

CSC110 Project Proposal: The negative consequences of COVID-19 on NAICS categorized industries in Canada

Paul He, Xiling Zhao, Jiawei Yu

Friday, November 5, 2021

Problem Description and Research Question

Since the beginning of 2020, human beings throughout the globe have been co-existing with COVID-19 for nearly two years. Undoubtedly, almost every industry has been affected by COVID to some extent. The purpose of this project is to visualize the effects of COVID on different industries in Canada.

Everything has two sides and the pandemic is no exception. Some industries experience more severe consequences than others, such as commerce, real estate, transposition etc. However, some industries are growing faster than before because of the pandemic. For example, Zoom. Very few people used it before the pandemic, but since 2020, almost everyone has it on their computers. On the other side, due to COVID precautions, construction workers are required to stop working, which lead to many real estates failing to finish their constructions on time, causing these companies to bankruptcies.

Being shocked by the failures of many rising companies which we speculated that would be successful, we want to understand which type of companies are sensitive to the impact of COVID. We want to analyse the extent to which different companies are affected by COVID based on real world data.

We will use North American Industry Classification System (NAICS) to group companies into different industry sectors. In this project, we categorize industries to 20 different sectors, each with a unique code as in NAICS.

This has led us to propose our research question: **Which industry in NAICS endure the negative consequences of COVID-19 in Canada the most?**

We use GDP which stands for Gross Domestic Product to see if an industry is developed or not. Gross domestic product (GDP) is the total monetary or market value of all the finished goods and services produced within a country's borders in a specific time period(Investopedia, 2021). Since the currency inflates throughout the years, it needs a standard value for Canadian Dollars to show the real value of GDP. For this project, the currency will refer to 2012's value.

Dataset Description

1. Dataset 1: Covid

- (a) Source: Health-infobase Canada (Organization)
- (b) Format: CSV File
- (c) Description: A csv file which shows new COVID cases, deaths, recoveries per day from March 2020 to October 2021 in Canada. We just need the data of Canada instead of other provinces. In other words, we only use rows that start with '1' in the first column. Also, the column, we just use the data in 'numtoday' column.

2. Dataset 2: Industry GDP

- (a) Source: Statistics Canada (Government Website)
- (b) Format: CSV file
- (c) Description: A csv File which shows each industry and its GDP for each month between March 2020 to September 2021 in Canada. We will just use the data between 14th row and 49th row.

Computational Overview

1. What our program does

We use two scores to indicate the extent that COVID has affected different industries.

By our observation, nearly every industry was severely damaged at the first wave of COVID: from April to June in 2020. Thus, our first score is to measure the severeness of this damage. We would refer to this impact as a drop-down in GDP for a particular industry in the rest of the report. The higher the first score (Score1) is means that COVID makes the GDP of this industry drop more during the first wave.

Our second score (Score2) is used to measure how fast an industry can recover from the impact of COVID during the first wave. The larger the value, the longer it takes for this industry to recover.

We then sum up each industry's score 1 and score 2, the total score of one industry is used to compare the total scores with other industries'. The industry with the largest score is the industry that is affected by COVID the most.

Notice that we also distinguish 'regular industries' and 'special industries' when we calculate the score 1 and score 2 of industries. When calculating score 1, if the GDP of an industry overgrows its pre-COVID GDP, instead of falling down, we mark it as 'special'. When calculating score 2, if the GDP drops below the industry's pre-COVID GDP, we mark it as 'special'. This is because that we observe that most of the industries' GDP would continue growing after recovering from the first COVID wave.

2. Extract Data

(a) Covid Data

We would store data in a dictionary called `covid_cases`. The key for `covid_cases` is a datetime type value indicating the month. It is of the format (yyyy/mm/01). The 'day' argument does not have any significant meaning, it exists because python requires this format. The value for `covid_cases` will be the total new cases in that month and will be calculated by a function which adds up all the new cases per day for that month.

(b) Industry GDP Data(for a certain industry: a)

We would store the data in a list. To make processing industry data more convenient, we create a new dataclass called `industry`. It has four attributes: name, GDP (a dictionary that maps a datetime.date to an int, which means its GDP value at that month), score1 and score2. Thus, the datas would be stored as a list of `Industry`.

3. Process Data

(a) Covid data

- 1: Extract the covid data and store the data as described in **2.(a)**
- 2: Pass the data to the `main` file.

(b) GDP for Industry a

- 1: Extract the covid data and store the data as described in **2.(b)**
- 2: Pass the data to the `scoring` file. Note that we always solely focus on the 'regular industries' since we would individually analyse the graph of 'special industries' in the **Discussion** section.

(c) Scoring system

For score 1, we use the lowest GDP of this industry during the first COVID-wave to divide its pre-COVID GDP. The resulting value, p , would be used to minus one. In other words, we use $p - 1$ to indicate how much an industry is damaged in the first wave of the COVID. We then take the average of these different p s for different industries, and get our result mean p_{mean} . Then, we divide the p for each industry, to see how an industry performed in the first wave. The resulting value, $\frac{p-1}{p_{\text{mean}}}$, would be multiplied by 10 and this value would be the score 1 for each industry.

For score 2, we calculate the number of days each industry takes to re-achieve its pre-COVID GDP, d . Then we calculate the average time for industries to recover, d_{mean} . Then, we calculate $\frac{d}{d_{\text{mean}}}$ to see how an industry performs at recovery from COVID. The scorer 2 of each industry is 10 times the resulting value.

(d) Visualise our results

Graph 1: COVID-Cases line graph

We use pandas to format our data set into a data frame represented as a table, with column one being the date and column two being the total new cases for that month. This is then used by plotly to create a line graph. The x-axis represents the dates and the y-axis represents the number of cases.

Graph 2: Industry GDP line graph

We use pandas to format our data set into a data frame represented as a table, with column one being the industry and column two being the date and column three being the GDP of that industry of that month. We then use plotly to graph this as a line graph. Since there are multiple industries with their own line graph, there is a list on the right side of the table that allows users to choose which industry to be shown.

Graph 3: Industry Score bar chart

We use pandas to format our data set into a data frame represented as a table. The first column is the industry, the second is Score 1 and their column is Score 2. Plotly will use this to graph into an interactive bar chart. When both scores are selected on the right it shows the sum of the two scores, the user will be able to choose which score to be chosen and plotly automatically scales the bar chart.

4. Visualize data

We use the the library `plotly` to plot 3 graphs.

- (a) The first one is the graph showing the number of new cases by time(monthly) using the dictionary `A`. We use a line chart to present this.
- (b) The second one is the graph showing the GDP change of each industry by time(monthly). Since there are about 20 industries in the dataset, we allow users to select which industry to be shown on the graph. Plotly has a special function for this purpose called 'Line Charts' in 'Dash' that we would use to create this graph.
- (c) The third one is the graph showing the score of each industry. We use a bar chart to present this. The x-axis is the name of the industries, and the y-axis is their scores.

Instructions

1. From Pycharm menu bar, preferences, python interpreter, click the + button and install all Python libraries listed under. This can also be seen in `requirements.txt` with version details. If an error occurs, you might need to update pip to the newest version
`python.ta`
`pytest`
`plotly`
`pandas`
2. Download dataset from the following places:
COVID DATA: <https://health-infobase.canada.ca/covid-19/epidemiological-summary-covid-19-cases.html>
INDUSTRY GDP DATA: <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3610043401>
NOTE: statcan.gc.ca might not respond, if that is the case load data from the files we provided.
3. Name COVID DATA into `covid19.csv`
4. Name INDUSTRY GDP DATA into `gdp_data.csv`
5. Make sure that all the files are in the same folder, in pycharm, mark the folder as sources route.
6. open `main.py` and you should be able to run the file. Three tabs will appear on your browser.
7. You may select / deselect any industries or scores you wish to see. Double click on an industry to show its graph.

Changes

1. Research Question:
Based on TA's feedback, we will focus more on the negative consequences of the impact of COVID-19 on Canadian Industries classified under NAICS.

2. Scoring System:
Now our Scoring System is purely on GDP data because most industries swiftly adapt to COVID. And in 2021, most industries GDP continue to grow after recovering from the first COVID wave.
3. No longer using Dash:
Dash allows us to categorize data and create interfaces allowing the user to choose which category to be displayed. However, the only thing that we want to let our users to choose is the industries to be displayed. Plotly has this feature already.
4. Separate all industries into two groups and analysis separately:
We divide all industries in the GDP data set into two groups: regular industry and special industry. We analyse these two groups separately. We use our project code to analyse the regular industries. We will analyse special industries individually.

Discussion

- Recall our research question **Which industry in NAICS endure the negative consequences of COVID-19 in Canada the most?** Generally Our program solves our research question. From our **the bar chart** we can see that in regular industries, **Arts, entertainment and recreation [71]** endures the negative consequences of COVID-19 in Canada the most which can be verified using **graph 2**. We can see that there is a big drop down during the first COVID wave, also it not recover by September 1 2021. Also, notice that **information and communication technology sector[T017] 4** has the lowest total score so this industry suffers the least from the COVID, which verifies our assumption in the introduction. Notice that **Arts, entertainment and recreation [71]** is only the regular industry that suffers the most from COVID. There might be industries that are marked as special, and who suffer more in COVID.
- Among all the special industries, we notice some special graphs that need more attention. We use our COVID cases graph as a reference to analyse the behaviour of GDP graphs:

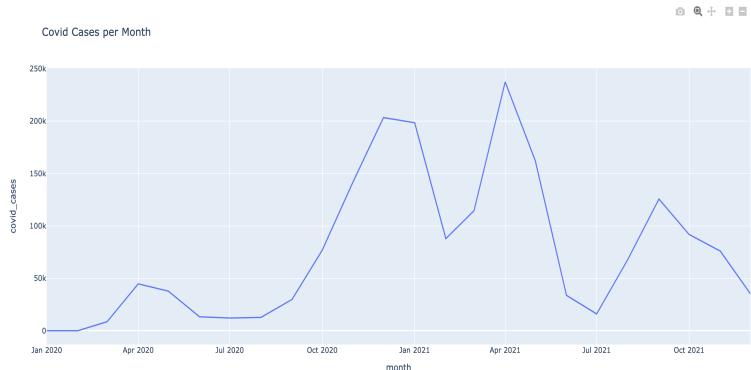


Figure 1: Cases

The first one to be noticed is the graph for Cannabis Sector, we can see that its graph achieves its first complex along with the graph of COVID, this might be due to the serious lockdown caused by COVID. People use Cannabis as a way to release their mental stress.

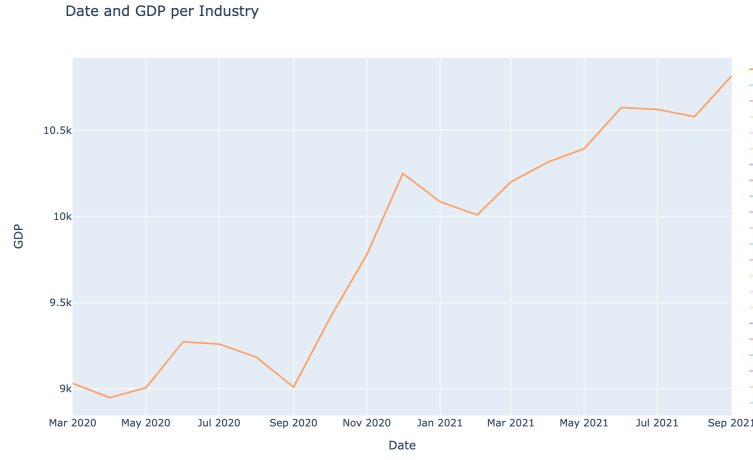


Figure 2: Cannabis

We also notice that the graph of **Finance and insurance** [52] does not have drop down no matter how critical the COVID cases becomes.

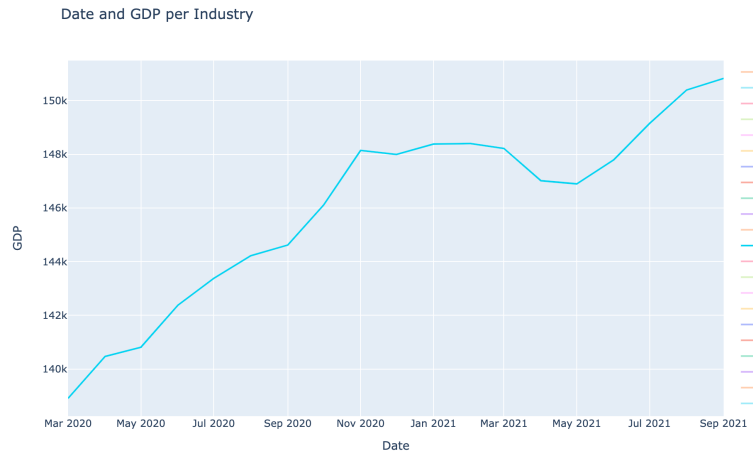


Figure 3: Finance and Insurance

- The most significant limitation of our algorithm is that we might misjudge some industries as special industry.

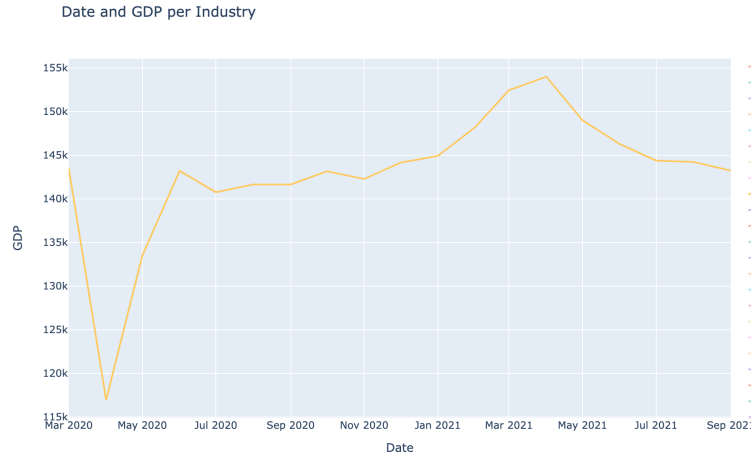


Figure 4: Construction

For example: The graph for **construction** is similar to most graphs of regular industries. However, it is marked as special only because it slightly dropped down below pre-COVID GDP value in July 2020. Similarly, the graph of **Accommodation and food services**, **Accommodation and food services** [72] and **Mining, quarrying and oil and gas extraction** [21] also fits in the shape of other regular industries.

- One possible way to solve this problem is that we can lower the threshold of being marked as special to, for example, 95% of its pre-COVID GDP value. By doing so, only noticeable drop-down would make the industry marked as special.
- There might also be delay effects. For example, an increase in COVID cases in March might cause the GDP drop down in April or even later. Thus, we might take the delay effects into account. A further improvement to our program can be made to allow users to decide what is the time period between an increase in COVID cases and its effects.

References

1. Canada, P. H. (2021, May 28). COVID-19 daily epidemiology update. Retrieved from <https://health-infobase.canada.ca/covid-19/epidemiological-summary-covid-19-cases.html>
2. Fernando, J. (2021, November 3). Gross domestic product (GDP). Investopedia. Retrieved November 4, 2021, from <https://www.investopedia.com/terms/g/gdp.asp>.
3. (2020). Statcan.gc.ca. <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3610043401>