**Data Analysis Report: Investigation of Abuse and Neglect**

Samantha Paige Minchew

Professor Satvika Marrapu

*Data Analysis and Knowledge Discovery, INFO 4670, Section 001*

Assignment 3

31 October 2023

To begin, I changed my data set from “Ramen Ratings” to “Investigation Report of Abuse and Neglect.” The reason was for the lack of data in my original data set and because I had no more questions on the report. It was becoming difficult to analyze such a small data set with a lack of information. I selected this data set because it had a plethora of information that would give me the freedom to analyze it how I wanted. The date, region, type of report, and count of allegations conclude the information I have collected in this new data set.

**DATA METHODS/PRE-PROCESSING:**

Since it is an extensive data set with new reports constantly being updated, this report is listed as continuous data. The analysis report covers 11 different regions across Texas and categorizes them between county and region. Having a category and subcategory helped create a map of investigation reports while being able to dig deeper for more information. I had many questions about my data set which I was able to answer using Excel and Tableau. In this report, I wanted to analyze the crime rate between different areas in Texas, look at the growth or decline of investigative abusive reports over 11 consecutive years, and discover the most common types of allegations in Texas. I was able to categorize the type of abuse by year or region which created a clear picture to answer the questions I wanted to solve in this data report.

**OBJECTIVES:**

To be able to identify types of abuse or neglect by region, date, and allegation to determine if the crime rate is increasing or decreasing and what areas have higher incident reports. Using this data set, I will be able to analyze and cross-examine information to categorize and understand the evidence outlined. This analysis account will answer the following questions I had before examining the investigative analysis report:

1. **DATES**
   1. *What years have higher/lower crime rates?*
   2. *Does crime increase or decrease throughout the years?*
   3. *What crimes escalated/deescalated throughout time?*
2. **REGIONS**
   1. *What regions/counties have higher/lower crime rates?*
   2. *Do certain regions have a consistent type of abuse over other areas?*
   3. *What regions have fewer reports and what do those areas have in common?*

**DATA ANALYSIS:**

The data set reported 217,686 different investigations of abuse and neglect. Using Excel made it easy to categorize per year by adding slicers to a detailed pivot chart. Figure 1.a is calculated by year and count of confirmed allegations on the X-Y axis.

***Figure 1.a***

A screenshot of a computer

Description automatically generated

Listed, is a pivot chart detailing all occurrences or reported abuse categorized by color on the X-Y axis of year and count. Using the slicer, you can view data based on year, region, or type of abuse. Using this function, you could analyze data from a selected year, county, or reported abuse to identify what area has common crime and over what years.

***Figure 1.b***

A screenshot of a computer

Description automatically generated

In 2008, the highest type of abuse was medical neglect. Now, you can search for differences in various cities and compare them to each other.

***Figure 1.c***

A screenshot of a computer

Description automatically generated

Compared to 2008, 2018’s highest reported abuse was neglectful supervision. Like Figure 1.b, you can now select which region of reported abuse wanted to investigate and compare and contrast them to other cities in that year.

***Figure 1. d***

A graph showing a line of different colors

Description automatically generated with medium confidence

Using Tableau, this area chart analyses data throughout the year by categorizing types of abuse per color. You can see inflation and deflation of certain reported allegations compared to each year.

***Figure 1.e***

A graph of different colored lines

Description automatically generated

In Tableau, you can take the same information but focus on the year and type of abuse. For this sheet, I did not include region because I wanted to compare the overall count of occurrences over the years. As you can see in the graph, crime was consistent then certain types of abuse rose in 2015. This Gantt view in Tableau can categorize reported abuse per year but does not focus on total abuse throughout these years. This graph is only helpful if you want to analyze specific types of data. To know if all abuse has risen or declined, you’d have to analyze a Pareto chart.

***Figure 1.f***

A screenshot of a graph

Description automatically generated

Using a Pareto chart, we can identify the crime rate per year and determine whether it increases or decreases overall. While some types of reported allegations increased, overall abuse decreased from 2008 to 2018.

***Figure 2.a***

A graph of different colored squares

Description automatically generated

With the same information, I created a histogram chart on Tableau that categorizes the information by region. As depicted in the graph, Austin has the highest reported allegations of abuse while El Paso has the lowest. Listed in color, you can see the type of reported abuse along with the count and date in each city (when selecting which graph).

***Figure 2.b***

A map of texas with different colored squares

Description automatically generated

Analyzing the type of abuse by region, this graph lists the most common type of abuse in each area. However, this graph does not include the count of occurrences.

***Figure 2.c***

A map of texas with blue squares

Description automatically generated

Figure 2.c depicts the region by count of crime. Compared to Figure 2.b, this graph allows us to identify which cities have higher abuse rates per capita, regardless of the type of allegation.

***Figure 2.d***

A screenshot of a graph

Description automatically generated

This scatter chart is also a good way to analyze the number of allegations by comparing the count of the county to confirmed allegations. Categorized by color, it’s easy to see that neglectful supervision is the most common type of abuse regardless of year.

***Figure 2.e***

A screenshot of a graph

Description automatically generated

Examining a pie chart gives us the same answer as the scatter chart, just without viewing the total amount and area of crime. If you wanted you could add in date, count, and region, but for this graph, I wanted to solely analyze the most common and least common types of abusive allegations from 2008 to 2018. It’s easy to read that this graph lists neglect as 60% of the most common type of abuse in Texas, while abandonment is at the lowest.

**CONCLUSION:**

Overall, this data set was extremely detailed. Using Excel and Tableau, I was able to analyze and cross-examine data referencing dates, regions, and types of allegations. Using charts, graphs, tables, and slicers creates a detailed and easy-to-read visualization of data. Creating visualizations of data helps us analyze big data sets thoroughly while also finding answers quickly. In this report, I analyzed which years have higher crime rates, and investigated whether crime increased or decreased while also finding out which type of crime was the most reported and the lowest. Creating area graphs and histograms categorized crime rates so that we could discover which regions have higher crime than others and which areas have which type of reported allegations.