Penn Data Science Boot Camp

Project 1 – Analysis and Conclusion

Eric Croston

5/22/2024

**Analysis Summary** This analysis addresses two questions about global COVID-19 data trends. First, what are the trends and relationships between new cases, active cases, and deaths worldwide? Second, how do these trends differ in the countries with the highest COVID-19 case counts compared to global patterns? The data for this evaluation is sourced from <https://www.kaggle.com/datasets/josephassaker/covid19-global-dataset>.

Key findings include:

1. A positive correlation exists between new cases, active cases, and deaths globally and in all countries.
2. New cases and active cases show a strong positive correlation both globally and in most countries.
3. Deaths exhibit a weak positive correlation with new and active cases worldwide.
4. The correlation between deaths and cases varies by country, ranging from weak to strong.

**Conclusion** The positive correlations suggest that increases in new cases typically correspond with rises in active cases and deaths. The globally weaker correlation between deaths and cases indicates that, on average, deaths are lower relative to cases.

The variation in death-case correlation across countries is notable. In some countries, death rates closely track case counts, while in others, deaths remain significantly lower. Analyzing the factors behind these differences could provide valuable insights into mitigating fatalities in future pandemics.