# Abstract

The aim of this project is to analyse Covid-19 data in Ontario for 2021. All data used was from `https://data.ontario.ca/dataset/`. The data being analysed is confirmed positive Covid-19 cases per day, total hospitalizations per day, patients in the ICU due to Covid-19 per day, patients in the ICU due to Covid-19 requiring the ventilator per day, as well as total number of people who have been vaccinated for at least 14 days. The available vaccination data begins from January 2021. Routine VOC PCR testing of positive Covid-19 samples ceased on November 12th 2021 [1], and so to compare the effect of vaccinations to positive Covid-19 cases, we use analyse data from 1st January 2021 to 31st October 2021. We investigate the effect increased positive Covid-19 cases has on hospitalizations, on requirement of the ICU and ventilators. We also investigate the effect of the first dosage of vaccination on positive Covid-19 cases and hospitalization numbers. In order to investigate the effect of the vaccine, cumulative data for the vaccination is used (i.e. how many total people are vaccinated by a given day). We also shift the vaccination data by 14 days, in order to have our data represent number of people who have been vaccinated for at least 14 days, to account for the delay between receiving the vaccination and its overall affect on the population.

# Data Acquisition and Filtering

The data from the data.ontario website was downloaded and plotted as discrete time series. Before using techniques to compare different data sets (e.g. auto-correlations etc), the data was filtered using techniques from class, to remove any unwanted noise. In order to filter the data, techniques from lab 3 were used, where our original time-series was de-trended by removing a numpy.polyfit line (of sufficiently high degree, at least degree 7).The de-trended data was Fourier Transformed (using Numpy's built in FFT), with the Fourier spectrum in the frequency domain set to zero for high frequencies. This filtered de-trended time-series was then inverse Fourier transformed back to the time domain (using Numpy fft.ifft), with the original trend finally added back in. Initially, filtering this way led to a bit of an unexpected effect; an unnatural uptick was introduced to the end of the data. It appears that this uptick was due to the periodic nature of the inverse Fourier transform, leading to the end values of the data being mirrored (and then offset due to frequency changes when filtering). In order to rectify this, the data used was taken for up-to November 14th 2021, with a boxcar window applied before filtering (setting data after October 31st 2021 to 0). The data is then only plotted up-to October 31st 2021. The data for number of total first dose vaccine administered had minimal high frequency noise (as this data was cumulative, not day-to-day), and so this data was not filtered. The data was however shifted, in order for it to represent total people who had been vaccinated for 14 days. The plots of all data used (raw versus processed) is visible on the next page.

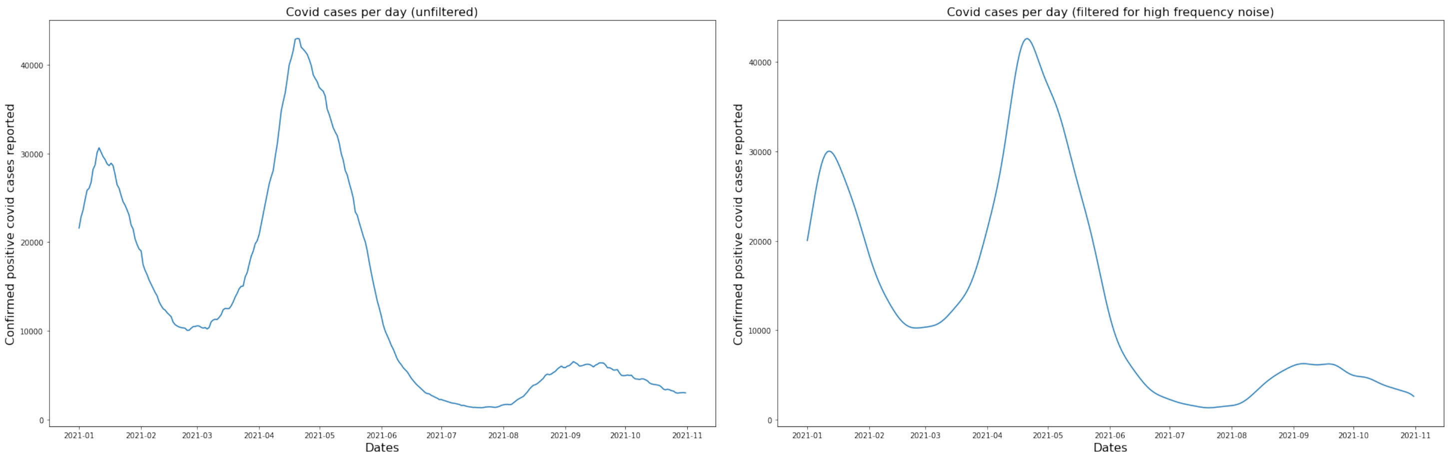


**Figure 1: Covid-19 cases f-domain filtering comparison; unfiltered (left) versus filtered (right)**

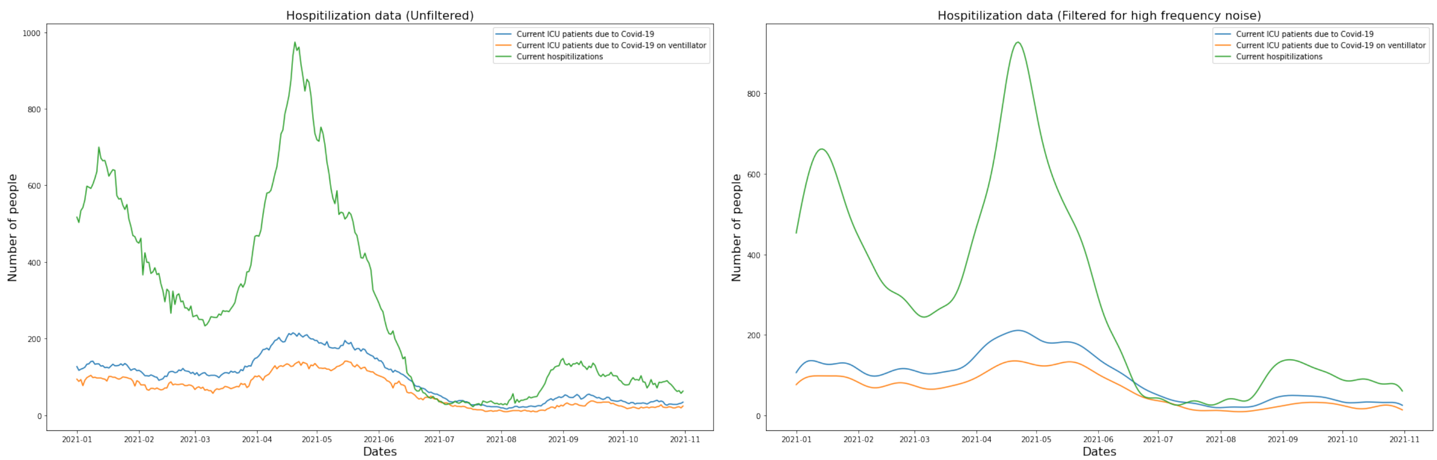


**Figure 2: Hospitalization data f-domain filtering comparison; unfiltered (left) versus filtered (right)**

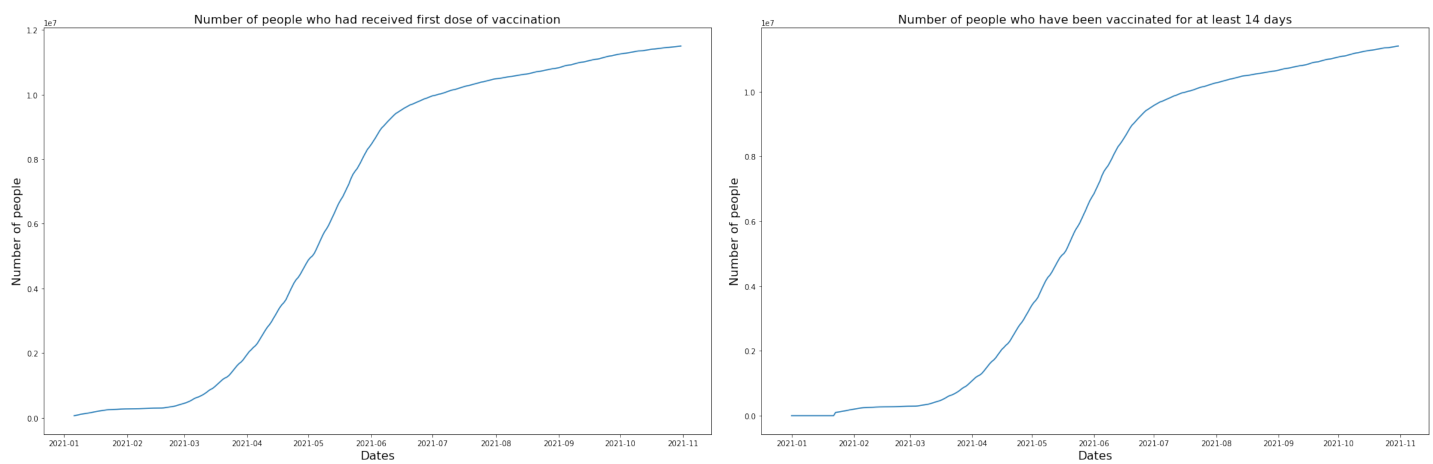


**Figure 3: First dose vaccination data; unshifted (left), shifted (right)**

# Analysis

## Positive Covid-19 cases and Hospitalizations

The data for positive Covid-19 cases and hospitalizations is not cumulative but day to day; hence we see an oscillatory trend (where on some days more people test positive, other days less people ; and similar for hospitalizations). In order to analyse similarity in data trends here, we use cross-correlation. This will show us how one series' oscillation overlaps with the others (i.e. how changing case numbers changes hospitalization data etc). We do this using Numpy's auto-correlate function (i.e. numpy.correlate with mode = 'same'). We use this function directly rather than computation directly from definition as we showed in lab 4 that these are identical. As we only care about comparison of trends, all data is normalized before cross correlations are taken.

For the cross-correlation graphs between positive Covid-19 cases and the different hospitalization data, we except to see a maximum at 0 (close to 1 i.e. max correlation), with the graph decreasing to zero away from the origin. The reason for this is that cross-correlation is a measure of overlap between two time-series. This overlap is computed as one series is shifted across the other (hence the time*lag* axis) The maximum occurs when the series maximally overlap. We expect this peak to occur at t = 0, as these series have similar trends without any shifting (which is what we except; positive Covid-19 cases lead to immediate hospitalizations and increased ICU/ventilator usage).

The cross-correlations are computed and the graphs shown below:

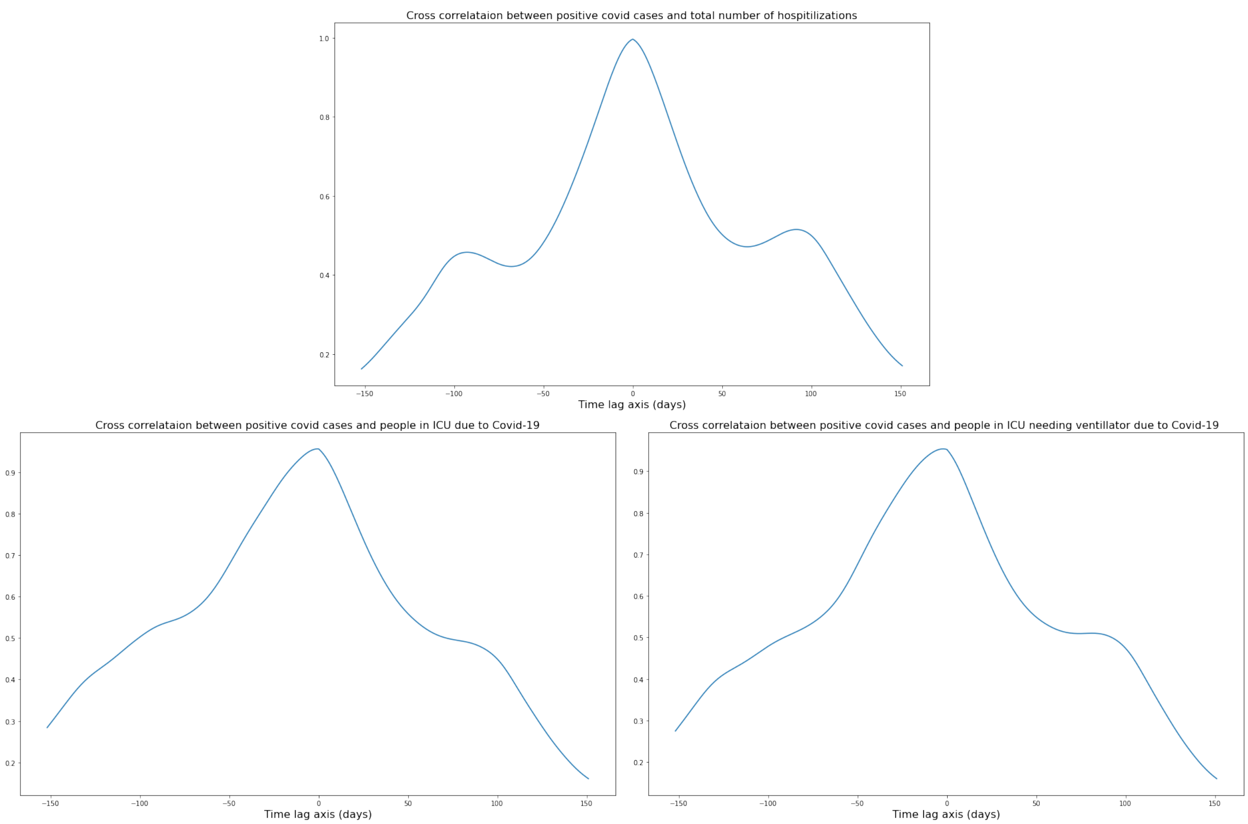


**Figure 4: Normalized auto-correlations of positive Covid-19 cases and hospital data**

As expected, all three graphs have a maximum close to 0 that fades down moving away from the origin. This confirms our hypothesis that rising Covid-19 cases almost immediately leads to increased hospitalizations, as well as increased requirement of the ICU and ventilators in the ICU. It should be noted that there are a total of three peaks in the cross-correlations (most pronounced in the positive Covid-19 cases versus hospitalizations graph). The reason for this is the three spikes of Covid-19 cases we see in our original data; each of these corresponds to an increase in hospitalizations, i.e. there are three times when Covid-19 cases spike up, and hence three times the number hospitalization increases. Because these other peaks are smaller, we see that not only do Covid-19 cases increasing cause in increased hospitalization, but a larger number of cases means a larger number of hospitalization (that is why these peaks are lower; the overlap follows a similar trend but is of different magnitude when the hospitalization data is shifted). It is interesting to note that these extra peaks are not as pronounced in the ICU/ventilator data. This suggests that, while increased cases leads to increased requirement of ICU and ventilators, this is only after a certain threshold of positive cases is reached; that is why there is no maximal correlation when shifting the ICU/ventilator data (the lower spikes in Covid-19 case numbers does not significantly contribute to ICU/ventilator numbers).

## Vaccinations and Covid-19 cases/hospitalizations

Our vaccination data is cumulative (it is total amount of people who have been vaccinated after each given day), and so its time-series is strictly non-decreasing function, unlike the oscillatory behaviour of positive Covid-19 tests and hospitalization data that varies day by day. Thus a cross correlation here will not provide too much info; there is no similar signal shape to observe any delay in overlap between the two (for completeness sake, these cross-correlations are calculated and plotted at the end of this report, but they are exactly as expected). Hence, to analyse the effect of vaccination with positive Covid-19 cases and hospitalization data, we instead plot normalized versions of both sets of data on the same graph, to qualitatively observe trends.

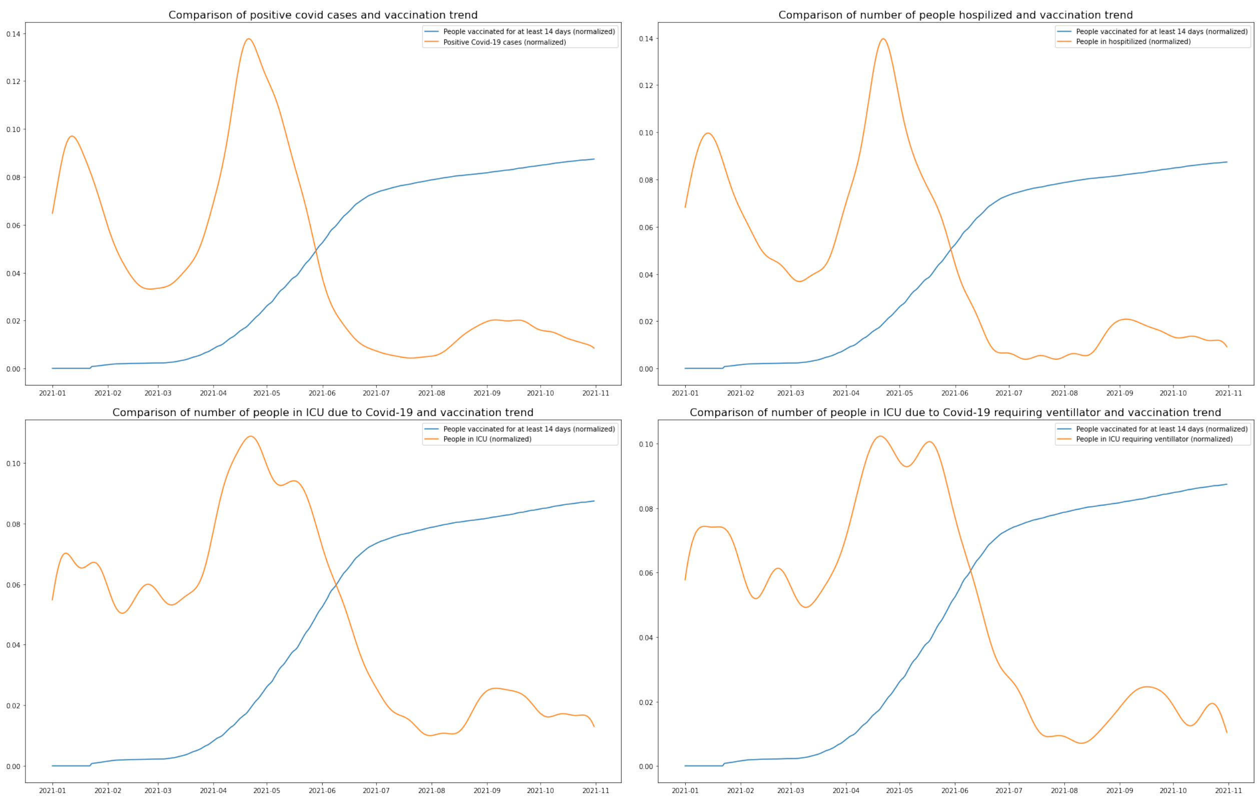


**Figure 5: Vaccination numbers versus cases and hospitalizations (normalized)**

The effect of vaccinations is clear; an upward trend in vaccinations leads to a decline in positive Covid-19 cases and hospitalizations. There is a noticeable delay between ICU and ventilator numbers dropping as vaccinations increase, an effect not observed with the positive case numbers and hospitalizations. After the rate of vaccinations approaches a constant, the Covid-19 cases and all hospitalization numbers (including ICU and ventilator numbers) remain low. There is a small surge in cases, and hence hospitalizations (this correlation observed in the previous analysis) around August 2021, but the maximum of this surge is much lower (about a fifth) of the surge before the vaccination.

# Conclusions

There are a number of conclusions drawn from this analysis. The first being the directly proportional relationship between positive Covid-19 cases and hospitalization numbers, which include requirement of the ICU and ventilator. This analysis did reveal however that surges in ICU and ventilator numbers are not only dependant in surges of positive Covid-19 cases, but require that the number of such cases be beyond a certain threshold before ICUs and ventilators are used in increasing amounts.

Analysis of vaccination data showed a very clear inverse relation between vaccination numbers and positive Covid-19 cases and hospitalizations. Most importantly, the data showed that (at least for a few months) post vaccination, positive cases and hospitalizations do not rise back up to the numbers shown pre-vaccination.

## Limitations and possible improvements

The biggest limitation to this analysis was the lack of data available. As routine VOC PCR testing ceased in November, analysis on different variants of Covid-19 (most notably Omicron) was not possible. It is also unclear how long the effects of vaccination keeping hospitalizations and positive case numbers low lasts. Analysing the effects of the second and third dosage of the vaccine would also be very useful, but as the time these were made available to the public was close to the end of VOC PCR testing, this data was not studied.

# Misc. figures

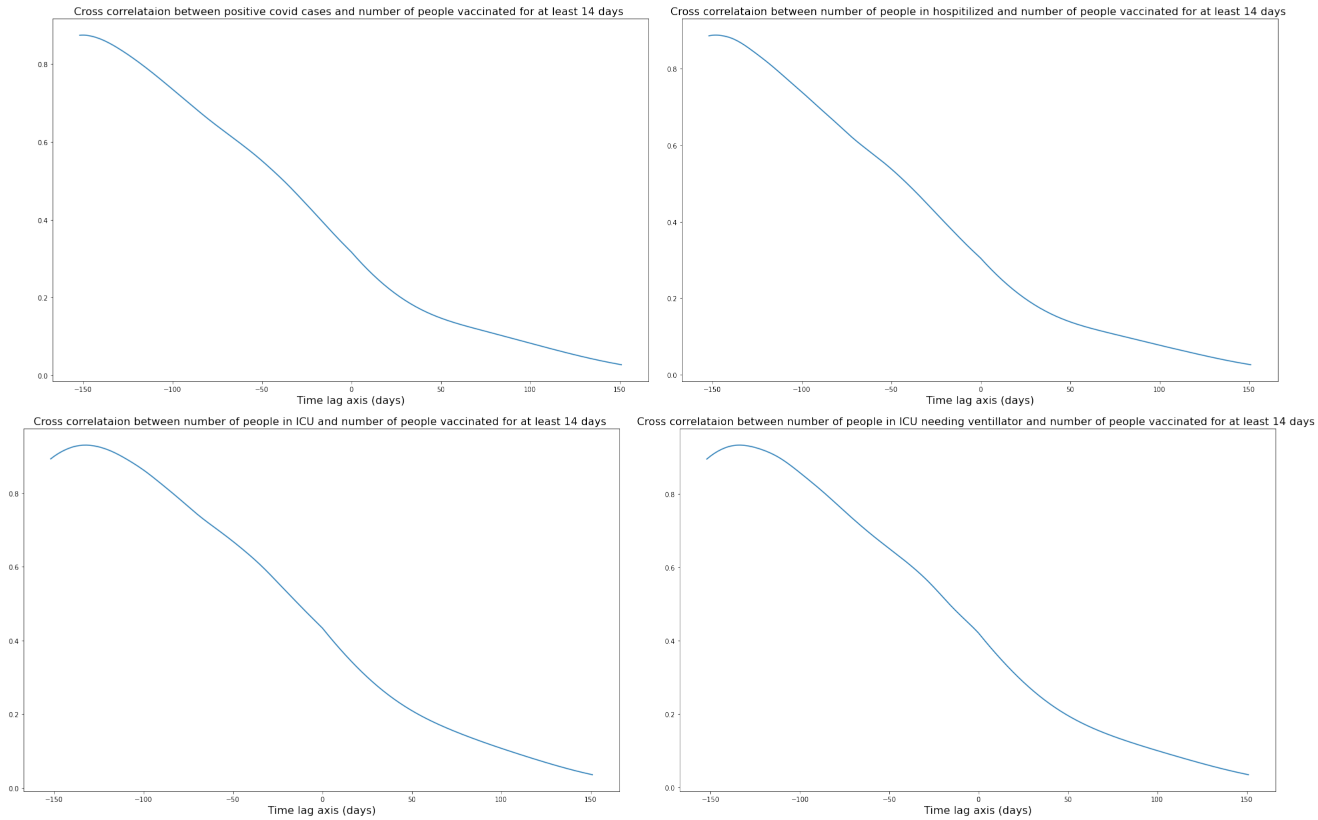


**Figure 6: Vaccination numbers versus cases and hospitalizations cross-correlations (normalized)**

# References

[1] data.ontario.ca. Status of covid-19 cases in ontario. https://data.ontario.ca/dataset/status-of-covid-19-cases-in-ontario/resource/ed270bb8-340b-41f9-a7c6-e8ef587e6d11, 2022. [Online; accessed 19-April-2022].