Analyzing Crime in Los Angeles

August 25, 2024

Los Angeles, California . The City of Angels. Tinseltown. The Entertainment Capital of the World!

Known for its warm weather, palm trees, sprawling coastline, and Hollywood, along with producing some of the most iconic films and songs. However, as with any highly populated city, it isn't always glamorous and there can be a large volume of crime. That's where you can help!

You have been asked to support the Los Angeles Police Department (LAPD) by analyzing crime data to identify patterns in criminal behavior. They plan to use your insights to allocate resources effectively to tackle various crimes in different areas.

0.1 The Data

They have provided you with a single dataset to use. A summary and preview are provided below.

It is a modified version of the original data, which is publicly available from Los Angeles Open Data.

1 crimes.csv

Column	Description
`'DR_NO'`	Division of Records Number: Official file number made up of a 2-digit year, area ID, and 5 digits.
`'Date Rptd'`	Date reported - MM/DD/YYYY.
`'DATE OCC'`	Date of occurrence - MM/DD/YYYY.
`'TIME OCC'`	In 24-hour military time.
`'AREA NAME'`	The 21 Geographic Areas or Patrol Divisions are also given a name designation that references a landmark or the surrounding community that it is responsible for. For example, the 77th Street Division is located at the intersection of South Broadway and 77th Street, serving neighborhoods in South Los Angeles.
`'Crm Cd Desc'`	Indicates the crime committed.
`'Vict Age'`	Victim's age in years.
`'Vict Sex'`	Victim's sex: `F`: Female, `M`: Male, `X`: Unknown.

| `'Vict Descent'` | Victim's descent:

`A` - Other Asian

`B` - Black

`C` - Chinese

`D` - Cambodian

`F` - Filipino

`G` - Guamanian

`H` - Hispanic/Latin/Mexican

`I` - American Indian/Alaskan Native

`J` - Japanese

`K` - Korean

`L` - Laotian

`O` - Other

`P` - Pacific Islander

`S` - Samoan

`U` - Hawaiian

`V` - Vietnamese

`W` - White

`X` - Unknown

`Z` - Asian Indian

| | 'Weapon Desc' | Description of the weapon used (if applicable). | | 'Status Desc' | Crime status. | | 'LOCATION' | Street address of the crime. |

[39]:
```
# Re-run this cell
# Import required libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
crimes = pd.read_csv("crimes.csv", parse_dates=["Date Rptd", "DATE OCC"],␣
 ↪dtype={"TIME OCC": str})
crimes.head()
```

[39]:

	DR_NO	Date Rptd	DATE OCC	TIME OCC	AREA NAME	Crm Cd Desc	Vict Age	Vict Sex	Vict Descent	Weapon Desc	Status Desc	LOCATION
0	220314085	2022-07-22	2020-05-12	1110	Southwest	THEFT OF IDENTITY	27	F	B	NaN	Invest Cont	2500 S SYCAMORE AV
1	222013040	2022-08-06	2020-06-04	1620	Olympic	THEFT OF IDENTITY	60	M	H	NaN	Invest Cont	3300 SAN MARINO ST
2	220614831	2022-08-18	2020-08-17	1200	Hollywood	THEFT OF IDENTITY	28	M	H	NaN	Invest Cont	1900 TRANSIENT
3	231207725	2023-02-27	2020-01-27	0635	77th Street	THEFT OF IDENTITY	37	M	H	NaN	Invest Cont	6200 4TH AV
4	220213256	2022-07-14	2020-07-14	0900	Rampart	THEFT OF IDENTITY	79	M	B	NaN	Invest Cont	1200 W 7TH ST

[45]:
```
crimes.info()
```

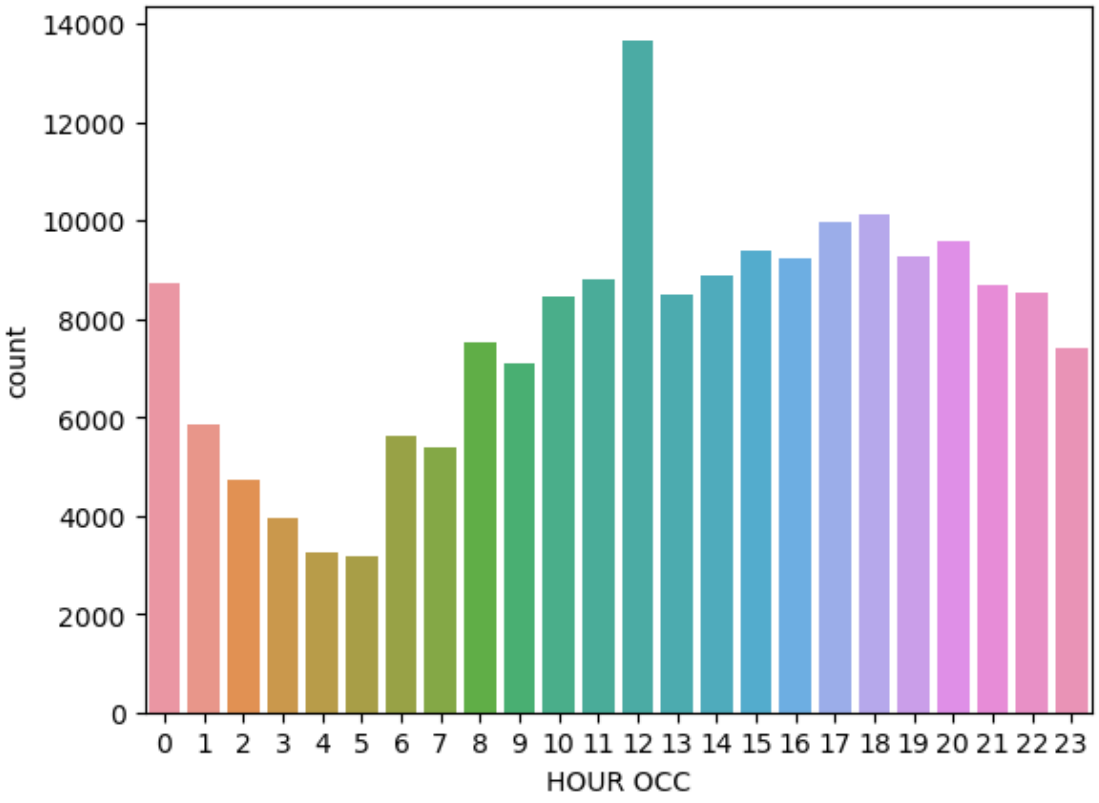
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 185715 entries, 0 to 185714
Data columns (total 12 columns):
 #   Column        Non-Null Count   Dtype
---  ------        --------------   -----
 0   DR_NO         185715 non-null  int64
 1   Date Rptd     185715 non-null  datetime64[ns]
 2   DATE OCC      185715 non-null  datetime64[ns]
 3   TIME OCC      185715 non-null  object
 4   AREA NAME     185715 non-null  object
 5   Crm Cd Desc   185715 non-null  object
 6   Vict Age      185715 non-null  int64
```

```
 7   Vict Sex       185715 non-null  object
 8   Vict Descent   185715 non-null  object
 9   Weapon Desc    185715 non-null  object
 10  Status Desc    185715 non-null  object
 11  LOCATION       185715 non-null  object
dtypes: datetime64[ns](2), int64(2), object(8)
memory usage: 17.0+ MB
```

[43]:
```
crimes.fillna(0,inplace=True)
```

[47]:
```
# Extract the first two digits from "TIME OCC", representing the hour,and␣
 ↪convert to integer data type
crimes["HOUR OCC"] = crimes["TIME OCC"].str[:2].astype(int)
```

[51]:
```
# Produce a countplot to find the largest frequency of crimes by hour
sns.countplot(data=crimes, x="HOUR OCC")
plt.show()
```



[53]:
```
# Midday has the largest volume of crime
peak_crime_hour = 12
```

[55]:
```
crimes.columns
```

[55]:
```
Index(['DR_NO', 'Date Rptd', 'DATE OCC', 'TIME OCC', 'AREA NAME',
       'Crm Cd Desc', 'Vict Age', 'Vict Sex', 'Vict Descent', 'Weapon Desc',
       'Status Desc', 'LOCATION', 'HOUR OCC'],
      dtype='object')
```

[91]:
```
crimes['HOUR OCC'].value_counts()
```

[91]:
```
HOUR OCC
12    13663
18    10125
17     9964
20     9579
15     9393
19     9262
16     9224
14     8872
11     8787
0      8728
21     8701
22     8531
13     8474
10     8440
8      7523
23     7419
9      7092
1      5836
6      5621
7      5403
2      4726
3      3943
4      3238
5      3171
Name: count, dtype: int64
```

[97]:
```
night_crime = crimes[crimes['HOUR OCC'].isin([22,23,0,1,2,3])]
```

[99]:
```
night_crime
```

[99]:
```
            DR_NO  Date Rptd   DATE OCC TIME OCC    AREA NAME  \
8       231207476 2023-02-27 2020-08-15     0001  77th Street
10      221711184 2022-06-15 2020-05-15     0155   Devonshire
30      221314362 2022-07-11 2020-04-07     0001       Newton
33      231307252 2023-03-03 2020-07-05     2305       Newton
36      221614254 2022-11-13 2020-01-01     0001     Foothill
…             …          …          …        …           …
```

```
185687  231306158 2023-02-11 2023-02-08     2200       Newton
185695  231212224 2023-05-17 2023-05-17     2300  77th Street
185700  231300825 2023-06-07 2023-06-07     2203       Newton
185701  230908346 2023-04-19 2023-04-18     2243     Van Nuys
185704  231710498 2023-05-29 2023-05-29     0200   Devonshire

                              Crm Cd Desc  Vict Age Vict Sex Vict Descent  \
8                                BURGLARY        72        M            B
10                     THEFT OF IDENTITY        27        M            B
30                     THEFT OF IDENTITY        53        F            H
33                     THEFT OF IDENTITY        22        F            B
36                     THEFT OF IDENTITY        22        F            H
…                                    …       …      …          …
185687              BURGLARY FROM VEHICLE        38        M            H
185695    VIOLATION OF RESTRAINING ORDER        35        F            H
185700           BATTERY - SIMPLE ASSAULT        45        F            B
185701                 THEFT OF IDENTITY        34        F            B
185704  INTIMATE PARTNER - SIMPLE ASSAULT        27        F            H

                                          Weapon Desc  Status Desc  \
8                                                    0  Invest Cont
10                                                   0  Invest Cont
30                                                   0  Invest Cont
33                                                   0  Invest Cont
36                                                   0  Invest Cont
…                                                  …          …
185687                                               0  Invest Cont
185695                                               0  Invest Cont
185700                     UNKNOWN WEAPON/OTHER WEAPON  Invest Cont
185701                                               0  Invest Cont
185704  STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)  Invest Cont

                                    LOCATION  HOUR OCC
8        8800    HAAS                   AV         0
10       8300    WHITE OAK              AV         1
30       1600 E  OLYMPIC                BL         0
33                         6600 S  BROADWAY        23
36      10200    TELFAIR                AV         0
…                                     …       …
185687   1700 E  16TH                   ST        22
185695   8200 S  MAIN                   ST        23
185700   2300    WALL                   ST        22
185701   4800    COLDWATER CANYON       AV        22
185704  11400    PORTER RANCH           DR         2

[39183 rows x 13 columns]
```

[151]:
```
peak_night_crime_location = (night_crime.groupby("AREA␣
↪NAME",as_index=False)["HOUR OCC"].count()
                             .sort_values(by='HOUR OCC',ascending=False)).
↪iloc[0]["AREA NAME"]

# Print the peak night crime location
print(f"The area with the largest volume of night crime is␣
↪{peak_night_crime_location}")
```

```
The area with the largest volume of night crime is Central
```

[153]:
```
## Identify the number of crimes committed against victims by age group (0-17,␣
↪18-25, 26-34, 35-44, 45-54, 55-64, 65+)
## Save as a pandas Series called victim_ages
# Create bins and labels for victim age ranges
age_bins = [0, 17, 25, 34, 44, 54, 64, np.inf]
age_labels = ["0-17", "18-25", "26-34", "35-44", "45-54", "55-64", "65+"]

# Add a new column using pd.cut() to bin values into discrete intervals
crimes["Age Bracket"] = pd.cut(crimes["Vict Age"],
                               bins=age_bins,
                               labels=age_labels)

# Find the category with the largest frequency
victim_ages = crimes["Age Bracket"].value_counts()
print(victim_ages)
```

```
Age Bracket
26-34    47470
35-44    42157
45-54    28353
18-25    28291
55-64    20169
65+      14747
0-17      4528
Name: count, dtype: int64
```

[ ]:
```

```