

TANGLATEC SCHOOLSTRAAT 8 9883, BELLEM, BELGIUM +32486257049 info@tanglatec.com www.tanglatec.com VAT: BE0644.780.972

DETAILED WEEKLY REPORT

MACHINE LEARNING AND DATA SCIENCE WEEK 1

Submitted by Nyuysemo Calglain

Under the supervision of

TANGLATEC

19/7/2024

**Task**

**with the provided data set create a model that will plot the two D data in python**

Contents

[**1.1** **Introduction** 3](#_Toc172310626)

[1.2 Daily activities summarized 3](#_Toc172310627)

[**1.3** **Daily activities** 3](#_Toc172310628)

[**2** **Day 1: Task identification and setting up the development environment.** 3](#_Toc172310629)

[**2.1** **Difficulties day 1** 4](#_Toc172310630)

[3 **Day 2: Data format conversion and upl**oad to jupyter notebook 4](#_Toc172310631)

[**3.1** **Difficulties day 2** 4](#_Toc172310632)

[4 **Day 3:** 4](#_Toc172310633)

[5 **Data upload to jupyter notebook, importation of libraries and display of data**. 4](#_Toc172310634)

[**5.1** **Achievements** 4](#_Toc172310635)

[**6** **Procedure:** 5](#_Toc172310636)

[**Before using pandas, I used NumPy to load data from the datasheet document then derived the x and y components from the uploaded data. this implementation can be seen on the screenshot below.** 5](#_Toc172310637)

[**6.1**  **screenshot** 5](#_Toc172310638)

[**7** **Day 4:** **Data upload to jupyter notebook, importation of libraries and display of data.** 5](#_Toc172310639)

[**7.1** **Achievements** 5](#_Toc172310640)

[**8** **Screenshots:** 6](#_Toc172310641)

[**8.2** **Conclusion** 7](#_Toc172310642)

## **Introduction**

This report contains a detailed weekly process of how I worked, the difficulties encountered, success and failures and lessons learned.

The task that was assigned to me for internship week 1 was to be able to identify the data set provided to me and come out with a visual (graphical) representation of the data.

## Daily activities summarized

* **Day 1:** Task identification and environmental set-up
* **Day 2:** Data format conversion and upload to jupyter notebook
* **Day 3:** Data upload to jupyter notebook, importation of libraries and display of data.
* **Day 4**: Data upload to jupyter notebook, importation of libraries and display of data.
* **Day 5**: Report writing, code arrangement and upload to GitHub.

## **Daily activities**

# **Day 1: Task identification and setting up the development environment.**

The provided data has both the input and the expected output data hence it’s a supervised learning problem and since the output is voltage it’s continues data hence a type of **supervised learning called regression.**

To **set-up the environment** to perform the given task, I downloaded and **installed anaconda** then lunched it and created a new environment using the command

conda create --name week1task.

I then activate my environment by typing

conda activate week1task.

I then installed Jupiter notebook, pandas, NumPy, matplotlib, Scikit-learn in my environment using the command format

conda install nameOfApp.

I then imported the installed extensions in the jupyter notebook.

## **Difficulties day 1**

was unable to import the dataset due to its nature and hence will look for a way to separate the X input from the expected output Y.

# **Day 2: Data format conversion and upl**oad to jupyter notebook

To upload the given data set to jupyter notebook, simply open the word document, copy the data, paste in Notepad and while saving it I choose all formats then write the name of the file with a “.tsv” extension. I could also use “.csv” extension if I want.

## **Difficulties day 2**

Due to poor internet connection here, I was unable to lunch my work space and upload my data to jupyter notebook.

# **Day 3:**

# **Data upload to jupyter notebook, importation of libraries and display of data**.

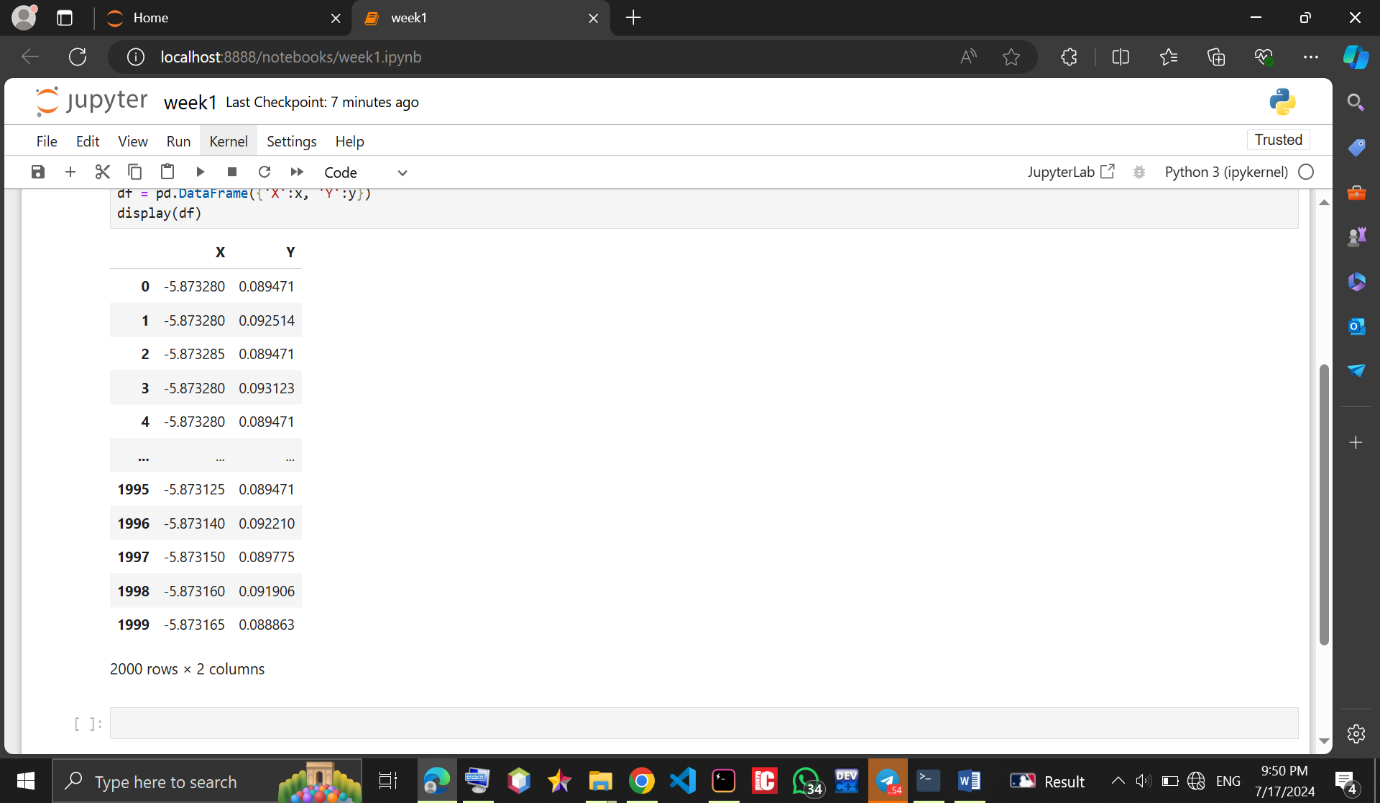
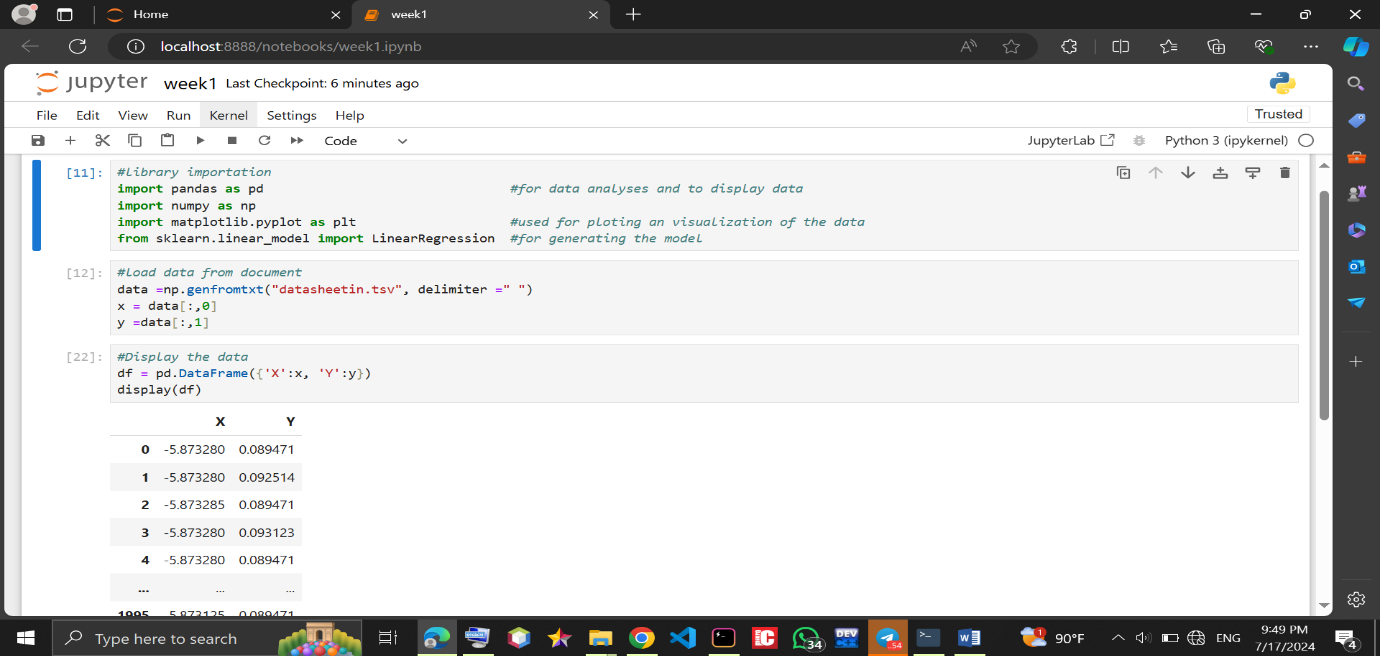
## **Achievements**

After studies, I was able to upload my data to jupyter notebook, import the necessary libraries and used pandas to display data in a tabular format.

# **Procedure:**

# **Before using pandas, I used NumPy to load data from the datasheet document then derived the x and y components from the uploaded data. this implementation can be seen on the screenshot below.**

## **screenshot**



# **Day 4:** **Data upload to jupyter notebook, importation of libraries and display of data.**

## **Achievements**

I was able to learn and implement data description and basic data analysis and also did a scattered and a histogram plot of the problem.

#### **Procedure**

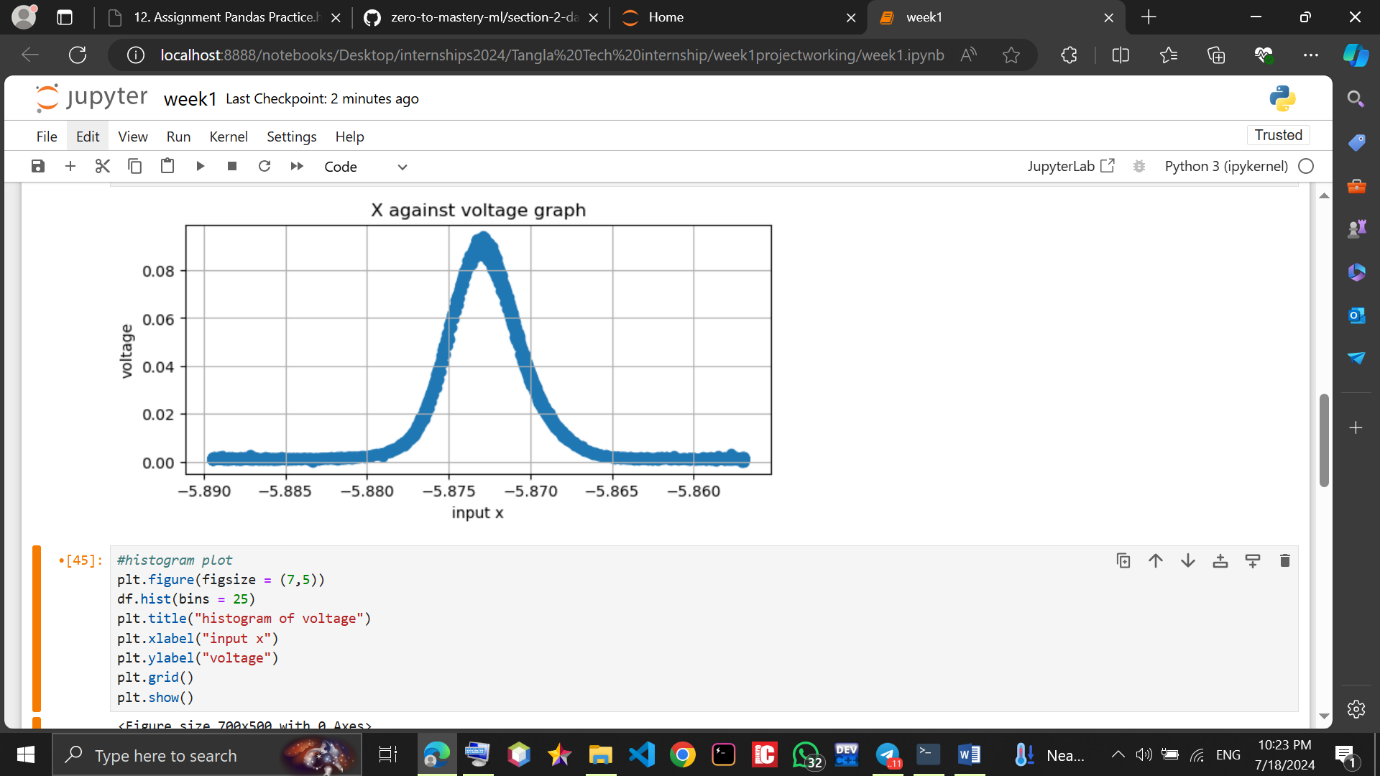
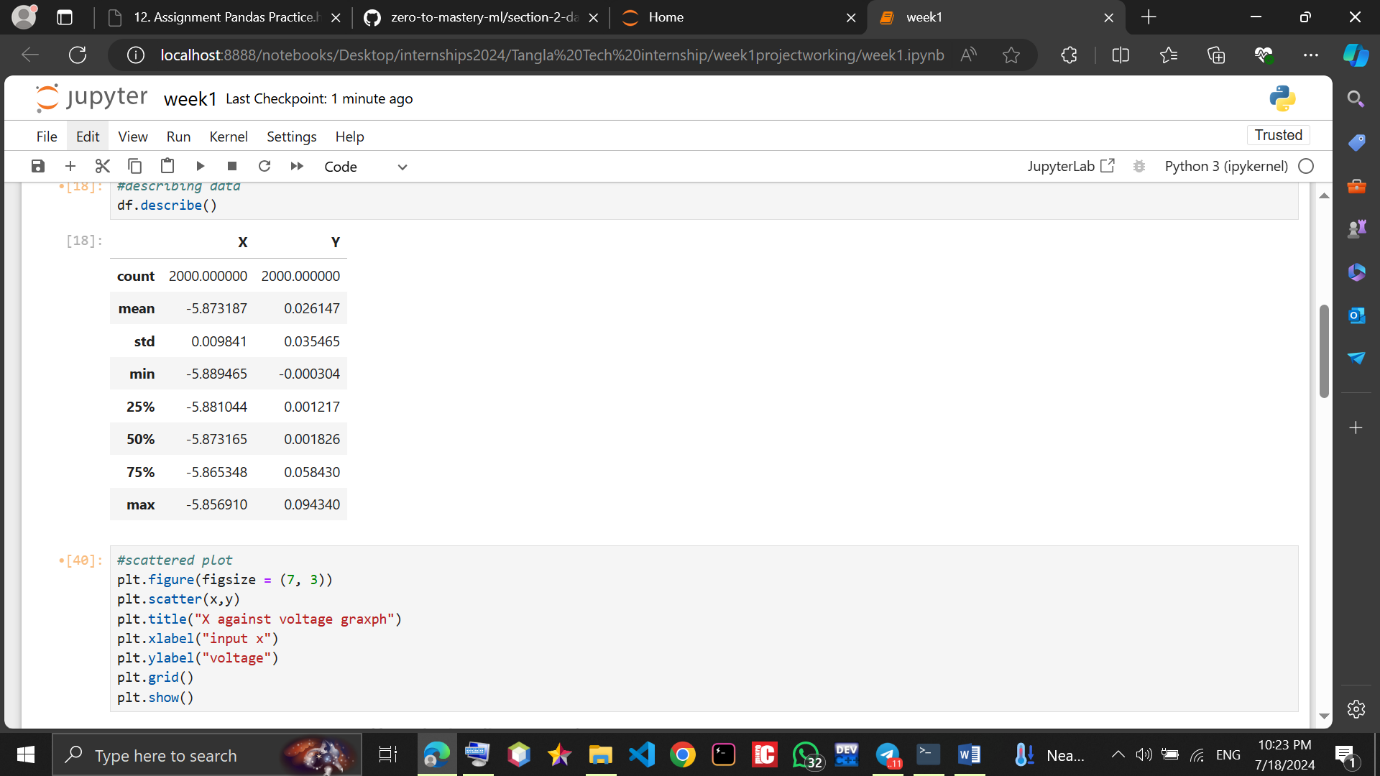
From where we ended the previous day I went forward to use the **pandas describe** () function which **displays the mean, standard deviation, min and max values of the datasheet**.

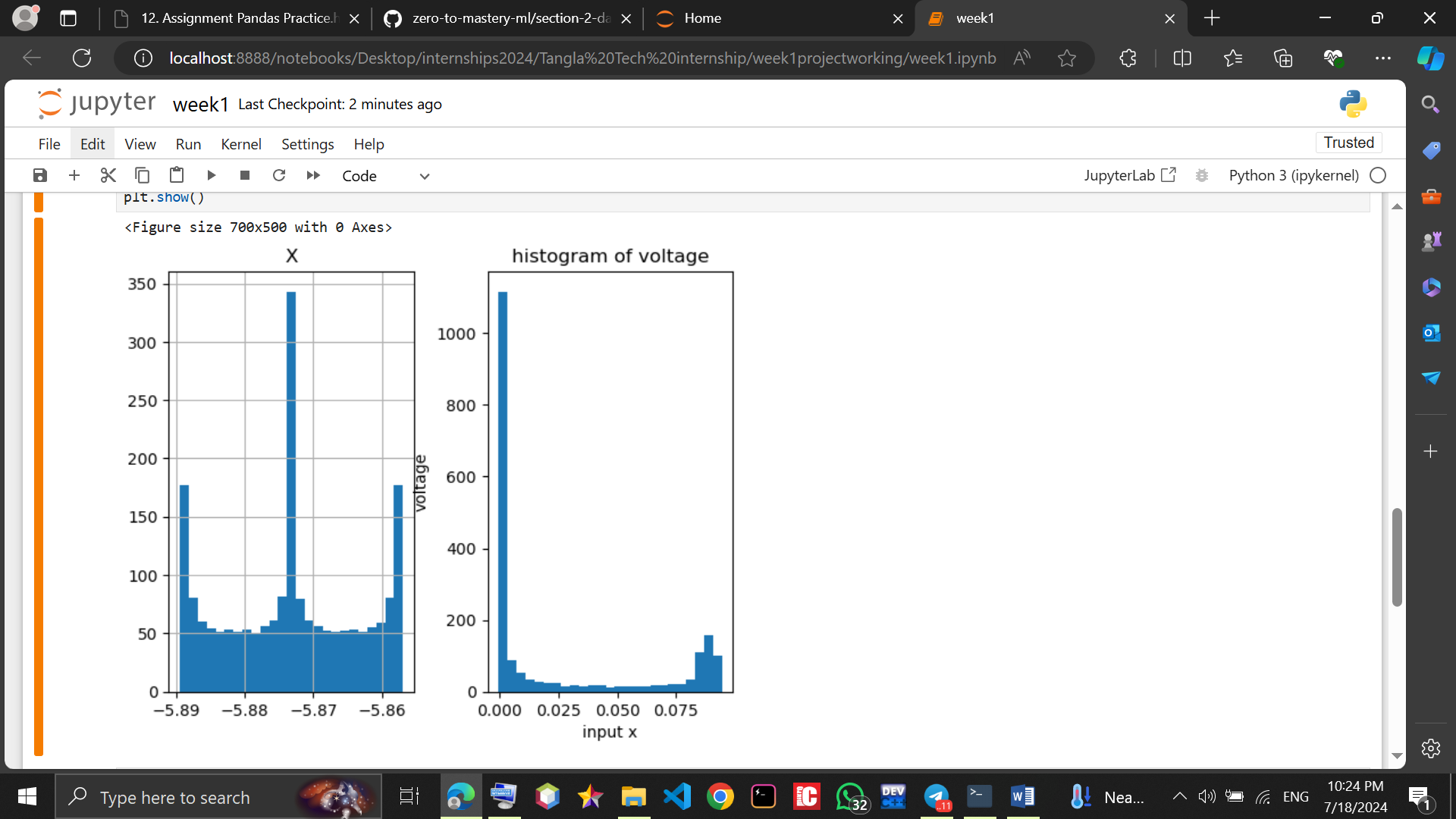
Syntax: data.describe()

I also used sklearn.pyplot(plt) to visualize the data graphically as seen in the screenshots below

#### 

# **Screenshots:**





## **Conclusion**

Above is a brief explanation of my achievements and difficulties. As a newbie to this field I’ll say I’ve learnt many new things and becoming progressively familiar with the jupyter notebook environment. I believe I can do better.

Unfortunately for me I could not build the model before today but I’ll put in more learning to become better.