Crime Analysts

Datasets

-Crimes: last year to present
-Chicago Map
-Population 2022

Jibreel Mohamed

Sajal Chandra

Hypothesis

Before beginning our analysis, we came up with the following hypotheses:

- Crimes related to Theft are the highest type of crime
- Crimes related to drug consumption take place in secluded regions away from downtown

```
crimes.isna().sum().sum()
```

```
18363
```

```
print("Missing values = ", crimes.isna().sum().sum()*100/crimes.shape[0],'%')
```

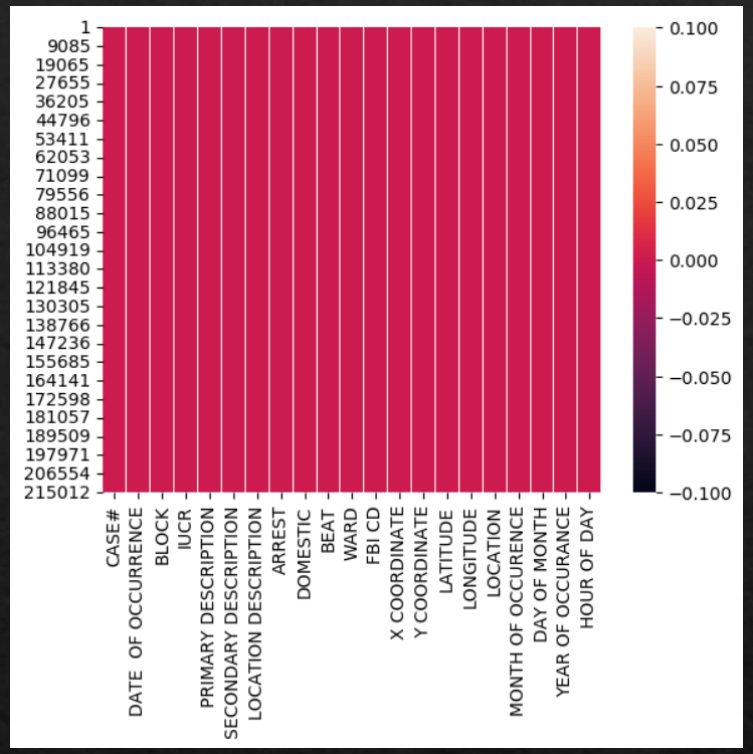
```
Missing values =  8.217430995596606 %
```

```
crimes.dropna(inplace=True)
```

```
crimes.isna().sum().sum()
```

```
0
```

```
sns.heatmap(pd.isna(crimes))
```



Checking and clearing NAN values

What we have in the main dataset

- CASE#
- DATE OF OCCURRENCE
- BLOCK
- IUCR
- PRIMARY DESCRIPTION
- SECONDARY DESCRIPTION
- ARREST
- DOMESTIC
- BEAT
- WARD
- FBI CD
- X COORDINATE
- Y COORDINATE
- LATITUDE
- LONGITUDE
- LOCATION

What we have in the main dataset

- ~~CASE#~~
- DATE OF OCCURRENCE
- ~~BLOCK~~
- ~~IUCR~~
- PRIMARY DESCRIPTION
- ~~SECONDARY DESCRIPTION~~
- ARREST
- DOMESTIC
- ~~BEAT~~
- ~~WARD~~
- ~~FBI CD~~
- ~~X COORDINATE~~
- ~~Y COORDINATE~~
- LATITUDE
- LONGITUDE
- ~~LOCATION~~

- YEAR
- MONTH
- DAY
- HOUR

- DESCRIPTION

Make broad categories for easier understanding

```
crimes['PRIMARY DESCRIPTION'].nunique()
```

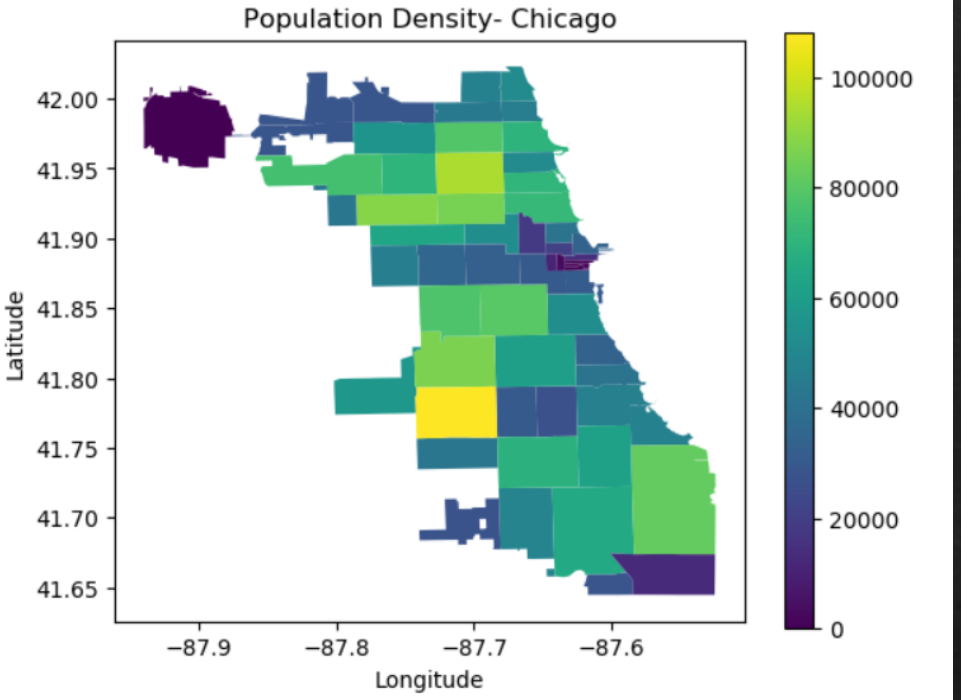
```
31
```

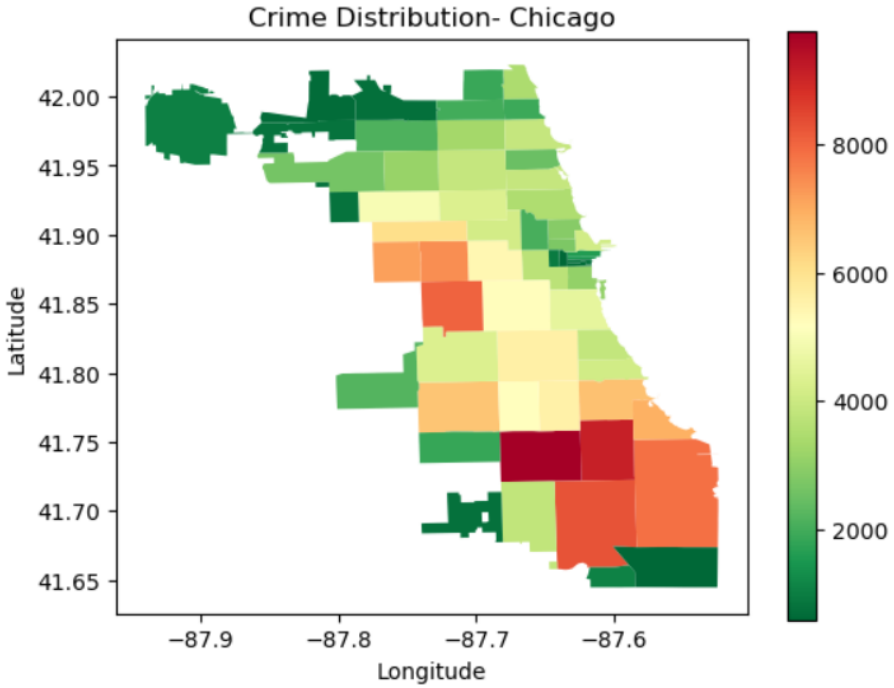
```
phy = "PHYSICAL HARM"
theft = "THEFT"
nhc = "NON-HARM CRIMES"
drug = "DRUG ABUSE"
other = 'OTHER'
prop = 'PROPERTY DAMAGE'
kidnap = 'KIDNAPPING/TRAFFICKING'
pros = "PROSTITUTION"
primary_desc = {
    'ASSAULT': phy,
    'BATTERY': phy,
    'CRIMINAL SEXUAL ASSAULT': phy,
    'HOMICIDE': phy,
    'SEX OFFENSE': phy,
    'CRIMINAL DAMAGE': prop,
    'ARSON': prop,
    'WEAPONS VIOLATION': nhc,
    'CRIMINAL TRESPASS': nhc,
    'DECEPTIVE PRACTICE': nhc,
    'STALKING': nhc,
    'CONCEALED CARRY LICENSE VIOLATION': nhc,
    'PROSTITUTION': pros,
    'OBSCENITY': nhc,
    'PUBLIC INDECENCY': nhc,
    'GAMBLING': nhc,
    'LIQUOR LAW VIOLATION': nhc,
    'PUBLIC PEACE VIOLATION': nhc,
    'INTERFERENCE WITH PUBLIC OFFICER': nhc,
    'THEFT': theft,
    'MOTOR VEHICLE THEFT': theft,
    'BURGLARY': theft,
    'ROBBERY': theft,
    'NARCOTICS': drug,
    'OTHER NARCOTIC VIOLATION': drug,
    'OTHER OFFENSE': other,
    'NON-CRIMINAL': other,
    'OFFENSE INVOLVING CHILDREN': other,
    'INTIMIDATION': other,
    'KIDNAPPING': kidnap,
    'HUMAN TRAFFICKING': kidnap
}

crimes['DESCRIPTION'] = crimes['PRIMARY DESCRIPTION'].map(primary_desc)
```

Merging the population with the map

```
crime_map = gpd.GeoDataFrame(crimes, geometry=gpd.points_from_xy(crimes['LONGITUDE'], crimes['LATITUDE']))
```

```
populationF = population[population['city'] == 'Chicago']
chicago['zip'] = chicago['zip'].apply(lambda x : int(x))
merged = pd.merge(chicago,population ,on='zip')
pop = gpd.GeoDataFrame(merged, geometry='geometry')
```

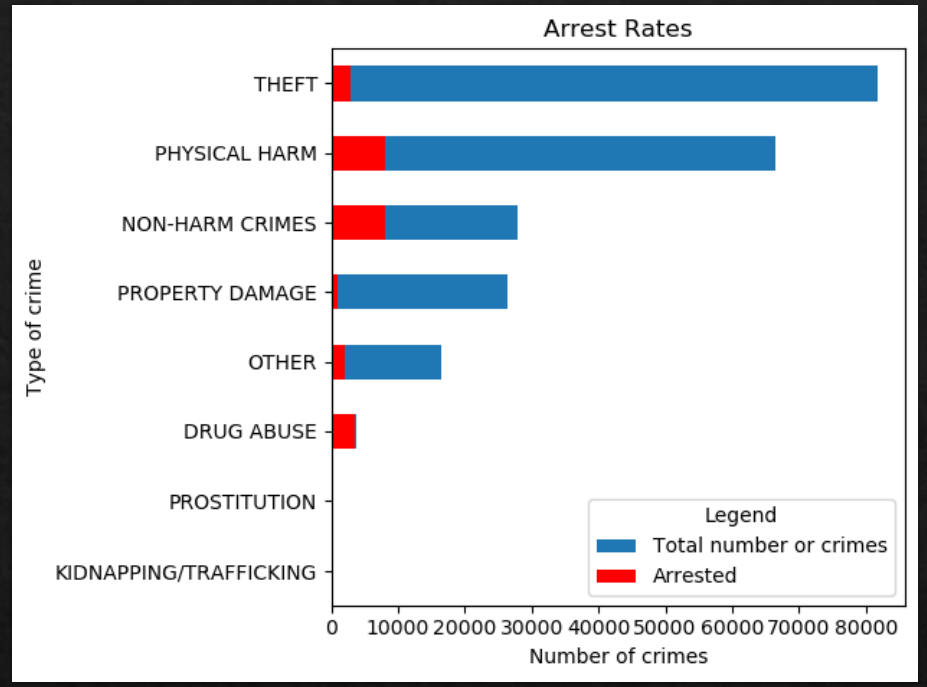


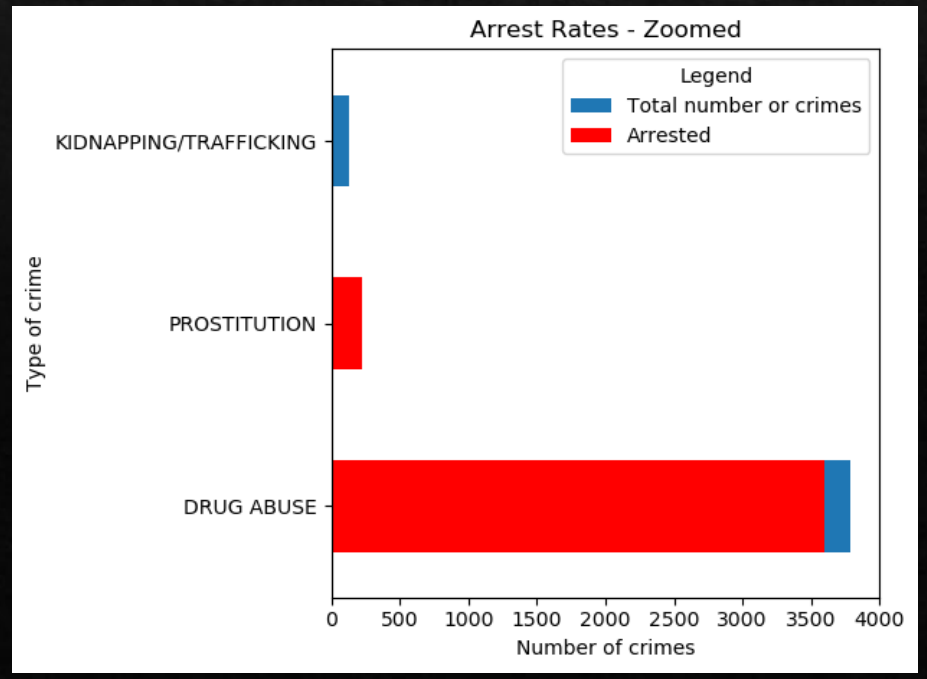


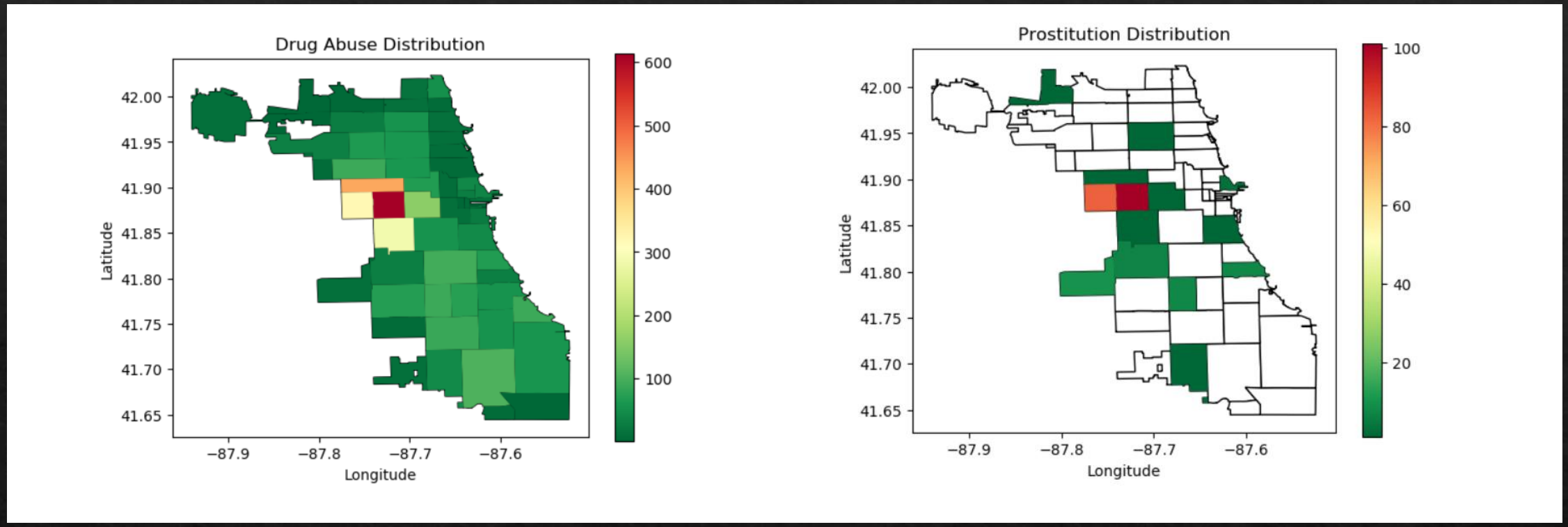
Comparing the population with crimes in Chicago

Arrest rates of different crimes

- Theft has an arrest rate of only 3.5%
- Drug abuse has an arrest rate of over 95%
- Prostitution has an arrest rate of over 98%
- Kidnapping/Trafficking has an arrest rate of less than 8%.







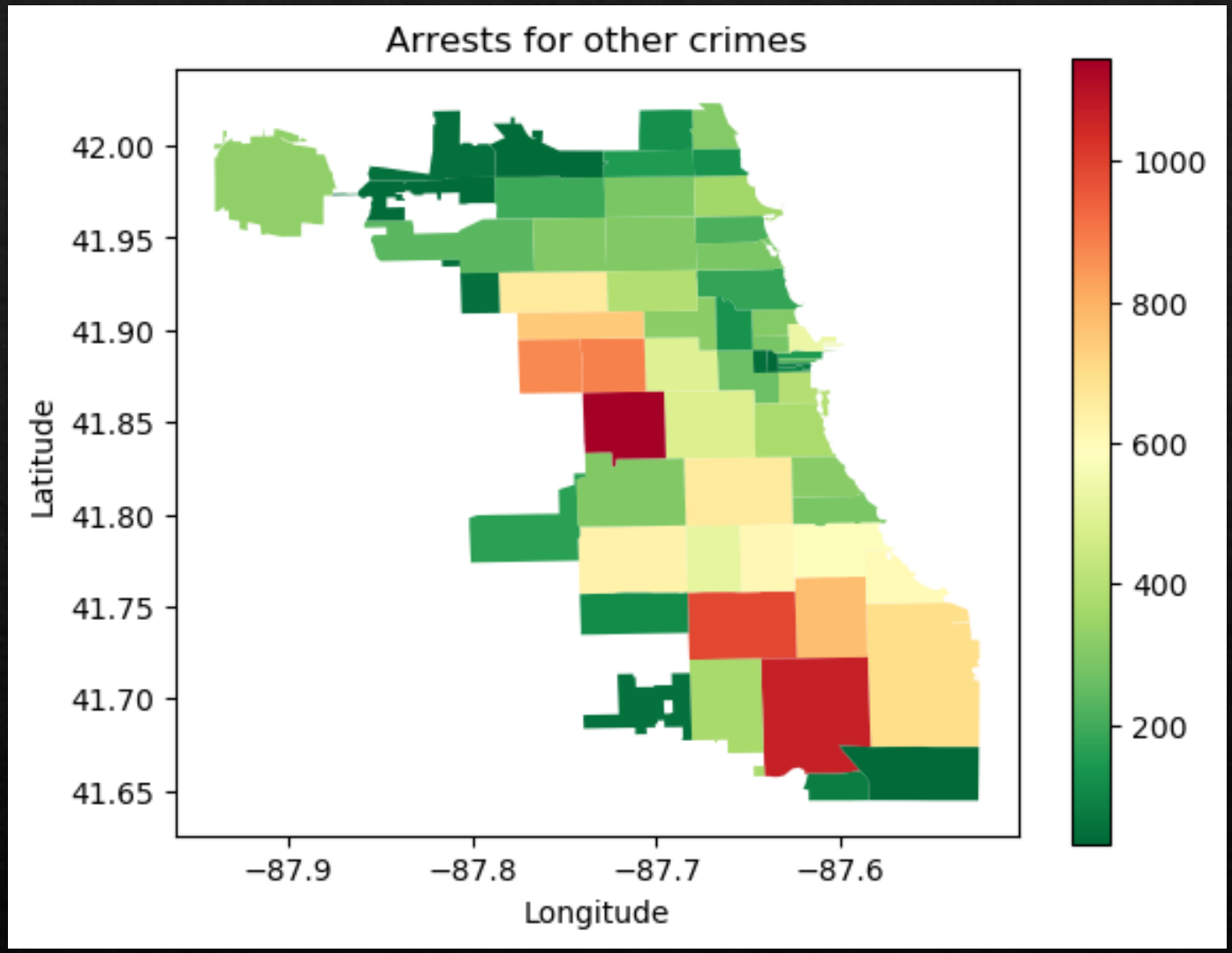
Comparing the distribution of Prostitution and Drug Abuse cases

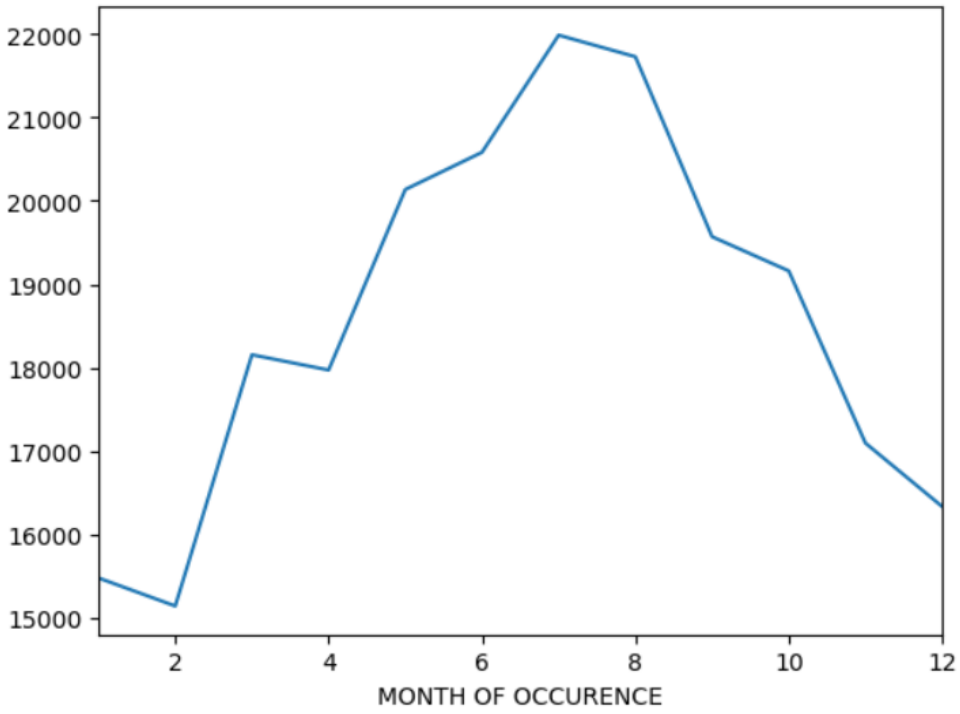
Hypothesis

The areas with high Prostitution and Drug Abuse should have higher arrest rates for other types of crimes compared to the neighboring areas.

Arrests for crimes other than Drug Abuse and Prostitution

- Areas with high Drug Abuse and Prostitution have significantly higher arrest rates for other crimes too
- Our Hypothesis is CORRECT!





Crime Rates by Months

Revisiting our Initial Hypotheses

- Crimes related to Theft are the highest type of crime: TRUE!
- Crimes related to drug consumption take place in secluded regions away from downtown: TRUE!

THANK YOU