Programming with Data

CEIS110

MONSIFF ZEIZAE

Introduction

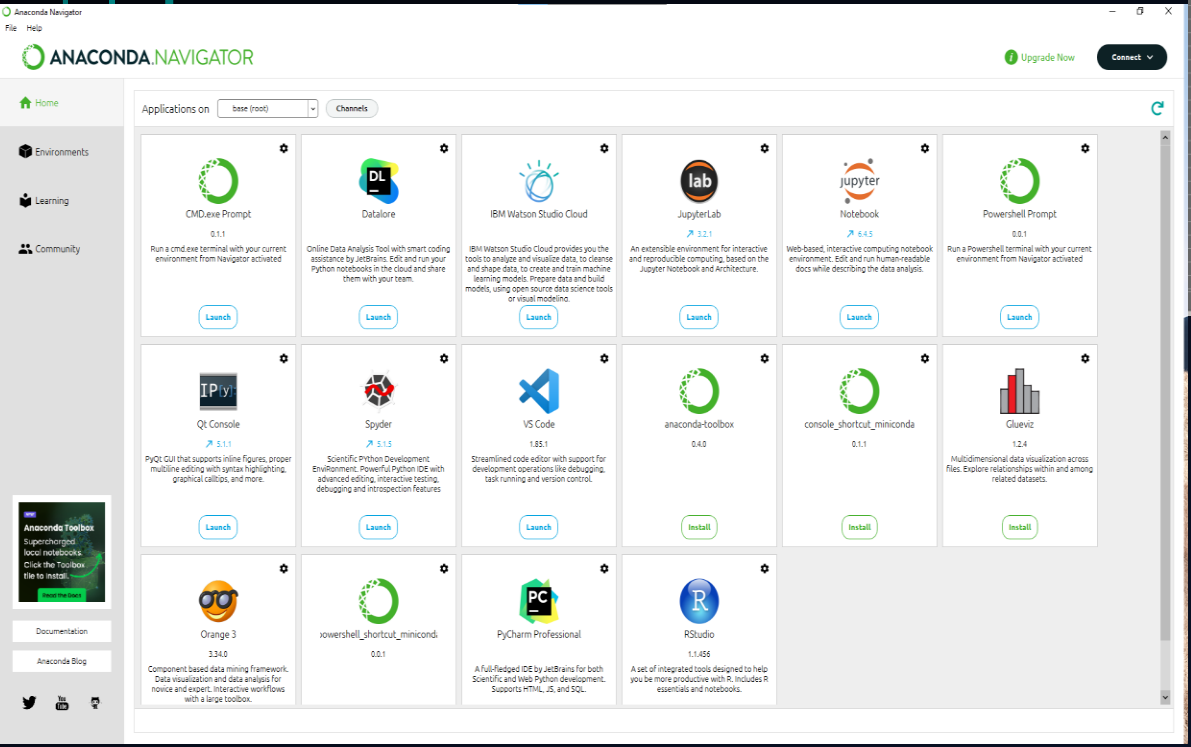
This project will demonstrate how we can use Python to analyze data that is downloaded from a cloud date source.

Software Inventory

The two softwares that will be used is Python and Microsoft Excel

Examples of the software used

Anaconda Navigator with Spyder



CMD.exe Prompt
0.1.1
Run a cmd.exe terminal with your current environment from Navigator activated
Launch

Datalore
Online Data Analysis Tool with smart coding assistance by JetBrains. Edit and run your Python notebooks in the cloud and share them with your team.
Launch

IBM Watson Studio Cloud
IBM Watson Studio Cloud provides you the tools to analyze and visualize data, to cleanse and shape data, to create and train machine learning models. Prepare data and build models, using open source data science tools or visual modeling.
Launch

JupyterLab
↗ 3.2.1
An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.
Launch

Notebook
↗ 6.4.5
Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.
Launch

Powershell Prompt
0.0.1
Run a Powershell terminal with your current environment from Navigator activated
Launch

Qt Console
↗ 5.1.1
PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.
Launch

Spyder
↗ 5.1.5
Scientific PYthon Development EnviRonment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features
Launch

VS Code
1.85.1
Streamlined code editor with support for development operations like debugging, task running and version control.
Launch

anaconda-toolbox
0.4.0
Install

console_shortcut_miniconda
0.1.1
Install

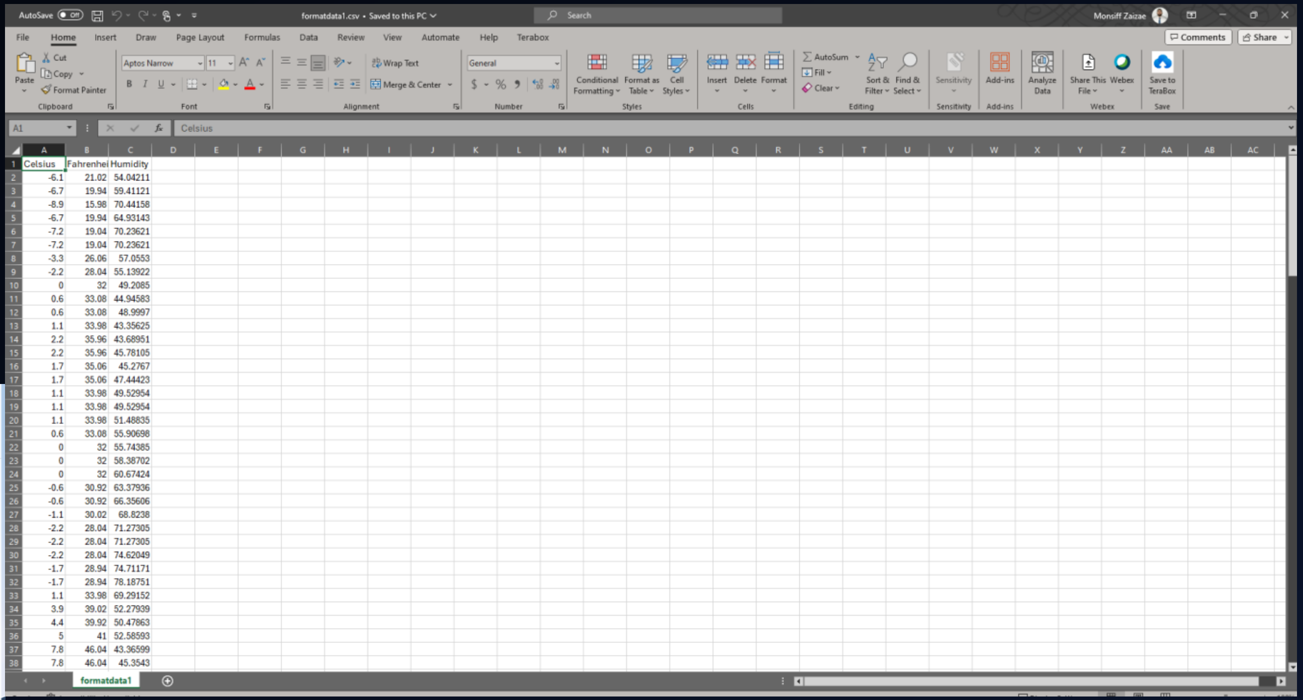
Glueviz
1.2.4
Multidimensional data visualization across files. Explore relationships within and among related datasets.
Install

Orange 3
3.34.0
Component based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflows with a large toolbox.

›owershell_shortcut_minicondi
0.0.1

PyCharm Professional
A full-fledged IDE by JetBrains for both Scientific and Web Python development. Supports HTML, JS, and SQL.

RStudio
1.1.456
A set of integrated tools designed to help you be more productive with R. Includes R essentials and notebooks.



	A	B	C
1	Celsius	Fahrenhei	Humidity
2	-6.1	21.02	54.04211
3	-6.7	19.94	59.41121
4	-8.9	15.98	70.44158
5	-6.7	19.94	64.93143
6	-7.2	19.04	70.23621
7	-7.2	19.04	70.23621
8	-3.3	26.06	57.0553
9	-2.2	28.04	55.13922
10	0	32	49.2085
11	0.6	33.08	44.94583
12	0.6	33.08	48.9997
13	1.1	33.98	43.35625
14	2.2	35.96	43.68951
15	2.2	35.96	45.78105
16	1.7	35.06	45.2767
17	1.7	35.06	47.44423
18	1.1	33.98	49.52954
19	1.1	33.98	49.52954
20	1.1	33.98	51.48835
21	0.6	33.08	55.90698
22	0	32	55.74385
23	0	32	58.38702
24	0	32	60.67424
25	-0.6	30.92	63.37936
26	-0.6	30.92	66.35606
27	-1.1	30.02	68.8238
28	-2.2	28.04	71.27305
29	-2.2	28.04	71.27305
30	-2.2	28.04	74.62049
31	-1.7	28.94	74.71171
32	-1.7	28.94	78.18751
33	1.1	33.98	69.29152
34	3.9	39.02	52.27939
35	4.4	39.92	50.47863
36	5	41	52.58593
37	7.8	46.04	43.36599
38	7.8	46.04	45.3543

Microsoft Excel

Planning & Design

To begin the planning and design of this project, a flow chart was created

It was vital to initiate the flow chart as it gave guidance on how to carry out the project

We will discuss what a flowchart is and how it can be beneficial

Expanding on Flowcharts

Simply put, a flowchart allows us to graphically illustrate a process, and

Workflows. Our flowchart displayed the process and output of the software development prokect

Flowchart

Processes in Our Flowchart:

a. Run a basic python program

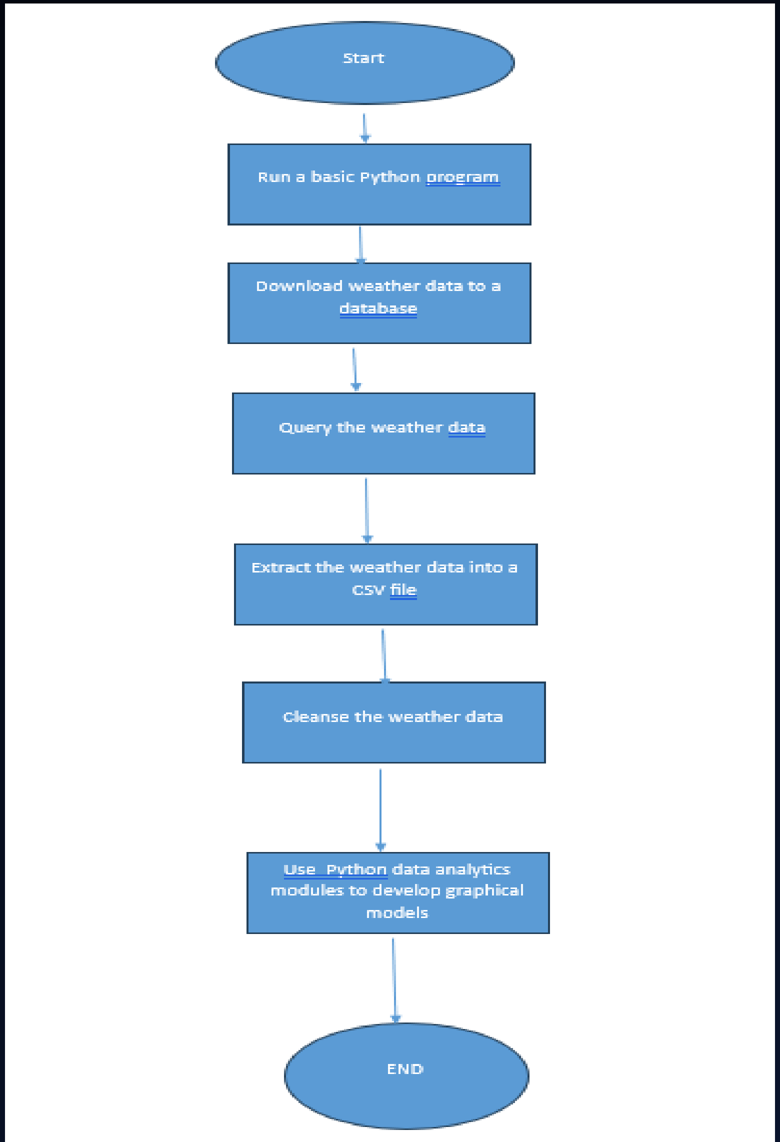
b. Download weather data to a database.

c. Query the weather data

d. Extract weather data from database into a comma separated file with python

e. Cleanse weather data

f. Use python data analytics modules to develop graphical models



Start

Run a basic Python program

Download weather data to a database

Query the weather data

Extract the weather data into a CSV file

Cleanse the weather data

Use Python data analytics modules to develop graphical models

END

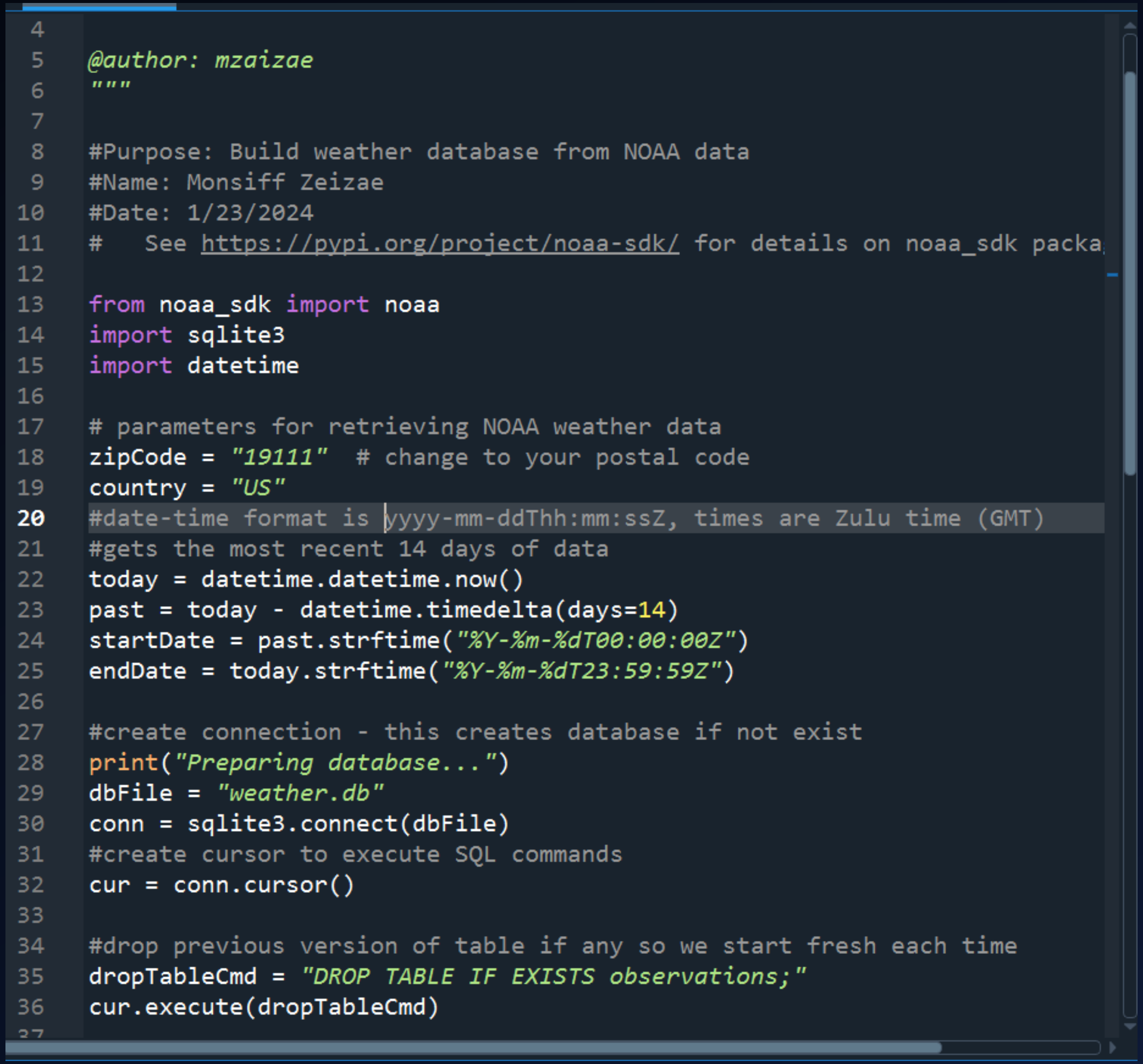
Introduction to Python

Python is a high-level, interpreted programming language known for its simplicity and readability. It was created by Guido van Rossum and first released in 1991. Python supports multiple programming paradigms, including procedural, object-oriented, and functional programming styles.

BuildWeatherDb.py Code (Screenshot)

This code will create a table named observations with the fields: timestamps, windspeed, temperature, relativeHumidity, windDirection, barometricPressure, visibility, & textDescription

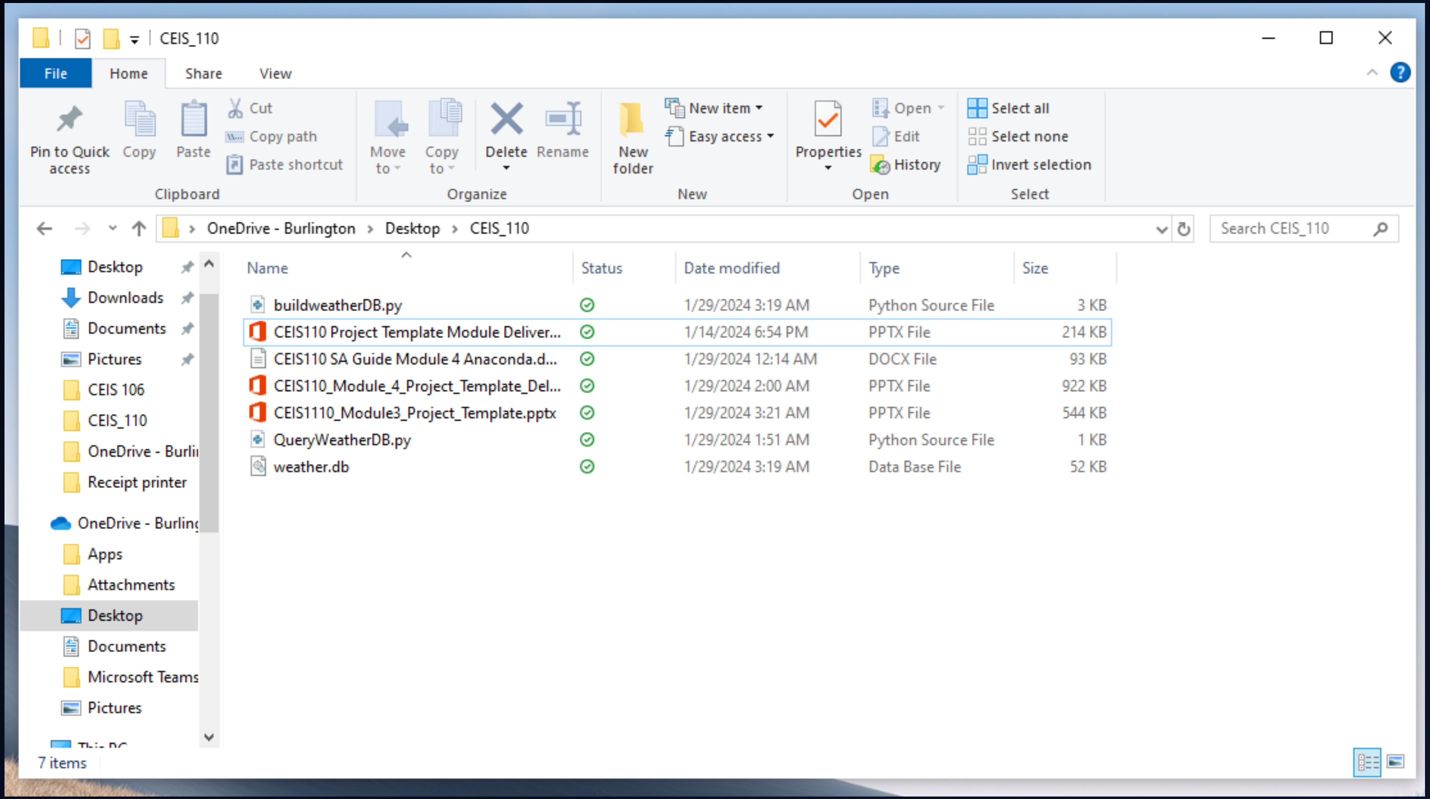
Furthermore, the database will be named weather.db and store in the same directory as the python code.



```
@author: mzaizae
"""

#Purpose: Build weather database from NOAA data
#Name: Monsiff Zeizae
#Date: 1/23/2024
#   See https://pypi.org/project/noaa-sdk/ for details on noaa_sdk packa

from noaa_sdk import noaa
import sqlite3
import datetime

# parameters for retrieving NOAA weather data
zipCode = "19111"  # change to your postal code
country = "US"
#date-time format is yyyy-mm-ddThh:mm:ssZ, times are Zulu time (GMT)
#gets the most recent 14 days of data
today = datetime.datetime.now()
past = today - datetime.timedelta(days=14)
startDate = past.strftime("%Y-%m-%dT00:00:00Z")
endDate = today.strftime("%Y-%m-%dT23:59:59Z")

#create connection - this creates database if not exist
print("Preparing database...")
dbFile = "weather.db"
conn = sqlite3.connect(dbFile)
#create cursor to execute SQL commands
cur = conn.cursor()

#drop previous version of table if any so we start fresh each time
dropTableCmd = "DROP TABLE IF EXISTS observations;"
cur.execute(dropTableCmd)
```

WEATHER.DB FILE (SCREENSHOT)

Screenshot of Windows Explorer showing database file Weather.db was created



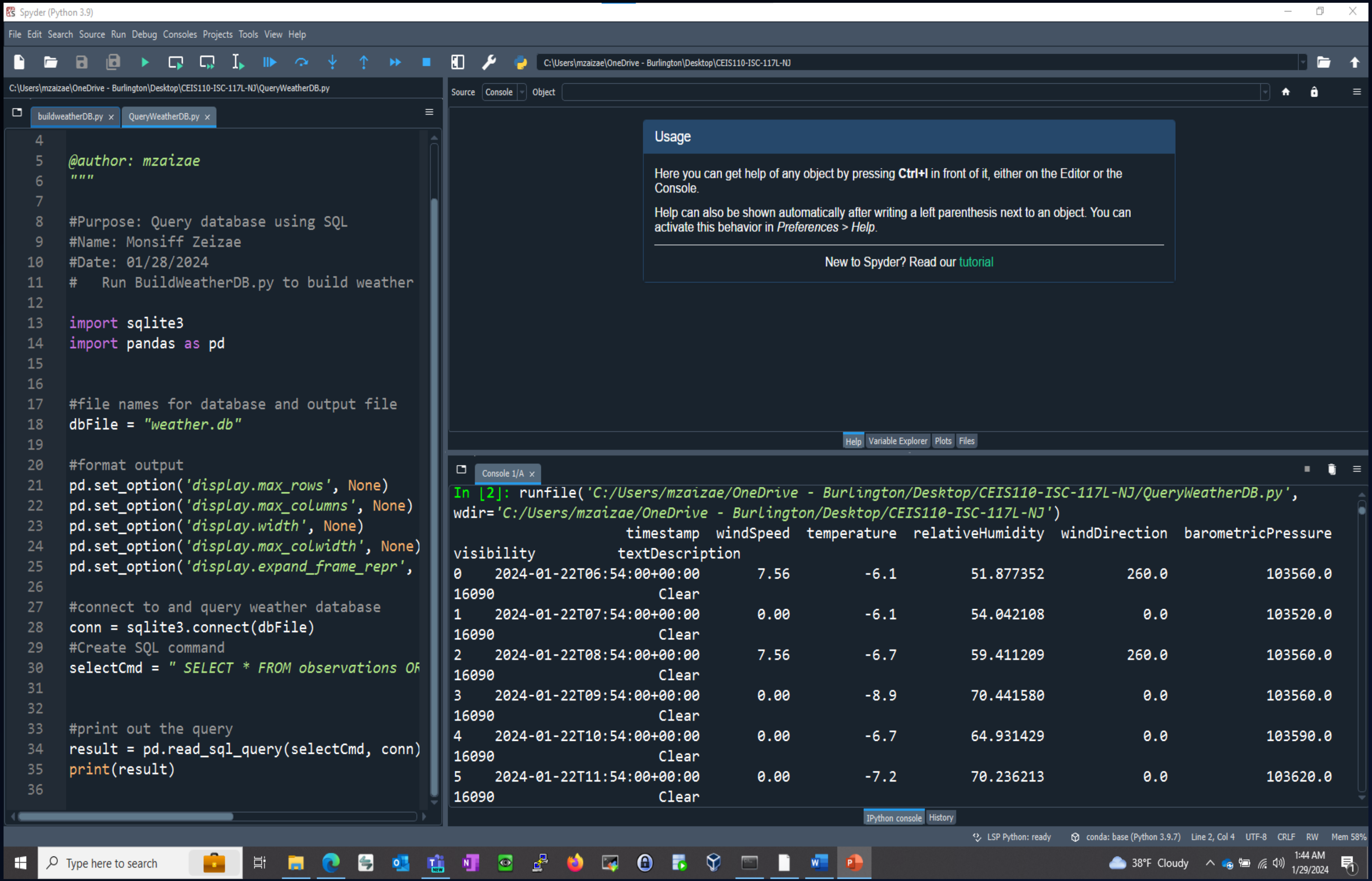
Name	Status	Date modified	Type	Size
buildweatherDB.py	✓	1/29/2024 3:19 AM	Python Source File	3 KB
CEIS110 Project Template Module Deliver...	✓	1/14/2024 6:54 PM	PPTX File	214 KB
CEIS110 SA Guide Module 4 Anaconda.d...	✓	1/29/2024 12:14 AM	DOCX File	93 KB
CEIS110_Module_4_Project_Template_Del...	✓	1/29/2024 2:00 AM	PPTX File	922 KB
CEIS1110_Module3_Project_Template.pptx	✓	1/29/2024 3:21 AM	PPTX File	544 KB
QueryWeatherDB.py	✓	1/29/2024 1:51 AM	Python Source File	1 KB
weather.db	✓	1/29/2024 3:19 AM	Data Base File	52 KB

Querying the Database

SQL or structured query language is a programming language that's used when working with relational databases

SQL was used to query the database and view the results

Query to retrieve all columns and all rows (Screenshot)


Spyder (Python 3.9)
File Edit Search Source Run Debug Consoles Projects Tools View Help
C:\Users\mzaizae\OneDrive - Burlington\Desktop\CEIS110-ISC-117L-NJ
C:\Users\mzaizae\OneDrive - Burlington\Desktop\CEIS110-ISC-117L-NJ\QueryWeatherDB.py
buildweatherDB.py
QueryWeatherDB.py
Source
Console
Object
Help
Variable Explorer
Plots
Files
Console 1/A
IPython console
History
LSP Python: ready
conda: base (Python 3.9.7)
Line 2, Col 4
UTF-8
CRLF
RW
Mem 58%
Type here to search
38°F Cloudy
1:44 AM
1/29/2024

```python

@author: mzaizae
"""

#Purpose: Query database using SQL
#Name: Monsiff Zeizae
#Date: 01/28/2024
#   Run BuildWeatherDB.py to build weather

import sqlite3
import pandas as pd


#file names for database and output file
dbFile = "weather.db"

#format output
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
pd.set_option('display.width', None)
pd.set_option('display.max_colwidth', None)
pd.set_option('display.expand_frame_repr',

#connect to and query weather database
conn = sqlite3.connect(dbFile)
#Create SQL command
selectCmd = " SELECT * FROM observations OR


#print out the query
result = pd.read_sql_query(selectCmd, conn)
print(result)

```

Usage

Here you can get help of any object by pressing **Ctrl+I** in front of it, either on the Editor or the Console.

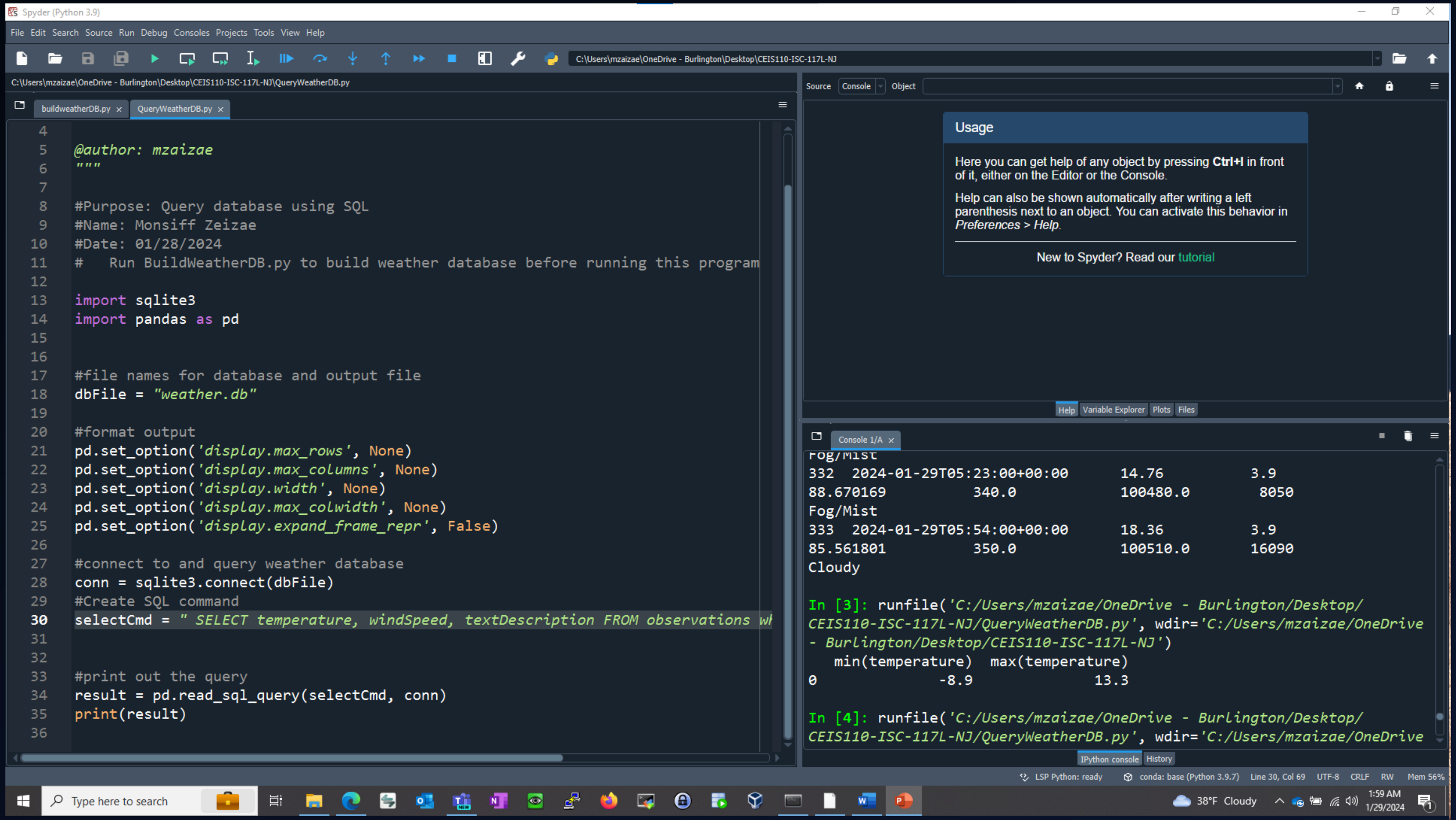
Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in *Preferences > Help*.

New to Spyder? Read our tutorial

```
In [2]: runfile('C:/Users/mzaizae/OneDrive - Burlington/Desktop/CEIS110-ISC-117L-NJ/QueryWeatherDB.py',
wdir='C:/Users/mzaizae/OneDrive - Burlington/Desktop/CEIS110-ISC-117L-NJ')
                   timestamp  windSpeed  temperature  relativeHumidity  windDirection  barometricPressure
visibility           textDescription
0    2024-01-22T06:54:00+00:00       7.56         -6.1         51.877352          260.0            103560.0
16090                      Clear
1    2024-01-22T07:54:00+00:00       0.00         -6.1         54.042108            0.0            103520.0
16090                      Clear
2    2024-01-22T08:54:00+00:00       7.56         -6.7         59.411209          260.0            103560.0
16090                      Clear
3    2024-01-22T09:54:00+00:00       0.00         -8.9         70.441580            0.0            103560.0
16090                      Clear
4    2024-01-22T10:54:00+00:00       0.00         -6.7         64.931429            0.0            103590.0
16090                      Clear
5    2024-01-22T11:54:00+00:00       0.00         -7.2         70.236213            0.0            103620.0
16090                      Clear
```

Query to retrieve lowest and highest temperatures (Screenshot)

The min and mac temperature were retrieved. These temperatures are reported in celsius


Spyder (Python 3.9)
File Edit Search Source Run Debug Consoles Projects Tools View Help
C:\Users\mzaizae\OneDrive - Burlington\Desktop\CEIS110-ISC-117L-NJ
C:\Users\mzaizae\OneDrive - Burlington\Desktop\CEIS110-ISC-117L-NJ\QueryWeatherDB.py
buildweatherDB.py
QueryWeatherDB.py
Source Console Object
Help Variable Explorer Plots Files
Console 1/A
IPython console History
LSP Python: ready conda: base (Python 3.9.7) Line 30, Col 69 UTF-8 CRLF RW Mem 56%
Type here to search
38°F Cloudy
1:59 AM
1/29/2024

```python

@author: mzaizae
"""

#Purpose: Query database using SQL
#Name: Monsiff Zeizae
#Date: 01/28/2024
#   Run BuildWeatherDB.py to build weather database before running this program

import sqlite3
import pandas as pd


#file names for database and output file
dbFile = "weather.db"

#format output
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
pd.set_option('display.width', None)
pd.set_option('display.max_colwidth', None)
pd.set_option('display.expand_frame_repr', False)

#connect to and query weather database
conn = sqlite3.connect(dbFile)
#Create SQL command
selectCmd = " SELECT temperature, windSpeed, textDescription FROM observations wh


#print out the query
result = pd.read_sql_query(selectCmd, conn)
print(result)

```

Usage

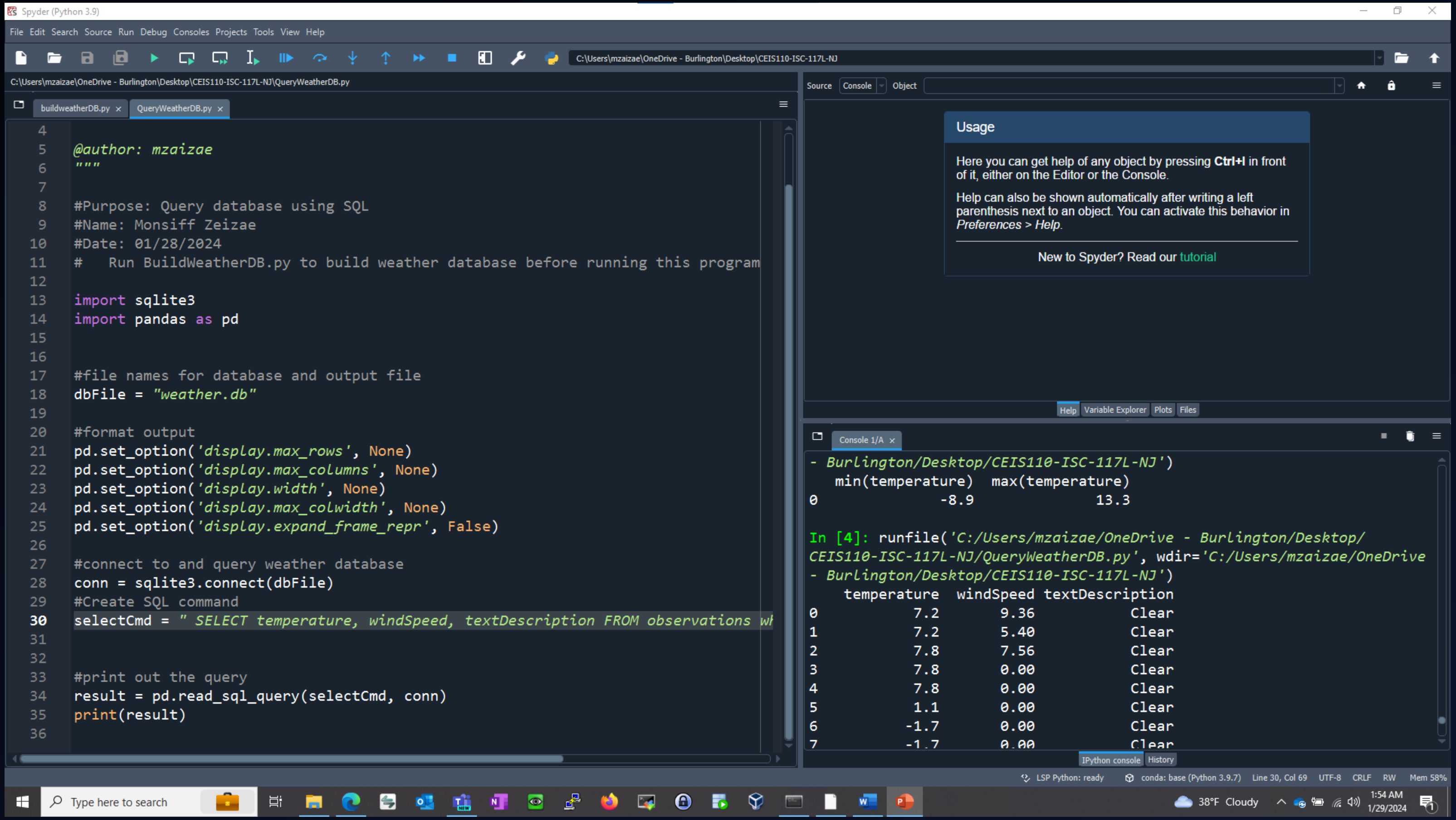
Here you can get help of any object by pressing **Ctrl+I** in front of it, either on the Editor or the Console.

Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in *Preferences > Help*.

New to Spyder? Read our tutorial

```
Fog/Mist
332  2024-01-29T05:23:00+00:00      14.76          3.9
88.670169          340.0           100480.0        8050
Fog/Mist
333  2024-01-29T05:54:00+00:00      18.36          3.9
85.561801          350.0           100510.0       16090
Cloudy

In [3]: runfile('C:/Users/mzaizae/OneDrive - Burlington/Desktop/
CEIS110-ISC-117L-NJ/QueryWeatherDB.py', wdir='C:/Users/mzaizae/OneDrive
- Burlington/Desktop/CEIS110-ISC-117L-NJ')
   min(temperature)  max(temperature)
0              -8.9              13.3

In [4]: runfile('C:/Users/mzaizae/OneDrive - Burlington/Desktop/
CEIS110-ISC-117L-NJ/QueryWeatherDB.py', wdir='C:/Users/mzaizae/OneDrive
```

Query to retrieve all clear days (Screenshot)



Spyder (Python 3.9)
File Edit Search Source Run Debug Consoles Projects Tools View Help
C:\Users\mzaizae\OneDrive - Burlington\Desktop\CEIS110-ISC-117L-NJ
C:\Users\mzaizae\OneDrive - Burlington\Desktop\CEIS110-ISC-117L-NJ\QueryWeatherDB.py
buildweatherDB.py
QueryWeatherDB.py
Source
Console
Object

```python

@author: mzaizae
"""

#Purpose: Query database using SQL
#Name: Monsiff Zeizae
#Date: 01/28/2024
#   Run BuildWeatherDB.py to build weather database before running this program

import sqlite3
import pandas as pd


#file names for database and output file
dbFile = "weather.db"

#format output
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
pd.set_option('display.width', None)
pd.set_option('display.max_colwidth', None)
pd.set_option('display.expand_frame_repr', False)

#connect to and query weather database
conn = sqlite3.connect(dbFile)
#Create SQL command
selectCmd = " SELECT temperature, windSpeed, textDescription FROM observations wh


#print out the query
result = pd.read_sql_query(selectCmd, conn)
print(result)

```

Usage

Here you can get help of any object by pressing **Ctrl+I** in front of it, either on the Editor or the Console.

Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in *Preferences > Help*.

New to Spyder? Read our tutorial

```
- Burlington/Desktop/CEIS110-ISC-117L-NJ')
   min(temperature)  max(temperature)
0              -8.9              13.3

In [4]: runfile('C:/Users/mzaizae/OneDrive - Burlington/Desktop/
CEIS110-ISC-117L-NJ/QueryWeatherDB.py', wdir='C:/Users/mzaizae/OneDrive
- Burlington/Desktop/CEIS110-ISC-117L-NJ')
    temperature  windSpeed textDescription
0           7.2       9.36           Clear
1           7.2       5.40           Clear
2           7.8       7.56           Clear
3           7.8       0.00           Clear
4           7.8       0.00           Clear
5           1.1       0.00           Clear
6          -1.7       0.00           Clear
7          -1.7       0.00           Clear
```

Data Cleansing

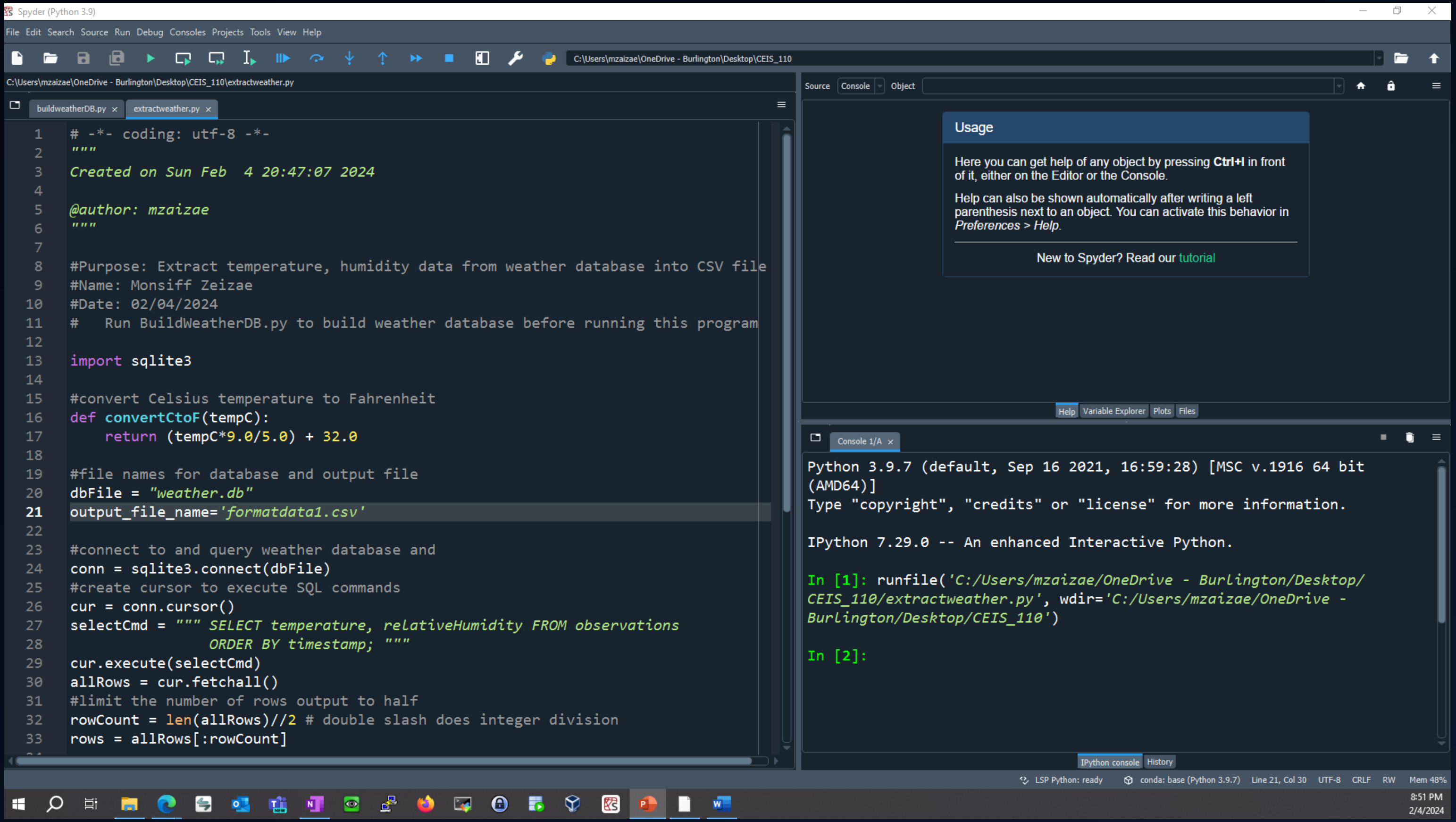
Data output from machines can posses errors. To offset this, programs can automatically put it in the format needed to be read by the other programs when cleansing the data.

A python program is reading in 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 error or missing values altogether in a dataset.

ExtractTempHumidity.py Python code screenshot

The code used will only retrieve the temperature and humidity values & report them to a csv file.



Spyder (Python 3.9)
File Edit Search Source Run Debug Consoles Projects Tools View Help
C:\Users\mzaizae\OneDrive - Burlington\Desktop\CEIS_110
C:\Users\mzaizae\OneDrive - Burlington\Desktop\CEIS_110\extractweather.py
buildweatherDB.py
extractweather.py
Source
Console
Object

```python
# -*- coding: utf-8 -*-
"""
Created on Sun Feb  4 20:47:07 2024

@author: mzaizae
"""

#Purpose: Extract temperature, humidity data from weather database into CSV file
#Name: Monsiff Zeizae
#Date: 02/04/2024
#   Run BuildWeatherDB.py to build weather database before running this program

import sqlite3

#convert Celsius temperature to Fahrenheit
def convertCtoF(tempC):
    return (tempC*9.0/5.0) + 32.0

#file names for database and output file
dbFile = "weather.db"
output_file_name='formatdata1.csv'

#connect to and query weather database and
conn = sqlite3.connect(dbFile)
#create cursor to execute SQL commands
cur = conn.cursor()
selectCmd = """ SELECT temperature, relativeHumidity FROM observations
                ORDER BY timestamp; """
cur.execute(selectCmd)
allRows = cur.fetchall()
#limit the number of rows output to half
rowCount = len(allRows)//2 # double slash does integer division
rows = allRows[:rowCount]
```

Usage

Here you can get help of any object by pressing **Ctrl+I** in front of it, either on the Editor or the Console.

Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in *Preferences > Help*.

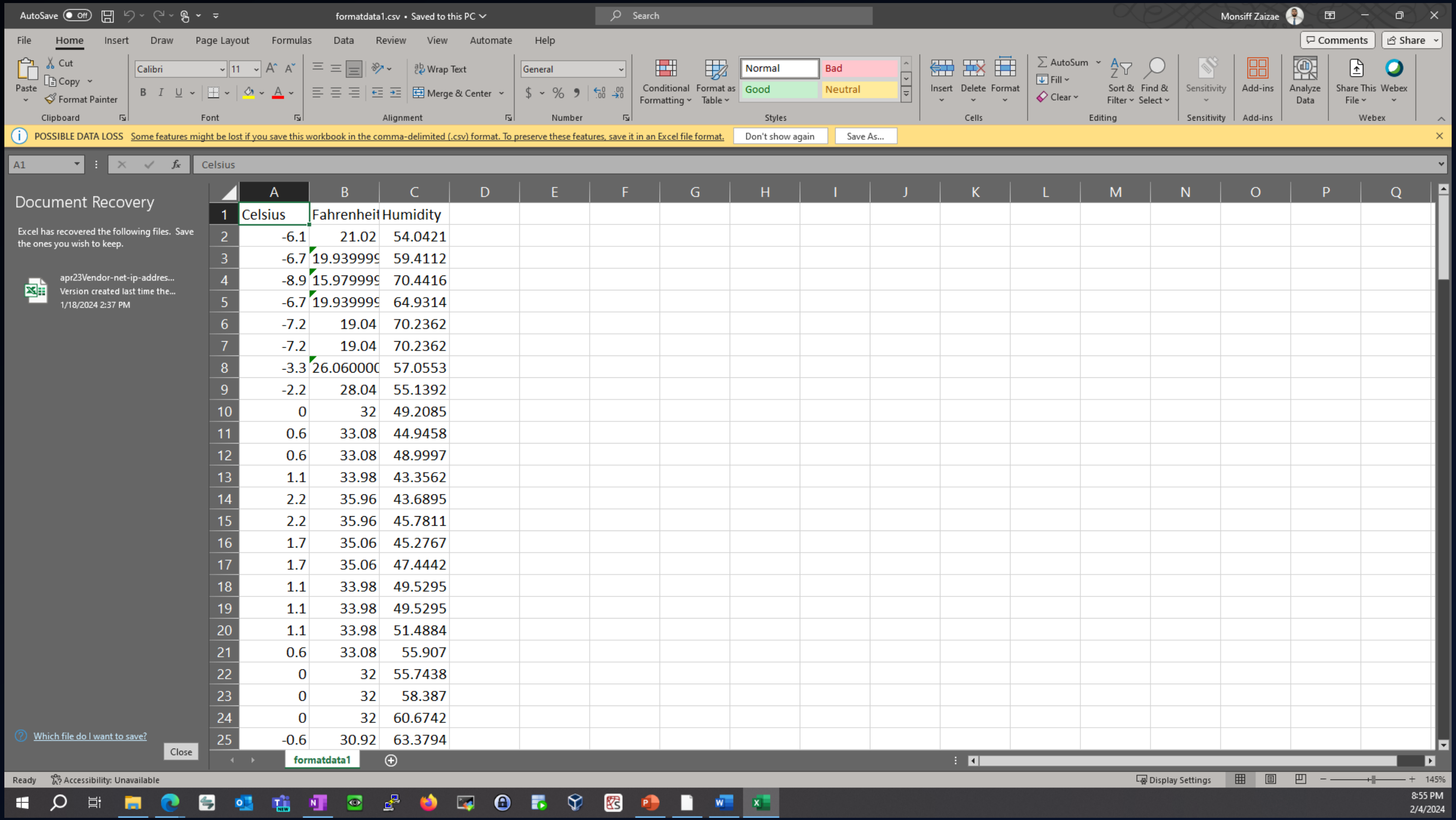
New to Spyder? Read our tutorial

```
Python 3.9.7 (default, Sep 16 2021, 16:59:28) [MSC v.1916 64 bit
(AMD64)]
Type "copyright", "credits" or "license" for more information.

IPython 7.29.0 -- An enhanced Interactive Python.

In [1]: runfile('C:/Users/mzaizae/OneDrive - Burlington/Desktop/
CEIS_110/extractweather.py', wdir='C:/Users/mzaizae/OneDrive -
Burlington/Desktop/CEIS_110')

In [2]:
```

Retrieve and Convert Data to CSV Format (Screenshot)

Our python code created a csv file called formdata

This file contains data for Celsius, Fahrenheit and humidity



formatdata1.csv • Saved to this PC

POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format.

Document Recovery

Excel has recovered the following files. Save the ones you wish to keep.

apr23Vendor-net-ip-addres...
Version created last time the...
1/18/2024 2:37 PM

Which file do I want to save?

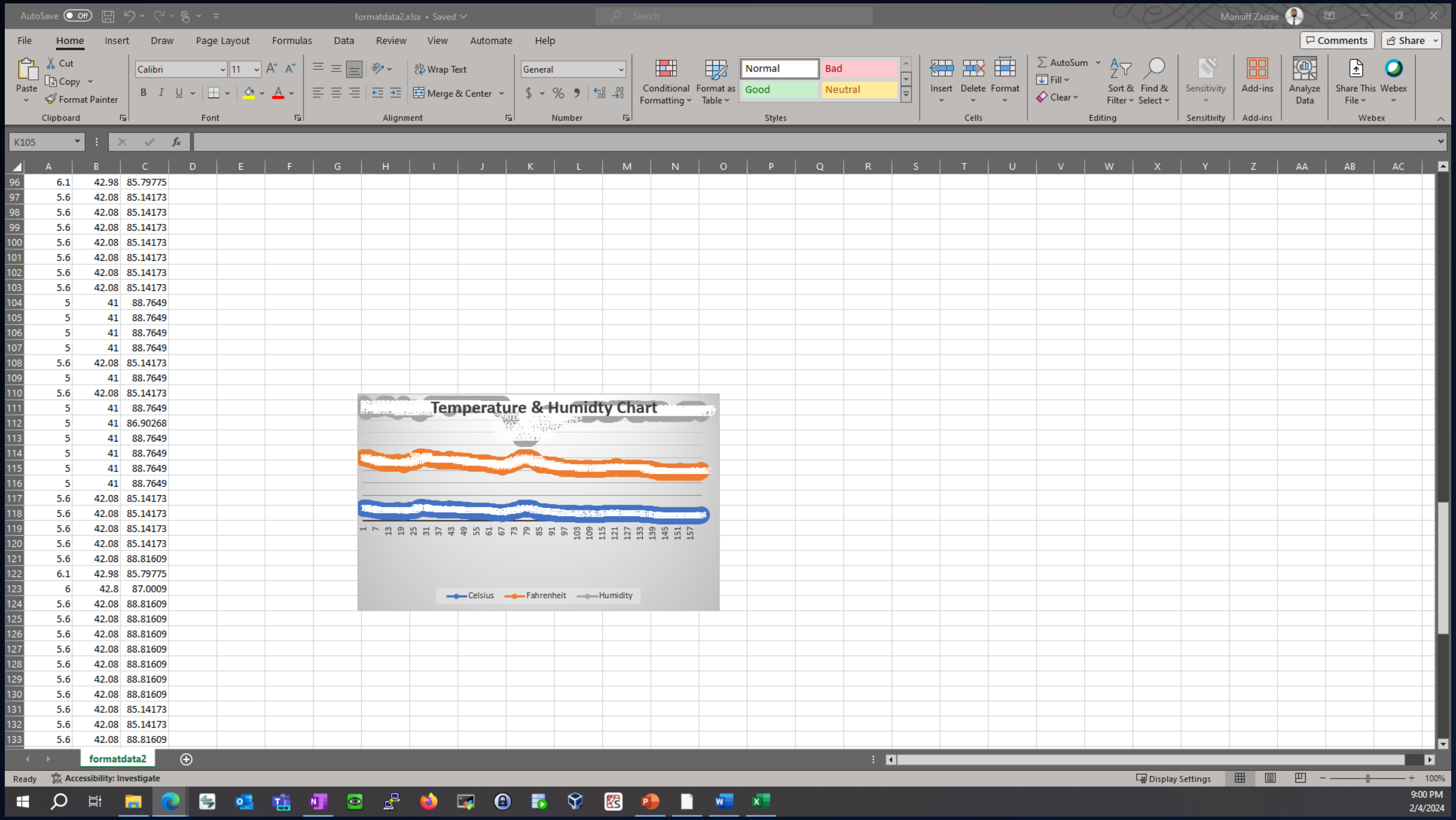
A1 Celsius

	A	B	C
1	Celsius	Fahrenheit	Humidity
2	-6.1	21.02	54.0421
3	-6.7	19.939999	59.4112
4	-8.9	15.979999	70.4416
5	-6.7	19.939999	64.9314
6	-7.2	19.04	70.2362
7	-7.2	19.04	70.2362
8	-3.3	26.060000	57.0553
9	-2.2	28.04	55.1392
10	0	32	49.2085
11	0.6	33.08	44.9458
12	0.6	33.08	48.9997
13	1.1	33.98	43.3562
14	2.2	35.96	43.6895
15	2.2	35.96	45.7811
16	1.7	35.06	45.2767
17	1.7	35.06	47.4442
18	1.1	33.98	49.5295
19	1.1	33.98	49.5295
20	1.1	33.98	51.4884
21	0.6	33.08	55.907
22	0	32	55.7438
23	0	32	58.387
24	0	32	60.6742
25	-0.6	30.92	63.3794

formatdata1

Temperature and Humidity Chart (Graph)

Excel chart based on temperature and humidity data from database



	A	B	C
96	6.1	42.98	85.79775
97	5.6	42.08	85.14173
98	5.6	42.08	85.14173
99	5.6	42.08	85.14173
100	5.6	42.08	85.14173
101	5.6	42.08	85.14173
102	5.6	42.08	85.14173
103	5.6	42.08	85.14173
104	5	41	88.7649
105	5	41	88.7649
106	5	41	88.7649
107	5	41	88.7649
108	5.6	42.08	85.14173
109	5	41	88.7649
110	5.6	42.08	85.14173
111	5	41	88.7649
112	5	41	86.90268
113	5	41	88.7649
114	5	41	88.7649
115	5	41	88.7649
116	5	41	88.7649
117	5.6	42.08	85.14173
118	5.6	42.08	85.14173
119	5.6	42.08	85.14173
120	5.6	42.08	85.14173
121	5.6	42.08	88.81609
122	6.1	42.98	85.79775
123	6	42.8	87.0009
124	5.6	42.08	88.81609
125	5.6	42.08	88.81609
126	5.6	42.08	88.81609
127	5.6	42.08	88.81609
128	5.6	42.08	88.81609
129	5.6	42.08	88.81609
130	5.6	42.08	88.81609
131	5.6	42.08	85.14173
132	5.6	42.08	85.14173
133	5.6	42.08	88.81609

Data Analytics

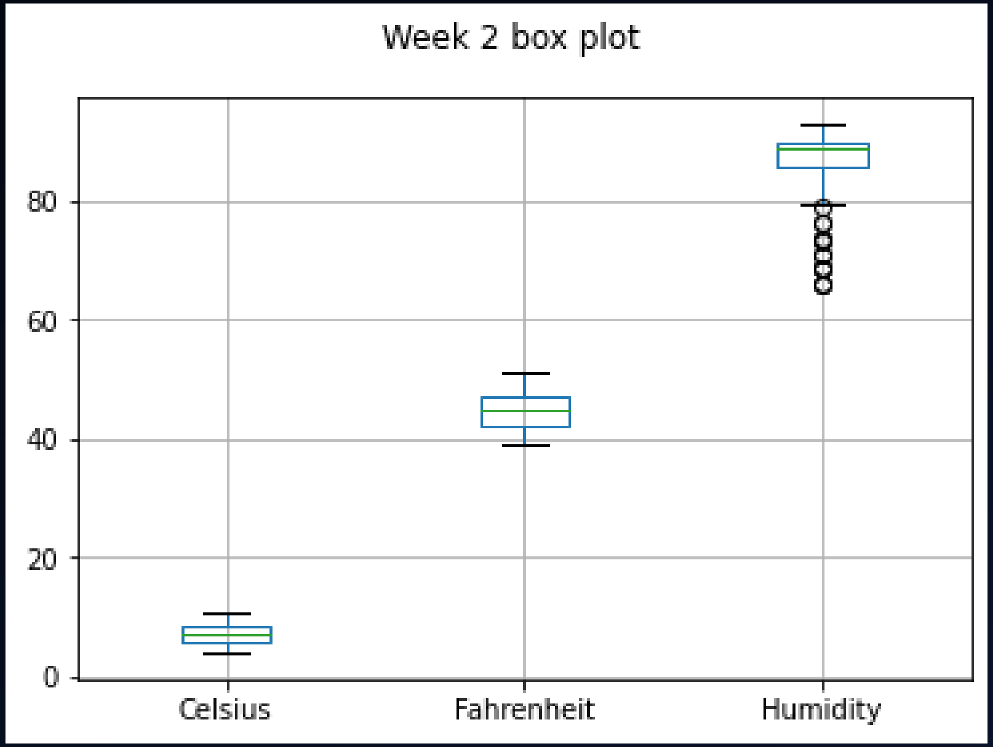
Python contains data analytics modules that allow us to create charts and graphs that represent our data

Plots were generated analyzing both humidity and temperature

By downloading Anaconda, we are able to call upon these modules when needed as they are available within Anaconda

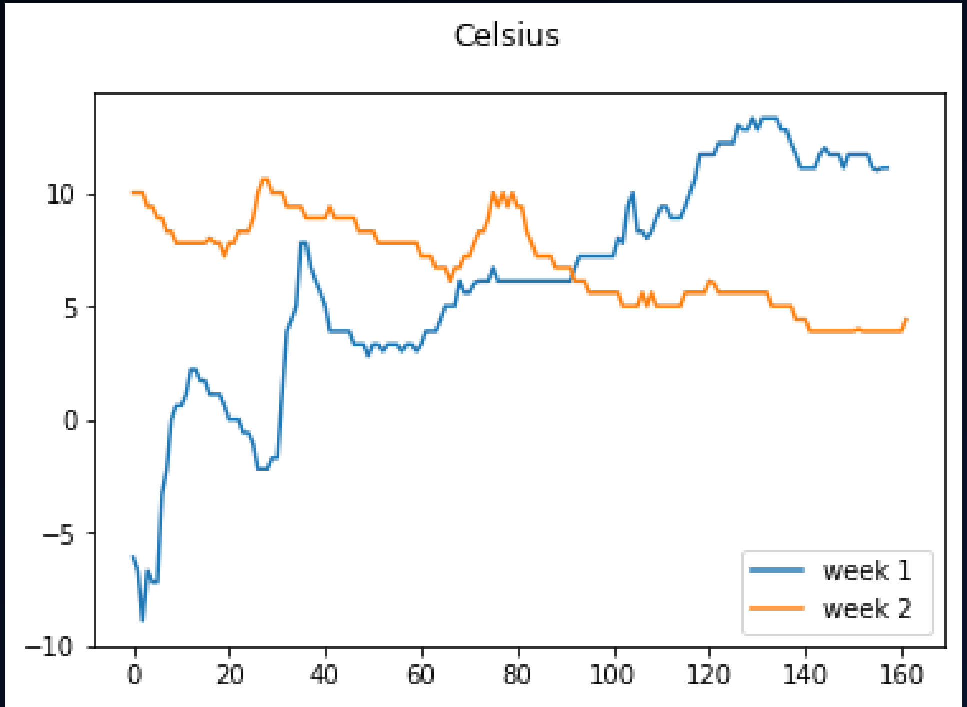
Plot 1

```
#Purpose: Create box plot for week 2 data
#Name: Monsiff Zeizae
#Date: 02/10/2024
import pandas as pd
import matplotlib.pyplot as plt
df2 = pd.read_csv("formatdata2.csv")
df2.boxplot(); plt.suptitle('Week 2 box plot')
#plt.show()
print(df2.info())
print(df2.describe())
print("The median is",df2.median())
```



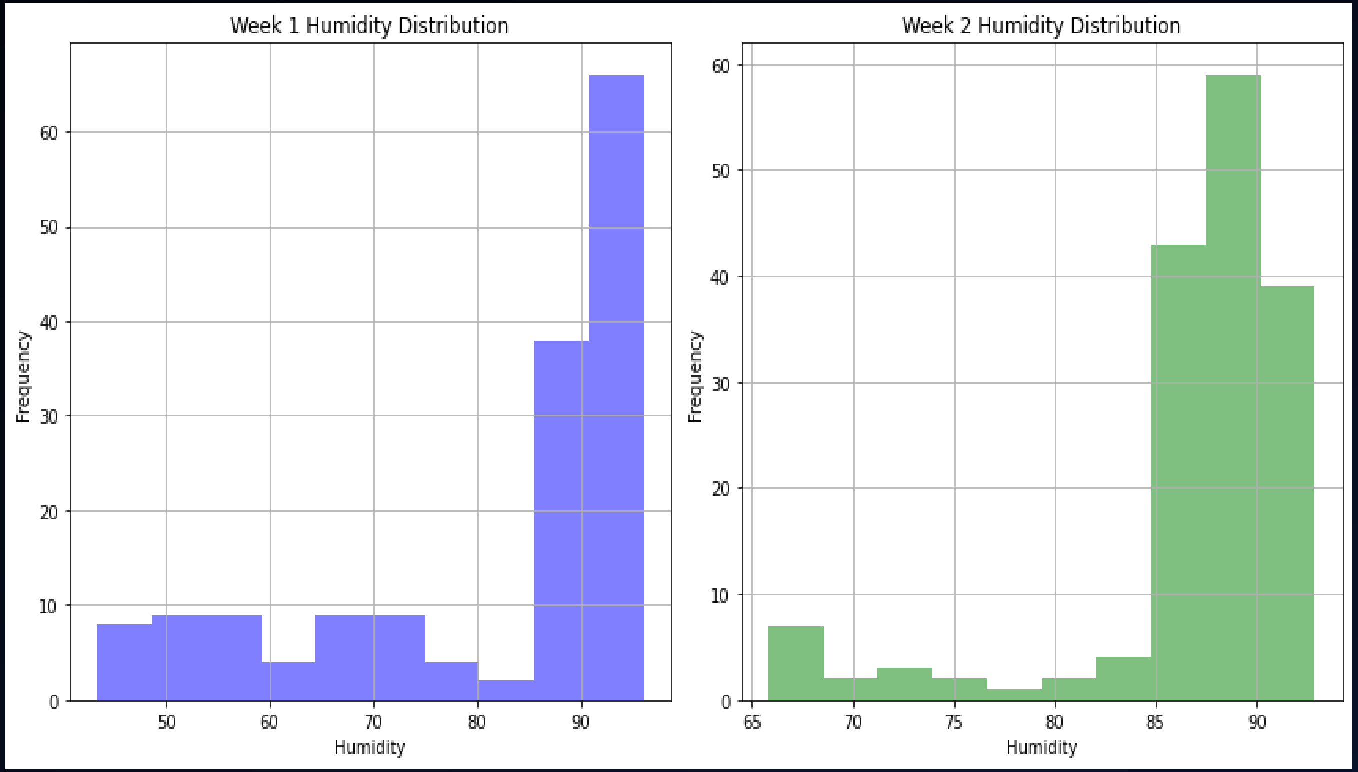
Plot #2

```
#Purpose: Create Celsius plot comparing period 1 and period 2 (week 1 and week 2)
#Name: Monsiff Zeizae
#Date: 02/10/2024
import pandas as pd
import matplotlib.pyplot as plt
df1 = pd.read_csv("formatdata1.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 = 'week 1 ');
df2.Celsius.plot(label = 'week 2');
plt.legend(loc='best'); plt.suptitle('Celsius')
plt.show()
```



Analysis

- Think of your own question and create a chart/graph to answer it
  - In the video we adjusted our code for humidity so I decided to build on that & compare humidity levels from the two different files/weeks



Challenge

On module 3 of the project I faced a challenge when running the pip install noaa_sdk in the console window. For some reason it would not execute and for that reason I was not able to successfully complete the project. After some instructions from the professor, I uninstalled anaconda altogether and re-installed it which resolved the issue.

Career Skills Gained

Ability to program using Python

Ability to retrieve data from relational databases

Ability to utilize a flowchart to better organize thoughts and program outlines

Ability to accurately analyze data, and utilize the graphing function of excel to represent our data

Conclusion

In conclusion, this project successfully demonstrated programming with data that was obtained from a cloud service, and also performing analysis on that data

After extensive work on this project, I feel fully equipped and able to further practice manipulating data using the python programming language