Puja Roy 3/4/22

CET 4900 - OL60

Internship Journal Entry #6

Throughout this week of my internship, I worked on analyzing and visualizing the Bay View site dataset of real-time data collected from the New York Urban Hydro-Meteorological Testbed (NY-uHMT) weather station. I printed the pwd command which prints the working directory of a pathname or file that you are working with or want to work with in a virtual environment. I imported the dataset of Bay View (Site 10). Since I am mostly working with data science and analysis, I mostly worked with Python (Numpy, Pandas, Matplotlib and Seaborn) to analyze and visualize the data. As shown below in Fig 1, I imported the Bay View dataset into a DataFrame object by importing pandas as pd. Pandas is a Python library and it is mostly used to analyze data. Then, I printed the dataframe (df) which showed the output of my data. The Bay View data has 89,211 rows and 8 columns. One of the major goals of my technical engineering internship is to analyze the trends of the variables in each dataset which is time consuming. If any part of the code runs into errors, I spend time troubleshooting and testing various codes and algorithms to process the data efficiently. I also research which Python libraries would be essential to visualize the datasets.

Import the Data

```
In [3]: df= pd.read_csv('C:\\Users\\prati\\UHMTProject\\Data\\Site10_BayView_Fifteen.txt')
        print(df)
```

```
                  TIMESTAMP RECORD VWC     T             P        VWC_2   T_2           P_2      VWC_3   T_3                 P_3
                  TS        RN     m^3/m^3 Deg F         unitless m^3/m^3 Deg F         unitless m^3/m^3 Deg F           unitless
                  NaN       NaN    Smp     Smp           Smp      Smp     Smp           Smp      Smp     Smp                   Smp
                  2017-04-03 12:15:00 0      NAN     14241.87      0        NAN     14241.87      0        NAN     14241.87        0
                  2017-04-03 12:30:00 1      NAN     9.635455E+007 0        NAN     9.635455E+007 0        NAN     9.635455E+007   0
                  ...                                                                                                             ...
                  2021-04-13 16:15:00 124433 0.1964  71.16908      10.421   0.2152  73.6547       11.4257  0.1961  74.57396       10.4061
                  2021-04-13 16:30:00 124434 0.1957  71.82644      10.3847  0.2153  74.62868      11.4283  0.1955  75.89462       10.3701
                  2021-04-13 16:45:00 124435 0.1958  70.6604       10.3859  0.2146  74.5988       11.3879  0.1947  72.554         10.3304
                  2021-04-13 17:00:00 124436 0.1944  68.41688      10.3129  0.2124  68.14598      11.2734  0.1941  66.52615       10.2973
                  2021-04-13 17:15:00 124437 0.1936  64.28606      10.2755  0.2125  63.7097       11.2759  0.1941  63.71456       10.2998

                                                                                                                      Site10_BayView
                  \
                  TIMESTAMP RECORD VWC     T             P        VWC_2   T_2           P_2      VWC_3   T_3                 VWC_4
                  TS        RN     m^3/m^3 Deg F         unitless m^3/m^3 Deg F         unitless m^3/m^3 Deg F           m^3/m^3
                  NaN       NaN    Smp     Smp           Smp      Smp     Smp           Smp      Smp     Smp                   Smp
                  2017-04-03 12:15:00 0      NAN     14241.87      0        NAN     14241.87      0        NAN     14241.87        NAN
                  2017-04-03 12:30:00 1      NAN     9.635455E+007 0        NAN     9.635455E+007 0        NAN     9.635455E+007   NAN
```

Figure 1 – Importing the Bay View Dataset into a dataframe using Python Pandas

I also worked on creating time series graphs. Time series graphs are created to plot aggregated values of each column of the dataset which displays trends of intervals. As shown below in Fig 2-3, I created time series graphs of variables such as relative humidity, air temperature Fahrenheit, total rainfall, and soil moisture to analyze the trends in the data and to predict the possible outcomes of the data. To make plots, it is necessary to import the Python library matplotlib since it is used

for creating plots and graphs. As shown below in Fig 2-3, it is evident that relative humidity increased and decreased over time similarly in Fig 4-5.

Time Series Graphs

```
In [17]: import matplotlib.pyplot as plt
```

```
In [18]: df.BAY_RH.plot(figsize=(20,5), title= "BayView Weather Station (BAY)")
         plt.xlabel("TIMESTAMP")
         plt.ylabel("Relative Humidity")
         plt.show()
         # xlabel..   Timestamp
         #  y label .. relative humidity
         # decide label tick
```

Figure 2 – Creating time series graphs using matplotlib to create plots of the relative humidity column

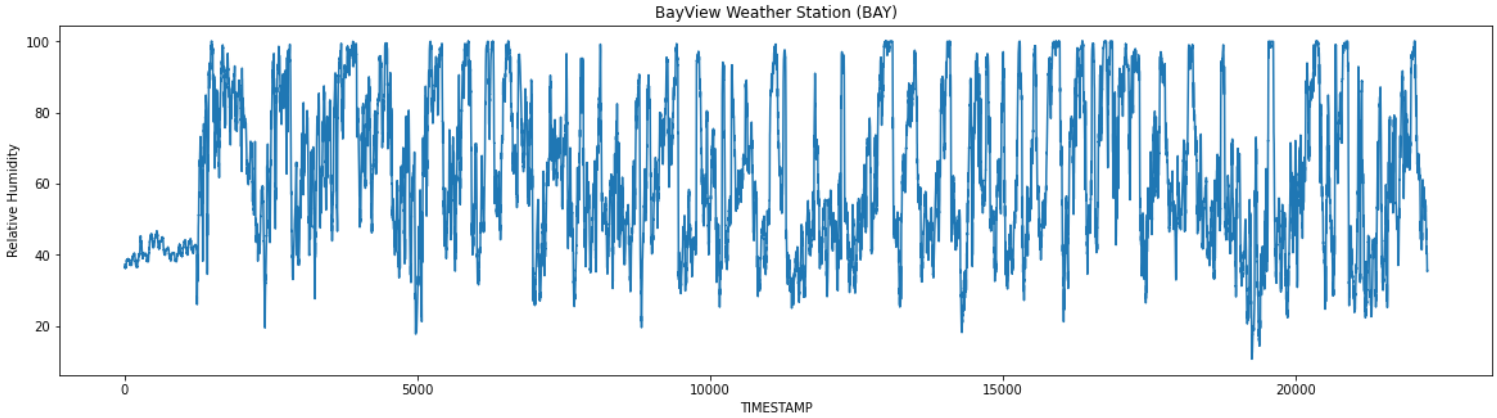

Figure 2 – Result of the time series code shown above in Fig 2 displaying trends of relative humidity in the Bay View site

```
In [21]: df.BAY_VWC1.plot(figsize=(20,5), title= "BAY_VWC1")
         plt.xlabel("TIMESTAMP")
         plt.ylabel("Soil Moisture")
         plt.show()
```

Figure 4 – Code for creating time series graph plots using matplotlib to visualize data

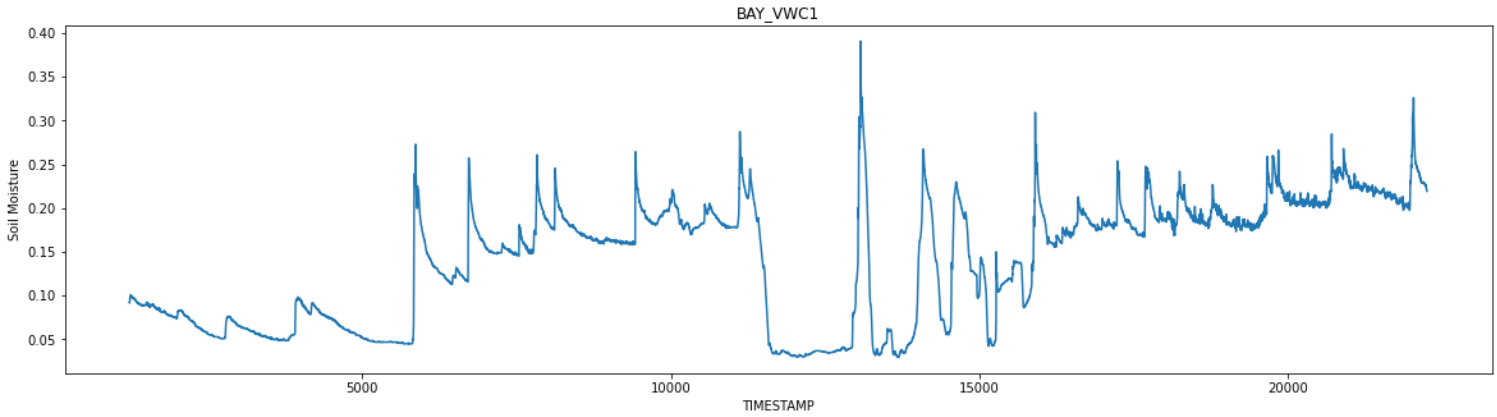

Figure 5 - Time series graph of Bay View's soil moisture column