

# CSC110 Project Report: Impact of Covid-19 on Mental Health and Unemployment

Amirali Tolooei, Vinayak Maharaj, Prahlad Ranjit, Madhav Thenappan

Monday, December 13, 2021

## Problem Description and Research Question

The pandemic saw a massive decrease in the profit margins of many major corporations and the only way that the corporations dealt with that was to fire the non-essential employees and underpay the remaining. This was the case with the majority of the population of Canada during the peak of the pandemic, either out of a job or severely underpaid in their current one.

At that rate, even if the virus did not get them, the emotional overload certainly would.

We've all had to give some of our major exams and had to do our college applications in the pandemic, which was a pretty bizarre experience. Some of us were impacted negatively due to this as the stress and uncertainty of COVID on top of our exams and decisions was overwhelming. We thought that if we found it so mentally draining, what about the individuals who have to earn their own bread and couldn't because of COVID, how much more stressful would that be, how would that affect mental health.

It is from this thought process that our idea originated and we chose to correlate the jobs lost and the wages reduced due to Covid-19 to the mental health of individuals in Canada. We wanted to get a better understanding on how Covid-19 has caused major mental pressure on people, especially because covid caused many different financial struggles in our world. Due to the nature of unemployment being a bad circumstance, we believe that mental health is more greatly affected by it than people may think. The pressure on the finance world caused by covid-19 we believe surely caused a lot of unemployment, thus causing more deteriorated mental health. This is what we want to display within our project.

The research question that my group has chosen is: **How did the jobs lost and the wages reduced due to COVID impact the mental health of individuals in Canada.**

## Dataset Description

We have found three separate data sets to use for our project. One dataset is from Fred Reserve Economic Data called *Unemployment Rate: Aged 15 and Over: All Persons for Canada*. The format of the data is in csv. The Date and LRUNTTTTTCAM156S columns from the csv file were used as the axes for plotting the line graph. Another dataset is from the Government of Canada website called *Public Health Infobase - Data on COVID-19 in Canada*. The Public Health Infobase - Data on COVID-19 in Canada csv file was downloaded from this source. The data was collected from January 31st. The data includes the number of cases per day categorized by province and Canada as a whole. The values displayed in the table are provided by the Public Health Infobase, managed by the Health Promotion and Chronic Disease Prevention Branch (HPCDPB) of the Public Health Agency of Canada (PHAC). The date and numconf columns in the csv file were used as the axes for the line graph. A dataset from Statistics Canada called *Your Mental Health Public Use Microdata File*. The format of the data is in a folder containing datafiles in the forms of ASCII, sas7bdat, sav, and dta. We have, however, provided a conversion file with instructions in the docstring on converting from dta format to csv (which is the data type used in the functions). The basic description of this data set is that it was a crowdsourcing project that collected data from many Canadians on how their mental health was affected by Covid-19 (these are rated in a "ratio" kind of variable called "pumfact"). The columns used were the MH\_5, MH\_10, MH\_15A - MH\_15G, MH\_20, PLM\_30 (which we used as lm\_30 for simplicity sake), and the pagegr (which we used as age\_responses for simplicity sake).

## Computational Overview

We have first extracted different types of data such as *unemployment percentages from March 2020 to October 2021*, *number of covid cases from March 11th 2020 to December 10th 2021*, and *answers to over 9 mental health questions in 2020*.

For *unemployment percentages* and *number of covid cases* we just plot them in a graph with respect to time with pandas and attempt to extract a regression graph from the data as well. With the *mental health data*, however, to correlate it to the 2 former factors, we plotted a multi-column bar graph with pandas with the different age-groups of people being the x-axis and the *mental-health* of *employed*, and *unemployed* people. The *mental health data* is in a scale of 1-14 (1 being the best and 14 being the worst). Our main computations *unemployment percentages* and *number of covid cases* were to extract the data from the .csv files, create a line graph, and a regression slope that could be used in a function to discover the regression model of that graph. For mental health, we have 9 different questions, namely mh\_5, mh\_15a to mh\_15g, and mh\_20. For questions mh\_15a to mh\_15g, we have a 1-4 scale (1 being the best and 4 being the worst) and we to an average of all the responses for the different categories (employed, unemployed) in different age groups. Then for mh\_5 and mh\_20, the both had a scale of 1-5 (1 being the best and 5 being the worst), here we just calculated the score for the 2 categories over the different age groups and added the respective scores to the 1-4 scores that we got from the parts of mh\_15. This gives us a scale of

$$mh_{15} + mh_5 + mh_{20} = 4 + 5 + 5 = 14$$

We have 4 python files, namely main.py, covid.py, mental\_health.py, and unemployment.py. The main.py file is basically the manager file as in it is the interface that the user interacts with and it calls all the other classes and functions from the other files.

In covid.py, we have created a dataclass Covid. The covid-19 dataset, covid-19\_download.csv was filtered in the covid.py file to only include data from the March onward until December 10th 2021 and extracting records for only Canada. The columns for date and numconf were used to make a line graph using the pandas library with number of covid-19 cases on the y axis and dates on the x axis, this is demonstrated after running the plot\_data function in the covid.py file.

In mental\_health.py, we first define some variables variables *mh\_5-responses*(3), *mh\_10-responses*(4), *mh\_15-responses*(5-11), *mh\_20-responses*(12), *lm\_30-responses*(37), *age-responses*(32) to convert the data extracted from the given .csv file into numbers to make them more workable. We then create a dataclass MentalHealth with instance attributes *all\_data*. The *all\_data* attribute contains all the relevant data like the answer to the mental health questions, age groups of the individuals, and employment status of the individuals. We then have major functions to separate the employment data (*separate\_on\_employment*), separate the age groups (*separate\_on\_age\_group*), calculate the average mental health ratio (*avg\_mental\_health\_ratio*). Then finally we have a function to plot and display the multi-column bar chart (*multiple\_bar\_graph\_mental\_health*). We then also have a (*count\_on\_mh10*) which separates the number of people that selected each response within the question (from much better to much worse at 5 levels) and a (*pie\_chart\_mental\_effect*) which correlates the data from the previous function and creates a pie chart out of it comparing the percentages of the different answers.

In unemployment.py, we have 2 functions, *load\_data* and *graph\_data*. The *load\_data* function is extracting the covid percentages which are dated after 3rd March, 2020 or in 2021 from the given .csv file and adding them to a list *employment\_percentages* with the corresponding dates being added to the list *dates*. The function *graph\_data* then plots and displays a line graph with the x-axis as the time frame (in months) and the y-axis as the unemployment percentage.

We are using 2 new python libraries, *pandas*, *plotly*. *pandas* is being used to plot the extracted data (turn the data into comprehensible variables, then turn those into points on a graphs), and *plotly* is being used to actually display the graphs. With *pandas* we were also able to convert the time strings to plotable timestamps, and group different data's to a dataframe (which is used to plot the points). We were able to use *plotly* to make our graphs looks a little cleaner than the normal look by changing colours, font sizes, font distances, and more!

# Instructions

1.Install all python libraries stated in the requirements.txt file submitted on MarkUs

2.Download the dataset files from these URLs:

ICC\_RCC\_MH\_2020\_EN.DTA from <https://abacus.library.ubc.ca/dataset.xhtml?persistentId=hdl:11272.1/AB2/RHP5H5>  
Click "Access Data", then click "original format zip". You will then have a popup. You may read the terms and then click Accept. Once downloaded, extract the folder into an accessible directory. Then enter the "Data" folder and inside extract (*ICC\_RCC\_MH\_2020\_Stata\_dta.zip*). Cut the .dta file and place it in the same directory as all the other python files. You may either convert the file using the provided python file (instructions on the docstrings) or some other way, however, make sure that the data format is in .csv before usage within the functions

LRUNTTTTCAM156S.csv from <https://fred.stlouisfed.org/series/LRUNTTTTCAM156S>

You may either download the file as is or change the start date (WARNING: no matter how much time before 2020-03-01 you download, the file will only run the graph for the time frame between 2020-03-01 and the final available date up to the end of 2021) (this was to accommodate for the covid data only beginning on 2020-03-11 so the graphs could be comparable)

covid19-download.csv from <https://open.canada.ca/data/en/dataset/261c32ab-4cfd-4f81-9dea-7b64065690dc>  
Click download on the (Public Health Infobase - Data on COVID-19 in Canada) link

3. Download all the uploaded .py files on MarkUs TO THE SAME FOLDER (the program will not run if all four of the main files are not in the same folder/directory!)

4. Make sure all the datasets are saved within the same folder as the other python files (or remember the the path to them from the directory you saved the python files to.

5. Run the main.py file, there you will receive the prompt asking for which graph the user wants. It also lists the available options. First the user must input one of the available options. Next the user must input the file location (if all data and files are in the same location, this is entered simply by typing in the data file name .csv. e.g: dataset.csv) for the corresponding graph by entering either the location relative to the main.py file, or by referring to the absolute path. Only entering the unemployment data for the unemployment graphs will work, only the covid data will work with the covid graphs, and only the mental health data will work with the mental health graphs (provided these all must be csv files)

\*It is advised to always maximise the graph window to see the x-axis labels.

\*We can also pan the graph view using the 4-arrow option in the graph window.

**Here are sample screenshots of the possible graphs:**

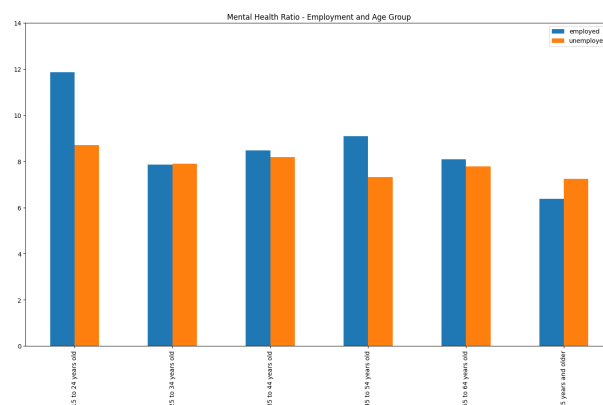


Figure 1: Mental Health Bar Chart

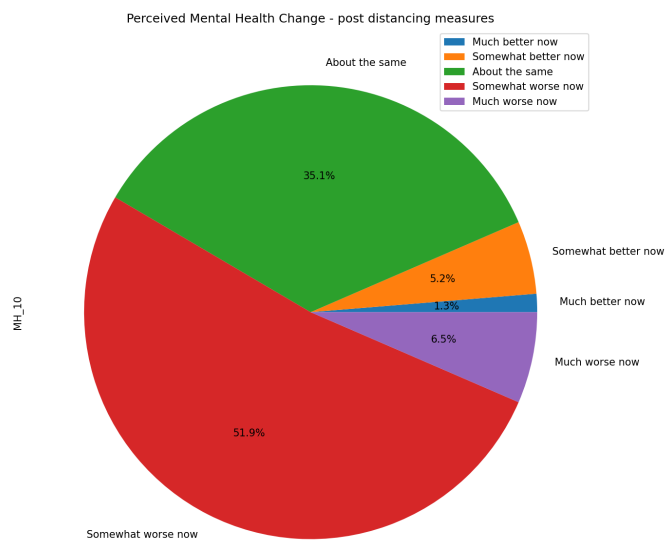


Figure 2: Mental Health Pie Chart

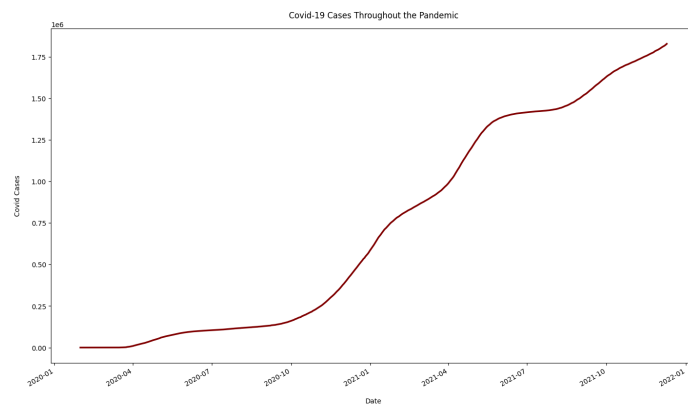


Figure 3: Covid 19 Cases

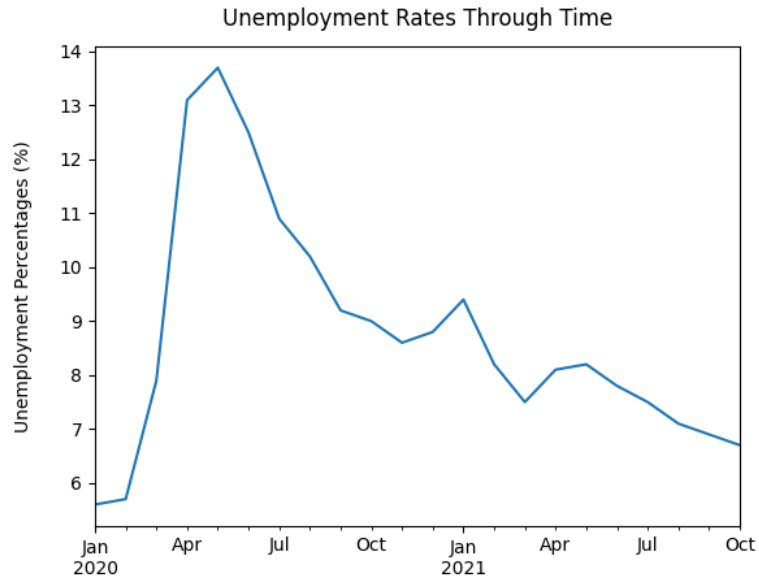


Figure 4: Unemployment Rates

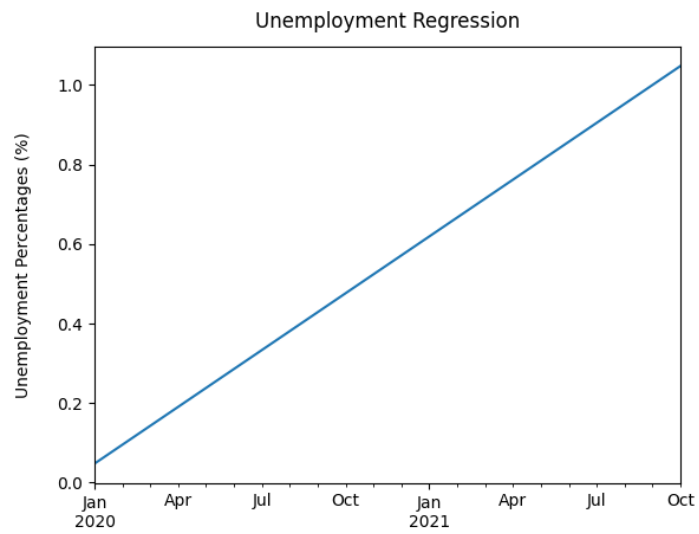


Figure 5: Unemployment Regression

## Changes from Proposal to Final Report

We are now using 3 data sets instead of 2, now also including a data set for covid-19 so that we can plot a line graph of the number of cases throughout the pandemic. We also decided to use more graphs including a regression model for unemployment and a regression model for covid cases, line graphs for unemployment rates and covid cases, pie chart representing perceived mental and a bar graph showing a comparison of the mental health between employed and unemployed persons grouped by age range. This was done due to feedback on the Proposal requesting more graphs and complexity in our analysis. We have also converted the worded responses to numeric values and used that to make the bar chart as it was also suggested in the proposal feedback.

## Discussion

The graphs and data analysis employed for this project did answer the research question but also raised other questions as well. From the Mental Health Ratio bar graph, we identified that employed persons in 2020 had a lower quality of mental health if they are aged 54 years or younger. This can be due to younger persons having to balance having a job and school whereas older persons have less factors that can possibly lead to better mental health.

Firstly, we observe from the Perceived Mental Health Change Pie Chart (data from Jun 2020) that, of the 46000 people surveyed, compared to before physical distancing, 51.9% of respondents said that their mental health is worse, and 6.5% of respondents said that their mental health is much worse now. Thus, we see that COVID-19 has had a huge impact on the mental health of the general public, and for a more in-depth analysis, we look at the responses in the Mental Health Ratio - Employment and Age Groups multiple bar graph. Here, we observed that younger respondents in the range 15 to 44 stand most affected. In an unexpected twist, employed respondents reported worse levels of mental health, measured by the mental health ratio (the higher the ratio, the worse the mental health), compared to their un-employed counterparts. While this stands true for the younger respondents (age 15 - 44), in the older respondents, the unemployed respondents showed worse levels of mental health compared to their counterparts. Hence, with these two graphs we conclude that COVID-19 has had a significant impact on the mental health of the general public, especially the young and employed population. Thus, we establish a correlation of unemployment corresponding to comparative better perceived mental health during the peak of the first wave of the pandemic. These may be related to the increased job insecurity that was already present in the people and consequently, took a toll on the mental health of people. We can make the argument that people with jobs had worse mental health due to the chronic anxiety that they faced everyday due to the fear of losing their job.

Some limitations that we experienced throughout this project included the Statistics Canada website being inaccessible, the covid-19 dataset having missing data prior to March 11th and lack of further information on whether the employed persons in the mental health survey worked at home or in person. Since the Statistics Canada website was inaccessible, we had to use a new source for our data set which led to a smaller pool of data being collected for this project. Since the dataset had data missing prior to March 11th we had to 'clean' the data in order to have a more accurate model to draw conclusions from. Another limitation of this data is that it does not identify what is the makeup of the structure of the population who are employed, whether they are working at home or if they are working in person. This can lead to vast differences with the factors affecting mental health in the employment population for example, persons working at offices have to worry about possibly contracting covid-19 and transmitting it to others whereas persons at home have possible worries like maintaining a work environment while keeping control of external factors like making food for their kids.

In conclusion, while it is true that job losses due to covid-19 have impacted the mental health of individuals, it is not in the way that people predicted, instead persons who are unemployed generally report being less stressed and are comparatively mentally healthier during the pandemic than those who have retained their jobs. Further insight into this statement can be done by doing more research and breaking down the makeup of persons working at home vs in workplaces and whether there is more job satisfaction while working in-person or while working online.

## References

Canada, Statistics. “Crowdsourcing: Impacts of the Covid-19 on Canadians – Your Mental Health Public Use Microdata File, [2020].” Abacus Data Network, Abacus Data Network, 15 June 2020, [abacus.library.ubc.ca/dataset.xhtml?persistentId=hdl%3A11272.1%2FAB2%2FRHP5H5](https://abacus.library.ubc.ca/dataset.xhtml?persistentId=hdl%3A11272.1%2FAB2%2FRHP5H5).

Organization for Economic Co-operation and Development, Unemployment Rate: Aged 15 and Over: All Persons for Canada [LRUNTTTTCAM156S], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/LRUNTTTTCAM156S>, December 13, 2021.

Public Health Infobase - Data on COVID-19 in Canada. ”covid19-download.csv”, <https://open.canada.ca/data/en/dataset//261c32ab-4cfd-4f81-9dea-7b64065690dc>

Pandas Library Documentation. “Chart Visualization.” Chart Visualization - Pandas 1.3.4 Documentation, 17 Oct. 2021, [https://pandas.pydata.org/docs/user\\_guide/index.html](https://pandas.pydata.org/docs/user_guide/index.html)

W3Schools Tutorials. “Pandas Tutorial.” Pandas Tutorial, [www.w3schools.com/python/pandas/default.asp](http://www.w3schools.com/python/pandas/default.asp).

Joshi, Suraj. “Pandas Plot Multiple Columns on Bar Chart with Matplotlib.” Delft Stack, 14 Nov. 2020, [www.delftstack.com/howto/matplotlib/pandas-plot-multiple-columns-on-bar-chart-matplotlib/](http://www.delftstack.com/howto/matplotlib/pandas-plot-multiple-columns-on-bar-chart-matplotlib/).

Friedman, Dan. “Line Plot Using Pandas.” Dan Friedman’s Data & Programming Knowledge Base, 10 Mar. 2018, [dfrieds.com/data-visualizations/line-plot-python-pandas.html](https://dfrieds.com/data-visualizations/line-plot-python-pandas.html).