ASSIGNMENT 2
REPORT
DATA VISUALISATION USING TABLEAU
(Dataset used - NSW Domestic Violence Quarterly Report)

Introduction:

"A picture is worth ten thousand words" The world-famous quote by Fred Barnard establishes the foundation for the study of visualisation. Numerous amounts of data are generated every second but it is only viable, if it can be integrated, observed, studied and then provide some meaningful insights. Being simple to understand, humans consume knowledge more efficiently with graphical representation. For this, we need tools and techniques to acquire and represent the data in the most efficient form.

Heitzman (2019) defines data visualisation as the process of acquiring data or statistics that may then be transformed into graphs and maps to enhance perception and knowledge of the data. It is typically connected to any graphic component that presents information in any field or field of study. (Kelleher & Wagener 2011). Businesses rely on data visualisation to present vast volumes of complex data in the clearest possible terms. (Qin et al. 2020, Bayoumi et al. 2018 and Moer & Purchase 2011).

In this report, Visualisation tools and techniques are used to graphically represent the NSW Domestic Violence Quarterly report, issued by the NSW Bureau of Crime Statistics and Research. Insights from visualisation provide an easy understanding of data and are represented understandably so the statistical analysis can be performed to find the trends and furthermore.

Visualisation Software:

For our visualisation, we are going to use "**Tableau**" software. Easy to use interface helps the implementation of visualisation simple and it provides various visualisation techniques. As mentioned in tableau websites, it is a powerful, flexible, and secure data visualisation platform. Dataflair (2019) listed further benefits of using this tool, including its high-quality visual image capabilities, reliable and consistent performance, and linkage to multiple information sources. This report will provide insights on the visualisation created on mentioned data and the discussion on the steps taken to achieve that visualisation.

Dataset:

NSW Domestic Violence Quarterly report, issued by NSW Bureau of Crime Statistics and Research provides the domestic offences stats in NSW state for the complete year of 2021. Before starting with the dataset, it was cleaned and reviewed, which

avoids errors and hence some data will not be represented. The data is divided into three types of content, first is Summary and Detailed Offence information, which provides statistical data of the time of incidents (day and time), Location, Premise type, and alcohol-related crimes. The second type of content provided is Victims' information such as age, gender, and aboriginality and the third type of content is Offenders' Information (same age, gender, and aboriginality).

Visualisation implementation and Analysis:

For our visualisation, we will start with Aboriginality of Victims' and offenders' information. Aboriginality of offenders provides types of crime such as Abduction and Kidnapping, Attempted Murder, Breach of Apprehended Violence Order, Domestic violence related assault, Intimidation, stalking and harassment, Malicious damage to property, Murder, Sexual assault, and Sexual touching, sexual act, and other sexual offences. The Statistics of these 9 types of crimes are mentioned over on the aboriginality of the offender as Aboriginal, non-aboriginal, Unknown and Total.

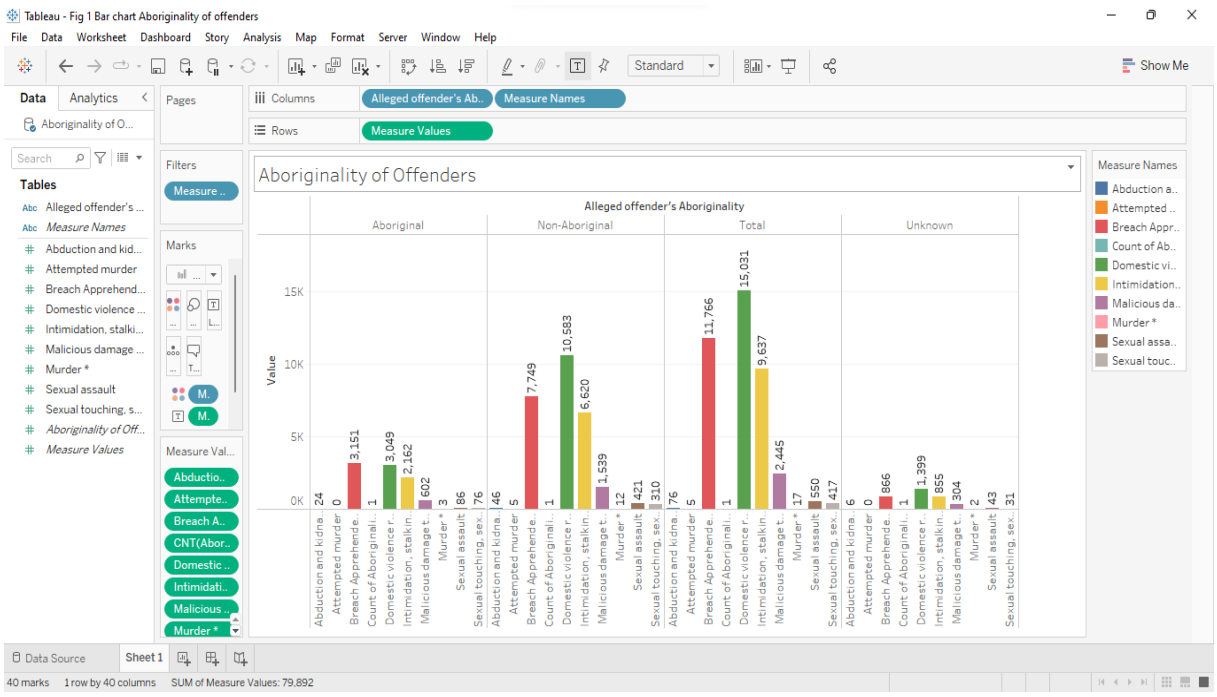

Figure 1: Aboriginality of Offenders

We implemented the aboriginality of the offender which is a subset of the NSW Crime data set with the use of a bar graph. Here all 9 types of crime are mentioned with differentiation of aboriginal, non-aboriginal, unknown, and total. The bar

graph along with types of crimes shows a proper number of types of crime, also providing the exact number of crimes that happened in every 4 categories. We acquired this data by inputting 2 measures in rows and 1 dimension in the column. The first measure is the types of crimes and the other is the Aboriginal category. The only dimension in the column is the values of the measures which depict the numbers in every category. This Visualisation provides complete knowledge about the Aboriginality of offenders along with the type of offence they committed. Segregation into aboriginal and non-aboriginal data provides offences committed by almost every offender around the state. Different colour bar graphs help in differentiation and exact number around that data.

Non-aboriginal offenders are 2 to 3 times more in number in comparison with aboriginals. Domestic violence-related assault is more for offenders in both categories. Out of 15031, 3039 offences were committed by Aboriginals while 10583 offences were committed by non-aboriginals and 1399 were unknown.

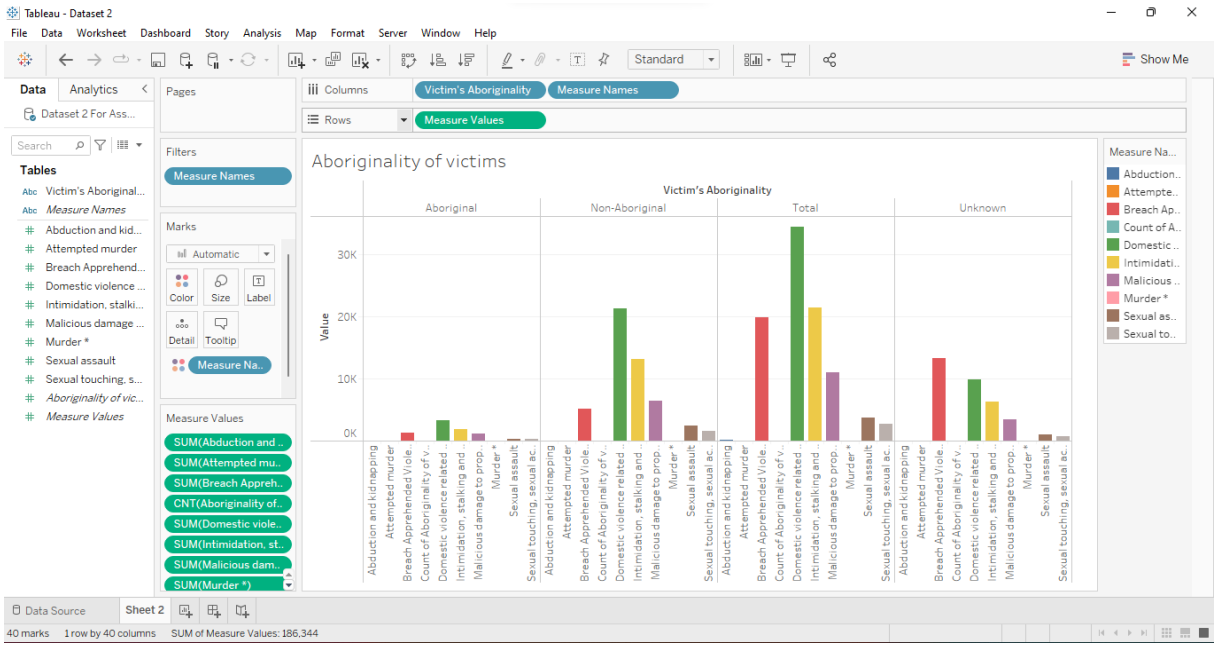

Figure 2: Aboriginality of Victims

Same way as for offenders, the Aboriginality of victims' visualisation is also implemented. The number of victims of Domestic violence-related assault crime type is more. Non-aboriginal victims' count is almost 5 times that of aboriginal victims' count in Domestic violence-related assault.

Instead of implementing different visualisation for Offenders and Victims, it is better if it can be generated in single visualisation abiding it should provide all relevant information as it was provided earlier. To do this we coupled two sub-datasets as it was possible as both have the same crime type division. Doing such a union allowed us to represent data on a single page and with all information that can be extracted.

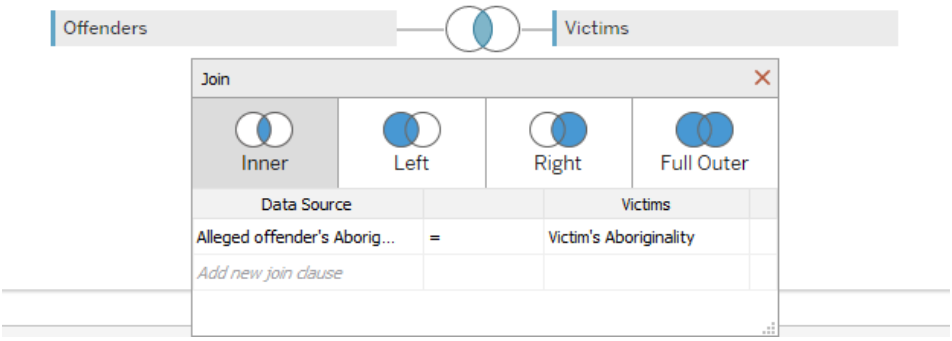

Figure 3: Coupling Offenders' dataset with Victims' dataset.

On coupling these two datasets many errors contained cells were cleared. Along with that, we changed the headers' names of columns which was causing unnecessary confusion. After this following visualisation was the result.

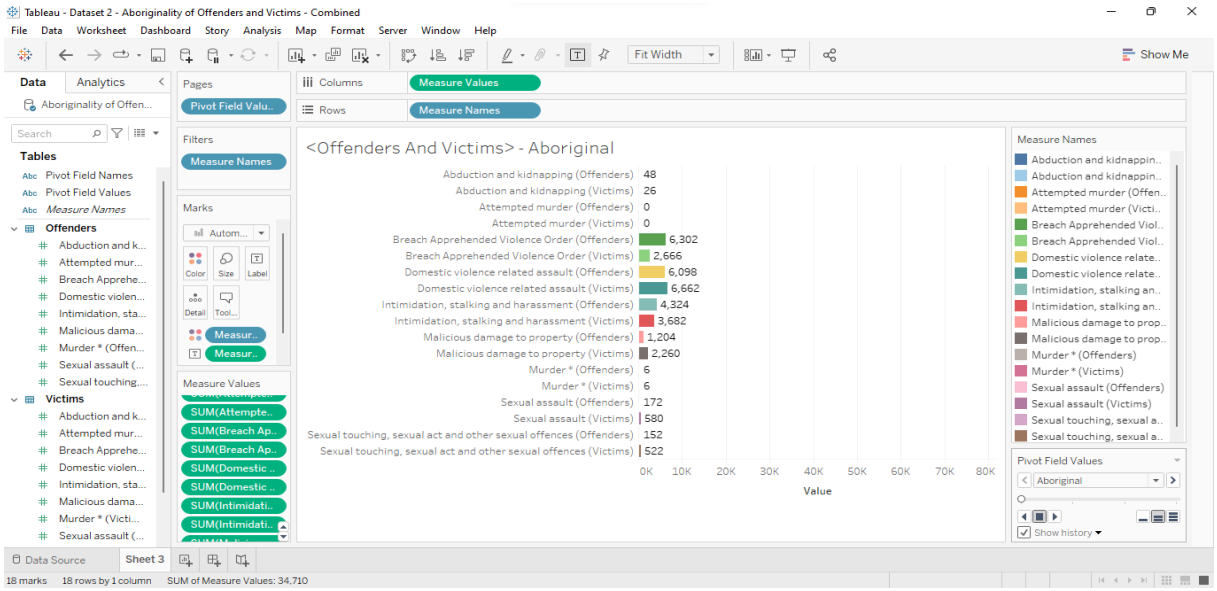

Figure 4: Aboriginal Offenders and Victims.

By combining two datasets, visualisation is now easier and provides more information in a single interface. Previously we worked on using victims' and

offenders' data separately. Now all crime data along with the aboriginality of both victims and offenders are shown on a single page. Combining 2 datasets enable us to create a pivot field of crime names and crime number which are now merged and playing in a loop plays the 4 screens, the first one of this in figure 4.

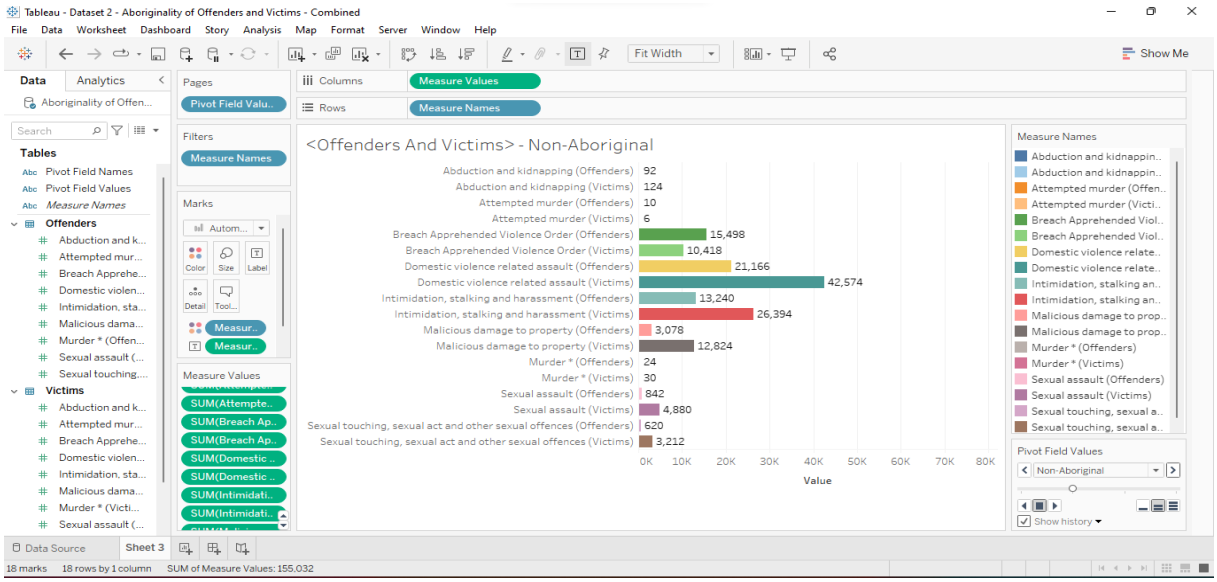

Figure 5: Non-Aboriginal Offenders and Victims.

Non-aboriginal Offenders and Victims are shown on a single screen which helps in comparing offender and victim data instantly. Mentioning count on each bar graph with different colours for different types of crimes gives complete stats on a single screen.

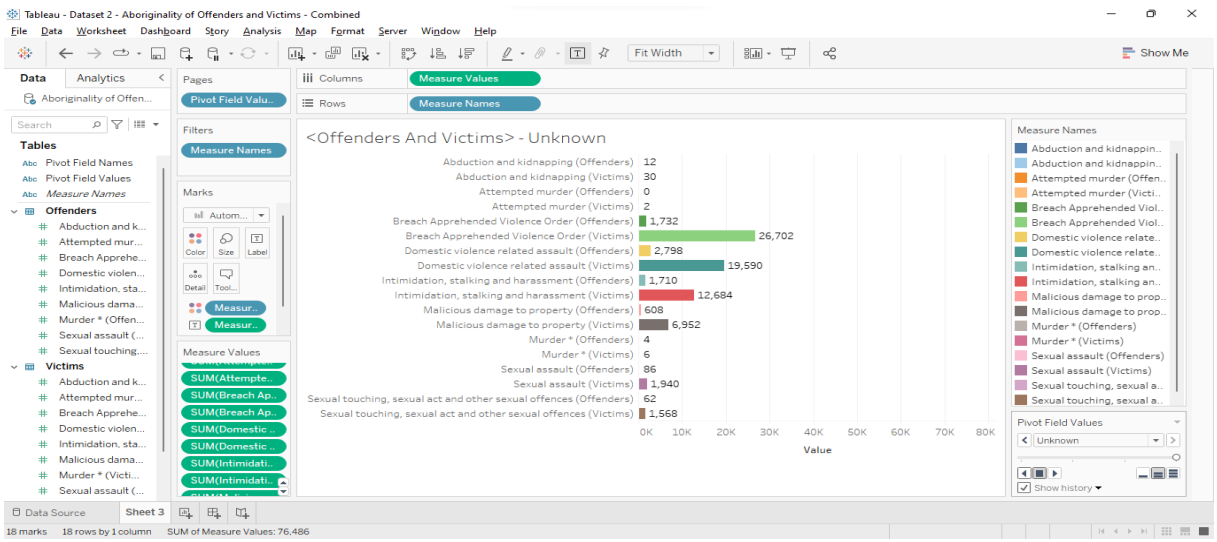

Figure 6: Unknown Offenders and Victims

Figure 6, provides a count of each offence type of unknown category. More victims are in Breach Apprehended Violence Order for the unknown category. Though offenders are very less for the same offence type.

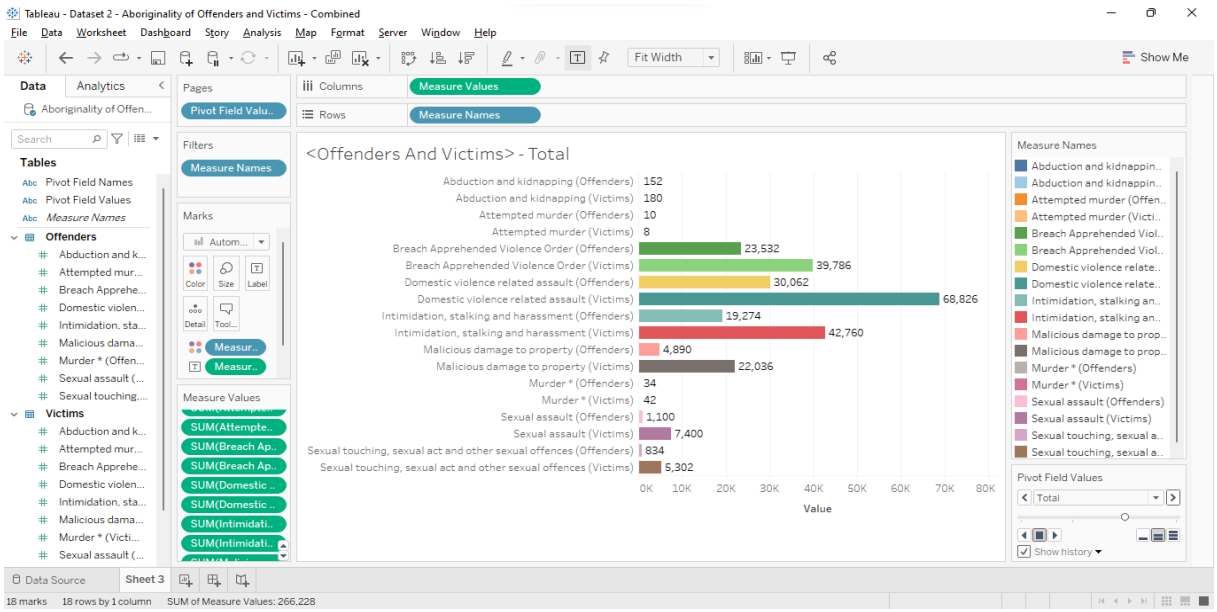

Figure 7: Total number of Offenders and Victims

Finally, the total count of victims and offenders is shown in figure 7, which shows Domestic Violence related assaults have a huge number of offenders and the same crime type.

With the help of Pivot and the union of two datasets in a single workbook, we can achieve all information mentioned in both datasets. Selecting loop playback will keep repeating each screen and show data categorise into Aboriginal, Non-aboriginal, Unknown and Total.

As we are dealing with limited statistical data, Bar graphs are the most efficient and from the above visualisation, we can provide the premise for the same. Though bar graphs do not provide good visualisation on large datasets having many measures or parameters but still for a nominal number of parameters bar graph outperforms in explaining the details more easily. As we move to the next data set where month-wise stats of crime are given it shows many offence types (same as previous) that happened in the span of 12 months in 2021. Mentioning data from Jan 2021 to Dec 2021 following is the visualisation created with help of filters, colour marks, and more.

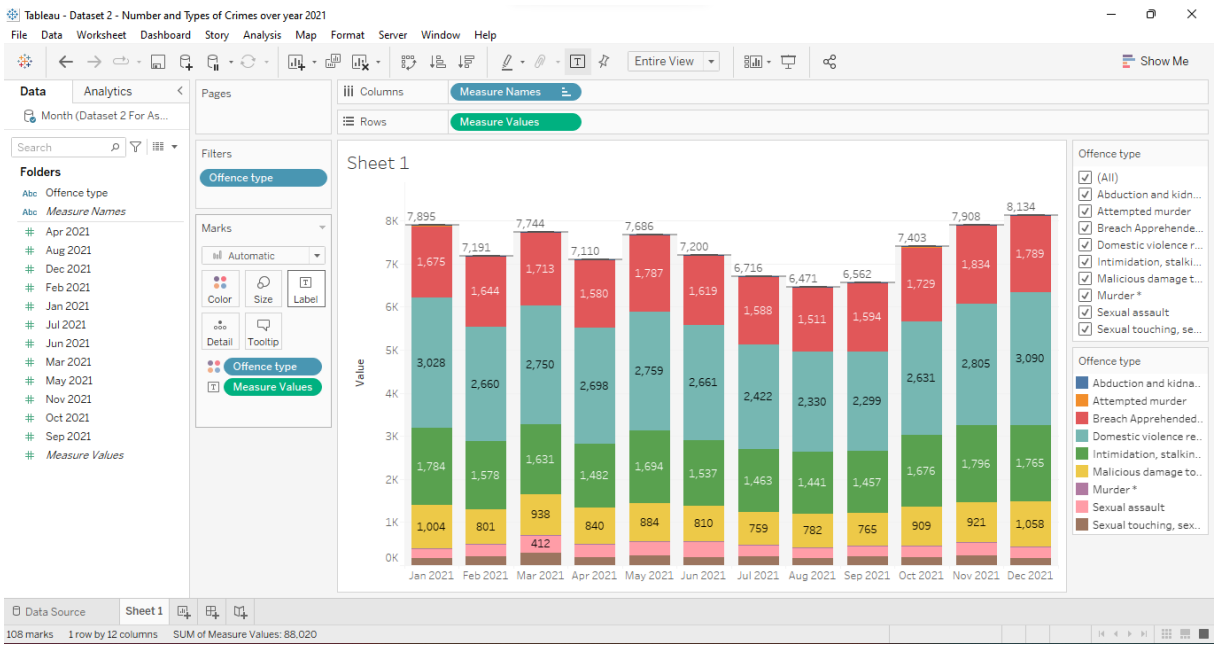

Figure 8: Incidents by offence type and Month

As shown in figure 8, by inserting the Measure name in Columns we got all months of 2021. Sorting done on Measure names to arrange month in data source ordered (in sequence from Jan to Dec 2021). In the Filter section inserting the offence type allow us to filter the offence type to gain more insights about selective offence type. As some offence counts are very low such as Abduction and kidnapping, Attempted murder and murder, by selecting only those in the filter their count can be obtained which is shown in figure 9.

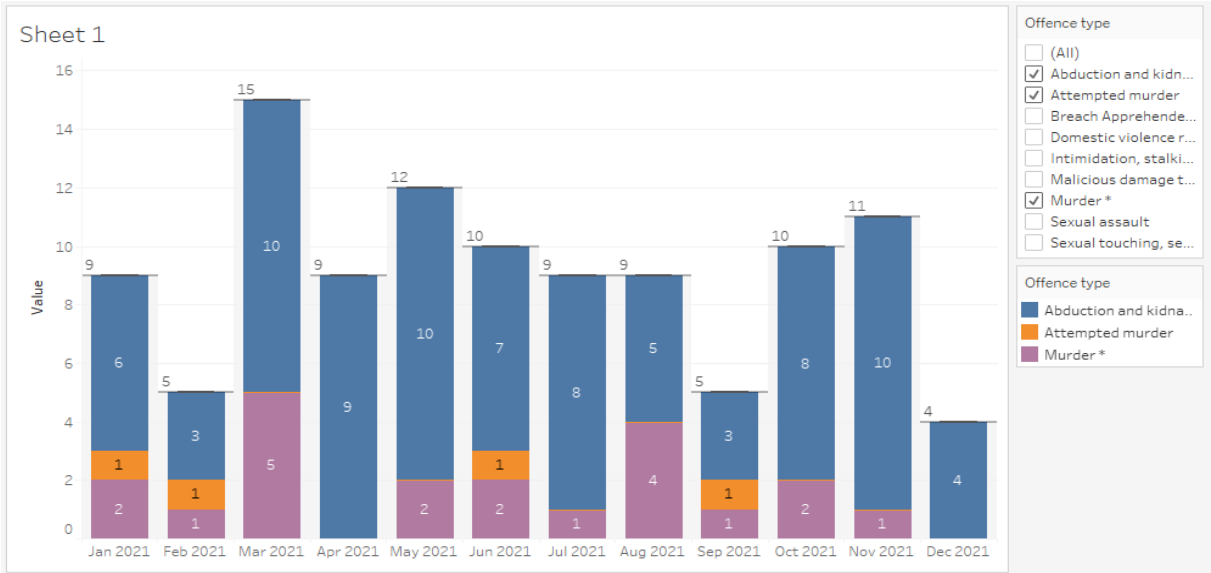

Figure 9: Incidents by offence type and Month – Filtering the data

Inserting offence type in colour mark, colourful bar graph representing specific colour for offence type allows identifying its type and its intensity. Further adding measured values data in the label mark, provides the count of each offence in a particular month. This makes it easy to understand data from every parameter available.

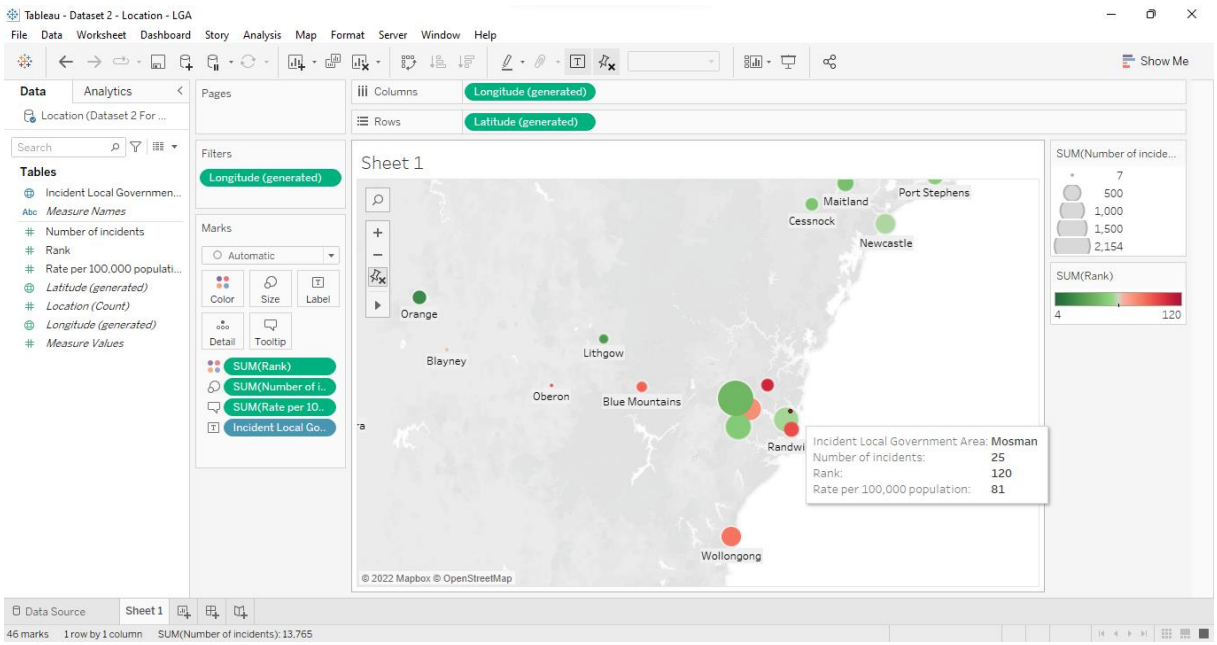

Figure 10: Incident by Local Government Area using Geo Map

As shown in the above figure we moved to another dataset that provides Data on domestic violence incidents that occurred in 2021 in every Local Government Area in NSW. It provides the number of incidents, rate (Per 100k population) and Rank. We first created longitude and latitude values by selecting the geographic role of the city. As all LGAs do not have city names, it excludes most of the data. Still implementing it on the map it shows LGA names and data only after hovering the cursor over the tag marked for LGAs. We tried inserting rank values in the colour mark, several incidents in the size mark and rate in the tooltip mark it provides the visualisation as shown in figure 10. It is not good visualisation as the size of the circle is small, we try for more area and if we zoom in, some circles overlap with another circle of LGA. Hence to better implement the visualisation we used multiple graphs showing every parameter across every LGA with help of a bar and line graph.

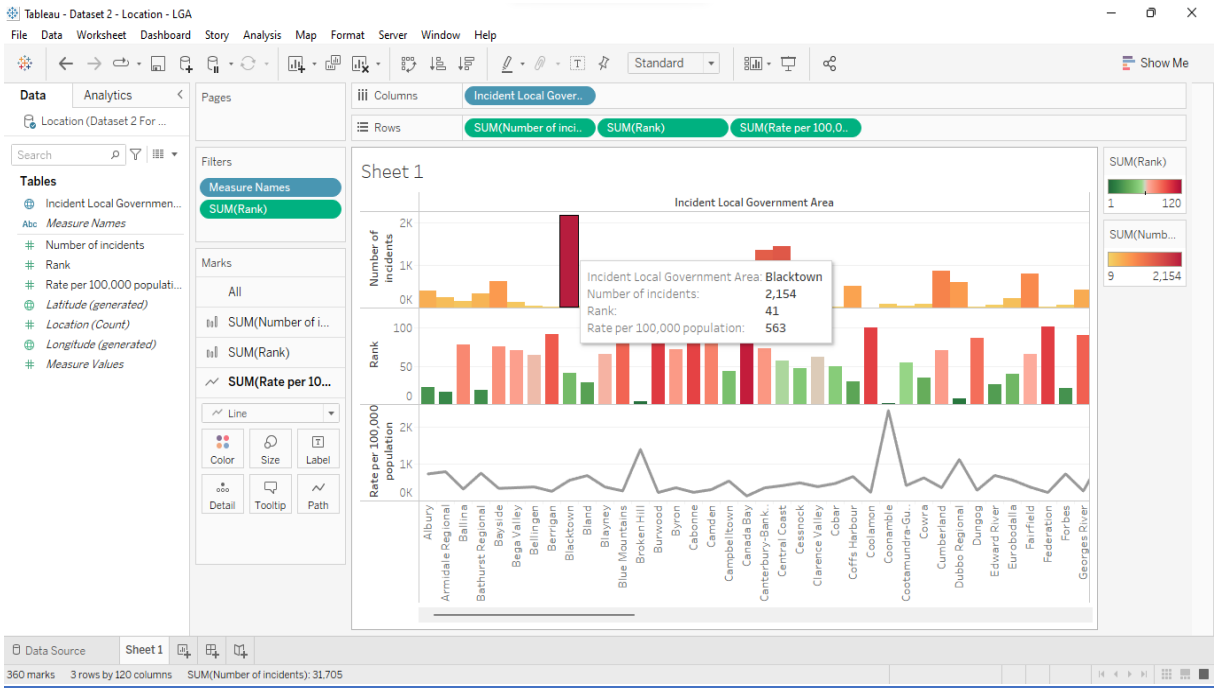

Figure 11: Incident by Local Government Area using Bar and Line Graph

The same data with help of bar and line graph as visualisation technique provides the drastic result. All 3 parameters are easily distinguished now, and details can be noted easily. Hovering over any part provides complete data of LGA under the cursor. Inserting all three parameters in rows allows marks where we select the Bar and Line graph. For rank and number of incidents, a bar graph with a colour mark is used while for rate only a line graph is used.

Now we further move to another dataset where 9 types of offences and their occurrence in the day and on time range are mentioned. This data provides information on which weekday most offences take place and around which time range (6hour range). We worked on using a bubble map to implement visualisation and the result is shown in figure 12.

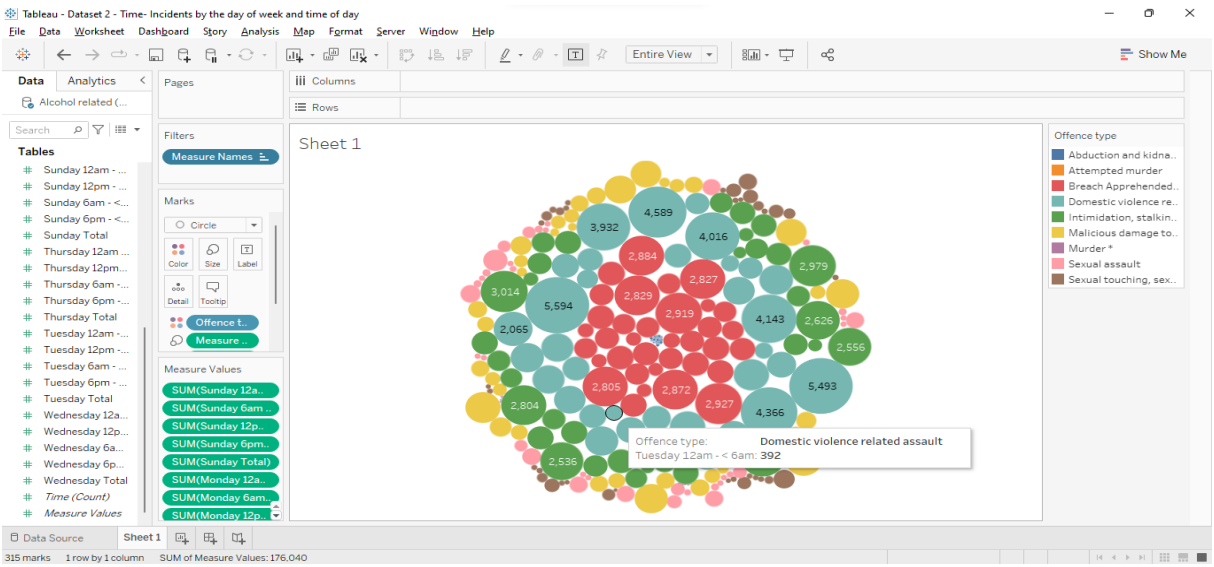

Figure 11: Incidents by Weekday and time range using a Bubble map

We achieved this visualisation by selecting the circle in shape mark and inserting offence type in colour mark and measuring the value in size mark. Colour-wise distribution was observed with a bigger circle indicating a greater number of incidents in that day and time range. This allows for extracting preliminary data but does not provide more insights about the trends or actual count of incidents on day and time range. Hence try using a Simple tree map which is shown in figure 12.

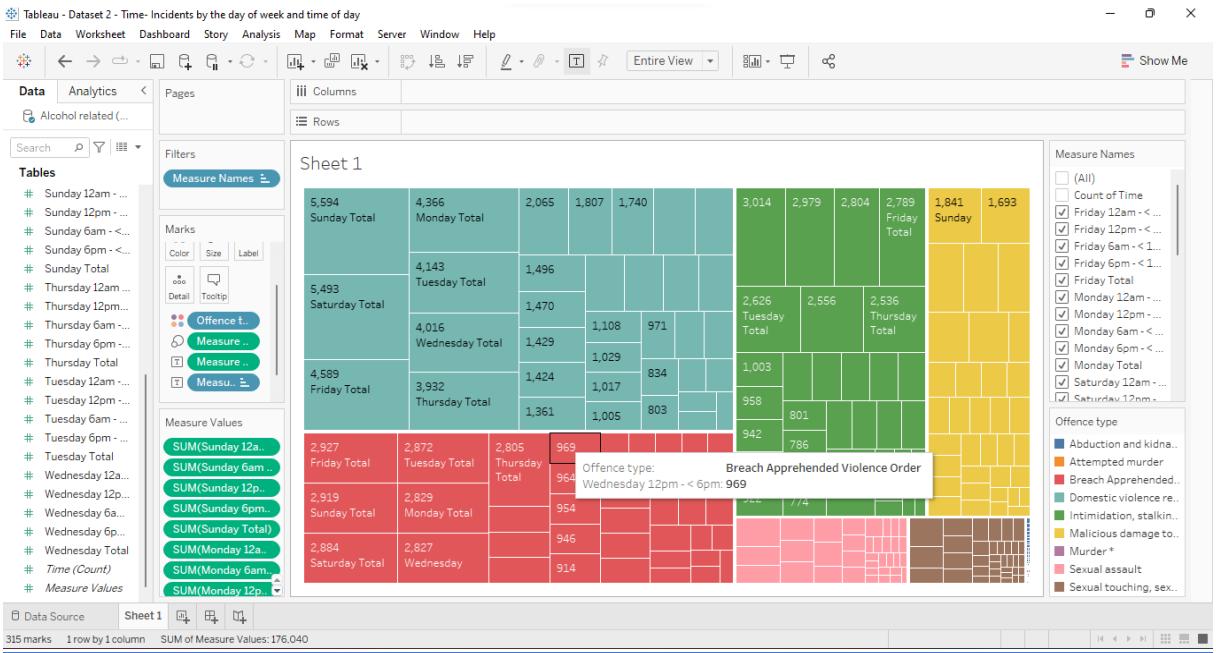

Figure 12: Incident by weekday and time – Treemap

Same as a bubble map this treemap shows data only from those who have high numbers and does not able to provide any trend-related data. Hence we tried implementing it with a bar graph and the following were the result.

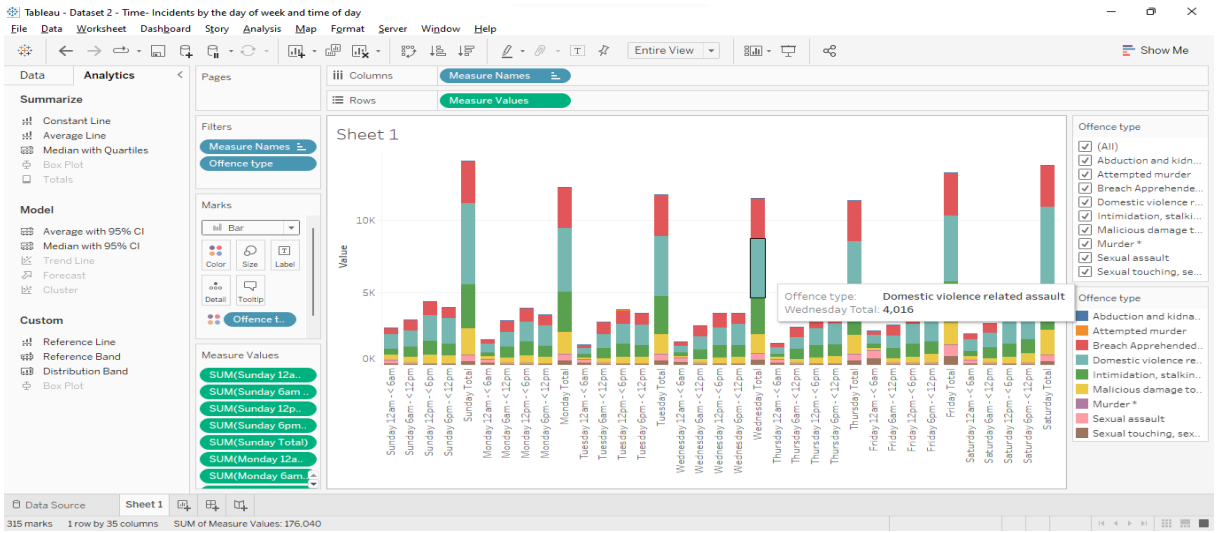

Figure 13: Incident by weekday and time – Bar Graph

Easily implemented by assigning measure names and measure values in columns and rows respectively provides data for all weekdays with 6 hours time range. Colour mark for offence type provides colour differentiation for identifying offence type. Also filtering offence type we can be able to do the selection so the data can be adjusted for the parameter. But to represent this data more efficiently, an area graph is used which creates good visualisation explaining all details mentioned in the dataset.

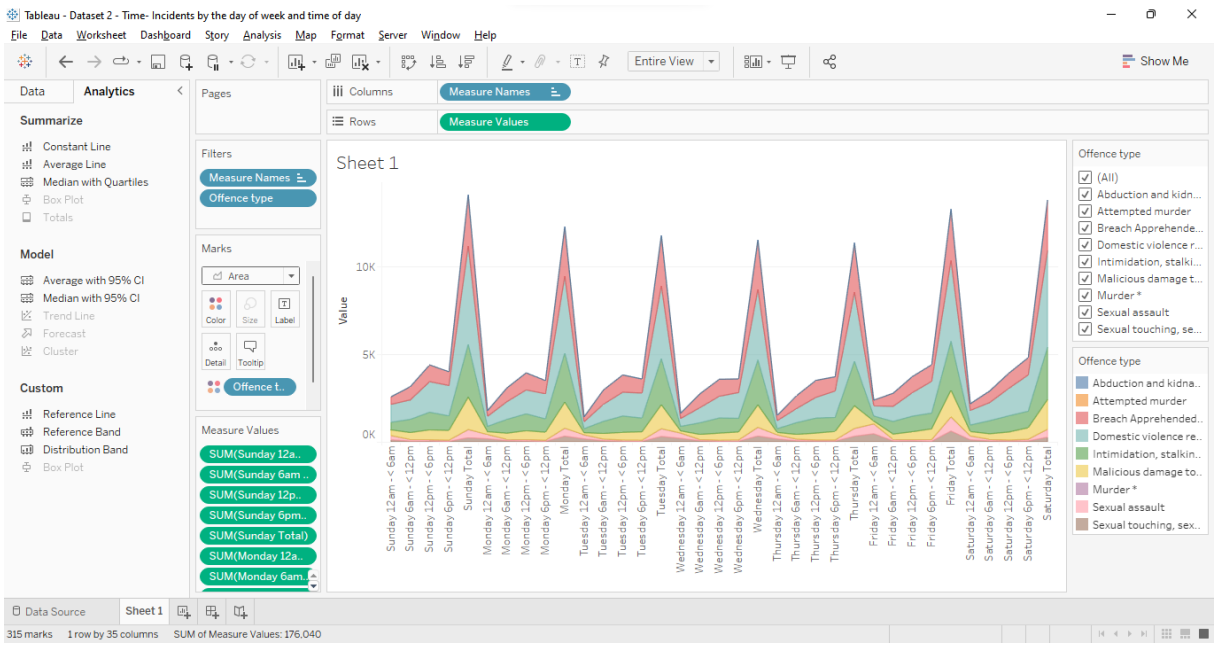

Figure 14: Incident by weekday and time – Area Graph

This area graph helps in representing trends which tells more about the data. As we can see in figure 10, Domestic Violence related assaults are higher in comparison with any other offence type. Also, more crime occurs at weekends rather than on weekdays. Sunday's 12 pm to 6 pm range shows more number incidents along with Saturday's 6 pm to 12 am. Hence it is easy to read and understand statistical data in the area chart graph. Filter on the right side allows selection of offence types which helps in gaining insight about a low number of incidents in offence types. The total number of incidents in a day can be spotted by clicking any weekday total column. This area graph allows more insights that any other previous visualisation shown for this dataset.

Conclusion:

By implementing various techniques in tableaus such as pivot, union, mark and filtering we have achieved visualisation on various datasets of the NSW Domestic Violence Quarterly report. Statistical data with limited parameters is easy and efficient when implemented in bar graphs and area graphs. Though there are numerous techniques and tools to always improve the visualisation. We tried implementing visualisation from the techniques we learn hence being able to provide more insights and analysis.

References:

Bayoumi, S., Alghamdi, K., Alqusair, D. and Alfutamani, A., 2018, April. Visualization of Fire Accidents in Saudi Arabia. In 2018 21st Saudi Computer Society National Computer Conference (NCC) (pp. 1-6). IEEE.

Fry, B. 2008, July. Data Visualisation. *O' Reilly Media.*

Heitzman, A 2019, 'Data visualization: what it is, why it's important & How to use it for SEO', SEJ, 29 January 2019, viewed 27 July 2022.

Kelleher, C and Wagener, T 2011. Ten guidelines for effective data visualization in scientific publications. *Environmental Modelling & Software,* 26(6), pp.822-827.

Liang, J, Nguyen, QV, Simoff, S & Huang, ML 2015, 'Divide and conquer treemaps: visualizing large trees with various shapes', *Journal of Visual Languages and Computing*, vol. 31, pp 104-127.

Moore, A.V. and Purchase, H., 2011. On the role of design in information visualization. *Information Visualization*, 10(4), pp.356-371.