Uncovering Pattern in Electrical Operation Data

Jenny Yu

Introduction

The Galvanize Datathon is an “Open” datathon where the participants have the flexibility to formulate their own questions and solutions.

New York Independent System Operator(ISO) is an entity that manages New York’s electrical grid and wholesale energy market. It provides the public dataset for pricing, load and other operational time series data.

For the purpose of the Datathon, I analyzed the price and load data from January - July 2021, to explore trends that can help optimize/lower electricity cost for the user or buyer.

Data Analysis

- Factors affecting the price (LBMP): Load/demand, outage, transmission congestion.
- The time series data for Price and Load are collected in a 5-minute interval, so they can be merged based on `Time Stamp` and `Name` of the zone.
- There's a total of 15 zones
- Extract time series features an statistical measures using `Kats`: entropy, stability, lumpiness. seasonality
    - Use entropy as a measure of how unpredictable the price is.

Insights

- Across the zones, the lowest load/demand is the month of April.
- MILLWD has the lowest load across the time period, and NYC has the highest.
- Uncertainty in pricing peaks at April and dips in June. High uncertainty corresponds to relatively low load.
- The PCA component cluster plot shows some significant outliers for the month of July in zones LONGIL, MILLWD, and OH