The Exploration of the Relationship between COVID-19 and Per Capita Income

Author: Alise Pedersen

Key Findings:

The K-means clustering analysis categorized counties into three distinct groups based on death rates and per capita income in 2020:

1. **Cluster 0**: This group, comprised of 73 counties, had the highest median per capita income of all groups at $73,639, which is more than double the national median of $35,384. Despite this higher per capita income, Cluster 0 experienced a median death rate of 0.051897, which is higher than that of Cluster 1.

2. **Cluster 1**: With 437 counties, Cluster 1 had a median per capita income of $48,586—over $10,000 above the national median. The median death rate in this cluster was 0.036874, the lowest among the three clusters.

3. **Cluster 2**: This group included 152 counties with a median per capita income of $47,521 and a median death rate of 0.097121. While the median income in Cluster 2 was slightly lower than that in Cluster 1, it still exceeded the national median. However, Cluster 2 had the highest death rate.

Challenges:

The dataset has 1,128 missing values in the county column, which means that not all counties are represented in the analysis. This missing data could potentially impact the overall findings and interpretations.

Conclusion:

The analysis suggests that higher per capita income does not necessarily correlate with lower COVID-19 death rates or better outcomes. Despite having significantly higher median per capita incomes, counties in Cluster 0 experienced worse COVID-19 outcomes compared to those in Cluster 1. Interestingly, the lowest death rates were observed in counties with a median income just above the national median, rather than in the wealthiest counties. This disparity could be influenced by various factors such as healthcare access, public health infrastructure, and demographic differences, including age. Further research is needed to explore these factors in depth and understand their impact on COVID-19 outcomes.