Name:A.Rajeswari RollNo:195214142 Course:I M.Sc.,Computer Science(A)

**MACHINE LEARNING PROJECT**

In this project I have used **LinearRegression** and **SGDRegressor** Machine

Learning model for Insurance.csv dataset from **Kaggle website .**

**PROBLEM STATEMENT:**

This is a Insurance Dataset is calculate the insurance charges done by using **LinearRegression and SGDRegressor.Then traing and testing the dataset used by** x\_train,y\_train,x\_test,y\_test.

**ML METHODOLOGY:**

**Linear Regression** is a **machine learning** algorithm based on supervised **learning**. **Linear regression** performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this **regression** technique finds out a **linear** relationship between x (input) and y(output).

**Stochastic Gradient Descent (SGD)** is a simple yet very efficient approach to discriminative learning of linear classifiers under convex loss functions such as (linear) [Support Vector Machines](https://en.wikipedia.org/wiki/Support_vector_machine) and [Logistic Regression](https://en.wikipedia.org/wiki/Logistic_regression). Even though SGD has been around in the machine learning community for a long time, it has received a considerable amount of attention just recently in the context of large-scale learning.

**DATASET DESCRIPTION:**

* Age: age of primary beneficiary
* Sex: insurance contractor gender, female, male
* bmi: Body mass index, providing an understanding of body, weights that are relatively high or low relative to height,  
  objective index of body weight (kg / m ^ 2) using the ratio of height to weight, ideally 18.5 to 24.9
* Children: Number of children covered by health insurance / Number of dependents
* Smoker: Smoking
* Region: the beneficiary's residential area in the US, northeast, southeast, southwest, northwest.
* Charges: Individual medical costs billed by health insurance

**PRE-PROCESSING:**

In this Insurance Dataset used for pre-processing methods are,

%matplotlib inline

import matplotlib.pyplot as plt

import numpy as np

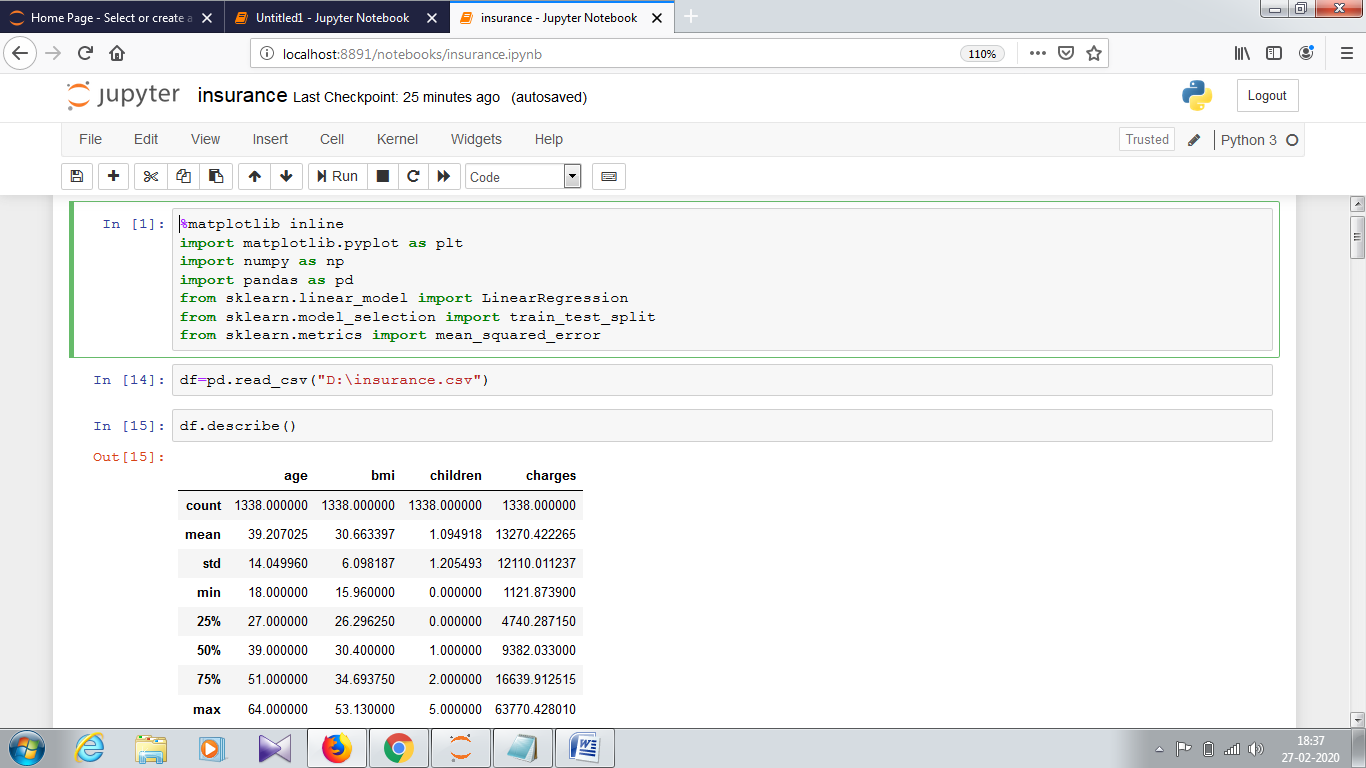
import pandas as pd

from sklearn.linear\_model import LinearRegression

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

from sklearn.metrics import mean\_squared\_error

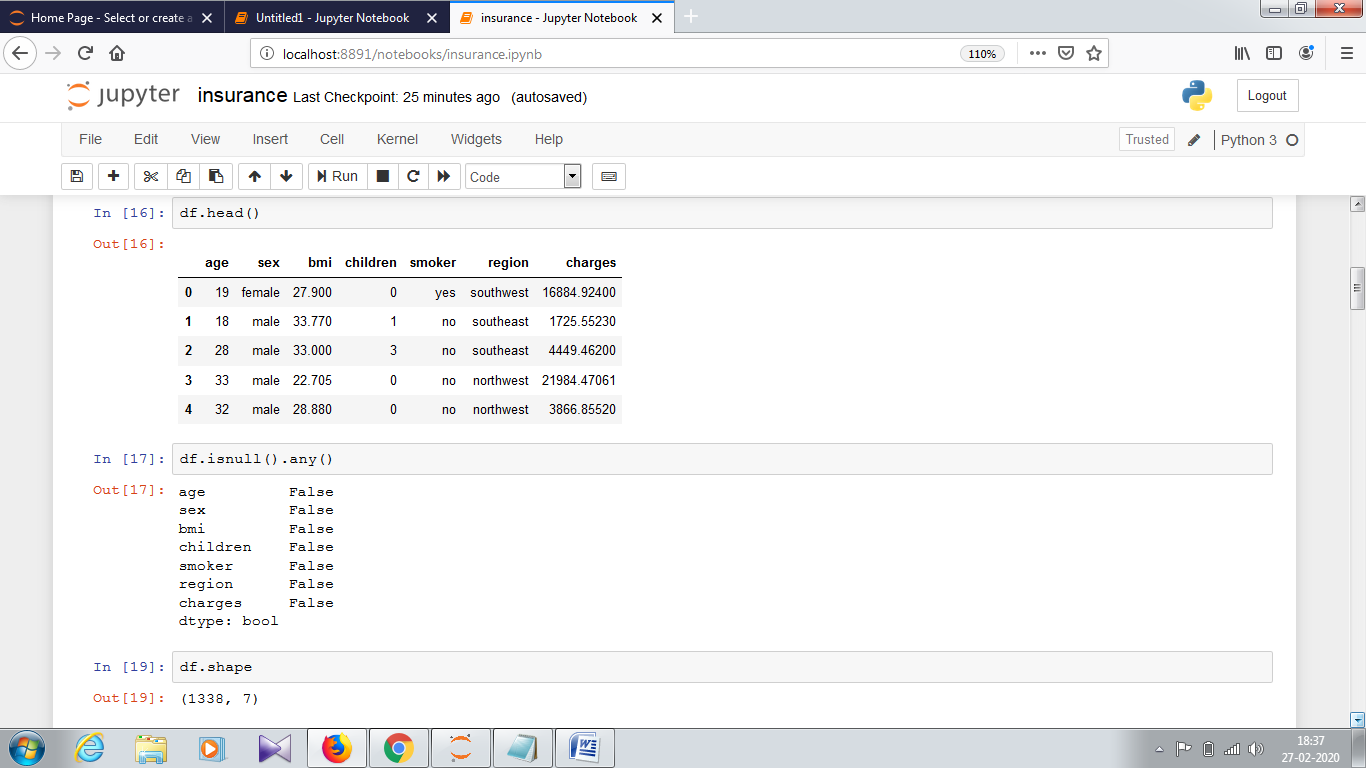
df.describe()

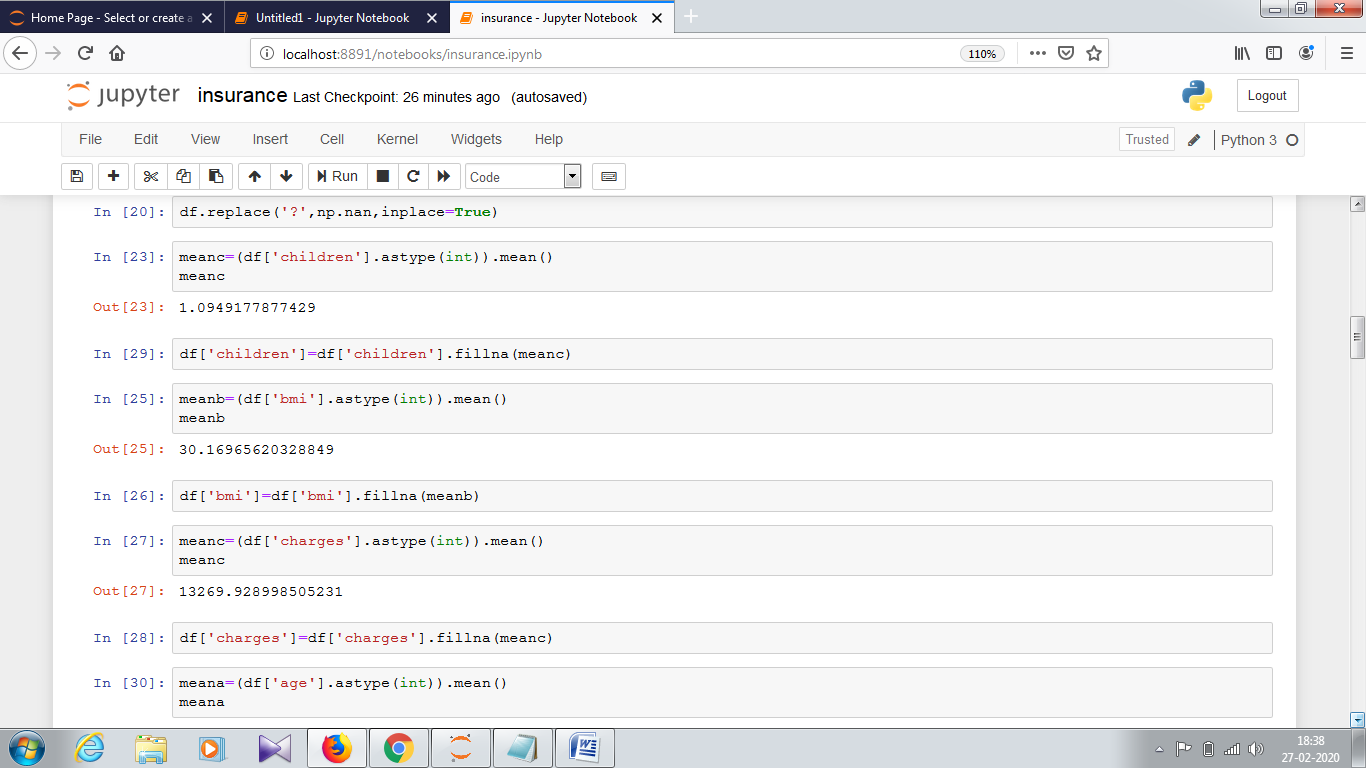


df.head()

df.isnull().any()

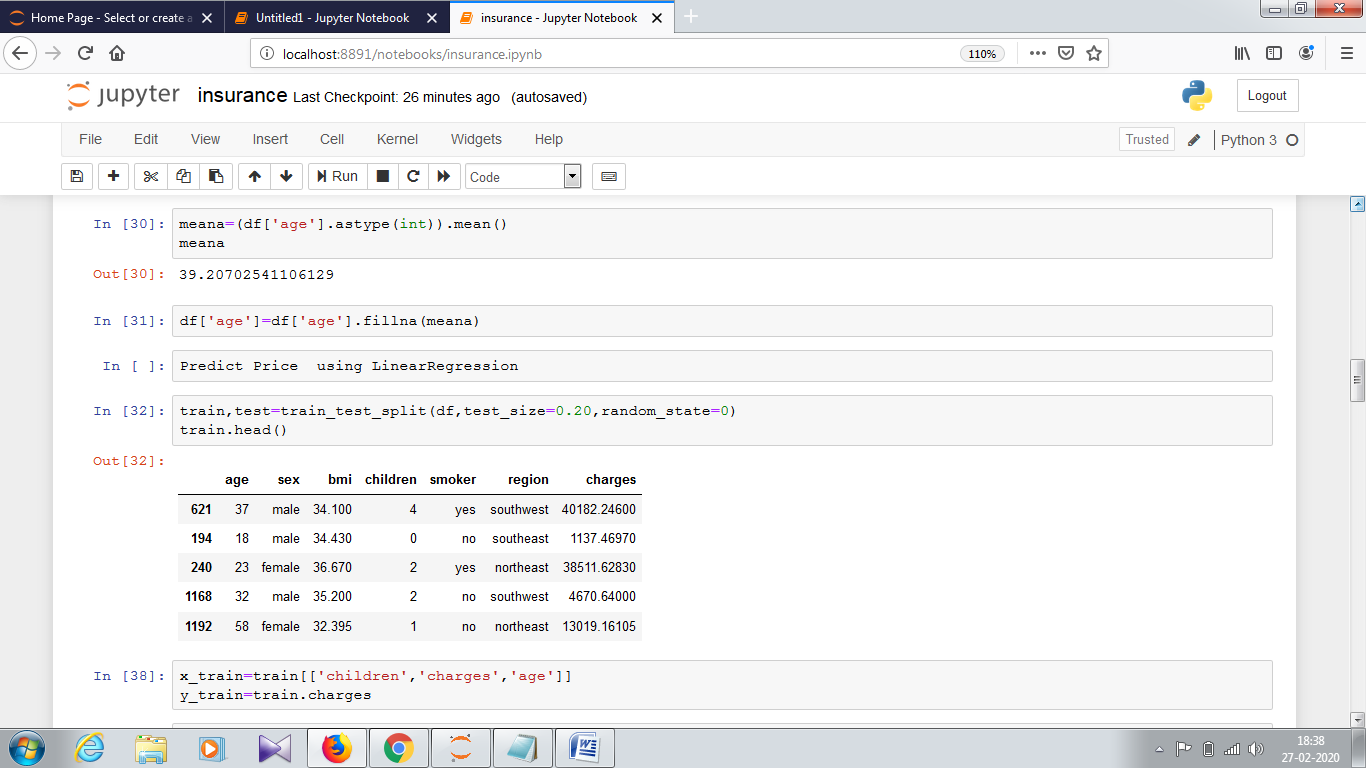
df.shape





train,test=train\_test\_split(df,test\_size=0.20,random\_state=0)

train.head()



print(x\_train.shape)

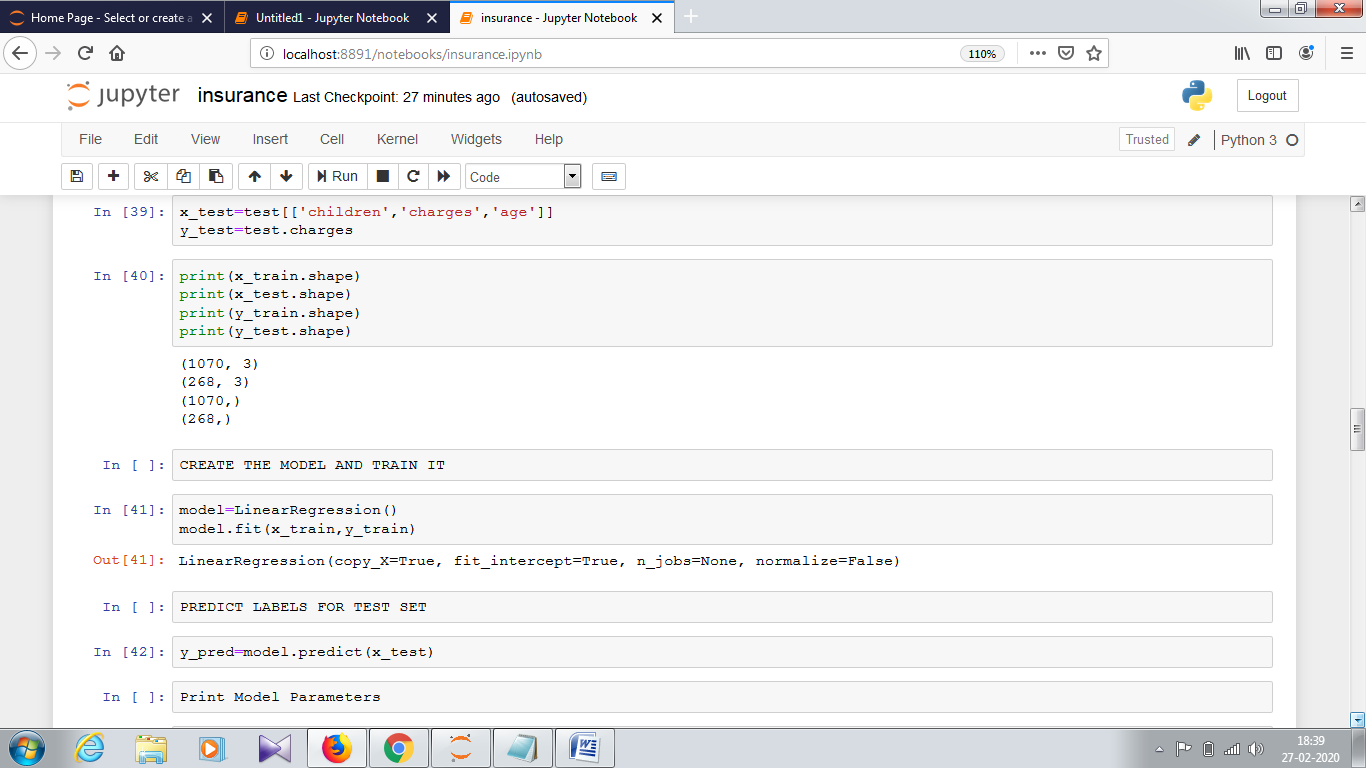
print(x\_test.shape)

print(y\_train.shape)

print(y\_test.shape)

model=LinearRegression()

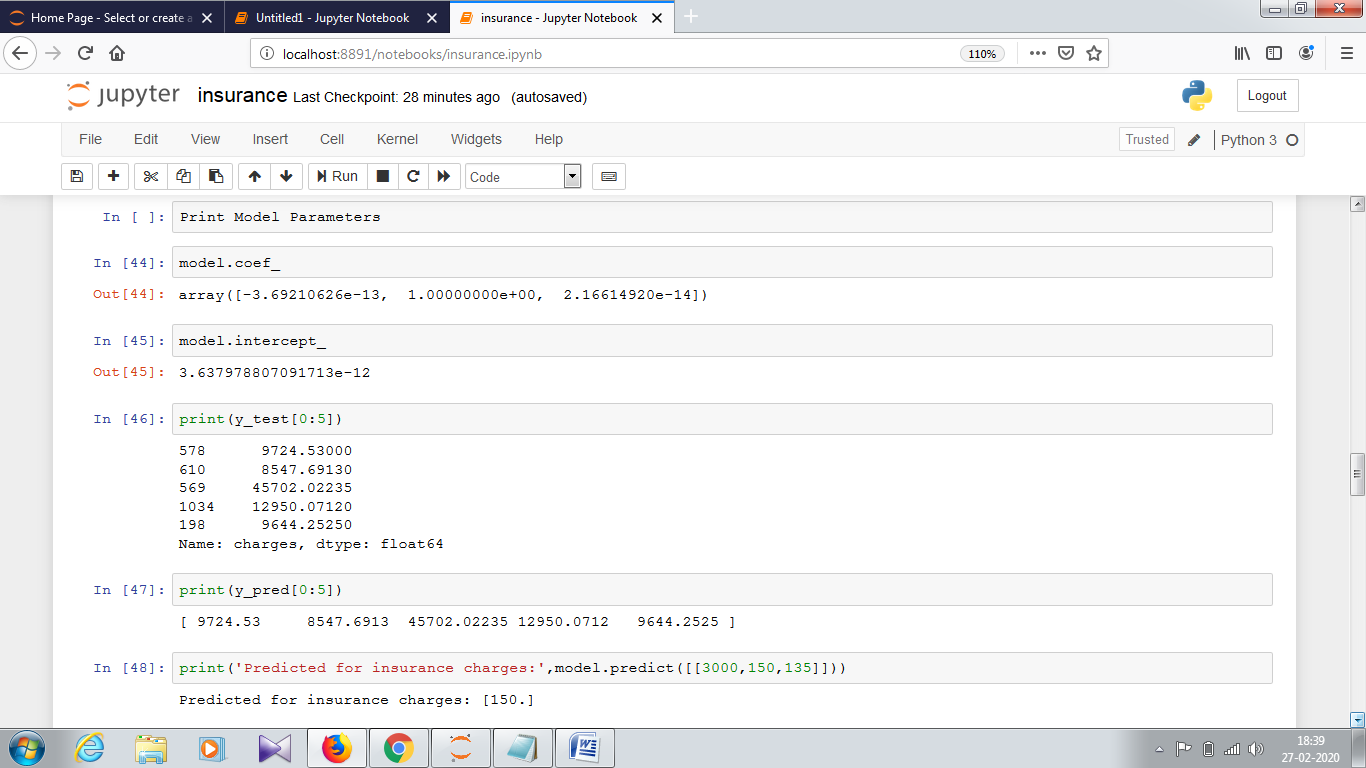
model.fit(x\_train,y\_train)

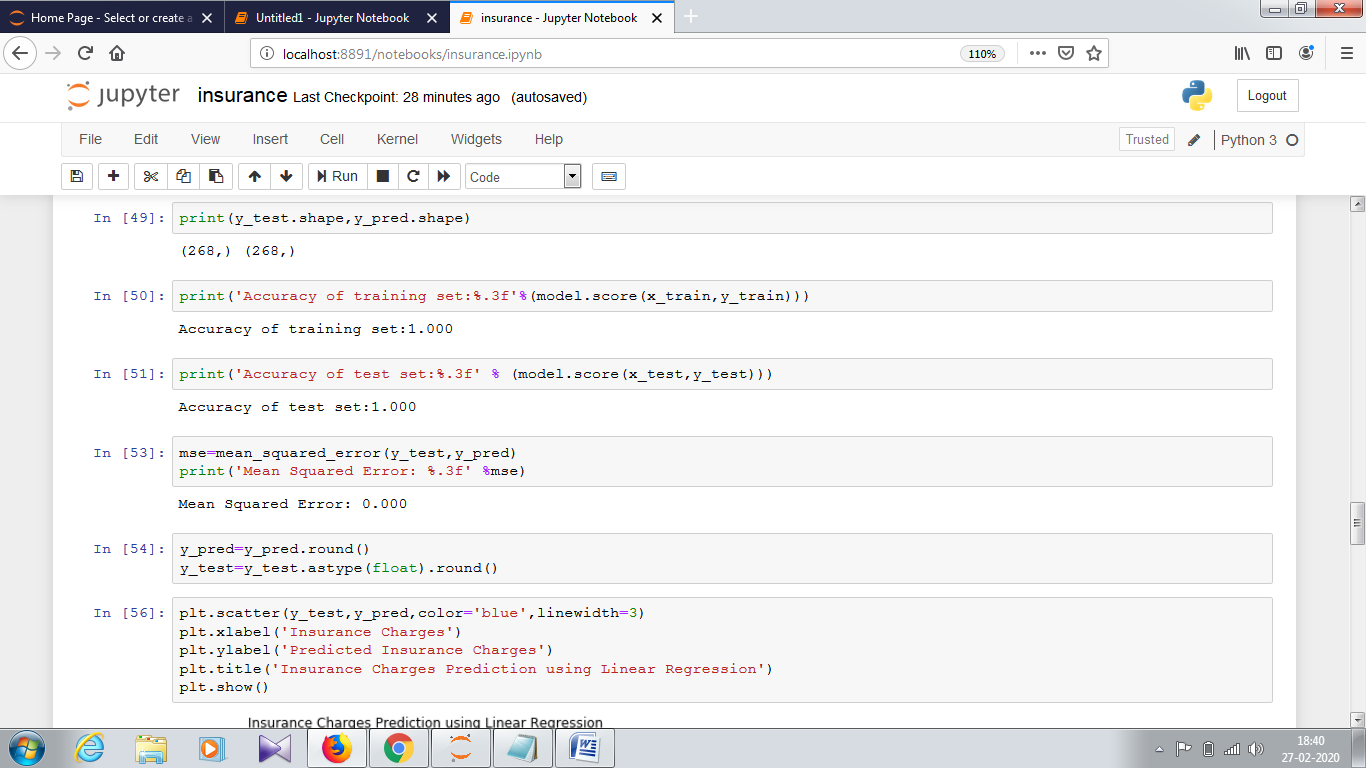


model.coef\_

model.intercept\_

print('Predicted for insurance charges:',model.predict([[3000,150,135]]))





plt.scatter(y\_test,y\_pred,color='blue',linewidth=3)

plt.xlabel('Insurance Charges')

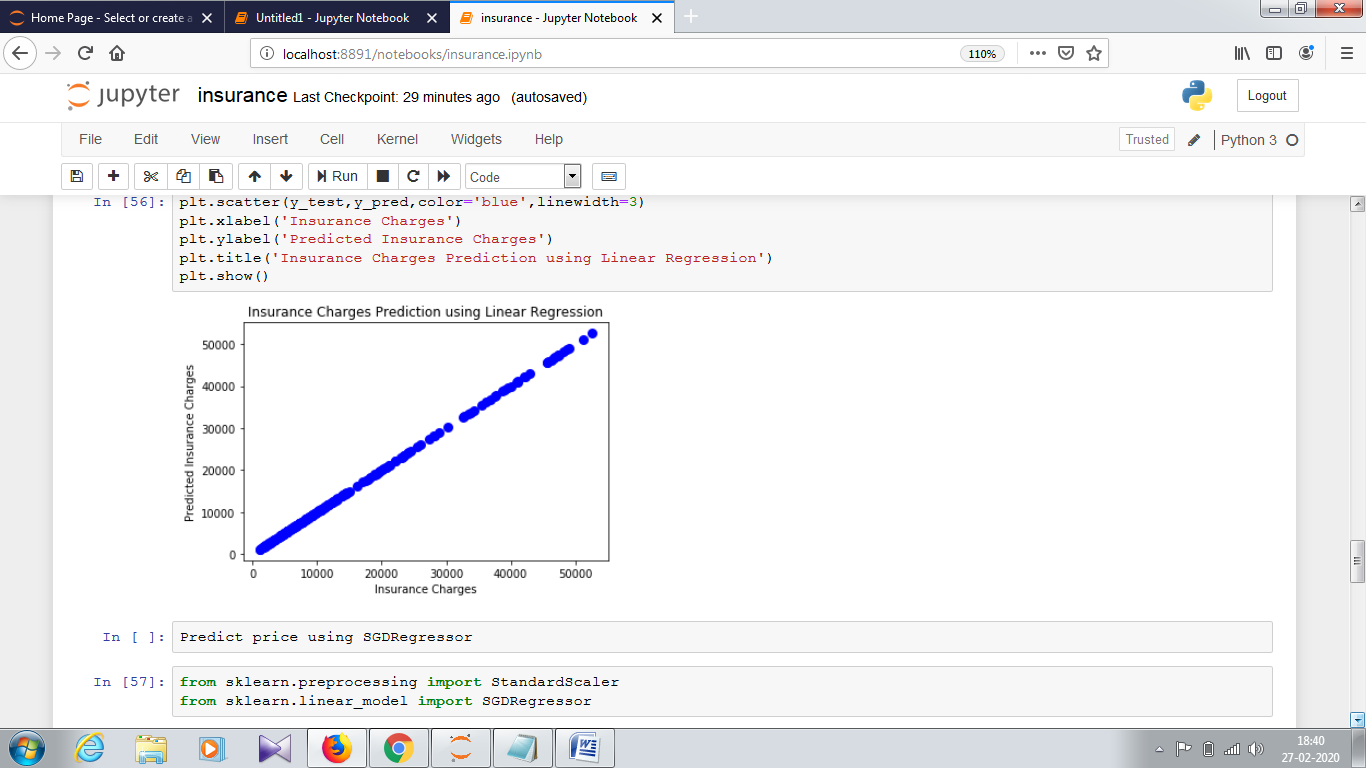
plt.ylabel('Predicted Insurance Charges')

plt.title('Insurance Charges Prediction using Linear Regression')

plt.show()

from sklearn.preprocessing import StandardScaler

from sklearn.linear\_model import SGDRegressor



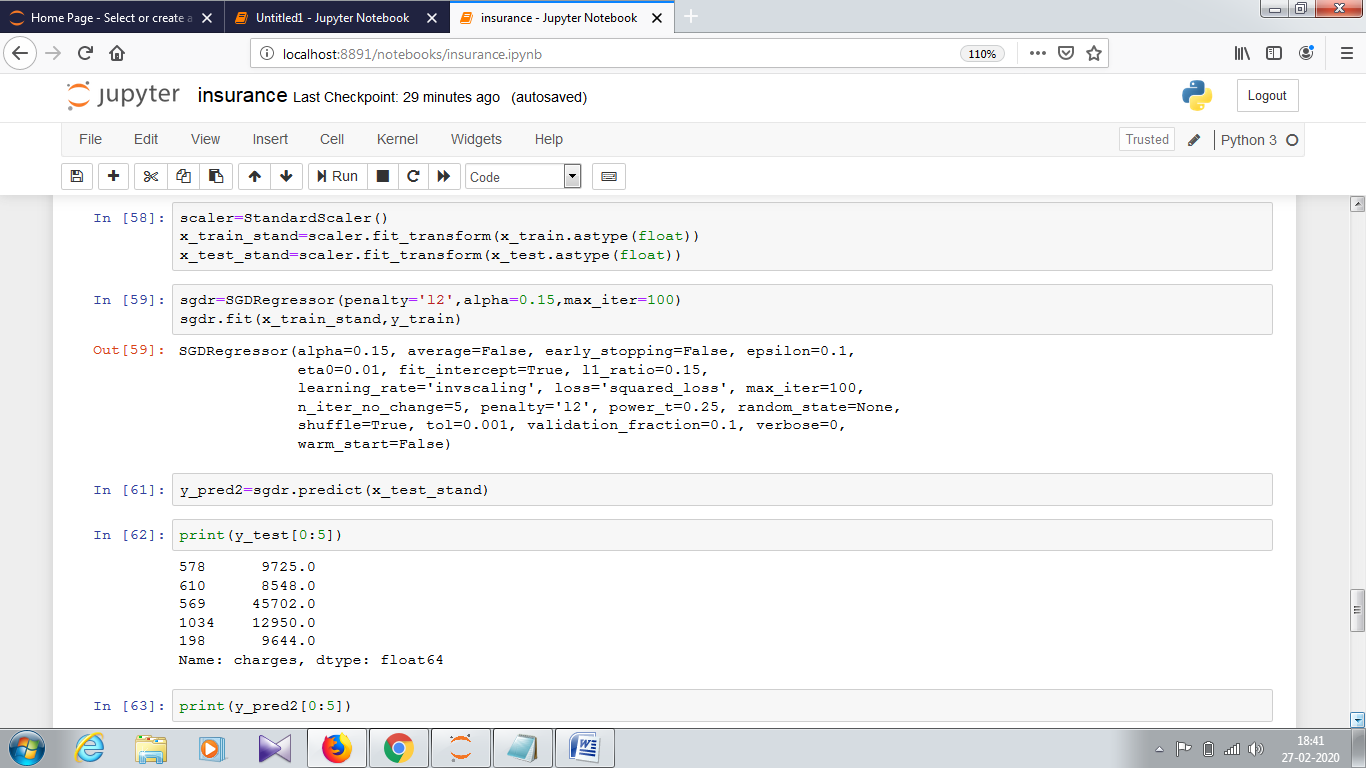
scaler=StandardScaler()

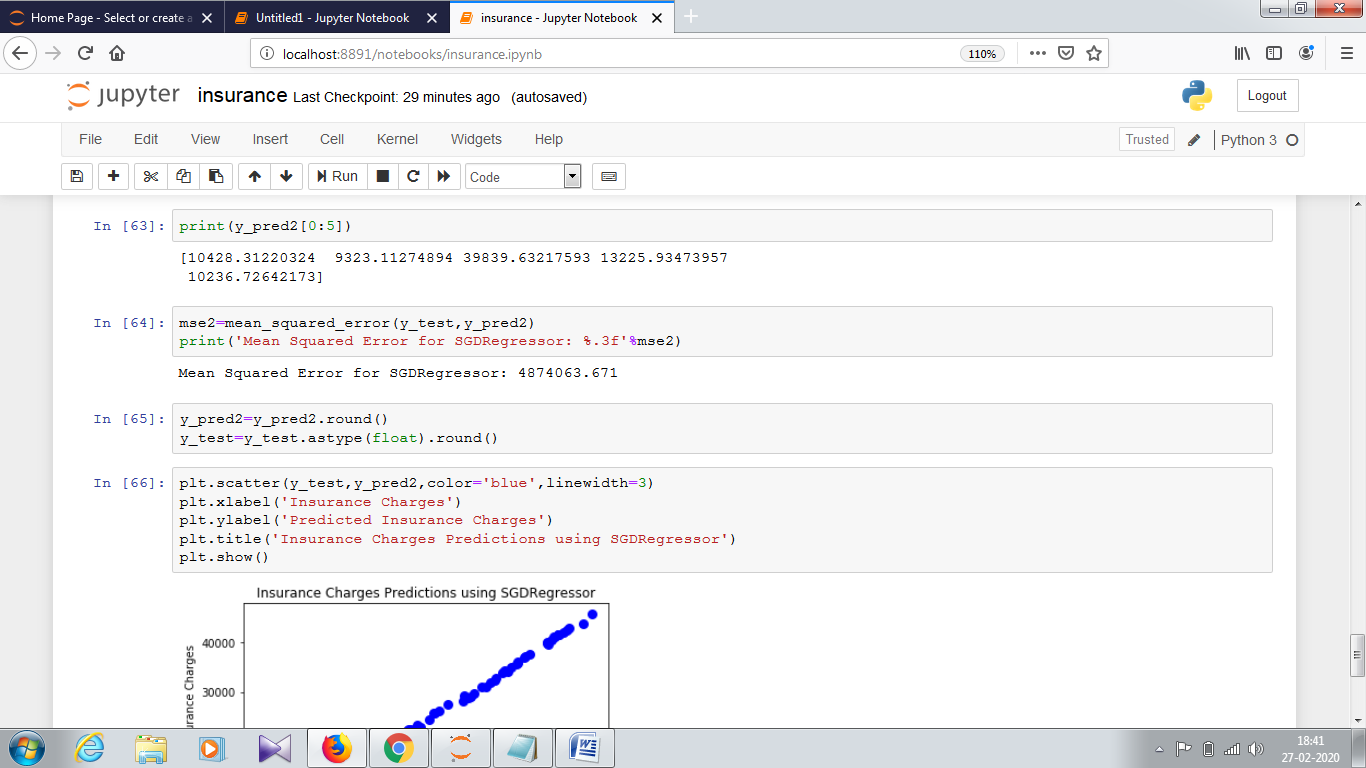
x\_train\_stand=scaler.fit\_transform(x\_train.astype(float))

x\_test\_stand=scaler.fit\_transform(x\_test.astype(float))

sgdr=SGDRegressor(penalty='l2',alpha=0.15,max\_iter=100)

sgdr.fit(x\_train\_stand,y\_train)





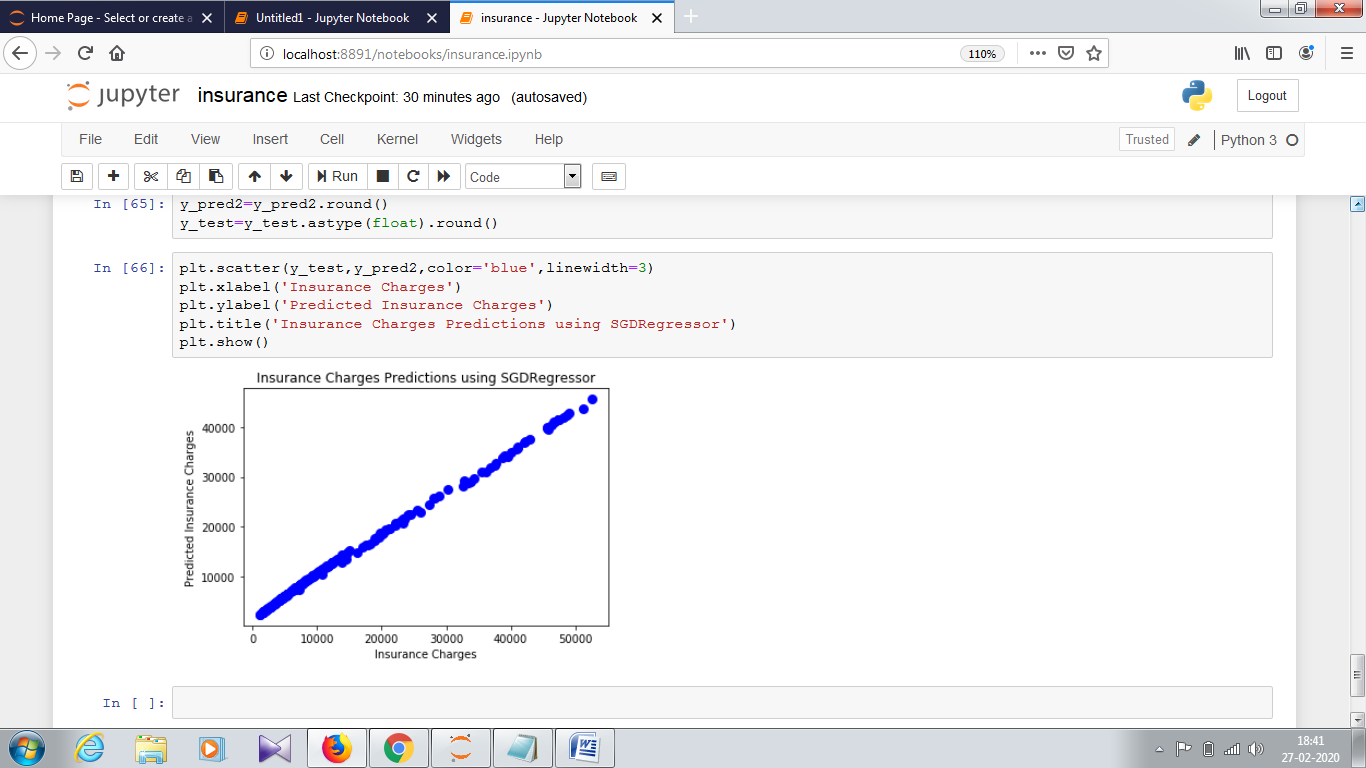
plt.scatter(y\_test,y\_pred2,color='blue',linewidth=3)

plt.xlabel('Insurance Charges')

plt.ylabel('Predicted Insurance Charges')

plt.title('Insurance Charges Predictions using SGDRegressor')

plt.show()



**CONCLUSION:**

Insurance charges are calculated by using Linear Regression and SGDRegressor.Then calculated the training and testing dataset.