

## GROUP (Team: JOBS) PROJECT REPORT

### Group Members:

Jamie Barnes  
Oscar Toohey  
Bach Vu  
Seth Gunn  
Angus Loader



**Topic Selection:** COVID-19

**Project Question:** Has COVID-19 impacted the cost of living in New Zealand?

### **Project Summary:**

This report is meant to show how we collaborated on the DATA201 group project. Originally, our idea was to look at the effects of covid on tourism number and the great walks bookings, however, this fell through as DoC keeps almost all of their data private, apart from some basic surface-level information. This made it nigh on impossible to scrape the data from them so we abandoned the idea. After feeling the effects of COVID, we decided to look at the impacts of covid on the cost of living due to the recent inflation bubble that we are currently experiencing as well as the minimum wage increase that we've had in the last year from \$18.90 to \$20.00. This report will show the data sources we used and why we used them.

### **What data sources we used:**

We used 3 datasets in this project. They were COVID-19 case numbers, Imports and exports of New Zealand (cost and tonnage), and the Consumer Price Index(CPI) Indicators. The case numbers and imports/exports of New Zealand datasets were wrangled from the Statistics New Zealand Application Programming Interface (API) and exported as CSV. The Consumer Price Index dataset was downloaded as a CSV from the Statistics New Zealand website.

Data Source Links:

<https://api.stats.govt.nz/>

<https://www.stats.govt.nz/>

### **Why did we choose these data sources:**

After seeing COVID case numbers on the news and seeing how it was affecting New Zealand, we wanted to find out where we could acquire official data that was accessible for the public. This was located on the Statistics New Zealand website. We know that this is a reliable source as they are New Zealand's official data agency. This means we are confident regarding the validity of the data we collected. We decided to directly download the CPI (Consumer Price Index) data as a Comma Separated Values (.CSV) and used the Statistics New Zealand API for the COVID case and Shipping data. This was done to show two different ways of acquiring data from the same website.

### **What target did we choose (i.e., what is the intended use of the data, ...):**

We intend to use our data to see what effects covid has caused on households in New Zealand. We hope our data could help improve future government spending as our national debt, which has increased by almost 100% during the duration of covid, is at 127 billion and continues to climb.

### **What difficulties we had to overcome to wrangle the data sources into the target data model:**

The problem we faced with the Stat's NZ data was that it was so comprehensive, when we first looked at it the data was very confusing. For example, for covid cases there were 3 different measurements for the number of cases which sounds simple but can be very difficult when the data comes in a list and no explanation.

Another issue we had with understanding the data was the information about the data was very limited. They don't have any explanation for the columns or units of data which makes it hard to wrangle and graph. We also had missing days in our data to overcome this we averaged out the week to get a more accurate number, this was also not an issue because most economic data is done in quarters so the data doesn't have to be that accurate.

The graphs were very hard to interpret for importing and exporting numbers. So we countered this by doing an 8 month rolling average. This helps to show trends related to economic downfall due to covid but still accurate enough to not obscure the data from the reality of covid.

### Techniques we used:

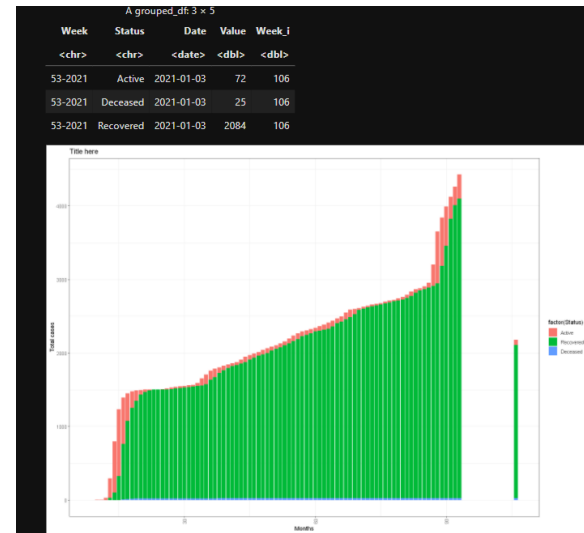
We used the StatsNZ api to scrape the covid-19 and shipping imports/exports data.

We used Github to keep track of all of our files so that we could easily collaborate on multiple computers. Github is a version control system that accommodates multiple users working on the same project. We used a separate branch for each person so that when someone committed their changes it wouldn't affect the main branch, the main branch would then be merged with all of the separate branches once someone was done editing their branch.

We used Trello to keep track of the tasks we needed to complete. Trello is a web service application that assists in keeping track of tasks that need to be completed for a project. Multiple people can assign themselves to a task and then mark when it is complete.

### What we managed to achieve and what you failed to do:

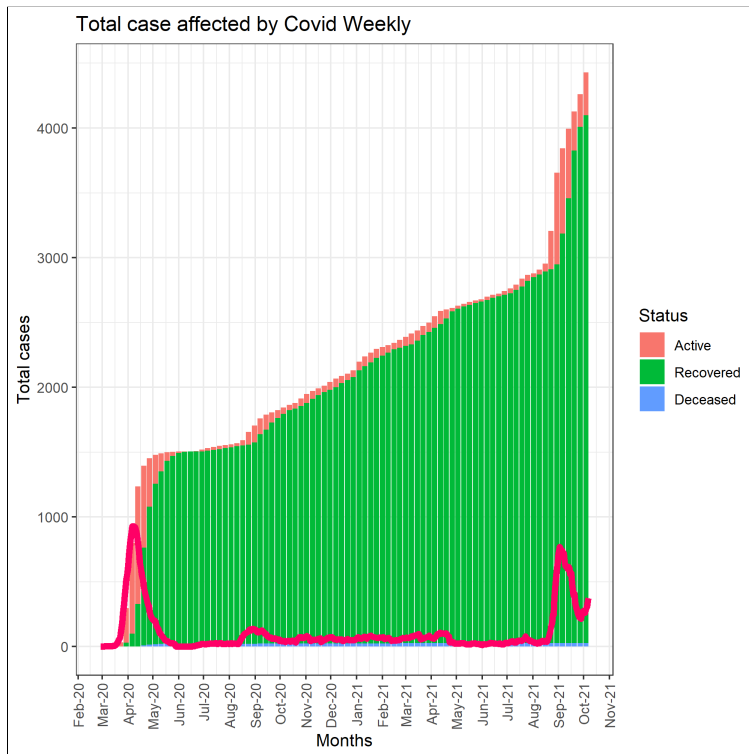
- Failed with Great Walks prototype due to the fact that DoC are very private with their data post 2017 and because the StatsNZ api wouldn't let us scrape the immigration data properly as we were unable to access the correct pages through the api, and the information was not available on other web pages.
- Failed conversion by R (expecting 01-2021) - removed outlier column, only Sunday in dataset where daily covid cases were published. Maybe done to provide people covid information especially around the New Years period.
- We successfully scraped the StatsNZ api for our Covid and Shipping data sets.
- We successfully found an accurate NZ CPI data set.
- We successfully wrangled our data so that it could be easily compared.
- We linked wrangled data's separate csvs by their dates so that they could be easily compared.



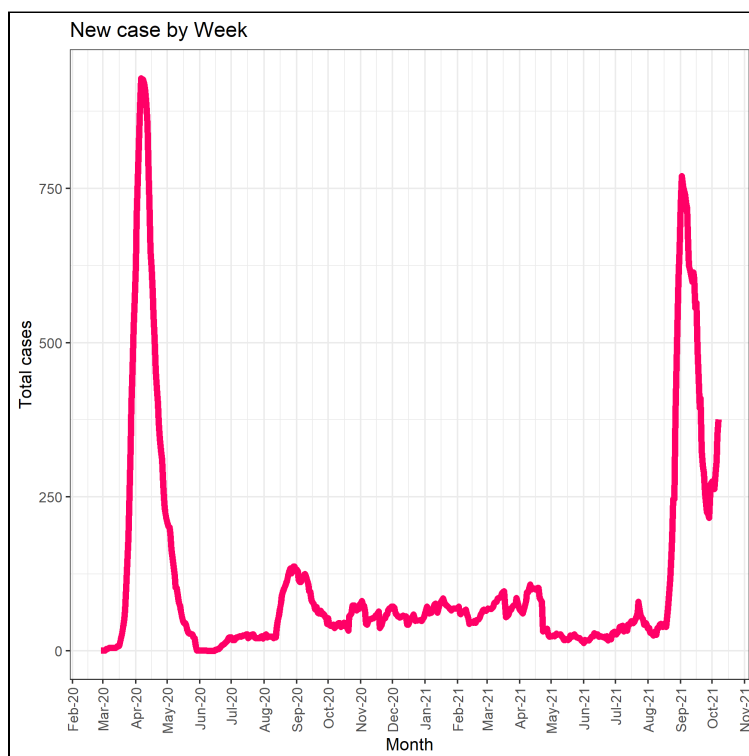
## Results:

The following graphs were made in R (ggplot) using the data that we wrangled.

### COVID Graphs:



The graph to the left is a stacked bar chart with time on the x-axis (months) and the number of COVID-19 cases on the y axis. The red bars show weekly the number of active cases of COVID-19 in New Zealand, the green bars show the number of cases that have recovered, and the blue bars show the number of COVID-related fatalities. The pink trend line shows the number of new cases per week. It is good to see that New Zealand hasn't seen many deaths in proportion to the thousands of people that have contracted COVID. You can also see how with this delta variant that there are becoming more and more active cases compared to those that have recovered.

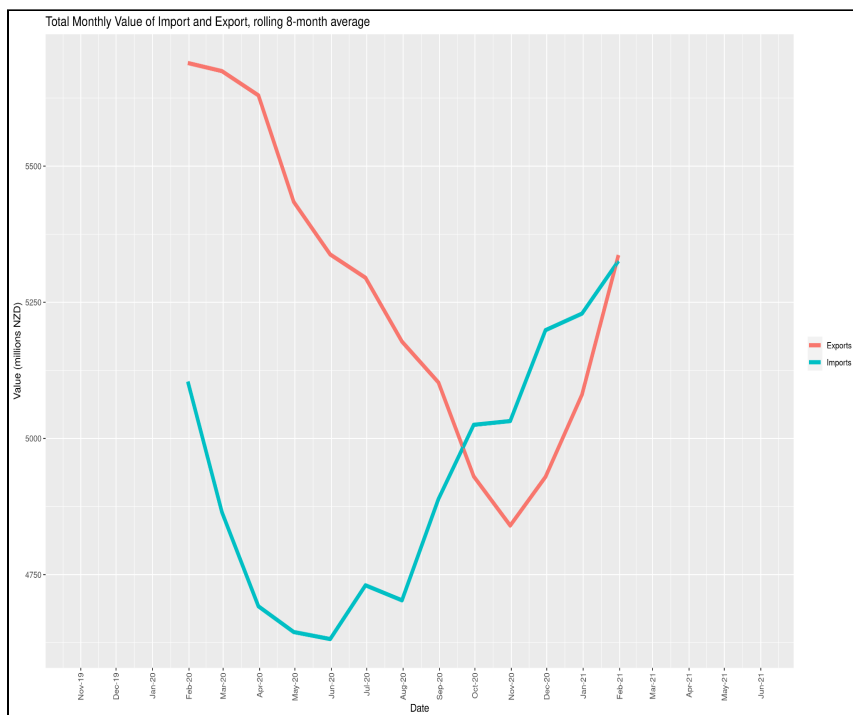


The graph to the left is a line graph showing the number of new cases per week with time on the x-axis (months) and the number of COVID-19 cases on the y axis. There are two major peaks, one at the start of the pandemic in 2020 and the Delta variant outbreak in 2021. Currently, the line is very steep as daily covid cases are increasing as we speak.

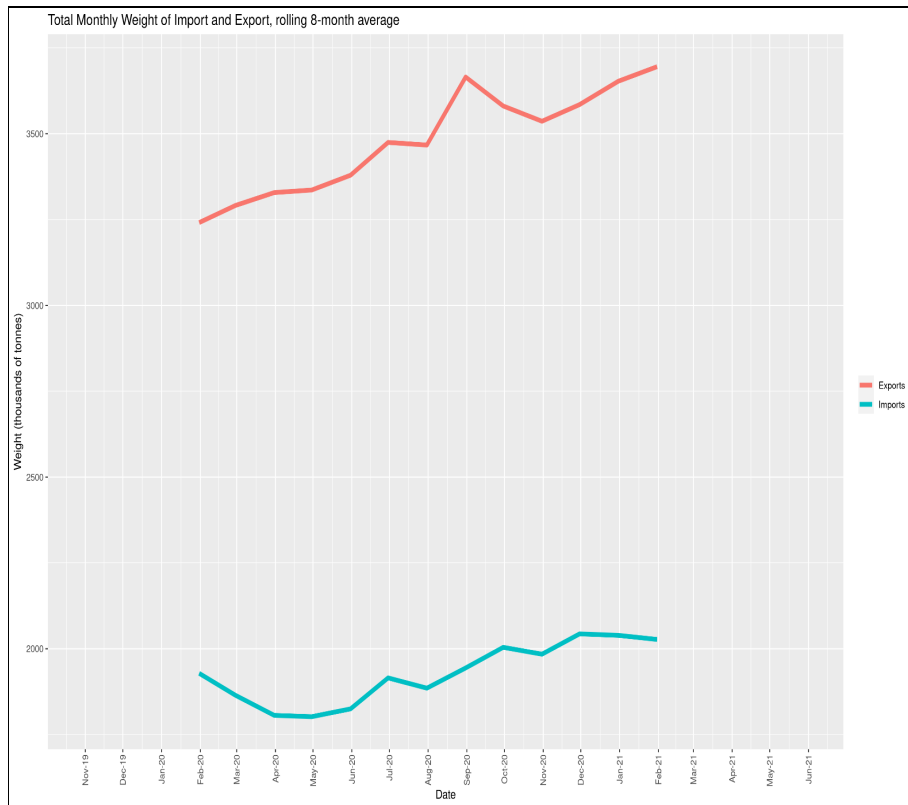
## Overseas Cargo Graphs:



The line graph to the left shows an 8-month rolling average for exports (red) and imports (blue) value for New Zealand cargo with the date (year) on the x axis and value in NZD millions on the y axis. The scale of the value is in millions which means for example 4500 million is in fact 4.5 billion. This will become important when plotting the combined graphs. This data starts in mid-2015 and mainly shows a strong upwards trend until 2020 (start of COVID) where there is a sudden drop in export (~5750 million down to ~4875 million) and import (down to ~4625) value. Even though import and export values have increased since COVID they are not at the level seen since the pre-COVID peak. You can definitely see that COVID has affected imports and exports value for New Zealand.

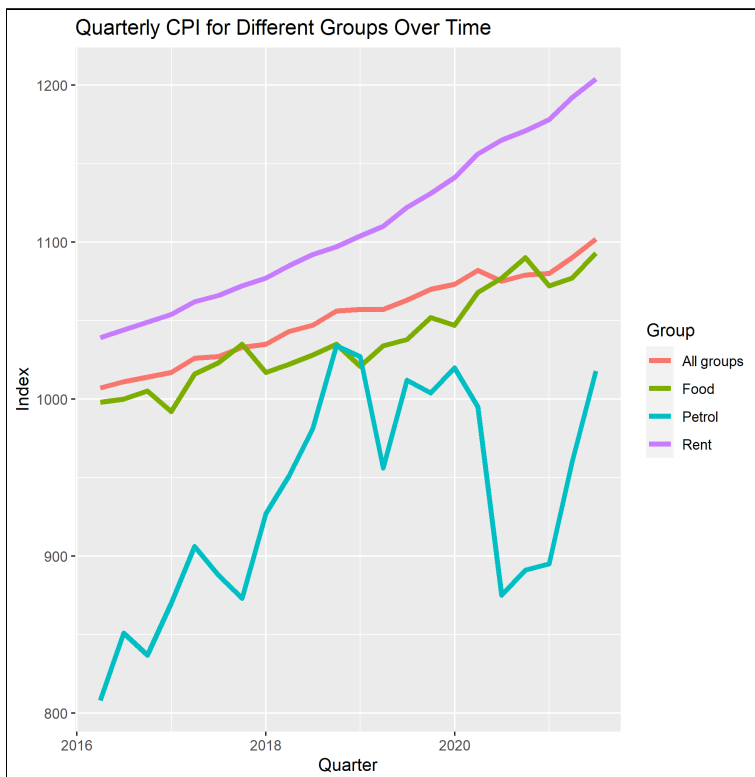


The line graph to the left shows an 8-month rolling average for exports (red) and imports (blue) value for New Zealand cargo with the date (months) on the x axis and value in NZD millions on the y axis. This time the data starts from the start of COVID (first cases in NZ) in 2020 till February 2021 (based on Final Values). This makes it easier to see the effect that the COVID era is having on our import and export values. Presently they are nearly in line with each other at ~5300 million New Zealand Dollars (NZD). The effects of the first lockdown are evident in the graph but the delta variant lockdown effects haven't been seen due to there not being as many supply chain issues.



The line graph to the left shows an 8-month rolling average for exports (red) and imports (blue) weight for New Zealand cargo with the date (months) on the x axis and weight in NZD millions on the y axis. This data starts from early 2020 (start of COVID in NZ) till February 2021 (based on Final Values). You can see that Import weight was affected at the start of COVID by the drop-off in import weight. This was because of supply chain issues during the first few months of the pandemic. Weight of exports however increased during COVID and this will be likely due to our agricultural exports like milk products or meat. The scale of the weight is in millions which means for example 2000 million tonnes is in fact 2 billion tonnes. This will become important when plotting the combined graphs.

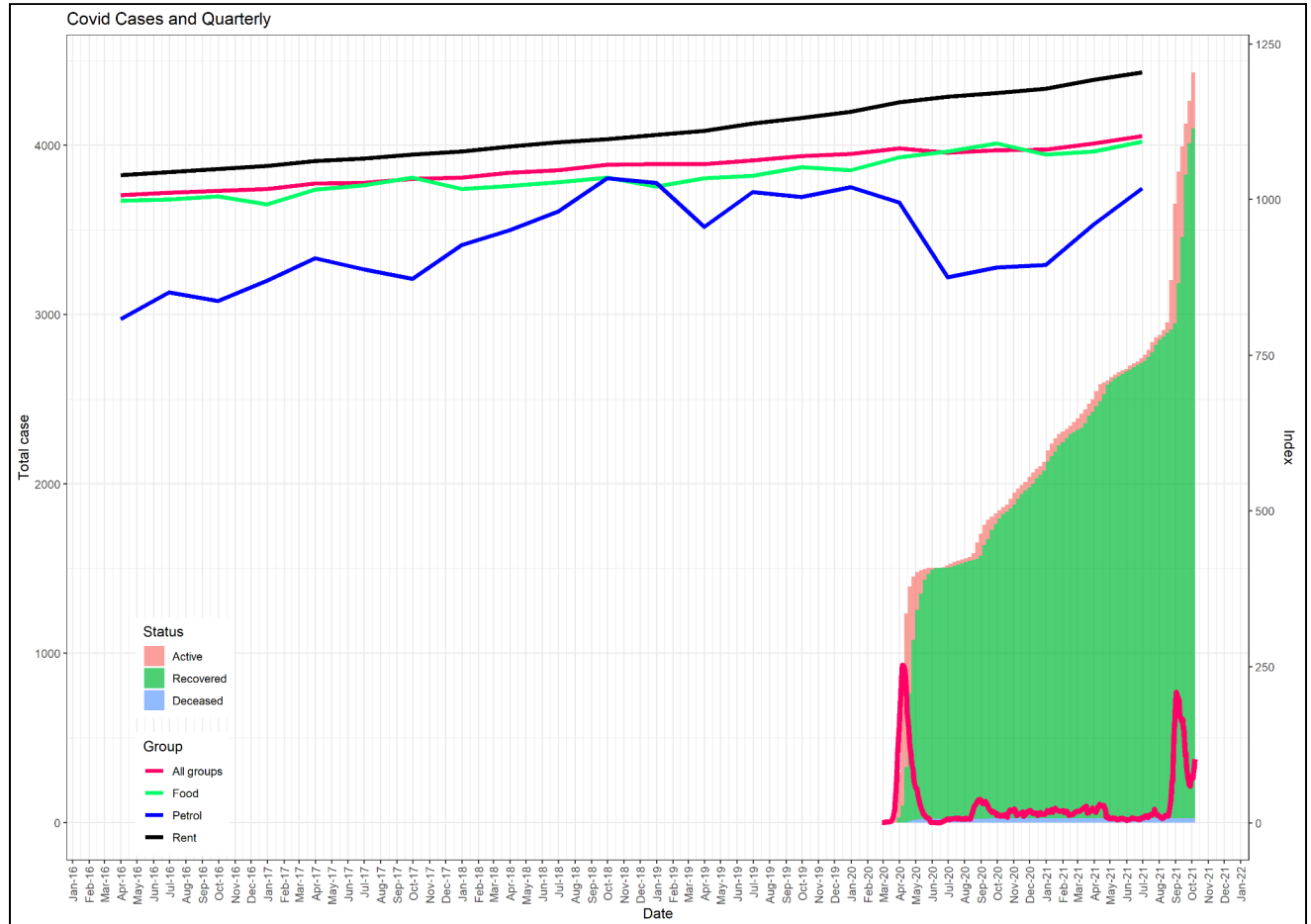
### CPI Graph:



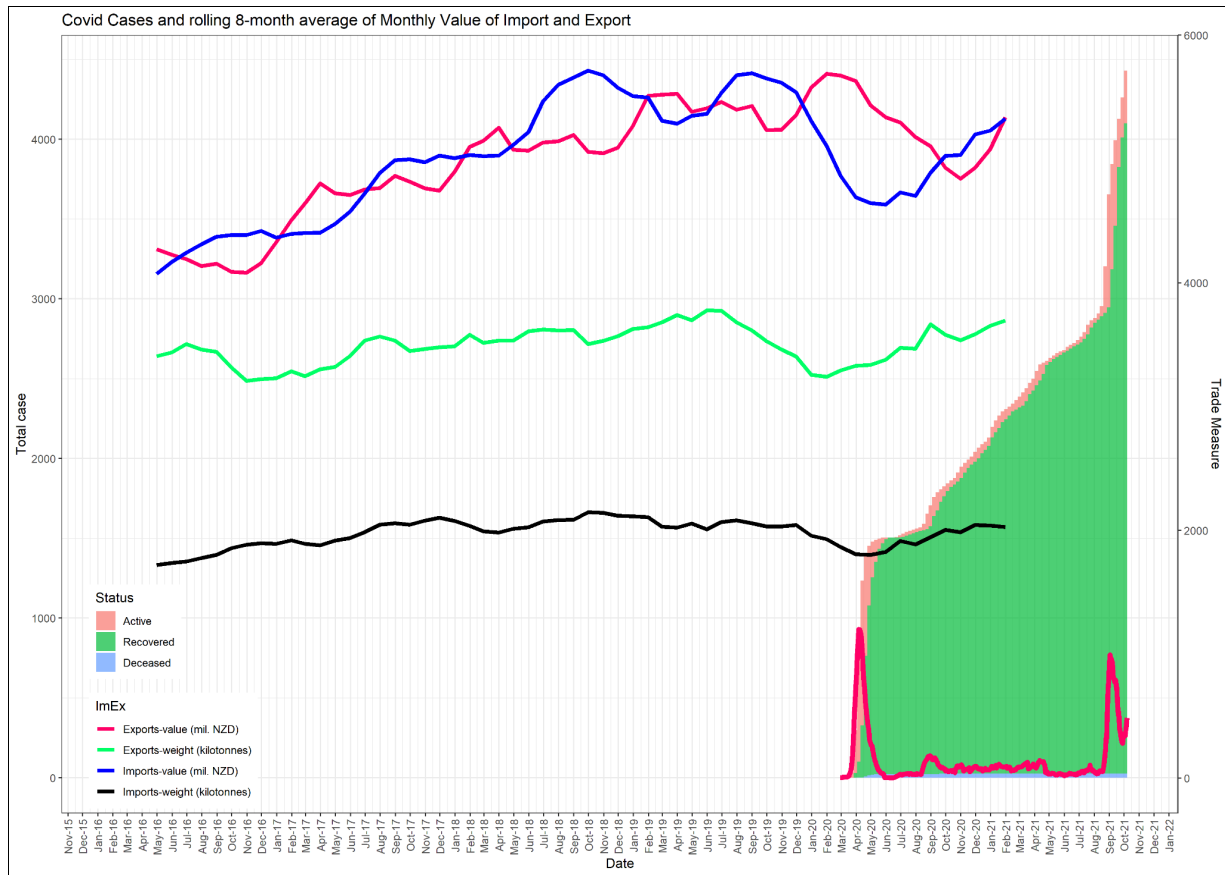
The line graph to the left shows some key CPI (Consumer Price Index) items. The purple line shows Rent, the green line shows food, the blue line shows petrol, and the red line shows all options averaged. On the x axis is the Quarter, as CPI values are released every 3 months. The y axis is measured in points which is a relative term from which a percentage change can be derived. Overall Consumers are paying more for their goods and services even before the pandemic. However, 2020 and 2021 have made the curve a lot steeper and we as consumers are likely to see this increase further. Rent seems to have a positive linear trend that is relatively unaffected by COVID due to the fact that we still have a housing shortage. Petrol has been heavily affected by our lockdowns because no one is out driving as much and the global oil price was on a steady decline because of this reason. Finally, Food seems to have fluctuated during COVID with a steep increase followed by a drop just before 2021. It's now increasing again.



## Combined Graphs:



The above combined graph shows the covid cases (first graph) along with the CPI's (sixth graph). The left scale is for the number of covid cases and the right scale is for CPI values. By combining these two graphs we can see that the price of petrol has actually increased about halfway through 2020 when daily cases were low (from near the end of 2020 to midway through 2021). Having these two graphs combined into one allows us to see what is happening to covid cases and the CPI at the same time.



This combined graph shows the covid cases (first graph) along with New Zealand exports and imports value (third graph) and weight (fifth graph). On the x axis is date in months and on the y axis is the number of cases, Import/Export value, and Import/Export weight. This is able to happen because we have scaled the weight and value of cargo data such that it's displayed as thousands of millions (normally displayed in billions) so that we are able to see the covid case data and New Zealand cargo data at a similar scale. If we didn't do this then it would be hard to visually see the comparisons and trends of the two datasets.



**Final Dataset:**









Combining the Trade, CPI, and Covid datasets into a long dataset by time allows for the data to exist with different entry frequencies without having a large quantity of NA values. This compiled dataset has a time span of about 5 years from the start of 2016 to October 2021, however there is no trade data after May 2021, and no CPI data after June 2021.

	Date	key	value
357	2020-03-29	Active	457
358	2020-03-30	Deceased	1
359	2020-03-30	Recovered	63
360	2020-03-30	Active	525
361	2020-03-31	Deceased	1
362	2020-03-31	Recovered	74
363	2020-03-31	Active	572
364	2020-03-31	Exports_weight_kiloto...	3591.309
365	2020-03-31	Imports_weight_kiloto...	2278.137
366	2020-03-31	Exports_value_mil_nzd	6244.932
367	2020-03-31	Imports_value_mil_nzd	5250.431
368	2020-03-31	All groups	1082
369	2020-03-31	Food	1068
370	2020-03-31	Petrol	995
371	2020-03-31	Rent	1156

### **Github:**

**<https://github.com/oscar-toohey/DATA201-group-project>**

We will be releasing the information that we have found on a new public github so that others can use our work in other projects.

Graph	Description	Date	Author
	 BachV <i>origin</i> Merge branch 'main' into BachV	26 Oct 2021 00:59	Bach Vu
	Merge branch 'BachV' of https://github.com/oscar-toohey/DATA201-group-proj...	26 Oct 2021 00:58	Bach Vu
	Add default plot save size	26 Oct 2021 00:57	Bach Vu
	Update Notebook_Final.ipynb	26 Oct 2021 00:51	Bach Vu
	 main <i>origin</i>  AngusL <i>origin</i>  origin/HEAD Merge branch 'mai...	26 Oct 2021 00:00	Angus L
	Uploading draft presentation and report	25 Oct 2021 23:56	Angus L
	Update save plot function	25 Oct 2021 23:38	Bach Vu
	Fixed a number of issues and added more comments	25 Oct 2021 19:41	JamieBarnes964
	 JamieB <i>origin</i> Created Wrangled_covid.csv	25 Oct 2021 17:26	JamieBarnes964
	 SethG <i>origin</i> Merge branch 'main' into SethG	25 Oct 2021 17:04	Bach Vu
	Finished the readme	25 Oct 2021 17:01	Ehst-SG
	Updated stuff	25 Oct 2021 16:32	Ehst-SG
	Update README.md	25 Oct 2021 16:31	Ehst-SG
	Work on the readme	25 Oct 2021 16:26	Ehst-SG
	 OscarT <i>origin</i> Oppes 3 files missed	22 Oct 2021 02:28	Bach Vu

### **Ethical considerations:**

We are showing only a few factors that contribute to the cost of living in New Zealand; this could skew someone's views towards our current government. Our simplified analysis of the cost of living shows very few factors which show that their cost of living hasn't changed or gotten cheaper while covid. The problem is there are other factors that we cannot show in a csv as well as a relatively short data set which can't show the long term trends. We have mitigated our interpretation risks involved with our data by having comments in our code and a readme file.