Where I found the data set I am analyzing:

<https://www.kaggle.com/datasets/pritish509/s-and-p-500-esg-risk-ratings>

My primary question

What factors affect the ESG total score?

What I did first

My first thing I wanted to see is that if there is a relationship between the Total Workers and the Total ESG because to me it shouldn’t matter how many workers a corporation has the ESG is affected by other factors.

I ran a Regression Summary in R and I printed out the result. The scatter plot and the summary demonstrates a low correlation between the Total Workers and Total ESG scores with the R^2 value being 0.0002995744. Proving my initial assumption.

For the jupyter file I added I decided to run the elbow method and I couldn’t identify a K value. So, I ran an algorithm using the Silhouette Coefficient to determine the number of clusters. For K clustering and I made the scatter plot label which industries the data points are associated with. I noticed that overall the industries score all over the place.

Next Time

Investigate individual industry performances

Investigate other columns in the file with regression and report on it.

A graph showing a line

Description automatically generated

A screen shot of a computer

Description automatically generated