Victoria Crime Data Visualization 2019-2022 March quarter

# Project Report

Team: Berta Devenyi, Phuong Tieu

## Project Proposal

2017, Victoria reported the highest home burglary rate in Australia. In 2019, we had harsh bush fire, in 2020 we had the COVID-19 outbreak, and the state was on and off lockdown until the end of September 2021. In 2022, we gradually have had a more normal life. With these major changes last few years, there were impacts to our livelihood household wise, income and mental changes, we would like to see how crime and safety changed between 2019 and 2022 in the state of Victoria.

## Pre- Processing

### Data source:

Data was obtained from the online public dataset:

* Victoria Crime Statistics Agency (https://www.crimestatistics.vic.gov.au/harm-and-crime)

## Scope of the process:

The data we are proposing for the project covers the period 2019-2022 March quarter (for this period the data is complete). Open-source data has some limitations, and only provides access to historical data.

# Data Extraction

For the extraction process, four different datasets were used from the data. The following datasets were used as below:

* Data\_Tables\_LGA\_Criminal\_Incident\_Year\_Ending\_March\_2022.xlsx has 5 table worksheets



# Data Transformation

Data cleaning: we needed to drop leading space for police region name, dropped total rows for police region table.

Created data frames for the data columns we intended to use.

Exported dataframes to csv files for data quality checking.

An ERD diagram was developed after evaluating the data columns information and its correlation. In considering the scope and time constraint, we investigated table1, 2 and added a new summary table for easier data manipulation. From that, the team created 3 tables connecting them through concatenated keys, e.g. Year and region\_name; or year, region\_name and lga\_name. The diagram has been filed in the repository.

Heroku Postgres Database Setup

Logged onto Heroku, created an account, selected Resources and Heroku Postgres database. After database setup, in Credentials view, copied the URI string, saved it into db\_conn.py file. The URI string will be required in ‘app.py’.

# Data Load

Used flask and SQLAlchemy we manage to import the data into the Heroku Postgres database.

Data Visualization Webpages

Compare Regions

This webpage shows a group bar chart (compare\_region.html) that has up to 4 year regional incidents for user to compares (created using chartjs), hover over a bar will show region name and incident count;

Search Regions

This webpage provides options users to search region (search\_region.html) by selecting on year and region. The screen will show a bar chart (created using chartjs) that lists regions with incidents in descending order, hover over a bar will show incident count; a bubble chart (created by plotly) that shows regions with incident and rate per 100,000,000 population, hover over the bubble can see incident number and rates.

Police Data

This webpage for offence summary (policedata.html) that is providing a list of the different offence types for each year for each local government area (LGA). This least can be filtered by different fields (Local Government Area and Year). Using the filters on the same page and Plotly, we display a bar chart representing the different offence types for a specific Local Government area for a specific year (both filter has to be selected). We used various technologies for displaying this data, including JavaScript, HTML, CSS, D3 and Plotly (policedata.html, policadata.js, policedata.css).

Additionally, with the help of Jupyter notebook, we analysed the 10 most dangerous LGAs based on the number of committed crimes against people 2019, 2020 & 2021. This is not part of the Flask app, we used Python, Pandas, Matplotlib (policedata.ipynb).   
This analysis led us to two of our findings:

1. If we compare individual LGAs with the highest number of crimes before COVID and during COVID, we can see that the ranking hasn’t significantly changed. The number of crimes did change, however.
2. The number of crimes has dropped a lot in 2022 compared to the number of crimes in 2020.

# Considerations and others

Considerations in source data:

* Victoria Crime Statistics data were released quarterly each year; this project is using March 2022 data.
* Had tried to search Victorian Income data to do data correlation, but could not find the compactable year data, the latest open source dataset from Australia Tax Office is 2017.
* ‘Incident’ is used at the police region level (higher level), ‘offence’ is used at the LGA level.
* Removed ‘Total’ rows from police region table.
* With constraints of Heroku’s free data capacity allowance, for regional data (high level) we cover year range of 2019-2022, for LGA detail level we cover year 2022.

Others:

* For ‘Compare Region’ webpage currently showing group bar chart, had tried to build 4 plotly pie charts for regional comparison on 2 rows where each row shows 2 charts, chart labels got truncated when used ‘outside’ label option; when used ‘inside’ label option, only 4 LGAs showing, not showing the Justice and Unincoperated as they had much less incidents , need more time to investigate, therefore group bar chart is used.
* In ‘Compare Region’ and ‘Search Region’ webpages, ‘Incident count per 100’ and ‘Rate per 100,000,000’ were the products of descaling by 100 and 1000 respectively, so that charts can show nicely; particularly for bubble chart, with descaling orange bubbles were shown, otherwise the chart would show a rectangular orange canvas.