**Energy Analytics Project**

**Background**

Natural gas is delivered to customers in the Northeast (Connecticut, Massachusetts, Vermont, New Hampshire, Rhode Island, Maine) on the Algonquin (AGT) pipeline. AGT publishes the daily natural gas demand by category – Residential & Commercial and demand from natural gas fired power plants. The pipeline’s “gas day” runs from 10AM EPT (Eastern Prevailing Time) to 10AM EPT the following day. As an example, the reported gas demand for January 1, 2023 would be the demand from Jan 1, 2023 10:00 AM EPT to Jan 2, 2023 10:00 AM EPT. One of the key factors that determines the

demand for natural gas each day is the region’s temperature. Boston is the largest city in the region thus its temperature has a large impact on AGT gas demand each day.

**Python Assessment**

- Complete the below tasks in Python version 3.7 or higher. Please do not use any custom packages that cannot be installed from PIP or conda. Provide **one** .py file with all of your code.

1. Scrape historical hourly temperature data for Boston from NOAA ISD Lite Dataset.

The data is located at the below link. For this exercise, you will be pulling Boston hourly temperature data for 2019-2023. Note: you will need to replace the year in the below link to download the remaining years.

<https://www.ncei.noaa.gov/pub/data/noaa/isd-lite/2019/725090-14739-2019.gz>

This is a fixed width file. The column names and units are provided below:

|  |  |
| --- | --- |
| **Column Name** | **Units** |
| Year | UTC Time |
| Month | UTC Time |
| Day | UTC Time |
| Hour | UTC Time |
| Air temperature\*\* | degrees Celsius \* 10 |
| Dew point temperature\*\* | degrees Celsius \* 10 |
| Sea level pressure | hectopascals |
| Wind direction | angular degrees |
| Wind speed\*\* | meters per second \* 10 |
| Total cloud cover | coded, see format documentation |
| One-hour accumulated precipitation | millimeters |
| Six-hour accumulated precipitation | millimeters |

\*\*NOAA scaled the units by 10. To get back to Celsius you will need to divide this by 10.

Further documentation can be found at the below links, but is not necessary for completion of this assignment <https://www1.ncdc.noaa.gov/pub/data/noaa/isd-lite/isd-lite-technical-document.pdf>

<https://www.ncei.noaa.gov/pub/data/noaa/isd-format-document.pdf>

1. Convert the air temperature from Celsius to Fahrenheit
2. Once you have collected and converted the data, please answer the following questions:
   * **How many missing hourly temperature observations are in the dataset? (Provide the dates and hours missing if there are any and interpolate any missing values)**
   * **What is the mean air temperature for all observations in July 2021?**
3. Generate Gas Day temperatures

The gas day starts at 10AM (EPT) Eastern Prevailing Time and runs to 10AM EPT the following day. The gas day temperature for that day is the average of the 24 hourly air temperature observations. The gas day temperatures should be reported in Eastern Time and in Fahrenheit. You will be aggregating the 24 hourly observations into one average temperature for that gas day.

* + **What is the Boston gas day temperature for July 4th, 2020?**

1. Collect Demand Data from provided Excel file (agt.xlsx)

This file contains daily demand data for Residential/Commercial and Power Plant users.

1. Build a model to forecast Total AGT demand (Res/Comm + Power Plant)
2. The daily gas day temperature data should be used in your model
3. You may use other variables from the NOAA weather dataset or variables that you find important
4. Your model should forecast AGT demand for 2022

Provide the following:

* **Mean Squared Error (MSE) for your forecast vs actual AGT total demand for 2022**
* **A plot of your daily 2022 forecast vs actual AGT total demand**
* **A plot of your daily 2022 forecast vs the gas day temperature**

1. Provide a qualitative assessment of your work:
   * **What assumptions did you make and why?**
   * **Was there anything interesting about the dataset?**
   * **Why did you choose to construct your model the way you did?**
   * **What other data or information do you think would improve your forecast?**

**What to submit**

* Please summarize your findings in a word document named Lastname\_Firstname\_ Assessment.doc.
  + This should include answers to all of the above **bolded** questions notated with a check
* Include all of your python code in one file named Lastname\_Firstname\_ Assessment.txt. Please ensure you rename the **.py** to **.txt** or else if will get caught in our email filters. We should be able to install any required packages and run your code without any other changes.
* Send both files back to the HR Representative that originally sent you the test by the given deadline.