

Andrew Zhou

A5: Extension of Analysis Plan

1. My analysis is going to focus on unemployment and uninsured rate. Uninsured rate is the number of people who do not have healthcare insurance. Texas is known to have very high uninsured rate, and I would like to see how COVID impacted that rate. Most people have their insurance through their employment, and by having a high unemployment rate, I also want to see if that amplified the uninsured rate even further. Through this study, I hope I can see the impact of COVID and uninsured rate. My main human centered data science question is: how does COVID impact the uninsured rate of healthcare insurance, and how does this affect peoples' ability to get healthcare they need.
2. Research Question: How was unemployment rate impacted by the pandemic? How was the uninsured rate in Dallas, Texas impacted by the pandemic?

Hypothesis: I hypothesize that as daily new cases of COVID increase, then unemployment rate will also increase, which in turn results in a very big increase in the uninsured rate.

3. I will be looking at 4 datasets for this project. The first two are the COVID confirmed cases by John Hopkins University and the Mask Mandates from CDC. The other two external data sets are for unemployment rate and uninsured rate. For unemployment rate, I have data from the Bureau of Labor Statistics that includes data of unemployment rates for Dallas, Texas from 2011 to 2021. I will be using this data set to see if the number of COVID cases increases, then what happens to the unemployment rate (the correlation between COVID cases and unemployment rate). My second external data set is the uninsured rates, which were found from the U.S. Census Bureau. The data includes uninsured rates from 2010 to 2019, and I had used US Department of Health and Human Services to obtain uninsured data for 2020 and 2021. I want to see the correlation of unemployment and uninsured rates and also COVID cases and uninsured rates.

Links:

https://data.bls.gov/timeseries/LAUMT481910000000003?amp%253bdata_tool=XGtable&output_view=data&include_graphs=true
<https://www.census.gov/>

4. My data external data sets are unfortunately only having data that is monthly and annually. For uninsured rates, I only have annual data, which would only account for 2 data points since COVID (2020 and 2021). It will be difficult to test correlation for this variable. For unemployment, I have monthly data which is slightly better.

My analysis could also include other factors that impact unemployment and uninsured rates, which my data and analysis may be unable to capture. I should be aware of potential factors like this, such as President Biden's unemployment benefits. His benefits allow quite a few people to continue be unemployed, rather than looking for jobs.

5. I plan on plotting the time series into one graph and see if there are any correlations between COVID cases, mask mandates, unemployment rate, and uninsured rate. I will be including a time series plot in my report. I may try an ARIMA model with the uninsured rate as the dependent variable and the other variables as independent variables. I want to see if the data will come out correlated, and check for any assumptions such as normality, independence, and constant variance. I also want to try an ARIMA model on unemployment to see if I can predict unemployment rate to some extent using COVID cases. If these models turn out to be significant, I would like to add these to my report and plot the ARIMA onto my plot. I may also try a regression to ignore of uninsured rates and unemployment rate to see if there is any correlation without time being a factor.
6. For this A5, I have already collected the necessary extra data for this project. If I find it necessary to find more data, I hope to find data by the end of next week. I believe exploratory data analysis and initial time series visualization should be completed first. I will attempt to complete this by next week. I hope to complete at least one ARIMA model and improve it by the second week. By the third week, I hope to have all my models ready and conclusions completed. By the fourth week, I hope to have my presentation ready and final documentation and repository ready for submission.