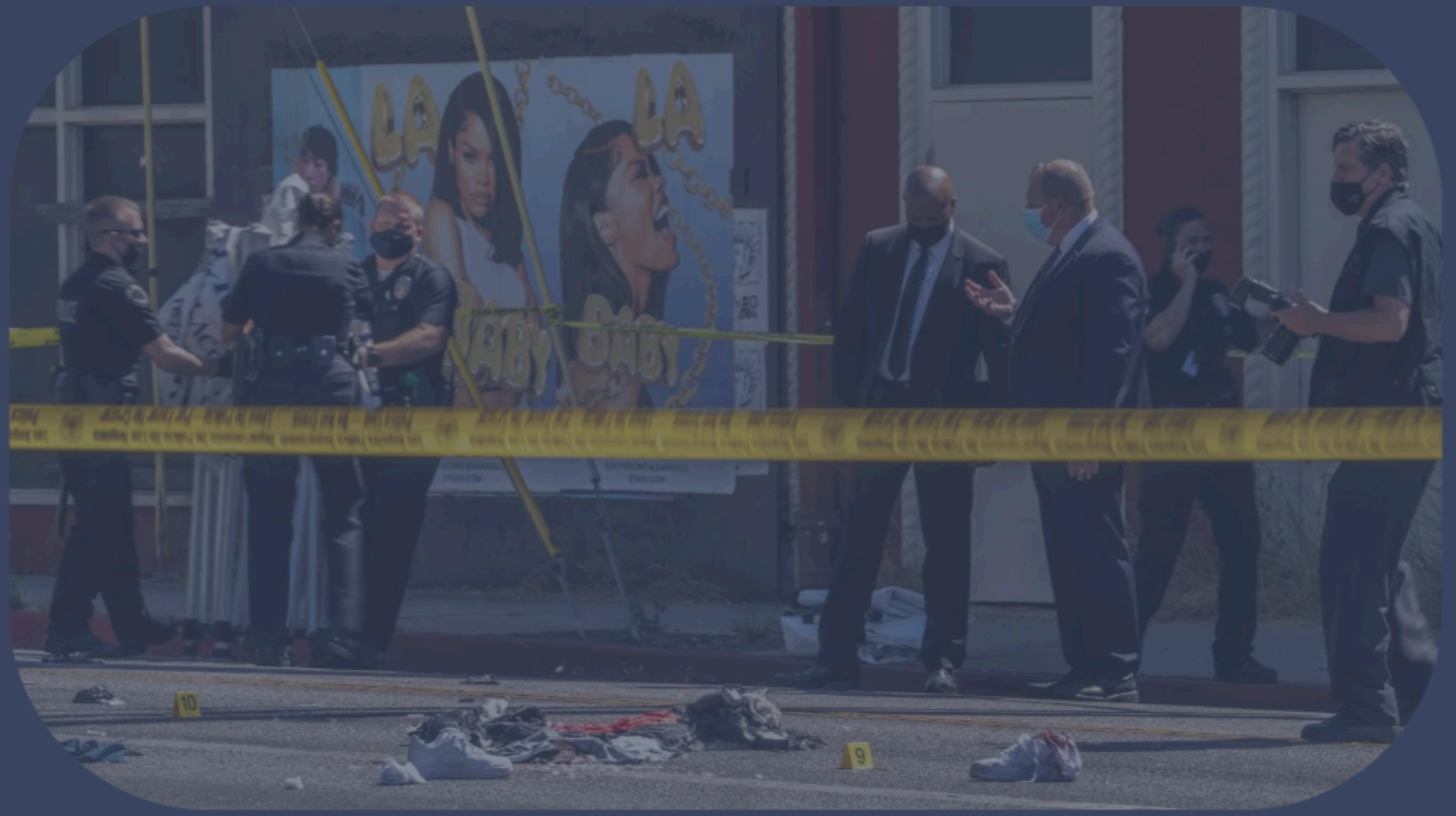
LA Crime Analysis



Presented by: Jason Warner
and Charles Hill

OVERVIEW

BACKGROUND

Business Problem

California Crime - Along with the rest of the country, California (specifically LA) has experienced high crime rates since the 2010's. Local law enforcements have struggled to maintain the resurgence, particularly in terms of allocating resources. By working with this dataset, we hope to analyze trends that can help agencies better predict the nature and volatility of future crime.

Purpose

Our objective is to find and understand relationships between some of the variables in our data set, and determine how we can use these relationships to make inferences about potential crime in California. We aim to apply these findings in a broader context to help agencies make more informed decisions about their allocation of resources.

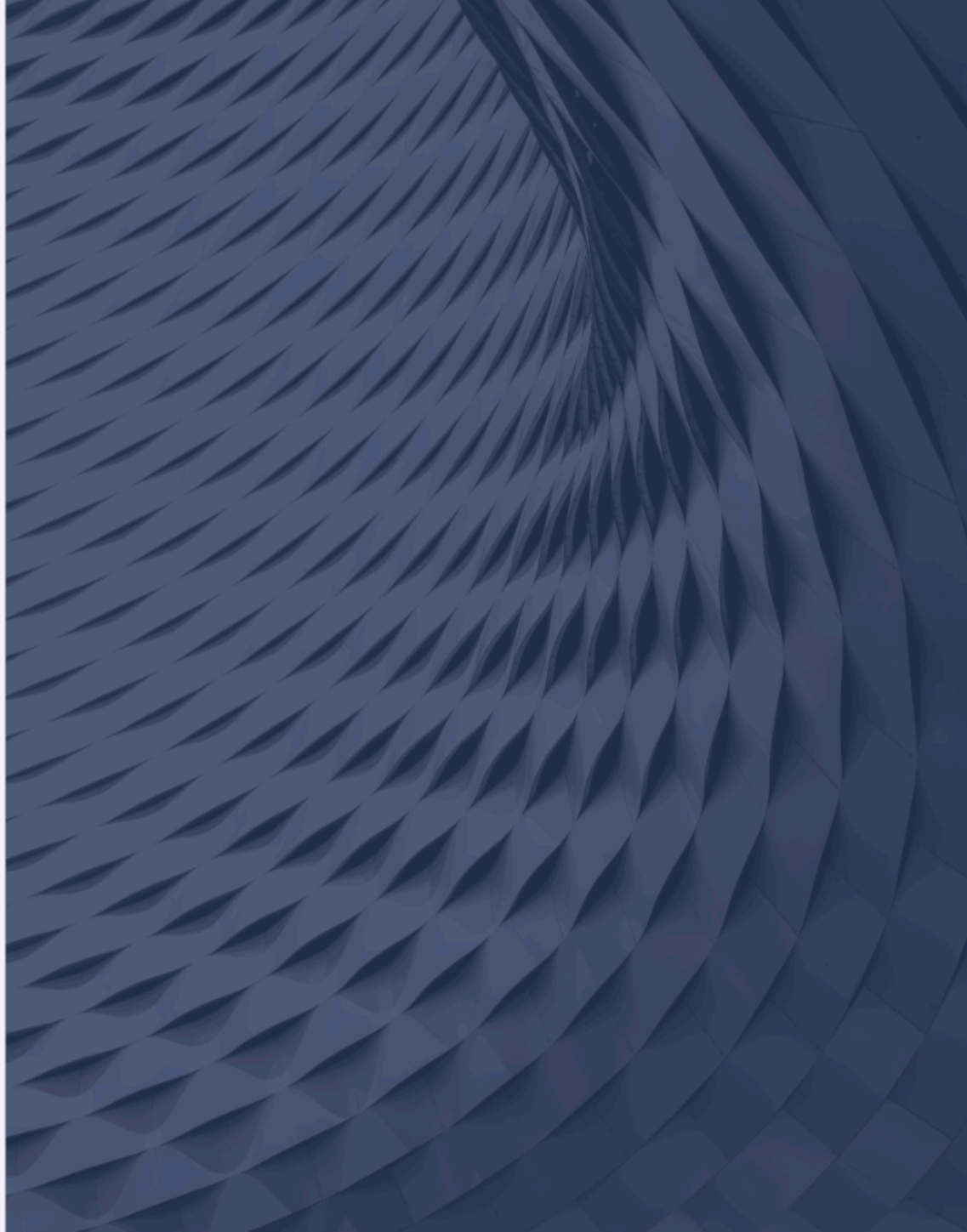
Project Overview

Hypothesis

We predict that the highest rates of crime occur in the later times of the day, and that more crime tends to occur in higher rates in more concentrated areas of the city. We also are hoping to see a connection between the status of the case and the severity of the crime.

Why?

After looking at our data set, we think that certain variables will show correlation between each other. Because of these connections, we could produce graphs that show these connections, and show the frequency of connections to show which are more likely. This will give us information need to form analysis from our data set.



Methodology Significance

DETERMINE THE MOST COMMONLY USED WEAPONS AGAINST SPECIFIC DEMOGRAPHICS

By associating weapons with certain crimes, agencies can better assess and prepare for threats

UNDERSTAND WHAT PREMISES AND AREAS ARE MORE LIKELY TO EXPERIENCE CERTAIN CRIMES

Identifying premises and areas with a higher likelihood of certain crimes helps in pinpointing crime hotspots

ANALYZE WHAT CRIMES ARE MOST LIKELY TO OCCUR AT CERTAIN TIMES PER DAY

Analyzing crime occurrence at different times per day helps identify times for targeted interventions

Our original data set had ~816,000 rows and 28 columns. The data consits of crime reports from the LA area 2020-Present.

Packages:

Pandas- Helps with the cleaning of our data set

Matplotlib.py- Creates the interactive visuals and charts.

Numpy- Creates arrays, and helps with computing

File importer- Used to import our data set into google collab

Original Data

	DR_NO	Date Rptd	DATE OCC	TIME OCC	AREA	AREA NAME	Rpt Dist No	Part 1-2	Crm Cd	Crm Cd Desc	...	Status	Status Desc	Crm Cd 1	Crm Cd 2	Crm Cd 3	Crm Cd 4	LOCATION	Cross Street	LAT	LON
0	10304468	01/08/2020 12:00:00 AM	01/08/2020 12:00:00 AM	2230	3	Southwest	377	2	624	BATTERY - SIMPLE ASSAULT	...	AO	Adult Other	624.0	NaN	NaN	NaN	1100 W 39TH PL	NaN	34.0141	-118.2978
1	190101086	01/02/2020 12:00:00 AM	01/01/2020 12:00:00 AM	330	1	Central	163	2	624	BATTERY - SIMPLE ASSAULT	...	IC	Invest Cont	624.0	NaN	NaN	NaN	700 S HILL ST	NaN	34.0459	-118.2545
2	200110444	04/14/2020 12:00:00 AM	02/13/2020 12:00:00 AM	1200	1	Central	155	2	845	SEX OFFENDER REGISTRANT OUT OF COMPLIANCE	...	AA	Adult Arrest	845.0	NaN	NaN	NaN	200 E 6TH ST	NaN	34.0448	-118.2474
3	191501505	01/01/2020 12:00:00 AM	01/01/2020 12:00:00 AM	1730	15	N Hollywood	1543	2	745	VANDALISM - MISDEAMEANOR ($399 OR UNDER)	...	IC	Invest Cont	745.0	998.0	NaN	NaN	5400 CORTEEN PL	NaN	34.1685	-118.4019
4	191921269	01/01/2020 12:00:00 AM	01/01/2020 12:00:00 AM	415	19	Mission	1998	2	740	VANDALISM - FELONY ($400 & OVER, ALL CHURCH VA...	...	IC	Invest Cont	740.0	NaN	NaN	NaN	14400 TITUS ST	NaN	34.2198	-118.4468
5	200100501	01/02/2020 12:00:00 AM	01/01/2020 12:00:00 AM	30	1	Central	163	1	121	RAPE, FORCIBLE	...	IC	Invest Cont	121.0	998.0	NaN	NaN	700 S BROADWAY	NaN	34.0452	-118.2534
6	200100502	01/02/2020 12:00:00 AM	01/02/2020 12:00:00 AM	1315	1	Central	161	1	442	SHOPLIFTING - PETTY THEFT ($950 & UNDER)	...	IC	Invest Cont	442.0	998.0	NaN	NaN	700 S FIGUEROA ST	NaN	34.0483	-118.2631
7	200100504	01/04/2020 12:00:00 AM	01/04/2020 12:00:00 AM	40	1	Central	155	2	946	OTHER MISCELLANEOUS CRIME	...	IC	Invest Cont	946.0	998.0	NaN	NaN	200 E 6TH ST	NaN	34.0448	-118.2474
8	200100507	01/04/2020 12:00:00 AM	01/04/2020 12:00:00 AM	200	1	Central	101	1	341	THEFT-GRAND ($950.01 & OVER)EXCPT,GUNS,FOWL,LI...	...	IC	Invest Cont	341.0	998.0	NaN	NaN	700 BERNARD ST	NaN	34.0677	-118.2398
9	201710201	06/19/2020 12:00:00 AM	05/26/2020 12:00:00 AM	1925	17	Devonshire	1708	1	341	THEFT-GRAND ($950.01 & OVER)EXCPT,GUNS,FOWL,LI...	...	AO	Adult Other	341.0	NaN	NaN	NaN	11900 BALBOA BL	NaN	34.2864	-118.5021

Cleaned Data

	Case Number	Date Reported	Date Occured	Time of Crime	Crime Code	Crime Description	Victim Age	Victim Sex	Victim Race	Premis Code	Premis Description	Weapon Code	Weapon Description	Status	Status Desc	Latitude	Longitude
0	10304468	2020-01-08	2020-01-08	2230	624	BATTERY - SIMPLE ASSAULT	36	F	B	501.0	SINGLE FAMILY DWELLING	400.0	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AO	Adult Other	34.0141	-118.2978
1	190101086	2020-01-02	2020-01-01	330	624	BATTERY - SIMPLE ASSAULT	25	M	H	102.0	SIDEWALK	500.0	UNKNOWN WEAPON/OTHER WEAPON	IC	Invest Cont	34.0459	-118.2545
2	200110444	2020-04-14	2020-02-13	1200	845	SEX OFFENDER REGISTRANT OUT OF COMPLIANCE	0	X	X	726.0	POLICE FACILITY	NaN	NaN	AA	Adult Arrest	34.0448	-118.2474
3	191501505	2020-01-01	2020-01-01	1730	745	VANDALISM - MISDEAMEANOR ($399 OR UNDER)	76	F	W	502.0	MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)	NaN	NaN	IC	Invest Cont	34.1685	-118.4019
4	191921269	2020-01-01	2020-01-01	415	740	VANDALISM - FELONY ($400 & OVER, ALL CHURCH VA...	31	X	X	409.0	BEAUTY SUPPLY STORE	NaN	NaN	IC	Invest Cont	34.2198	-118.4468
5	200100501	2020-01-02	2020-01-01	30	121	RAPE, FORCIBLE	25	F	H	735.0	NIGHT CLUB (OPEN EVENINGS ONLY)	500.0	UNKNOWN WEAPON/OTHER WEAPON	IC	Invest Cont	34.0452	-118.2534
6	200100502	2020-01-02	2020-01-02	1315	442	SHOPLIFTING - PETTY THEFT ($950 & UNDER)	23	M	H	404.0	DEPARTMENT STORE	NaN	NaN	IC	Invest Cont	34.0483	-118.2631
7	200100504	2020-01-04	2020-01-04	40	946	OTHER MISCELLANEOUS CRIME	0	X	X	726.0	POLICE FACILITY	NaN	NaN	IC	Invest Cont	34.0448	-118.2474
8	200100507	2020-01-04	2020-01-04	200	341	THEFT-GRAND ($950.01 & OVER)EXCPT,GUNS,FOWL,LI...	23	M	B	502.0	MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)	NaN	NaN	IC	Invest Cont	34.0677	-118.2398
9	201710201	2020-06-19	2020-05-26	1925	341	THEFT-GRAND ($950.01 & OVER)EXCPT,GUNS,FOWL,LI...	0	X	X	203.0	OTHER BUSINESS	NaN	NaN	AO	Adult Other	34.2864	-118.5021

Edits: Dropping irrelevant columns, renaming for readability, changing DateTime formats, etc.

Variables

TIME OF CRIME

When crime occured

CRIME CODE / DESCRIPTION

Code identifying what type of crime occurred

VICTIM AGE/SEX/RACE

Demographics

PREMIS CODE/DESCRIPTION

Code identifying what type of location the crime occured

WEAPON CODE/DESCRIPTION

Code identifying what type of weapon was used

STATUS

Investigation Ongoing, Arrest, Other

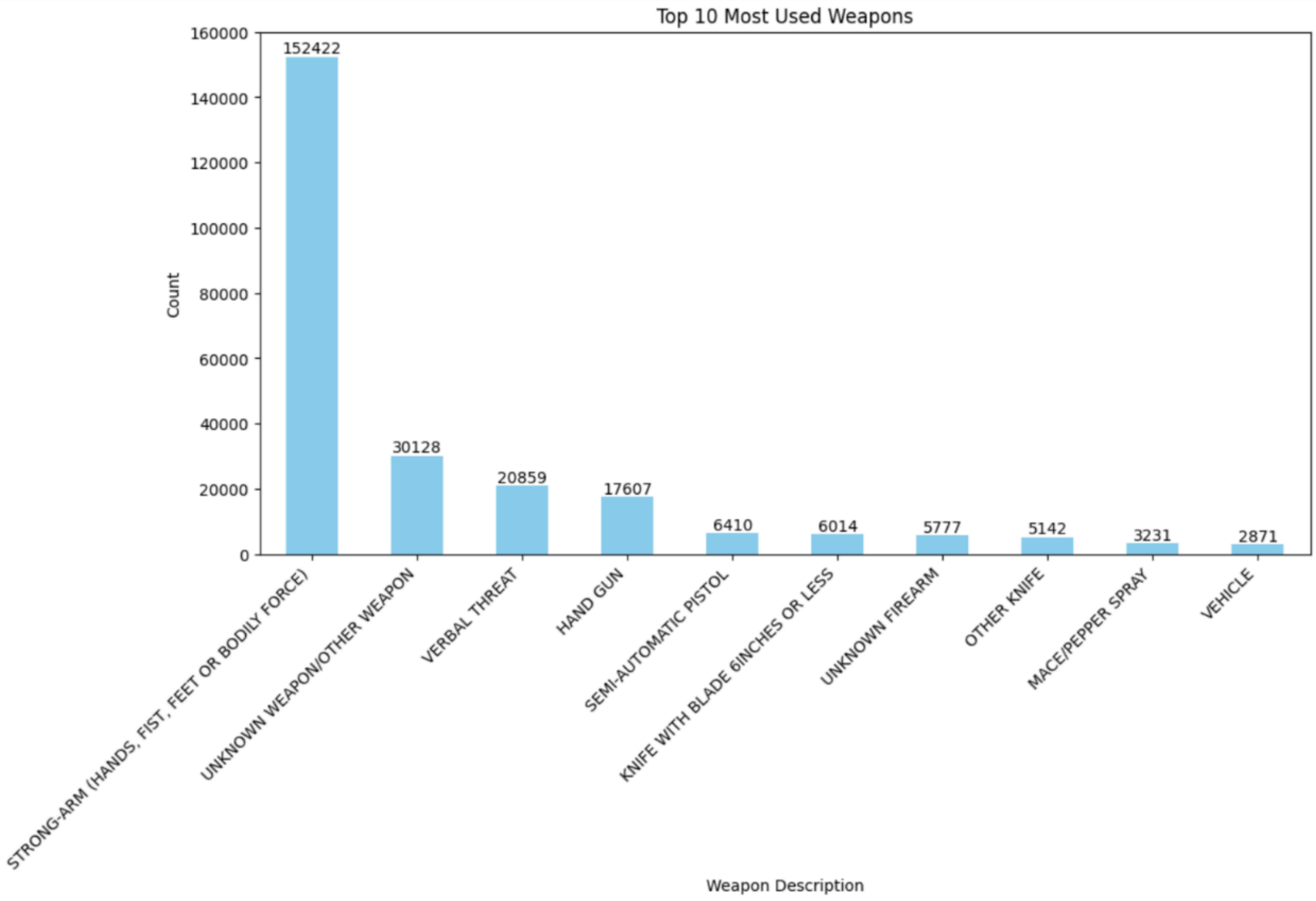
DATE OCCURED

When initial crime occured

LONGITUDE / LATITUTDE

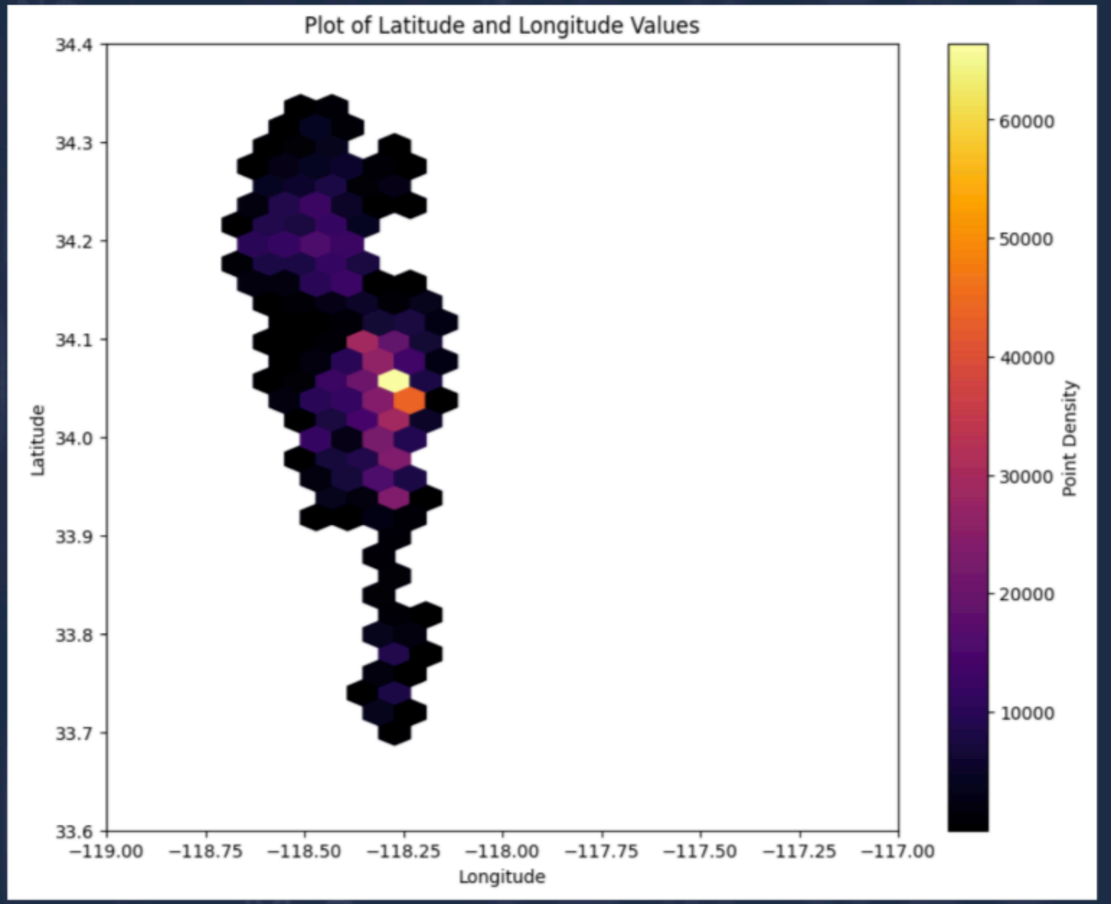
Relative Location

Summary


Top 10 Most Used Weapons
Count
Weapon Description
160000
140000
120000
100000
80000
60000
40000
20000
0
152422
30128
20859
17607
6410
6014
5777
5142
3231
2871
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)
UNKNOWN WEAPON/OTHER WEAPON
VERBAL THREAT
HAND GUN
SEMI-AUTOMATIC PISTOL
KNIFE WITH BLADE 6INCHES OR LESS
UNKNOWN FIREARM
OTHER KNIFE
MACE/PEPPER SPRAY
VEHICLE

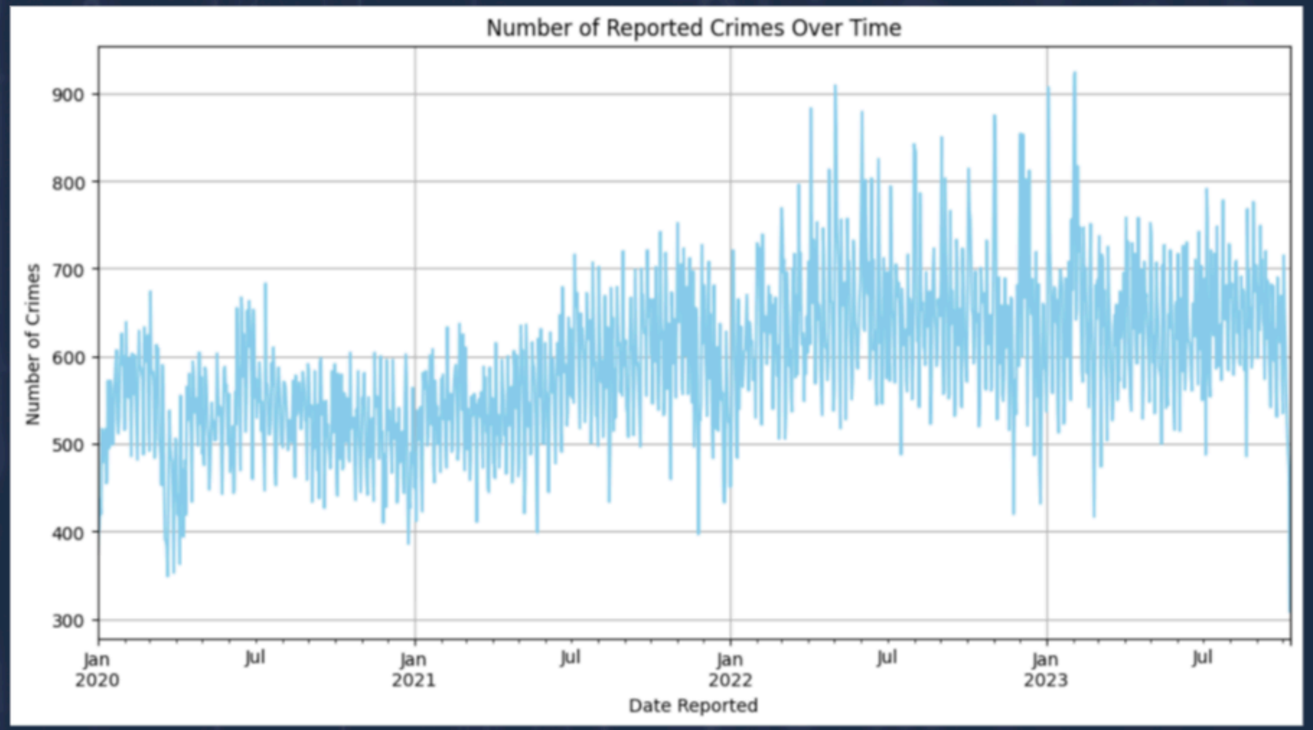
- This heat map took the range of our longitude and latitude to show the frequency of each relativelongitude and latitude.
- From this graph we attempted to find out the locations where crimes were most frequent. This too into consideration any crime, and solely relied on the location of crime.
- This helps our final analysis because we are able to connect crimes committed towards locations.

```
plt.figure(figsize=(10, 8))
LaMap = [-119, -117, 33.6, 34.4]
plt.hexbin(df['Longitude'], df['Latitude'], gridsize=1500, cmap='inferno', mincnt=1)
plt.colorbar(label='Point Density')
plt.xlabel('Longitude')
plt.ylabel('Latitude')
plt.title('Plot of Latitude and Longitude Values')
plt.xlim(LaMap[0], LaMap[1])
plt.ylim(LaMap[2], LaMap[3])
plt.show()
```

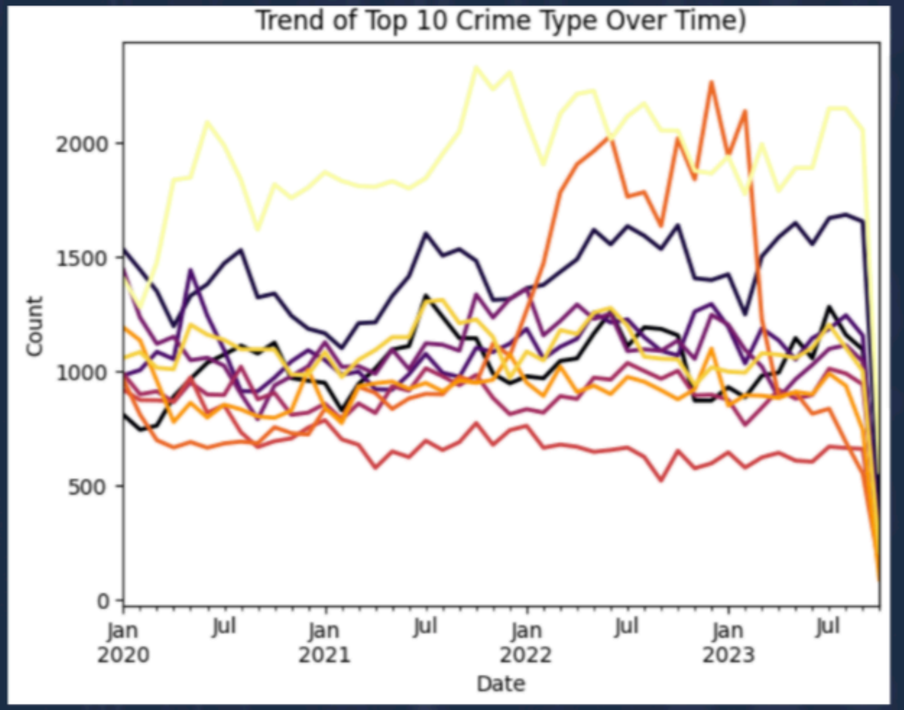


- This bar chart analyzes the trend between number of crimes reported and the Date in which the crimes where reporting
- The main takeaways from this analysis were a substantial dip during the onset of the pandemic, followed by a higher trend as the year went on.
- It has since increased to a somewhat steady rate of ~650 crimes reported per day in the greater LA area

```
df['Date Reported'] = pd.to_datetime(df['Date Reported'])
crime_counts_by_date = df.groupby('Date Reported').size()
plt.figure(figsize=(12, 6))
crime_counts_by_date.plot(kind='line', color='skyblue')
plt.title('Number of Reported Crimes Over Time')
plt.xlabel('Date Reported')
plt.ylabel('Number of Crimes')
plt.grid(True)
plt.show()
```



- We wanted to find the correlation from the crimes committed and their occurrences over time. We thought this could be important to see if certain crimes had timetables on where they were more frequent and may not be relevant at this time.
- We would be looking for the crimes that trending in the positive direction, as they would be the ones we need to attempt to limit as there are becoming more of an issue.
- Located these trends could allow authorities to be on the look out for certain crimes, that we show are occurring more frequently than previous times.




```
df['Date Occured'] = pd.to_datetime(df['Date Occured'])

top_n_crimes = 10
top_crimes = df['Crime Description'].value_counts().head(top_n_crimes).index

df_filtered = df[df['Crime Description'].isin(top_crimes)]

crime_date_counts = df_filtered.groupby(['Date Occured', 'Crime Description']).size().unstack()

plt.figure(figsize=(14, 8))
resampled_data.plot(kind='line', colormap='inferno', lw=2)

plt.title(f'Trend of Top {top_n_crimes} Crime Type Over Time)')
plt.xlabel('Date')
plt.ylabel('Count')
plt.legend(title='Crime Description', bbox_to_anchor=(1, 1), loc='upper left')
plt.show()
```

Findings

- While we still have more relationships to analyze, so far we have discovered that a significant amount of crime is centered within a very specific longitude and latitude.
- We have also determined that the majority of crimes committed involve the use of bodily-force, and crimes tend to happen later in the day.
- We are still working to establish stronger conclusions, but even now, we believe agencies can take more preventative approaches with handling crime in their cities by allocating their resources to where crime has a higher probability of occurring.

Moving Forward

While we have made substantial progress in our initial data analysis, there are still many more relationships we would like the understand the relationship between. Some of these include:

TIME AND TYPE OF CRIME

Does the time of day affect the likelihood that certain weapons are used?

AREA AND SEVERITY

Which areas of LA see the most crimes, and which types?

STATUS AND TYPE OF CRIME

What types of crime see the highest arrest rate?

TIME AND PREMIS

Are certain locations more likely to be a target at certain times of day?