

Question: Which regions in Victoria should there be more emphasis on vaccination/COVID-safety?

1. What is the research question and how is it related to the theme of understanding the liveability, inclusiveness, health and sustainability of communities in Victoria?

Our main focus is on which Local Government Area (LGA) in Victoria should have more emphasis on vaccination. We want to see whether certain areas have higher or lower vaccination rates in relation to the population in the area. We hope to highlight areas that have low vaccination rates but have high active Covid cases so we can provide more emphasis on vaccinating the people in that area.

This will look at the health and inclusivity of the communities that reside in each area as having higher vaccination rates across the state will only increase the health and safety of people. As a result, it will decrease the pressure on our current healthcare system. The data presented also does not discriminate against age or gender, meaning it is synonymous to being inclusive of all people.

It is a useful tool for people to use this data as it shows how their area as a whole is performing against metrics of vaccination rates. This will motivate them to get vaccinated and increase the possibility of movement and faster return to COVID Normal. Hence, increasing liveability and overall happiness - with social interactions resuming. This also prevents overfunding areas which already have high vaccination rates and allocates it to other areas that may need it as an attempt to be resourceful and sustainable with our limited resources.

What we hope to find is that areas with lower vaccination will require more emphasis and may require more funding from the government.

2. What are the datasets you've used and how have you linked them together?

We have pulled datasets from multiple websites, comparing the relevance and accuracy of the data as we go. We want the most recent and detailed data that we can find. Our time frame of reference for our data is October 2021. We gathered data for the percentage of vaccinated people per LGA and also the population of each LGA. This way, we can calculate the percentage of people vaccinated in each LGA and create a heatmap. This will show us (visually) which areas are less vaccinated and where our emphasis on vaccination should go.

- Our data sets
 - Percentage of vaccinated people per LGA
 - <https://www.health.gov.au/resources/collections/covid-19-vaccination-geo-graphic-vaccination-rates-lga>
 - Exposure sites per population for each suburb
 - <https://discover.data.vic.gov.au/dataset/victorian-coronavirus-data/resource/bc71e010-253a-482a-bdbc-d65d1befe526>
 - <https://www.planning.vic.gov.au/land-use-and-population-research/victoria-in-future>
 - Geographical shape files for LGA in VIC

- <https://data.gov.au/dataset/ds-dga-bdf92691-c6fe-42b9-a0e2-a4cd716fa811/details?fbclid=IwAR1ifKTg5-9NcC1xvQ5nGtmcjEOya7ocgjZexXgFILT4UgZR-2hyzbS-O-Q>

3. What wrangling and analysis methods have you applied? Why have you chosen these methods over other alternatives?

We first converted all our raw source datasets, some of which in .xlsx files into csv so that we could process it with python.

We then cleaned our csv files by loading them into a pandas dataframe then removed description text at the top, unnecessary columns, filtering rows based on states and removing non numeric parts of our data values using regex.

Added a new column by dividing active cases by population to get per capita data.

Using geopandas we were then able to load the Victoria LGA shape files and graph a heat map on top of the actual map of Victoria but first as LGA is a common characteristic throughout the datasets, we needed to ensure the representation (spelling and case) was consistent before merging.

We did this using regex and Hamming distance text similarity. We used regex because it was simple and effective in datasets that had fairly consistent formatting. In cases where regex was not possible, we had to use Hamming distance text distance. We chose to use Hamming distance as it is a simple and fast algorithm. Hamming similarity was paramount for stitching the region shapefile

We also included a bar chart, as this will give a good indication of areas that have lower vaccination rates. This clearly shows the data in a clear visualisation that can identify outliers immediately.

We put all of our data and workspace on Github (<https://github.com/COMP20008/assignment-2-comp20008-assignment-2-group-29>) so that we could work ssin a single repository and have a record of our process. This helped us especially if there was something not working in our implementation, we could easily see where and when the problem occurred. This meant debugging and problem solving was more productively accomplished.

4. What are the key results your research has obtained?

Through our results, we were able to establish that where there were active cases, the percentage of vaccinations in the region is closely correlated as seen in figure 1. In our heatmap, it used a legend to show the saturation of the colour in each map, showing that certain areas have very low vaccination rates whilst others have higher vaccination rates. It is important to note that we do not have any 100% vaccinated areas.

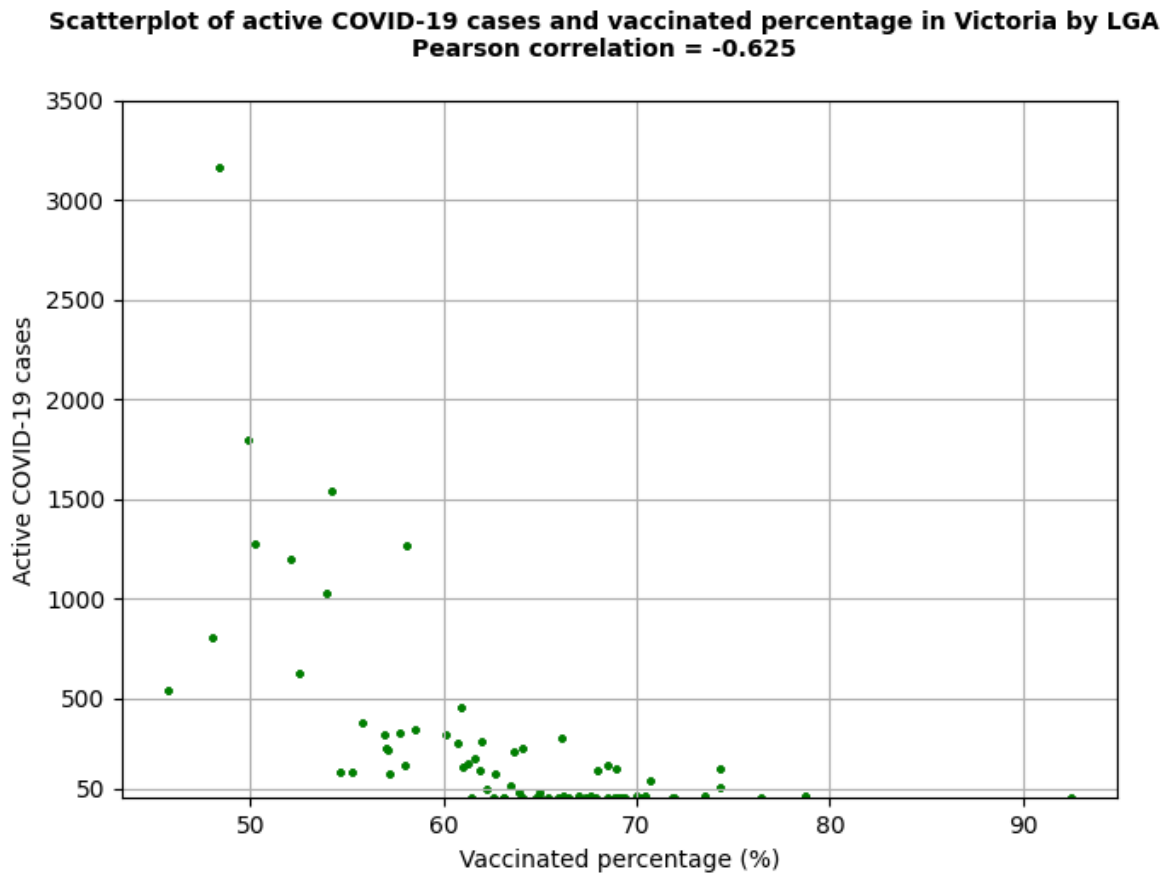


Figure 1: Scatterplot of Active COVID-19 Cases and Vaccinated Percentage

Our scatterplot (Figure 1) showed a negative linear relationship and so we used the Pearson Correlation to calculate. According to Jacob Cohen, >-0.5 is a large negative correlation. Therefore, our scatterplot shows a strong negative correlation because our score exceeds -0.5 , at -0.625 .

In Figures 2 and 3 (joined in Figure 4) heat maps we can see a cluster of high active cases and low vaccination rates in the **Northern suburbs - Northern Central and North Western, therefore these regions in Victoria need more emphasis on vaccination and covid safety.** Specifically named LGAs are **Hume, Whittlesea and Casey.**

To merge the two heatmaps together, we used a weighted average mark formula, giving a weight of 5 to percentage of active covid cases per capita and -1 to vaccination percentage.

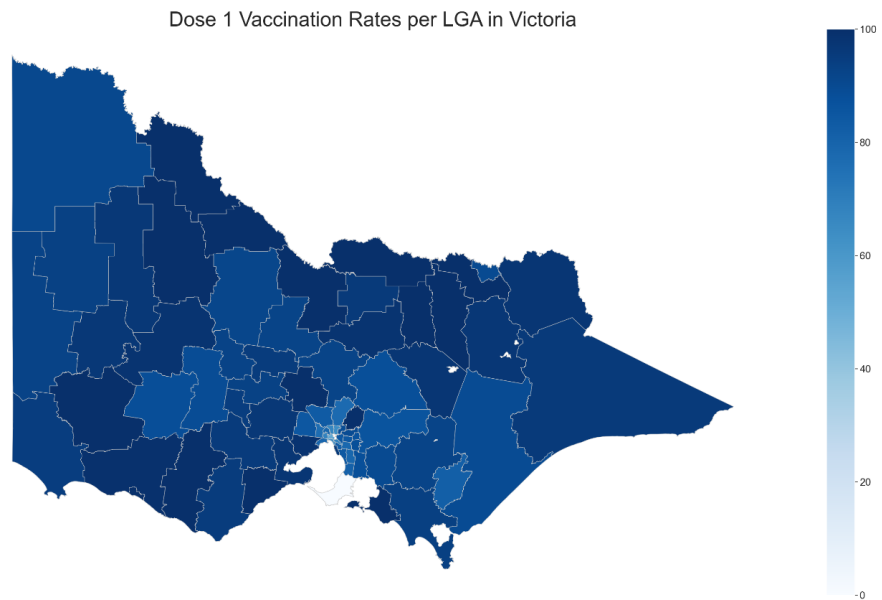


Figure 2: Heatmap for Dose 1 Vaccination Percentage

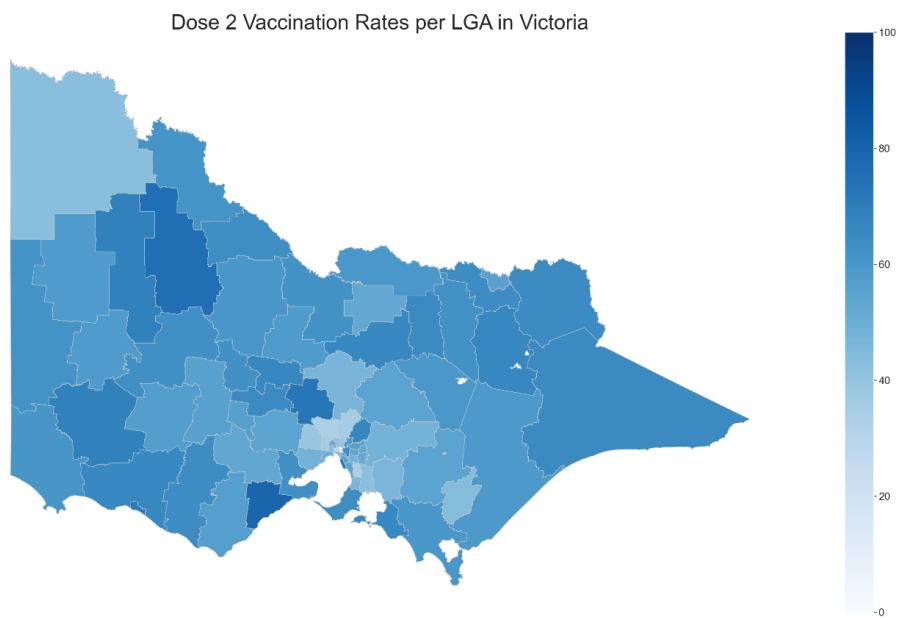


Figure 3: Heatmap for Dose 2 Vaccination Percentage

Victoria LGAs that need more emphasis on vaccines (Proportional to other LGAs)

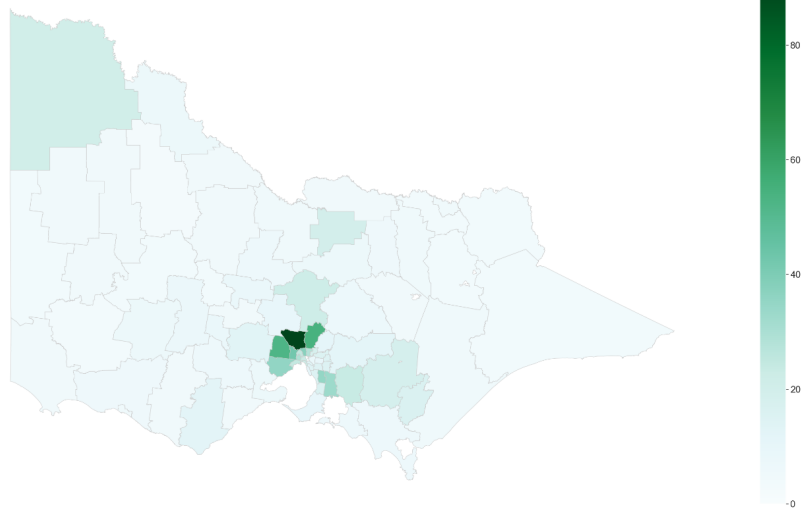


Figure 4: Merged Heatmap for Dose 2 Vaccination Percentage

Victoria map with LGA boundaries and names for reference:





5. Why are your results significant and valuable?

As our results show specific areas and clusters of covid cases and low vaccination rate, this not only helps the government know which areas are lacking in vaccination support and possible funding, but also motivates more people to get vaccinated. These results can document the progress so far and help to propel future growth in vaccination rates going forward.

We mainly focused on Local Government Area (LGA) because although suburb is a smaller subset and may show a more concentrated view, we feel it is best to target an area where we feel most heard. In every LGA, there is a local council, this means that the Minister of each local council is able to take responsibility and relay the message more effectively. When there is somebody who is taking the responsibility and initiative, people are more likely to listen and take action. Councils can also distribute pamphlets with statistics and information on how they can not only contribute to fast tracking COVID Normal, but also create a sense of community along the way. If people feel like they are not alone, they are more likely to act.

6. What are the limitations of your results and how can the project be improved for the future?

The limitations of our project lies in the reliant nature of updated data. Without current data, this will decrease the accuracy of the results as it is a project focused on how certain areas, with low vaccination rates can obtain help presently. Having outdated data would not contribute to solving the problems we have at hand now. As a tool that looks at vaccination rates, it can be used in a multitude of settings and health related illnesses. This means that it is versatile, as long as there is updated data. Where current data is not readily available, means that the

accuracy of the results will be. For example, the census data collected in July 2021 would not be readily available for the public until June 2022, meaning our current population data could be misleading and inaccurate and therefore accurate vaccination rates can't be provided in LGA with strong population growth.

In the future, it would be worthwhile to spend time finding more datasets and be resourceful in our data collection. There could also be more elements that would add to the filtering of smaller subsets like age, gender and other demographic features. This will provide a better user experience and allow people better access to what they are looking for.

More Graphs:

Percentage of people vaccinated against COVID-19 in Victoria boxplot

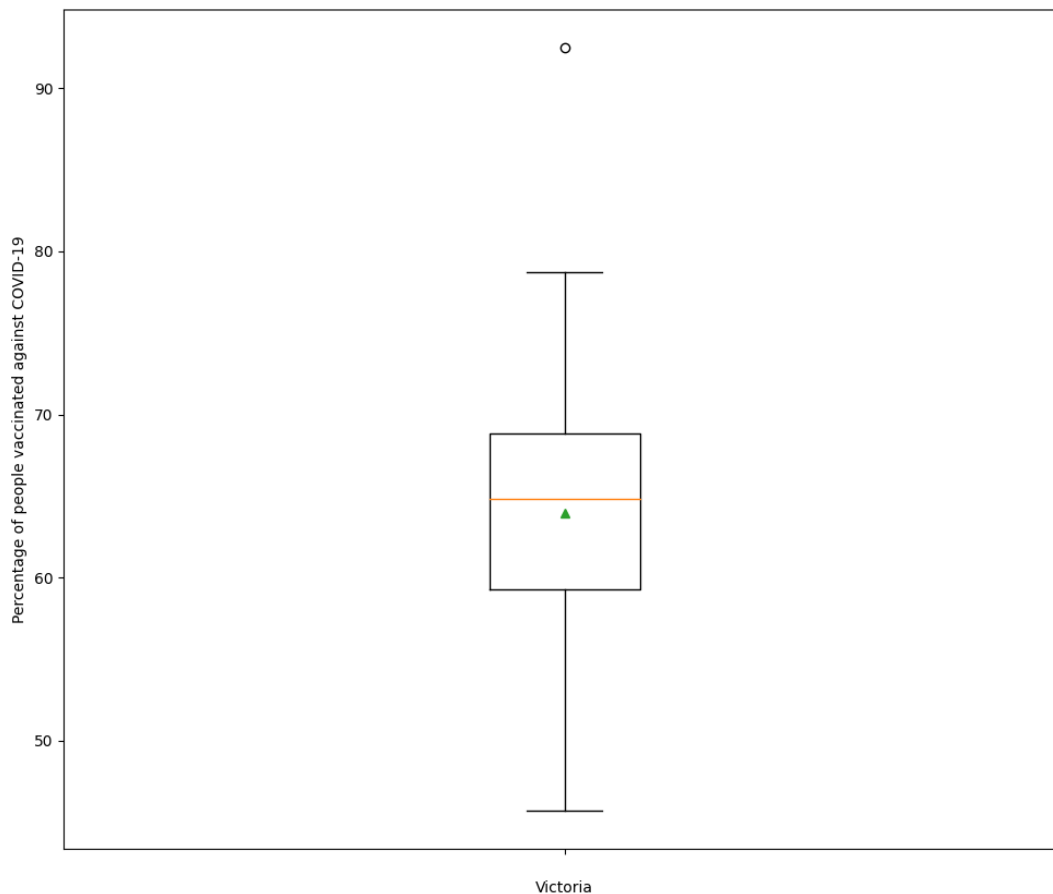


Figure 5: Boxplot of percentage of people vaccinated against COVID-19

The boxplot of percentage of people fully vaccinated against COVID-19 in Victoria has a mean of 64.5%. The outlier is LGA Queenscliff which has a fully vaccinated rate of 92.5% due to their low population of ~1300 people.

Number of active COVID-19 cases in Victoria box plot

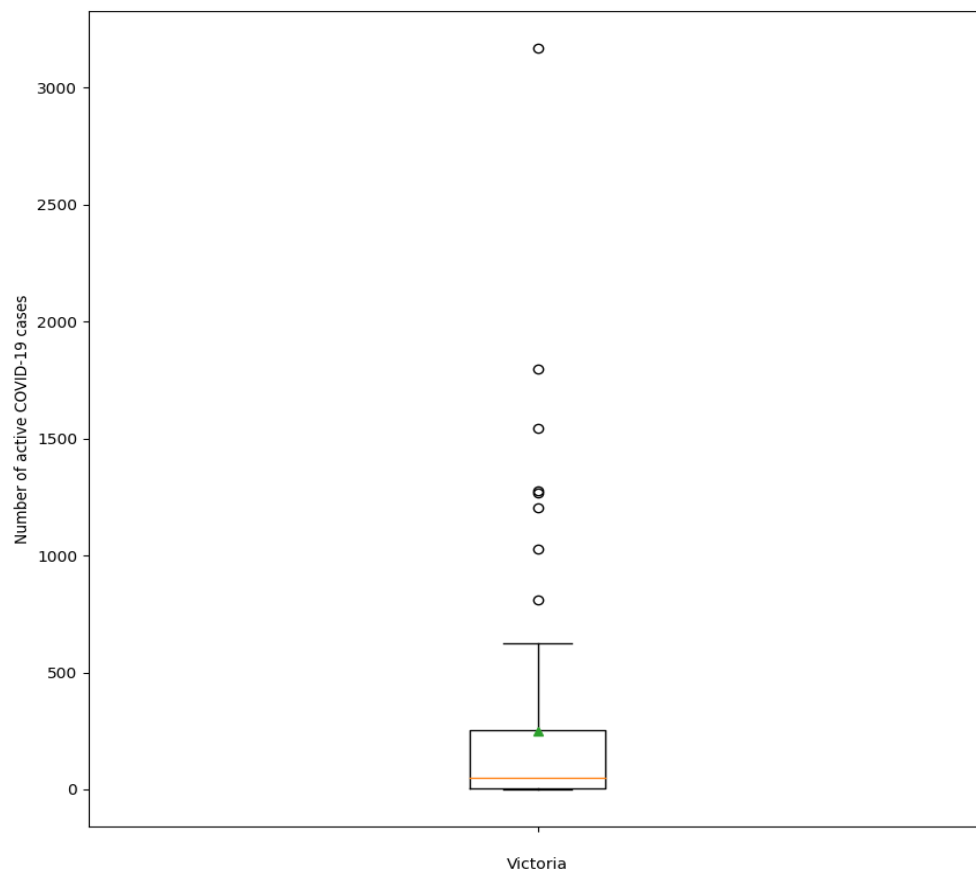
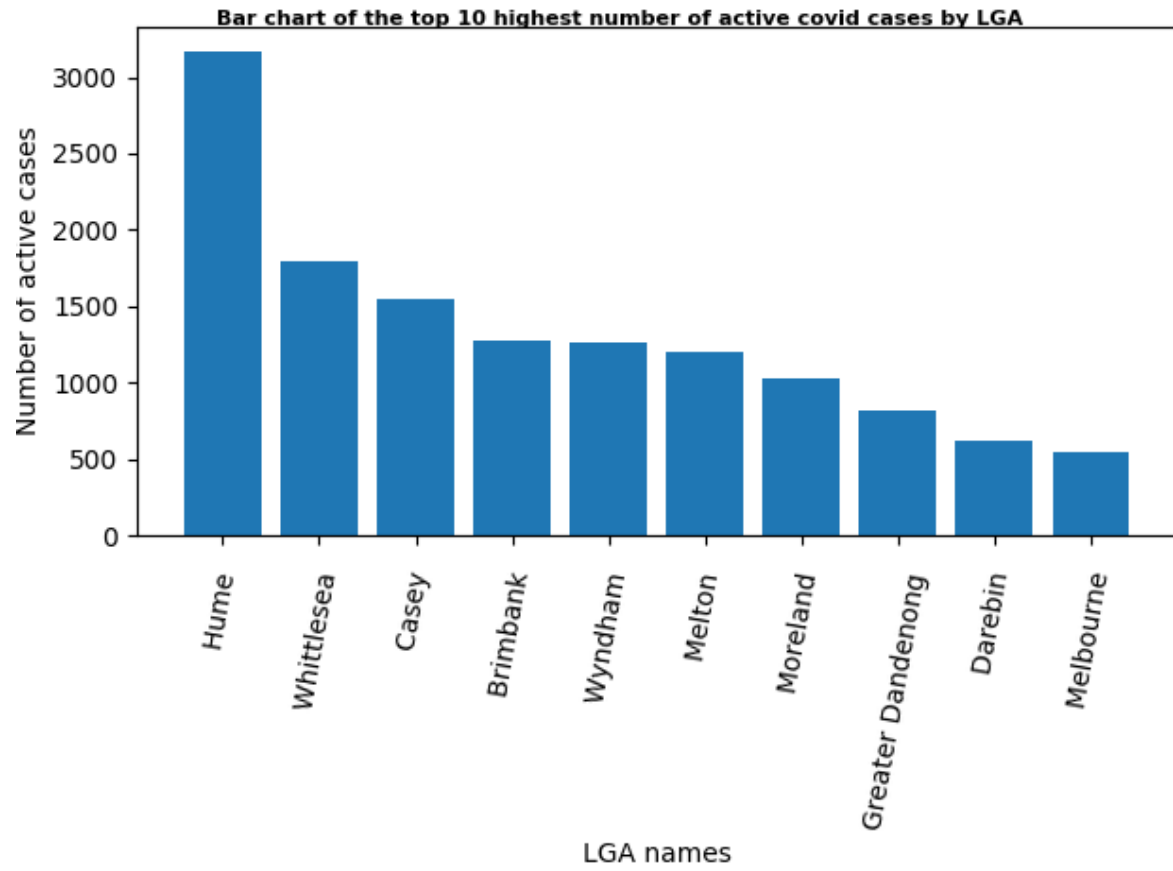


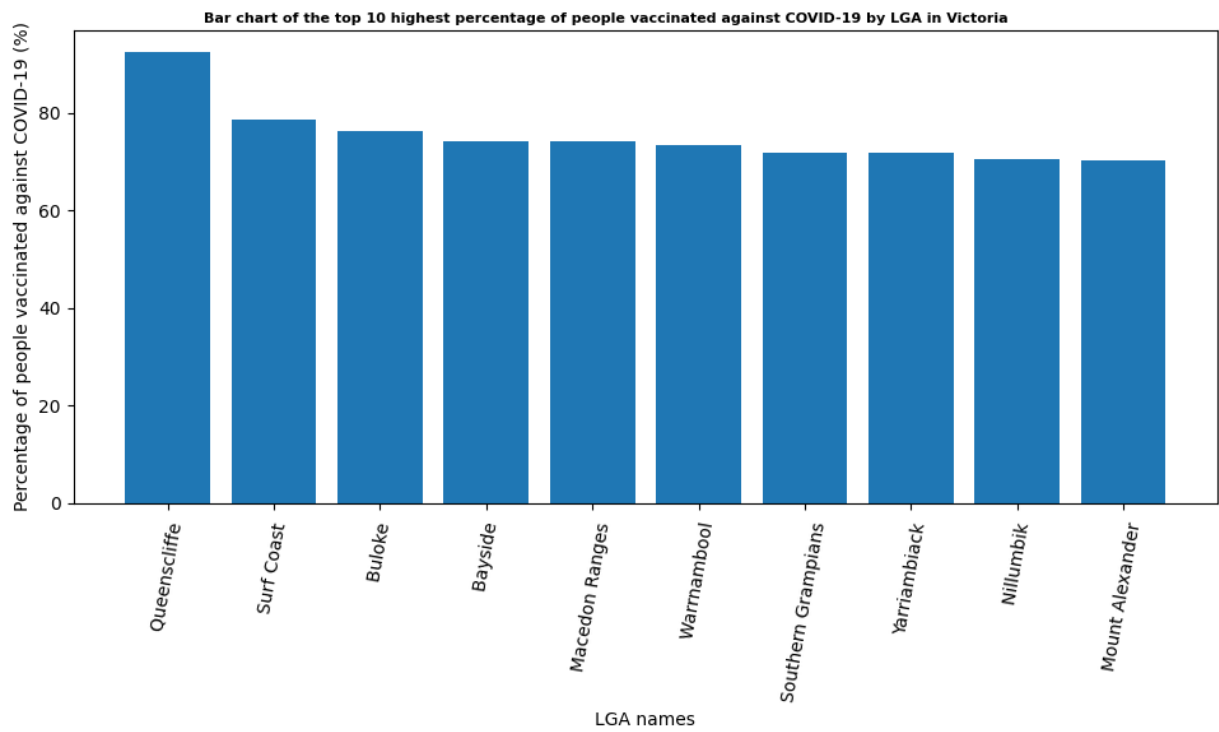
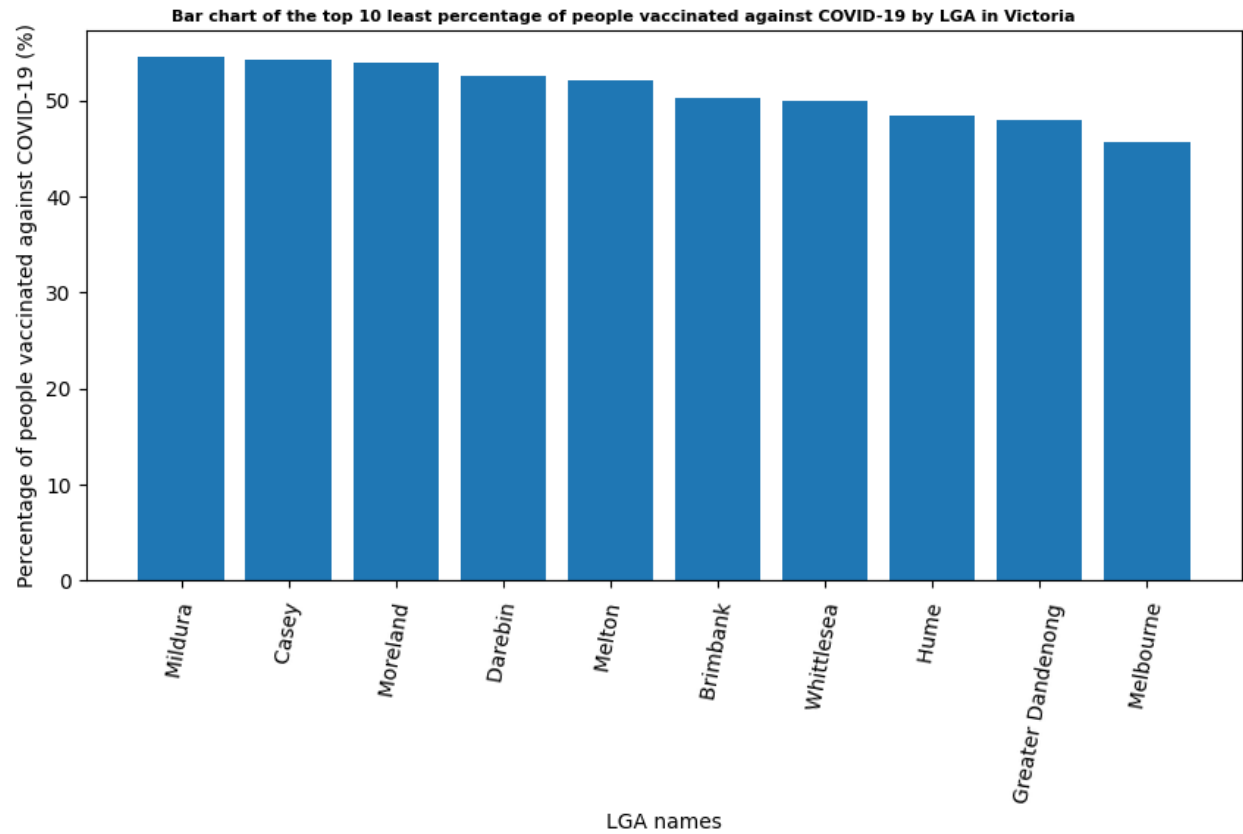
Figure 6: Number of active COVID-19 cases in Victoria

The boxplot of active COVID-19 cases in Victoria has a mean of 249.3 active COVID-19 cases. The first outlier is the LGA Hume, which became the center of the delta variant outbreak.

Source:

<https://www.theguardian.com/australia-news/2021/sep/18/theyre-terrified-how-melbournes-city-of-hume-became-the-centre-of-victorias-delta-outbreak>





We then constructed several bar charts so we can observe the outliers.

Most of the top active cases by LGA also happen to be the least percentage of people fully vaccinated.