

Wind Power Forecasting

Nigel Sjölin Grech

Note to reader: this should be read in combination with the accompanying tableau story.

Active Power Meets Expectation

A bit of context on the data. It is all numerical data collected from a single wind turbine and was intended as a challenge to predict the active power produced by the turbine.

Looking at the mean KWh produced we see that it follows a rather stable and seasonal pattern. Nothing here is surprising and it is as we expect it to be.

Patterns in features

Looking at the features of the data set we see that they all follow similar seasonal patterns and are very correlated to one another, some looking almost identical at first glance. Again this is expected and a different result would be indicative of an issue with the Turbine or the data or both.

Further Investigation

The lack of dimensions in the data set made making interesting observations between groups challenging, so I relied on using time and power production as dimensions. Here I am showing that trend again of the data increasing in the summer months between June and August, and dropping to lows in October and April.

A power cap?

Nothing out of the expected until I looked at the plots between power production and wind speed. There seemed to be a cut off point where increase in wind speed had no effect on the power being produced. The plot on top, where it is more pronounced, shows the hourly averages and the one on the bottom the daily ones. This I was not expecting.

A Hypothesis

This can be tested and based on the plot I assumed that there would be no difference in the means of power production at higher wind speed compared to lower ones. I tried a number of ranges, this, 9.5-10 vs >10, seemed the most reasonable.

Rethinking

The result of the t-test tells another story! That there is a difference in the means of these two groups and my initial assumption based on the scatter plot was wrong.

Looking at the plot on the bottom we see the plot of average hourly Active Power per day against wind speed zoomed in on that plateau. Perhaps the reason there is a difference in the means is that the variance of the point in the white portion of the graph is higher than the grey.

So the moral of this story is that plots are a great place to get ideas and hypotheses about our data but they should not be solely where we draw our facts from.