CS-2120-Arailym Yntalbekova

STEP #1: Project Endterm

- Analyzing a real dataset related to law enforcement using the following data analysis tools: Pandas, NumPy, SQL.
- This project aims to analyze the dataset and identify patterns or trends that can help law enforcement agencies with their operations.
- The Chicago Crime dataset contains a summary of reported crimes in the City of Chicago from 2001 through 2017.
- Datasource: https://www.kaggle.com/currie32/crimes-in-chicago

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from fbprophet import Prophet

chicago_df_1 = pd.read_csv('E:/DL and ML Practical Tutorials -
Package/Project
3/Chicago_Crimes_2005_to_2007.csv',error_bad_lines=False)
chicago_df_2 = pd.read_csv('E:/DL and ML Practical Tutorials -
Package/Project
3/Chicago_Crimes_2008_to_2011.csv',error_bad_lines=False)
chicago_df_3 = pd.read_csv('E:/DL and ML Practical Tutorials -
Package/Project
3/Chicago_Crimes_2012_to_2017.csv',error_bad_lines=False)
```

```
b'Skipping line 533719: expected 23 fields, saw 24\n'
b'Skipping line 1149094: expected 23 fields, saw 41\n'
```

```
chicago_df_1.shape
```

```
(1872343, 23)
```

```
chicago_df_2.shape
```

```
(2688710, 23)
```

```
chicago_df_3.shape
```

```
(1456714, 23)
```

```
chicago_df = pd.concat([chicago_df_1,chicago_df_2,chicago_df_3])

chicago_df.shape
```

```
(6017767, 23)
```

THE DataSet

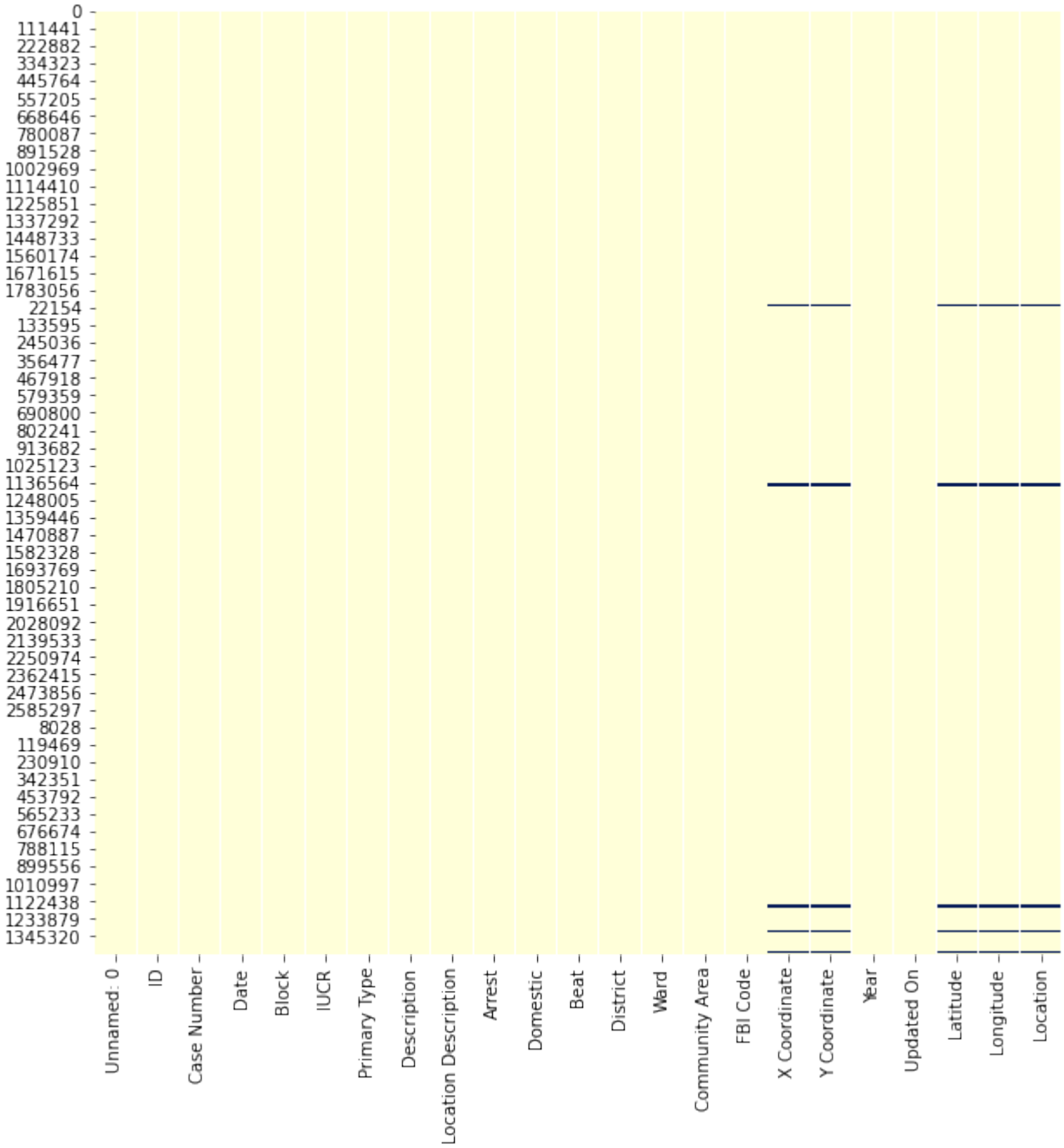
```
chicago_df.head(5)
```

```
   Unnamed: 0       ID Case Number                    Date  \
0           0  4673626    HM274058  04/02/2006 01:00:00 PM
1           1  4673627    HM202199  02/26/2006 01:40:48 PM
2           2  4673628    HM113861  01/08/2006 11:16:00 PM
3           4  4673629    HM274049  04/05/2006 06:45:00 PM
4           5  4673630    HM187120  02/17/2006 09:03:14 PM

                  Block  IUCR   Primary Type
Description  \
0    055XX N MANGO AVE  2825  OTHER OFFENSE       HARASSMENT BY
TELEPHONE
1   065XX S RHODES AVE  2017      NARCOTICS
MANU/DELIVER:CRACK
2      013XX E 69TH ST  051A        ASSAULT           AGGRAVATED:
HANDGUN
3  061XX W NEWPORT AVE  0460        BATTERY
SIMPLE
4      037XX W 60TH ST  1811      NARCOTICS  POSS: CANNABIS 30GMS OR
LESS

  Location Description  Arrest  ...  Ward  Community Area  FBI Code  \
0            RESIDENCE   False  ...  45.0            11.0        26
1             SIDEWALK    True  ...  20.0            42.0        18
2                OTHER   False  ...   5.0            69.0       04A
3            RESIDENCE   False  ...  38.0            17.0       08B
4                ALLEY    True  ...  13.0            65.0        18

   X Coordinate  Y Coordinate  Year              Updated On   Latitude
\
0     1136872.0     1936499.0  2006  04/15/2016 08:55:02 AM  41.981913

1     1181027.0     1861693.0  2006  04/15/2016 08:55:02 AM  41.775733

2     1186023.0     1859609.0  2006  04/15/2016 08:55:02 AM  41.769897

3     1134772.0     1922299.0  2006  04/15/2016 08:55:02 AM  41.942984

4     1152412.0     1864560.0  2006  04/15/2016 08:55:02 AM  41.784211


   Longitude                          Location
0 -87.771996  (41.981912692, -87.771996382)
1 -87.611920  (41.775732538, -87.611919814)
2 -87.593671  (41.769897392, -87.593670899)
```

```
3 -87.780057  (41.942984005, -87.780056951)
4 -87.716745   (41.784210853, -87.71674491)

[5 rows x 23 columns]
```

```
plt.figure(figsize = (10,10))
sns.heatmap(chicago_df.isnull(),cbar = False, cmap = 'YlGnBu')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x28725243dc0>
```



```
chicago_df.drop(['Unnamed: 0','Case Number','ID','IUCR','X
Coordinate','Y Coordinate','Updated On','Year','FBI
Code','Beat','Ward','Community
Area','Location','Latitude','Longitude','District'], inplace
=True,axis =1)

chicago_df
```

```
                          Date                  Block     Primary Type
\
0        04/02/2006 01:00:00 PM     055XX N MANGO AVE    OTHER OFFENSE

1        02/26/2006 01:40:48 PM    065XX S RHODES AVE        NARCOTICS

2        01/08/2006 11:16:00 PM       013XX E 69TH ST          ASSAULT

3        04/05/2006 06:45:00 PM   061XX W NEWPORT AVE          BATTERY

4        02/17/2006 09:03:14 PM       037XX W 60TH ST        NARCOTICS

...                         ...                   ...              ...

1456709  05/03/2016 11:33:00 PM       026XX W 23RD PL          BATTERY

1456710  05/03/2016 11:30:00 PM   073XX S HARVARD AVE  CRIMINAL DAMAGE

1456711  05/03/2016 12:15:00 AM       024XX W 63RD ST          BATTERY

1456712  05/03/2016 09:07:00 PM  082XX S EXCHANGE AVE          BATTERY

1456713  05/03/2016 11:38:00 PM       001XX E 75TH ST    OTHER OFFENSE


                           Description            Location Description
Arrest  \
0             HARASSMENT BY TELEPHONE                       RESIDENCE
False
1                  MANU/DELIVER:CRACK                        SIDEWALK
True
2                 AGGRAVATED: HANDGUN                           OTHER
False
3                              SIMPLE                       RESIDENCE
False
4        POSS: CANNABIS 30GMS OR LESS                           ALLEY
True
...                                ...                             ...
...
1456709       DOMESTIC BATTERY SIMPLE                       APARTMENT
True
1456710                    TO PROPERTY                       APARTMENT
True
```

```
1456711           AGGRAVATED: HANDGUN                           SIDEWALK
False
1456712       DOMESTIC BATTERY SIMPLE                           SIDEWALK
False
1456713       OTHER WEAPONS VIOLATION  PARKING LOT/GARAGE(NON.RESID.)
True

         Domestic
0           False
1           False
2           False
3           False
4           False
...           ...
1456709      True
1456710      True
1456711     False
1456712      True
1456713     False

[6017767 rows x 7 columns]
```

```
# Assembling a datetime by rearranging the dataframe column "Date".

chicago_df.Date = pd.to_datetime(chicago_df.Date, format='%m/%d/%Y %I:
%M:%S %p')

chicago_df.Date
```

```
0         2006-04-02 13:00:00
1         2006-02-26 13:40:48
2         2006-01-08 23:16:00
3         2006-04-05 18:45:00
4         2006-02-17 21:03:14
                  ...
1456709   2016-05-03 23:33:00
1456710   2016-05-03 23:30:00
1456711   2016-05-03 00:15:00
1456712   2016-05-03 21:07:00
1456713   2016-05-03 23:38:00
Name: Date, Length: 6017767, dtype: datetime64[ns]
```

```
# setting the index to be the date
chicago_df.index = pd.DatetimeIndex(chicago_df.Date)

chicago_df
```

```
                                    Date                  Block  \
Date
2006-04-02 13:00:00 2006-04-02 13:00:00     055XX N MANGO AVE
2006-02-26 13:40:48 2006-02-26 13:40:48    065XX S RHODES AVE
```

```
2006-01-08 23:16:00 2006-01-08 23:16:00       013XX E 69TH ST
2006-04-05 18:45:00 2006-04-05 18:45:00   061XX W NEWPORT AVE
2006-02-17 21:03:14 2006-02-17 21:03:14       037XX W 60TH ST
...                                 ...                   ...
2016-05-03 23:33:00 2016-05-03 23:33:00       026XX W 23RD PL
2016-05-03 23:30:00 2016-05-03 23:30:00   073XX S HARVARD AVE
2016-05-03 00:15:00 2016-05-03 00:15:00       024XX W 63RD ST
2016-05-03 21:07:00 2016-05-03 21:07:00  082XX S EXCHANGE AVE
2016-05-03 23:38:00 2016-05-03 23:38:00       001XX E 75TH ST

                        Primary Type                   Description  \
Date
2006-04-02 13:00:00    OTHER OFFENSE       HARASSMENT BY TELEPHONE
2006-02-26 13:40:48        NARCOTICS            MANU/DELIVER:CRACK
2006-01-08 23:16:00          ASSAULT           AGGRAVATED: HANDGUN
2006-04-05 18:45:00          BATTERY                        SIMPLE
2006-02-17 21:03:14        NARCOTICS  POSS: CANNABIS 30GMS OR LESS
...                               ...                          ...
2016-05-03 23:33:00          BATTERY       DOMESTIC BATTERY SIMPLE
2016-05-03 23:30:00  CRIMINAL DAMAGE                   TO PROPERTY
2016-05-03 00:15:00          BATTERY           AGGRAVATED: HANDGUN
2016-05-03 21:07:00          BATTERY       DOMESTIC BATTERY SIMPLE
2016-05-03 23:38:00    OTHER OFFENSE       OTHER WEAPONS VIOLATION

                               Location Description  Arrest  Domestic

Date

2006-04-02 13:00:00                         RESIDENCE   False     False

2006-02-26 13:40:48                          SIDEWALK    True     False

2006-01-08 23:16:00                             OTHER   False     False

2006-04-05 18:45:00                         RESIDENCE   False     False

2006-02-17 21:03:14                             ALLEY    True     False

...                                               ...     ...       ...

2016-05-03 23:33:00                         APARTMENT    True      True

2016-05-03 23:30:00                         APARTMENT    True      True

2016-05-03 00:15:00                          SIDEWALK   False     False

2016-05-03 21:07:00                          SIDEWALK   False      True

2016-05-03 23:38:00  PARKING LOT/GARAGE(NON.RESID.)    True     False
```

```
[6017767 rows x 7 columns]
```

```
chicago_df['Primary Type'].value_counts()
```

```
THEFT                                1245111
BATTERY                              1079178
CRIMINAL DAMAGE                       702702
NARCOTICS                             674831
BURGLARY                              369056
OTHER OFFENSE                         368169
ASSAULT                               360244
MOTOR VEHICLE THEFT                   271624
ROBBERY                               229467
DECEPTIVE PRACTICE                    225180
CRIMINAL TRESPASS                     171596
PROSTITUTION                           60735
WEAPONS VIOLATION                      60335
PUBLIC PEACE VIOLATION                 48403
OFFENSE INVOLVING CHILDREN             40260
CRIM SEXUAL ASSAULT                    22789
SEX OFFENSE                            20172
GAMBLING                               14755
INTERFERENCE WITH PUBLIC OFFICER       14009
LIQUOR LAW VIOLATION                   12129
ARSON                                   9269
HOMICIDE                                5879
KIDNAPPING                              4734
INTIMIDATION                            3324
STALKING                                2866
OBSCENITY                                422
PUBLIC INDECENCY                         134
OTHER NARCOTIC VIOLATION                 122
NON-CRIMINAL                              96
CONCEALED CARRY LICENSE VIOLATION         90
NON - CRIMINAL                            38
HUMAN TRAFFICKING                         28
RITUALISM                                 16
NON-CRIMINAL (SUBJECT SPECIFIED)           4
Name: Primary Type, dtype: int64
```

```
chicago_df['Primary Type'].value_counts().iloc[:15]
```

```
THEFT                          1245111
BATTERY                        1079178
CRIMINAL DAMAGE                 702702
NARCOTICS                       674831
BURGLARY                        369056
OTHER OFFENSE                   368169
ASSAULT                         360244
```

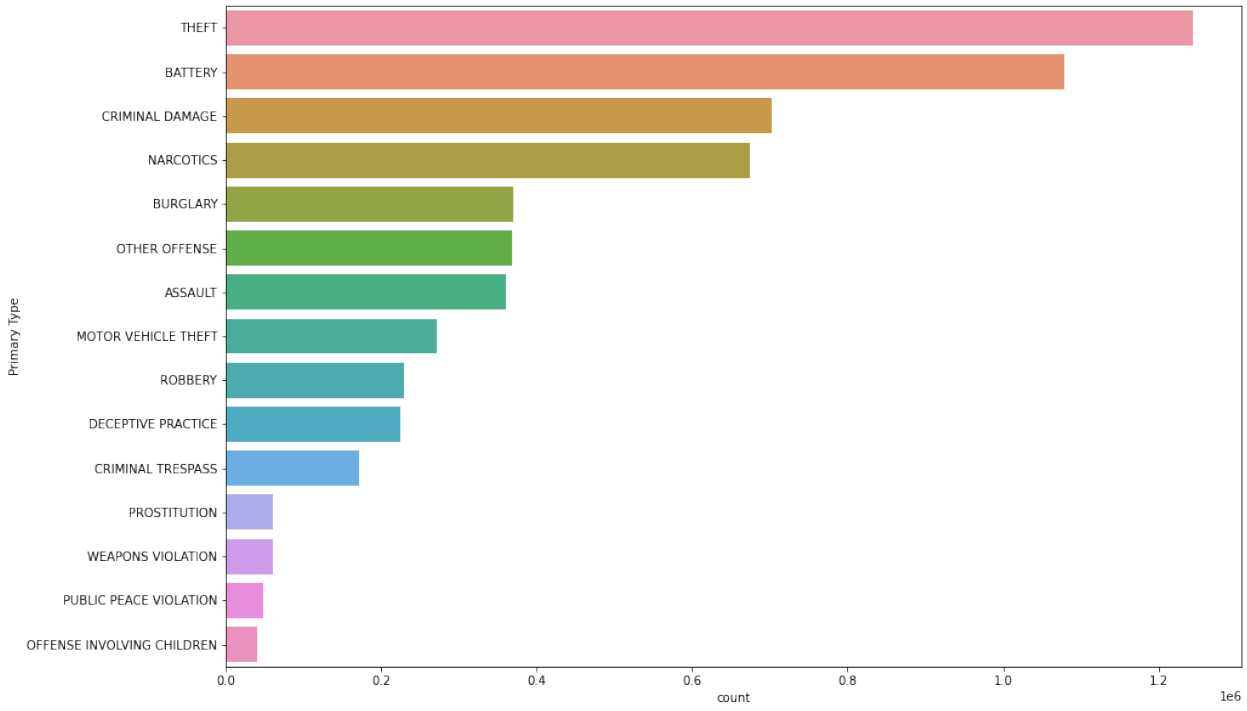
```
MOTOR VEHICLE THEFT            271624
ROBBERY                        229467
DECEPTIVE PRACTICE             225180
CRIMINAL TRESPASS              171596
PROSTITUTION                    60735
WEAPONS VIOLATION               60335
PUBLIC PEACE VIOLATION          48403
OFFENSE INVOLVING CHILDREN      40260
Name: Primary Type, dtype: int64
```

```
order_data = chicago_df['Primary Type'].value_counts().iloc[:15].index
order_data
```

```
Index(['THEFT', 'BATTERY', 'CRIMINAL DAMAGE', 'NARCOTICS', 'BURGLARY',
       'OTHER OFFENSE', 'ASSAULT', 'MOTOR VEHICLE THEFT', 'ROBBERY',
       'DECEPTIVE PRACTICE', 'CRIMINAL TRESPASS', 'PROSTITUTION',
       'WEAPONS VIOLATION', 'PUBLIC PEACE VIOLATION',
       'OFFENSE INVOLVING CHILDREN'],
      dtype='object')
```

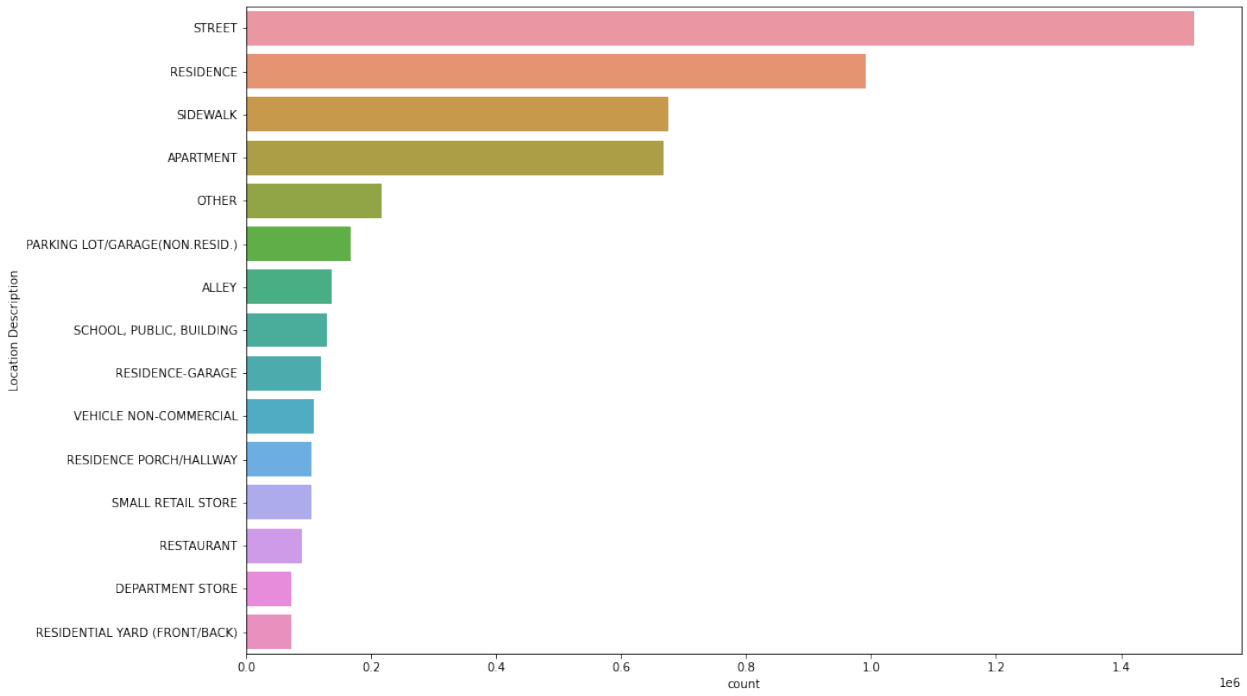
```
plt.figure(figsize = (15,10))
sns.countplot(y = 'Primary Type', data = chicago_df, order =order_data
)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x287826d3730>
```



```
plt.figure(figsize = (15,10))
sns.countplot(y = 'Location Description', data = chicago_df, order
=chicago_df['Location Description'].value_counts().iloc[:15].index)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x28794e52af0>
```




```
chicago_df.resample('Y').size()
```

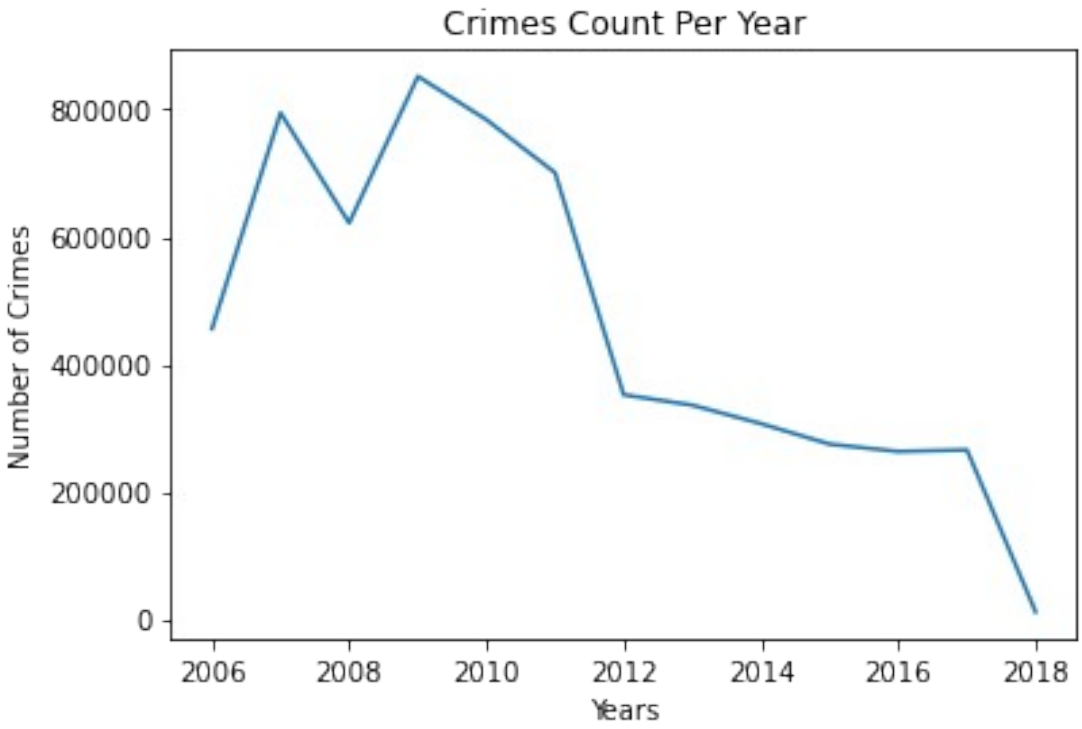
```
Date
2005-12-31    455811
2006-12-31    794684
2007-12-31    621848
2008-12-31    852053
2009-12-31    783900
2010-12-31    700691
2011-12-31    352066
2012-12-31    335670
2013-12-31    306703
2014-12-31    274527
2015-12-31    262995
2016-12-31    265462
2017-12-31     11357
Freq: A-DEC, dtype: int64
```

```
plt.plot(chicago_df.resample('Y'))
```

```
# Resample is a Convenience method for frequency conversion and
resampling of time series.
```

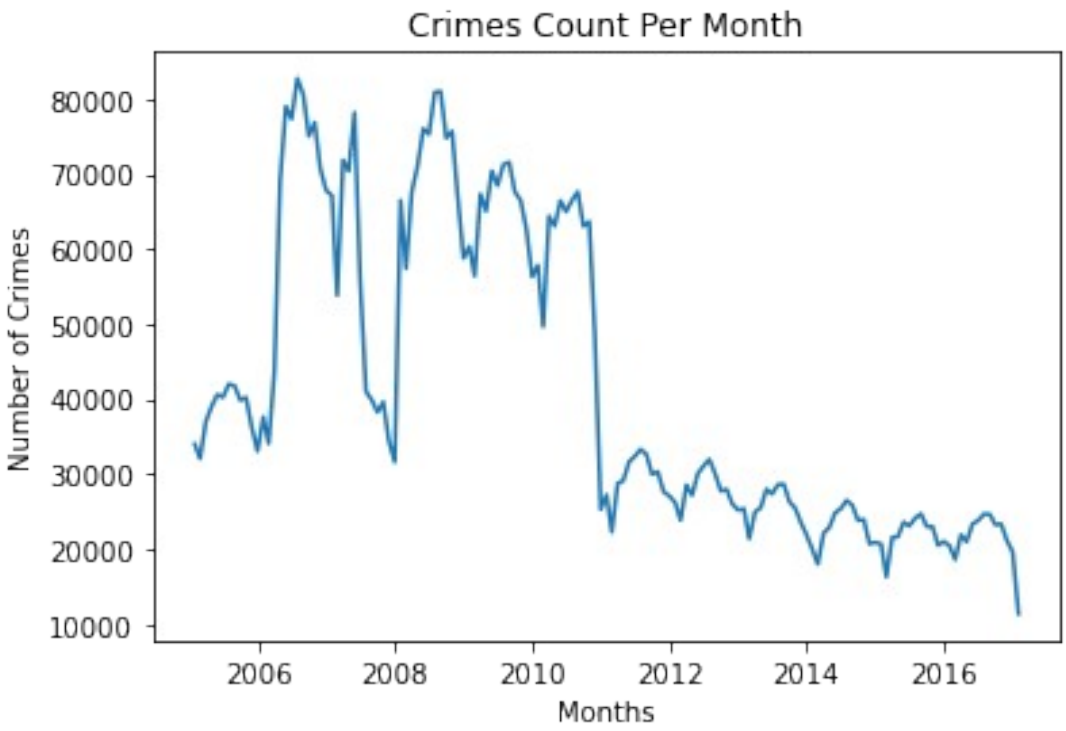
```
plt.plot(chicago_df.resample('Y').size())
plt.title('Crimes Count Per Year')
plt.xlabel('Years')
plt.ylabel('Number of Crimes')
```

```
Text(0, 0.5, 'Number of Crimes')
```

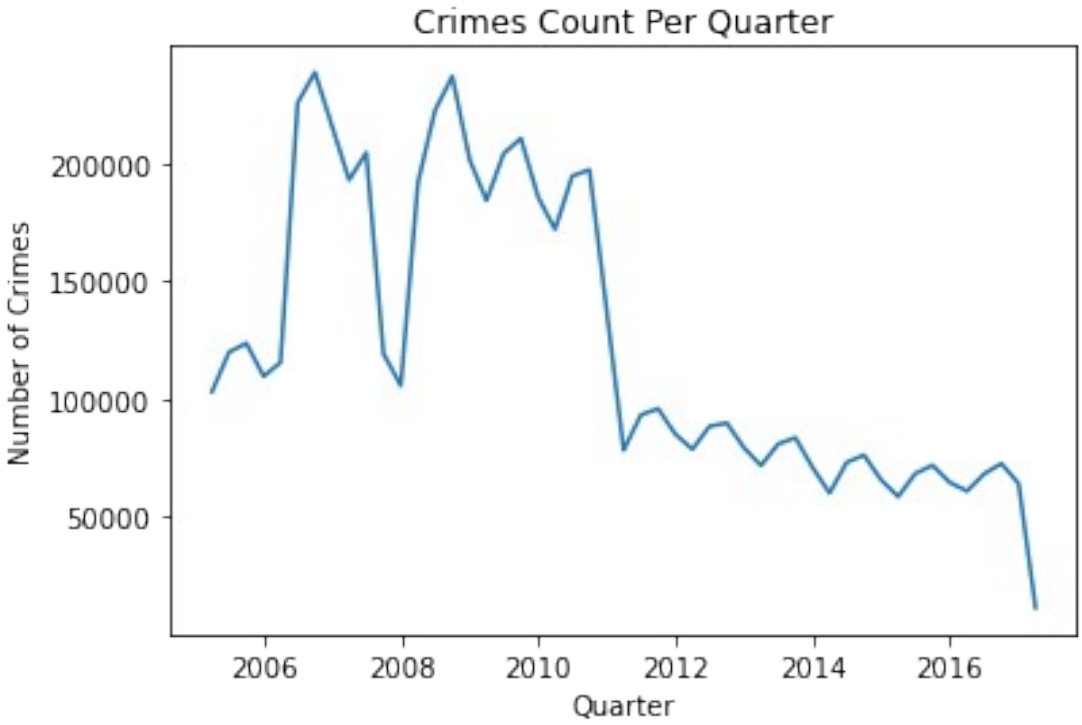



```
plt.plot(chicago_df.resample('M').size())
plt.title('Crimes Count Per Month')
plt.xlabel('Months')
plt.ylabel('Number of Crimes')
```

```
Text(0, 0.5, 'Number of Crimes')
```




```
plt.plot(chicago_df.resample('Q').size())
plt.title('Crimes Count Per Quarter')
plt.xlabel('Quarter')
plt.ylabel('Number of Crimes')

Text(0, 0.5, 'Number of Crimes')
```




```
chicago_prophet = chicago_df.resample('M').size().reset_index()

chicago_prophet
```

```
          Date      0
0   2005-01-31  33983
1   2005-02-28  32042
2   2005-03-31  36970
3   2005-04-30  38963
4   2005-05-31  40572
..         ...    ...
140 2016-09-30  23235
141 2016-10-31  23314
142 2016-11-30  21140
143 2016-12-31  19580
144 2017-01-31  11357

[145 rows x 2 columns]
```

```
chicago_prophet.columns = ['Date', 'Crime_counts']

chicago_prophet
```

```
          Date  Crime_counts
0   2005-01-31         33983
1   2005-02-28         32042
2   2005-03-31         36970
3   2005-04-30         38963
```

```
4   2005-05-31         40572
..         ...           ...
140 2016-09-30         23235
141 2016-10-31         23314
142 2016-11-30         21140
143 2016-12-31         19580
144 2017-01-31         11357

[145 rows x 2 columns]
```

```
chicago_prophet_df_final = chicago_prophet.rename(columns = {'Date':
'ds', 'Crime_counts': 'y'})

chicago_prophet_df_final
```

```
            ds      y
0   2005-01-31  33983
1   2005-02-28  32042
2   2005-03-31  36970
3   2005-04-30  38963
4   2005-05-31  40572
..         ...    ...
140 2016-09-30  23235
141 2016-10-31  23314
142 2016-11-30  21140
143 2016-12-31  19580
144 2017-01-31  11357

[145 rows x 2 columns]
```

MAKE PREDICTIONS

```
m  = Prophet()
m.fit(chicago_prophet_df_final)
```

```
INFO:fbprophet:Disabling weekly seasonality. Run prophet with
weekly_seasonality=True to override this.
INFO:fbprophet:Disabling daily seasonality. Run prophet with
daily_seasonality=True to override this.

<fbprophet.forecaster.Prophet at 0x287824acc40>
```

```
# Forcasting into the future
future = m.make_future_dataframe(periods=1000)
forecast = m.predict(future)

forecast
```

```
             ds         trend    yhat_lower    yhat_upper
trend_lower  \
0    2005-01-31  60379.720403  38941.159281  71982.422114
```

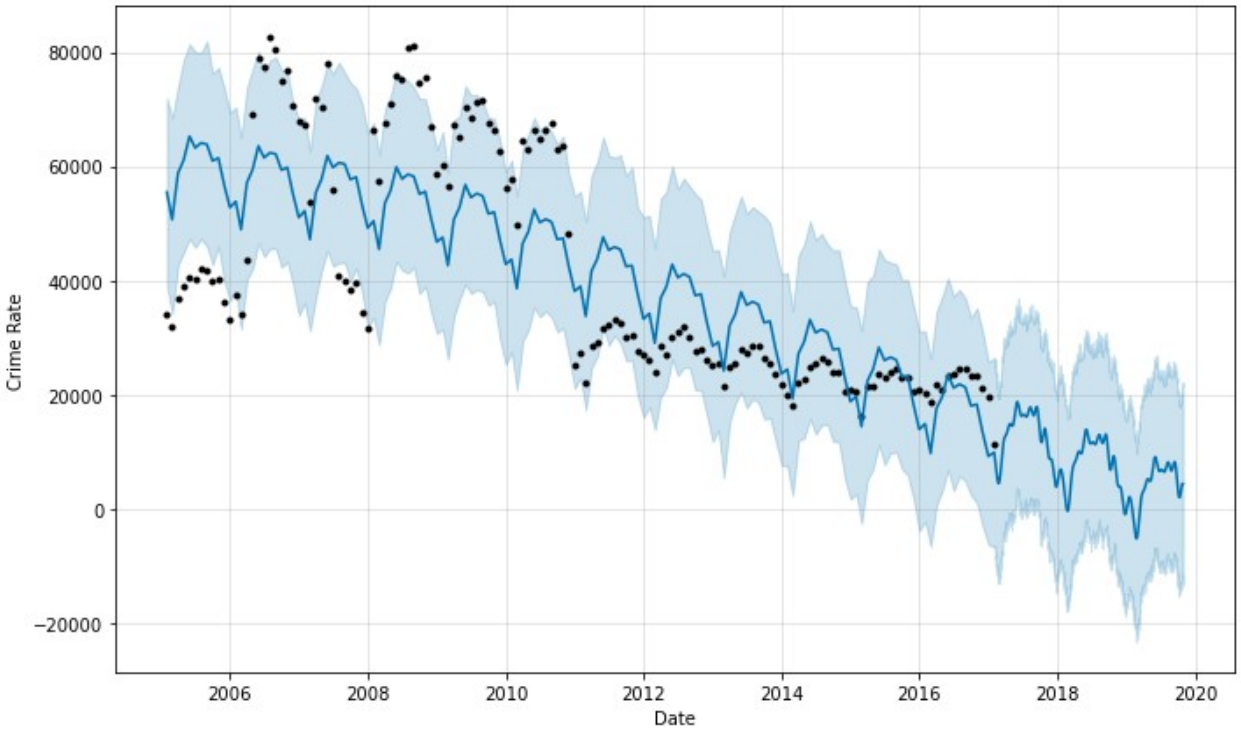
```
60379.720403
1    2005-02-28  60249.704311  34165.573635  68462.552422
60249.704311
2    2005-03-31  60105.757924  42655.432616  74030.080695
60105.757924
3    2005-04-30  59966.454969  44973.102872  78769.127031
59966.454969
4    2005-05-31  59822.508582  47334.657440  81416.439739
59822.508582
...         ...           ...           ...           ...           ..
.
1140 2019-10-24   1645.467200 -12607.346885  22145.681729
917.518239
1141 2019-10-25   1632.246809 -13499.713030  21345.965855
901.834742
1142 2019-10-26   1619.026419 -11907.586826  22232.520866
887.457189
1143 2019-10-27   1605.806028 -13009.673234  20657.235898
873.298780
1144 2019-10-28   1592.585637 -11691.571269  20229.780677
859.227398

       trend_upper  additive_terms  additive_terms_lower  \
0     60379.720403    -4888.559945          -4888.559945
1     60249.704311    -9556.480896          -9556.480896
2     60105.757924    -1267.695327          -1267.695327
3     59966.454969     1143.949615           1143.949615
4     59822.508582     5464.443944           5464.443944
...            ...             ...                   ...
1140   2439.652311     2693.371392           2693.371392
1141   2428.827282     2782.011034           2782.011034
1142   2416.571843     2831.398202           2831.398202
1143   2404.306421     2839.437402           2839.437402
1144   2392.040999     2805.048280           2805.048280

      additive_terms_upper       yearly  yearly_lower  yearly_upper  \
0             -4888.559945 -4888.559945  -4888.559945  -4888.559945
1             -9556.480896 -9556.480896  -9556.480896  -9556.480896
2             -1267.695327 -1267.695327  -1267.695327  -1267.695327
3              1143.949615  1143.949615   1143.949615   1143.949615
4              5464.443944  5464.443944   5464.443944   5464.443944
...                    ...          ...           ...           ...
1140           2693.371392  2693.371392   2693.371392   2693.371392
1141           2782.011034  2782.011034   2782.011034   2782.011034
1142           2831.398202  2831.398202   2831.398202   2831.398202
1143           2839.437402  2839.437402   2839.437402   2839.437402
1144           2805.048280  2805.048280   2805.048280   2805.048280

      multiplicative_terms  multiplicative_terms_lower  \
0                      0.0                         0.0
```

```
1                          0.0                           0.0
2                          0.0                           0.0
3                          0.0                           0.0
4                          0.0                           0.0
...                        ...                           ...
1140                       0.0                           0.0
1141                       0.0                           0.0
1142                       0.0                           0.0
1143                       0.0                           0.0
1144                       0.0                           0.0
```

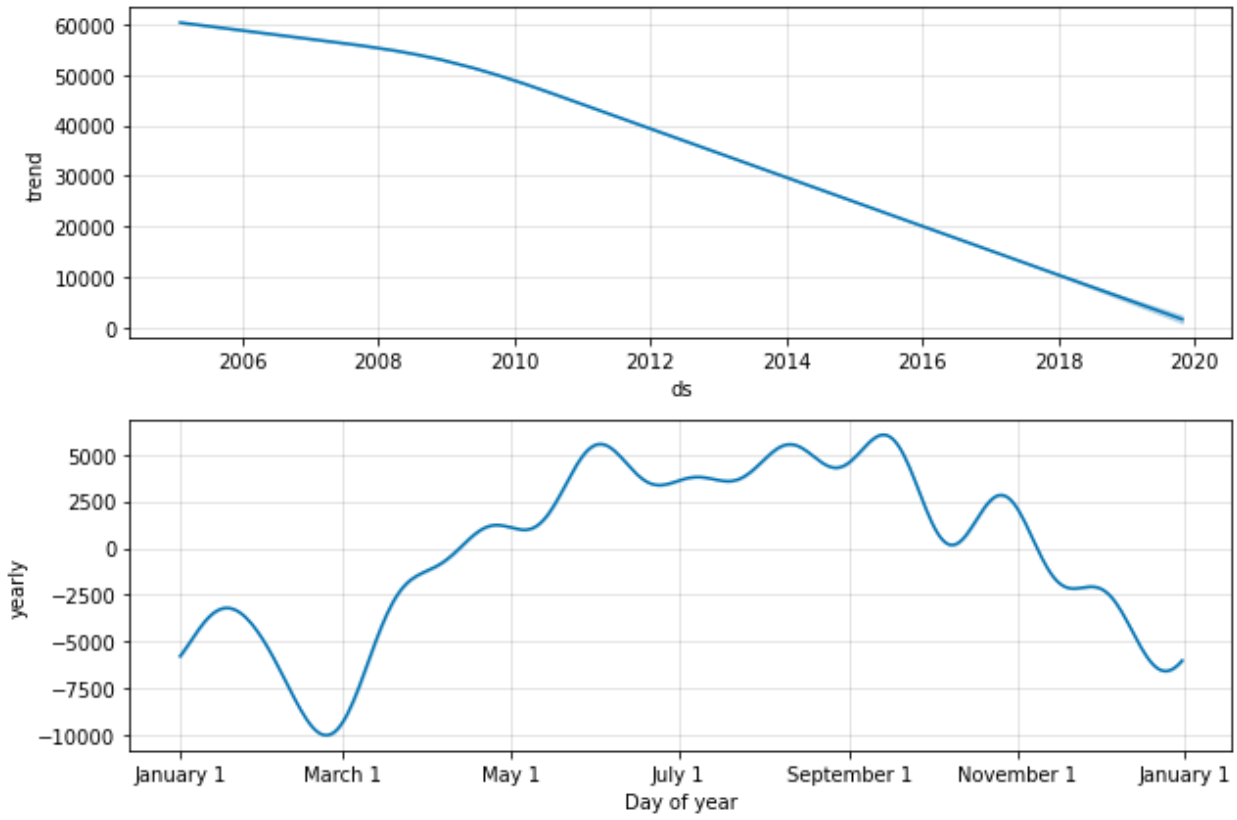
	multiplicative_terms_upper	yhat
0	0.0	55491.160458
1	0.0	50693.223415
2	0.0	58838.062597
3	0.0	61110.404583
4	0.0	65286.952525
...	...	...
1140	0.0	4338.838593
1141	0.0	4414.257844
1142	0.0	4450.424620
1143	0.0	4445.243430
1144	0.0	4397.633917

[1145 rows x 16 columns]

```
figure = m.plot(forecast,xlabel= 'Date',ylabel= 'Crime Rate')
```



```
figure3 = m.plot_components(forecast)
```



Conclusion

To summarize, the Chicago Police Department dataset is proving valuable for understanding and analyzing criminal incidents in the city. A variety of dataset features covering crime types, locations, arrests, and domestic incidents allow for a variety of analyses. Researchers and law enforcement can use it to understand crime patterns and develop targeted strategies. Redacting specific location information strikes a balance between privacy and usefulness, emphasizing responsible data use. Overall, this dataset has significant potential to improve public safety research and policing strategies in Chicago.

https://github.com/Arailymmmmmmmmmmmm/Endterm