Young Yu

Sharikh Irfaan Mohammed

ADTA 5250.501

December 9, 2023

Executive Summary Report

## Introduction:

Commissioned by Dallas City officials, UNT Data & Analytics conducted a study on crime trends in Dallas, focusing on identifying crucial areas for consideration in the upcoming 2024 Annual Operating Plan (AOP) budget planning meetings. The primary objective of this study was to provide data-driven insights to assist officials in making informed decisions about funding allocations towards programs and education targeting specific demographic groups. The study analyzed various aspects of crime, including trends over time, demographic profiles of arrestees, and patterns related to the time and severity of offenses. The findings are expected to play a pivotal role in guiding the allocation of resources, aiming to reduce crime rates effectively and efficiently. The insights gained from this study are instrumental for Dallas City Officials in identifying the most impactful areas for funding, thereby facilitating targeted and strategic efforts to improve public safety and community well-being.

## Scope of Work:

UNT Data & Analytics services are an end-to-end solution. As agreed, UNT Data & Analytics will provide data services, time frame analysis, demographic analysis, a crime severity assessment, and reports.

**Details of these tasks:**

* Data Services: Collect and analyze data on various types of crimes and arrest demographics. Clean data to increase the accuracy while maintaining the integrity of the information.
* Time Frame Analysis: Examine crime trends over relevant years leading up to 2024.
* Demographic Analysis: Investigate the demographic profiles of arrestees, focusing on age, race, ethnicity, and sex.
* Severity Assessment: Analyze the severity of crimes and its distribution across different demographic groups.

**Our deliverables include:**

* Comprehensive Report: Detailed report of findings, including data analysis, graphs, and insights.
* Executive Summary: A summary highlighting key findings and recommendations.
* Presentation: A formal presentation of the study's findings for the 2024 AOP meeting.

## Schedule:

The study will commence on August 27, 2023, and conclude with the submission of the final report by December 11, 2023.

## Quality Assurance:

UNT Data & Analytics is committed to ensure data accuracy, confidentiality, and compliance with relevant laws and ethical guidelines. Though Texas does not have a law addressing data protection, like California Consumer Privacy Act (CCPA), due diligence was exercised in ensuring no Personal Identifiable Information (PII) was exposed. Public datasets were used for this project. Links to the datasets can be found in the Appendix of this document.

## Data:

UNT Data & Analytics core value is our commitment to the highest standards of data accuracy through audits, validation, and regular data cleaning processes. The use reliable sources and methods ensures the integrity and precision of the results. We uphold our dedication to quality and trustworthiness in all our data-driven endeavors through industry best practices and transparency.

**About the Dataset:**

Two datasets were inner joined to create the main dataset used for this project. Police\_Arrest\_Charges.csv (136.2k rows) contained the demographic information and Police\_Arrest.csv (103.2k Rows) held details about the crimes committed. There was a difference of ~33k rows between the two sets. After the inner join, 98.8k rows had the complete information to do the analysis.

* 27% of the data in Police\_Arrest\_Charges.csv did not match with the join.
* 96% of the information in Police\_Arrest.csv matched and was used for the analysis.

We analyzed the unmatched data for patterns that could explain the unusable data and concluded that there were a few contributing factors. Many of the inconsistencies and errors were due to data entry by the arresting officer(s) or the reviewing supervisor. We recommend providing comprehensive training for staff on data entry procedures and establish clear guidelines to minimize errors. Another factor is changes in law affect the classification and severity of crimes. An example of this is the decriminalization of marijuana. While the recreational use of the drug is still illegal, local law enforcement may choose not to enforce this law. The amount in possession also has different degrees of severity, which has also changed over the last decade. Lastly, court decisions on these matters can be delayed, so this would explain the gap in matched records.

**Cleaning Procedures:**

Tableau was identified as our customers preferred tool for data preparation and visualizations. The raw files in \*.csv, the Tableau Prep Builder \*.itl will all be provided as part of the deliverables. Our procedures are documented more in-depth in the \*.itl file.

High Level Summary of Data Cleaning:

Remove Non-Essential Categories

* 92 columns were removed the dataset.

Verify and Correct Data Types and Format

* ArrestYr was, by default, imported as an integer. It was converted to a string to allow grouping of incidents by year.

Identify Null values and select a fitting imputation method

* There were several null values in the Sex category. We attributed these nulls to the culture change in how we identify. The structurally missing data is not available in the system officers use for data entry. The choices are limited to Male, Female, and “ “. We chose to create a new category “Unknown” and imputed the nulls with the new value.

Correct Data Entry Errors

* Several typos that created new bins were corrected. For example, “Dals”, “DAL”, and “Dalas” were all corrected to “Dallas”.

Simplify

* Several of the crimes had multiple categories due to severity. For example, Criminal Mischief had twelve iterations. Each had a financial value range that differentiated it from the next iteration of the crime. A more holistic approach was taken by grouping like crime types into common bins.

**Tools Used**

* Tableau Data Prep
* Tableau Desktop
* Excel
* Anaconda

## Overview

Dallas's budget for the 2023-24 fiscal year has increased to $4.63 billion, up from last year's $4.51 billion. The General Fund now has $1.8 billion, which is $137.8 million more than before. An initiative led by Dallas City leaders to curb criminal activity by 20% has sparked the question, where does the funding need to be distributed to make the most impact. In partnership with UNT Data & Analytics, an analytical approach was taken to answer these questions.

1. What are the most common crimes in Dallas?
2. Which demographic group is most at risk?
3. Are there specific times of the day, days of the week, or months of the year when crimes are more likely to occur?

Public intoxication has the highest number of arrests in Dallas. The tables are a breakdown of the probabilities for race, age, and sex for public intoxication. For example, we took a random sample of 100 arrests from all types of arrests, we would expect 9 of the observations to be black males being arrested for public intoxication. 3 of the 8 would be from the 30-39 age group.

A screenshot of a computer screen

Description automatically generated

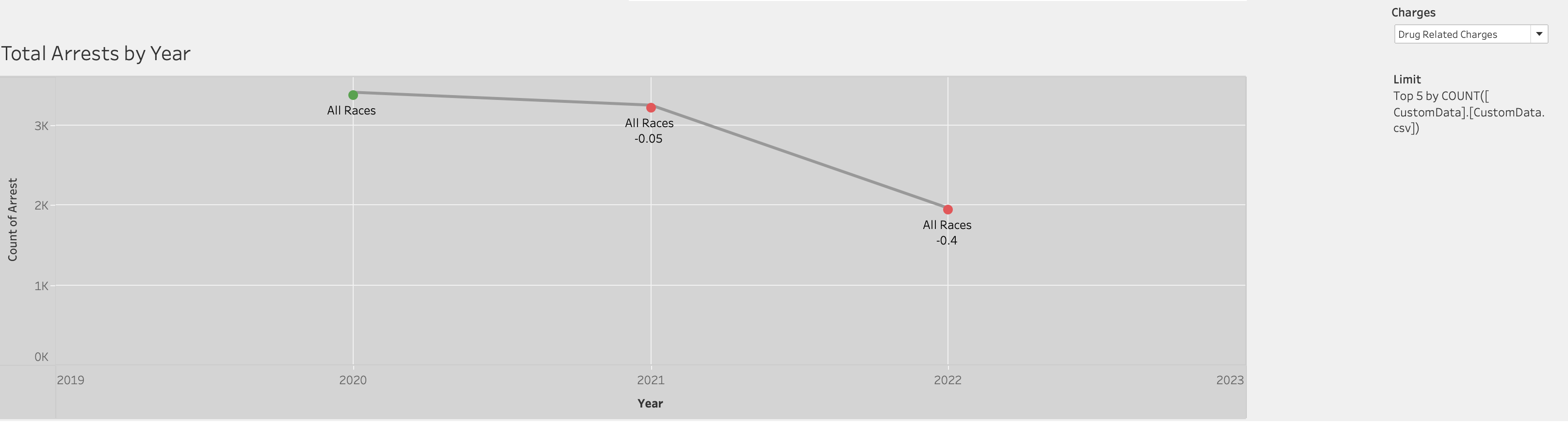
*Figure 1: Top 5 crimes 2014-2023*

|  |  |  |
| --- | --- | --- |
| Race | Count | Probability |
| Black | 12,862 | 36.62% |
| Hispanic | 11,304 | 32.19% |
| White | 10,226 | 29.12% |
| Asian | 313 | 0.89% |
| Middle Eastern | 217 | 0.62% |
| American Indian | 136 | 0.39% |
| Native Hawaiian | 48 | 0.14% |
| Unknown | 14 | 0.04% |

|  |  |  |
| --- | --- | --- |
| Age | Count | Probability |
| 18-29 | 10,412 | 29.65% |
| 30-39 | 10,968 | 30.47% |
| 40-49 | 6,944 | 19.78% |
| 50-59 | 5,265 | 14.99% |
| 60-69 | 1,624 | 4.63% |
| 70-79 | 144 | 0.41% |
| 80+ | 25 | 0.07% |

|  |  |  |
| --- | --- | --- |
| Sex | Count | Probability |
| Male | 29,021 | 82.63% |
| Female | 6,094 | 17.35% |
| Unknown | 5 | 0.01% |

Drug related arrests came in second. As mentioned earlier, changes in laws can impact arrests categories and severities. Using the decriminalization of marijuana example from earlier, 37% of all drug-related arrests in this dataset were for marijuana of various amounts. In 2020, there were 3.4k arrests followed by a downtrend of 5% in 2021, and then a significant drop of 47% in 2022. We excluded 2023 from this analysis since the focus was by year and the data is incomplete.



*Figure 2: Drug-related Arrests 2020-2022*

|  |  |  |
| --- | --- | --- |
| Race | Count | Probability |
| Black | 13,455 | 52.95% |
| Hispanic | 7,771 | 30.60% |
| White | 3,965 | 15.62% |
| Asian | 98 | 0.39% |
| Middle Eastern | 60 | 0.24% |
| American Indian | 21 | 0.08% |
| Native Hawaiian | 16 | 0.06% |
| Unknown | 16 | 0.06% |

|  |  |  |
| --- | --- | --- |
| Age | Count | Probability |
| 18-29 | 10,365 | 40.82% |
| 30-39 | 8,251 | 32.49% |
| 40-49 | 3,882 | 15.29% |
| 50-59 | 2,065 | 8.13% |
| 60-69 | 688 | 2.71% |
| 70-79 | 82 | 0.32% |
| 80+ | 2 | 0.01% |
| Sex | Count | Probability |
| Male | 21,662 | 85.31% |
| Female | 3,641 | 14.34% |
| Unknown | 32 | 0.13% |

Lastly, driving while intoxicated. Using a heat map, we can visually see a clear pattern of arrests and this pattern was consistent throughout all the various demographic groups. Saturday and Sunday had the highest probability, so we combined the two into a “Weekend” category and look at the various demographic groups.

A screenshot of a computer screen

Description automatically generated

*Figure 3: Distribution of DWI arrests by day of the week*

|  |  |  |
| --- | --- | --- |
| Day of the Week | Count | Probability |
| Monday | 1,685 | 13.10% |
| Tuesday | 1,065 | 8.20% |
| Wednesday | 1,027 | 7.99% |
| Thursday | 1,360 | 10.60% |
| Friday | 2,020 | 15.70% |
| Saturday | 2,839 | 22.00% |
| Sunday | 2,853 | 22.20% |

|  |  |  |
| --- | --- | --- |
| Race | Count | Probability |
| Hispanic | 3,212 | 56.40% |
| Black | 1,243 | 21.83% |
| White | 1,113 | 19.54% |
| Asian | 79 | 1.39% |
| Middle Eastern | 19 | 0.33% |
| American Indian | 16 | 0.28% |
| Native Hawaiian | 9 | 0.16% |
| Unknown | 4 | 0.07% |

|  |  |  |
| --- | --- | --- |
| Age | Count | Probability |
| 18-29 | 2,231 | 39.26% |
| 30-39 | 1,695 | 29.83% |
| 40-49 | 976 | 17.17% |
| 50-59 | 548 | 9.64% |
| 60-69 | 200 | 3.52% |
| 70-79 | 31 | 0.55% |
| 80+ | 2 | 0.04% |

|  |  |  |
| --- | --- | --- |
| Sex | Count | Probability |
| Male | 4,609 | 80.93% |
| Female | 1,082 | 19.00% |
| Unknown | 4 | 0.07% |

## Recommendations:

**Public Intoxication**

With public intoxication leading to the highest number of arrests, and given the racial and age distribution, targeted intervention programs for at-risk demographics could be beneficial. This could include alcohol education programs, substance abuse counseling, and community support initiatives.

**Drugs**

The significant decrease in drug-related arrests, particularly for marijuana, following decriminalization policies, suggests that continued evaluation and adjustment of drug laws could further impact arrest categories and severity. Lawmakers and law enforcement agencies may consider focusing resources on more severe drug offenses and exploring alternative measures for lower-level offenses, such as education programs and community service.

**Driving While Intoxicated**

Since DWI arrests are most prevalent on weekends, there could be an increased emphasis on DWI prevention during these times. Strategies might include public awareness campaigns about the dangers of drunk driving, increased patrols and checkpoints, and ride-sharing incentives during peak times for DWI offenses. Hispanic males, ages 18-39, are disproportionally at a higher risk of being arrested for DWI on the weekend at 7.92%.

**Closing Thoughts**

In summarizing our findings, it's important to recognize that the analysis conducted provides only an introductory perspective on the complexities of crime dynamics within Dallas. The dataset utilized, though insightful, is not exhaustive and our examination has been limited by time constraints.

A comprehensive understanding necessitates a richer array of data and a broader scope of inquiry. Significant questions remain, particularly regarding the effectiveness of social programs in crime mitigation. Further investigation is required to evaluate the correlation between social interventions and their potential to decrease crime rates.

The raised represent initial steps towards a more profound analysis. It has the potential to inform policy development and promote community well-being. The insights gained could be pivotal in developing strategies that foster safer environments. This endeavor, though complex, is an essential component of societal advancement.

## Data Sources:

Police\_Arrests\_Charges.csv

<https://www.dallasopendata.com/Public-Safety/Police-Arrest-Charges/9u3q-af6p>

Police\_Arrests.csv

<https://www.dallasopendata.com/Public-Safety/Police-Arrests/sdr7-6v3j>

aareststatscnty23.xlsx

<https://www.dps.texas.gov/sites/default/files/documents/administration/crime_records/pages/aareststatscnty23.xlsx>

aareststatscnty22.xls

<https://www.dps.texas.gov/sites/default/files/documents/administration/crime_records/pages/aareststatscnty22.xls>

arreststatscnty21.xlsx

<https://www.dps.texas.gov/sites/default/files/documents/administration/crime_records/pages/arreststatscnty21.xlsx>

FY2023-Proposed-Budget.pdf

<https://www.dallascounty.org/Assets/uploads/docs/budget/fy2023/FY2023-Proposed-Budget.pdf>.

**Individual Contribution:**

**Appendix:**

|  |  |
| --- | --- |
| **Team Member Name** | **Participation in doing Report** |
| Young Yu, SharikhIrfaan | Together discussed and wrote the Prep for the Official report jointly. |
| Young Yu, SharikhIrfaan | Together provided the business insights and unanticipated findings. |
| Young Yu, SharikhIrfaan | Together provided business insights as well as unexpected findings for 2nd chart. |
| Young Yu, SharikhIrfaan | Together provided business insights as well as unexpected findings for 3rd chart. |
| Young Yu, SharikhIrfaan | Finally reviewed the full document and formatting before uploading. |