# Final Project Check In Data 603

By: Hafsa Chaudhry

# My Problem Statement

Since the beginning of 2020 to present, there has been roughly 500,000 reported crime cases in Los Angeles alone. The majority of victims being the ages of 23-25, and female. How can we lower the likelihood of people in LA being victim of crime ?

# My Source Data, Along with some initial data exploration

Kaggle. Kaggle has a good variety of data sets on machine learning. It requires registration but is worth it.

https://www.kaggle.com/datasets/hemil26/crime-in-los-angeles/code

INITIAL DATA EXPLORATION IN NEXT SLIDES

```
df = pd.DataFrame(data)
```

```
df.head()
```

| DR_NO | DATE OCC | TIME OCC | AREA | AREA NAME | Rpt Dist No | Crm Cd | Crm Cd Desc | Mocodes | Vict Age | Vict Sex | Vict Descent | Premis Desc | Weapon Used Cd | Weapon Desc | Status | Status Desc | LOCATION |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10304468 | 08/01/2020 | 22:30 | 3 | Southwest | 377 | 624 | BATTERY - SIMPLE ASSAULT | 0444 0913 | 36 | F | B | SINGLE FAMILY DWELLING | 400.0 | STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE) | AO | Adult Other | 1100 W 39TH PL |
| 190101086 | 01/01/2020 | 03:30 | 1 | Central | 163 | 624 | BATTERY - SIMPLE ASSAULT | 0416 1822 1414 | 25 | M | H | SIDEWALK | 500.0 | UNKNOWN WEAPON/OTHER WEAPON | IC | Invest Cont | 700 S HILL ST |
| 191501505 | 01/01/2020 | 17:30 | 15 | N Hollywood | 1543 | 745 | VANDALISM - MISDEAMEANOR ($399 OR UNDER) | 0329 1402 | 76 | F | W | MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC) | 500.0 | UNKNOWN WEAPON/OTHER WEAPON | IC | Invest Cont | 5400 CORTEEN PL |
| 191921269 | 01/01/2020 | 04:15 | 19 | Mission | 1998 | 740 | VANDALISM - FELONY ($400 & OVER, ALL CHURCH VA... | 0329 | 31 | X | X | BEAUTY SUPPLY STORE | 500.0 | UNKNOWN WEAPON/OTHER WEAPON | IC | Invest Cont | 14400 TITUS ST |
| 200100501 | 01/01/2020 | 00:30 | 1 | Central | 163 | 121 | RAPE, FORCIBLE | 0413 1822 1262 1415 | 25 | F | H | NIGHT CLUB (OPEN EVENINGS ONLY) | 500.0 | UNKNOWN WEAPON/OTHER WEAPON | IC | Invest Cont | 700 S BROADWAY |

```
df.info()
```

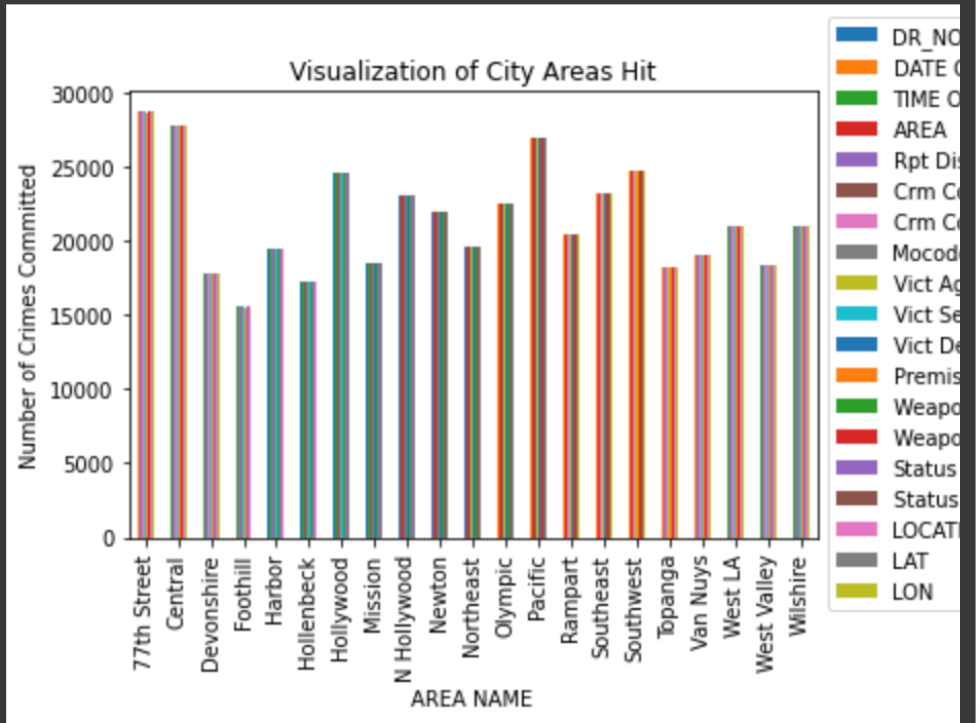
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 449630 entries, 0 to 449629
Data columns (total 20 columns):
 #   Column          Non-Null Count   Dtype  
---  ------          --------------   -----  
 0   DR_NO           449630 non-null  int64  
 1   DATE OCC        449630 non-null  object 
 2   TIME OCC        449630 non-null  object 
 3   AREA            449630 non-null  int64  
 4   AREA NAME       449630 non-null  object 
 5   Rpt Dist No     449630 non-null  int64  
 6   Crm Cd          449630 non-null  int64  
 7   Crm Cd Desc     449630 non-null  object 
 8   Mocodes         449630 non-null  object 
 9   Vict Age        449630 non-null  int64  
 10  Vict Sex        449630 non-null  object 
 11  Vict Descent    449630 non-null  object 
 12  Premis Desc     449446 non-null  object 
 13  Weapon Used Cd  449630 non-null  float64
 14  Weapon Desc     449630 non-null  object 
 15  Status          449630 non-null  object 
 16  Status Desc     449630 non-null  object 
 17  LOCATION        449630 non-null  object 
 18  LAT             449630 non-null  float64
 19  LON             449630 non-null  float64
dtypes: float64(3), int64(5), object(12)
memory usage: 68.6+ MB
```

```
print("CRIME CODE DESCRIPTION")
df["Crm Cd Desc"].describe()
df["Crm Cd Desc"].head
```

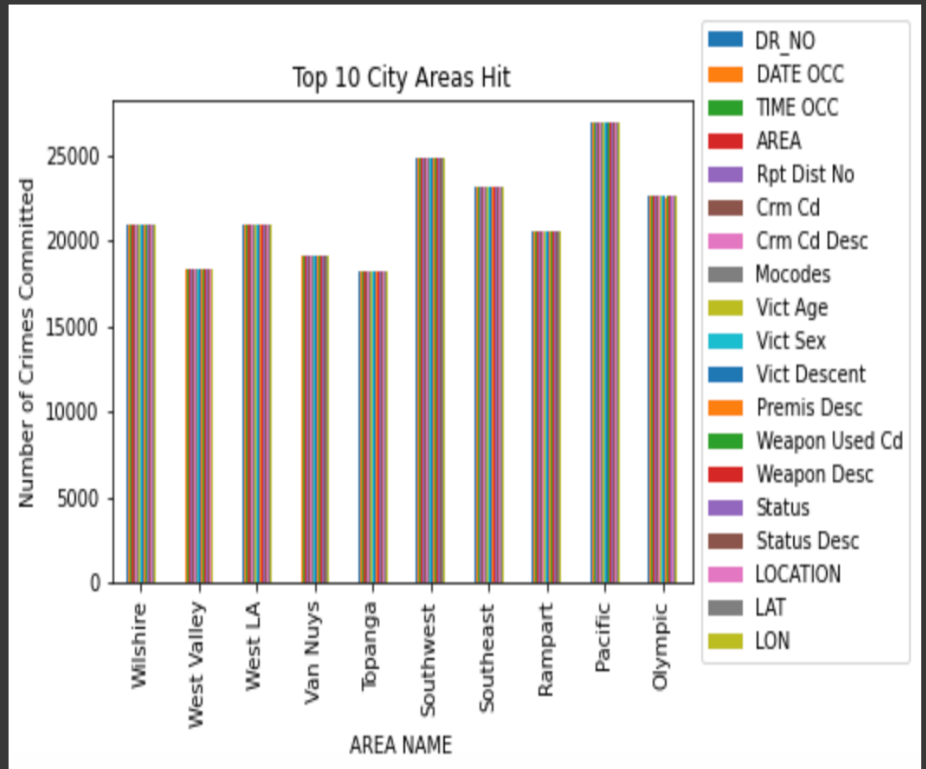
```
CRIME CODE DESCRIPTION
<bound method NDFrame.head of 0                                      BATTERY - SIMPLE ASSAULT
1                                    BATTERY - SIMPLE ASSAULT
2                  VANDALISM - MISDEAMEANOR ($399 OR UNDER)
3         VANDALISM - FELONY ($400 & OVER, ALL CHURCH VA...
4                                          RAPE, FORCIBLE
                                ...
449625                      VIOLATION OF RESTRAINING ORDER
449626                  THEFT FROM MOTOR VEHICLE - ATTEMPT
449627                                    VEHICLE - STOLEN
449628                                    VEHICLE - STOLEN
449629    THEFT FROM MOTOR VEHICLE - GRAND ($950.01 AND ...
Name: Crm Cd Desc, Length: 449630, dtype: object>
```

```python
import matplotlib.pyplot as plt

city_count = df.groupby('AREA NAME').count()
city_count.plot.bar()
plt.ylabel('Number of Crimes Committed')
plt.title('Visualization of City Areas Hit')
plt.legend(loc='center left', bbox_to_anchor=(1, 0.5))
plt.show()
```




```python
top_10 = city_count.sort_values(by="AREA NAME", ascending=False).
top_10.plot.bar()
plt.ylabel('Number of Crimes Committed')
plt.title('Top 10 City Areas Hit')
plt.legend(loc='center left', bbox_to_anchor=(1, 0.5))
plt.show()
```



# MY PROPOSED SOLUTION

My solution is purely data analysis:

By analyzing data for highest crime committed regions per timeframe, we can increase the presence of LAPD officers to monitor and prevent crime activity. (hope to prove that you can predict at certain times, crime is more likely in certain areas)

By creating visualization aid of types of crime and ages of victims correlating to area of crime, we can help create awareness to civilians in the area. (hope to prove that you can predict that type of crime and ages of victims are more correlated in certain areas)