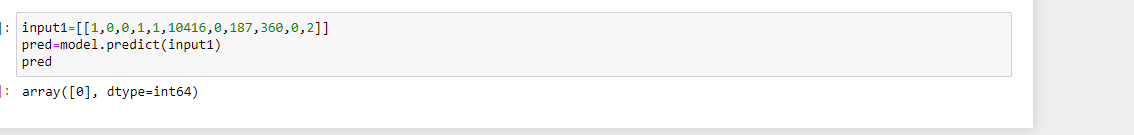












CONCLUSION :-

WE HAVE MULTIPLE ALOGORITHMS FOR TRAINING PURPOSES LIKE DECISION TREE, SVC, LOGISTIC REGRESSION….

HERE I USED **LOGISTIC REGRESSION** FOR PREDICTING THE VALUES WHICH HAS THE **ACCURACY SCORE 0.83** PERCENTAGE.

FEATURE ENGINEERING HELPED ME TO INCREASE MY MODEL ACCURACY.

1. PERCENTAGE OF MARRIED PEOPLE HAVE GOT THEIR LOAN APPROVED IS HIGHER THAN NON-MARRIED.

2. PERCENTAGE OF APPLICANTS WITH EITHER 0 0R 2 DEPENDENTS HAVE GO THEIR LOAN APPROVEDIS HIGHER.

3. PERCENTAGE OF GRADUATES HAVE CHANCE OF GETTING LOAN HIGGER COMPARED TO WHO ARE NOT GRADUATES.

4. DOESNT MATTER IS THE PERSON IS SELF\_EMPLOYED OR NOT.

5. THERE IS A HIGH CHANCE FOR TOTAL INCOME BASED PREDICTION WHICH IS BOTH APPLICANT INCOME AND COAPPLICANT INCOME.