

# Project Report

## Part 4 – Course Project

### Introduction

Meta recently laid off 13% of their staff, 11,000 employees. Twitter sacked about 50% of their employees. 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 and livelihood of multiple workers residing in **Hudson County**, trying to earn a living for their families, during the unforgiving COVID-19 pandemic.

The primary purpose of this analysis is to question the popular belief, at the start of the pandemic, that the unemployment rate will rise with the increase in the number of COVID cases. I would like to argue that the number of deaths due to COVID had a more direct correlation with the unemployment rate, rather than the number of COVID cases. Using Pearson correlation analysis, I have successfully been able to verify this claim, while also considering its limitations.

### Background

March and April 2020 marked the brink of the downfall of the US markets owing to the COVID-19 pandemic. As a result, the US saw large-scale layoffs and a soaring unemployment rate. The Washington Post [2] claims that the unemployment rate in April 2020 was the worst since the “Depression Era”. An article published by CNN Business [1] claims that the US economy lost 20.5 million jobs in April itself, which is *“by far the most sudden and largest decline since the government began tracking the data in 1939.”* The article also claims that the number of jobs lost in March and April 2020 was double of 8.7 million jobs lost during the previous financial crisis.

Another article by The Guardian [3], dated as early as March 26, 2020, claimed that a record 3.3 million Americans had filed for unemployment. The previous record was 695,000 in October 1982.

The above articles portray this unemployment crisis to be one of the worst in the history of the US. With many articles speculating what factor was most closely related to this rise in unemployment, I have tried to answer this question in this project. From the many articles published in March and April 2020, it is evident that most people believe that the increasing number of COVID cases would trigger the rise in unemployment. However, having a retrospective view I would like to argue that the number of deaths due to COVID had a more direct correlation with the unemployment rate, rather than the number of COVID cases.

## Research Questions

- 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?

## Methodology

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.

### Step 1: Data collection

- Download the required CSV files from all the Data Sources mentioned below.
- Rename the column names to maintain consistency between data frames.

### Step 2: Data selection and sub-setting

- Filter the data for the data range 02/01/2020 to 10/01/2021.
- Aggregate the COVID cases/deaths monthly.
- Create time series data frames for COVID cases/deaths and the unemployment rate reported monthly.

### Step 3: Correlation Analysis

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

1. Pearson correlation [4]
2. Time Lagged Cross Correlation (TLCC) [5]
3. Dynamic Time Warping (DTW) [6]
4. Instantaneous phase synchrony [7]

Reason for shortlisting these techniques:

- These allow us to analyze the correlation between two time series.
- The above set is a good mix between methods that require stationary time series and that don't.
- We want to study correlations of an overall trend eliminating biases that may occur at certain points during the pandemic.

Ethical considerations:

- During the pandemic, certain sections of society were targeted, such as Asian Americans and African Americans.
- An example would be hostility shown towards Asian Americans, by claiming that they were responsible for this pandemic.
- None of the datasets used in this project segregates data by race or other attributes. Hence, we do not know whether the unemployment rate was biased towards or against any particular subset of people.
- However, since we only wish to study the overall correlations, we should not bias our study to certain groups and look at the overall unemployment rate and the number of COVID cases/deaths.

### Step 4: Visual Analysis

- Plot the COVID cases/deaths as well as the unemployment rate superimposed on the same chart. Any obvious correlations would be evident in such a graph.
- We also align this graph with another plot (sharing the same x-axis), indicating the Pearson correlation coefficient for each month, calculated using a 3-month frame.

## Findings and Discussions

From the above 4 correlation analysis techniques, only the first method (Pearson correlation analysis) provided us with understandable, intuitive, and explainable results.

### Pearson Correlation Analysis

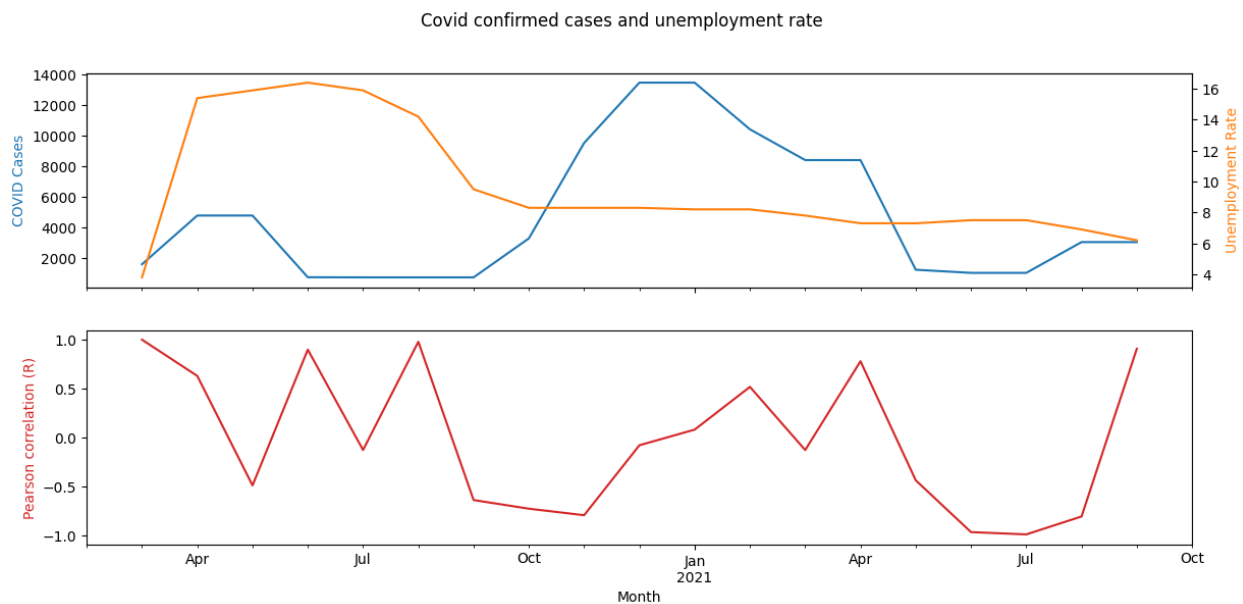
#### *COVID Confirmed Cases vs Unemployment Rate*

Overall Pearson R = **0.024714571151235428**

P-value = 0.917625354716538

Confidence Interval: low=-0.42242616402559624, high=0.46218061729684407

The first graph plots the visual comparison of the number of confirmed cases and the unemployment rate for the duration of the pandemic. We can also visualize how the Pearson correlation coefficient between the number of confirmed cases and the unemployment rate has changed over time.



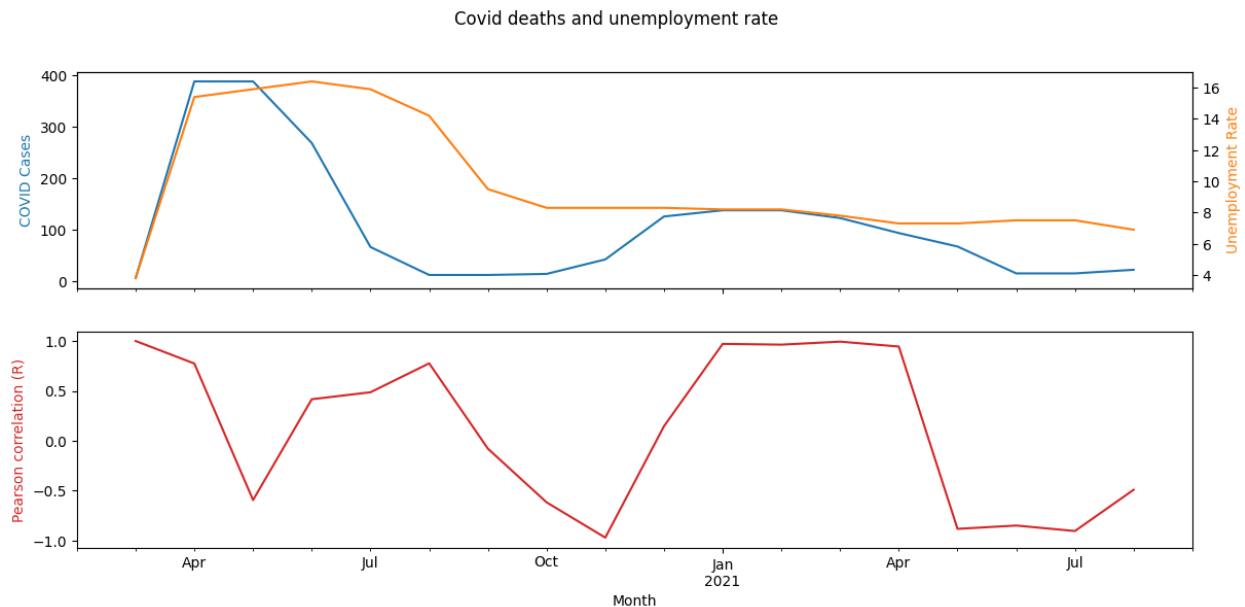
We can see some correlation during the beginning of the first wave. However, overall, we do not see any significant correlation. The frame-wise correlation is not stable and varies a lot during the pandemic. Analytically,  $R=0.024$  with a p-value of 0.92 indicates almost no correlation between the two time series. Hence, we can conclude there is no significant evidence for any evidence of a correlation between COVID confirmed cases and the unemployment rate.

### *COVID Deaths vs Unemployment Rate*

Overall Pearson R = **0.5901300165453667**

P-value = 0.0061616943614491585

Confidence Interval: low=0.19978100403594903, high=0.8188201716160485



$R=0.59$  indicates a strong correlation between the COVID death and the unemployment rate. A p-value of 0.006 ( $< 0.05$ ) strengthens the claim of correlation.

Visually, we notice a much stronger correlation between the unemployment rate and the COVID deaths as compared to the COVID cases. Both the graphs follow a very similar trend. We notice a high correlation, especially during the months of the first and second waves of the pandemic. However, there are a few months where the correlation is absent or highly negative.

We can conclude that there is significant evidence for a correlation between the COVID deaths and the unemployment rate.

Clearly, the unemployment rate is more closely correlated to COVID deaths, rather than COVID Cases.

## Limitations

The above analysis has certain limitations that must be considered:

1. Small data size
  - a. We have only analyzed data for 20 months.
  - b. We have used a rolling window of just 3 months while calculating the frame-wise Pearson correlation.
  - c. For the second comparison, we conclude that we have a correlation. In fact, visually we notice just two periods for which there is a high correlation – The first and second waves of the pandemic.
2. Confounding factors
  - a. The rise in unemployment coincides with the start of the pandemic and the second wave. Both were periods of uncertainty. The uncertainty could be the main reason for disturbance in the markets resulting in layoffs and unemployment. However, it is likely that future waves may not bring uncertain times again. The next time the number of deaths increases, it may be possible that employment is not affected.
3. Implicit biases in the data
  - a. We have not considered that the affected population due to employment may disproportionately belong to a particular race or occupation type. It is possible that most jobs affected were those that could not be carried out from home, such as laborious jobs. Or maybe the rising racial discrimination during the pandemic caused employees of a particular race to be laid off. Hence, our analysis may not apply to all employees equally.

## Conclusion

### Pearson Correlation Analysis:

- Covid Confirmed Cases vs Unemployment Rate:
  - $R = 0.024714571151235428$
- Covid Deaths vs Unemployment Rate:
  - $R = 0.5901300165453667$

Using Pearson Correlation Analysis, analytically and visually we can conclude that the unemployment rate is more closely correlated to COVID deaths, rather than COVID Cases.

In the situation of rising COVID cases, you need not necessarily worry much about losing your job or having to search for a job in a bare job market. However, if the number of deaths due to COVID is also increasing, it should be a matter of concern.

This conclusion comes with its limitations of small data size, presence of confounding factors, and implicit biases in the data.

## References

- [1] <https://www.cnn.com/2020/05/08/economy/april-jobs-report-2020-coronavirus/index.html>
- [2] <https://www.washingtonpost.com/business/2020/05/08/april-2020-jobs-report/>
- [3] <https://www.theguardian.com/business/2020/mar/26/us-unemployment-rate-coronavirus-business>
- [4] [https://en.wikipedia.org/wiki/Pearson\\_correlation\\_coefficient](https://en.wikipedia.org/wiki/Pearson_correlation_coefficient)
- [5] <https://www.sciencedirect.com/science/article/abs/pii/S0375960114012766>
- [6] [https://en.wikipedia.org/wiki/Dynamic\\_time\\_warping](https://en.wikipedia.org/wiki/Dynamic_time_warping)
- [7] [https://en.wikipedia.org/wiki/Instantaneous\\_phase\\_and\\_frequency](https://en.wikipedia.org/wiki/Instantaneous_phase_and_frequency)
- [8] <https://towardsdatascience.com/four-ways-to-quantify-synchrony-between-time-series-data-b99136c4a9c9>

## Data Sources

- John Hopkins University COVID-19 data
  - <https://www.kaggle.com/datasets/antgoldbloom/covid19-data-from-john-hopkins-university>
- Masking Mandates by County
  - <https://data.cdc.gov/Policy-Surveillance/U-S-State-and-Territorial-Public-Mask-Mandates-Fro/62d6-pm5i>
- Mask Compliance Survey
  - <https://github.com/nytimes/covid-19-data/tree/master/mask-use>
- The unemployment rate in Hudson County, NJ
  - <https://fred.stlouisfed.org/series/NJHUDS7URN>