**ASSIGNMENT SUBMISSION COVER SHEET**

**Student Id: BIDA23-017**

**Student names: KGOSI B. LETSHWAO**

**Student email: BIDA23-017@THUTO.BAC.AC.BW**

**Cohort: -**

**Assignment title: BI 201 INTRODUCTION TO DATA ANALAYTICS**

**Date of submission: -**

**Programme of Study: BUSINESS INTELLIGENCE AND DATA ANALYTICS**

**Year of Study:2023**

**Intellectual property statement By checking the box below, I certify that this assignment is my own work and is free from plagiarism. I understand that the assignment may be checked for plagiarism by electronic or other means and may be transferred and stored in a database for the purposes of data-matching to help detect plagiarism. The assignment has not previously been submitted for assessment in any other unit or to any other institution. I have read and understood the Botswana Accountancy College plagiarism guidelines policy.**

**☐ Agree**

**Signature……………………………………. Date………………………………………….**

**TASK 1**

a)

* Area name: Selecting the area name as an element of interest will help identify the areas with the highest crimes prevailing. This will help guide the police department pinpoint areas that require an increase in police officer patrol. This element of interest is of qualitative data type, in which a bar chart could be used from the data derived to show areas which have the most crime occurring at.
* Crime committed: Knowing which crimes are mostly committed aids the police chief to know which areas to sensitize the public on so that they are aware of how to better protect themselves against being victims. Crime committed data is categorical and once projected as a pie chart, the most prevailing crimes in LA could be easily identified.
* Date occurred: This element will be helpful to show which months have the highest crime rates, thus giving the police chief and their department a better picture of which seasons to intensify patrols on. Date occurred is categorical and a line graph could be used to determine which seasons have the most crimes being committed.
* Victim sex: Patterns could be drawn from this victim sex element to display which crimes affect which sex mostly. The police chief will now have a better understanding of where to allocate resources to counter the situation. The data collected about victim sex is quantitative therefore, a bar graph could be used to display how various crimes affect each sex.

b)

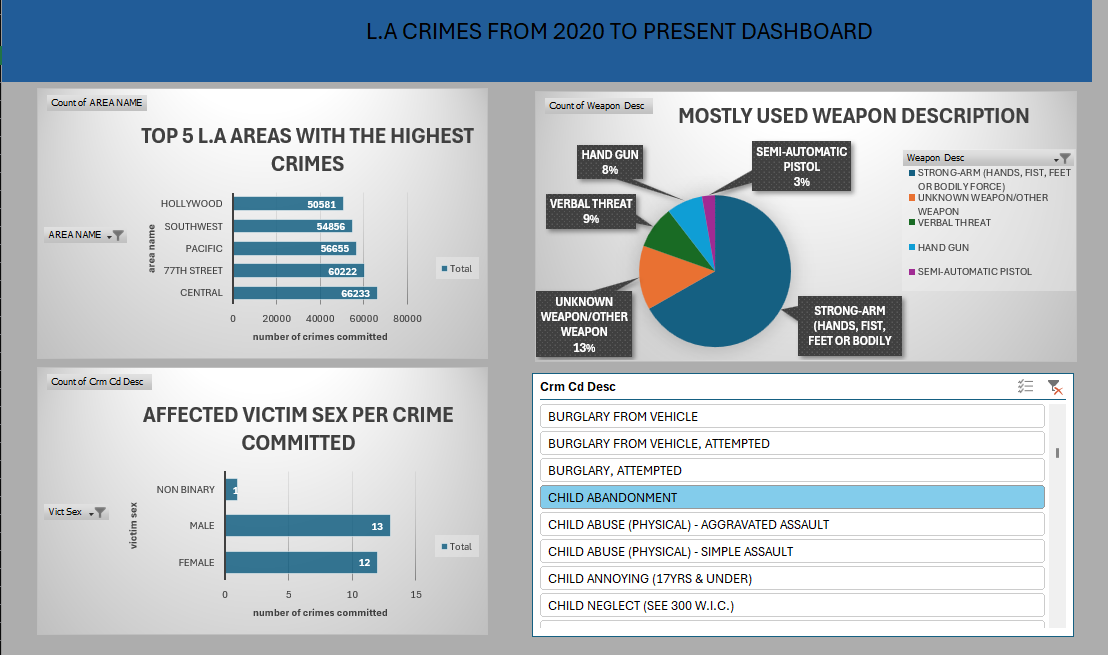
* Missing data error: there was evidence of values not recorded when the data was collected. To clean the data, the missing data value was replaced with the mean as per the recommendation from the text book "Practical Statistics for Data Scientists: 50 Essential Concepts" by Peter Bruce and Andrew Bruce (2017). Quoting the book, "Replacing missing values with the median preserves the distribution's central tendency without being overly influenced by extreme values" (p. 37).
* Inconsistency in data recording: As records were collected there was an inconsistency in entering the records with a value pertaining to another written in a different format. To rectify the inconsistency the value was rewritten as a single value as per the recommendation of Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer. It reads, "Standardizing and normalizing data helps in aligning different scales, particularly when the data originates from various sources or is recorded in different units, ensuring consistency across the dataset." (Chapter 14, Section 14.3).
* Outliers were evident: in the collection of data and projecting it in the form of scatter graphs outliers were within the dataset. To solve for this defect, extreme values were replaced for in order to limit their effect. This was as per the recommendation of "Introduction to the Practice of Statistics" by David S. Moore, George P. McCabe, and Bruce A. Craig. It reads, “Outliers can often distort statistical analyses. Winsorization, which caps extreme values, is one way to manage outliers without excluding data entirely” (Moore, McCabe, & Craig, 2017, p. 138).

c)

* CHART TO DISPLAY THE MOST AFFECTED VICTIM SEX PER CRIME
* CHART TO DISPLAY THE MOST COMMONLY USED WEAPON DURING CRIMES

* CHART TO DISPLAY THE TOP 5 AREAS WITH THE HIGHEST NUMBER OF CRIMES COMMITTED

d) DASHBOARD



e)Inferences:

1. Central Los Angeles has the highest crime rate: among the top five areas, having a maximum of 66,233 reported crimes. This suggests that Central may have higher risk factors or challenges contributing to its elevated crime rates compared to other regions in Los Angeles.
2. Hollywood, while known for its tourism and nightlife, ranks fifth: in this list with 50,581 crimes, serving as the area with minimum crimes committed, which could imply that crime is not isolated to the busiest or most popular areas. Crime is significant across various zones, each with different social and economic characteristics.

Recommendations and Solutions:

1. Increase Police Presence and Surveillance in High-Crime Areas:

* Implement more foot and car patrols in Central, 77th Street, and Pacific areas, as these zones have the highest crime rates.
* Install more surveillance cameras in critical locations to deter potential criminals and facilitate quicker incident responses.

2. Community Engagement Programs:

* Create initiatives that involve local residents in crime prevention, such as neighbourhood watch programs and workshops on personal safety and security measures.
* Establish partnerships with local organizations to address social issues that may contribute to crime, like unemployment and substance abuse.

3.Enhance Youth and Recreational Programs:

* Implement youth programs in these neighborhoods to engage young people in positive activities, reducing the likelihood of criminal involvement.
* Invest in community centers and sports facilities that provide safe spaces for recreation and mentorship programs so to keep youth occupied.