USA Power Plant CO2 Emissions Analysis

USA CO2 Emissions 2016-2019

2.13 bn tons	2.04 bn tons	2.07 bn tons	1.90 bn tons
2016	2017	2018	2019

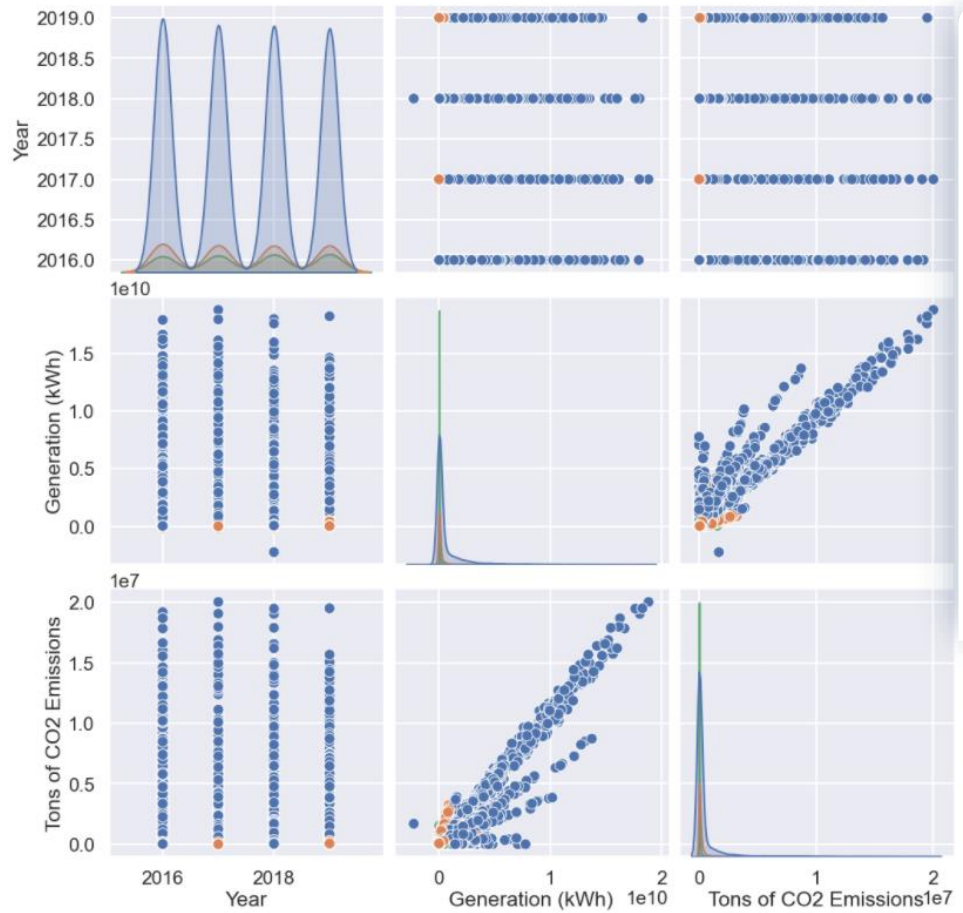
Challenge

This Power Plant CO2 Emissions analysis focused on finding demonstrable **statistical evidence** showing that electric power generation has a direct impact on CO2 emissions, but the following questions were also posed:

Key Questions

- How much CO2 emissions are associated with electricity generation?
- How much electricity is produced in each sector group, and how much fuel is required to produce this?
- Which state, sector group, and NERC regions produce the highest total emissions?
- Which electric plants emit the most CO2 emissions, and in which state, sector group, and NERC region are they located in?

Finding Relationships in the Data



Electric Power 49%

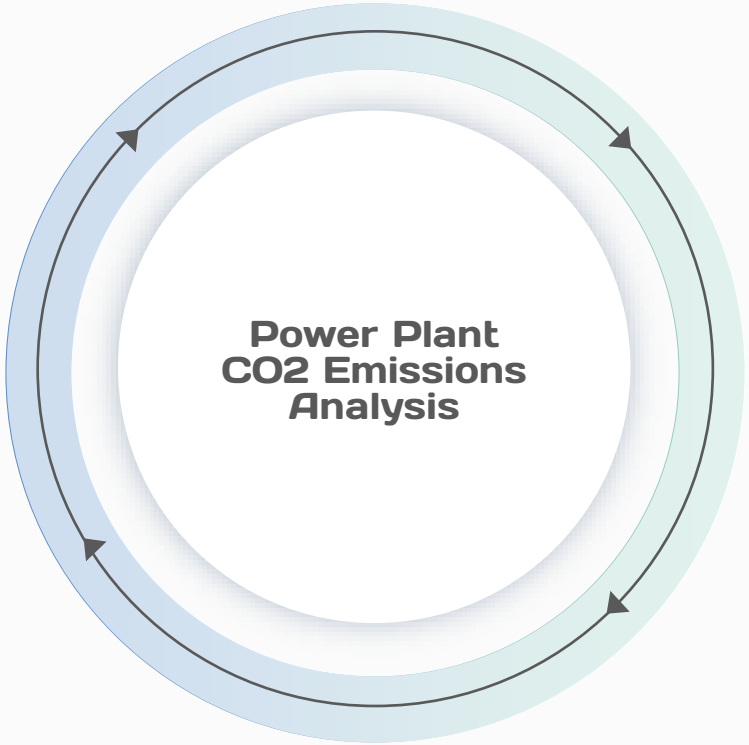
Commercial 27%

Industrial 14%

Project Life Cycle

Data Preparation

- Sourced public data from a reliable source
- Used **Python** to assess data accuracy, quality, consistency, completeness, uniqueness, and timeliness
- **Wrangled**, cleaned, transformed and merged the data
- Profiled the data using basic **statistical** methods



Data Exploration

- Explored data relationships using correlation heatmaps, scatterplots, pair plots, and categorical plots
- Performed **linear regression** analysis and **cluster analysis** using the elbow technique and k-means clustering
- Used shapefiles in a geospatial analysis
- Explored geographic variables using choropleth maps to draw early insights
- Performed **time series** analysis, including decomposition, tests for stationarity, and the stationarization of the data

Statistical *Analysis*

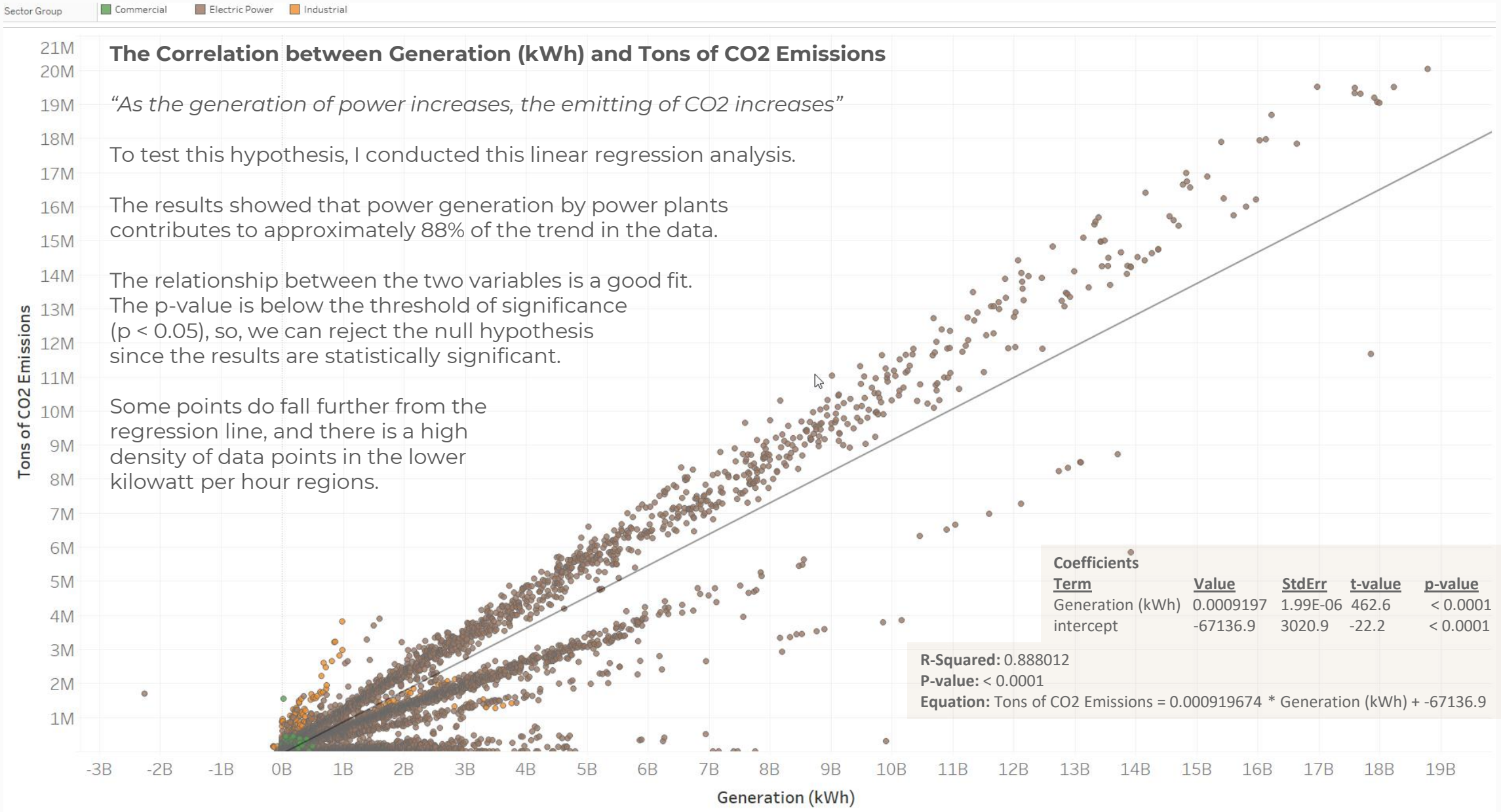
- The results of exploring relationships, performing regression and cluster analysis, and use of **geospatial analysis** makes it clear that there is a close relationship between power generation and higher CO2 emissions.
- This allows us to use the cluster model on other variables to gain greater insights

Data sources:
EIA (Energy Information Administration) & USA Shapefile

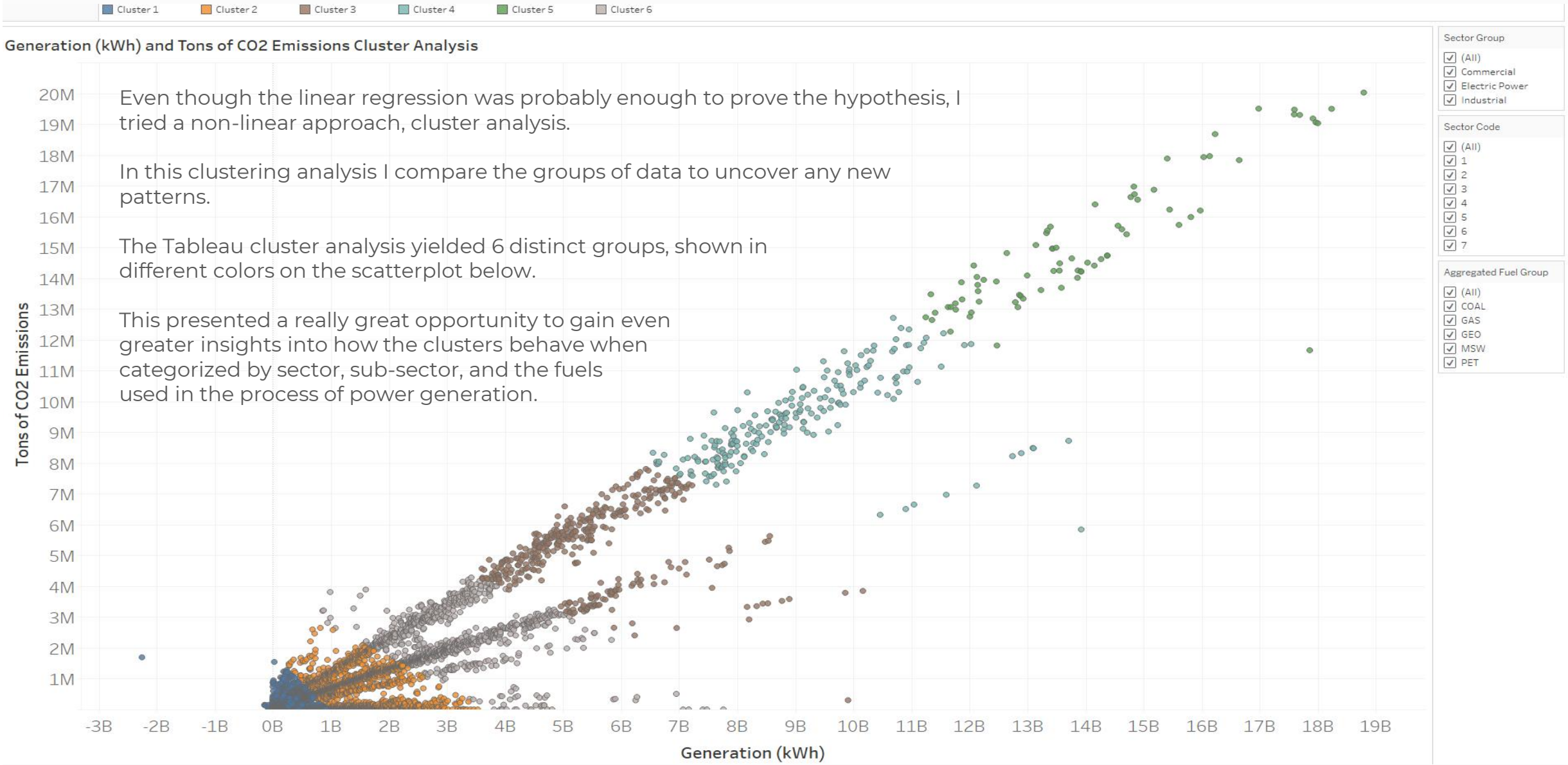
Challenges

- It is likely that the stationarized data is not sufficient for **fore-casting** since the EIA (Energy Information Administration) has only annual data, and any monthly or daily imputations would insert possible bias into the data.

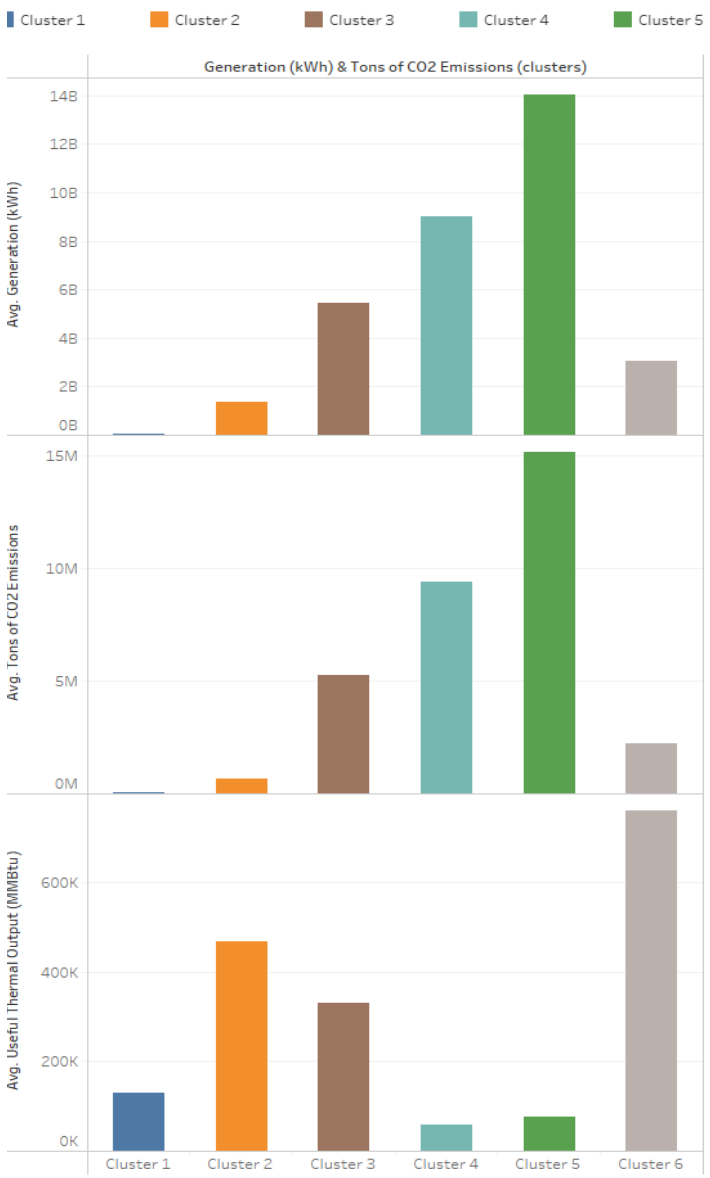
Linear Regression Analysis



Cluster Analysis



Cluster Analysis Results



Looking at the first **2 charts on the left**, we see that Cluster 5 (green), on average, generates the greatest amount of power while also producing the most CO2 emissions.

This is clearly apparent in all clusters and supports the hypothesis.

However, when I applied the cluster model to see if thermal output would follow the same trend (**3rd chart on the left**), it did not, as we can see that in this variable, greater power generation does not always mean greater thermal output, and we can see this behavior in the other clusters too.

Turning our focus to the **2 charts on the right**. Here, I added the sector code category to the analysis, and we see the same pattern where greater power generation means higher CO2 emissions.

Conclusions and Recommendations

Now that we've demonstrably and statistically answered the most important question about power plants and CO2 emissions, we can ask and answer additional question related to the topic at hand:

- How much CO2 emissions are associated with electricity generation?
- Which state's electric plants emit the most CO2 emissions, and in which sector group, and NERC region are they located in?
- Which fuels produce the highest CO2 emissions?

Explore the interactive dashboard for the answers & my GitHub repository for the Python code!

Limitations of the case study:
- The EIA (Energy Information Administration) creates biannual estimates of air emissions from power plants.
- Complete information is not collected from all power plants, so statistical imputation is also used.

Next steps:
- Gather more data points for these variables.
- Run a classification algorithm to see whether we can predict future CO2 emissions.
- Analyze the impact of additional variables on CO2 emissions.