CEIS110 Programming with Data

Introduction

Data is growing

This project uses a cloud system to gather temperature and humidity data

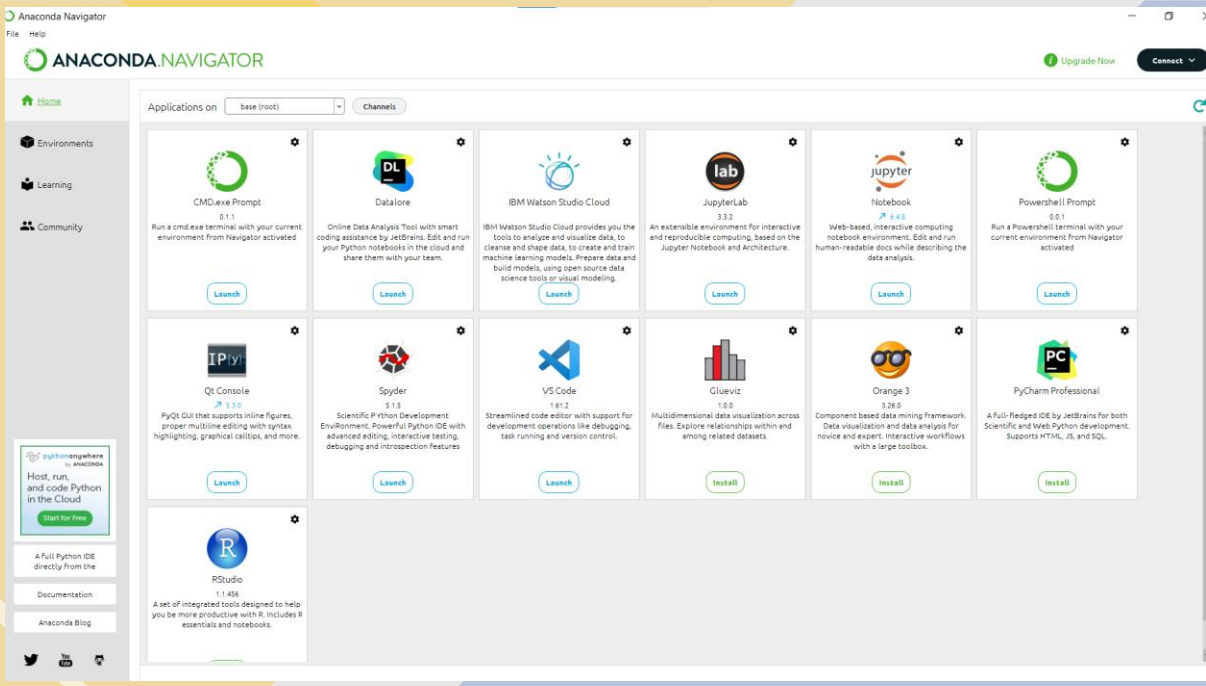
Them the data is analyzed using programming data

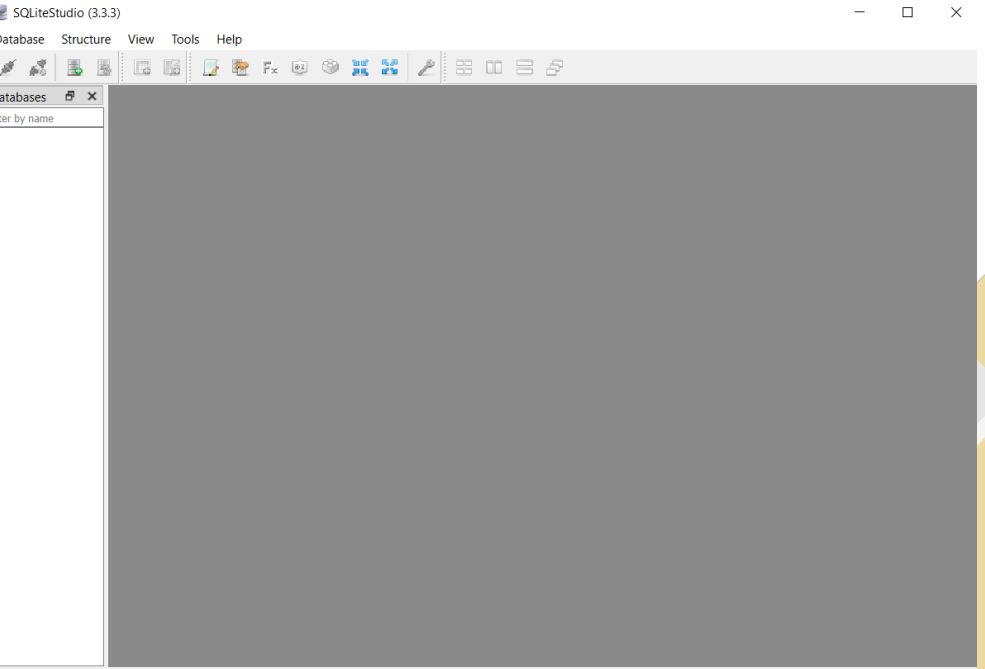
Software Inventory

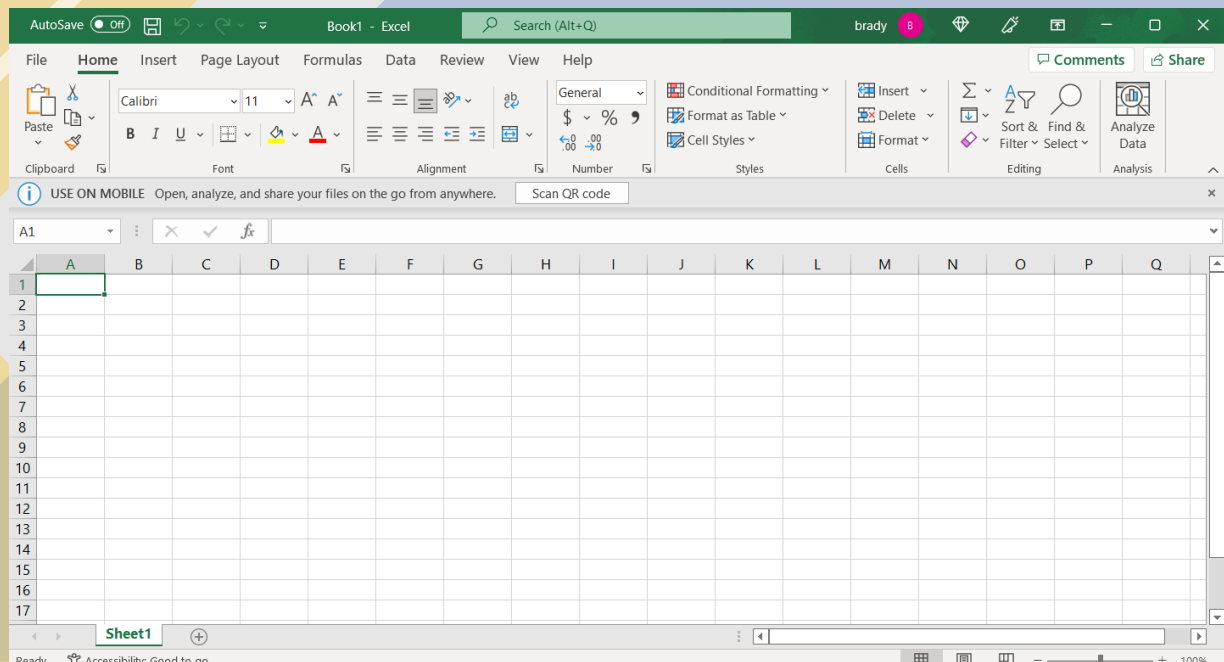
- Before developing programming with data project, all software must bed downloaded and installed
- Software needed included SQLite studio, Excel, and Python IDE - anaconda

Software:

The Software needed for this project includes Microsoft Excel SQLite anaconda Navigation with Spyder

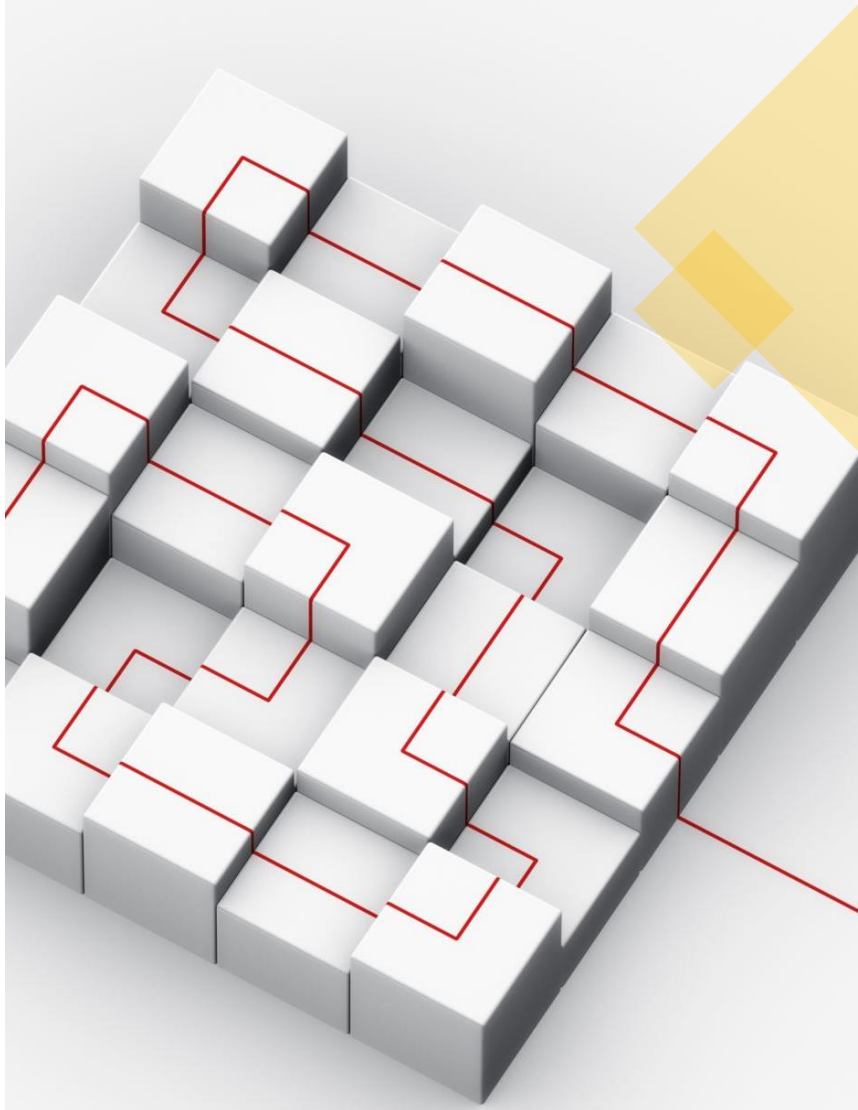






Planning and Design

- After performing of the software needed for the project, a plan was created for the data and temperature project.
- To plan the design of the project, a flowchart was generated.
- Planning and design of the project are crucial steps to understanding the development process.



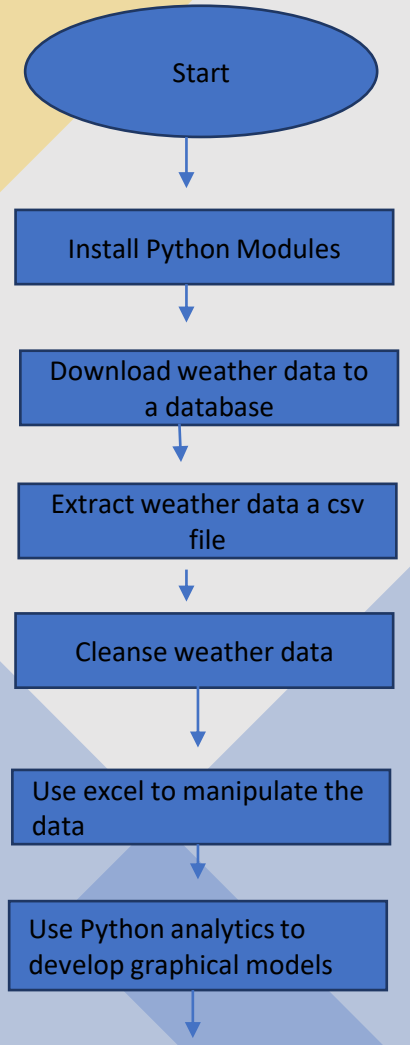
What are Flowcharts?

- Flowcharts are graphical representations of processes of workflows, and the flowchart developed illustrates the process and the output of this software development project.
- Many companies use flowcharts as simple narratives

Flowchart

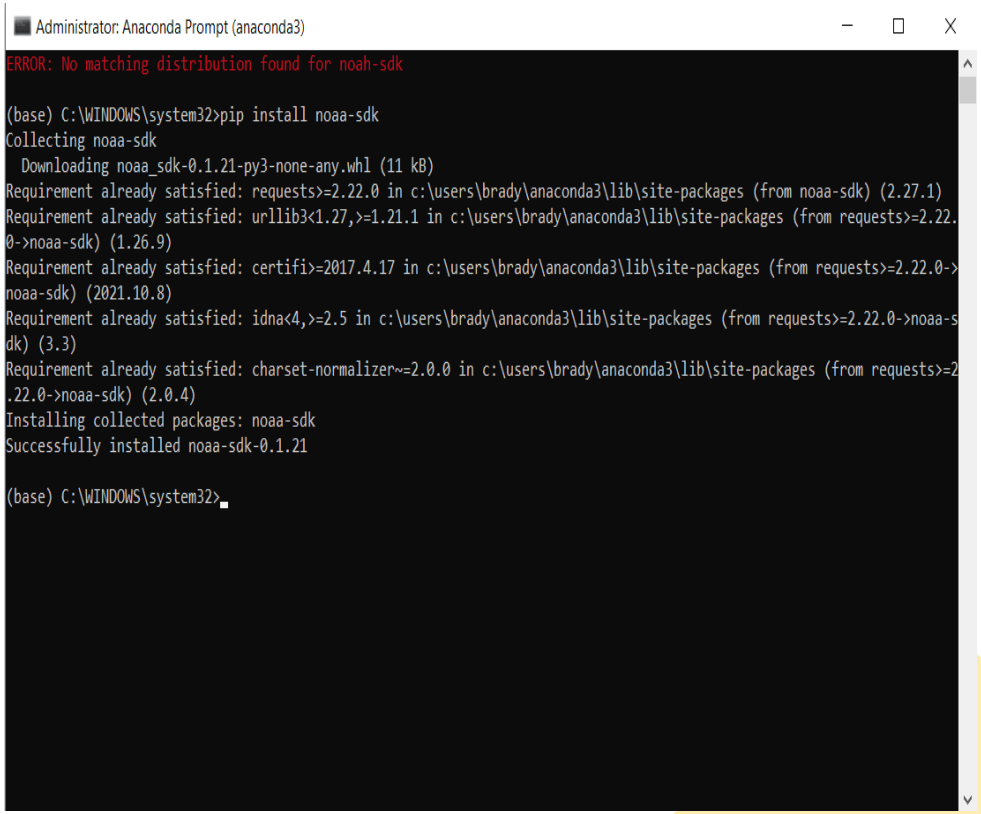
Include the following processes:

- Install python
- Download weather data to a database.
- Extract weather data from database into a comma separated file with python
- Cleanse weather data
- Use Excel to manipulate data
- Use python data analytics modules to develop graphical models



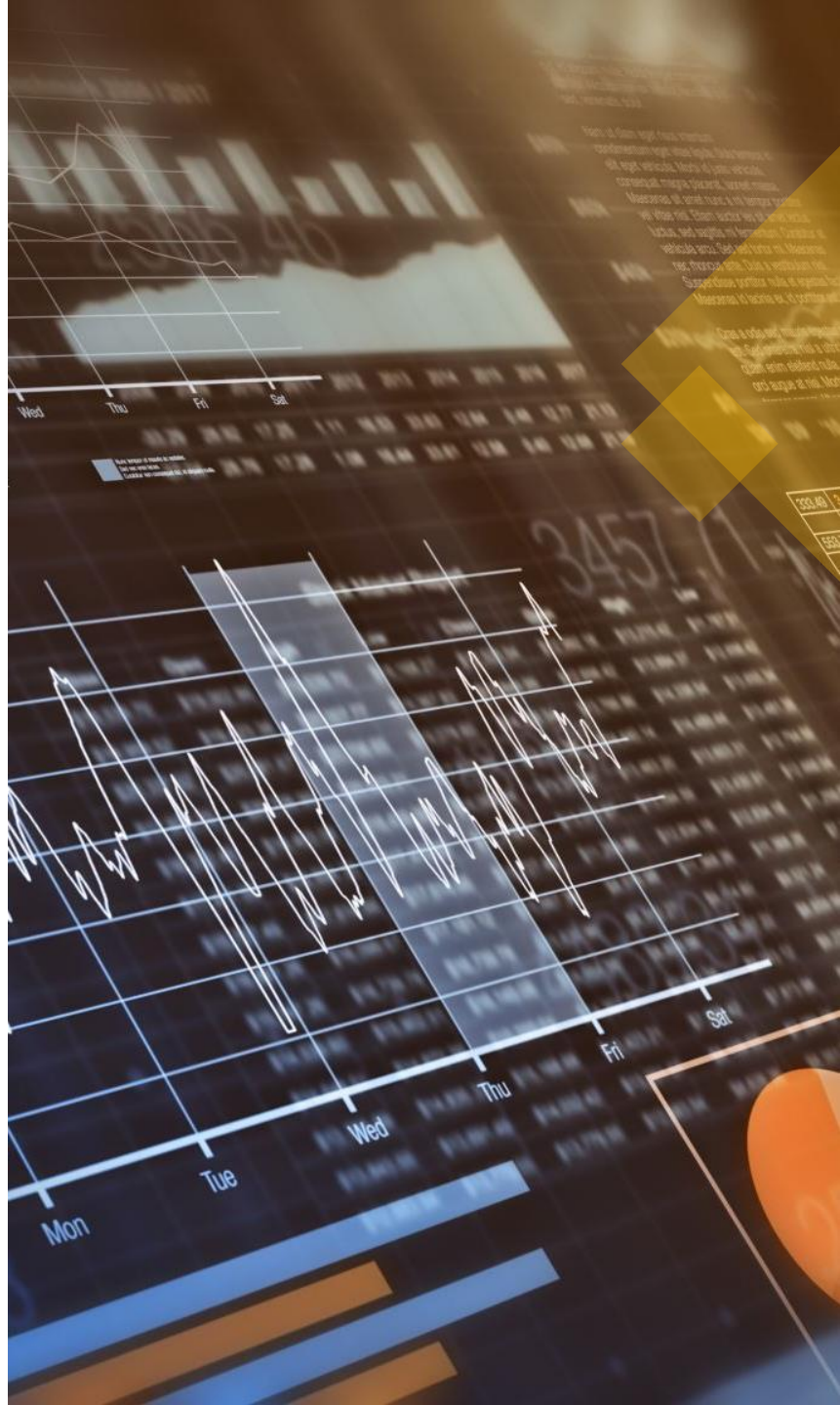
Adding Library

- In order for python to connect to the US government Nation oceanic and Atmospheric Administration (NOAA) weather data service using a cloud-based Application Programming Interface (API), a library module must be installed to use this service.
- The screenshot shows the NOAA-SDK library installed.



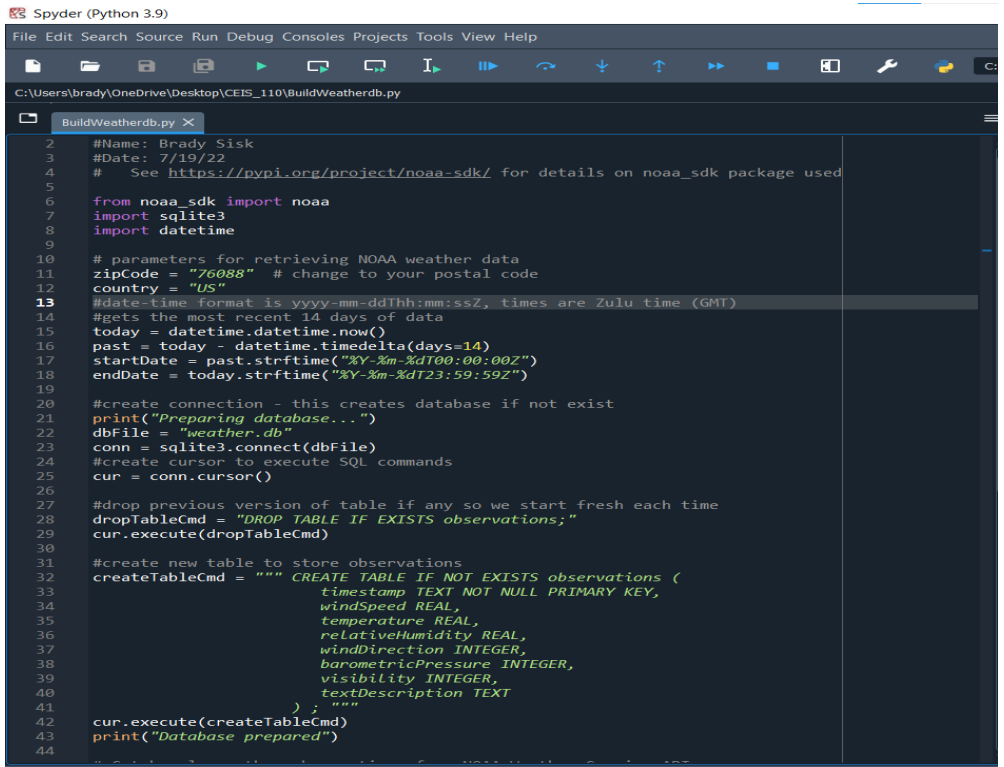
Gathering Temperature and Humidity data

- After planning and design, the code was developed to download a set of weather observations
- This data was stored on local database in table for later analysis

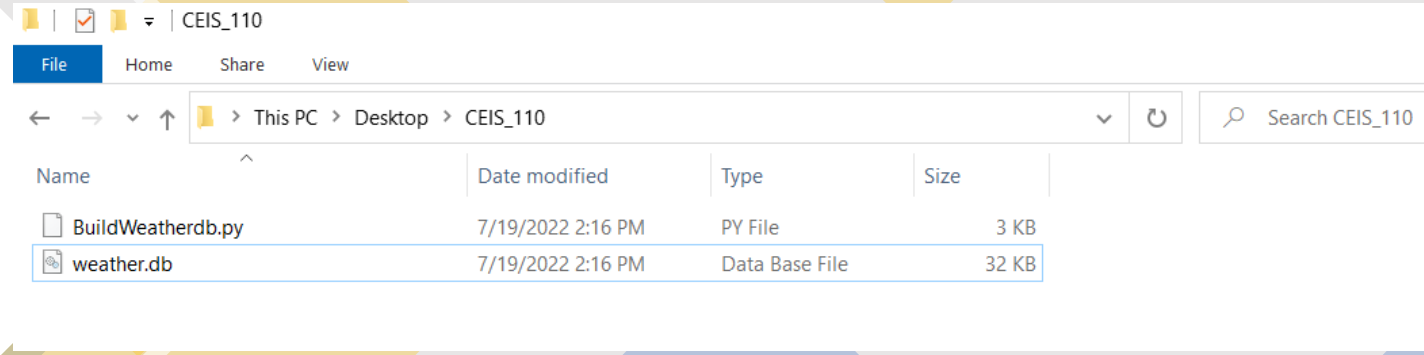


BuildWeatherDb.py

- Screenshot of code in Spyder
- The code will create a table named Observations, the fields timestamp, windspeed, temperature, relativeHumidity, windDirection, barometricPressure, visibility, and textDescription.
- The database will be named weather.db and stored in the same directory as the python code



Weather db File



Screenshot of Windows explorer showing database wether db was created.

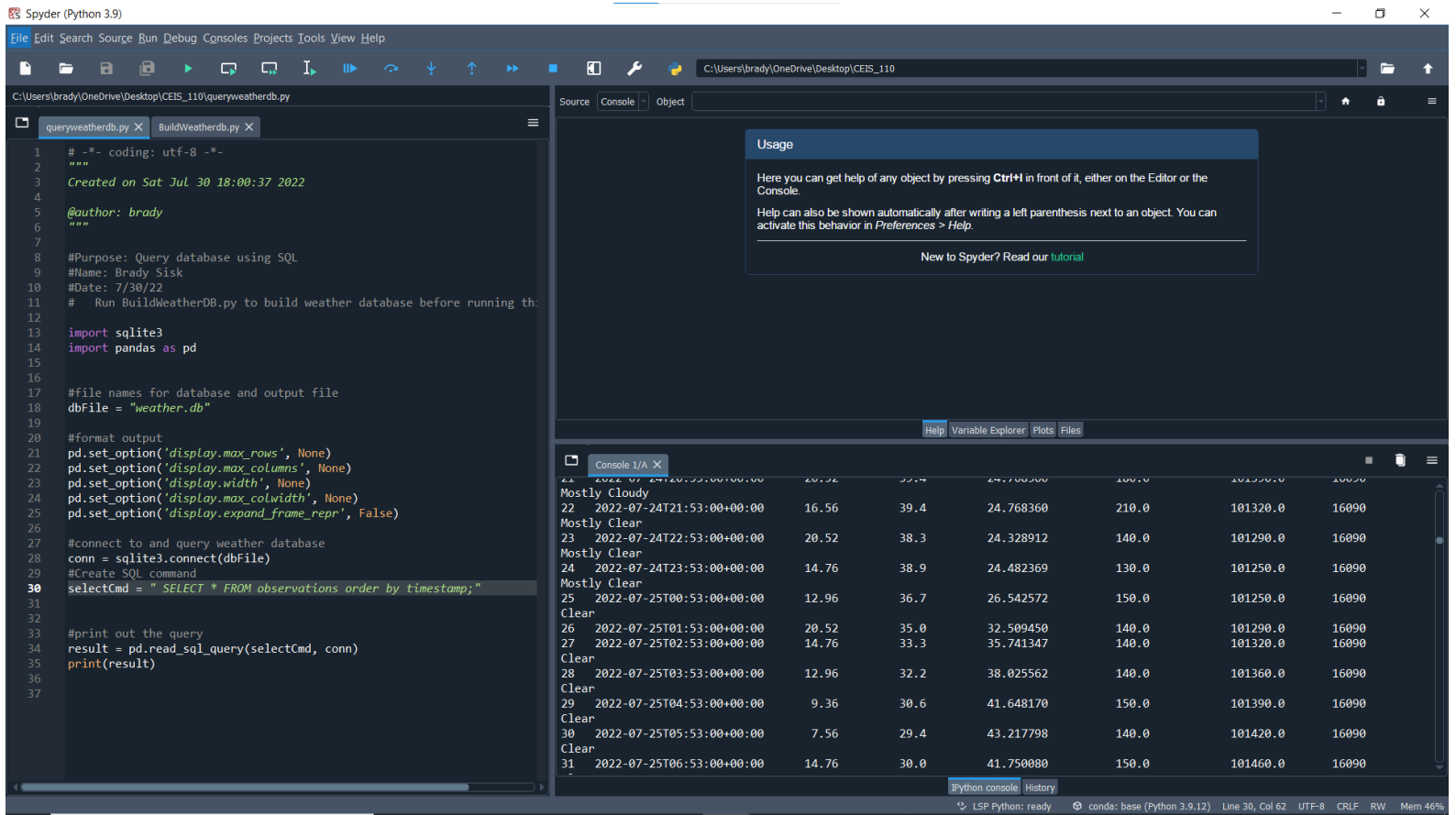
Querying the Database

Querying the Database

- Structured Query Language is a programming language used for working with relational database
- SQLiteStudio was sued to query the database and view the results

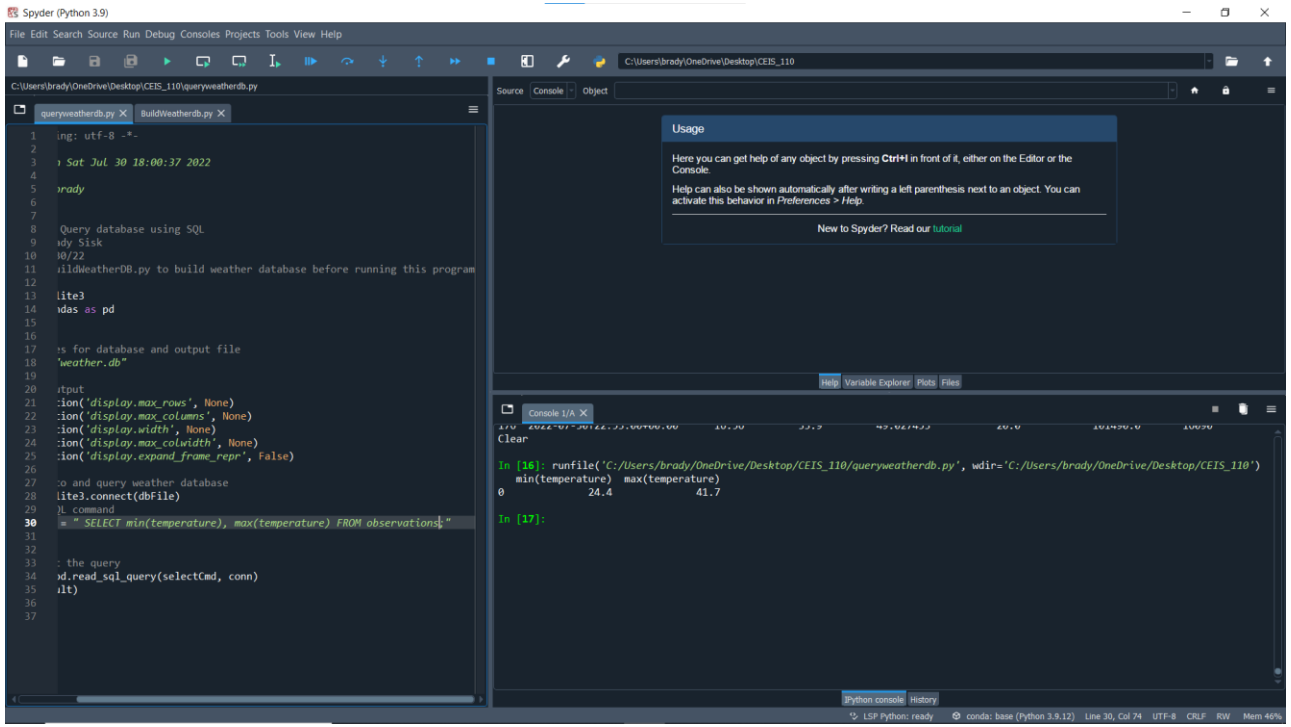
Query to retrieve all columns and all rows

- The SQL command “select * from observations” was executed to retrieve all rows and columns from the observations table.



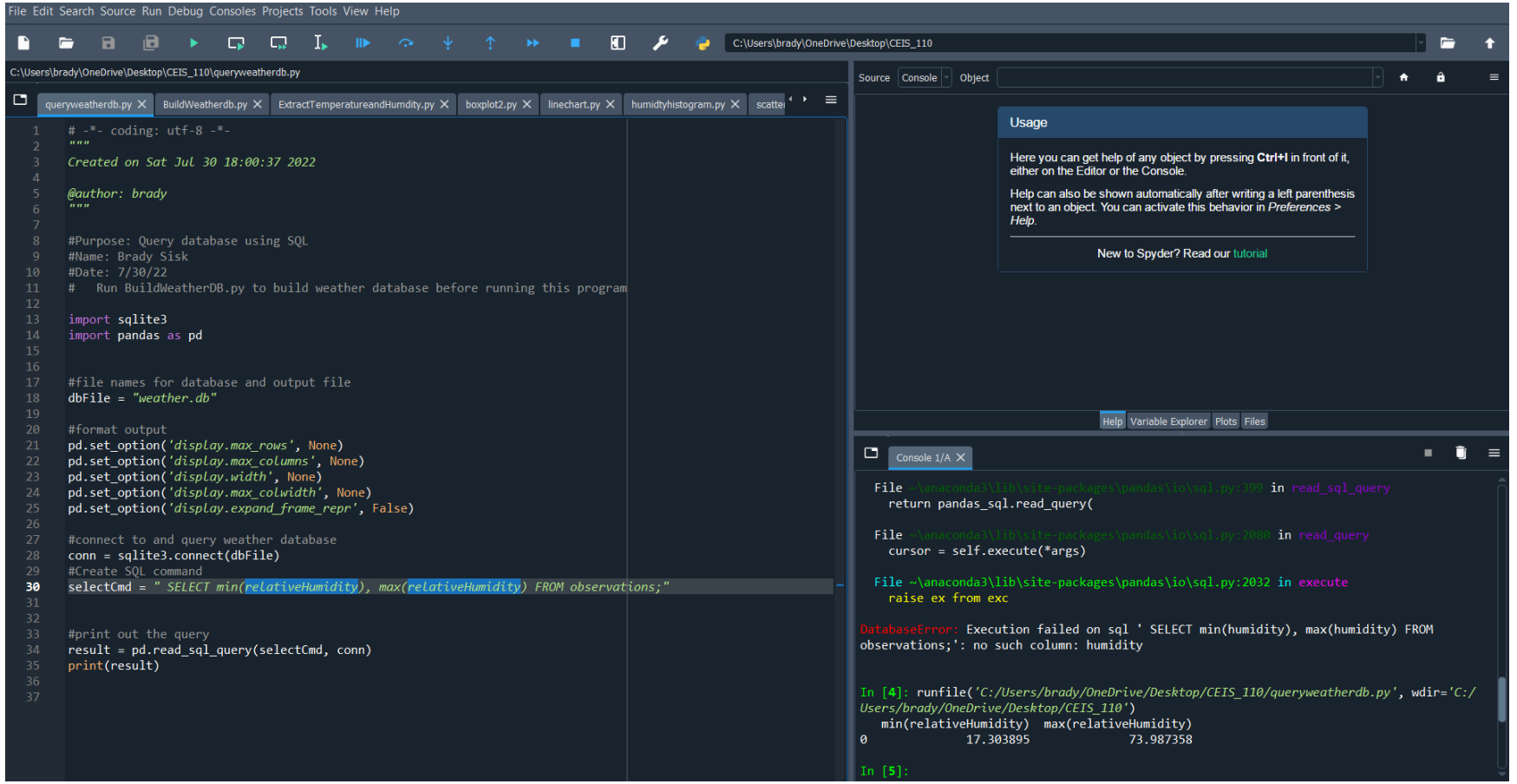
Query to retrieve lowest and highest temperature

- The min and max temperatures were retrieved. These temperatures are captured based on the Celsius scale.



Query to retrieve lowest and highest humidity

- Another query performed to retrieve the lowest and highest relative humidity. The relativeHumidity column was used to retrieve this information.

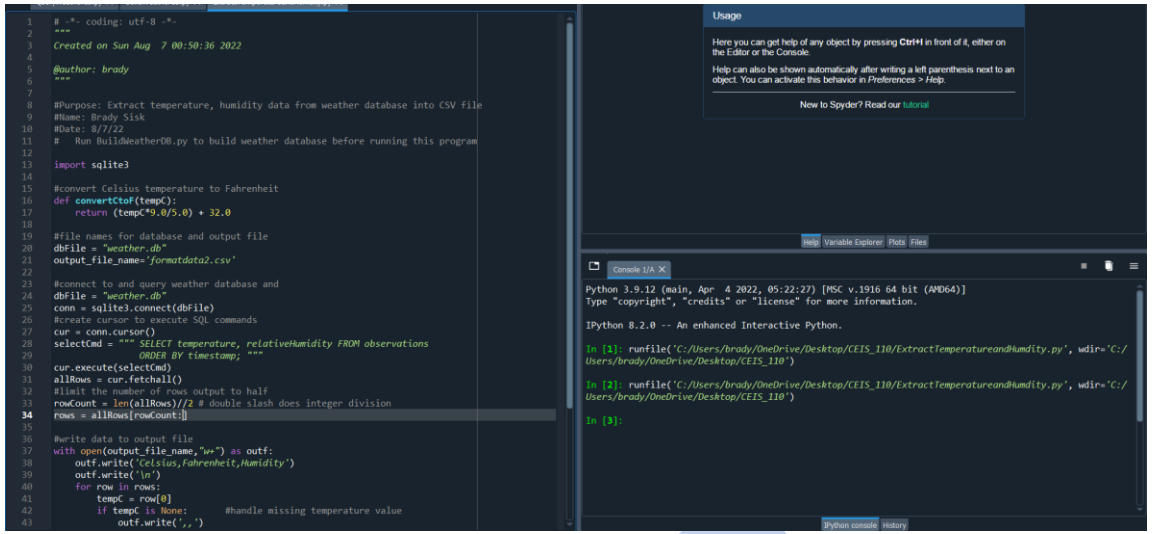


Data Cleansing

- Data output from machines may have errors or extraneous data. When cleansing the data, programs can automatically put it in the format needed to be read by other programs.
- A python program is reading the data output by the python program and saving it in a csv file so that it can be read by Excel.
- Often data must be cleansed of spurious or missing values in a dataset. The data must be complete, valid, and standardized.

Extracting Temperature and Humidity using Python code

- The weather.db database may contain null or missing values. The code used retrieves only the temperature and humidity and writes them to a comma separated values (CSV) file. Two files are created – a formatdata.csv and formatdata.2 that each contain half of the rows. Missing or invalid values are not written to the file.



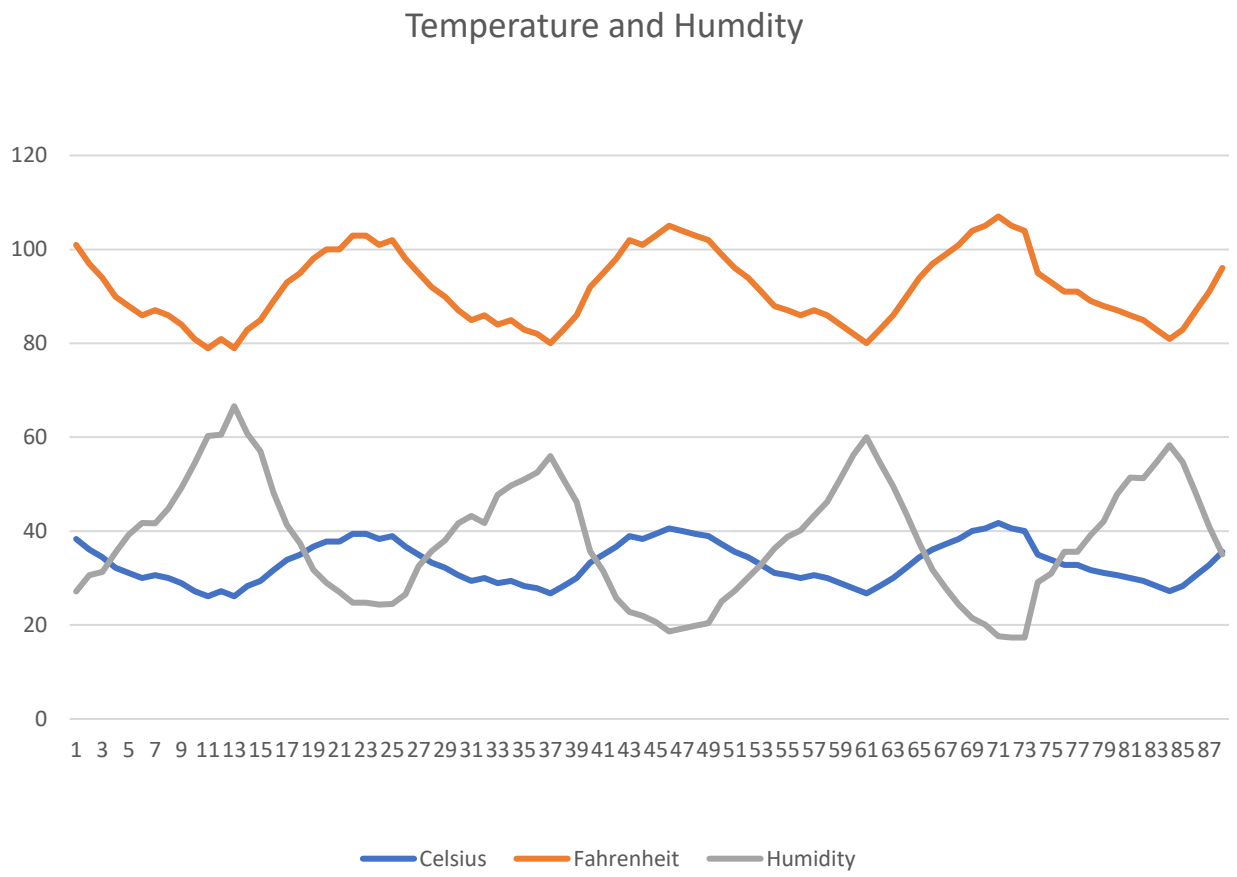
Data Formatted in an Excel Spreadsheet

- The python program created a formatdata.csv file
- This file contains 3 columns: Celsius, Fahrenheit, and Humidity
- Statistics can be performed on this spreadsheet

	A	B	C
1	Celsius	Fahrenheit	Humidity
2	38.3	100.94	27.1349
3	36.1	96.98	30.59447
4	34.4	93.92	31.32396
5	32.2	89.96	35.43829
6	31.1	87.98	39.2029
7	30	86	41.75008
8	30.6	87.08	41.64817
9	30	86	44.78357
10	28.9	84.02	49.25382
11	27.2	80.96	54.38442
12	26.1	78.98	60.2634
13	27.2	80.96	60.5158
14	26.1	78.98	66.60937
15	28.3	82.94	60.76613
16	29.4	84.92	57.01373
17	31.7	89.06	48.15548
18	33.9	93.02	41.24052
19	35	95	37.36307
20	36.7	98.06	31.74628
21	37.8	100.04	28.9657
22	37.8	100.04	27.00366
23	39.4	102.92	24.76836
24	39.4	102.92	24.76836
25	38.3	100.94	24.32891

Data Visualization

A line chart was developed in Excel showing the Temperature and Humidity over Period 1

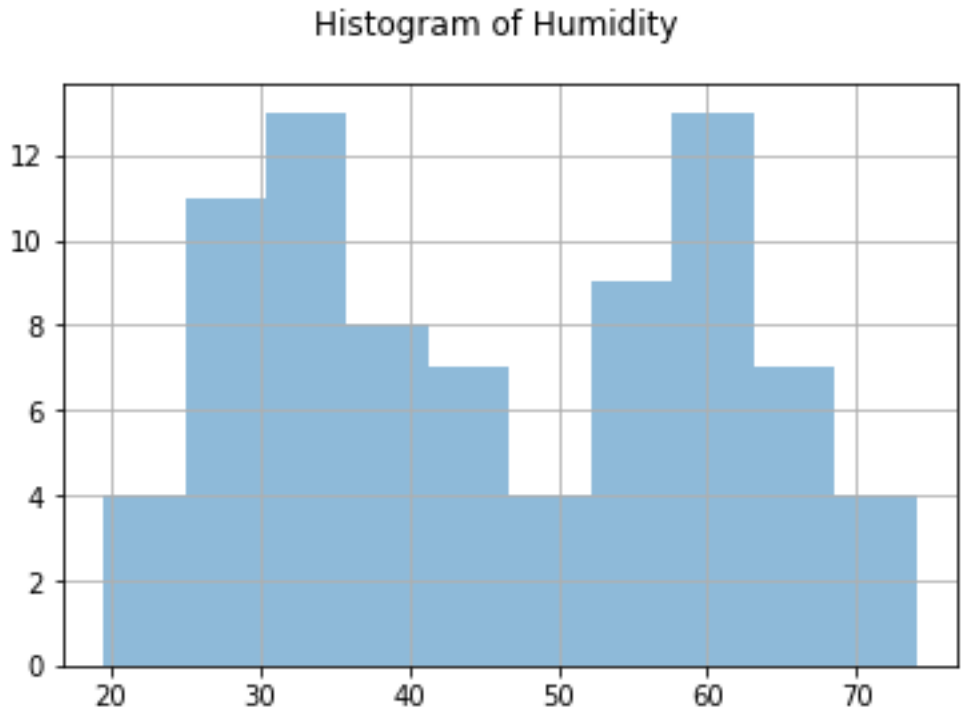


Data Analytics

- Python Data modules allows users to develop charts and graphs depicting data.
- The data sets can be manipulated as well and saved into tabular format
- The data analytics modules are available as part of Anaconda Several plots were generated looking at humidity and temperature
- Then a prediction was made about the data

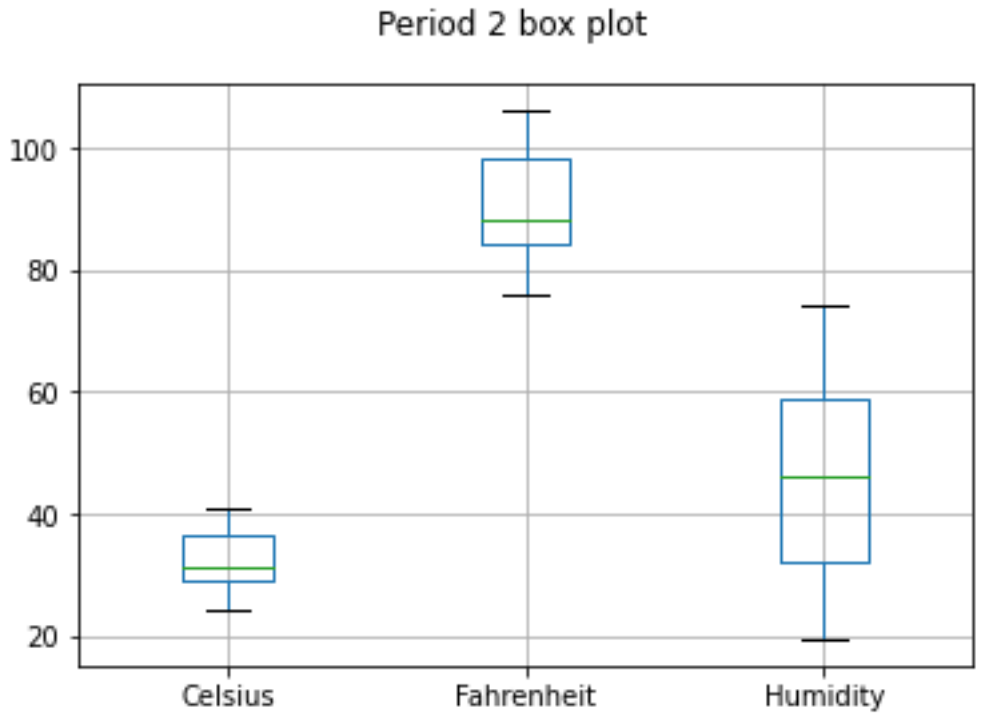
Histogram of Humidity

```
#Purpose: Create a histogram of humidity data from the second period
#Name: Brady Sisk
#Date: 8/13/22
import pandas as pd
import matplotlib.pyplot as plt
df1 = pd.read_csv("formatdata.csv")
df2 = pd.read_csv("formatdata2.csv")
df2['Humidity'].hist(bins=10, alpha=0.5); plt.suptitle('Histogram of Humidity')
```



Period 2 Box Plot

```
#Purpose: Create box plot for period 2 data
#Name: Brady Sisk
#Date: 8/13/22
import pandas as pd
import matplotlib.pyplot as plt
df2 = pd.read_csv("formatdata2.csv")
df2.boxplot(); plt.suptitle('Period 2 box plot')
plt.show()
```



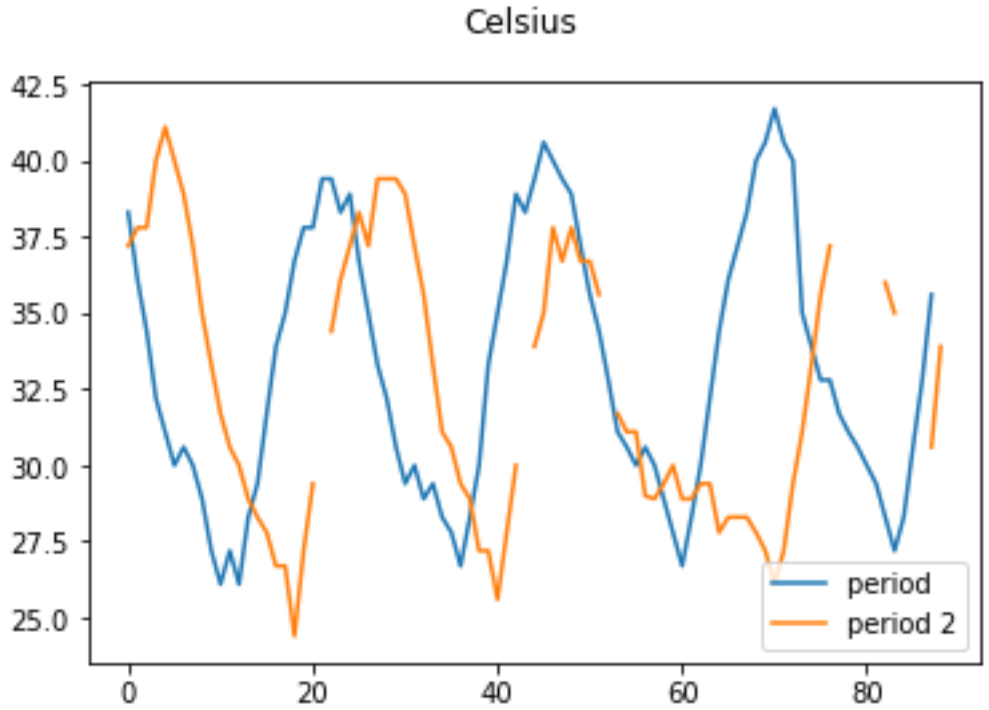
Analysis

- Think of your own question and create a chart/graph to answer it

- Your own question:

- Was there a big difference between period 1 and
- Period 2?

- Answer supported by Chart:
- No, judging by the chart I would say the periods were very similar



Code

- #Purpose: Create Celsius plot comparing period 1 and period 2
- #Name: Brady Sisk
- #Date: 8/13/22
- import pandas as pd
- import matplotlib.pyplot as plt
- df1 = pd.read_csv("formatdata.csv") #baseline data is period 1 (older)
- df2 = pd.read_csv("formatdata2.csv") #data for period 2 (more recent)
- plt.figure(); df1.Celsius.plot(label = 'period '); df2.Celsius.plot(label = 'period 2'); plt.legend(loc='best'); plt.suptitle('Celsius')
- plt.show()

Prediction

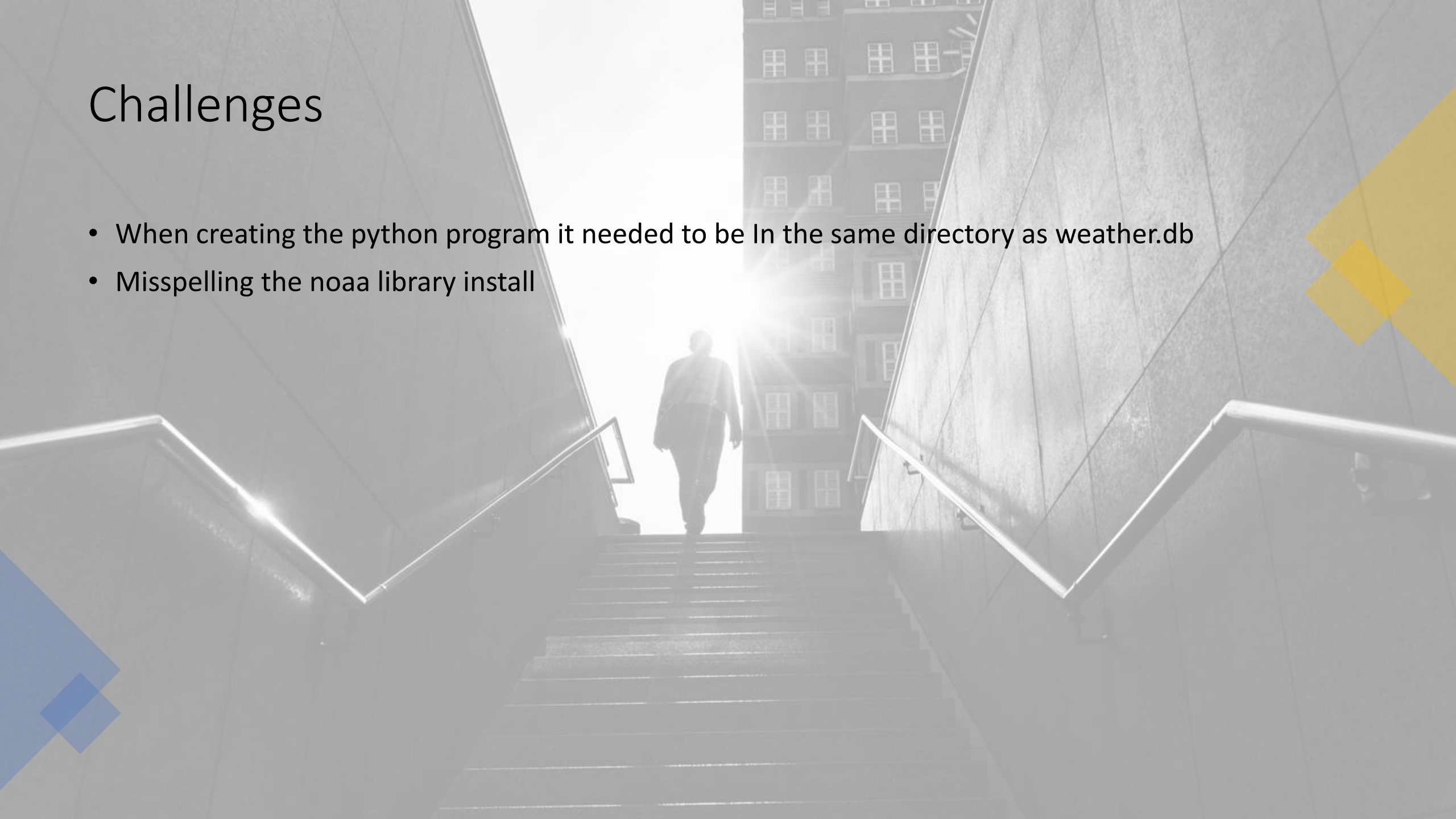
- Develop a prediction based on the data. What variations in temperature and humidity do you expect over the next few hours or days? How would humidity change if temperature goes up or down?

- Over the next few hours I would predict that the temperature will drop and the humidity rise.

- “As air temperature increases, air can hold more water molecules, and its relative humidity decreases. **When temperatures drop, relative humidity increases**”

Challenges

- When creating the python program it needed to be In the same directory as weather.db
- Misspelling the noaa library install



Career Skills

- Several career skills were gained in this project.
  - Communication – using flowcharts to depict the plan of a project
  - Database Development
  - Programming using python
  - Troubleshooting errors in the code and data cleansing
  - Analysis – Reviewing and graphs to make prediction on the data

Conclusion

- This project covered the fundamental topics of programming with data by using data gathered from cloud service to perform data analytics operations
- Building this project provided a hands on learning opportunity to put into practice the topics covered in the course.