Peer-Graded Assignment: Milestone 3: Beyond Descriptive Stats

Dive Deeper

Look deeper into the features you are investigating. Consider:

- Relationships/correlation, Pearson correlation
- Linear regression for future prediction (if the relationship is linear)
- Textual analysis for TF-IDF (term frequency-inverse document frequency; row-based and column-based, stop-word removal?)

Specify 1-2 correlations you discovered. List the fields that you found to be correlated and describe what you learned from these correlations.

I discovered some correlations between the vaccination status and mask-wearing habits of the survey participants. The majority of participants who were vaccinated also responded that they wore a mask or face covering when in stores or other businesses “all or most of the time” (2,311). However, the majority of unvaccinated participants also indicated that they wore a mask or face covering “all or most of the time” (6,974) when in stores or other businesses. This finding was unexpected, but this could be because mask mandates were in place at the time.

I also created a correlation heat map using the seaborn library to view correlations between vaccination status, mask wearing, and various demographic variables. I found positive correlations between vaccination status and religion and between vaccination status and political party affiliation.

Go Broader

Expand the features you are investigating. Look for connections/relationships that you may have initially missed.

1. What jumps out at you now?
2. Use the descriptive stats to point you to features that you may now want to consider.

What key terms did you discover in any text analysis? For whom? Any themes? If you are not analyzing text, summarize what other things you are considering in your analysis.

I am looking at other factors that may be correlated with vaccination status, such as mask-wearing habits and other demographic information.

I also calculated weighted values for certain questions to see how the survey responses can be generalized to the greater population.

New Metric

Create 1 or 2 new metrics to track relationships of data you discovered. Explain why you created them.

I created new metrics to see the percentages of subgroups (for example, what percentage of vaccinated participants usually wear masks in public). I also added columns to dataframes showing the weighted vs. unweighted values of the survey responses. This helped me understand how the responses represent the greater population and helped me analyze the data in a new way.