LA CRIME DATA ANALYSIS (2020- PRESENT)

CALIFORNIA STATE UNIVERSITY LOS ANGELES

CIS 5270- BI PROJECT

SUBMITTED TO: Dr. Shilpa Balan

SUBMITTED BY: Shailja Pandit (401971060)

TABLE OF CONTENTS

INTRODUCTION

The Los Angeles Police Department (LAPD), officially known as the City of Los Angeles Police Department, is the municipal police department of Los Angeles, California. It was formed in 1869, after the city of Los Angeles was incorporated. It was founded by Charles A. Rowland. It is the Third-Largest Municipal police department in the US, after NYPD and Chicago PD. Some of their most famous cases include solving the Black Dahlia murder case and arresting OJ Simpson for murder.

This dataset reflects incidents of crime in the City of Los Angeles dating back to 2020. The dataset focuses on different categories of crimes committed in different parts of the city of Los Angeles. The department operates from four regional headquarters and 45 area stations. There are four major bureaus in Central, Valley, South, and West, with a total of about 21 divisions within these bureaus. Our goal is to identify the most occurring crimes, high prone areas, and time. For the purpose of Data Analysis in this Project we will be using Tableau for Data Visualization and MS Excel for Data Cleaning.

MOTIVATION- The LAPD have been collecting crime data for decades and we have lots of data for the crimes happening in Los Angeles. But Crime Data is very difficult to access and understand. This analysis would help analyze and understand the data and trends better, so that people can make informed decisions about where to live and work. It will help citizens to get a better understanding of what is happening in their city and to help identify hot spots for crime and safety issues. This way, the residents can see how their neighborhood compares with other neighborhoods in LA when it comes to crime rates. This data analysis can also be used by many organizations like LA Times, UCLA, and various agencies to create visualizations, maps, and charts on different topics such as theft and shootings.

DATA DESCRIPTION

Dataset URL: https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8
It is a huge dataset with around 935K records and about 26 columns. This data is transcribed from original crime reports that are typed on paper and therefore there may be some inaccuracies within the data. Agencies across the country participate in the FBI Producing a monthly summary report on the UCR (Uniform Crime Reporting) program. Part I offenses include but are not limited to murder, rape, robbery, Exacerbated assault, trafficking-commercial sexual activity, trafficking-involuntary Bondage, robbery, car theft and Arson. Crimes such as Simple Assaults, Forgery/Counterfeiting, Embezzlement/Fraud, Receiving Stolen Property, Weapon Violations, Prostitution, Sex Crimes, Crimes Against Family/Child, Narcotic Drug Laws, Liquor Laws, Drunkenness, Disturbing the Peace, Disorderly Conduct, Gambling, DUI and Moving Traffic Violations are considered as Less Serious offenses and are referred as Part II crimes. The **UCR-COMPSTAT062618.pdf** in https://data.lacity.org/Public-Safety/Crime-Data-from-2010-to-2019/63jg-8b9z can be referred for getting Crime Codes. The different columns of the dataset consist of the values listed below:

Column Name	Description	Type
DR_NO	Division of Records Number: Official file number made up of a 2 digit year, area ID, and 5 digits	Plain Text
Date Rptd	Date Reported. MM/DD/YYYY Format	Date & Time
DATE OCC	Date of Occurrence .MM/DD/YYYY Format	Date & Time
TIME OCC	Time of Occurrence. In 24 hour military time.	Plain Text
AREA	The LAPD has 21 Community Police Stations referred to as Geographic Areas within the department. These Geographic Areas are sequentially numbered from 1-21.	Plain Text
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 77th Street Division is located at the intersection of South Broadway and 77th Street, serving neighborhoods in South Los Angeles.	Plain Text
Rpt Dist No	A four-digit code that represents a sub-area within a Geographic Area. All crime records reference the "RD" that it occurred in for statistical comparisons.	Plain Text
Part 1-2	1- Severe Crime 2- Less Severe Crime	Number
Crm Cd	Crime Code. Indicates the crime committed. (Same as Crime Code 1)	Plain Text
Crm Cd Desc	Crime Code Description. Defines the Crime Code provided.	Plain Text
Mocodes	Modus Operandi: Activities associated with the suspect in commission of the crime.	Plain Text
Vict Age	Victim Age. Two character numeric	Plain Text
Vict Sex	Victim Sex. F - Female M - Male X - Unknown	Plain Text
Vict Descent	Victim Descent Code: 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	Plain Text
Premis Cd	Premis Code. The type of structure, vehicle, or location where the crime took place.	Number
Premis Desc	Premis Description. Defines the Premise Code provided.	Plain Text
Weapon Used Cd	Wepon Used Code. The type of weapon used in the crime.	Plain Text
Weapon Desc	Wepon Description. Defines the Weapon Used Code provided.	Plain Text
Status	Status of the case. (IC is the default)	Plain Text
Status Desc	Defines the Status Code provided.	Plain Text
Crm Cd 1	Indicates the crime committed. Crime Code 1 is the primary and most serious one. Crime Code 2, 3, and 4 are respectively less serious offenses. Lower crime class numbers are more serious.	Plain Text
Crm Cd 2	May contain a code for an additional crime, less serious than Crime Code 1.	Plain Text
LOCATION	Street address of crime incident rounded to the nearest hundred block to maintain anonymity.	Plain Text
Cross Street	Cross Street of rounded Address	Plain Text
LAT	Latitude	Number
LON	Longtitude	Number

DATA CLEANING

The data which we get from the official sites is usually raw data. This data needs to be cleaned to make it easier to analyze data. We need to remove incorrect, duplicate, or incomplete data. Various data cleaning methods can be used for Data cleaning, these may vary from Dataset to dataset. Listed below are some of the Data Clearing methods which has been used to Clean the Crime Data set. The Data has been cleaned using Microsoft Excel.

1. **Illegal Values**

    Fixing Illegal values in a dataset is very important to Improve overall productivity. In out dataset, The **STATUS** field contains Illegal value as highlighted in the screenshot below. The Value After Cleaning, is replaced by 'AO' based on Status Description. AO stands for Adult Other.

    Before Cleaning:

R	S	T	
Weapon Desc	Status	Status Desc	Cr
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	??!	Adult Other	
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AO	Adult Other	
	AO	Adult Other	
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	IC	Invest Cont	
UNKNOWN FIREARM	IC	Invest Cont	
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AO	Adult Other	
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AO	Adult Other	
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AA	Adult Arrest	
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AO	Adult Other	
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AA	Adult Arrest	
	AO	Adult Other	
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AA	Adult Arrest	
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AO	Adult Other	
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AA	Adult Arrest	
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AO	Adult Other	
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	??!	Adult Other	
	IC	Invest Cont	
	IC	Invest Cont	
ROPE/LIGATURE	AA	Adult Arrest	

After Cleaning:

Weapon Desc	Status	Status Desc	Crm C
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AO	Adult Other	
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AO	Adult Other	
	AO	Adult Other	
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	IC	Invest Cont	
UNKNOWN FIREARM	IC	Invest Cont	
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AO	Adult Other	
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AO	Adult Other	
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AA	Adult Arrest	
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AO	Adult Other	
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AA	Adult Arrest	
	AO	Adult Other	
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AA	Adult Arrest	
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AO	Adult Other	
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AA	Adult Arrest	
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AO	Adult Other	
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AO	Adult Other	
	IC	Invest Cont	
	IC	Invest Cont	

2. **Missing Values**

Data cleaning can be handled in many ways, one such way is fixing missing values.

The Crm Cd (Crime Code) column values are missing. I have entered values based on

Crm Cd Description (Crime Code Description). For Example, in the Screenshot below.

As highlighted in Yellow the Crm Cd for '**Child Abuse' is 627** and Crm Cd for

'Violation of Court Order' is 900. The UCR-COMPSTAT062618.pdf in

https://data.lacity.org/Public-Safety/Crime-Data-from-2010-to-2019/63jg-8b9z can be

referred for getting Crime Codes.

Before Cleaning:

AREA NAME	Rpt Dist No	Part 1-2	Crm Cd	Crm Cd Desc	Mocod	Vict Ag	Vict Sex
Foothill	1641	2	624	BATTERY - SIMPLE ASSAULT	1822 0444	4	M
Mission	1961	2	930	CRIMINAL THREATS - NO WEAPON DISPLAYED	0421 2004	4	F
Southwest	393	1	230	ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT	0302 0334	4	M
Southeast	1863	2	627	CHILD ABUSE (PHYSICAL) - SIMPLE ASSAULT	0552 0913	4	M
Harbor	524	2	920	KIDNAPPING - GRAND ATTEMPT	1257 2004	4	M
Van Nuys	901	2	354	THEFT OF IDENTITY	1822 0917	4	M
Rampart	256	2		CHILD ABUSE (PHYSICAL) - SIMPLE ASSAULT	1257 1251	4	M
Harbor	551	2	627	CHILD ABUSE (PHYSICAL) - SIMPLE ASSAULT	0913 0602	4	M
Mission	1981	1	815	SEXUAL PENETRATION W/FOREIGN OBJECT	1258 0360	4	F
Southwest	326	2	237	CHILD NEGLECT (SEE 300 W.I.C.)	0360 0553	4	F
Southwest	392	2	237	CHILD NEGLECT (SEE 300 W.I.C.)	1257 0554	4	F
Southeast	1864	2	930	CRIMINAL THREATS - NO WEAPON DISPLAYED	0913 0552	4	F
Southwest	354	2	813	CHILD ANNOYING (17YRS & UNDER)	1251 1257	4	M
Olympic	2029	2		VIOLATION OF COURT ORDER	1242 1257	4	F
Rampart	221	2	812	CRM AGNST CHLD (13 OR UNDER) (14-15 & SUSP 10 YRS OLDER)	1822 1257	4	F
N Hollywood	1585	2	627	CHILD ABUSE (PHYSICAL) - SIMPLE ASSAULT	0400 0416	4	F
Mission	1961	2	812	CRM AGNST CHLD (13 OR UNDER) (14-15 & SUSP 10 YRS OLDER)	1402 1257	4	F

After Cleaning:

	AREA NAME	Rpt Dist No	Part 1-2	Crm Cd	Crm Cd Desc	Mocod	Vict Ag	Vict Sex	V
342	Foothill	1641	2	624	BATTERY - SIMPLE ASSAULT	1822 0444	4	M	C
343	Mission	1961	2	930	CRIMINAL THREATS - NO WEAPON DISPLAYED	0421 2004	4	F	H
344	Southwest	393	1	230	ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT	0302 0334	4	M	B
345	Southeast	1863	2	627	CHILD ABUSE (PHYSICAL) - SIMPLE ASSAULT	0552 0913	4	M	B
346	Harbor	524	2	920	KIDNAPPING - GRAND ATTEMPT	1257 2004	4	M	H
347	Van Nuys	901	2	354	THEFT OF IDENTITY	1822 0917	4	M	H
348	Rampart	256	2	627	CHILD ABUSE (PHYSICAL) - SIMPLE ASSAULT	1257 1251	4	M	H
349	Harbor	551	2	627	CHILD ABUSE (PHYSICAL) - SIMPLE ASSAULT	0913 0602	4	M	H
350	Mission	1981	1	815	SEXUAL PENETRATION W/FOREIGN OBJECT	1258 0360	4	F	H
351	Southwest	326	2	237	CHILD NEGLECT (SEE 300 W.I.C.)	0360 0553	4	F	B
352	Southwest	392	2	237	CHILD NEGLECT (SEE 300 W.I.C.)	1257 0554	4	F	C
353	Southeast	1864	2	930	CRIMINAL THREATS - NO WEAPON DISPLAYED	0913 0552	4	F	B
354	Southwest	354	2	813	CHILD ANNOYING (17YRS & UNDER)	1251 1257	4	M	B
355	Olympic	2029	2	900	VIOLATION OF COURT ORDER	1242 1257	4	F	C
356	Rampart	221	2	812	CRM AGNST CHLD (13 OR UNDER) (14-15 & SUSP 10 YRS OLDER)	1822 1257	4	F	H
357	N Hollywood	1585	2	627	CHILD ABUSE (PHYSICAL) - SIMPLE ASSAULT	0400 0416	4	F	H
358	Mission	1961	2	812	CRM AGNST CHLD (13 OR UNDER) (14-15 & SUSP 10 YRS OLDER)	1402 1257	4	F	H
359	Mission	1918	2	812	CRM AGNST CHLD (13 OR UNDER) (14-15 & SUSP 10 YRS OLDER)	0913 1817	4	F	H
360	77th Street	1213	2	812	CRM AGNST CHLD (13 OR UNDER) (14-15 & SUSP 10 YRS OLDER)	0400 1257	4	F	H

3. Misspellings

It is very important to fix the Misspellings in a data set, as this may cause ambiguity. As shown in the screenshot below the 'Area Name'- Harbor is spelled Incorrectly. Hence replaced the incorrect value of **Harbor** to Correct Value, for the whole Column- 'Area Name'

Before Cleaning:

DR_NO	Date Rptd	Date Occ	TIME OCC	AREA	AREA NAME	Rpt Dist No	Part 1-2	Crm Cd
200512216	7/23/2020	5/1/2020	1	5	Harbour	589	1	820
200514757	10/1/2020	10/1/2020	2130	5	Harbour	532	1	230
200915549	10/6/2020	10/3/2020	2000	9	Van Nuys	923	1	815
201415102	8/7/2020	5/3/2020	1900	14	Pacific	1409	1	820
201223873	10/31/2020	10/31/2020	1600	12	77th Street	1243	1	230
201226944	12/22/2020	12/8/2020	1200	12	77th Street	1249	1	310
201315268	8/10/2020	8/10/2020	1130	13	Newton	1351	1	420
201800852	12/17/2020	11/13/2020	1200	18	Southeast	1831	1	235
201710758	7/4/2020	7/4/2020	2230	17	Devonshire	1751	1	230
201207539	2/20/2020	2/19/2020	1030	12	77th Street	1267	1	815

After Cleaning:

DR_NO	Date Rptd	Date Occ	TIME OCC	AREA	AREA NAME	Rpt Dist No	Part 1-2
200512216	7/23/2020	5/1/2020	1	5	Harbor	589	1
200514757	10/1/2020	10/1/2020	2130	5	Harbor	532	1
200915549	10/6/2020	10/3/2020	2000	9	Van Nuys	923	1
201415102	8/7/2020	5/3/2020	1900	14	Pacific	1409	1
201223873	10/31/2020	10/31/2020	1600	12	77th Street	1243	1
201226944	12/22/2020	12/8/2020	1200	12	77th Street	1249	1
201315268	8/10/2020	8/10/2020	1130	13	Newton	1351	1
201800852	12/17/2020	11/13/2020	1200	18	Southeast	1831	1

4. Duplicated Records

Duplicate records cause confusion., It also causes unnecessary wastage of memory, Hence the Duplicate rows and columns should be removed in a data set. The values under Column '**Crm Cd' and 'Crm Cd 1' are the same**. Hence, this is the case of duplicate columns. For removing redundancy, I would be removing 'Crm Cd' Column. Crm Cd stands for Crime Code.

Before Cleaning:

Part 1-2	Crm Cd	Crm Cd Desc	Vict Age	Vict Sex	Vict Descent	Status Desc	Crm Cd 1	Crm Cd 2	Crm Cd 3	LOCATION
1	815	SEXUAL PENETRATION W/FOREIGN OBJECT	3	F	H	Adult Other	815	998		14800 ARCHWOOD ST
1	820	ORAL COPULATION	3	M	O	Adult Other	812	820		9700 REGENT ST
1	230	ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT	3	F	B	Invest Cont	230			DENKER AV
1	310	BURGLARY	3	F	B	Invest Cont	310			6200 S HOOVER ST
1	420	THEFT FROM MOTOR VEHICLE - PETTY ($950 & UNDER)	3	M	H	Invest Cont	420			4200 S BROADWAY
1	235	CHILD ABUSE (PHYSICAL) - AGGRAVATED ASSAULT	3	M	H	Invest Cont	235			500 W 120TH ST
1	230	ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT	3	X	X	Invest Cont	230			10200 FARRALONE AV
1	815	SEXUAL PENETRATION W/FOREIGN OBJECT	3	F	B	Adult Other	812	815	998	8300 S VERMONT AV
1	820	ORAL COPULATION	3	F	H	Adult Other	812	820		3100 DREW ST
1	815	SEXUAL PENETRATION W/FOREIGN OBJECT	3	M	B	Adult Other	812	815		23300 BASSETT ST
1	815	SEXUAL PENETRATION W/FOREIGN OBJECT	3	F	H	Adult Other	812	815	998	2200 E IMPERIAL HY
1	815	SEXUAL PENETRATION W/FOREIGN OBJECT	3	F	H	Juv Arrest	812	815	998	600 W 113TH ST
1	341	THEFT-GRAND ($950.01 & OVER)EXCPT,GUNS,FOWL,LIVESTK,PROD	3	F	W	Invest Cont	341			3300 MANNING AV
1	820	ORAL COPULATION	3	F	W	Invest Cont	812	820		12600 MATTESON AV
1	230	ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT	3	X	X	Invest Cont	230			3400 DREW ST

After Cleaning:

Part 1-2	Crm Cd Desc	Vict Age	Vict Sex	Vict Descent	Status Desc	Crm Cd 1	Crm Cd 2	Crm Cd 3	LOCATION
1	SEXUAL PENETRATION W/FOREIGN OBJECT	3	F	H	Adult Other	815	998		14800 ARCHWOOD ST
1	ORAL COPULATION	3	M	O	Adult Other	812	820		9700 REGENT ST
1	ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT	3	F	B	Invest Cont	230			DENKER AV
1	BURGLARY	3	F	B	Invest Cont	310			6200 S HOOVER ST
1	THEFT FROM MOTOR VEHICLE - PETTY ($950 & UNDER)	3	M	H	Invest Cont	420			4200 S BROADWAY
1	CHILD ABUSE (PHYSICAL) - AGGRAVATED ASSAULT	3	M	H	Invest Cont	235			500 W 120TH ST
1	ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT	3	X	X	Invest Cont	230			10200 FARRALONE AV
1	SEXUAL PENETRATION W/FOREIGN OBJECT	3	F	B	Adult Other	812	815	998	8300 S VERMONT AV
1	ORAL COPULATION	3	F	H	Adult Other	812	820		3100 DREW ST
1	SEXUAL PENETRATION W/FOREIGN OBJECT	3	M	B	Adult Other	812	815		23300 BASSETT ST
1	SEXUAL PENETRATION W/FOREIGN OBJECT	3	F	H	Adult Other	812	815	998	2200 E IMPERIAL HY
1	SEXUAL PENETRATION W/FOREIGN OBJECT	3	F	H	Juv Arrest	812	815	998	600 W 113TH ST
1	THEFT-GRAND ($950.01 & OVER)EXCPT,GUNS,FOWL,LIVESTK,PROD	3	F	W	Invest Cont	341			3300 MANNING AV
1	ORAL COPULATION	3	F	W	Invest Cont	812	820		12600 MATTESON AV
1	ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT	3	X	X	Invest Cont	230			3400 DREW ST
1	ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT	3	F	B	Adult Other	230			ARLINGTON
1	ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT	3	M	H	Invest Cont	230			MARTIN LUTHER KING JR BL

5. Inconsistent Date Format

As we can see from the Screenshot below, The Date formats of the Highlighted columns 'Date Rptd' i.e. Date Reported and 'Date Occ' i.e. the date on which the incident occurred , are not the same . Hence, I will be changing the Date format for 'Date Rptd' Column to match that of 'Date Occ' Column.

Before Cleaning:

DR_NO	Date Rptd	Date Occ	TIME OCC	AREA	AREA NAME	Rpt Dist No
200512216	7/23/2020 0:00	5/1/2020	1	5	Harbor	589
200514757	10/1/2020 0:00	10/1/2020	2130	5	Harbor	532
200915549	10/6/2020 0:00	10/3/2020	2000	9	Van Nuys	923
201415102	8/7/2020 0:00	5/3/2020	1900	14	Pacific	1409
201223873	10/31/2020 0:00	10/31/2020	1600	12	77th Street	1243
201226944	12/22/2020 0:00	12/8/2020	1200	12	77th Street	1249
201315268	8/10/2020 0:00	8/10/2020	1130	13	Newton	1351
201800852	12/17/2020 0:00	11/13/2020	1200	18	Southeast	1831
201710758	7/4/2020 0:00	7/4/2020	2230	17	Devonshire	1751
201207539	2/20/2020 0:00	2/19/2020	1030	12	77th Street	1267

After Cleaning:

DR_NO	Date Rptd	Date Occ	TIME OCC	AREA	AREA NAME	Rpt Dist No
200512216	7/23/2020	5/1/2020	1	5	Harbor	589
200514757	10/1/2020	10/1/2020	2130	5	Harbor	532
200915549	10/6/2020	10/3/2020	2000	9	Van Nuys	923
201415102	8/7/2020	5/3/2020	1900	14	Pacific	1409
201223873	10/31/2020	10/31/2020	1600	12	77th Street	1243
201226944	12/22/2020	12/8/2020	1200	12	77th Street	1249
201315268	8/10/2020	8/10/2020	1130	13	Newton	1351
201800852	12/17/2020	11/13/2020	1200	18	Southeast	1831
201710758	7/4/2020	7/4/2020	2230	17	Devonshire	1751

DATA VISUALIZATION

We have created few Visualizations in the form of graphs and charts. These visualizations help us identify which crimes occur the most, People of which ethnicity are more vulnerable to crimes in which area, when does the Crime occur the most and so on. Graph Visualization is a vital tool for exploring and understanding crime data at scale. For the purpose of Data Visualization, we have used Tableau in this Project. Two screenshots are attached to each Visualization. One screenshot shows the graph and other screenshot gives us the details about the graph, such as what dimensions and measures are being used to create that graph.

1. **Which Crime has Occurred the Most?**
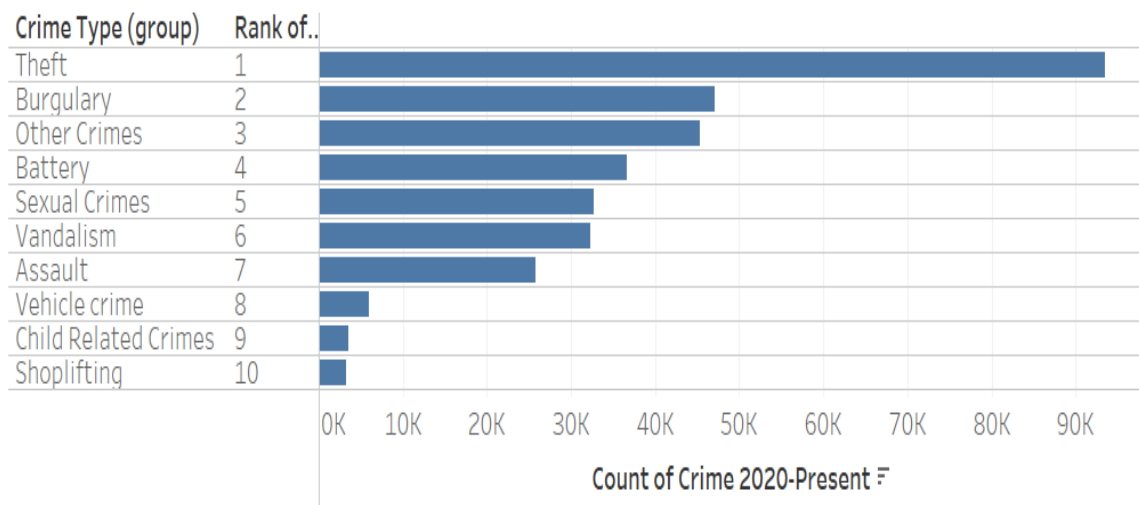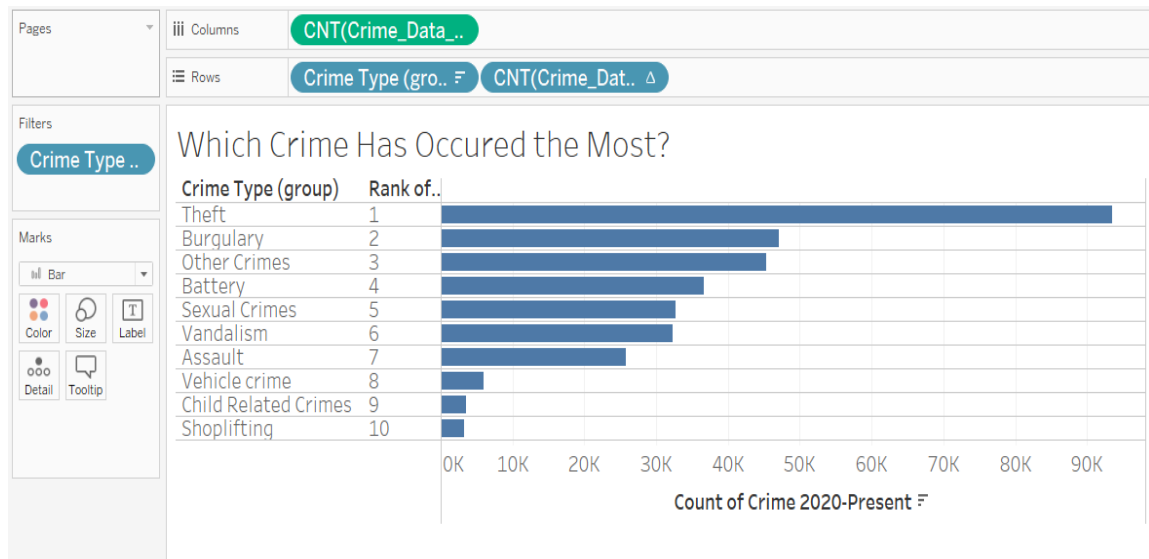
Application used: Groups, Rank, Filter

This analysis shows the crime breakdown structure. It helps us analyze which crime has occurred the greatest number of times and the Crimes have been ranked High to Low based on there occurrence. A filter is being used on Crime Type, to display Top 10 Crime Types. This analysis shows that the probability of Theft is the most, hence the residents of LA are more prone to theft and should be aware of the surroundings.

Also, in this analysis, various types of similar crimes have been categorized under one Heading. This is done using the one of the functionalities of Tableau which is called Groups. The example of how the groups is being formed is shown below.

Edit Group [Crime Type (group)]

Field Name: Crime Type (group)

Groups: Add to:

- ∨ Assault
  - ARSON
  - ASSAULT WITH DEADLY WEAPON ON POLICE OFFICER
  - ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT
- › Theft
- ∨ Battery
  - BATTERY - SIMPLE ASSAULT
  - BATTERY ON A FIREFIGHTER
  - BATTERY POLICE (SIMPLE)
  - BATTERY WITH SEXUAL CONTACT
- › Other Crimes
- ∨ Vehicle crime
  - BIKE - ATTEMPTED STOLEN
  - BIKE - STOLEN
  - RECKLESS DRIVING
  - VEHICLE - ATTEMPT STOLEN
  - VEHICLE - MOTORIZED SCOOTERS, BICYCLES, AND WHEELCHAIRS
  - VEHICLE - STOLEN
- ∨ Burgulary
  - BURGLARY
  - BURGLARY FROM VEHICLE
  - BURGLARY, ATTEMPTED
- › Child Related Crimes
- ∨ Drugs
  - DRUGS, TO A MINOR
  - DRUNK ROLL
- › Sexual Crimes
- › Shoplifting
- › Vandalism

2. Which ethnicity has been more vulnerable to Crime in LA?

Which ethnicity has been more vulnarable to crime in LA?

Victim Descent
- American Indian
- Asian Indian
- Black
- Cambodian
- Chinese
- Filipino
- Guamanian
- Hawaiian
- Hispanic/Latin?Mexican
- Japanese
- Korean
- Laotian
- Other
- Other Asian
- Pacific Islander
- Samoan
- Vietnamese
- White

© 2022 Mapbox © OpenStreetMap
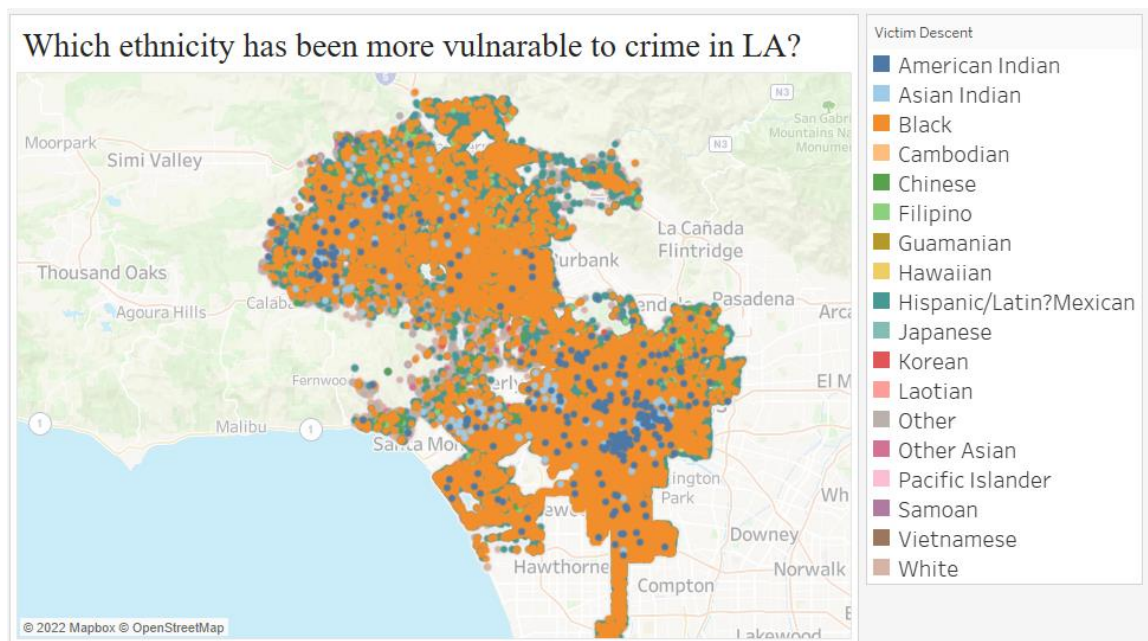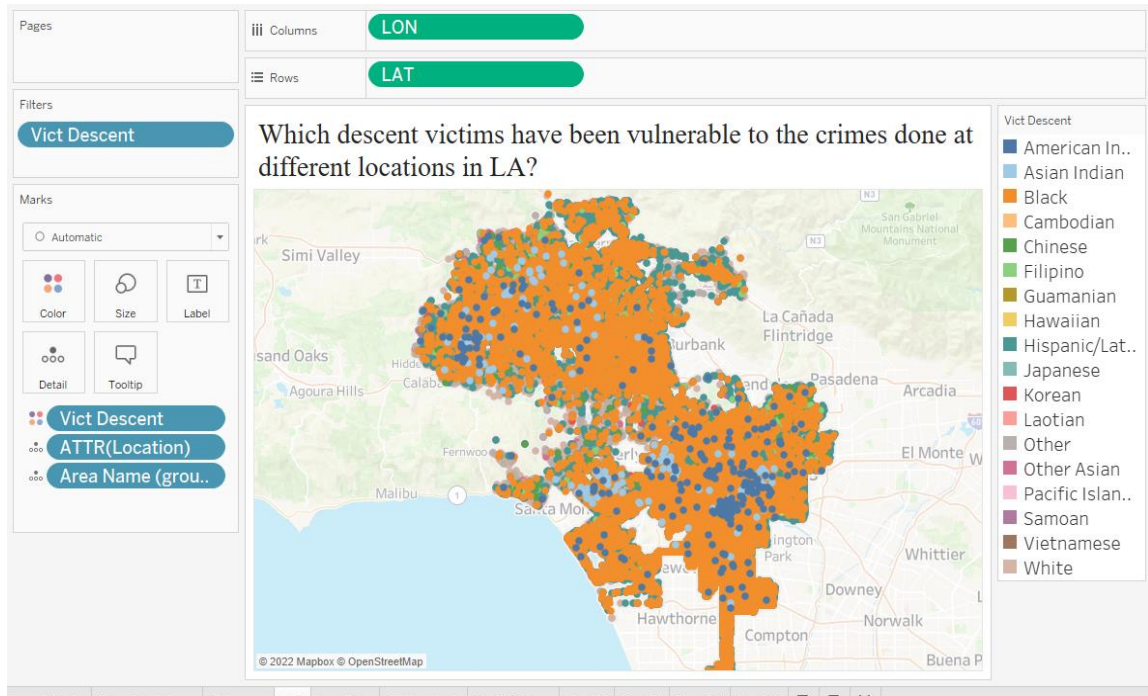
Map 1
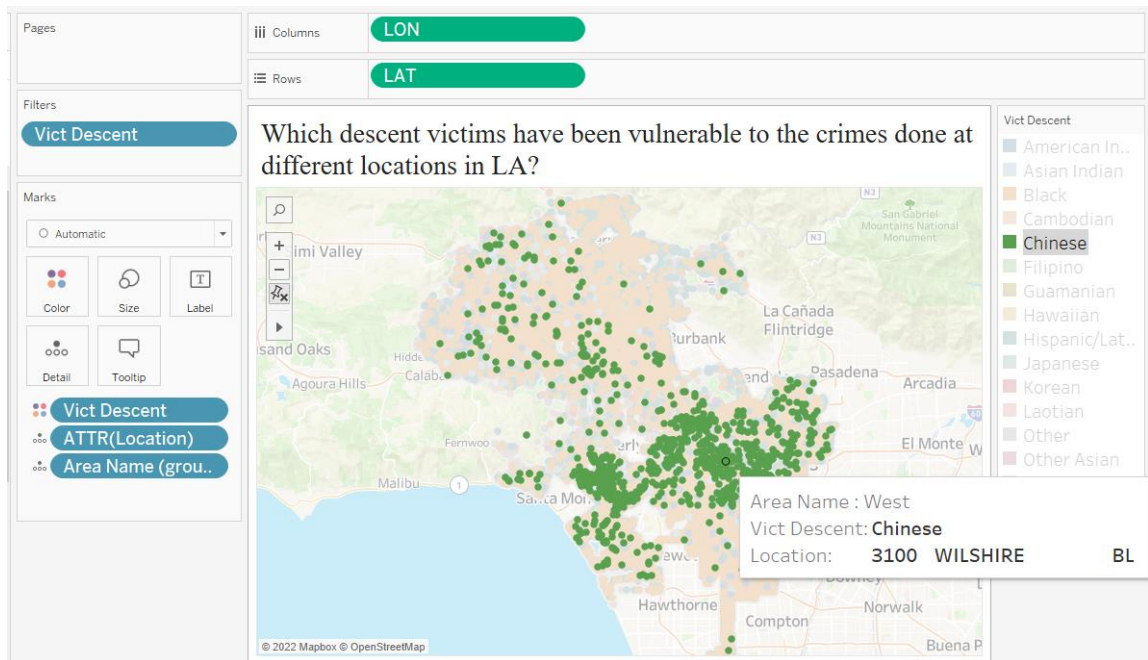
Map 2

Analysis Used: Geographic Maps, Alias, Groups

This is the map analysis of ethnicity of the Victims that are more vulnerable to crime in a particular area. For the purpose of this Visualization 21 area names have been grouped under the 4 bureaus i.e., Central, West, Valley South which were mentioned above. This was done using Group option in Tableau. For this analysis we can Click on the 'Victim Descent' category for which the analysis needs to be made from the Righthand Legend. For Example: In Map 2 , We can see that people of 'Chinese' descent are more vulnerable to the Crimes in West Area when Compared to 'North', 'South' or 'Valley'. Based on this analysis the Chinese people can make a comparison among different areas and can choose their place of residence based on the occurrence of crime. Also in the first look at Map 1 we can see that the map appears to be orange, which means that the percentage of Black Victims out of the total Victims is the most. Certain measures should be taken by the LAPD and the City to stop the crime against People of Color.

Field Name: Area Name (group)

Groups:

- ⌄ 📎 South
  - 77th Street
  - Harbor
  - Southeast
  - Southwest
- ⌄ 📎 Central
  - Central
  - Hollenbeck
  - Newton
  - Northeast
  - Rampart
- ⌄ 📎 Valley
  - Devonshire
  - Foothill
  - Hollywood
  - Mission
  - N Hollywood
  - Topanga
  - Van Nuys
  - West Valley
- ⌄ 📎 West
  - Olympic
  - Pacific
  - West LA
  - Wilshire

Group	Rename	Ungroup

3. Victims of which age group are being targeted the most in LA ?
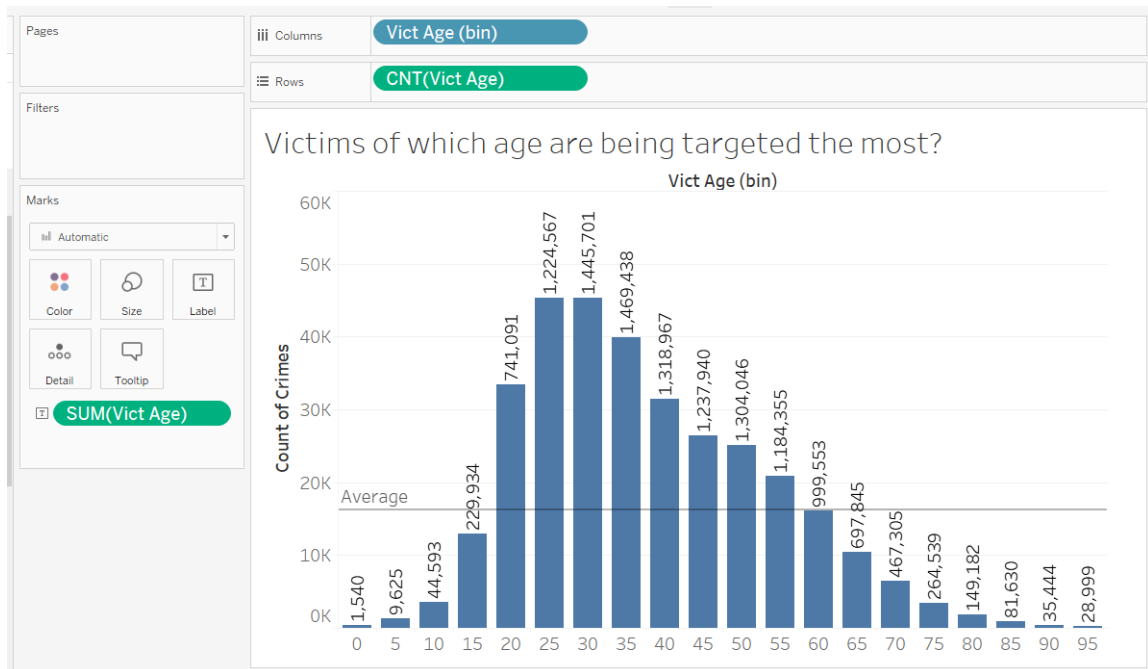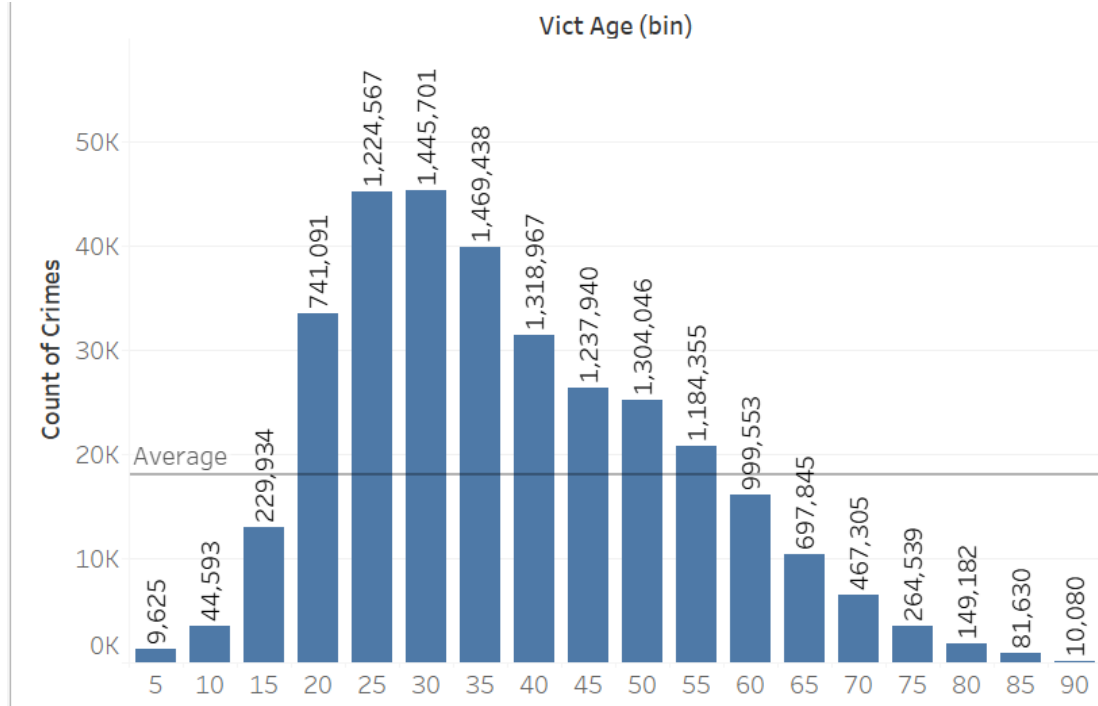
Analysis Used- Histogram, Reference Line

Using the graph, we can analyze Victims of which age are more vulnerable to crimes. For the purpose of analysis, a Histogram is being used. Histogram is similar to bar chart, the only difference being that it groups the values (Age in this case) into continuous ranges (0-5,5-10 etc.). Each bar in histogram represents the height of the number of values present in the range. It is quite evident from the graph that People with Average Age of 25-30 are more prone to Crime. And the Victims that are least prone to crime fall in age group 0-10 and 85-95. While creating histograms, we have age bin size of 5. The Average Count of Crimes is shown by a Referenced Line and is 16301. Above each bar also contains the total number of Victims that fall under a particular age group. Such Visualization does definitely make data analysis easier.

4. **What is the trend of Arrests and Unarrests occurred during last two years in LA?**

Count of Arrests Vs Count of Unarrests
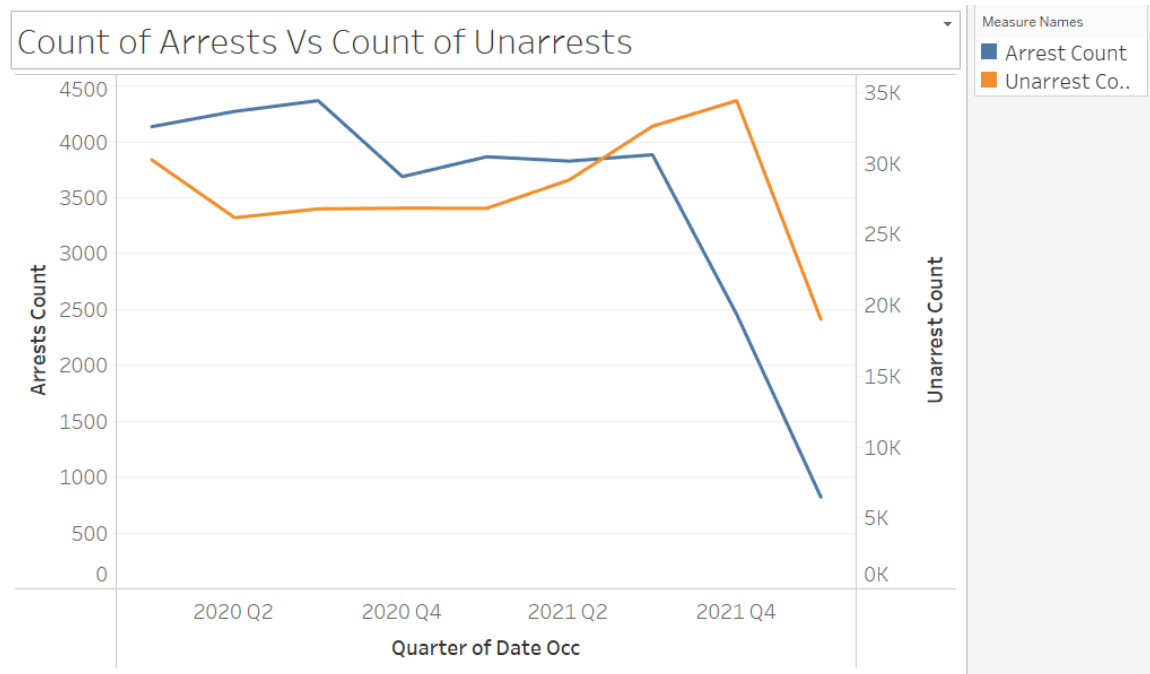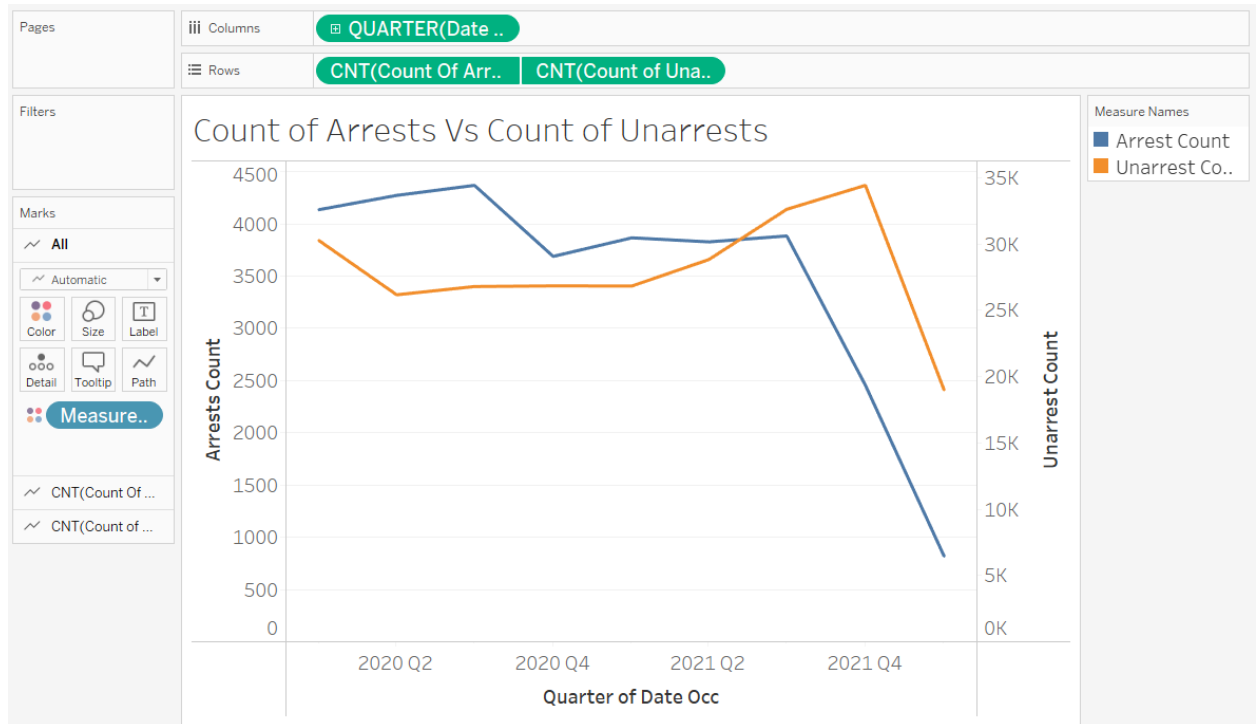
Analysis Used- Calculated Fields, Dual Chart.

This analysis shows the trends for number of Criminals Arrested vs the Number of Criminals Not Arrested. This Visualization is done using Dual axis chart. The left y-axis shows the Count of Criminals Arrested and the right y-axis shows the Count of Criminals that are not arrested. For the purpose of this visualization, we had created a Calculated field to get the Count of Criminals Arrested and the Count of Criminals that are not arrested on the basis of 'Status' Field. As it is clear from the Visualization the Number of Criminals that are not arrested is much higher than the number of Criminals that are arrested. Also, it is quite evident from the graph that the 'Arrest Count' decreases significantly. I believe the LAPD needs to make sure that proper Investigation should be done, and serious action is being taken to arrest the Criminals.

5. On which day does the crime occur the most?

On which day does the crime occur the most

On which day does the crime occur the most

Upper Whisker:	23,556
Upper Hinge:	22,878.5
Median:	22,201
Lower Hinge:	12,533
Lower Whisker:	2,865
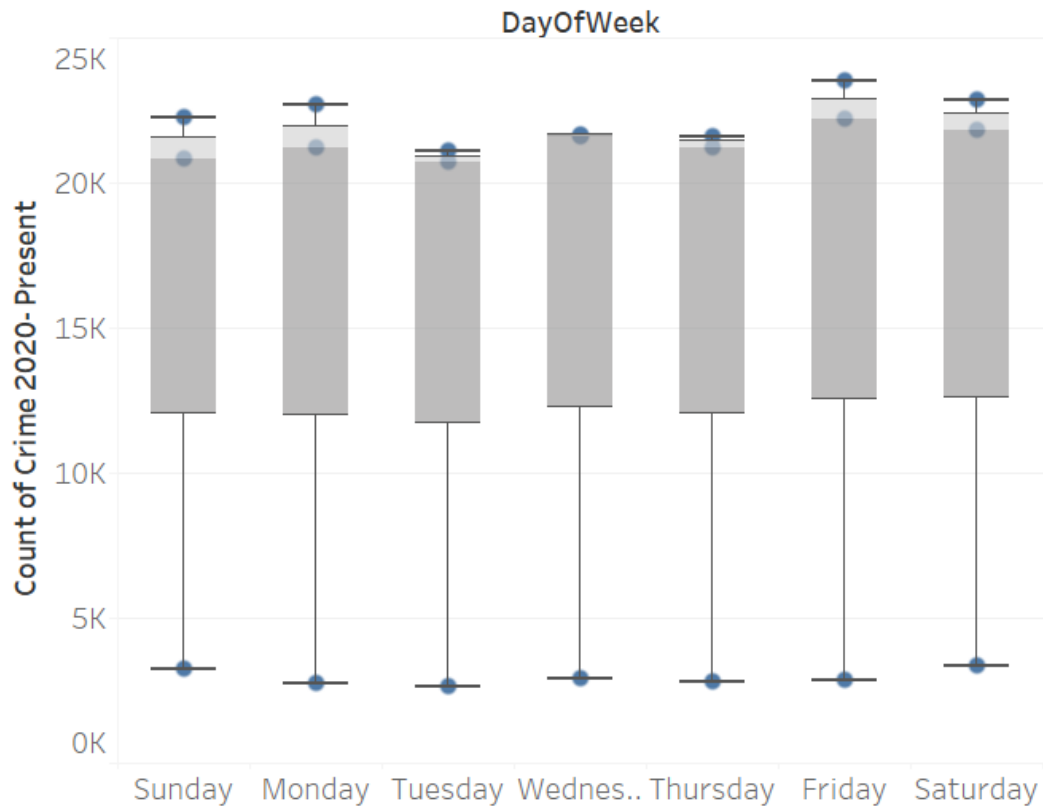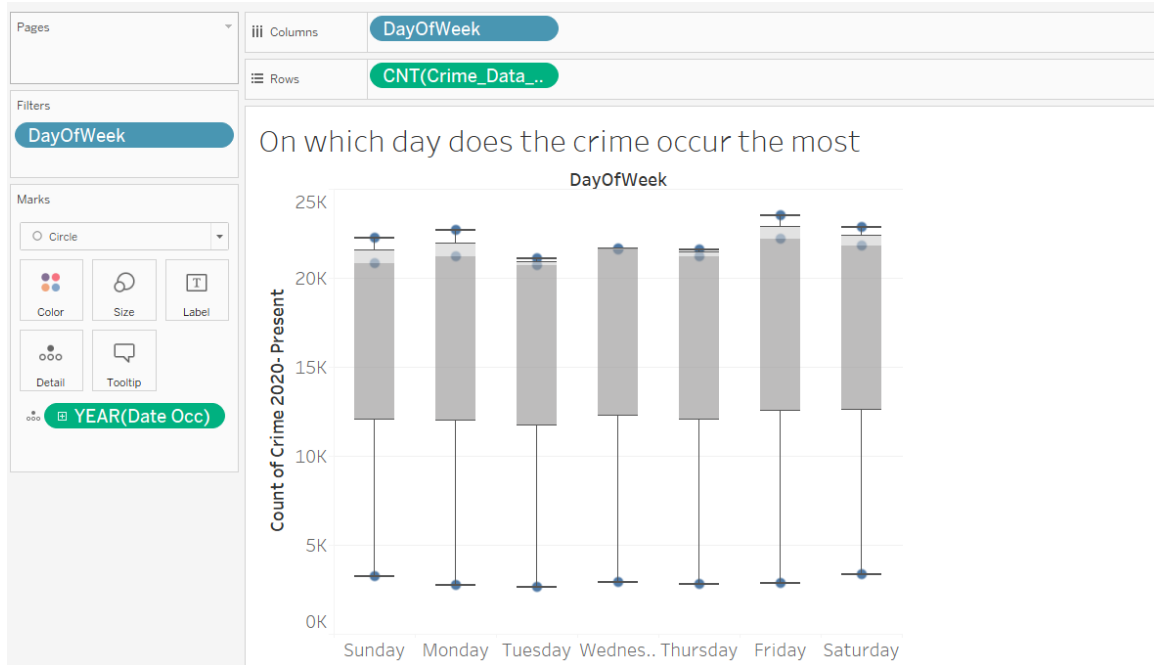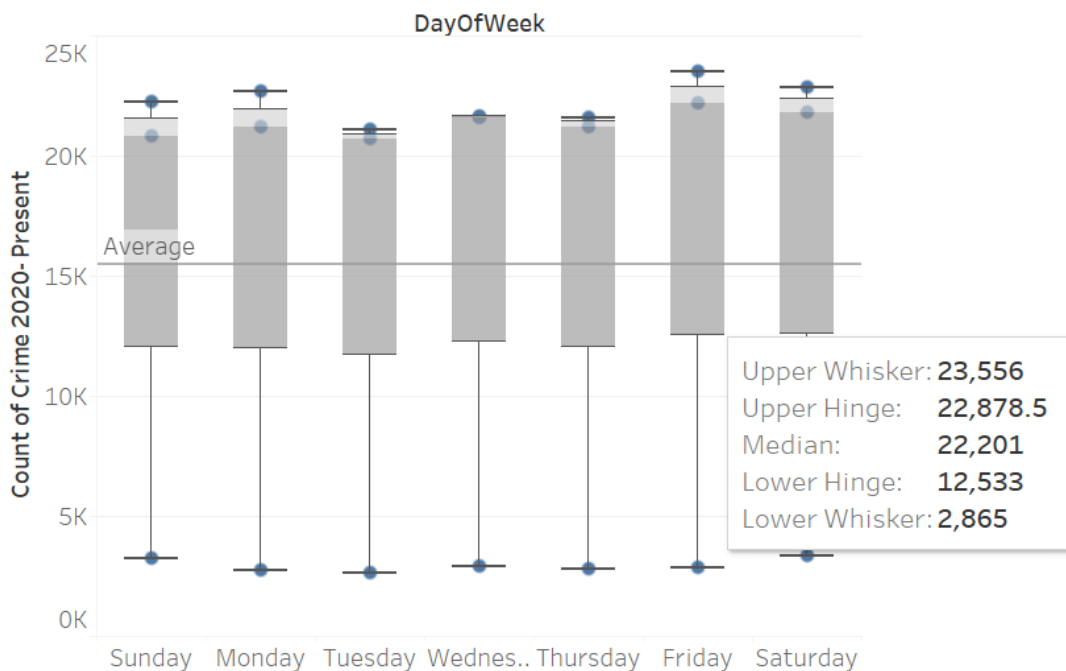
Analysis used: Calculated Fields, Box and Whisker, Reference Line

This visualization depicts the number of crimes based on different days of the week . The

day on which a crime has been occurred, is calculated on the basis of Date of Occurrence

which was provided to us in the data set. It is clear from the Box and Whisker Plot that the maximum number of crimes take place on Friday and the least crime occurs on Tuesday. One should be more vigilant during those times. Box and Whisker Plot shows the distribution of the values along the axis. The boxes indicate the middle 50 percent of the data which is, the middle two quartiles of the Crime data's distribution. We can identify Median, Maximum, Minimum and quartile ranges using box and whisker plot. For example, In the figure above the Median for Count of Crimes on Friday is 22, 201 with 23,556 being the highest no of crimes recorded in the year 2021 and lowest being 2,865 for the year 2022(first 3 months). The reference line shows the average of total crimes. The average number of crimes occurred each day is about 15,524.

DASHBOARD

A dashboard is a special feature of Tableau which is a combination of several views that you can arrange for presentation or to monitor. It helps you compare a variety of data simultaneously. Creating a dashboard would help you display all the views at once, rather that navigating to separate worksheets. Analyzing and comparing different worksheets at once helps create better insights. The dashboard prepared is shown below:
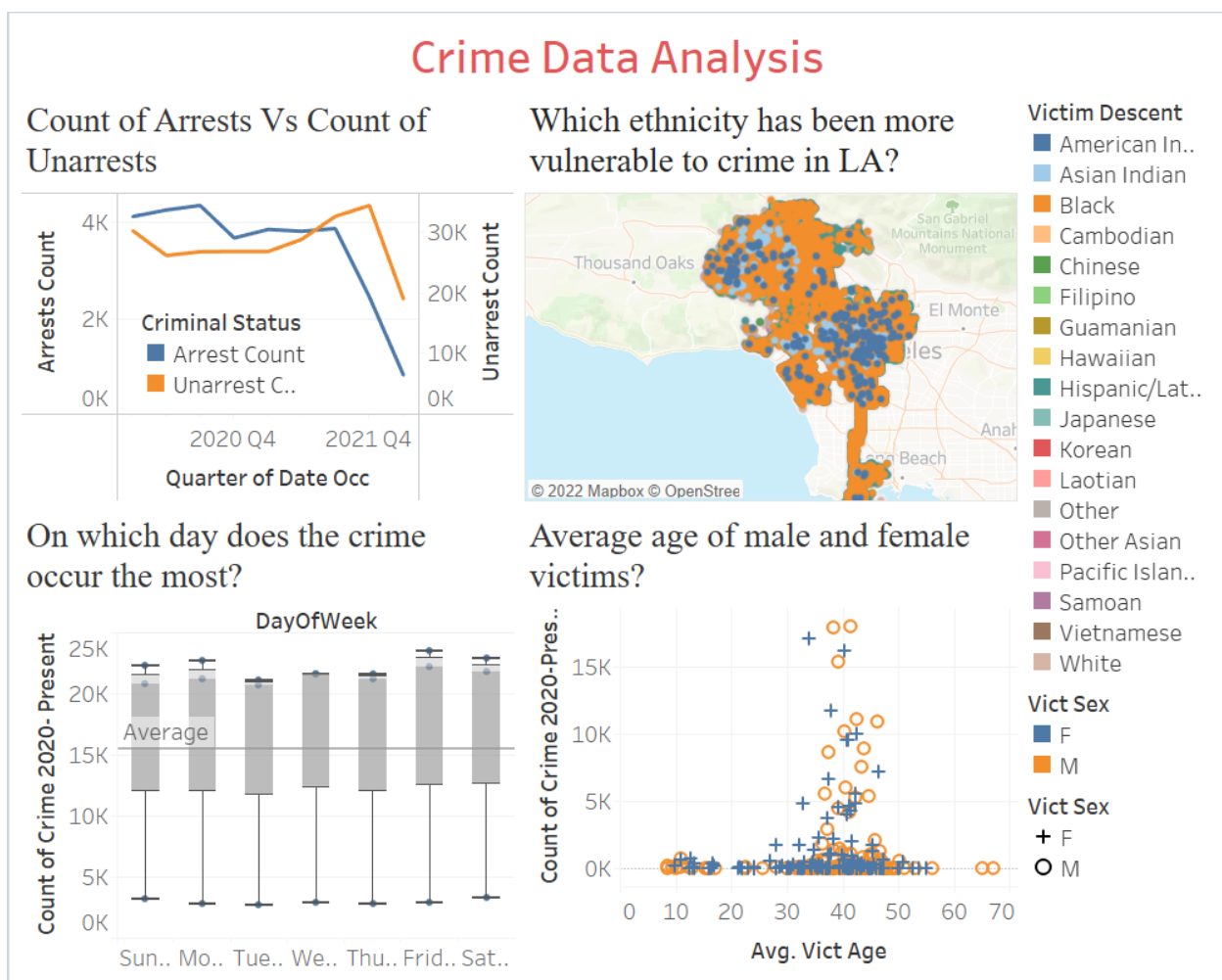
The dashboard created is interactive, this will enable a user to visualize the data , filter on demand and simply click to dig deeper into the underlying data. The insights are not only fast but

its also fun to play around with the interactive dashboard. Let us now dig deep into each of the Visualizations present in the Dashboard and understand its purpose.

1. **Status of Criminals for past 2 years**

   Analysis Used: Dual Axis, Calculated Field

   This Dual Axis Chart shows the number of Criminals Arrested Versus The number of Criminals which are not arrested. We can understand from the graph that the Arrest Count has significantly decreased over the years. Also, on analyzing this graph we can also observe that the number of Un-arrested Criminals are way more than those of Arrested ones.

2. **On which Day Does the crime Occur the most?**

   Analysis Used: Box-and Whisker, Reference Line

   This Box – and -whisker plot gives us the insight of how many Crimes occur on a particular day. Using this analysis residents of LA can be more vigilant during the days in which the probability of crimes is the most. There might be a possibility that the criminals follow a particular pattern. We can see from the plot that the Number of Crimes is more during the weekend and days before and after it. The average number of crimes seems to be low during the Weekdays i.e. Tuesday, Wednesday, Thursday.

3. **Which ethnicity is more Vulnerable to Crime in LA?**

   Analysis Used: Maps , Filters

   Tableau has a unique and aesthetic feature to Visualize, that is using Geographical Maps. Using this analysis, you can during this analysis you can identify people of which ethnicity are more vulnerable to crimes at which particular area in LA.

4. **Average Age of male and Female Victims?**
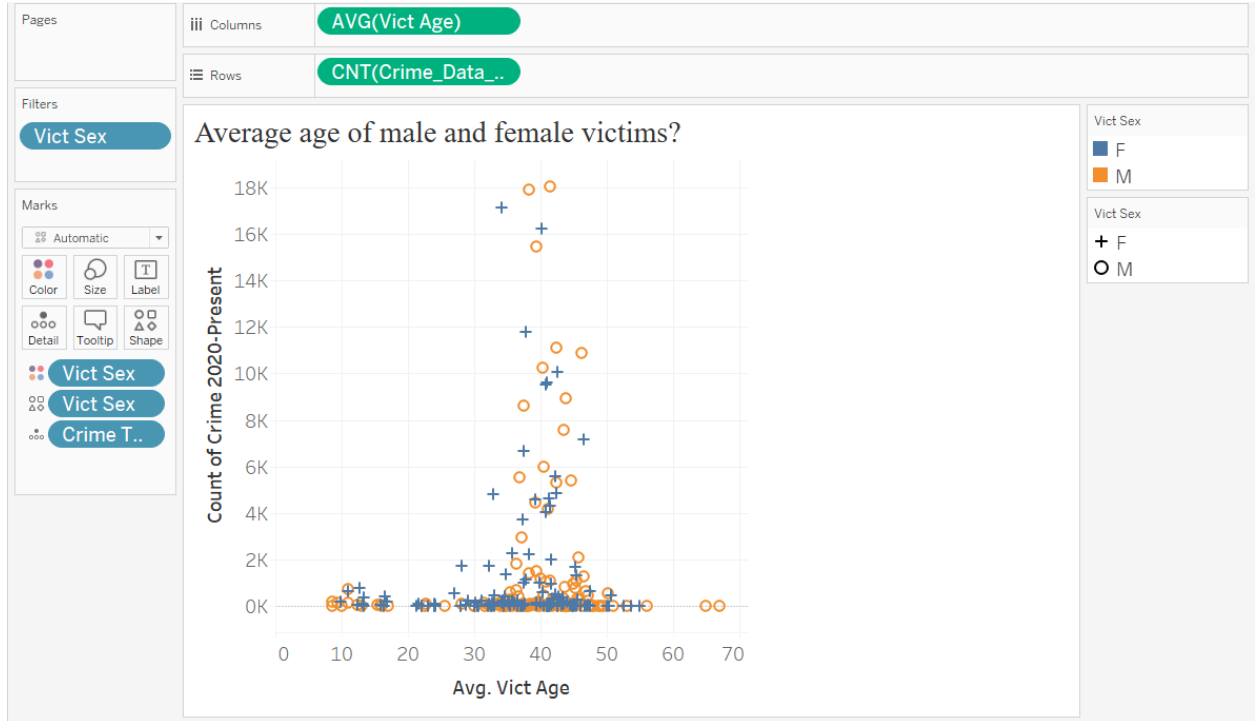
Analysis Used: Scatter Plot, Filters, Color

This visualization shows the dependency of two variables i.e., the number of crimes and the average victim age. This scatterplot shows the average age of males and female victims that are being targeted by the criminals in LA. In this graph the blue represents the number of females, and the orange color represents the number of males. As we can see from the graph that the number of females and the number of males is more in the age group of 30 to 45. Using this analysis, we can also conclude that the number of females and males are more prone to assaults when compared to crimes like theft and robbery. This visualization helps us helps us analyze people of what age group are prone to what type of crimes and the people can be more vigilant.

STORY TELLING

This LAPD (Los Angeles Police Department) crime dataset consist of all the crimes which occurred from year 2020 to present. A data story consisting of a series of Visualizations analyzing Crime trends in LA City with the aim of providing LAPD with valuable and practical insights to promote safer areas. The LA has its own fight against criminal activity, but the ability to analyze criminal and victim data helps the LAPD unleash its greatest asset in the fight for justice. One of the unique features of Tableau is it storytelling feature. In tableau, a story is a sequence of visualizations that work together to convey information we can create stories to tell a data narrative, provide context, demonstrate how decisions relate to outcomes or to simply make a compelling case. Some of the interesting analysis made on this data are discussed below with supporting images from Tableau story.

Be aware of Thieves!
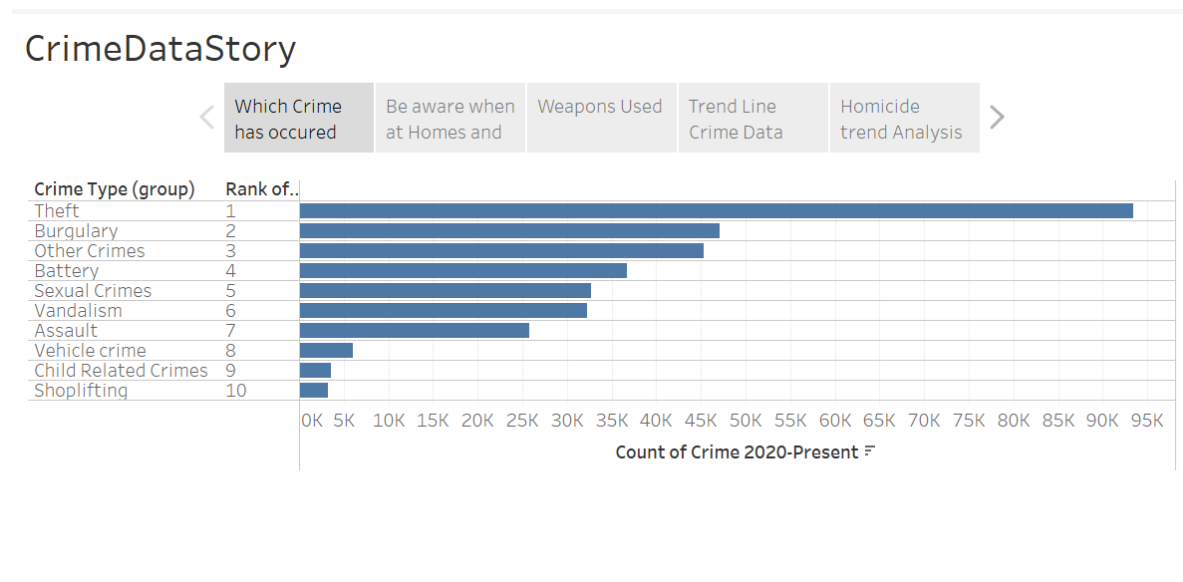
The above **bar graph** shows the **rank** of all the crime type groups. According to the visualization most of the crimes add around theft odd burglary end probability of shoplifting and children related crimes are the least this analysis helps us identify that if we plan to reside in LA, which crime has the maximum probability of occurrence. This analysis gives us an idea about the most prevalent Crime type in the Los Angeles. All the Crime Types have been Ranked based on its Occurrence. The article by Peter Kiefer published in LA magazine "How Bad is Crime in L.A.?" also talks about the increasing cases of Theft and Robbery in LA. *(Losangeles magazine)* It emphasis on the fact that how scared the people of LA are , the residents of LA are scared to leave their homes, due to the fear of being followed. According to the article , one of the residents of LA says that she makes sure that she is not being followed by checking the rear view mirror again and again and she also complaints that she feels danger staying in LA and does not leave her house with jewelry or nice bags. There is paranoia among the commons. The paranoia is not confined to only the privileged, but gang violence has also been increased among the masses. Let us dig deeper and analyze the trends for two of the Crime Types- Homicide and Vehicle Theft. These two types of Crimes are particularly chosen for analysis because they had been in news lately.

Increase in Homicide and Vehicle Theft.
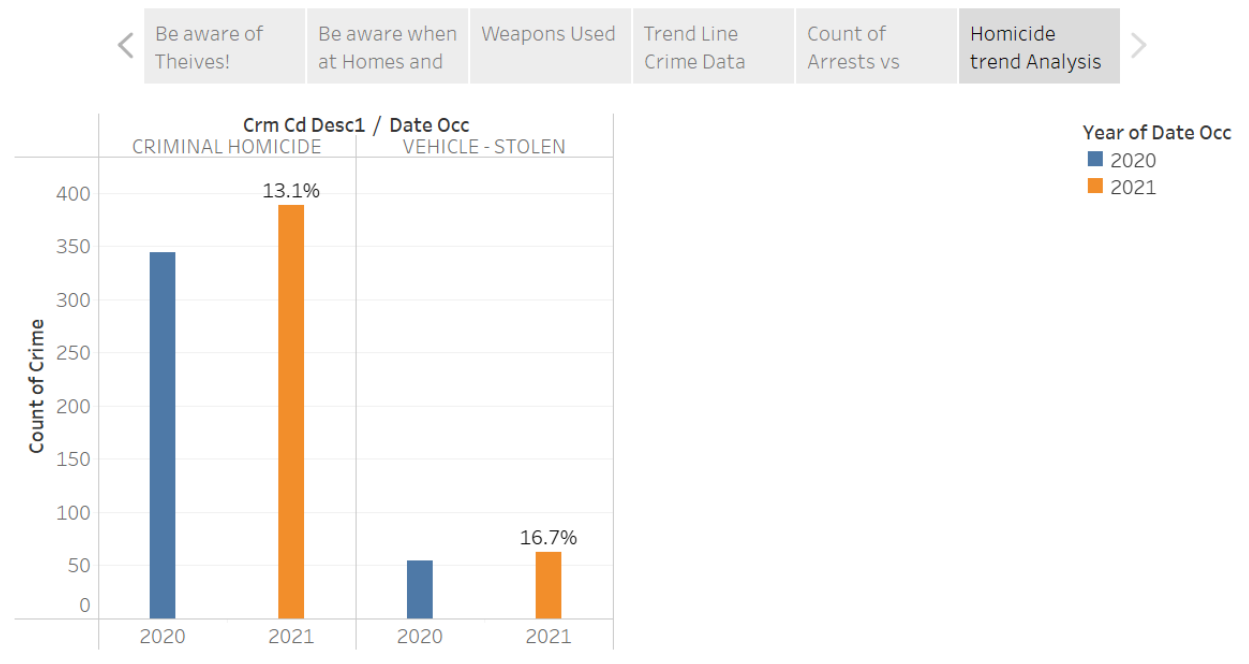
Graph shows the number of crimes versus the year of occurrence. For this analysist we are analyzing only the criminal homicide and the vehicle stolen crime type as it is quite evident from the graph that the number of criminal homicides has significantly increased from year 2020 to year 2021. A little increase in the number of crimes for vehicle theft have also been increased from year 2021 to year 2020. In order to get the exact increase from year 2020 to 2021, we have performed percentage calculation of each crime based on the previous year. Using this analysis, we can determine that Homicide has increased by approximately 13% relative to the previous year and the Vehicle Theft has increased by approximately 17% relative to year 2020. *(Spectrum News1, 2022)* This article from Spectrum news1 published on January 13, 2022, also talks about the increase in homicides and Vehicle thefts in 2021. Now that we know what Crimes are majorly occur in the City of Los Angeles, let us also see which Area/ Bureau is most dangerous and open to crimes.

<u>When in Valley Area- Be aware of the Surroundings!</u>
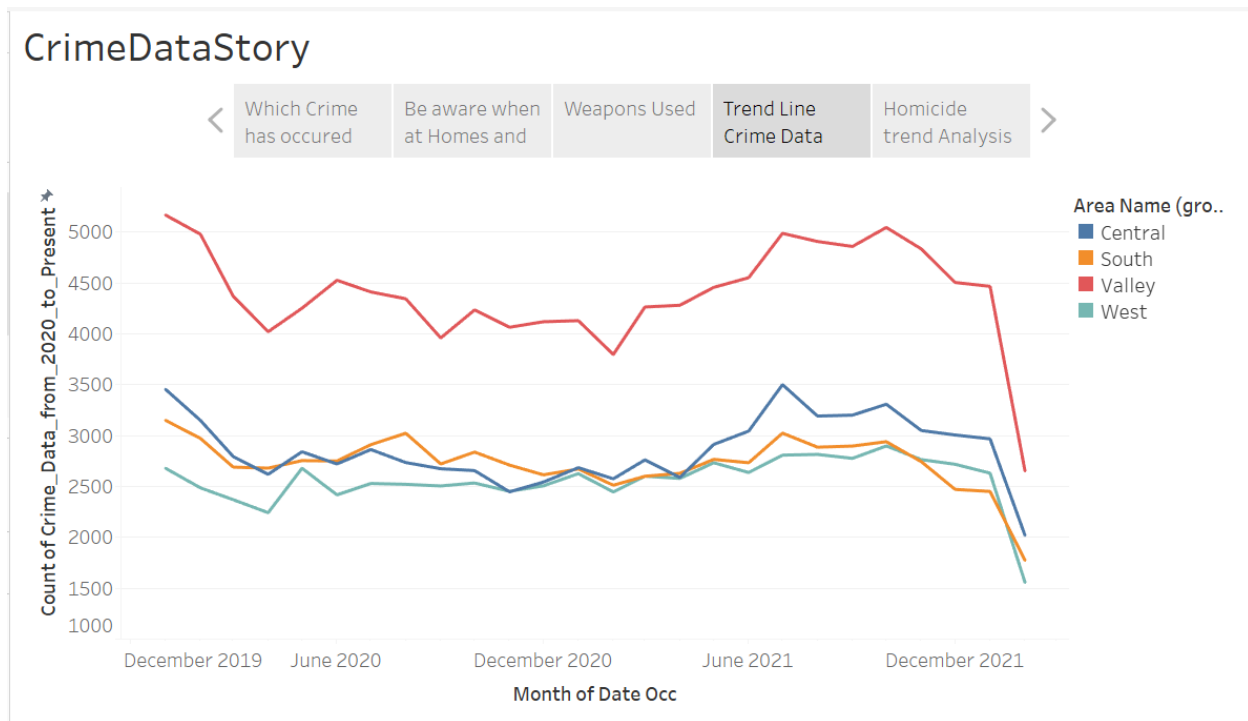
This **line chart** shows the number of crimes that have hope happened over the past few quarters.
For the purpose of this analysis, we have **grouped** 21 areas into four major bureaus - Central,
Valley, South, and West. The analysis is based on the major areas of L.A. Bing central, South,
valley, West. It is quite evident from the graph that most of the crimes have happened in the
valley and the vest is least vulnerable to the crimes. This analysis can help people those who are
shifting to L.A. To get the better understanding of the area and can choose the location of their
interest to reside in. In just a glance It can also be concluded from the graph that the probability
of Crime in Central, South and West areas is approximately the same, and is very low when
compared to that of the Valley. It can also be derived from the graph that the no of crimes
occurred in year 2020 is less in comparison to the year after. Such visualizations make data
analysis easier. The above analysis would also help local public to help them if they are looking
to transfer to different areas in the city. Premise is the most prone to Crime. Let us now dig

deeper, and try to analyze places where the probability of the Crime is the most, so that we can stay more alert at those places.

Be aware When at Homes and Streets!
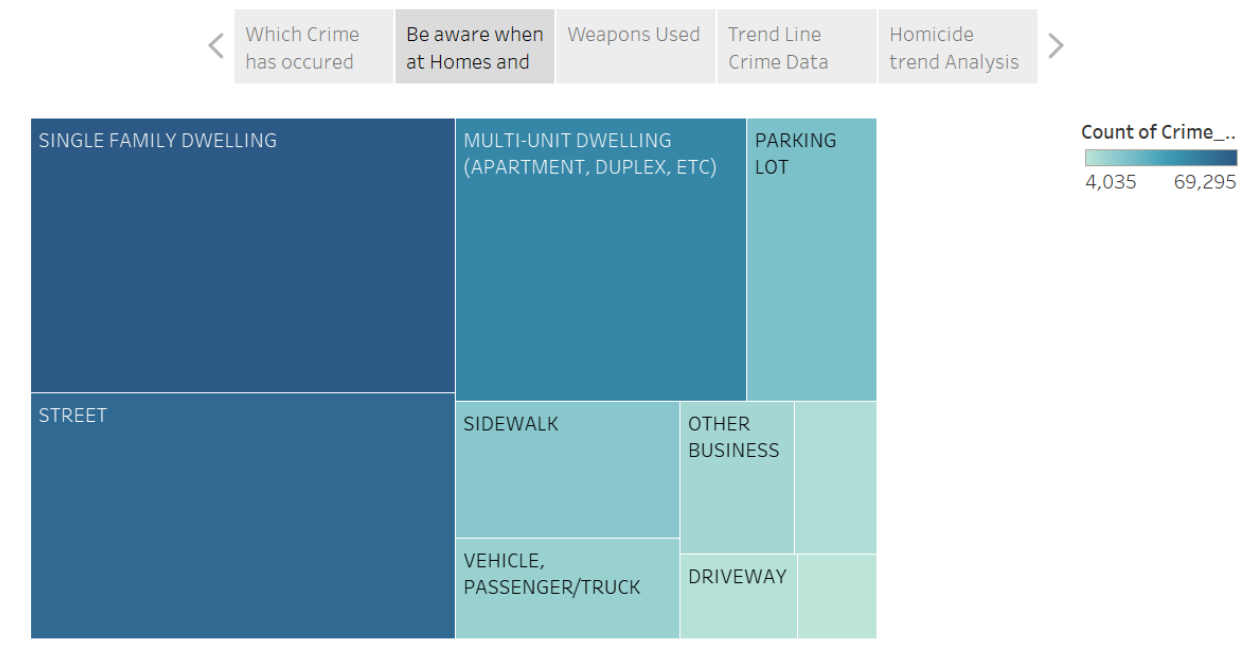
This Analysis shows the count for each of the top 10 most vulnerable premises, this analysis is done using a **tree map**. It shows us that most of the crimes happen in the houses add on streets. Very few crimes happen outdoors or in public. This visualization helps us understand that a person should we more aware when at homes and while he's walking on the streets alone. A filter has been used to **filter** out the top 10 most vulnerable Premise which together make up about 70 % of the crimes with recorded premises. To eliminate the risk of crimes at home, the residents should be more vigilant and should make sure the doors and windows are properly always locked. One should also keep an eye of what is happening in their surroundings when they are

walking on the streets. After analyzing the location, which is most prone to crimes, we also need to understand what kind of weapons are being used for majority of the crimes.

Weapons Used- Cautious of Strong Men!
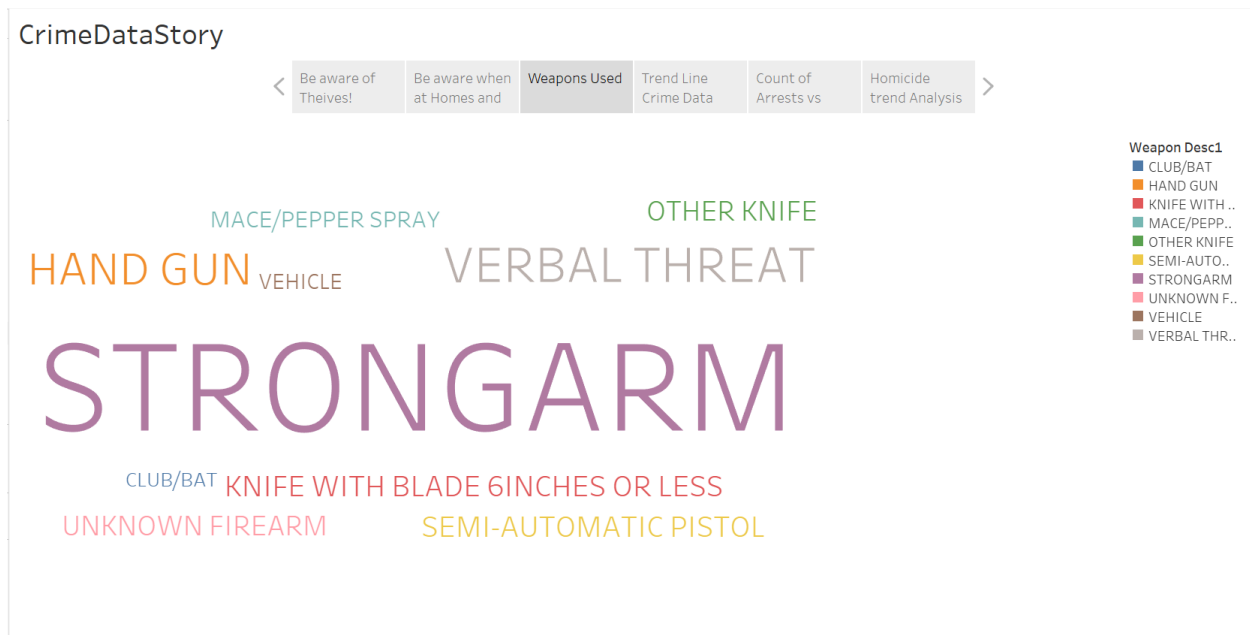
When we talk about Crimes and Violence, we cannot ignore Weapons. The Weapon can be as complicated as a machine gun and as simple as a physical force. A Verbal Threat is also considered as a Weapon. In simple words, anything which harms the Victim Physically or Mentally can be termed as a weapon. Abusing, humiliating , or Verbal abuse is no less a crime.

A **Word Cloud** has used in this analysis. This analysis which shows us which weapons were the most being used by the criminals. The larger the Text Size, the more its probability. As per the visualization a person uses his bodily force as a weapon, the most. The top 10 Weapons Used are being displayed, this is done using the **Filter** feature present in Tableau. Different colors have been assigned for different Weapon types. The color corresponding to weapon is mentioned in the legend on the right, for better visualization. Now that we are aware of the Types of weapons

which are being used the most, places where the crimes happen the most , Won't it be exciting to know which Ethnicity is the most prone to Crimes in LA.

Ethnicity more prone to crime
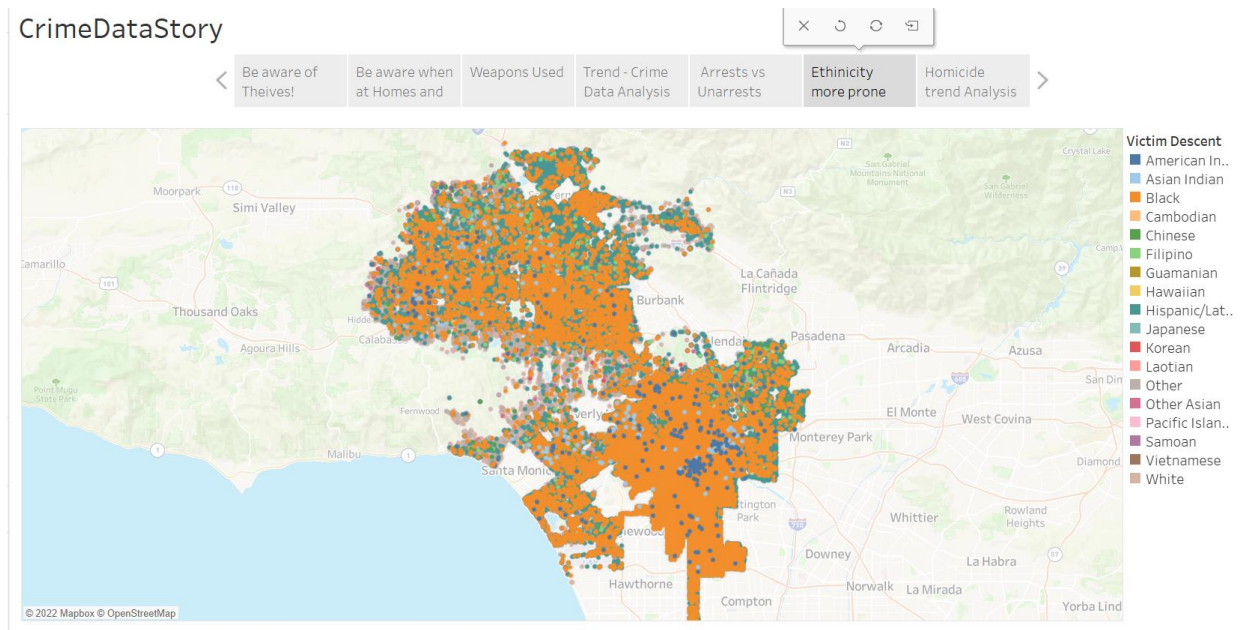
An analysis can be done for people pertaining to different ethnic background. This map depicts that which ethnicity is more prone to attacks. The below two screenshots shows significant difference in the proportion of Black or Hispanic/Latino and the Whites. By selecting Victim Descent from the filter on the right, we can view the probability of crime among each ethnicity.

According to the article by Los Angeles Times *(Los Angeles Times)* the Latino and Black victims account for nearly all of the L.A.'s surge in the homicides. It states that there is a surge in homicides in LA since the start of the pandemic and majority of the victims are Back or Latino. According to the impact of violence to gangs have been felt by all, victims old and young, sitting in their cars and working a shift, housed and homeless. The figures shows that this is really a

concerning situation, and trauma for the communities of color. If this continues, the People of

color will start feeling unsafe and start leaving LA.
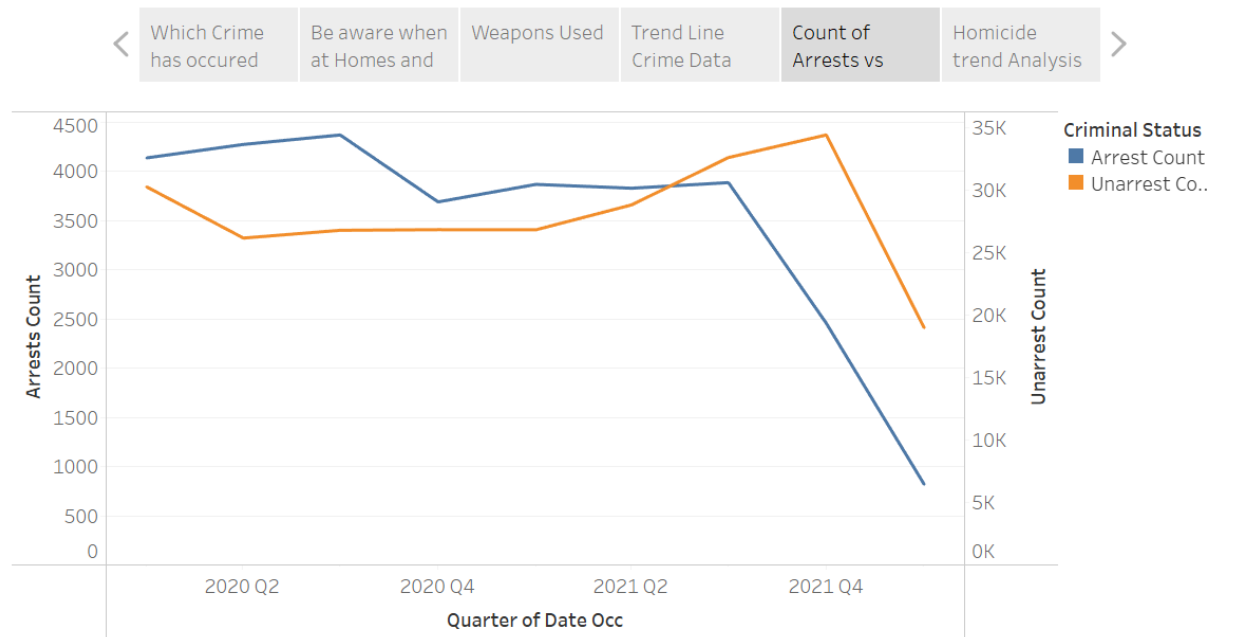
Map for Hispanic/Latino
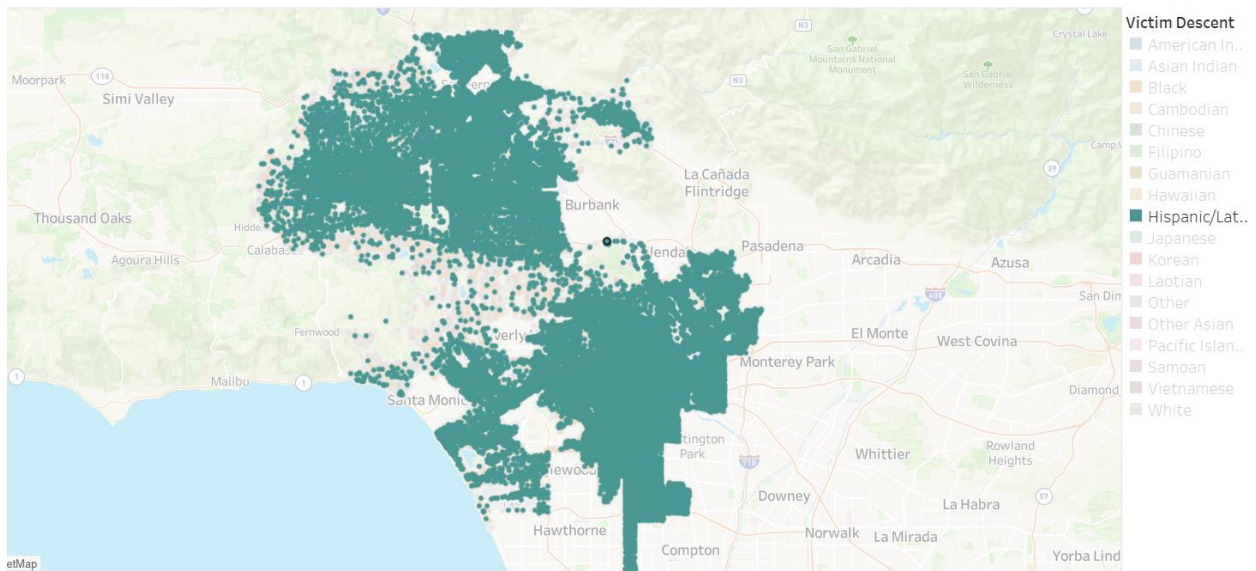
Map for Black Ethnicity
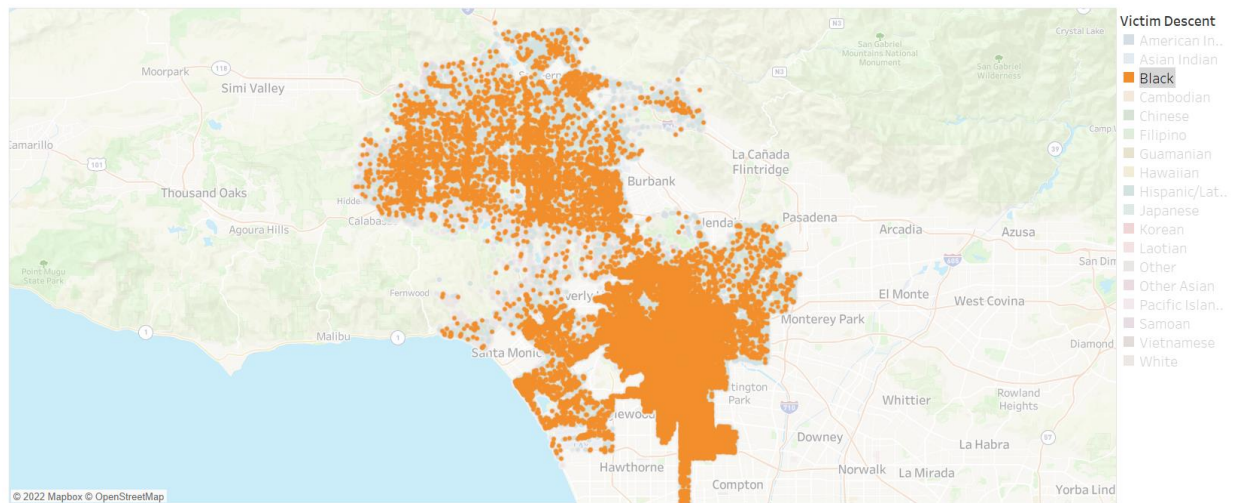
It is important to know the status of the Criminals. i.e., if proper Investigation is being done by the LAPD and are all the Criminals given the punishment they deserve. We can expect a safe surrounding only when the criminals are behind bars and not let out loose.

<u>Watchout! The Criminals are still roaming out loose!</u>

This is a dual access chart which helps us identify the status of the criminals i.e., how many criminals have been arrested for the crime and how many are not being arrested. Two different colors have been used for this analysis. The Orange color represents the number of Criminals not arrested and Blue being the number of Criminals arrested. It is quite evident from the graph that the number of criminals not arrested is very high when compared to number of criminals that are arrested. Also see that the number of Everest count starts gradually decreasing after the quarter two of 2021. The nonarrest count obtains its peak during 2021 quarter 4 and then we can see a sudden downfall in the number of criminals that are not arrested. At a single glance we can see that the count of Criminals not arrested had always been less than those arrested until Q2 2021, but after that we could see the increase in the numbers.

CONCLUSION

Crimes cannot be completely removed from our society; however, it can be controlled through the efforts of judiciary and law enforcement agencies and citizens. We as citizens are responsible for creating a safe living environment and one of the steps to follow carefully is to report a crime when it is encountered. We don't have to be victims but reporting crimes on the spot can prevent others from becoming one.

- Majority of the Crimes were recorded in Valley Area. Increased Surveillance and Patrolling is required in and around this area.

- Criminals arrested Count is very Low. Proper Investigation needs to be done for every crime. So that no criminal can roam free as it can cause threat to the public.

- Theft Crimes have topped the list of crimes reported and it seems to increase with every passing year. Proper measures need to be taken and the LAPD needs to keep a check on the plausible reasons.

- Most of the crimes take place at the Residence and streets. To avoid such miss happening the public should be more aware when walking on streets and at homes. They should make sure the Doors and Windows of their houses are well- locked and have security cameras in place. One should also be alert while walking on the streets.

- There is the rise in crimes around the weekend. These Crimes can be reduced by tightening the security and alert public.

REFERENCES

1. Spectrum News1 City officials: LA homicides, vehicle thefts, officer-involved shootings

   increased in 2021 *Carpenter, Jan 2022.* Retrieved on April 5, 2022

   From https://spectrumnews1.com/ca/la-west/public-safety/2022/01/13/homicides--
   vehicle-thefts--officer-involved-shootings-increased-in-la-in-2021

2. Losangeles Magazine: How Bad is the Crime in L.A? *Peter Kiefer*, January 2022.

   Retrieved on April 5, 2022 https://www.lamag.com/citythinkblog/how-bad-is-crime-in-l-
   a/

3. Los Angeles Times: Latino and Black Victims account for nearly all of the L.A.'s surge in

   homicides. August 13,2021. Retrieved on April 5 , 2022

   https://www.latimes.com/california/story/2021-08-13/la-homicides-since-covid-take-
   mostly-black-latino-victims