

Analyzing Los Angeles Crime Data from 2010 to September 2017

DATASCI 200 Project 2

Sineli Gamage, Shruti Jain, Himaja Jangle

Github Repository: [Project2_Gamage_Jain_Jangle](#)

Dataset: [Data from week we downloaded](#) (Consistent with our analysis), [Kaggle](#) (Updated weekly)

Github: https://github.com/UC-Berkeley-I-School/Project2_Gamage_Jain_Jangle

Introduction

The dataset we are studying focuses on crime in the LA area from 2010 through September 2017. It has 1584316 rows and 26 columns, and we are specifically interested in this subset of columns: Date of Occurred, Area Name, Status Description, Crime Description, Victim age/sex/descent, and Weapon Description. This dataset has been put together with information from crime reports and the source acknowledges that some inputs may be incorrect.

We are looking to see if we can find trends in crime in LA by looking at a variety of factors and how they interact with each other, such as: Common areas of crime in LA, Common types of crimes, Common age range and race of victims, Common times of crimes occurred, Common dates of crimes occurred, and Arrests made

Data Cleaning and Sanity Checks

The initial dataset has 1584316 rows, 26 columns. After cleaning the dataset has 1584316 rows, 16 columns. Data Cleaning consisted of dropping columns that were outside the scope of our data analysis and converting columns 'Date Reported' and 'Date Occurred' into datetime objects. We also mapped the columns 'Victim Sex' and 'Victim Descent' so that their values were more descriptive. For example, changing 'M' in the column 'Victim Sex' to 'Male', or changing 'V' in the column 'Victim Descent' to 'Vietnamese'.

In terms of sanity checks, we ensured that the 'Date Reported' and 'Date Occurred' columns were within the range of 2010 to 2017, and that the 'Time Occurred' column did not exceed the range of 1-2359 (24-hour time). We also made sure the Crime code and its corresponding Description matched up, as well as for the Area ID and Area name columns.

Research Questions

1. What is the most prevalent type of crime?

To understand the correlation between crime, area and weapon usage, there are a few sub-questions that can help to specify and understand the different types of weapons and different areas in LA with corresponding crimes.

a. What is the most common weapon used?

The top five weapons involved in reported incidents include "STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)," "VERBAL THREAT," "UNKNOWN WEAPON/OTHER WEAPON," "HAND GUN," and "VEHICLE." The most common weapon used in Los Angeles is "STRONG-ARM (HANDS, FIST, FEET, OR BODILY FORCE)." These findings suggest a diversity of weapon types, with physical force being the most prevalent. This is the most common weapon possibly because people often use physical force during everyday arguments or fights. These situations can quickly escalate, and without any planned weapons, individuals resort to using their bodies. The spontaneous and easily accessible nature of strong-arm tactics makes them prevalent, especially in simple assaults.

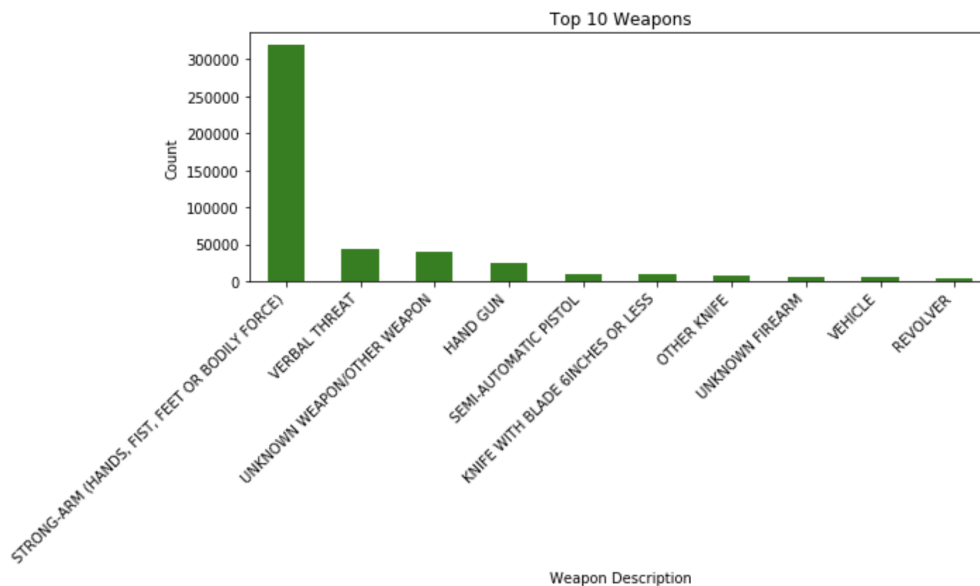


Figure 1a. Bar Chart showing the value counts of the top 10 weapons

b. For the top 10 crimes, what is the top weapon for each?

Among the top 10 crimes, the most common weapon is "Strong-arm" (involving hands, fists, feet, or bodily force), particularly associated with offenses such as Battery - Simple Assault, Intimate Partner - Simple Assault, Robbery, and Attempted Robbery. However, crimes like Assault with Deadly Weapon, Aggravated Assault, and Brandish Weapon show a distinct association with specific weapons such as Hand Gun. Criminal Threats with no weapon displayed are often linked to verbal threats, while Burglary tends to involve an unknown or other types of weapons. Intimate Partner - Aggravated Assault follows the trend of simple assault with a preference for strong-arm methods.

Crime		Top Weapon
0	BATTERY - SIMPLE ASSAULT	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)
1	INTIMATE PARTNER - SIMPLE ASSAULT	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)
2	ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT	HAND GUN
3	ROBBERY	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)
4	CRIMINAL THREATS - NO WEAPON DISPLAYED	VERBAL THREAT
5	ATTEMPTED ROBBERY	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)
6	BRANDISH WEAPON	HAND GUN
7	INTIMATE PARTNER - AGGRAVATED ASSAULT	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)
8	BURGLARY	UNKNOWN WEAPON/OTHER WEAPON
9	BATTERY WITH SEXUAL CONTACT	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)

Figure 1b. Table showing the top weapon for the top 10 crimes

c. Do certain areas in Los Angeles exhibit a higher frequency of specific crime codes?

The top crimes in various areas of the city exhibit a notable pattern. "Battery - Simple Assault" emerges as the predominant offense in several regions, including 77th Street, Central, Hollywood, Olympic, Rampart, Southeast, and Southwest. Meanwhile, property crimes such as "Burglary," "Burglary

from Vehicle," and "Vehicle - Stolen" are prevalent in areas like Devonshire, N Hollywood, Topanga, Van Nuys, West LA, and West Valley. Interestingly, the Pacific area stands out for incidents categorized as "Theft Plain - Petty (\$950 & Under)." This suggests a mixture of both violent and property-related offenses across different neighborhoods, showcasing diverse crime profiles and enforcement challenges in the city.

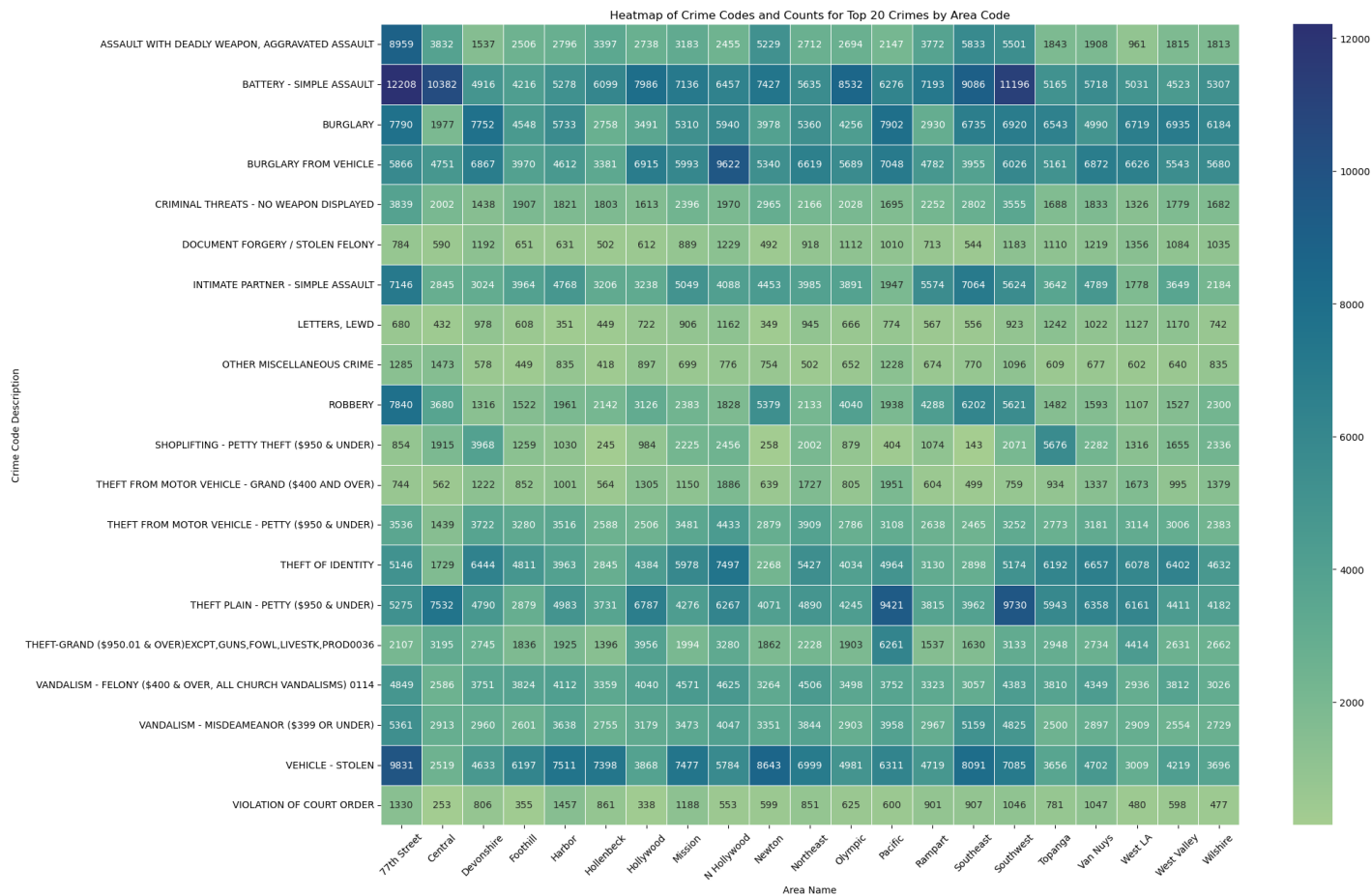


Figure 1c. Heatmap showing the Top 20 Crimes and their count in each LA area

d. Which areas in Los Angeles experience the highest incidence of crimes?

Of the reported crimes in various areas of Los Angeles, the 77th Street area records the highest total crimes, reaching 110,605 incidents. This is followed closely by Southwest with 102,259 and North Hollywood with 86,405. Conversely, the Hollenbeck, Foothill, and Wilshire areas report comparatively lower total crimes, with counts ranging from 57,592 to 63,598. Crime rates in different areas of Los Angeles vary due to a mix of socio-economic factors, demographics, geography, law enforcement resources, community programs, and historical context. Higher poverty levels, specific demographics, and geographic features may contribute to increased crime. Additionally, law enforcement presence, community engagement, and historical factors play roles in shaping crime patterns.

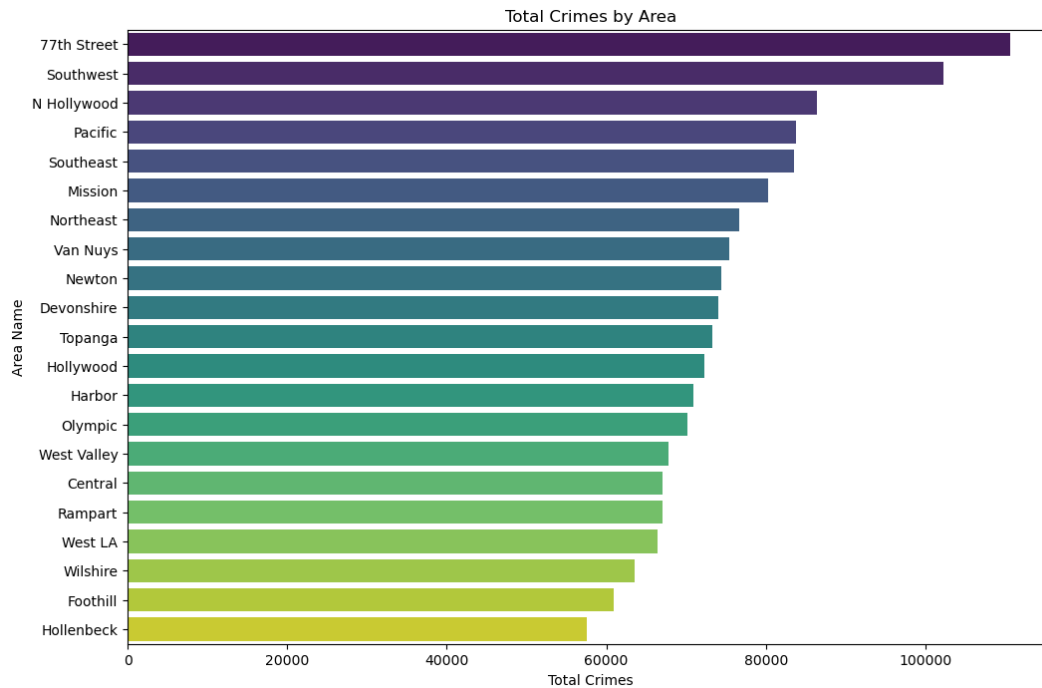


Figure 1d. Horizontal Bar Chart showing the total crimes by area

2. At what times are crimes most frequently occurring?

To understand the correlation between time and crime, there are a few sub-questions that can help to specify and understand the different types of time and corresponding crime occurrence.

a. During which month or year did the highest number of crimes occur?

We started off on a larger scale of looking into each year and month and their corresponding crime rate. A heatmap showcases the month with the highest crime occurrence for each year easily. At first glance, the darkest cells are January 2010, October 2016 and July 2017.

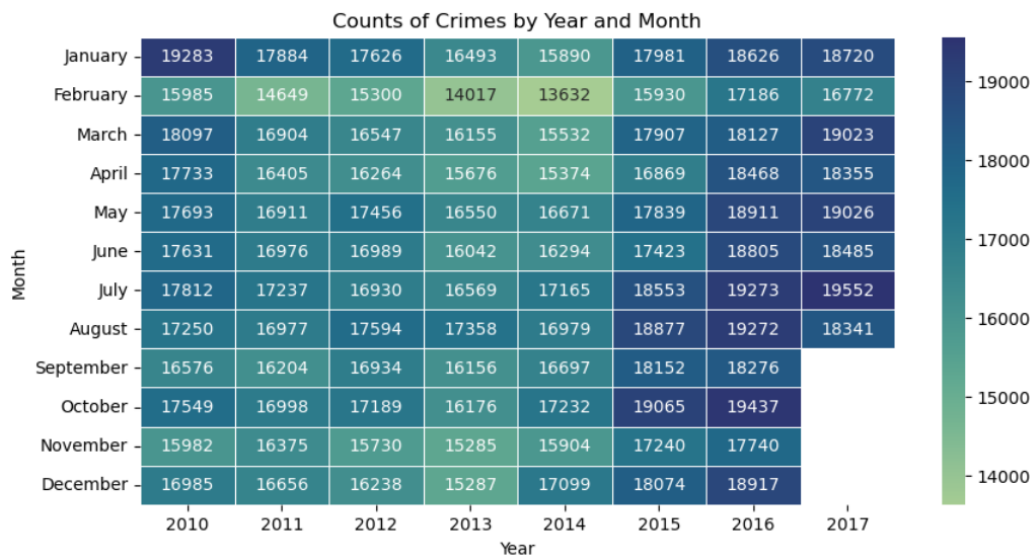


Figure 2a. Heatmap showing Crime frequency for each month in 2010-September 2017.

More closely, we can see that from 2010-2012 and from 2014-2016, the highest number of crimes occurred in January and October, respectively. Higher crime in January and October could suggest special

events such as New Years and Halloween. Although it is hard to determine any specific events during that time that may have triggered an increase in crime, certain events and activities that took place could be a possibility.

Year	Max Crime Count	Month
2010	19283.0	January
2011	17884.0	January
2012	17626.0	January
2013	17358.0	August
2014	17232.0	October
2015	19065.0	October
2016	19437.0	October
2017	19552.0	July

Table 2a. Highest crime count for each year and corresponding month

b. Do specific days of the week exhibit higher rates of criminal activity?

Narrowing our scope, we looked into if there was a specific day of the week that had an effect on crime rate. We plotted the count of crimes for each day and saw quite an increase for Friday. Friday marks the start of the weekend, and many people may engage in social events where alcohol and drug consumption can occur. Friday's are popular for nightlife and entertainment, and these areas experience higher crime rates. Impulsive behaviors from alcohol and drugs can lead to assault, public intoxication, and DUI (Driving Under the Influence). When people are out at these events and activities on Friday's, burglaries and other property crimes can occur as well.

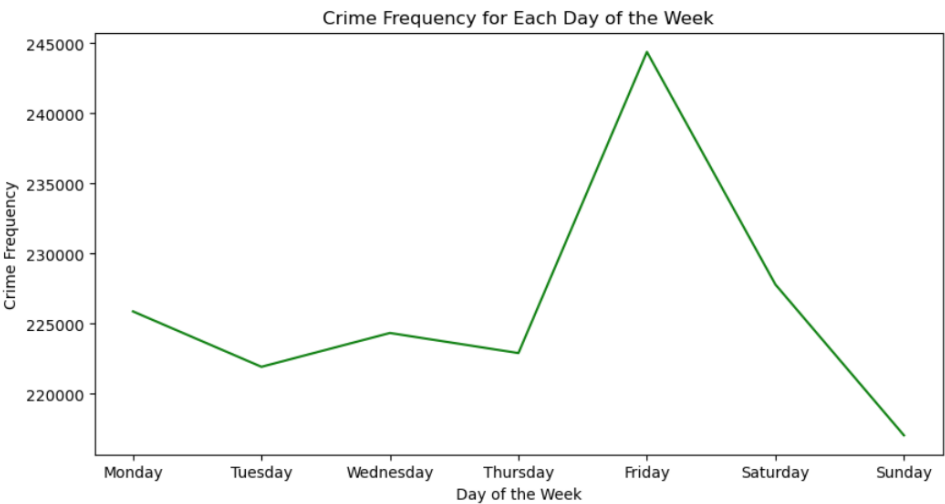


Figure 2b. Line graph depicting crime frequency for each day of the week.

c. Do crimes show a higher frequency during particular times of the day?

Looking more specifically at the time of day, we see a large jump in crime rate between 12pm to 1pm. This time is lunchtime for many individuals and can be a cause of increased crime. The lunchtime rush brings many people out of work and on foot finding their spot for a meal, and are more likely to be distracted and less prepared to respond to criminal activity. This vulnerability can possibly cause an increase in theft and pickpocketing. Retail areas can also experience an increase in theft as more people are visiting for shopping and eating. In addition, homes are also empty at this time as individuals are at work or school, and therefore burglaries and other property crimes may occur.

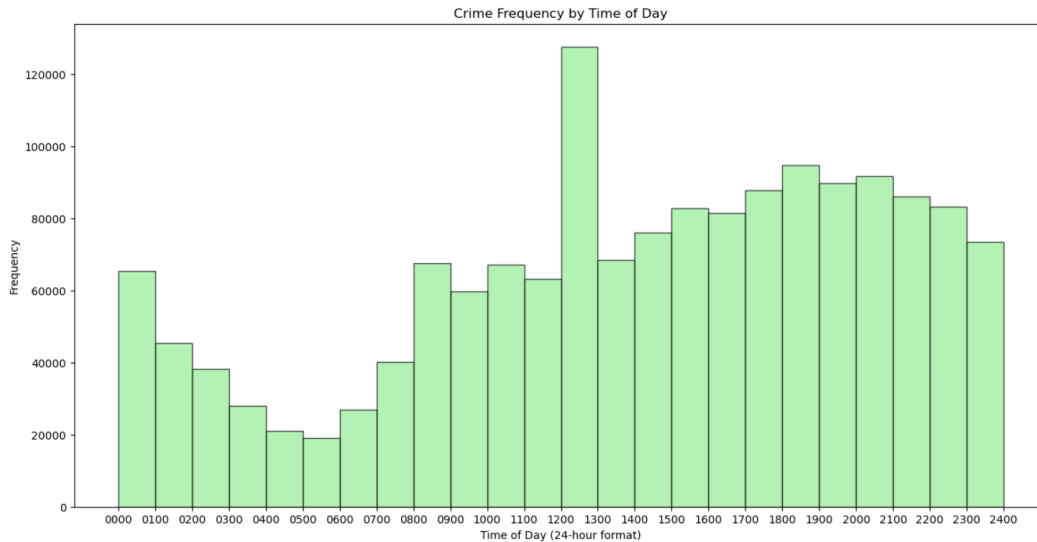


Figure 2c. Histogram showing the crime frequency based on time of day.

d. Do certain types of crimes tend to have a higher likelihood of being ongoing?

There are approximately 1227180 ongoing investigations, with some open and evidence still being gathered, and others that will continue being unsolved. With 'Burglary from vehicle' coming at highest crime count and 'Vehicle - Stolen' at almost the same, we can see that crimes involving vehicles are likely to be unsolved. These types of crimes are also quite common as cars parked in parking lots, residential areas, or public spaces are easily accessible. Due to Los Angeles having a higher population density, there is an insufficient law enforcement presence and also a lack of visibility with little to no security cameras in the area. Also, people are very likely to hide valuable inside car compartments, and these items can be sold quickly for money. Vehicles and their parts also have a large resale value and can be quick money.

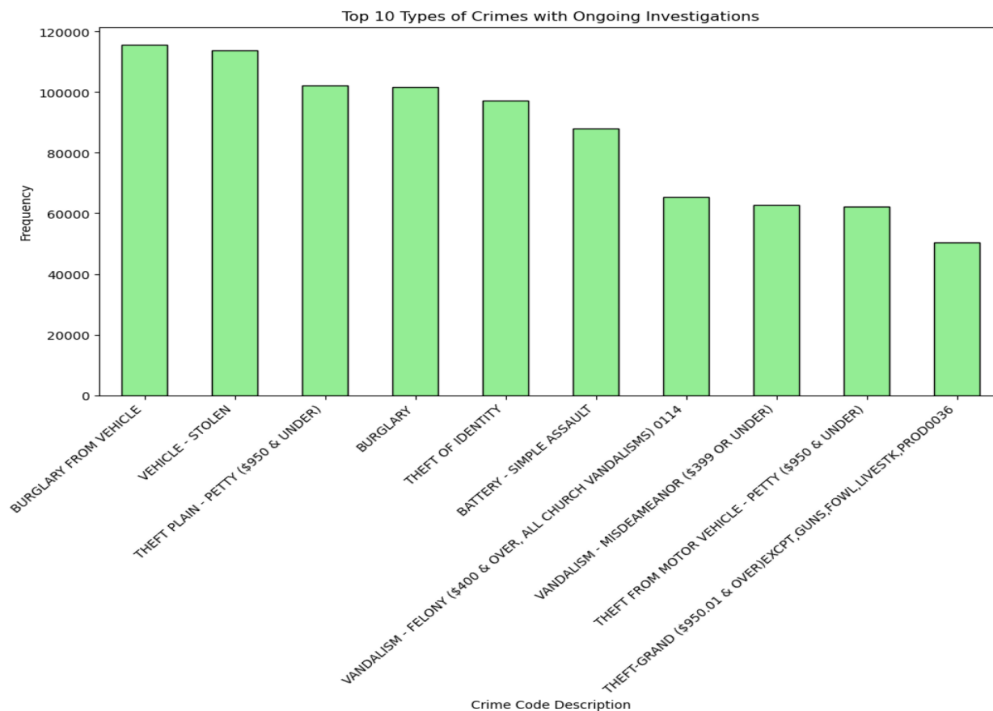


Figure 2d. Bar graph showcasing the top 10 crimes that remain ongoing.

3. Who are the most frequently identified victims or perpetrators?

In order to further understand this question, we are going to be diving deeper into different breakdowns of victims and perpetrators by factors such as age, sex, area of crime, and descent.

a. Is there a particular age group or gender that is more frequently targeted in incidents?

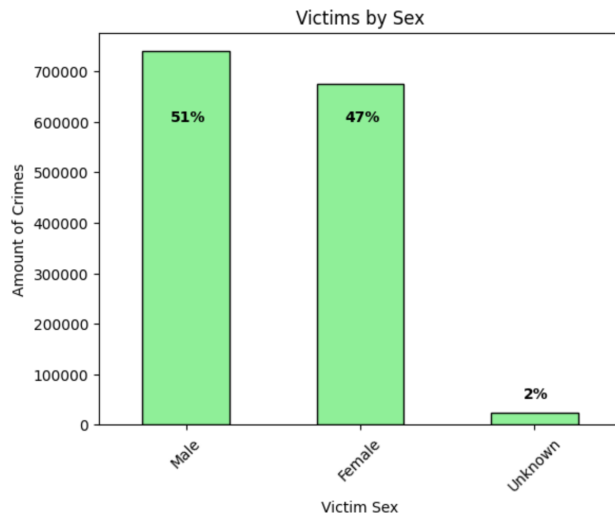


Figure 3a. Bar graph showcasing victim sex breakdown

Based on the graph generated above, it seems as if victims are almost equally split between male and female with just slightly more males as victims (51%). This tells us that whether crimes were targeted or not targeted, perpetrators did not care about the sex of their victim. There are likely other factors that perpetrators took into account when finding victims other than Victim Sex.

	Date Occurred	Percentage of Total
Victim Sex		
Male	739581	51.39
Female	675402	46.93
Unknown	24080	1.67

Figure 3b. Chart showcasing victim sex breakdown by percentage

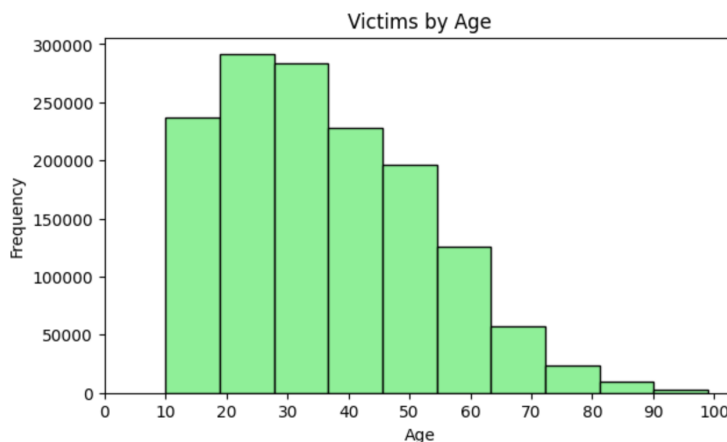


Figure 3c. Bar graph showcasing victim age breakdown

Based on the graph above, we can see that most victims are on the younger side, in the ages of 20-40. This could be either because those are common targets, or that is reflective of the age breakdown of the population in the Los Angeles area. After doing some research on age breakdowns of the LA population (LA Almanac) we find that ages 21-34 make up only about ~21% of the LA population while they make up much more of the victims in our crime dataset which is 92% for ages 21-34. This shows that younger victims were unproportionally targeted. This could be for a variety of factors such as them being careless or having valuable items for perpetrators to steal in the case of robberies.

	Victim Age	Victim Gender	Count
0	25.0	Female	20402
1	24.0	Female	20231
2	23.0	Female	19823
3	26.0	Female	19433
4	27.0	Female	18905
5	22.0	Female	18597
6	28.0	Female	18288
7	29.0	Female	17954
8	30.0	Female	17572
9	21.0	Female	17158

Figure 3d. Table showcasing top victim age/gender combinations

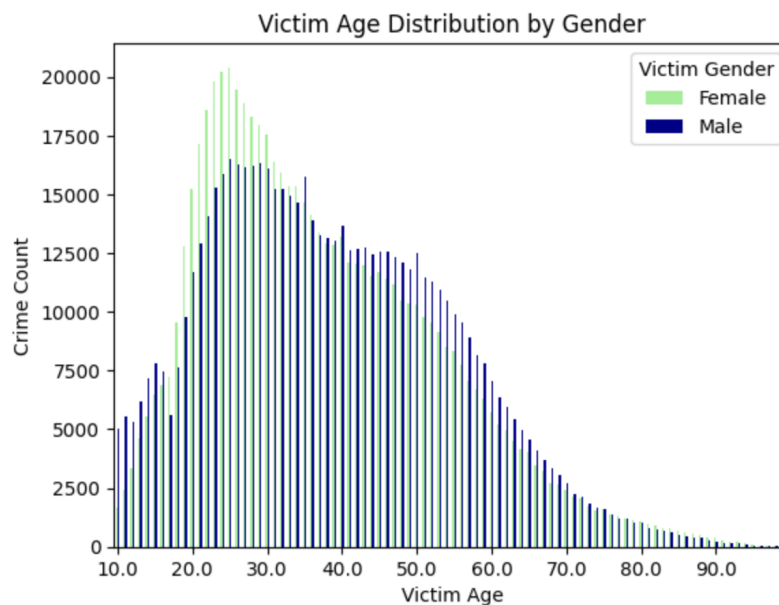


Figure 3e. Graph showcasing crime counts by broken out by victim age and gender

By looking at this breakdown, it is interesting to see how the crime counts by gender change by victim age. The top 10 victims are females ages 21-30, but interestingly enough, males still make up 51% of the entire victim pool. It is interesting to see that as women get older, they are not as “targetable” but men still are. This suggests that women in their 20s need to be extra vigilant in LA, as they seem to be

unproportionally targeted. This could be because they may be perceived as “easier” targets or tend to have high value items on them such as luxury purses.

b. Do specific areas in Los Angeles show a higher incidence of targeting a particular ethnic group compared to the demographic distribution of those areas?

We first looked at the breakdown of Victim Descents and the counts of crime associated with each one. Looking at this breakdown below, we can see that most of the victims are either Black, White, or Hispanic/Latin/Mexican. This helped narrow down the specific ethnicities to look at when breaking this down further by area. We were interested in seeing if specific areas have a higher count of crime against a specific ethnicity.

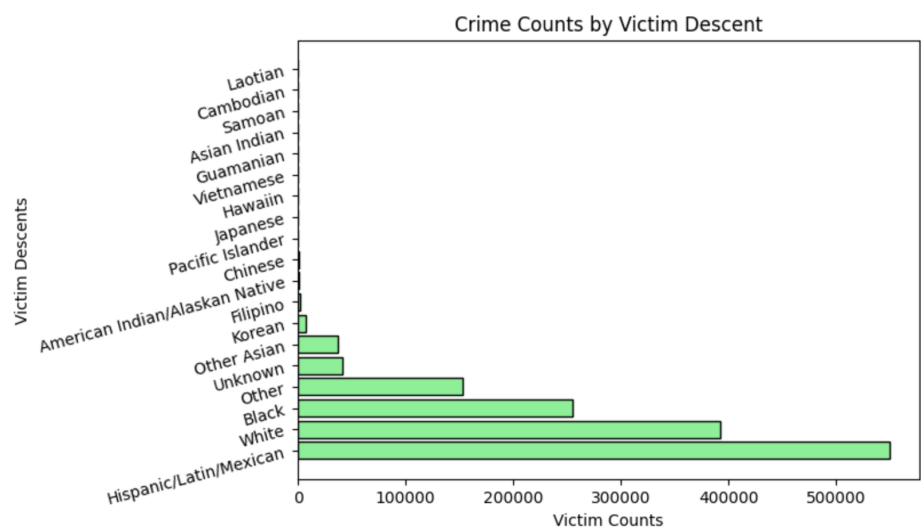


Figure 3f. Bar graph showcasing victim ethnicity breakdown

Looks like the largest group of victims are Hispanic/Latin/Mexican and White. This is interesting because if we compare it with the population (U.S. Census Bureau), these two ethnic groups make up 93%. Black makes up 8.6% of LA population but are 16% of victims which is pretty much double. This tells us that this ethnic group is unproportionally targeted for crimes in LA.

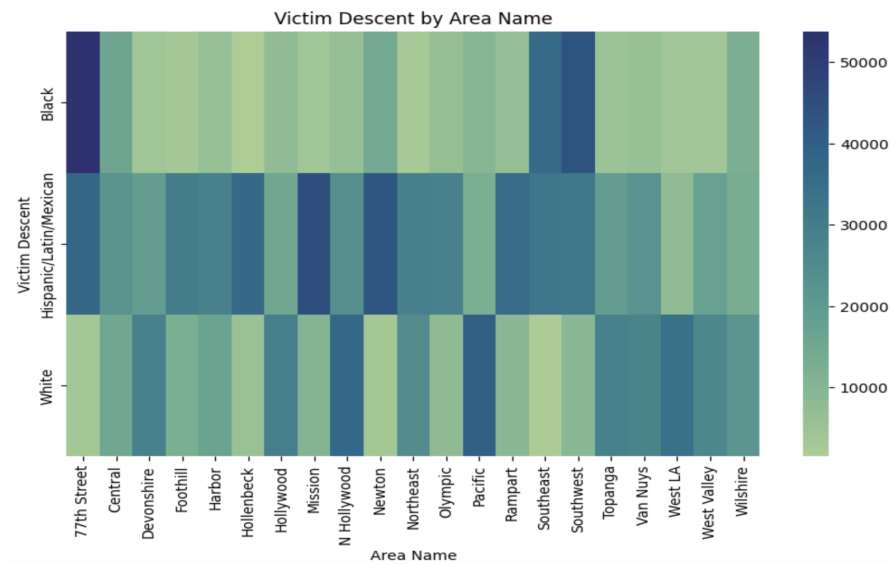


Figure 3g. heatmap victim sex breakdown by area

In this further breakdown, we can see that there are certain areas where more crimes occur against specific ethnicities. For example, we can see that Black individuals are most heavily targeted in the 77th Street area and significantly less in the other areas except for Southeast and Rampart. The other two top ethnicities have relatively similar distributions across all areas in our dataset. This suggests 77th Street can be a pretty dangerous area where specific groups are targeted and likely should be avoided for safety purposes by Black individuals, but also suggests wider issues.

c. What is the distribution of crimes committed by adults versus juveniles in various locations, considering only closed cases?

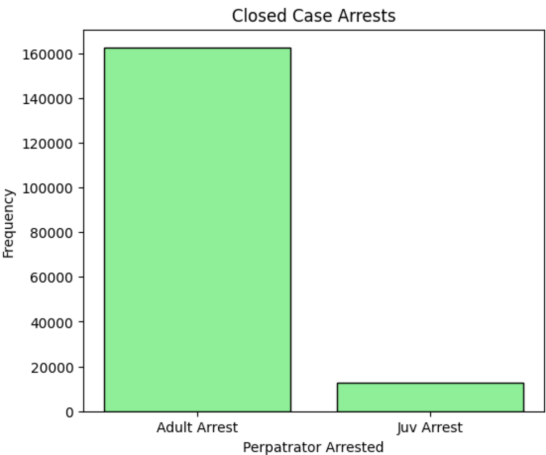


Figure 3h. Bar graph showcasing perpetrator breakdown of closed cases

Note that only 11% of the cases noted in this dataset led to arrests. 93% were adults which makes sense as Juveniles are individuals less than 18 years old. The population of Adults is much higher generally, so this is in line with what we expect. We did want to dive deeper to see if there were specific areas where Juvenile Arrests were more common, maybe in areas with schools or libraries. Looking at the heatmap below, we see a spike in Juvenile Arrests in the Southeast area without a similar spike in Adult Arrests. This area seems to be a higher area of Juvenile crime, likely one that many minors are at commonly, or are simpler crimes that can be committed by minors.

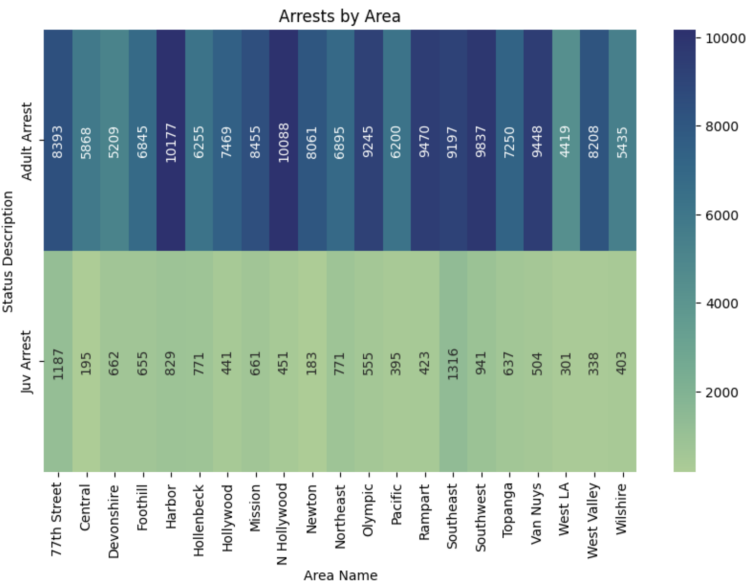


Figure 3i. Heatmap showcasing perpetrator type by area

Conclusion

The analysis of crime data in Los Angeles with regards to crime, weapons and area reveals interesting insights into the city's crime landscape. The most prevalent type of crime is "STRONG-ARM (HANDS, FIST, FEET, OR BODILY FORCE)," indicating a use of physical force, often due to spontaneous altercations. Verbal threats are seen as the primary weapon associated with the top 10 crimes, emphasizing the role of communication in various offenses, from bomb scares to criminal threats. Geographically, different areas exhibit distinct crime patterns, with "Battery - Simple Assault" dominating several regions. The areas experiencing the highest incidence of crimes, such as 77th Street and Southwest, underscore the complex interplay of socio-economic factors, demographics, and law enforcement resources influencing crime rates across Los Angeles.

From the visualizations of the crime data and crime occurrence at different times in the year/month/day, it was found that there are particular times that crime is more likely to happen. Crime incidence increases in January or October, on Friday's and between 12pm-1pm.

In general most victims are of Hispanic/Latin/Mexican, Black, or White descent. Additionally, while the split in victims is close to half for both males and females, the top 10 targeted groups were females ages 21-30, suggesting they should be extra vigilant.

Understanding these data points is crucial for crafting targeted strategies to address specific crime challenges and enhance community safety in the Los Angeles area.

Appendix

Data Structure

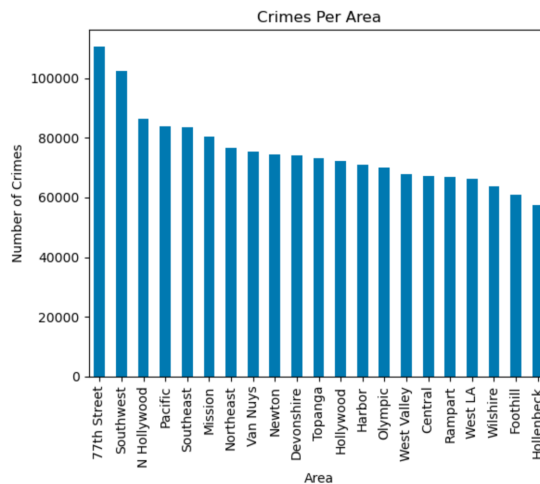
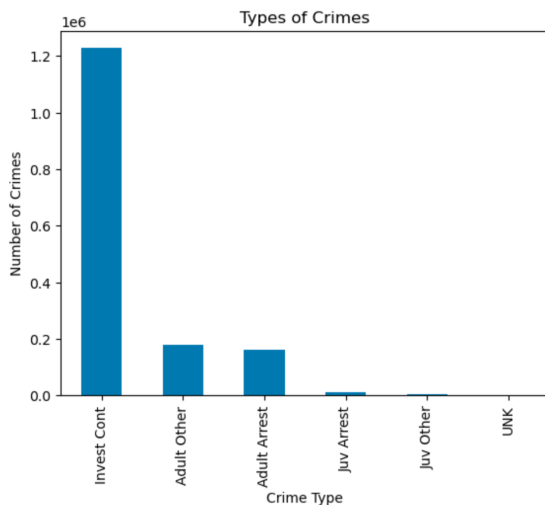
1584316 rows, 26 columns

Column Name	Example Value	Column Description
DR Number	1208575	Division of Records Number - Official file number made up of a 2 digit year, area ID, and 5 digits
Date Reported	03/14/2013	MM/DD/YYYY
Date Occurred	03/11/2013	MM/DD/YYYY
Time Occurred	1800.0	In 24 hour military time
Area ID	12.0	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	77th Street	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
Reporting District	1241.0	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
Crime Code	626.0	Indicates the crime committed
Crime Code Description	INTIMATE PARTNER - SIMPLE ASSAULT	Defines the Crime Code above
MO Codes	100	Modus Operandi - Activities associated with the suspect in commission of the crime. Defined in MO_Codes.csv
Victim Age	30.0	Two character numeric
Victim Sex	F	F - Female, M - Male, X - Unknown
Victim Descent	W	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
Premise Code	502.0	The type of structure, vehicle, or location where the crime took place
Premise Description	MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)	Defines the Premise Code above.
Weapon Used Code	400.0	The type of weapon used in the crime

Weapon Description	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	Defines the Weapon Used Code above
Status Code	AO	Status of the case. (IC is the default)
Status Description	Adult Other	Defines the Status Code above
Crime Code 1	510.0	Indicates the crime committed. Crime Code 1 is the primary and most serious one. Crime Code 2, 3, and 4 are respectively less serious offenses. Lower crime class numbers are more serious
Crime Code 2	626.0	May contain a code for an additional crime, less serious than Crime Code 1
Crime Code 3	626.0	May contain a code for an additional crime, less serious than Crime Code 1
Crime Code 4	626.0	May contain a code for an additional crime, less serious than Crime Code 1
Address	6300 BRYNHURST AV	Street address of crime incident rounded to the nearest hundred block to maintain anonymity
Location	(33.9829, -118.3338)	Latitude, Longitude of crime location

Initial exploration plots/figures/tables:

We are going to be looking at the variables in the Table above. In our initial exploration, we first built bar graphs and histograms to look at the most common values visually, to get an understanding of our data.



Works Cited

“Population by Sex & Age Los Angeles County.” Population by Sex and Age in Los Angeles County, California, www.laalmanac.com/population/po09.php. Accessed 11 Dec. 2023.

U.S. Census Bureau Quickfacts: Los Angeles City, California, www.census.gov/quickfacts/fact/table/losangelesciticacalifornia/PST045222. Accessed 12 Dec. 2023.