

# Crime in LA

Jack Liu, Ellie Xie

Summary of research questions and results:

1. Due to safety concerns when people are doing their daily activities, we want to investigate and find out what time period has the highest crime rate in LA based on the data.
  - a. By analyzing the data, we found that the time period has the highest crime rate in LA is 12 p.m.. However, the crime rate still remains high until midnight.
2. Due to safety concerns, we want to investigate how the crimes are distributed among different places so that people can have more awareness when they perform daily activities there.
  - a. By analyzing the data, we found that the top 5 places that crime took place are street, single family dwelling, multi-unit dwelling, parking lot, and other business places.
3. We concern if there is any relationship between the victims' age and sex with the crime rate? We want to analyze the certain type of people that are more easily targeted by crime.
  - a. By analyzing the data, we found that 35 years male are most likely being targeted.
4. We are concerned about what types of weapons that were being used in crimes, so that people could be more prepared and aware.
  - a. By analyzing the data, we found that the most used weapons are strong-arm, verbal threat, hand gun, semi-automatic pistol, and knife with blade.

Motivation:

We care about the topic “Crime in LA” because safety is always the priority concern of society, and it is related to every citizen. However, as we all know that it is impossible to eliminate the occurrence of crimes. Therefore, we want to use the crime data of LA to answer our research questions, and by providing visual graphs and summaries so that we could help citizens and government departments to be better prepared and aware during daily activities.

We care about the first question because as citizens, we always need to go out for different activities, such as schooling, working, shopping, and partying at night. However, people always don’t know what is the most dangerous time to go out statistically, therefore we want to provide the truth based on the dataset. With the result we found that the most dangerous time is not at midnight as we thought but at noon. We hope that the result could help people to be more aware and prepared even during the daytime, which could decrease the future crime rate.

We care about the second question because by understanding the most dangerous places in LA, citizens would know where they should go less frequently during daily activities, and be more aware and prepared if they still need to be there. Furthermore, the government should also have the idea of where they should strengthen administration to improve the environment's safety.

We cared about the third question because we worry if there is a certain type of people that are more likely to get targeted by crime? And with the data we want to find factors that contributed to this result. And by the analysis and conclusion that 35 years old male are most dangerous to be exposed to crime, so that they should be more aware and be prepared to protect themselves.

We cared about the fourth question because by analyzing the data of weapons used in crime and the distribution of weapons in different areas, people could also be more aware of the dangers of the surroundings and be prepared for emergencies.

Dataset:

The data we use is called “Los Angeles Crime Data” which shows committed crime in LA during 2020 to 2022. It included the criminal's gender, race, time that occurred, etc.

[https://www.kaggle.com/hemil26/crime-in-los-angeles?select=crime\\_in\\_la.csv](https://www.kaggle.com/hemil26/crime-in-los-angeles?select=crime_in_la.csv)

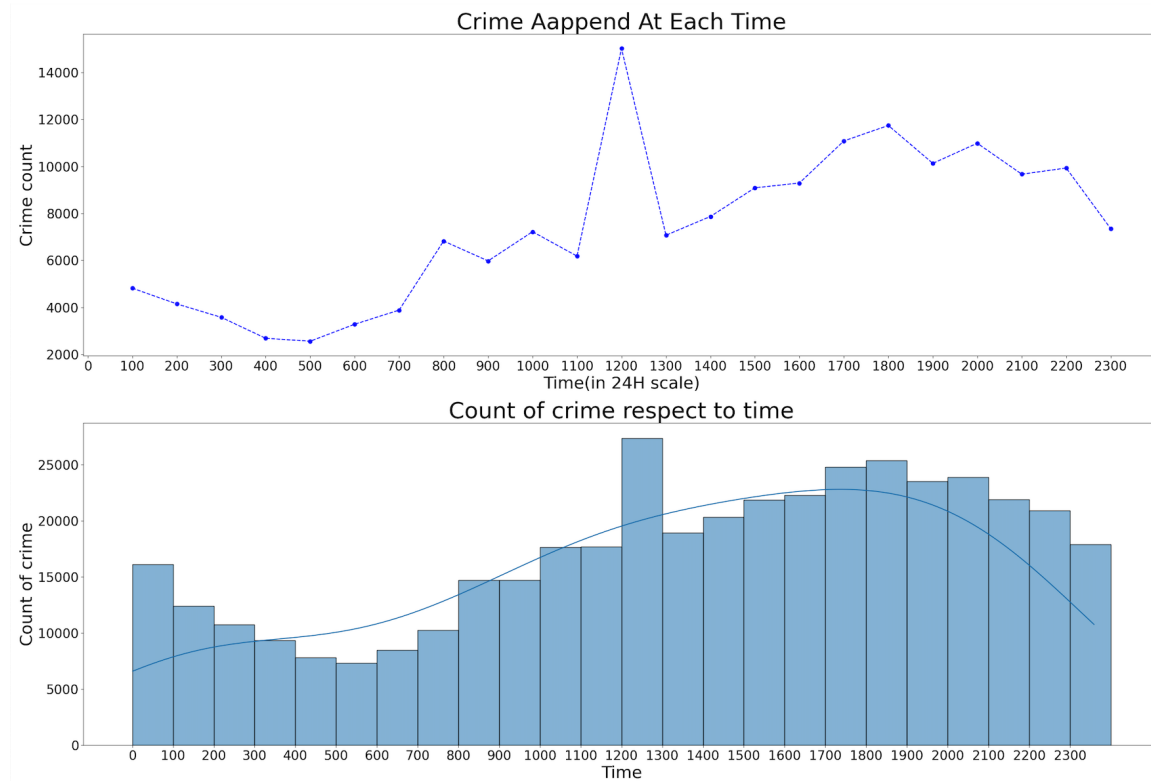
Method:

1. First use a jupyter notebook as a playground and do some data processing and to decide which part of data is useful.
2. I will groupby for the column "Time Occ" and from there use count to see how many cases happened in one year. Since time is in 24H (23:00) format, we need to convert it to numbers, and make it round up to integer because we want to see the time period in hours, then we can make a histogram with seaborn. From the histogram we can see what time period has the highest value.
3. Use column "Premis Desc" and groupby to count how many cases for each place, then make a histogram. Also use zip to put "LAT" and "LON" together to plot cases in the map.
4. First groupby the "Vict Sex" column to count and plot. Then, filter sex to be "M" then groupby the age and count, also do the same thing with "F". Then use plotly to plot the graph. Then we try to pass some feature like vict's sex, age, location and time to ML to predict crime type but result is not really satisfy.
5. We are trying to give top 10 to 15 weapons that were involved in the crime. Also plot them in the map to see the region that is more "dangerous" use plotly.
6. Find the top 7 of the crime type and plot with the location that involved those common crimes.

Result:

For the first research question, we were surprised by the result that 12 p.m. has the highest crime rate in LA. We were expecting to see that night time has a higher crime rate than day time, but the result shows an opposite outcome with our expectation. However, I think there are few factors contributing to this result: 1. People are more likely to go out during daytime, therefore there are more conflicts during the daytime than night time; 2. People are more alert when they go out at night than in the daytime, so there are more chances for criminals to commit a crime during day time; 3.

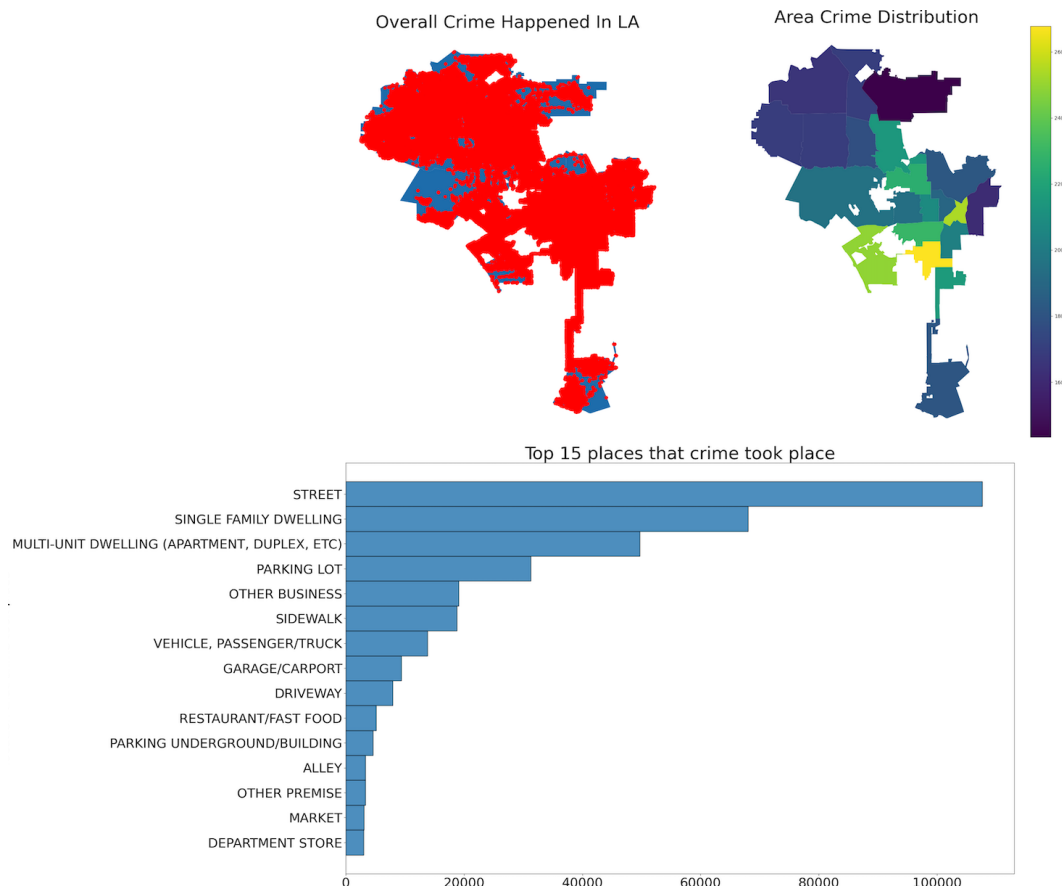
It might be difficult to report a criminal immediately to the police during the night time due to darkness; 4. Criminals find it more difficult to target victims during night time since there are not many people walking around at night, and criminals also need to take a rest at night.



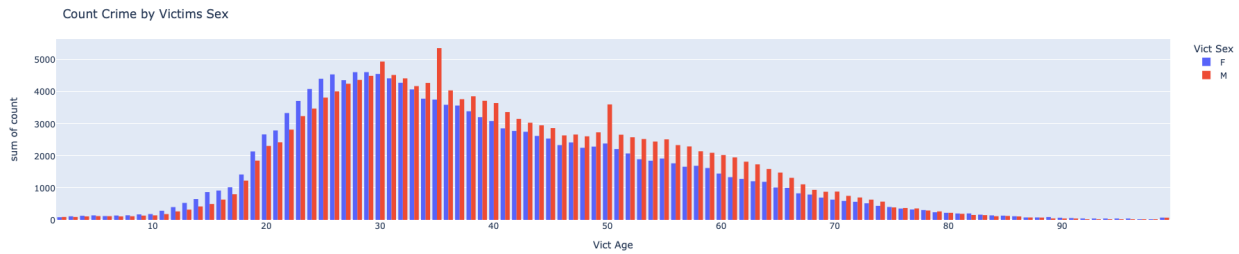
For the second research question, we found it important to note the result that the top 5 places that crime took place are street, single family dwelling, multi-unit dwelling, parking lot, and other business places. We might be able to understand that streets, parking lots or other business places are dangerous since those are public locations, but why the houses that we live in, especially single family dwelling and multi-unit dwelling, are so dangerous. I think there are few reasons that could explain it:

1. Family issues, people are having conflicts inside the family that result in fights verbally or even physically;
2. The family has less power to fight for criminals when there is only a single grown up person in the house, and criminals are more likely to target single family dwelling. Reasons that criminals are also very likely to happen in multi-unit

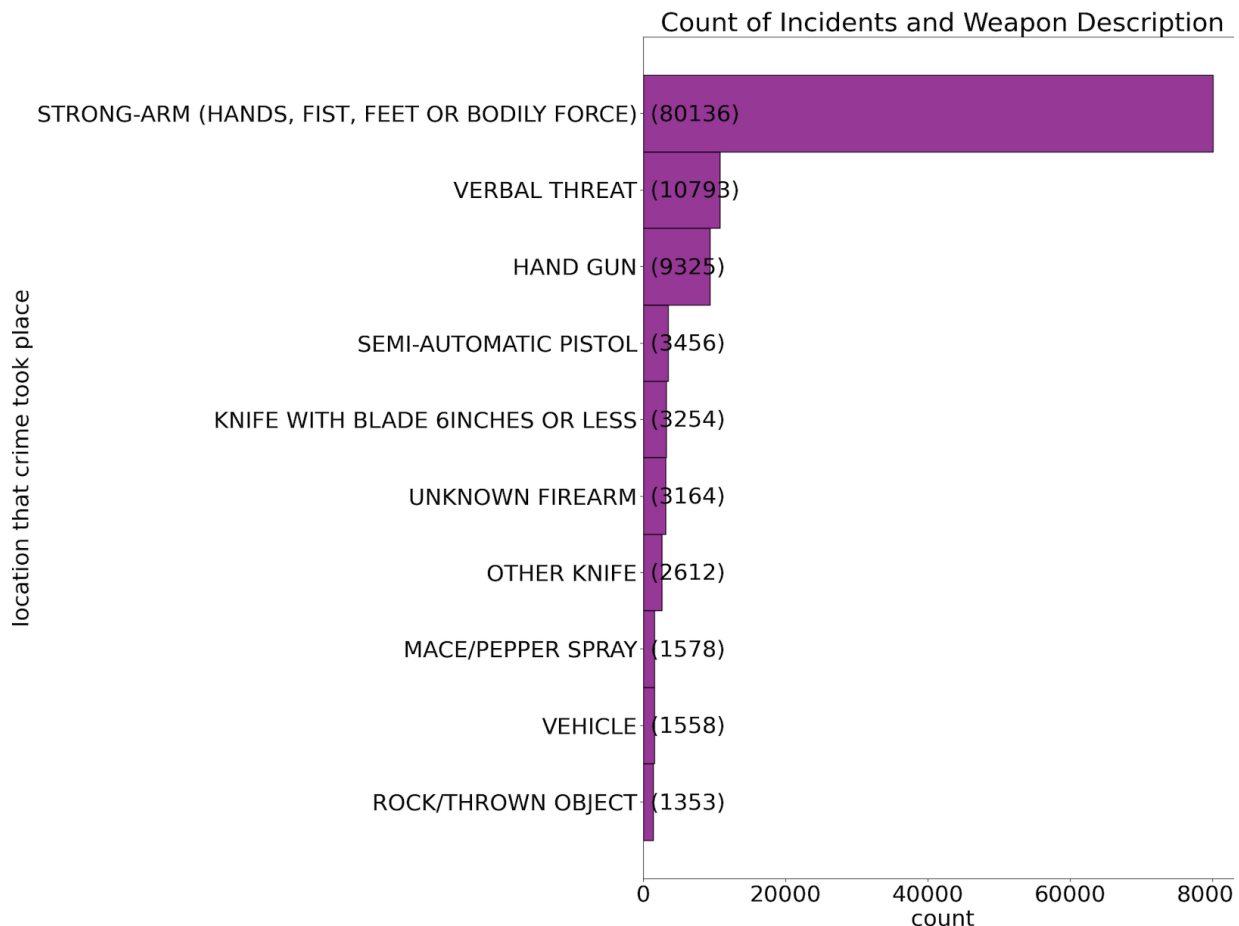
dwelling might be there are more conflicts occurring when people are living too close to each other, and conflicts would bring hate, and finally result in crimes.



For the third research question, we were surprised that 35 years male are most likely being targeted. We were expecting that female teenagers or elder people are more likely to be targeted since they don't have that much strength to fight back when they are being attacked. However the data shows an opposite result: that the strong mid-age male are more likely to be attacked. I think the explanation for this outcome might be: 1. There are more 35 years old males that need to go out to work, so there are more chances for them to be attacked; 2. Compared to female teenagers or elder people, a mid-age male is more likely to fight back against criminals instead of just running away, which might cause them to be injured.



For the fourth research question, we found that the most used weapons are strong-arm, verbal threat, hand gun, semi-automatic pistol, and knife with blade. We felt surprised by the result since we all expected that handguns and knives might be used the most during conflicts. However the result shows the strong-arm, such as hands, fist, feet or bodily force, are contributing the most during criminals. Some explanation to the result might be: 1. Many conflicts are not planned, so criminals didn't bring weapons with them when conflicts happened;. It is difficult for criminals to get access to weapons like guns and knives.



### Impact and Limitations:

As mentioned in the motivation section, we investigate the topic “Crime in LA” because we care about the safety issues in the society. Therefore, some implications we hope our research could bring might be that citizens could be more aware about some safety issues even during the day time regardless their sex or age, and the government could reinforce their public security level in locations like street, single family dwelling, multi-unit dwelling, parking lot, and other business places. We believe that our research would bring positive influence to the society, because citizens and government are going to benefit from our research since it would show them some ideas and suggestions that could keep them safe. We don’t think there are any biases in the data that might impact our results, and I think everyone could use our conclusion since the data set is reliable and it’s addressing the truth.

### Challenge goals:

During constructing our coding, we realized that our data is a bit hard to use, so that we need to pre process in order to use it. We separate the data processing part in different file named `data_processing.py`

Also to make our visualization better, we find a new data set contains all the geometric information about Los Angeles. So we decided to join two datasets together so that our analysis can have more sense. So we decided to join two datasets together so that our analysis can have more sense. Some notes while we were writing code: we found out that most geometries of a map come as shapefile, and we need to download the whole package including `.shx`, `.shp`, `.prj`, `.dbf`, `.cpg` files in order to let `geoPandas` read shape file. They should all have the same name beside file type and in the same folder, then we are able to use `gdp.read("file_name.shp")` to get all the geometry.

To make our graphs more clear and organized, we used a lot of attributes in the `plt` that did not talk in class, such as **`fig.add_gridspec`**, **`constrained_layout`**, **`ax.tick_params`**, and put numbers on histogram.

Furthermore, we found out that **ML** might not be the best way that could help us to analyze our data as we planned before, so we looked for and tried several models, and finally found out that prediction is way off from expectations. We think it might be because there is not much of a relationship between crimes or we did not find out the correct model yet. But learning a new model is really difficult so we decided to do something else to make our analysis better.

Finally we decided to use a new library: plotly to plot some of our graphs. We plotted the crimes committed in a real LA map, and found a data set about the LA map and combined our crime data with it. We also plotted the crimes committed in a real LA map. And to make the map better visualized, we give different depths of color and give a more general map.

#### Work Plan Evaluation:

The proposed work plan estimates are not really accurate. Since we do not have instruction on what to do, some easy tasks become really difficult since we need to look into data and find a way to represent something. We underestimate the difficulty of each part of the work. For example, when we want to plot a histogram, we need to decide how to filter data, how to adjust figures to look nice, how to set axes and how to set legend size. Each part sounds easy but it takes a long time to make the result as expected. Also ML does not really work as we expect, so we find some other challenging goals.

#### Test:

Mainly, we test our code in a jupyter notebook by plotting the data in a graph and to see if the graph has expected behavior. But we are just looking at general features because the data set is too large to tell if the number is exact. We use assert equal to test our data processing is correct and test a really small part of the data to see if numbers is correct

#### Collaboration:



We look up online a lot to figure out how to use ploty and try to find a good ML model. We also talk to grader TA for proposals to receive suggestions such as adding a dataset about maps.