Penn Data Science Boot Camp

Project 1 – Analysis and Conclusion

Eric Croston

5/22/2024

* Analysis Summary

This analysis considers two questions about daily COVID-19 data. What are the trends and relationships between new cases, active cases and deaths worldwide? What are these trends and relationships in the countries with the most COVID cases and how do they compare worldwide? The data used to evaluate these questions is a csv file found at <https://www.kaggle.com/datasets/josephassaker/covid19-global-dataset>.

This analysis found:

1. There is a positive correlation between new cases, active cases, and deaths worldwide and within all countries.
2. There is a strong positive correlation between new cases and active cases worldwide and within most countries.
3. There is a weak positive correlation between deaths and both new and active cases worldwide.
4. The positive correlation between deaths and both new and active cases ranges from weak to moderate to strong in different countries.

* Conclusion

The positive correlation across new cases, active cases and deaths indicates that as new cases rise or fall so do active cases and deaths. The worldwide weak correlation of deaths to new and active cases shows that the number of deaths is lower than the number of cases.

The variation in the correlation of deaths to new and active cases between countries is a very interesting finding. This means that in some countries the number of deaths was very similar to the number of cases while in others the number of deaths was much lower than the number of cases. Determining what caused some countries to have a very low correlation between deaths and cases could yield very valuable insights into preventing deaths in a future pandemic.