LA Crime 2020 - 2023

May 6, 2024

1 Unveiling Patterns of Crime in Los Angeles: An Analysis of Crime Incidents from 2020 to 2023

1.1 Abstract:

This analysis explores a comprehensive dataset of crime incidents in the City of Los Angeles from 2020 to 2023, sourced from original crime reports. The dataset offers a wealth of information for understanding the trends, patterns, and dynamics of criminal activity within the city. By leveraging data-driven insights, this analysis aims to uncover valuable insights that can inform law enforcement strategies, resource allocation, and public safety initiatives.

The study begins by performing data cleaning and preprocessing steps to ensure data integrity and consistency. This includes handling missing values, removing irrelevant columns, converting data types, and filtering out invalid data points. Despite potential inaccuracies inherent in manually recorded crime reports and privacy measures limiting address specificity, the cleaned dataset remains a reliable resource for examining crime patterns in Los Angeles.

Through exploratory data analysis and visualization techniques, this study investigates various aspects of crime in the city. It examines the temporal distribution of crime incidents, revealing trends over time and seasonal variations. The analysis also delves into the demographic characteristics of crime victims, shedding light on the age distribution and identifying vulnerable populations.

Furthermore, the study explores the geographical distribution of crime, identifying areas with the highest crime rates during different time periods, such as nighttime hours. By uncovering these spatial patterns, the analysis provides actionable insights for targeted law enforcement interventions and community-based crime prevention strategies.

The findings of this analysis have significant implications for public safety, law enforcement resource allocation, and policy development. By understanding the temporal, demographic, and spatial dimensions of crime in Los Angeles, stakeholders can make data-driven decisions to enhance public safety, optimize law enforcement efforts, and develop targeted interventions in high-risk areas.

For these reasons, this analysis contributes to a deeper understanding of the crime landscape in Los Angeles and serves as a valuable resource for researchers, policymakers, and law enforcement agencies in their efforts to combat crime and promote public safety in the city.

```
[1]: import pandas as pd
     import matplotlib.pyplot as plt
     crime = pd.read_csv('../data_course/Crime_Data_from_2020_to_Present.csv')
```

```
# Suppress the SettingWithCopyWarning
pd.options.mode.chained_assignment = None
```

2 Understand and Clean the Data

In this section, we perform initial data exploration and cleaning steps. This includes examining the data schema, checking for null values, removing irrelevant columns, converting data types, and filtering out invalid data points.

2.1 Key observations:

- The dataset contains 27 columns with information about crime incidents.
- Several columns have high null values and are removed for analysis.
- Date columns are converted to datetime format for easier manipulation.
- Invalid victim ages (negative or 0) are filtered out.

```
[2]: crime.head()
```

```
[2]:    division_number date_reported      date_occurred  area    area_name  \
    0         10304468   2020-01-08  2020-01-08 22:30:00     3    Southwest
    1        190101086   2020-01-02  2020-01-01 03:30:00     1      Central
    2        200110444   2020-04-14  2020-02-13 12:00:00     1      Central
    3        191501505   2020-01-01  2020-01-01 17:30:00    15  N Hollywood
    4        191921269   2020-01-01  2020-01-01 04:15:00    19      Mission


       reporting_district  part  crime_code  \
    0                 377     2         624
    1                 163     2         624
    2                 155     2         845
    3                1543     2         745
    4                1998     2         740


                                    crime_description  modus_operandi  … \
    0                        BATTERY - SIMPLE ASSAULT       0444 0913  …
    1                        BATTERY - SIMPLE ASSAULT  0416 1822 1414  …
    2         SEX OFFENDER REGISTRANT OUT OF COMPLIANCE            1501  …
    3            VANDALISM - MISDEAMEANOR ($399 OR UNDER)       0329 1402  …
    4  VANDALISM - FELONY ($400 & OVER, ALL CHURCH VA…            0329  …


       status status_description crime_code_1  crime_code_2 crime_code_3  \
    0     AO         Adult Other        624.0           NaN          NaN
    1     IC         Invest Cont        624.0           NaN          NaN
    2     AA        Adult Arrest        845.0           NaN          NaN
    3     IC         Invest Cont        745.0         998.0          NaN
    4     IC         Invest Cont        740.0           NaN          NaN


       crime_code_4                                    location cross_street  \
```

```
0          NaN   1100 W  39TH                    PL         NaN
1          NaN    700 S  HILL                    ST         NaN
2          NaN    200 E  6TH                     ST         NaN
3          NaN   5400    CORTEEN                  PL         NaN
4          NaN  14400    TITUS                    ST         NaN

   latitude  longitude
0  34.0141   -118.2978
1  34.0459   -118.2545
2  34.0448   -118.2474
3  34.1685   -118.4019
4  34.2198   -118.4468

[5 rows x 27 columns]
```

[3]: `crime.columns`

```
[3]: Index(['division_number', 'date_reported', 'date_occurred', 'area',
            'area_name', 'reporting_district', 'part', 'crime_code',
            'crime_description', 'modus_operandi', 'victim_age', 'victim_sex',
            'victim_descent', 'premise_code', 'premise_description', 'weapon_code',
            'weapon_description', 'status', 'status_description', 'crime_code_1',
            'crime_code_2', 'crime_code_3', 'crime_code_4', 'location',
            'cross_street', 'latitude', 'longitude'],
           dtype='object')
```

[4]: `crime.shape`

[4]: `(852950, 27)`

[5]: `crime.dtypes`

```
[5]: division_number          int64
     date_reported           object
     date_occurred           object
     area                     int64
     area_name               object
     reporting_district       int64
     part                     int64
     crime_code               int64
     crime_description       object
     modus_operandi          object
     victim_age               int64
     victim_sex              object
     victim_descent          object
     premise_code           float64
     premise_description     object
```

```
weapon_code              float64
weapon_description        object
status                    object
status_description        object
crime_code_1             float64
crime_code_2             float64
crime_code_3             float64
crime_code_4             float64
location                  object
cross_street              object
latitude                 float64
longitude                float64
dtype: object
```

[6]: ```
#Check for Null values
crime.isna().sum()
```

[6]: ```
division_number               0
date_reported                 0
date_occurred                 0
area                          0
area_name                     0
reporting_district            0
part                          0
crime_code                    0
crime_description             0
modus_operandi           118311
victim_age                    0
victim_sex               112606
victim_descent           112614
premise_code                 10
premise_description         518
weapon_code              556202
weapon_description       556202
status                        0
status_description            0
crime_code_1                 11
crime_code_2             790429
crime_code_3             850837
crime_code_4             852888
location                      0
cross_street             717289
latitude                      0
longitude                     0
dtype: int64
```

```
[7]: #Remove high null and irrelevant columns
     df = crime.drop(labels = ['modus_operandi', 'weapon_description',\
                               'weapon_code', 'crime_code_2',\
                               'crime_code_3', 'crime_code_4',\
                               'cross_street'], axis = 1)
```

```
[8]: df
```

```
[8]:         division_number date_reported      date_occurred  area    area_name  \
     0              10304468    2020-01-08  2020-01-08 22:30:00     3    Southwest
     1             190101086    2020-01-02  2020-01-01 03:30:00     1      Central
     2             200110444    2020-04-14  2020-02-13 12:00:00     1      Central
     3             191501505    2020-01-01  2020-01-01 17:30:00    15  N Hollywood
     4             191921269    2020-01-01  2020-01-01 04:15:00    19      Mission
     ...                 ...           ...                  ...   ...          ...
     852945        231606525    2023-03-22  2023-03-22 10:00:00    16     Foothill
     852946        231210064    2023-04-12  2023-04-12 16:30:00    12   77th Street
     852947        230115220    2023-07-02  2023-07-01 00:01:00     1      Central
     852948        230906458    2023-03-05  2023-03-05 09:00:00     9     Van Nuys
     852949        230319786    2023-11-10  2023-11-09 23:00:00     3    Southwest

             reporting_district  part  crime_code  \
     0                      377     2         624
     1                      163     2         624
     2                      155     2         845
     3                     1543     2         745
     4                     1998     2         740
     ...                    ...   ...         ...
     852945                1602     1         230
     852946                1239     1         230
     852947                 154     1         352
     852948                 914     2         745
     852949                 395     1         331

                                        crime_description  victim_age  \
     0                            BATTERY - SIMPLE ASSAULT          36
     1                            BATTERY - SIMPLE ASSAULT          25
     2              SEX OFFENDER REGISTRANT OUT OF COMPLIANCE        0
     3                 VANDALISM - MISDEAMEANOR ($399 OR UNDER)      76
     4         VANDALISM - FELONY ($400 & OVER, ALL CHURCH VA…      31
     ...                                              ...         ...
     852945        ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT    25
     852946        ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT    29
     852947                                        PICKPOCKET        24
     852948            VANDALISM - MISDEAMEANOR ($399 OR UNDER)      53
     852949  THEFT FROM MOTOR VEHICLE - GRAND ($950.01 AND …      38
```

```
       victim_sex victim_descent  premise_code  \
0               F              B         501.0
1               M              H         102.0
2               X              X         726.0
3               F              W         502.0
4               X              X         409.0
…             …              …             …
852945          F              H         102.0
852946          M              B         222.0
852947          F              H         735.0
852948          F              H         502.0
852949          M              W         501.0


                              premise_description status  \
0                           SINGLE FAMILY DWELLING     AO
1                                         SIDEWALK     IC
2                                  POLICE FACILITY     AA
3            MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)     IC
4                              BEAUTY SUPPLY STORE     IC
…                                              …      …
852945                                    SIDEWALK     IC
852946                                  LAUNDROMAT     IC
852947                NIGHT CLUB (OPEN EVENINGS ONLY)     IC
852948   MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)     IC
852949                      SINGLE FAMILY DWELLING     IC


       status_description  crime_code_1  \
0              Adult Other         624.0
1              Invest Cont         624.0
2              Adult Arrest        845.0
3              Invest Cont         745.0
4              Invest Cont         740.0
…                      …             …
852945         Invest Cont         230.0
852946         Invest Cont         230.0
852947         Invest Cont         352.0
852948         Invest Cont         745.0
852949         Invest Cont         331.0


                                location  latitude  longitude
0          1100 W   39TH               PL   34.0141  -118.2978
1           700 S   HILL               ST   34.0459  -118.2545
2           200 E   6TH                ST   34.0448  -118.2474
3          5400     CORTEEN            PL   34.1685  -118.4019
4         14400     TITUS              ST   34.2198  -118.4468
…                        …             …       …          …
852945    12800     FILMORE            ST   34.2790  -118.4116
```

```
852946   6100 S   VERMONT                    AV   33.9841   -118.2915
852947    500 S   MAIN                       ST   34.0467   -118.2485
852948  14500     HARTLAND                   ST   34.1951   -118.4487
852949   4100 S   HOBART                     BL   34.0091   -118.3078

[852950 rows x 20 columns]
```

[9]: `df.sort_values(by = 'date_reported')`

[9]:
```
        division_number date_reported        date_occurred  area     area_name  \
39025         201004026    2020-01-01  2020-01-01 18:23:00    10   West Valley
137157        202100503    2020-01-01  2020-01-01 13:15:00    21       Topanga
2389          200304056    2020-01-01  2020-01-01 14:37:00     3     Southwest
103594        201804032    2020-01-01  2020-01-01 12:50:00    18     Southeast
25891         200104434    2020-01-01  2020-01-01 02:50:00     1       Central
...                 ...           ...                  ...   ...           ...
770235        230125557    2023-12-04  2023-12-04 06:50:00     1       Central
692345        231917421    2023-12-04  2023-12-03 09:45:00    19       Mission
835683        232117291    2023-12-04  2023-03-01 12:00:00    21       Topanga
825607        231116467    2023-12-04  2023-12-01 22:30:00    11     Northeast
676338        231116462    2023-12-04  2023-11-17 12:00:00    11     Northeast

        reporting_district  part  crime_code  \
39025                 1091     1         310
137157                2189     1         343
2389                   329     2         888
103594                1823     1         210
25891                  128     1         330
...                    ...   ...         ...
770235                 111     1         330
692345                1936     1         420
835683                2125     2         354
825607                1141     1         331
676338                1173     2         354

                                         crime_description  victim_age  \
39025                                             BURGLARY          44
137157         SHOPLIFTING-GRAND THEFT ($950.01 & OVER)          75
2389                                           TRESPASSING          19
103594                                            ROBBERY          50
25891                                 BURGLARY FROM VEHICLE          26
...                                                    ...         ...
770235                                BURGLARY FROM VEHICLE           0
692345      THEFT FROM MOTOR VEHICLE - PETTY ($950 & UNDER)           0
835683                                   THEFT OF IDENTITY           0
825607  THEFT FROM MOTOR VEHICLE - GRAND ($950.01 AND …           0
676338                                   THEFT OF IDENTITY           0
```

```
       victim_sex victim_descent  premise_code  \
39025           M              O         504.0
137157          M              X         402.0
2389            X              X         255.0
103594          M              B         102.0
25891           F              W         108.0
…             …              …           …
770235        NaN            NaN         717.0
692345        NaN            NaN         101.0
835683        NaN            NaN         502.0
825607        NaN            NaN         101.0
676338        NaN            NaN         502.0

                                    premise_description status  \
39025                                   OTHER RESIDENCE     IC
137157                                           MARKET     AA
2389    AUTO DEALERSHIP (CHEVY, FORD, BMW, MERCEDES, E…     IC
103594                                         SIDEWALK     IC
25891                                      PARKING LOT     AO
…                                                     …      …
770235                                    HEALTH SPA/GYM     IC
692345                                            STREET     IC
835683      MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)     IC
825607                                            STREET     IC
676338      MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)     IC

       status_description  crime_code_1  \
39025          Invest Cont         310.0
137157         Adult Arrest        343.0
2389           Invest Cont         888.0
103594         Invest Cont         210.0
25891          Adult Other         330.0
…                  …             …
770235         Invest Cont         330.0
692345         Invest Cont         420.0
835683         Invest Cont         354.0
825607         Invest Cont         331.0
676338         Invest Cont         354.0

                                     location  latitude  longitude
39025   3800    WINFORD                    DR   34.1386  -118.5525
137157  20000 W  VENTURA                   BL   34.1719  -118.5684
2389    3300 S   FIGUEROA                   ST   34.0225  -118.2796
103594                          CENTURY   33.9456  -118.2652
25891    300 E   2ND                       ST   34.0498  -118.2400
…                                         …        …         …
```

```
770235    700 W   CESAR E CHAVEZ           AV    34.0606   -118.2439
692345   11400     AMBOY                   AV    34.2767   -118.4477
835683   21400     SATICOY                 ST    34.2119   -118.6017
825607       N     KENMORE                 AV    34.1018   -118.2973
676338    1900     ALLESANDRO              ST    34.0900   -118.2587

[852950 rows x 20 columns]
```

[10]:
```python
#Convert objects to datetimes where relevant and update datatypes
df['date_reported'] = pd.to_datetime(df['date_reported'], format = '%Y-%m-%d')
df['date_occurred'] = pd.to_datetime(df['date_occurred'], format = '%Y-%m-%d')
```

[11]:
```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 852950 entries, 0 to 852949
Data columns (total 20 columns):
 #   Column               Non-Null Count   Dtype
---  ------               --------------   -----
 0   division_number      852950 non-null  int64
 1   date_reported        852950 non-null  datetime64[ns]
 2   date_occurred        852950 non-null  datetime64[ns]
 3   area                 852950 non-null  int64
 4   area_name            852950 non-null  object
 5   reporting_district   852950 non-null  int64
 6   part                 852950 non-null  int64
 7   crime_code           852950 non-null  int64
 8   crime_description    852950 non-null  object
 9   victim_age           852950 non-null  int64
 10  victim_sex           740344 non-null  object
 11  victim_descent       740336 non-null  object
 12  premise_code         852940 non-null  float64
 13  premise_description  852432 non-null  object
 14  status               852950 non-null  object
 15  status_description   852950 non-null  object
 16  crime_code_1         852939 non-null  float64
 17  location             852950 non-null  object
 18  latitude             852950 non-null  float64
 19  longitude            852950 non-null  float64
dtypes: datetime64[ns](2), float64(4), int64(6), object(8)
memory usage: 130.1+ MB
```

[12]:
```python
df.describe(include = ['float64', 'int64'])
```

[12]:
	division_number	area	reporting_district	part
count	8.529500e+05	852950.000000	852950.000000	852950.000000
mean	2.166969e+08	10.707354	1117.165490	1.412575
std	1.100081e+07	6.097178	609.716073	0.492298

9

```
        min      8.170000e+02      1.000000        101.000000     1.000000
        25%      2.102184e+08      6.000000        615.000000     1.000000
        50%      2.202184e+08     11.000000       1141.000000     1.000000
        75%      2.301094e+08     16.000000       1617.000000     2.000000
        max      2.399306e+08     21.000000       2199.000000     2.000000

                   crime_code     victim_age     premise_code    crime_code_1  \
        count  852950.000000  852950.000000   852940.000000   852939.000000
        mean      500.746338      29.742191      305.974292      500.486350
        std       207.705242      21.799470      216.950442      207.493864
        min       110.000000      -3.000000      101.000000      110.000000
        25%       331.000000       5.000000      101.000000      331.000000
        50%       442.000000      31.000000      203.000000      442.000000
        75%       626.000000      45.000000      501.000000      626.000000
        max       956.000000     120.000000      976.000000      956.000000

                   latitude      longitude
        count  852950.000000  852950.000000
        mean       33.983232    -118.040106
        std         1.756263       6.089068
        min         0.000000    -118.667600
        25%        34.014100    -118.429700
        50%        34.058500    -118.321500
        75%        34.163200    -118.273900
        max        34.334300       0.000000
```

[13]:
```python
# Extract the 'victim_age' column into a new DataFrame
df['victim_age'].value_counts()
```

[13]:
```
0        211842
30        19421
35        19008
31        18603
29        18552
           …
97           63
-1           60
-2           13
 120          1
-3            1
Name: victim_age, Length: 103, dtype: int64
```

[14]:
```python
# Filter out invalid negative ages and age 0
df_age = df[df['victim_age'] > 0]
```

[15]:
```python
df_age['victim_age'].value_counts()
```

```
[15]: 30      19421
      35      19008
      31      18603
      29      18552
      28      18266
              …
      95         89
      96         88
      98         67
      97         63
      120         1
      Name: victim_age, Length: 99, dtype: int64
```

```
[16]: #Adding year and month columns
      df['year'] = df['date_occurred'].dt.year
      df['month'] = df['date_occurred'].dt.month
```

```
[17]: df.sort_values(by = 'date_reported')
```

```
[17]:         division_number date_reported        date_occurred  area   area_name  \
      39025          201004026    2020-01-01  2020-01-01 18:23:00    10  West Valley
      137157         202100503    2020-01-01  2020-01-01 13:15:00    21      Topanga
      2389           200304056    2020-01-01  2020-01-01 14:37:00     3    Southwest
      103594         201804032    2020-01-01  2020-01-01 12:50:00    18    Southeast
      25891          200104434    2020-01-01  2020-01-01 02:50:00     1      Central
      …                    …            …                    …     …          …
      770235         230125557    2023-12-04  2023-12-04 06:50:00     1      Central
      692345         231917421    2023-12-04  2023-12-03 09:45:00    19      Mission
      835683         232117291    2023-12-04  2023-03-01 12:00:00    21      Topanga
      825607         231116467    2023-12-04  2023-12-01 22:30:00    11    Northeast
      676338         231116462    2023-12-04  2023-11-17 12:00:00    11    Northeast

              reporting_district  part  crime_code  \
      39025                 1091     1         310
      137157                2189     1         343
      2389                   329     2         888
      103594                1823     1         210
      25891                  128     1         330
      …                       …    …           …
      770235                 111     1         330
      692345                1936     1         420
      835683                2125     2         354
      825607                1141     1         331
      676338                1173     2         354

                                       crime_description  victim_age  …  \
      39025                                     BURGLARY          44  …
```

```
137157        SHOPLIFTING-GRAND THEFT ($950.01 & OVER)              75  …
2389                                          TRESPASSING          19  …
103594                                           ROBBERY           50  …
25891                              BURGLARY FROM VEHICLE            26  …
…                                                        …          …  …
770235                             BURGLARY FROM VEHICLE             0  …
692345    THEFT FROM MOTOR VEHICLE - PETTY ($950 & UNDER)            0  …
835683                                 THEFT OF IDENTITY             0  …
825607  THEFT FROM MOTOR VEHICLE - GRAND ($950.01 AND …             0  …
676338                                 THEFT OF IDENTITY             0  …

        premise_code                        premise_description  \
39025          504.0                            OTHER RESIDENCE
137157         402.0                                     MARKET
2389           255.0  AUTO DEALERSHIP (CHEVY, FORD, BMW, MERCEDES, E…
103594         102.0                                   SIDEWALK
25891          108.0                                PARKING LOT
…                …                                           …
770235         717.0                              HEALTH SPA/GYM
692345         101.0                                     STREET
835683         502.0        MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)
825607         101.0                                     STREET
676338         502.0        MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)

        status status_description crime_code_1  \
39025       IC        Invest Cont        310.0
137157      AA        Adult Arrest       343.0
2389        IC        Invest Cont        888.0
103594      IC        Invest Cont        210.0
25891       AO        Adult Other        330.0
…            …                  …           …
770235      IC        Invest Cont        330.0
692345      IC        Invest Cont        420.0
835683      IC        Invest Cont        354.0
825607      IC        Invest Cont        331.0
676338      IC        Invest Cont        354.0

                                    location  latitude longitude  year  \
39025    3800   WINFORD                   DR  34.1386  -118.5525  2020
137157   20000 W  VENTURA                 BL  34.1719  -118.5684  2020
2389     3300 S  FIGUEROA                 ST  34.0225  -118.2796  2020
103594                             CENTURY   33.9456  -118.2652  2020
25891    300 E  2ND                       ST  34.0498  -118.2400  2020
…                                          …        …         …    …
770235   700 W  CESAR E CHAVEZ            AV  34.0606  -118.2439  2023
692345   11400  AMBOY                     AV  34.2767  -118.4477  2023
835683   21400  SATICOY                   ST  34.2119  -118.6017  2023
```

```
825607          N   KENMORE                    AV    34.1018  -118.2973  2023
676338    1900      ALLESANDRO                 ST    34.0900  -118.2587  2023


         month
39025        1
137157       1
2389         1
103594       1
25891        1
...        ...
770235      12
692345      12
835683       3
825607      12
676338      11


[852950 rows x 22 columns]
```

2.2 Feature Understanding

2.2.1 In this section, we explore and visualize various features of the crime data to gain insights into patterns and trends.

2.2.2 Key findings:

- The count of crime incidents varies by month, with higher counts generally observed in the middle months of the year.
- There is an increasing trend in the number of crime incidents from 2020 to 2022.
- The distribution of victim ages shows a peak in the 20-40 age range.
- The top 20 most frequent crimes are identified, with theft-related crimes being the most prevalent.

```python
[18]:  # Plot the count of crime incidents by month
       df['month'].value_counts().sort_index().plot()
```

```
[18]: <AxesSubplot:>
```
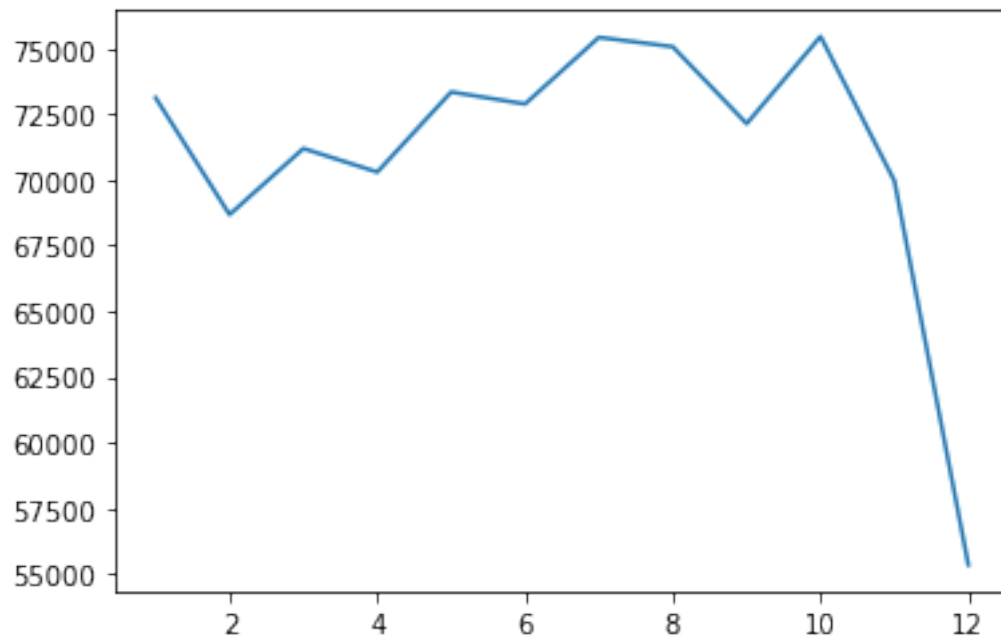
```
[19]:  # Plot the count of crime incidents by year as a bar chart
       df['year'].value_counts().sort_index().plot(kind = 'bar')
```

[19]: <AxesSubplot:>

14
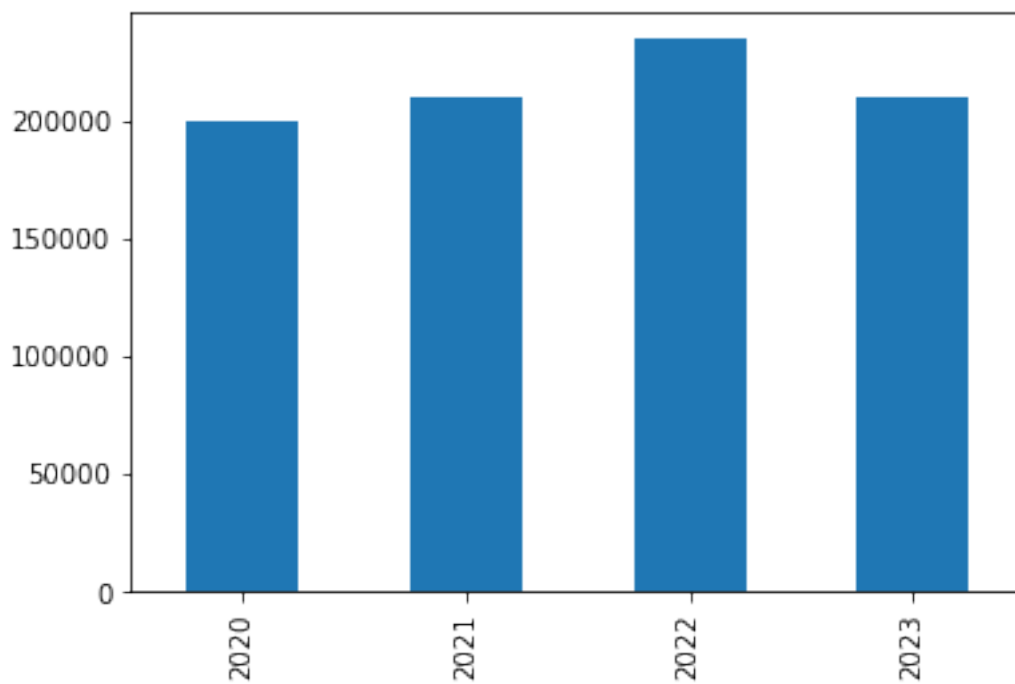
```
[20]:  # Remove rows where month is 12 and year is 2023 (incomplete data)
       df = df[~((df['month'] == 12) & (df['year'] == 2023))]

[21]:  # Sort the DataFrame by the 'date_occurred' column
       df.sort_values(by = 'date_occurred')

[21]:          division_number date_reported       date_occurred  area    area_name  \
       168264         201810982   2020-05-11  2020-01-01 00:01:00    18    Southeast
       196441         220706832   2022-03-16  2020-01-01 00:01:00     7     Wilshire
       1559           210708670   2021-05-11  2020-01-01 00:01:00     7     Wilshire
       199361         221818077   2022-09-28  2020-01-01 00:01:00    18    Southeast
       129124         201704365   2020-01-02  2020-01-01 00:01:00    17   Devonshire
       ...                  ...          ...                  ...   ...          ...
       672120         230717815   2023-12-01  2023-11-30 23:45:00     7     Wilshire
       766816         230917064   2023-12-04  2023-11-30 23:48:00     9     Van Nuys
       770137         230221843   2023-12-01  2023-11-30 23:50:00     2     Rampart
       726894         231116332   2023-12-01  2023-11-30 23:50:00    11    Northeast
       743649         232018618   2023-12-01  2023-11-30 23:59:00    20      Olympic

              reporting_district  part  crime_code  \
       168264                1822     2         810
       196441                 782     2         668
       1559                   775     2         812
       199361                1801     1         820
       129124                1761     2         740
       ...                    ...   ...         ...
       672120                 702     1         310
       766816                 984     1         310
       770137                 216     1         510
       726894                1132     2         888
       743649                2044     2         930

                                          crime_description  victim_age  …  \
       168264  SEX,UNLAWFUL(INC MUTUAL CONSENT, PENETRATION W…          17  …
       196441          EMBEZZLEMENT, GRAND THEFT ($950.01 & OVER)           0  …
       1559     CRM AGNST CHLD (13 OR UNDER) (14-15 & SUSP 10 …          13  …
       199361                              ORAL COPULATION           6  …
       129124  VANDALISM - FELONY ($400 & OVER, ALL CHURCH VA…          23  …
       ...                                               ...         ...  …
       672120                                        BURGLARY          30  …
       766816                                        BURGLARY           0  …
       770137                                VEHICLE - STOLEN           0  …
       726894                                     TRESPASSING           0  …
       743649          CRIMINAL THREATS - NO WEAPON DISPLAYED          55  …
```

```
        premise_code                    premise_description  status  \
168264         501.0                  SINGLE FAMILY DWELLING      AO
196441         203.0                         OTHER BUSINESS      IC
1559           710.0                          OTHER PREMISE      AO
199361         501.0                  SINGLE FAMILY DWELLING      AO
129124         122.0                VEHICLE, PASSENGER/TRUCK      IC
...              ...                                    ...     ...
672120         501.0                  SINGLE FAMILY DWELLING      IC
766816         501.0                  SINGLE FAMILY DWELLING      IC
770137         101.0                                 STREET      IC
726894         203.0                         OTHER BUSINESS      IC
743649         502.0  MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)   IC

       status_description crime_code_1  \
168264        Adult Other        810.0
196441        Invest Cont        668.0
1559          Adult Other        812.0
199361        Adult Other        812.0
129124        Invest Cont        740.0
...                   ...          ...
672120        Invest Cont        310.0
766816        Invest Cont        310.0
770137        Invest Cont        510.0
726894        Invest Cont        888.0
743649        Invest Cont        930.0

                          location  latitude  longitude  year  \
168264    400 W   CENTURY                BL   33.9456  -118.2808  2020
196441   1800 S   FAIRFAX               AV   34.0431  -118.3692  2020
1559     1700 S   LONGWOOD              AV   34.0437  -118.3440  2020
199361    500 W   92ND                  ST   33.9528  -118.2827  2020
129124   9500     OWENSMOUTH            AV   34.2427  -118.6021  2020
...                                      ...      ...       ...   ...
672120    700 N   VISTA                 ST   34.0836  -118.3523  2023
766816  14600     SUTTON                ST   34.1477  -118.4530  2023
770137   1000     CORONADO              TR   34.0777  -118.2688  2023
726894   3200     RIVERSIDE             DR   34.1143  -118.2695  2023
743649    900 S   KINGSLEY              DR   34.0559  -118.3031  2023

        month
168264      1
196441      1
1559        1
199361      1
129124      1
...        ...
672120     11
```

```
766816      11
770137      11
726894      11
743649      11

[851405 rows x 22 columns]
```

[22]:
```python
# Plot the count of crime incidents by year and month
crime_over_time = df.groupby(['year', 'month']).size().plot()
crime_over_time
```

[22]: <AxesSubplot:xlabel='year,month'>
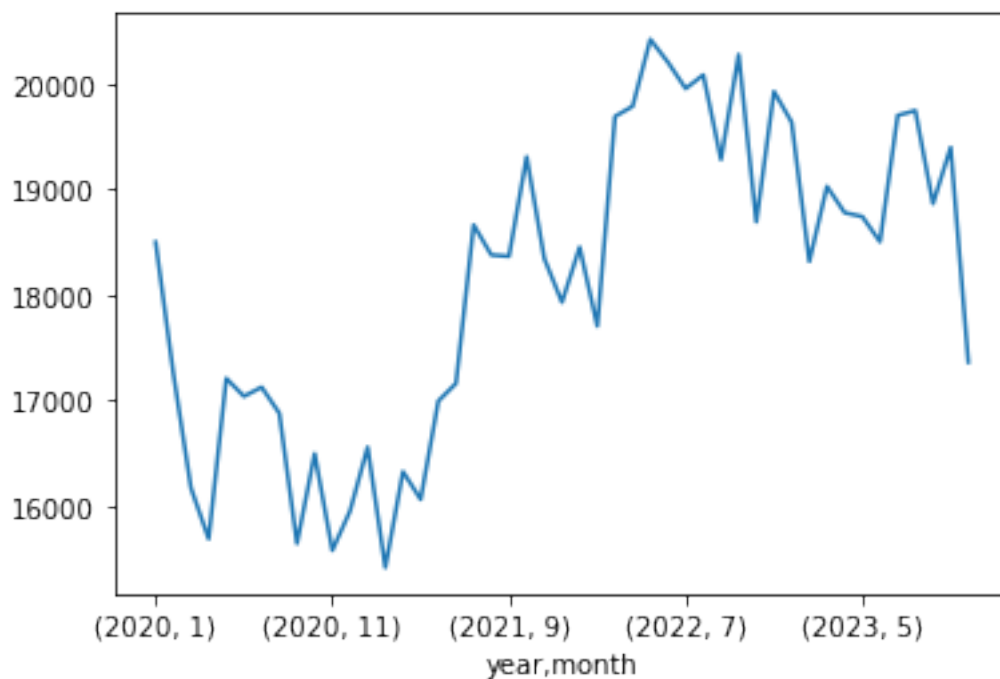
[23]:
```python
# Plot a histogram of victim ages
plt.hist(df_age['victim_age'], edgecolor = 'black', bins = 15)
```

[23]: (array([3.96800e+03, 1.86630e+04, 9.81120e+04, 1.44858e+05, 1.22469e+05,
        8.64850e+04, 7.20760e+04, 4.59690e+04, 2.94150e+04, 1.28000e+04,
        4.50300e+03, 1.27400e+03, 4.41000e+02, 0.00000e+00, 1.00000e+00]),
 array([2.   ,   9.86666667, 17.73333333, 25.6   ,
         33.46666667, 41.33333333, 49.2   , 57.06666667,
         64.93333333, 72.8   , 80.66666667, 88.53333333,
         96.4   , 104.26666667, 112.13333333, 120.   ]),
 <BarContainer object of 15 artists>)
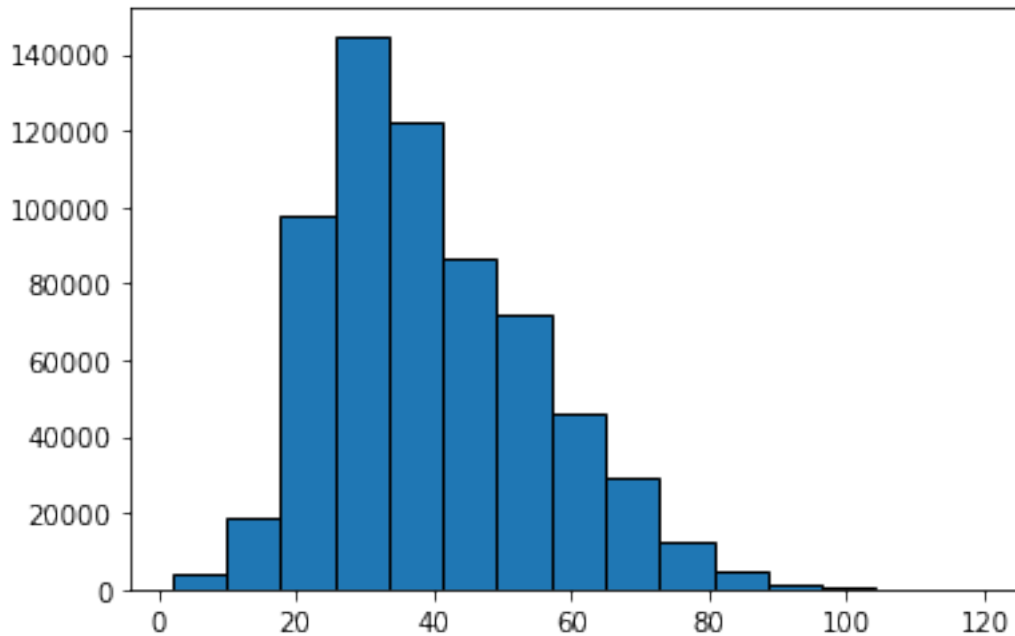
```
[24]: df['date_occurred_no_time'] = df['date_occurred'].dt.date
```

```
[25]: #Calculate average daily count of crimes
      crime_count_per_day = df.groupby(['date_occurred_no_time',␣
       ↪'crime_description']).size()
```

```
[26]: #Calculate top 20 daily crimes
      top_20_crimes = crime_count_per_day.groupby('crime_description').mean()\
      .reset_index(name = 'average_daily_count').\
      sort_values(by = 'average_daily_count', ascending = False).head(20)
```

```
[27]: top_20_sorted = top_20_crimes.sort_values(by = 'average_daily_count', ascending␣
       ↪= True)
```

```
[28]: #Plot the top 20 cri'0omes
      plt.barh(top_20_sorted['crime_description'],␣
       ↪top_20_sorted['average_daily_count'])
```

```
[28]: <BarContainer object of 20 artists>
```
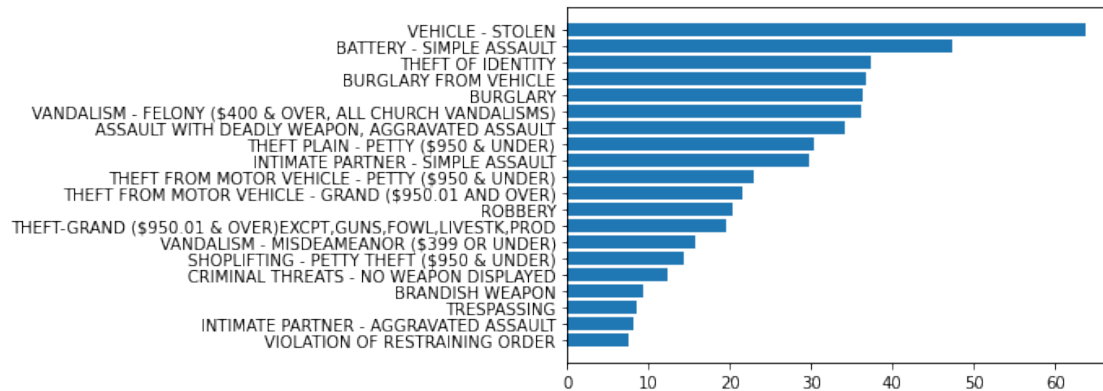
```
[29]: plt.style.use('ggplot')

[30]: plt.figure(figsize = (12, 8))

      # Plot 1, line chart
      plt.subplot(2, 2, 1)
      crime_plot = df.groupby(['year', 'month']).size().plot(color = 'green',␣
        ↪linewidth = 2)
      crime_plot.set_xlabel('Year - Month', fontsize = 10)
      crime_plot.set_title('Crime Over Time')
      # Plot 2, histogram of victim ages
      plt.subplot(2, 2, 2)
      plt.hist(df_age['victim_age'], edgecolor = 'black', bins = 15)
      plt.xlabel('Age', fontsize = 10)
      plt.tight_layout()
      plt.title('Distribution of Victim Age')
      #Plot 3, top 20 crimes
      plt.figure()
      plt.barh(top_20_sorted['crime_description'],␣
        ↪top_20_sorted['average_daily_count'], color = 'orange')
      plt.title('Top 20 Crimes')

[30]: Text(0.5, 1.0, 'Top 20 Crimes')
```

Crime Over Time

Distribution of Victim Age

Top 20 Crimes
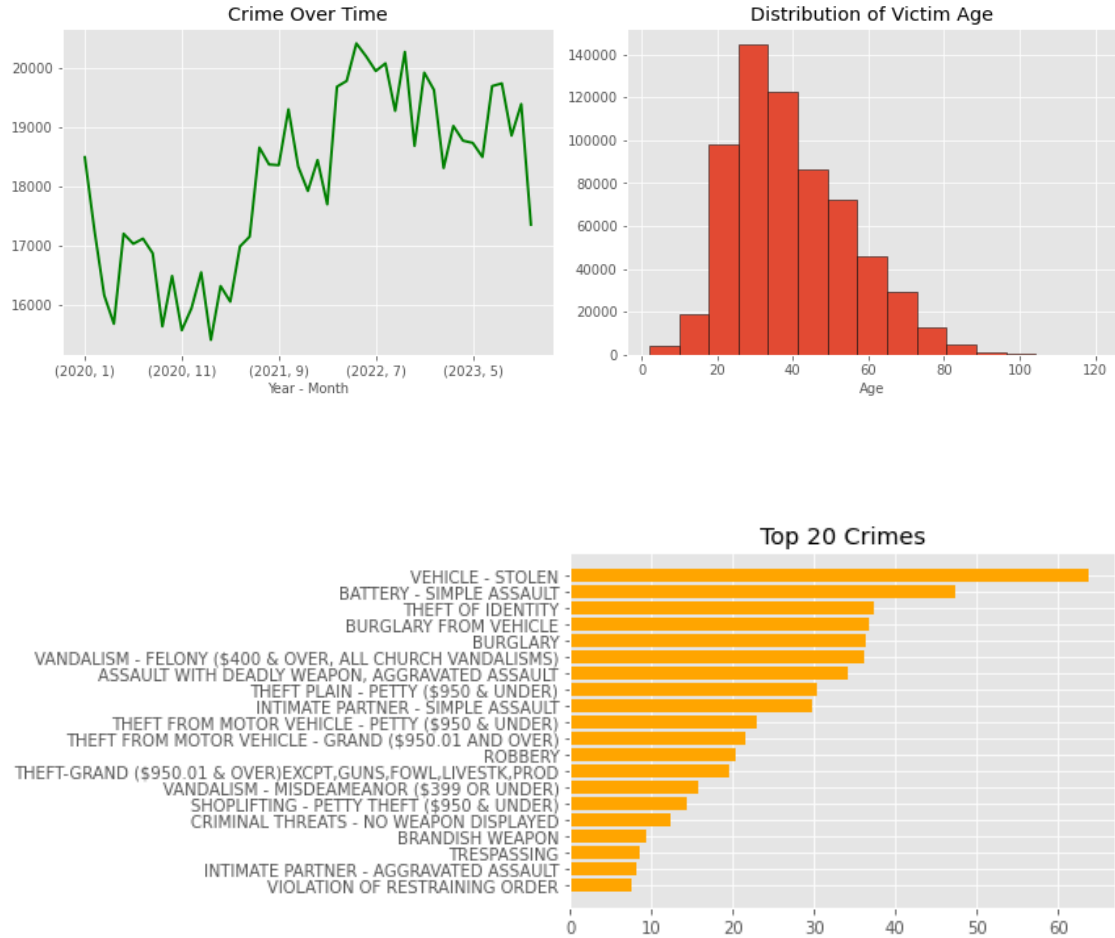
2.3 Time Taken to Report Crime

2.3.1 In this section, we analyze the time difference between when a crime occurred and when it was reported.

2.3.2 Key observations:

- A large number of crimes are reported on the same day they occur (time difference of 0 days).
- Some crime types have longer average reporting times than others.

```
[31]: # Show the difference in time for crime reported vs occurred
      df['Time to report'] = (pd.to_datetime(df['date_reported']) - pd.
       ↪to_datetime(df['date_occurred_no_time'])).dt.days
```

```
[32]: df
```

```
[32]:        division_number date_reported        date_occurred  area   area_name  \
      0             10304468   2020-01-08  2020-01-08 22:30:00     3   Southwest
      1            190101086   2020-01-02  2020-01-01 03:30:00     1     Central
```

20

```
2             200110444    2020-04-14 2020-02-13 12:00:00   1      Central
3             191501505    2020-01-01 2020-01-01 17:30:00  15  N Hollywood
4             191921269    2020-01-01 2020-01-01 04:15:00  19      Mission
…                 …           …              …   …          …
852945        231606525    2023-03-22 2023-03-22 10:00:00  16     Foothill
852946        231210064    2023-04-12 2023-04-12 16:30:00  12  77th Street
852947        230115220    2023-07-02 2023-07-01 00:01:00   1      Central
852948        230906458    2023-03-05 2023-03-05 09:00:00   9     Van Nuys
852949        230319786    2023-11-10 2023-11-09 23:00:00   3    Southwest

        reporting_district  part  crime_code  \
0                      377     2         624
1                      163     2         624
2                      155     2         845
3                     1543     2         745
4                     1998     2         740
…                        …     …           …
852945                1602     1         230
852946                1239     1         230
852947                 154     1         352
852948                 914     2         745
852949                 395     1         331

                                crime_description  victim_age  … \
0                         BATTERY - SIMPLE ASSAULT          36  …
1                         BATTERY - SIMPLE ASSAULT          25  …
2             SEX OFFENDER REGISTRANT OUT OF COMPLIANCE      0  …
3               VANDALISM - MISDEAMEANOR ($399 OR UNDER)    76  …
4       VANDALISM - FELONY ($400 & OVER, ALL CHURCH VA…    31  …
…                                               …          …  …
852945     ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT   25  …
852946     ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT   29  …
852947                                       PICKPOCKET     24  …
852948          VANDALISM - MISDEAMEANOR ($399 OR UNDER)    53  …
852949  THEFT FROM MOTOR VEHICLE - GRAND ($950.01 AND …    38  …

        status status_description  crime_code_1  \
0           AO       Adult Other          624.0
1           IC       Invest Cont          624.0
2           AA      Adult Arrest          845.0
3           IC       Invest Cont          745.0
4           IC       Invest Cont          740.0
…            …                 …             …
852945      IC       Invest Cont          230.0
852946      IC       Invest Cont          230.0
852947      IC       Invest Cont          352.0
852948      IC       Invest Cont          745.0
```

```
852949      IC          Invest Cont         331.0

                                        location latitude longitude  year  \
0        1100 W  39TH                         PL  34.0141 -118.2978  2020
1         700 S  HILL                         ST  34.0459 -118.2545  2020
2         200 E  6TH                          ST  34.0448 -118.2474  2020
3         5400    CORTEEN                      PL  34.1685 -118.4019  2020
4        14400    TITUS                        ST  34.2198 -118.4468  2020
...         ...                               ...      ...       ...   ...
852945   12800    FILMORE                      ST  34.2790 -118.4116  2023
852946    6100 S  VERMONT                      AV  33.9841 -118.2915  2023
852947     500 S  MAIN                         ST  34.0467 -118.2485  2023
852948   14500    HARTLAND                     ST  34.1951 -118.4487  2023
852949    4100 S  HOBART                       BL  34.0091 -118.3078  2023

        month  date_occurred_no_time  Time to report
0           1             2020-01-08               0
1           1             2020-01-01               1
2           2             2020-02-13              61
3           1             2020-01-01               0
4           1             2020-01-01               0
...       ...                    ...             ...
852945      3             2023-03-22               0
852946      4             2023-04-12               0
852947      7             2023-07-01               1
852948      3             2023-03-05               0
852949     11             2023-11-09               1

[851405 rows x 24 columns]
```

[33]: `df['Time to report'].value_counts().head(10)`

```
[33]: 0    419620
      1    185029
      2     52808
      3     30894
      4     19932
      5     14416
      6     10837
      7      9722
      8      7053
      9      5660
      Name: Time to report, dtype: int64
```

[34]:
```python
#Group by crime description and reporting times
df.groupby(['crime_description'])['Time to report'].mean().
 ↪sort_values(ascending = False).head(10)
```

```
[34]: crime_description
      CRM AGNST CHLD (13 OR UNDER) (14-15 & SUSP 10 YRS OLDER)    128.181013
      SEX OFFENDER REGISTRANT OUT OF COMPLIANCE                   123.779056
      SEX,UNLAWFUL(INC MUTUAL CONSENT, PENETRATION W/ FRGN OBJ    110.913921
      LEWD/LASCIVIOUS ACTS WITH CHILD                              97.324675
      DISHONEST EMPLOYEE ATTEMPTED THEFT                           87.000000
      BIGAMY                                                       77.666667
      SEXUAL PENETRATION W/FOREIGN OBJECT                          66.092946
      ORAL COPULATION                                             62.078616
      EMBEZZLEMENT, PETTY THEFT ($950 & UNDER)                     61.322581
      EMBEZZLEMENT, GRAND THEFT ($950.01 & OVER)                  59.657109
      Name: Time to report, dtype: float64
```

2.3.3 Identify the top 3 crimes with the highest average victim age

```python
[35]: df_age.groupby('crime_description')['victim_age'].mean().reset_index().
      ↪sort_values(by = 'victim_age', ascending = False).head(3)
```

```
[35]:                       crime_description  victim_age
      45  DISHONEST EMPLOYEE ATTEMPTED THEFT   60.000000
      12        BLOCKING DOOR INDUCTION CENTER   54.666667
      77                              LYNCHING   53.000000
```

2.4 Areas With the Highest Crime Rates at Night

In this section, we identify the areas with the highest crime rates during nighttime hours (8 PM to 6 AM).

2.4.1 Key finding:

The top five areas with the highest nighttime crime rates are:

- Central (55,315 crimes)
- 77th Street (51,137 crimes)
- Pacific (47,889 crimes)
- Southwest (45,718 crimes)
- Hollywood (42,763 crimes)

The areas with the lowest nighttime crime rates are Foothill (27,267 crimes) and Hollenbeck (30,549 crimes). There is a significant variation in nighttime crime rates across different areas of Los Angeles.

```python
[36]: crimes_at_night = df[(df['date_occurred'].dt.hour) >= 20 | (df['date_occurred'].
      ↪dt.hour <= 6)]
```

```python
[37]: crimes_at_night['area_name'].value_counts().reset_index()
```

```
[37]:        index  area_name
      0    Central      55315
```

```
1    77th Street      51137
2        Pacific      47889
3      Southwest      45718
4      Hollywood      42763
5      Southeast      41080
6        Olympic      41050
7         Newton      40779
8    N Hollywood      40628
9       Wilshire      39196
10       Rampart      38310
11       West LA      37575
12      Northeast      35174
13      Van Nuys      34651
14   West Valley      34383
15       Topanga      33475
16        Harbor      33435
17     Devonshire      33340
18        Mission      32630
19     Hollenbeck      30549
20       Foothill      27267
```

[38]:
```python
crimes_at_night_counts = crimes_at_night['area_name'].value_counts().
 ↪reset_index()
crimes_at_night_counts.columns = ['area_name', 'count']

plt.figure(figsize=(10, 8))
plt.barh(crimes_at_night_counts['area_name'], crimes_at_night_counts['count'])
plt.xlabel('Number of Crimes')
plt.ylabel('Area Name')
plt.title('Number of Crimes at Night by Area')
plt.tight_layout()
plt.show()
```

Number of Crimes at Night by Area
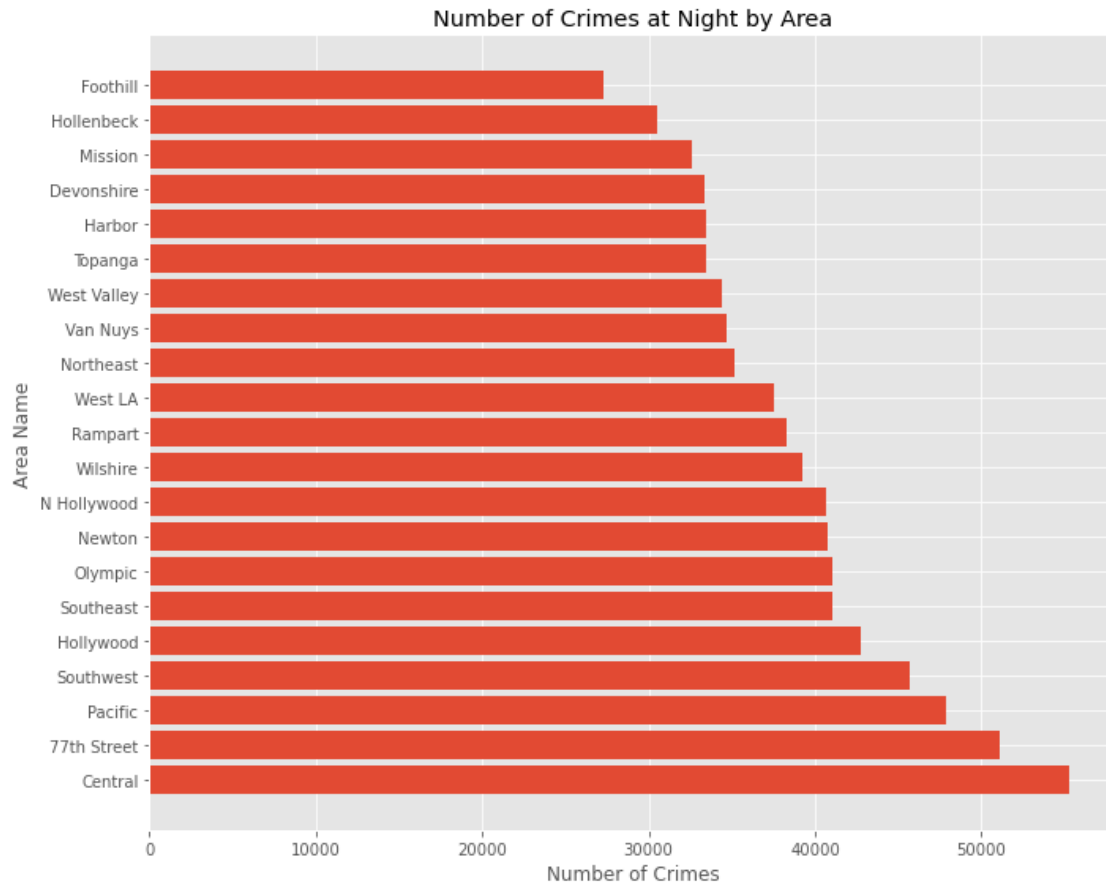
2.5 Crime by the Hour

In this section, we analyze the distribution of crimes across different hours of the day.

2.5.1 Key observations:

- There is a notable peak in crime incidents around noon (12 PM).
- Identity theft is a significant contributor to the crime count at this hour.

```
[39]: hourly_crime = df['date_occurred'].dt.hour.value_counts().sort_index().
      ↪reset_index()
      hourly_crime.columns = ['hour', 'crime_count']
```

```
[40]: hourly_crime
```

```
[40]:    hour   crime_count
      0     0        35061
      1     1        25740
      2     2        21801
      3     3        19040
```

```
4      4      16099
5      5      14829
6      6      19484
7      7      21985
8      8      31428
9      9      30812
10     10     36461
11     11     36967
12     12     57683
13     13     38422
14     14     41494
15     15     44394
16     16     44754
17     17     49492
18     18     50606
19     19     46701
20     20     47399
21     21     43270
22     22     41704
23     23     35779
```

[41]: ```python
hourly_crime.columns
```

[41]: ```
Index(['hour', 'crime_count'], dtype='object')
```

[42]: ```python
#Crimes including identity theft
plt.bar(hourly_crime['hour'], hourly_crime['crime_count'])
```

[42]: ```
<BarContainer object of 24 artists>
```
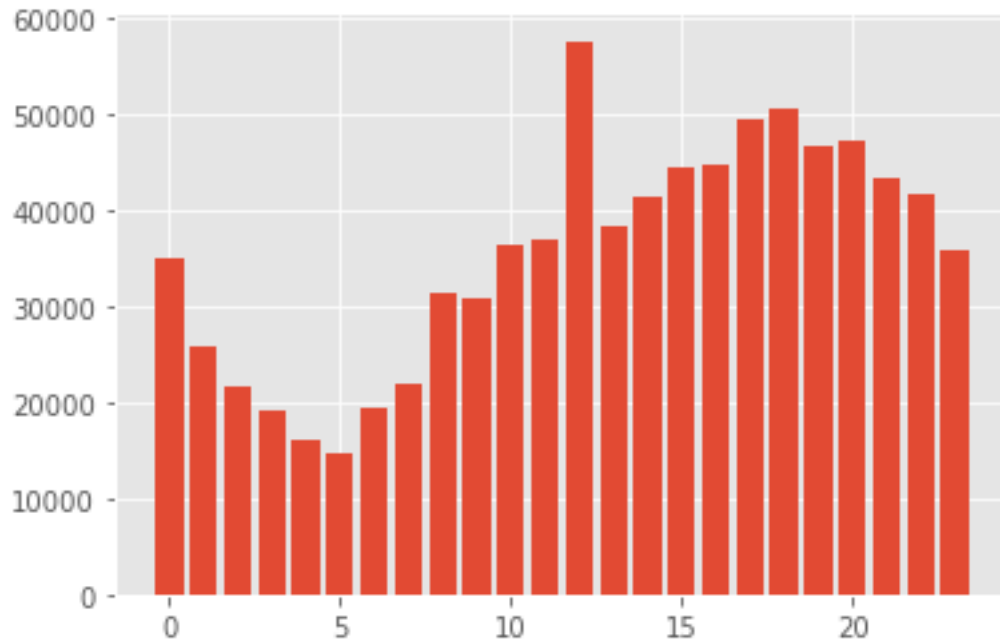
```
[43]: df[df['date_occurred'].dt.hour == 12]['crime_description'].value_counts()
```

```
[43]: THEFT OF IDENTITY                            9098
      VEHICLE - STOLEN                             4048
      BATTERY - SIMPLE ASSAULT                     3958
      THEFT PLAIN - PETTY ($950 & UNDER)           3785
      BURGLARY                                     2945
                                                    ...
      DRUGS, TO A MINOR                               1
      THEFT, COIN MACHINE - GRAND ($950.01 & OVER)    1
      BIGAMY                                          1
      GRAND THEFT / INSURANCE FRAUD                   1
      INCITING A RIOT                                 1
      Name: crime_description, Length: 124, dtype: int64
```

```
[44]: #Filter out identity theft to focus on active crime reporting
      no_id_theft = df[~(df['crime_description'] == 'THEFT OF IDENTITY')]
```

```
[45]: new_hourly_count = no_id_theft['date_occurred'].dt.hour.value_counts().
       ↪sort_index().reset_index()
```

```
[46]: new_hourly_count.columns = ['hour', 'crime_count']
```

```
[47]: plt.figure(figsize=(10, 6))
      plt.bar(new_hourly_count['hour'], new_hourly_count['crime_count'])
      plt.xlabel('Hour')
```

```
plt.ylabel('Crime Count (Excluding Identity Theft)')
plt.title('Crime by the Hour (Excluding Identity Theft)')
plt.xticks(new_hourly_count['hour'])
plt.show()
```


Crime by the Hour (Excluding Identity Theft)
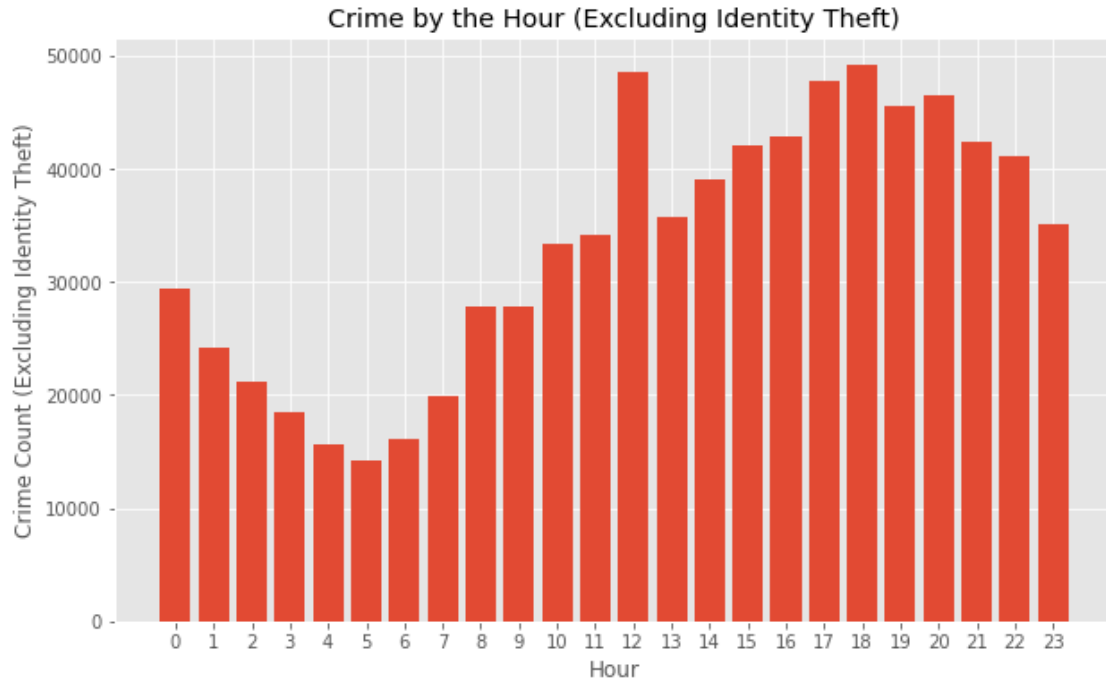
3 Conclusion

3.0.1 This analysis provides a comprehensive overview of crime patterns in Los Angeles from 2020 to 2023. Key insights include:

- An increasing trend in crime incidents over the years.
- Seasonal variations in crime counts, with higher incidents in the middle months.
- Theft-related crimes being the most prevalent.
- Specific areas experiencing higher crime rates during nighttime hours.
- A significant peak in crime incidents around noon, largely driven by identity theft.

These findings can inform law enforcement strategies, resource allocation, and crime prevention initiatives in the City of Los Angeles. Further analysis could explore additional factors such as location-specific trends, demographic correlations, and the impact of socioeconomic conditions on crime rates.

3.1 Further Insights and Recommendations

The analysis of crime incidents in Los Angeles from 2020 to 2023 reveals several noteworthy patterns and trends that warrant further attention. While the overall increasing trend in crime incidents is

concerning, it is crucial to delve deeper into the underlying factors contributing to this rise. Socioeconomic conditions, such as unemployment rates, income inequality, and housing affordability, may play a significant role in shaping crime patterns across different areas of the city. Investigating these factors and their correlation with crime rates could provide valuable insights for targeted interventions and support programs.

Moreover, the prominence of theft-related crimes highlights the need for enhanced crime prevention strategies. Strengthening community policing efforts, increasing surveillance in high-risk areas, and promoting public awareness campaigns could help deter potential offenders and reduce the incidence of theft. Additionally, the peak in crime incidents around noon, largely attributed to identity theft, underscores the importance of digital security and fraud prevention measures. Collaborating with financial institutions, educating the public about online safety, and enforcing stricter penalties for identity theft could help mitigate this specific crime category.

To further enhance the effectiveness of law enforcement efforts, it is recommended to conduct a more granular analysis of crime patterns at the neighborhood level. Identifying specific hotspots within each area and understanding the unique characteristics and challenges of these locations can enable more targeted and efficient deployment of resources. Furthermore, analyzing the temporal patterns of crime, such as seasonal variations and peak hours, can help optimize patrol schedules and ensure a more proactive approach to crime prevention.

Lastly, it is crucial to recognize the importance of community engagement and collaboration in addressing crime. Strengthening partnerships between law enforcement agencies, community organizations, and local residents can foster trust, encourage information sharing, and develop tailored solutions to address the specific needs and concerns of each neighborhood. By adopting a holistic and data-driven approach, the City of Los Angeles can work towards creating safer communities and reducing the overall impact of crime on its residents.

[ ]: