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 • `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	Crm Cd Desc	Vict Age	Vict Sex
0	220314085	2022-07-22T00:00:00.000	2020-05-12T00:00:00.000	1110	Southwest	THEFT OF IDENTITY	27	F
1	222013040	2022-08-06T00:00:00.000	2020-06-04T00:00:00.000	1620	Olympic	THEFT OF IDENTITY	60	M
2	220614831	2022-08-18T00:00:00.000	2020-08-17T00:00:00.000	1200	Hollywood	THEFT OF IDENTITY	28	M
3	231207725	2023-02-27T00:00:00.000	2020-01-27T00:00:00.000	0635	77th Street	THEFT OF IDENTITY	37	M
4	220213256	2022-07-14T00:00:00.000	2020-07-14T00:00:00.000	0900	Rampart	THEFT OF IDENTITY	79	M

5 rows ⬇

```python
# Which hour has the highest frequency of crimes? Store as an integer variable called peak_crime_hour.
crim_temp = crimes.copy()

crim_temp['hour'] = crim_temp['TIME OCC'].str[:2]
crim_temp['hour'].fillna('00', inplace=True)
crim_temp['hour'] = crim_temp['hour'].astype(int)

max_hour = crim_temp['hour'].max()

# mh_rs = crim_temp[crim_temp['hour'] == max_hour].iloc[0]['hour']

# peak_crime_hour = int(mh_rs)

crim_by_freq = crim_temp.groupby('hour')['DR_NO'].count()

peak_crime_hour = crim_by_freq.sort_values(ascending=False).reset_index()['hour'][0]

peak_crime_hour
```

12

```python
# Which area has the largest frequency of night crimes (crimes committed between 10pm and 3:59am)? Save as a string variable called peak_night_crime_location.

crim_temp_night = crim_temp[(crim_temp['hour'] >= 10) | (crim_temp['hour'] <= 3)]

crim_temp_night = crim_temp_night.groupby('AREA NAME')['DR_NO'].count()

peak_night_crime_location = crim_temp_night.sort_values(ascending=False).reset_index()['AREA NAME'][0]

peak_night_crime_location
```

'Central'

```python
# Identify the number of crimes committed against victims of different age groups. Save as a pandas Series called victim_ages, with age group labels "0-17", "18-25", "26-
34", "35-44", "45-54", "55-64", and "65+" as the index and the frequency of crimes as the values.
bins = [0, 17, 25, 34, 44, 54, 64, 99]
age_groups = ["0-17", "18-25", "26-34", "35-44", "45-54", "55-64", "65+"]
```

```
crim_temp['age group'] = pd.cut(crim_temp['Vict Age'], bins=bins, labels=age_groups)

victim_ages = crim_temp.groupby('age group')['DR_NO'].count().sort_values(ascending=False)
victim_ages
```

age group	DR_NO
26-34	
35-44	
45-54	
18-25	
55-64	
65+	
0-17	

7 rows

```
crim_temp['age group'] = pd.cut(crim_temp['Vict Age'], bins=bins, labels=age_groups)

victim_ages = crim_temp.groupby('age group')['DR_NO'].count().sort_values(ascending=False)
victim_ages
```

age group	DR_NO
26-34	
35-44	
45-54	
18-25	
55-64	
65+	
0-17	

7 rows