Slide 1

Los Angeles Crime Rates 2020 to Present

Team:
Dusko
Eva
Shayla
Julie

Julie
Welcome everyone,
For our project, we have chosen to look at data related to crimes occurring in the greater city of Los Angeles in the US during 2020.

Your presenters for project group 1 tonight are:

- Eva who will talk about the overall number of crimes committed based on the Types of Crimes
- Dusko who will talk about the monthly distribution of those crimes as well as the resulting standard deviation along with some location-based analysis.
- Shayla's analysis will focus on the demographics around the victims of crime.
- And myself, Julie where I will talk about the averages around the number of days it takes to report a crime after it has occurred and then zero in on the top 3 for a bit of deeper analysis.

But first, Eva will start with the questions we sought to answer and what drove us to answer them…Eva

Slide 2

1. Questions that you found interesting and what motivated you to answer them

- Total number of crimes committed by Type of Crime
- Total number of crimes committed per month in 2020
- High crime areas of Los Angeles
- Victims of Crime - Men v Women
- The number of crimes committed by type of crime and age group
- Average days to report crimes
- Days range for top 3 types of crime averages

Eva

In a live business sense, you seek out data to answer specific questions, as well as make discoveries based on available data.

Our approach was to formulate questions based on discovery driven by the available data.

The questions formulated were:
What were the total number of crimes committed by Type of Crime and per month?

What areas of Los Angeles did the most crimes occur during 2020?

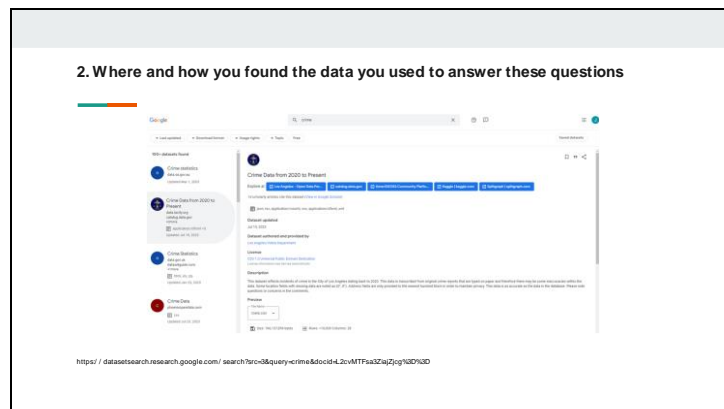Who is more likely to be a victim of crime? Men or Women?

What were the number of crimes committed by type of crime and age group?

What was the average number of days between the date the crime occurred and the date the crime was reported?
and,
What was the range of days between occurrence and reporting for the top 3 types of crimes with the largest average?
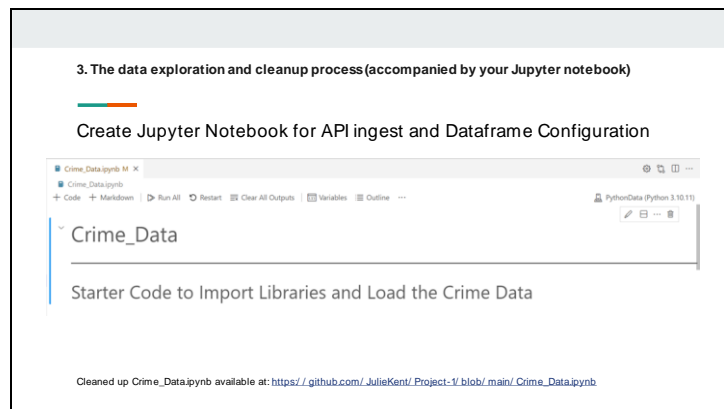
Slide 3

Eva
As mentioned earlier, our initial search for data was driven by what we could locate and access freely.

So, an initial google search of free API's led the team to the 'Crime Data from 2020 to Present' API.

Slide 4

3. The data exploration and cleanup process (accompanied by your Jupyter notebook)

Create Jupyter Notebook for API ingest and Dataframe Configuration

Crime_Data

Starter Code to Import Libraries and Load the Crime Data

Cleaned up Crime_Data.ipynb available at: https://github.com/JulieKent/Project-1/blob/main/Crime_Data.ipynb

Eva
Our first step was to set up the Jupyter notebook to facilitate the ingestion of the API data for configuration.

The cleaned up data notebook is available to view via the link on this slide.

Now, Dusko will run through some of the clean up process….Dusko

Slide 5

Dusko
Along with installing and importing various functions in the dependency set up, we also needed to sign up to the Los Angeles Open Data website to enable access to the API data.

Slide 6

3. The data exploration and cleanup process (accompanied by your Jupyter notebook)

Ingest and review JSON results

```
[
    {
        "area": "03",
        "area_name": "Southwest",
        "crm_cd": "624",
        "crm_cd_1": "624",
        "crm_cd_desc": "BATTERY - SIMPLE ASSAULT",
        "date_occ": "2020-01-08T00:00:00.000",
        "date_rptd": "2020-01-08T00:00:00.000",
        "dr_no": "010304408",
        "lat": "34.0141",
        "location": "1100 W 39TH                    PL",
        "lon": "-118.2978",
        "mocodes": "0444 0913",
        "part_1_2": "2",
        "premis_cd": "501",
        "premis_desc": "SINGLE FAMILY DWELLING",
        "rpt_dist_no": "0377",
        "status": "AO",
        "status_desc": "Adult Other",
        "time_occ": "2230",
        "vict_age": "36",
        "vict_descent": "B",
        "vict_sex": "F",
        "weapon_desc": "STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)",
        ...
        "vict_descent": "B",
        "vict_sex": "F"
    }
]
```
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings.

Dusko
The next step was to get the API results, print the results to see what fields were available and convert the JSON to a pandas dataframe.

Slide 7

3. The data exploration and cleanup process (accompanied by your Jupyter notebook)

Look for missing values

```
Date reported     10000
Date Occurred     10000
Time Occurred     10000
Type of Crime     10000
Area Name         10000
Victim Age        10000
Victim Gender      9139
Street No_Name    10000
lat               10000
lon               10000
dtype: int64
```

Dusko
A count of the column values assisted in the decision for how to treat missing values, while maintaining the integrity of the data for later analysis.

Now, Shayla will take you through the rest of the data clean up process....Shayla

Slide 8

3. The data exploration and cleanup process (accompanied by your Jupyter notebook)

Drop rows with missing data and view resulting dataframe

	Date reported	Date Occurred	Time Occurred	Type of Crime	Area Name	Victim Age	Victim Gender	Street No_Name	lat	lon
0	2020-01-08	2020-01-08	22:30	BATTERY - SIMPLE ASSAULT	Southwest	36	F	1100 W 39TH PL	34.0141	-118.2978
1	2020-01-02	2020-01-01	03:30	BATTERY - SIMPLE ASSAULT	Central	25	M	700 S HILL ST	34.0459	-118.2545
2	2020-04-14	2020-02-13	12:00	SEX OFFENDER REGISTRANT OUT OF COMPLIANCE	Central	0	X	200 E 6TH ST	34.0448	-118.2474
3	2020-01-01	2020-01-01	17:30	VANDALISM - MISDEAMEANOR ($399 OR UNDER)	N Hollywood	76	F	5400 CORTEEN PL	34.1685	-118.4019
4	2020-01-01	2020-01-01	04:15	VANDALISM - FELONY ($400 & OVER, ALL CHURCH VA...	Mission	31	X	14400 TITUS ST	34.2198	-118.4468

Shayla
At this point, we started to develop our questions based on the data we could see in each field.

Slide 9

3. The data exploration and cleanup process (accompanied by your Jupyter notebook)

Consolidate Types of Crime names of similar value

	Date reported	Date Occurred	Time Occurred	Type of Crime	Area Name	Victim Age	Victim Gender	Street No_Name	lat	lon
0	2020-01-08	2020-01-08	22:30	ASSAULT	Southwest	36	F	1100 W 39TH PL	34.0141	-118.2978
1	2020-01-02	2020-01-01	03:30	ASSAULT	Central	25	M	700 S HILL ST	34.0459	-118.2545
2	2020-04-14	2020-02-13	12:00	OTHER MISCELLANEOUS CRIME	Central	0	X	200 E 6TH ST	34.0448	-118.2474
3	2020-01-01	2020-01-01	17:30	VANDALISM	N Hollywood	76	F	5400 CORTEEN PL	34.1685	-118.4019
4	2020-01-01	2020-01-01	04:15	VANDALISM	Mission	31	X	14400 TITUS ST	34.2198	-118.4468
...	...	...	...	...	...	...	...	...	...	...
9995	2020-11-27	2020-11-27	12:04	THEFT	Wilshire	70	M	W OLYMPIC BL	34.0582	-118.3639
9996	2020-09-21	2020-08-30	18:00	THEFT	Harbor	69	F	1200 W 27TH ST	33.7144	-118.2997
9997	2020-12-11	2020-11-16	08:00	THEFT	Devonshire	76	F	9700 AQUEDUCT AV	34.2464	-118.474
9998	2020-10-17	2020-10-15	10:05	THEFT	Northeast	64	M	2700 W AVENUE 34	34.1137	-118.2353
9999	2020-12-25	2020-12-24	21:45	THEFT	Central	30	F	100 N JUDGE JOHN AISO ST	34.0504	-118.2409

Shayla
We decided that it would be of greater value to base the analysis on the groupings of types of crime as opposed to the entire raw list.

So, we worked to export a list of the unique types of crime, make a shared judgement of their grouping, replace the old names with our newly defined group names, and lastly, create a new resulting dataframe.

Slide 10

4. The analysis process (accompanied by your Jupyter notebook)

Divide and conquer

📓 Dusko.ipynb
📓 Eva.ipynb
📓 Julie.ipynb
📓 Shayla.ipynb

Consolidated Crime_Analysis.ipynb available at: https://github.com/JulieKent/Project-1/blob/main/Crime_Analysis.ipynb

Shayla
We each started with a separate notebook to develop our plots to conduct our target analysis, each using the import jupyter notebook function.

This is to ensure we were all drawing on the same, cleaned up dataframe without the need for repetition of code.

Eva will be kicking off first with her plot and analysis…Eva

Slide 11

4. The analysis process (accompanied by your Jupyter notebook)

	Type of Crime	count
0	ARSON	20
1	ASSAULT	1504
2	BOMB SCARE	10
3	BUNCO, ATTEMPT	3
4	CHILD RELATED	35
5	CONTEMPT OF COURT	3
6	COUNTERFEIT	2
7	CRIMINAL HOMICIDE	6
8	CYBER CRIME	7
9	DISCHARGE FIREARMS	128

	Type of Crime	count
0	THEFT	4578
1	ASSAULT	1504
2	ROBBERY	1213
3	VANDALISM	722
4	IDENTITY THEFT	294
5	THREAT	154
6	DISCHARGE FIREARMS	128
7	OTHER MISCELLANEOUS CRIME	125
8	TRESPASSING	98
9	RESTRAINING ORDER	94

Eva
The first step, was to Group by "Type of Crime" and, count occurrences followed by, sorting the count in descending order to establish the right cut of data to utilise.

A cut of the top 14 was representative of crimes with a count greater than 20.

Nex, we have Dusko to talk about his analysis and plots….Dusko

Slide 12

4. The analysis process (accompanied by your Jupyter notebook)

- Calculate the Crime Incidents by month and number of crimes
- Calculate standard deviation
- Bar plot to visualise results

Dusko
The first step in my analysis was to create a simple, Crime Incidents by month.

Following that, the data was grouped by the Number of Crimes.

I included a standard deviation calculation with the plot as well.

Slide 13

4. The analysis process (accompanied by your Jupyter notebook)

- Calculate number of crimes by area name and number of crimes.
- Heatmap to visualise results

Dusko
The next step was to calculate the number of crimes by area.

Finally, the heatmap was created to visualise the results.

Up next is Shayla to talk through her analysis approach…Shayla

Slide 14

4. The analysis process (accompanied by your Jupyter notebook)

- Drop values that were not representative of victim gender; i.e. 'X', 'H' values
- Create pivot table to separate 'F' and 'M' columns to show number of victims per gender per crime
- Plot bar graph for visualisation

Shayla
The first step in creating my graph that seeks to explore the relationship of victim gender and type of crime, was to drop values that are not relevant to the analysis.

The number of victims per gender, per type of crime was determined using a pivot table. To produce a concise graph, the focus is on the six types of crime that comprise at least 50 victims for each gender.

Slide 15

Shayla
In order to analyse the relationship between victim age and type of crime, it was necessary to create bins for age groups for victim age, as well as a count for how many victims are in each group for each type of crime.

Similar to the first analysis, the focus was, types of crime with at least 70 victims where the intention of excluding certain data groups is to shift focus to types of crime that are most impactful.

The bins created were then converted into a pivot table in order to visualise the data.

Now, Julie will go through her analysis process…Julie

Slide 16

Julie
In determining the number of days it took to report each type of crime, the first step was to create a simple date difference calculation.

Following that, the data was grouped by the Type of Crime and the number of days were averaged for each type of crime.

There were a number of crimes that had quite a small count, so I arbitrarily filtered the results for a count per type of crime above 10 results.

Slide 17

4. The analysis process (accompanied by your Jupyter notebook)

- Create a filter for the Top 3 averages
- Calculate the IQR to determine if there are any potential outliers
- Plot the results in a Box Plot

Julie
The next analysis was largely driven by the results of my first analysis results.

So, the first step in the second analysis process was to create a filter of the Type of Crime, to limit the dataframe to the desired list, being the top 3 averages.

Wanting to look deeper at the days difference for these specific crimes, the next step was to create a list for plotting and calculate the interquartile range to determine if there are any potential outliers.

Finally, the boxplot was generated to visualise the results.

Eva will now take you through her analysis results….Eva

Slide 18
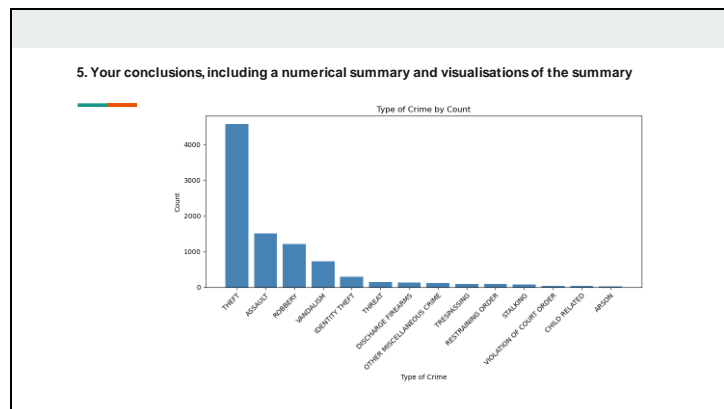
Eva
Theft was by far the most common crime committed with a count of 4578.

This included crimes like, Grand Theft (stealing property valued at >= $300), Motor Vehicle Theft, Embezzlement (misappropriation of funds), Attempted Theft etc.

The next three types of crime; Assault, Robbery and Vandalism, have significantly lower counts.

These four categories seem to represent the bulk of crimes committed in 2020.

Dusko will now talk about the results around the crimes by month and area….Dusko
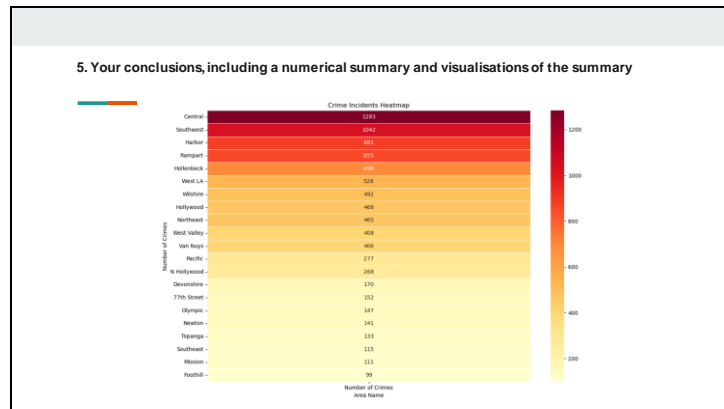
Slide 19
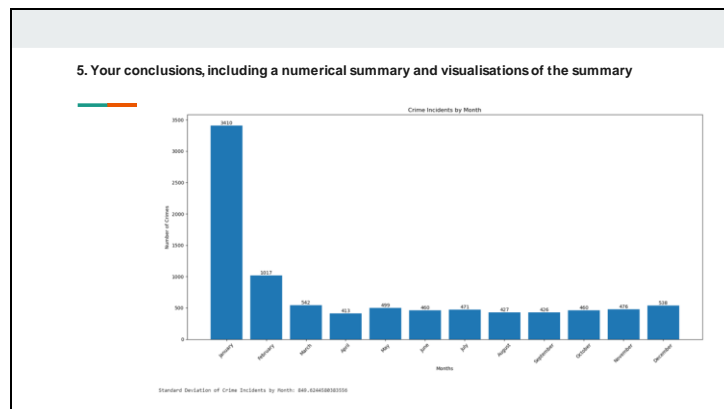
Dusko
Plotting out the overall number of crimes that occurred per month over the course of 2020 resulted in a somewhat interesting, and kind of expected result given the external environmental factors happening at the same time.

What we see in the chart is a significant drop in the number of crimes after January which, given the world had at the time, commenced COVID lockdowns, it could easily be seen as a key contributor to the sharp reduction.

However, additional data and analysis would be required to substantiate this assumption.

The standard deviation came out at approximately 848.90, which means that the data points are quite spread out from the mean. The significant drop in the number of crimes from February onwards, goes a long way to explain the result.

Slide 20

5. Your conclusions, including a numerical summary and visualisations of the summary
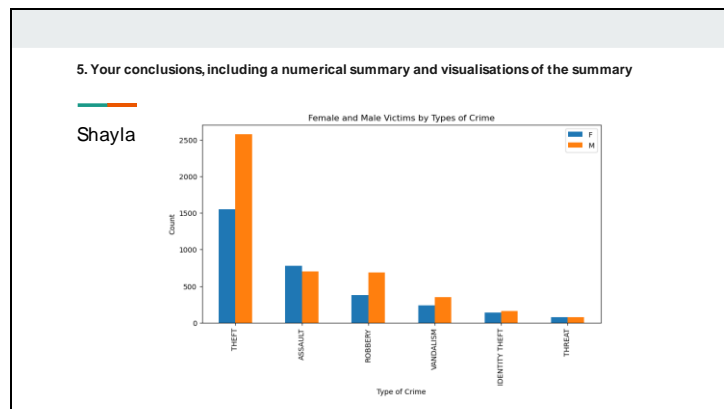
Dusko
The crime rates are not evenly distributed across Los Angeles.

Some areas have significantly higher crime rates than others.

Crime rates can vary widely across different neighbourhoods in Los Angeles.

Central, Southwest, Harbor and Rampart have experienced significantly higher crime rates than most other areas of greater Los Angeles.

Shayla will now talk about her analysis of the victim demographics…Shayla

Slide 21

5. Your conclusions, including a numerical summary and visualisations of the summary

Shayla

Female and Male Victims by Types of Crime

Shayla
The focus is on the relationship between the type of crime and the victim's sex.

Theft comprises a significantly larger number of male victims (2584) in comparison to female victims (1557).

Interestingly, the majority of these crimes show that male victims are more prevalent than female victims. Except for the case of assault; this is the only crime category where the number of female victims is higher.

This dataset is not as extensive as to provide an explanation for this, which is a potential area for further analysis.

5. Your conclusions, including a numerical summary and visualisations of the summary
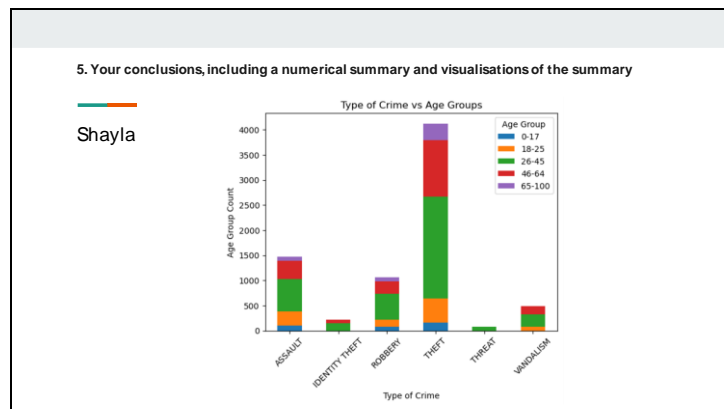
Shayla
This graph demonstrates the number of victims per age group according to which type of crime.

The crime of theft comprises of the largest number of victims, with the adult age group of 26-45 being the primary target; this age group also appears to be the primary target amongst other crimes.

Looking at identity theft, it is observed that this crime has victimised only two age groups; 26-45 and 46-64 for this year of data.

Coincidentally, these two age groups also represent the majority of the victims across all crimes in comparison to other age groups.

Julie will now take you through the analysis of the days to report crimes….Julie
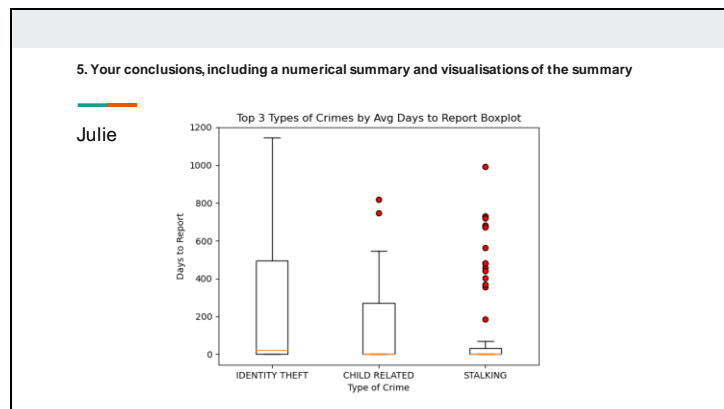
Slide 23
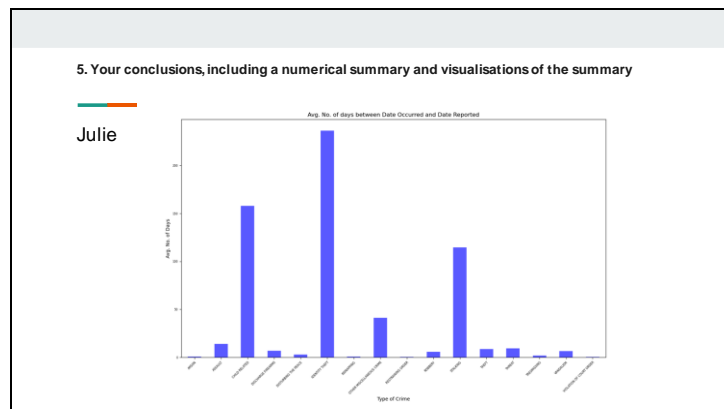
Julie
When plotting the average number of days between when a crime occurred, and when it was reported by the Type of Crime, there are 3 stand out results;

Child Related, Identity Theft and Stalking.

Anecdotally, identity theft and stalking are thought to generally happen over a longer period of time so, we assume that there would be a greater number of days between when it started and when it is eventually reported.

However, in our efforts to consolidate similar crimes based on our own interpretations, the results should be reviewed and potentially re-cut, particularly in relation to child related crimes to get a more accurate result.

Slide 24

5. Your conclusions, including a numerical summary and visualisations of the summary

Julie

Top 3 Types of Crimes by Avg Days to Report Boxplot

Julie
Surprisingly though, the previously held concerns around the grouping of child related crimes appears to be unfounded as indicated by the box plot results, which show a concentration in the 3rd quartile with only a couple of outliers.

What's more interesting is the number of outliers in relation to stalking indicating a lack of consistency in the number of days and that the outliers can be seen as the primary contributing factor in the elevated average for the group. Challenging the previously held assumptions.

What we do see consistently in this plot, is the mean which sits quite low in all 3 types of crimes.

Slide 25

6. The implications of your findings: what do your findings mean?

- Data analysis used to drive efficient allocation of resources
- We have more questions
- We need more data

Julie
Conducting analysis like that of crime rates, would serve to enable local government (in the case of the US) to gain efficiency in the allocation of policing resources in areas that could drive better outcomes.

What we found is, the more we looked at the data, the more questions we had but also found how easy it was to start attributing results to factors not included in the data but very evident in known external factors.
As with any dataset, these external factors remind us to interpret the findings within a broader context and consider additional sources of information to attain a comprehensive understanding of crime dynamics in Los Angeles neighbourhoods.
While we have provided a very brief analysis of the data in this presentation, a more in depth analysis is available in our Crime_Analysis Jupyter Notebook.

Thanks for coming