2020-12-19

Covid 19 Cases in Ontario Public Schools

Kevin Wang	Ziyu Wang	Kathy Zhu
1007099523	1007048874	1007309832
wangke69	wangz965	zhukathy

Table of Contents

Background

As anyone living in 2020 should know, Covid-19 is a newly discovered infectious disease that has created a global pandemic. Having spread to Canada, it is affecting everyone living in it, including young children who go to public schools. Our dataset is a list of over ten thousand schools, complete with the exact number of covid cases in each of these schools, and the location of these schools. By summing up the number of cases per area, a rough estimate of the dangers of sending your child to school can be obtained. For example, a parent living in downtown Toronto may want to forgo in-person schooling for homeschooling, online schooling, or simply a year off since the cases are just so high. While on the other hand, a parent in Clarington may not have to make this decision since there are close to 0 cases of Covid in their area.

Providing an easy to read, accurate visualisation of every city in Ontario would be ideal, but it is unfortunately unrealistic with our available medium.

There are over a hundred cities in Ontario, and it would be unfeasible to graph over one hundred bars on a bar graph, especially when the return would be so little, with most of the minor regions having 0 cases or simply a very low population. That is why only the top 10 most prominent Ontario cities are featured in the visualisation, and the remaining cities are grouped as one. This too is not perfect, as it will soon be apparent. The sum of the remaining cities' covid cases far exceed just the initial 10, heavily skewing the graph, and making some bars extremely small. As a tentative solution, by including two graphs, one with and one without the minor cities of Ontario, to provide a slightly better look. The graphs can be seen below in Results, with only the top 10, (fig 1.) and with all cities included (fig 2.)

Problem Statement

The aim of this project is to provide a visualization tool for determining the correlation between the number of covid cases among Ontario schools and the municipalities they are in. We will identify which municipality has the most school covid cases and predict the trend for future months. With these data we can determine which areas need improvement in its system of containing the spread (e.g increase safety measures), identify which schools need to be closed as well as help help provide an estimate of how much/percentage of the vaccine will need to be allocated to each municipality (or the general surrounding areas).

Solution

High Level:

Our dataset was first a .xlsx file, consisting of 8 columns. In order, they were the date, the date again, the name of the school, the municipality, confirmed student cases, confirmed staff cases, confirmed cases of other types, and lastly, total cases. The total cases is the most important piece of information here for a census, and so we decided to sum up the cases corresponding to each municipality. Then once we have aggregated the data into tangible form, i.e. numbers, we can consider ways of representing the data visually. This first required us to investigate the capabilities of both ourselves and the tool being used, MatPlotLib.

Low Level:

The first step was to convert the numbers in the file and aggregate them, as the numbers in each municipality ranged from around 1 to 2. By summing them up per region, we could get some more interesting information. Using string comparison, for each row entry of column 3, which was the column containing the region of the indicated school, we could add the entry of column 7 of the same row, containing the number of cases. Then store the sum in a list of length 11, with each index representing a region of Ontario, the last one representing a large collection of them. Which index represents which is somewhat ambiguous, only reading the comments can help discern this. Another way of doing this could be with a dictionary's key-value pairs, but I did not choose to do this for this project. After the data has been sorted into a list, all that remained was to graph it. By consulting MatPlotLib's documentation, we were able to learn the syntax required to generate a bar graph of our desired specifications, and so the graphs were created.

The code structure is fairly simple, with a method "read_csv" for reading, and then a section for reading the csv and inputting data into a list, and finally a section to turn the values in the list into a graph using MatPlotLib's imported functions.

Results

With all the schools categorized into their general region, we have determined that the top 10 cities with the highest cases among their schools- ordered from having the most confirmed cases to least- were as follows: Toronto, Ottawa, Brampton, Mississauga, Hamilton, Vaughan, Oakville, Markham, London, Waterloo. The numbers between the first 3 cities had the biggest decrease in change, though it began to slow down continuing down the list.

It should also be noted that, when added together, the total number of cases of these 10 cities made up 43.69% of all cases recorded from all Ontario schools, leaving the other 130 municipalities with 56.31%.

One interesting detail also discovered was that the cities according to their total number of cases was actually not completely aligned with the order of ranking from most to least populated. The municipalities Oakville and Waterloo are also considered part of the more notable anomalies, as they are not even part of the 10 most populated cities in Ontario.

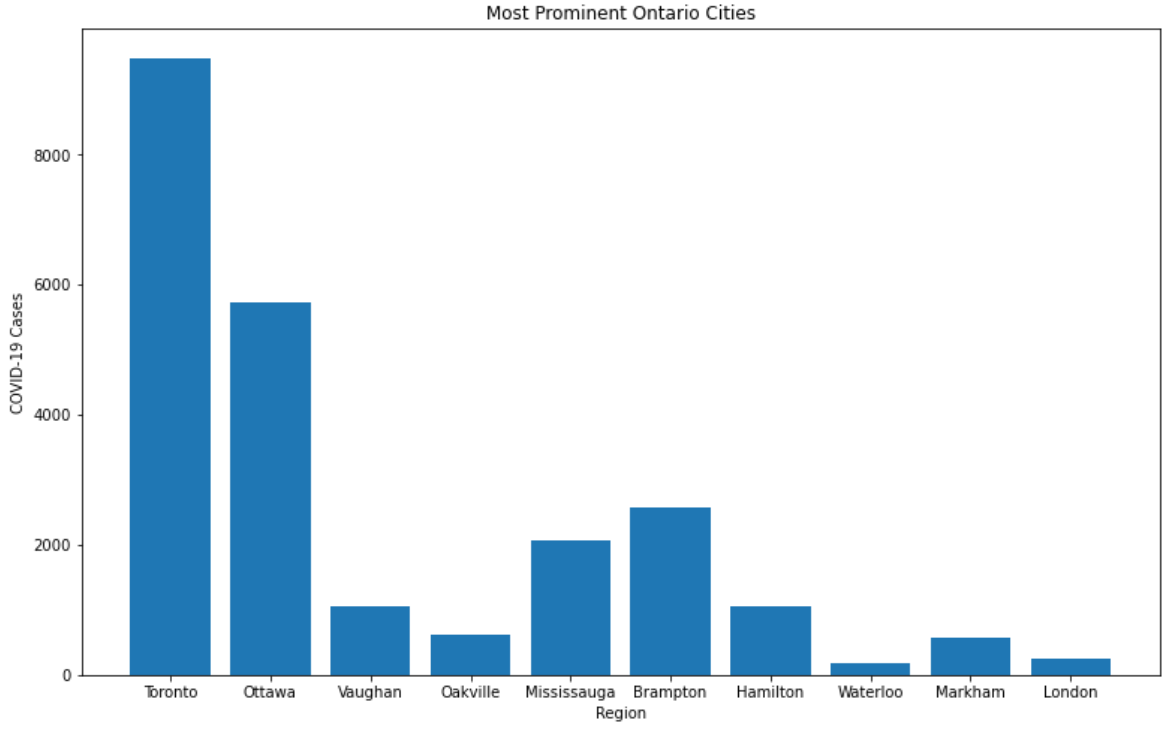


Fig. 1

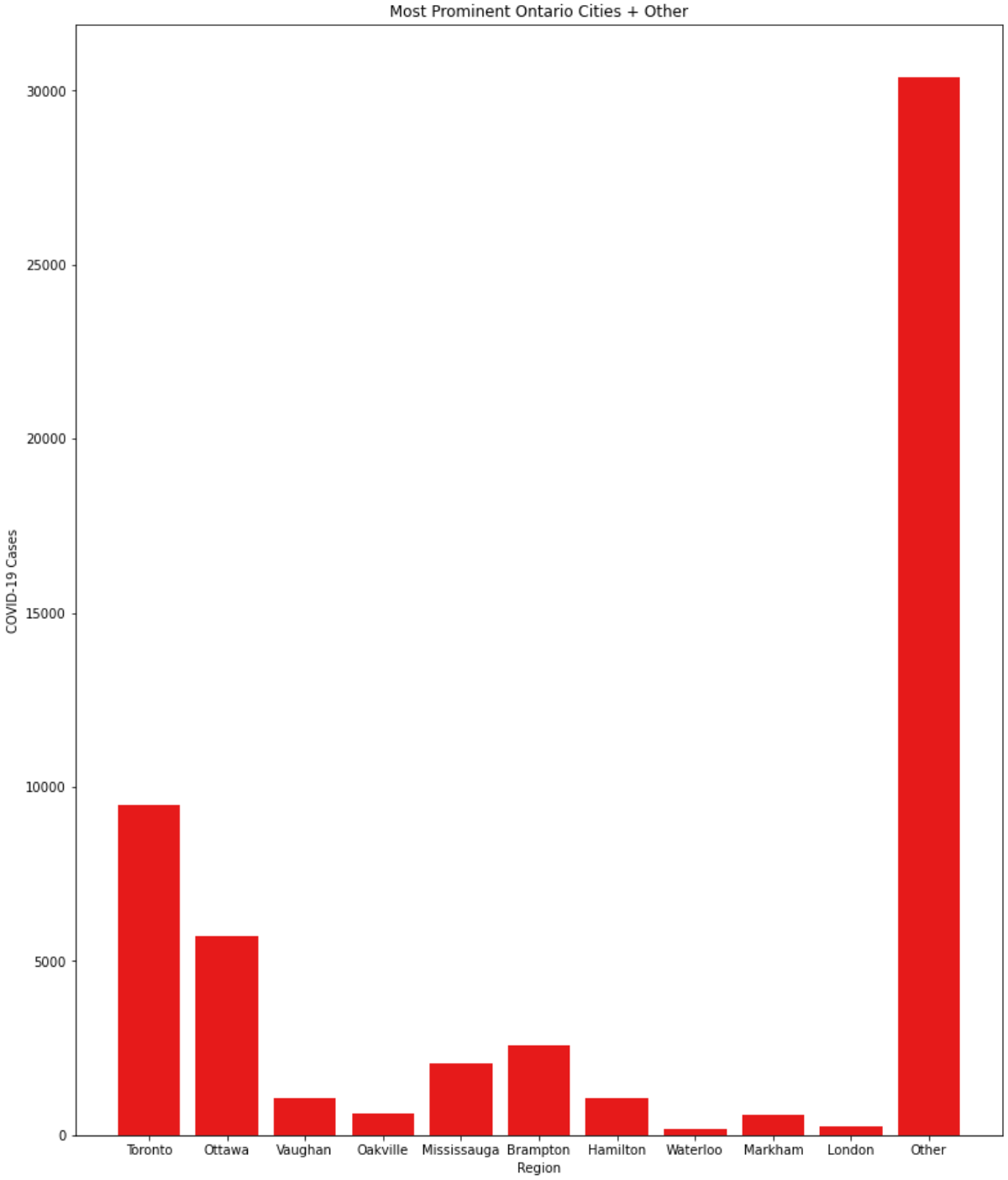

Most Prominent Ontario Cities + Other
COVID-19 Cases
0
5000
10000
15000
20000
25000
30000
Toronto
Ottawa
Vaughan
Oakville
Mississauga
Brampton
Hamilton
Waterloo
Markham
London
Other
Region

Fig. 2

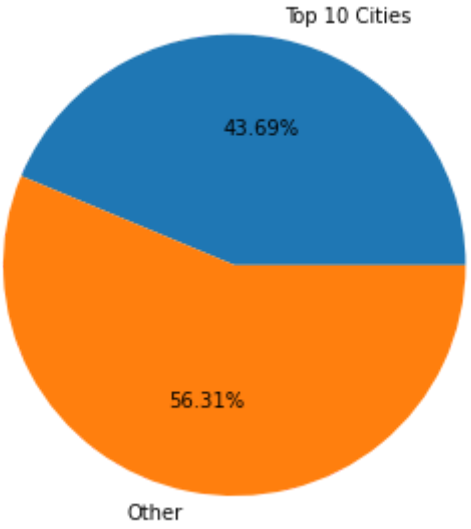


Fig. 3

Challenges and Future Work

Our dataset was initially in a .xlsx file format, which was not readable by the given functions. By saving the file as a .csv however, we discovered that there were further issues due to it being saved as in UTF-8 and was causing errors in python further still. By changing some more settings related to this issue, we were able to resolve it and end up with the fully readable .csv file we now use.

Conclusion

From a data set of the confirmed COVID-19 cases in over 1800 Ontario schools grouped into around 150 municipalities, we have been able to conclude and produce visualizations of the fact that the 10 ten cities with the most cases were Toronto, Ottawa, Brampton, Mississauga, Hamilton, Vaughan, Oakville, Markham, London, Waterloo.

Being from data between the time frame of September 10th, 2020 to November 19th, 2020, this information could really only be used as part of a larger collection of research for predicting future trends within the entire province, as well as providing ideas on how to take the next steps such as placing up new restrictions or allocating the adequate amounts of vaccine to the different regions.

It could also be used for analyzing the behaviour of the school communities within these municipalities. For example, it could be of interest to understand how and why Oakville has had more cases than Markham when it has fewer schools.

At the end of the day, the product from this project can simply contribute as a piece to the bigger, overall picture of understanding how to overcome COVID-19.

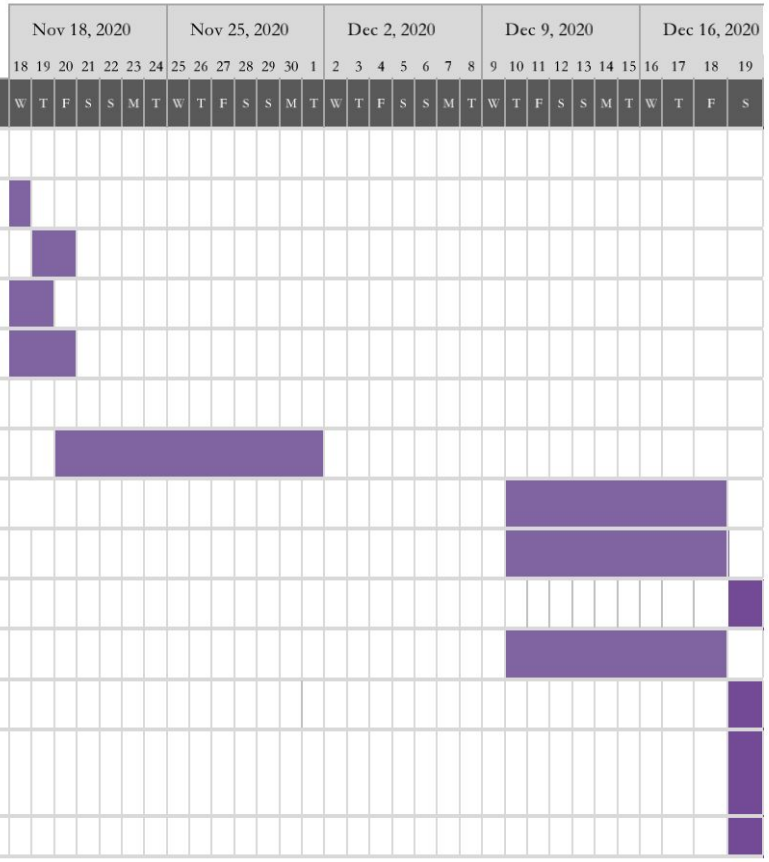
Project Planning and Work Allocation

Kevin: Background, Code to generate visualizations, Solution
Ziyu: Background, Problem Statement
Kathy: Results, Conclusion

CSC104 Project

Project Start: 18/11/2020
Project End: 19/12/2020

TASK	ASSIGNED TO	START	END
Interim report			
Project topic	Ziyu Wang	11/18/20	11/18/20
Background	Ziyu Wang	11/19/20	11/20/20
Data Loading	Kevin Wang	11/18/20	11/19/20
Preliminary Information	Kevin Wang	11/17/20	11/20/20
Final Report			
Graphs	Kevin Wang	11/20/20	12/1/20
Background	Kevin Wang	12/10/20	12/18/20
Problem Statement	Kevin Wang	12/10/20	12/18/20
Problem Statement	Ziyu Wang	12/18/20	12/19/20
Solution	Kevin Wang	12/10/20	12/18/20
Results	Kathy Zhu	12/19/20	12/19/20
Challenges and Future Work	Kathy Zhu	12/19/20	12/19/20
Conclusion	Kathy Zhu	12/19/20	12/19/20



Project End: 19/12/2020

TASK	ASSIGNED TO	START	END
Charts	Ziyu Wang	12/19/20	12/19/20