

Saatvik Sandal 114378631

One thing I noticed with my data was that certain variables tend to group in a certain way. For example in the bar graph, where you can select a bar to generate colors, if you select energy_% and select the highest bar it matches with the biplot and parallel coordinate patterns of being distinct against something like acoustics. Other variables like the release day was completely random and randomly colored everything in.

For the k means clustering, the elbow method found that 5 clusters is the best option. After applying that to the visualizations, I would say it clustered pretty well.

The area distinctions are very apparent in MDS plot as well as the bi plot.

Cluster 1 and 3 were kind of jumbled up in the biplot but were very different in the MDS plot.

This might be because of the large dimensionality of the data or the existence of non-linear relationships in the dataset, which was seen in the previous labs.