DSE 315: Data Science in Practice

Assignment 1

Web Scrapping Report

In this assignment, I have scrapped the weather data from the Easter Park Weather website. As we know how vital weather data is for us. Our whole economy is dependent on the weather to a considerable extent. I have scrapped the data with the help of the beautiful soup library of python. First, I chose the month I wanted the data and then used the bs4 library to make the soup object and its HTML parser to parse the data.

After getting the HTML content of the website through the soup object, I searched all the table classes in that and found the table data tag. After that, I iterate over all the <td> tags to parse their content and store it into a dictionary object.

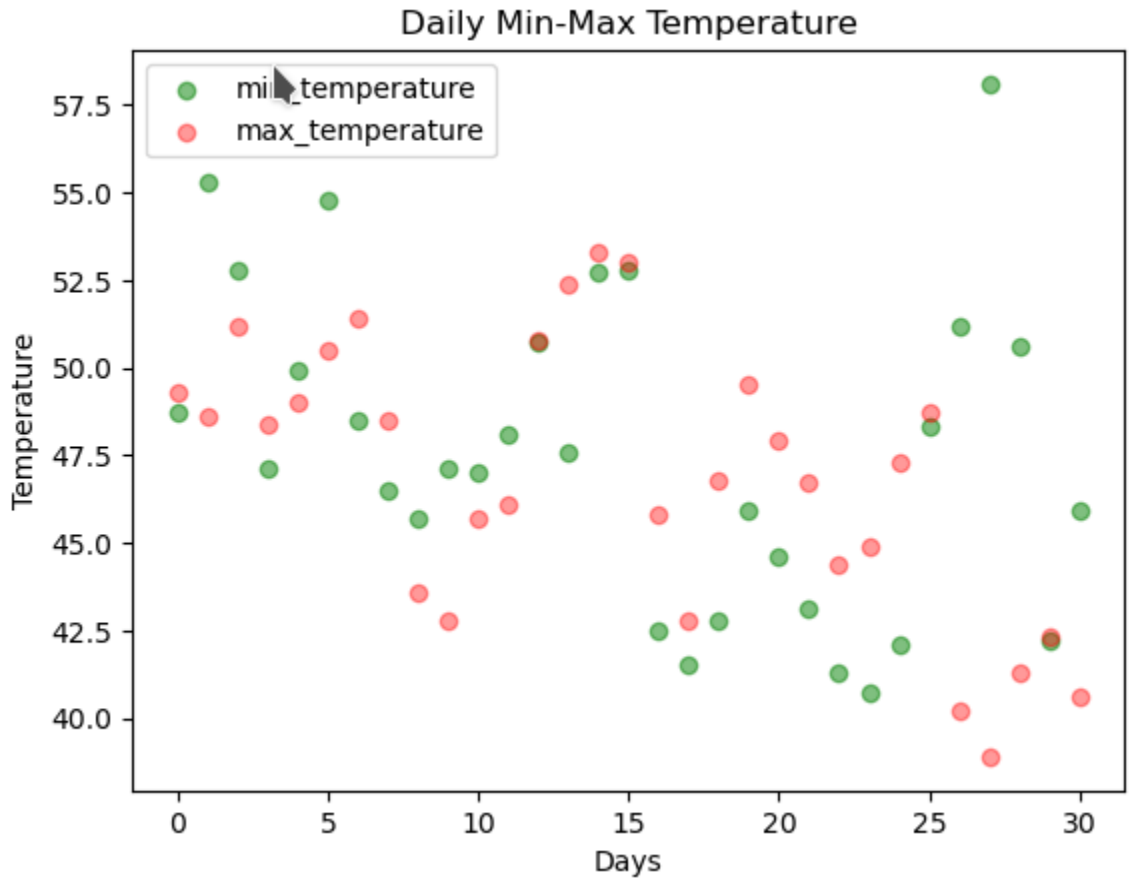
Then that dictionary is converted into a pandas data frame, and that data of the data frame is stored in a CSV file.

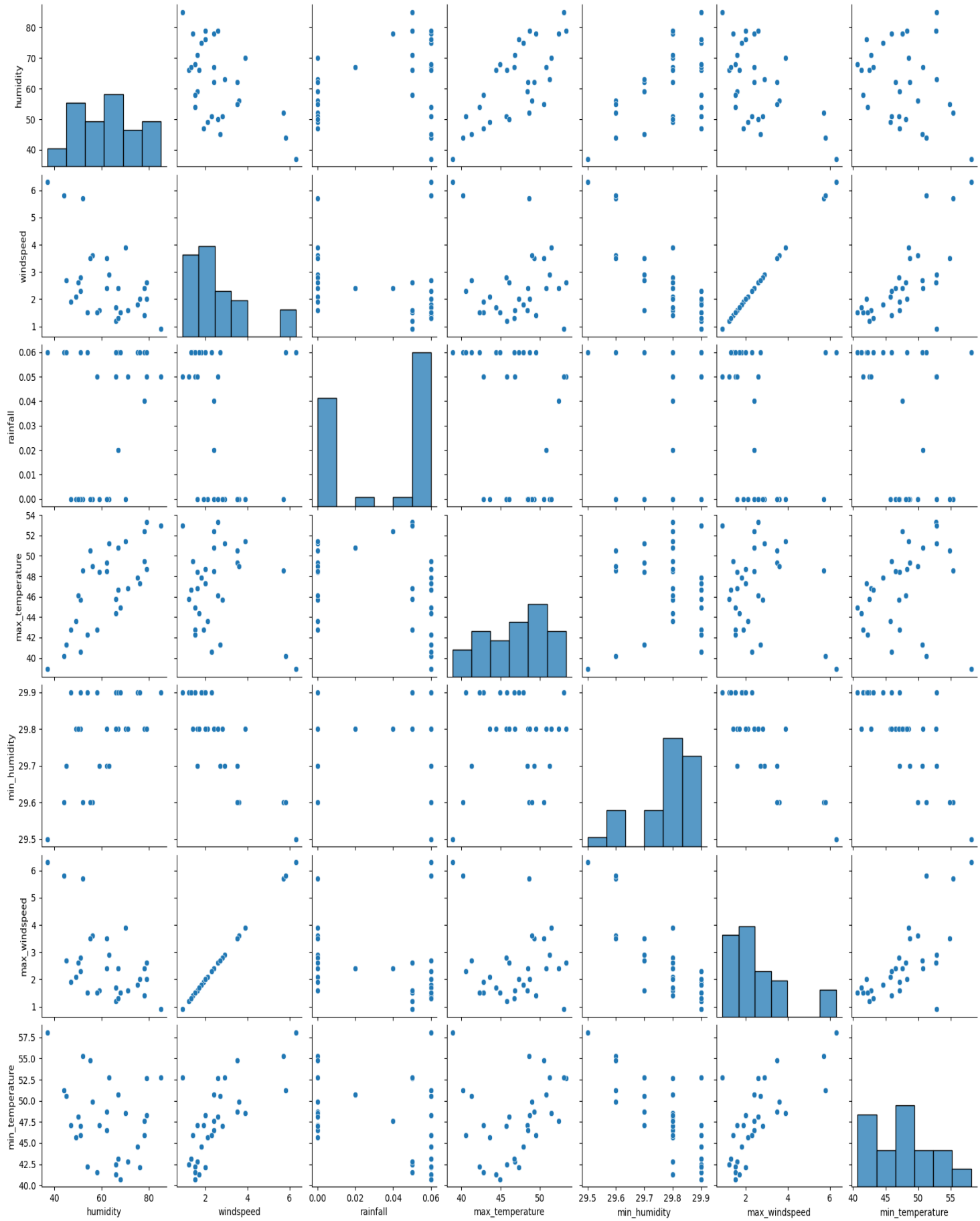
After cleaning the data and getting the data in a CSV file, the Exploratory Data Analysis (EDA) begins.

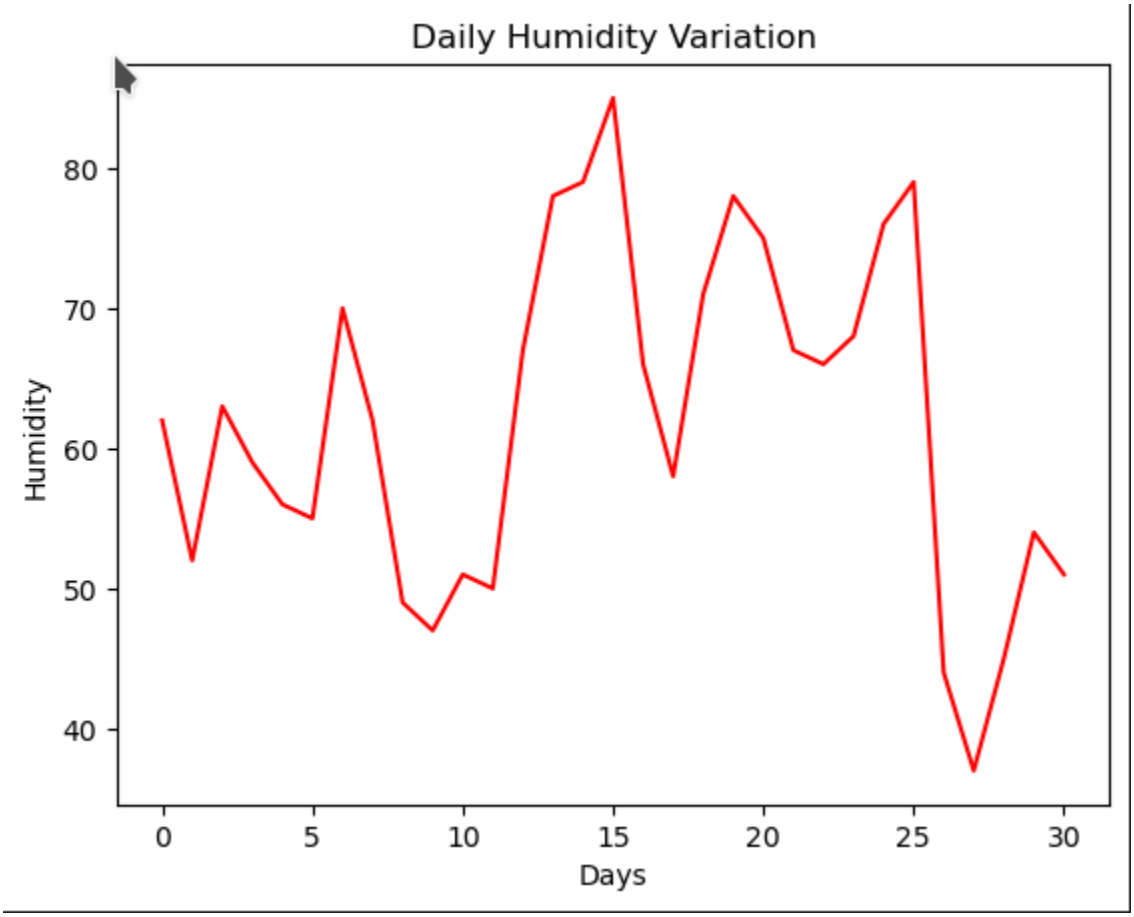
In EDA, I have found that if the data contain any null values or has any missing values. Then I found each data column's mean, median and maximum values to get insight into the data.

I also plotted some data against each other to know if there was some trend in the data. There are some exciting trends, but most are random due to daily weather conditions.

Some plots which I found are shown below:




humidity
windspeed
rainfall
max_temperature
min_humidity
max_windspeed
min_temperature



Thank you for reading.

By

Divyansh Tripathi
(Department of Physics, 20108)