GREAT LEARNING

Crime LA Project

Data Analytics using SQL

Project designer: **DURGESH KUMAR JHA**

Session: 2024-25

Problem Statement

People throughout Los Angeles are concerned about recent reports of crimes in numerous locations. The mayor of Los Angeles has established a new Criminal Investigation Division to study how and why crime is on the rise, as well as the elements that contribute to it, so that officials may take the necessary steps to keep the city's residents safe.

Objective

As a member of the Analytics Division, you are aware that numerous questions must be answered utilizing CID data. Import the dump file that contains various tables that are present in the database. Using the data, react to the questions and write a detailed report for the authorities to utilize in taking action against crimes in Los Angeles.

Business Questions

Question 1: What is the total number of crimes for each crime status?

Solution Query:

SELECT

   case_status_desc,

   COUNT(*) AS case_count

FROM report_t

GROUP BY case_status_desc;

Output:

```
SQL queries passed

Query 1 (Passed): SELECT case_status_desc, COUNT(*) AS case_count FROM report_t GROUP BY case_status_desc
Output:
Showing 3 rows
```

case_status_desc	case_count
Adult Arrest	94
Adult Other	38
Invest Cont	1186

Adult Arrest (94 cases): This category indicates a significant number of adult arrests, which may reflect trends in law enforcement or social issues.

Adult Other (38 cases): This smaller count suggests that there are fewer cases that fall outside the standard arrest category, possibly indicating less frequent incidents or more specialized cases.

Investigation Continued (1186 cases): This is the largest category by far, suggesting that ongoing investigations are prevalent. It might imply a backlog or a high level of complexity in cases that require extended inquiry.

High Volume of Ongoing Investigations: The dominance of the "Investigation Continued" status could indicate that many cases are complex, leading to prolonged investigations. It may be worth analyzing the nature of these cases to understand the reasons behind the delays.

Potential Resource Allocation Issues: The disparity in case counts suggests that resources may need to be evaluated. If investigations are taking longer than expected, it could be beneficial to assess whether additional support is needed.

Question 2: Which was the most frequent crime committed each week?

Solution Query:

SELECT

week_number, crime_type, crimes_reported

FROM (

SELECT week_number, crime_type,

COUNT(*) AS crimes_reported,

RANK() OVER(PARTITION BY week_number ORDER BY COUNT(*) DESC) AS high_crime_reported

FROM report_t

GROUP BY week_number, crime_type

) AS wk_crime

WHERE wk_crime.high_crime_reported = 1;

Output:

Query 2 (Passed): SELECT week_number, crime_type, crimes_reported FROM (SELECT week_number, crime_type, COUNT(*) AS crimes_reported, RANK() OVER(PARTITION BY week_number ORDER BY COUNT(*) DESC) AS high_crime_reported FROM report_t GROUP BY week_number, crime_type) AS wk_crime WHERE wk_crime.high_crime_reported = 1
Output:
Showing 4 rows

week_number	crime_type	crimes_reported
1	BURGLARY FROM VEHI...	24
2	BURGLARY FROM VEHI...	40
3	BATTERY - SIMPLE AS...	48
4	BURGLARY FROM VEHI...	59

Observation and Insights:

Week 1: Burglary from Vehicle (24 crimes) *This indicates that vehicle-related burglaries were the most common crime reported in the first week.*

Week 2: Burglary from Vehicle (40 crimes) *The trend continues with an increase in vehicle burglaries, suggesting this may be a recurring issue.*

Week 3: Battery - Simple Assault (48 crimes) *A shift occurs in Week 3, with simple assaults becoming the most reported crime, indicating a potential spike in violent incidents during that period.*

Week 4: Burglary from Vehicle (59 crimes) *Burglary from vehicles reemerges as the dominant crime type, with the highest count in this dataset.*

Recurring Crime Type: The repeated appearance of "Burglary from Vehicle" as the highest reported crime in multiple weeks suggests a consistent issue. This could prompt law enforcement to focus on preventative measures or increased patrols in areas known for such crimes.

Trend Analysis: The fluctuation between different types of crimes highlights the importance of ongoing monitoring. Tracking these trends over a longer period can help identify whether these are anomalies or part of a larger pattern.

Question 3: Does the existence of CCTV cameras deter crimes from happening?

SELECT area_name,

 cctv_count,

  COUNT(*) AS cases_reported

  FROM report_t as r JOIN location_t as l ON l.area_code = r.area_code

  GROUP BY area_name, cctv_count ORDER BY cctv_count;
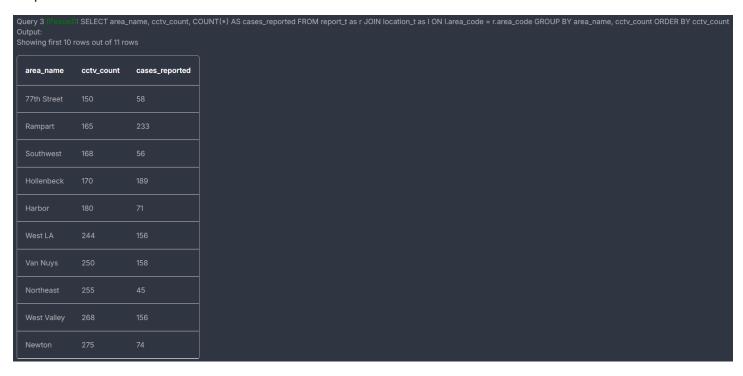
Output:

Query 3 (Passed): SELECT area_name, cctv_count, COUNT(*) AS cases_reported FROM report_t as r JOIN location_t as l ON l.area_code = r.area_code GROUP BY area_name, cctv_count ORDER BY cctv_count
Output:
Showing first 10 rows out of 11 rows

area_name	cctv_count	cases_reported
77th Street	150	58
Rampart	165	233
Southwest	168	56
Hollenbeck	170	189
Harbor	180	71
West LA	244	156
Van Nuys	250	158
Northeast	255	45
West Valley	268	156
Newton	275	74

High CCTV Coverage Areas: Hollywood has the highest number of CCTV cameras (280), followed closely by Newton (275) and West Valley (268). These areas are heavily monitored.

Consistent High Numbers: West Valley and Van Nuys both report over 150+ cases despite having a high number of CCTV cameras.

CCTV Effectiveness: The presence of CCTV cameras in an area does not directly correlate with a reduction in reported crimes. Van Nuys and West Valley, with high numbers of cameras, still show high crime rates.

Question 4: How much footage has been recovered from the CCTV at the crime scene?

Solution Query:

SELECT SUM(cctv_count) AS total_cctv_installed,

SUM(CASE WHEN cctv_flag = 'Yes' THEN 1 ELSE 0 END) AS total_cctv_footage_available

FROM report_t as r

JOIN location_t as l ON l.area_code = r.area_code

GROUP BY area_name;

Output:

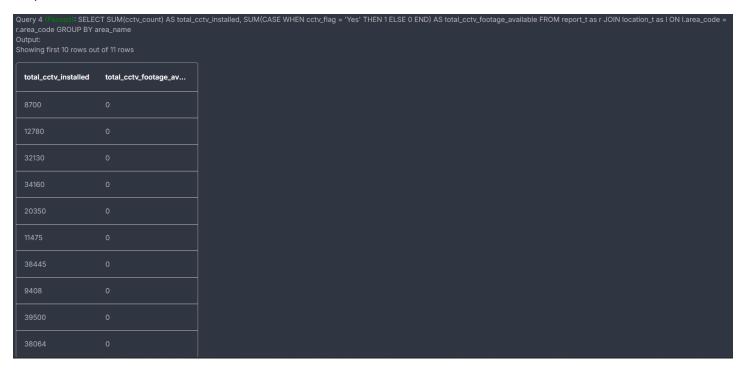Query 4 (Passed): SELECT SUM(cctv_count) AS total_cctv_installed, SUM(CASE WHEN cctv_flag = 'Yes' THEN 1 ELSE 0 END) AS total_cctv_footage_available FROM report_t as r JOIN location_t as l ON l.area_code = r.area_code GROUP BY area_name
Output:
Showing first 10 rows out of 11 rows

total_cctv_installed	total_cctv_footage_av...
8700	0
12780	0
32130	0
34160	0
20350	0
11475	0
38445	0
9408	0
39500	0
38064	0

Total CCTV Installed: *The numbers range from 8,700 to 38,064 across different areas.*

Total CCTV Footage Available: *All areas show 0 footage available, indicating that while there are many cameras installed, none are currently providing recorded footage.*

Question 5: What is the frequency of various complaint types?

Solution Query:

SELECT

complaint_type,

COUNT(*) AS cases_reported

FROM report_t

GROUP BY complaint_type;

Output:

complaint_type	cases_reported
Email	62
In-Person	446
Phone	810

Email Complaints (62 cases): A relatively low number of complaints reported via email indicates that this method may not be the preferred choice for most individuals. It might suggest a lack of awareness of this option or possibly a preference for more immediate forms of communication.

In-Person Complaints (446 cases): A significant number of complaints were made in person, reflecting a strong preference for direct interaction. This may indicate that individuals feel more comfortable discussing their concerns face-to-face or believe it leads to more effective resolutions.

Phone Complaints (810 cases): The highest number of complaints came via phone, suggesting that this is the most accessible and preferred method for reporting issues. It may also imply that people value the immediacy and convenience of phone communication.

Preference for Direct Communication: The data shows a clear preference for in-person and phone complaints over email. This could inform how resources are allocated for customer service, suggesting an emphasis on staffing for phone lines and in-person interactions.

Question 6: Is crime more likely to be committed by the relation of victims or strangers?
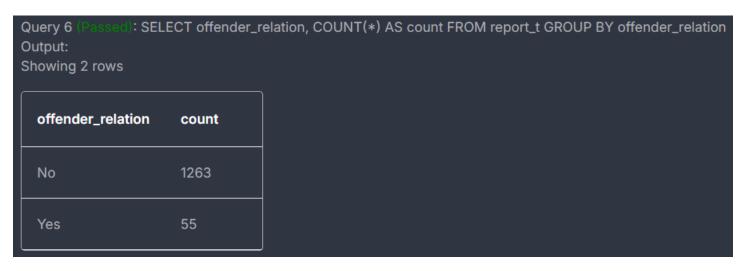
Solution Query:

SELECT

    offender_relation,

    COUNT(*) AS count

FROM report_t

GROUP BY offender_relation;

Output:

Query 6 (Passed): SELECT offender_relation, COUNT(*) AS count FROM report_t GROUP BY offender_relation
Output:
Showing 2 rows

offender_relation	count
No	1263
Yes	55

No (1263 cases): *This category indicates that the majority of cases (96.6%) involve offenders who do not have a known relationship with the victims. This suggests that many incidents may be stranger-related or involve individuals who are not connected in any significant way.*

Yes (55 cases): *A small number of cases (4.4%) involve offenders who are known to the victims. This could include relationships such as acquaintances, family members, or partners, indicating a different dynamic in those incidents.*

Prevalence of Unknown Offenders: *The significant proportion of cases with no known relationship could indicate broader social issues, such as random acts of violence or crime. This might inform community safety initiatives or law enforcement strategies.*

Focus on Relationship Dynamics: *The smaller number of cases involving known offenders highlights the importance of understanding the context in which these crimes occur.*

Resource Allocation: *Given the overwhelming number of cases without a known relationship, law enforcement might consider strategies focusing on community outreach and crime prevention initiatives that address issues contributing to stranger-related crimes.*

Question 7: Is crime more prevalent in areas with a higher population density, fewer police personnel, and a larger precinct area?

Solution Query:

SELECT precinct_code,

SUM(l.population_density) AS pop_density,

COUNT(DISTINCT l.area_code) AS total_areas,

COUNT(DISTINCT o.officer_code) AS total_officers,

COUNT(r.report_no) AS cases_reported

FROM report_t AS r

JOIN location_t AS l ON l.area_code = r.area_code

JOIN officer_t AS o ON precinct_code = o.precinct_code

GROUP BY precinct_code

ORDER BY precinct_code LIMIT 5;

Output:

precinct_code	pop_density	total_areas	total_officers	cases_reported
1	53304468	11	7	9226
2	68534316	11	9	11862
3	83764164	11	11	14498
4	76149240	11	10	13180
5	45689544	11	6	7908

Precinct Codes: The data covers precincts 1 through 5.

Population Density: Ranges from about 45.7 million to 83.8 million, indicating significant variation in density across precincts

Total Areas: Each precinct covers 11 distinct areas

Total Officers: The number of officers varies from 6 to 11 across precincts.

Cases Reported: The number of reported cases ranges from 7,908 to 14,498, with Precinct 3 having the highest case count

Question 8: At what parts of the day is the crime rate at its peak? Group this by the type of crime. Use the following mapping to divide the day into five parts.

SELECT

*   dayparts, crime_type, crimes_reported*

FROM (

*   SELECT*

*   CASE*

*   WHEN strftime('%H:%M', r.incident_time) >= '00:00' AND strftime('%H:%M', r.incident_time) < '05:00' THEN 'Midnight'*

*   WHEN strftime('%H:%M', r.incident_time) >= '05:01' AND strftime('%H:%M', r.incident_time) < '12:00' THEN 'Morning'*

*   WHEN strftime('%H:%M', r.incident_time) >= '12:01' AND strftime('%H:%M', r.incident_time) < '18:00' THEN 'Afternoon'*

*   WHEN strftime('%H:%M', r.incident_time) >= '18:01' AND strftime('%H:%M', r.incident_time) <= '21:00' THEN 'Evening'*

*   ELSE 'Night'*

*   END AS dayparts,*

*   crime_type,*

*   COUNT(*) AS crimes_reported,*

*   RANK() OVER(PARTITION BY CASE*

WHEN strftime('%H:%M', r.incident_time) >= '00:00' AND strftime('%H:%M', r.incident_time) < '05:00' THEN 'Midnight'

        WHEN strftime('%H:%M', r.incident_time) >= '05:01' AND strftime('%H:%M', r.incident_time) < '12:00' THEN 'Morning'

        WHEN strftime('%H:%M', r.incident_time) >= '12:01' AND strftime('%H:%M', r.incident_time) < '18:00' THEN 'Afternoon'

        WHEN strftime('%H:%M', r.incident_time) >= '18:01' AND strftime('%H:%M', r.incident_time) <= '21:00' THEN 'Evening'

        ELSE 'Night'

    END ORDER BY COUNT(*) DESC) AS high_crime_reported

  FROM report_t AS r

  JOIN location_t AS l ON l.area_code = r.area_code

  GROUP BY dayparts, crime_type

) AS wk_crime

WHERE wk_crime.high_crime_reported = 1;

Output:

```
Query 8 (Passed): SELECT dayparts, crime_type, crimes_reported FROM ( SELECT CASE WHEN strftime('%H:%M', r.incident_time) >= '00:00' AND strftime('%H:%M', r.incident_time) < '05:00' THEN 'Midnight' WHEN strftime('%H:%M', r.incident_time) >= '05:01' AND strftime('%H:%M', r.incident_time) < '12:00' THEN 'Morning' WHEN strftime('%H:%M', r.incident_time) >= '12:01' AND strftime('%H:%M', r.incident_time) < '18:00' THEN 'Afternoon' WHEN strftime('%H:%M', r.incident_time) >= '18:01' AND strftime('%H:%M', r.incident_time) <= '21:00' THEN 'Evening' ELSE 'Night' END AS dayparts, crime_type, COUNT(*) AS crimes_reported, RANK() OVER(PARTITION BY CASE WHEN strftime('%H:%M', r.incident_time) >= '00:00' AND strftime('%H:%M', r.incident_time) < '05:00' THEN 'Midnight' WHEN strftime('%H:%M', r.incident_time) >= '05:01' AND strftime('%H:%M', r.incident_time) < '12:00' THEN 'Morning' WHEN strftime('%H:%M', r.incident_time) >= '12:01' AND strftime('%H:%M', r.incident_time) < '18:00' THEN 'Afternoon' WHEN strftime('%H:%M', r.incident_time) >= '18:01' AND strftime('%H:%M', r.incident_time) <= '21:00' THEN 'Evening' ELSE 'Night' END ORDER BY COUNT(*) DESC) AS high_crime_reported FROM report_t AS r JOIN location_t AS l ON l.area_code = r.area_code GROUP BY dayparts, crime_type ) AS wk_crime WHERE wk_crime.high_crime_reported = 1
Output:
Showing 5 rows
```

dayparts	crime_type	crimes_reported
Afternoon	BATTERY - SIMPLE AS...	51
Evening	BURGLARY FROM VEHI...	54
Midnight	BATTERY - SIMPLE AS...	19
Morning	BATTERY - SIMPLE AS...	49
Night	BURGLARY FROM VEHI...	39

Afternoon: Battery - Simple Assault (51 cases) *The most reported crime during the afternoon hours, suggesting that this time period may see increased interpersonal conflicts or disputes leading to assaults.*

Evening: Burglary from Vehicle (54 cases) *This indicates that evenings may be a peak time for vehicle-related burglaries, potentially linked to reduced visibility and more opportunities for theft.*

Midnight: Battery - Simple Assault (19 cases) *A lower number of reported assaults at midnight compared to the afternoon, but still significant, which may indicate late-night disturbances or conflicts.*

Morning: Battery - Simple Assault (49 cases) *Similar to the afternoon, assaults remain prevalent in the morning, suggesting ongoing issues related to personal conflicts or disturbances.*

Night: Burglary from Vehicle (39 cases) *Like the evening, the night hours also see a considerable number of vehicle burglaries, indicating a pattern of criminal activity during these hours.*

Prevalence of Battery - Simple Assault: The consistent reporting of battery across multiple dayparts suggests that interpersonal conflicts are a significant issue at different times of day. This may warrant targeted intervention strategies to reduce such incidents, especially in community outreach and education on conflict resolution.

Vehicle Burglaries: The spike in burglaries from vehicles during the evening and night suggests that these times are vulnerable for such crimes. Increased patrols or community awareness campaigns focused on securing vehicles may help reduce these incidents.

Time-Based Crime Patterns: Understanding when crimes are most likely to occur can help law enforcement allocate resources more effectively. For example, enhancing visibility and patrols in the identified high-crime periods could deter potential offenders.

Question 9: At what point in the day do most crimes occur in different localities? Use the same mapping provided in Question 8 to divide the day into five parts.

Solution Query:

```
SELECT
    area_name, dayparts, cases_reported
FROM (
    SELECT
        area_name,
        CASE
            WHEN strftime('%H:%M', r.incident_time) >= '00:00' AND strftime('%H:%M', r.incident_time) < '05:00' THEN 'Midnight'
            WHEN strftime('%H:%M', r.incident_time) >= '05:01' AND strftime('%H:%M', r.incident_time) < '12:00' THEN 'Morning'
            WHEN strftime('%H:%M', r.incident_time) >= '12:01' AND strftime('%H:%M', r.incident_time) < '18:00' THEN 'Afternoon'
            WHEN strftime('%H:%M', r.incident_time) >= '18:01' AND strftime('%H:%M', r.incident_time) <= '21:00' THEN 'Evening'
            ELSE 'Night'
        END AS dayparts,
        RANK() OVER(PARTITION BY l.area_name ORDER BY COUNT(*) DESC) AS high_crime_reported,
        COUNT(*) AS cases_reported
    FROM report_t AS r
    JOIN location_t AS l ON l.area_code = r.area_code
    GROUP BY l.area_name, dayparts
) AS d_case
WHERE d_case.high_crime_reported = 1;
```

Output:

Query 9 (Passed): SELECT area_name, dayparts, cases_reported FROM (SELECT area_name, CASE WHEN strftime('%H:%M', r.incident_time) >= '00:00' AND strftime('%H:%M', r.incident_time) < '05:00' THEN 'Midnight' WHEN strftime('%H:%M', r.incident_time) >= '05:01' AND strftime('%H:%M', r.incident_time) < '12:00' THEN 'Morning' WHEN strftime('%H:%M', r.incident_time) >= '12:01' AND strftime('%H:%M', r.incident_time) < '18:00' THEN 'Afternoon' WHEN strftime('%H:%M', r.incident_time) >= '18:01' AND strftime('%H:%M', r.incident_time) <= '21:00' THEN 'Evening' ELSE 'Night' END AS dayparts, RANK() OVER(PARTITION BY l.area_name ORDER BY COUNT(*) DESC) AS high_crime_reported, COUNT(*) AS cases_reported FROM report_t AS r JOIN location_t AS l ON l.area_code = r.area_code GROUP BY l.area_name, dayparts) AS d_case WHERE d_case.high_crime_reported = 1
Output:
Showing first 10 rows out of 12 rows

area_name	dayparts	cases_reported
77th Street	Evening	15
77th Street	Afternoon	15
Harbor	Afternoon	23
Hollenbeck	Afternoon	52
Hollywood	Afternoon	37
Newton	Afternoon	18
Northeast	Afternoon	15
Rampart	Afternoon	71
Southwest	Afternoon	18
Van Nuys	Afternoon	43

77th Street *has a notable number of cases reported in both the Evening and Afternoon, each with 15 cases.*

Hollenbeck *leads with 52 cases in the Afternoon.*

Rampart *has the highest count overall with 71 cases, also in the Afternoon*

*Other areas like **Hollywood** and **Van Nuys** also show significant Afternoon activity, with 37 and 43 cases, respectively*

Question 10: Which age group is more likely to fall victim to crimes at certain points in the day? Use the same mapping provided in Question 8 to divide the day into five parts. Additionally, use the following mapping to divide the age group.

Age 0 to 12: kids

13 to 23: teenage

24 to 35: middle age

36 to 55: adults

56 to 120: old

Solution Query:

SELECT

CASE

 WHEN strftime('%H', incident_time) >= '00' AND strftime('%H', incident_time) < '05' THEN 'Midnight'

WHEN strftime('%H', incident_time) >= '05' AND strftime('%H', incident_time) < '12' THEN 'Morning'

WHEN strftime('%H', incident_time) >= '12' AND strftime('%H', incident_time) < '18' THEN 'Afternoon'

WHEN strftime('%H', incident_time) >= '18' AND strftime('%H', incident_time) <= '21' THEN 'Evening'

ELSE 'Night'

 END AS dayparts,

CASE

WHEN victim_age >= 0 AND victim_age <= 12 THEN 'Kids'

WHEN victim_age > 12 AND victim_age <= 23 THEN 'Teenage'

WHEN victim_age > 23 AND victim_age <= 35 THEN 'Middle age'

WHEN victim_age > 35 AND victim_age <= 55 THEN 'Adults'

WHEN victim_age > 55 THEN 'Old'

ELSE 'Unknown'

END AS age_cat,

COUNT(*) AS cases_reported

FROM report_t AS r

JOIN victim_t AS v ON v.victim_code = r.victim_code

GROUP BY dayparts, age_cat

ORDER BY age_cat, dayparts;

Output:

Query 19 (Passed): SELECT CASE WHEN strftime('%H', incident_time) >= '00' AND strftime('%H', incident_time) < '05' THEN 'Midnight' WHEN strftime('%H', incident_time) >= '05' AND strftime('%H', incident_time) < '12' THEN 'Morning' WHEN strftime('%H', incident_time) >= '12' AND strftime('%H', incident_time) < '18' THEN 'Afternoon' WHEN strftime('%H', incident_time) >= '18' AND strftime('%H', incident_time) <= '21' THEN 'Evening' ELSE 'Night' END AS dayparts, CASE WHEN victim_age >= 0 AND victim_age <= 12 THEN 'Kids' WHEN victim_age > 12 AND victim_age <= 23 THEN 'Teenage' WHEN victim_age > 23 AND victim_age <= 35 THEN 'Middle age' WHEN victim_age > 35 AND victim_age <= 55 THEN 'Adults' WHEN victim_age > 55 THEN 'Old' ELSE 'Unknown' END AS age_cat, COUNT(*) AS cases_reported FROM report_t AS r JOIN victim_t AS v ON v.victim_code = r.victim_code GROUP BY dayparts, age_cat ORDER BY age_cat, dayparts
Output:
Showing first 10 rows out of 25 rows

dayparts	age_cat	cases_reported
Afternoon	Adults	229
Evening	Adults	183
Midnight	Adults	110
Morning	Adults	163
Night	Adults	57
Afternoon	Kids	41
Evening	Kids	25
Midnight	Kids	13
Morning	Kids	20
Night	Kids	5

Adults as Victims: Adults (ages 36 - 55) are the most frequent victims of crime, especially in the afternoon (229 cases) and evening (183 case).

Children's Vulnerability: Kids (ages 0 - 12) show a notable number of cases in the afternoon (41 cases) and evening (25 cases), but significantly fewer during the night (5 cases).

Midnight to Morning Crime: Both adults and kids are less likely to be victims of crime during the midn ight hours, though there are still notable instances (110 for adults, 13 for kids).

Resource Focus: Law enforcement should allocate resources to protect adults and kids especi ally during the afternoon and evening times, which are peak times for victimization.

Adult Safety Programs: For adults, afternoon and evening crime prevention programs could be effective, focusing on awareness and personal safety strategies.

Crime Metrics Overview

No. of Precinct	Total Crimes Reported	No. of Areas	No. of Offenders
54	1318	11	1318
No. of Officers	**Total Population**	**No. of CCTV's**	**No. of Cases where Status is IC**
54	62683	2405	1186

Recommendations

Investigate Ongoing Cases: *Analyse the complexity of cases marked as "Investigation Continued" to address potential backlog issues and improve resolution rates.*

Enhance CCTV Utility: *Assess the functionality of CCTV systems to ensure footage is available when needed. Increased maintenance and checks may help.*

Resource Allocation: *Allocate more officers to high-crime areas, particularly during peak times identified (Afternoon and Evening).*

Targeted Outreach for Adults: *Given the high number of adult victims, consider targeted outreach and education programs focusing on conflict resolution and safety*

Community Engagement: *Given the high incidence of battery and vehicle-related crimes, implement community awareness programs focusing on personal safety and crime prevention strategies.*