9]:

**Data Dictionary**

* **Congestion** - is when the lowest-priced energy is prevented from flowing freely to a specific area on the grid because heavy electricity use is causing parts of the grid to operate near their limits. <https://learn.pjm.com/-/media/about-pjm/newsroom/fact-sheets/congestion-fact-sheet.ashx>

**DEA**

* Most frequent voltage amounts
* LMP is autoregressive, a time series is said to be AR when present value of the time series can be obtained using previous values of the same time series i.e the present value is weighted average of its past values.
  + But the influence of weighted averages doesn’t reach very far back, only 2.
* DEA suggests that time-series is additive
  + **y(t) = Level + Trend + Seasonality + Noise**
  + Seasonality at the level of a week. LMP peaked at Thursday/Friday would plummet through weekend, surge at beginning of week from weekend avgs, and steadily move towards the Thursday/Friday peak
    - The frequency (width) and amplitude (height) of cycles were linear
  + Trend is a straight line (avg at beginning avg at end)
* Diceky Fuller p-value shows that p-value is < 0.05 so i can reject null hypothesis that time-series is non-stationary. P-value = 0

A given time series is thought to consist of three systematic components including level, trend, seasonality, and one non-systematic component called noise.

These components are defined as follows:

* Level: The average value in the series.
* Trend: The increasing or decreasing value in the series.
* Seasonality: The repeating short-term cycle in the series.
* Noise: The random variation in the series.

**Feature Engineering**

* Used Standard Scaler on all numerical columns.
  + Many different unit variances: mw, currency amounts, ect.

**Modeling**

* Baseline Persistence Model was w/ lagged datatset
  + Score was MSE: 0.642

**ARIMA Model**

* PDQ
  + 10,0,10 returned lowest MSE
  + Determined by for loop
* Split of hourly median dataframe
  + Train
    - 9 weeks (July 1st – Sept. 8th)
    - 1656 rows
    - MSE Train ARIMA(10,0,10) ... 186.61
    - R-sq Train ARIMA(10,0,10) ... 0.4042
  + Test
    - 3 weeks (Rest of September)
    - 529 rows
    - MSE Test ARIMA(10,0,10) ... 202.42
    - R-sq Test ARIMA(10,0,10) ... -0.0508