Crime Data

SQL Project

danielokoli448@gmail.com
OKOLI NNAMDI NNAEMEKA | 09067260556

Professional Background

With over three years of versatile experience in both office and freelance settings, I have honed my skills as a data analyst. My impactful tenure at Guinness Nigeria was marked by the optimization of supply chain operations through strategic data analysis, resulting in a notable reduction in processing time.

As a National Youth Service Corps (NYSC) graduate, I served as a Data Entry personnel, utilizing Microsoft Excel to navigate extensive datasets and presenting insights through clear Excel tables and PowerPoint presentations. Transitioning to the role of IT Assistant at Delta Broadcasting Service (DBS) Warri, I further strengthened my data management skills over six months.

My commitment to continuous learning is evident in my proficiency in SQL, MS Excel, and Python. Additionally, I bring expertise in visualization tools such as Tableau and Power BI, with ongoing efforts to enhance my skills in R language. Practical applications include the creation of interactive dashboards using Power BI and experimentation with R language in controlled environments.

Notably, my analytical approach focuses on tangible outcomes and quantifiable achievements. At Guinness Nigeria, my data-driven strategies had a direct and positive impact on supply chain efficiency. Eager to contribute this expertise to a remote data analyst role, I am poised to deliver results through a combination of technical proficiency and strategic problem-solving.

Table of Contents

Introduction

This dataset reflects incidents of crime in the City of Los Angeles dating back to 2020. This data is transcribed from original crime reports that are typed on paper and therefore there may be some inaccuracies within the data. Some location fields with missing data are noted as (0°, 0°). Address fields are only provided to the nearest hundred block in order to maintain privacy.

The problem is to better understand the why there is crime in the city of Los Angeles to try and reduce crime rates in the city.

I used the dataset Crime_Data to analyze and answer this problem.

I applied SQL commands to analyse data: JOIN, ORDER BY, AS, WHERE, AND, OR, SUM(), COUNT(), GROUP BY, LIKE, HAVING.

Also, I used Root Cause Analysis to understand the problem and ask right questions.

As a result, I have found out crucial insights of provided data sets, prepared visualisations, and report for my team.

Root Cause Analysis

The problem i want to analyze in this dataset is to better understand the why there is crime in the city of Los Angeles to try and reduce crime rates in the city but first, what is root cause analysis?

Root cause analysis (RCA) is the process of discovering the root causes of problems in order to identify appropriate solutions. RCA assumes that it is much more effective to systematically prevent and solve for underlying issues rather than just treating ad hoc symptoms and putting out fires. Root cause analysis can be performed with a collection of principles, techniques, and methodologies that can all be leveraged to identify the root causes of an event or trend. Looking beyond superficial cause and effect, RCA can show where processes or systems failed or caused an issue in the first place. Therefore, I'm trying to discover the root causes of the problem from this dataset.

To understand the problem, I need to analyze the existing database of crime in the city of LA. I will also present some crucial numbers and visualization of the datasets. So, I decided to ask some questions.

- How many recorded crimes do we currently have in our database?
- What timeframe does this dataset cover?
- How old is the oldest victim?
- How much crime was committed each year in the database?
- Where is the safest area to live in LA?

Also, I decided to apply Root Cause Analysis to the problem to figure out the underlying issues in order to identify appropriate solutions.

1. Why is there crime in the city of Los Angeles?
   Ans. There are so much unsolved crimes in LA. Perpetrators of these crimes aren't frequently caught.

2. Why are there so much unsolved crimes in the city of LA?
   Ans. Most unsolved crimes are petty crimes so the police or law enforcement agency of the city may not give top priority to them.

3. Why do so many petty crimes occur in LA?
   Ans. The victims of these crimes are mostly focused on youths between the age of 20 and 40.

4. Why is crime focused mainly on the youths?
   Ans. Poor security in the neighborhoods is the reason for this.

5. Why is there poor security in the city of LA?
   Ans. There could be a variety of reasons for this. Inadequate policing, poor sense of security from youths, absence of security cameras in key areas ranging from streets to individual homes.

Insights from the Analysis

I made use of a relational databases Crime_data to answer the problem.

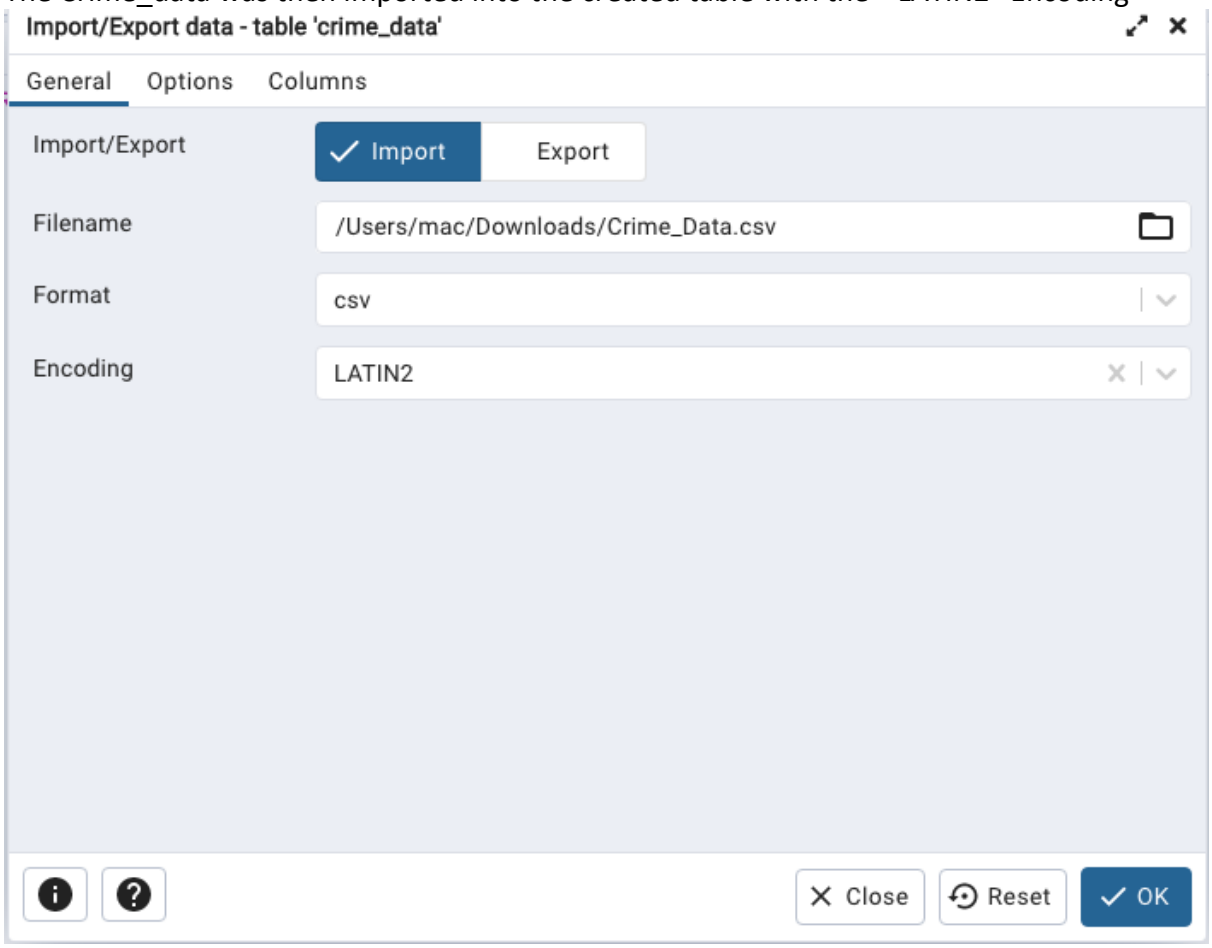
POSTGRESQl Database Management System was used to find out main insights.

CRIME_DATA includes such data:
- DR_NO
- Date Rptd
- DATE OCC
- TIME OCC
- AREA
- AREA NAME
- Rpt Dist No
- Crm Cd
- Crm Cd Desc
- Vict Age
- Vict sex
- Vict Descent
- Premis Cd
- Premis Desc
- Weapon Used Cd
- Weapon Desc
- Status
- Status Desc
- Crm Cd 1
- Crm Cd 2
- LOCATION
- Cross Street
- LAT
- LON

A crime_data table was created using Postgresql.

```
CREATE TABLE crime_data (
    DR_NO SERIAL PRIMARY KEY,
    Date_Rptd TIMESTAMP,
    DATE_OCC TIMESTAMP,
    TIME_OCC INTEGER,
    AREA INTEGER,
    AREA_NAME VARCHAR(40),
    Rpt_Dist_No INTEGER,
    Crm_Cd INTEGER,
    Crm_Cd_Desc VARCHAR(100),
    Vict_Age INTEGER,
    Vict_sex CHAR(1),
    Vict_Descent CHAR(1),
    Premis_Cd INTEGER,
    Premis_Desc VARCHAR(100),
    Weapon_Used_Cd INTEGER,
    Weapon_Desc VARCHAR(100),
    Status CHAR(2),
    Status_Desc VARCHAR(100),
    Crm_Cd_1 INTEGER,
    Crm_Cd_2 INTEGER,
    LOCATION VARCHAR(100),
    Cross_Street VARCHAR(50),
    LAT REAL,
    LON REAL
);
```

The Crime_data was then imported into the created table with the "LATIN2" Encoding



SELECT statement was used to fetch data from a database.

```
29  SELECT * FROM crime_data;
```

To find the recorded number of crimes with the COUNT() Function, I used a command like this:

```
31  SELECT COUNT(dr_no) FROM crime_data;
```

To get the timeframe of the dataset, I used the MIN() and MAX():

```
33  SELECT MAX(DATE_OCC)
34  FROM crime_data;
35
36  SELECT MIN(DATE_OCC)
37  FROM crime_data;
```

To know the age of the oldest victim, I used MAX():

```
39  SELECT MAX(Vict_Age)
40  FROM crime_data;
```

I used the COUNT() function to find the number of crimes committed each year:

```
45  SELECT COUNT(dr_no) FROM crime_data
46  WHERE DATE_OCC < '2021-01-01';
47
48
49  SELECT COUNT(dr_no) FROM crime_data
50  WHERE DATE_OCC BETWEEN '2021-01-01' AND '2021-12-31';
51
52  SELECT COUNT(dr_no) FROM crime_data
53  WHERE DATE_OCC BETWEEN '2022-01-01' AND '2022-12-31';
54
55  SELECT COUNT(dr_no) FROM crime_data
56  WHERE DATE_OCC BETWEEN '2023-01-01' AND '2023-11-11';
```

I used the GROUP BY function to know the safest area to live in LA.

```
58  SELECT AREA_NAME, COUNT(*)
59  FROM crime_data
60  GROUP BY AREA_NAME
61  ORDER BY COUNT(*) DESC;
```

I used the GROUP BY function to find the number of victims under each status description. Also, I used the AS command to give an alias to COUNT(*). The ORDER BY query was used to organize the data:

```
63  SELECT Status_Desc, COUNT(*) AS Number_of_victims
64  FROM crime_data
65  GROUP BY Status_Desc
66  ORDER BY COUNT(*) DESC;
```

I used the GROUP BY function to find the number of victims under each crime description. Also, I used the AS command to give an alias to COUNT(*). The ORDER BY query was used to organize the data. The HAVING clause restricts the query to 10000 victims or more.

```
84  SELECT Crm_Cd, Crm_Cd_Desc, COUNT(*) AS Number_of_victims
85  FROM crime_data
86  WHERE Status_Desc = 'Invest Cont'
87  GROUP BY Crm_Cd, Crm_CD_Desc
88  HAVING COUNT(*) > 10000
89  ORDER BY COUNT(*) DESC;
```

To know the number of victims under each crime description where investigation was not completed and difference in age.

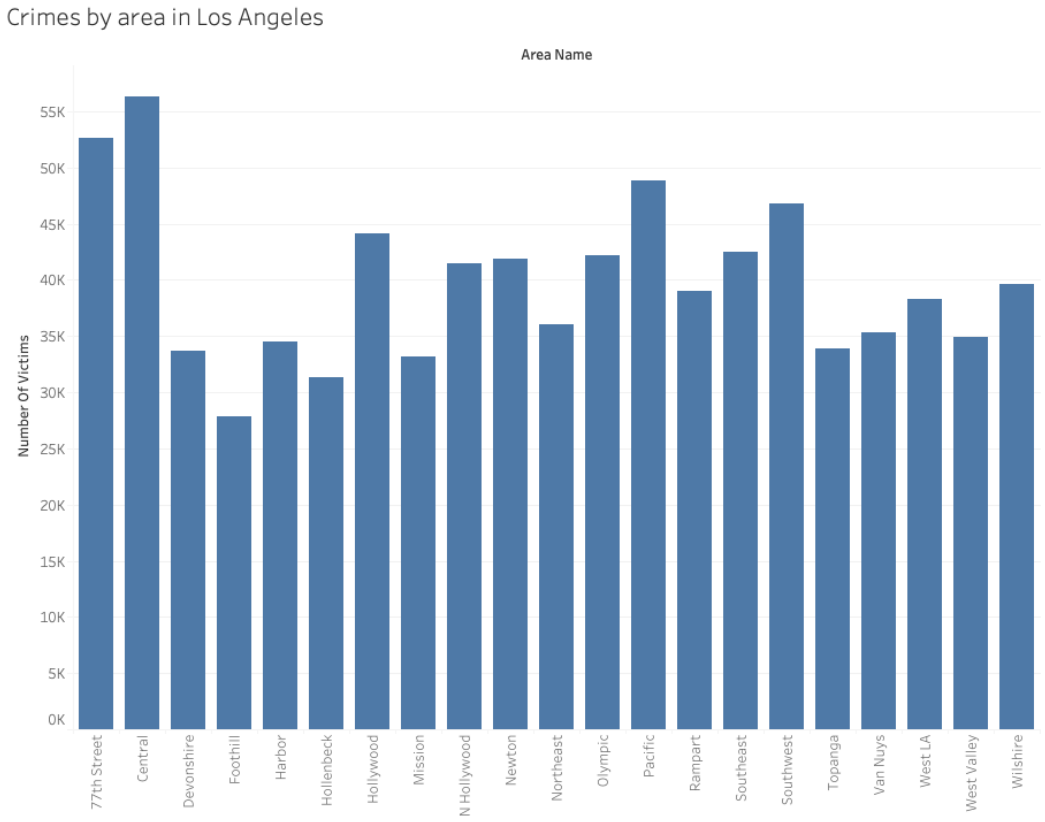
```
84  SELECT Crm_Cd, Crm_Cd_Desc, COUNT(*) AS Number_of_victims
85  FROM crime_data
86  WHERE Status_Desc = 'Invest Cont'
87  AND Vict_age BETWEEN 20 AND 40
88  GROUP BY Crm_Cd, Crm_CD_Desc
89  HAVING COUNT(*) > 10000
90  ORDER BY COUNT(*) DESC;
91
92  SELECT Crm_Cd, Crm_Cd_Desc, COUNT(*) AS Number_of_victims
93  FROM crime_data
94  WHERE Status_Desc = 'Invest Cont'
95  AND Vict_age > 70
96  GROUP BY Crm_Cd, Crm_CD_Desc
97  ORDER BY COUNT(*) DESC;
```

To know the number of victims under each crime description where investigation was not completed, the difference in age and premise description.

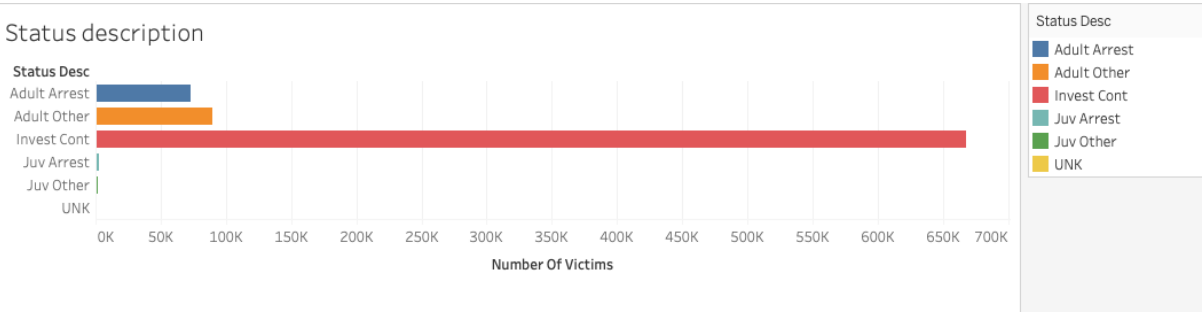
```
100 SELECT Premis_Cd, Premis_Desc, COUNT(*) ,
101 FROM crime_data
102 WHERE Status_Desc = 'Invest Cont'
103 AND Vict_age BETWEEN 20 AND 40
104 GROUP BY Premis_Cd, Premis_Desc
105 ORDER BY COUNT(*) DESC
106 limit 5;
107
108 SELECT Premis_Cd, Premis_Desc, COUNT(*)
109 FROM crime_data
110 WHERE Status_Desc = 'Invest Cont'
111 AND Vict_age > 70
112 GROUP BY Premis_Cd, Premis_Desc
113 ORDER BY COUNT(*) DESC
114 limit 5;
```

Tableaus was used as a very powerful tool for data analysis and Visualization for better understanding of the data.

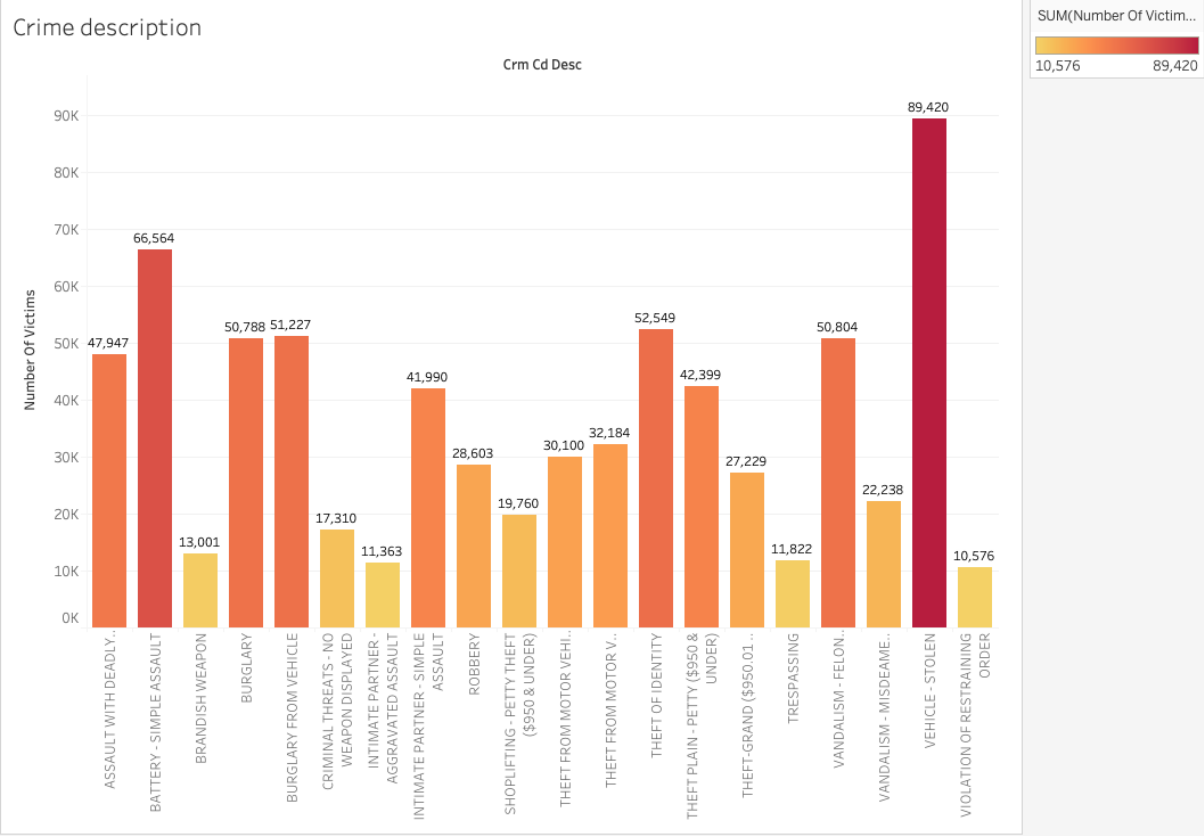
This Tableau clearly shows that majority of crimes are occur in the central part of LA. While, Foothill is the safest area in LA with the least crimes committed.



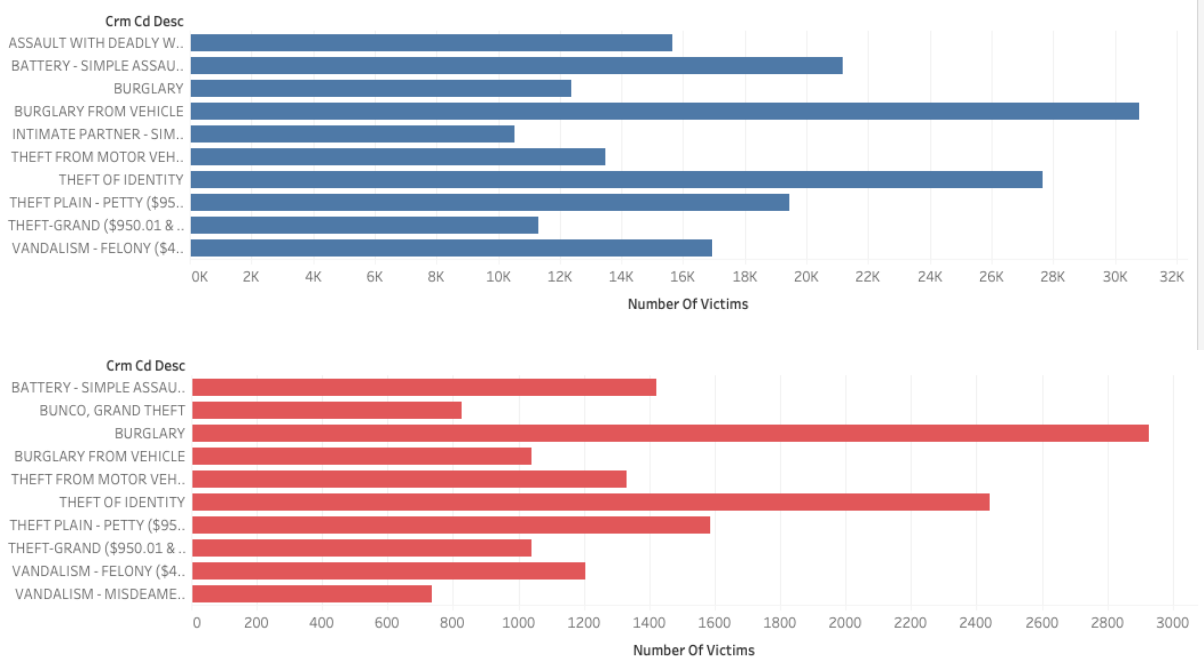
This visualization shows the status of all crimes in the dataset.



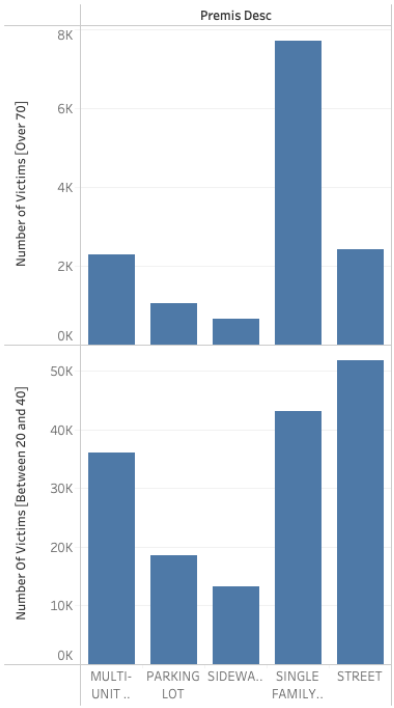
The visualization below shows top 20 most common crimes in LA.



These visualizations show the top 10 crimes on victims between the age of 20 and 40 and over 70 respectively



This visualization shows the premise most crimes are committed on victims between the age of 20 and 40 and over 70 respectively



Findings and Recommendations

Here are the results of the data set analysis:

- The total number of recorded crimes we have in the database is 834,320
- The dataset covers over 3 years of crimes in LA between 01/01/2020 and 06/11/2023
- The oldest victim is 120
- 199,439 crimes was recorded in 2020, 209,401 crimes was recorded in 2021, 234,311 crimes was recorded in 2022, 191,169 crimes was recorded in 2023
- The safest area to live in is Foothill.

Record of crimes committed in all areas in LA

area_name	number_of_victims
Central	56268
77th Street	52602
Pacific	48832
Southwest	46757
Hollywood	44177
Southeast	42457
Olympic	42156
Newton	41879
N Hollywood	41490
Wilshire	39669
Rampart	39063
West LA	38298
Northeast	36018
Van Nuys	35353
West Valley	34867
Harbor	34481
Topanga	33891
Devonshire	33722
Mission	33218
Hollenbeck	31321
Foothill	27801

Thus, we can see the number of crimes vary heavily across each area.

Status of each crime from arrests to continued investigation.

status_desc	number_of_victims
Invest Cont	667644
Adult Other	89864
Adult Arrest	72652
Juv Arrest	2710

Juv Other	1446
UNK	4

We can see that around 80% of crimes in LA are still under investigation. Meaning, a large number of these crime perpetrators aren't caught and left to commit more crimes.

The count of different types of crimes in LA over 10,000:

crm_cd	crm_cd_desc	number_of_victims
510	VEHICLE - STOLEN	89420
624	BATTERY - SIMPLE ASSAULT	66564
354	THEFT OF IDENTITY	52549
330	BURGLARY FROM VEHICLE	51227
740	VANDALISM - FELONY ($400 & OVER, ALL CHURCH VANDALISMS)	50804
310	BURGLARY	50788
230	ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT	47947
440	THEFT PLAIN - PETTY ($950 & UNDER)	42399
626	INTIMATE PARTNER - SIMPLE ASSAULT	41990
420	THEFT FROM MOTOR VEHICLE - PETTY ($950 & UNDER)	32184
331	THEFT FROM MOTOR VEHICLE - GRAND ($950.01 AND OVER)	30100
210	ROBBERY	28603
341	THEFT-GRAND ($950.01 & OVER)EXCPT,GUNS,FOWL,LIVESTK,PROD	27229
745	VANDALISM - MISDEAMEANOR ($399 OR UNDER)	22238
442	SHOPLIFTING - PETTY THEFT ($950 & UNDER)	19760
930	CRIMINAL THREATS - NO WEAPON DISPLAYED	17310
761	BRANDISH WEAPON	13001
888	TRESPASSING	11822
236	INTIMATE PARTNER - AGGRAVATED ASSAULT	11363
901	VIOLATION OF RESTRAINING ORDER	10576

We can note that there are a lot more petty crimes committed at a higher rate in LA.

Top 10 crimes against youth victims (Between 20 and 40 years old)

crm_cd	crm_cd_desc	number_of_victims
330	BURGLARY FROM VEHICLE	30793
354	THEFT OF IDENTITY	27666
624	BATTERY - SIMPLE ASSAULT	21189
440	THEFT PLAIN - PETTY ($950 & UNDER)	19456

740	VANDALISM - FELONY ($400 & OVER, ALL CHURCH VANDALISMS)	16927
230	ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT	15656
331	THEFT FROM MOTOR VEHICLE - GRAND ($950.01 AND OVER)	13467
310	BURGLARY	12364
341	THEFT-GRAND ($950.01 & OVER)EXCPT,GUNS,FOWL,LIVESTK,PROD	11322
626	INTIMATE PARTNER - SIMPLE ASSAULT	10533

Top 10 crimes against older victims (over 70 years old)

crm_cd	crm_cd_desc	number_of_victims
310	BURGLARY	2927
354	THEFT OF IDENTITY	2442
440	THEFT PLAIN - PETTY ($950 & UNDER)	1587
624	BATTERY - SIMPLE ASSAULT	1423
331	THEFT FROM MOTOR VEHICLE - GRAND ($950.01 AND OVER)	1331
740	VANDALISM - FELONY ($400 & OVER, ALL CHURCH VANDALISMS)	1204
341	THEFT-GRAND ($950.01 & OVER)EXCPT,GUNS,FOWL,LIVESTK,PROD	1041
330	BURGLARY FROM VEHICLE	1040
662	BUNCO, GRAND THEFT	827
745	VANDALISM - MISDEAMEANOR ($399 OR UNDER)	734

Here, we can see the difference in crimes against the older and younger victims.

Top 5 premises crimes occur against youth victims (Between 20 and 40 years old)

premis_cd	premis_desc	number_of_victims
101	STREET	51839
501	SINGLE FAMILY DWELLING	43132
502	MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)	36078
108	PARKING LOT	18486
102	SIDEWALK	13220

Top 5 premises crimes occur against youth victims (over 70 years old)

premis_cd	premis_desc	count
501	SINGLE FAMILY DWELLING	7725

101	STREET	2432
502	MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)	2295
108	PARKING LOT	1064
102	SIDEWALK	662

Here, we can see the top 5 places crimes occurs the most between this two age group.

Conclusion

I have analysed the dataset Crime_Data to help reduce the crime rate in Los Angeles to improve security and safeguard lives and properties.

Hence, our victims are very dissimilar people. They live in different areas in LA, they are of different age groups, they are of different descent, the time crimes are committed varies, they fall victim to these crimes in different places like homes, streets, parking lots etc.

However, there are some crucial points that we need to count and try to use to curb crimes in the city of Los Angeles.

To begin with, I found out that most crimes are still under investigation meaning that the perpetrators of these crimes are still out there left to participate in more crimes. This should be immediately dealt with as these individuals need to be caught to reduce crime in the city.

Furthermore, I noticed a significant number of the crimes in LA are petty crimes i.e non-major crimes. So, the law enforcement agencies in the city may not see them as a priority when they are reported encouraging more individuals to participate in more of these types of crimes because they feel they won't be caught.
Therefore, priority should also be placed on these kinds of crimes to reduce the number of them that occurs.

Moreover, I observed that the level of security in LA is low. Crimes frequently occur in homes, parking lots and even the open street. Better security would help curb crime in the city. Thus, security cameras should be placed in public places like streets and parking lots and individuals should be encouraged to have them in their homes too. Security personnel should also be assigned to these streets or areas with high crime rates to help reduce crime.

In conclusion, we must concentrate to improve security in the city, making petty crimes a priority and placing heavier penalties on those who commit these petty crimes. Also, a greater effort should be put to catch these individuals to get more arrest so they would think twice before attempting to commit a crime. There is lack of valuable information in the dataset. More research should be made to improve this.