

## **Assignment 2:**

### **COMP20008 Assignment 2 Final Report**

Baichuan Yu  
Jie Xie  
Kate Zhang  
Lihang Zhang

## **Introduction**

This report explores and investigates the research question, *'Which suburb in Victoria provides the strongest sense of security to local residents?'*

Liveability can be described as the sum of the factors that add up to a community's quality of life. People's sense of security is a significant reference for evaluating the living standards in their area. Thus the liveability of a community is tightly connected with local people's sense of security. In 2020, sense of security became a popular topic due to the global outbreak of COVID-19 that brought the world into an era of fear and anxiety.

## **Dataset**

In carrying out our research, we adopted two datasets. One of them is the offence rate in each local government area (LGA) in Victoria from 2010 to 2014. The other is about the community strength in each LGA in Victoria for the year of 2014. We chose these datasets because the community strength dataset contains subjective information given by the residents and the offence rate dataset contains objective information that can be used to assess the sense of security offered by a particular area. Looking closer into the datasets, the offence rate dataset records the offence rate per 100,000 population in each LGA, while the other dataset measures community strengths of a certain area from a range of factors, including disagree with accepting other cultures, felt discrimination in past 12 months, support in time of crisis, provide support to other relatives, feel safe walking alone in local area after dark, voluntary work in past 12 months. For each factor, the unit of measurement is the age standardised rate per 100 population. The datasets are linked through their shared column LGA name.

## Data Wrangling and Analysis Methodologies

In preparation for analysis, we first only kept the 2014 data points in the offence rate dataset to cooperate with the community strength dataset. The crime rate dataset was in the xlsx format which could not be directly transferred to a dataframe in pandas. Hence, we manually copied the relevant data into a csv, `offence_rate.csv`. We cleaned out irrelevant data points in the community strength dataset, including a row that represents the area “Unincorporated Vic” which contains no information. In this project we are interested in data points measured as rates, as this will not penalise LGAs with large populations. Although our two datasets were measured in rates already, we had to unify the unit of measurement across the two datasets, because the rate in the community strength dataset is measured per 100 population whereas the rate in the Crime dataset is measured per 100,000 population. Finally, we merged the community strength and offence rate dataset by the LGA name. During the merging process, the names of the LGAs weren’t matching initially, as the names in the community strength dataset include a bracket with extra information. We removed this extra information using the `strip` function in python to obtain matching columns and linked the two datasets through these columns. The pre-processed data was stored in `vic-security.csv`.

The analysis methodologies include both visual and numerical analysis. Numerically, we calculated a score in assessing how well an area is offering its residents a sense of security relative to other areas. The score for each factor is the Z-score multiplied by 100 to give numbers that mostly range between -100 and 100, which are more interpretable. For this project, we defined a higher score as an indication of better performance for that particular factor in comparison to other areas. For example, if a LGA has a score of 0, that means this area represents the average performance of the LGAs in Victoria. If a LGA has a score above 0, that means this area has higher performance than average, and vice versa for a score below 0. We calculated an overall score for each LGA by taking the mean of the scores across all factors, which measures the overall performance of an LGA relative to other LGAs by taking all the factors into account. The overall score is stored in a separated csv `overall-score.csv`.

Since the scores for each factor is indicative of how well a LGA is doing for that particular factor, we modified the column names of the dataframe from the original dataframe in `vic-security.csv` to better communicate their contextual meaning in terms of scores. The dataframe with the scores and modified column names is stored in `vic-security-score.csv`.

Visually, we create a bar chart with the overall scores on the y-axis and their corresponding LGA names on the x-axis, which presents the rankings of the performance of each LGA visually with the height of each bar. The spread of overall score also becomes more transparent with a bar chart. Apart from the bar chart, we also constructed a choropleth. LGAs with darker blue colour on the choropleth are considered to have better overall performance. An advantage of choropleth is its geographic interpretation of the data, which can be helpful for new immigrants who are not familiar with the LGAs’ names. The choropleth also allows people to view the situations of their surrounding LGAs conveniently. We also performed clustering on the choropleth that groups the LGAs into three clusters - high, median and low performance, which provides more general and straightforward information on performance geographically. Furthermore, we constructed a heatmap to display the individual scores of each factor for all LGAs, so that all details displayed clearly in front of the audience. The lighter the colour of each cell, the better the performance of the LGA for that particular factor. People who have special interest for a certain factor can easily use heatmap to assess each LGA when it comes to their factor of interest.

## Result

lga_name	overall score
Bayside	96.2
Boroondara	80.3
Stonnington	71.1
Moyne	68.3
Surf Coast	61.5

Figure1: Top 5 LGAs form overall-score.csv.

Greater Shepparton	-75.6
Brimbank	-81.0
Melton	-81.8
Hume	-82.4
Greater Dandenong	-89.2

Figure2: Bottom 5 LGAs form overall-score.csv.

Looking at the table for overall score, we see that Bayside, Boroondara and Stonnington are the three top performing LGAs in Victoria that provide the strongest sense of security to local residents, with outstanding overall scores of 96.2, 80.3, and 71.1 respectively. The three worst performing LGAs are Greater Dandenong, Hume and Melton, with overall scores of -89.2, -82.4 and -81.8 respectively.

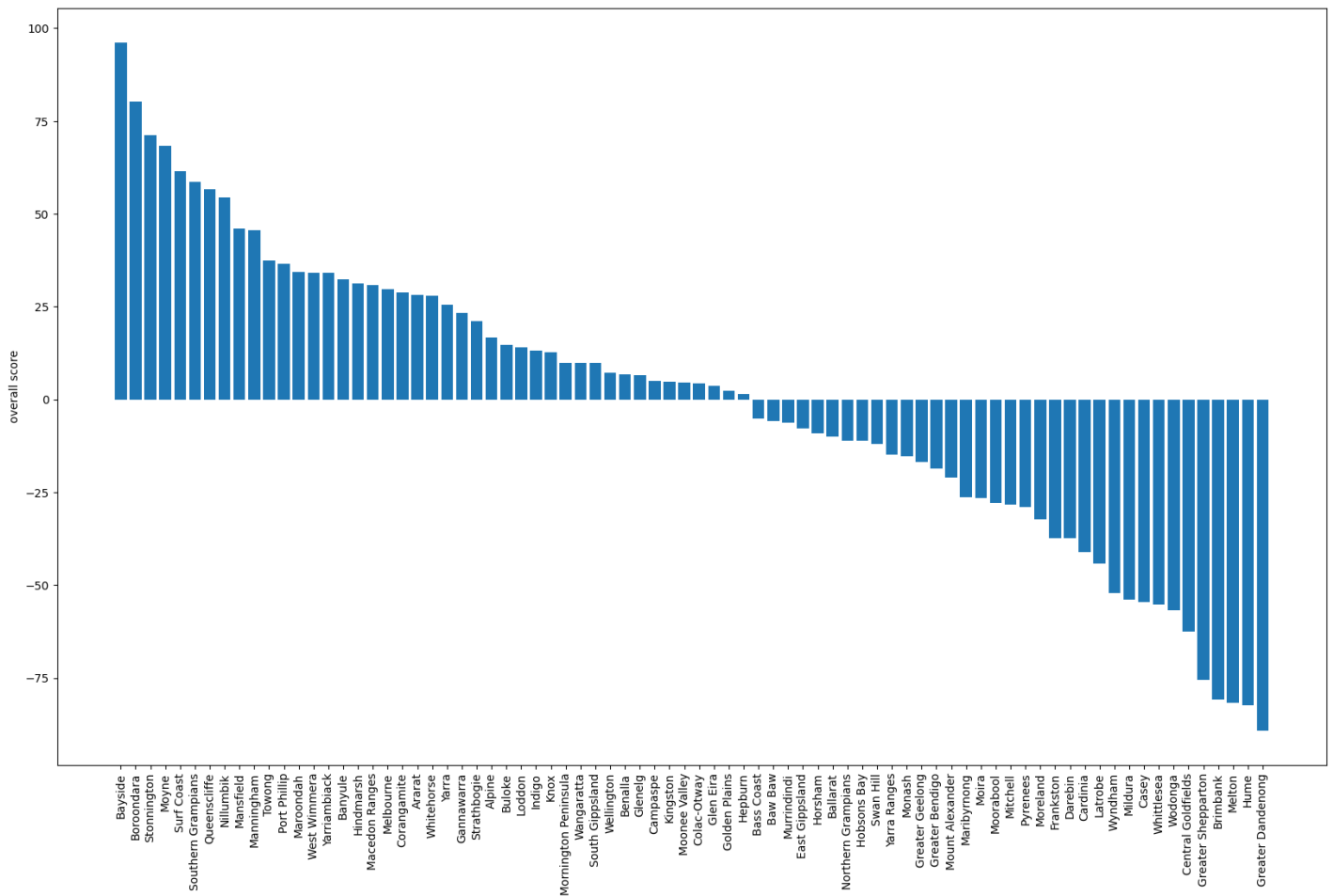


Figure3: Bar Chart – overall score ranking

From the bar chart, we clearly see the rankings of all LGAs and the fact that the number of LGAs with performance below and above the average is very similar. This demonstrates that there's no significance skewness in the overall performance of the LGAs

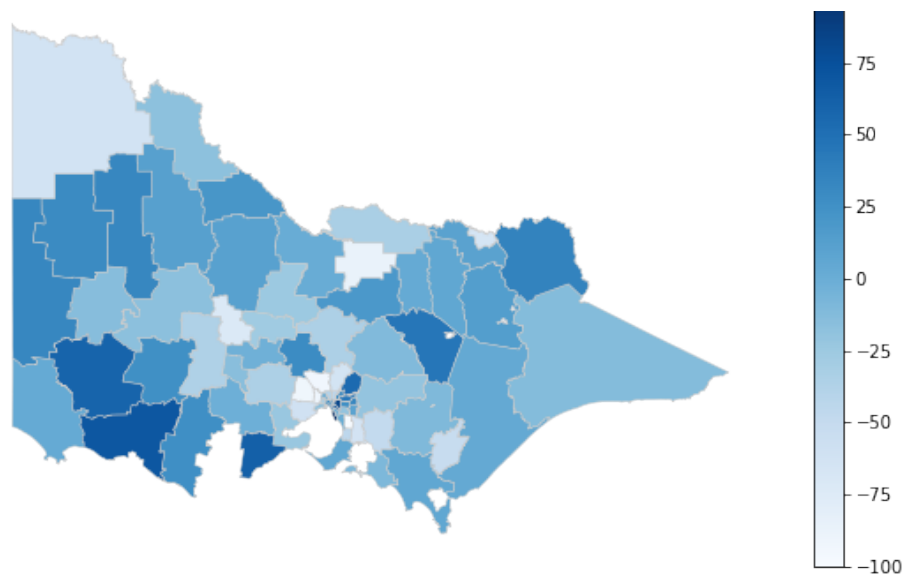


Figure4: Choropleth – Overall Performances of LGAs in Victoria

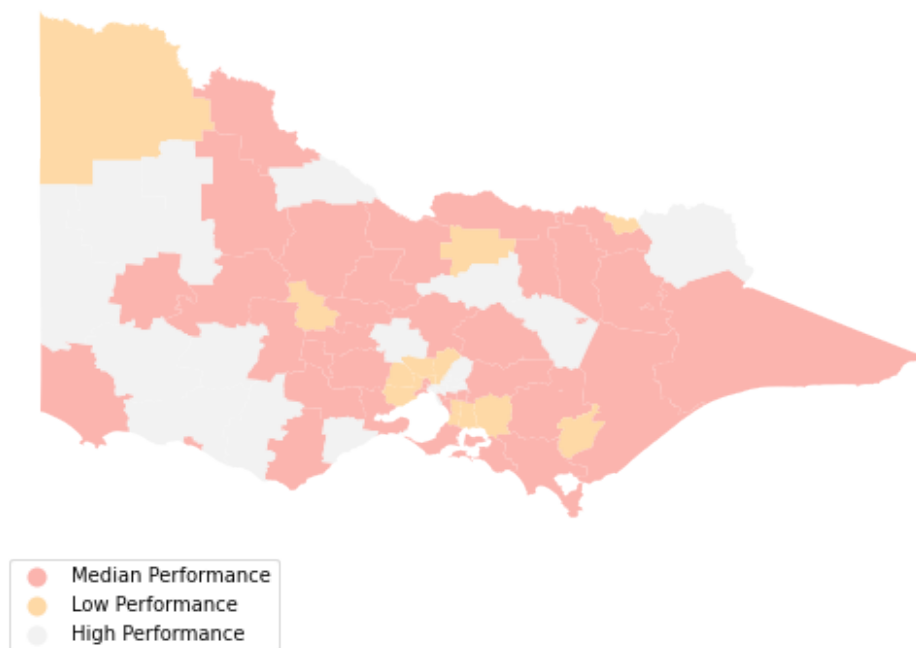
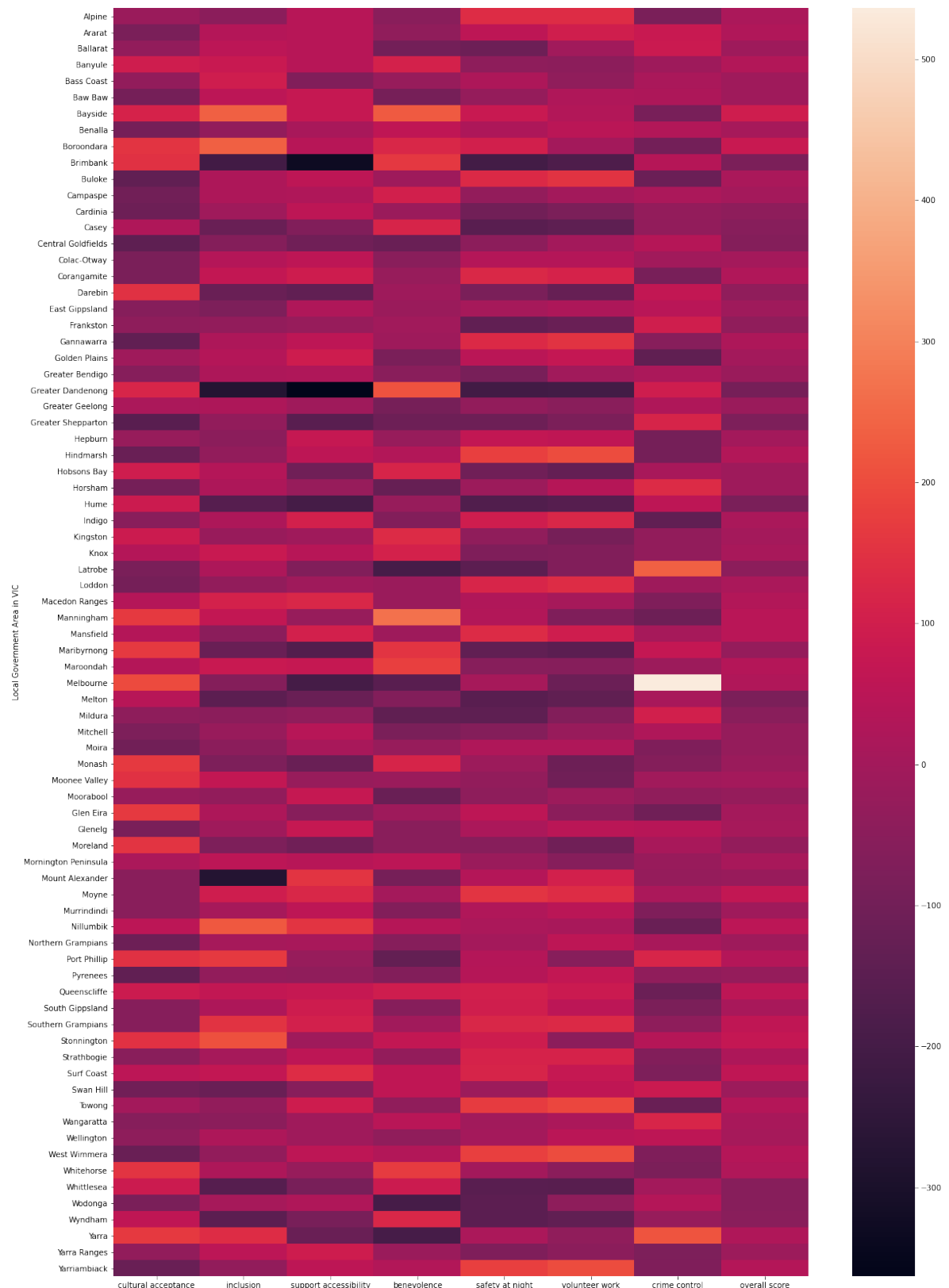


Figure5: Choropleth with Clustering

The choropleth and choropleth with clustering demonstrate that southwestern Victoria tends to give its residents a higher sense of security. However, the north-western corner of Victoria and a few LGAs in Melbourne are doing poorly in terms of giving their residents a sense of security.



Going beyond the overall score, the heatmap contains key information regarding the LGAs' performances for each factor. For example, looking at Mount Alexander with Moreland, although their overall scores are similar, Mount Alexander is doing much worse than in Moreland when it comes to inclusion.

Our findings will be valuable to immigrants, local residents, real estate agencies and the local government of each LGA. For immigrants and local residents, the scores and ranking for each LGA will help them to find an ideal residence. The information is also of interest to real estate agencies and the local government as it's related to their institutional obligation. Our result also provides insightful information as we assessed each LGA with subjective information. Existing research on the topic is more objective and mostly focuses on assessing how safe each area is, whereas our analysis focuses on both objective facts such as offence rate and the subjective information by looking into how secure the residents feel in their respective areas.

## Discussion

The major limitation of our result is the timeframe of our dataset (2014) as it's difficult to find other suitable dataset due to its subjective nature. It would be ideal if we can obtain datasets from a more recent timeframe as much has changed due to increased migrants and the impact of COVID-19, which will allow us to offer more valuable information to our audience. However, through our research we have developed a set of methodologies to rate and rank each LGA, therefore once any recent data is available from the same source, the set of methodology can be easily reused to draw new conclusions. Additionally, the fact that the score of each LGA is derived using Z-score is another limitation as this scoring system is susceptible to skewed data. The fact that the final overall score has a range between -100 and 100 might lead to misconception to the audience, as the audience might think the score is a rating of the LGA's performance whereas the score actually measures the relative performance. In the future, a more sophisticated weighted scoring system can be introduced to improve the result as the factors have different levels of impact on people's sense of security and at the same time make the scores more comprehensible.

## References for Datasets

*Year ending 31 December 2014 | Crime Statistics Agency Victoria.* (2015). [Dataset]. Crime Statistics Agency.

<https://www.crimestatistics.vic.gov.au/crime-statistics/historical-crime-data/year-ending-31-december-2014>

Torrens University Australia - Public Health Information Development Unit. (2015). *LGA15 Community*

*Strength - 2014 - AURIN Data* [Dataset]. AURIN. [https://data.aurin.org.au/dataset/tua-phidu-tua-phidu-2015-](https://data.aurin.org.au/dataset/tua-phidu-tua-phidu-2015-lga-aust-community-strength-2014-lga2011)

[lga-aust-community-strength-2014-lga2011](https://data.aurin.org.au/dataset/tua-phidu-tua-phidu-2015-lga-aust-community-strength-2014-lga2011)