**Course Project**

**DeVry University**

**College of Engineering and Information Sciences**

**Course Number: CEIS110**

# Module 6: Develop Graphical Models and Interpret Results

# Objectives

* To learn how to use Python data analytics modules
* To compare data at different times.
* Use data analytics to interpret and analyze data from a database

# Introduction

IoT devices, websites, social media, and other sources generate a significant amount of data. This data must be gathered, stored, cleansed, processed, and analyzed. Each step in this project has focused on gathering and processing the data. This part of the project focuses on data analysis with pandas, the data analytics module in Python. The pandas module is included as part of the Anaconda distribution of Python.

You should have two CSV files and at least one xlsx file. Each CSV file probably contains about one week of weather data (this might be different for your location). The first CSV file contains data for the first (older) time period and the second CSV file contains data for the second (more recent) time period.

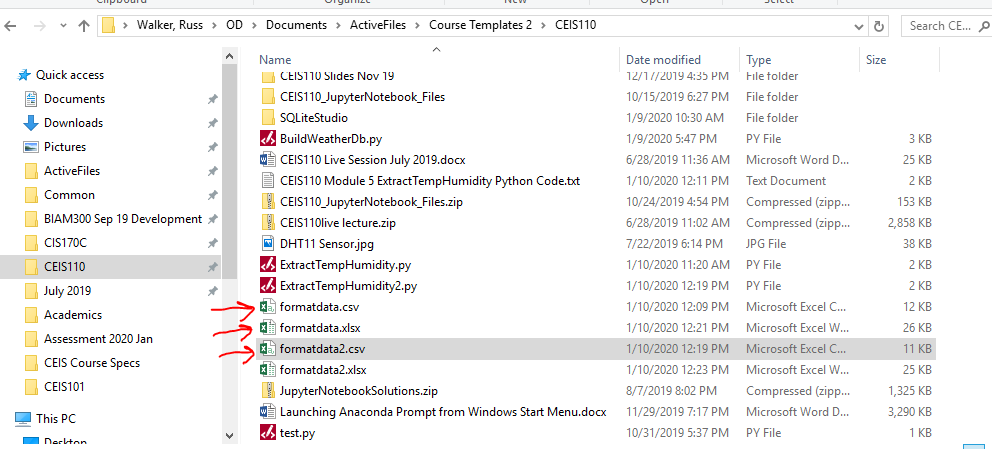


Figure 1: Files from previous step

We can create some plots to review and analyze the data. There are many different plots you can use in matplotlib and pandas. You can use the read\_csv command to read in all of the data into a dataframe, which is the basis for data analysis. Several plots can be generated based on the data. A few are listed below with the code associated.

**The boxplot() command creates a box plot.** A box plot is useful to display the distribution of data. The code to create this plot is below.

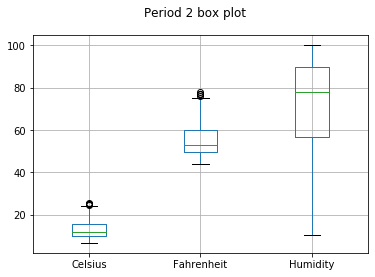


Figure 2: Box plot

#Purpose: Create box plot for period 2 data

#Name: Your name

#Date: Your date

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()

**The plot() command creates a line chart with a x and y axis.** The plot() command is useful to compare two sets of data. The code to create the plot is below.

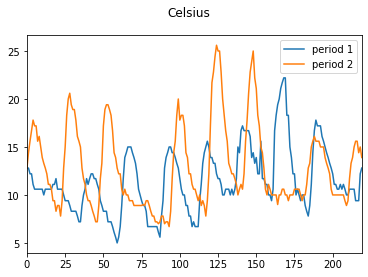


Figure 3: Plot command

The above plot is created with the following code:

#Purpose: Create Celsius plot comparing period 1 and period 2

#Name: Your name

#Date: Your date

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()

**The hist() command creates a histogram** which is a graphical distribution of data. Data is grouped into ranges and plotted in the histogram.

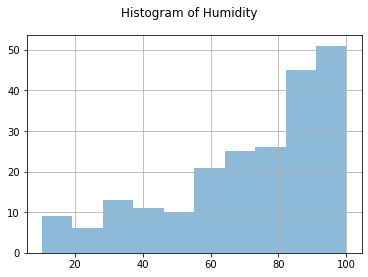


Figure 4: Histogram

#Purpose: Create a histogram of humidity data from the second period

#Name: Your name

#Date: Your date

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')

plt.show()

**The scatter() command is used for scatter plots.**

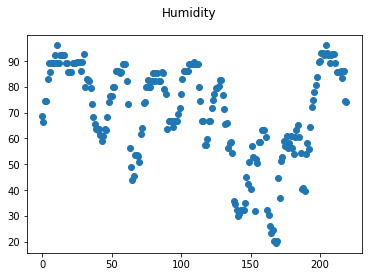


Figure 5: Scatter plot

#Purpose: Create scatter plot of humidity for period 1. Can replace df1 to df2 to display second period data

#Name: Your name

#Date: Your date

import pandas as pd

import matplotlib.pyplot as plt

df1 = pd.read\_csv("formatdata.csv")

df2 = pd.read\_csv("formatdata2.csv")

plt.scatter(df1.index.values,df1['Humidity']); plt.suptitle('Humidity')

plt.show()

Period 1 was the older data and period 2 contained the more recent data. To view the differences between these periods, think about some of the following:

1. Were there differences in temperature and humidity between the two periods? Was one period warmer or cooler, or dryer or wetter, than the other? Generate a plot showing temperature from period 1 and period 2 on the same plot and another plot showing humidity from period 1 and period 2 on the same plot.
2. Use a box plot to determine for Celsius, Fahrenheit, and Humidity where the greatest variation in ranges of numbers occurred for both trials. Did humidity or temperature vary more during the first or second periods?
3. Find the approximate average value from Period 1 and Period 2 for Humidity.
4. Review the data and analyze it. Think of your own question and create a chart/graph to answer it.
5. Develop a prediction based on the data. You have looked at temperature and humidity data for about the past two weeks. Based on this, what variations in temperature and humidity do you expect to see in the next few hours or next few days? If the temperature goes up or down, what might be expected to happen to humidity?

# Deliverables Module 6

* Complete the Course Project Presentation Template
* Create two plots
* Analyze data
* Create a prediction based on the data