

Problem:

We are tasked with predicting the influent flow rates into a fictitious water treatment facility in Palo Alto, CA. Water flows into this plant from many sources. The plant doesn't have any control over the amount of water coming in. Being able to predict flow rates will allow the plant operators to better control their plant leading up to and during high flow events.

Data Description:

The manager has provided us with their influent flows from early 2013 through 2016 at an hourly interval. This data can be found in the file '*influent_flow.csv*'. The column labeled 'INF_FLOW' is the influent flow rate in million gallons per day (MGD).

Since we suspect that the weather has an influence on the influent flow, we have pulled historical weather data from a nearby weather station from 2013 through early 2018. The weather data can be found in the file '*weather.csv*'. We were not provided with the description or the units for the columns. For your problem, you can assume that weather forecasts and historical weather data are equivalent (i.e. you can assume weather forecasts have zero error).

We also know from past experience that nearby creek levels tend to be good predictors of influent flow. We believe that nearby gage level of the San Francisquito Creek (USGS site 11164500) would be a good feature for the model. This data can be found in the file '*usgs_san_francisquito.txt*'. For more information on the USGS data format: <https://help.waterdata.usgs.gov/faq/about-tab-delimited-output>.

Tasks:

1. Explore the data received using Python. Create any tables, plots, or images you deem valuable to understand the data.
2. Create a slide deck to be presented to the client. Some possible things to include (but feel free to add/subtract as you wish):
 - a. Can the flow be predicted?
 - i. If so, how well? Do you think it would be sufficient?
 - b. What are the strongest predictors of the flow?
 - i. Any intuition as to why?
 - c. If a product is developed from this, how would you present the results to the client?
 - d. Any interesting insights?
 - e. What would you suggest as next steps?

Expectations:

- Clearly state any assumptions that are made.
- Data analysis and any model development should be run from Python. Data parsing, cleaning, etc. can be done using any method you prefer.
- We prefer Jupyter notebooks for your coding scripts are fine as well.