We were to produce a very convincing EDA and visualization showing insights and patterns from open datasets related to Los Angeles City ranging from LAPD, COVID,Weather etc.

Task1:

Starting with loading all the crime datasets and taking info about the datasets.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 317854 entries, 0 to 317853
Data columns (total 28 columns):
 #   Column          Non-Null Count   Dtype  
---  ------          --------------   -----  
 0   DR_NO           317854 non-null  int64  
 1   Date Rptd       317854 non-null  object 
 2   DATE OCC        317854 non-null  object 
 3   TIME OCC        317854 non-null  int64  
 4   AREA            317854 non-null  int64  
 5   AREA NAME       317854 non-null  object 
 6   Rpt Dist No     317854 non-null  int64  
 7   Part 1-2        317854 non-null  int64  
 8   Crm Cd          317854 non-null  int64  
 9   Crm Cd Desc     317854 non-null  object 
 10  Mocodes         274531 non-null  object 
 11  Vict Age        317854 non-null  int64  
 12  Vict Sex        276448 non-null  object 
 13  Vict Descent    276443 non-null  object 
 14  Premis Cd       317849 non-null  float64
 15  Premis Desc     317746 non-null  object 
 16  Weapon Used Cd  116477 non-null  float64
 17  Weapon Desc     116477 non-null  object 
 18  Status          317854 non-null  object 
 19  Status Desc     317854 non-null  object 
 20  Crm Cd 1        317851 non-null  float64
 21  Crm Cd 2        25981 non-null   float64
 22  Crm Cd 3        880 non-null     float64
 23  Crm Cd 4        30 non-null      float64
 24  LOCATION        317854 non-null  object 
 25  Cross Street    56977 non-null   object 
 26  LAT             317854 non-null  float64
 27  LON             317854 non-null  float64
dtypes: float64(8), int64(7), object(13)
memory usage: 67.9+ MB
```

Looking at the info(), we can see every crime dataset has 28 columns.

We concatenated all the crime datasets into a single DataFrame as all the columns were same and it also helps in visualizing data in a single go.

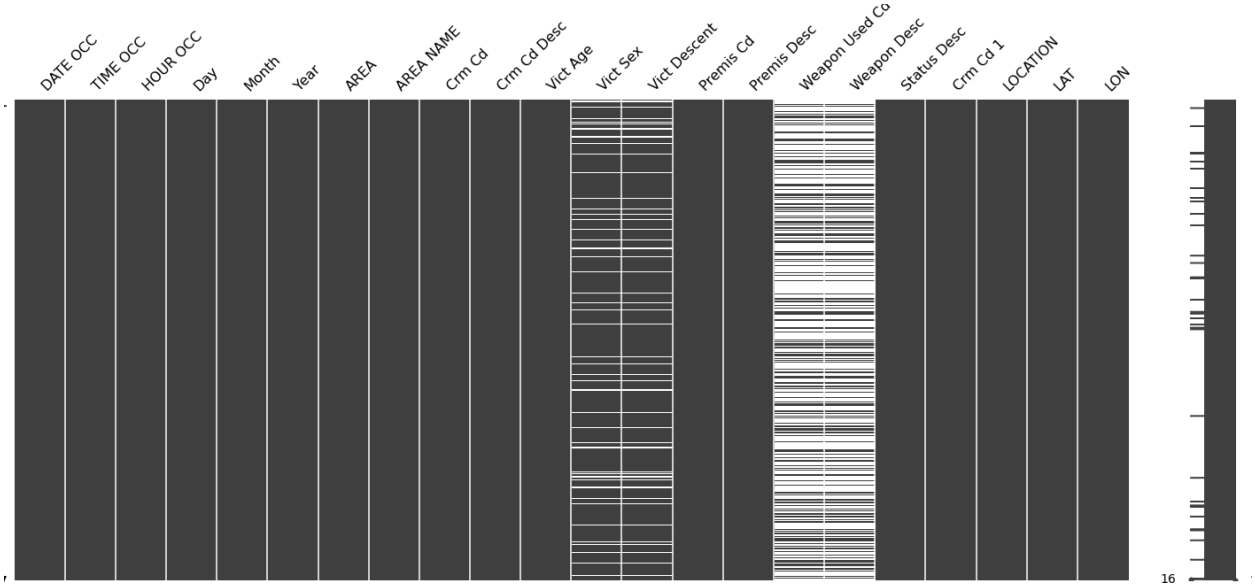
```
[4]  df= pd.concat([df,df1])

[5]  df= pd.concat([df,df2])

[6]  df.info()

     <class 'pandas.core.frame.DataFrame'>
```

There is just a single data column in the crime dataset, we extracted name of day,month name as well 4 digit year into different columns

```
df.head(2)
```

DATE OCC	TIME OCC	HOUR OCC	Day	Month	Year	AREA	AREA NAME	Crm Cd	Crm Cd Desc	Vict Age	Vict Sex	Vict Descent	Premis Cd	Premis Desc	Weapon Used Cd	Weapon Desc	Status Desc	Crm Cd 1	LOCATION	LAT	LON
2020-01-08	2230	22	Wednesday	January	2020	3.0	Southwest	624	BATTERY - SIMPLE ASSAULT	36	F	B	501	SINGLE FAMILY DWELLING	400	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	Adult Other	624	1100 W 39TH PL	34.0141	-118.2978
2020-01-01	1730	17	Wednesday	January	2020	15.0	N Hollywood	745	VANDALISM - MISDEAMEANOR ($399 OR UNDER)	76	F	W	502	MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)	MISSING	MISSING	Invest Cont	745	5400 CORTEEN PL	34.1685	-118.4019

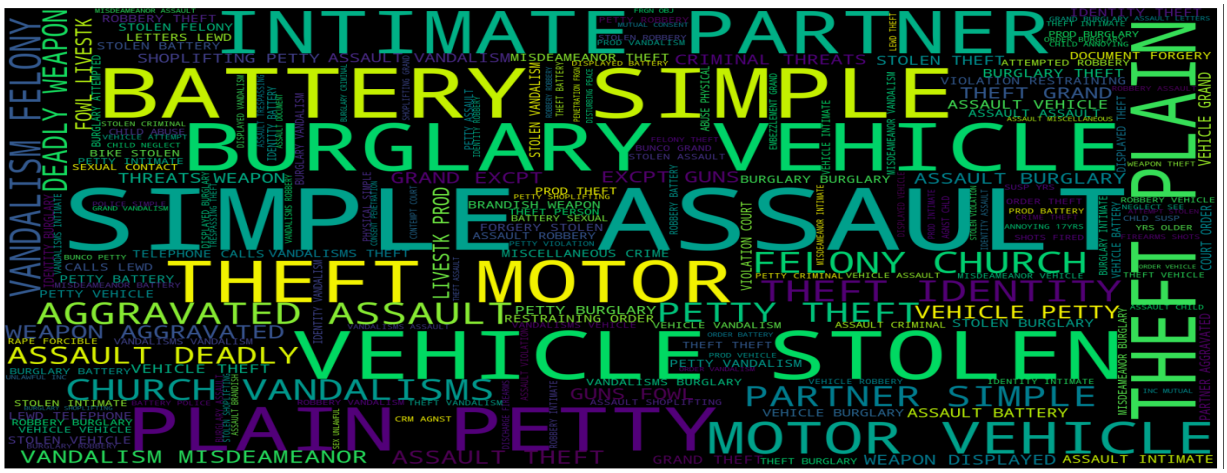
Visualizing the Missing Data:



We used a data visualization library for visualizing missing data from DataFrame just like above, it gives an overview look of what's going on tn the DataSet in a blink of an eye.

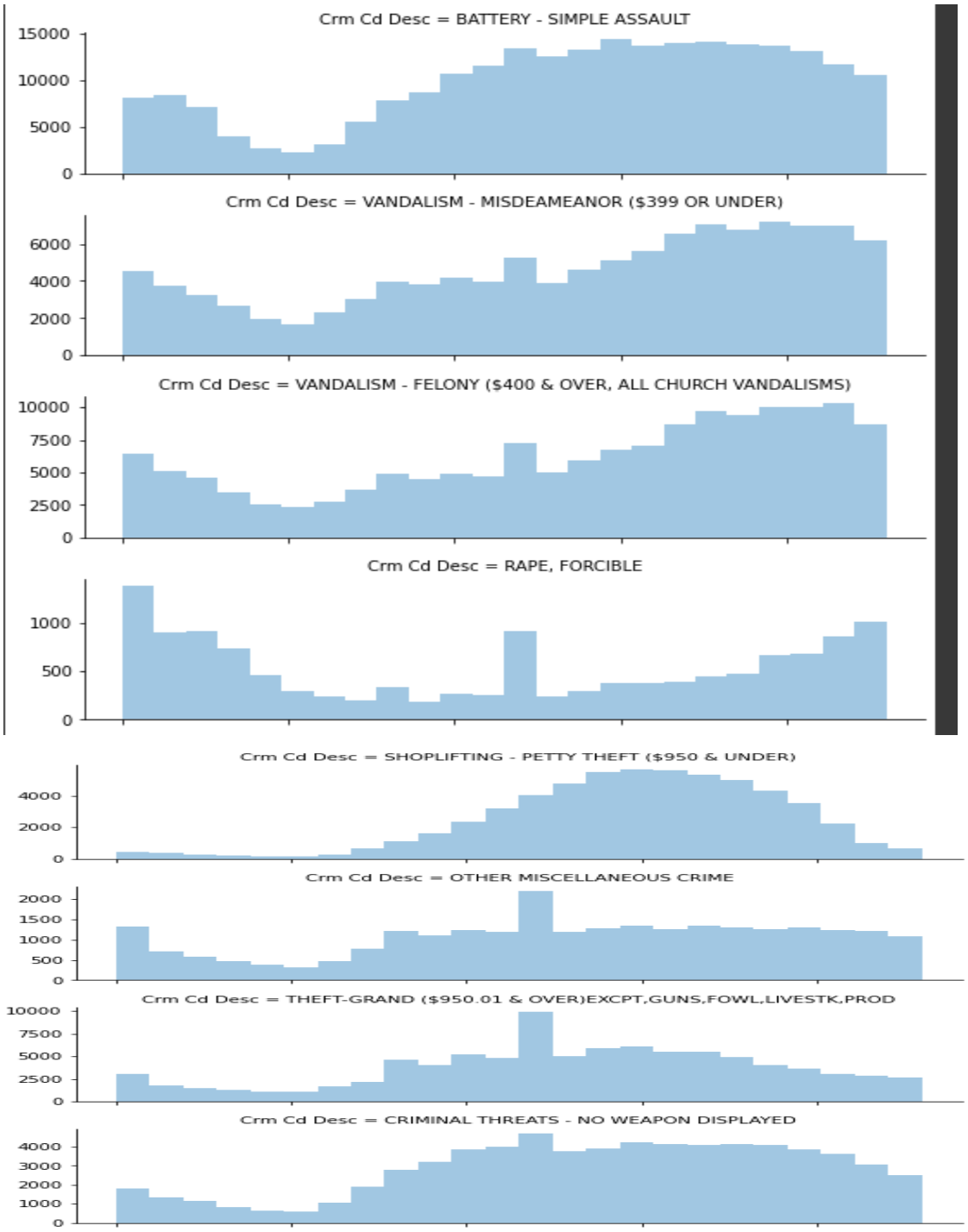
```
import missingno as msno
msno.matrix(df)
plt.show()
```

Crimes Description visualization:



World cloud is the classic example of visualizing what is the most occurring and frequent type of variable/vector in any form in a column.

Crimes Peak time in a day:



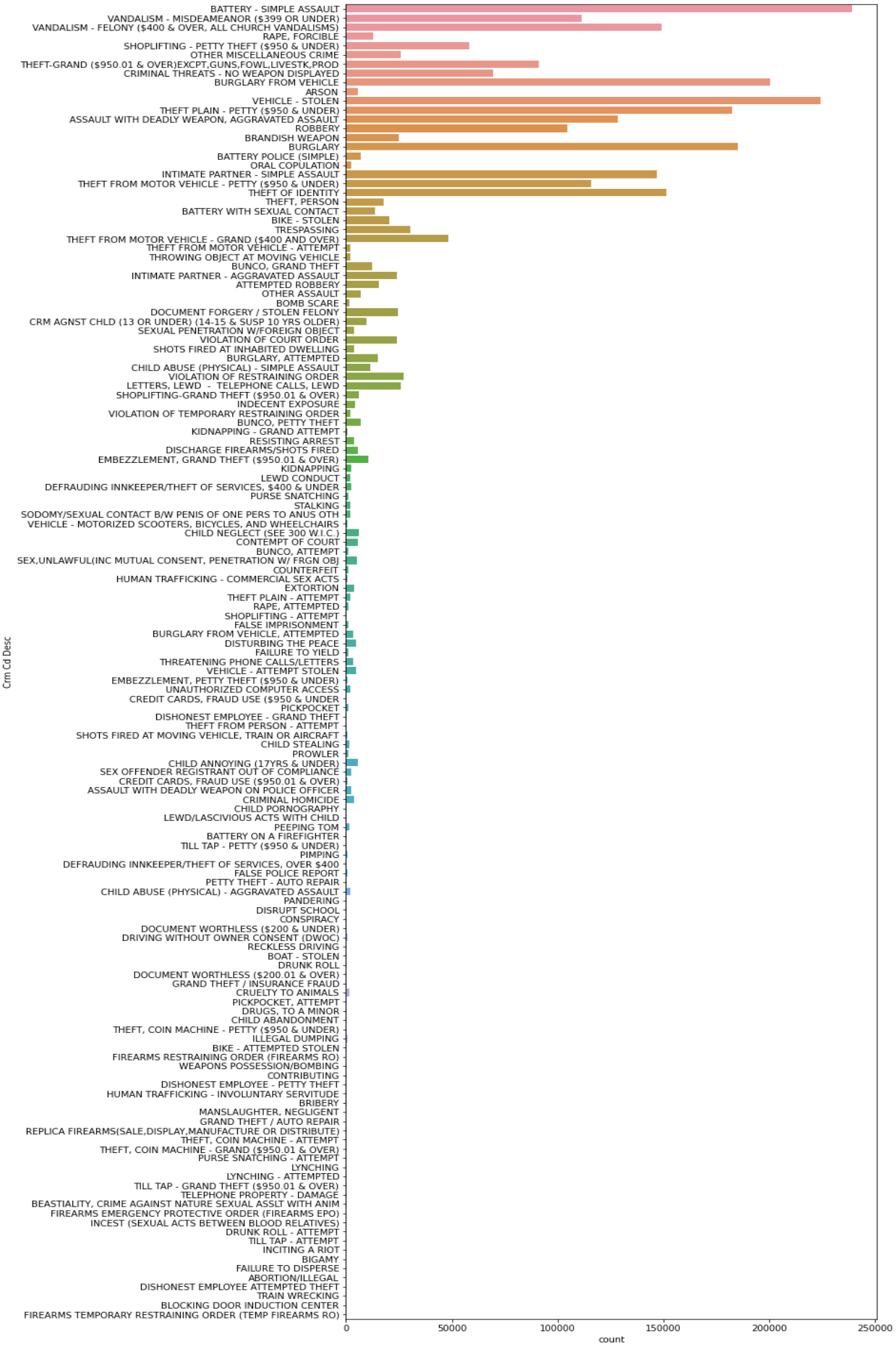
We applied seaborn FacetGrid to get graphs for every crime and then plot them for the **simple question that at which time of the day crime is at the peak.** Crime is at peak @ around 12 and after 12 it picks pace in most cases.

Question: Which day is most hot for crimes to happen?

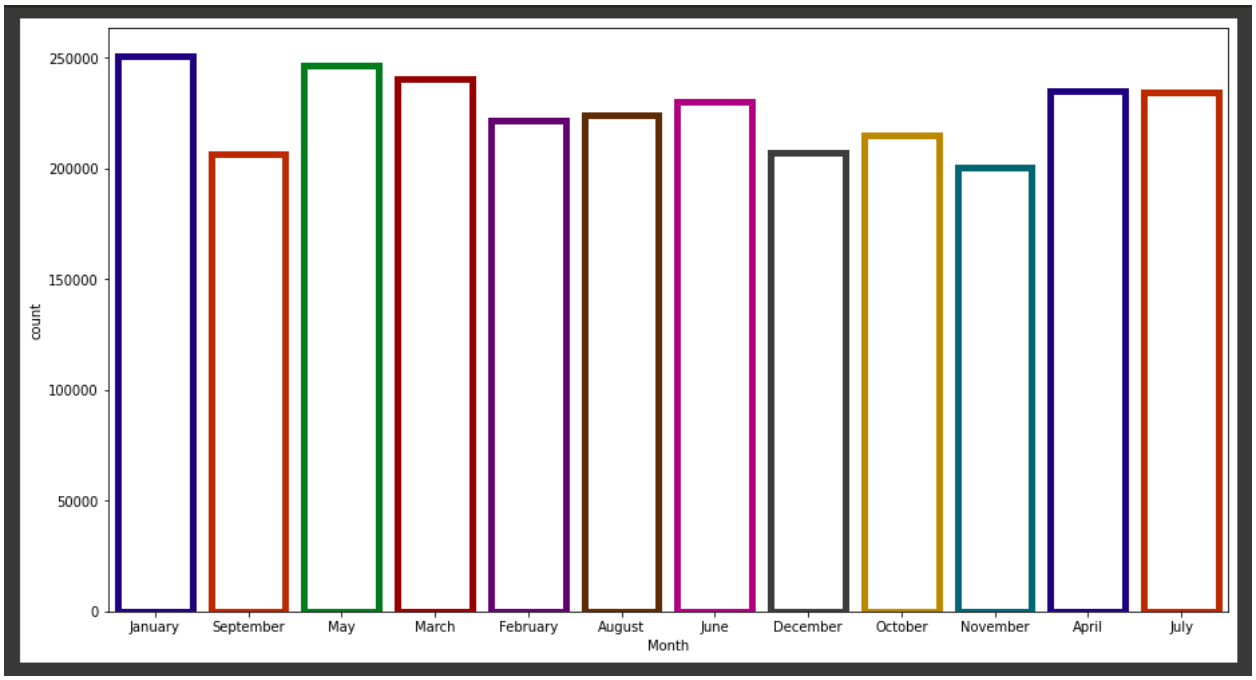
There is no apparent difference between the frequency of crimes but still Friday sees most crimes in LA and while Sunday remains the most calm day of the week. The simpe reason inferred can be that on Friday most people are out for weekend as they get free from offices and schools,colleges so criminals are also active on that day, while on Sunday most people like to chill @ home and hence the low crime rate.

Crimes and difference in the frequency:
Next visualization shows frequencies of all the crimes that are in the dataset. We inferred that:

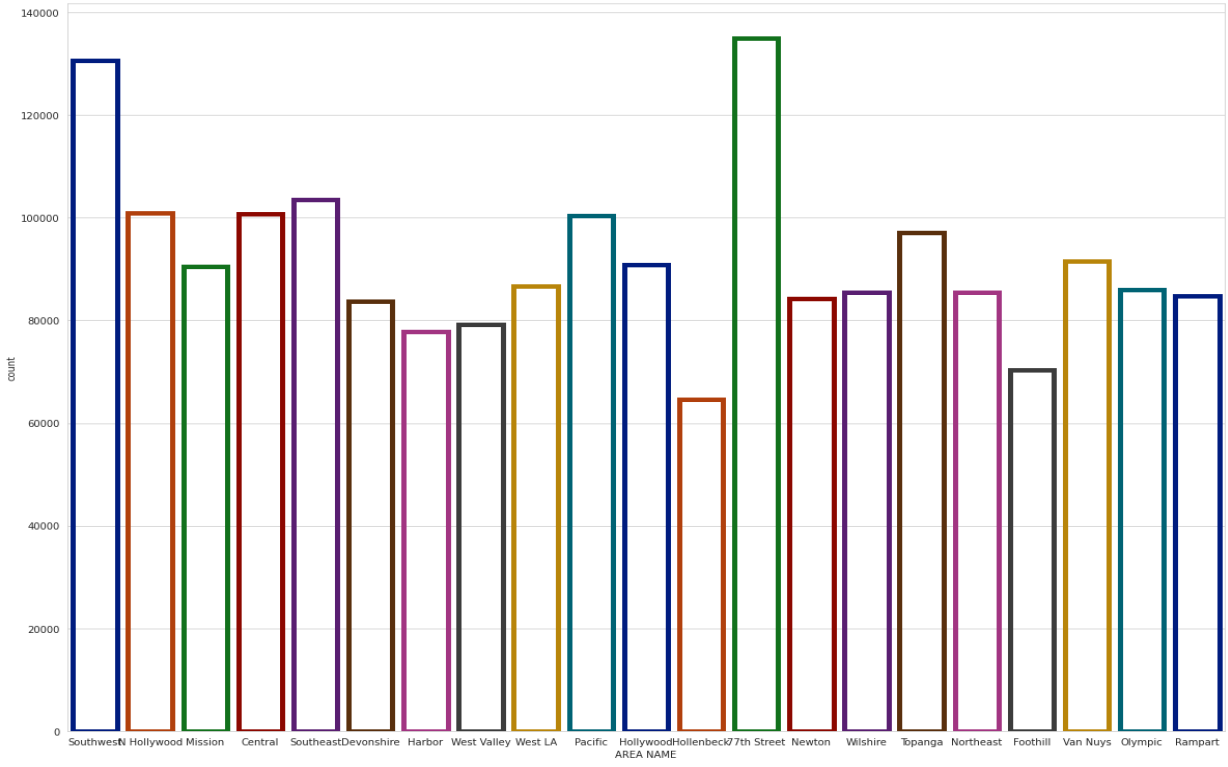
1) Battery Assault is most prevalent in LA county and its been on the rise.
2) Vehicle related crimes are behind Battery.
3) Cruelty to animals is among the least occurring crime


BATTERY - SIMPLE ASSAULT
VANDALISM - MISDEAMEANOR ($399 OR UNDER)
VANDALISM - FELONY ($400 & OVER, ALL CHURCH VANDALISMS)
RAPE, FORCIBLE
SHOPLIFTING - PETTY THEFT ($950 & UNDER)
OTHER MISCELLANEOUS CRIME
THEFT-GRAND ($950.01 & OVER)EXCPT,GUNS,FOWL,LIVESTK,PROD
CRIMINAL THREATS - NO WEAPON DISPLAYED
BURGLARY FROM VEHICLE
ARSON
VEHICLE - STOLEN
THEFT PLAIN - PETTY ($950 & UNDER)
ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT
ROBBERY
BRANDISH WEAPON
BURGLARY
BATTERY POLICE (SIMPLE)
ORAL COPULATION
INTIMATE PARTNER - SIMPLE ASSAULT
THEFT FROM MOTOR VEHICLE - PETTY ($950 & UNDER)
THEFT OF IDENTITY
THEFT, PERSON
BATTERY WITH SEXUAL CONTACT
BIKE - STOLEN
TRESPASSING
THEFT FROM MOTOR VEHICLE - GRAND ($400 AND OVER)
THEFT FROM MOTOR VEHICLE - ATTEMPT
THROWING OBJECT AT MOVING VEHICLE
BUNCO, GRAND THEFT
INTIMATE PARTNER - AGGRAVATED ASSAULT
ATTEMPTED ROBBERY
OTHER ASSAULT
BOMB SCARE
DOCUMENT FORGERY / STOLEN FELONY
CRM AGNST CHLD (13 OR UNDER) (14-15 & SUSP 10 YRS OLDER)
SEXUAL PENETRATION W/FOREIGN OBJECT
VIOLATION OF COURT ORDER
SHOTS FIRED AT INHABITED DWELLING
BURGLARY, ATTEMPTED
CHILD ABUSE (PHYSICAL) - SIMPLE ASSAULT
VIOLATION OF RESTRAINING ORDER
LETTERS, LEWD - TELEPHONE CALLS, LEWD
SHOPLIFTING-GRAND THEFT ($950.01 & OVER)
INDECENT EXPOSURE
VIOLATION OF TEMPORARY RESTRAINING ORDER
BUNCO, PETTY THEFT
KIDNAPPING - GRAND ATTEMPT
RESISTING ARREST
DISCHARGE FIREARMS/SHOTS FIRED
EMBEZZLEMENT, GRAND THEFT ($950.01 & OVER)
KIDNAPPING
LEWD CONDUCT
DEFRAUDING INNKEEPER/THEFT OF SERVICES, $400 & UNDER
PURSE SNATCHING
STALKING
SODOMY/SEXUAL CONTACT B/W PENIS OF ONE PERS TO ANUS OTH
VEHICLE - MOTORIZED SCOOTERS, BICYCLES, AND WHEELCHAIRS
CHILD NEGLECT (SEE 300 W.I.C.)
CONTEMPT OF COURT
BUNCO, ATTEMPT
SEX,UNLAWFUL(INC MUTUAL CONSENT, PENETRATION W/ FRGN OBJ
COUNTERFEIT
HUMAN TRAFFICKING - COMMERCIAL SEX ACTS
EXTORTION
THEFT PLAIN - ATTEMPT
RAPE, ATTEMPTED
SHOPLIFTING - ATTEMPT
FALSE IMPRISONMENT
BURGLARY FROM VEHICLE, ATTEMPTED
DISTURBING THE PEACE
FAILURE TO YIELD
THREATENING PHONE CALLS/LETTERS
VEHICLE - ATTEMPT STOLEN
EMBEZZLEMENT, PETTY THEFT ($950 & UNDER)
UNAUTHORIZED COMPUTER ACCESS
CREDIT CARDS, FRAUD USE ($950 & UNDER
PICKPOCKET
DISHONEST EMPLOYEE - GRAND THEFT
THEFT FROM PERSON - ATTEMPT
SHOTS FIRED AT MOVING VEHICLE, TRAIN OR AIRCRAFT
CHILD STEALING
PROWLER
CHILD ANNOYING (17YRS & UNDER)
SEX OFFENDER REGISTRANT OUT OF COMPLIANCE
CREDIT CARDS, FRAUD USE ($950.01 & OVER)
ASSAULT WITH DEADLY WEAPON ON POLICE OFFICER
CRIMINAL HOMICIDE
CHILD PORNOGRAPHY
LEWD/LASCIVIOUS ACTS WITH CHILD
PEEPING TOM
BATTERY ON A FIREFIGHTER
TILL TAP - PETTY ($950 & UNDER)
PIMPING
DEFRAUDING INNKEEPER/THEFT OF SERVICES, OVER $400
FALSE POLICE REPORT
PETTY THEFT - AUTO REPAIR
CHILD ABUSE (PHYSICAL) - AGGRAVATED ASSAULT
PANDERING
DISRUPT SCHOOL
CONSPIRACY
DOCUMENT WORTHLESS ($200 & UNDER)
DRIVING WITHOUT OWNER CONSENT (DWOC)
RECKLESS DRIVING
BOAT - STOLEN
DRUNK ROLL
DOCUMENT WORTHLESS ($200.01 & OVER)
GRAND THEFT / INSURANCE FRAUD
CRUELTY TO ANIMALS
PICKPOCKET, ATTEMPT
DRUGS, TO A MINOR
CHILD ABANDONMENT
THEFT, COIN MACHINE - PETTY ($950 & UNDER)
ILLEGAL DUMPING
BIKE - ATTEMPTED STOLEN
FIREARMS RESTRAINING ORDER (FIREARMS RO)
WEAPONS POSSESSION/BOMBING
CONTRIBUTING
DISHONEST EMPLOYEE - PETTY THEFT
HUMAN TRAFFICKING - INVOLUNTARY SERVITUDE
BRIBERY
MANSLAUGHTER, NEGLIGENT
GRAND THEFT / AUTO REPAIR
REPLICA FIREARMS(SALE,DISPLAY,MANUFACTURE OR DISTRIBUTE)
THEFT, COIN MACHINE - ATTEMPT
THEFT, COIN MACHINE - GRAND ($950.01 & OVER)
PURSE SNATCHING - ATTEMPT
LYNCHING
LYNCHING - ATTEMPTED
TILL TAP - GRAND THEFT ($950.01 & OVER)
TELEPHONE PROPERTY - DAMAGE
BEASTIALITY, CRIME AGAINST NATURE SEXUAL ASSLT WITH ANIM
FIREARMS EMERGENCY PROTECTIVE ORDER (FIREARMS EPO)
INCEST (SEXUAL ACTS BETWEEN BLOOD RELATIVES)
DRUNK ROLL - ATTEMPT
TILL TAP - ATTEMPT
INCITING A RIOT
BIGAMY
FAILURE TO DISPERSE
ABORTION/ILLEGAL
DISHONEST EMPLOYEE ATTEMPTED THEFT
TRAIN WRECKING
BLOCKING DOOR INDUCTION CENTER
FIREARMS TEMPORARY RESTRAINING ORDER (TEMP FIREARMS RO)
Crm Cd Desc
0
50000
100000
150000
200000
250000
count

On average for the last 10 years, **January** is the most violent time of the year. The reason being most of people are enjoying their New Year's and Christmas holidays so does criminals.

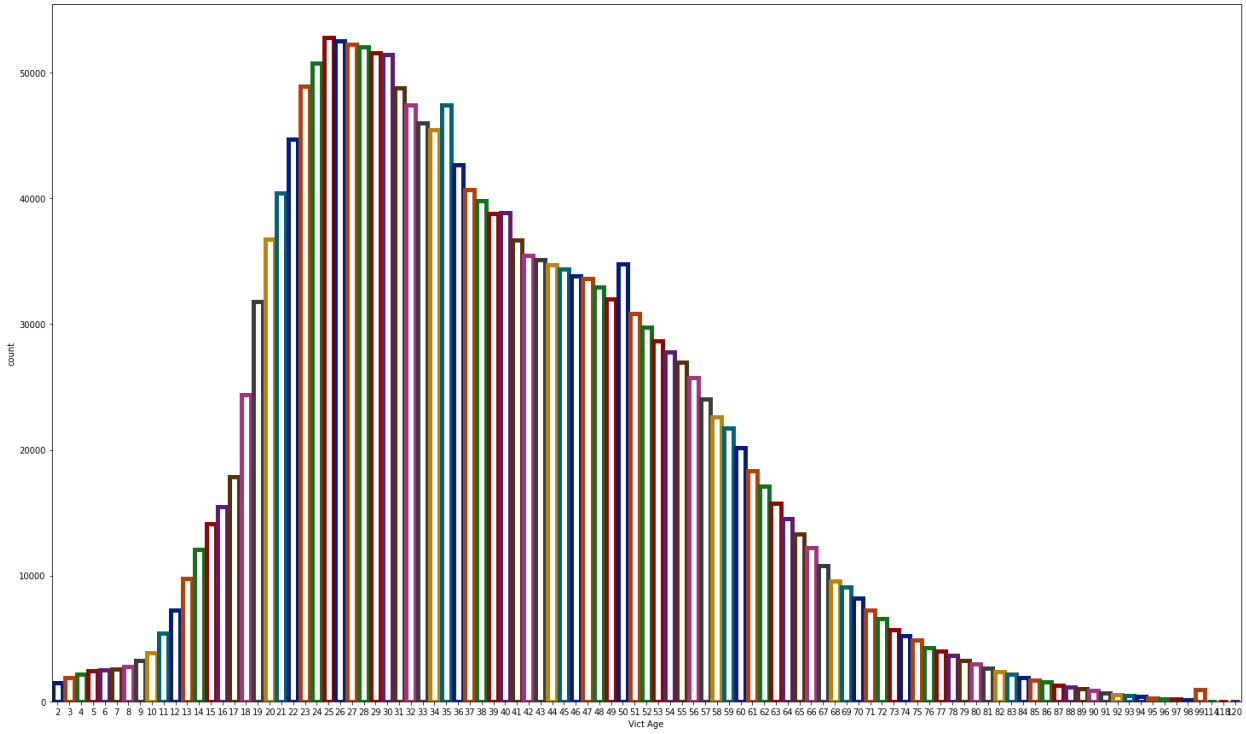


LA divisions most violent street is the 77th as most crimes occur here. **South LA** is also notorious for GangWars.



Victims Age:

Age wise crimes are mostly against people from 20-50 years of age as they are also the people which are in earning position.

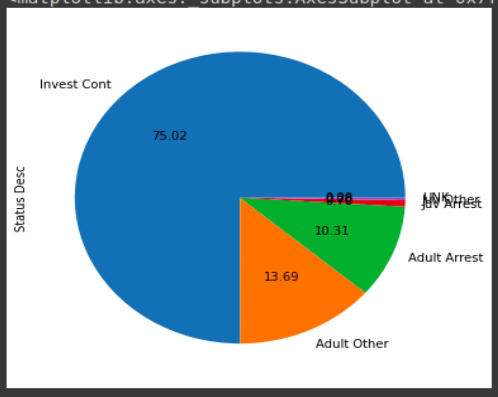


Crimes Case Status:

Status of Cases

```
[33]  df['Status Desc'].value_counts().plot(kind = 'pie',autopct='%.2f',figsize=(6,6))
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f1442ad1c50>
```

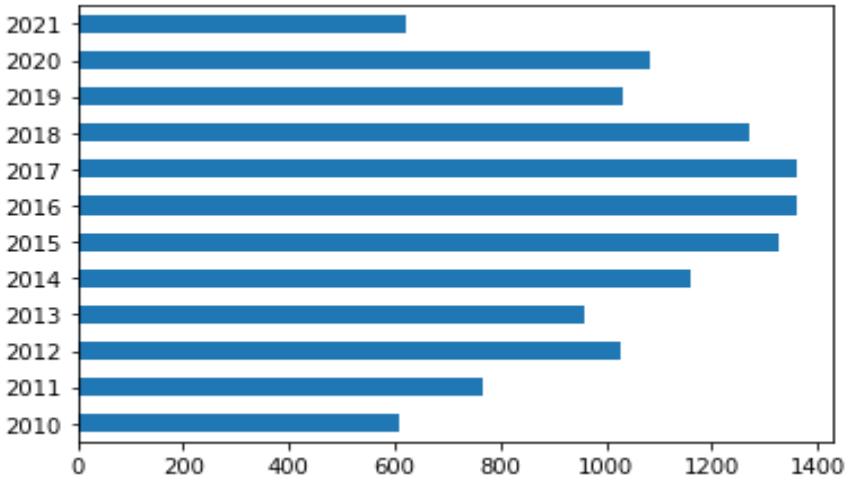



```
for status in df['Status Desc'].value_counts().keys():
    temp = df[df['Status Desc'] == status]['Crm Cd Desc'].value_counts()
    print("Most common crime with {} is {}".format(status,temp.keys()[0]))
```

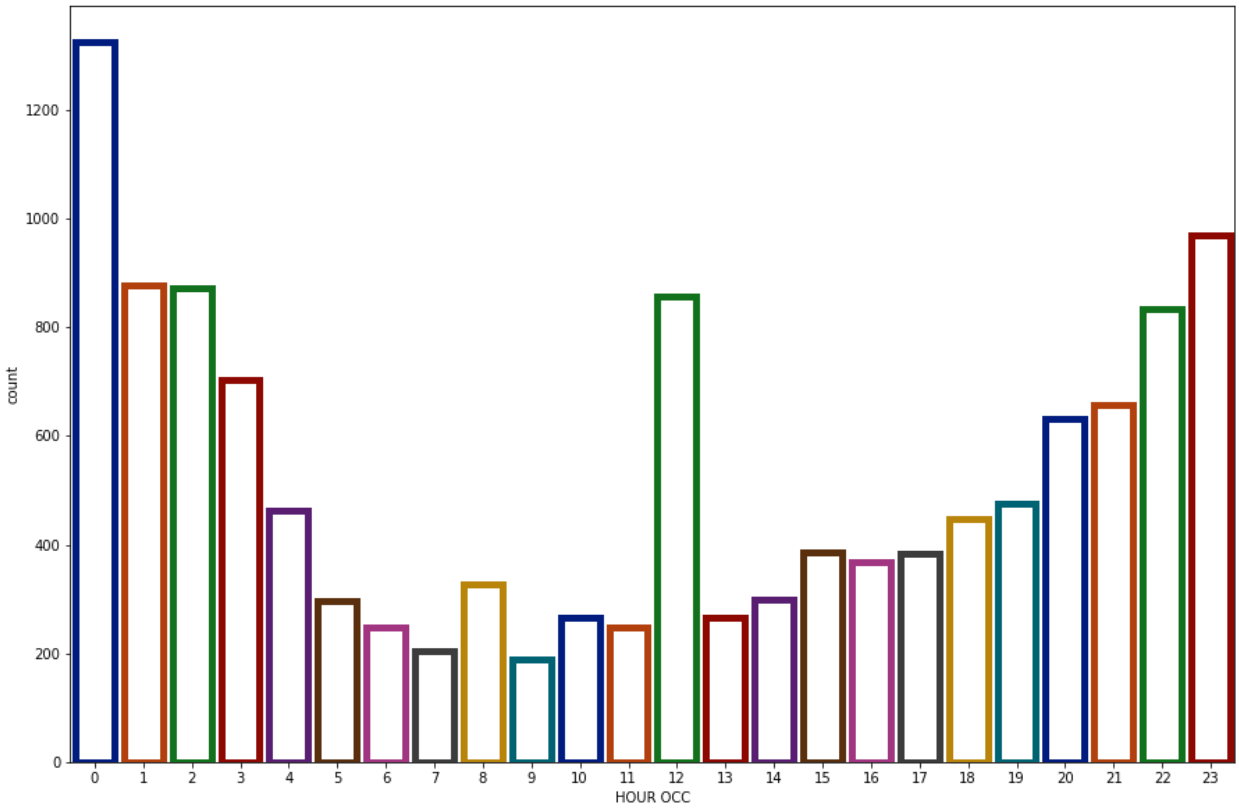
```
Most common crime with Invest Cont is BURGLARY FROM VEHICLE
Most common crime with Adult Other is BATTERY - SIMPLE ASSAULT
Most common crime with Adult Arrest is INTIMATE PARTNER - SIMPLE ASSAULT
Most common crime with Juv Arrest is BATTERY - SIMPLE ASSAULT
Most common crime with Juv Other is BATTERY - SIMPLE ASSAULT
Most common crime with UNK is THEFT PLAIN - PETTY ($950 & UNDER)
```

Rape Data Stats:

Rape cases were on the rise but COVID slowed its pace and hope it gets lower from there as well. The reduction can be attributed to the fact that mostly people were not allowed out of the house.



Most frequent time:



According to the data provided 12 @ midnight is the most frequent for RAPE, so LAPD has to deploy more people at night for the safety of population.

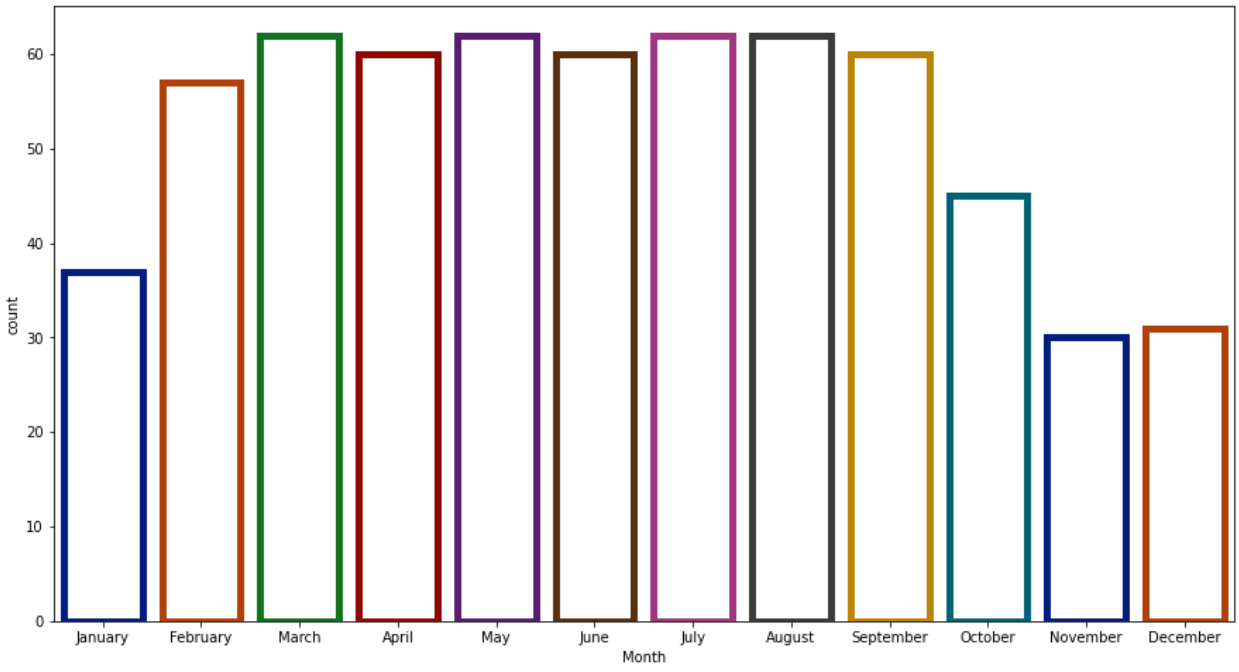
COVID DATA:

COVID is comprised of 3 files, but only county file was required so we loaded that and extracted all the data related to **Los Angeles**

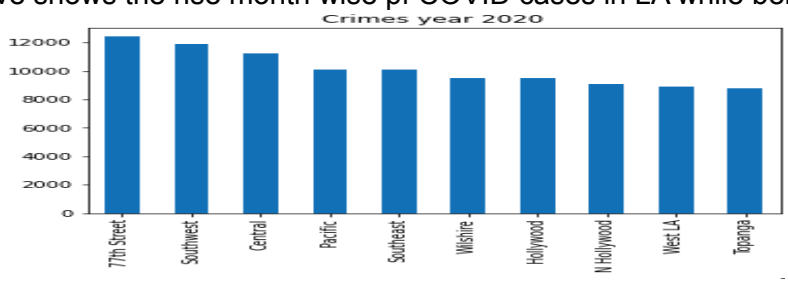
```
[52]  LACovidDF
```

	date	county	state	fips	cases	deaths	Day	Month	Year
9	2020-01-26	Los Angeles	California	6037.0	1	0.0	Sunday	January	2020
14	2020-01-27	Los Angeles	California	6037.0	1	0.0	Monday	January	2020
19	2020-01-28	Los Angeles	California	6037.0	1	0.0	Tuesday	January	2020
24	2020-01-29	Los Angeles	California	6037.0	1	0.0	Wednesday	January	2020
29	2020-01-30	Los Angeles	California	6037.0	1	0.0	Thursday	January	2020
...	...	...	...	...	...	...	...	...	...
1800572	2021-10-10	Los Angeles	California	6037.0	1471533	26312.0	Sunday	October	2021
1803821	2021-10-11	Los Angeles	California	6037.0	1472349	26338.0	Monday	October	2021
1807072	2021-10-12	Los Angeles	California	6037.0	1473123	26346.0	Tuesday	October	2021
1810321	2021-10-13	Los Angeles	California	6037.0	1474114	26362.0	Wednesday	October	2021
1813570	2021-10-14	Los Angeles	California	6037.0	1475222	26379.0	Thursday	October	2021

We inferred the last columns for the Day,Month and Year extracted from date.



This graph above shows the rise month wise pf COVID cases in LA while below we can see the drop in crimes .



This below is a map with overlaying points where crime is more prevalent