CSC110 Final Project: Evaluating the impact of COVID-19 on Canadians' spending habits

Ajinkya Bhosale, Kowshik Mazumdar, Rafael Gacesa, Raiyan Raad

Friday, November 5, 2021

Introduction

The COVID-19 pandemic has caused severe financial stress for many Canadians. A national poll from October of 2020 reported that nearly 1 in 3 Canadians are worried they may never financially recover from the pandemic (Montgomery). Roughly half of Canadians under the age of 35 have borrowed money from institutions to make ends meet, while roughly the same amount have taken advantage of the Government's emergency assistance (Montgomery). This is largely due to two major reasons.

First, the pandemic drove the unemployment rate as high as 13.7 percent in May of 2020, an all time high, leaving many Canadians with reduced household income or without a source of income at all (Government of Canada, Statistics Canada). Secondly, the price of almost everything has been driven up by the pandemic - for example, Canada's Food Price Report is estimating an increase of anywhere between 3 to 5 percent in the 2021 calendar year largely due to transport difficulties caused by border and facility closures, employment difficulties, and shortages of goods all caused by the pandemic (Charlebois).

Employment shortages have significantly slowed the economy, especially in shipping companies like Fedex, who cited that one example facility was only operated at 65 percent capacity due to labour shortages (Balasubramanian). These types of shortages are causing prices of goods and services to skyrocket (Ziady). Furthermore, rent prices across Canada are continuing to recover after dropping, with September being the sixth consecutive month of average rental price increasing (Landau).

With all of these financial constraints, we wanted to evaluate the direct impact on Canadians - has the pandemic caused Canada to be overly unaffordable for its inhabitants? Have the rising costs of food caused Canadians to sacrifice other expenditures? We seek to evaluate the impact of the pandemic on the spending habits of the Canadian population. **How exactly has the pandemic affected how much Canadians spend and what Canadians are spending their money on?**

Datasets

Each dataset can be downloaded from the clickable hyperlink 'download link' next to each dataset title.

1. Monthly adjusted consumer expenditure basket weights: download link (rename to: weighted-baskets_dataset.csv)

- This data was sourced from Statistics Canada (csv format). It shows the fluctuations/growth of consumer expenditure in various products that vary from food, clothing, transport, internet and even more. For the purpose of this project we will be focusing in on the changes that take place for food items particularly.

- From this dataset, the first column containing the date, the fourth column containing the name of the commodity, and the tenth column containing the value are used. The rows which contain relevant months (specified inside the program) are used to produce visuals.

2. Consumer Price Index, monthly, not seasonally adjusted: download link (rename to: consumer-price-index_dataset.csv)

- This data was sourced from Statistics Canada (csv format). Shows the growth in goods and services pricing compared to 2002, where 2002 is 1.0 (ratio)

- From this dataset, the first column containing the date, the second column containing the region, the fourth column containing the name of the commodity, and the tenth column containing the value are used. The rows which contain relevant months (specified inside the program) are used to produce visuals.

3. Epidemiologic data on the COVID 19 Outbreak in Canada: download link (rename to: covid-19_dataset.csv)

- This data was sourced from Statistics Canada (csv format). We will use epidemiological surveillance data to show how the COVID-19 situation is evolving in Canada and how it correlates to consumer expenditure.

- From this dataset, the first column containing the region ID, the fourth column containing the date, and the sixteenth column containing the daily number of cases are used. The rows which contain relevant months (specified inside the program) are used to produce visuals.

4. Canadian unemployment data: download link (rename to: unemployment-rate_dataset.csv)

- This data was sourced from the Federal Reserve Economic Data (csv format). We will use monthly unemployment data to correlate the effect Covid-19 has had on unemployment and the correlation of it towards the price index.

- From this dataset, both of the two columns, containing date and unemployment rate are used. The rows which contain relevant months (specified inside the program) are used to produce visuals.

Computational Overview

The lowest-level code is contained in the module `backend.py`, which directly interacts with the datasets. The main function of the module is `load_data()` which transforms the raw data into a list of `Month` objects, which is a dataclass defined inside of the module. This dataclass contains the data for one month from the four .csv datasets. The computational highlights of this module are the four transformation functions, one for each dataset, the CSI computation method, and the prediction method.

The four transformation functions read in raw CSV data in nested lists and produce filtered values based on the provided date range. CSI, our Consumer Spending Index, is also computed in this module. The index is a composite of Statistics Canada Consumer Price Index and spending weighted baskets, which provide data on the relative cost of goods and the relative proportion of Canadian spending on each commodity. By creating a composite of these numbers, as seen in the `compute_csi()` method, we can track exactly how much Canadians are spending relative to each month.

The `load_data()` function also contains a flag which allows for the filling of missing CSI values using prediction, as the Statistics Canada data is incomplete. This is done by the `complete_dataset()` method in the backend module, which mutates the main data list using predictions generated by Predictor objects which are created in the method. These Predictor objects are defined in `prediction.py`, which makes use of sklearn's LinearRegression objects to take data from the backend module and produce predictions, as seen in the `generate_predictor()` and `make_prediction()` methods of the Predictor class.

- some globals were required for use (constants in graph module and pygame related in visualization)

Instructions For Use

1. Collect each of the modules, images, and the requirements file and place them in the same directory.

2. Install each package listed in `requirements.txt`.

3. Download each dataset from the provided links in the Datasets section, and rename them as specified. (Note: when downloading from Statistics Canada, make sure to extract and rename the 18100XXX.csv file from the .zip and not the metadata). Place the datasets in the same directory as the modules.

4. Run `main.py`. It should take anywhere from 15-30 seconds to initialize and load the data.

5. Once initialization has completed, you will be prompted in the console to enter a date. Do NOT interact with the pygame window until you have entered start and end dates for the data you wish to analyse into the console in the specified format.

6. Once you have entered the dates, you should see an interactive panel appear with options for creating a graph!

7. (Optional) In the main menu, you can change the colour theme of the panel using the button in the top left. You can also read more about the application in the 'About' section.

8. To use the interactive panel to make graphs, click 'Graph'. (Note: If on MacOS or Linux, you must use right click to interact with the GUI. On Windows, left click works fine. This is a compatibility issue with pygame.)

9. Select which metrics you would like to be included on the graph (plotted against time), and press 'Submit', or 'Animate' if you would like an animated graph. (Note: if plotting the subcategories of CSI, a bar chart will be created instead as that is a more appropriate visualization)

10. A browser pane should open with the requested graph.

Notable Differences Between the Proposal and the Project

There were several differences between the proposal and the main project. First, using TA feedback from the proposal, the sklearn LinearRegression model was implemented to fill in gaps in our data, as data was missing for crucial months (February to April 2021) for the weighted baskets spending data. Secondly, data normalization was implemented using methods from the sklearn library and the pandas library, to make graphs more useful. No sorting, searching, or recursive algorithms were used (sorry!). Lastly, no table of values was outputted as through group discussions we decided that a table of values added little benefit to analysis - there are far too many datapoints for a table of values to be readable.

Discussion

The beginning of this analysis section will observe the produced graphs and discuss visible trends, attempting to find out exactly how Canadians have changed their spending habits over the last 18 months. The most notable graphs include the CSI breakdown, which shows each spending category and how it changed over time, the unemployment graph, and the COVID cases vs total CSI graph.
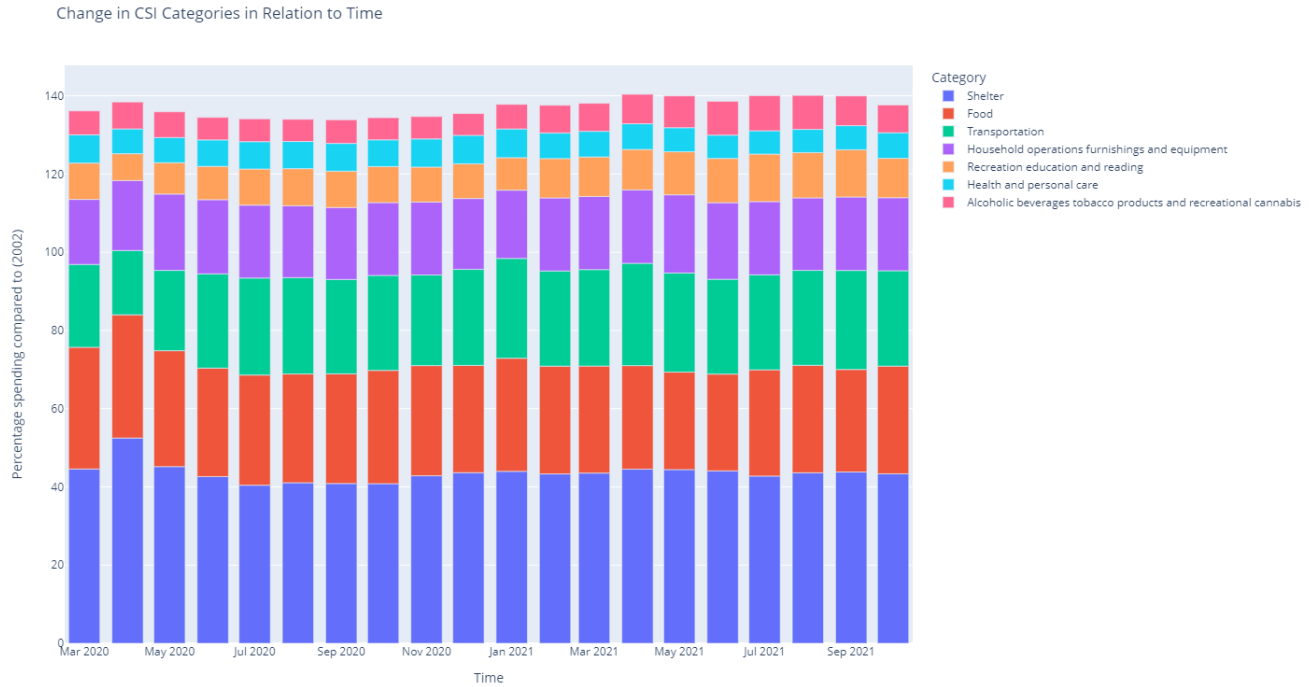
Figure 1: CSI category breakdown from March 2020 to October 2021

First and foremost, in Figure 1, it can be seen that Canadians spent within the same 5-point window throughout the pandemic, from a CSI of 135 to 140. Notably, as seen in Figure 2, there is some correlation between CSI and COVID cases, when normalized on a scale of 1-10. The initial wave of COVID cases in March and April of 2020 caused quite an uptick in CSI, as well as future waves in September through January and March through May. Generally, there is most certainly a correlation between COVID cases and Canadian spending, demonstrated by the increase in spending during each COVID 'wave'. A further analysis could attempt to find causation.
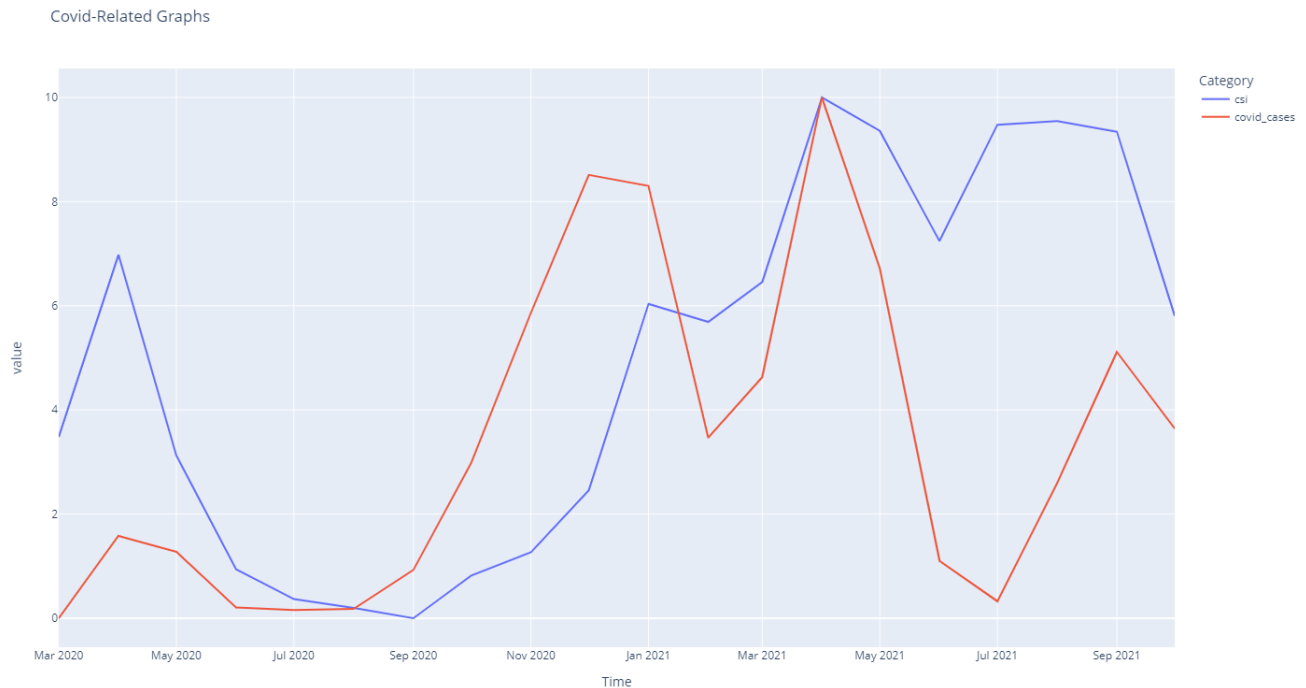
Figure 2: CSI and COVID cases from March 2020 to October 2021, normalized on a scale from 1-10

Some interesting trends can also be picked out from the specific categories. For example, spending on alcohol, tobacco, and cannabis increased by a maximum of 33 percent, which is a significant difference. It can be hypothesized that this is due to the stress imposed by the pandemic, with more Canadians resorting to substances. Other interesting trends are an overall 3-point increase in recreation and educational spending, and a 2-point increase in household spending, which could be a result of the increased amount of time spent inside due to the pandemic.
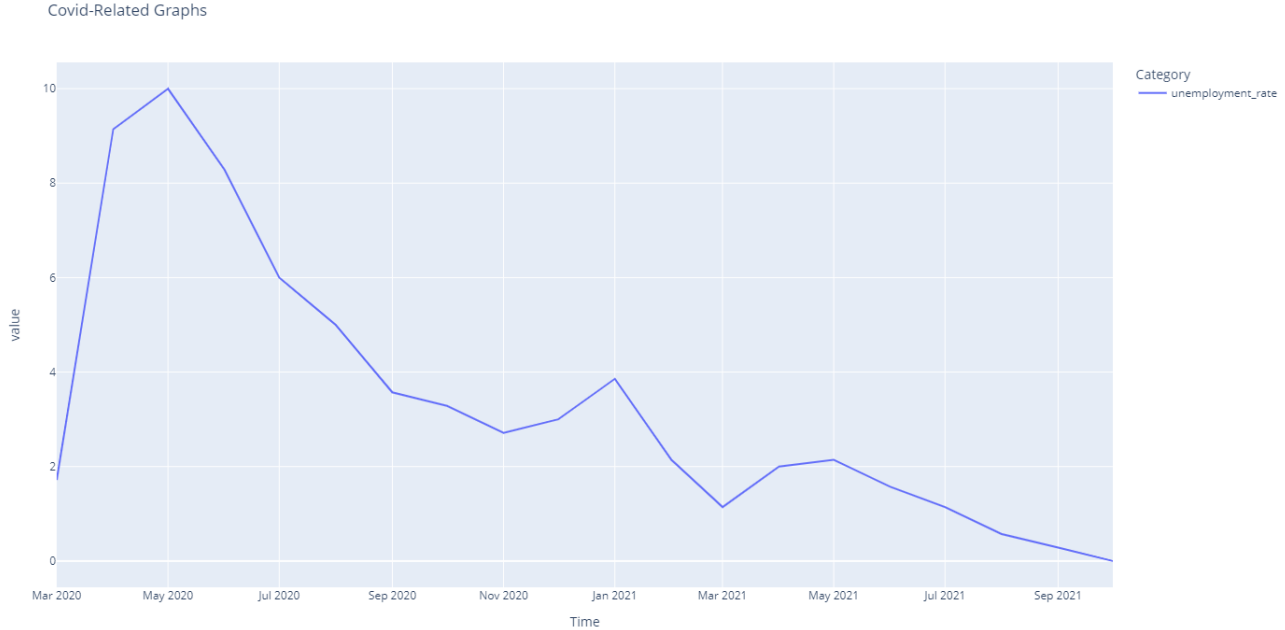

Figure 1: Unemployment rate from March 2020 to October 2021, normalized on a scale from 1-10

One last notable feature of the plots are the immediate changes in CSI for many categories from March to April of 2020, the initial onset of the pandemic. The largest difference in spending can be seen in the Shelter category, with a near 8-point increase. This could be due to the large spike in unemployment rate during the initial pandemic onset as seen in Figure 3, causing many to move into more affordable housing - perhaps a spike due to more security deposits or upfront payments occurring in April 2020. A similar trend can be seen in food spending, with a large initial uptick, perhaps due to supply chain issues or panic buying. Lastly, an interesting initial 3-point decrease can be seen in recreational and educational spending, possibly due to a lack of disposable income (caused by the unemployment uptick as seen in Figure 3).

However, it is impossible to draw any scientific conclusions from the data presented in this paper, besides empirical statements. This is because causation cannot be drawn from correlation - while we can discuss the changes in CSI and specific categories, the multi-variable nature of the data makes it impossible to assume COVID as the cause. This connects with another one of the limitations we faced with the project - the sklearn LinearRegression model used to fill in the gaps of missing data cannot accurately predict CSI values, as it assumes total correlation between CSI and cases. This assumption of causation is inherently wrong; as stated earlier, spending and prices of goods rely on thousands of variables, and require a much more complex model than the linear regression we used in the sklearn library. Lastly, the general lack of data is another major limitation we faced during the project. Consumer basket data pre-March 2020 would be invaluable to our analysis - it would provide a true control case of pre-COVID Canadian spending.

While we faced severe limitations, they do not render the results of our work completely useless. The raw data we present in the graphs provide many further research questions, which could attempt to find actual causation. For example, we noted an increase in spending on recreational cannabis and alcohol - is the pandemic truly causing Canadians to resort to substances? This could lead to correlations with mental health and substance abuse. Many similar further research questions of interest can be generated from our data analysis.

References

Balasubramanian, M.D. Covid-19 Continues To Propagate Dire Shortages Across Multiple Industries. *Forbes*, Sept. 2021. https://www.forbes.com/sites/saibala/2021/09/24/covid-19-continues-to-propagate-dire-shortages-across-multiple-industries/..

Charlebois, Sylvain. Canada's Food Price Report 2021. *Dalhousie University*, Dec. 2020. https://www.dal.ca/sites/agri-food/research/canada-s-food-price-report-2021.html.

For Economic Co-operation, Organization and Development. Unemployment Rate: Aged 15 and Over: All Persons for Canada. *FRED*, Oct. 2021. https://fred.stlouisfed.org/series/LRUNTTTTCAM156S.

Formazione. pygame button: A class to make buttons within pygame. *GitHub*, Sept. 2020. https://github.com/formazione/pygame_button.

Government of Canada, Statistics Canada. *Labour force characteristics, monthly, seasonally adjusted and trend-cycle, last 5 months*, Oct. 2021. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410028701.

———. *Monthly adjusted consumer expenditure basket weights*, Apr. 2021. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1810026401.

———. *Consumer Price Index, monthly, not seasonally adjusted*, Oct. 2021. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1810000401.

Landau, Jack. Toronto area rent prices just keep skyrocketing further from attainability. *blogTO*, Oct. 2021. https://www.blogto.com/real-estate-toronto/2021/10/toronto-area-rent-prices-just-keep-skyrocketing-further-attainability/..

Montgomery, Marc. COVID causing major financial stress for Canadians: surveys. *RCI*, Oct. 2020. https://www.rcinet.ca/en/2020/10/28/covid-causing-major-financial-stress-for-canadians-surveys/.

Of Canada, Public Health Agency. Coronavirus disease (COVID-19): Outbreak update. *Canada.ca*, Sept. 2021. https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection.html.

Plotly. Displaying Figures. *Plotly*, https://plotly.com/python/renderers/.

Ziady, Hanna. The perfect storm making everything you need more expensive. *CNN*, May 2021. https://www.cnn.com/2021/06/09/business/rising-prices-inventories-post-pandemic/index.html.