# Hourly Electric Load Forecast for PJMI - Pittsburgh, USA - For 7 Days

Modeling and forecasting hourly electric load by

multiple linear regression with interactions

Sept 1st, 2021

By -

KF009 - Abhishek Janjalkar

KF008 - Mohammed Hasan Khan

KF020 - Kaustubh Sawant

# Hourly Electric Load Forecast for PJMI - Pittsburgh, USA - For 7 Days

## Abstract

* Short term electric load modeling and forecasting.
* Demand Side Management (DSM) with beyond when and how much demand will be and - why is electricity being consumed?
* Factors driving the fluctuations of the electric load at a particular time period.
* Model hourly demand and investigate causality of the consumption of electric energy
* Terms - Load forecasting, load management, load modeling, multiple linear regression
* Both the datasets combined on common datetime attribute -Resulting dataset has 45250 rows, containing hourly data for years 2012-2017 of which data for the years 2012-2016 is used for training and data for the year 2017 is used for forecasting

### Datasets used -

* [Hourly Electric Load Data](https://www.kaggle.com/robikscube/hourly-energy-consumption)
* [Hourly Temperature Data](https://www.kaggle.com/selfishgene/historical-hourly-weather-data)

## Hourly Electric Load Data

PJM Interconnection LLC (PJM) is a regional transmission organization (RTO) in the United States. It is part of the Eastern Interconnection grid operating an electric transmission system serving all or parts of Delaware, Illinois, Indiana, Kentucky, Maryland, Michigan, New Jersey, North Carolina, Ohio, Pennsylvania, Tennessee, Virginia, West Virginia, and the District of Columbia.

The hourly power consumption data comes from PJM's website and are in megawatts (MW).

The regions have changed over the years so data may only appear for certain dates per region. Hourly Temperature Data The dataset contains ~5 years of high temporal resolution (hourly measurements) data of various weather attributes, such as temperature, humidity, air pressure, etc.

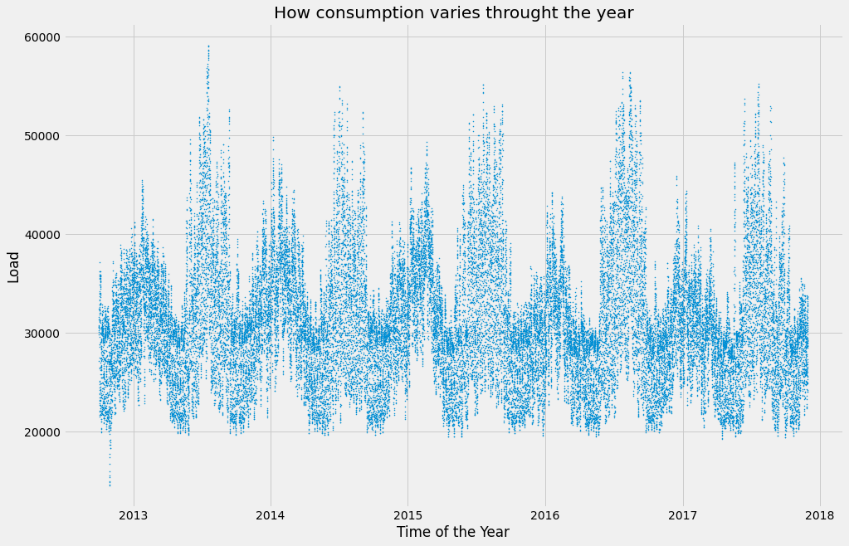
This data is available for 30 US and Canadian Cities, as well as 6 Israeli cities. We've organized the data according to a common time axis for easy use. Each attribute has it's own file and is organized such that the rows are the time axis (it's the same time axis for all files), and the columns are the different cities (it's the same city ordering for all files as well). Additionally, for each city we also have the country, latitude and longitude information in a separate file.

## Multiple Linear Regression in the Context

* Used to model relationship between several independent and a dependent variable as a linear function
* Y = b0 + b1x1 + b2x2 + e where Y is dependent variable, x1, x2 are independent variables, b are parameters to estimate and e is error term
* Coefficients of independent variables represent how Y will change when that specific independent variable changes while all other variables are held constant

## Looking at the data

### How the load changes throught the year

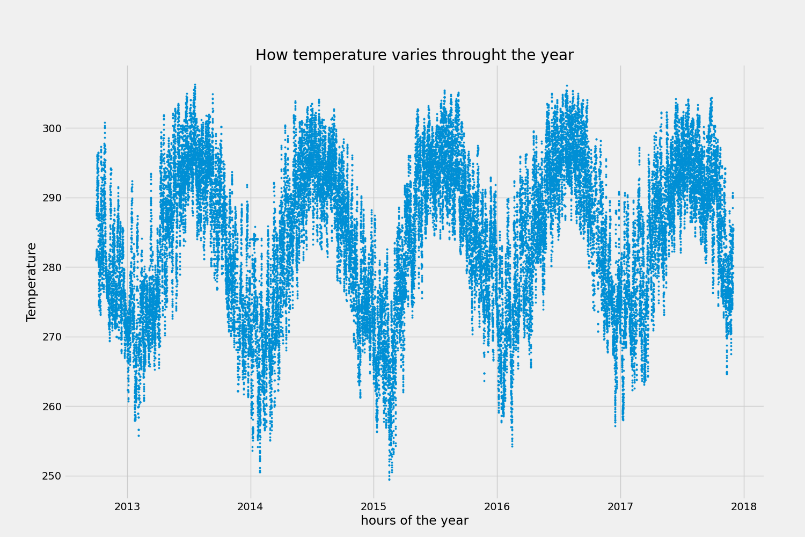
Plot of the Load for the Years 2012-2017

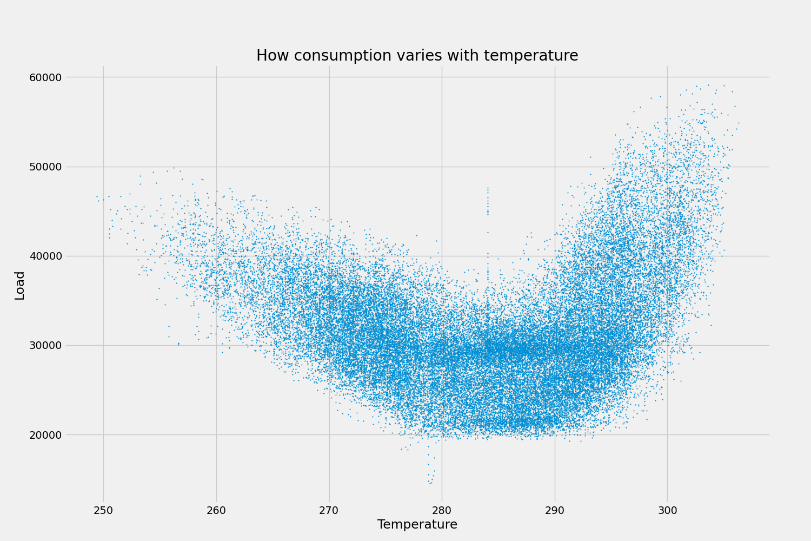
By looking at the above plot, following can be observed - overall increasing trend year by year, possible reasons are - human activities/increase of temperature.

### How the temperature changes throught the year

Plot for How the temperature changes throught the year - Pattern of seasonality is visible. - Winter peak load during valley temperature period in winter and summer peak load during summer peak temperature period in summer - Summer peak higher than winter peak - Less need of AC in Fall and Spring

### How the temperature changes throught the year

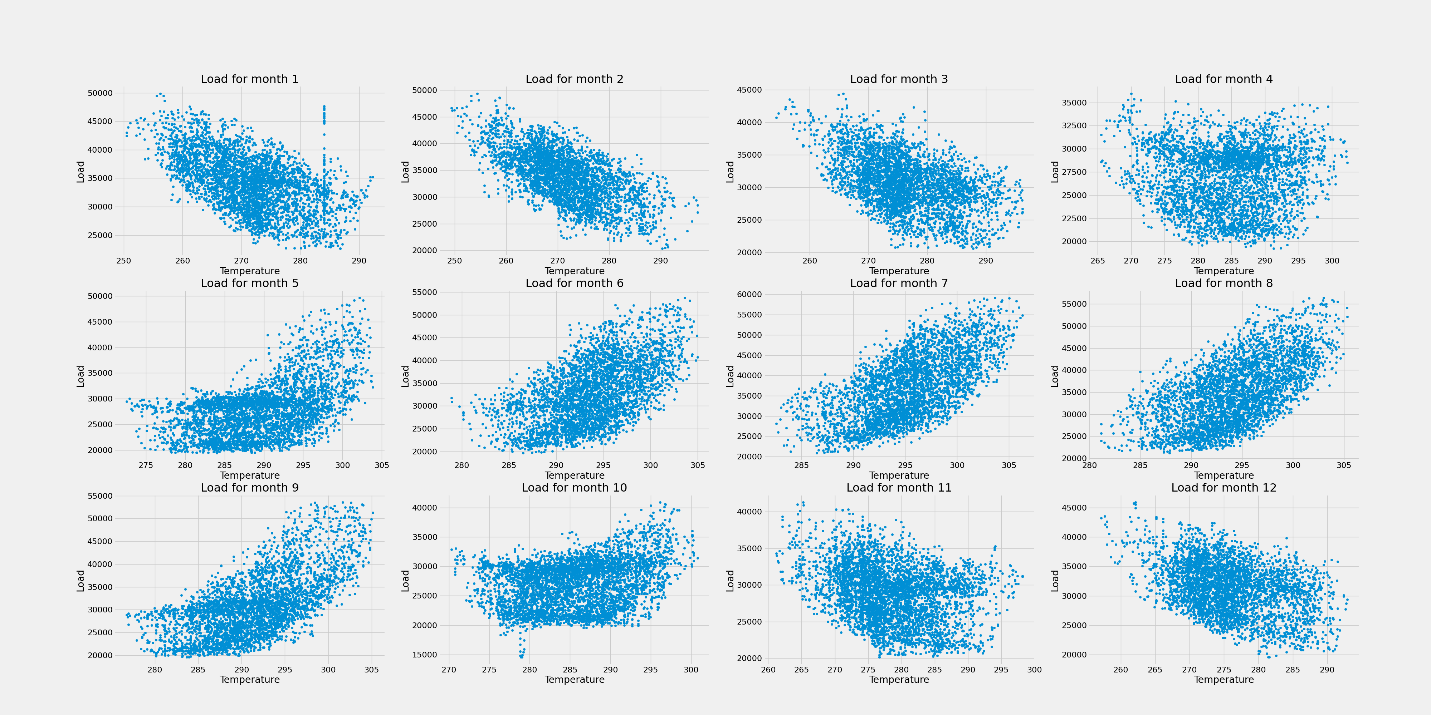




With the above plot following points can be observed - Upper boundary of load-temperature plot shows nonsymmetrical V-shape while lower boundary shows nonsymmetrical U-shape - Piecewise nonlinear function is preferred with dataset cut-off at 284K introducing following dummy and quantitative variables - TMPID = 1 for all temperature < 285 - TMPID = 2 for all temperature >= 285 - TMP = temperature - TMP2 = square of temperature

### Seasonality by months throught the year

Plot for Seasonality by months throught the year

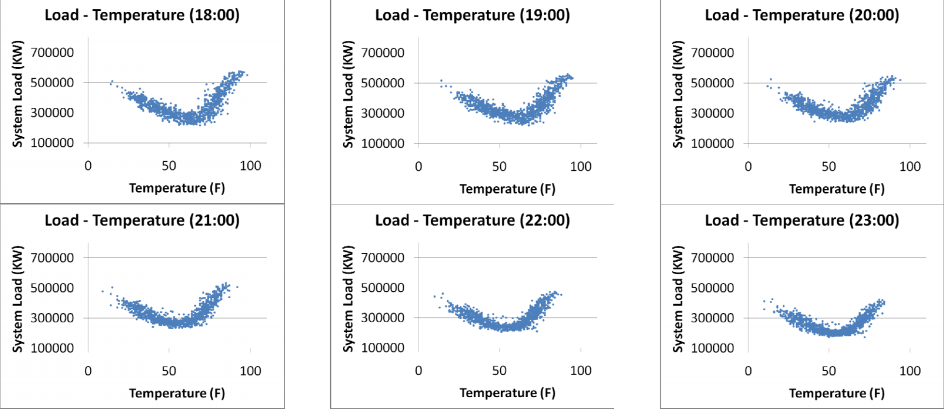


#### With above plot Following points can be observed

* Higher levels of load in March than that in April despite similar temperature temperature - explained by different human activities
* Model relationship between load and temperature for each month individually
* Dummy variable - Month
* Interaction Terms - TMP*TMPID*Month
* TMP2*TMPID*Month

### Seasonality by Hour throught the year

#### Following points can be observed

* Due to sunrise and sunset, time and temperature are correlated
* 24 plots combined construct previous load-temperature plot suggesting the consideration of piecewise quadratic function
* Vertical levels of plot appear to be different
* Horizontal positions of plots appear to be different by hour suggesting consideration of different functions to model load-temperature for - different hours
* Heat build-up effect
* Hence, hour of the day as dummy variable (Hour), interaction terms - TMP*TMPID*Hour and TMP2*TMPID*Hour
* Heat build effect interaction term - DTMP*TMPID*Hour where DTMP = current temperature - previous hour temperature
* No statistically significant differences among weather over days in a week, hence load is affected by human activities.
* Different human activities on Weekdays and Weekends hence dummy variable D1 - weekdays, saturdays, sundays, D2 - mondays, other weekdays, saturdays, sundays

### Training

For training, we've used XGBoost library with following configuration - n\_estimators = 1000 - early\_stopping\_rounds=50

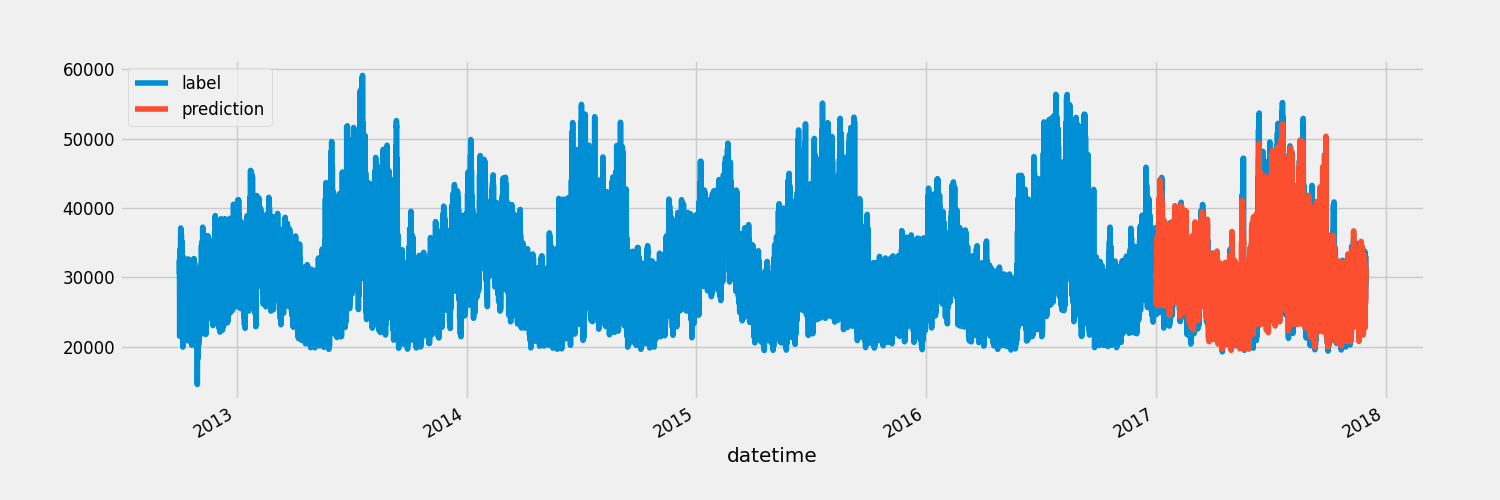
## Importance of feature-variables

Importance of feature-variables

## Importance of feature-variables

## Predictions for year 2017

Predictions for year 2017



Performance of the Model

|  |  |
| --- | --- |
| Metric | Value |
| Mean Squared Error | 5722792.126897162 |
| Mean Absolute Error | 1802.4898734634994 |
| Mean Absolute Percentage Error | 5.856947648551624 |

The trained model has achieved accuracy of about 95%.

## Load Forecasting Using weather.gov api for hourly temperature forecast

Load Forecasting Using weather.gov api for hourly temperature forecast

## Integrating the resulting model with the weather.gov api for hourly tmperature forecastLoad Forecasting Using weather.gov api for hourly temperature forecastIntegrating the resulting model with the weather.gov api for hourly tmperature forecast

Integrating the resulting model with the weather.gov api for hourly tmperature forecast

## Tools and Technologies

* Languages used - HTML, Javascript, Python, Markdown
* Frameworks Used - Flask, ChartJS
* Weather.gov API - [Hourly Weather Forecast](https://www.weather.gov/documentation/services-web-api#/default/get_gridpoints__wfo___x___y__forecast_hourly)

## Specification

* This project plots attempts to predict the Electric load on PJMI grid in pennsylvania for next 7 days.
* It fetches hourly forecasted weather info for 7 days using following api
* It calls for the weather api on the page refresh and attempts to predict the load for next week
* Plots the resulting prediction using ChartJS Library

## Future Scope and Improvements

* Proposed regression used to generate short-term hourly demand forecast for next week
* Good interpretability of the behavior electricity consumption which helps understand system load profile when performing DSM and T&D
* Load forecast with satisfying accuracy provides good reference for energy purchasing purpose
* Considering cross effects among explanatory variables, performance improved with MAPE of 5.857%
* Model can be improved by considering holidays and enhancing quality of data

## References

* T. Hong, M. Gui, M. E. Baran and H. L. Willis, "Modeling and forecasting hourly electric load by multiple linear regression with interactions," IEEE PES General Meeting, 2010, pp. 1-8, doi: 10.1109/PES.2010.5589959.
* [Stack Overflow](https://stackoverflow.com/)
* [Hourly Electric Load Data](https://www.kaggle.com/robikscube/hourly-energy-consumption)
* [Hourly Temperature Data](https://www.kaggle.com/selfishgene/historical-hourly-weather-data)