U.S. HOURLY ELECTRICITY FORECASTS

DIPESH MANANDHAR,
SHANDIZ MONTAZERI
ANDREW DEVRIES
JOSH BODE

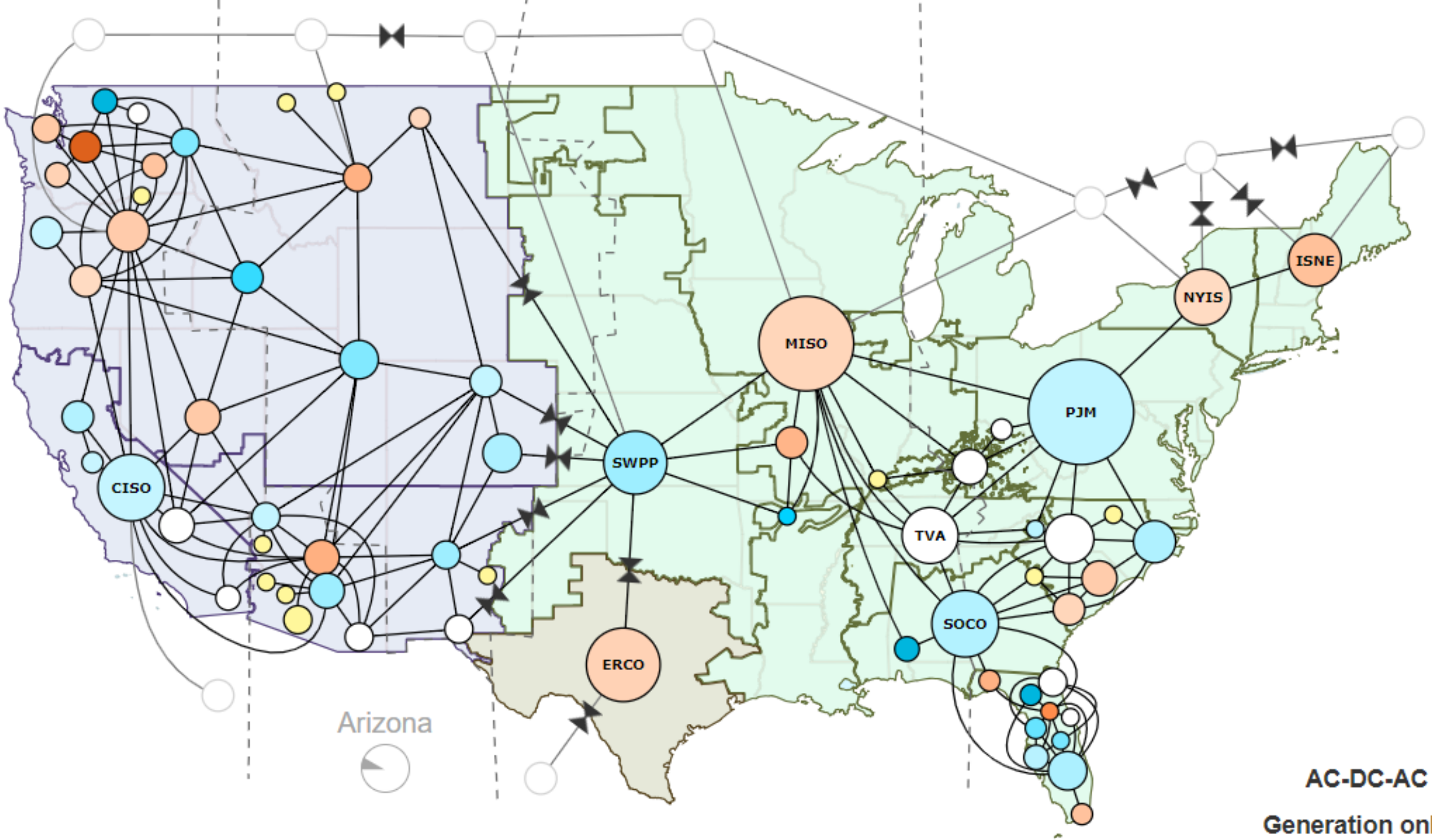
STUDY OBJECTIVES

WHAT?

- Pull multiple years of historical hourly demand data for each electric systems in the US
- Analyze the historical patterns
- Develop predictive models to forecast the hourly demand for the next and week

WHY?

- The electric market is a $350 billion business per year in the U.S. alone.
- Peak demand drives infrastructure investments
- Peak demand lead to market price spikes
- Currently, each electric system develop its own forecasts.


CISO
SWPP
MISO
PJM
NYIS
ISNE
TVA
SOCO
ERCO
Arizona
AC-DC-AC tie
Generation only,
no demand

DATA SOURCES

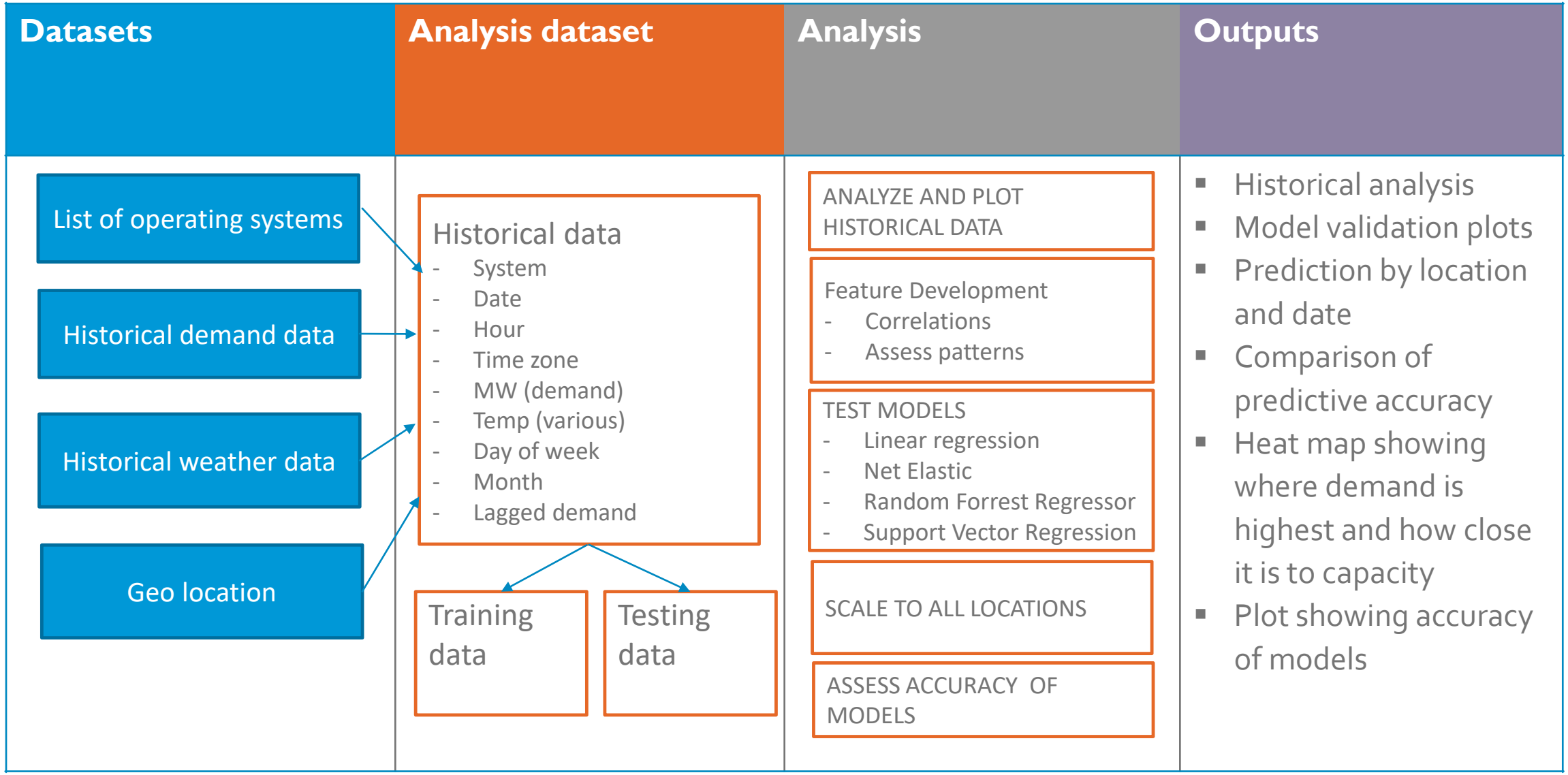
Data	Description	Link
Map of electric systems	Lat/long for all the major US electric grids	https://www.eia.gov/realtime_grid/#/status?end=20181013T13
Historical hourly actual data by operating system	▪ System ▪ Date ▪ Hour ▪ MW (demand)	https://www.eia.gov/opendata/qb.php?category=2122628
Historical hourly 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	https://www1.ncdc.noaa.gov/pub/data

Electric data

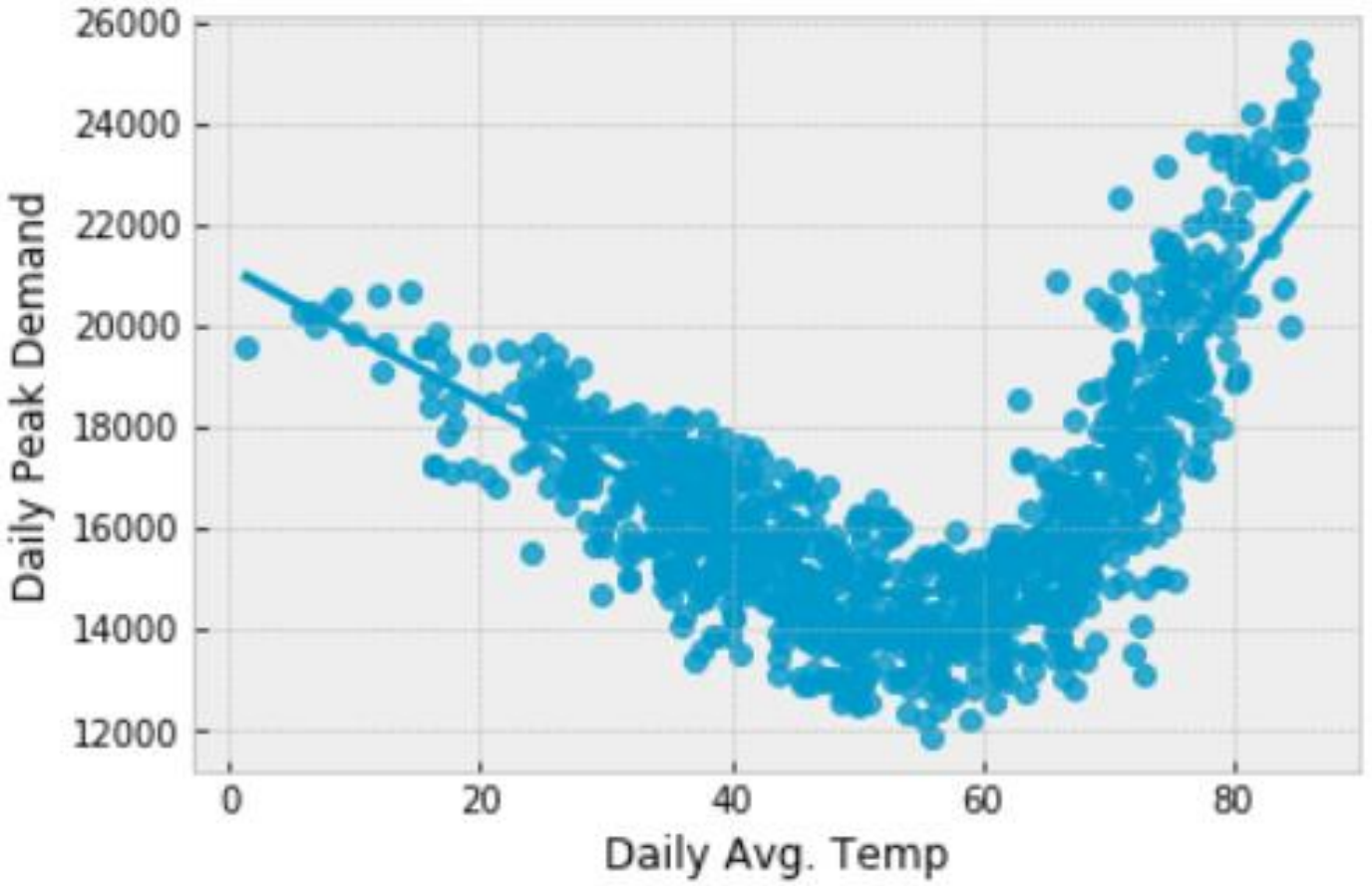
- 51 systems
- 3 year of data each
- 1.2 M observations

Weather Data

- 261 weather stations
- 3 years of data
- 6M+ observations

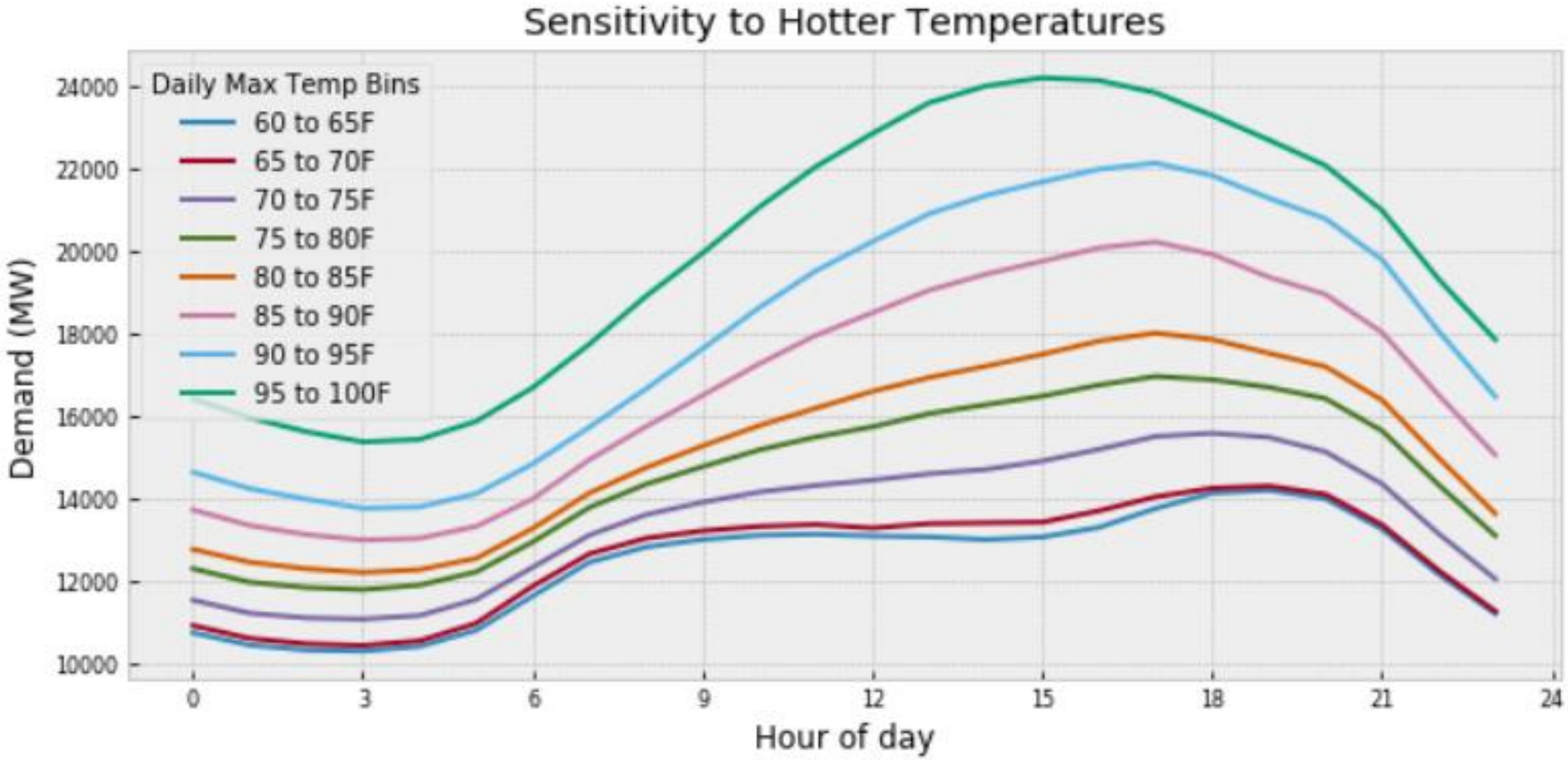

Datasets
Analysis dataset
Analysis
Outputs
List of operating systems
Historical demand data
Historical weather data
Geo location
Historical data
- System
- Date
- Hour
- Time zone
- MW (demand)
- Temp (various)
- Day of week
- Month
- Lagged demand
Training data
Testing data
ANALYZE AND PLOT HISTORICAL DATA
Feature Development
- Correlations
- Assess patterns
TEST MODELS
- Linear regression
- Net Elastic
- Random Forrest Regressor
- Support Vector Regression
SCALE TO ALL LOCATIONS
ASSESS ACCURACY OF MODELS
Historical analysis
Model validation plots
Prediction by location and date
Comparison of predictive accuracy
Heat map showing where demand is highest and how close it is to capacity
Plot showing accuracy of models

DATA EXPLORATION - DAILY PEAK DEMAND VARIES WITH WEATHER

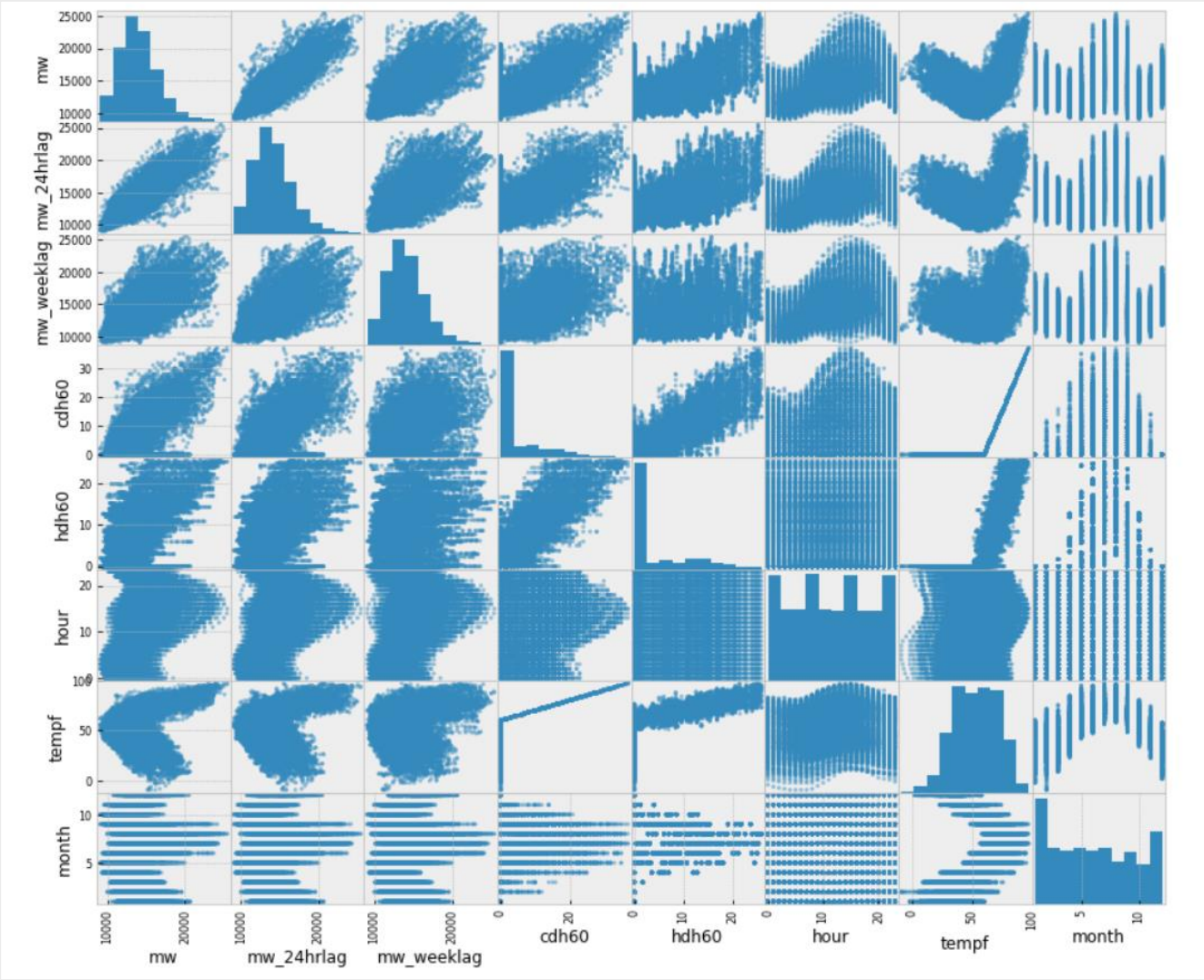


- Non-linear, asymmetric pattern
- When it gets very hot, we use more electricity (AC)
- When it get very cold, we use more electricity (electric heating)
- The varies slightly by location because of
  - Differences in weather
  - Difference in air conditioning and electric penetration
  - Differences in the mix of residential, commercial and industrial customers

DATA EXPLORATION – DEMAND ALSO VARIED BY HOUR DUE TO OCCUPANCY PATTERNS



FEATURE DEVELOPMENT - CORRELATIONS



- Assess correlation with different of variables
- Plot out the relationship due to non-linearity
- Include some lags
- Split temperature at point where no one is using heating /cooling
- Decision separately model hours and weekday/weekend days

MODELS

- Modeling techniques tested
  - Linear regression with polynomials
  - Net Elastic Regression - identifies the best features and drops out irrelevant ones
  - Random Forest Regression – similar to random forest trees
  - Support Vector Regression - Takes way too long and did not yield improvements
- Initial testing was done for one location
- Wrote function that automated the process

- 7,344 Models
  - 51 systems
  - 3 modeling techniques
  - 48 model per systems (by hour and weekday/weekend)

LETS LOOK AT SOME RESULTS!