

School of Information Technologies

Faculty of Engineering & IT

ASSIGNMENT/PROJECT COVERSHEET - GROUP ASSESSMENT

Unit of Study:	ISYS5050 Knowledge Management System	าร
Assignment name:	Final Group Project	
	om = 9·00 nm Tutor name:	
Tutorial time: 8:00 p	<u>om – 9:00 pmTutor name:</u>	

DECLARATION

We the undersigned declare that we have read and understood the <u>University of Sydney Academic Dishonesty and Plagiarism in Coursework Policy</u>, 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 Sydrey 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 to	am members		
Stu	ident name	Student ID	Participated	Agree to share	Signature
1.	Chen Cui	470109940	Yes / No	Yes/No	Chen Già
2.	Chunliang Pan	470160239	Yes / No	Yes No	Churliang Pan
3.	Xiao Yang	480285562	Yes No	Yes No	AL.
4.	Gleb Tsoy	480451167	Yes No	Yes / No	Gleb Tsoy
5.	Minghao Wu	470210095	Yes / No	Yes No	Minghao Wu
6.			Yes / No	Yes / No	
7.			Yes / No	Yes / No	
8.			Yes / No	Yes / No	
9.			Yes / No	Yes / No	
10.			Yes / No	Yes / No	

Table of Contents

Introduction	2
Dataset Description	2
Related to the crime	2
Task Summary	2
Question assessment	3
Question 1	3
Problem Statement	3
Assessment and Visualization	3
Question 2	4
Problem statement	4
Assessment and Hypothesis	4
Question 3	5
Problem statement	5
Seasonal pattern	6
Holiday season	6
Question 4	9
Problem statement	9
Assessment and Hypothesis	9
Further findings and discussions	14
Hypothesis 1	16
Hypothesis 2	18
Hypothesis 3	21
Reference list	23

Report

Introduction

The goal of this assignment is to analyze the NSW crime dataset, derive interesting and meaningful visualizations in order to find rates, trends, and patterns.

Dataset Description

Local Government Area (LGA) - This is referred to suburbs/regions in NSW. For example, Albury, Ballina, Bega Valley, etc.

Offence Category - Different offence types, including homicide, assault, robbery, etc., in a total of 21 different categories

Offence Subcategory - A more detailed description of offence category. For example, robbery without a weapon, and robbery with a firearm under robbery category

The dataset contains a time span of more than 20 years, from January 1995 to December 2017, including monthly data of NSW crime recorded by the police.

Related to the crime

The Oxford dictionary had defined the crime as an action or omission which constitutes an offence and is punishable by law. However, the elements leading to crime is complex. Such as poverty, family environment or even certain psychological problems may possibly lead people to break the law. According to a recent report published by the British Broadcasting Corporation (BBC, n.d.), three categories of people are more likely to commit crimes worldwide. Therefore, When the report analyzes tendency, relevant demographics can be focused on to facilitate the prediction of results.

Task Summary

We first transformed the dataset using the OLAP operations to find the top and bottom five suburbs/regions in terms of the total number of different types of offences committed. We used different dimensions to find seasonal patterns if there is any; we tried to find any law related to economy, population, and other potential factors and background, to explain the reason behind the patterns or phenomenons.

Second, we tried to discover offence patterns over time, including the holiday season, weather season, etc. To ensure the integrity and credibility of the results, background search and relative statistics will be used as a reference.

Finally, we will find out if there is a tendency to more or less crime over time in the future. Different techniques and more data will be used to provide a credible, justified and data-driven result.

Question assessment

Question 1

Problem Statement

This problem is focusing on data pre-processing before we analyze. Since the dataset contains monthly data, in another word, the number of different categories of offences in each month is presented, a transformation is necessary thus a dimension for the date is created for us to apply OLAP operations.

Assessment and Visualization

We first extract all the monthly data, and simply do a pivoting. In this way, a new dimension of "date" will be created and it will be easier for future analysis. Figure 1-3 shows the whole ETL process.

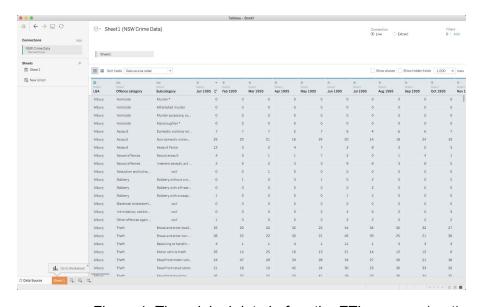


Figure 1: The original data before the ETL process (part)

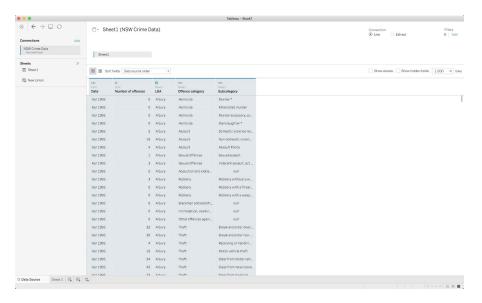


Figure 2: Transform all monthly data using Pivot (part)

Date			Date			Date	
Apr 1995	40,754	^	Jan 2007	60,526	^	Oct 2005	58,13.
Apr 1996	46.126		Jan 2008	60,338		Oct 2006	58,154
pr 1997	51,160		Jan 2009	60,939		Oct 2007	57,813
pr 1998	55,717		Jan 2010	60,986		Oct 2008	60,13
pr 1999	55.592		Jan 2011	61,949		Oct 2009	56,98
			Jan 2012	59,832		Oct 2010	56,16
pr 2000	59,460		Jan 2013	61,163		Oct 2011	57,35
pr 2001	62,896		Jan 2014	62,032		Oct 2012	58,00
pr 2002	63,103		Jan 2015	64,661		Oct 2013	57,43
pr 2003	61,048		Jan 2016	62,546		Oct 2014	59,49
pr 2004	56,347		Jan 2017	63,520		Oct 2015	59,93
pr 2005	55,323		Jul 1995	42,970		Oct 2016	60,09
pr 2006	54,823		Jul 1996	49,808		Oct 2017	61,21
pr 2007	56.862		Jul 1997	51,333		Sep 1995	45,40
pr 2008	56,458		Jul 1998	57,607		Sep 1996	50,35
pr 2009	54,222		Jul 1999	57,294		Sep 1997	54,52
pr 2010	54,219		Jul 2000	64,400		Sep 1998	58,87
			Jul 2001	66,909		Sep 1999	59,55
pr 2011	56,729		Jul 2002	62,481		Sep 2000	60,76
pr 2012	55,623		Jul 2003	60,468		Sep 2001	65,30
pr 2013	55,500		Jul 2004	54,771		Sep 2002	63,00
pr 2014	54,739		Jul 2005	55,272		Sep 2003	59,74
pr 2015	58,832		Jul 2006	56,808		Sep 2004	54,34
pr 2016	60,159		Jul 2007	54,940		Sep 2005	54,85
pr 2017	57,675		Jul 2008	55,746		Sep 2006	55,46
ug 1995	45.077		Jul 2009	52,390		Sep 2007	55,07
ug 1996	51,367		Jul 2010	52,926		Sep 2008	57,26
200000000000000000000000000000000000000			Jul 2011	52,861		Sep 2009	53,93
lug 1997	53,317		Jul 2012	54.304		Sep 2010	54,170

Figure 3: Load the new data (part)

Question 2

Problem statement

For this problem, we're aiming at finding the highest 5 and lowest 5 LGAs having the top 5 offence categories.

Assessment and Hypothesis

In order to find the top and bottom five LGAs by a number of crimes committed throughout the entire period we have used roll-up operation on the number of crimes measure and then sliced to leave only top 5, and another time for bottom 5. Then we sliced the data further and left only the top 5 offence categories by the sum of the number of offences and sorted the data in ascending order of crime numbers by LGA and offences committed by category. The result is represented in two bar charts (see Figure 4).

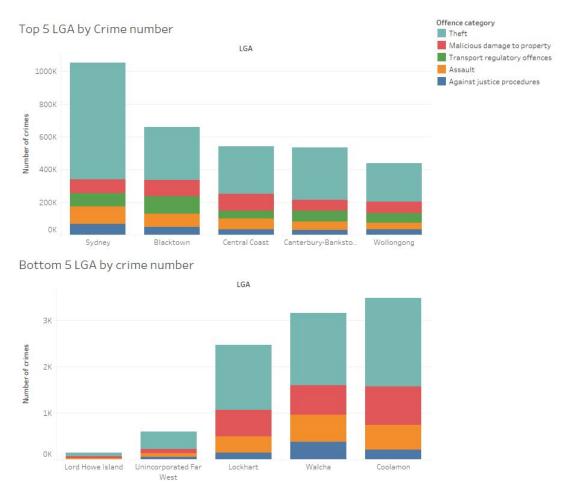


Figure 4: Bottom 5 and Top 5 LGAs with top 5 crime categories

It is clear from the bar charts which LGA have the highest and the lowest offences committed. Curiously theft is the category of crime with the biggest numbers in all of the regions. Bottom 5 has very small or almost no crimes as transport regulatory offences, while it is ranked the third in crime in the top 5 LGAs. As we will find out in further research, this might be related to the population, unemployment and the degree of economic development situation in different LGAs, for example, the bottom 5 LGAs are all rural area with low economic development and population, therefore the transport regulatory crime disappeared in these bottom areas because they do not even drive. (further analyzation see Question 4).

Question 3

Problem statement

In this task, using OLAP operations to perform analysis, we have an insight into seasonal patterns over time, if there is any. We also tried to analyze from additional seasonal type, including the holiday season.

Assessment and Hypothesis

Seasonal pattern

Crime amount of 4 seasons in every year

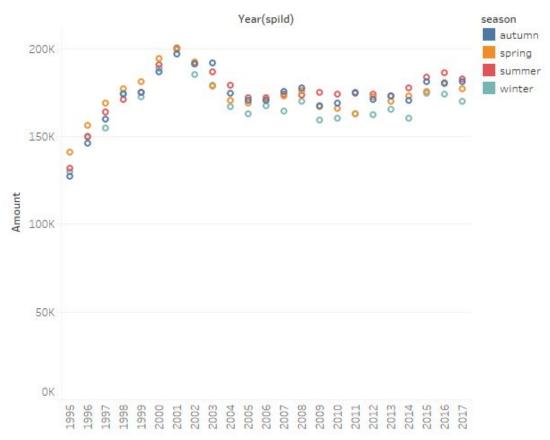


Figure 5: Total number of the offence of 4 seasons from 1995 to 2017

Figure 5 shows the distribution situation of the seasons in multiple years. We assume that the amount of crime is seasonally distributed, so we selected the circle views graph in order to distinguish four seasons clearly. In this graph, the two dimensions are seasons and years and the measure is the crime amount. It can be clearly seen from figure 5 that spring is the season that has the most crime numbers in every year before 2001. Then, the data in 2001 shows that the crime situation of seasons in this year is almost even. However, after 2002 (2003 to 2017), in the following 15 years, summer became the season that has the most crimes (10/15), the rest 5 years are occupied by autumn. Another information should be mentioned here is in winter, which has the lowest crime numbers in 16 of 23 years, in a percentage of 69.57%. Although there are some interesting findings listed above, the possible reason for these findings cannot be easily deduced from this graph. The following content in this part is the assumption we drew which is based on some further analyzation.

Holiday season

We add all offence values grouped by year and month from 1995 to 2017 to discover the offence pattern (see Figure 6). Using the roll-up operation in terms of month, where the month with the highest number of crimes is in January, around 1400k, followed by March

and December. It is noted that the common feature of these months is crime occurred over the holiday period.

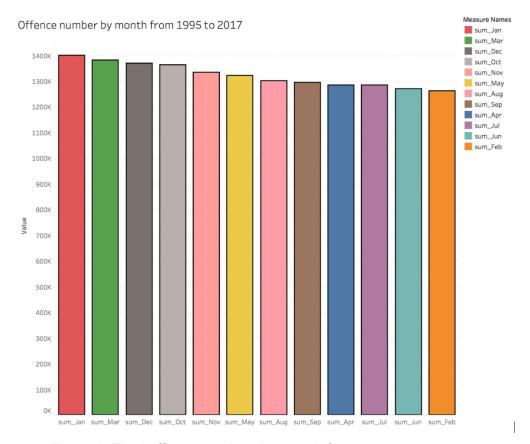


Figure 6: Total offence numbers by month from 1995 to 2017

New Year's Day and Australia Day are both in January, which is the national holidays of Australia with the several days break. In addition, Christmas Day at December 25th is also the grand festival in every region. March is also occupied the major number of offence before the Easter holiday in April.

To come up with the accurate reason, our group extracts the offence category in these top 3 months (see Figure 7).

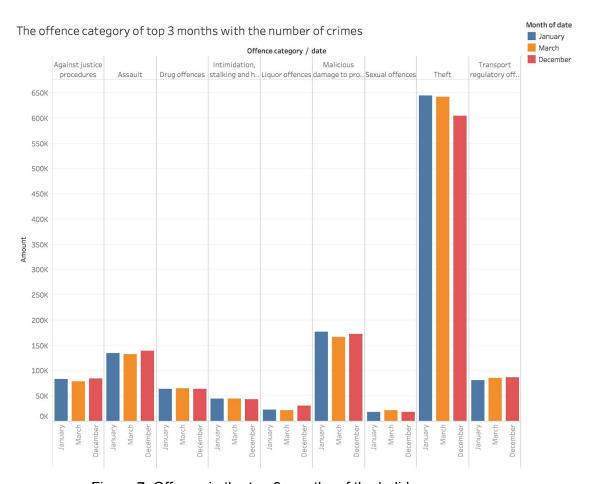


Figure 7: Offence in the top 3 months of the holiday season

As the figure displays that theft is keeping the highest number compared with other categories. With all the Christmas shopping and hustle bustle of the holiday season, it is no wonder that the theft has more opportunity to steal money and valuables. Especially, the number of assaults, liquor offences as well as malicious damage to property occurs in January and December exceeds these offences happened in March. According to our survey and analysis, one of the main reasons is "stress". For the majority of people, the Christmas holidays and New Year's Day means to spend time with family and immersed in overall good cheer. However, for others, it means stress and worse time full of fear and loneliness. Therefore, they would like to do some negative behaviours to express their grievances like malicious damage to property. Another reason is people spending more time together than they normally do, which leads to more conflicts with family. As Nina Funnell in The Age highlights that from "December 2009 to January 2010 police responded to more than 5000 domestic-violence related complaints" in NSW alone. This is also the reason why the number of assaults occurs in December and January is higher than that in March.

It is noted that the more crime occurred over the Christmas holidays and New Year's break, especially the assault, malicious damage to property and theft. The top 3 months of the number of offences have the common feature that is occurred over the holiday period or the early stage of the holiday. Therefore, there exist the relationship between the holiday season and offences occurred number.

Question 4

Problem statement

In this problem, we are required to find out if there is a tendency towards "more", "less" or "no real change" crime over time in NSW. Using the data, with the range of OLAP and other analysis, based on adjustments and a range of factors that could affect the changes in different categories of offence over time, we will find the answer with analysis, tables, and visualizations.

Assessment and Hypothesis

Consider the following graph that shows the number of crimes over the whole period (see Figure 8).

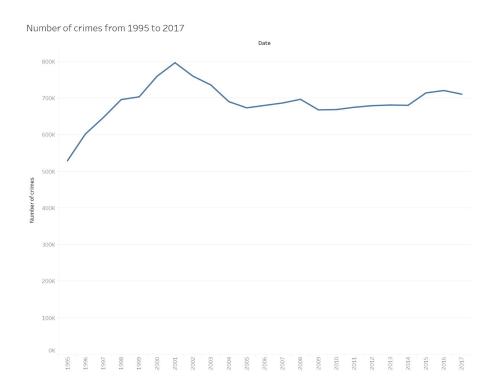


Figure 8: Total number of offence from 1995 to 2017

It is immediately apparent which years had an increase and decrease in the number of crimes.

The line chart (Figure 9) shows the trend of crime rate from 1995 to 2017. Since the population data is stored in another data file *NSW Estimated Population*, we have to connect two data sets together. The common dimension of the two data set is the year, so we built a new connection between the year after spilling the dimension in two columns as years and months. After connection, there are some duplicated data in the population column which will influence our calculation. In order to remove duplicate data, we create a new calculated field with *{Fixed [Year]: MIN([NSW Estimated Population])}*. The crime rate is calculated by *(sum (crime amount(year))/ removed population(year)*100%*. We can see that the peak of the

chart is 2001 where the rate is 0.13155, after that the crime rate keeps decreasing and reaches the lower rate at 2017 (0.09321). Our team predicts the crime rate in future 3 years, which is a stable trend compared to the previous years.

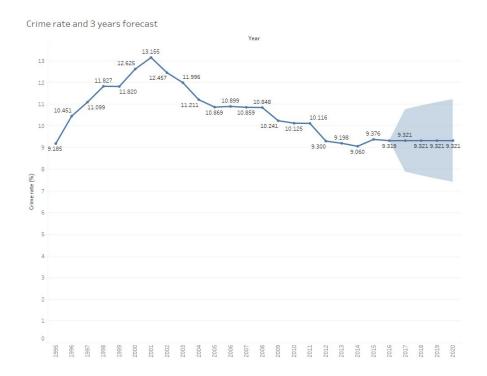


Figure 9: Crime Rate from 1995 to 2017 with 3 Years Forecast

We have gathered some additional information about the population demographics for NSW for some of these years. In particular, we found demographics of age groups beginning from 2001 from Healthstats NSW website (NSW Government, n.d.). The areas of interest here were the high crime number and low crime number periods. So we decided to compare the years 2001 and 2016 as years with a high number of offences as opposed to the years 2009 and 2005 with lowest number offences in the given period. Unfortunately, the data does not include the years before 2001 (see Figure 10).

As seen from the tables years 2001 and 2016 are comprised overall of younger age groups, as opposed to 2005 and 2016 with older age groups. From this comparison, we can make an observation that the younger population leads to more offences committed relatively and should be taken into consideration when making inferences in future research about crime in NSW.

2009		2001	-
Age (years)	Age (ye	ars)
35-39	517,370	35-39	504,45
25-29	509,797	40-44	496,49
45-49	504,974	30-34	490,16
20-24	495,991	25-29	476,36
30-34	483,556	5-9	451,07
2005		2016	
Age (years		Age (ye	ars)
Age (years 40-44	506,261	Age (ye 25-29	ars) 575,96
Age (years 40-44 30-34	506,261 504,412	Age (ye 25-29 30-34	ars) 575,96 570,89
2005 Age (years 40-44 30-34 35-39 45-49	506,261	Age (ye 25-29	ars) 575,96

Figure 10: Age groups

To further optimize the scope of analytic data, we defined the different criminal category based on the classification of crime types from the US Department of Justice (National Incident-Based Reporting System (NIBRS), 2012). In the report, crime has been defined in three categories, namely the crime against property, the crime against society and the crime against the person. For example, theft, arson, blackmail, and extortion, etc. will be considered as crimes against property; betting and gambling offences, drug offences, etc. will be categorized as crimes against the person.

We made a more advanced analysis to find out the Compound Annual Growth Rate (CAGR) of the total number of crimes (see Figure 11 - Figure 13). According to the figures, the overall results have the highest increasing rate of 13.78% comparing to 1995 in 1996, the rate keeps decreasing, with a small margin of wave in 2000, even though it is always positive comparing to 1995. Our hypothesis is that in 2000, the Olympic is the factor behind the increasing wave. Due to the population growth, immigration, and refugees, the absolute number of crimes is increasing with no doubt, nonetheless, there is a tendency that the crime is being stabilized, even though it's keep growing, the increasing speed year on year is slowing down. We draw on the three-category classification to bring up the next discussion by combining the compound annual growth rates.

For property crimes, these offences categories have a similar trend with the Compound Growth Rate of crime from 1995 to 2017 (Figure 11). These 5 offences categories within property crimes are occupied in the majority proportion compared with the social crimes and crimes against the person, which interprets the number of crimes against property is the principal factor to influence the Compound Growth Rate of crime.

For crimes against society, many subcategories are outliers. For example, 3 subcategories in this classification have declined over 10% in 1997, which is against the increasing tendency overall. We also observe for gambling and prostitution, a rebound occurred after the big decline and reached a peak around the year 2000 to 2004. We assume it might also be caused by the Olympic, due to foreign visitors and diversified demand.

From the crime against person graph, the crime rate changes of seven categories in this aspect are shown clearly. The crime rates of 5 categories were significantly decreased from 1996 to 1998 (exclude abduction & kidnapping and other offences). Since 1999, the crime rates of most categories keep decreasing and going to stabilize with a lower rate value, but there are two special categories should be mentioned separately. The crime rate of homicide kept decreasing since 2001 and reached a negative value in 2004 and in all rest years. This may be caused by the police have increased the emphasis and punishment on homicide after 2000. Another one is the other offences which have a negative increasing value from 1996 to 2000 but keep an increasing trend in most of the years.

Further findings regarding the crime types are analysed, with interesting facts for certain types of crime come out. Property crimes take account in a big percentage towards all crimes, therefore, other categories of crime fluctuate less dramatically in the overall trend.

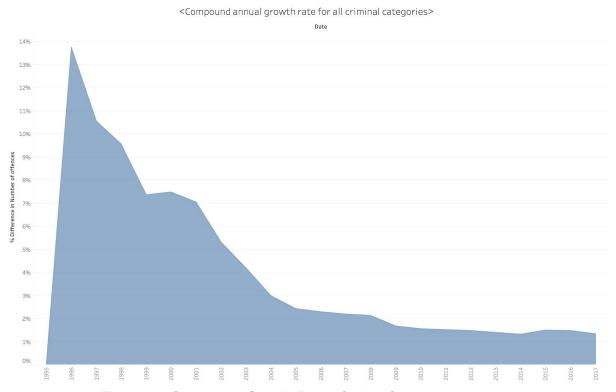


Figure 11: Compound Growth Rate of crime from 1995 to 2017

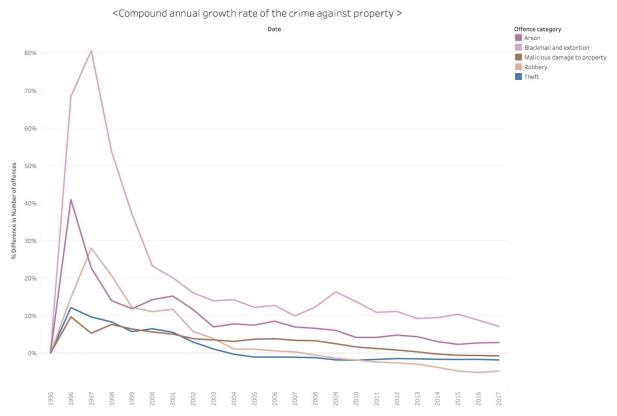


Figure 12: Compound Growth Rate of Property Crime from 1995 to 2017

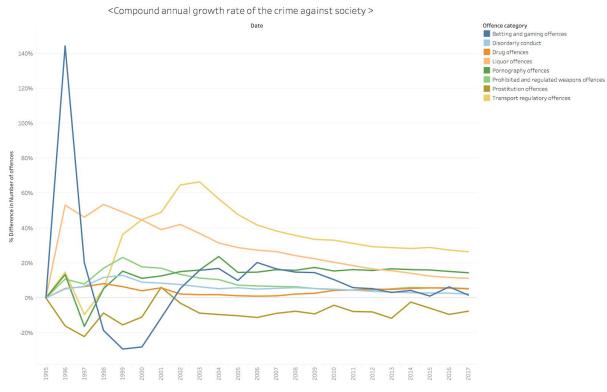


Figure 13: Compound Growth Rate of Society Crime from 1995 to 2017

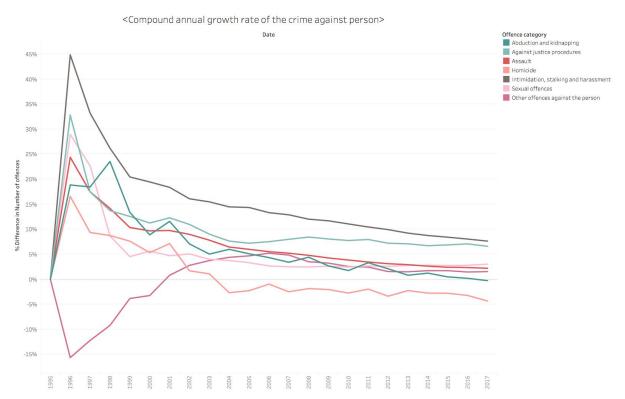


Figure 14: Compound Growth Rate of Person Crime from 1995 to 2017

Further findings and discussions

In order to analyse more trends with data-oriented analysis, we have analysed the top 10 and the bottom 10 criminals LGAs in the past 22 years as the analysis sample and proposed some hypotheses to guide the analysis.

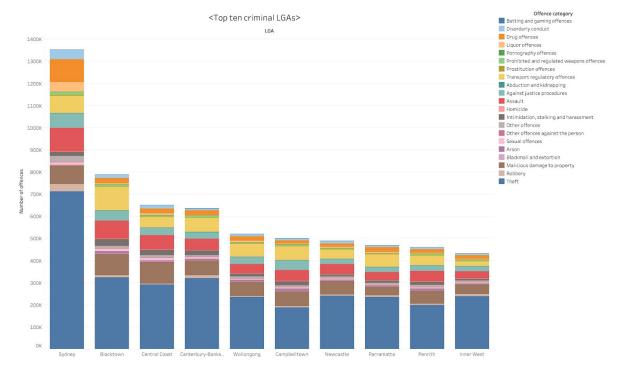


Figure 15: Top 10 LGAs with the Most Crimes (numbers)

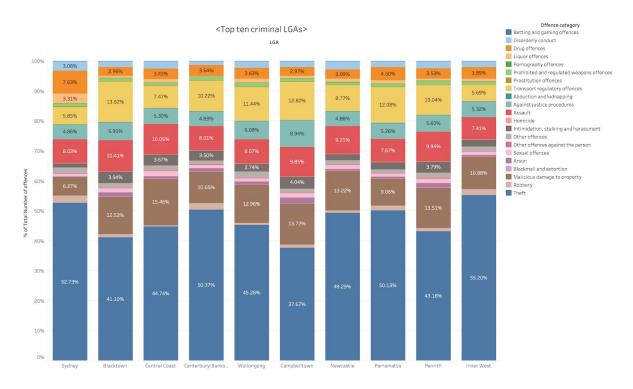


Figure 16: Top 10 LGAs with the Most Crimes (percentage)

The charts shown above (Figure 15 & Figure 16) are the top 10 criminals LGAs with the number of crimes the proportion of each crime. In order to play a role in comparison, and make the data more convincing, we rearranged the bottom 10 LGAs to generate the bar chart as follows (Figure 17 & Figure 18).

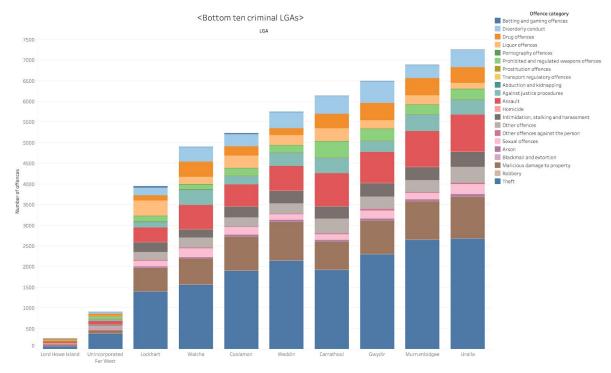


Figure 17: Bottom 10 LGAs with the Most Crimes (numbers)

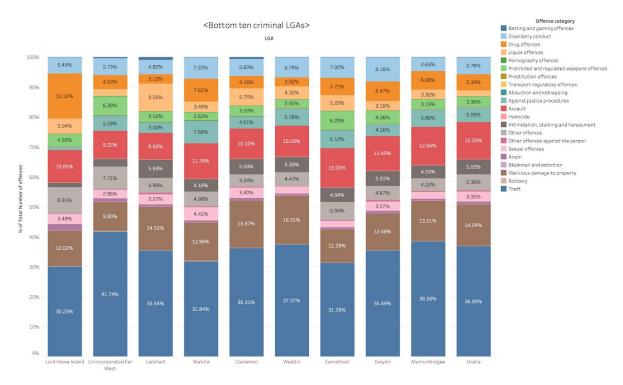


Figure 18: Bottom 10 LGAs with the Most Crimes (percentage)

Hypothesis 1

The crime against property is correlated with the income level. In other words, lower income level areas may be more prone to illegal activities, whereas wealthy areas may be opposite. In order to analyze the relationship between the income and crime rate, we extracted data from the Australian Bureau of Statistics and generated the following charts based on the census data (Australian Bureau of Statistics, n.d.).



Figure 19: Average Weekly Income in Top 10 Criminal LGAs

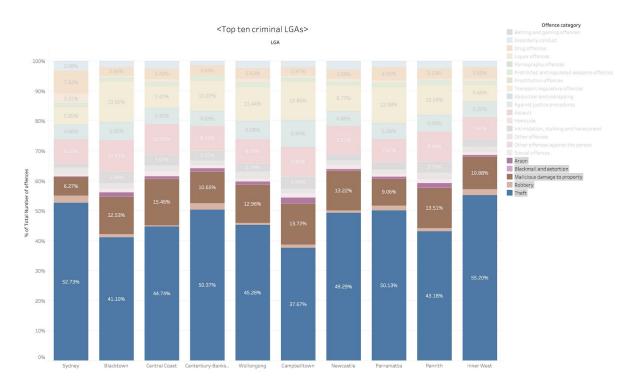


Figure 20: Percentage of Property Crimes in Top 10 Criminal LGAs

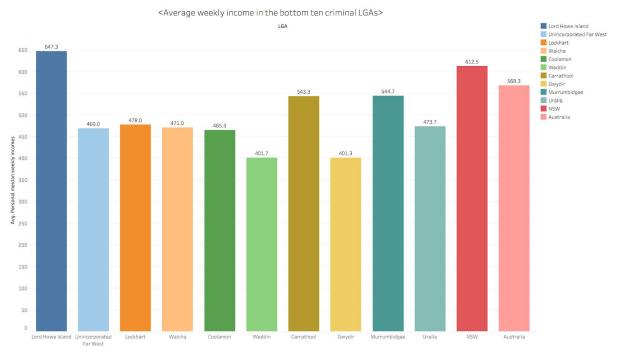


Figure 21: Average Weekly Income in Bottom 10 Criminal LGAs

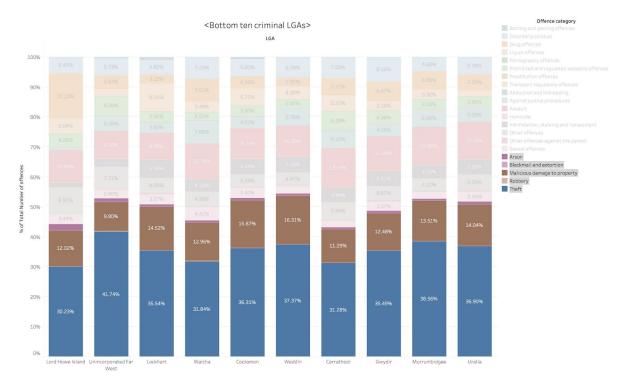


Figure 22: Percentage of Property Crimes in Bottom 10 Criminal LGAs

Of the top 10 regions in the list, only four LGAs were significantly lower than the average income level, while the other regions earned more than the median income of the NSW state. In addition, 80% of regional income on the bottom 10 bar chart is significantly lower than the median income of the NSW state. This also shows that the top 10 regions are more affluent than the bottom 10 regions after the crime.

From the figures, we could easily see the percentage of property crimes takes more weight in top 10 criminal LGAs rather than the bottom 10 LGAs. It's understandable with higher income, comes higher property crimes rate.

Hypothesis 2

The crime against society may be correlated with the unemployment rate.

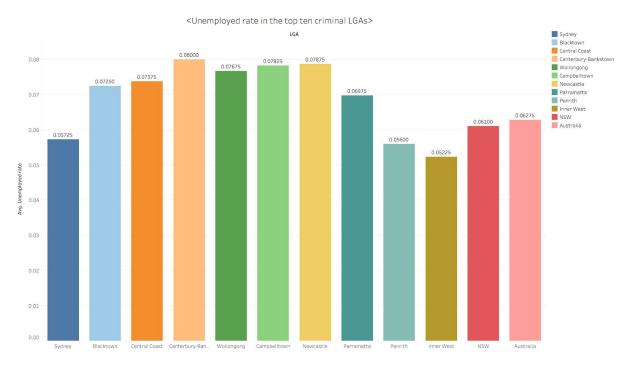


Figure 23: Unemployment Rate in Top 10 Criminal LGAs (numbers)

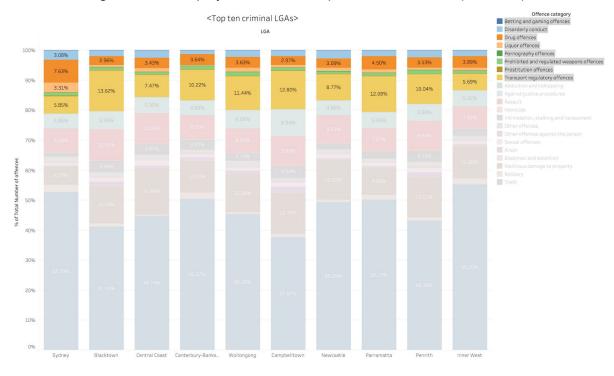


Figure 24: Unemployment Rate in Top 10 Criminal LGAs (percentage)

From the above data, the unemployment rate of the top ten LGAs is generally high, and 70% of the regions exceed the state's average unemployment rate. Correspondingly, in the map of crime types, the ratio of crimes related to traffic clearly dominates.

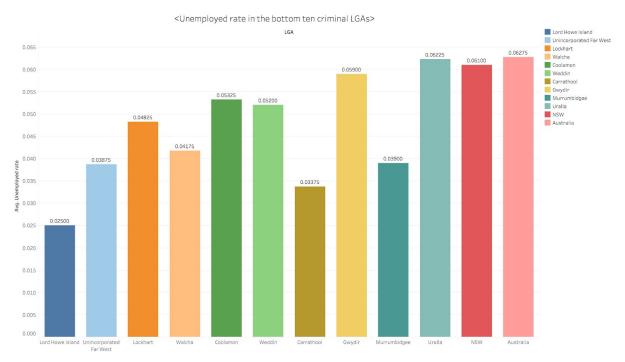


Figure 25: Unemployment Rate in Bottom 10 Criminal LGAs (numbers)



Figure 26: Unemployment Rate in Bottom 10 Criminal LGAs (percentage)

When we analyze the lowest crime group, we can find that eight out of ten regions have a lower unemployment rate than the state average unemployment rate. At the same time, the types of crimes have and the proportion of the crime against society caused by certain individual acts (pornography, alcoholism and pornography) has risen sharply.

From the above data, we can confirm the hypothesis that the unemployment rate does have a relationship with criminal behaviour. And based on this we can infer that in areas where

the unemployment rate is higher, people's behaviour is more inclined to hedonism, which is why the proportion of crimes related to alcoholism, drugs and sex is higher.

Hypothesis 3

The crime against the person is correlated with the geographic location of criminal LGAs.

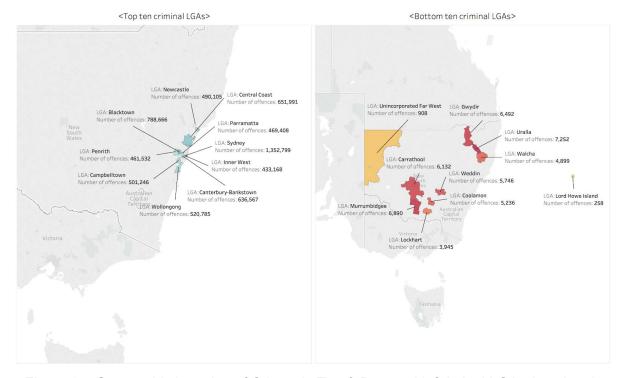


Figure 27: Geographic Location of Crimes in Top & Bottom 10 Criminal LGAs (numbers)

Based on the geographic location, the analysis group has found that the top ten criminal LGA areas are located in the surrounding area of the Great Sydney, which all clustered near the east coast of Sydney. In addition, another ten LGAs with the lowest number of crimes in the past 22 years is more rural and less urbanized area.

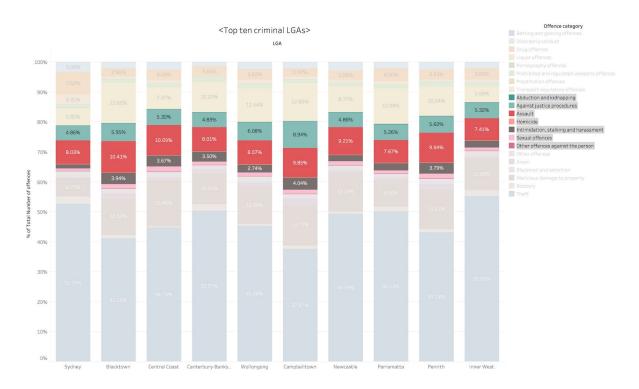


Figure 28: Geographic Location of Crimes in Top 10 Criminal LGAs (percentage)

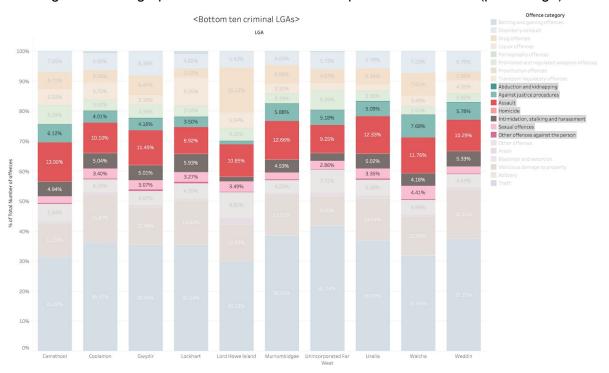


Figure 29: Geographic Location of Crimes in Bottom 10 Criminal LGAs (percentage)

When we compare the proportion of crime types based on the geographic location, surprisingly, crime against the person has more weight in less urbanized area than Greater Sydney. We can make bold inferences based on this phenomenon, which the human is social animals, people who live in isolation tend to be aggressive and are more likely to have a vicious crime against the person.

Reference list

NSW Government. (n.d.). *Population by Local Government Area*. Retrieved from HealthStats NSW: http://www.healthstats.nsw.gov.au/Indicator/dem_pop_lgamap

BBC News. (n.d.). Crime - Revision 3 - National 5 Modern Studies - BBC Bitesize. Retrieved May 25, 2019, from https://www.bbc.com/bitesize/quides/zgb2pv4/revision/3

National Incident-Based Reporting System (NIBRS). (2012). *Crimes Against Persons, Property, and Society* (the United States, U.S. Department of Justice, Federal Bureau of Investigation)

Australian Bureau of Statistics (n.d.). Census. Retrieved May 27, 2019, from https://www.abs.gov.au/websitedbs/D3310114.nsf/Home/census

Why Crime Increases Over the Holiday Period. (2017). Retrieved May 27, 2019, from https://lylawyers.com.au/why-crime-increases-over-the-holiday-period/