# Springboard--DSC Program

# Capstone Project 2 - Project Proposal

## Acea Smart Water Analytics

By: Ellen A. Savoye

February 06, 2021

**Business problem:**
Water is an important part of everyday life. It's easy to take the, seemingly, endless supply of water for granted. However, the water supply isn't endless. To sustain our usage, water supply companies need to be able to forecast the amount of water in a given body of water under their purview to handle user consumption. These bodies of water could be a water spring, a lake, a river, or an aquifer. Each is affected by the changing of the seasons. Typically, bodies of water are replenished in the cooler seasons (autumn, winter) while warmer months (spring, summer) have a decrease in water level. Regardless of the season, it is important to be able to predict water availability, in terms of level or flow. Can we build a model to predict water availability?

**Client:**
The client is an Italian multiutility operator known as *Acea Group.* The company manages and develops water and electricity networks and environmental services. *Acea* is the foremost Italian operator in the water services sector supplying 9 million inhabitants in Lazio, Tuscany, Umbria, Molise, and Campania[1].

**Data:**
The data used for this project is provided by *Acea.* Each dataset, a CSV file, is independent and not linked to each other. I will be looking at the datasets for aquifers and water springs. There are 4 aquifer datasets and 3 water spring datasets. The only similarities between the 7 tables is 2 variables: date and rainfall. The date, in most cases, is daily. However, date coverage prior to 2002 is not guaranteed to be daily. The target for aquifers and water springs is depth to groundwater and flow rate, respectively.

**Problem Approach:**
The 7 CSV files comprise my data. Given that they are independent from one another, all of them will be explored and analysed separately. During exploratory data analysis (EDA), the

[1] https://www.kaggle.com/c/acea-water-prediction/overview
[2] https://github.com/facebook/prophet

datasets will be analyzed to determine if there are data issues such as missing data, duplicate data, and outliers. After EDA and further statistical inferences are examined, a baseline model can be built using either regression, Pyramid auto ARIMA, or Prophet[2]. The baseline may be improved using whichever 2 models are not selected for the baseline. We will analyze the performances of the different models and suggest how to use these models in the context of the business problem.

**Deliverables:**
The deliverables will include all code developed with each step contained in it's own Jupyter Notebook, a written final report, and a written presentation slide deck.

[1] https://www.kaggle.com/c/acea-water-prediction/overview
[2] https://github.com/facebook/prophet