LOS ANGELES CRIME DATA ANALYSIS (2020-2023)

CIS 5270 BUSINESS INTELLIGENCE PROJECT

SPRING 2023

Author:

Dhwani Vaishnav
Pratiksha Yadav
Manimozhi Neethinayagam

Table of Contents

A. Introduction:

The Los Angeles Police Department (LAPD) provides a publicly available dataset that covers crime incidents in the City of Los Angeles from 2020. The department was formed in 1869 and by 2020, it had over 12,000 officers with a motto to protect and serve. *[1]*

Crime – it is an act committed in violation of law prohibiting or omitted in violation of ordering it *[2]*. Crime data can be used to evaluate the effectiveness of crime prevention programs and policies.

In this analysis, we aim to explore the dataset and gain insights into crime patterns in Los Angeles. The goal of this study is to make a clear distinction between which type of crime results in the highest victimization rate and how this relates to ethnicity, age, and gender. Another intention of this analysis is to understand the crime category which is more prevalent in the area.

The spatial analysis of crime data is important to understand the category and pattern of crime. The LAPD crime dataset covers the period from 2020 to 2023 (Jan, Feb 2023) and provides insights on various categories of crimes committed in different locations throughout the city of LA.

Using the given data, we will begin by analyzing the spatial distribution of crimes and determine the areas with the highest and lowest crime rates. We will examine the overall trend of crime incidents in the city and identify any significant changes in the frequency of different types of crimes over time. Additionally, we will explore the relationship between crime and various demographic factors, such as age, gender, and race.

According to the article “Types of crime analysis” *[3]*, tactical analysis involves analyzing the crime trend and pattern. The goal is to support short term operational strategies to address the

trend, pattern, high risk- area, hot –spots. The tactical method includes mapping and spatial analysis. Strategic Analysis involves analyzing the trend pattern for the long- term and identifying unusual trends in the increase or decrease of crime that is observed and recorded.

By using data visualizations such as charts, graphs, and maps, complex information is presented in a clear and concise manner, making it easier to identify patterns and trends. The crime analysis using an interactive dashboard provides precise understanding of crime patterns and trends. To make our analysis more engaging and impactful, we incorporated a storytelling section in our report.

By adopting this approach, we hope to enhance the delivery of our findings and increase their effectiveness in shaping future policies and initiatives related to crime in LA. Additionally, this approach will also help citizens better understand what is happening in their city and identify hot spots of crime and safety issues. Especially, as Cal state LA students, it is important to be aware of the crime trends in the surrounding area to ensure personal safety and take necessary precautions. Knowledge of crime patterns can help students make informed decisions about where to live, travel, and socialize. Furthermore, being aware of crime trends can help students recognize potential dangers and avoid becoming victims of crime. By staying informed about the crime statistics and trends in the area, Cal state LA students can take steps to stay safe and contribute to a safer community overall.

By comprehensively analyzing this dataset, our aim is to enhance the understanding of crime in Los Angeles and inform policy decisions to reduce crime rates and improve public safety.

B. Data Description:

URL: https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8

The dataset presents LAPD crime data spanning from 2020 to 2023, comprising approximately 686,000 records and 22 columns. This data is transcribed from original crime reports that are typed on paper and therefore there may be some inaccuracies within the data. Each row represents a single crime incident. Missing location data is indicated as (0°, 0°), and to protect privacy, address information is only provided up to the nearest hundred block.

The columns Date & Time occurred gives us details on date and time of the crime occurred, using which we can find out which day and which time of the day was most sensitive.

The column labeled "status" provides information on the status of the crime, such as AA, AO, CC, IC, JO, and JA. The JO and JA statuses are updates related to juvenile crimes, and the presence of these statuses in the dataset enables us to differentiate between crimes committed by adults and those committed by juveniles.

The "location", “latitude” and “longitude” columns detail regarding the location of the crime.

Additionally, to properly represent the areas of crime, we have categorized the Bureaus as per *[4]* LAPD comprises four significant bureaus - Central, Valley, South, and West Bureau - and a total of around 21 divisions operating under these bureaus. These regions have been divided into four bureaus using the "create group" function in Tableau.

Area Name	Bureau
Central	Central
	Hollenbeck
	Newton
	Northeast
	Rampart
Valley	Devonshire
	Foothill
	Mission
	North
	Topanga
	Van Nuys
	West Valley
South	77th Street
	Harbor
	Southwest
	Southeast
West	Hollywood
	Pacific
	West LA
	Wilshire
	Olympic

Columns Description:

Column Name	Description
Date Rptd	mm/dd/yyyy
DATE OCC	mm/dd/yyyy
TIME OCC	In 24-hour, military time.
AREA	The LAPD has 21 Community Police Stations referred to as Geographic Areas within the department. These Geographic Areas are sequentially numbered from 1-21.
AREA NAME	The 21 Geographic Areas or Patrol Divisions are also given a name designation that references a landmark or the surrounding community that it is responsible for. For example, 77th Street Division is located at the intersection of South Broadway and 77th Street, serving neighborhoods in South Los Angeles.
Rpt Dist No	A four-digit code that represents a sub-area within a Geographic Area. All crime records reference the "RD" that it occurred in for statistical comparisons. Find LAPD Reporting Districts on the LA City GeoHub at
Crm Cd	Indicates the crime committed. (Same as Crime Code 1)
Crm Cd Desc	Defines the Crime Code provided.
Vict Age	Two character numeric
Vict Sex	F - Female M - Male X - Unknown

Vict Descent	Descent Code: A - Other Asian B - Black C - Chinese D - Cambodian F - Filipino G - Guamanian H - Hispanic/Latin/Mexican I - American Indian/Alaskan Native J - Japanese K - Korean L - Laotian O - Other P - Pacific Islander S - Samoan U - Hawaiian V - Vietnamese W - White X - Unknown Z - Asian Indian
Premis Cd	The type of structure, vehicle, or location where the crime took place.
Premis Desc	Defines the Premise Code provided.
Weapon Used Cd	The type of weapon used in the crime.
Weapon Desc	Defines the Weapon Used Code provided.
Status	Status of the case. (IC is the default)
Status Desc	Defines the Status Code provided.
Crm Cd 2	May contain a code for an additional crime, less serious than Crime Code 1.
LOCATION	Street address of crime incident rounded to the nearest hundred block to maintain anonymity.
Cross Street	Cross Street of rounded Address
LAT	Latitude
LON	Longitude

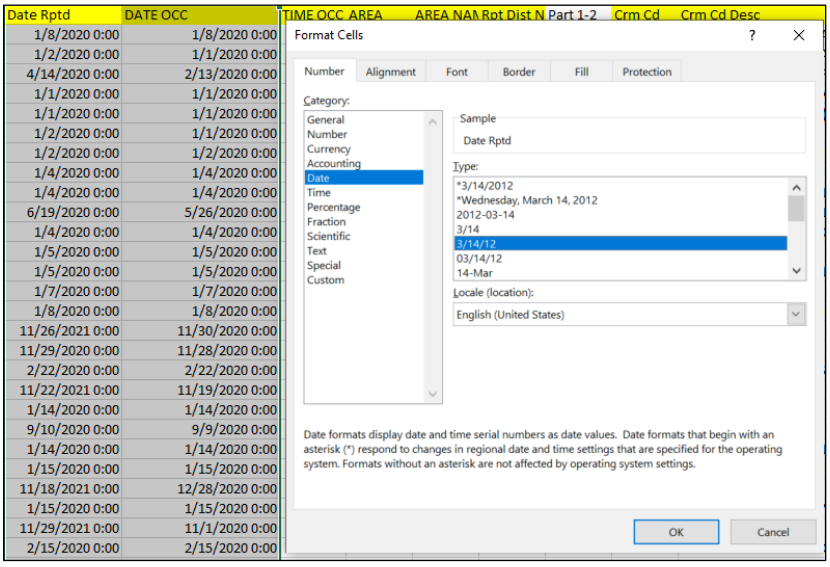
C. Data Cleaning

The data which we get from the official sites is usually raw data. This data needs to be cleaned to make it easier to analyze data. We need to remove duplicates, combine columns, change the incorrect date format etc. Various data cleaning methods can be used for Data cleaning, these may vary from Dataset to the dataset. Listed below are some of the Data Clearing methods which have been used to clean the LA Crime Dataset. The Data has been cleaned using Microsoft Excel.

1. Embedded values in the field - Date Format:

For the columns “Date Rptd” & “Date OCC”, there are two different values written – date & “0:00”. We have removed the embedded time “0:00” portion so that the entire field has only single type of values. By doing this, we can simplify the data and make it easier for analysis using the date format. Furthermore, leaving the time portion in the date format can cause issues when sorting or grouping data by date.

Pre-Cleaning screenshot:



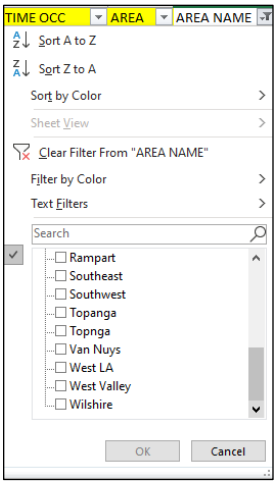
Post-Cleaning Screenshot:

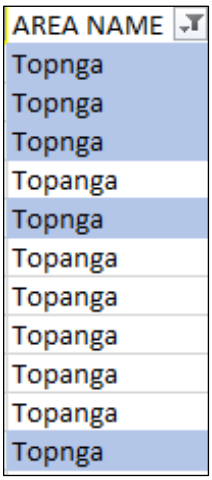
Date Rptd	DATE OCC
1/8/20	1/8/20
1/2/20	1/1/20
4/14/20	2/13/20
1/1/20	1/1/20
1/1/20	1/1/20
1/2/20	1/1/20
1/2/20	1/2/20
1/4/20	1/4/20
1/4/20	1/4/20
6/19/20	5/26/20
1/4/20	1/4/20
1/5/20	1/5/20
1/5/20	1/5/20
1/7/20	1/7/20

2. Misspellings:

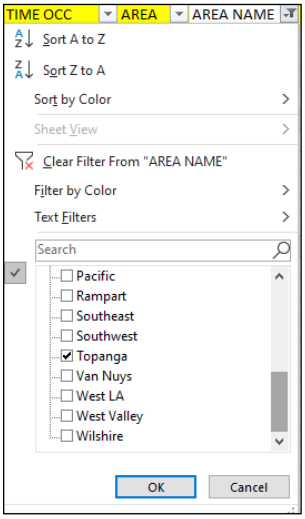
It is very important to fix the Misspellings in a data set, as this may cause ambiguity. As shown in the screenshot below, the column "Area Name" has a value – "**Topnga"** is spelled Incorrectly. Hence, we replaced the incorrect value "**Topnga"** with the Correct Value **"Topanga",** for the whole Column- 'Area Name'. After correcting the Misspelled value, we can see only one filter with the 'Area Name'- **Topanga** in the post-cleaning screenshot.

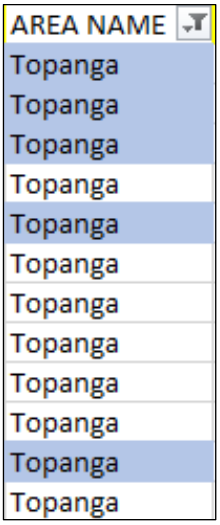
Pre-Cleaning Screenshots:





Post-Cleaning Screenshots:





3. Duplicated Records:

Duplicate records cause confusion while analyzing the data as well as visualizing it leading to unnecessary wastage of memory. Therefore, removing duplicate records is essential before conducting any data analysis to ensure accurate and reliable results.

Here, we have 2 Fields “Crm Cd” & “Crm Cd1”, both contain the same data. Hence, we have removed the column “Crm Cd1” and kept the other column to avoid the duplicate records.

Pre-cleaning Screenshot:

Crm Cd	Crm Cd 1	Crm Cd 2
624	624	
624	624	
845	845	
745	745	998
740	740	
121	121	998
442	442	998
946	946	998
341	341	998
341	341	
330	330	
930	930	
341	341	
648	648	998
442	442	
626	626	
626	626	
440	440	624

Post-cleaning Screenshot:

Crm Cd 1	Crm Cd 2
624	
624	
845	
745	998
740	
121	998
442	998
946	998
341	998
341	
330	
930	
341	
648	998
442	
626	
626	
440	624

4. Combining Columns:

Combining columns helps to standardize the data with similar data types or formats and make it more meaningful.

Here, we have two fields “LOCATION” & “Cross Street”, which can be merged and give us the precise location. So have merged these two columns by using the CONCAT formulas in excel. We have also used “TRIM” formulas to remove unwanted blank spaces from the “LOCATION” column.

Pre-cleaning Screenshots:

LOCATION	Cross Street
1100 W 39TH	PL
700 S HILL	ST
200 E 6TH	ST
5400 CORTEEN	PL
14400 TITUS	ST
700 S BROADWAY	
700 S FIGUEROA	ST
200 E 6TH	ST
700 BERNARD	ST
11900 BALBOA	BL
15TH	OLIVE
800 N ALAMEDA	ST
800 S OLIVE	ST
700 W 7TH	ST
100 S LOS ANGELES	ST
14200 BERG	ST

LOCATION	Cross Street	=U1&" "&V1
1100 W 39TH	PL	
700 S HILL	ST	
200 E 6TH	ST	
5400 CORTEEN	PL	
14400 TITUS	ST	
700 S BROADWAY		
700 S FIGUEROA	ST	
200 E 6TH	ST	
700 BERNARD	ST	
11900 BALBOA	BL	
15TH	OLIVE	
800 N ALAMEDA	ST	
800 S OLIVE	ST	
700 W 7TH	ST	
100 S LOS ANGELES	ST	
14200 BERG	ST	
3200 W AVENUE 32		
PACIFIC COAST	VERMONT	

LOCATION	Cross Street	LOCATION	Cross Street
1100 W 39TH	PL	1100 W 39TH	PL
700 S HILL	ST	700 S HILL	ST
200 E 6TH	ST	200 E 6TH	ST
5400 CORTEEN	PL	5400 CORTEEN	PL
14400 TITUS	ST	14400 TITUS	ST
700 S BROADWAY		700 S BROADWAY	
700 S FIGUEROA	ST	700 S FIGUEROA	ST
200 E 6TH	ST	200 E 6TH	ST
700 BERNARD	ST	700 BERNARD	ST
11900 BALBOA	BL	11900 BALBOA	BL
15TH	OLIVE	15TH OLIVE	
800 N ALAMEDA	ST	800 N ALAMEDA	ST
800 S OLIVE	ST	800 S OLIVE	ST
700 W 7TH	ST	700 W 7TH	ST
100 S LOS ANGELES	ST	100 S LOS ANGELES	ST
14200 BERG	ST	14200 BERG	ST
3200 W AVENUE 32		3200 W AVENUE 32	
PACIFIC COAST	VERMONT	PACIFIC COAST VERMONT	
14700 FRIAR	ST	14700 FRIAR	ST
7TH	HILL	7TH HILL	
13600 LEADWELL	ST	13600 LEADWELL	ST

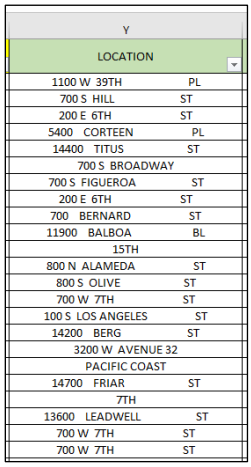
Post-Cleaning Screenshot:

Location	
1100 W 39TH	PL
700 S HILL	ST
200 E 6TH	ST
5400 CORTEEN	PL
14400 TITUS	ST
700 S BROADWAY	
700 S FIGUEROA	ST
200 E 6TH	ST
700 BERNARD	ST
11900 BALBOA	BL
15TH OLIVE	
800 N ALAMEDA	ST
800 S OLIVE	ST
700 W 7TH	ST
100 S LOS ANGELES	ST
14200 BERG	ST
3200 W AVENUE 32	
PACIFIC COAST VERMONT	
14700 FRIAR	ST
7TH HILL	
13600 LEADWELL	ST

5. Trimming extra “blanks” from the string:

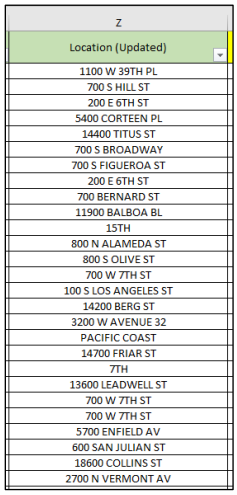
Trimming extra blanks from a string is important because it can help to improve the readability and usability of the data. Extra blanks can be added accidentally during data entry, data manipulation or file formatting, and can make it difficult to analyze and process the data accurately. Trimming extra blanks from strings can help to ensure that the data is accurate, consistent, and easily processed.

Pre-cleaning Screenshot:



Y
LOCATION
1100 W 39TH PL
700 S HILL ST
200 E 6TH ST
5400 CORTEEN PL
14400 TITUS ST
700 S BROADWAY
700 S FIGUEROA ST
200 E 6TH ST
700 BERNARD ST
11900 BALBOA BL
15TH
800 N ALAMEDA ST
800 S OLIVE ST
700 W 7TH ST
100 S LOS ANGELES ST
14200 BERG ST
3200 W AVENUE 32
PACIFIC COAST
14700 FRIAR ST
7TH
13600 LEADWELL ST
700 W 7TH ST
700 W 7TH ST

Post-cleaning Screenshot:



Z
Location (Updated)
1100 W 39TH PL
700 S HILL ST
200 E 6TH ST
5400 CORTEEN PL
14400 TITUS ST
700 S BROADWAY
700 S FIGUEROA ST
200 E 6TH ST
700 BERNARD ST
11900 BALBOA BL
15TH
800 N ALAMEDA ST
800 S OLIVE ST
700 W 7TH ST
100 S LOS ANGELES ST
14200 BERG ST
3200 W AVENUE 32
PACIFIC COAST
14700 FRIAR ST
7TH
13600 LEADWELL ST
700 W 7TH ST
700 W 7TH ST
5700 ENFIELD AV
600 SAN JULIAN ST
18600 COLLINS ST
2700 N VERMONT AV

6. Standardizing the Time Format:

Time data becomes difficult to analyses if it is not in standard format. Time-related data can be complex and small errors or inconsistencies will affect the accuracy in analysis. Cleaning the data ensures it is in a more reliable and usable format. We have used Formulas to convert the numeric format to Time format.

We have a column “TIME OCC” which contains the value in military 24 Hr. time format. It is OK for the time like 22:30 (having all 4 digits). However, with values like “330” (03:30) and “30” (12:30), it is difficult to perform visualization. Hence, we have changed the format of the values to the 24 hrs. format.

Pre-Cleaning Screenshot:

TIME OCC
2230
330
1200
1730
415
30
1315
40
200
1925
2200
955

Post-Cleaning Screenshots:

We have used two Formulas, first (Time (left (c2, LEN(c2)-2), RIGHT(c2,2),0)) to convert the 4- or 3-digit integers to Time format. Second (A7/1440) to convert the 1 or 2-digit integers into time format.

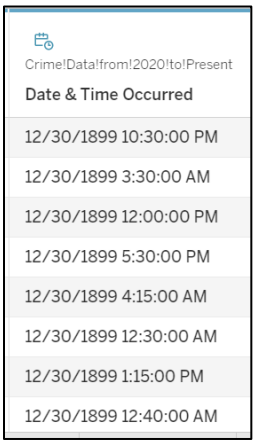
Old Time	Time Occurred
2230	10:30 PM
330	3:30 AM
1200	12:00 PM
1730	5:30 PM
415	4:15 AM
30	#VALUE!
1315	1:15 PM

30	0.020833333
40	0.027777778
1	0.000694444
55	0.038194444
30	0.020833333
40	0.027777778
15	0.010416667
1	0.000694444
1	0.000694444

30	12:30 AM
40	12:40 AM
1	12:01 AM
55	12:55 AM
30	12:30 AM
40	12:40 AM
15	12:15 AM
1	12:01 AM

1638	16:38
1805	18:05
730	7:30
2018	20:18
1900	19:00
1200	12:00
1330	13:30
1735	17:35
1730	17:30
1445	14:45
1	0:01

For visualization in the Tableau, these values are taken in “datetime” datatype as below:



To extract the default date “12/30/1899”, we have added some calculated fields in Tableau to split “date” & “time” and we have used “time” part in our visualization.

7. Rephrasing the Column Names:

The given Column Names in the dataset were more meaningful & understandable for LAPD employees. However, for us, the names were unclear & ambiguous, it was difficult to determine - what data represents. Hence, we have renamed the column names for better interpretation of data, we have made them more descriptive and informative, which can help to clarify the meaning and purpose of the data.

For example, we have renamed the existing columns:

“Crm Cd” to “Primary Crime Code”, “Status” to “Crime Status”, “Vict Age” to “Victim Age”, “LAT” to “Latitude”, “LON” to “Longitude”

Pre-Cleaning Screenshot:

H	O	R	AB	AC
Crm Cd	Status	Vict Age	LAT	LON
624	AO	36	34.0141	-118.2978
624	IC	25	34.0459	-118.2545
845	AA	0	34.0448	-118.2474

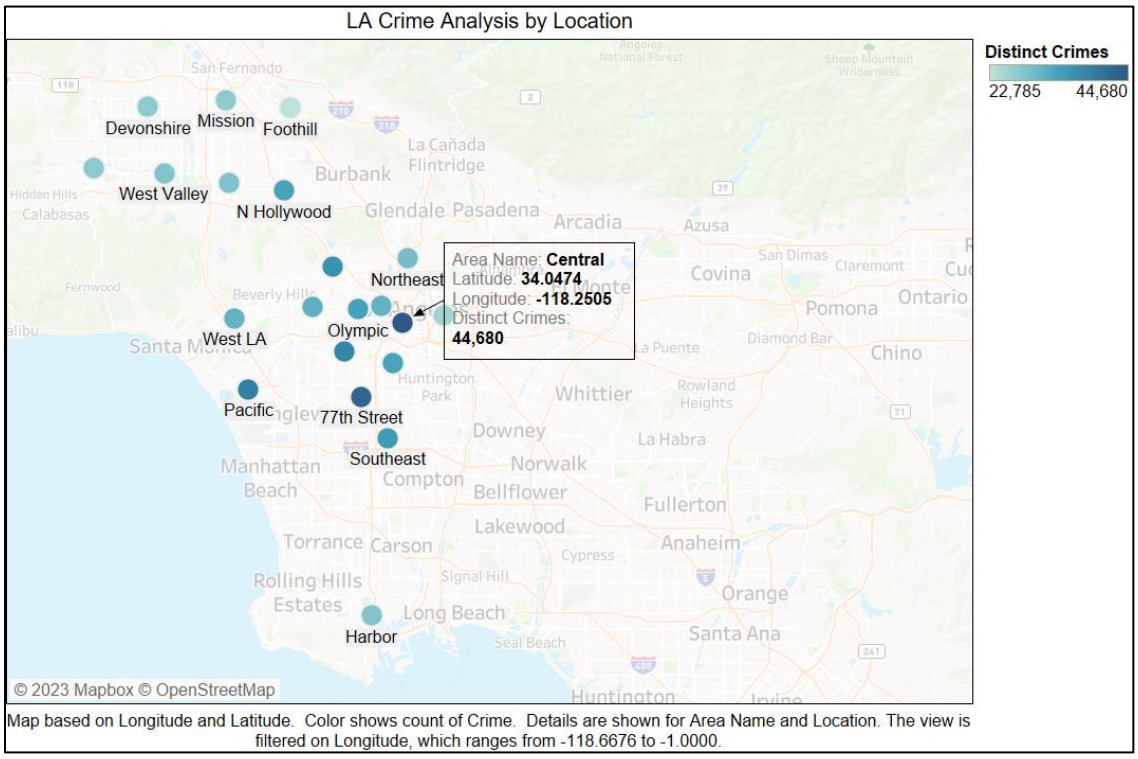
Post-Cleaning Screenshots:

I	P	R	AC	AD
Primary Crime Code	Crime Status Code	Victim Age	Latitude	Longitude
624	AO	36	34.0141	-118.2978
624	IC	25	34.0459	-118.2545
845	AA	0	34.0448	-118.2474

D. Data Visualizations

The information in the LA Crime Dataset has been examined to determine the number of crimes that occurred each year, categorized by bureau & Areas, types of crime, the reported status of the crimes & victim's demographic data. This analysis enables us to gain insight into the LAPD's effectiveness over time in establishing a secure environment for the community.

1. What parts of Los Angeles experience elevated levels of criminal activity? Display the distribution of crime throughout Los Angeles.



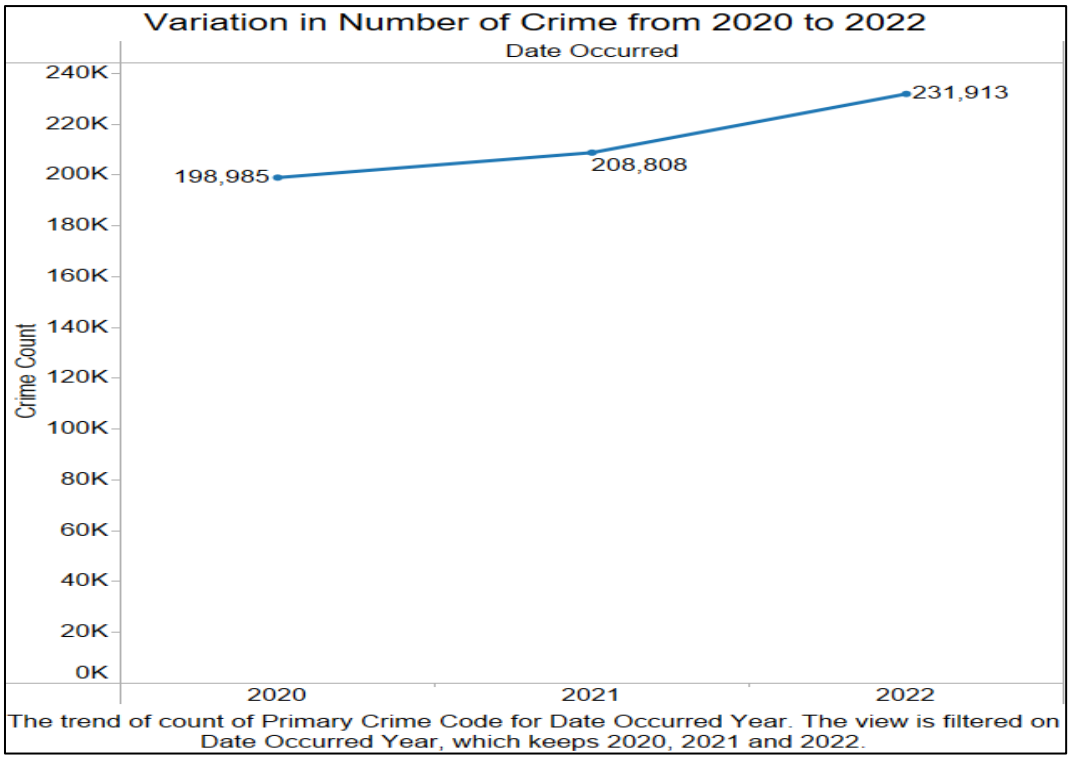
Categories used: Geographic Maps, Calculated Fields

Using the geographical coordinates provided in the dataset, we created a visualization in Tableau that displays the distribution of crime in the nearby vicinity. The process was made easier due to the availability of location data in terms of latitude & longitude.

The density of crime in a particular area is represented by the darkness of the markers, with clusters of closely located crime data points forming a group. **The existence of several groupings in the central region of LA implies a greater occurrence of unlawful acts in that vicinity.** As individuals attending Cal State LA, the knowledge we have acquired about the whereabouts of criminal activity is assisting us in pinpointing potentially more secure pathways throughout the metropolis.

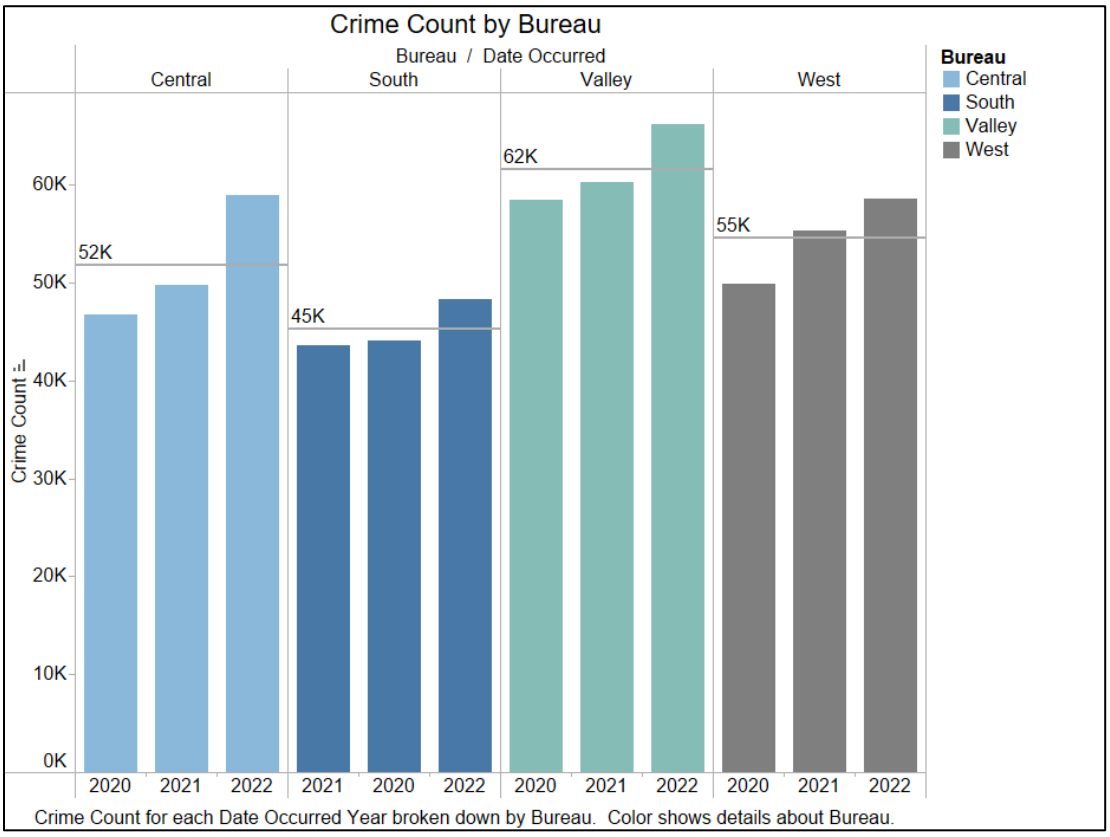
Moreover, the zones identified by the lighter density demonstrate a lower frequency of criminal incidents, such as the vicinity encompassing the Foothill area. This indicates that fewer unlawful activities occur in this specific location compared to other areas in the city.

2. What is the current state of the trend in criminal activity? Display the year-over-year rise or fall of crime for each bureau.



Categories used: Line graph, Dates

Through an examination of the quantity of criminal offenses committed throughout the years, we can assess whether the frequency of crime has risen or fallen. A line chart has been employed to illustrate the changes in crime rates visually, spanning from 2020 to 2022. The greatest quantity of criminal activities, nearly 199K, were recorded in 2020, with the number escalating over the following two years. By 2022, the number of reported crimes had risen to approximately 232K.

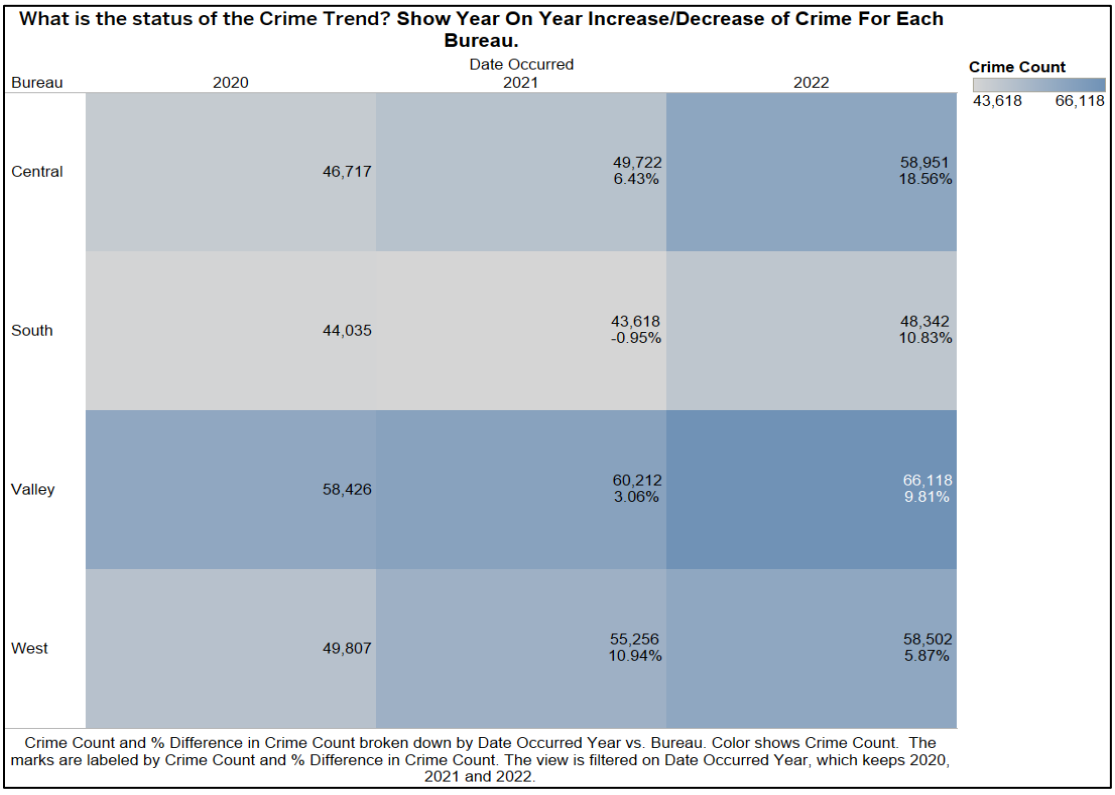


Categories used: Bar chart, Dates, Groups, Calculated Fields, Reference Lines

The chart offers an understanding of the variation in crime rates across various bureaus, highlighting those with the highest and lowest levels of criminal activity. To simplify comparisons, a reference line has been included on the graph for each year, representing the average total number

of crimes committed during that period. This line assists in identifying the bureaus that have more criminal incidents than the average.

The analysis highlights that the Valley Bureau had the highest crime rate in 2022, with an approximate average count of 62,000 for all three years. This indicates that criminal activity has been consistently high in this area over the years. On the other hand, the South Bureau had the lowest average number of crimes, suggesting that it is comparatively safer than other areas in Los Angeles.

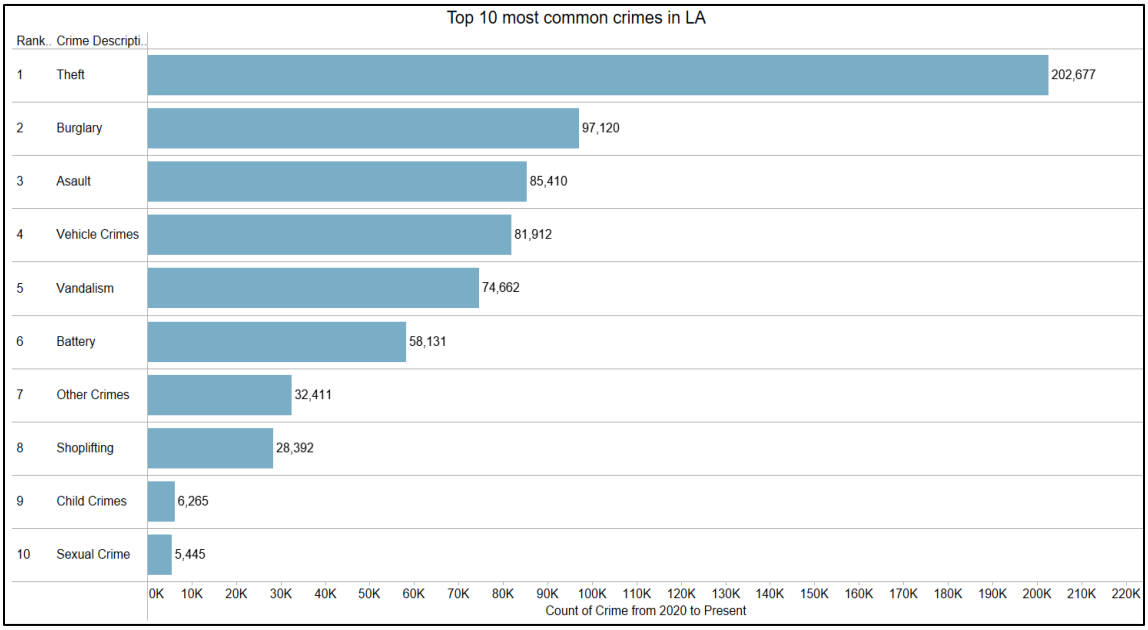


Categories used: Heatmap, Dates-Year, Groups

We have determined the exact fluctuation in crime rates over time for each bureau by calculating the percentage difference between consecutive year values. The percentage difference indicates the change in value in relation to the preceding year. The Heatmap employs various shades of color

to enable us to easily identify the bureau that had the most significant decrease or increase in crimes each year. For instance, in 2022, the Central bureau experienced a notable increase in crimes, approximately 19%, compared to the previous year, 2021. Conversely, in 2021, the West bureau had a noteworthy rise in crimes. In contrast, the South bureau observed a decline in criminal incidents, approximately 0.95%, in 2021 compared to 2020. Visual representations, like this one, simplify the analysis of data.

3. What are the ten most frequently occurring crimes in Los Angeles from 2020 to the present?

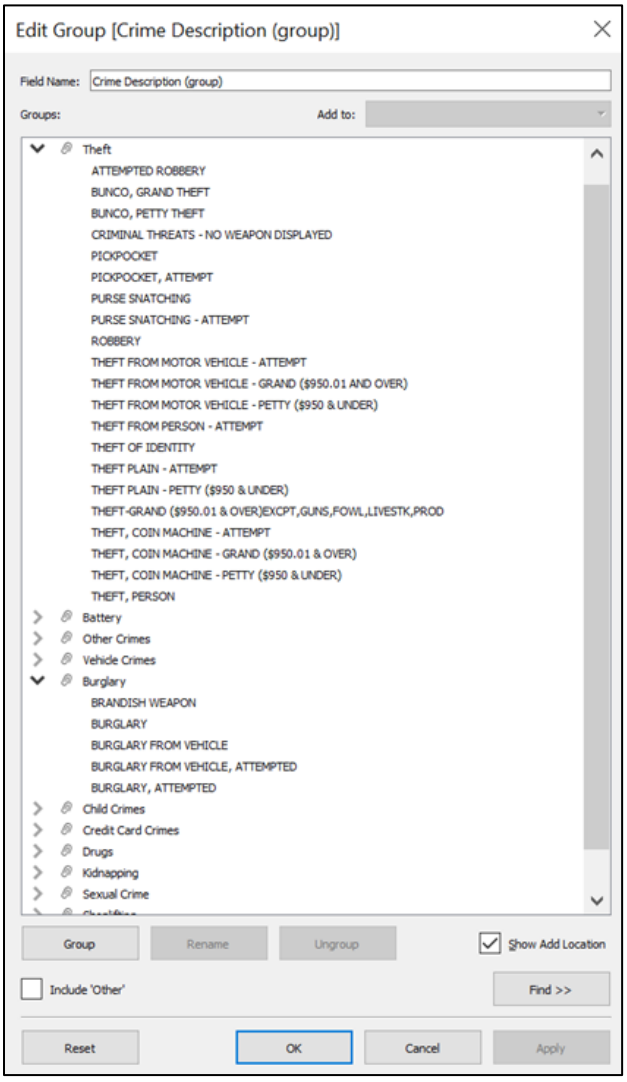


Categories Used: Filters, Groups, Rank, Count

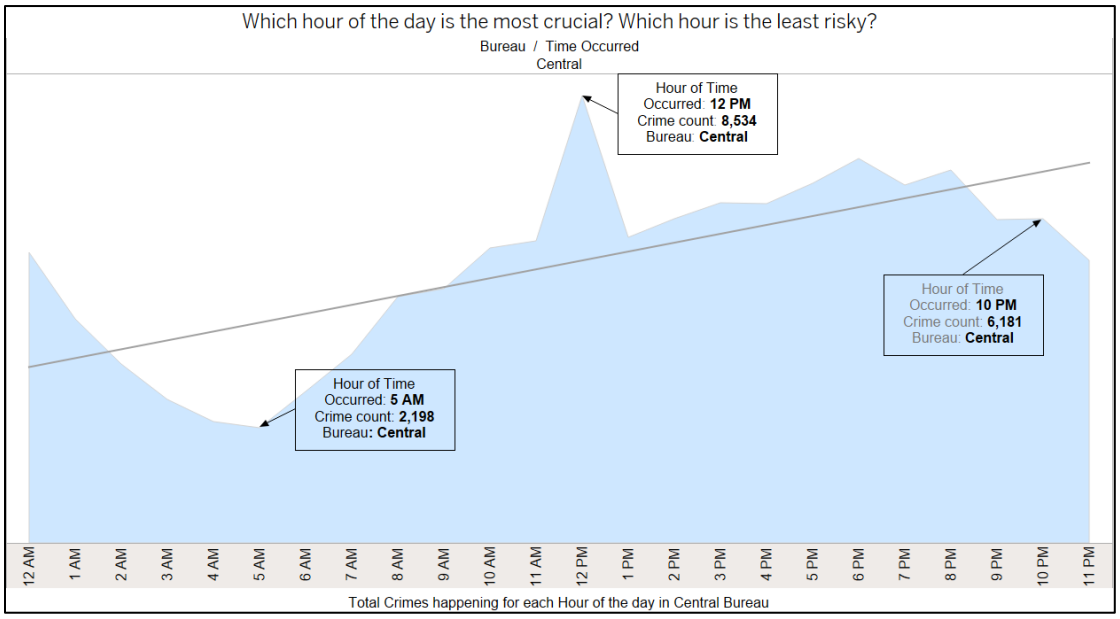
In the above graph, we have used a bar chart to visualize the Top 10 most common crimes in the city of Los Angeles in the period 2020 to the present. This analysis depicts the breakdown of crimes in a structured manner, allowing us to determine which crimes occur most frequently. By

filtering for the Top 10 crime types, we can rank them in order of occurrence from high to low. The results indicate that theft is the most prevalent crime, indicating that the residents of LA are at a higher risk of experiencing theft and should take caution in their surroundings. Additionally, this type of analysis can be useful for law enforcement agencies, LAPD, and the public in understanding the nature and extent of criminal activity.

Additionally, this analysis utilizes Tableau's grouping functionality to categorize similar types of crimes under one heading. The below screenshot shows how the groups are formed for the Crime Description.



4. In the Central Bureau (the region with the high crime rate) - which hour of the day is the most crucial? Which hour is the least risky?



Categories Used: Area chart, Dates – the Hour part, Groups, Forecast Trend Lines

The time of day is a significant factor in criminal activity. We have generated a chart above that displays the hours of the day in 12-hour format and highlights the most critical and least dangerous times of day, accompanied by the number of crimes that occurred during those hours.

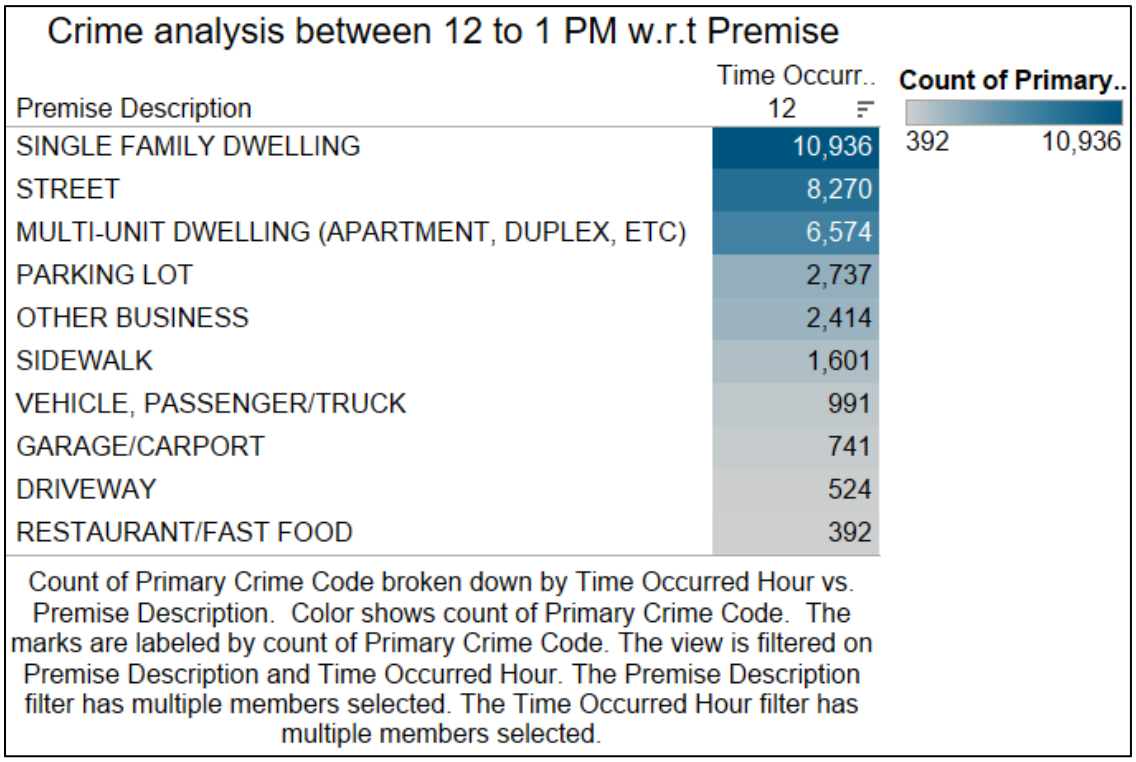
According to the analysis, there is a decrease in criminal activities at 5 am when most people are sleeping. This outcome is not surprising since it is commonly known that fewer people are out on the streets during the early morning hours, resulting in fewer opportunities for criminal activities.

On the other hand, we anticipated an increase in criminal incidents at 10 pm, a time when there are fewer people around to witness the crimes, and the darkness provides a cover for the perpetrators to conceal their identity and actions. This outcome is expected since nighttime hours

are usually considered more dangerous due to the lower visibility and less presence of people on the streets.

However, what caught the analysis off guard was the surge in criminal activities during midday, a time when the streets are more crowded, and there are more opportunities for potential witnesses to observe and report criminal incidents. The rise in criminal activity during midday was almost reaching its peak level, which was unexpected and could be attributed to several factors, such as increased foot traffic, more unlocked homes during this time, and other societal and economic factors that might have affected the criminal behavior patterns.

5. **Premises analysis for the crucial hour of the day (12 PM to 1 PM) based on previous analysis**

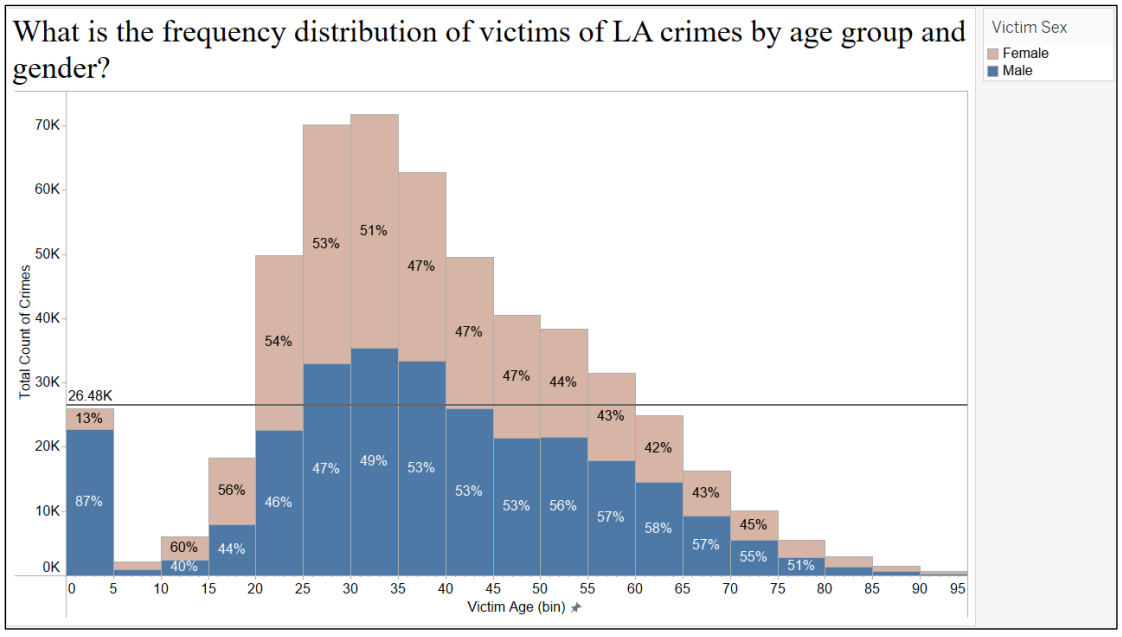


After conducting a deeper analysis of the LA crime data, we found that single-family dwelling apartments were the most targeted type of residence during the identified crucial hour. This

suggests that individuals who live in single-family apartments may be at a higher risk of being victimized during this time. The second most common location for crime during this period was on the street, which is not surprising given that it is a public space. However, it is important to note that crime can occur anywhere, and caution should be exercised in all areas.

It is also worth noting that restaurants ranked only 10th out of all premises where crimes take place. This may be due to several factors, such as the presence of security measures, surveillance cameras, or the fact that many restaurants are located in areas with high foot traffic and visibility.

6. What is the distribution of LA crime victims based on their age group and gender?

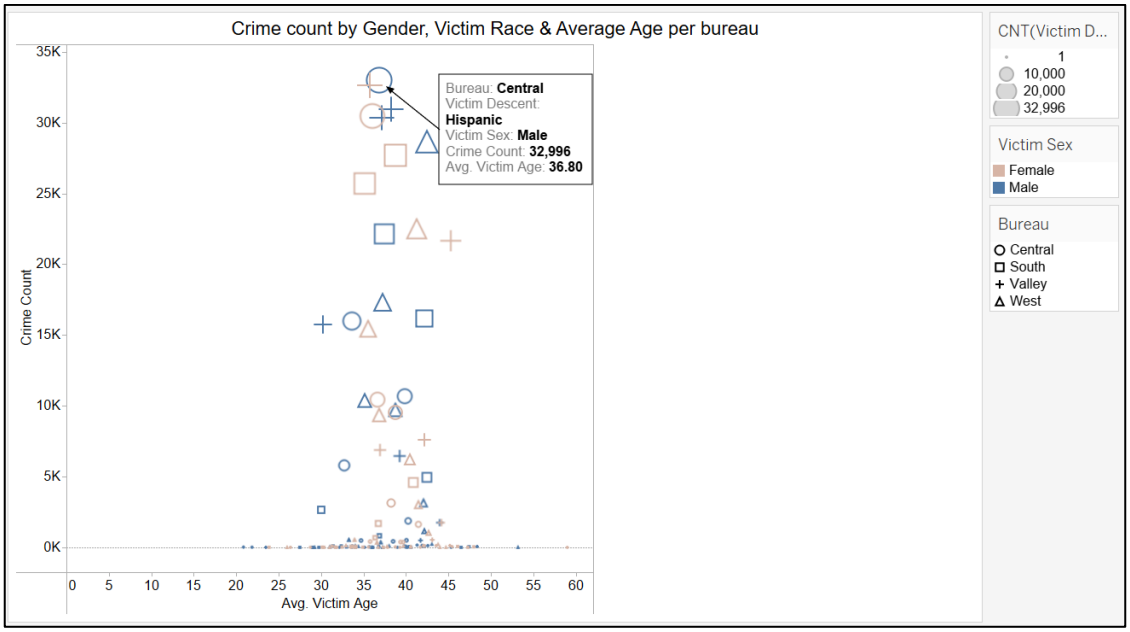


Categories Used: Histogram, Reference Line, Filter, Bin

We have used histogram to analyze victims of which age group based on gender are vulnerable to crime. Histogram is like a bar chart. The only difference is that it groups the value of Age into a continuous range (0-5, 5-10, and so on). Each bar represents the number of victims within a

specific age range and gender group. The height of the bar indicates the frequency or the count of victims in that age and gender group, while the width of the bar indicates the range of ages that the group encompasses. It is quite evident from the graph that people in the age group 25-40 are more prone to crime. The data indicates that within the age group of 25-35, the percentage of female victims is slightly higher than that of males, implying that females are slightly more likely to be victims of crimes in this age range. In contrast, within the age range of 35-40, the percentage of male victims is higher than that of females, suggesting that males are more likely to be victims of crimes in this age range. While creating the histogram we set the bin size to 5. The Average count of crime is shown by Reference Line and is 26.48K. This reference Line serves as a point of comparison to the bras of the histogram. The Age group 20-60 is above the reference line which indicates the crime count for this age group is higher than the average count.

7. What is the demographic group that is most impacted by crime in the bureau? Display the information by age and gender of the victims.

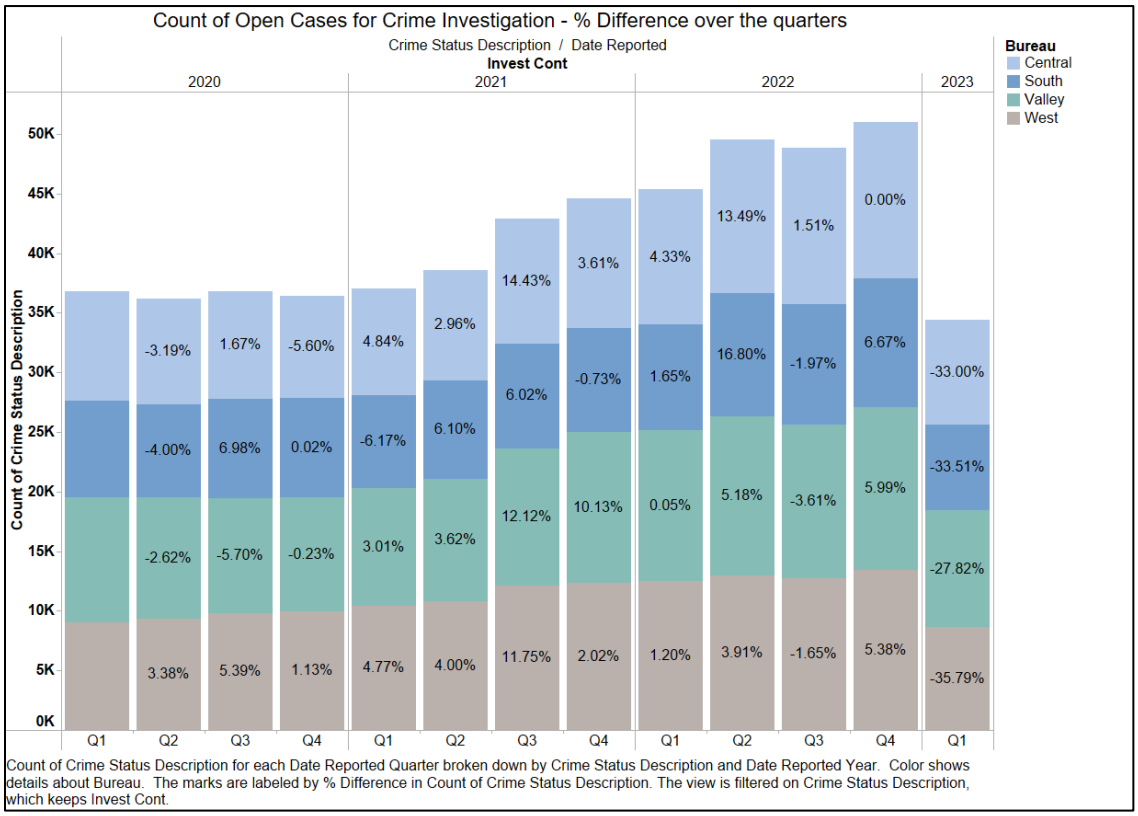


Categories Used: Scatter plot, Group, Filters

An analysis has been conducted to investigate the crime count pattern based on the victim's age, gender, and ethnic background. This analysis can help us identify which age group is more vulnerable to crime, and by incorporating the bureau as a dimension, we can determine the areas with a higher risk of victimization.

A Scatter plot is used to graphically represent the count of crimes and the average age of the victim. The different shapes on the graph represents the number of crimes that occurred in the bureau, while different colors distinguish between male and female victims. The data shows that the age group with the highest number of victims falls between 30 to 40, and the Central bureau has the highest number of male victims, while the Valley bureau has the highest number of female victims. Our analysis indicates that victims of White and Hispanic descent have the highest incidence of crime.

8. How many open cases are there for criminal investigations, and what is the percentage difference for each quarter between the years 2020 and 2023?



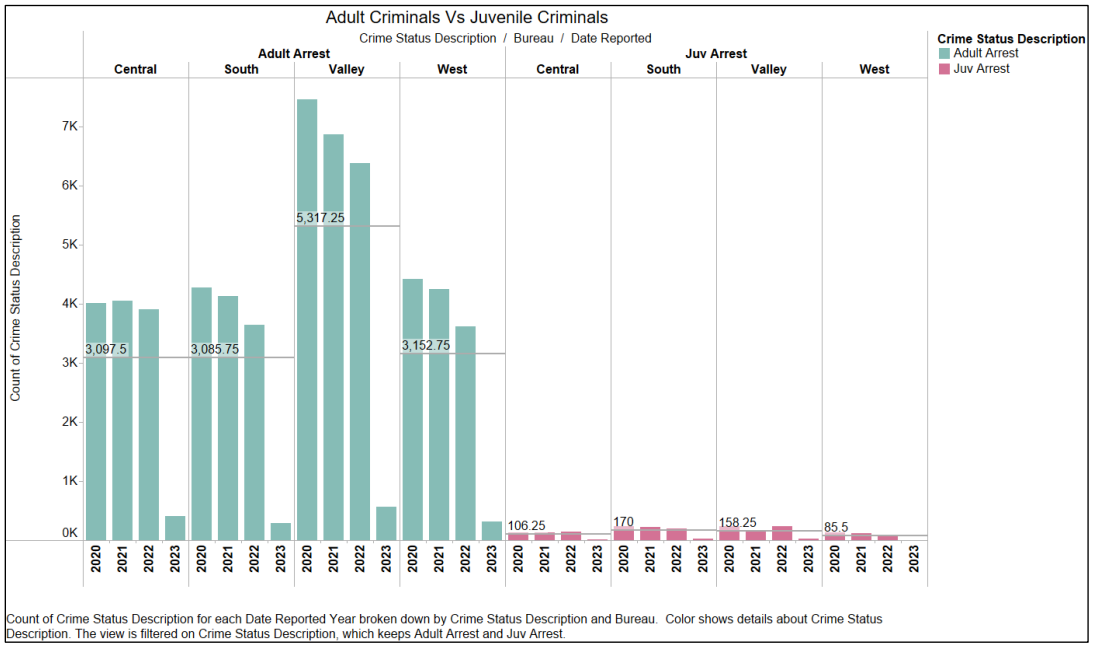
Categories Used: Date – Year, Quarters, Group, Stacked Bars

An Analysis is made to determine how many open cases are there in each bureau and for each quarter. This analysis will be helpful to allocate more resources and to prioritize the cases.

We have used stacked bars and applied colors to differentiate each bureau. The difference in percentage is calculated by comparing the count of open crime status with the previous quarter. From the analysis it is clear the count of open cases increases gradually and the year 2021 Q3, except for south bureau all the other bureau open crime counts increased and decreased again for Q4, whereas in 2022 Q2 South bureau open cases were increased from 1.65 % to 16.80%. In 2022

Q4, the central bureau has the same number of open cases compared to the previous quarter Q3. This type of analysis provides insight to law enforcement team to solve cases more effectively and build trust with the community.

9. Which Bureau has the highest Adult and Juvenile Arrest? Compare by year on year.



Categories Used: Groups, Reference Line, Date - Year

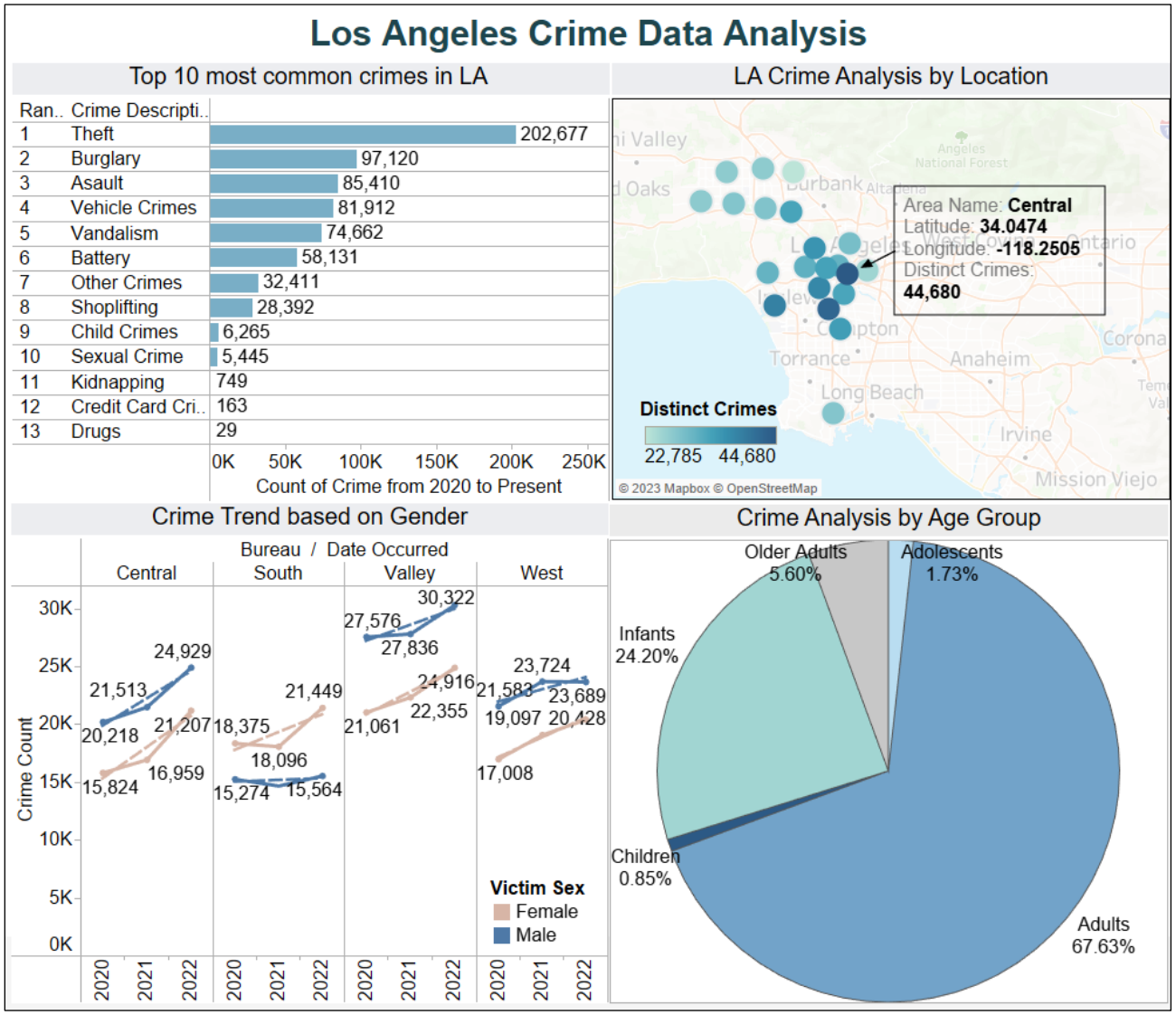
We have created a side-by-side bar chart to compare adult and juvenile arrest data, which was grouped by bureau using the create option based on their division. Analyzing and comparing the arrest rates between juveniles and adults is crucial in understanding crime patterns and identifying risk factors for each group. Adults and juveniles may have different reasons and motivations for committing crimes, and they may require different approaches for rehabilitation and reintegration into society. By understanding the differences between adult and juvenile offenders, law

enforcement agencies can tailor their response and interventions to best address the needs of each group, ultimately leading to better outcomes for both the offenders and the community.

We generated a visual representation to gain a clear understanding of the bureau with the highest crime rate and the year with the most significant number of arrests. The analysis reveals that the Valley Bureau had the most significant number of adult and juvenile arrests. Furthermore, we observed that the year 2020 recorded the highest number of adult arrests, while in the case of juveniles, there was a change in trend. In 2021, the South bureau had the most substantial number of juvenile arrests, whereas the Valley bureau had the highest number of arrests in 2022. We also included a reference line to estimate the average crime rate in the bureau over the years.

E. Dashboard

Tableau is a powerful data visualization tool that allows users to easily create interactive and informative dashboards. A Tableau dashboard is a collection of multiple views (charts, graphs, maps, etc.) that are presented on a single page and are designed to provide a clear and concise overview of data. The process of analyzing and comparing multiple worksheets simultaneously can provide more comprehensive insights. As a result of this analysis, a dashboard has been created, which is presented below.

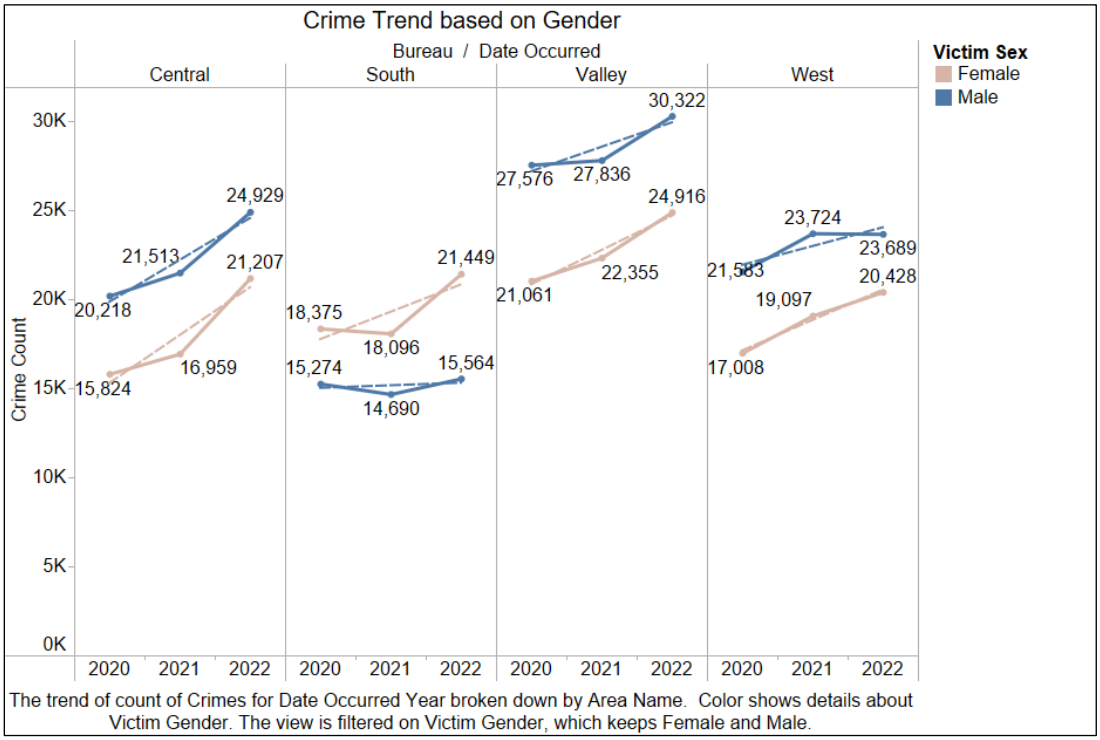


To understand the Los Angeles Crime Analysis from the year 2020 to the present, we will now take a closer look at each of the visualizations present in the dashboard.

The Bar chart, which reveals LA's most common crimes, determined by a tableau grouping function and filtered to the Top crimes, shows that “Theft” is the most prevalent crime, posing a risk to residents who should exercise caution. The analysis benefits law enforcement agencies and the public by providing insights into criminal activity, enabling further precautions.

Additionally, the geographical map of crime distribution in LA displays high and low crime areas through plotted longitude and latitude, allowing residents to identify areas where crime rates are particularly high or low. This analysis benefits law enforcement agencies and the public by providing insights into criminal activity and enabling targeted strategies for prevention, such as increased police presence and surveillance technology.

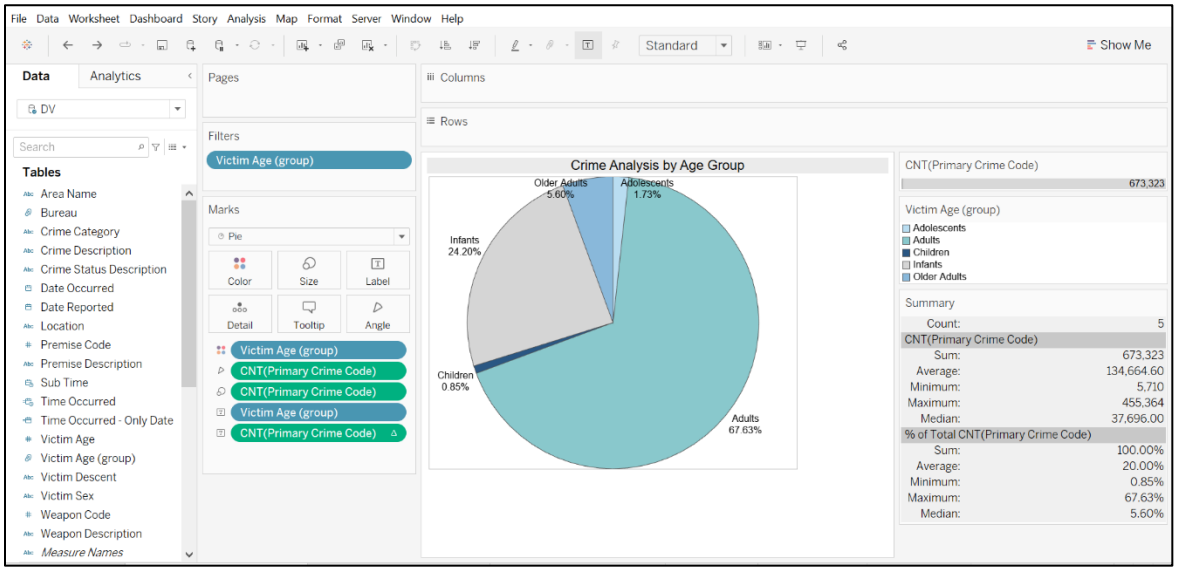
The trend in the number of reported crimes based on the gender of the victim over the course of the year is being analyzed:



Categories Used: Trend Lines, Date, Group

To understand the pattern of male and female victimization in different bureau, we used continuous line graph to show the trend between them. From the visualization, we can see for the bureau Central, South, Valley both male and female count decreased in the year 2021 and then again raised in 2022.On the contrary the male and female count got increased for the West bureau in the year 2021 and the count dropped again in 2022. We have used Trend line to understand the pattern of data set and the direction of data set. This analysis indicates the trend involved in male and female victim and over all landscape of victimization. It's quiet surprise to learn that male and female are equally vulnerable and their victim count rate very close.

Analysis on specific age groups which are more frequently targeted by criminal activity:



Categories Used: Pie chart, Count, Group, Filter

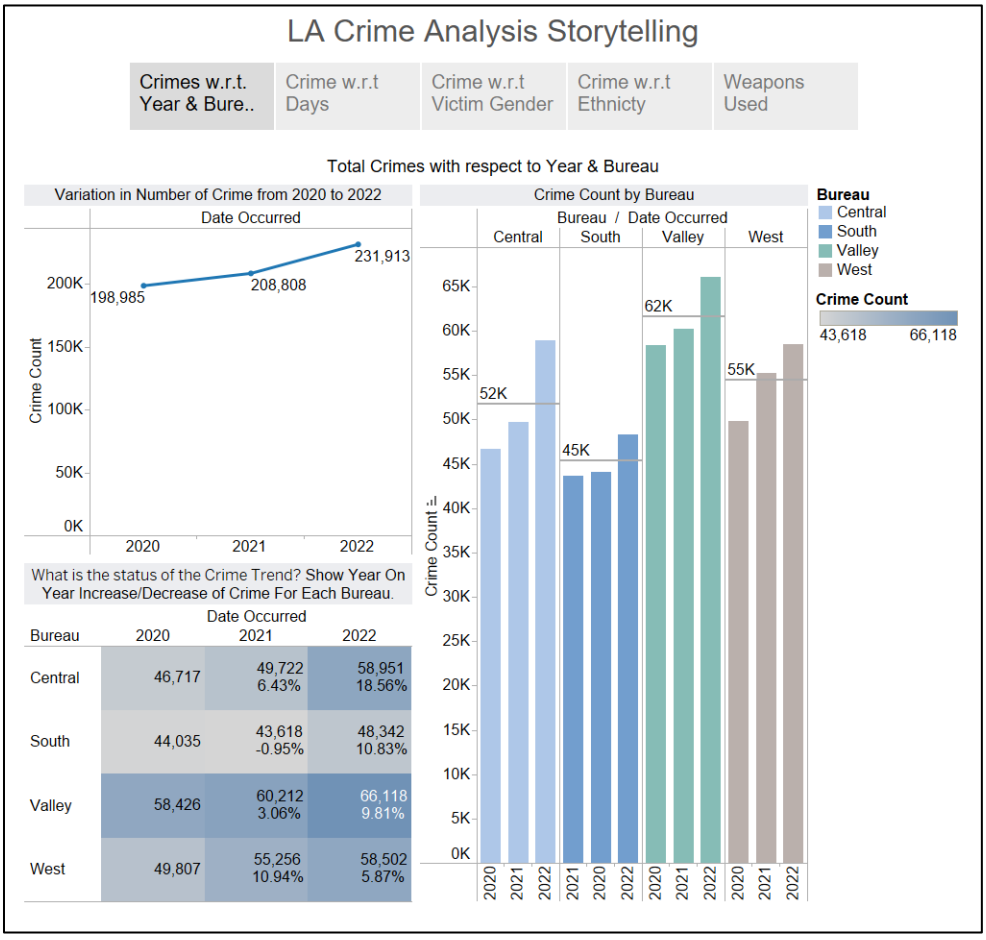
The pie chart above illustrates an analysis of crime categorized by age group. The age groups are established using tableau group functionality and then filtered accordingly. The largest proportion of victims, which accounted for 67.63% of the total, are adults aged 18-64, indicating that this age group is more vulnerable to certain types of crimes, such as property and violent crimes. The second-highest percentage is infants aged 0-1 years at 24.20%, although they are unlikely to be direct victims of crime, they could be at risk of harm in households with high crime or violence rates. The elderly aged 65 and above represented only 5.60%. The smallest percentage of 0.85% is for children aged 2-12 years, who are at risk of gunshots or active shooter attacks based on recent occurrences of shootings at school premises.

This data analysis indicates that crime can impact different age groups in distinct ways, and policymakers, law enforcement, and community organizations can use this information to create

targeted strategies for preventing and addressing crime's impact on various age groups within the population.

F. Story Telling

The story is to detail the analysis conducted on the LA crime dataset, which encompasses all the crimes reported between 2020 and the first two months of 2023. This dataset contains various pieces of information such as the date when the crime was reported, the date and time of the incident and the area where the crime occurred. It also includes details such as the crime code, description, and status (whether the criminal has been arrested or not), as well as the victim's demographic information like age, gender, and race, the crime location, and the latitude/longitude information. To properly analyze the data and create visualizations, additional categorization of crimes has been performed, such as grouping crime areas by the bureau and categorizing crimes by their type.

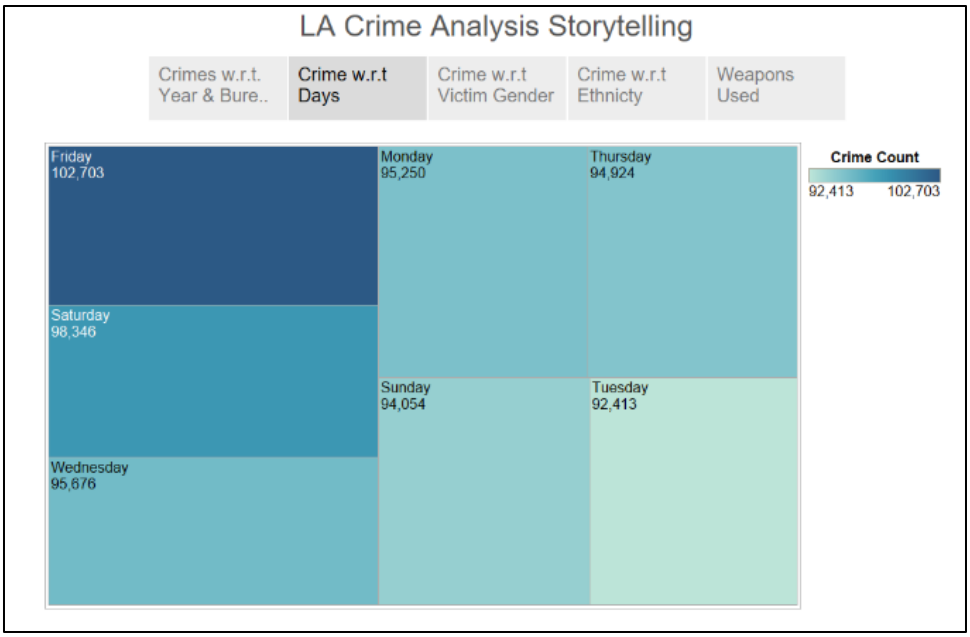


Insights into crime patterns by area and bureau are revealed through the generated visualizations shown above. A **line chart** exhibits the overall count of crimes across the years, with the least number of crimes recorded in 2020 (around 199K) and the highest in 2022 (approximately 232K). This evidently corroborates the notion that the occurrence of crimes has been on the rise each year.

The **bureau and yearly graphs** demonstrate the number of reported crimes per bureau per year and reveal fluctuations in crime rates. Average crime rates are used as reference lines to compare the crime rates of each bureau. Based on the analysis, it was found that the Valley Bureau had the highest crime rate in 2022, with an approximate average count of 62,000 for all three years. On the other hand, the South Bureau had the lowest average number of crimes.

The crime analysis also includes assessing the **percentage change** in crime rates compared to previous years, to identify any upward or downward trends in crime rates within each bureau. The analysis reveals that the year 2022 had the highest increase in crime rates, with a rise of approximately 12% in all bureaus. On the other hand, the year 2021 showed a decrease in crime rates, particularly in the South bureau, with a decline of 0.95% in the crime count.

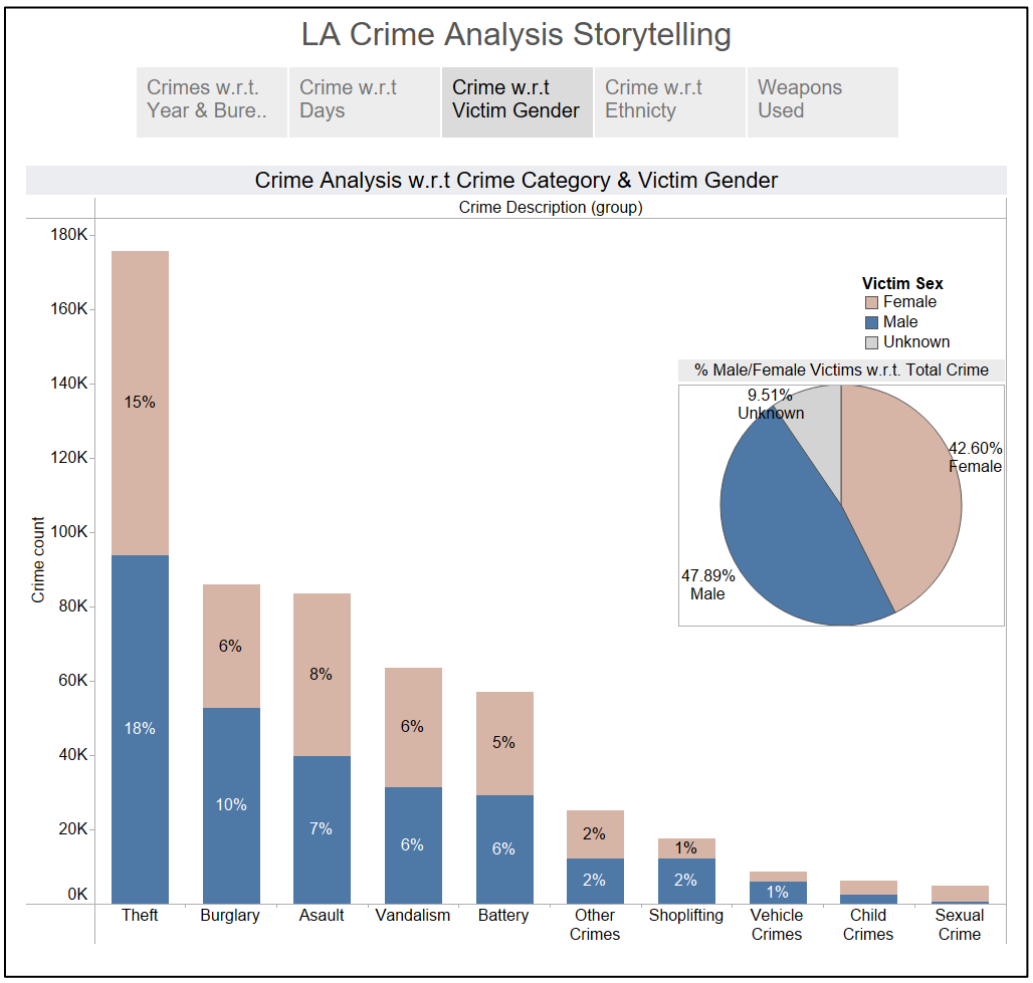
Furthermore, as per the March 20th publication of USA Today, it has been reported that the crime rate in Los Angeles rose by 11% in 2022, with 60 reported crimes per 1,000 residents as opposed to 54 per 1,000 residents in 2019.*[5]*



The preliminary information provides an understanding of crime trends based on location and department, while a **Tree map** is utilized to analyze data and present crime occurrences across different days of the week from 2020 until now. To focus on the specific days of the week, a filter is applied to exclude the other dimensions and only display data for the days of the week. According to the analysis, weekends are trouble! Friday accumulated the most number of crimes around 102,703, followed by Saturday with approximately 98,346 cases. In contrast, the lowest number of crimes occurred on Tuesday, with about 92,413 cases. An article "LAPD Launches 24-Hour Crime Analysis Center *"[6]* talks about how useful it is to know the crimes that occurred on the days of the week. This visualization is helpful in identifying patterns in crime occurrences based on the days of the week. It can be advantageous for law enforcement agencies and policymakers to allocate resources and implement measures to enhance public safety on the identified days.

After identifying the days of the week with the highest occurrence of crimes, let's shift our focus to the most common crime types based on victim demographics. Specifically, we conducted an

analysis of the gender of the victim, as this demographic category is considered the most vulnerable. This analysis was conducted for each year between 2020 and 2023.



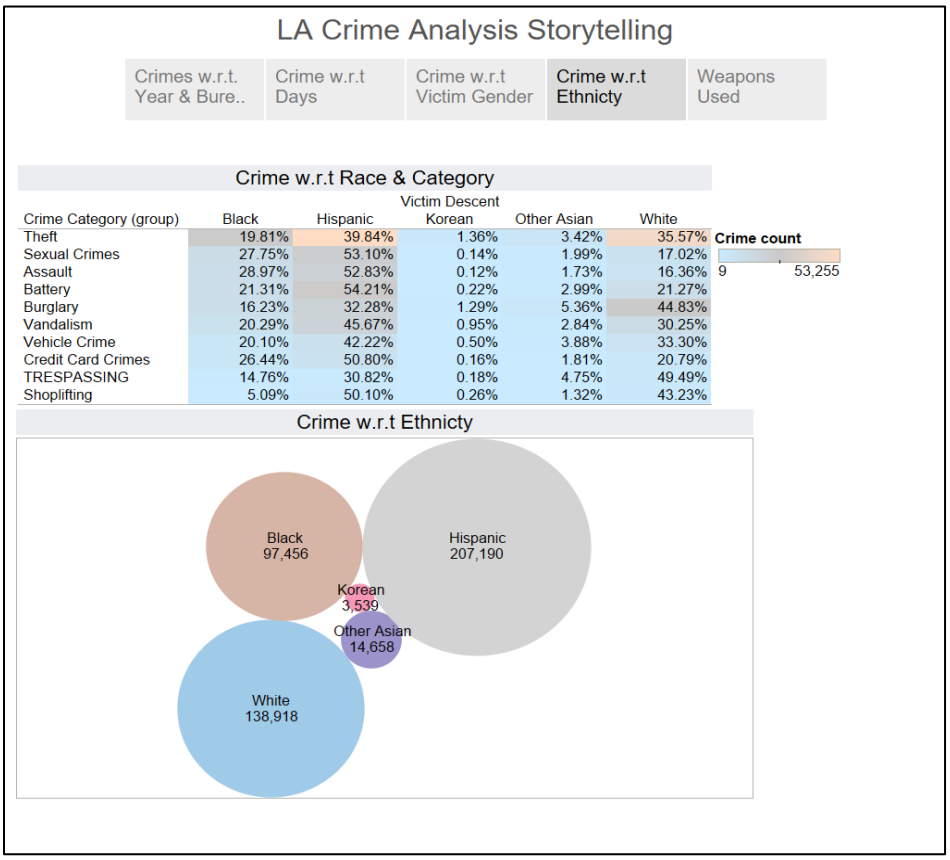
According to our analysis, there was nearly no difference between the number of male and female victims of crime. Out of all the reported crimes, approximately 49% of the victims were male and 43% were female.

The analysis of crimes against male and female victims revealed that "Theft" was the most frequent crime, comprising 18% of all reported male victim cases. Burglary (10%) was the next most common crime, followed by Assault (7%). Similarly, for female victims, the most prevalent crime

was "Theft," accounting for 15% of all reported crimes. "Assault" was the second most common crime (8%), followed by burglary (6%).

In general, our examination of LA's crime statistics concerning victim gender indicates that crime affects both males and females, with Theft being the most prevalent crime type for both. Moreover, Assault is also a noteworthy problem that affects both genders. This knowledge could assist law enforcement organizations in comprehending crime patterns and allocating their resources and efforts to lessen crime rates in the city.

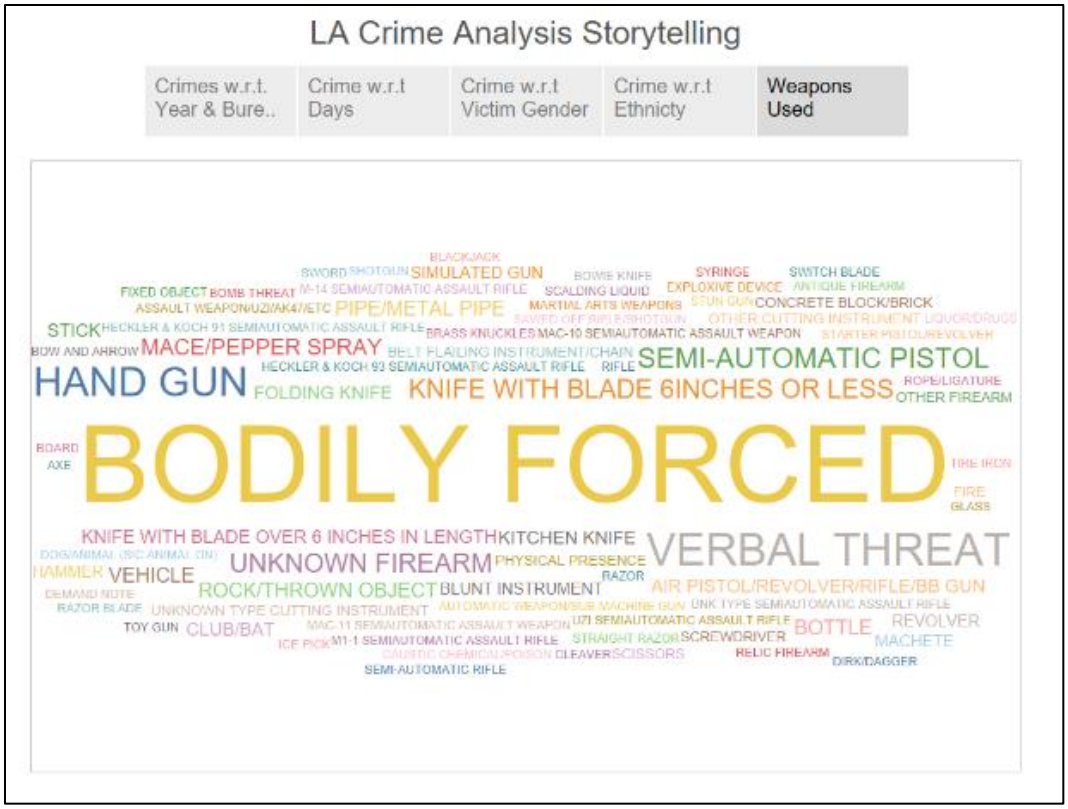
Moving on, to gain a better understanding of victimization patterns, we conducted an analysis based on ethnicity in Los Angeles, which has a diverse population. Previous analysis has already revealed the risk of victimization based on gender and age. However, ethnic differences within gender groups are more intricate.



Analysis of crime data reveals disparities in victimization among the ethnic group. We have used a **bubble chart** to display the Victim descent and used simple highlight tables to categories the crime group based on their descent. By Filtering the top 10 avoid cluttering visualization with too much information.

According to the data, Hispanics, Blacks, and White are the **most affected victims by crime** in terms of number of incidents reported. 39.84% of all theft victims were Hispanics which is a little high compared to white victims who has the percentage of 35.57%. White descent faced the highest Burglary percentage of 44.83% compared to all other races. Assault crime category was the highest crime happened to Black descent Individuals when compared to all other crime types. Hispanics were highly victimized with a percentage of 54.21% under the category of Battery. Korean was the least affected victim with the highest 1.39% faced for theft. Other Asian were highly victimized for Burglary crime. This Analysis also reveals the disparities in the victimization rates for different ethnicities within specific crime type. *[7]*

To further understand the nature of criminal activity, it is crucial to analyze the weapons used in crimes. Delving into the types of weapons used can help identify patterns and trends that could assist in preventing future incidents and keeping communities safe.



Weapons can come in various forms, ranging from high-powered machine guns to the use of mere physical force. Anything that causes physical or mental harm to humans can be considered a weapon, and it is crucial to identify the weapons used in crimes to understand the nature of criminal activity. To conduct our analysis, we utilized the **Word Cloud technique**, which enables us to identify the most common weapons used by criminals. The larger the text, the higher the likelihood of it being used as a weapon. Our analysis indicates that the most used weapon by criminals is bodily force, followed closely by verbal threats and guns. The visual representation of this analysis provides a clear understanding of the weapons commonly used in criminal activity, which can aid law enforcement organizations in devising efficient approaches to avert such occurrences and ensure the safety of communities.

Overall, our analysis of crime data in Los Angeles has revealed important insights into crime patterns and victimization. The central region of LA has several high crime areas, while lighter

areas have lower crime rates. The Foothill area has a lower frequency of criminal incidents compared to other areas. The number of reported crimes increased from approximately 199,000 in 2020 to roughly 232,000 by 2022. Weekends, particularly Friday and Saturday, have the highest number of crimes, while Tuesday has the lowest. Crime affects both genders with theft being the most common crime type, followed by assault. The age group of 25-40 has the highest number of victims affected by crime. Hispanics, Blacks, and Whites are the most affected victims by crime based on reported incidents. Bodily force was the most used weapon, followed by verbal threats and guns.

G. References:

1. “Los Angeles Police Department.” *Wikipedia*, Wikimedia Foundation, 24 Mar. 2023, https://en.wikipedia.org/wiki/Los_Angeles_Police_Department.
2. “Chapter 1 Introduction to Criminal Law.” *Introduction to Criminal Law*, https://saylordotorg.github.io/text_criminal-law/s05-introduction-to-criminal-law.html.
3. “Types of Criminal Analysis and Which of Them to Choose - 328 Words: Essay Example.” *Free Essays*, https://ivypanda.com/essays/types-of-criminal-analysis-and-which-of-them-to-choose/.
4. “LAPD Organization.” *LAPD Online*, 12 May 2022, https://www.lapdonline.org/lapd-organization/.
5. "Los Angeles sees 11% rise in crime rate, with 60 crimes per 1,000 residents in 2022." USA Today, Gannett Satellite Information Network, 20 Mar. 2023, https://www.usatoday.com/story/news/2023/03/20/los-angeles-crime-rates-map-safest-neighborhoods/11032761002/.
6. Daily News. “LAPD Launches 24-Hour Crime Analysis Center.” *Daily News*, Daily News, 29 Aug. 2017, https://www.dailynews.com/2006/01/18/lapd-launches-24-hour-crime-analysis-center/.
7. Allen J. Beck. “Race and Ethnicity of Violent Crime Offenders and Arrestees, 2018.” *Race and Ethnicity of Violent Crime Offenders and Arrestees, 2018*, https://bjs.ojp.gov/content/pub/pdf/revcoa18.pdf.