

## **BDA Term Project Plan**

### 1. Team Leader and Team Scribe

Team Leader - Nikita Nalawade

Team Scribe - Manasi Prabhune

### 2. Contact Information

Ashwini Kasbekar - (980)-365-9168

Eshan Bhatt - (704)-5474186

Manasi Prabhune - (704)-323-0402

Nikita Nalawade - (704)-222-9436

Prutha Shiroadkar - (704)-361-1270

Ripal Bhavsar - (980)-365-9121

### 3. Method of Communication

Email

Skype calls

Meetings every week

File sharing via Email and Google Drive

### 4. Meeting Dates

Friday - 10:30 pm

Monday and Wednesday -Online

### 5. Dataset

The dataset which we are planning to use is the CRIME dataset for the city of Los Angeles. The crime details are from 2010 to present.

<https://catalog.data.gov/dataset/crime-data-from-2010-to-present>

The search questions which we are planning to work on are

- The area in the city where the crime took place
- The weapon with which the crime was committed
- The type of crime and the time at which it was committed

## 6. BDA Plan

Project Task	Start Date	End Date	Duration	Status
Group Formation	10th January	16th January	6 days	
Project Plan	17th January	27th January	10 days	
Phase 1 Report	28th January	17th February	11 days	
Peer Evaluation	18th February	14th April	4 days	
Project Report (a-h)	5th April	14th April	9 days	
Peer Review	15th April	21st April	7 days	
Revision of Paper	22nd April	27th April	5 days	
PPT presentation	27th April	28th April	1 day	
Peer evaluation submission	28th April	29th April	1 day	

### **Business Case**

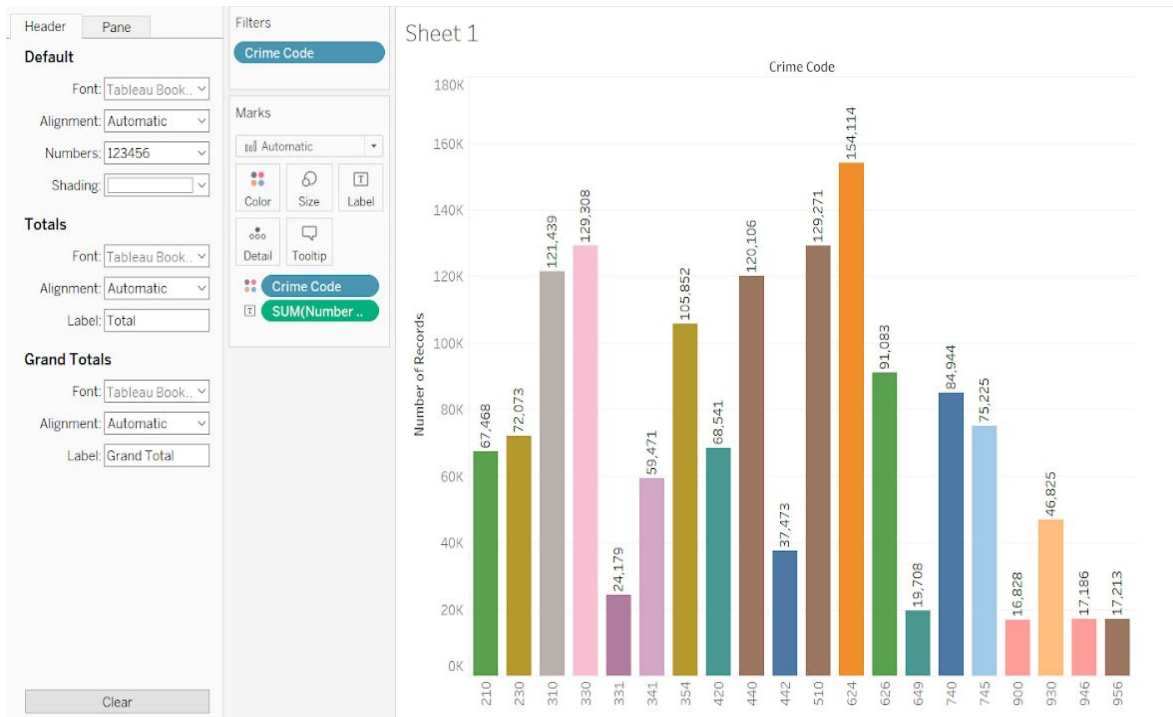
Mrs. D'souza, is the Deputy Commissioner of Los Angeles Police Department and she wants to understand the current crime scenario in the different areas of Los Angeles. Being the Deputy Commissioner, she needs to understand different aspects of the committed crimes, based on which she can take the necessary measures to reduce the crime rate.

Mrs. D'souza has provided us with the data of committed crimes for the years 2010-2018 and has asked us to use this data along with the analytical skills of the team, to generate insights using which Mrs. D'souza can take the necessary actions to minimize the crime rate in Los Angeles. Mrs D'souza has gave us certain criteria which needs to be satisfied:-

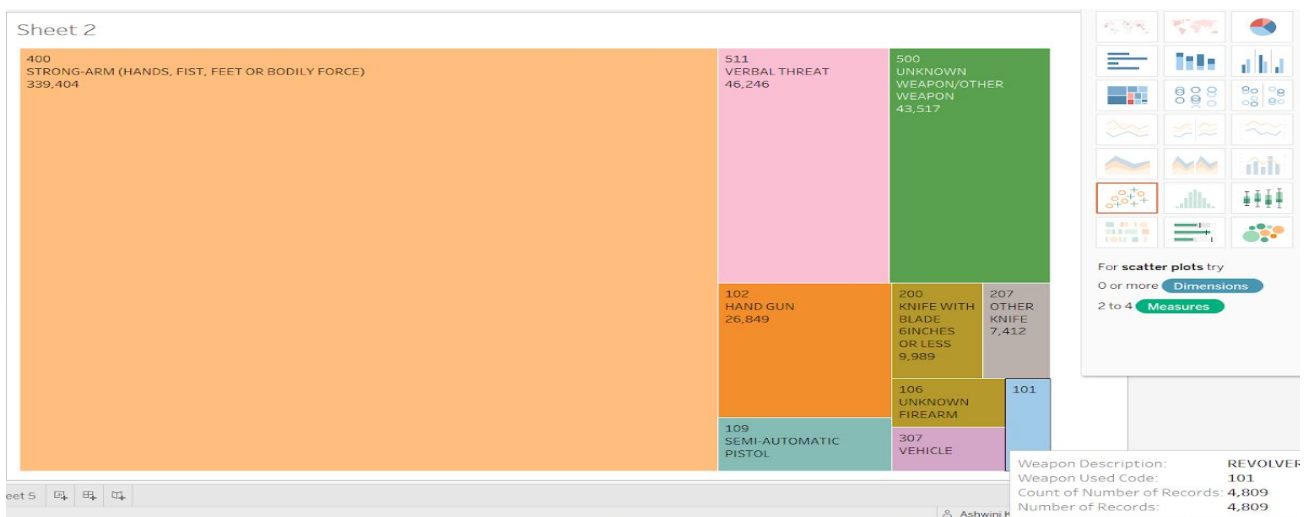
1. Analysis should include time-wise crime report
2. Identify the most used weapon
3. Maximum type of crimes in the city.
4. Most affected area.

## Visualizations:

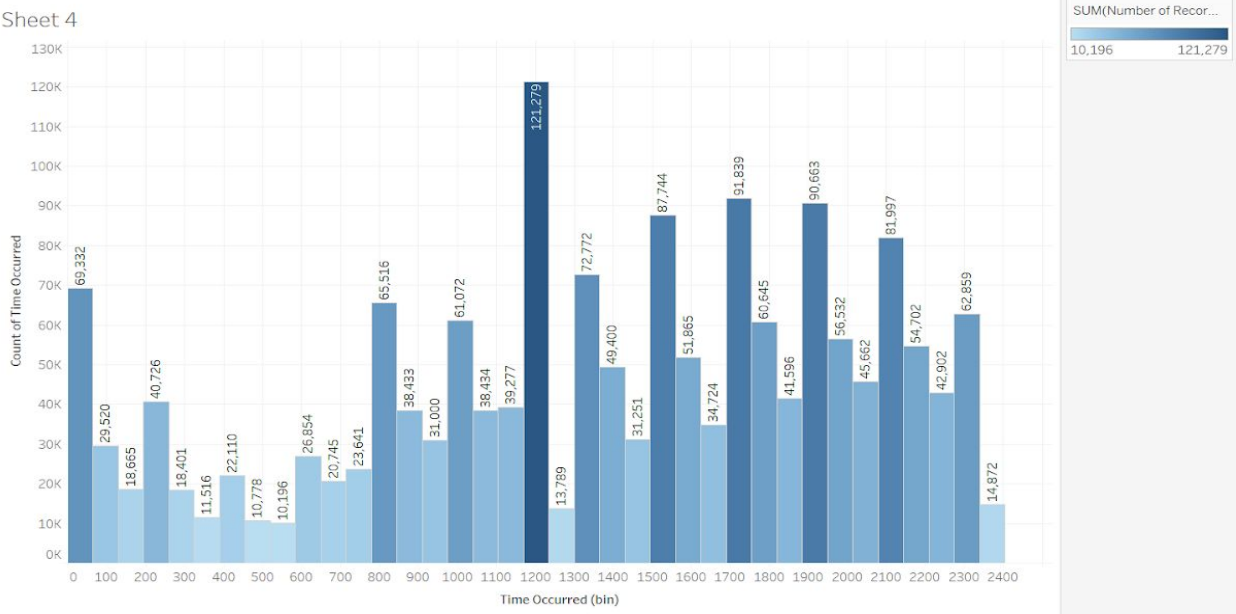
### 1. Top N crimes in Los Angeles[B]



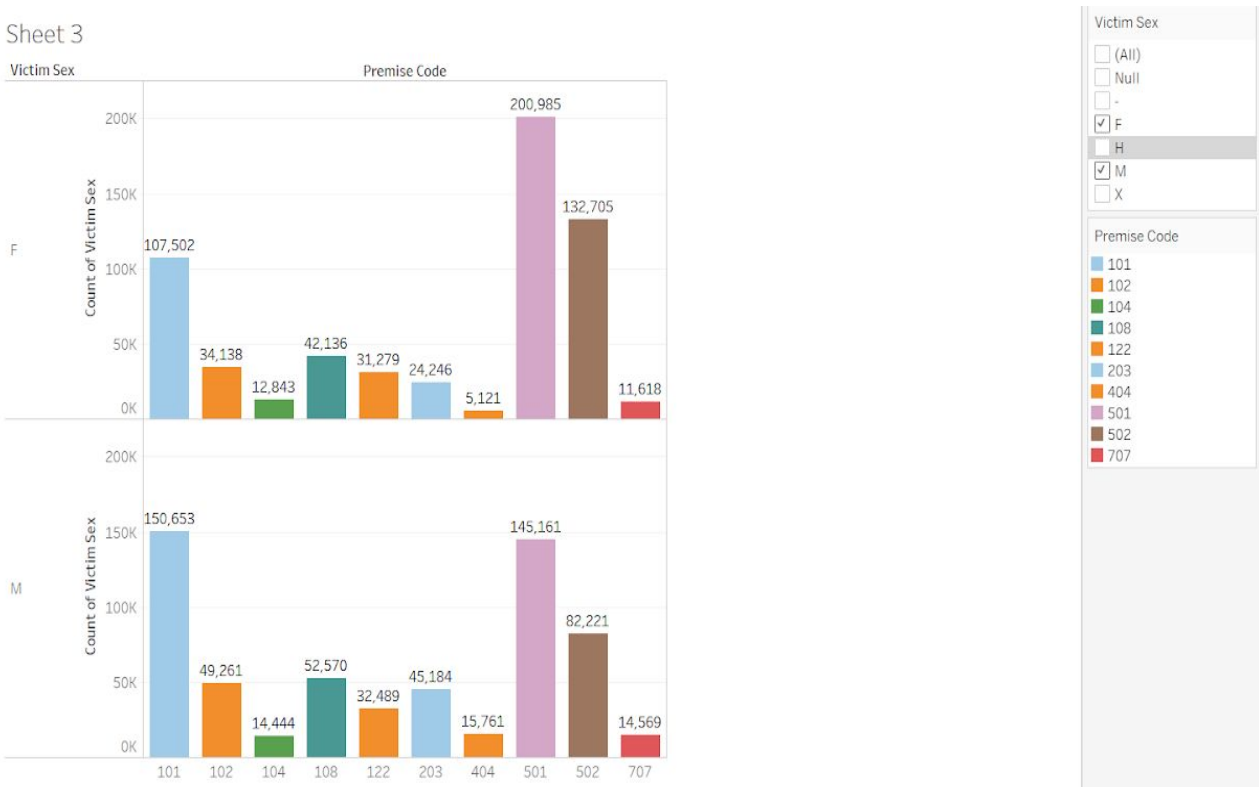
### 2. Top 10 used weapons[C]



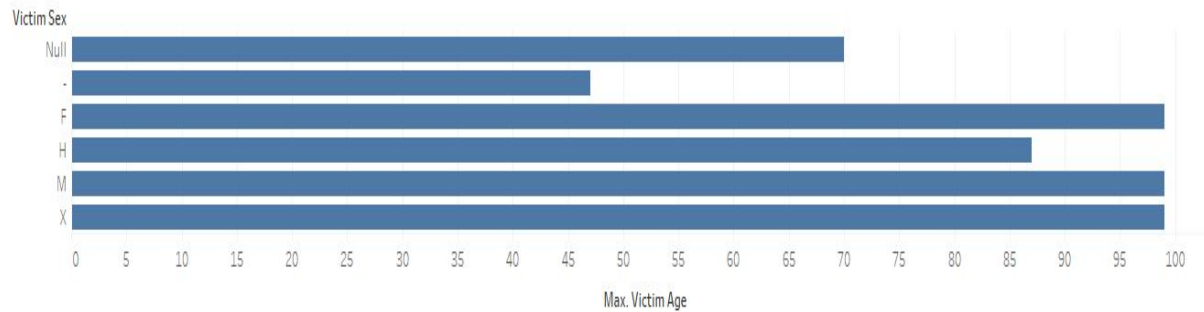
3. Time wise crime count[D]



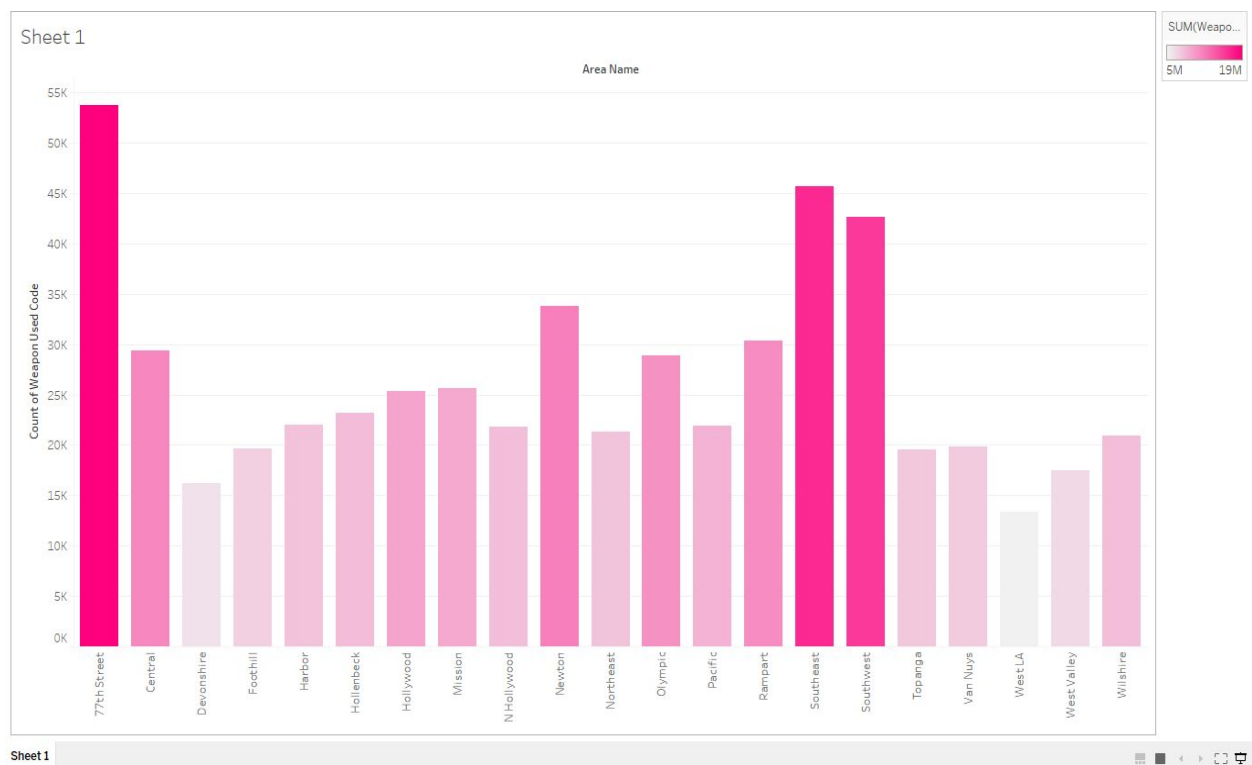
4. Premise wise count for male and female victims[E]



## 5. Victim age versus victim sex[F]



## G. Location verses the weapon used[G]



The research questions based on visualizations

- Timeline of crime data from 2010
- Premise wise crime analysis

**Data Preprocessing:**

We have performed following activities:

1. Replacing Missing Values with necessary values
2. Redundant Data Removal
3. Discretizing Appropriate Data (See Appendix A)

## Appendix:

[A] Victim Age: Discretize the victims age in 5 broad categories for a better and also a broad analysis on victims age with respect to crime occurred.

Time at which Crime Occurred: Discretized the time occurred with 4 classes; namely morning, afternoon, evening and night. This type of data can be used to analyse and identify the time of day during which most of the crimes occur; which Official can use for various tasks.

[B] The above visualization plot gives the count of top N crimes. We can see top 20 crimes.

On analysing the above graph we can see that crime code 624 i.e (BATTERY - SIMPLE ASSAULT) has occurred 154,114 times. This crime has occurred maximum times. Least occurred crime is Violation of court order (900) which occurred 16,828 times.

[C] Above is a treemap view of top 10 most used weapons. Preliminary analysis here shows that Strong-arm (Hands, fist, feet or bodily force) was used to attack victims 339,404 times. The least weapon used to attack victims was revolver used 4809 times. Rest weapons used most of the times were: hand gun, semi automatic pistol, knife with blade, vehicle etc.

[D] Bar graph below gives the details of number of crimes occurring at particular time. We can see maximum crimes occurred at 1200 hrs (count : 1,21,279 ) and least crime count was at 0600 hrs with a count of 10,196.

Bar graph below gives the details of number of crimes occurring at particular time. We can see maximum crimes occurred at 1200 hrs (count : 1,21,279 ) and least crime count was at 0600 hrs with a count of 10,196.

[E] Graph below gives count of male and female victim count based on premises. These are top 10 premises. We can see that premise 501 (Single Family Dwelling) has highest male and female victim ratio with 200,985 female victims and 145,161 male victim counts. Department store (premise code 404) is the location with least female victim count and Driveway (premise code 104) is the premise for least male count.

[F] The horizontal bars graph gives us the record details for the maximum gender and corresponding age group of the victims.

From the graph below , we can figure out that the maximum victim age is 99 with maximum of females, males and x gender in it.

[G] The visualisation shows the graph of different locations in our dataset for different weapons used to commit crime in that area. From this we can conclude, 77th street, Southeast and Southwest areas are the maximum committed crime areas in Los Angeles.

