PROJECT: ANALYZING CRIME IN LOS ANGELES

The Data

crimes.csv

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 • `K` - Korean • `L` - Laotian • `O` - Other • `P` - Pacific Islander • `S` - Samoan • `U` - Hawaiian • `V` - Vietnamese • `W` - White • `X` - Unknown

Column	Description
	• `Z` - Asian Indian
`'Weapon Desc'`	Description of the weapon used (if applicable).
`'Status Desc'`	Crime status.
`'LOCATION'`	Street address of the crime.

```
# 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", dtype={"TIME OCC": str})
crimes.head()
```

	D	D…	D…		AR…	Crm Cd Desc
0	220314085	2022-07-22	2020-05-12	1110	Southwest	THEFT OF IDENTITY
1	222013040	2022-08-06	2020-06-04	1620	Olympic	THEFT OF IDENTITY
2	220614831	2022-08-18	2020-08-17	1200	Hollywood	THEFT OF IDENTITY
3	231207725	2023-02-27	2020-01-27	0635	77th Street	THEFT OF IDENTITY
4	220213256	2022-07-14	2020-07-14	0900	Rampart	THEFT OF IDENTITY

Rows: 5 Expand

```
# Validating and cleaning
crimes.info()
print(crimes.isna().sum())
crimes['Weapon Desc'] = crimes['Weapon Desc'].fillna('Unknown')
crimes['Vict Sex'] = crimes['Vict Sex'].str.replace('H', 'X')
crimes.head()
```

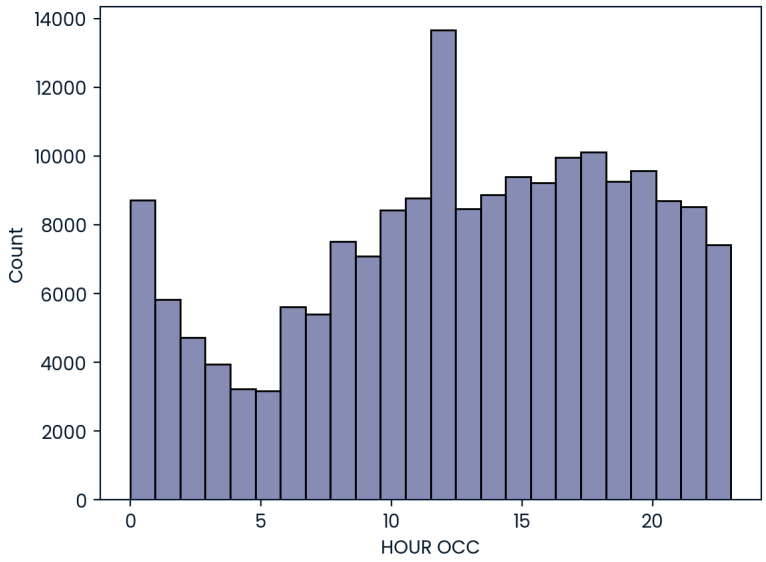
```
 #   Column        Non-Null Count   Dtype
---  ------        --------------   -----
 0   DR_NO         185715 non-null  int64
 1   Date Rptd     185715 non-null  object
 2   DATE OCC      185715 non-null  object
 3   TIME OCC      185715 non-null  object
 4   AREA NAME     185715 non-null  object
 5   Crm Cd Desc   185715 non-null  object
 6   Vict Age      185715 non-null  int64
 7   Vict Sex      185704 non-null  object
 8   Vict Descent  185705 non-null  object
 9   Weapon Desc   73502 non-null   object
 10  Status Desc   185715 non-null  object
 11  LOCATION      185715 non-null  object
dtypes: int64(2), object(10)
memory usage: 17.0+ MB
DR_NO               0
Date Rptd           0
DATE OCC            0
TIME OCC            0
AREA NAME           0
Crm Cd Desc         0
Vict Age            0
Vict Sex           11
Vict Descent       10
Weapon Desc     112213
Status Desc         0
LOCATION            0
dtype: int64
```

	D	D…	D…		AR…	Crm Cd Desc
0	220314085	2022-07-22	2020-05-12	1110	Southwest	THEFT OF IDENTITY
1	222013040	2022-08-06	2020-06-04	1620	Olympic	THEFT OF IDENTITY
2	220614831	2022-08-18	2020-08-17	1200	Hollywood	THEFT OF IDENTITY
3	231207725	2023-02-27	2020-01-27	0635	77th Street	THEFT OF IDENTITY
4	220213256	2022-07-14	2020-07-14	0900	Rampart	THEFT OF IDENTITY

Rows: 5 Expand

```
#finding which hour is peak crime hour (highest freq of crimes)
crimes['HOUR OCC'] = crimes['TIME OCC'].str[:2]
crimes['HOUR OCC'] = crimes['HOUR OCC'].astype(int)
crimes.info()
sns.histplot(data=crimes, x='HOUR OCC',bins=24) #bins24 for 24 hour mark
peak_crime_hour = 12
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 185715 entries, 0 to 185714
Data columns (total 13 columns):
 #   Column        Non-Null Count   Dtype
---  ------        --------------   -----
 0   DR_NO         185715 non-null  int64
 1   Date Rptd     185715 non-null  object
 2   DATE OCC      185715 non-null  object
 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      185704 non-null  object
 8   Vict Descent  185705 non-null  object
 9   Weapon Desc   185715 non-null  object
 10  Status Desc   185715 non-null  object
 11  LOCATION      185715 non-null  object
 12  HOUR OCC      185715 non-null  int64
dtypes: int64(3), object(10)
memory usage: 18.4+ MB
```




```
#finidng which area has th largest freq of night crimes between 10pm and 3:59am (3:59 - 22:00)
between = ['3','22']
night_hours = crimes[['AREA NAME','LOCATION']].groupby('AREA NAME').value_counts().sort_values(ascending=False).reset_index(name='Total')
peak_night_crime_location = night_hours.loc[0,'AREA NAME']
print('Area that has largest frequency of night crimes is:',peak_night_crime_location)
sns.barplot(data=night_hours, x="AREA NAME", y="Total", ci=None)
plt.xticks(rotation=45)
plt.show()
```

```
Area that has largest frequency of night crimes is: Central
```