datala

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.

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.

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

Column	Description
	• 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.

```python
# 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()
```

|	DR_NO	Date Rptd	DATE OCC	TIME OCC	AREA NAME
0	220314085	2022-07-22T00:00:00.000	2020-05-12T00:00:00.000	1110	Southwest
1	222013040	2022-08-06T00:00:00.000	2020-06-04T00:00:00.000	1620	Olympic
2	220614831	2022-08-18T00:00:00.000	2020-08-17T00:00:00.000	1200	Hollywood
3	231207725	2023-02-27T00:00:00.000	2020-01-27T00:00:00.000	0635	77th Street
4	220213256	2022-07-14T00:00:00.000	2020-07-14T00:00:00.000	0900	Rampart

5 rows

```python
# 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)
crimes.head()
```

	DR_NO	Date Rptd	DATE OCC	TIME OCC	AREA NAME
0	220314085	2022-07-22T00:00:00.000	2020-05-12T00:00:00.000	1110	Southwest
1	222013040	2022-08-06T00:00:00.000	2020-06-04T00:00:00.000	1620	Olympic
2	220614831	2022-08-18T00:00:00.000	2020-08-17T00:00:00.000	1200	Hollywood
3	231207725	2023-02-27T00:00:00.000	2020-01-27T00:00:00.000	0635	77th Street
4	220213256	2022-07-14T00:00:00.000	2020-07-14T00:00:00.000	0900	Rampart

5 rows

```python
#countplot to find the largest frequency of crimes by the hour
sns.countplot(data = crimes, x= "HOUR OCC")
plt.show()
```
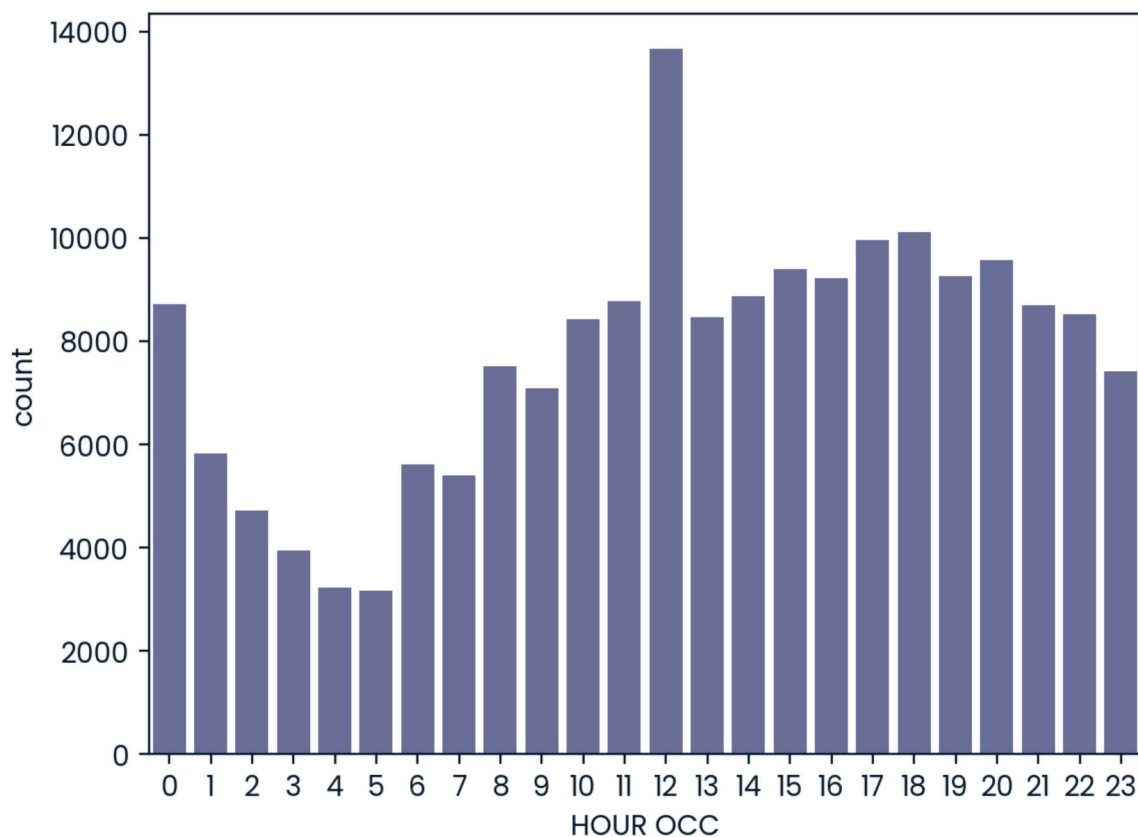


```python
#Midday has the highest volume of crime
peak_crime_hour = 12
```

```python
# Area with the largest frequency of night crimes (crimes between 10pm and 3:59 am)
#filtering for the night time hours
# 0 = midnight , 3 = crimes between 3:00 am - 3:59 am (doesn't include 4am)
night_time = crimes[crimes["HOUR OCC"].isin([22,23,0,1,2,3]) ]
```

```python
# Group by "Area Name" and count occurences, filtering for the largest value and saving the
area name
```

```
peak_night_crime_location = night_time.groupby("AREA NAME", as_index= False)["HOUR
OCC"].count().sort_values("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

```
## 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)
```

```
26-34    47470
35-44    42157
45-54    28353
18-25    28291
55-64    20169
65+      14747
0-17      4528
```