**PROJECT REPORT**

Predicting apparent temperature in given climatic conditions

*­Concept: Linear Regression in Multiple Variables*

Introduction:

A dataset consisting of the temperature and various climatic conditions has been provided and the task is to predict the apparent temperature based on the different climatic conditions provided by the user.

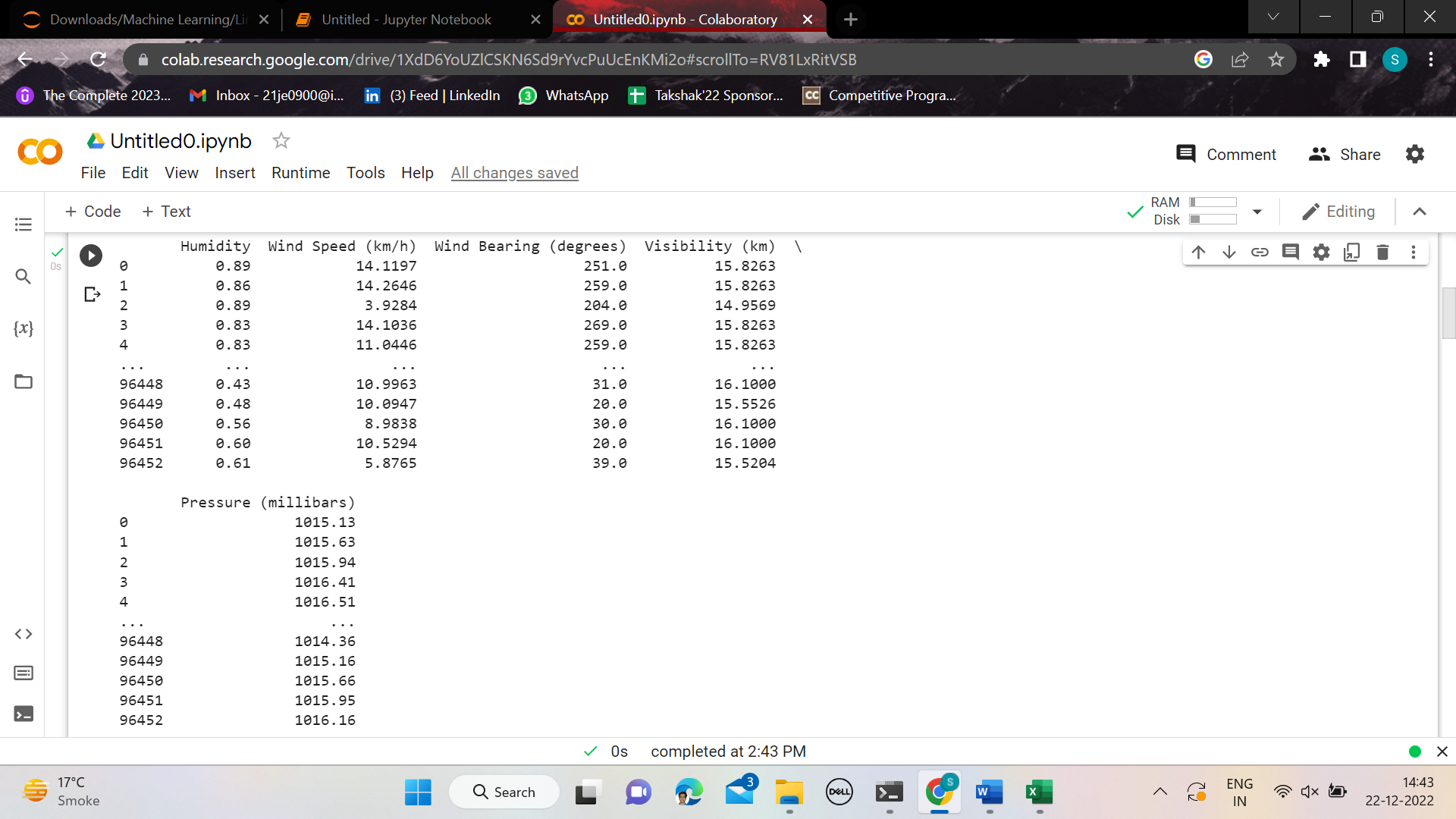
**The various climatic conditions are:**

* Humidity
* Wind Speed (in km/h)
* Wind Bearing (in degrees)
* Visibility (in km)
* Pressure (in millibars)

The output column in the training set consists of the recorded temperature in degrees Celsius.

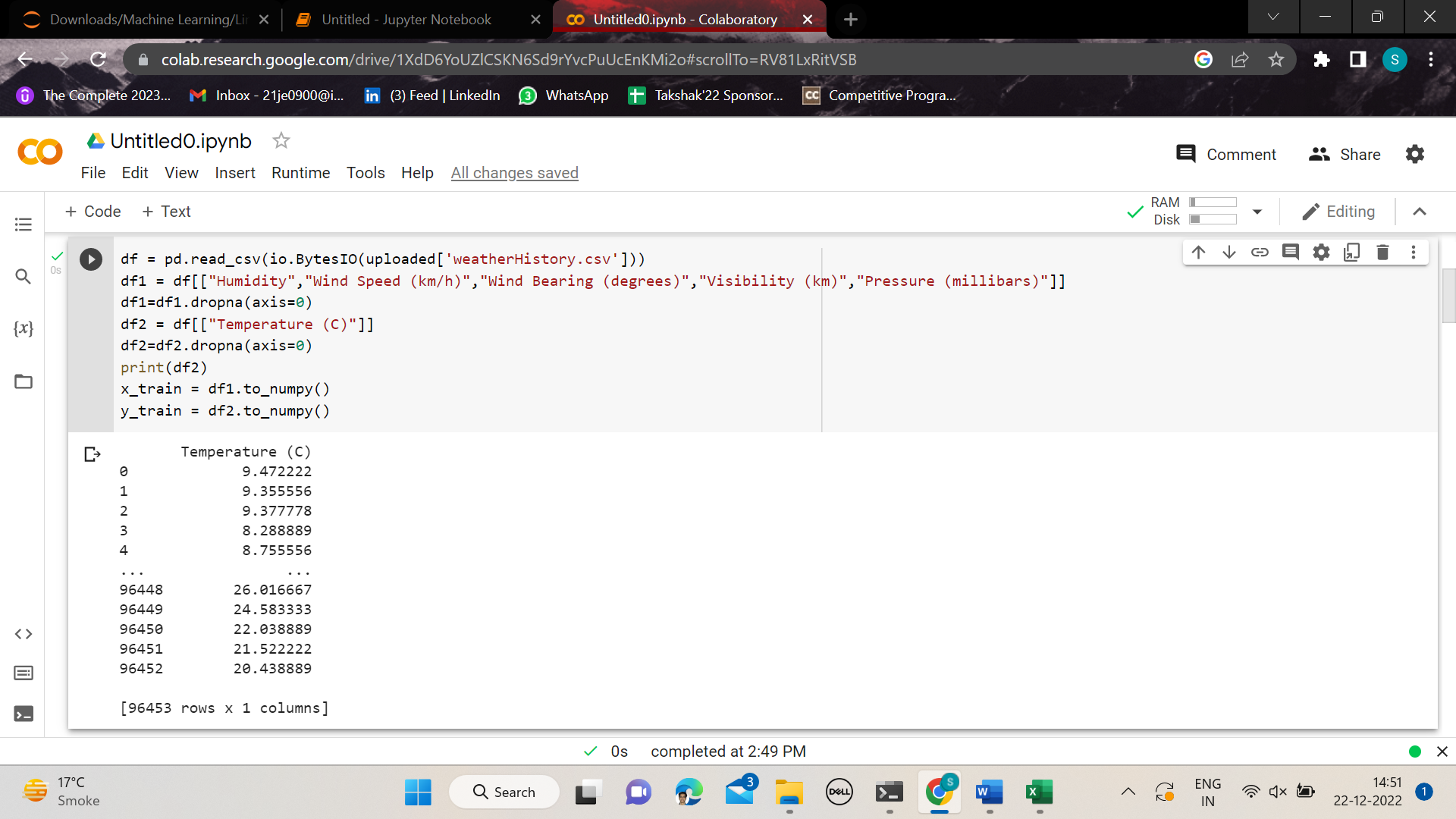
Overview of the dataset:

Input:



96453 rows x 5 columns

Output:



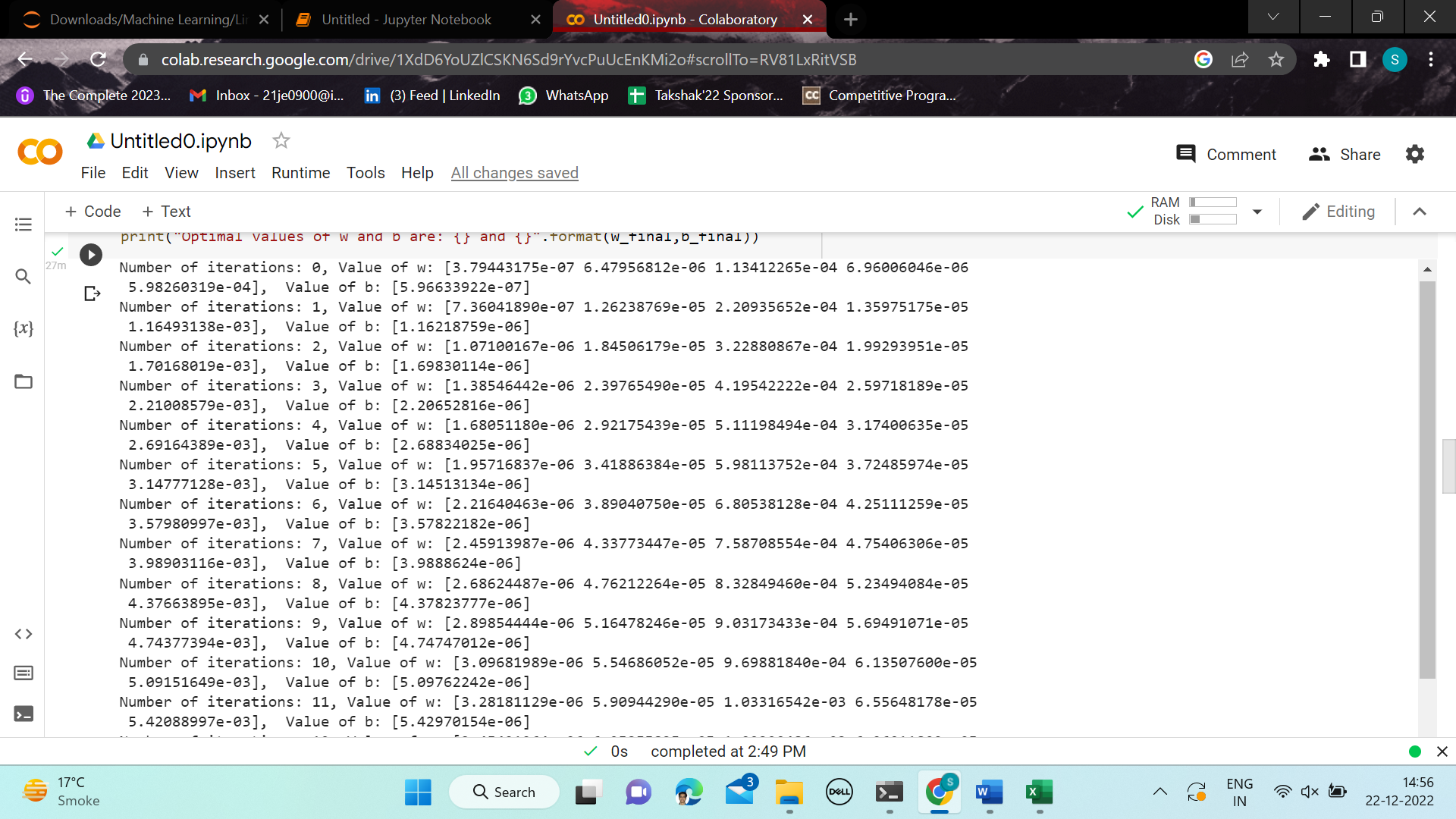
96453 rows x 1 columns

**Procedure:**

* All the data were stored in a CSV (Comma Separated Values) file.
* The CSV file was loaded into the Python programme with the help of Pandas library.
* The input and output values were respectively stored in 2 NumPy arrays.
* The data was cleaned so as to drop any rows consisting of None values.
* Initial values of the parameters to be calculated were set to 0.
* A function was created to compute cost (Cost Function)

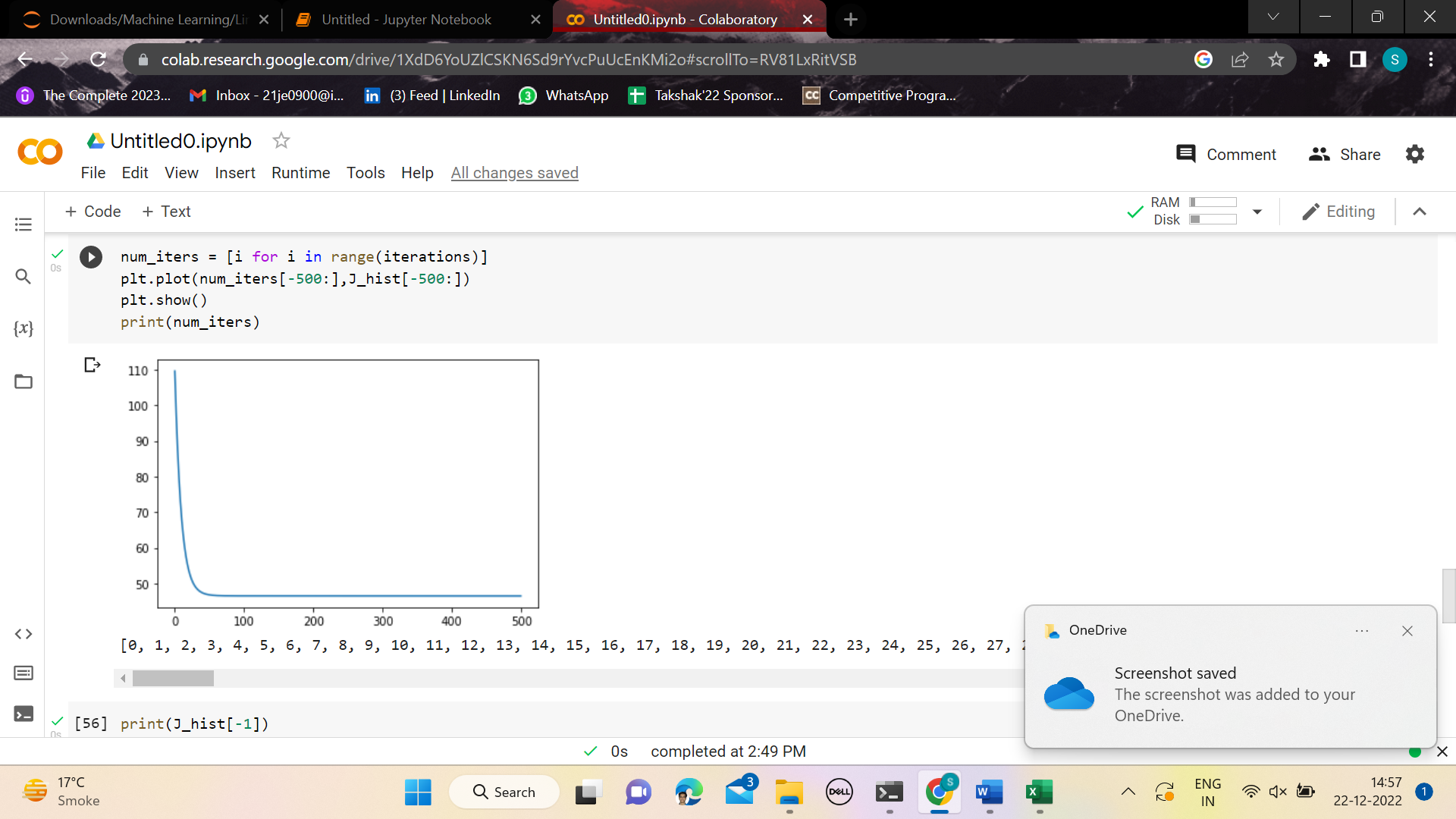
**Cost Function:** It is the measure of the squared mean difference between the predicted value and the exact value. Our task is to minimize this cost function therefore leading to more accuracy.

* A function was created to compute the gradients i.e., the measure of change of cost function with change in values of the parameters to be calculated.
* Then gradient descent function was formed to come up with the minimum value of cost function with a particular number of iterations. The implementation of gradient descent requires the implementation of both the above-mentioned functions.
* After this the gradient descent function was implemented.
* The parameters with which we get the minimum value of the cost-function (thus the maximum accuracy) are the desired parameters which would be used to predict values.



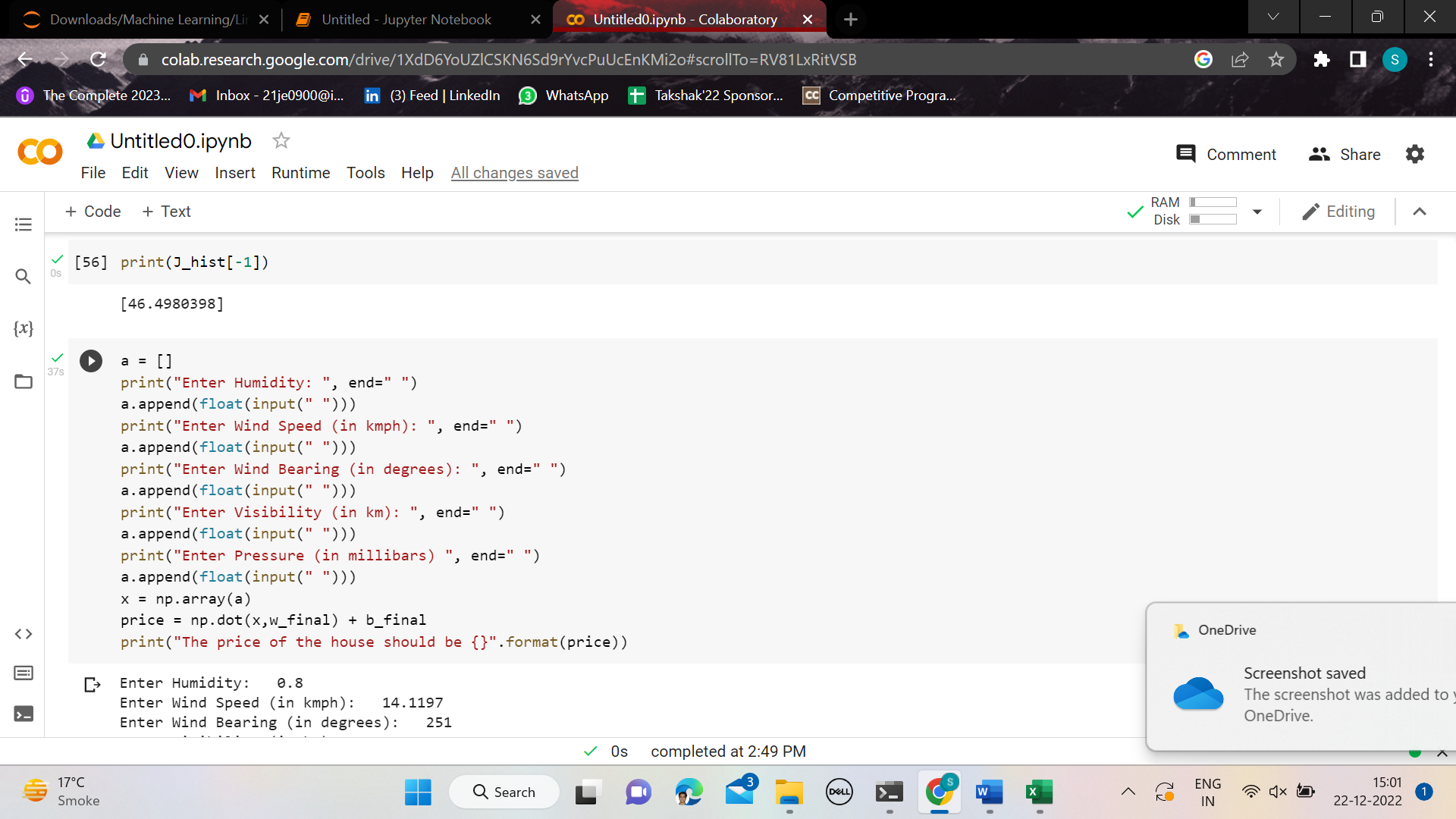
**Observations:**

The plot of cost-function with the number of iterations:



Clearly, we can see that the cost-function is continuously decreasing with increasing the number of iterations.

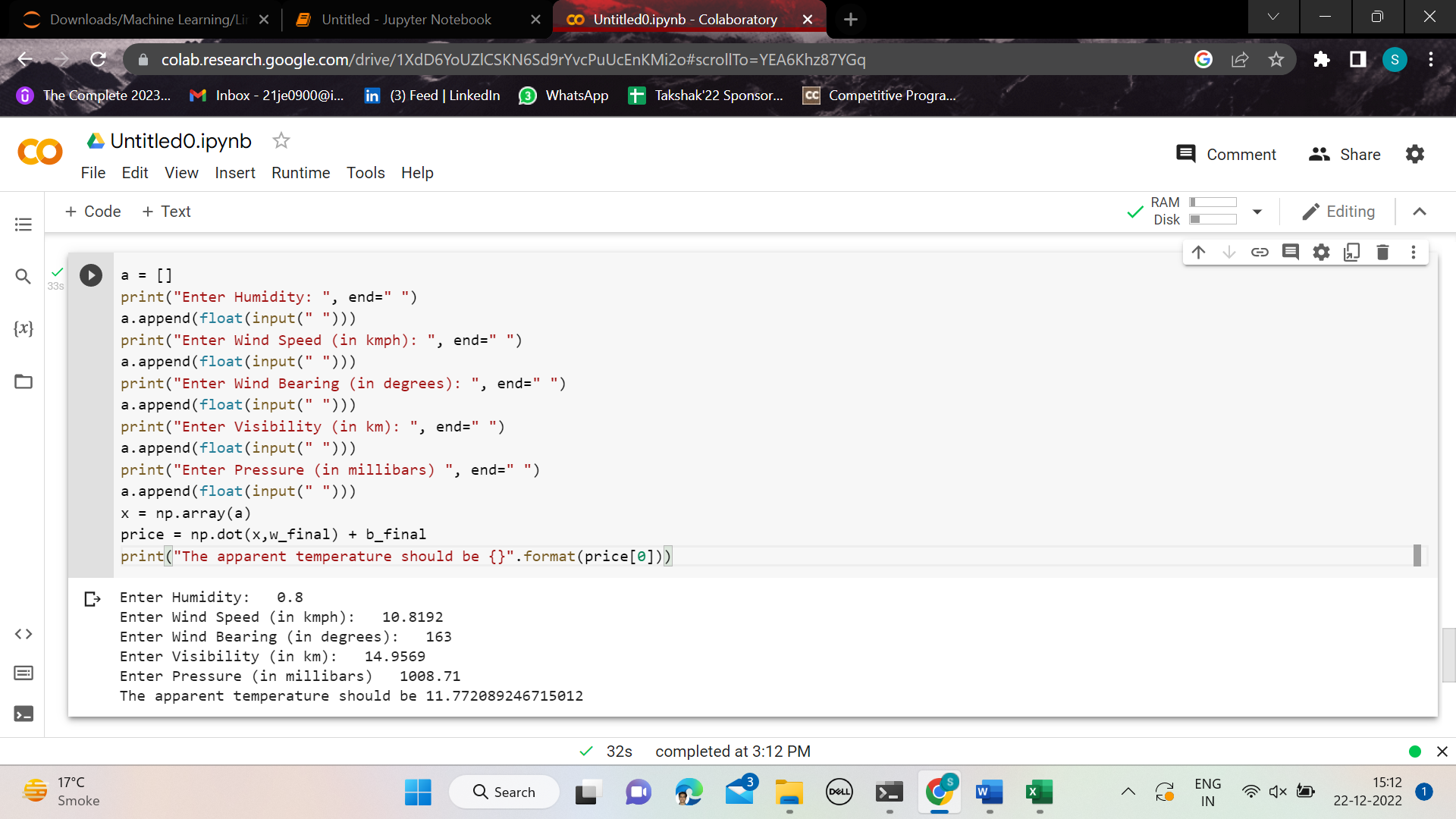
Final value of the cost function:

****

**Output:**

We have to give 5 parameters as input which are respectively the above-mentioned climatic conditions.

The output will be the temperature in degrees Celsius.



The actual temperature as per the dataset for the above values is 11.1833 degrees Celsius.

To improve the accuracy:

We can increase the number of iterations so that the cost function decreases.