Proposal for Final Project

UCB Data Analytics Bootcamp

REPORT



**Prepared by:**

**Shandiz**

**Andrew DeVries**

**Dipesh Manandhar**

**Josh Bode**

**October 14, 2018**

Hourly Electric Forecasts for Electric Grids in the U.S.

# 

Table of Contents

[Topic and Rationale 2](#_Toc527284112)

[Datasets 2](#_Toc527284113)

[Screen shots of the data 3](#_Toc527284114)

[Map of electric systems in US 3](#_Toc527284115)

[Historical actual data 3](#_Toc527284116)

[Historical Day Ahead Forecast Data (Our competition) 4](#_Toc527284117)

[Weather station characteristics and locations 4](#_Toc527284118)

[Project workflow 5](#_Toc527284119)

[Scaling Strategy 5](#_Toc527284120)

[Team Member Roles 6](#_Toc527284121)

# Topic and Rationale

The objective for this project is to take multiple years of historical hourly demand data for multiple electric system in the US, analyze the historical patterns, and develop predictive models to forecast the hourly demand for the next and week.

The electric market is a $350 billion business per year in the U.S. alone. The hourly forecasts are critical because they drive infrastructure investments and market prices spike when demand is high and available capacity is low. Currently, each electric system (aka balancing authority) develop its own forecast. Both the historical actual and forecast hourly data is publicly available via and API.

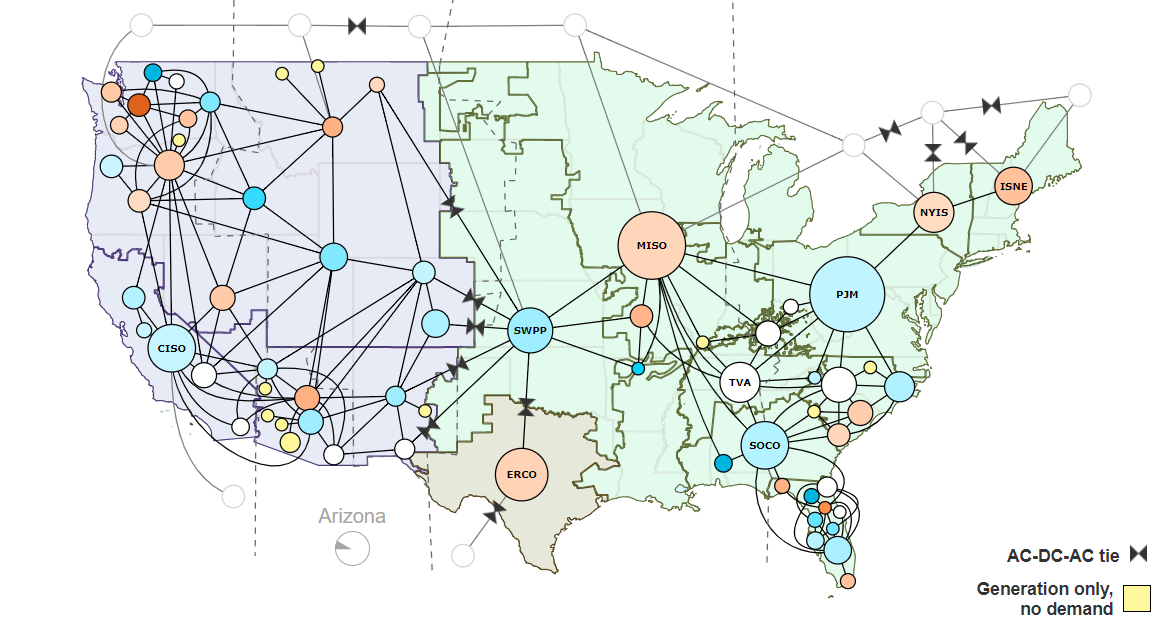
Electric demand is highly sensitive to weather, hour of day (occupancy schedule), day of week, and season. The weather varies by location but so does the weather sensitivity. Weather sensitivity is driven by the penetration of air conditioning, electric heating, and the mix of industrial/commercial facilities. As a result, the model will be different for each of the locations.

# Datasets

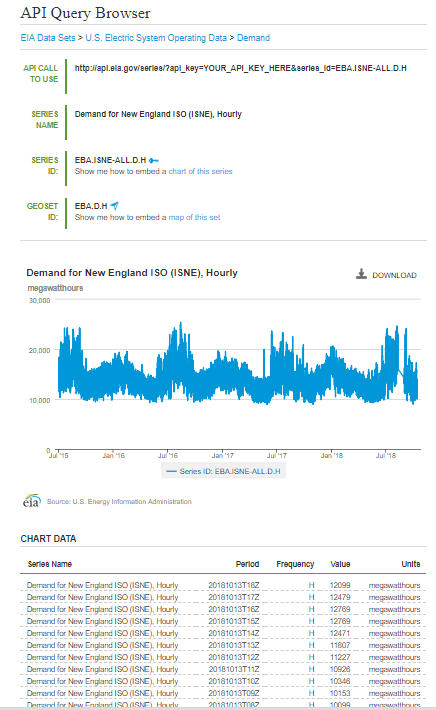
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data | Description | | Link | |
| Map of electric systems | | Lat/long for all the major US electric grids | | <https://www.eia.gov/realtime_grid/#/status?end=20181013T13> |
| Historical actual data by operating system | | * System * Date * Hour * MW (demand) | | <https://www.eia.gov/opendata/qb.php?category=2122628> |
| Historical forecast data by operating system | | * System * Date * Hour * MW (demand) | | <https://www.eia.gov/opendata/qb.php?category=2122627> |
| Weather station locations and characteristics | | * Name and IDs * Location, including lat/lon * Dates for which weather station data is available | | https://www1.ncdc.noaa.gov/pub/data/noaa/isd-history.csv |
| Historical weather data | | * Station * Date and hour * Temperature (dry and wet bulb) * Humidity * Precipitation (rain/snow) * Wind direction and speed | | <https://www1.ncdc.noaa.gov/pub/data> |
| Forecast weather data | | * Nearest station * Date and hour * Forecasted hourly temperature   (7 days of forecast data) | |  |

# Screen shots of the data

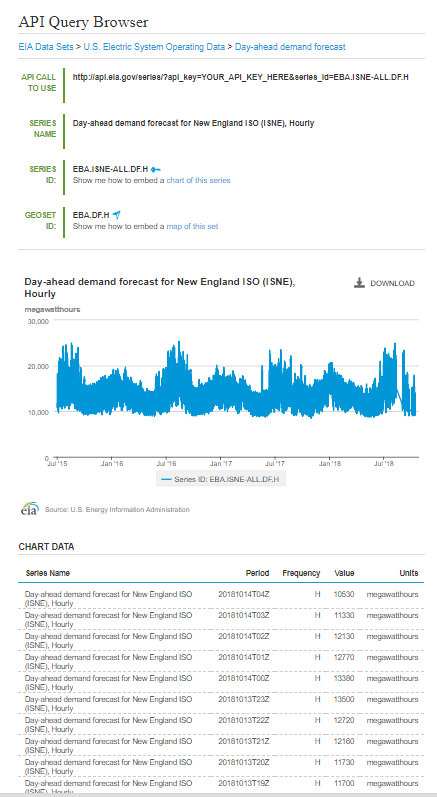
### Map of electric systems in US



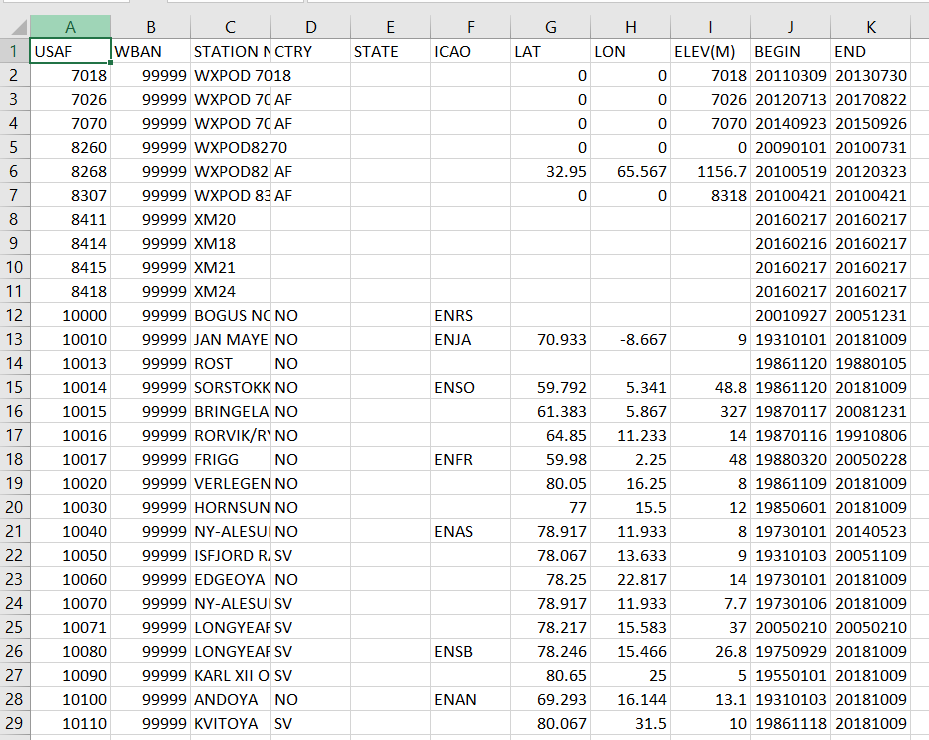
### Historical actual data



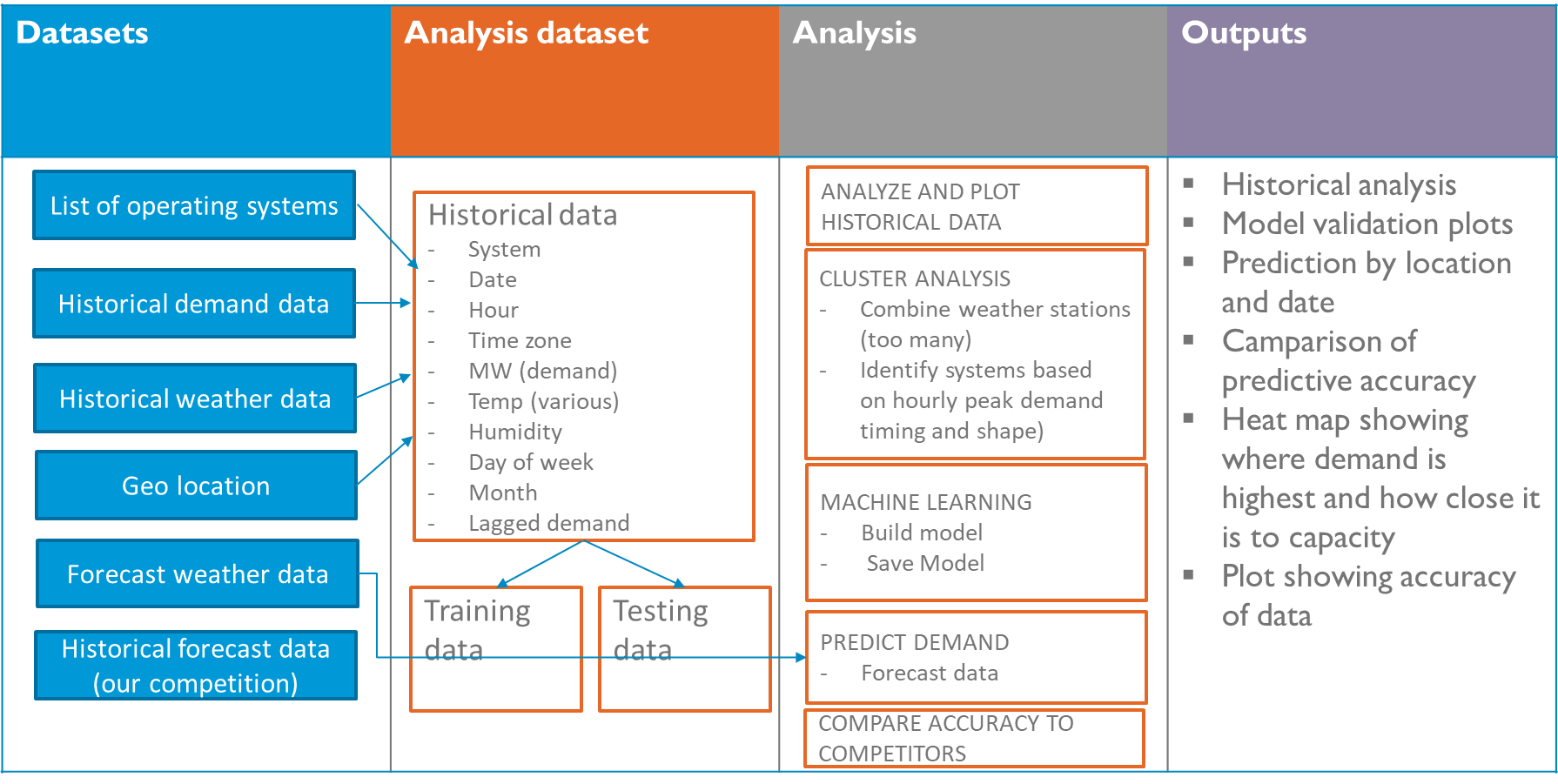
### Historical Day Ahead Forecast Data (Our competition)



### Weather station characteristics and locations



# Project workflow



# Scaling Strategy

* Phase 1 – Build out one location using static data
* Phase 2 – Run analysis for multiple systems
* Phase 3 – Automate pulling of updated historical data, forecast weather, and models
* Phase 4 – Convert into an app that automatically updates periodically

Initial visualization will be in Python and Tableau. If time permits, build out site.

# Team Member Roles

|  |  |  |  |
| --- | --- | --- | --- |
| Task | | Key Steps | Team Member |
| Data engineering | * Download all historical demand * Download historical weather data – 100 station in US for 5 years (in locations) * Download the historical forecast demand data * Standardize the time zone and daylight saving time issues * Produce feature for different weather metrics, day of week, etc. | | * Dipesh * Shandiz * Andrew * Dipesh/Shandiz/Andrew * Josh |
| Data Analysis | * Historical data analysis/visualization * Cluster analysis for weather * Cluster analysis of peak day shape * Machine learning model * Prediction using machine learning model * Accuracy metrics for our models versus competitor models | | * Andrew * Shandiz * Dispesh * Josh/Dispesh/Shandiz/Andrew * Josh/Dispesh/Shandiz/Andrew * Josh |
| Outputs | * Historical analysis * Model validation plots (predicted v actual) * Predicted demand by date hour and location * Maps showing where demand is highest and close it is to capacity * Plots comparing the accuracy our predictions to the competition | | * Andrew * Josh * Team effort * Andrew/Josh * Josh/Dispesh |
| Scaling and Automation | * Fully automate all aspects of data pulls * Schedule data extraction using the API’s * Schedule update to the machine learning models * Deploy | |  |