COMP20008 - Assignment 2 Final Report

Group 22

Andrew Ngadiman (1185855)

Fiona Agustina (1132633)

Nadya Aurelia Herryyanto (1185814)

Roseline R (1136445)

Preface

Australia had faced enormous shock due to the COVID-19 global pandemic. While Australia has been successful in handling outbreaks and has transitioned to the "New Normal" nationwide, the pandemic has immediate long lasting threats to its economy, and therefore to its labour force. Hence, this proposal begs the question: "Is there a correlation between the number of COVID-19 cases and number of people being employed in Victoria per month?". This research narrows down the scope for research to the state of Victoria, accounting for most of Australia's COVID-19 cases, to better evaluate the correlation of COVID-19 cases and employment numbers. Through the assessment of this scheme, this paper intends to examine the underlying health factors in Victoria as demonstrated through the number of COVID-19 cases, livability factors as indicated by employment numbers in Victoria, as well as Victoria's sustainability factors as stipulated by the number of people being employed in Victoria and how this number could impact or maintain economic growth in Victoria.

As mentioned earlier, Victoria is one of the most heavily impacted states in Australia due to its strict restrictions and lockdowns causing plenty of job sectors to be forced to close down. Consequently, employees and staff are obliged to work from home or are entirely dismissed, resulting in a presumably low employment rate statewide. Moreover, a low employment rate also suggests that a lot of people are idle as they struggle to find job opportunities to support lives leading to worsened mental conditions. Imposing work-from-home initiatives may also prompt both positive and negative outcomes, namely that although employees can now handle work and house chores simultaneously, the lack of socialization can also accumulate stress and reduce productivity.

It is important to note that this study does not imply any causal significance, but that this study endeavors to reveal undermining relationships between the employment numbers and new COVID-19 cases in Victoria.

Datasets and Data Wrangling

Accordingly, our group had managed to gather two secondary datasets to use for our research. The first dataset consists of the number of COVID-19 cases and new daily cases acquired in Victoria throughout the course of the pandemic through to February of 2021. The second dataset contains the number of employment and underemployed residents of each state in Australia from July 2014 onwards. To increase relevance and accuracy, our group has decided to standardize these datasets, focusing on datas relevant to the state of Victoria from January of 2020 (to which the first COVID-19 case had just initially entered the state) to February of 2021. It should be noted that the first Covid-19 case entered Victoria on January 25th, 2020 (Health.Vic, 2020), which means that a comparison can still be made before, during, and after the state of emergency. Thus, to link the two datasets to amass a conclusion, several preprocessing techniques must be done to take both datasets into the same timeframe from January 2020 to February 2021 to increase relevance and accuracy.

1. **COVID-19 Cases in Victoria Dataset**

   The COVID-19 dataset is in a csv format and initially consists of only dates (day and month) and the different categorical incidents occurring on each respective day. The first pre-processing technique conducted was to integrate the year to each date to account for the lack of year indicators and prevent overlapping datas from the year 2020 and 2021; for instance, there will be two datas for February 1st (01/02). Secondly, only the

new-cases category was extracted to accommodate for the scheme of our research. A group method was also utilized to group each date into their respective months, therefore aggregating the total number of new cases in Victoria per month. A csv file, 'covid_clean.csv', is produced for further analysis.

2. **Employment in Victoria Dataset**

In contrast, the employment dataset from the Australian Bureau of Statistics is in xls format and requires installation of the xlrd library to read the xls file. Extraction was undertaken to collect the number of people employed in Victoria per month between the months of January 2020 to February 2021. A csv file, 'employment_clean.csv', is produced for further analysis.

After the two datasets were cleaned, two new data-frames were created, namely 'employment_clean.csv' and 'covid_clean.csv'. Afterwards, the dataframes will be combined into a single dataset with three columns with Month/Year as the unique key. The new/merged data frame will then be used for data visualization where a line graph and scatter plot was created to better illustrate the data. The Pearson correlation coefficient along with its p-value will be calculated and tested with 95% confidence of the correlation between the two variables.
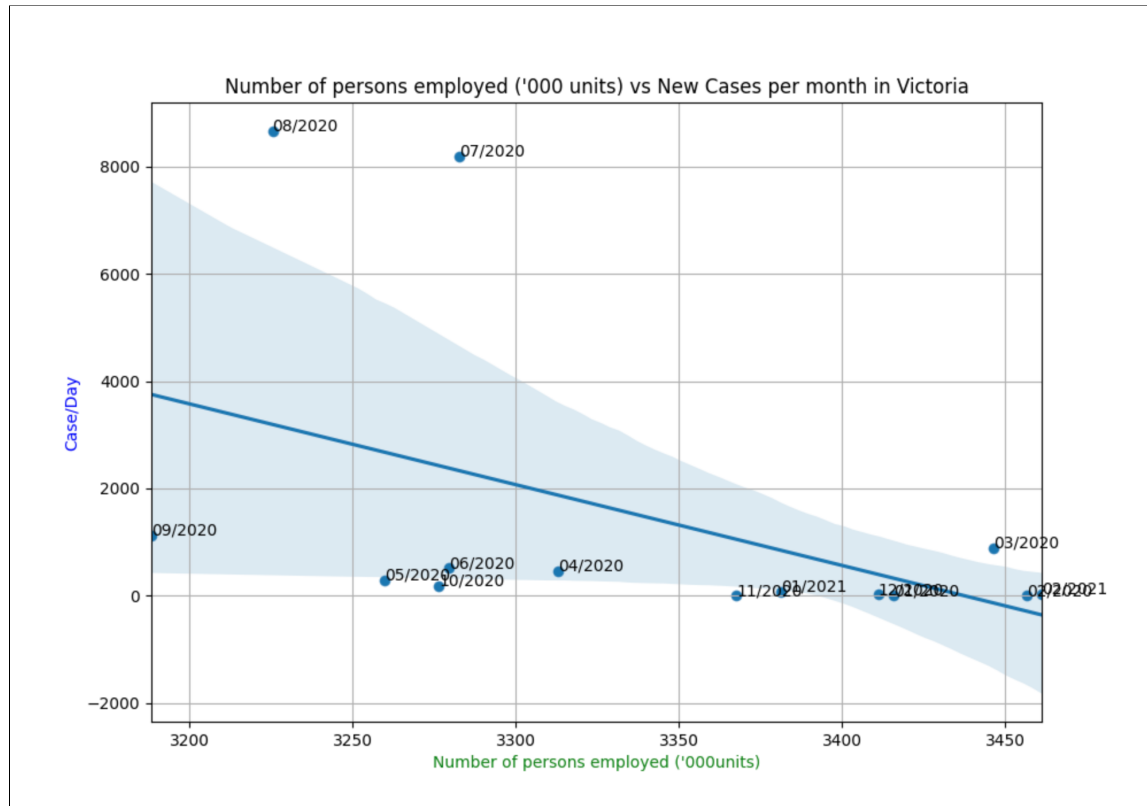
Visualisations and Results

Figure 1

Note. A scatter-plot with a best fit line to show correlation between number of employment and new cases

Based on figure 1, we can see that there is a negative correlation between the number of persons employed and cases per month. Although there are 2 outliers on the upper left side on the scatter-plot graph, the other plots are close enough to one another and scattered on the lower part of the graph. The result for the Pearson correlation coefficient is -0.46, showing a low negative correlation with a two-tailed p-value of 0.0984 (refer to appendix 1). Since the p-value is above 0.05, we cannot conclude that the correlation between the number of persons employed and the number of new cases per day are statistically significant.
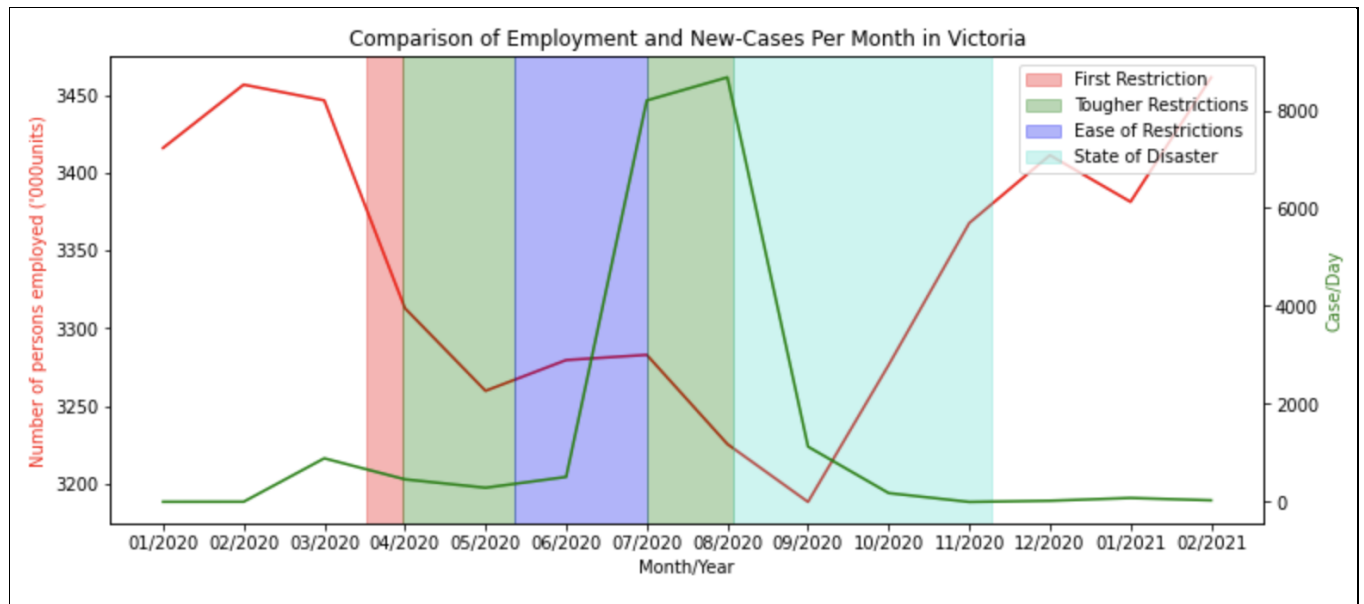
Figure 2

Note. A line graph to indicate the occurrences based on the specified timeline

In contrast, based on figure 2, a general conclusion can be drawn in which the number of employment and new covid cases are negatively correlated. Specifically, when the number of cases spiked up, the employment figure decreased respectively. The stage of emergency was also added to enhance the timeline for better evaluation, in which, during the second 'tougher restriction' up through the first quarter of 'state of disaster', employment units decreased significantly from July to September 2020. However, as restrictions ease down and new cases remained low/unpresent from September 2020 onwards, employment rate continuously surged, with only a slight drop after the new years.

Overall, the results visualised in figure 1 and 2 can be useful as future references for multiple stakeholders statewide, or even nationwide. For instance, state government bodies can take these values and research results into consideration when deciding future policies and interventions in relation to the Victorian labour force, or perhaps when coming up with

COVID-19 restriction responses. Additionally, the results of this study could be useful as reference for future studies on the related fields. Finally, the results generated from this research is significant and valuable as it contradicts general public understanding that the COVID-19 global pandemic, in the grand scheme of things, has detrimental effects on employment rates globally.

Limitations

According to the results stated above, several limitations and setbacks might have affected the overall accuracy of the outcome. Firstly, despite COVID-19 being a significant factor, external circumstances, such as pension and aging, international travelers, and existing government policies on labour might also affect the employment rate. Secondly, because the time scope of the COVID-19 global pandemic is relatively small (only around 1 year), the data might not present the relationship of the number of cases and employment in Victoria accurately. In addition, the COVID-19 pandemic is a recent occurrence, hence relevant datas can be difficult to comeby. Furthermore, using the research result for future use may not be relevant as Covid-19 is a current occurrence that is still growing daily.

Improvements

Future studies on Victoria's COVID-19 situation and its implications on statewide labour force could account for a more detailed approach when describing the labour force. For instance, dividing Victoria's labour force into subgroups of full-time and part-time employees. In addition, future studies could take a more area-specified approach and separate the state of Victoria to

more focused subgroups – suburbs. This method of grouping could produce output that considers underlying external factors such as contained suburb outbreaks (would employment numbers trend in suburbs with contained COVID-19 outbreaks adhere to correlation results demonstrated in this report?). Finally, future studies on this subject matter could explore different sample areas to examine if the trend holds true for other parts of Australia, or other parts of the world.

Conclusion

The COVID-19 global pandemic inflicts substantial economic and societal stress in Victoria. Our study cannot demonstrate that there is any statistically significant correlation between the number of people employed in Victoria per month and the number of new COVID-19 cases per month in Victoria.

References:

Health.Vic. (2020). First novel coronavirus case in Victoria. Retrieved from

https://www2.health.vic.gov.au/about/media-centre/MediaReleases/first-novel-coronavirus
-case-in-victoria

Appendix:

Appendix 1

```
PS C:\Users\srz\Documents\GitHub\assignment-2-group-22> & C:/Users/srz/AppData/Local/Microsoft/WindowsApps/PythonSoftwareFoundation.Python.3.9_qbz5n2kfra8p0/python.
exe c:/Users/srz/Documents/GitHub/assignment-2-group-22/scatterplot.py
The pearson correlation coefficient and it's two-tailed p-value is:
(-0.45946209298790025, 0.09836789590865237)
```

Note. Calculating the Pearson correlation coefficient and the two-tailed p-value to measure the

correlation for the scatter plot.