**MACHINE LEARNING**

**DATASET:**

**Ecommerce Customers**

**NAME: D.VIGNESH KUMAR**

**ROLL NO: 195214174**

**CLASS:I M.Sc-CS-A**

**SHIFT: I**

**­­­**

**MACHINE LEARNING PROJECT**

In this project, I have used Linear Regression,Classification and clustering Machine Learning model for **Ecommerce Customers** dataset from Kaggle website.

**PROBLEM STATEMENT:**

We have a data of customers transaction details.We will use logistic regression (classification) algorithm ,here we are using **Ecommerce customer.csv** dataset.it has 8 columns and 488 rows.

**ML METHODOLOGY:**

Linear Regression is the methodology used for training and testing the dataset. Linear Regression is a method of modeling a target value based on independent predictors. This method is mostly used for forecasting and finding out cause and effect relationship between variables. Linear Regression techniques mostly differ based on the number of independent variables and the type of relationship between the independent and dependent variables.

**DATASET DESCRIPTION:**

**Columns:**

Email

Address

Avatar

Avg. Session Length

Time on App

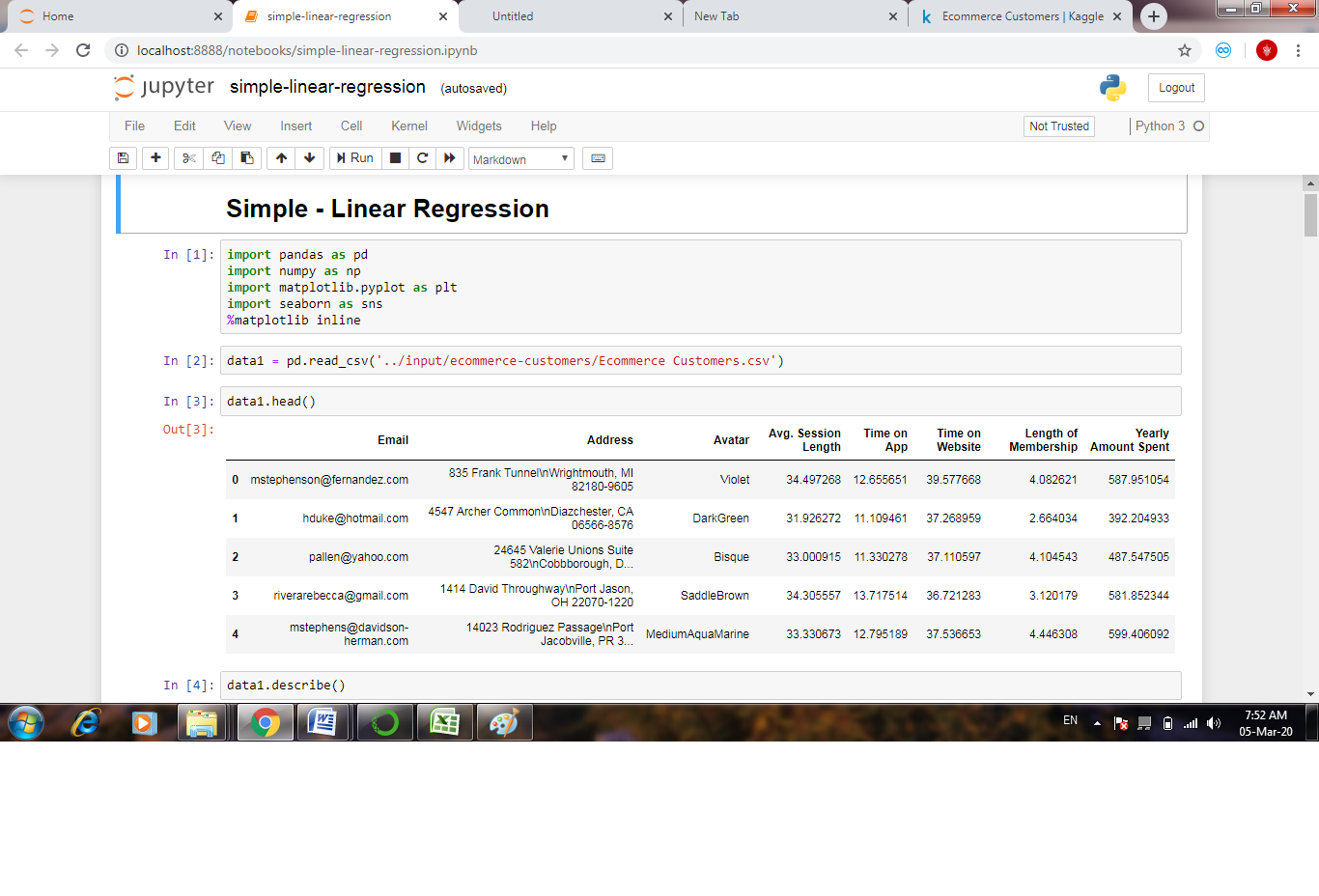
Time on Website

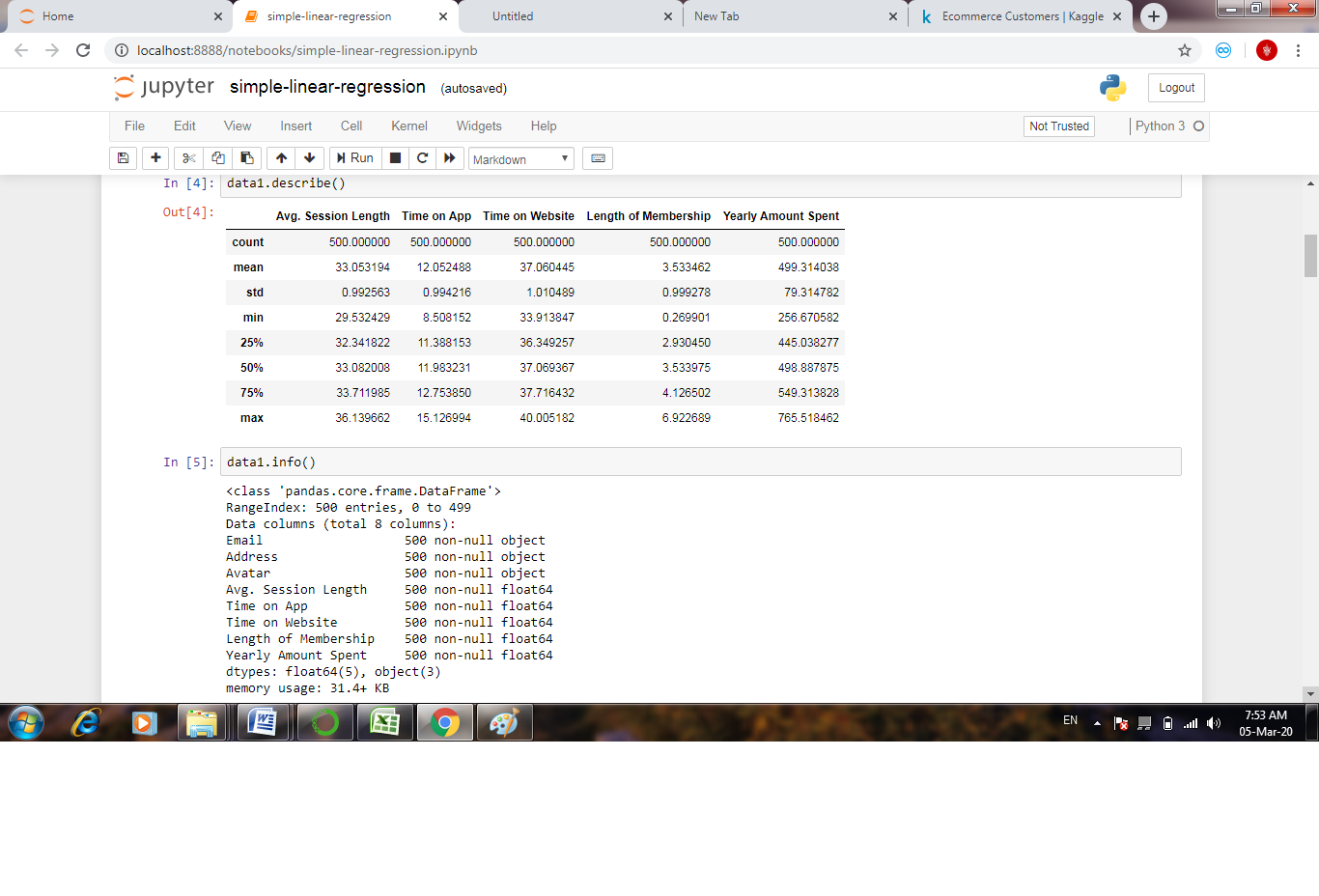
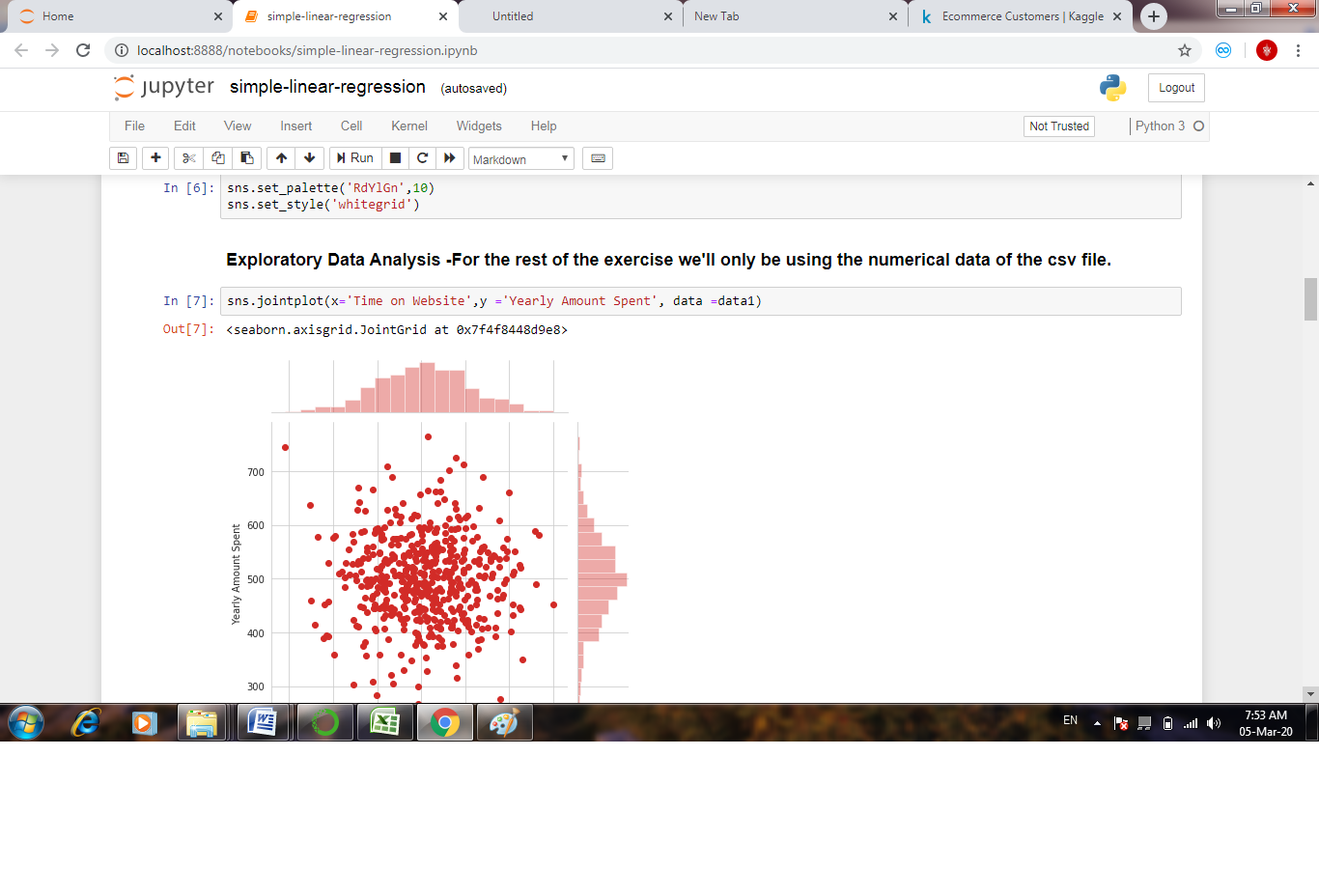
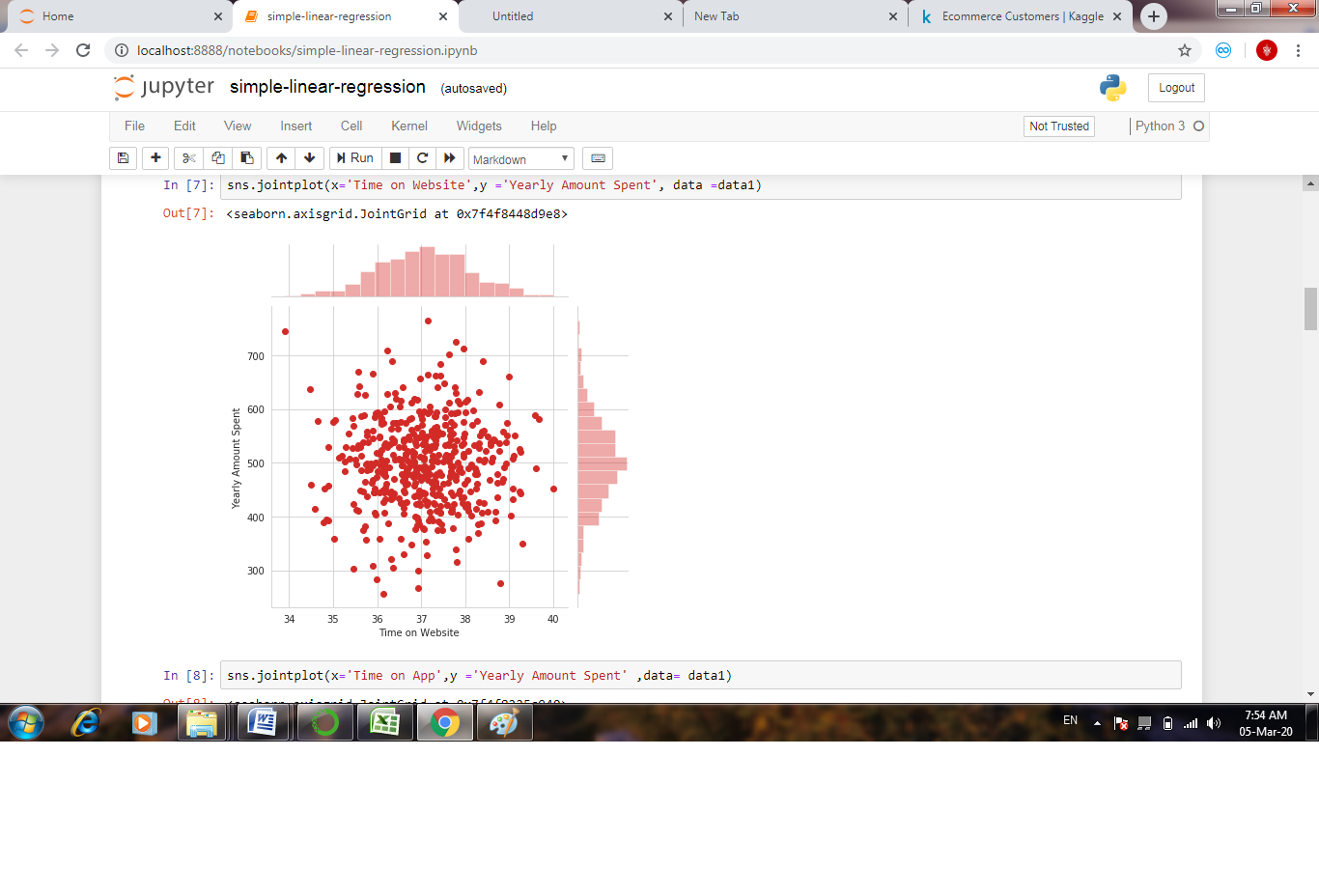
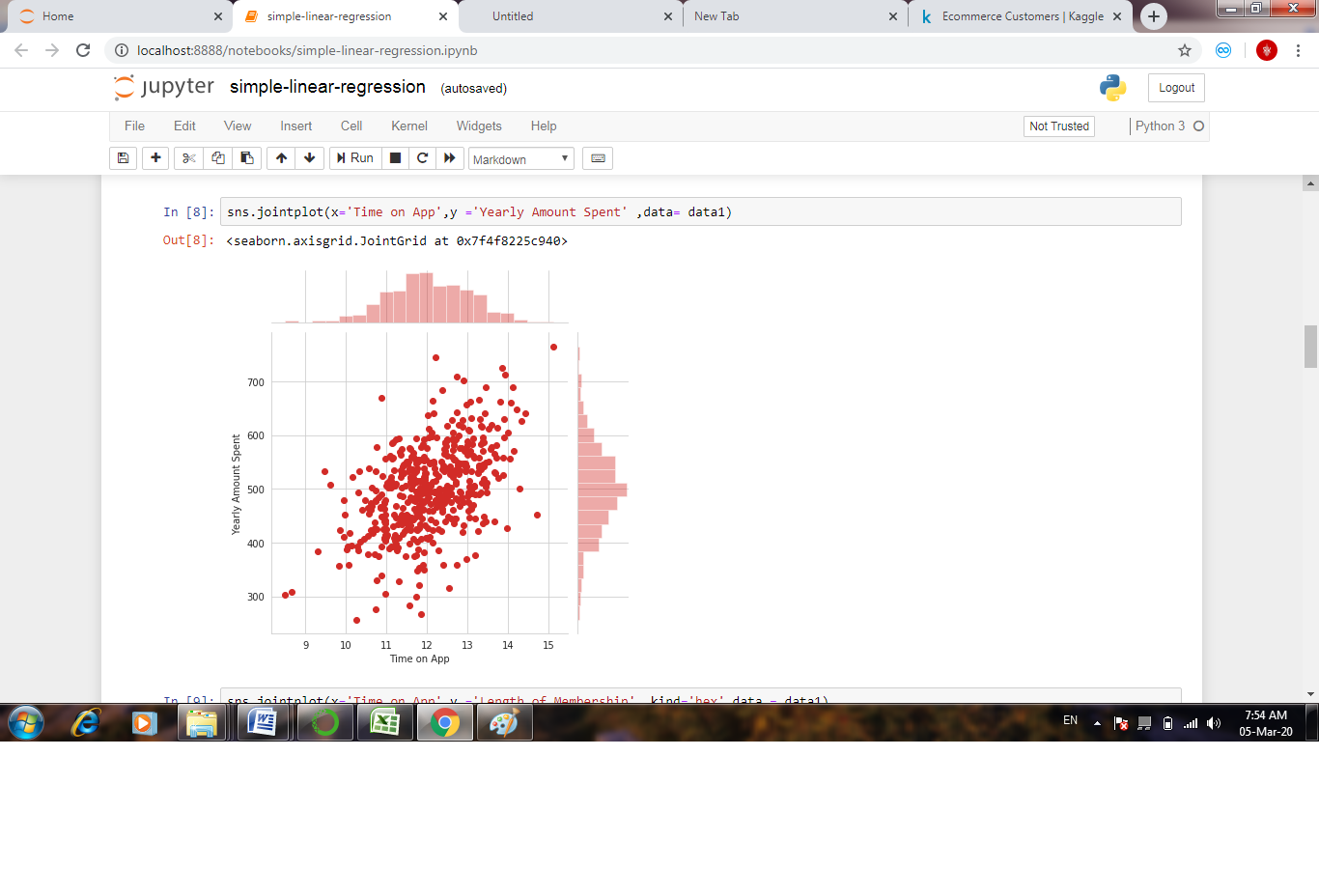
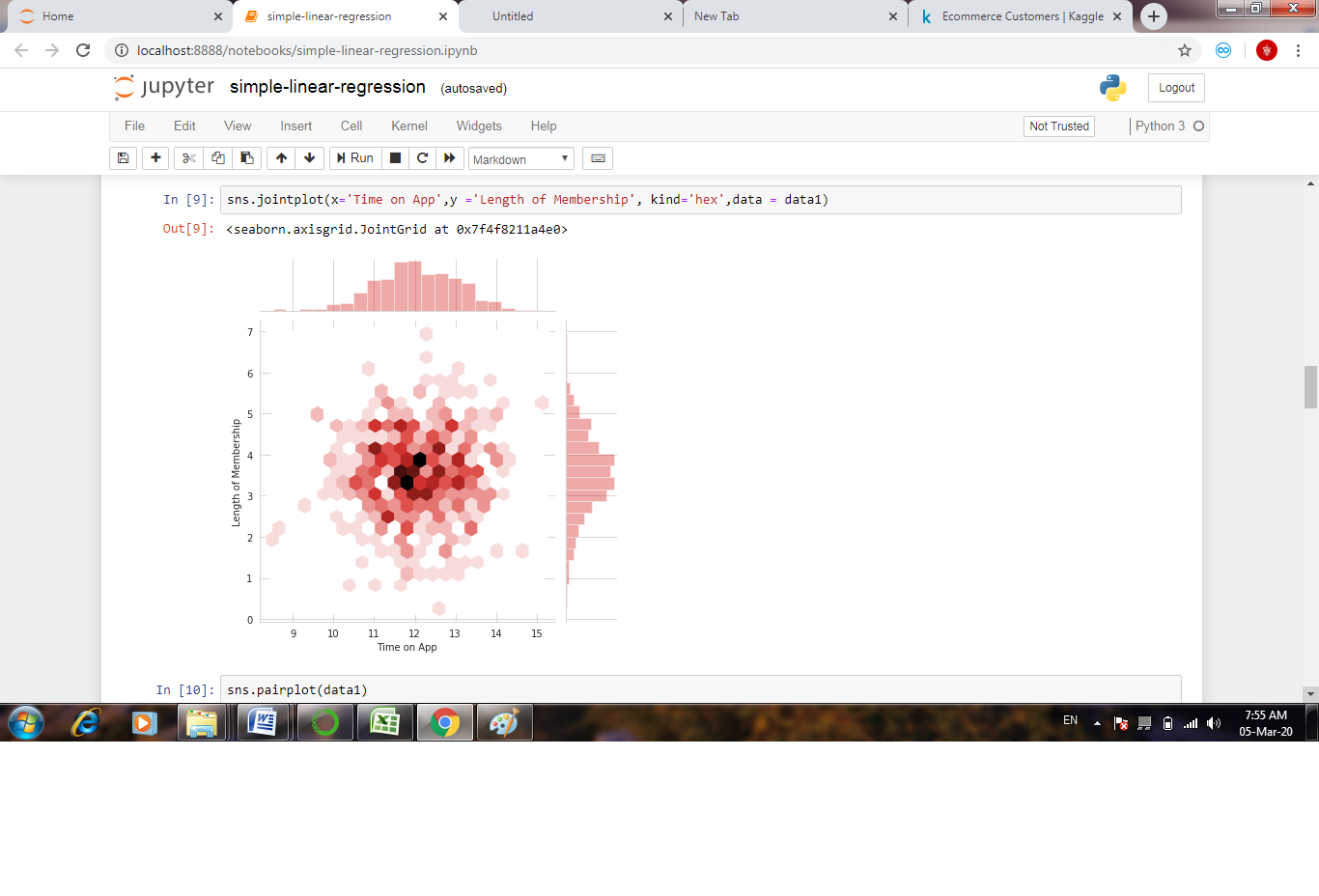
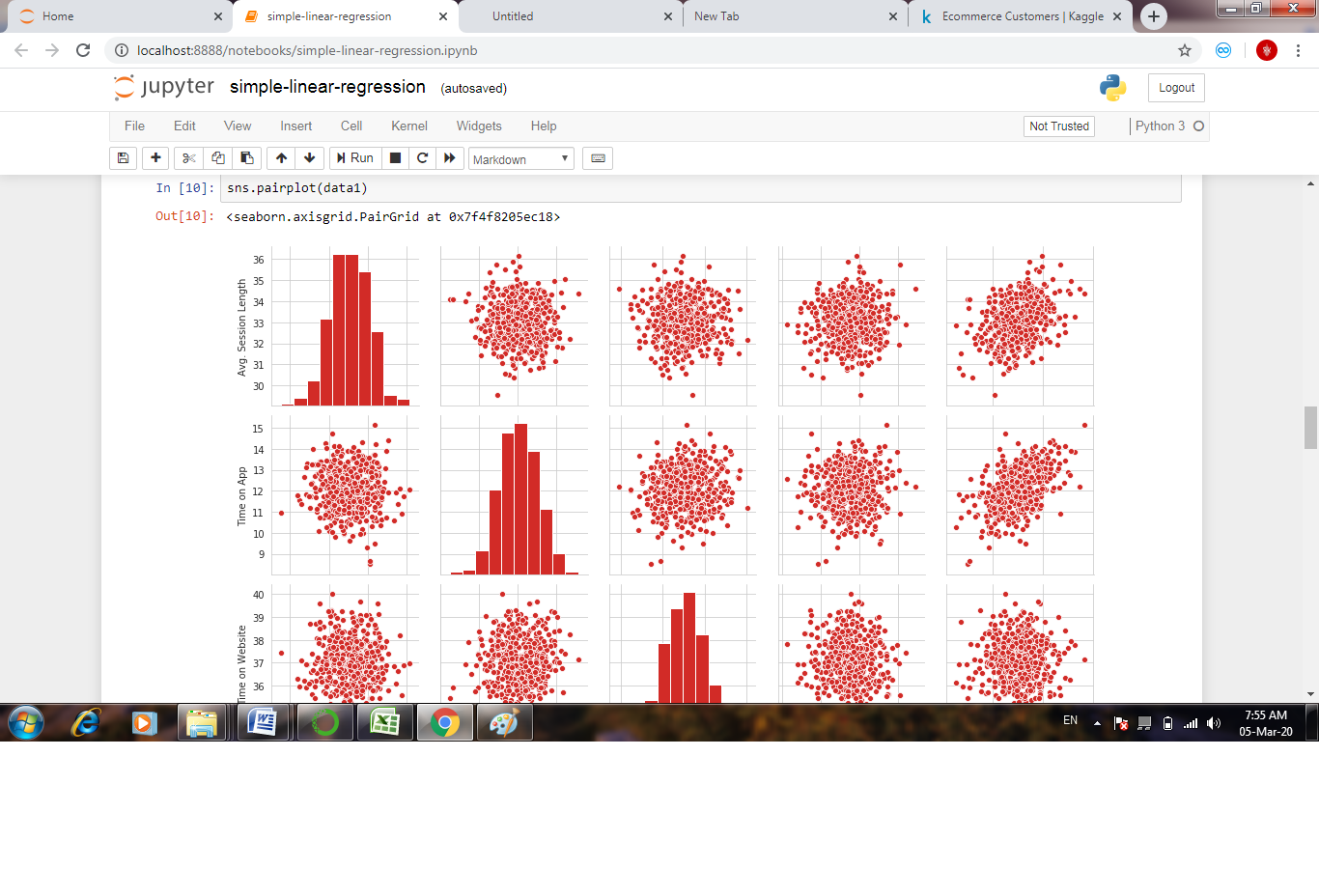
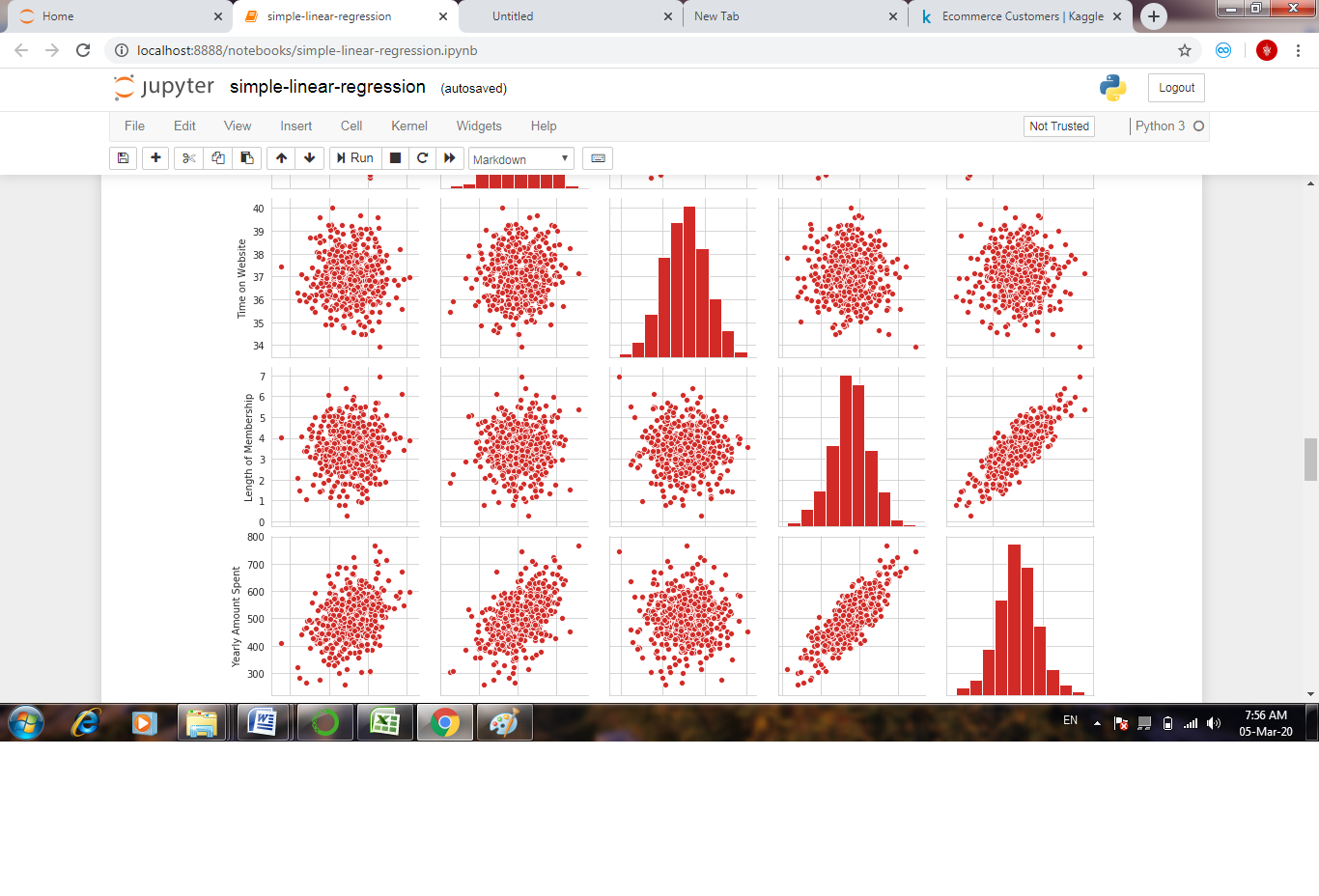
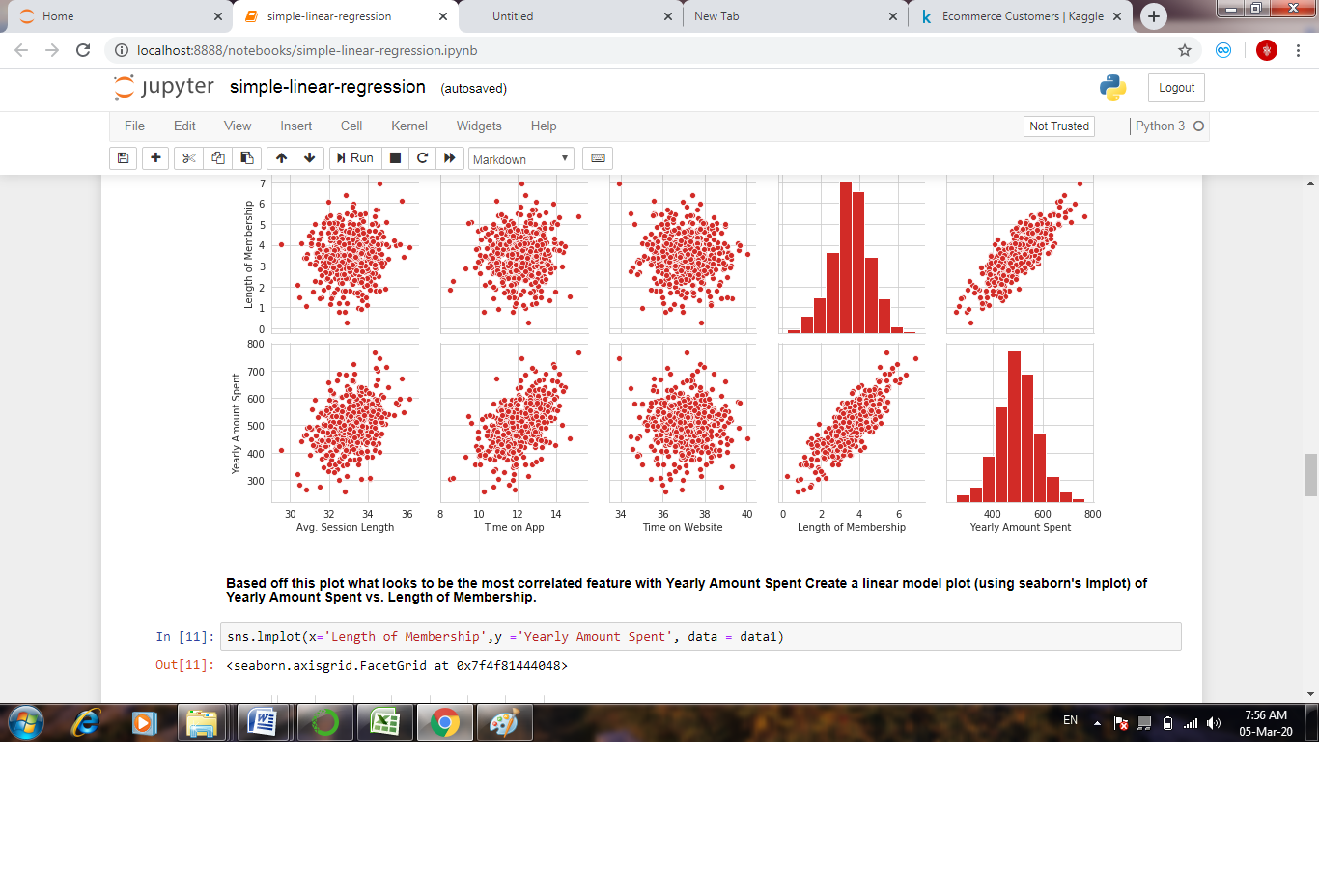
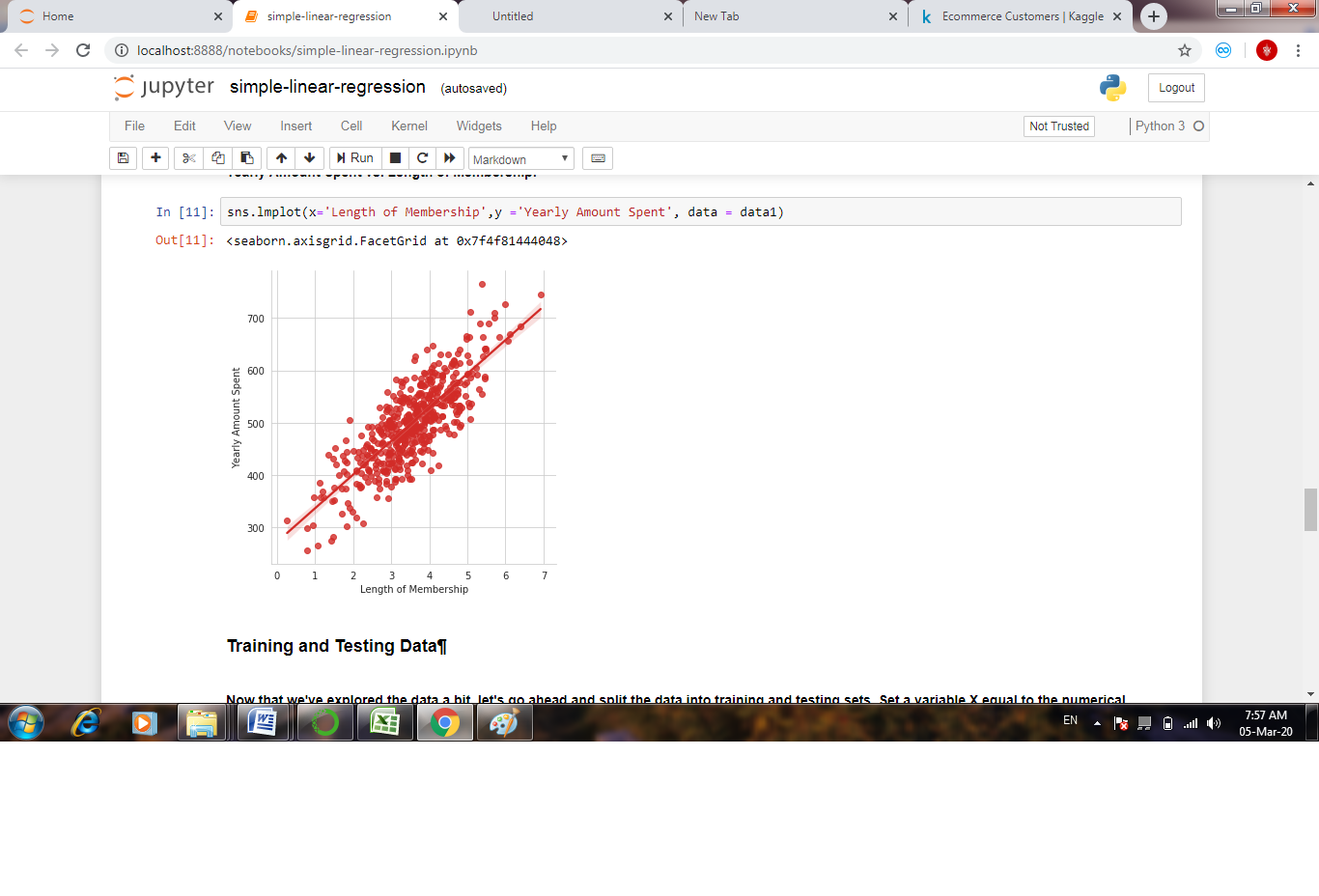
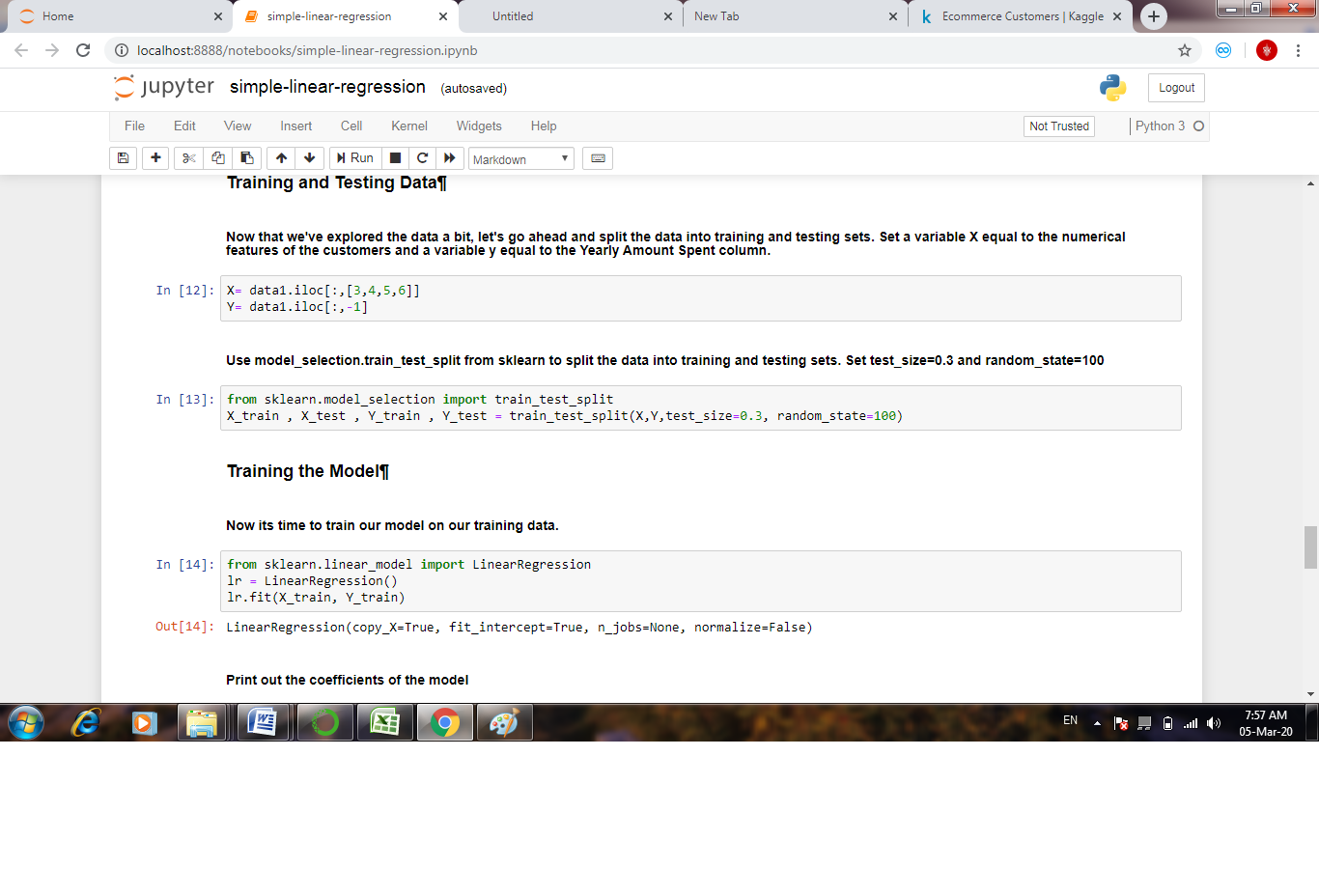
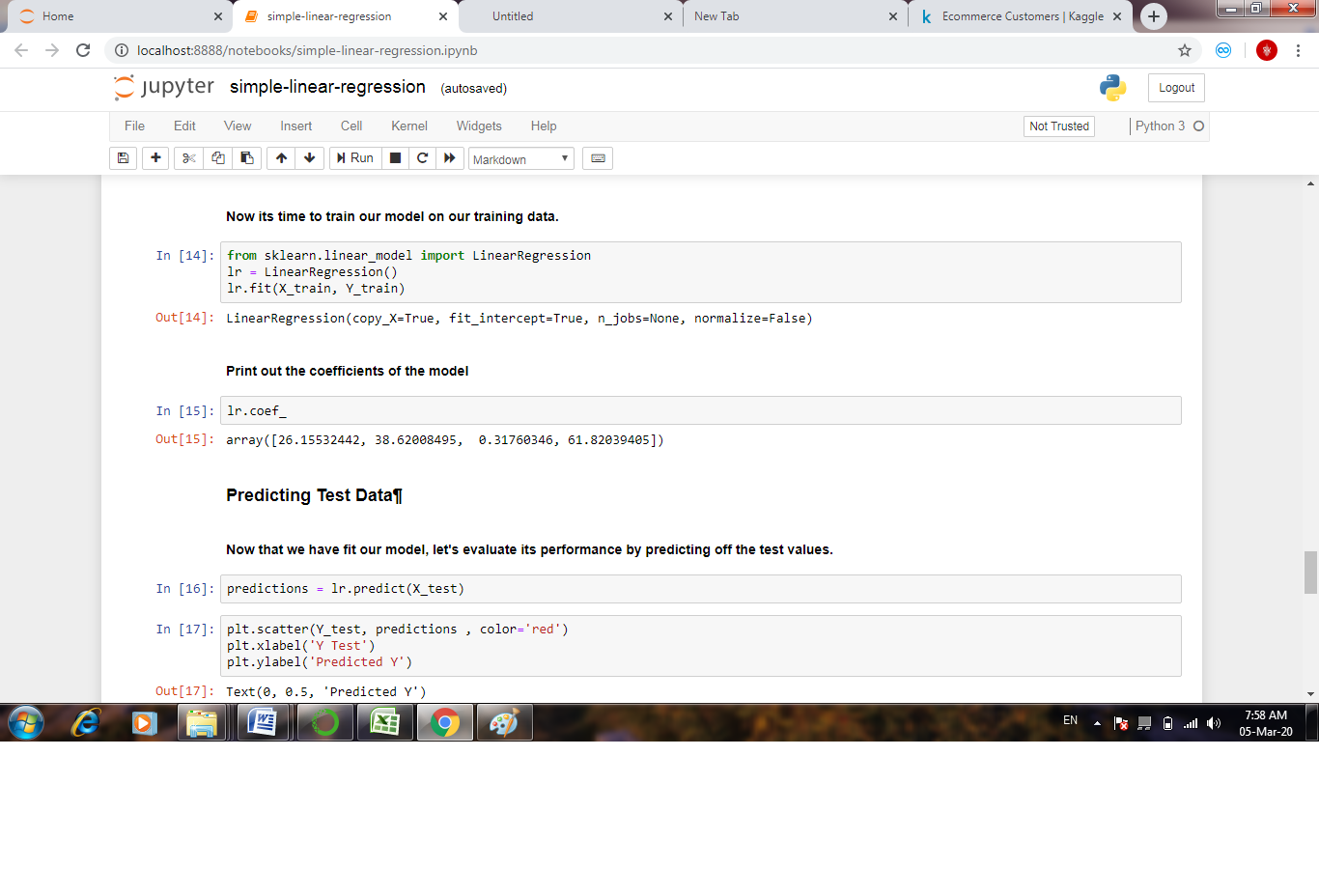
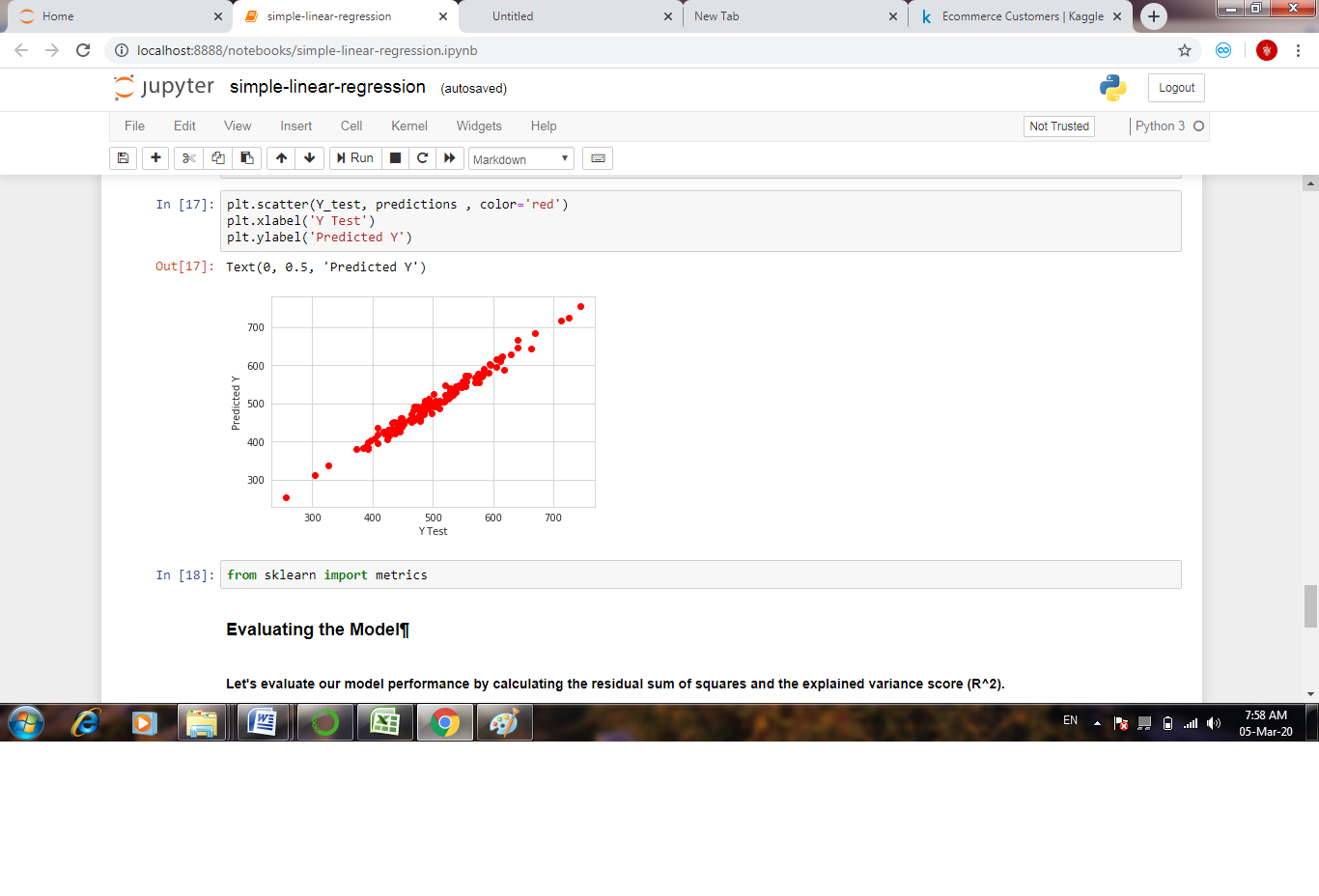
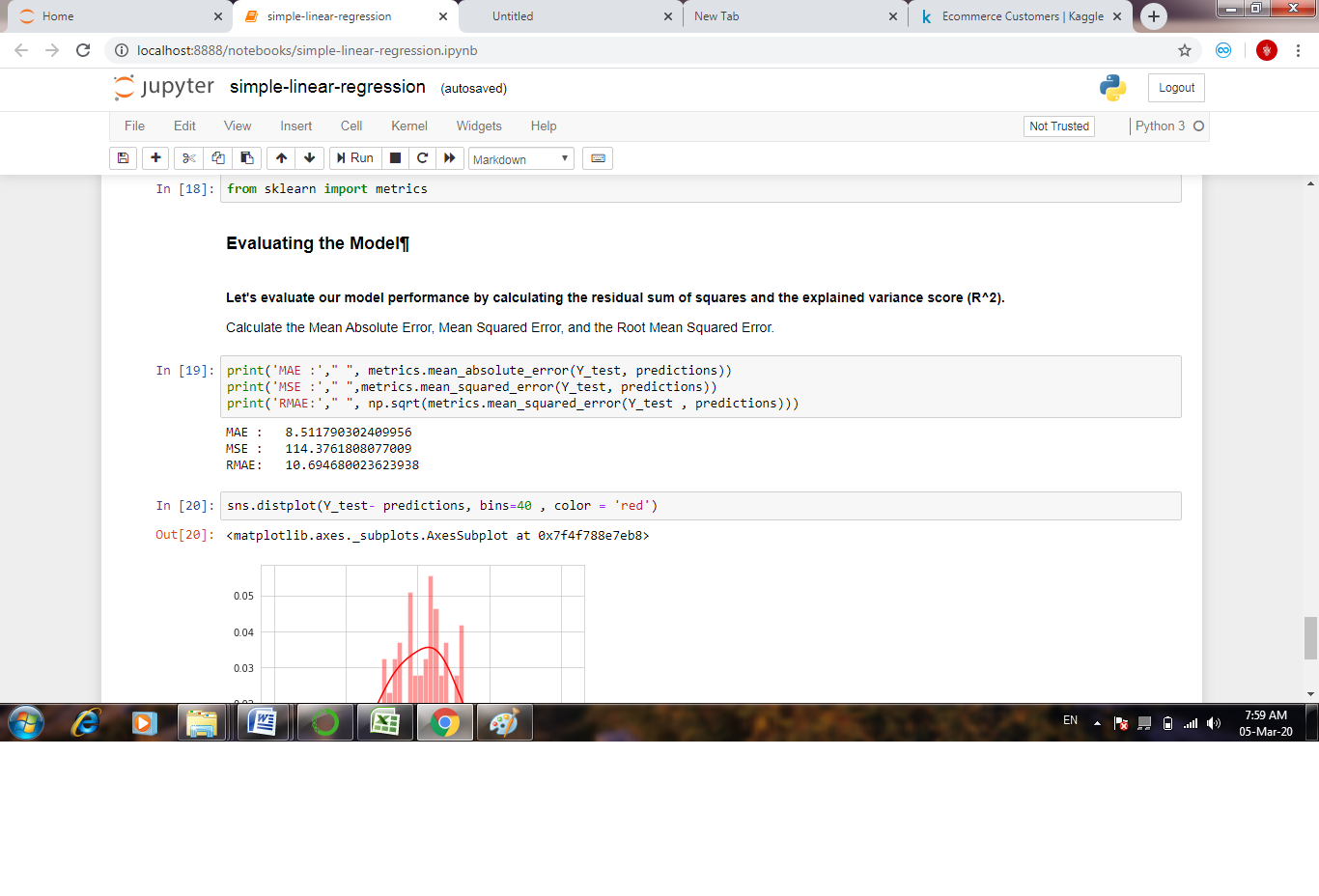
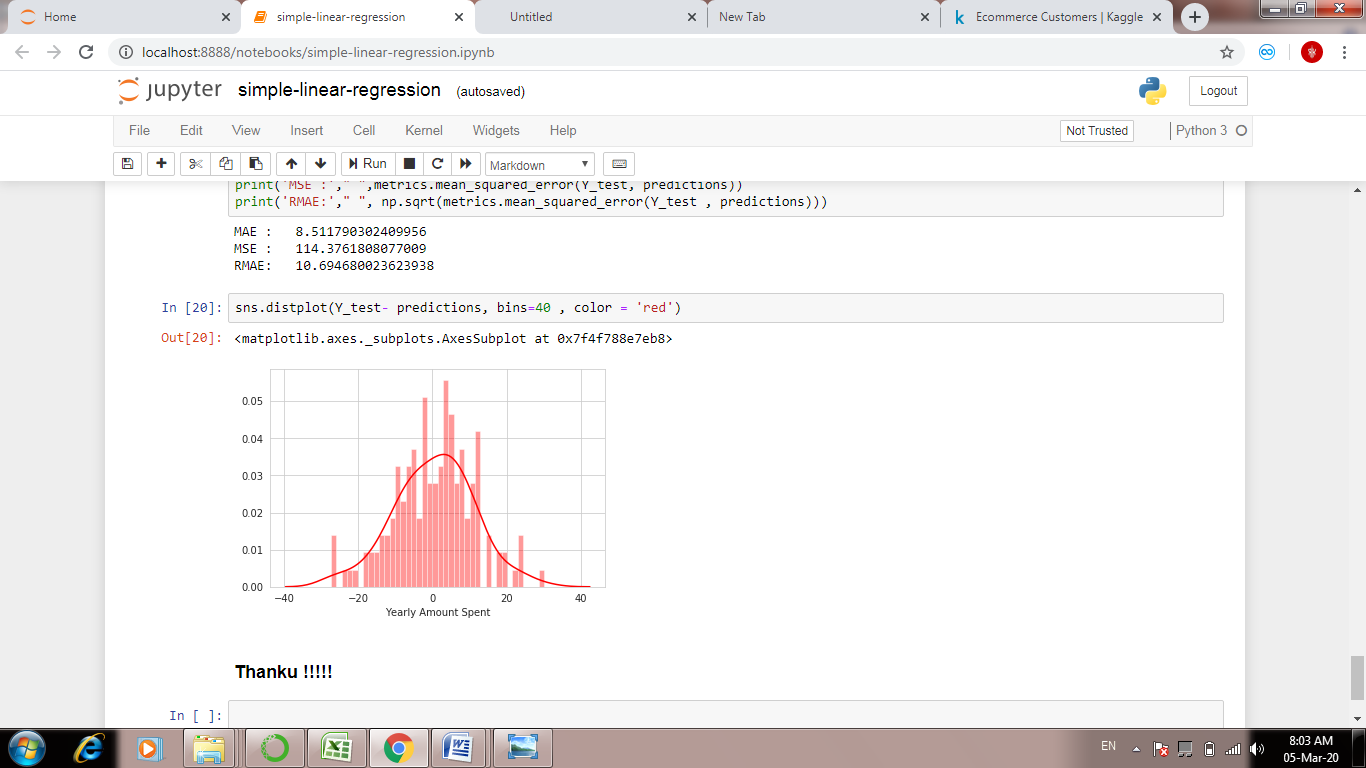
Length of Membership

Yearly Amount Spent

**PRE-PROCESSING:**

Pre-processing refers to the transformations applied to our data before feeding it to the algorithm.



**CODING:**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

%matplotlib inline

data1 = pd.read\_csv('../input/ecommerce-customers/Ecommerce Customers.csv')

data1.head()

data1.describe()

data1.info()

sns.set\_palette('RdYlGn',10)

sns.set\_style('whitegrid')

sns.jointplot(x='Time on Website',y ='Yearly Amount Spent', data =data1)

sns.jointplot(x='Time on App',y ='Yearly Amount Spent' ,data= data1)

sns.jointplot(x='Time on App',y ='Length of Membership', kind='hex',data = data1)

sns.pairplot(data1)

sns.lmplot(x='Length of Membership',y ='Yearly Amount Spent', data = data1)

X= data1.iloc[:,[3,4,5,6]]

Y= data1.iloc[:,-1]

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

X\_train , X\_test , Y\_train , Y\_test = train\_test\_split(X,Y,test\_size=0.3, random\_state=100)

from sklearn.linear\_model import LinearRegression

lr = LinearRegression()

lr.fit(X\_train, Y\_train)

lr.coef\_

predictions = lr.predict(X\_test)

plt.scatter(Y\_test, predictions , color='red')

plt.xlabel('Y Test')

plt.ylabel('Predicted Y')

from sklearn import metrics

print('MAE :'," ", metrics.mean\_absolute\_error(Y\_test, predictions))

print('MSE :'," ",metrics.mean\_squared\_error(Y\_test, predictions))

print('RMAE:'," ", np.sqrt(metrics.mean\_squared\_error(Y\_test , predictions)))

sns.distplot(Y\_test- predictions, bins=40 , color = 'red')

**CONCLUSION**

This is about iris dataset.all the dataset are collected and its pre-processed,trained,tested and reported successfully.this is development for simple linear regression to bring smarter one for people.

**URL:**