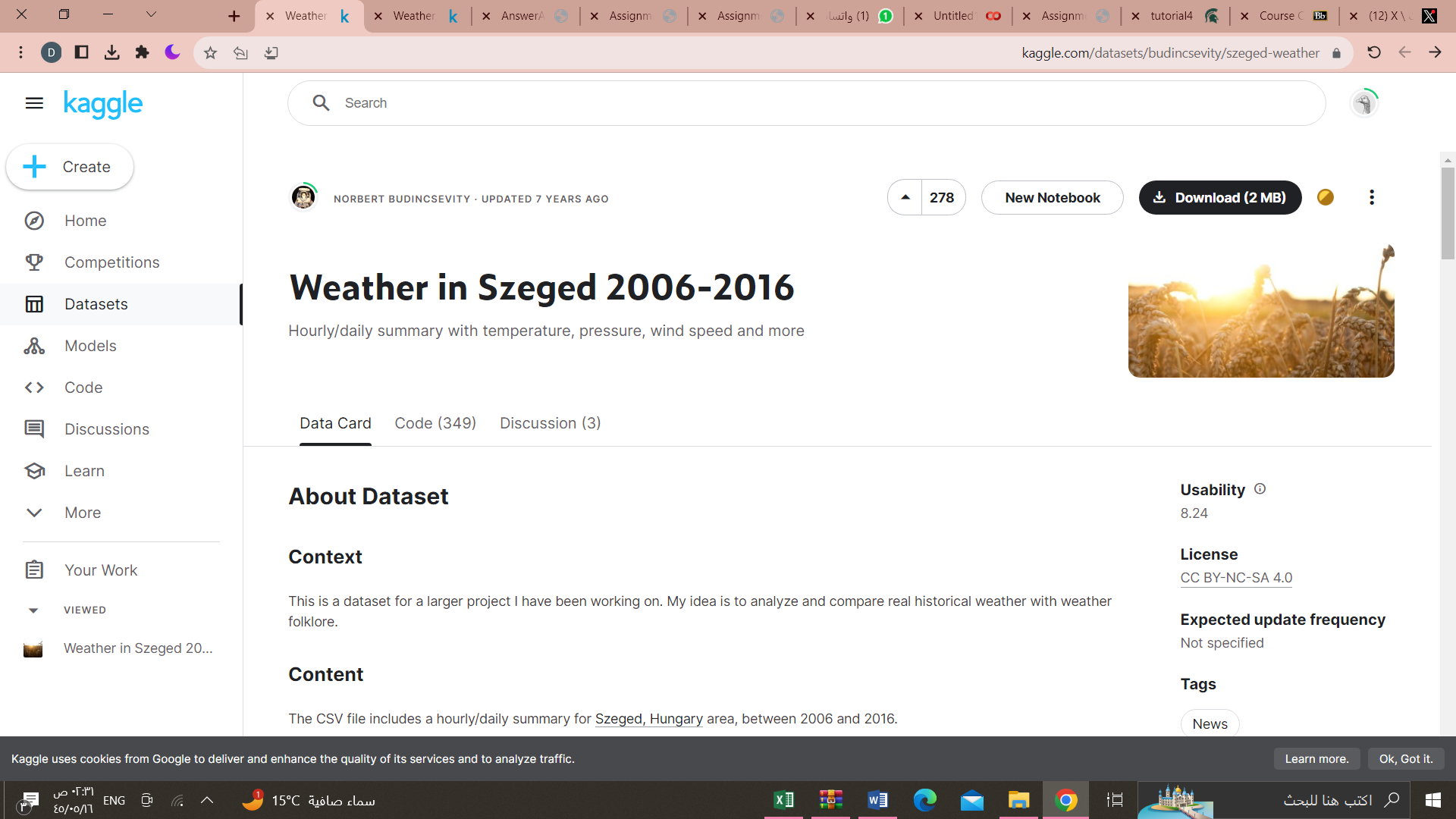
|  |  |  |
| --- | --- | --- |
| Assignment 5 | | |
| Regression | | |
| Student Name: saja eid | Section No: | Student ID 442050483 |

Use the weather dataset at the link <https://www.kaggle.com/datasets/budincsevity/szeged-weather>

to build a prediction model using the data provided.

* Download dataset

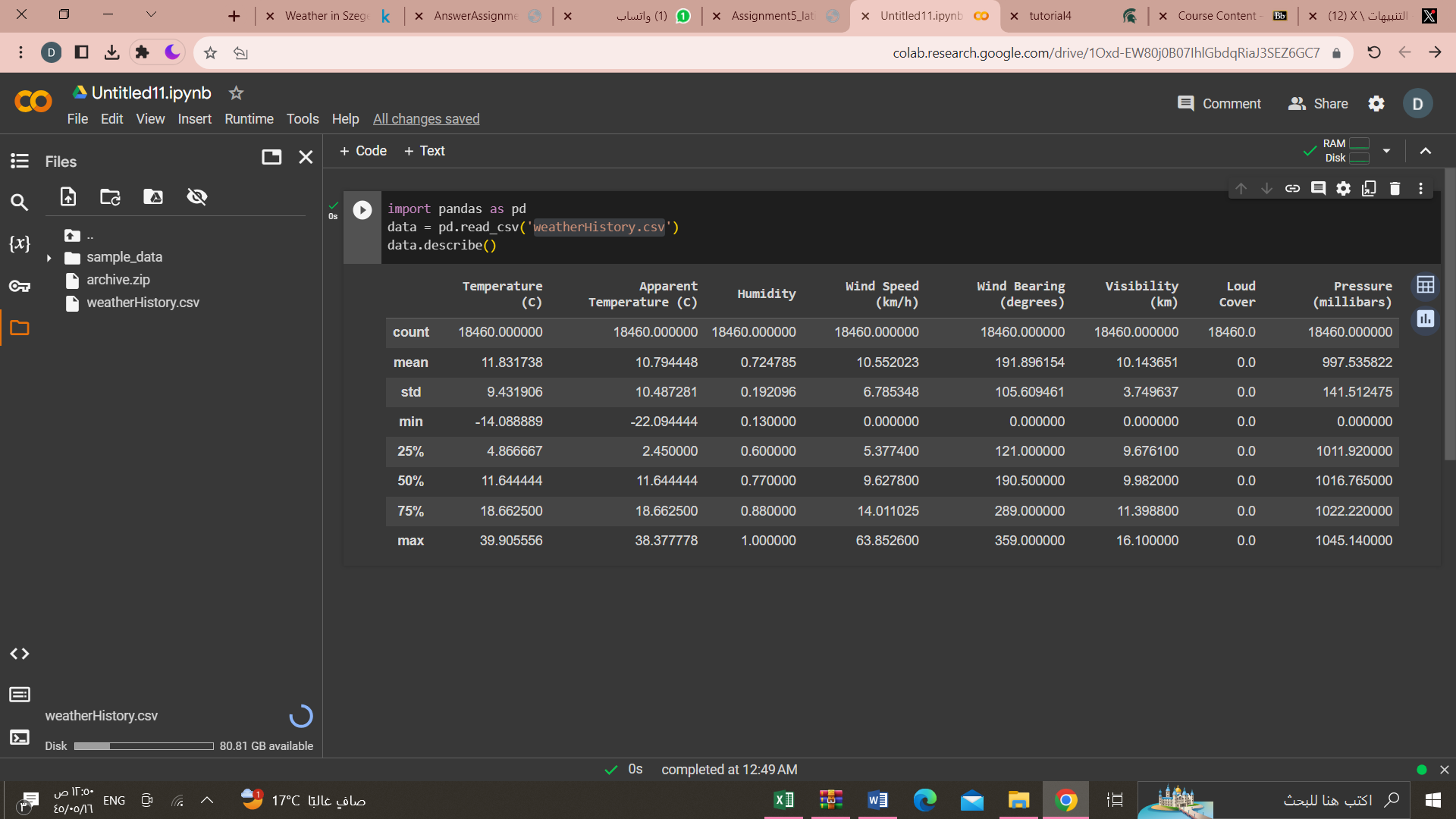


* Read the dataset

import pandas as pd

data = pd.read\_csv('weatherHistory.csv')

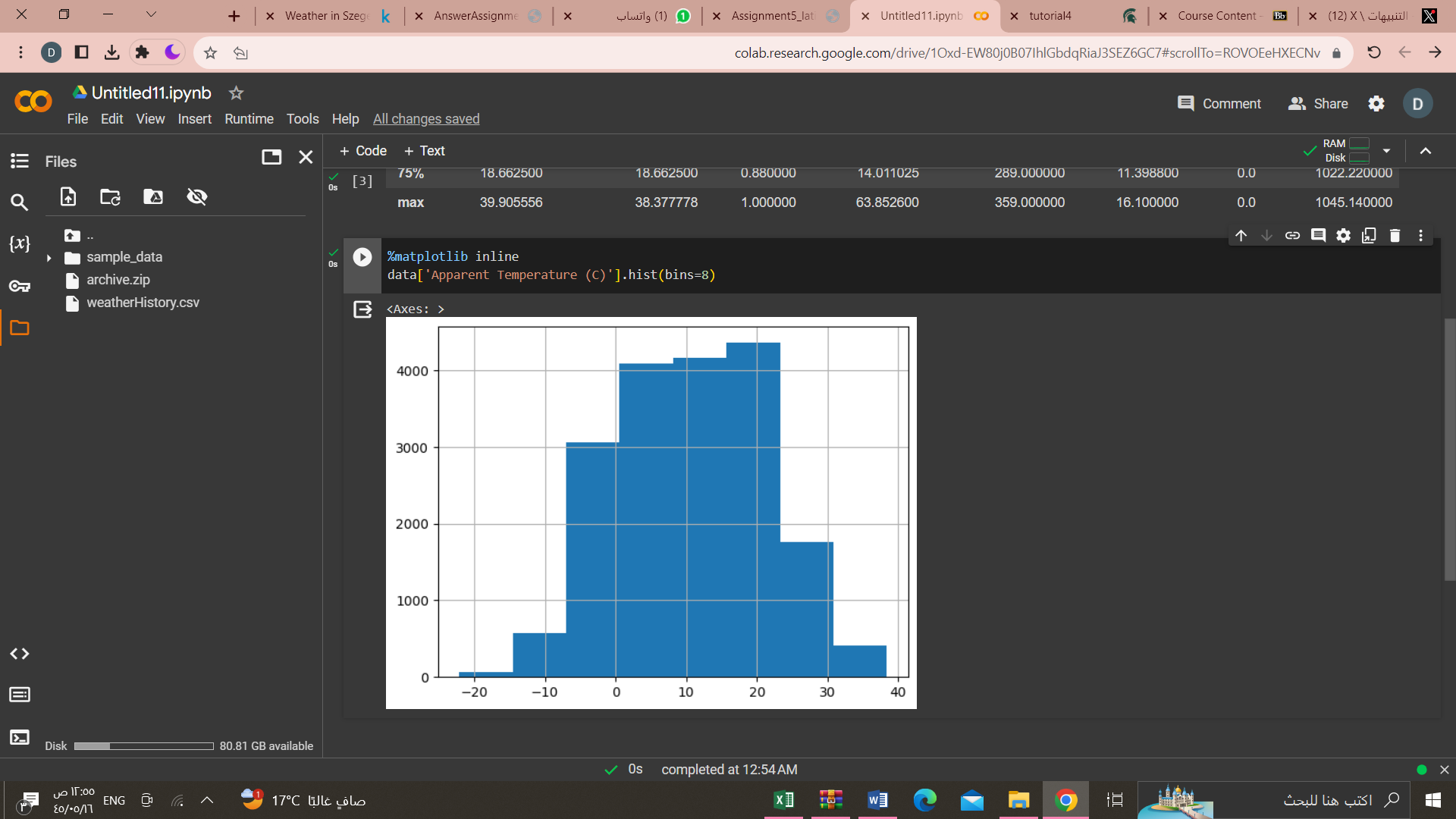
data.describe()



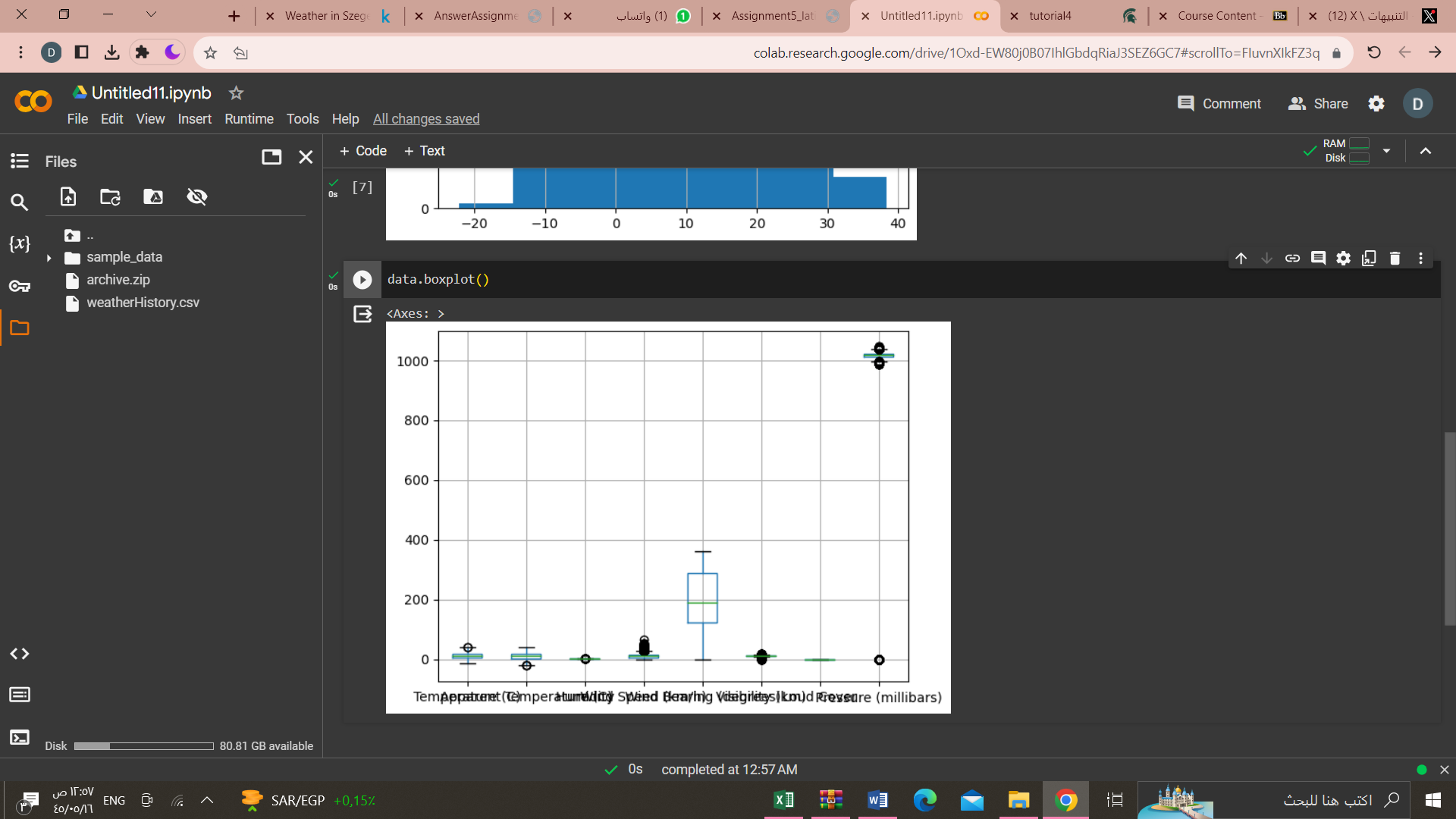
* Use visualization if possible.

%matplotlib inline

data['Apparent Temperature (C)'].hist(bins=8)



data.boxplot()

* 
* Split the dataset 80% for train-20% for test

import numpy as np

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

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_squared\_error

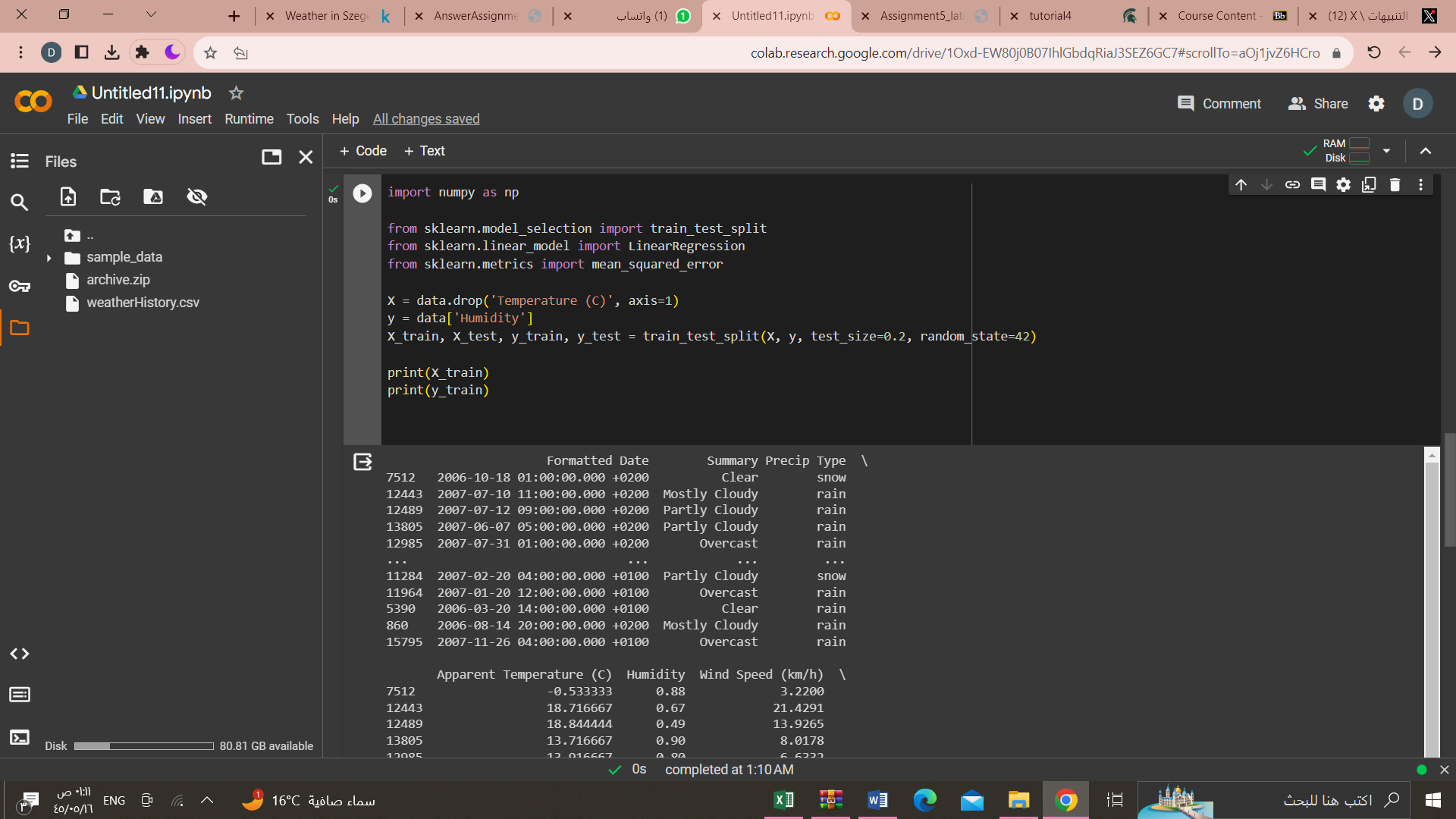
X = data.drop('Temperature (C)', axis=1)

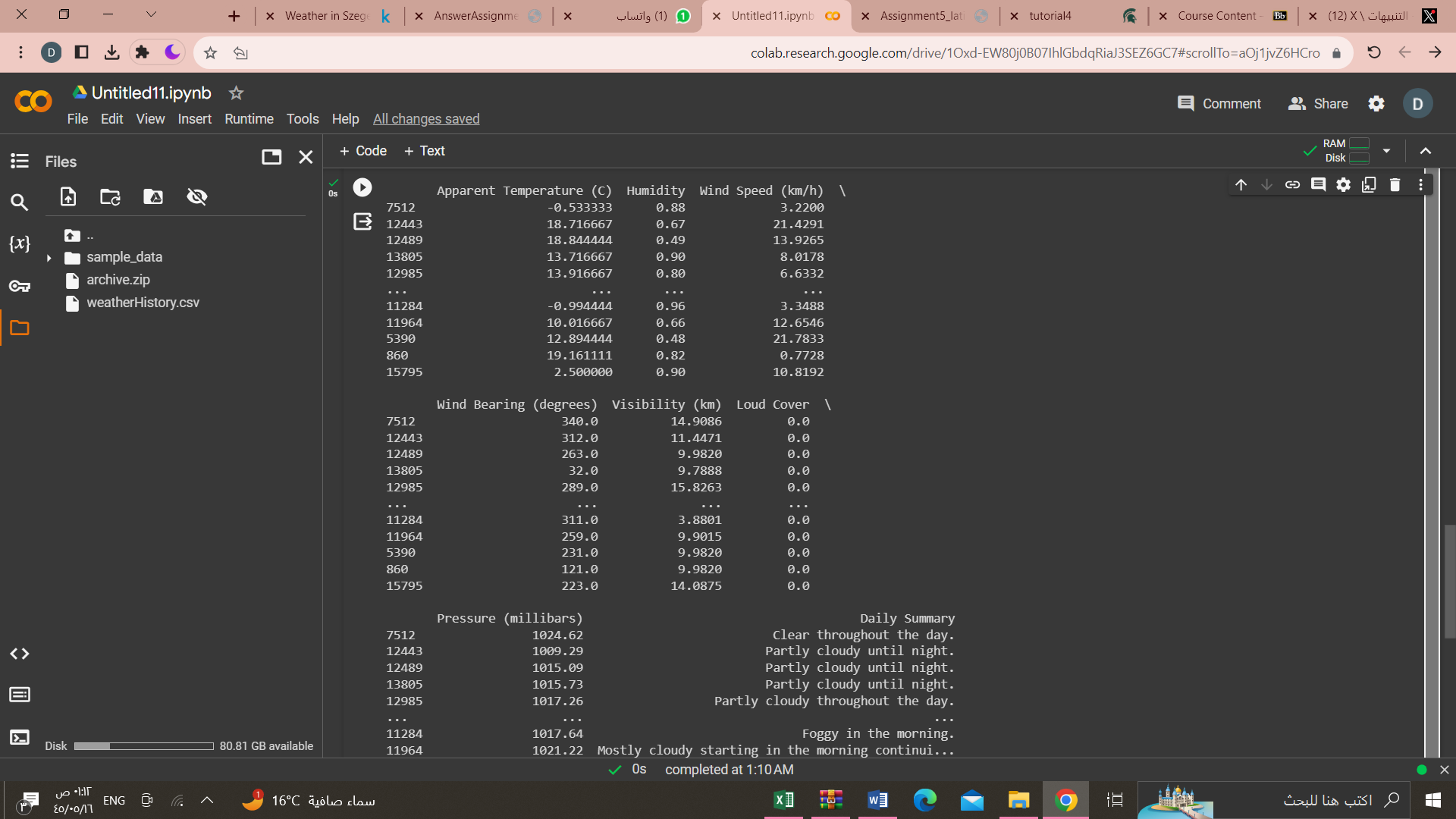
y = data['Humidity']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

print(X\_train)

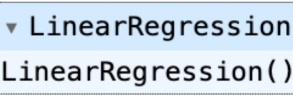
print(y\_train)



* 
* Build the model (linear regression, SVM) using the train dataset split.

linear regression

* X = data[['Temperature (C)','Humidity']]
* y = data['Apparent Temperature (C)']
* X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)
* model = LinearRegression()
* model.fit(X\_train, y\_train)



SVM

X = dataset[['Temperature (C)','Humidity']]

y = dataset['Apparent Temperature (C)']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2,

random\_state=43)

model = SVR()

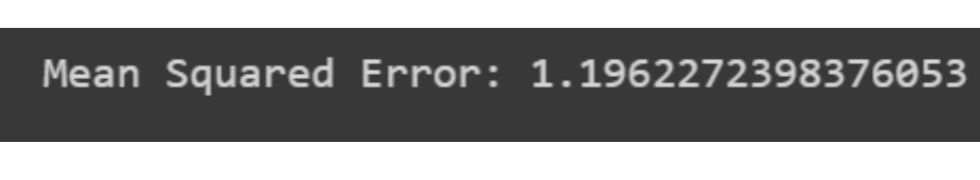
model.fit (X\_train, y\_train)



* Use the test dataset for evaluation.
* y\_pred=model.predict(X\_test)

mse = mean\_squared\_error(y\_test, y\_pred)

print('Mean Squared Error:', mse)

* 
* Show the performance of the model.
* model.score(X\_train, y\_train)
* model.score(X\_test, y\_test)

