Analysis of Sydney Suburbs

Introduction

1.1 Background

Sydney is the state capital of New South Wales in Australia and the most populous city in Australia and Oceania. Located on Australia's east coast, the metropolis surrounds Port Jackson and extends about 70 km (43.5 mi) on its periphery towards the Blue Mountains to the west, Hawkesbury to the north, the Royal National Park to the south and Macarthur to the south-west. Sydney is made up of 658 suburbs, 40 local government areas and 15 contiguous regions. Residents of the city are known as "Sydneysiders". As of June 2017, Sydney's estimated metropolitan population was 5,230,330 and is home to approximately 65% of the state's population.

Despite being one of the most expensive cities in the world, Sydney frequently ranks in the top ten of lists of the most liveable cities in the world. It is classified as an Alpha+ World City by Globalization and World Cities Research Network, indicating its influence in the region and throughout the world. Ranked eleventh in the world for economic opportunity, Sydney has an advanced market economy with strengths in finance, manufacturing and tourism. There is a significant concentration of foreign banks and multinational corporations in Sydney and the city is promoted as Australia's financial capital and one of Asia Pacific's leading financial hubs.

1.2 Problem

A large percentage of population of Sydney comprises of migrants from all over the world. There are migrants primarily from China, India, United Kingdom and a number of other countries. This has led to a rapid expansion of the metropolitan area of Sydney. Immigrants calling Sydney home are in increasing numbers are settling into newer suburb. This has resulted in rapid expansion in the number of smaller business and number of employment opportunities in these new as well in the older suburbs in Sydney.

It is therefore very critical to analyse the patterns in immigration, types of facilities, nature of business, type of job opportunities created as well the amount of crime to get a better picture of change in Sydney.

1.2 Interest

This detailed analysis would be great interest to local government bodies, the New South Wales state government and all its allied agencies. This analysis can aid it town planning and allocation of various community facilities depending on local requirements. Further this report can be also helpful for future planning and strategy.

2. Data

2.1 Data Sources

Data for this report has been web scaped form different website online. The primary source of this data is various NSW government agencies. These include NSW Crime Statistics, the Australian Bureau of Statistics, city of Sydney and Australia post websites. From here data relating to demographics, postcodes, geo coordinates, and crime, population and population trends are used.

2.2 Data survey and cleaning

The data acquired from the NSW crime statistics website contained details of recorded crimes in the state from January 1995 to December 2019. Recorded number of crimes were tabulated by month and type of crime for every suburb of the state.

About 2% values in the data state were missing. The missing data was replaced by appropriate mean values calculated from the available data.

Data acquired from Australia post contained geocoding information along with longitude and latitude coordinates for all post codes in Australia.

The data was tabulated depending on distribution type for a suburb. All distribution types in the suburb does have the same post code and therefore there were a number of duplicates in the post code and the corresponding longitude and latitude data. The duplicates were removed. Post codes belonging to the state of NSW was filtered from the entire set.

The crime data set and the post code data set was then merged matching suburb as the common column in both datasets. The data set was further refined by dropping all the suburbs containing geo coordinates beyond the geo coordinate limits of greater Sydney.

3. Exploratory Data Analysis

3.1 Crime Data

The crime data provides number of crimes reported and therefore recorded every month from January of 2005 to December 2019. These crimes were grouped under major categories and then their respective sub categories. Data for all suburbs of NSW was present in this data set.

This data was then groped together to sum up the number of reported crimes for every year over the 25yr period. Analysis was made to compare the number of reported crime under different crime categories.

The data was further analysed to see the suburbs with the maximum number of reported crimes in various categories and subcategories.
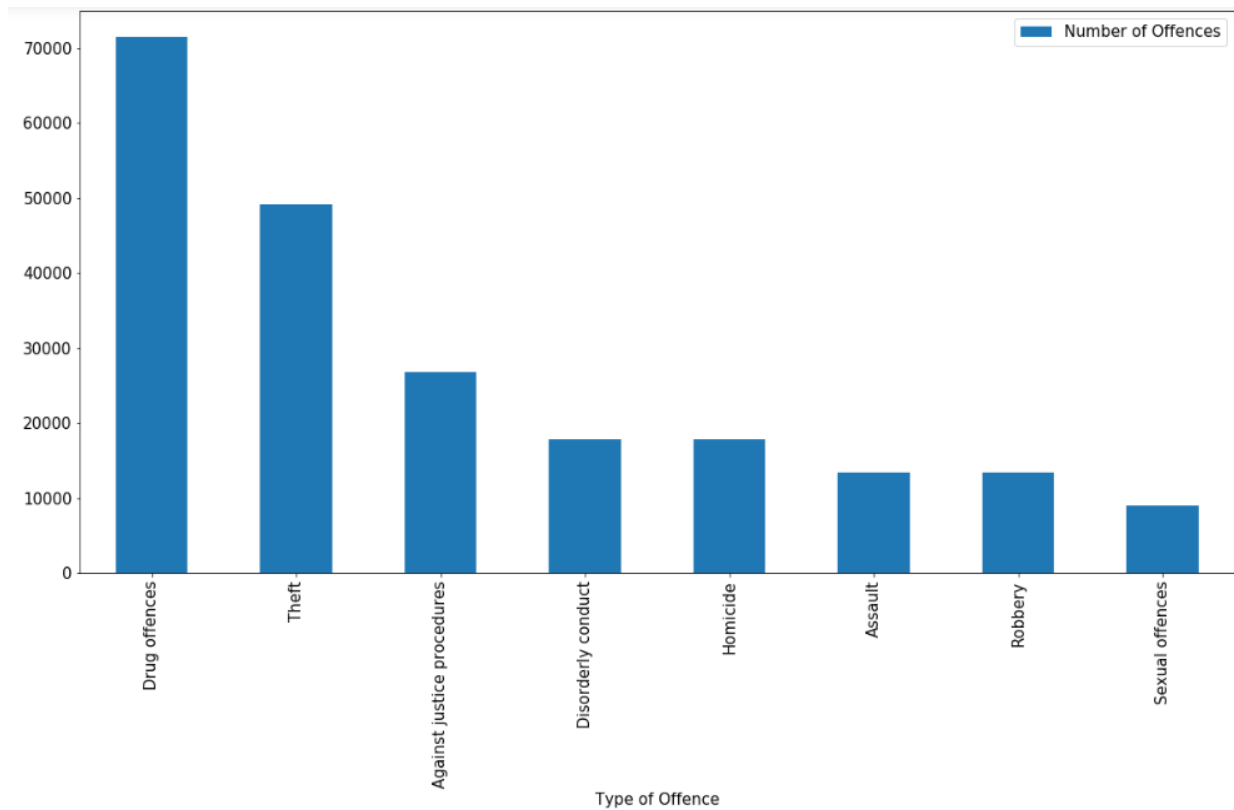
The three most types of crime as the fig above shows were related to drugs, theft and blocking judicial processes.

Similarly the figure below shows the total cumulative number of that were reported over the twenty five year period. The general trend has been a rise in the number of crimes of reported with every year. One of the many definite reasons for this is the increasing population of NSW.
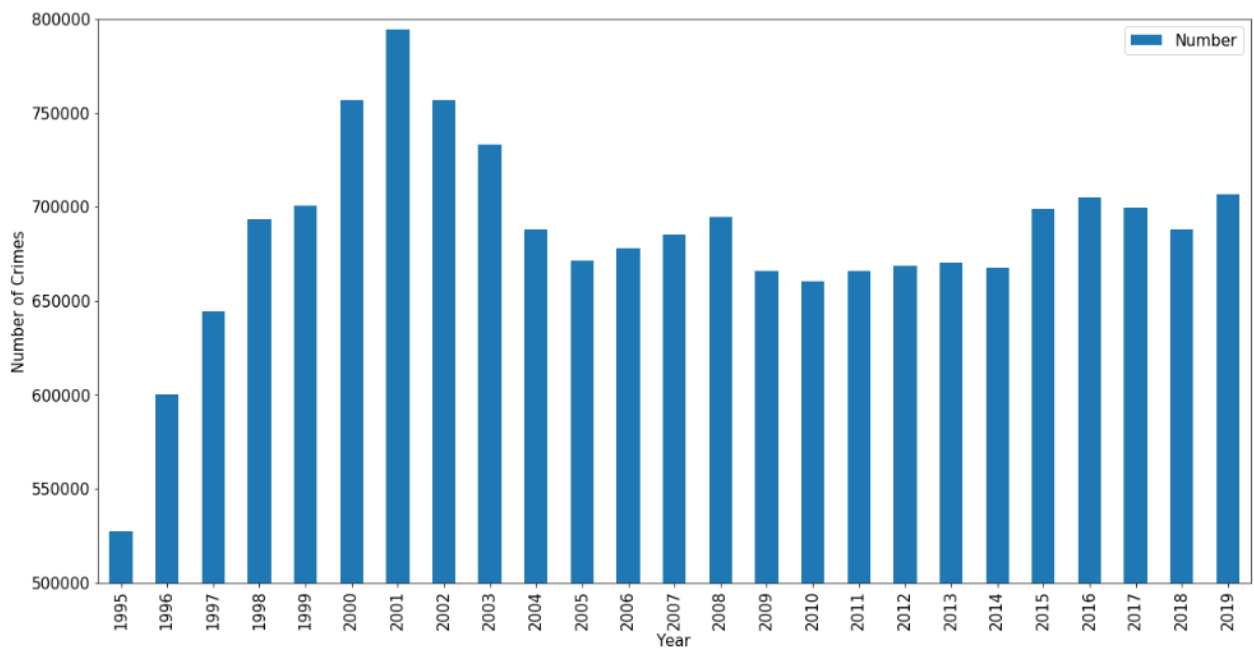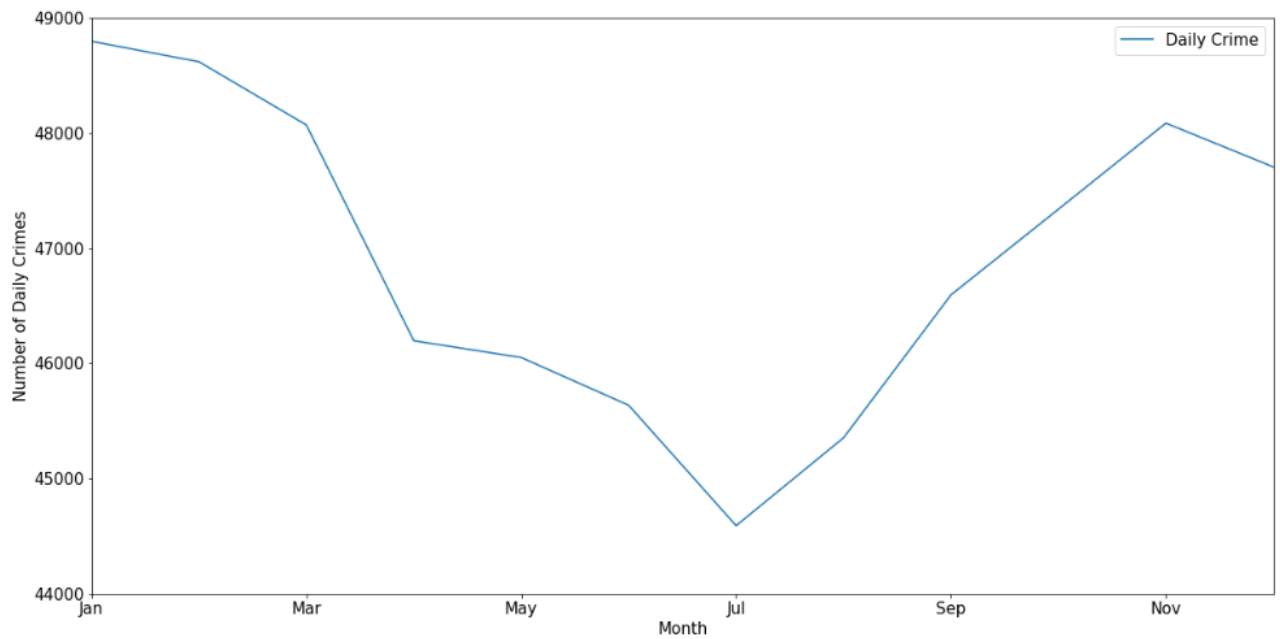
The figure above shows a monthly comparison of average number of crimes recorded daily every month for the entire 12 month period of the year. The graph above shows a clear dip in the number of crimes reported during the middle of the year. This period coincides with the winter season in NSW. The winter season is certainly marked with lesser human activities compared to the summer period which happens between the months of November to January

The fig below shows the suburbs of NSW with the highest number of reported crimes. Eleven out of the fifteen suburbs in the graph below are a part of the greater Sydney. The population of NSW is overwhelmingly metropolitan with 92% of the population of NSW residing in the greater Sydney area in the state. Higher population and population density in these suburbs explains the higher number of reported crimes.
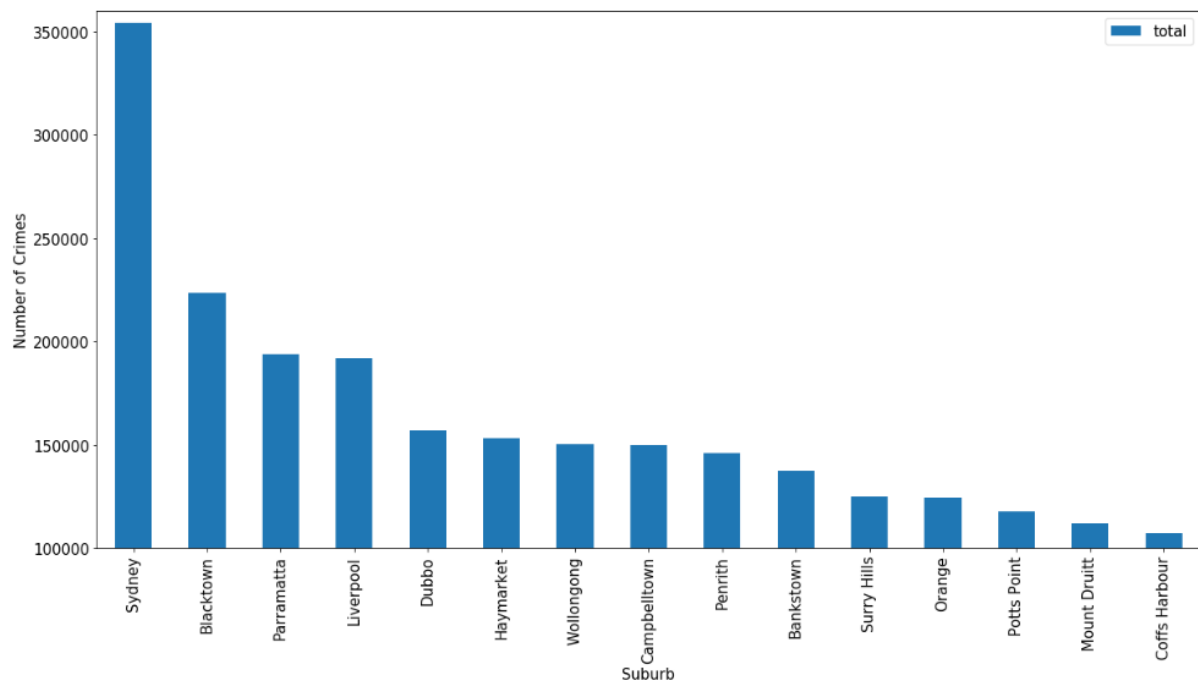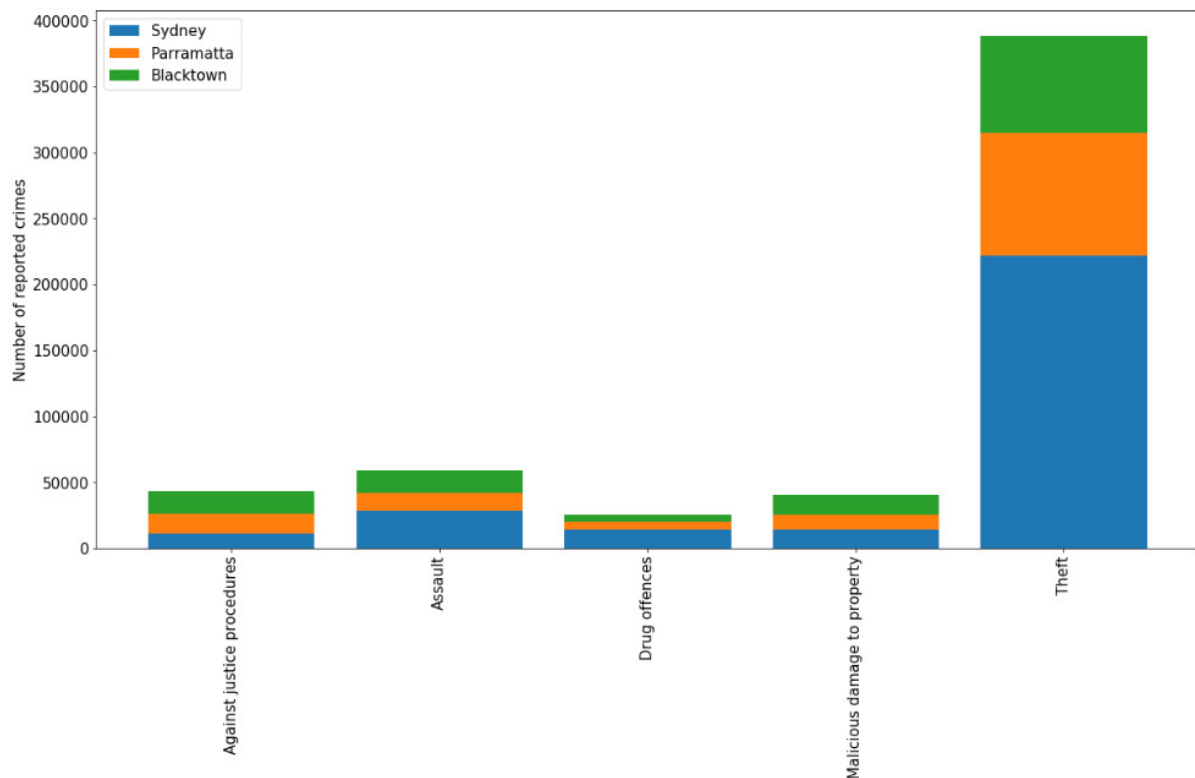
The figure above compares the five most common types of crimes recorded in the three suburbs of Sydney with high number of incidents of crime recorded. Theft by far is the most common type of crime reported. The other four most reported type of crimes have every similar number of incidents reported in the three suburbs. Sydney had the highest number of crimes reported. This is due to the fact that Sydney had the highest population and population density compared to all the suburbs in greater Sydney.

3.2 Demographics Data

The current population of Sydney is currently about 4.9 million. Population of Sydney has been increasing steadily by an average 1.2% over the last twenty years. A very large proportion of Sydney comprises of first generation immigrants. Immigrants move into Sydney primarily as

1) Skilled labour filling up white collar positions
2) Students taking up tertiary education in universities
3) Refugees fleeing violence in their home countries and regions

As the population of greater Sydney has grown, the size of the city has increased. Limits of greater Sydney have increase towards the North West, west and south west. Today the land area of Sydney is 4,195 sq. km. The population density is 1700 per sq. km. Population density reduces as you move towards the periphery of Sydney. Rapid expansion of Sydney has outpaced development of infrastructure and facilities in a number of newly settled suburbs around Sydney. As immigrants move into Sydney they choose to reside in suburbs which have a large percentage of people from they are home countries. As a result a number of regions of Sydney now have their own unique character and specialized commercial centres.
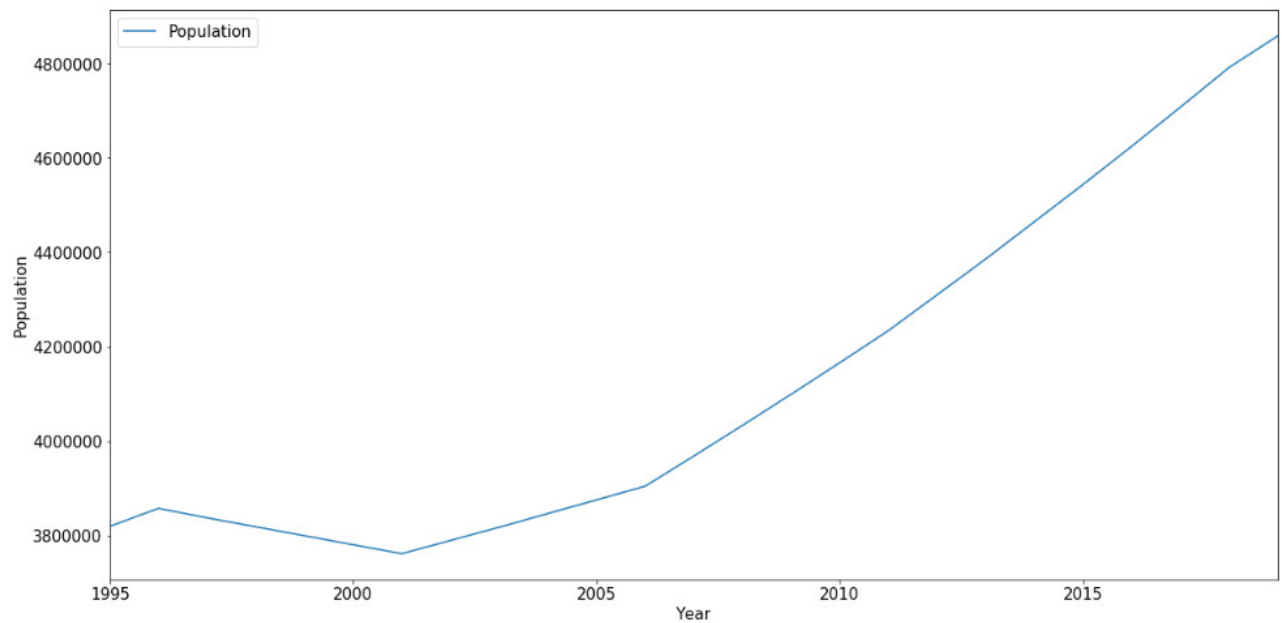
The fig above and below shows the population of Sydney over a twenty five year period from 1995 to 2019. The population of Sydney did show a slight dip in latter part of the decade of 90s. But since then it has been steadily rising at an average rate of 1.2%.
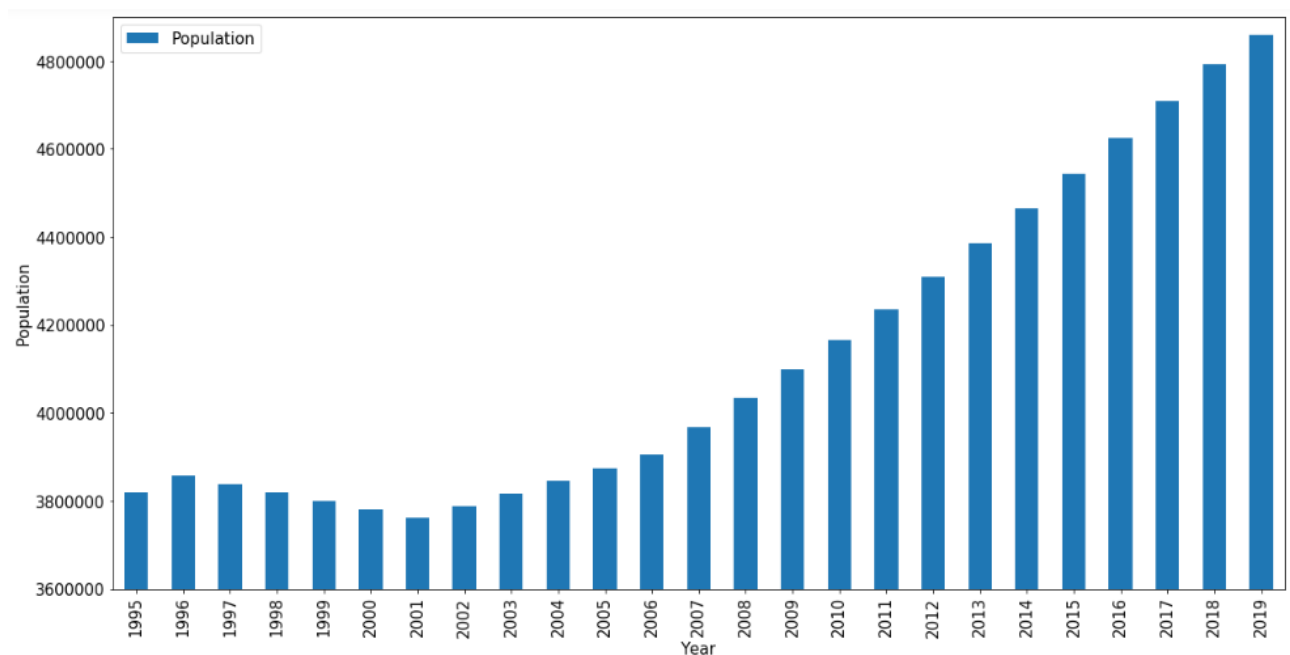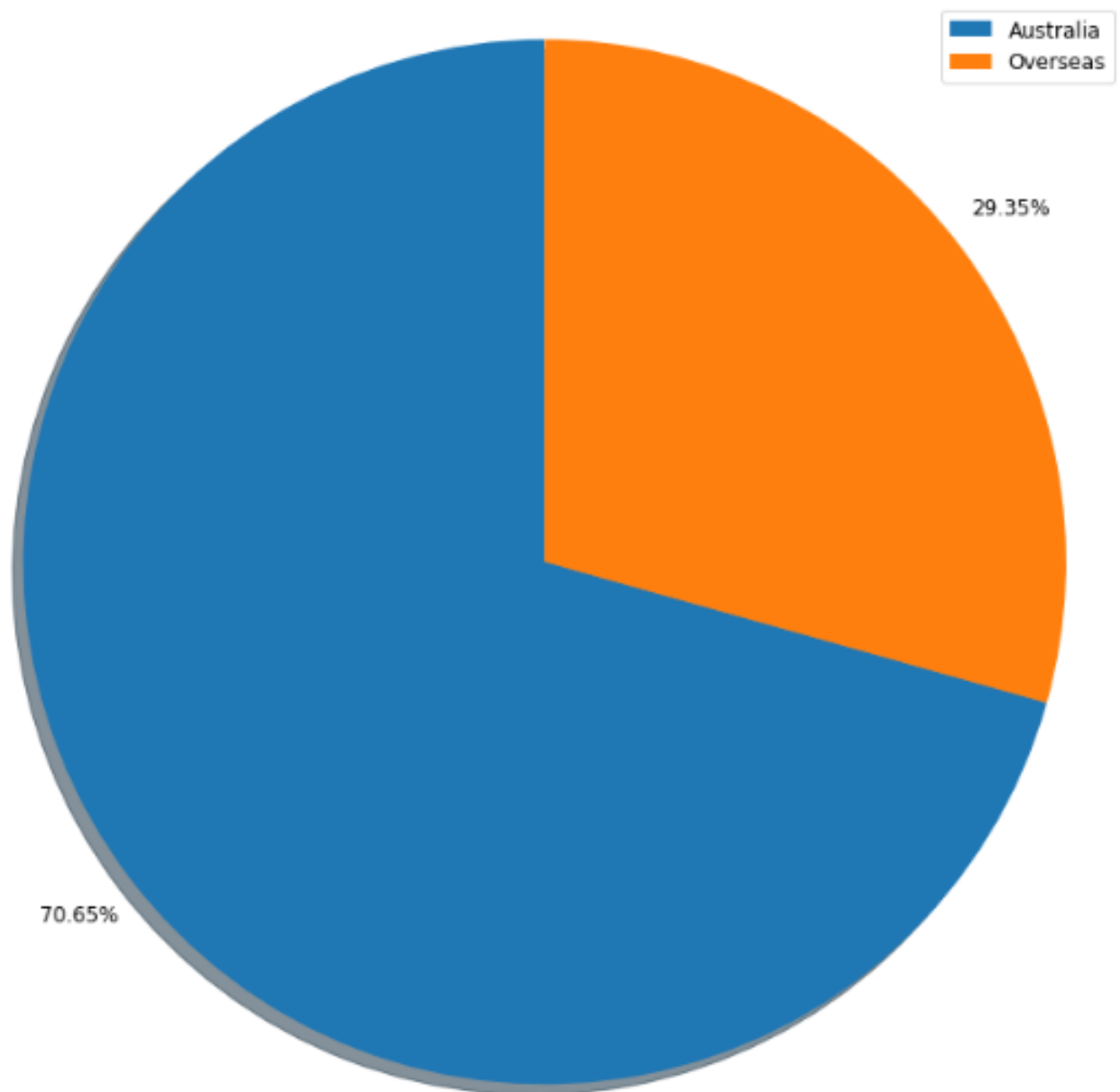
The pie chart above compares the number of people in Sydney who are born in Australia to those who are born overseas. Almost 3 out of every 10 individual in Sydney are born overseas. This clearly shows that immigrants form a sizeable proportion of the population of Sydney.
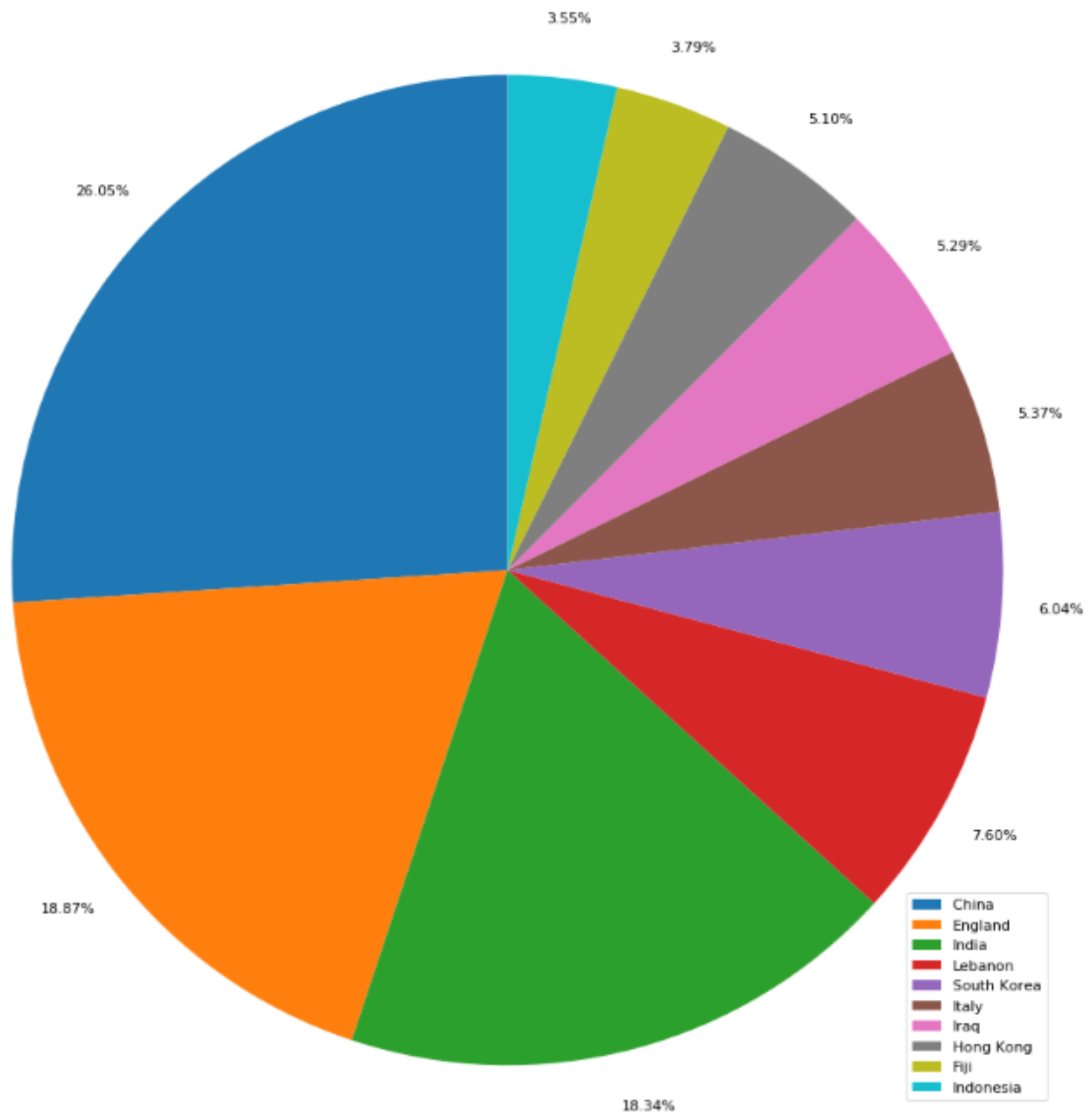
The pie chart above compares the amount of immigration from countries where from where maximum immigrants come to Sydney. Among the ten countries above the bulk of immigrants come from India, China and England. These trends have evolved with time with immigration from England and China reducing and immigration from India and Lebanon on the rise.

3.3 Venue and Business Data

Data for different types of venues and business for 575 suburbs in Sydney was acquired using the Foursquare geocoding API. The data received included information about name of venue, type of venue and its geo coordinates. Data for a total 6040 different venues and businesses under 368 venue categories was received from the API. The data was group by the venue category to see the most common type of venue and business in greater Sydney.
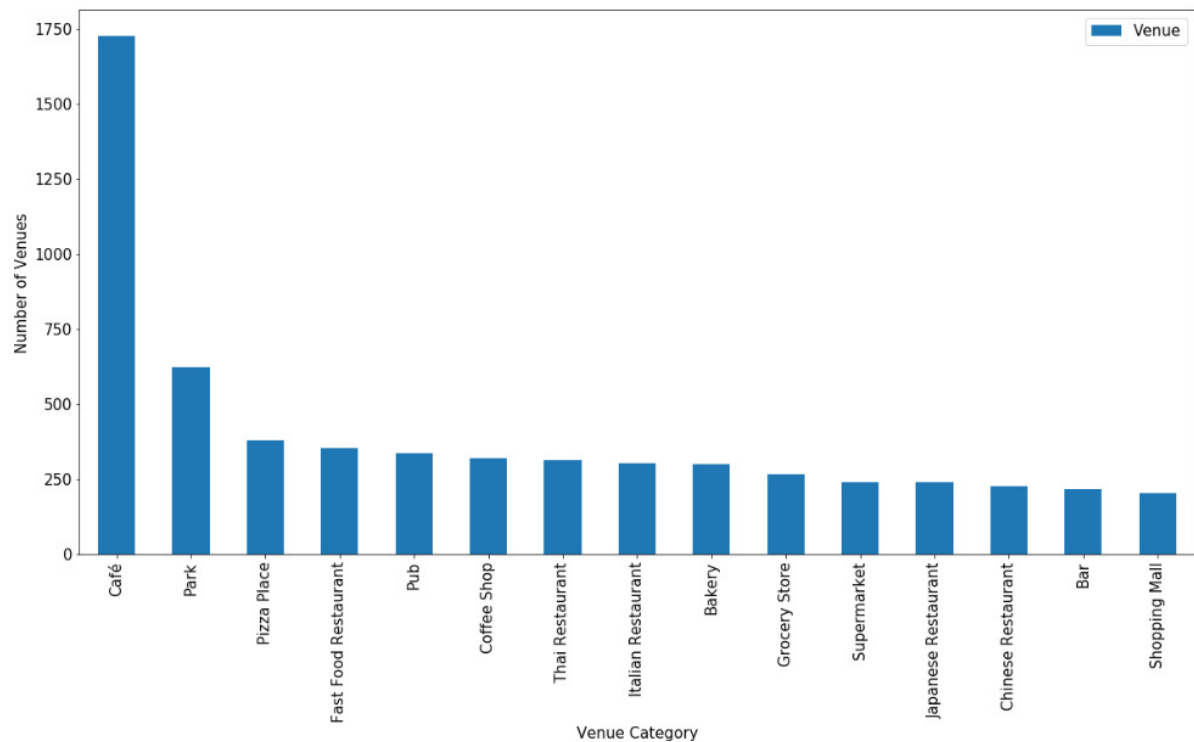
The figure above shows the most common type of venues in greater Sydney. Cafes and Parks are the most common and abundant type of venues in greater Sydney.

The fig below is a histogram of the number of suburbs vs the number of venues in Sydney. The max limit was number of venues for a suburb provided by the Foursquare API was 100.
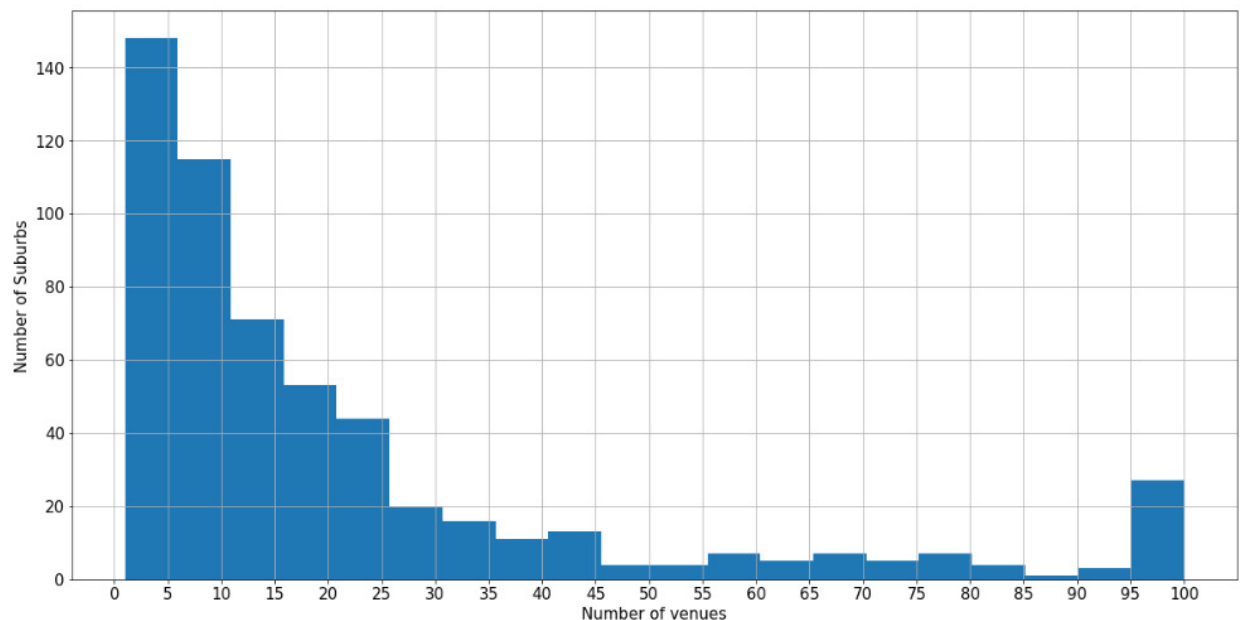
4. Data Modelling

4.1 Unsupervised learning models

The data received from Foursquare API was then processed further transferring information from a json file. The data was grouped together to find most common venues in all suburbs. It was then used train unsupervised machine learning models

Unsupervised training models such as K Means clustering is used to group suburbs having similar types and number of venues together. The data was clustered together in five different clusters. The K Means module from the sklearn library is used to cluster the data.
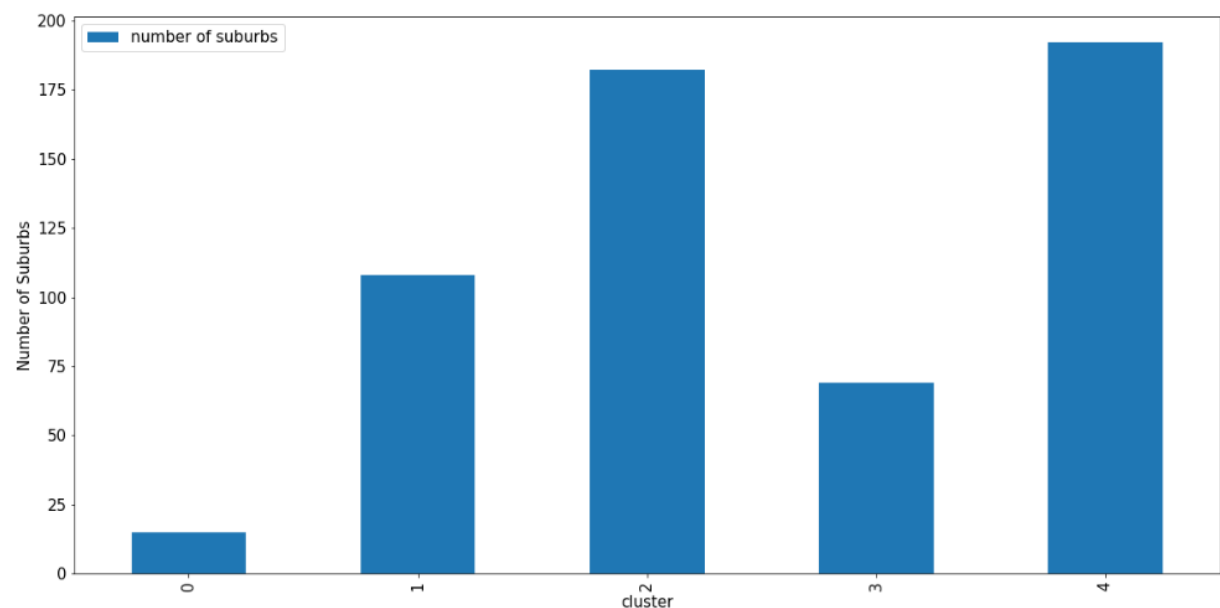

The fig above shows the number of suburbs in each cluster. Each cluster has similar type of venues in it.

The table below gives further details on each cluster.

Cluster no.	Number of Suburbs	Most Common Venues
0	15	Park, Filipino Restaurant, Field
1	108	Park, Café, Fast Food Restaurant, Field, Filipino Restaurant
2	182	Café, Park, Thai Restaurant, Bakery, Grocery Store, Italian Restaurant
3	69	Fast Food Restaurant, Pizza Place, Shopping Mall, Sandwich Place, Supermarket
4	192	Fast Food Restaurant, Filipino Restaurant, Cafes, Fish and Chip shops, Field

The processed data was then plotted on map using folium. This map contained information about the name of the suburb the number of crimes reported, number of venues and the cluster number.
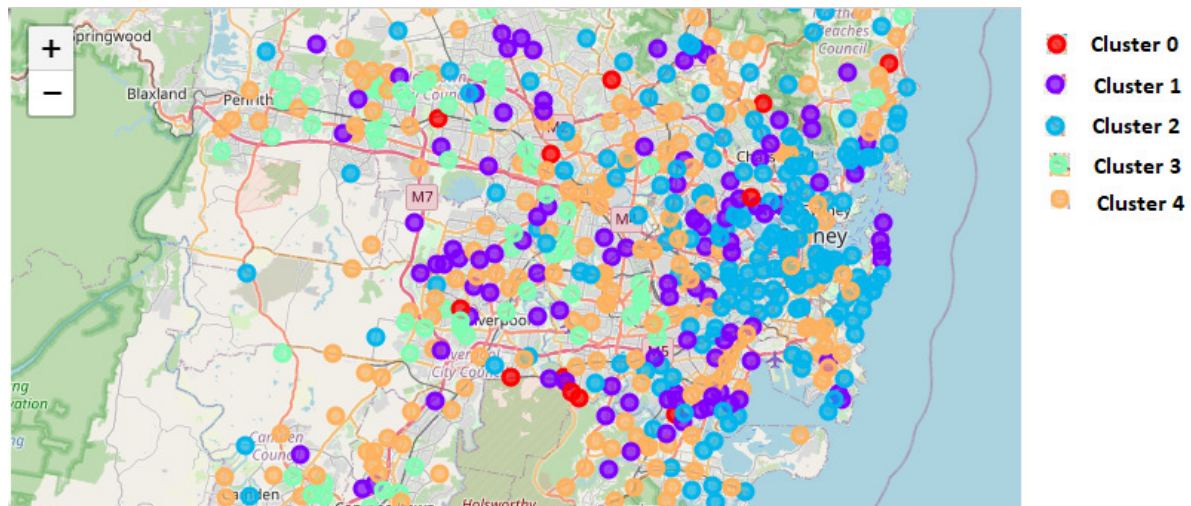
5. Conclusion

The map of greater Sydney above shows suburbs of Sydney having similar types of venues and businesses are scattered randomly all over the city.

Although Sydneysider do choose suburbs having a sizeable amount of population of similar background as their residence, venues are more randomly distributed with people travelling up to an average of twenty to twenty kilometres to venue of their choice.

The above analysis can be further expanded to model the relation between the types of venues in a suburb to the types of crimes reported from the suburb. This analysis is useful because it can then help provide some valuable insight into the trends of different types of crime in the suburbs of Greater Sydney.