**22.How to split training and test data set in simple linear regression using python?**

**Objective:**

* To split training and test data set in a simple linear regression model in python.

**Process:**

* Import regression model from scikit-learn library.
* Import train\_test\_split from scikit-learn library.
* Load the sample data set.
* Define independent(X) and dependent(Y) variable.
* Define the training data set from data sample.

**Input:**

* Data set(sample).

**Output:**

* Training data set.

**Source code:**

#import needed library

import matplotlib.pyplot as plt

import pandas as pd

import numpy as np

#import linear model and train\_test\_split

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

from sklearn.linear\_model import LinearRegression

#load the sample data

data={'salary':[100,200,300,400,500,400,300,200,100,50], 'age':[25,26,25,23,30,29,23,23,25,25],

'rating':[4,3.24,2.5,2.25,2,2.25,2.5,2.75,3.2,4.2], 'bonus':[2500,1200,900,3000,1800,1400,850,250,750, 1000]}

#create the data frame

df=pd.DataFrame(data)

print(df)

#Define x(independent variable)

X = df.iloc[:, :1].values

print("Value of X is",X)

#Define y(dependent variable)

y = df.iloc[:,2].values

print("Value of y is",y)

#declare the train data set

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.3, random\_state=0)

#build linear model

regressor = LinearRegression()

#pass the X-train and y\_train

t=regressor.fit(X\_train, y\_train)

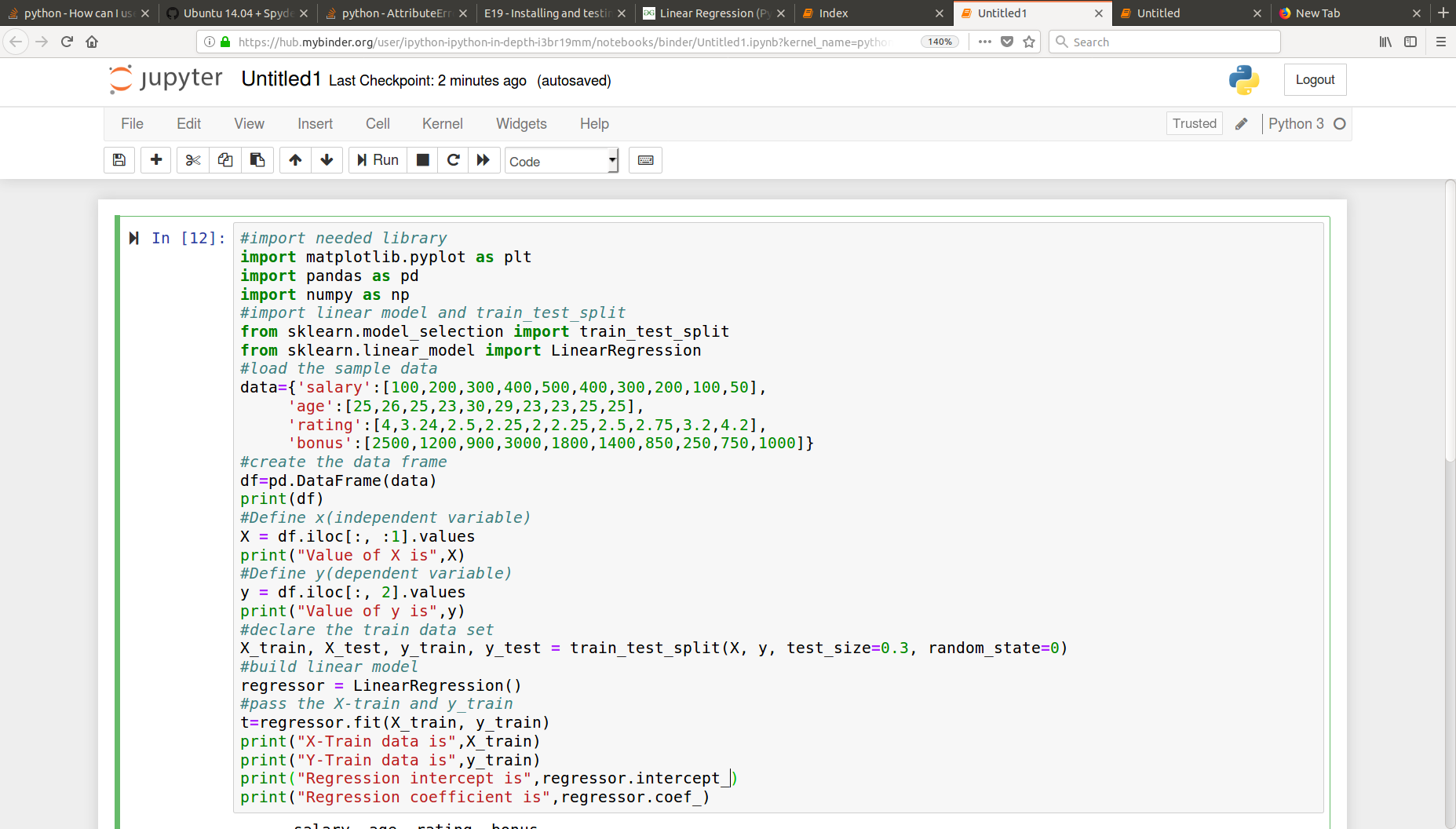
print("X-Train data is",X\_train)

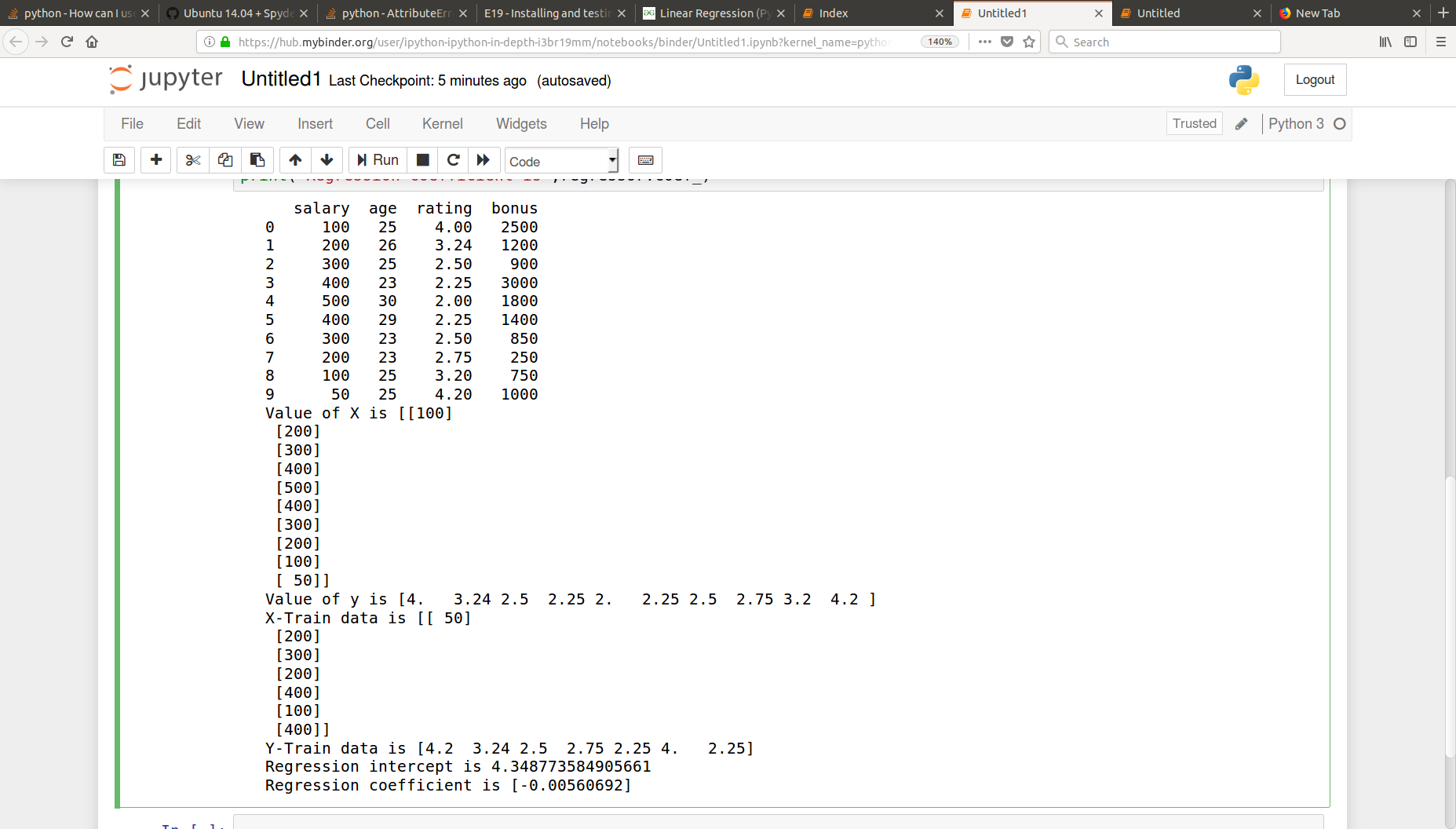
print("Y-Train data is",y\_train)

print("Regression intercept is",regressor.intercept\_)

print("Regression coefficient is",regressor.coef\_)

**Screen shot:**

****

****