

## Impact of COVID-19 on Work Habit, Stress Level Productivity Change, and Health Issues

Name: Jyoti Dave  
Master Of Science, Data Science  
Bellevue University  
DSC530 Data Exploration and Analysis  
Prof. Matthew Metzger

### Statistical/Hypothetical Question:

The central question for this analysis was to understand the impact of various factors, such as Stress Level, Productivity Change, and Health Issues, on Hours Worked Per Day. This inquiry aimed to explore whether higher stress levels or productivity changes were associated with longer work hours, and whether health issues play a significant role in determining how much time an individual spends working per day.

**Outcome of the Exploratory Data Analysis (EDA):** Through the exploratory data analysis (EDA), several trends and relationships were identified. Histograms of the key variables indicated that Hours Worked Per Day generally clustered around mid-range values, with relatively few extreme outliers. Correlation analysis revealed a very weak positive correlation between Hours Worked Per Day and Productivity Change (Pearson's correlation = 0.013), suggesting a minimal linear relationship between the two variables. Similarly, the analysis of Health Issue showed a similarly weak correlation (0.010), implying that health issues might not be as significant in influencing work hours.

The OLS Regression results also supported these findings, with Stress Level and Productivity Change having no statistically significant impact on the dependent variable (Hours Worked Per Day), based on the p-values and t-statistics.

**What was missing during the analysis?** One limitation of this analysis was the assumption that all the relationships between variables would be linear. Variables like Stress Level and Productivity Change may have non-linear effects on Hours Worked Per Day, which was not fully explored. Additionally, potential interactions between variables (e.g., between Stress Level and Health Issue) were not examined, and such interactions may provide more nuanced insights into the factors influencing work hours.

**Variables that could have helped:** Additional variables such as Age, Income, and Sector could have provided more context for the analysis. For instance, income could serve as a proxy for work pressure or job demands, while sector might reveal differing patterns based on industry-specific work culture. Including such variables may have strengthened the understanding of factors influencing work hours.

**Assumptions that may have been incorrect:** An assumption in this analysis was that variables like Stress Level and Productivity Change have a direct, linear relationship with Hours Worked Per Day. The regression model assumed that these factors would affect work hours in a straightforward manner, but other underlying factors, such as work-life balance, job type, and personal circumstances, may mediate these relationships in ways that were not captured here.

**Challenges faced and areas of confusion:** One challenge was the proper handling of outliers in the dataset. While some outliers were detected and acknowledged in the histograms, their impact on the analysis wasn't fully explored. The assumptions of normality in some variables, such as Hours Worked Per Day, also led to difficulties in selecting appropriate statistical models for analysis. Additionally, while performing the regression analysis, the results were inconclusive, which made it challenging to provide a strong recommendation regarding the variables influencing work hours.

Overall, this analysis highlighted the importance of considering additional variables and refining assumptions, especially regarding non-linear relationships and variable interactions. Future analysis could benefit from a more comprehensive approach, incorporating more diverse data and exploring alternative models.