**School of Information Technologies**
Faculty of Engineering & IT

## ASSIGNMENT/PROJECT COVERSHEET - GROUP ASSESSMENT

Unit of Study: _____ISYS 5050_____

Assignment name: _Knowledge Management Systems Final Project_

Tutorial time: Wed 8.00-9.00          Tutor name:

### DECLARATION

We the undersigned declare that we have read and understood the *University of Sydney Academic Dishonesty and Plagiarism in Coursework Policy*, an, and except where specifically acknowledged, the work contained in this assignment/project is our own work, and has not been copied from other sources or been previously submitted for award or assessment.

We understand that failure to comply with the *Academic Dishonesty and Plagiarism in Coursework Policy* can lead to severe penalties as outlined under Chapter 8 of the *University of Sydney By-Law 1999* (as amended). These penalties may be imposed in cases where any significant portion of my submitted work has been copied without proper acknowledgement from other sources, including published works, the internet, existing programs, the work of other students, or work previously submitted for other awards or assessments.

We realise that we may be asked to identify those portions of the work contributed by each of us and required to demonstrate our individual knowledge of the relevant material by answering oral questions or by undertaking supplementary work, either written or in the laboratory, in order to arrive at the final assessment mark.

| Project team members | | | | |
|---|---|---|---|---|
| **Student name** | **Student ID** | **Participated** | **Agree to share** | **Signature** |
| Wenjing Deng | 470349281 | Yes | Yes | Wenjing Dong |
| Kaiyu Wang | 460147130 | Yes | Yes | Kaiyu Wang |
| Jun Huang | 480244993 | Yes | Yes | Jun Huang |
| He Sun | 450105890 | Yes | Yes | *(signature)* |
| Di Li | 470485198 | Yes | Yes | Di Li |

# KNOWLEDGE MANAGEMENT SYSTEM FINAL PROJECT

Group24

| | |
|---|---|
| Wenjing Deng | 470349281 |
| Kaiyu Wang | 460147130 |
| Jun Huang | 480241993 |
| He Sun | 450105890 |
| Di Li | 470485198 |

# Content

# Introduction

With the rapid development of information technology, data analysis and management have been used in all aspects of life. In this assignment, the main target was to use the NSW crime dataset to quantify local crime events from 1995 to 2017.

As a software analysis technology, OLAP enables analysts to quickly, consistently and interactively observe information from all aspects to achieve a deep understanding of the data. Therefore, we decided to use this analysis technology to complete this assignment.
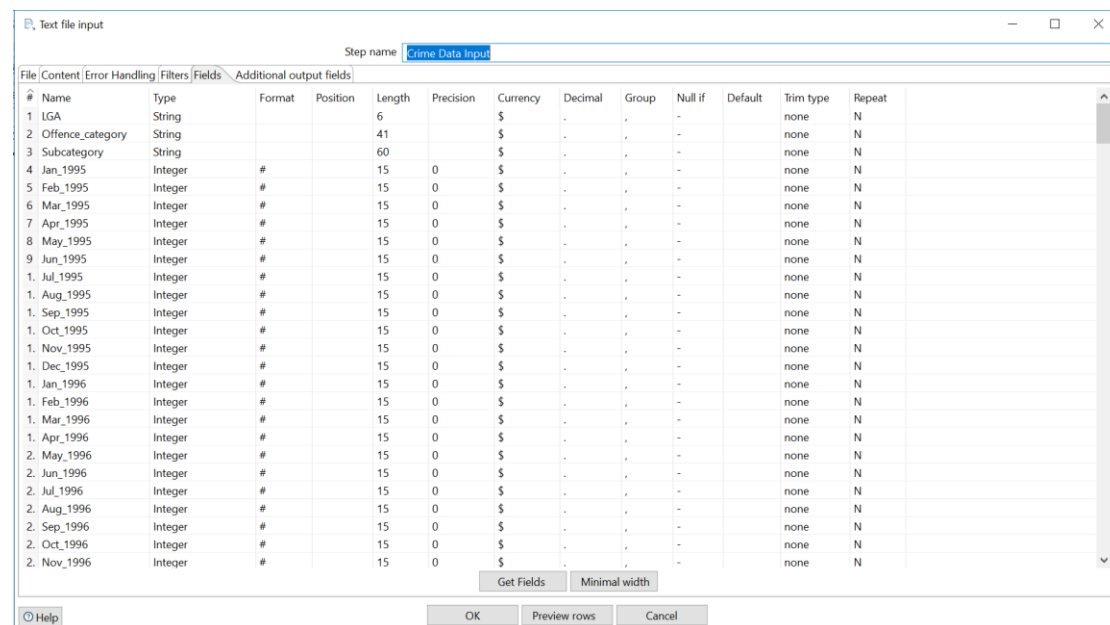
According to the requirements of the assignment, there are four tasks which we need to do. First, we do the pre-processing to transform data to the format which help to do the further work. The second one is analyzing the top five and the lowest five suburbs in terms of the total number of different types of offences committed. Then discovery the offence patterns over time to get seasonal patterns. Last, finding out if there is a tendency towards crime over time in the state.

# Part 1 Data pre-processing

In this part, we use two approaches to do the data pre-processing, one is use Pentaho Data Integration which is involved in ETL process and the other is use Python which help us to verify the correctness of our data pre-processing.

## 1.1 Pentaho Data Integration

In order to apply a professional OLAP operation, our group use Pentaho Data Integration tool to do ETL process. Firstly, use Microsoft Excel to change the format of origin dataset as csv, then select the Text File Input option to extract original raw dataset, and get fields from header to check the correctness.

It is easy to find that, the raw data has a very long header row to collect time information, and the purpose in transformation step is to simplify the header row with a date column and another related column for number of offences, so the group use row normaliser to apply transformation.
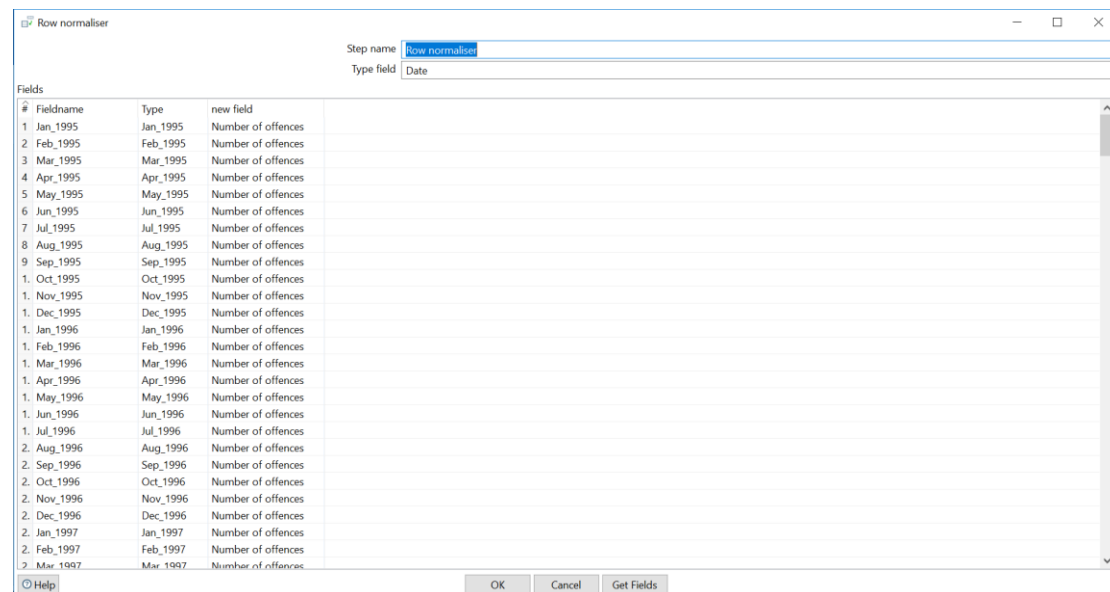


(Row Normaliser)

Finally, in the load data step, Choose Text File Output option with csv format to run the whole processing cycle. The expected data would have two million and twenty thousand rows with five columns: LGA, offence category, subcategory, date and number of offences.
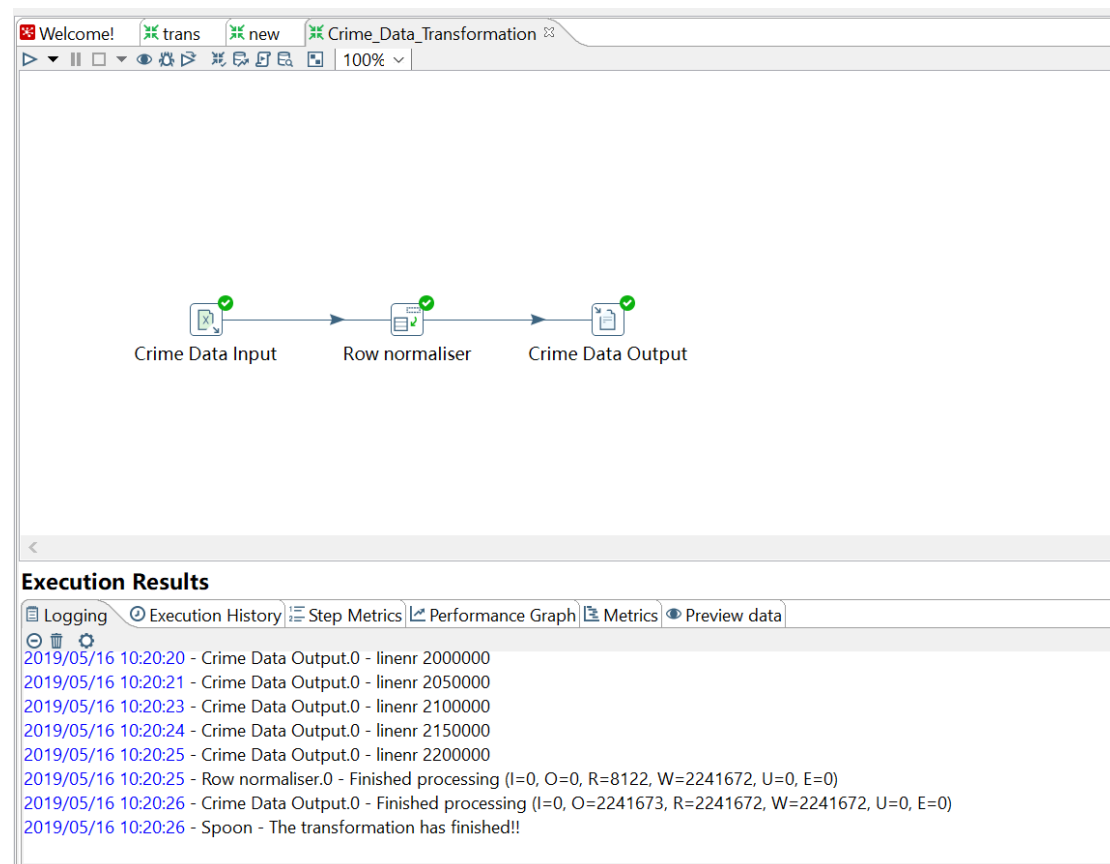
## 1.2 Python

In this part, we use Python to operate on the dataset columns and rows. The main purpose is to pivot the column to rows.

For example, here is a sample for the dataset before the pre-process.

| LGA | Offence cate | Subcategory | Jan-95 | Feb-95 | Mar-95 |
|---|---|---|---|---|---|
| Albury | Homicide | Murder * | 0 | 0 | 0 |
| Albury | Homicide | Attempted m | 0 | 0 | 0 |
| Albury | Homicide | Murder acce | 0 | 0 | 0 |
| Albury | Homicide | Manslaughte | 0 | 0 | 0 |
| Albury | Assault | Domestic vio | 7 | 7 | 7 |
| Albury | Assault | Non-domesti | 29 | 20 | 21 |

Our goal is to change the dataset with 5 columns which are the "LGA", "Offence category", "Subcategory", "date" and "count".
So, for each subcategory, we will generate several rows for different date related to its count.

| Albury | Homicide | Murder * | Jan-95 | 0 |
|---|---|---|---|---|
| Albury | Homicide | Murder * | Feb-95 | 0 |
| Albury | Homicide | Murder * | Mar-95 | 0 |
| Albury | Homicide | Attempted m | Jan-95 | 0 |
| Albury | Homicide | Attempted m | Feb-95 | 0 |
| Albury | Homicide | Attempted m | Mar-95 | 0 |
| Albury | Homicide | Murder acce | Jan-95 | 0 |
| Albury | Homicide | Murder acce | Feb-95 | 0 |
| Albury | Homicide | Murder acce | Mar-95 | 0 |
| Albury | Homicide | Manslaughte | Jan-95 | 0 |
| Albury | Homicide | Manslaughte | Feb-95 | 0 |
| Albury | Homicide | Manslaughte | Mar-95 | 0 |
| Albury | Assault | Domestic vio | Jan-95 | 7 |
| Albury | Assault | Domestic vio | Feb-95 | 7 |
| Albury | Assault | Domestic vio | Mar-95 | 7 |
| Albury | Assault | Non-domesti | Jan-95 | 29 |
| Albury | Assault | Non-domesti | Feb-95 | 20 |
| Albury | Assault | Non-domesti | Mar-95 | 21 |

We are using python to realize this approach and the code screenshot is shown below.

```python
import csv
import pandas as pd
with open('sample.csv', 'rt',encoding = 'utf-8') as f:
    reader = csv.reader(f)
    your_list = list(reader)
print(your_list)
new_list = []
count_of_date = len(your_list[0]) - 3
for items in your_list:
    if items[0] == "LGA":
        continue;
    else:
        for i in range(count_of_date):
            new_item = items[0:3]
            new_item.insert(len(new_item),your_list[0][i+3])
            new_item.insert(len(new_item),items[i+3])
            print(new_item)
            new_list.append(new_item)
# print(len(new_list))


item_dict = {}
for i in range(len(new_list)):
    item_dict.update({i:new_list[i]})


df = pd.DataFrame(
    item_dict
)
df = df.T
df.to_csv('sample1.csv',index = False)
print("process finished")
```

The origin csv file is read and a list with all rows is generated. After that, we count the rows of the date and saved it as the parameter "count_of_date". This parameter will be the number of rows for each subcategory after the shift. For each row, we first create the sub-list for only "LGA", "Offence category" and "subcategory", and then, for each column of date from original data, we generate a new item with it and append the count of crime on that date to it. The final result will be a list, every item in that list consist of 5 attributes, which are the "LGA", "Offence category", "Subcategory", "date" and "count". The list will be saved to a new csv file and the pivot process is finished.

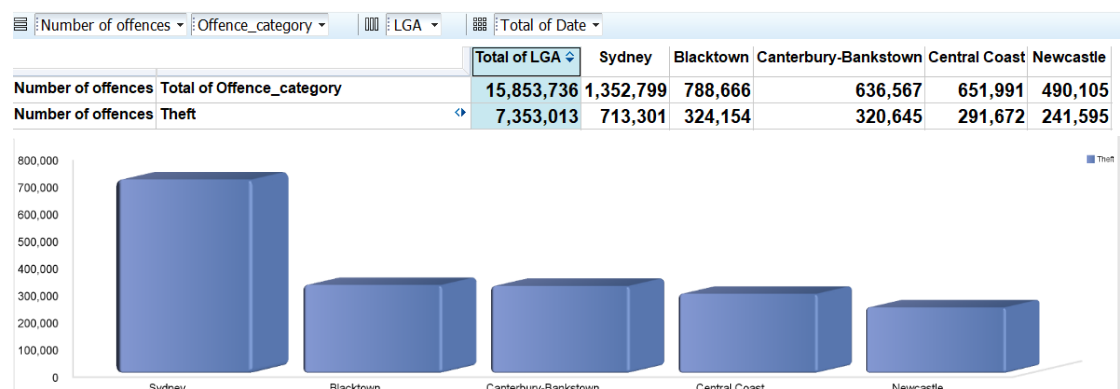# Part 2 The top five and the lowest five suburbs analysis

In this step we first slice the whole dataset in in top five categories with total number of offences. With the simple sort of aggregated dataset, through the table, it is easily to find that the top five categories of offence are: theft, malicious damage to property, assault, transport regulatory offence and against justice procedures.

| ☰ LGA ▾ | ▥ Number of offences ▾ | Offence_category ▾ | | ▦ Total of Date ▾ | | |
|---|---|---|---|---|---|---|
| | **Number of offences** | **Number of offences** | **Number of offences** | **Number of offences** | **Number of offences** | |
| | **Total of Offence_category** | **Theft** ⇕ | **Malicious damage to property** | **Assault** | **Transport regulatory offences** | **Against justice procedures** |
| **Total of LGA** ⟺ | 12,743,598 | 7,353,013 | 2,009,480 | 1,464,048 | 1,013,769 | 903,288 |

Then in each category we aggregate top five and lowest suburbs in number of offences below with table and bar chart analysis.

## 2.1 Theft

In term of the data for theft offence, the table illustrates the distribution of the value in top five suburbs with comparison with the total value of offences in an area. It can be seen that the number of theft category is nearly the half of the total offence in LGA. Through the bar chart below we can see the clear visualization of the horizontal comparison in value of number of theft offences, and it easily to find the total number of offences in Sydney is much more than any other four locations, which is Blacktown, Canterbury-Bankstown, Central Coast, and Newcastle.

| | | Total of LGA ⇕ | Sydney | Blacktown | Canterbury-Bankstown | Central Coast | Newcastle |
|---|---|---|---|---|---|---|---|
| Number of offences | Total of Offence_category | 15,853,736 | 1,352,799 | 788,666 | 636,567 | 651,991 | 490,105 |
| Number of offences | Theft | 7,353,013 | 713,301 | 324,154 | 320,645 | 291,672 | 241,595 |



The second table below is the summary of five suburbs with lowest number of theft offences. Then, use the same way to see the bar chart of the lowest five suburb data visualization. Obviously, the lowest five suburbs are Waicha, Lockhart, In Custody, Unincorporated Far West, and Lord Howe Island. However, the number of theft category is still occupied above 30% of the total number in major lowest five suburbs.
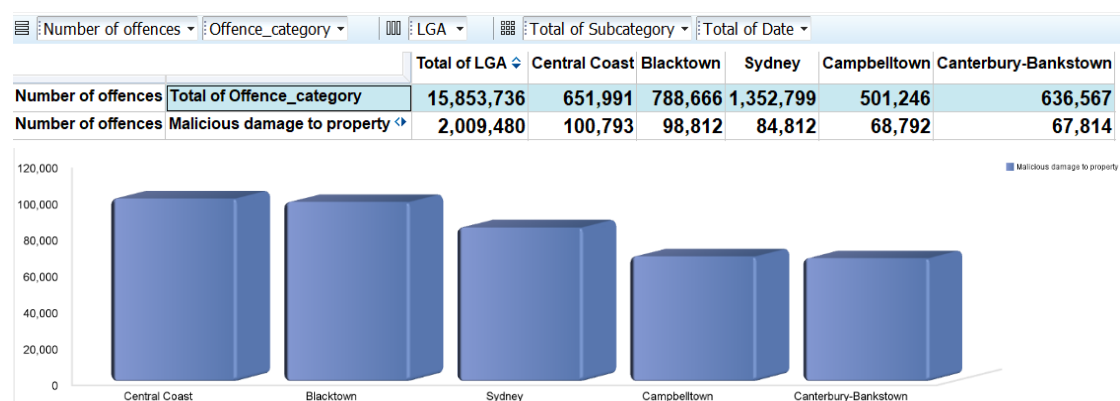
| | | Total of LGA ⇕ | Walcha | Lockhart | In Custody | Unincorporated Far West | Lord Howe Island |
|---|---|---|---|---|---|---|---|
| Number of offences | Total of Offence_category | 15,853,736 | 4,899 | 3,945 | 49,963 | 908 | 258 |
| Number of offences | Theft | 7,353,013 | 1,560 | 1,402 | 1,381 | 379 | 78 |
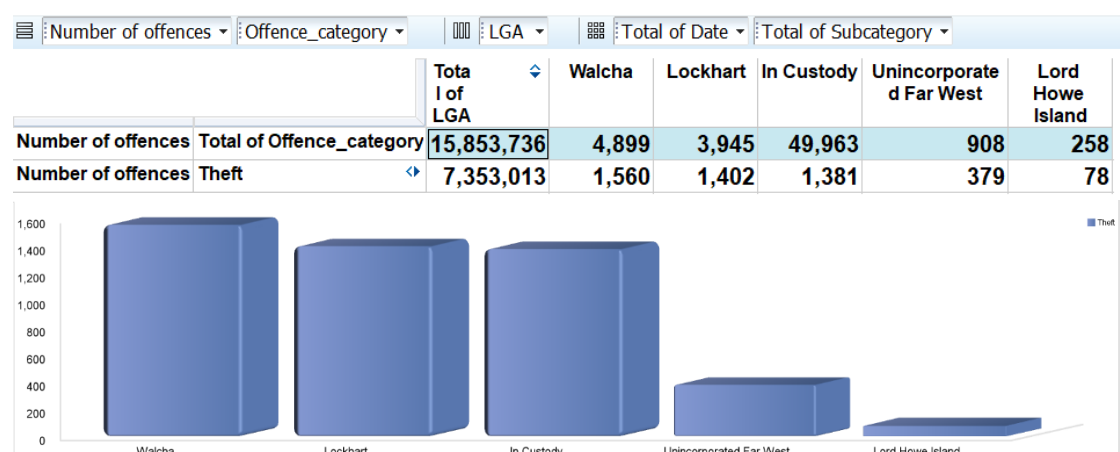


## 2.2 Malicious damage to property

As for the top five suburbs of malicious damage to property we can see from the table below. They are Central Coast, Blacktown, Sydney, Campbelltown, Canterbury-Bankstown. The vast majority are the same with theft except Campbelltown.

| | | Total of LGA ⇕ | Central Coast | Blacktown | Sydney | Campbelltown | Canterbury-Bankstown |
|---|---|---|---|---|---|---|---|
| Number of offences | Total of Offence_category | 15,853,736 | 651,991 | 788,666 | 1,352,799 | 501,246 | 636,567 |
| Number of offences | Malicious damage to property | 2,009,480 | 100,793 | 98,812 | 84,812 | 68,792 | 67,814 |

Then the lowest five suburbs in malicious damage of property is Carrathod, Walcha, Lockhart, Unincorporated Far West, and Lord Howe Island.

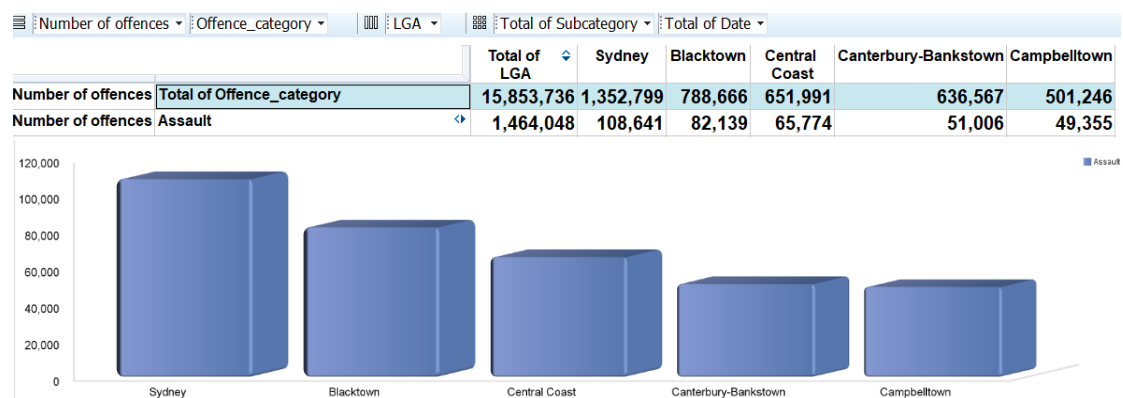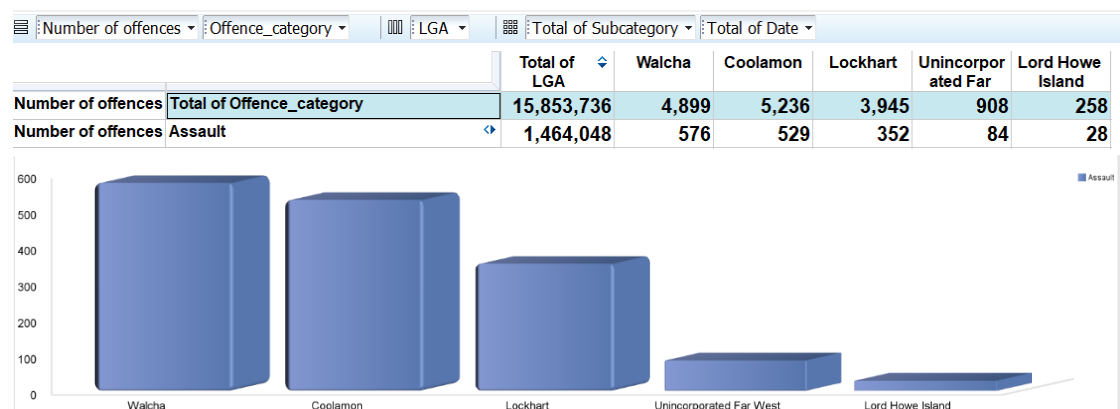| Number of offences ▾ | Offence_category ▾ | | LGA ▾ | | Total of Subcategory ▾ | Total of Date ▾ | | |
|---|---|---|---|---|---|---|---|---|
| | | | Total of LGA ⇕ | Carrathool | Walcha | Lockhart | Unincorpor ated Far | Lord Howe Island |
| Number of offences | Total of Offence_category | | 15,853,736 | 6,132 | 4,899 | 3,945 | 908 | 258 |
| Number of offences | Malicious damage to property ◄► | | 2,009,480 | 692 | 635 | 573 | 89 | 31 |
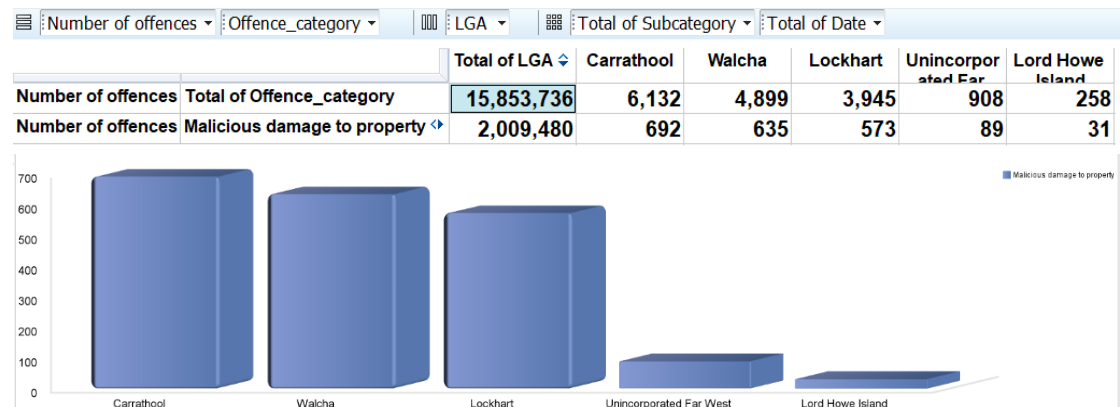


## 2.3 Assault

For assault category, the top five suburbs are Sydney, Blacktown, Central Coast, Canterbury-Bankstown, Campbelltown and these categories is totally the same with the Malicious damage to property.
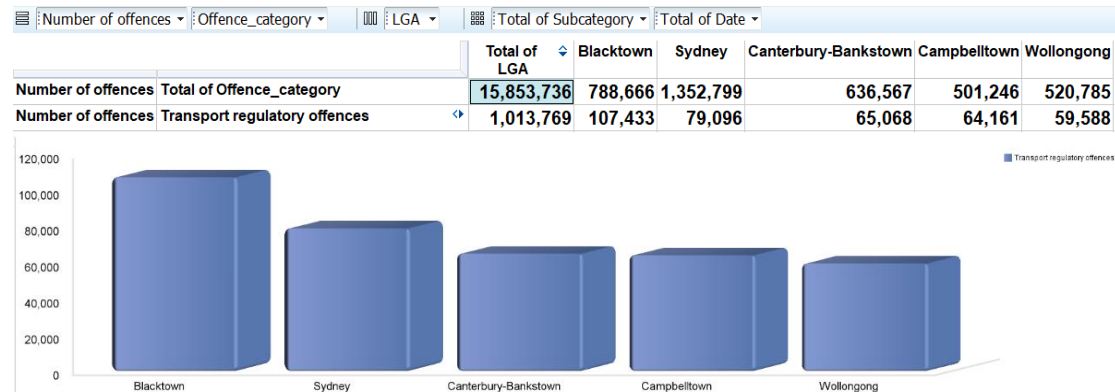
| Number of offences ▾ | Offence_category ▾ | | LGA ▾ | | Total of Subcategory ▾ | Total of Date ▾ | | |
|---|---|---|---|---|---|---|---|---|
| | | | Total of LGA ⇕ | Sydney | Blacktown | Central Coast | Canterbury-Bankstown | Campbelltown |
| Number of offences | Total of Offence_category | | 15,853,736 | 1,352,799 | 788,666 | 651,991 | 636,567 | 501,246 |
| Number of offences | Assault | ◄► | 1,464,048 | 108,641 | 82,139 | 65,774 | 51,006 | 49,355 |



The lowest five suburbs in assault is Walcha, Coolamon, Lockhart, Unincorporated Far West, and Lord Howe Island.

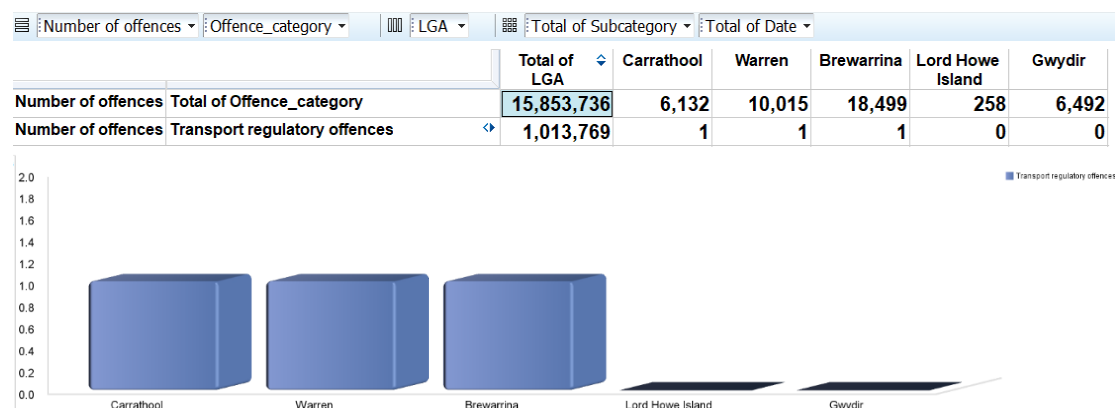| Number of offences ▾ | Offence_category ▾ | | LGA ▾ | | Total of Subcategory ▾ | Total of Date ▾ | | |
|---|---|---|---|---|---|---|---|---|
| | | | Total of LGA ⇕ | Walcha | Coolamon | Lockhart | Unincorpor ated Far | Lord Howe Island |
| Number of offences | Total of Offence_category | | 15,853,736 | 4,899 | 5,236 | 3,945 | 908 | 258 |
| Number of offences | Assault | ◄► | 1,464,048 | 576 | 529 | 352 | 84 | 28 |

## 2.4 Transport regulatory offence

In terms of transport regulatory offence, the top suburb is Blacktown, Sydney, Canterbury-Bankstown, Campbelltown and Wollongong. In this category, Wollongong is the first time happens in the top five.
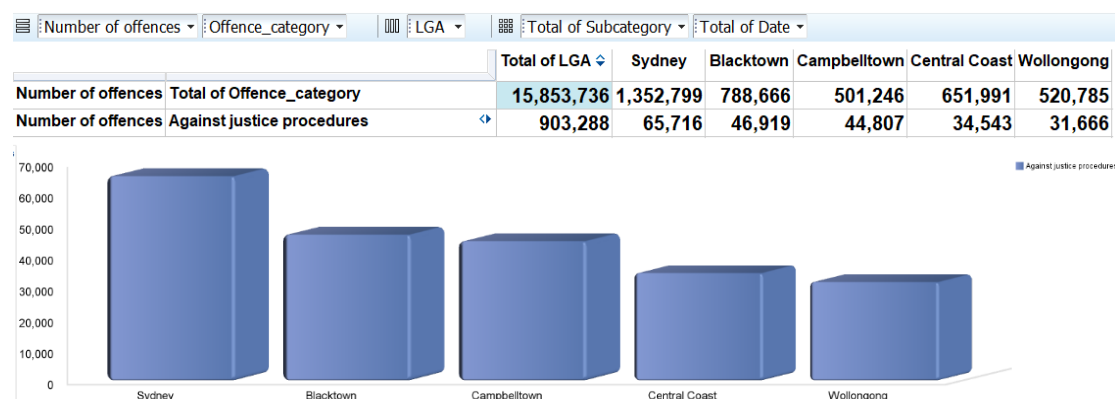
| Number of offences ▾ | Offence_category ▾ | | Total of LGA ⬍ | Blacktown | Sydney | Canterbury-Bankstown | Campbelltown | Wollongong |
|---|---|---|---|---|---|---|---|---|
| Number of offences | Total of Offence_category | | 15,853,736 | 788,666 | 1,352,799 | 636,567 | 501,246 | 520,785 |
| Number of offences | Transport regulatory offences | ◂▸ | 1,013,769 | 107,433 | 79,096 | 65,068 | 64,161 | 59,588 |



There are five lowest suburbs, which are Carrathool, Warren, Brewarrina, Lord Howe Island, and Gwydir which Lord Howe Island and Gwydir both are 0.

| Number of offences ▾ | Offence_category ▾ | | Total of LGA ⬍ | Carrathool | Warren | Brewarrina | Lord Howe Island | Gwydir |
|---|---|---|---|---|---|---|---|---|
| Number of offences | Total of Offence_category | | 15,853,736 | 6,132 | 10,015 | 18,499 | 258 | 6,492 |
| Number of offences | Transport regulatory offences | ◂▸ | 1,013,769 | 1 | 1 | 1 | 0 | 0 |



## 2.5 Against justice procedures

The top five suburbs of the against justice procedures are the same with the category of transport regulatory offence, though there is a different order between them.

| Number of offences ▾ | Offence_category ▾ | | Total of LGA ⬍ | Sydney | Blacktown | Campbelltown | Central Coast | Wollongong |
|---|---|---|---|---|---|---|---|---|
| Number of offences | Total of Offence_category | | 15,853,736 | 1,352,799 | 788,666 | 501,246 | 651,991 | 520,785 |
| Number of offences | Against justice procedures | ◂▸ | 903,288 | 65,716 | 46,919 | 44,807 | 34,543 | 31,666 |

When it comes to lowest five suburbs, they are Upper Lachian Shire, Coolamon, Lockhart, Unicorporated Far West, and Lord Howe Island.

| | | Total of LGA ⇕ | Upper Lachlan | Coolamon | Lockhart | Unincorpor ated Far | Lord Howe Island |
|---|---|---|---|---|---|---|---|
| Number of offences | Total of Offence_category | 15,853,736 | 8,717 | 5,236 | 3,945 | 908 | 258 |
| Number of offences | Against justice procedures ◇ | 903,288 | 264 | 210 | 138 | 47 | 3 |
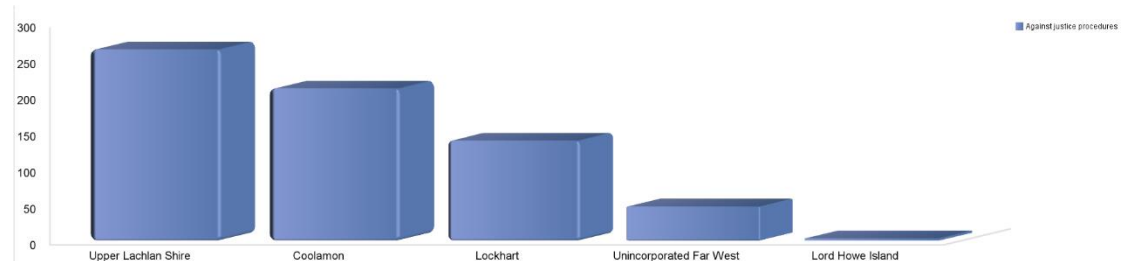


Through the information above, it is easily to find that in term of top five suburbs, Sydney as well as Blacktown emerges in each category as one of the top five suburb, and Lord Howe Island always belong to one of the lowest five suburbs with number of offences no matter which category it is.

After that, we roll up the dataset to see any differences in term of total number of number of offences.

The table below is the data of top five regions with first five most total crime numbers in NSW from 1995 to 2017. And the bar chart below is the visualization of number of offences in top five categories among those five regions. It is easily to see that, the number of offences in Sydney and Blacktown are top 2 among these regions which the number of theft category is occupied the most in Sydney, and the transport regulatory offences are the most in Blacktown.
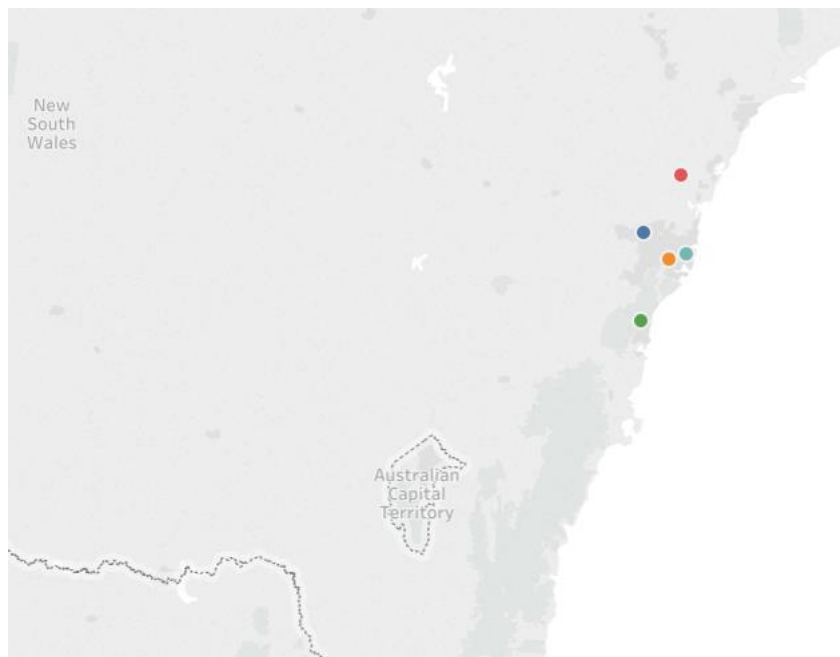
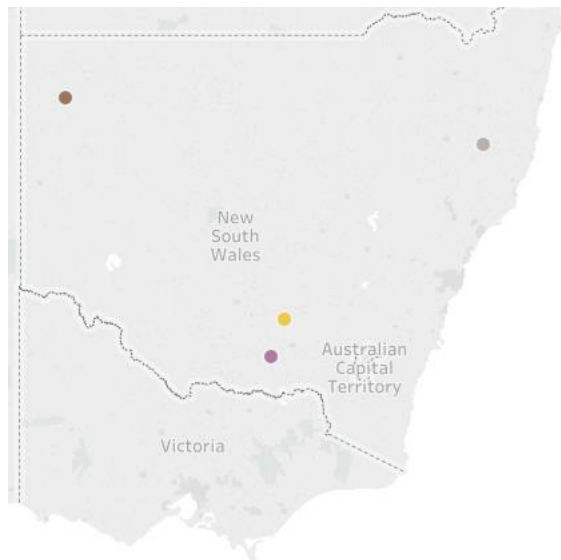| | Number of offences | Number of offences | Number of offences | Number of offences | Number of offences | Number of offences |
|---|---|---|---|---|---|---|
| | Total of Offence category ⇕ | Theft | Assault | Malicious damage to property | Transport regulatory offences | Against justice procedures |
| Sydney ◇ | 1,352,799 | 713,301 | 108,641 | 84,812 | 79,096 | 65,716 |
| Blacktown | 788,666 | 324,154 | 82,139 | 98,812 | 107,433 | 46,919 |
| Central Coast | 651,991 | 291,672 | 65,774 | 100,793 | 48,699 | 34,543 |
| Canterbury-Bankstown | 636,567 | 320,645 | 51,006 | 67,814 | 65,068 | 30,738 |
| Wollongong | 520,785 | 235,834 | 42,013 | 67,482 | 59,588 | 31,666 |
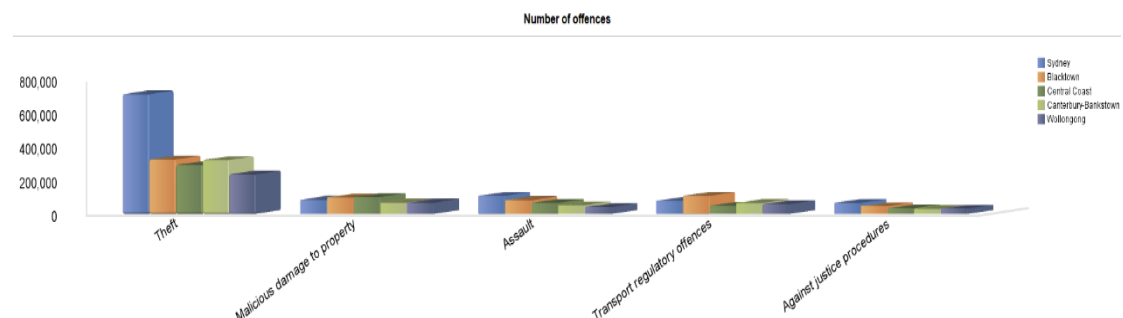


The table below is the data of lowest five regions with first five most total crime numbers in NSW from 1995 to 2017. And the bar chart below is the visualization of number of offences in lowest five categories among those five regions. It is easily to see that, the number of offences in Lord Howe Island has lowest number of total offences.

| | Number of offences | Number of offences | Number of offences | Number of offences | Number of offences | Number of offences |
|---|---|---|---|---|---|---|
| | Total of Offence_category ⬦ | Theft | Malicious damage to property | Assault | Liquor offences | Disorderly conduct |
| Lord Howe Island | 258 | 78 | 31 | 28 | 13 | 14 |
| Unincorporated Far West | 908 | 379 | 89 | 84 | 21 | 52 |
| Lockhart | 3,945 | 1,402 | 573 | 352 | 365 | 190 |
| Walcha | 4,899 | 1,560 | 635 | 576 | 171 | 354 |
| Coolamon | 5,236 | 1,901 | 831 | 529 | 301 | 293 |



Through comparison between top five and lowest five aggregated data we find, the five top five categories are different in these two tables. Although the top three categories in these two tables are theft, malicious damage to property and assault, the last two categories for top five offence suburb tables are transport regulatory offences and again justice procedures, and the last two categories for lowest five offence suburb tables are liquor offences as well as disorderly conduct. Therefore, we highlight the related suburbs on the map below. The first diagram is the highlight of top five suburbs in the geolocation.



Then, we use the same way to highlight the lowest five places by using tableau to show the geolocation on the map chart.

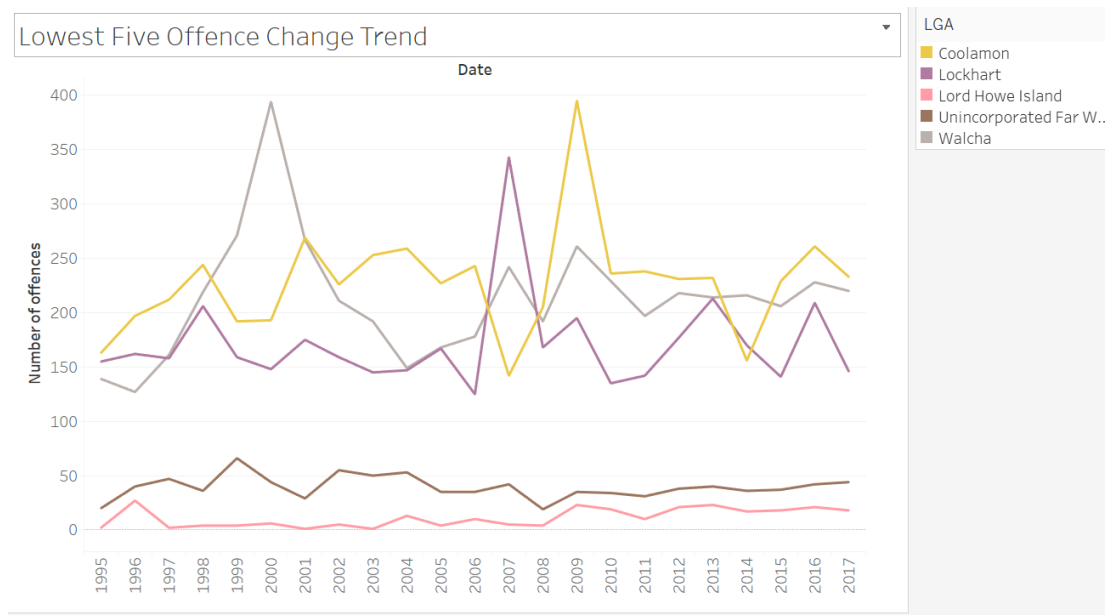| | Number of offences | Number of offences | Number of offences | Number of offences | Number of offences | Number of offences |
|---|---|---|---|---|---|---|
| | Total of Offence_category ⬦ | Theft | Malicious damage to property | Assault | Liquor offences | Disorderly conduct |
| Lord Howe Island | 258 | 78 | 31 | 28 | 13 | 14 |
| Unincorporated Far West | 908 | 379 | 89 | 84 | 21 | 52 |
| Lockhart | 3,945 | 1,402 | 573 | 352 | 365 | 190 |
| Walcha | 4,899 | 1,560 | 635 | 576 | 171 | 354 |
| Coolamon | 5,236 | 1,901 | 831 | 529 | 301 | 293 |

It can infer by comparison that the top five suburbs are located closely in southeast coastal areas where have advanced transportation and complete government services. However, the lowest five areas are located in different remote and backward areas, and that maybe the main cause of the differences in last two top five offence categories.

Apart from that, the exploration of the total number of offences change by year is also important to know the whole tendency. The first line chart below is the visualization of the number of offences in top five suburbs changing tendency over these 22 years.
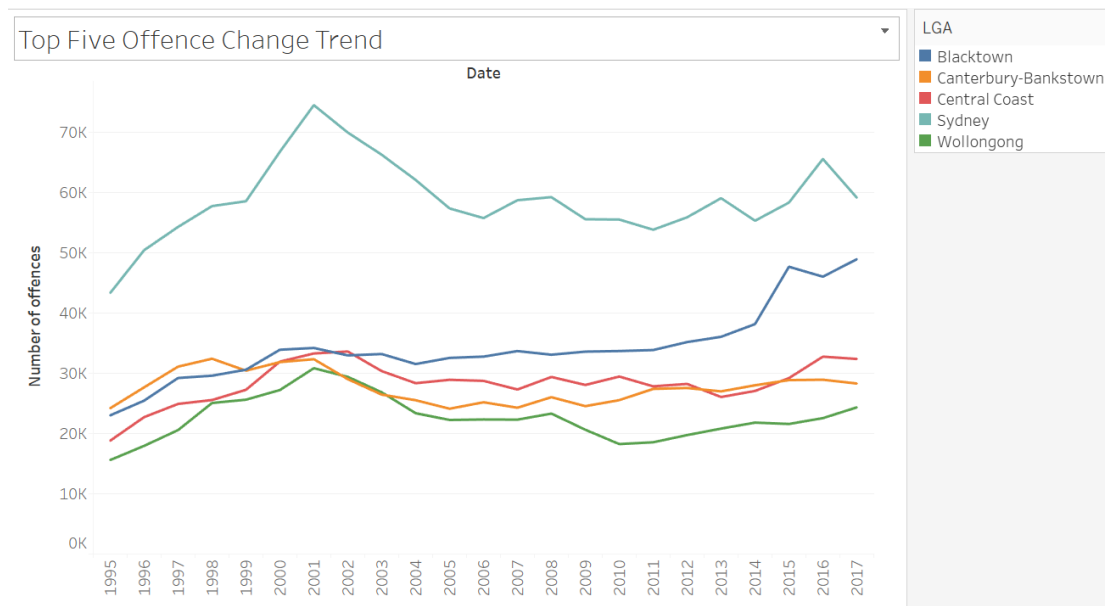


It is easily to find that although Sydney has the largest value of offence number, it has the same changing trend as Canterbury-Bankstown, Central Coast and Wollongong, which fluctuate with time went by and got peak at around 2000. However, the changing pattern in Blacktown is quite different, it kept growing in every single year, and may exceed the value of Sydney in the future.

Moving on to the visualization of 5 lowest number of offences in New South Wales. The changing trend is summarized by the line chart below.

The number of offences in Lord Howe and Unincorporated Far West are stable and low, but the value in another three places is changed sharply among these years.

## In Summary



From pie chart above, we can find that the value of number of offences in top five categories account for largest percentage of total proportion, and with the analysis of top five and lowest five suburbs in term of category and total number above, we could summarize that use theft, malicious damage to property and assault as three main types could predict and reflect the total offence situation in a location without any distribution, as the transport regulatory offences and against justice offence might be influenced by geographical location or local government control . Moreover, related experts should pay more attention to Sydney and Blacktown, which always have largest number of offences and quick increasing trend respectively.

# Part 3 Discover the patterns over time

This part is to use some OLAP operations to discover the patterns over time.
The first entry point for this topic is to seek some data characteristic between Christmas holiday and the non-holiday period to see if there is a pattern during holidays. So we chose January and December which is the Christmas holiday and March and April as the representation of non-holiday. The X axis represents the crime category and the Y axis represents the number of the crime. It can be easily shown in the two figures below, most of the offence categories has a greater number during Christmas holiday than non-holiday, such as the categories of theft, assault, and against justice procedure which are the top 3 number of the offence.
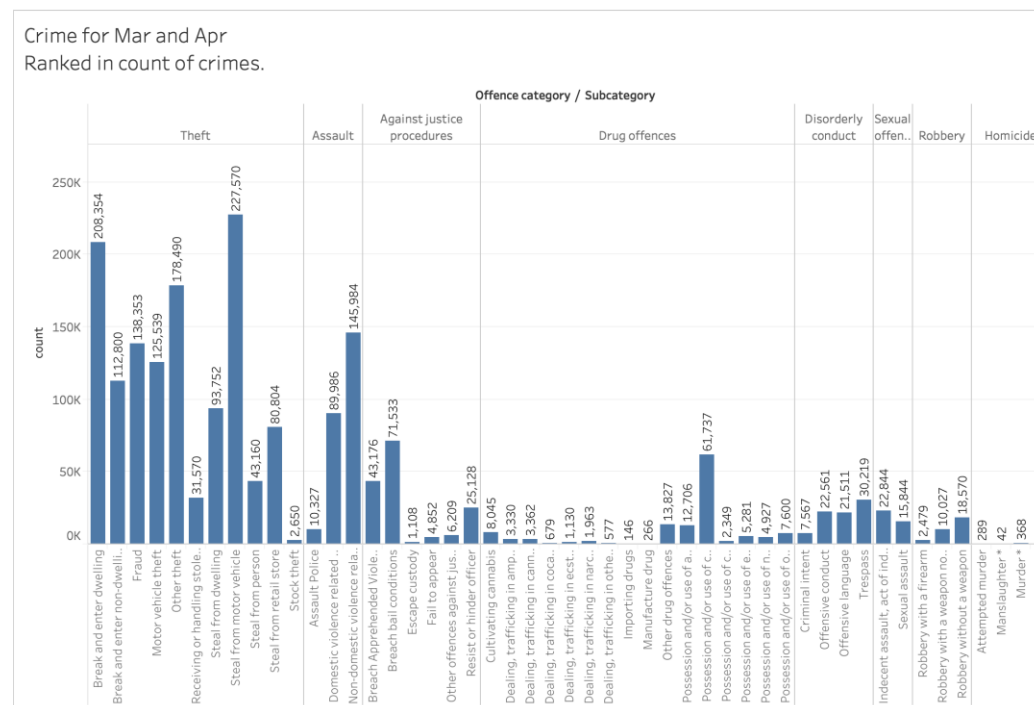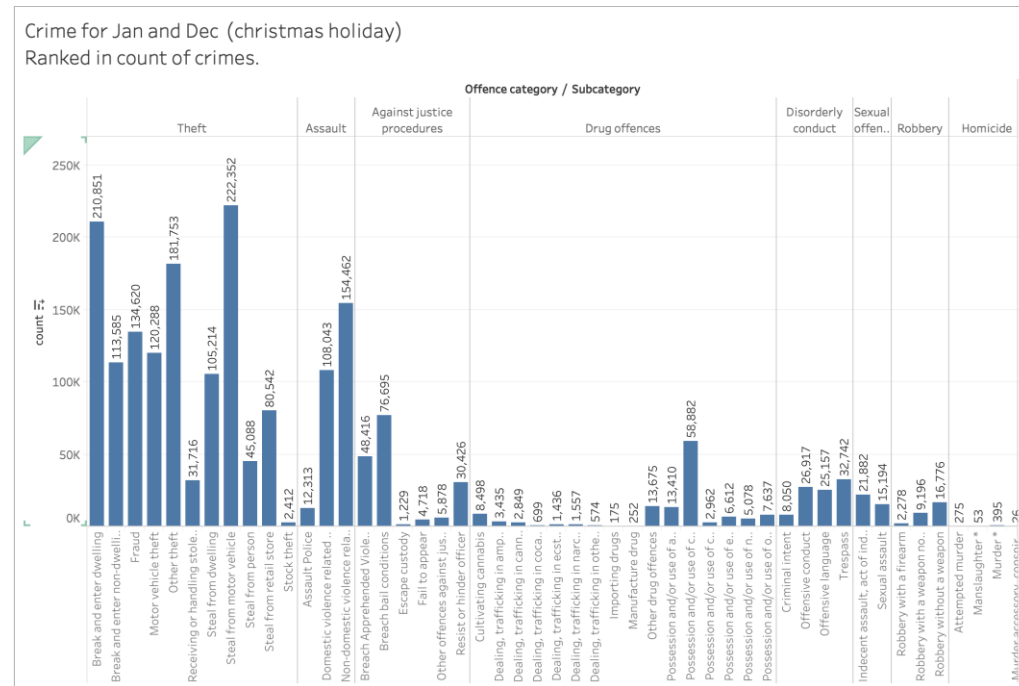
## Crime for Jan and Dec (christmas holiday)
## Ranked in count of crimes.

**Offence category**



| | | |
|---|---|---|
| Theft | 1,248,421 | |
| Assault | 274,818 | |
| Against justice p.. | 167,362 | |
| Drug offences | 127,731 | |
| Disorderly conduct | 92,866 | |
| Sexual offences | 37,076 | |
| Robbery | 28,250 | |
| Homicide | 749 | |

## Crime for Mar and Apr
## Ranked in count of crimes.

**Offence category**



| | | |
|---|---|---|
| Theft | 1,243,042 | |
| Assault | 246,297 | |
| Against justice p.. | 152,006 | |
| Drug offences | 127,925 | |
| Disorderly conduct | 81,858 | |
| Sexual offences | 38,688 | |
| Robbery | 31,076 | |
| Homicide | 716 | |
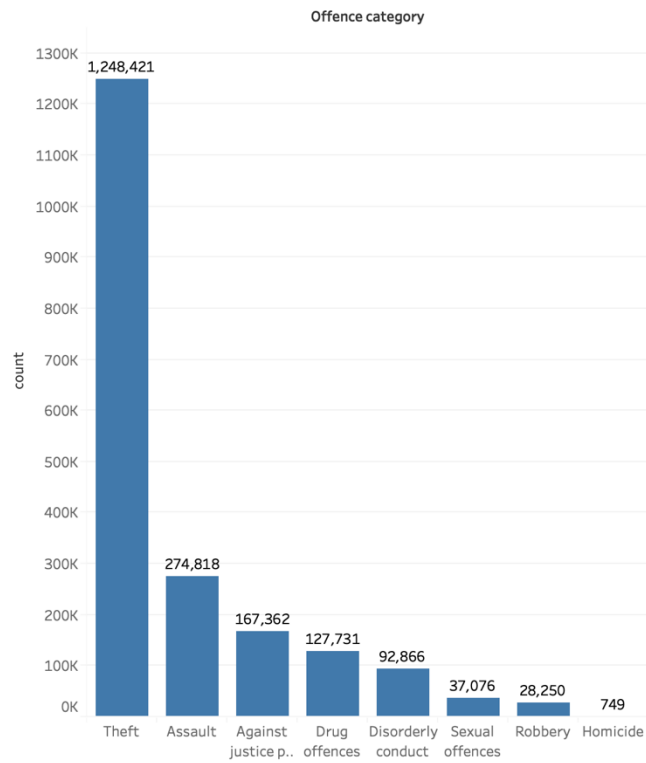
To get a detailed information between the trend change among this category during Christmas holiday and non-holiday, we use the drill-down operation to get such two figures below. Comparing between the two figures, a major of sub-categories has similar trend as their categories which has increased number when Christmas. It is clear that the holiday period has a higher crime frequency.



Crime for Jan and Dec (christmas holiday)
Ranked in count of crimes.



Crime for Mar and Apr
Ranked in count of crimes.

Then we divided all crimes over time via quarter to compared if there is a seasonal pattern in crime statistics. An obvious trend that the number of offences among all categories are increased sharply during quarter 1st and 4th in almost every year which is corresponding the Christmas holiday. Besides, the number of thefts is dramatic increasing in 2000 and 2001, and it gradually decreased in the next years. All categories except theft are nearly keeping smooth and steady. In addition, a new offence category which named transport regulator happened and slow grow in recent 10 years.

If we dice to a particular crime, which is theft. This graph shows different types of theft aggregated over years in quarters. As we can see, most of the trends is that the count increases in the 4th quarter.
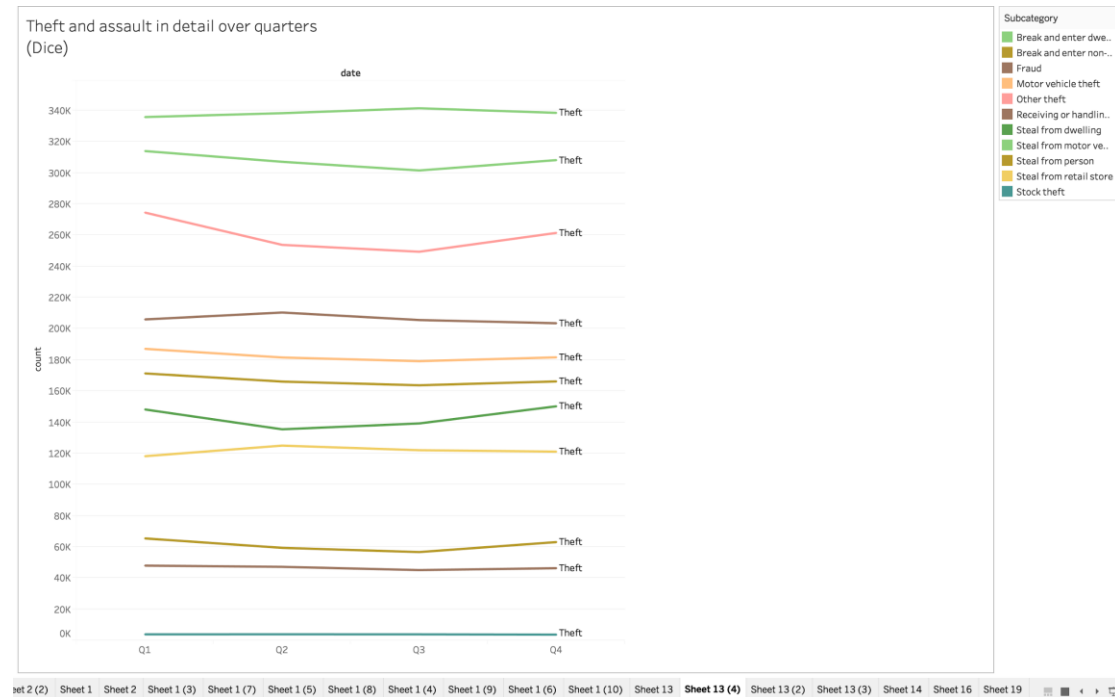


In this figure, almost all categories have the same trend which is quarter 1st and 4th are relatively higher that quarter 2nd and 3rd. We remove theft to show the other crimes' trend more obviously since the occurrence of theft is too large. A more apparent trend also can be seen in the second graph.
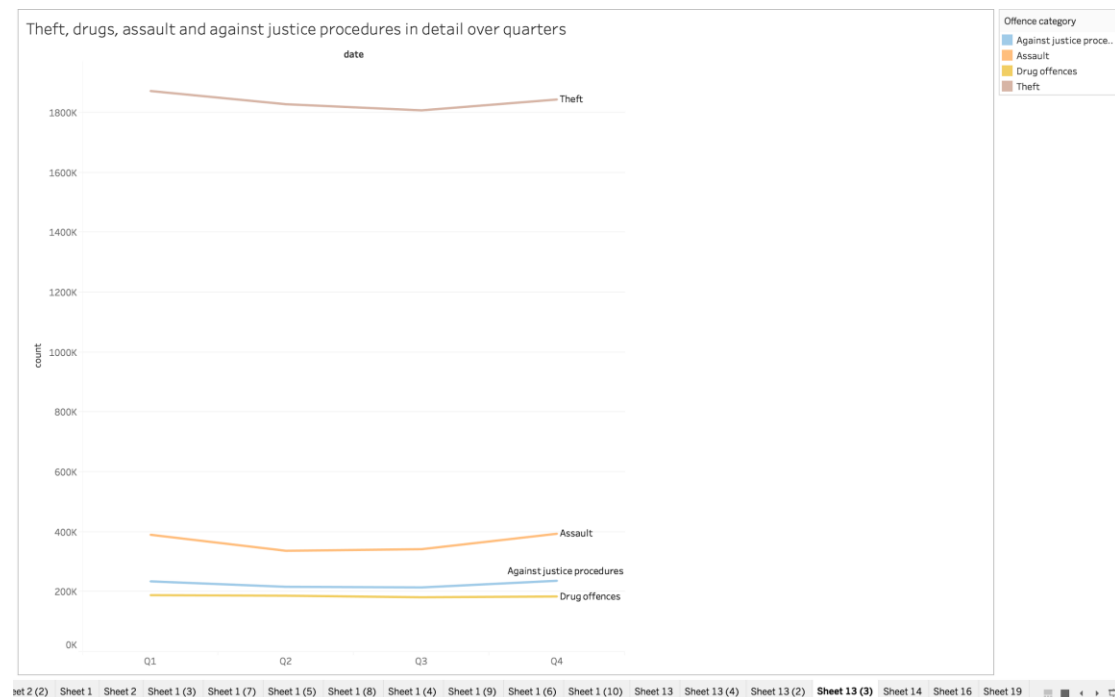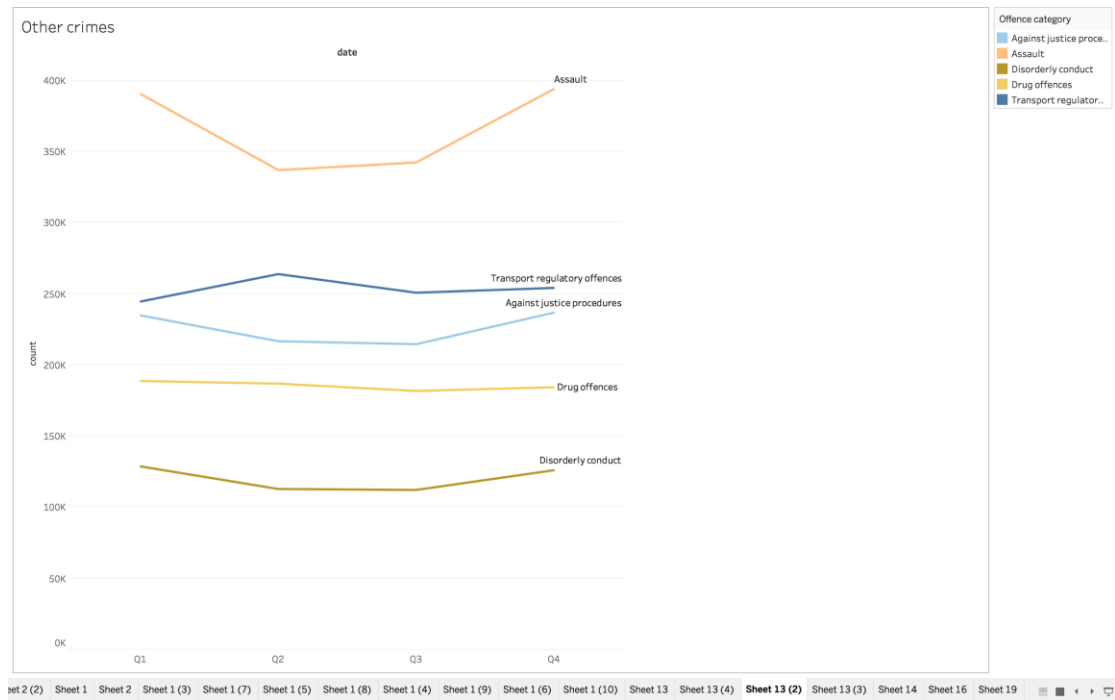
This figure is counting and clustering the crime count over months. Year-end months are divided into the same cluster which have more amount of offence than the months in the middle of the year. It also corresponds with the patterns of between the Christmas holiday and non-holiday.
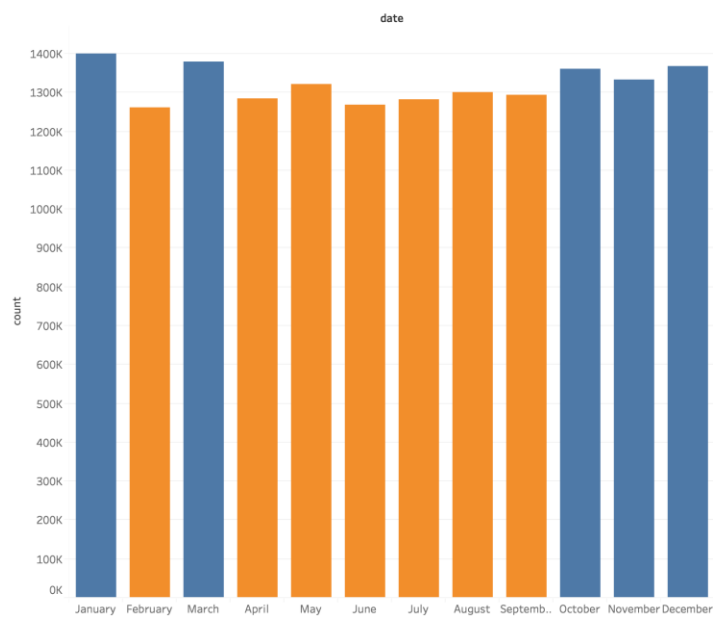
In the next graph, we compare all the categories in quarters. It can be easily seen that the category of theft has a beyond number than any other offence categories. Then the malicious damage to property is No.2 and the assault is No.3, and all the seasons have the similar crime proportion.
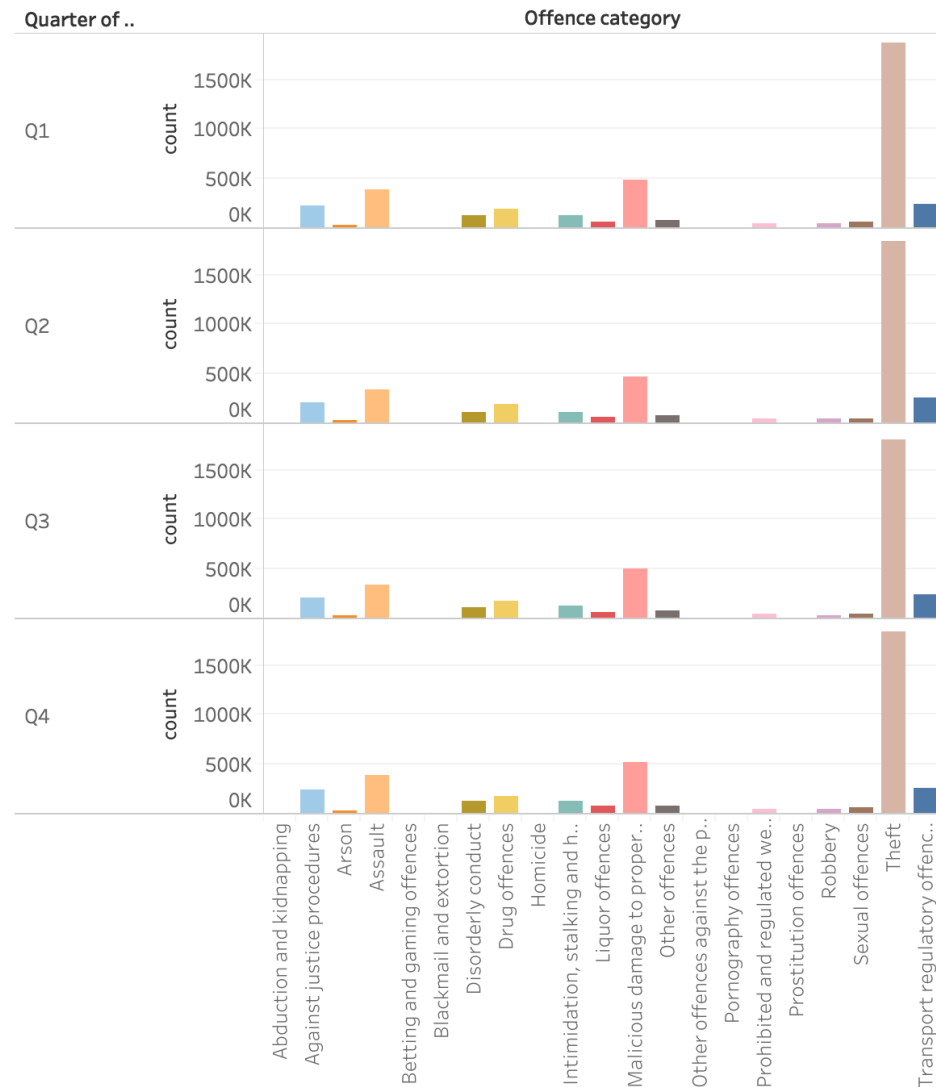


## In Summary

Overall, there is a seasonal pattern in crime statistics which can be expounded in two ways. For holiday and non-holiday, in our case, we focus on the Christmas holiday, which is the largest holiday during the year, the crime number increases during the holiday period, especially in the end of the year. For seasonal pattern, it is clear that quarter 1 and quarter 4 has a higher crime number than the other two quarters, which is also consistent with the Christmas holiday duration. It can also indicate that a colder temperature (winter season) may lead to a higher crime frequency.

# Part 4 Find out a tendency towards crime over time

In this part, we try to find out if there is a tendency towards crime over time in the state. There are a range of factors that may influence the changes in different categories of offence over time, such as population, the number of police force, government contribution, economy, and so on. Therefore, we analyze if these factors have the influence on the number of crime next.

## 4.1 Economy

Firstly, we analyze the relationship between offence number and estimated population, which is shown below. In this graph, the X axis represents the date whose unit is year, the left of Y axis represents the number of the offence whose unit is thousand, and the right Y axis represents the number of estimated people whose unit is million. It can be easily seen that the number of offences is changing constantly even though the population are continual increased. From the graph, an increase trend of offence number is continuing before the early of 21$^{st}$ century and after starting decreasing. There is no obvious tendency between offence number and estimated population, so we decided to analyze if it is related to the development of economy.
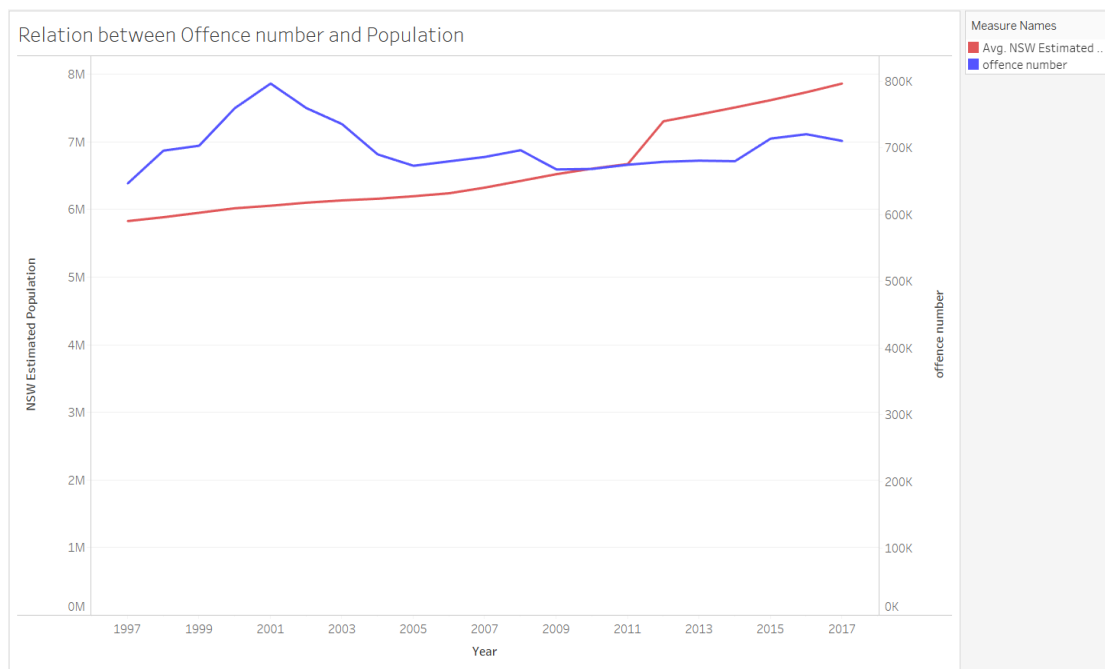


To find the relationship between economy and offence number, we analyze the tendency between crime rate and income gap. The crime rate is calculated by the offence number divided population, and in term of the income gap, we first aggregate income growth dataset per capita in New South Wales (Australian Government, 2018), then we calculate the gap by using the average income weekly of total population minus the average income weekly of bottom 40 percentage.

In this graph, the X axis represents the date, the left of Y axis represents the crime rate, and the right Y axis represents the income gap. With the economy developed fast, the income gap gradually increased in this recent 20 years and it can be seen that after 1999. Through the line chart, it is easily to find that with the increasing of income inequality the crime rate kept declining. However, it opposes to the common sense as far as people concerned, so we choose GDP data visualization of Australia from WorldBank (2019) as the measure of people living standard to see any relationship with crime rate.

Relation between Crime Rate and Income Gap
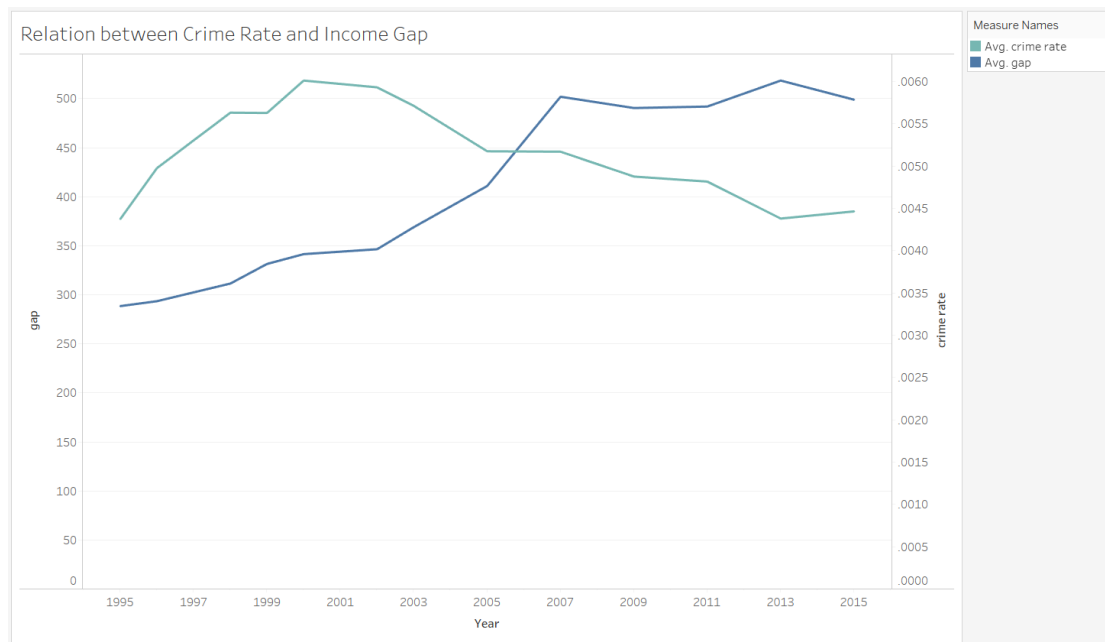
In this graph, the X axis represents the date, the Y axis represents GDP growth (annual %). Combine this graph to the previous one, before 21st century, the crime rate rose under the poor society though the GDP went up. Focus on the point shows a sharp decrease GDP growth happened in 2001, which can infer that the increasing crime rate is caused by. In addition, when the economy developed enough in the next few years, the crime rate is declined smoothly.



1995 - 2017
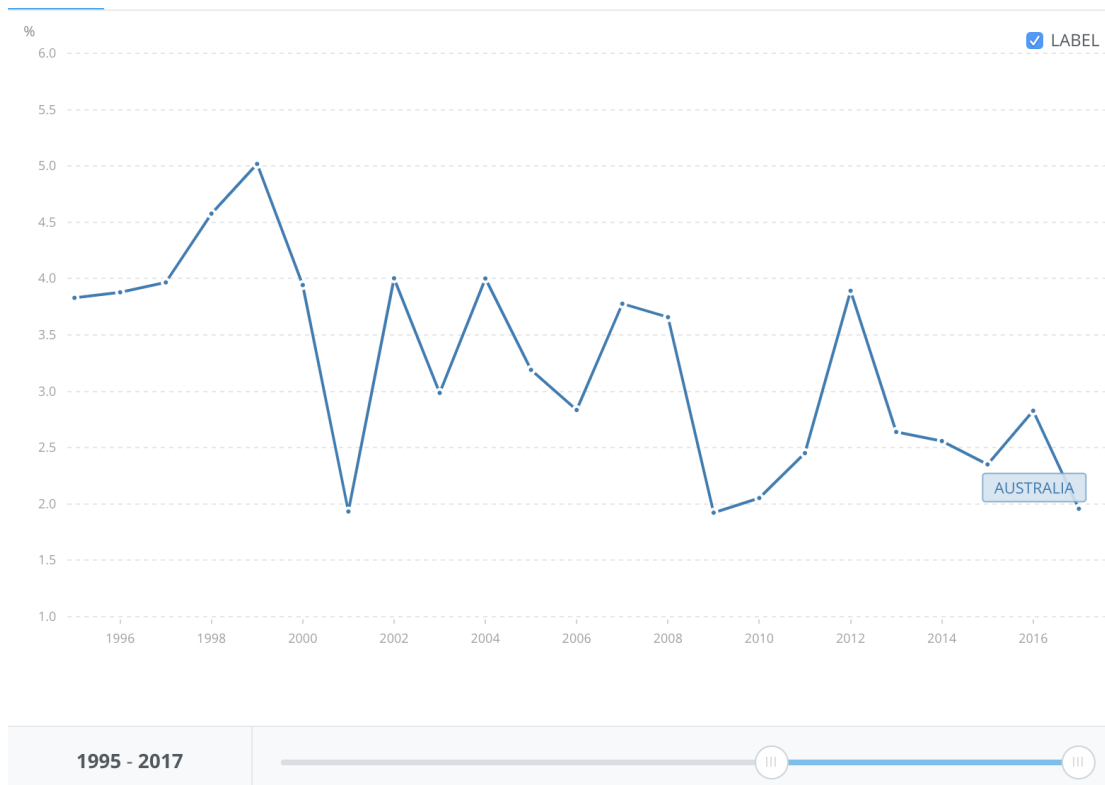
## 4.2 Police force

We also guess that the police force can influence the number of crimes. According to Opengov (2019), there are enough data about police force to do deep research. Thus, in this graph, the X axis represents the date, the left of Y axis represents the crime rate, and the right Y axis represents the police number. In 2001, the crime rate has a sharp fall which the police number increase gradually correspondingly. In addition, about 2004-2007, both the police number and crime rate changed smoothly which can see from the graph below, once the police number has continuing increase, the crime rate can under an effective control which have reduced in recent years. Therefore, the police force is a key factor of influencing the trend of crime over time.
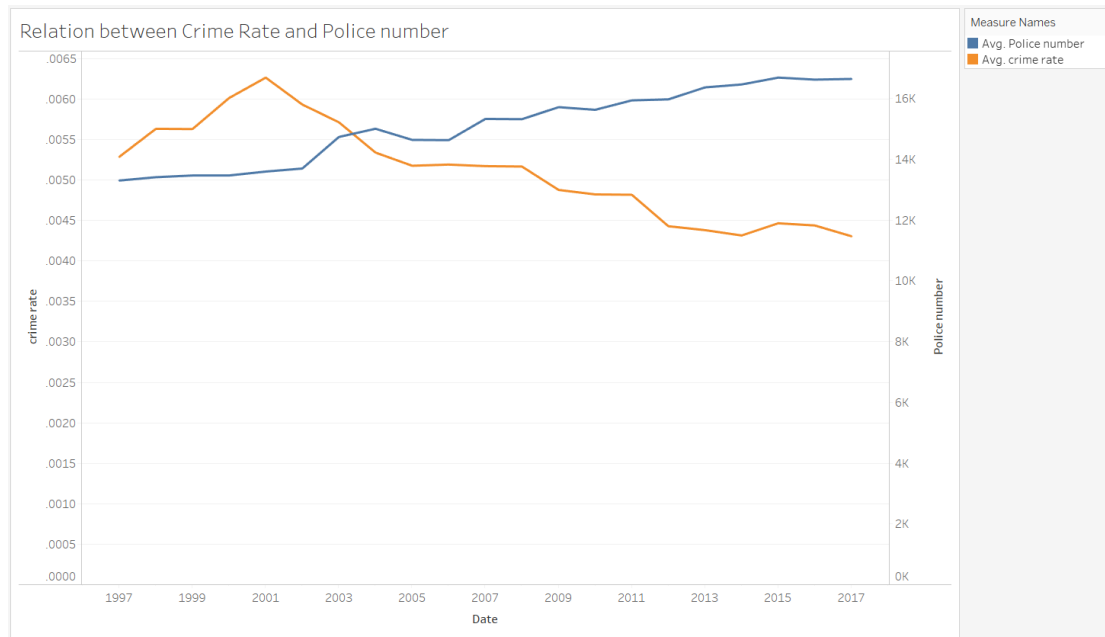


## 4.3 Government contribution

As we all known, government contribution has a strong effect on controlling the offence number. Therefore, we compare the trend between crime rate and government contribution which means the investment in managing offence. In this graph, the X axis represents the date, the left of Y axis represents the crime rate, and the right Y axis represents the government contribution which unit is dollar. Obviously, the government contribution is always ever-increasing, which can have a positive effect on controlling crime rate. In detail, looking at the change between 2011 and 2013, when the increase of government contribution is dramatic, the crime rate falls sharply at the same time. Thus it can be seen that the government contribution is also a key factor.

Relation between Crime Rate and Government Contribution ($)
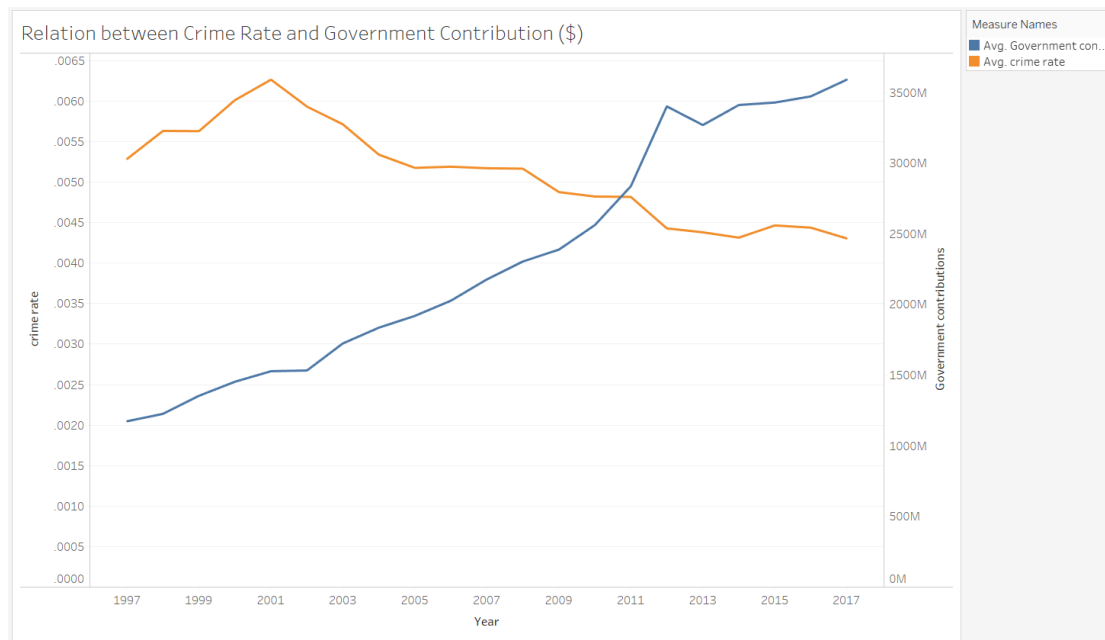
However, to get more relative and useful conclusion, we decided to analyze some specific categories which may show more obvious relationship between offence number and relative factors. According to the part 2, we know the top one category offence with most value is theft offence. Another especial category is transport because it keeps increasing all the time. The change number of these two categories can be seen in the graph below. In that way, we will analyze this two categories in detailed next.



Theft number

Transport regulatory offences number
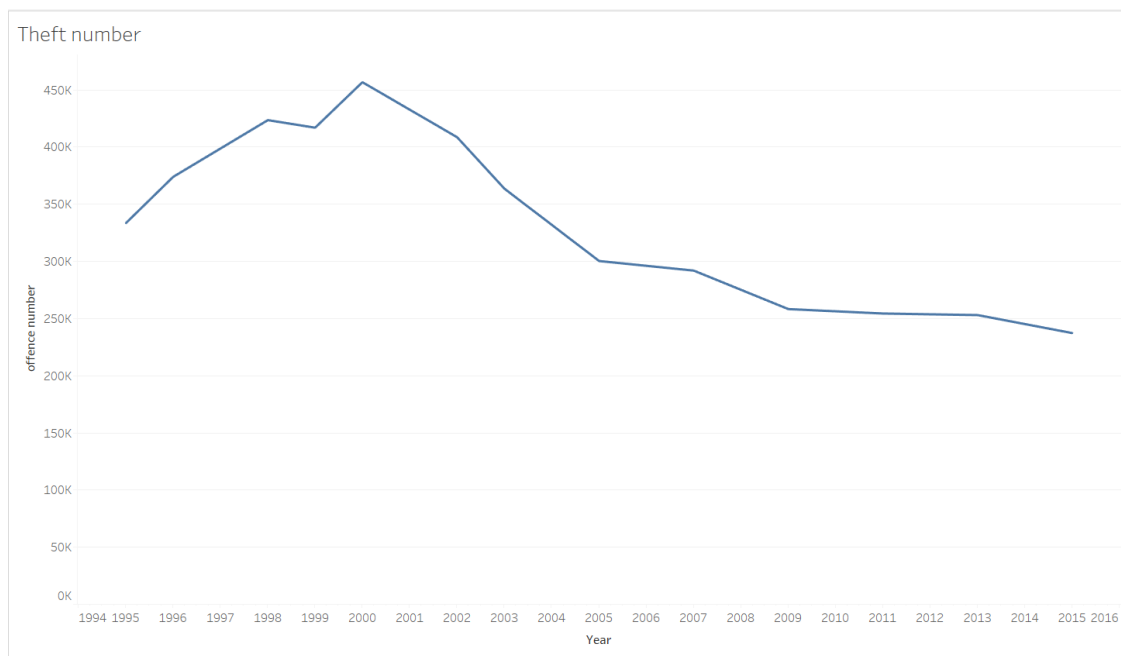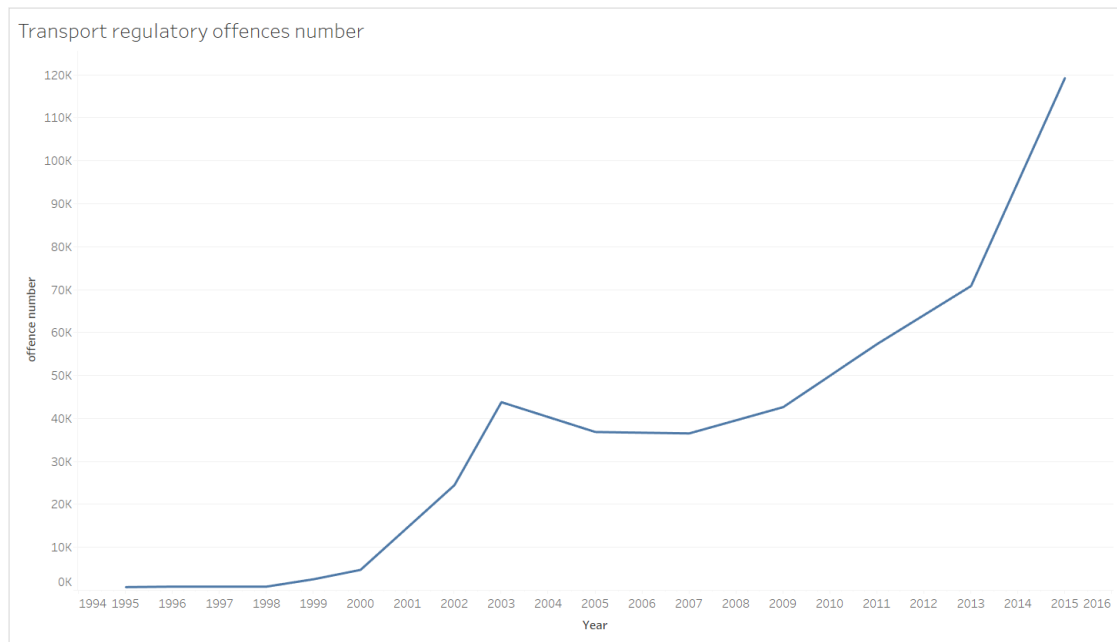
## 4.4 Factor for theft category

The offence category which named theft can represent the tendency of the whole offence. In this graph, the X axis represents the date whose unit is year, the left of Y axis represents income gap, and the right Y axis represents the number of the offence whose unit is thousand. From this graph, the number of theft offence got a peak at around 2000 and then decreased in the following year though the income gap increased all the time. To explain the reason why the declined happened, we found that according to Amanda Hoh (2017), the author explained three reason of theft falling. First, there is a better security in our society from 2000 and the NSW government invest $250,000 to prevent crime and fear of crime, such as installing closed circuit television, improving street lighting and so on. Second, the rise of large household appliances which makes reselling the appliances is cheap. In this way, the theft cannot get too much in stealing and reselling. Third, the reduction use of drug. The heroin use reduced from the late 90s has correlated with a decline in stealing.



Relation between Theft crime rate and income gap
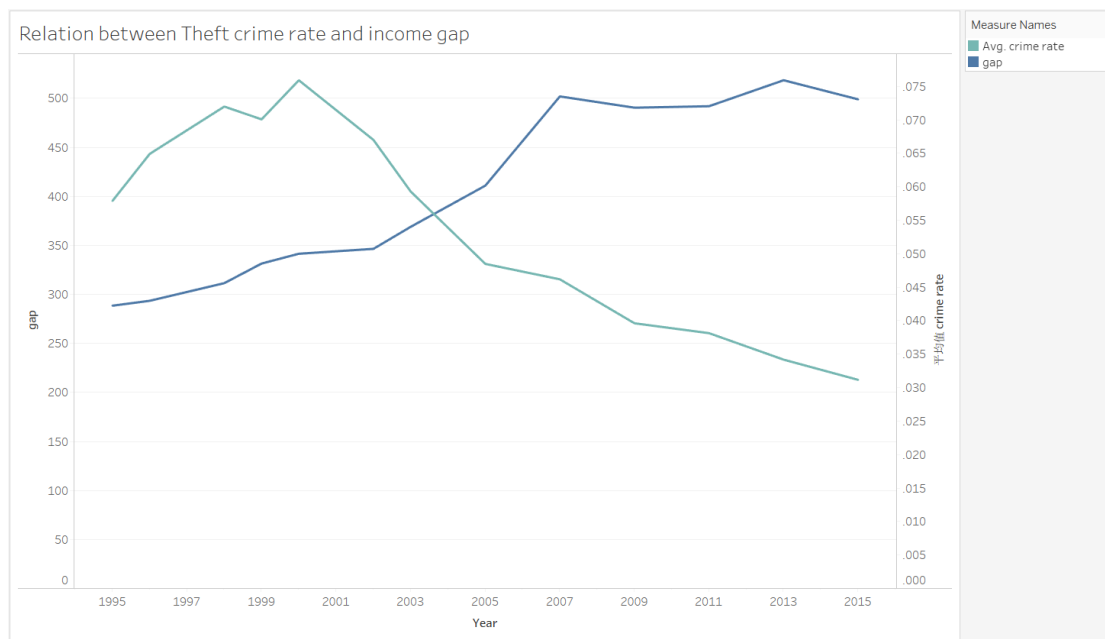
## 4.5 Factor for transport regulatory offence

Due to the number of transport offence change obviously, so we analyzed this category in detailed. We do OLAP operation as following. First, we do slice because we want to focus on only one offence-category--transport regulatory offence instead of whole offence number. Then, we pivot the Trip-Time.csv to make a dimension of date, and finally, we we roll-up the Trip-Time.csv. SUM() the trip times of each mode(bus, train, etc.). After combining our data set to the Trip-Times according to the TfNSW Open Data Hub and Developer Portal which include monthly trip times of each mode (bus, train, ect.) from 2016-2018.

The graph below shows the relation between transport offence and transport trip time, and the transport trip time dataset comes from Open Data (2019). There are only 16 months that can connected to our data, except two points which is the December in 2016 and 2017. Both of them are the end of year, and the number of transport offence slightly increased though the trip times decreased sharply. This phenomenon corresponding to the relationship we found in part 3, which the number of crimes is higher during the holiday. Then we look at a whole, the wave of both offence and trip times is nearly simultaneous change except the two months mentioned above which they go up and down together. According to this graph, we can infer that with the development of transport, public have more opportunities to trip, and the times of trip leads to the number of transport offence increased.
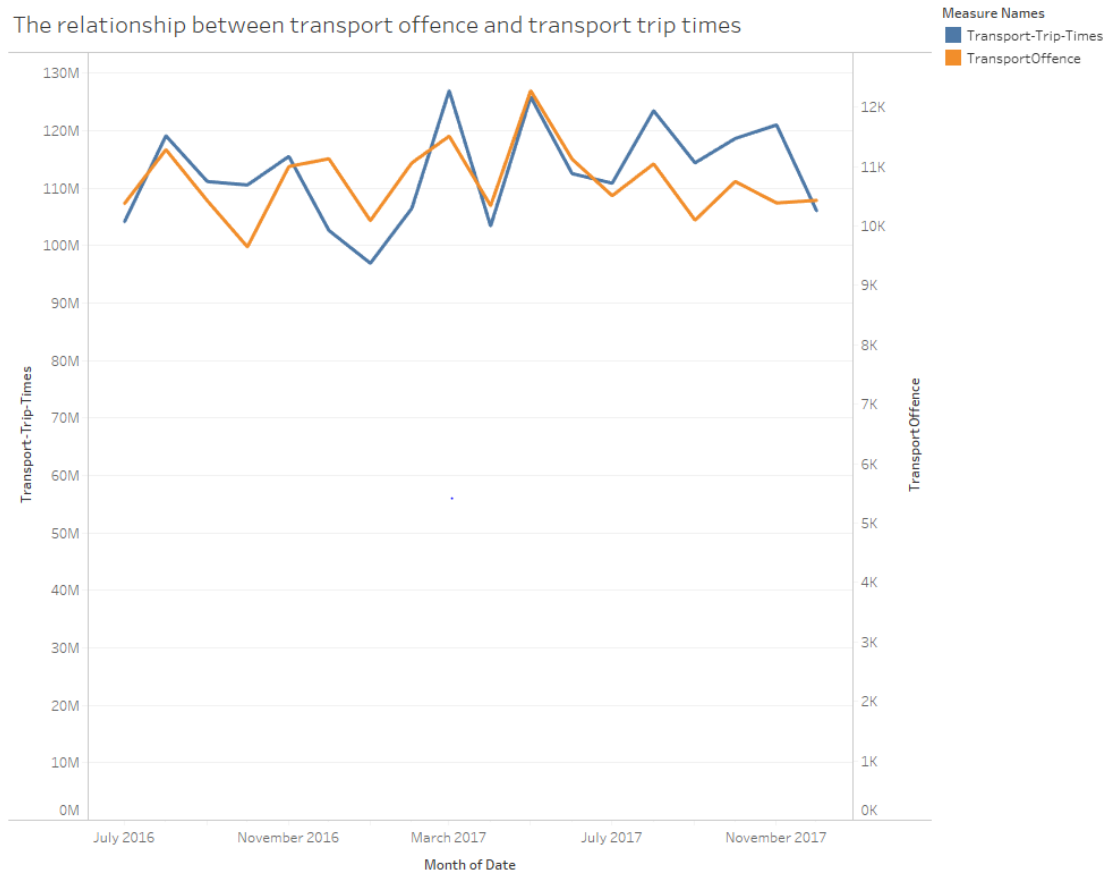
## In Summary

Criminal tendency is related to a variety of reasons, the population growth or income gap can influence the amount of criminal, but it not the conclusive factors. However, the factors about the living standard of people, the police force, and the government contribution have closely related to the change of offence. All of them are positively related with the tendency of offence number. In addition, the different offence category needs different specific analysis due to the factors influence the number of offences is different. For example, the theft category is influenced by government policy. The transport regulatory offence is related to the trip time of public strongly. But in general, the total number of offences has a decrease tendency with the development of society and economy. The government can focus more on control the offence and publish detailed related law to help reduce the crime number.

## References

Australian Government (2018). Growth rates income per capita of bottom 40 percent and total population.
https://data.gov.au/dataset/ds-dga-4e0fba19-cc15-4df0-81cd-6482de471ae5/details

Hoh, Amanda (2017). Sydney crime falling as prevention strategies yield results, says criminologist. *Investing in crime prevention.*
https://www.abc.net.au/news/2017-02-03/sydney-crime-falling-as-prevention-strategies-yield-results/8239014

Open Data (2019). Opal Trips – All Modes. *Opal trips by public transport by month.*
https://opendata.transport.nsw.gov.au/dataset/opal-trips-all-modes

Opengov.nsw.gov.au (2019). NSW Police Force Annual Report 96-17. *[online] Available at:*
https://www.opengov.nsw.gov.au/publications

The World Bank(2019). GDP growth (annual %). *WorldBank national accounts data, and OECD National Accounts data files.*
https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?end=2017&locations=AU&start=1995&view=chart