Data Science Job Postings Case Study

Drew Fingerhut

3/6/2022

**Background**

As a Data Analyst I am always seeking oppurtunites to increase my skill set and better myself. I have completed a variety of courses surrounding the Data Science field including machine learning, neural networks, NLP, and webscraping. As a way to show case this I took it upon myself to create a WebScraper using the BeautifulSoup library in Python.

**Overview**

In this case study I set out to search Indeed for Data Science job postings and gather a few key features from the listings. I thought it most important to gather the title, company, and salary if listed to compare listings in the current market. I did so by generating a WebScraper in Jupyter Notebook using the BeautifulSoup Library in Python. A spreadsheet was then generated and imported into Excel, where using the get data function, it was cleaned and organized for better interpretation. The cleaned data was then loaded into Tableau for visualization and analysis.

**Data Extracting**

Using power query the data for each power cabinet was pulled using a unique tag and timetable. As in previous case studies the data was pulled on an hourly basis and then summarized by day. These tables were create and later accessed for graphing the data. (1) The time range was also created using the current time and then another time that subtracts three months. That way everytime the file is refreshed it will show the most recent three months.

Text

Description automatically generated

**Figure 1: Power cabinet pulled data**

The data was plotted and a yearly average value was found. I then calculated a tolerance range for each cabinet and created a table for the range and lower and upper limit. (2)

Text, table

Description automatically generated

**Figure 2: Tolerance data table**

**Data Visualization**

It is then time to go in and create a custom combination chart. Plotting the data for the cabinet as well as creating a stacked area chart for the control limits creates the visual that I was looking for. (3)

Graphical user interface, application, table, Excel

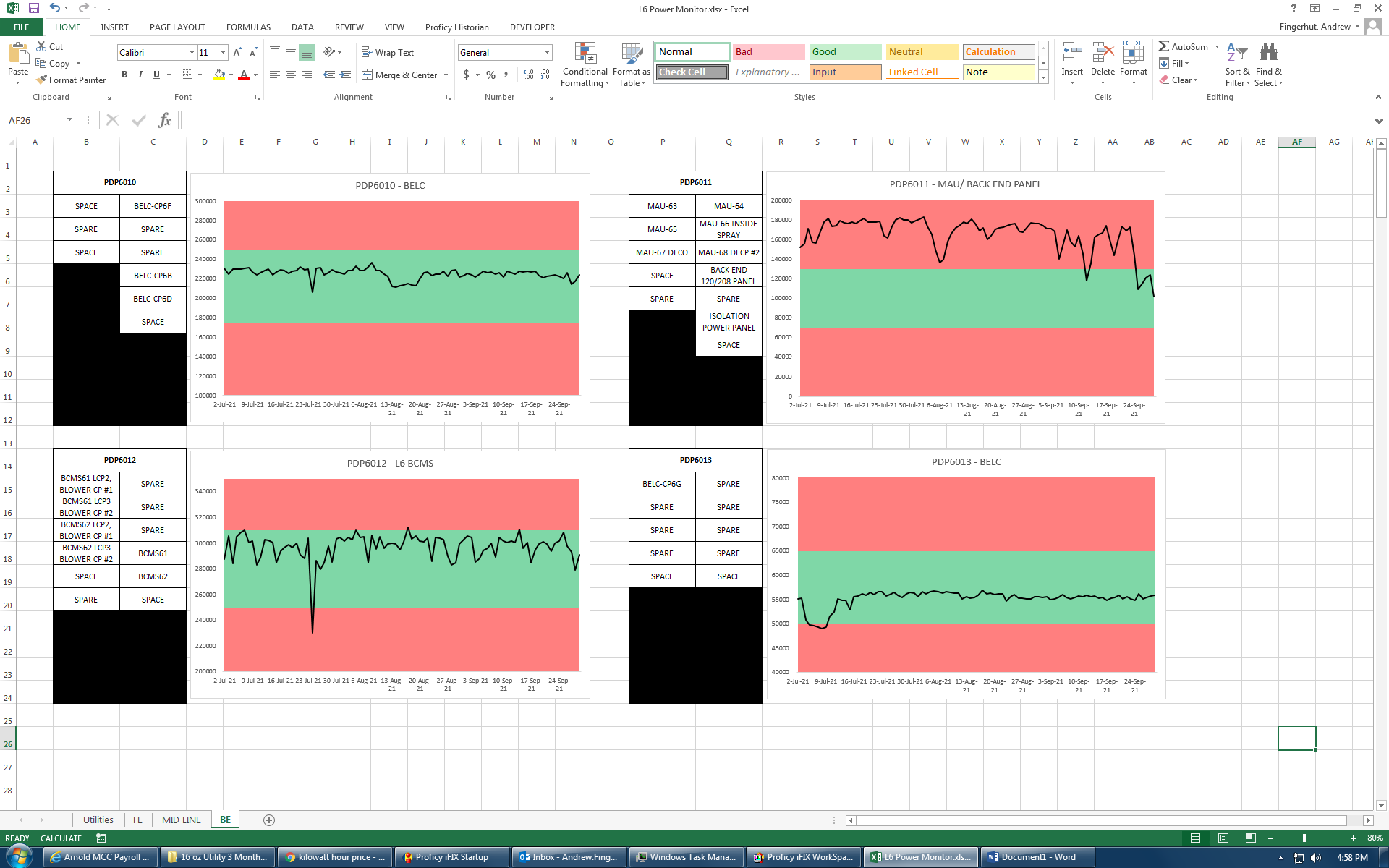
Description automatically generated

**Figure 3: Plotting the control limits**

**Reviewing Success**

Repeating these steps for all the power cabinets, the dashboard is created. This allows all cabinets to be reviewed in the matter of minutes. Data can be pulled up in the daily meeting and quickly be talked about. Action items are easily created and can be targeted at the issue cabinet. (4) In total this accomplishes the goals that were set out be achieved. The dashboard was put into use and adapted by the utilities department.

Graphical user interface

Description automatically generated with low confidence

**Figure 3: Breif look at the Dashboard**