Unveiling Patterns: An In-depth Data Analysis of Arrests and Crime in Chicago, IL

June 20, 2023

Jessica Andras
Saber Garibi
Andrew Rexford
Dennys Urdiales
Jennifer Alvarez

Agenda

- Purpose and Background
  - About our Data
- Questions Worth Asking
- Analysis 1
  - Chicago Districts
- Analysis 2
  - Fatalities
- Analysis 3
  - Cause of Arrest
- Limitations
- Interesting Takeaways & Things to Consider for Future Work

Purpose and Background

Purpose and Background

The objective of this project is to conduct a comprehensive data analysis of arrest and crime data in the city of Chicago, IL.

By examining the available datasets, the project aims to provide valuable insights into patterns, trends, and correlations related to arrests and crimes within the city.

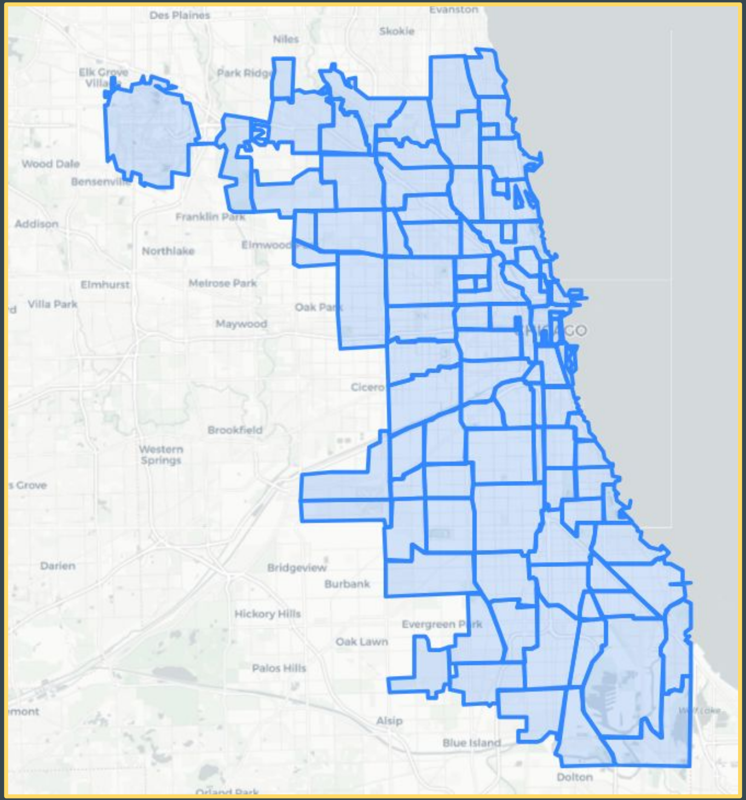


Image Source: Chicago Police Department

About Our Data

Resources/PublicReleaseArrestDataUPDATE.csv

ARR_DISTRICT	The Chicago Police district (geographic boundary) wherein the arrest took place
ARR_BEAT	The Chicago Police beat (geographic boundary) wherein the arrest took place
ARR_YEAR	The year in which the arrest took place
ARR_MONTH	The month in which the arrest took place
RACE_CODE_CD	The perceived race of the individual arrested
FBI_CODE	The crime type code (category) reported to the FBI conforming to IUCR standards
STATUTE	The specific statute or city ordinance which the person was charged with violating
STAT_DESCR	The plain text description (title) of the statute or ordinance which the person was charged with violating
CHARGE_CLASS_CD	Code representing the class of the charge in accordance with Illinois Compiled Statutes (ILCS) or Municipal Code of Chicago (MCC)
CHARGE_TYPE_CD	Code representing the level of the charge. “M” wherein the charge was a Misdemeanor. “F” wherein the charge was a Felony. Both in accordance with Illinois Compiled Statutes (ILCS).

Range of Data: 2014 - 2017

https://home.chicagopolice.org/statistics-data/public-arrest-data/

About Our Data

Resources/PublicReleaseArrestDataUPDATE.csv

ARR_DISTRICT	ARR_BEAT	ARR_YEAR	ARR_MONTH	RACE_CODE_CD	FBI_CODE	STATUTE	STAT_DESCR
10	1033	2017	8	BLK	18	720 ILCS 570.0/407-B-1	MFG/DEL COCAINE/SCH/PUB HS/PK
9	923	2017	8	WWH	WRT	725 ILCS 225.0/13	FUGITIVE FROM JUSTICE - OUT OF STATE WARRANT
10	1024	2017	8	BLK	WRT	725 ILCS 5.0/110-3	ISSUANCE OF WARRANT
11	1112	2017	8	BLK	18	720 ILCS 570.0/407-B-1	MFG/DEL HEROIN/SCH/PUB HS/PK
25	2524	2017	8	WHI	18	720 ILCS 570.0/402-C	PCS - POSSESSION - POSS AMT CON SUB EXCEPT (A)(D)
11	1122	2017	9	BLK	7	720 ILCS 5.0/21-2-A	CRIMINAL TRESPASS TO VEHICLES
16	1655	2017	9	BLK	26	720 ILCS 5.0/21-5-A	CRIM TRESPASS TO STATE LAND
24	2411	2017	8	WWH	26	720 ILCS 5.0/21-3-A-2	CRIMINAL TRESPASS TO LAND
24	2411	2017	8	WWH	26	720 ILCS 5.0/21-3-A-2	CRIMINAL TRESPASS TO LAND
24	2411	2017	8	BLK	26	720 ILCS 5.0/21-3-A-2	CRIMINAL TRESPASS TO LAND
11	1132	2017	8	BLK	18	720 ILCS 570.0/407-B-1	MFG/DEL HEROIN/SCH/PUB HS/PK
11	1111	2017	8	BLK	18	720 ILCS 570.0/402-C	PCS - POSSESSION - POSS AMT CON SUB EXCEPT (A)(D)
1	123	2017	8	BLK	6	720 ILCS 5.0/16-25-A-1	RETAIL THEFT/DISP MERCH/<$300
11	1121	2017	8	BLK	18	720 ILCS 570.0/402-C	PCS - POSSESSION - POSS AMT CON SUB EXCEPT (A)(D)
10	1022	2017	9	BLK	08B	720 ILCS 5.0/12-3.2-A-1	DOMESTIC BATTERY - BODILY HARM
3	311	2017	9	BLK	14	720 ILCS 5.0/21-1-A-1	CRIM DAMAGE TO PROPERTY <$300
		2017	9	BLK	26	720 ILCS 5.0/21-3-A-2	CRIMINAL TRESPASS TO LAND
5	523	2017	8	BLK	TRF	625 ILCS 5.0/4-104-A-4	POSS TITLE/REGISTRATION NOT AUTHORIZED ON VEHICLE
24	2423	2017	8	WHI	26	720 ILCS 5.0/21-3-A-2	CRIMINAL TRESPASS TO LAND
5	511	2017	9	BLK	TRF	625 ILCS 5.0/4-104-A-4	POSS TITLE/REGISTRATION NOT AUTHORIZED ON VEHICLE
25	2512	2017	8	BLK	TRF	625 ILCS 5.0/4-104-A-4	POSS TITLE/REGISTRATION NOT AUTHORIZED ON VEHICLE
11	1121	2017	8	BLK	26	720 ILCS 5.0/33A-2-A	ARMED VIOLENCE/CATEGORY I
17	1733	2017	8	WWH	21	625 ILCS 5.0/11-501-A	IVC - AGG DUI/LIC SUSP OR REVOKED
5	511	2017	8	BLK	08B	720 ILCS 5.0/12-3.2-A-2	DOMESTIC BATTERY - PHYSICAL CONTACT
7	711	2017	8	BLK	MCC	8-26-020-A	GUN OFFENDER DUTY TO REGISTER AND TO VERITY

Resources/PublicReleaseArrestDataUPDATE.csv

About Our Data

Resources/Crimes_2015toPresent.csv

Column Name	Description
ID	Unique identifier for the record.
Case Number	The Chicago Police Department RD Number (Records Division ...
Date	Date when the incident occurred. this is sometimes a best esti...
Block	The partially redacted address where the incident occurred, pl...
IUCR	The Illinois Unifrom Crime Reporting code. This is directly linke...
Primary Type	The primary description of the IUCR code.
Description	The secondary description of the IUCR code, a subcategory of ...
Location Description	Description of the location where the incident occurred.
Arrest	Indicates whether an arrest was made.
Domestic	Indicates whether the incident was domestic-related as define...
Beat	Indicates the beat where the incident occurred. A beat is the s...
District	Indicates the police district where the incident occurred. See t...
Ward	The ward (City Council district) where the incident occurred. S...
Community Area	Indicates the community area where the incident occurred. Chi...

Updated On	Date and time the record was last updated.
Latitude	The latitude of the location where the incident occurred. This l...
Longitude	The longitude of the location where the incident occurred. This...
Location	The location where the incident occurred in a format that allo...
Historical Wards 2003-2015	
Zip Codes	
Community Areas	
Census Tracts	
Wards	
Boundaries - ZIP Codes	
Police Districts	
Police Beats	

https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2

About Our Data

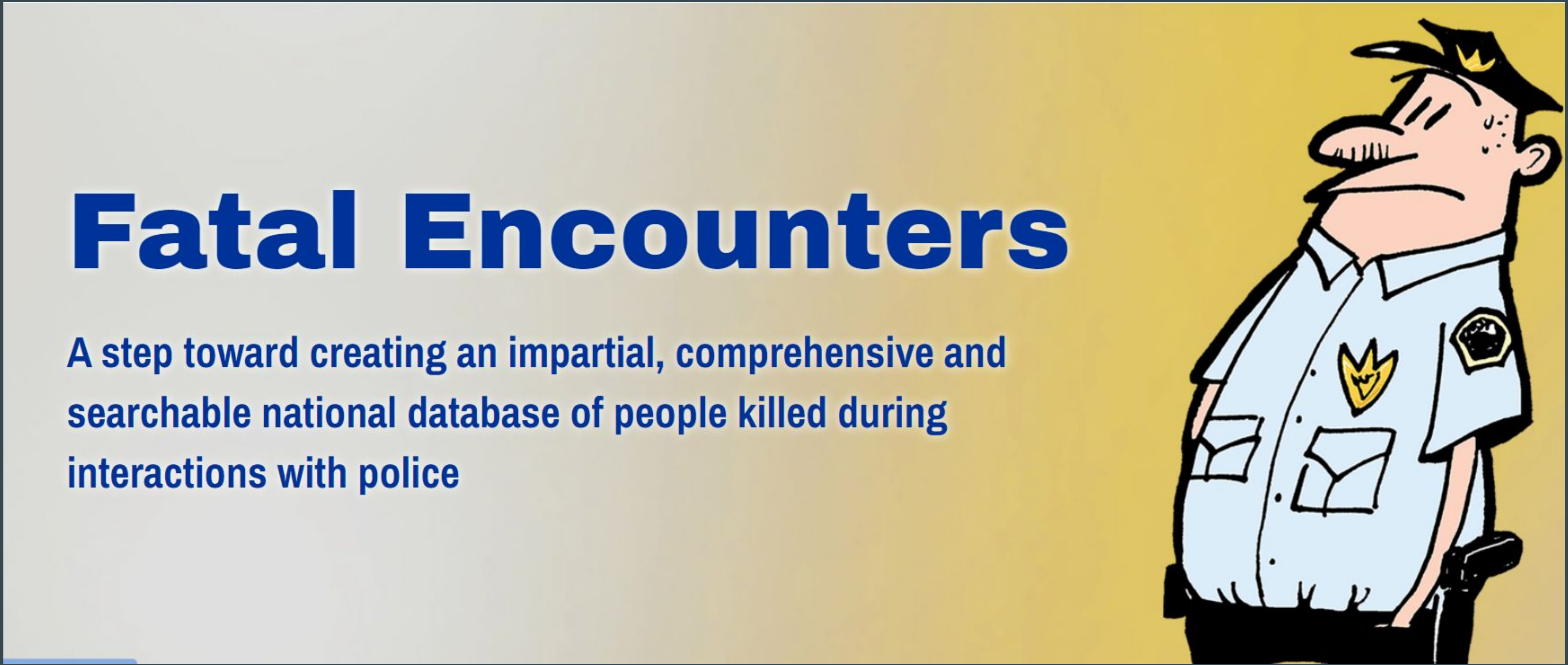
Resources/Crimes_2015toPresent.csv

District	Primary Type	Arrest	Date	IUCR	Description	Location Description	FBI Code
6	WEAPONS VIOLATION	TRUE	4/1/2023 2:00	143A	UNLAWFUL POSSESSION - HANDGUN	SIDEWALK	15
6	WEAPONS VIOLATION	TRUE	4/3/2023 21:12	143A	UNLAWFUL POSSESSION - HANDGUN	STREET	15
6	WEAPONS VIOLATION	TRUE	4/1/2023 20:54	1460	POSSESS FIREARM / AMMUNITION - NO FOID CARD	STREET	15
6	WEAPONS VIOLATION	TRUE	4/1/2023 20:20	143A	UNLAWFUL POSSESSION - HANDGUN	STREET	15
6	WEAPONS VIOLATION	TRUE	3/31/2023 23:50	143A	UNLAWFUL POSSESSION - HANDGUN	RESIDENCE	15
6	WEAPONS VIOLATION	TRUE	3/30/2023 15:59	143A	UNLAWFUL POSSESSION - HANDGUN	STREET	15
7	ASSAULT	TRUE	4/1/2023 18:54	520	AGGRAVATED - KNIFE / CUTTING INSTRUMENT	RESIDENCE	04A
7	ASSAULT	TRUE	3/30/2023 8:18	560	SIMPLE	APARTMENT	08A
17	NARCOTICS	TRUE	4/2/2023 6:15	2091	FORFEIT PROPERTY	RESIDENCE - YARD (FRO	18
17	NARCOTICS	TRUE	4/2/2023 23:30	1811	POSSESS - CANNABIS 30 GRAMS OR LESS	STREET	18
17	OTHER OFFENSE	TRUE	4/2/2023 20:10	502P	FALSE / STOLEN / ALTERED TRP	VEHICLE NON-COMMER	26
17	OTHER OFFENSE	TRUE	4/1/2023 21:40	502P	FALSE / STOLEN / ALTERED TRP	STREET	26
6	THEFT	TRUE	3/31/2023 11:40	860	RETAIL THEFT	DRUG STORE	6
16	WEAPONS VIOLATION	TRUE	3/31/2023 13:52	143A	UNLAWFUL POSSESSION - HANDGUN	AIRPORT TERMINAL UPI	15
7	BATTERY	TRUE	4/1/2023 17:53	460	SIMPLE	RESIDENCE	08B
7	BATTERY	TRUE	4/1/2023 11:15	486	DOMESTIC BATTERY SIMPLE	RESIDENCE	08B
7	CONCEALED CARRY LICENSE VIOLATION	TRUE	4/2/2023 16:10	1480	OTHER	STREET	15
17	BATTERY	TRUE	4/1/2023 21:44	486	DOMESTIC BATTERY SIMPLE	RESIDENCE	08B
18	ASSAULT	TRUE	4/2/2023 15:24	560	SIMPLE	DEPARTMENT STORE	08A
6	CRIMINAL TRESPASS	TRUE	4/1/2023 19:25	1360	TO VEHICLE	STREET	26
6	DECEPTIVE PRACTICE	TRUE	4/1/2023 15:20	1210	THEFT OF LABOR / SERVICES	CTA PLATFORM	11
6	INTERFERENCE WITH PUBLIC OFFICER	TRUE	4/1/2023 16:40	3731	OBSTRUCTING IDENTIFICATION	STREET	24
6	INTERFERENCE WITH PUBLIC OFFICER	TRUE	3/31/2023 18:38	3710	RESIST / OBSTRUCT / DISARM OFFICER	STREET	24
6	OTHER OFFENSE	TRUE	3/25/2023 23:55	5011	LICENSE VIOLATION	SMALL RETAIL STORE	26
6	NARCOTICS	TRUE	3/30/2023 23:05	1812	POSSESS - CANNABIS MORE THAN 30 GRAMS	CTA PLATFORM	18

https://github.com/jdmandras/Project-1/blob/main/Resources/Crimes_2015toPresent.csv

About Our Data

Resources/Fatal_complete.csv



https://fatalencounters.org/

About Our Data

Resources/Fatal_complete.csv

Unique ID	Name	Age	Gender	Race	Race with imputations	Imputation probability	URL of image (PLS NO HOTLINKS)	Date of injury resulting in death (month/day/year)	Loc
31495	Ashley Mc	28	Female	African-Ar	African-American/Black	Not imputed	https://fatalencounters.org/wp-con	12/31/2021	Sou
31496	Name withheld by pc		Female	Race unsp	NA	NA		12/31/2021	150
31497	Name withheld by pc		Male	Race unsp	NA	NA		12/31/2021	150
31491	Johnny C.	36	Male	Race unsp	NA	NA		12/30/2021	Ma
31492	Dennis Mc	44	Male	European-	NA	NA		12/30/2021	435
31493	Ny'Darius	21	Male	Race unsp	NA	NA		12/30/2021	Sta
31494	Timothy E	50	Male	European-	European-American/Whit	Not imputed	https://fatalencounters.org/wp-con	12/30/2021	Syk
31409	Name withheld by pc		Male	Hispanic/L	Hispanic/Latino	Not imputed		12/29/2021	Car
31410	Name withheld by pc		Female	Hispanic/L	Hispanic/Latino	Not imputed		12/29/2021	Car
31465	Christophe	49		African-Ar	African-American/Black	Not imputed	https://fatalencounters.org/wp-con	12/29/2021	152
31466	Osman Se	27	Male	African-Ar	African-American/Black	Not imputed		12/29/2021	Dar
31490	Theloniou	25	Male	African-Ar	African-American/Black	Not imputed	https://fatalencounters.org/wp-con	12/29/2021	99
31464	Dwayne M	62	Male	African-Ar	African-American/Black	Not imputed	https://fatalencounters.org/wp-con	12/28/2021	35
31408	Robert Mi	53	Male	European-	European-American/Whit	Not imputed	https://fatalencounters.org/wp-con	12/27/2021	100
31459	Christophe	22	Male	Race unsp	NA	NA		12/27/2021	200
31460	Kevin Dub	31	Male	Race unsp	NA	NA		12/27/2021	628
31461	Lyndon Ja	47	Male	European-	European-American/Whit	Not imputed	https://fatalencounters.org/wp-con	12/27/2021	S V
31462	Name withheld by pc		Male	Race unsp	NA	NA		12/27/2021	200
31463	Roberto C	39	Male	Race unsp	NA	NA		12/27/2021	E M
31407	Stanley "S.	13	Male	African-Ar	African-American/Black	Not imputed	https://fatalencounters.org/wp-con	12/26/2021	800
31453	Cesar Juar	34	Male	Race unsp	NA	NA		12/26/2021	Air
31454	Christophe	41	Male	Race unsp	NA	NA		12/26/2021	350
31455	Enrique Ru	33	Male	Race unsp	NA	NA		12/26/2021	290
31456	James Low	40	Male	African-Ar	African-American/Black	Not imputed	https://fatalencounters.org/wp-con	12/26/2021	150

Date Range: 2015-2021

https://github.com/jdmandras/Project-1/blob/main/Resources/Fatal_complete.csv

Questions Worth Asking

Questions Worth Asking

- Which districts have the highest number of arrests?
- What are the top 5 Zip Codes with the highest number of fatalities in the City of Chicago, and how have fatalities trends in these areas changed from 2015-2021?
- What is the breakdown of the most/least frequently violated laws and statutes, and which charges are the most commonly filed?
- What are the Common Causes of Arrest?

Analysis 1: Chicago Districts

Chicago Districts

Highest Crime to Highest Arrest (2015-2022)

```
# Most common district overall
most_common_district = data['District'].value_counts().idxmax()

# Filter data where Arrests is True
data_arrests_true = data[data['Arrest'] == True]

# Most common district with Arrests as True
most_common_district_arrests_true = data_arrests_true['District'].value_counts().idxmax()

print("Highest Recorded Crime Distict:", most_common_district)
print("District that ends the most in Arrest:", most_common_district_arrests_true)
```

✓ 0.0s

```
Highest Recorded Crime Distict: 25.0
District that ends the most in Arrest: 11.0
```

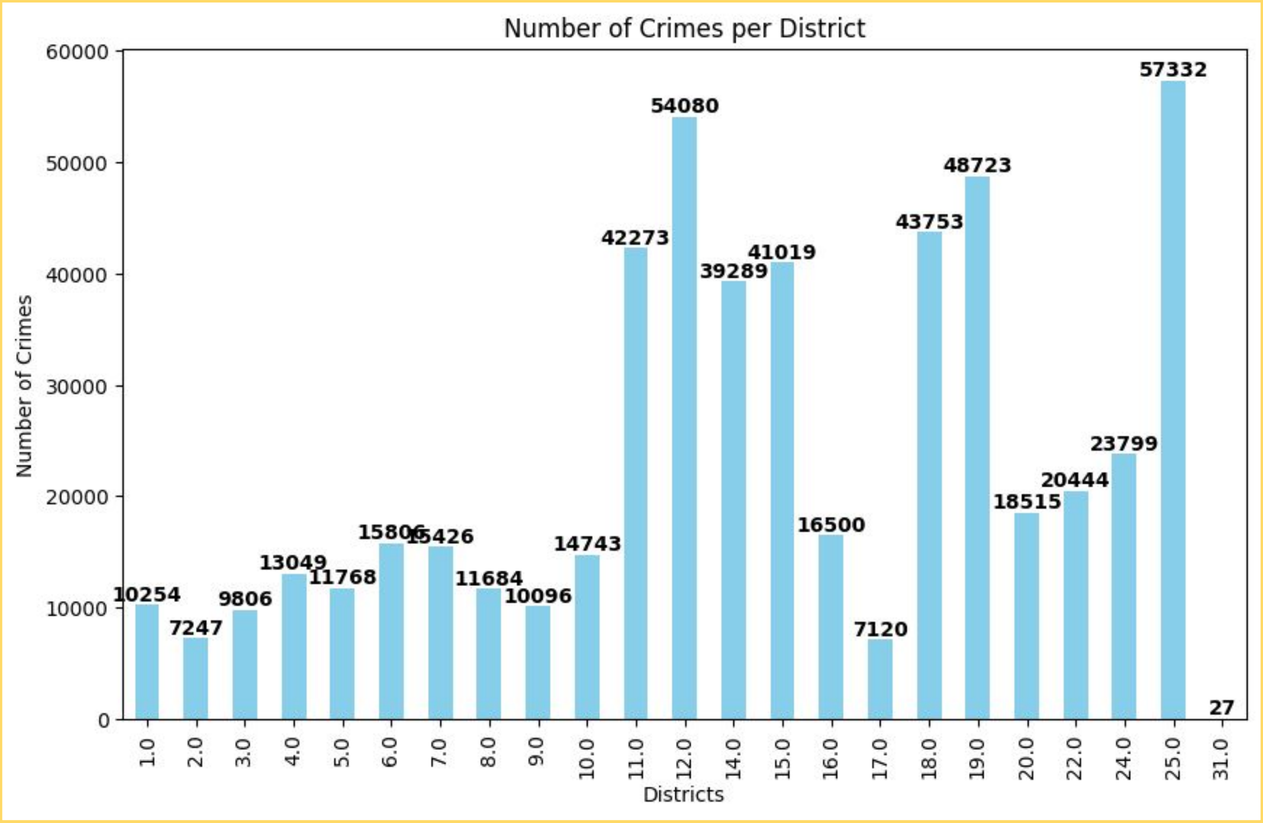
Chicago Districts: Number of Recorded Crime Per District

```
import pandas as pd
import matplotlib.pyplot as plt

# Read the CSV file
data = pd.read_csv('Resources/Crimes_2015toPresent.csv')

# Group the data by District and count the number of rows (crimes) in each district
district_counts = data['District'].value_counts().sort_index()

# plot
plt.figure(figsize=(10, 6))
ax = district_counts.plot(kind='bar', color='skyblue')
plt.xlabel('Districts')
plt.ylabel('Number of Crimes')
plt.title('Number of Crimes per District')

# Labeling each bar with the number of crimes
for i, v in enumerate(district_counts):
    ax.text(i, v, str(v), ha='center', va='bottom', fontweight='bold')

# display plot
plt.show()
```

Analysis1.ipynb

Chicago Districts

Number of Recorded Crime Per District



Analysis1.ipynb

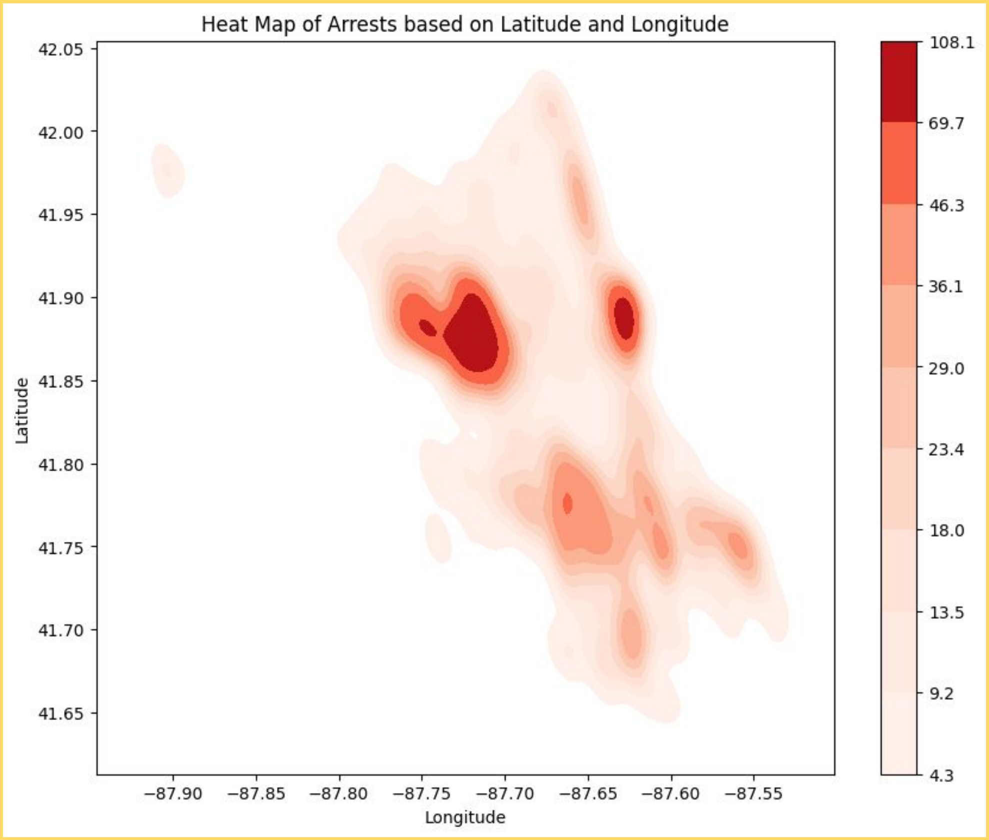
Making Heat Map of Arrests Made

```
# Filter data where Arrest is True and remove null/invalid values from Arrest, Latitude, and Longitude columns
arrest_data = data.dropna(subset=['Arrest', 'Latitude', 'Longitude'])
arrest_data = arrest_data[arrest_data['Arrest'] == True]

# The errors='coerce' argument tells the pd.to_numeric() function to convert any non-numeric values to NaN (Not a Number) instead of raising an error.
arrest_data[['Latitude', 'Longitude']] = arrest_data[['Latitude', 'Longitude']].apply(pd.to_numeric, errors='coerce')
arrest_data = arrest_data.dropna(subset=['Latitude', 'Longitude'])

# Set up the plot
plt.figure(figsize=(10, 8))
sns.kdeplot(data=arrest_data, x='Longitude', y='Latitude', cmap='Reds', fill=True, cbar=True)

# Set labels and title
plt.xlabel('Longitude')
plt.ylabel('Latitude')
plt.title('Heat Map of Arrests based on Latitude and Longitude')

plt.show()
```

✓ 5m 48.5s

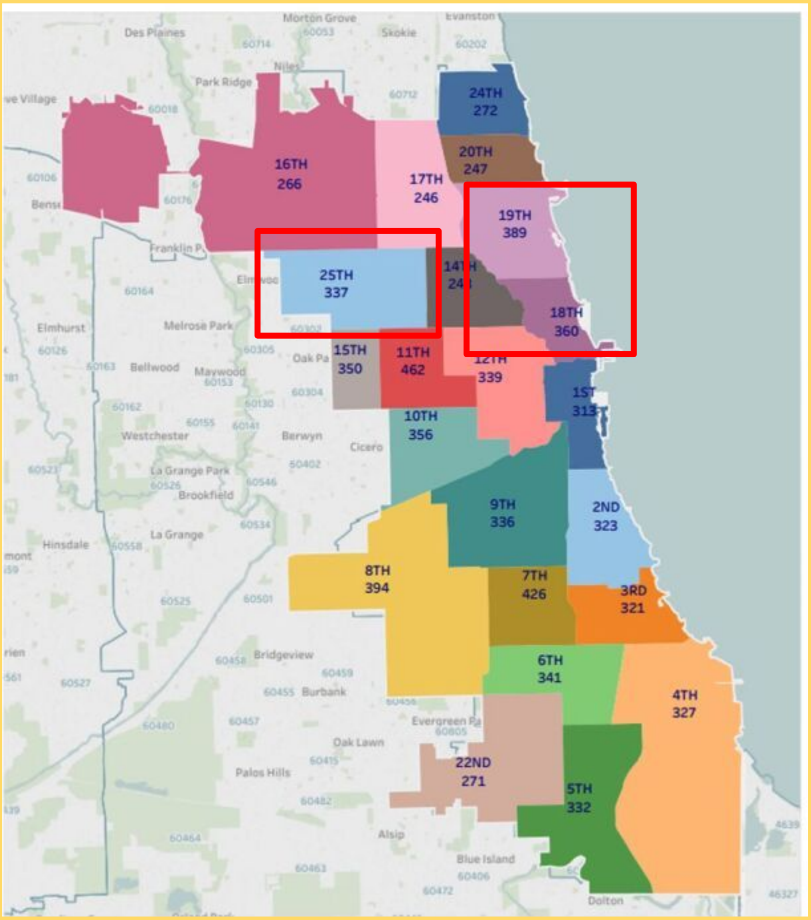
Analysis1.ipynb

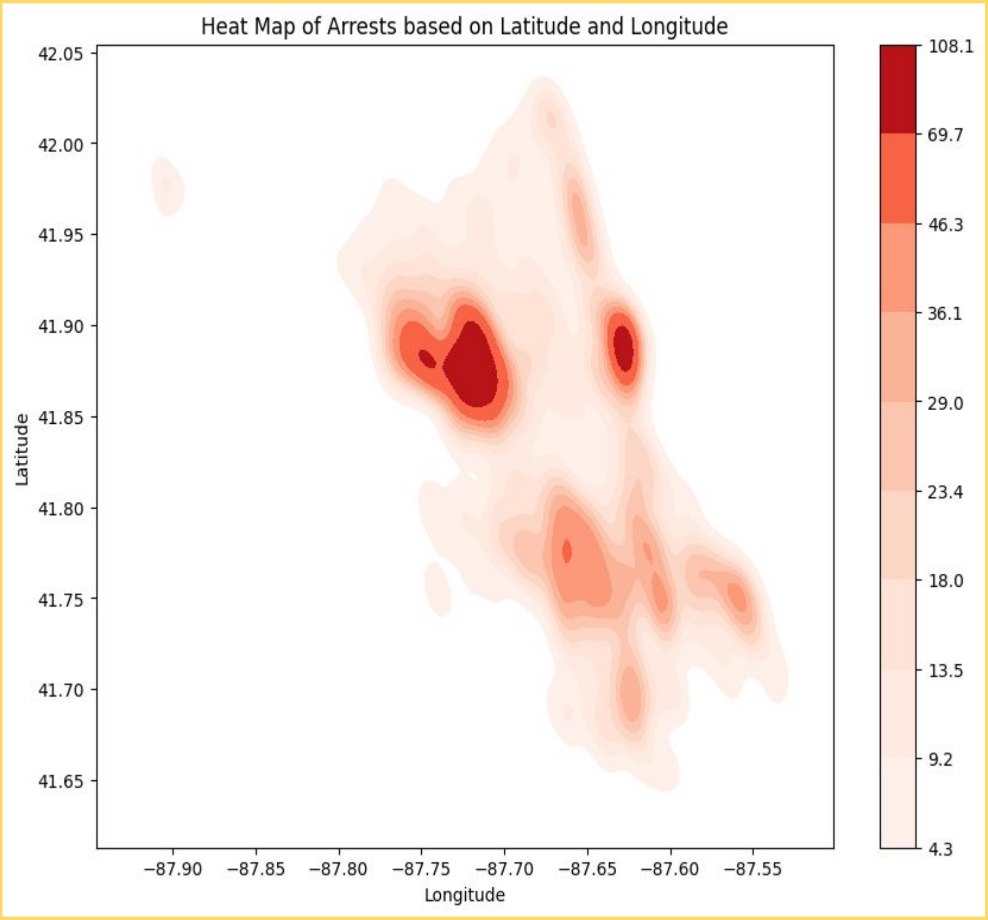
Heat Map of Arrests Made



The brighter red hotspots are the locations where more arrests are made.

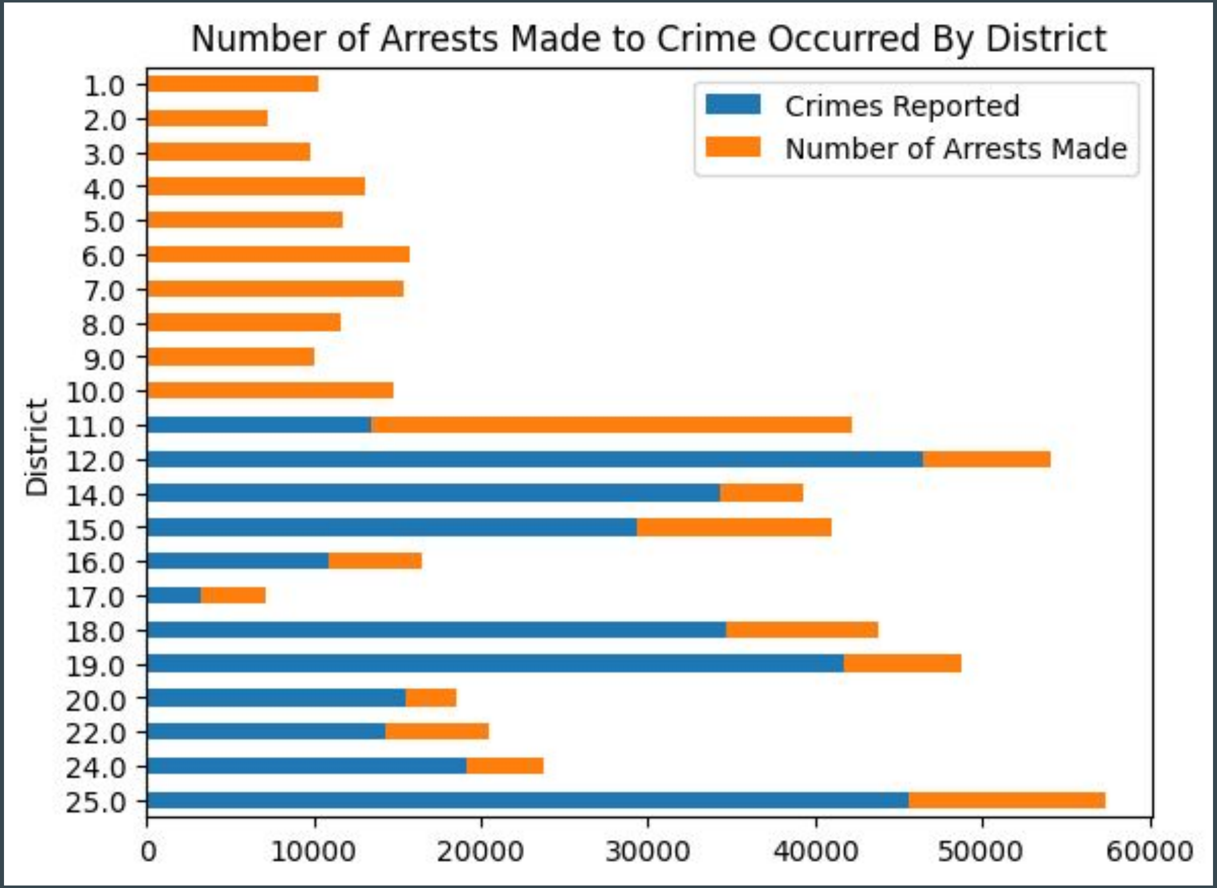
Analysis1.ipynb





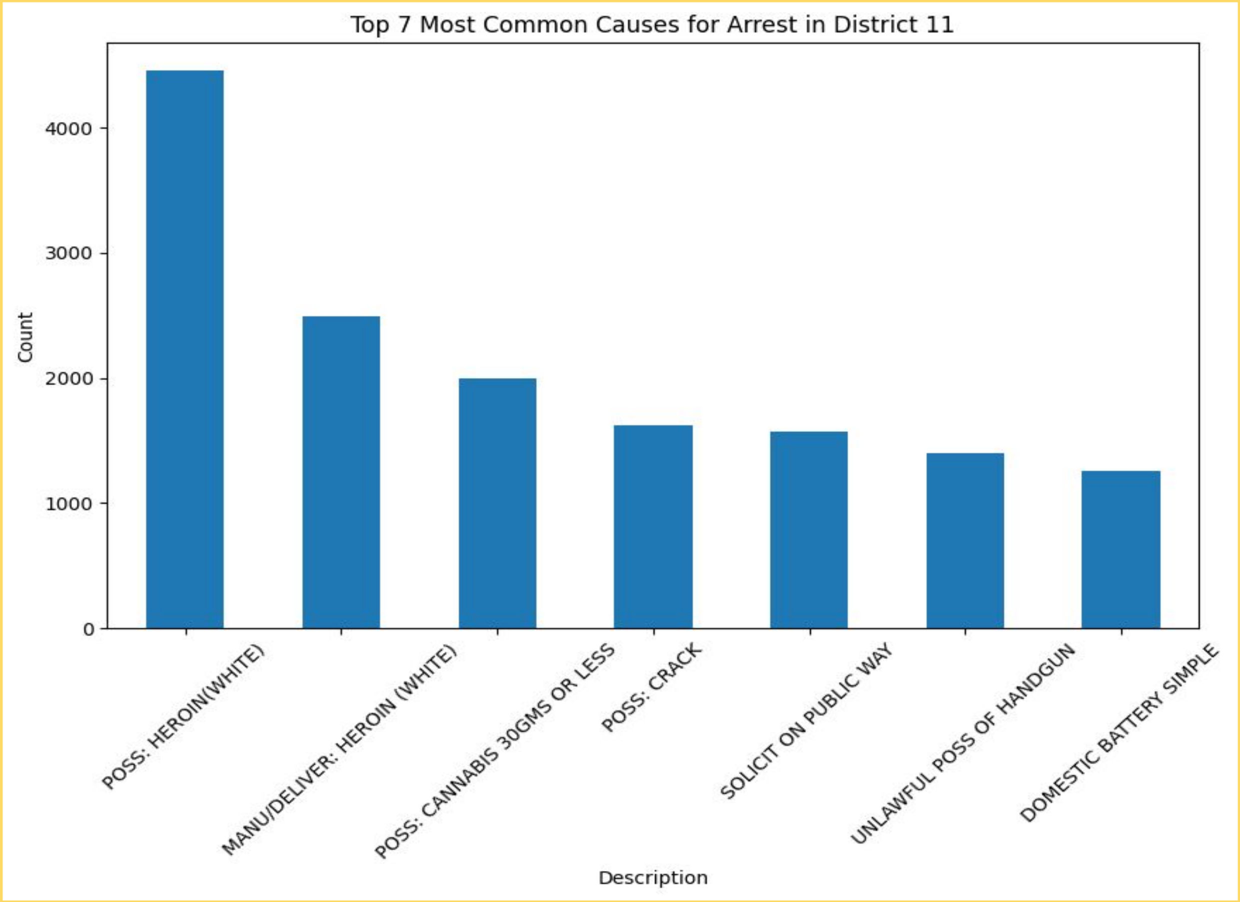
Chicago District Map
sg(rev).ipynb

```
import pandas as pd
import matplotlib.pyplot as plt

# Filter out District 31
District_df_filtered = District_df[District_df['District'] != 31]

# Reverse the order of the DataFrame
District_df_filtered = District_df_filtered.iloc[::-1]

# Stacked bar chart for Arrest:Crime by District
ax = District_df_filtered.plot(x="District", kind="barh", stacked=True,
title="Number of Arrests Made to Crime Occurred By District")
df_total = District_df_filtered["Number of Arrests Made"] +
District_df_filtered["Crimes Reported"]
df_rel = District_df_filtered[District_df_filtered.columns[1:]].div(df_total, 0)
* 100

plt.show()
```



District 11 has the lowest crimes reported but the highest number of arrests.

Resources/Crimes_2015toPresent.csv



Possession of Heroin is the most common cause for arrest in District 11.

Project-1-2/sg(rev).ipynb

```python
top_10_common_description = data['Description'].value_counts().head(10)

print(top_10_common_description)
```

[16] Python

```
SIMPLE                          49963
$500 AND UNDER                  46301
DOMESTIC BATTERY SIMPLE         43123
OVER $500                       29938
RETAIL THEFT                    29538
FROM BUILDING                   22155
TO VEHICLE                      19767
TO PROPERTY                     16485
AUTOMOBILE                      16035
POSS: CANNABIS 30GMS OR LESS    12765
Name: Description, dtype: int64
```

Most Popular Charge Description

Project1-2/sg(rev).ipynb
Crimes2015toPresent.csv

[33]

```
most_common_description_per_year
```

	Year	Description
0	2015	POSS: CANNABIS 30GMS OR LESS
1	2016	DOMESTIC BATTERY SIMPLE
2	2017	DOMESTIC BATTERY SIMPLE
3	2018	DOMESTIC BATTERY SIMPLE
4	2019	DOMESTIC BATTERY SIMPLE
5	2020	UNLAWFUL POSSESSION - HANDGUN
6	2021	UNLAWFUL POSSESSION - HANDGUN
7	2022	MANUFACTURE / DELIVER - CANNABIS OVER 10 GRAMS

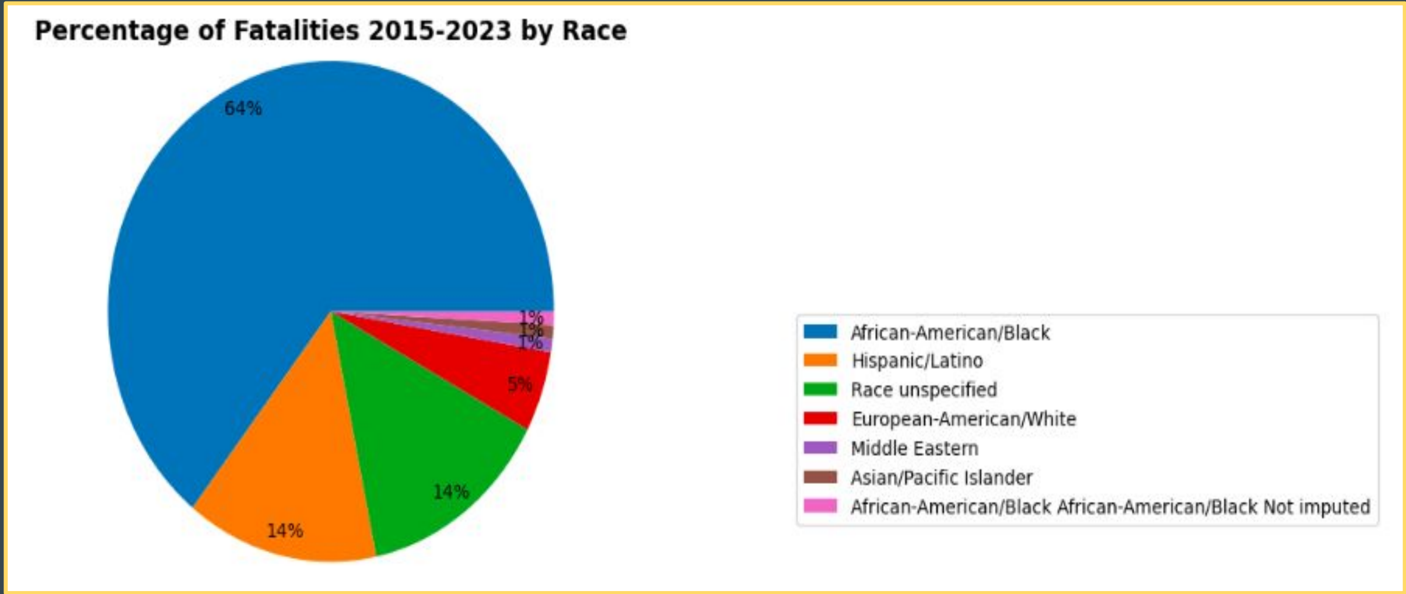
Most Popular Cause for a Charge by Year

Project1-2/sg(rev).ipynb
Crimes2015toPresent.csv

Analysis 2: Fatalities

Fatalities by Race Group

Years 2015-2021

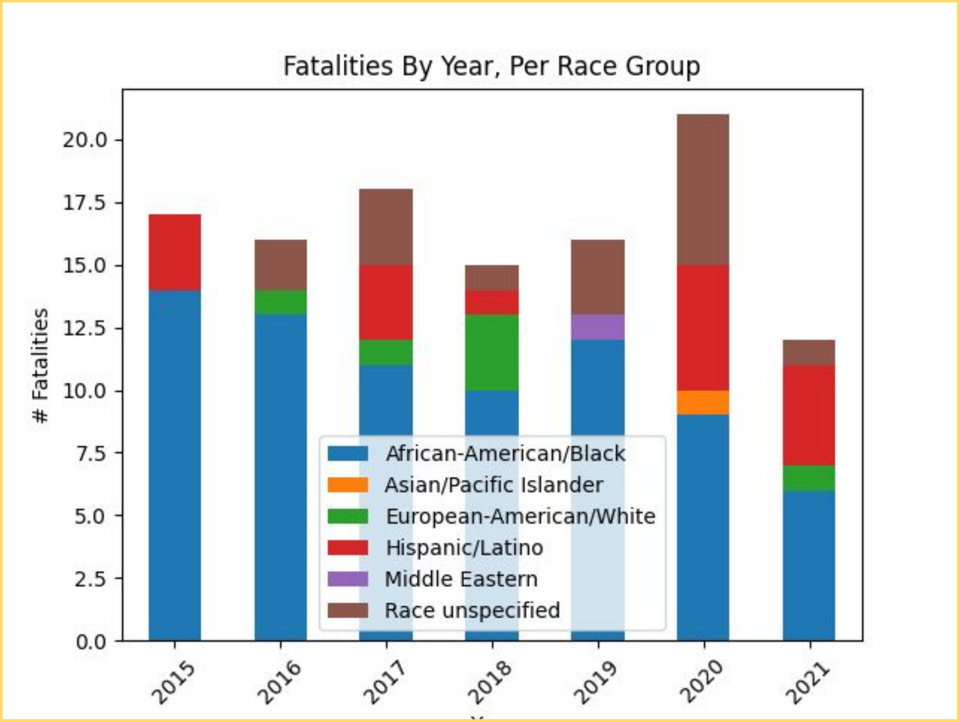


Fatalities by Race Group

Years 2015-2021

Throughout the eight-year period, the overwhelming majority of fatalities happened among African-American/Black people (blue, n = 75)

- Hispanic/Latino (red, n = 16)
- European-American/White (green, n = 6)
- Middle Eastern (purple, n = 1)
- Asian/Pacific Islander (orange, n = 1)
- Race Unspecified (brown, n = 16)



Top 5 Zip Codes with the Highest Number of Fatalities

Years 2015-2021

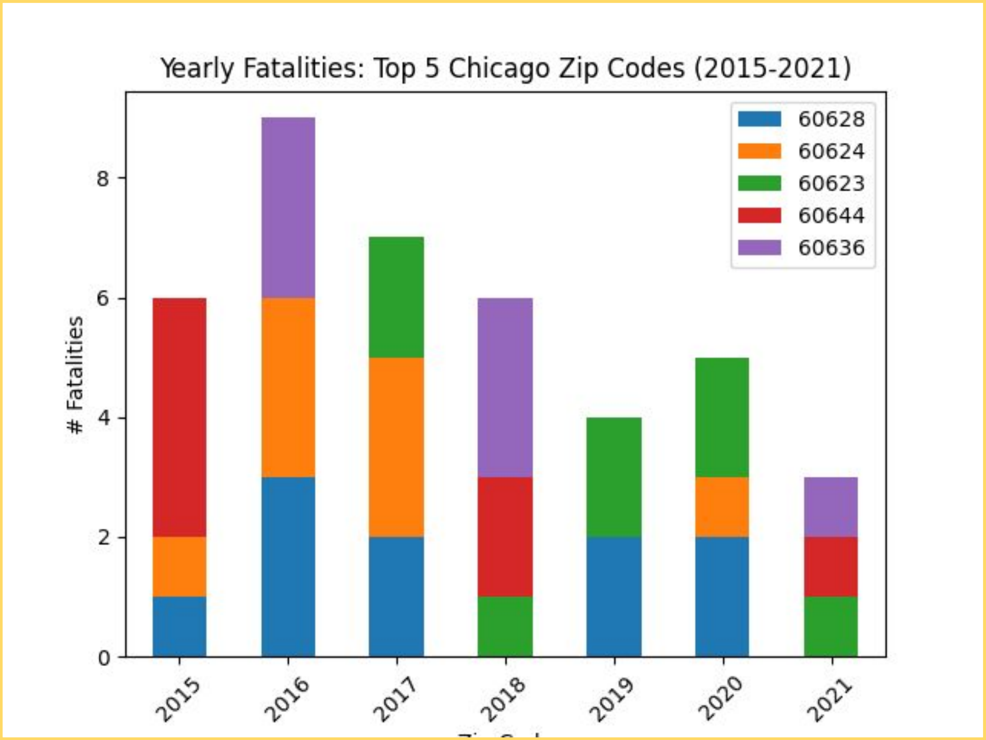
- Zip Code 60628, *Roseland* (n = 10)
- Zip Code 60624, *West Garfield Park* (n = 8)
- Zip Code 60623, *South Lawndale* (n = 8)
- Zip Code 60644, *Austin* (n = 7)
- Zip Code 60636, *West Englewood* (n = 7)

Top 5 Zip Codes with the Highest Number of Fatalities

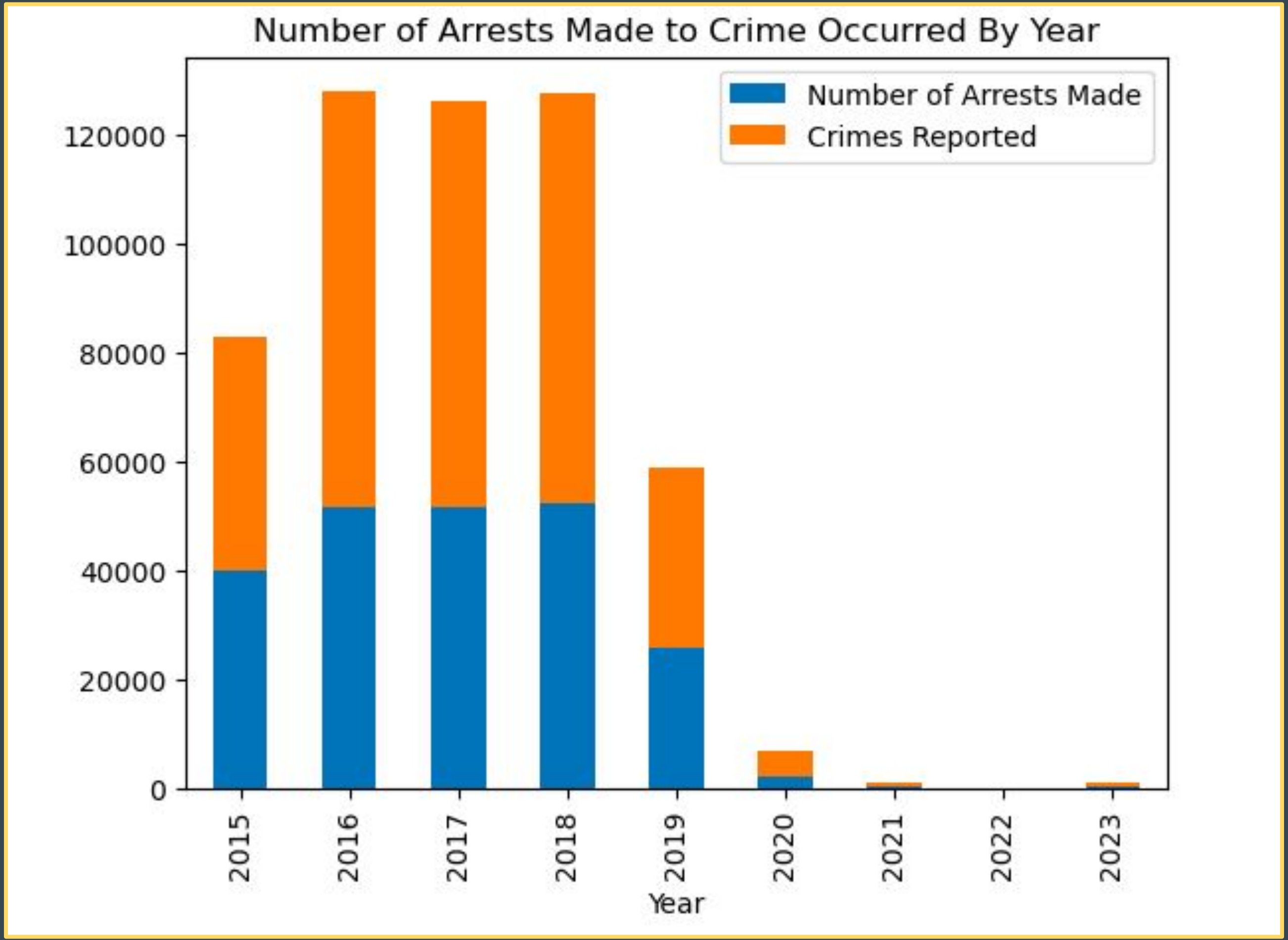
Years 2015-2021

Throughout the same period, the majority of fatalities occurred in Zip Code 60628, *Roseland* (blue, n = 10)

- Zip Code 60624, *West Garfield Park* (orange, n = 8)
- Zip Code 60623, *South Lawndale* (green, n = 8)
- Zip Code 60644, *Austin* (red, n = 7)
- Zip Code 60636, *West Englewood* (purple, n = 7)

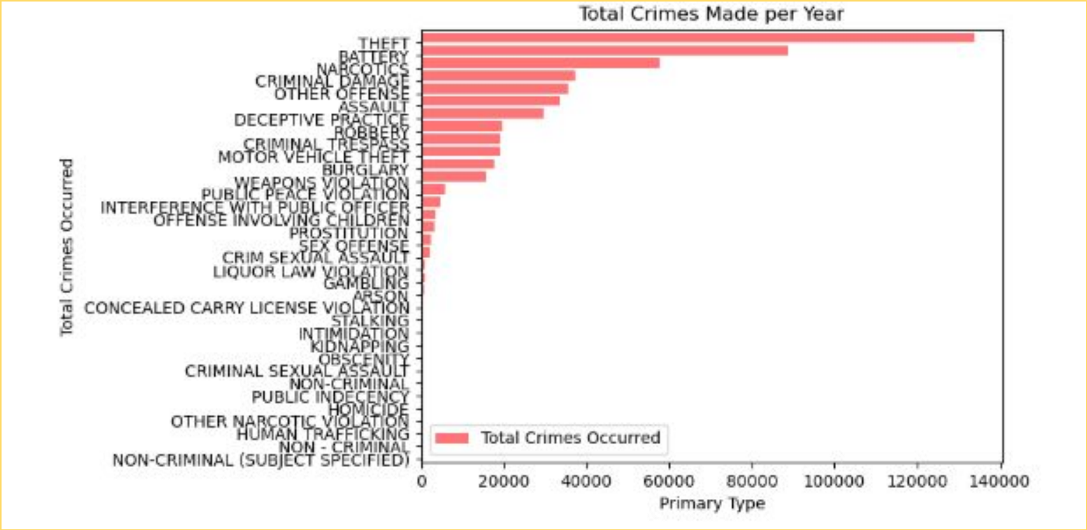


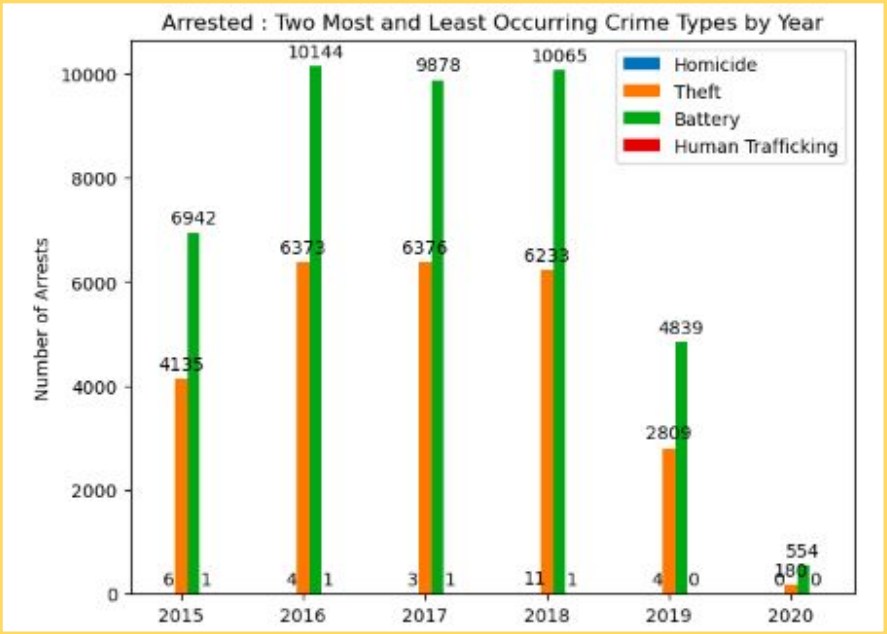
Analysis 3: Cause of Arrest


Number of Arrests Made to Crime Occurred By Year
Number of Arrests Made
Crimes Reported
0
20000
40000
60000
80000
100000
120000
2015
2016
2017
2018
2019
2020
2021
2022
2023
Year

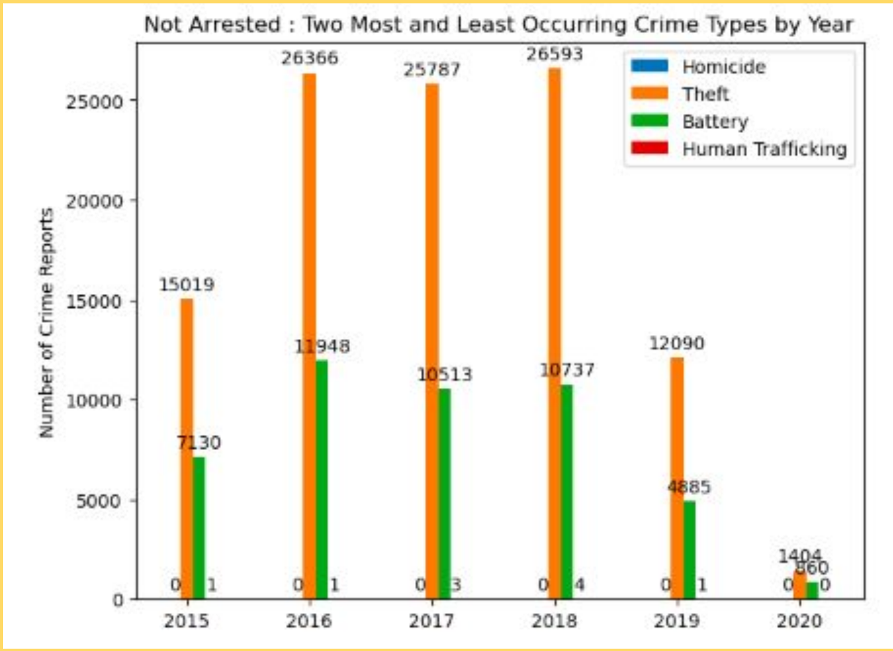
Analysis 3: Cause of Arrest

Primary Type	
NON-CRIMINAL (SUBJECT SPECIFIED)	2
NON - CRIMINAL	5
HUMAN TRAFFICKING	14
OTHER NARCOTIC VIOLATION	16
HOMICIDE	31
PUBLIC INDECENCY	47
NON-CRIMINAL	81
CRIMINAL SEXUAL ASSAULT	94
OBSCENITY	227
KIDNAPPING	250
INTIMIDATION	281
STALKING	343
CONCEALED CARRY LICENSE VIOLATION	368
ARSON	630
GAMBLING	813
LIQUOR LAW VIOLATION	967
CRIM SEXUAL ASSAULT	1952
SEX OFFENSE	2334
PROSTITUTION	3230
OFFENSE INVOLVING CHILDREN	3433
INTERFERENCE WITH PUBLIC OFFICER	4646
PUBLIC PEACE VIOLATION	5711
WEAPONS VIOLATION	15616
BURGLARY	17725
MOTOR VEHICLE THEFT	18944
CRIMINAL TRESPASS	19021
ROBBERY	19712
DECEPTIVE PRACTICE	29658
ASSAULT	33550
OTHER OFFENSE	35391
CRIMINAL DAMAGE	37204
NARCOTICS	57647
BATTERY	88872
THEFT	133938





Arrested : Two Most and Least Occurring Crime Types by Year
Number of Arrests
0
2000
4000
6000
8000
10000
2015
2016
2017
2018
2019
2020
Homicide
Theft
Battery
Human Trafficking
6942
4135
10144
6373
9878
6376
10065
6233
4839
2809
554
180



Not Arrested : Two Most and Least Occurring Crime Types by Year
Number of Crime Reports
0
5000
10000
15000
20000
25000
2015
2016
2017
2018
2019
2020
Homicide
Theft
Battery
Human Trafficking
15019
7130
26366
11948
25787
10513
26593
10737
12090
4885
1404
860

Safest District

(2015 - Present)

```
# District 31 is an outlier
filtered_data = data[data['District'] != 31]

# Count the number of rows for each district
district_counts = filtered_data['District'].value_counts()

# Find the district with the fewest number of rows
safest_district = district_counts.idxmin()

print("Safest District (Fewest Number of Rows):", safest_district)
```

✓ 0.0s

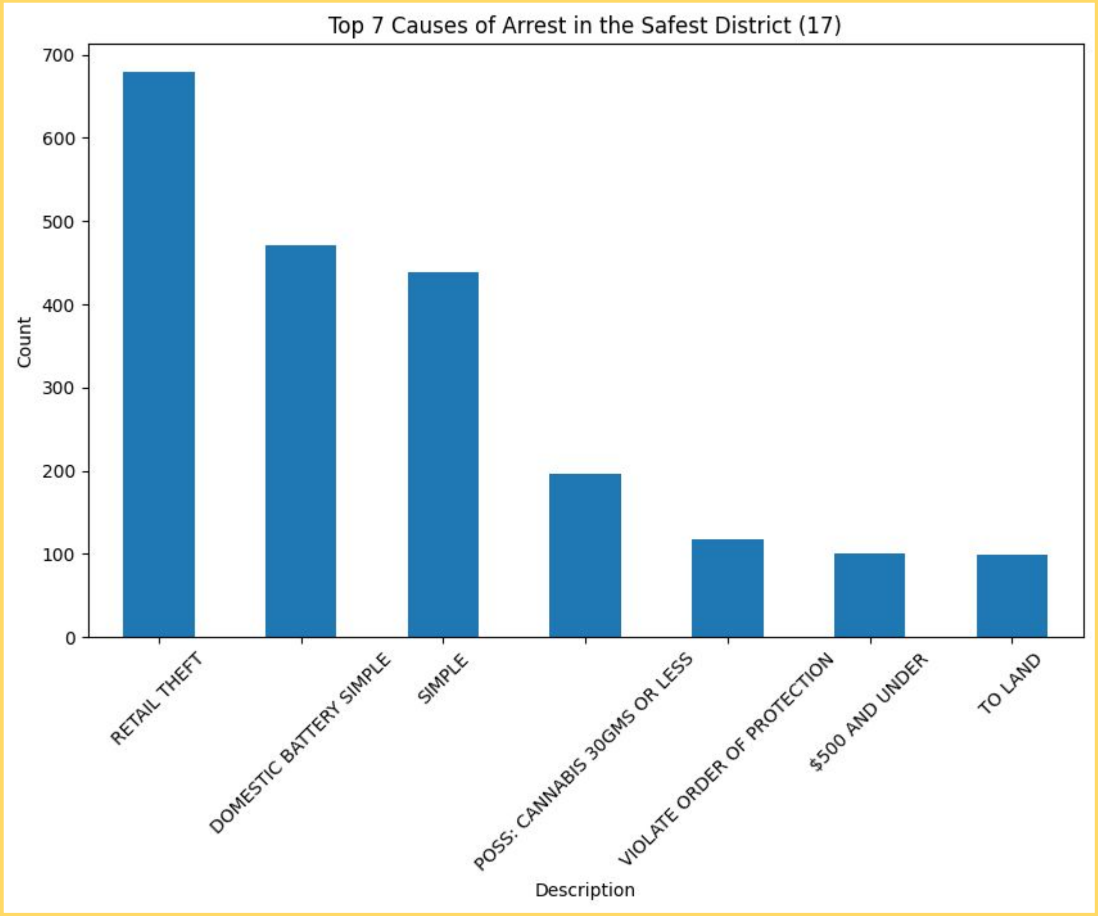
```
Safest District (Fewest Number of Rows): 17.0
```

sg(rev).ipynb

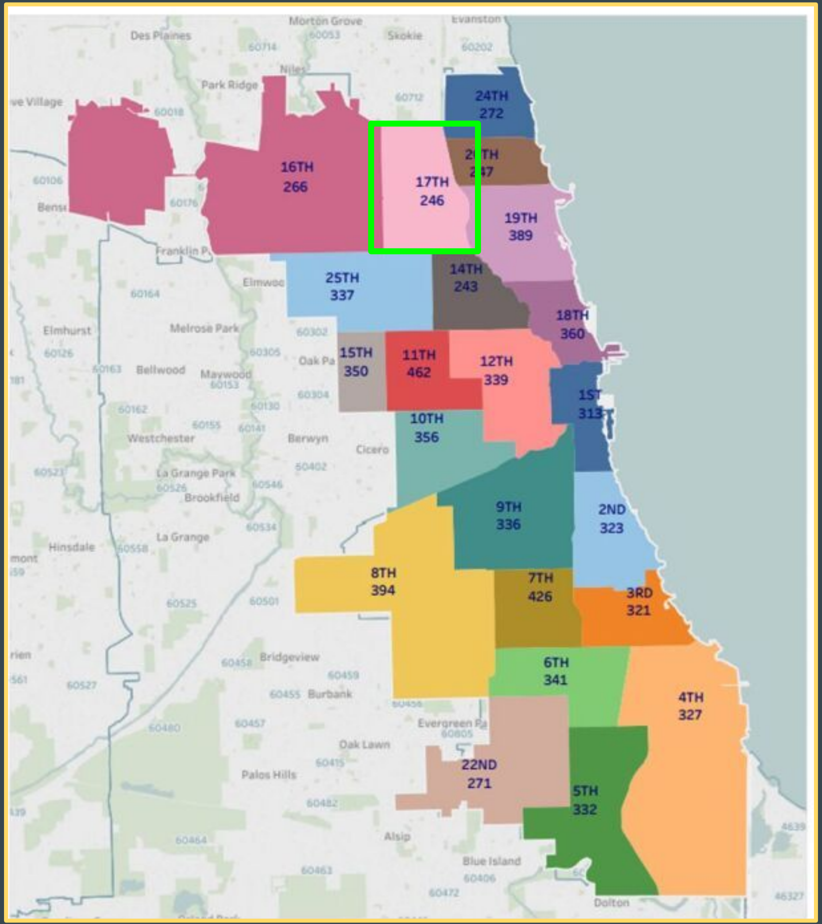
Top 7 Causes of Arrest in the Safest District (17)

```
# Filter data for Arrest == True and District 17
filtered_data = data[(data['Arrest'] == True) & (data['District'] == 17)]

# Get the top 7 descriptions by count
top_descriptions = filtered_data['Description'].value_counts().head(7)

# Create a bar plot
plt.figure(figsize=(10, 6))
top_descriptions.plot(kind='bar')
plt.xlabel('Description')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.title('Top 7 Causes of Arrest in the Safest District (17)')

plt.show()
```

Top 7 Causes of Arrest in the Safest District (17)



sg(rev).ipynb



District 17 (Safest District) on a Map

[Chicago District Map]

Common Causes of Arrest per Primary Type Overall (2015-Present)

```
#Arrests made per crime (Primary Type)
total_arrests_per_crime = crimes[crimes['Arrest'] == True].groupby('Primary Type').size()
total_arrests_per_crime
```

```
#sort primary type of arrests by greatest to smallest
arrest_primary_largest= total_arrests_per_crime.sort_values(ascending=False)
arrest_primary_largest
```

```
Primary Type
NARCOTICS                             57632
BATTERY                               42556
THEFT                                 26156
CRIMINAL TRESPASS                     16042
OTHER OFFENSE                         15530
ASSAULT                               14417
WEAPONS VIOLATION                     14077
CRIMINAL DAMAGE                        7015
PUBLIC PEACE VIOLATION                 4671
INTERFERENCE WITH PUBLIC OFFICER       4571
DECEPTIVE PRACTICE                     3526
ROBBERY                                3479
```

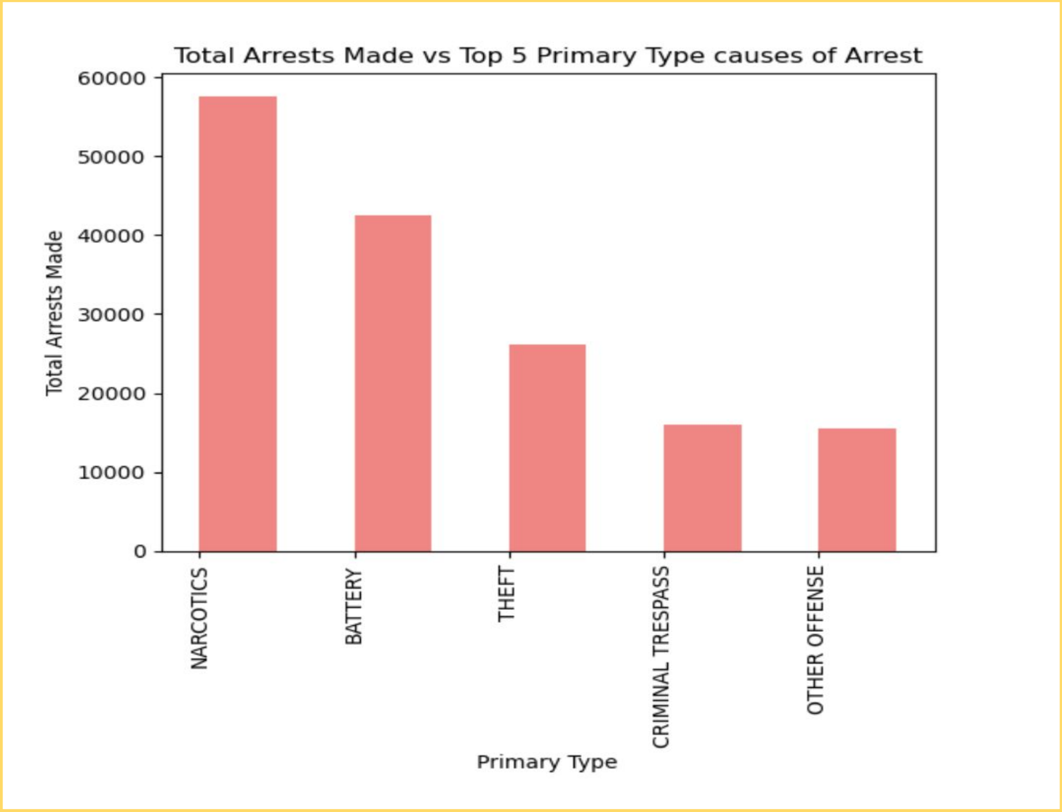
Narcotics was the overall highest cause of arrest for 2015 - Present, with Battery coming in second.

Common Causes of Arrest per Primary Type Overall (2015-Present)

```
#filter top 5 reasons for arrest by Primary type & graph
top5_primarytype=arrest_primary_largest.head(5)
top5_primarytype.plot(kind='bar', color= 'r', alpha=0.5, align="edge")
plt.xticks(rotation="vertical")
plt.xlabel("Primary Type")
plt.ylabel("Total Arrests Made")
plt.title("Total Arrests Made vs Top 5 Primary Type causes of Arrest")
plt.show
```

Plotting Primary Cause of Arrest

Common Causes of Arrest per Primary Type Overall (2015-Present)



Top 5 Common Causes of Arrest, Primary Type per Year

```
#sort by arrests made and year (2015-2022) by Primary Type
new_crimes_df_filtered=new_crimes_df[(new_crimes_df['Year']>=2015)&(new_crimes_df['Year']<=2022)]

arrests_made=new_crimes_df_filtered[new_crimes_df_filtered['Arrest'] == True]
top_primary_crimes=arrests_made.groupby('Year')['Primary Type'].value_counts().groupby(level=0).head()
grouped_top_primary=top_primary_crimes.to_frame()
grouped_top_primary
```

Year	Primary Type	
2015	NARCOTICS	12423
	BATTERY	6942
	THEFT	4135
	OTHER OFFENSE	2861
	CRIMINAL TRESPASS	2532
2016	NARCOTICS	13157
	BATTERY	10144
	THEFT	6373
	CRIMINAL TRESPASS	3694
	OTHER OFFENSE	3389
2017	NARCOTICS	11544
	BATTERY	9878
	THEFT	6376
	CRIMINAL TRESPASS	3996
	OTHER OFFENSE	3753
2018	NARCOTICS	13112
	BATTERY	10065
	THEFT	6233
	WEAPONS VIOLATION	3782
	CRIMINAL TRESPASS	3767
2019	NARCOTICS	6872
	BATTERY	4839
	THEFT	2809
	CRIMINAL TRESPASS	1928
	WEAPONS VIOLATION	1918
2020	BATTERY	554
	WEAPONS VIOLATION	429
	NARCOTICS	399
	ASSAULT	184
	THEFT	180
2021	WEAPONS VIOLATION	87
	BATTERY	67
	NARCOTICS	51
	ASSAULT	30
	THEFT	23
2022	NARCOTICS	26
	BATTERY	2
	DECEPTIVE PRACTICE	1
	HOMICIDE	1
	MOTOR VEHICLE THEFT	1

An Analysis was done to find the top 5 common causes of arrest per year
Narcotics taking the lead from 2015-2019

Common Causes of Arrest per Description

```python
data_filtered = data[(data['Year'] >= 2015) & (data['Year'] <= 2022)]

# Filter data where Arrests is True
data_arrests_true = data_filtered[data_filtered['Arrest'] == True]

# Find the most common description for each year
most_common_description_per_year = data_arrests_true.groupby('Year')['Description'].apply(lambda x: x.value_counts().idxmax()).reset_index()

most_common_description_per_year
```

[19] ✓ 0.3s Python

...

	Year	Description
0	2015	POSS: CANNABIS 30GMS OR LESS
1	2016	DOMESTIC BATTERY SIMPLE
2	2017	DOMESTIC BATTERY SIMPLE
3	2018	DOMESTIC BATTERY SIMPLE
4	2019	DOMESTIC BATTERY SIMPLE
5	2020	UNLAWFUL POSSESSION - HANDGUN
6	2021	UNLAWFUL POSSESSION - HANDGUN
7	2022	MANUFACTURE / DELIVER - CANNABIS OVER 10 GRAMS

Arrests Per Year (2015-2023)

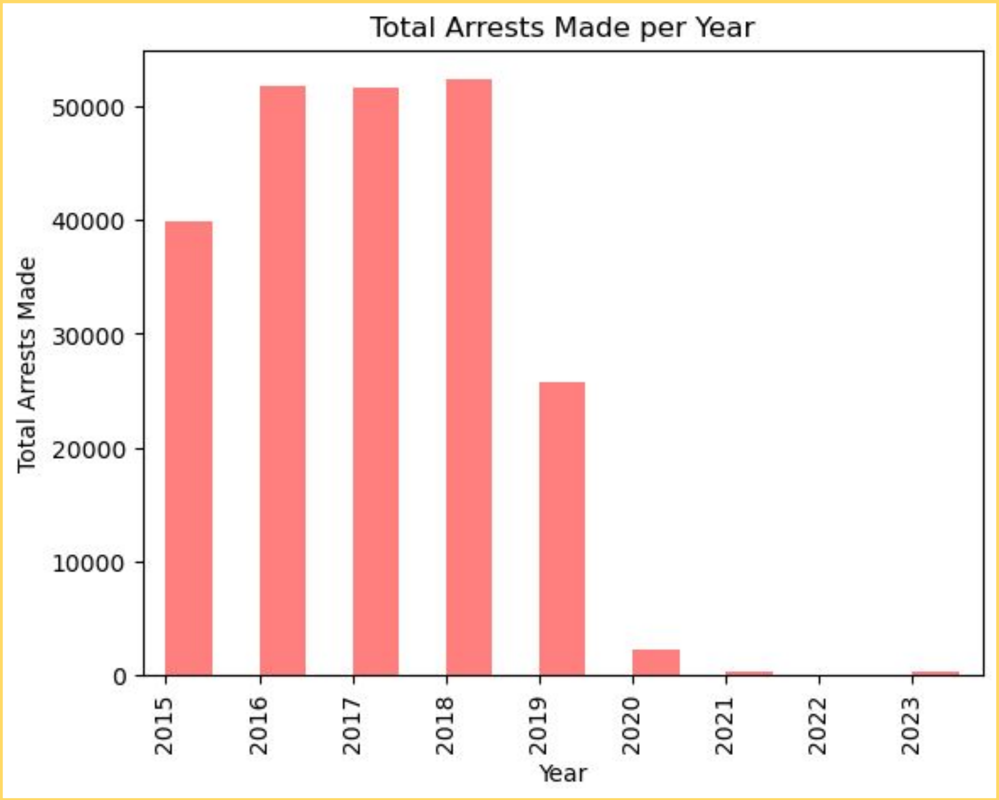
```
#count the number of arrests per year
total_arrests_per_year = crimes[crimes['Arrest'] == True].groupby('Year').size()
total_arrests_per_year
```

```
Year
2015    39908
2016    51806
2017    51592
2018    52314
2019    25715
2020     2194
2021      322
2022       33
2023      325
```

The most arrests were made in 2018 with a total of 52,314, with 2016 being the second highest of total arrests made, followed by 2017.

Total Arrests Made per Year

Years 2015 - 2023



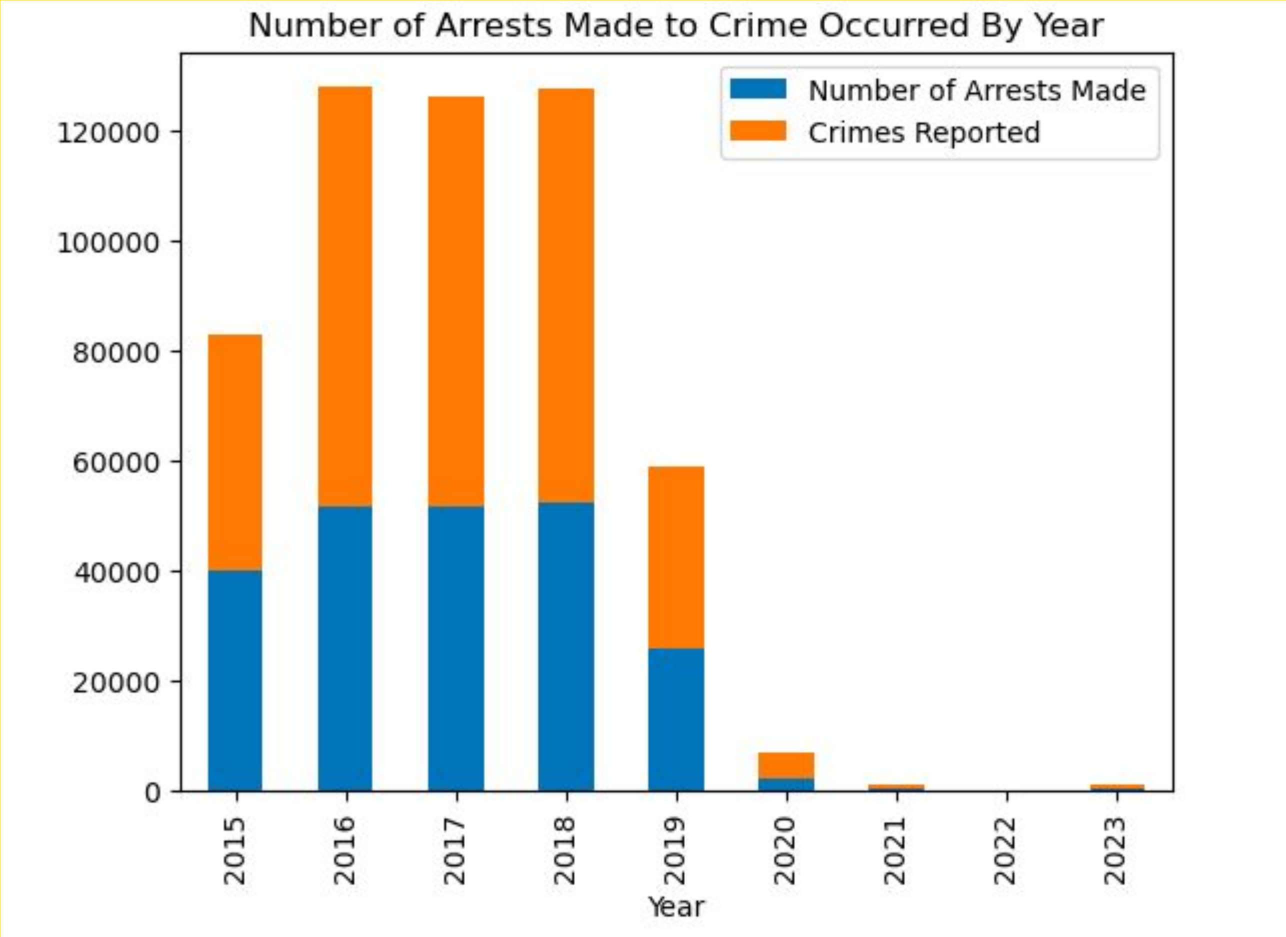
Independent T-Test

Independent T-Test

Based on Arrests Made in each District per Year

Objective - Find out if the average crime count between years 2015, 2016, and 2020 are statistically significant.

Before a test can be made the following criteria must be true.
- Data must be independent
- Homogenous (standard deviations must be mostly equal)
- Data must be normally distributed


Number of Arrests Made to Crime Occurred By Year
Number of Arrests Made
Crimes Reported
0
20000
40000
60000
80000
100000
120000
2015
2016
2017
2018
2019
2020
2021
2022
2023
Year

Create Data Frames for each year

In [9]:

```
#create a data frame for only 2015
year2015df = crimes[crimes["Year"] == 2015].reset_index()
year2015df.head()
```

Out[9]:

	index	District	Primary Type	Arrest	Date	IUCR	
0	0	1.0	ARSON	True	10/26/2015 07:06:00 PM	1090	ATTEN
1	416	7.0	ASSAULT	True	09/16/2015 10:33:00 PM	0560	
2	417	7.0	ASSAULT	True	09/16/2015 02:35:00 PM	0520	AGGRAVATED:KNIF
3	418	7.0	ASSAULT	True	09/13/2015 05:30:00 PM	0560	
4	419	7.0	ASSAULT	True	09/13/2015 10:35:00 AM	0560	

Arrests per District for the whole Dataset

In [6]:

```
#count the number of arrests per District
total_arrests_per_year = crimes.groupby('District').size()
total_arrests_per_year
```

Out[6]:

```
District
1.0     10254
2.0      7247
3.0      9806
4.0     13049
5.0     11768
6.0     15806
7.0     15426
8.0     11684
9.0     10096
10.0    14743
11.0    42273
12.0    54080
14.0    39289
15.0    41019
16.0    16500
17.0     7120
18.0    43753
19.0    48723
20.0    18515
22.0    20444
24.0    23799
25.0    57332
31.0       27
dtype: int64
```

Types of Crimes Committed in 2015

In [15]:

```
#Arrests made per crime in only (Primary Type)
crime_2015 = year2015df[year2015df['Arrest'] == True].groupby('Primary Type').size()
crime_2015 = pd.DataFrame(crime_2015).reset_index()
crime_2015.columns = ["Primary Type", "Primcount"]
crime_2015
```

Out[15]:

	Primary Type	Primcount
0	ARSON	33
1	ASSAULT	2399
2	BATTERY	6942
3	BURGLARY	481

Arrange the data in a normal distribution

In [16]:

```
#Re-arrange the "Primcount" column to be normally distributed
# Calculate mean and standard deviation
mean = crime_2015["Primcount"].mean()
std_dev = crime_2015["Primcount"].std()

# Generate a normal distribution with the same length as the DataFrame
normal_dist = np.random.normal(mean, std_dev, size=len(crime_2015))

# Sort the DataFrame by the "Primcount" column in descending order
crime_2015 = crime_2015.sort_values(by="Primcount", ascending=False)

# Update the "Primcount" column with the normal distribution values
crime_2015["Primcount"] = normal_dist

# Print the resulting DataFrame
print(crime_2015)
```

Calculate the mean of offense for 2015 and 2020

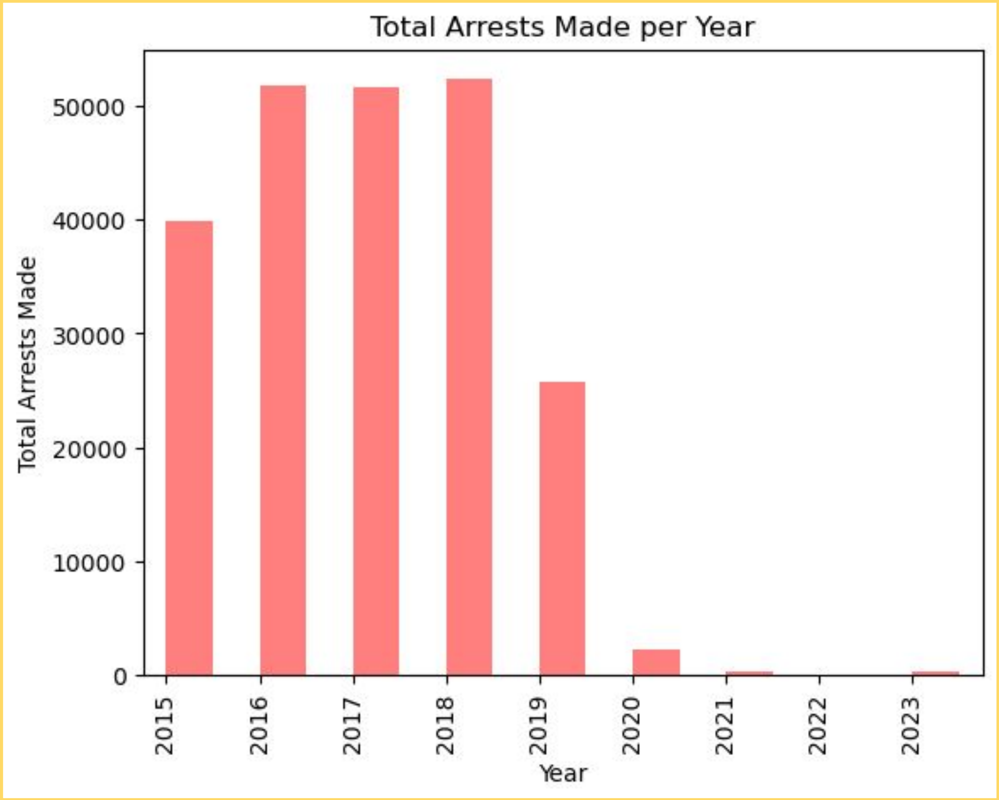
```
In [22]:  #Average of 2015 offense being true
          crime_2015.Primcount.mean()

Out[22]:  1048.5996316017163
```

```
In [23]:  #Average of 2016 offense being true
          crime_2016.Primcount.mean()

Out[23]:  1569.878787878788
```

```
In [27]:  #Average of 2020 offense being true
          crime_2020.Primcount.mean()

Out[27]:  109.7
```

Total Arrests Made per Year

Years 2017-2023



Test between 2015 to 2016

```
In [28]: # Note: Setting equal_var=False performs Welch's t-test which does
         # not assume equal population variance for 2015 AND 2016
         stats.ttest_ind(crime_2015.Primcount, crime_2016.Primcount, equal_var=False)

Out[28]: Ttest_indResult(statistic=-0.8156210993701961, pvalue=0.418070063240637)
```

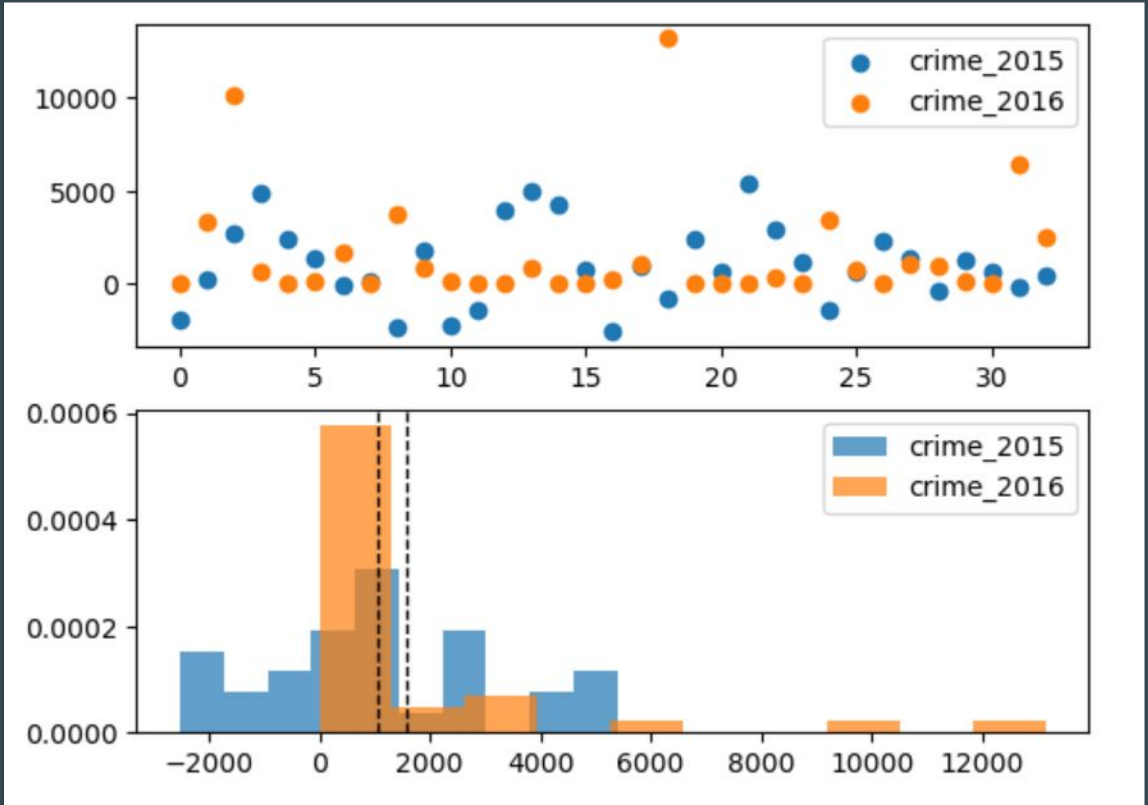
If the pvalue is less than 0.05 it is considered significant that which would suggest there is a relationship between 2015 and 2016, how ever not in this case. There is high variability in the data.

Test between 2015 to 2020

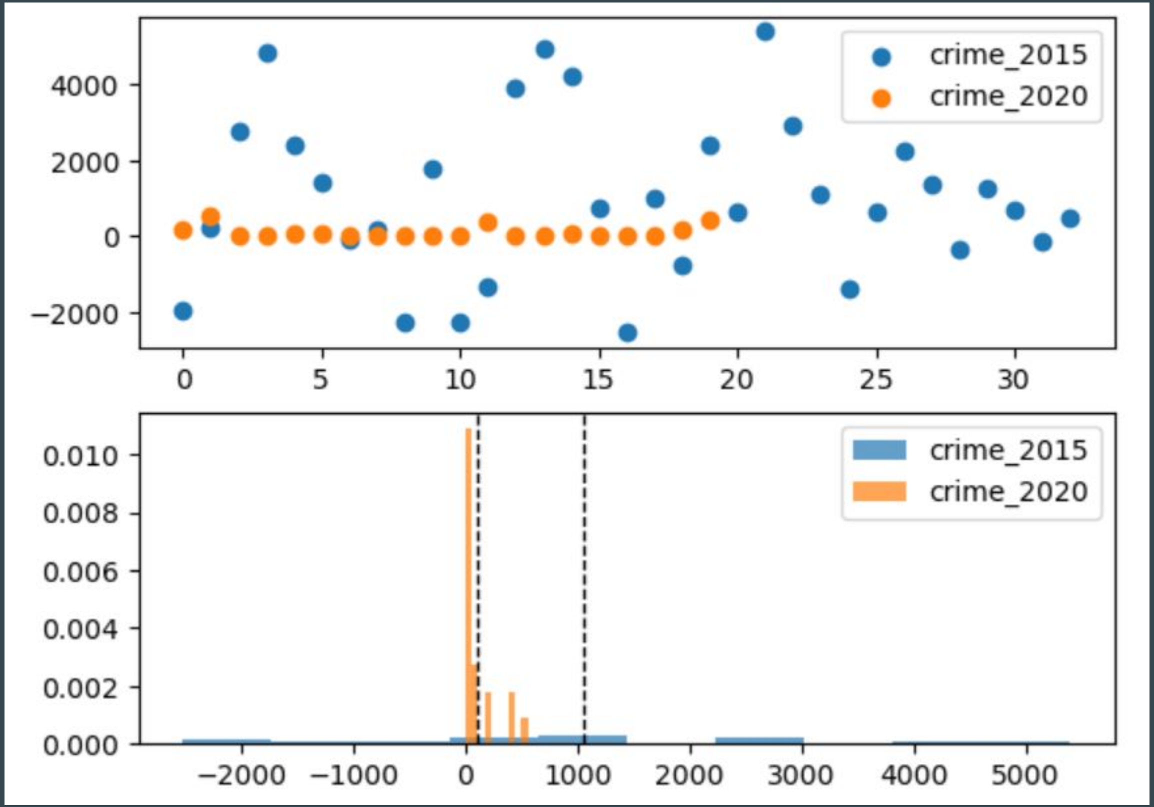
```
In [32]: # Note: Setting equal_var=False performs Welch's t-test which does
         # not assume equal population variance for 2015 AND 2020
         stats.ttest_ind(crime_2015.Primcount, crime_2020.Primcount, equal_var=False)

Out[32]: Ttest_indResult(statistic=2.5257629561752086, pvalue=0.016589045427415124)
```

If the pvalue is less than 0.05 it is considered significant that which would suggest there is a relationship between 2015 and 2020, how ever not in this case. There is high variability in the data.

Plot of T-Test for 2015-2016



Plot of T-Test for 2015-2020



Limitations

This data analysis project incorporates datasets from various resources, which introduces certain limitations that should be considered.

- Data pulled from various different resources
  - Quality and consistency of the dataset is not guaranteed.
  - Variations in data collection methodologies exist.
  - Data privacy and access restrictions can affect overall analysis.
- Scope and context
  - Analyzing data within a limited timeframe.
  - Findings may not be applicable to a broader population, or over a longer timeframe
- Data interpretation and bias
  - Different analysts may interpret data differently.
- Evolving nature of data
  - The way data is collected and categorized within these sources may continuously evolve at the source.

Interesting Takeaways & Things to Consider for Future Work

Interesting Takeaways & Things to Consider for Future Work

- Most Common Charges are Damage of $500 and Under, Domestic Battery and Retail Theft.
- The quality over quantity of data makes a huge difference.
- Narcotics was the most common cause of arrest until the legalization of cannabis happened in 2019.
- There is an overwhelmingly high number of police-related fatalities when interacting with African-Americans.
- Primary Type of crime occurrences analysis indicate a lack of record keeping from year to year - data points are only for cases opened and could be closed during the year or much later.
- District 11 has the lowest crimes reported but the highest number of arrests.
- Identified locations with high counts of arrests, and generalized crime.

Thank You!