Extension Plan

Part 2 – Course Project

Motivation / Problem Statement

Meta just laid off 13% of their staff, 11,000 employees, today morning. Twitter sacked about 50% of their employees last week. These are a few of the many companies that have laid off large fractions of their workforce. Many new graduates, including our cohort, will be searching for jobs in one of the barest job markets of recent times. However, this situation had been predicted a few months back and it was known that companies will start preparing for the incoming recession.

Around two and a half years back, many employees and new graduates found themselves in a similar situation. However, at that time the situation was unprecedented and caused due to the outbreak of the infamous COVID-19 pandemic. The beginning of the pandemic saw significantly large increases in unemployment rates around the world. For this project, I shall be analyzing the unemployment rates for Hudson County during the COVID-19 pandemic (2/1/2020 to 10/1/2021) in relation to the number of COVID-19 cases and deaths.

This problem statement is human-centered since it directly relates to the employment of multiple workers residing in **Hudson County**, trying to earn a living for their families, during the unforgiving COVID-19 pandemic.

Research Question

Q1. How did the increase in COVID-19 **cases** in Hudson County affect its unemployment rate?

Q2. How did the increase in COVID-19 **deaths** in Hudson County affect its unemployment rate?

Q3. Does the employment rate in Hudson County depend more on the number of COVID cases or the number of deaths due to COVID-19?

Data

I shall be making use of two time series data sets.

1. COVID-19 cases and deaths data from John Hopkins University
   This is the same dataset that we used in Part 1
   Link: https://www.kaggle.com/datasets/antgoldbloom/covid19-data-from-john-hopkins-university
   License: https://creativecommons.org/licenses/by/4.0/

This is a daily updating version of the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). The data updates every day at 6 am UTC, which updates just after the raw JHU data typically updates.
I'm making it available in both raw form (files with the prefix RAW) and convenient form (files prefixed with CONVENIENT).
The data covers:

- confirmed cases and deaths on a country level
- confirmed cases and deaths by US county
- some metadata that's available in the raw JHU data

2. The unemployment rate in Hudson County, NJ
   Link: https://fred.stlouisfed.org/series/NJHUDS7URN
   Source: U.S. Bureau of Labor Statistics
   Release: Unemployment in States and Local Areas (all other areas)

   The Civilian Labor Force includes all persons in the civilian noninstitutional population ages 16 and older classified as either employed or unemployed.

   Employed persons are all persons who, during the reference week (the week including the 12th day of the month), (a) did any work as paid employees, worked in their own business or profession or on their farm, or worked 15 hours or more as unpaid workers in an enterprise operated by a member of their family, or (b) were not working but who had jobs from which they were temporarily absent because of vacation, illness, bad weather, childcare problems, maternity or paternity leave, labor-management dispute, job training, or other family or personal reasons, whether or not they were paid for the time off or were seeking other jobs. Each employed person is counted only once, even if he or she holds more than one job.

   Unemployed persons are all persons who had no employment during the reference week, were available for work, except for temporary illness, and had made specific efforts to find employment sometime during the 4 weeks ending with the reference week. Persons who were waiting to be recalled to a job from which they had been laid off need not have been looking for work to be classified as unemployed.

   The unemployment rate is the unemployed percent of the civilian labor force [100 times (unemployed/civilian labor force)].

Unknowns / Dependencies

Hudson County, NJ is an urban county comprising major cities such as Jersey City, Hoboken, Secaucus, etc. Hudson County has high residential and commercial importance. Hence, reporting of COVID-19 cases/deaths as well as unemployment rates are done regularly and satisfactorily in Hudson County. Both the above-mentioned datasets have no missing data. Hence, we have no unknowns.

Since my analysis only relies on the association of COVID-19 cases/deaths with the unemployment rate in Hudson County, we have all the required data and have no dependencies.

It shall be possible to satisfactorily answer the mentioned research questions using the above data sources within the expected timeline.

Methodology

To compare the time series of COVID-19 cases/deaths with the unemployment rate in Hudson County I will be using four techniques:

1. Pearson correlation
   Source: https://en.wikipedia.org/wiki/Pearson_correlation_coefficient

2. Time Lagged Cross Correlation (TLCC)
   Source: https://www.sciencedirect.com/science/article/abs/pii/S0375960114012766

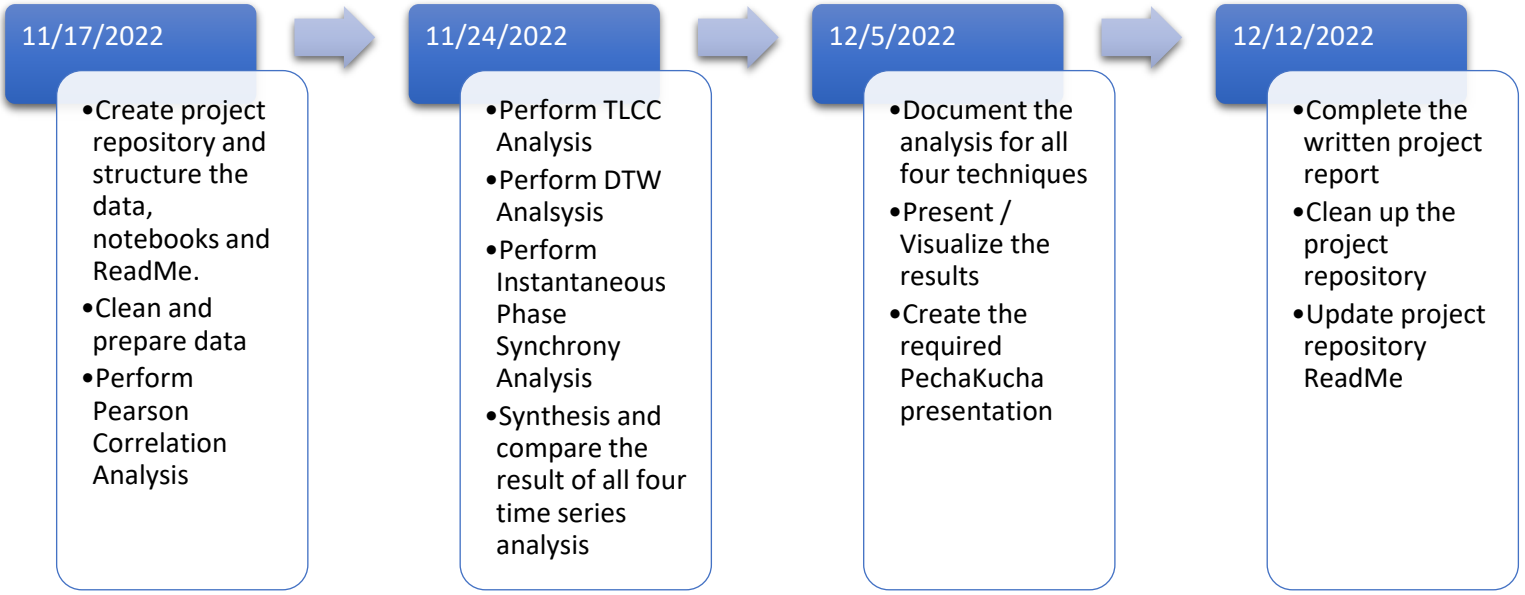
3. Dynamic Time Warping (DTW)
   Source: https://en.wikipedia.org/wiki/Dynamic_time_warping

4. Instantaneous phase synchrony
   Source: https://en.wikipedia.org/wiki/Instantaneous_phase_and_frequency

It is extremely difficult to define a statistical test to imply causation. However, I have found the above four techniques to compare two time series for any association. Some of these methods require stationary time series, while some work for non-stationary time series as well. I shall be experimenting with all four methods and comparing the results.

In addition, I shall also use these techniques to analyze whether the employment rate in Hudson County is associated with the number of COVID-19 cases or with the number of deaths due to COVID-19.

Timeline


11/17/2022
•Create project repository and structure the data, notebooks and ReadMe.
•Clean and prepare data
•Perform Pearson Correlation Analysis
11/24/2022
•Perform TLCC Analysis
•Perform DTW Analsysis
•Perform Instantaneous Phase Synchrony Analysis
•Synthesis and compare the result of all four time series analysis
12/5/2022
•Document the analysis for all four techniques
•Present / Visualize the results
•Create the required PechaKucha presentation
12/12/2022
•Complete the written project report
•Clean up the project repository
•Update project repository ReadMe