DATA 512 PROJECT PART 2

EXTENSION PLAN

I.   Motivation/problem statement:

Norfolk is a county in Massachusetts. It is currently made up of one city and twenty seven towns. The population of the county is 703, 740 and the employed population (as of 2020) is 376,643. I am interested in finding out how COVID 19 had an impact on the employment rate in the county. I would like to understand how many unemployment claims were filed during 2019-2020, whether there is any correlation between the unemployment rates and COVID-19, and if so how COVID 19 affected different genders, races, education levels, ethnicities. I chose this analysis because COVID 19 was a tough time for all of humanity but in general it was tougher for some more than others. I want to understand if there was an impact on unemployment due to COVID 19, were some marginalized communities more affected than the others. As an extension, if time permits, I would also like to explore whether there was any impact of masking enforcements on the employment/ unemployment rates. Is it possible that even with a rise in COVID-19 cases, given that masking was mandated, there was maybe a decrease in unemployment?

Human Centered Data Science focuses on fundamental principles of data science and its human implication. This includes doing research ethically, creating reproducible works and societal impacts of data science. I want to work on this problem statement as it helps me utilize the potential of data science in understanding the impact of COVID 19 on the livelihoods of the Norfolk county people. Since this work is focused on a small part of the US, I don't expect my work to be representative of the entire US population but it could be somewhat representative of the state. Making this work reproducible and open source can help me create a work that helps us understand what steps can be taken if ever another pandemic hits and how government resources could probably be distributed among different communities.

II.   Research questions and/or hypotheses:

Research Question:

1. How did the increase in COVID-19 cases impact employment in Norfolk County in 2020-2021?

2. Was there an increase in the unemployment claims as COVID-19 progressed in 2020-2021?
3. What proportion of unemployment claims to the population for gender, race/ethnicity to understand how much such demographic groups were impacted?
4. What was the average weekly claim prior to filing the claim during the COVID 19 progression? Was there any connection between increase of cases and average wages allotted?
5. How did educational background have an effect on unemployment claims and was there a group that was severely impacted due to COVID-19 between 2020-2021?

Further research questions/ Hypothesis (if time permits):

1. (Hypothesis) The number of unemployment claims were increased for all genders before and after COVID-10 peak cases in 2020-2021.
2. Is there an interesting connection between masking mandates and unemployment claims in Norfolk County between 2020-2021?

III.   Data to be used:

My primary source of additional data is as follows:

1. Weekly Unemployment Claims Data
   a. **Link to dataset:** https://lmi.dua.eol.mass.gov/lmi/ClaimsData
   b. This data is hosted on a website operated by the Department of Economic Research and was funded by a grant awarded by the U.S. Department of Labor's Employment and Training Administration.
   c. **Terms of Use:** https://www.mass.gov/terms-of-use-policy
   d. **Dataset Description:** Initial and continued weekly Unemployment Insurance claims by county. Includes claimant demographics (gender, race and ethnicity), industry, occupation, and education.
   e. **Purpose:** This data will help me answer the questions on increase in unemployment claims in connection to COVID 19, impact on unemployment claims in both genders before and after COVID 19 peak season and general analysis on education backgrounds and average weekly wages.

My secondary sources of additional data are:

IV. Employment and Wages Data:
   A. **Link to dataset:**
      https://lmi.dua.eol.mass.gov/LMI/EmploymentAndWages#
   B. This data is hosted on a website operated by the Department of Economic Research and was funded by a grant awarded by the U.S. Department of Labor's Employment and Training Administration.
   C. **Terms of Use**: https://www.mass.gov/terms-of-use-policy
   D. **Dataset Description:** Information on employment and wages is available by industry for counties in Massachusetts.
   E. **Purpose:** This data will help me answer the questions with respect to employment numbers during the COVID 19 cases increase.

V. Labor and Unemployment Data:
   A. **Link to Dataset:**
      https://lmi.dua.eol.mass.gov/LMI/LaborForceAndUnemployment#
   B. This data is hosted on a website operated by the Department of Economic Research and was funded by a grant awarded by the U.S. Department of Labor's Employment and Training Administration.
   C. **Terms of Use:** https://www.mass.gov/terms-of-use-policy
   D. **Dataset Description:** Information on the labor force, employment, unemployment, and unemployment rates for each county in Massachusetts.
   E. **Purpose:** This data has employment rates with respect to labor available and I thought this would help me with my analysis.

VI. Unknowns and dependencies:

The first assumption on the datasets are that they are reliable as they come from verified sources and that data collected was without any bias and represents the entire county. Another assumption is that it will be possible to find the demographic information of all years to be able to assess the proportion of unemployment across the demographics for the research. It is also important to keep in mind that no analysis in this report implies causation. Even if there is an

interesting connection between the rise of cases and the number of unemployment claims filed, it does not imply the COVID-19 cases are the cause for it.

VII.   Methodology:

I plan to investigate the research questions much like what I did for Project Part 1- Common Analysis. I will plot the weekly confirmed cases of COVID 19 along with the weekly data of employment numbers and unemployment claims to see if there is a trend or interesting patterns between the two. I will use Change Point Detection to understand how to understand points that can help determine the turn in the flow of events in the time series. I also plan to use exploratory data analysis and plot visualizations for different demographic and educational data points and try to find interesting trends or maybe spot anomalies which would address the issues that marginalized groups face. If time permits, then I would like to use the t-test to compare the average number of unemployment claims before the peak COVID 19 case rise with the average number of unemployment claims after COVID 19 cases for the hypothesis but I haven't given too much of a thought about it and may change the test if I find something better justified to handle the average claims at the different data points. The same goes for the research questions in finding interesting patterns between masking mandates and unemployment claims.

VIII.   Timeline to completion:

Timeline	Tasks
Nov 10th - Nov 16th	1. Explore the datasets more to filter for Norfolk county data and for the specific years 2. Find online papers/ resources that have done an analysis of unemployment and COVID-19 to understand if there is existing research that can help with my analysis. 3. Start the analysis notebook and exploring and cleaning the datasets
Nov 17th - Nov 23rd	1. Complete the first three research questions 2. Visualize the data and save the results

	3. Start the documentation process for the first three research questions and initial exploratory analysis.
Nov 23rd - Nov 29th	1. Complete the next two research questions and the hypothesis question 2. Visualize the data and save the results 3. Start the documentation process for the rest of the analysis. Stretch Goal: 1. Find relation between masking enforcements and employment rates 2. Document the process.
Nov 30th - Dec 7th	1. Research PechaKucha Style Presentations 2. Create a presentation of 12 slides 3. Prepare for Project presentation 4. Create GitHub repository
Dec 8th - Dec 12th	1. Upload all datasets, fully documented analysis code, visualization results into the repository 2. Add documentation - project report and intermediate data files to repository 3. Add README and Licenses 4. Submit repository