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.

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

...	D ...	D... ...	D... ...	...	AR... ...	Crm Cd Desc ...	...	...	Vict	We...
0	220314085	2022-07-22	2020-05-12	1110	Southwest	THEFT OF IDENTITY	27	F	B	null
1	222013040	2022-08-06	2020-06-04	1620	Olympic	THEFT OF IDENTITY	60	M	H	null
2	220614831	2022-08-18	2020-08-17	1200	Hollywood	THEFT OF IDENTITY	28	M	H	null
3	231207725	2023-02-27	2020-01-27	0635	77th Street	THEFT OF IDENTITY	37	M	H	null
4	220213256	2022-07-14	2020-07-14	0900	Rampart	THEFT OF IDENTITY	79	M	B	null

Rows: 5   ↗ Expand

```python
# Start coding here
# Use as many cells as you need

# 1. Finding the frequencies of crimes by the hour of occurrence

# To calculate the frequency by hour, I plan to create a column named hour.
# Then, I will group by the year column and calculate the frequency for each hour.
# After that, I will search for the maximum frequency and the corresponding year.

# I need to convert the military time to an integer, then extract the hour.

crimes["Hour"] = crimes["TIME OCC"].astype(int) // 100

max_data = crimes["Hour"].value_counts().sort_values(ascending = False)

print("Hour: ",max_data.index[0]," with a value of ",max_data.iloc[0])

peak_crime_hour = max_data.index[0]
```

```
Hour:  12  with a value of  13663
```

```python
# 2. Identifying the area with the most night crime


# I need to create a categorical variable that maps each hour to its corresponding period (e.g., "night", "morning", etc.).

# I am considering using np.select because it maps values to their corresponding labels based on conditions.

#Periods: "night"(22,03:59), "morning"(4,11:59), "afternoon"(12,17:59), "evening"(18,21:59)

categories = [
    "night",
    "morning",
    "afternoon",
    "evening"
]

conditions = [
    ((crimes["Hour"] > 22) | (crimes["Hour"] < 4)),   # night
    ((crimes["Hour"] >= 4) & (crimes["Hour"] < 12)),  # morning
    ((crimes["Hour"] >= 12) & (crimes["Hour"] < 18)), # afternoon
    ((crimes["Hour"] >= 18) & (crimes["Hour"] < 22))  # evening
]

crimes["Period"] = np.select(conditions,categories,default = "Unknown")

night_crimes = crimes[ crimes["Period"] == "night" ]
max_night_data = night_crimes["AREA NAME"].value_counts().sort_values(ascending = False)

print("Area: ",max_night_data.index[0]," with a value of ",max_night_data.iloc[0])
peak_night_crime_location = max_night_data.index[0]
```

```
Area:  Central  with a value of  2548
```

```python
# 3. Crimes by age group

# I will create an age_group column to map each age to its corresponding age group.

# This time, age is a continuous variable, so I can use pd.cut. Unlike the first problem,
# where the night period was split into two parts (22:00 to 23:59 and 00:00 to 03:59),
# here the values are continuous and can be binned easily.


labels = ["0-17","18-25","26-34","35-44","45-54","55-64","65+"]
bins = [0,17,25,34,44,54,64,float('inf')]
crimes["Age_group"] = pd.cut(crimes["Vict Age"],labels = labels, bins = bins)

victim_ages = crimes["Age_group"].value_counts().sort_values(ascending = False)
```