
Did covid-19 impact crime in Los Angeles?

**Group 7: Pratik Khadse, Hsuan-Ting (Tommy) Wu,
Nishkarsh Khokhar, Falak Jain, Yi-Ching (Millie) Lin**

Agenda

1. Questions
2. Datasets and Data Cleaning Process
3. Data Visualization
4. Analytics Process
5. Conclusion

Questions

1. How did covid-19 impact crime in Los Angeles? What other factors could lead to this impact?

→ **Regression**

2. Whether Asians were more likely to become victims after covid-19 in Los Angeles?

→ **Hypothesis**

Answering the questions - Data driven approach

Assumption 1:

- Population, Income, Unemployment Rate, Location and Covid-19 may have impact on the crime in LA.

Assumption 2:

- Asians were more likely to become victims after covid-19 in Los Angeles.

Ideal Experiment: Question 1

- We have two worlds: one with covid, one without covid
- Our variables are unemployment rate, population, median income, location, etc
- Controlling for these variables, is the crime rate different for the two worlds

Ideal Experiment: Question 2

- We have two worlds: one with covid, one without covid
- Is the proportion of crimes against Asians different for the two worlds.

Dataset

1. Los Angeles Crime Data 2010-2021

- Data source: Kaggle
- This dataset reflects incidents of crime in the City of Los Angeles from 2010 - 2019 as well as from 2020-2021 in two separate data sets
- This data is transcribed from original crime reports that are typed on paper and therefore there may be some inaccuracies within the data
- Address fields are only provided to the nearest hundred block in order to maintain privacy

	DR_NO	DATE_RPTD	DATE_OCC	TIME_OCC	AREA_NAME	CRM_1_SEVERITY	VICT_AGE	VICT_SEX	PREMIS_DESC
0	1307355	2010-02-20	2010-02-20	2021-11-21 13:50:00	Newton	2	48	M	SINGLE FAMILY DWELLING
1	11401303	2010-09-13	2010-12-09	2021-11-21 00:45:00	Pacific	3	0	M	STREET
2	70309629	2010-09-08	2010-09-08	2021-11-21 15:15:00	Newton	1	0	M	ALLEY
3	90631215	2010-05-01	2010-05-01	2021-11-21 01:50:00	Hollywood	2	47	F	STREET
4	100100501	2010-03-01	2010-02-01	2021-11-21 21:00:00	Central	9	47	F	ALLEY

Data Cleaning

Step1. Los Angeles Crime Data 2010-2021

- Removed leading zeros and converted datetime fields to datetime datatype
- Segmented crimes according to severity based on crime code

```
## bucket
df["CRM_CD"] = pd.cut(df["CRM_CD"], [0, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000],
                        labels = [1,2,3,4,5,6,7,8,9,10], ordered = True)

## change dtype
df["CRM_CD"] = df["CRM_CD"].astype("int")

## subtract for severity
df["CRM_CD"] = 11 - df["CRM_CD"]
```

- Maintaining consistency between codes and descriptions

Data Cleaning

Step2. Zip code from longitude/latitude

- The original data set consisted of coordinates
- Used reverse geocoding via a Google API
- Obtained zip code details for every set of coordinates
- Sample address returned from Google API Call:
 - 164 W Jefferson Blvd, Los Angeles, CA 90007, USA
- Using string formatting, we extract the zip code from this address
- Merged the zipcodes on the dataset in order to merge demographic data

Data Cleaning

Step3. Zip Code / Area

- Issue: Crimes committed in same zip codes handled by different police departments (areas)
- Solution: Find area of zip code with highest number of cases and use that as the area zip code belongs to

Before

	zipcode	AREA_NAME	DR_NO
85	90013.0	Central	32916
86	90013.0	Hollenbeck	276
87	90013.0	Newton	1101
88	90013.0	Pacific	1

After

	zipcode	AREA_NAME	DR_NO
67	90011.0	Newton	62307
73	90012.0	Central	25298
85	90013.0	Central	32916
90	90014.0	Central	19316

Dataset

2. US Census

- Data source: United States Census Bureau
- The dataset includes Unemployment rate / Median household income / Population in each ZCTA zip code in LA county
- Years: 2011 - 2019, based on American Community Survey data
- 27 csv files total

GEO_ID	NAME	S1901_C01_001E	S1901_C01_001M	S1901_C01_002E	S1901_C01_002M	S1901_C01_003E	S1901_C01_003M	S1901_C01_004E	S1901_C01_004M
id	Geographic Area Name	Households!!Estimate!!Total	Households!!Margin of Error!!Total	Households!!Estimate!!Less than \$10,000	Households!!Margin of Error!!Less than \$10,000	Households!!Estimate!!10,000to14,999	Households!!Margin of Error!!10,000to14,999	Households!!Estimate!!15,000to24,999	Households!!Margin of Error!!15,000to24,999
1 86000000US89010	ZCTA5 89010	149	42	18.1	12.7	0.0	19.5	14.8	10.8
2 86000000US89019	ZCTA5 89019	942	261	0.0	3.4	9.3	11.3	20.3	12.5
3 86000000US89060	ZCTA5 89060	4354	357	7.2	3.4	8.3	3.2	16.4	4.8
4 86000000US89061	ZCTA5 89061	1919	256	3.0	2.9	5.0	4.0	9.7	4.8

Data Cleaning

Step1. Census Data

- Choose specific columns from data set:
 - Households!!Estimate!!Median income (dollars)
 - Percent!!EMPLOYMENT STATUS!!Percent Unemployed
 - Estimate!!SEX AND AGE!!Total population
- Added year for year file
- 2010/2020/2021: Estimated using average percentage for each zipcode from the 9 years
- Extracted zip code from original format
- Combine population/unemployment rate/income into one

	Year	zipcode	Total population	Unemployment_rate	Median household income
0	2021	90001	65097.0	12.752946	43479.587251
1	2021	90002	57906.0	14.563990	39262.055882
2	2021	90003	79074.0	15.484133	38708.277385
3	2021	90004	71737.0	11.570812	50607.463793
4	2021	90005	45200.0	11.673085	39496.138261

Data Cleaning

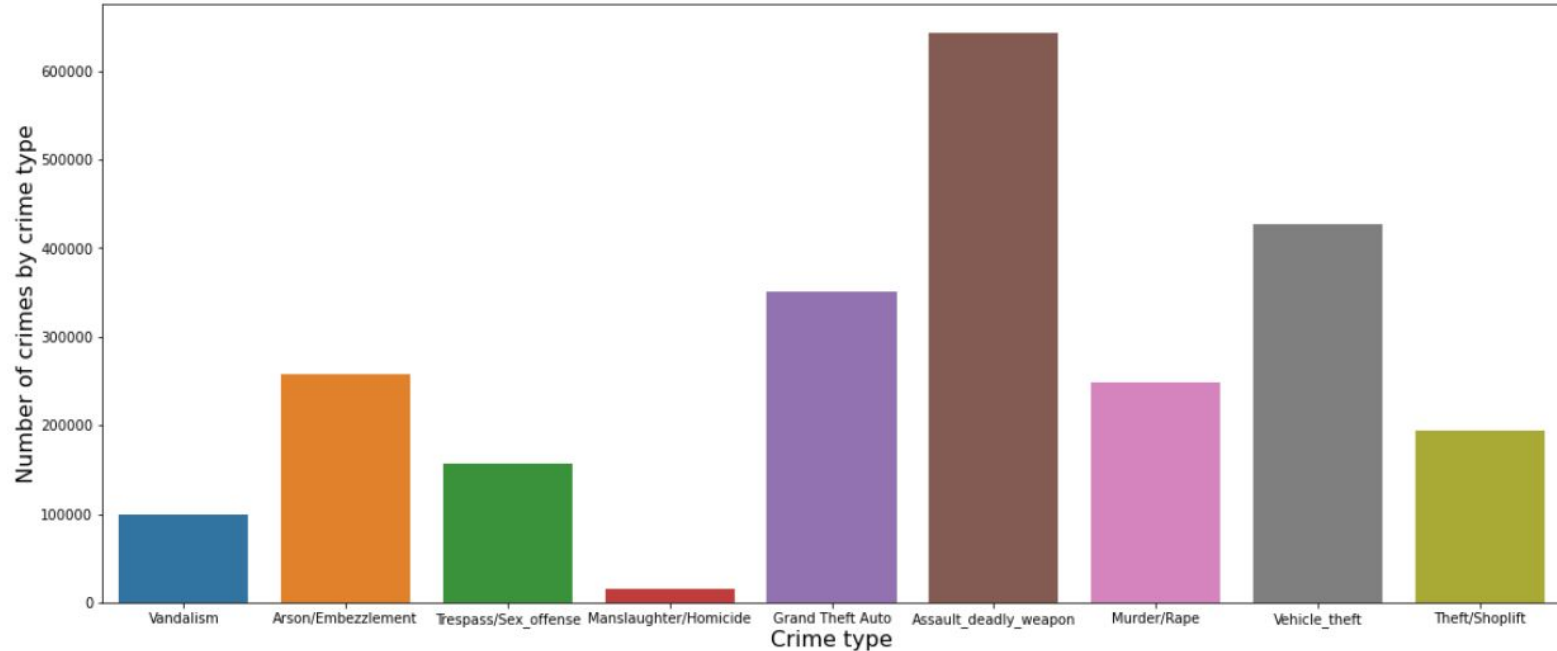
Step2. Connect crime with census data

- Census data left join zip/area dataframe
- Remove zip codes from census that did not have crime data
- Group by area and year
 - Population: Sum
 - Unemployment rate / Income: Average
 - Crime: From crime dataset, find average monthly number of crimes (2021 divide by 5)
- Combine into one dataframe

	AREA_NAME	Year	Population	Unemployment_rate	Median_household_income	Crimes_per_month
0	77th Street	2010	614727.0	15.898552	34668.071950	1203.333333
1	77th Street	2011	532941.0	12.154545	42531.909091	1187.333333
2	77th Street	2012	532445.0	12.490909	42105.818182	1186.916667
3	77th Street	2013	536127.0	13.745455	41582.090909	1145.083333
4	77th Street	2014	539384.0	13.481818	40561.181818	1170.833333

Data Visualization - Descriptive Statistics

1. Number of Crimes by Crime Type



→ Assault with deadly weapons are the most common crimes, followed by vehicle theft

Data Visualization - Descriptive Statistics

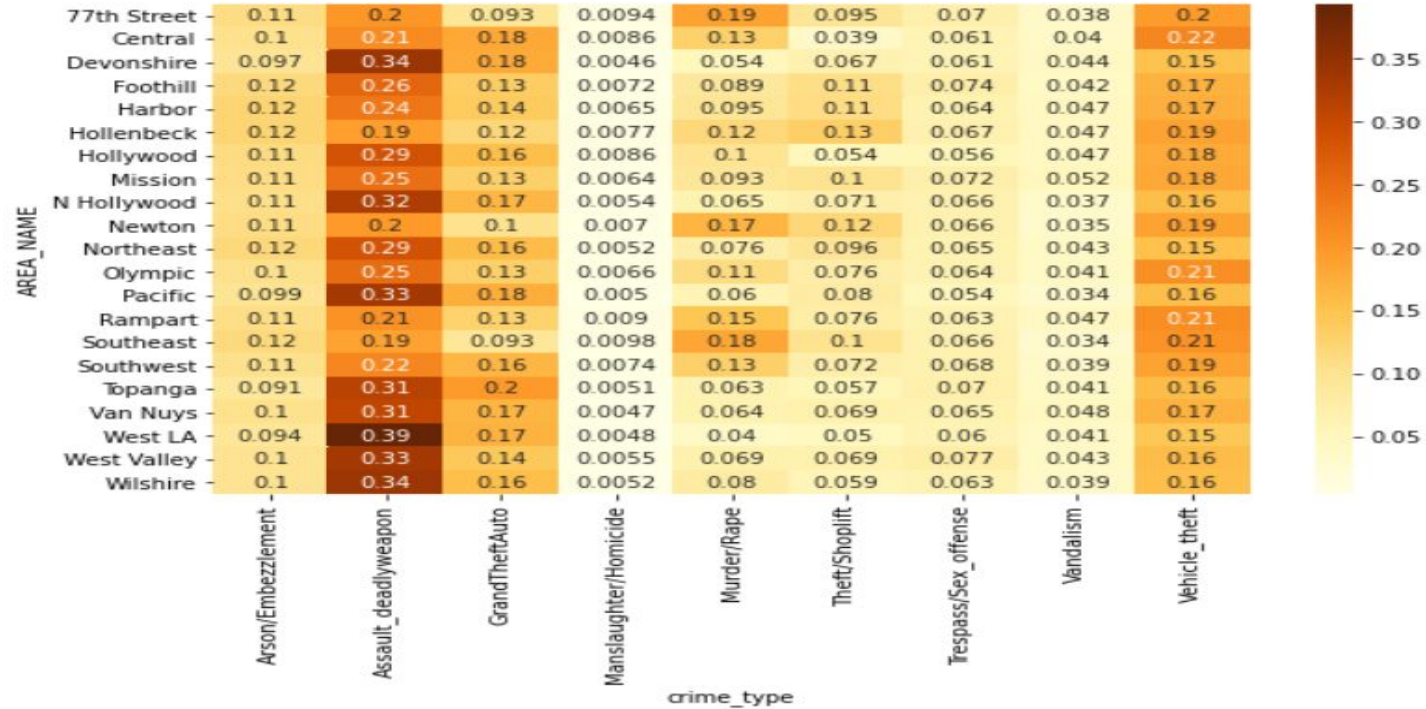
2. Proportion of Crime by Area Name and Crime Type

crime_type	Arson/Embezzlement	Assault_deadlyweapon	GrandTheftAuto	Manslaughter/Homicide	Murder/Rape	Theft/Shoplift	Trespass/Sex_offense	Vandalism	Vehicle_theft
AREA_NAME									
77th Street	11.12%	20.09%	9.35%	0.94%	18.72%	9.47%	7.01%	3.80%	19.50%
Central	10.38%	21.49%	18.12%	0.86%	13.05%	3.91%	6.05%	3.98%	22.15%
Devonshire	9.72%	34.18%	17.87%	0.46%	5.43%	6.71%	6.05%	4.44%	15.15%
Foothill	11.57%	25.97%	13.48%	0.72%	8.86%	10.87%	7.38%	4.21%	16.95%
Harbor	12.39%	23.99%	13.85%	0.65%	9.46%	11.11%	6.44%	4.70%	17.42%
Hollenbeck	12.38%	19.22%	12.13%	0.77%	11.96%	13.30%	6.75%	4.70%	18.78%
Hollywood	11.30%	28.64%	15.51%	0.86%	10.02%	5.43%	5.64%	4.69%	17.91%
Mission	11.11%	24.89%	13.16%	0.64%	9.29%	10.31%	7.16%	5.23%	18.20%
N Hollywood	10.78%	32.48%	16.51%	0.54%	6.53%	7.15%	6.60%	3.69%	15.71%
Newton	11.14%	19.87%	10.45%	0.70%	16.72%	12.07%	6.59%	3.53%	18.92%
Northeast	11.71%	28.66%	15.84%	0.52%	7.55%	9.59%	6.45%	4.33%	15.33%
Olympic	10.50%	25.31%	13.18%	0.66%	11.25%	7.59%	6.36%	4.12%	21.03%
Pacific	9.87%	33.39%	17.60%	0.50%	6.00%	7.99%	5.44%	3.42%	15.78%
Rampart	11.04%	20.75%	12.79%	0.90%	14.73%	7.64%	6.29%	4.66%	21.21%
Southeast	11.58%	18.89%	9.33%	0.98%	18.24%	10.31%	6.59%	3.44%	20.64%
Southwest	10.63%	22.14%	16.47%	0.74%	13.17%	7.22%	6.76%	3.92%	18.95%
Topanga	9.11%	31.40%	19.90%	0.51%	6.30%	5.68%	7.04%	4.13%	15.92%
Van Nuys	10.43%	30.76%	16.68%	0.47%	6.41%	6.91%	6.51%	4.81%	17.03%
West LA	9.40%	39.37%	17.15%	0.48%	4.03%	4.98%	6.00%	4.08%	14.52%
West Valley	10.15%	33.42%	14.28%	0.55%	6.86%	6.89%	7.73%	4.26%	15.85%
Wilshire	10.09%	34.05%	15.55%	0.52%	8.03%	5.88%	6.31%	3.93%	15.64%

➡ For each area in LA, deadly weapon assault(red) is the most common crime type, followed by vehicle theft(orange)

Data Visualization - Descriptive Statistics

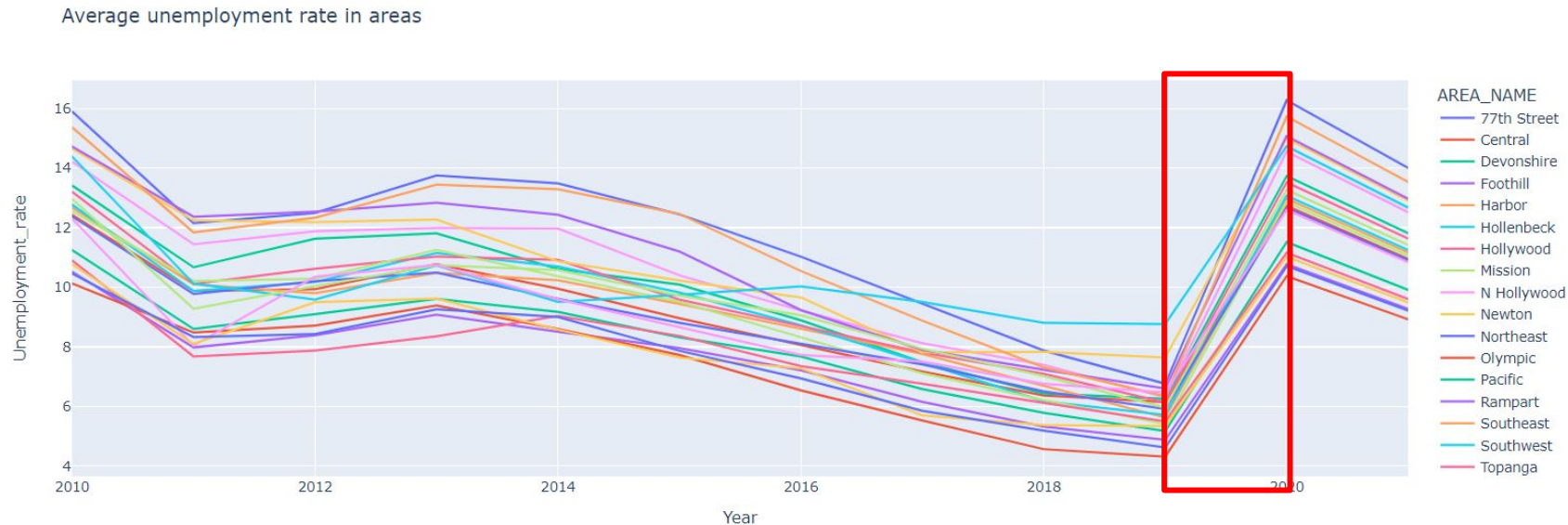
3. Heat Map for Crime Type by Area Name



Previous slide's information in a heatmap to highlight which areas are particularly dangerous(darker the worse)

Data Visualization - Descriptive Statistics

4. Average Unemployment Rate in LA Areas

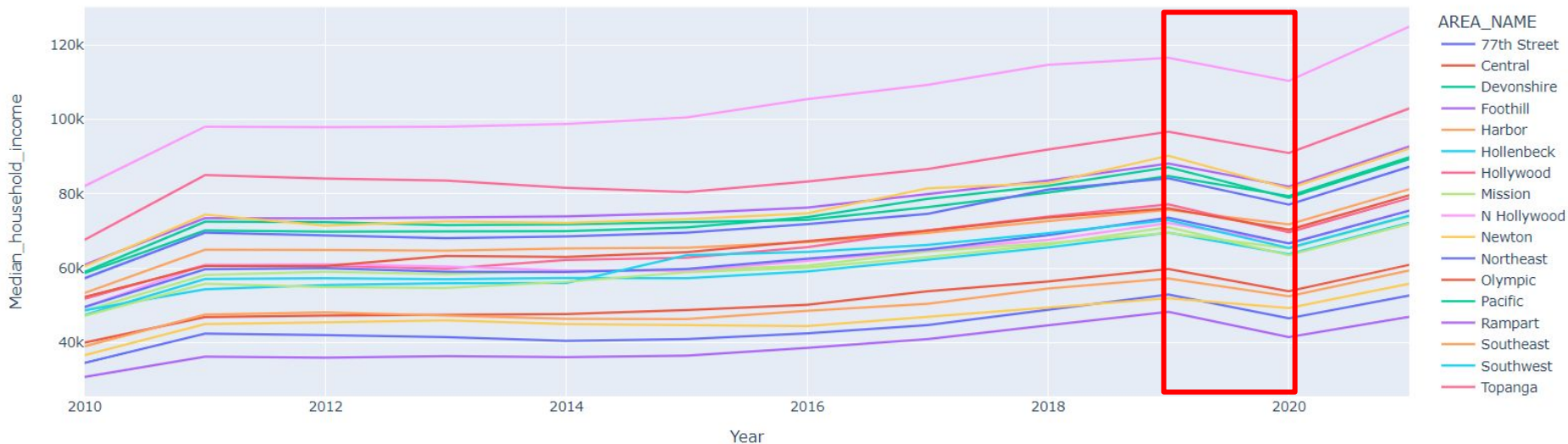


→ For the last 11 years, we see the unemployment rate drop on a consistent basis, until 2020 during covid where there was a significant spike in all areas, followed by a drop in 2021.

Data Visualization - Descriptive Statistics

5. Average Median Household Income in LA Areas

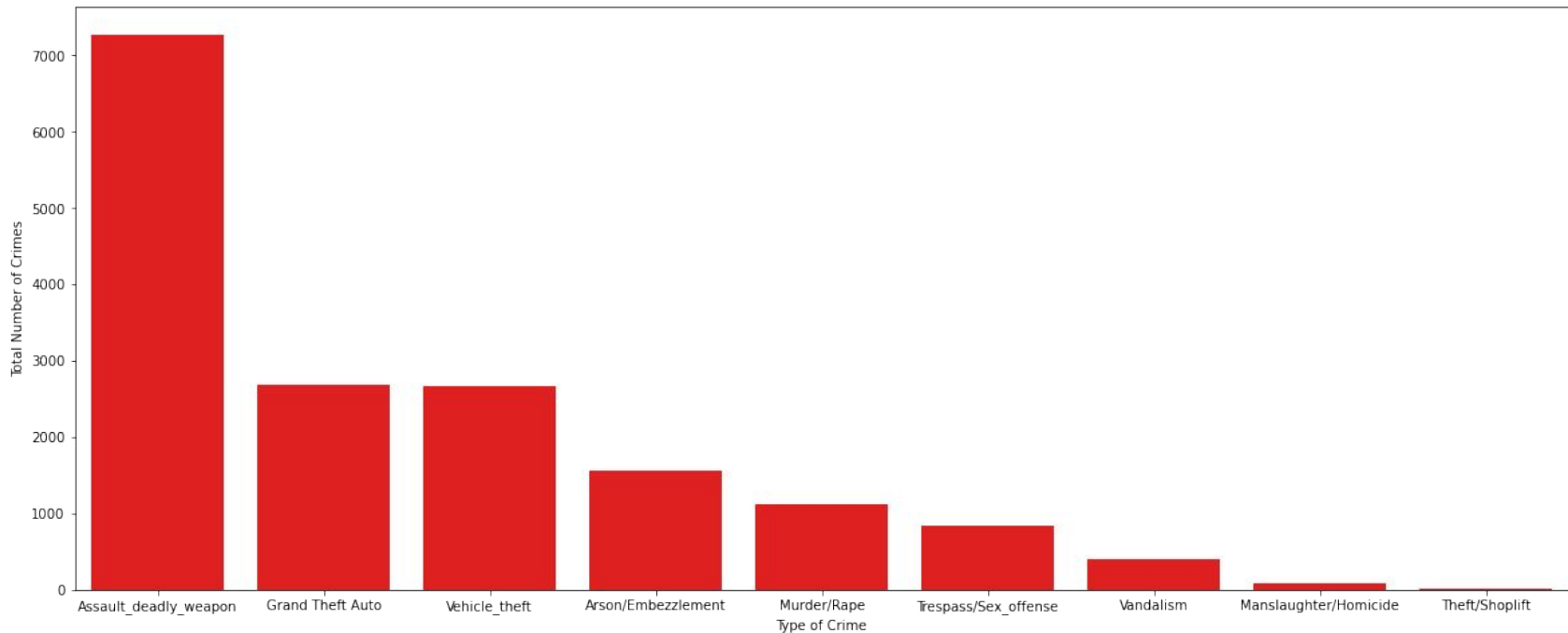
Average median household income in areas



For the last 11 years, we see the median household income rise on a consistent basis, until 2020 during covid where there was a big drop in all areas, followed by a sharp increase in 2021.

Data Visualization - Descriptive Statistics

6. Most crimes involving Asian victims are assault with deadly weapons

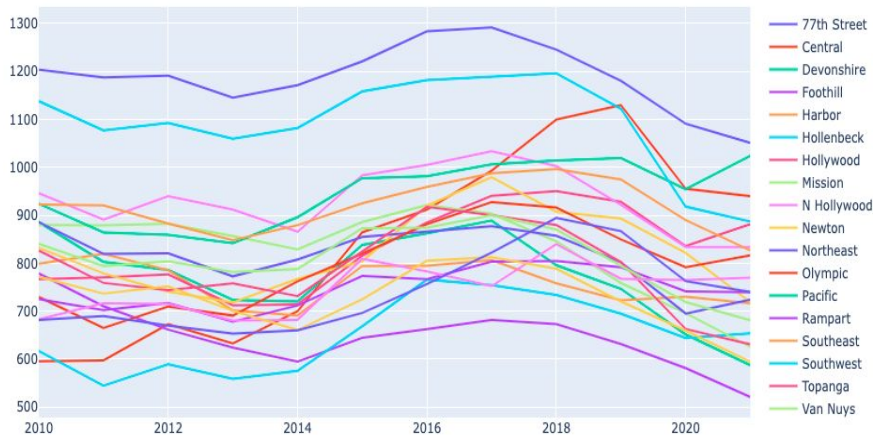


→ The chart above shows total number of crimes against Asians by crime type. Over the last 11 years, there have been more than 7000 deadly weapon assault cases against Asian community.

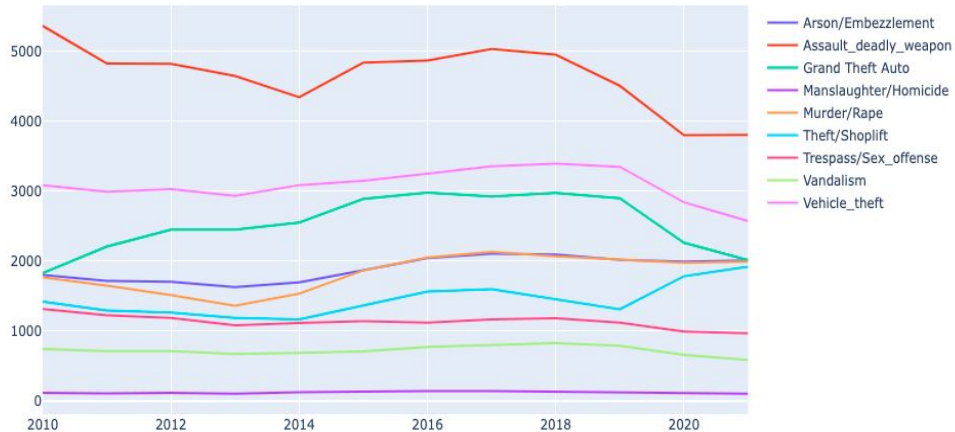
Analytics Process

Q1. How did covid-19 impact crime in Los Angeles? What other factors could lead to this impact?

Average # of monthly crime in each area over years



Average # of monthly crime for each type over years



Analytics Process

Q1. How does covid-19 impact crime in Los Angeles? What other factors could lead to this impact?

Linear Regression Model:

- $\log(\text{Average \# of monthly crimes each year in each area}) = b_0 + b_1 * \text{Covid} + b_2 * \log(\text{Population}) + b_3 * \log(\text{Unemployment Rate}) + \text{Area Dummies}$
- $\log(\text{Median household income})$ does not have impact on # of crime significantly and is therefore removed from the model

Analytics Process

Linear Regression Result:

- The population and unemployment rate have an impact on the number of crime significantly.
- Certain locations also significantly affect crime.
- After the pandemic, the average number of monthly crimes decreases by 10.4% compared to pre-covid periods.

OLS Regression Results						
Dep. Variable:	ln_Crimes_per_month	R-squared:	0.777			
Model:	OLS	Adj. R-squared:	0.754			
Method:	Least Squares	F-statistic:	34.47			
Date:	Thu, 02 Dec 2021	Prob (F-statistic):	2.47e-61			
Time:	13:42:09	Log-Likelihood:	268.98			
No. Observations:	252	AIC:	-490.0			
Df Residuals:	228	BIC:	-405.2			
Df Model:	23					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	-0.8141	2.133	-0.382	0.703	-5.018	3.390
Covid	-0.1099	0.025	-4.467	0.000	-0.158	-0.061
ln_Population	0.6360	0.162	3.914	0.000	0.316	0.956
ln_Unemployment_rate	-0.2083	0.026	-7.882	0.000	-0.260	-0.156
Central	-0.6519	0.071	-9.231	0.000	-0.791	-0.513
Devonshire	-0.1854	0.080	-2.314	0.022	-0.343	-0.028
Foothill	-0.5625	0.049	-11.529	0.000	-0.659	-0.466
Harbor	-0.8706	0.104	-8.350	0.000	-1.076	-0.665
Hollenbeck	-0.6721	0.037	-18.337	0.000	-0.744	-0.600
Hollywood	0.2912	0.179	1.626	0.105	-0.062	0.644
Mission	-0.2127	0.061	-3.467	0.001	-0.334	-0.092
N_Hollywood	-0.0071	0.075	-0.094	0.925	-0.155	0.141
Newton	0.2239	0.160	1.403	0.162	-0.090	0.538
Northeast	-0.4520	0.037	-12.197	0.000	-0.525	-0.379
Olympic	-0.2888	0.063	-4.586	0.000	-0.413	-0.165
Pacific	-0.5044	0.066	-7.686	0.000	-0.634	-0.375
Rampart	0.3197	0.207	1.543	0.124	-0.089	0.728
Southeast	-0.1058	0.054	-1.956	0.052	-0.212	0.001
Southwest	0.2394	0.095	2.511	0.013	0.052	0.427
Topanga	-0.0935	0.107	-0.876	0.382	-0.304	0.117
Van_Nuys	0.2926	0.186	1.574	0.117	-0.074	0.659
West_LA	-0.2018	0.087	-2.317	0.021	-0.373	-0.030
West_Valley	0.0470	0.160	0.293	0.770	-0.269	0.363
Wilshire	0.1132	0.174	0.650	0.516	-0.230	0.456
Omnibus:	16.219	Durbin-Watson:	0.898			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	22.812			
Skew:	-0.447	Prob(JB):	1.11e-05			
Kurtosis:	4.171	Cond. No.	5.21e+03			

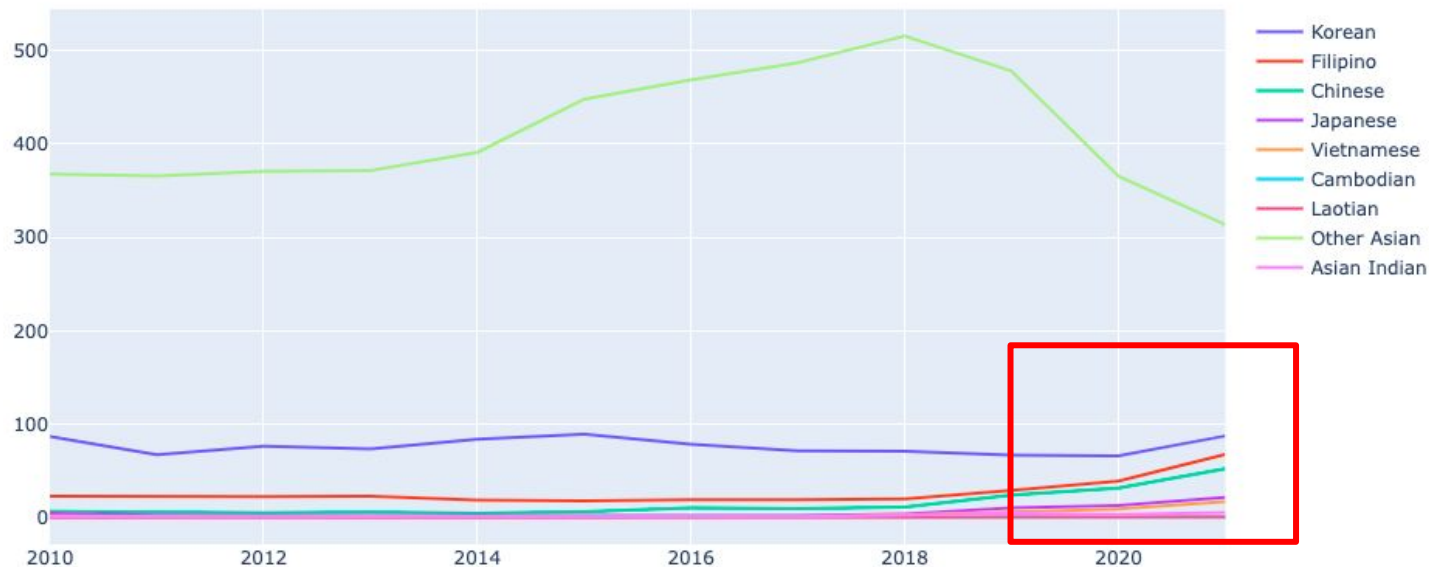
Notes:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 5.21e+03. This might indicate that there are strong multicollinearity or other numerical problems.

Analytics Process

Q2. Whether Asians were more likely to become victims after covid-19 in Los Angeles?

Average # of monthly victims in each race over years



Analytics Process

Q2. Whether Asian are more likely to become victims after covid-19 in Los Angeles?

Two sample t hypothesis test:

- Goal: To check whether the proportion of Asians victims (Chinese, Korean, Japanese, Filipino, Vietnamese, Cambodian, Laotian, Other Asian, Asian Indian) among all victims increases after covid-19 in LA.
- H_0 : The average proportion of Asian victims among all victims before covid-19 (μ_0) is the same as the average proportion after covid-19 (μ_1). $\mu_0 - \mu_1 = 0$
- H_a : The average proportion of Asian victims among all victims after covid-19 (μ_1) is larger than the average proportion before covid-19 (μ_0). $\mu_0 - \mu_1 < 0$

Analytics Process

Hypothesis Result:

- Test Statistics = -2.54
- Threshold: $t(df = 9, 0.05) = -1.833$
- Since $-2.54 < \text{Threshold}$, we can reject the null hypothesis.
- Inference: The average proportion of Asian victims among all victims after covid-19 is larger than the average proportion before covid-19.

The proportion of victims by race over years

CRIME_YEAR	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
VICTIM_RACE												
Other Asian	0.021112	0.021878	0.022084	0.023137	0.024009	0.024950	0.024950	0.025284	0.027038	0.026387	0.022285	0.019646
Black	0.162156	0.162203	0.166347	0.165829	0.168260	0.157521	0.151506	0.149697	0.153695	0.151229	0.142602	0.143362
Chinese	0.000387	0.000334	0.000298	0.000369	0.000271	0.000339	0.000550	0.000506	0.000586	0.001329	0.001916	0.003283
Cambodian	0.000019	0.000020	0.000005	0.000010	0.000010	0.000000	0.000009	0.000004	0.000017	0.000018	0.000030	0.000050
Filipino	0.001310	0.001325	0.001335	0.001412	0.001156	0.000998	0.001020	0.000991	0.001049	0.001610	0.002373	0.004235
Guamanian	0.000019	0.000040	0.000025	0.000016	0.000056	0.000060	0.000049	0.000052	0.000044	0.000037	0.000076	0.000088
Hispanic/Latin/Mexican	0.351275	0.352468	0.348238	0.345770	0.350645	0.344815	0.342798	0.337637	0.330280	0.330514	0.308179	0.302548
American Indian/Alaskan Native	0.000363	0.000379	0.000352	0.000457	0.000368	0.000469	0.000532	0.000424	0.000446	0.000672	0.000488	0.000839
Japanese	0.000258	0.000154	0.000124	0.000125	0.000153	0.000139	0.000124	0.000104	0.000205	0.000580	0.000793	0.001341
Korean	0.004967	0.004031	0.004556	0.004590	0.005136	0.004971	0.004179	0.003710	0.003728	0.003684	0.004015	0.005475
Laotian	0.000000	0.000015	0.000000	0.000016	0.000005	0.000005	0.000009	0.000000	0.000009	0.000028	0.000025	0.000038
Pacific Islander	0.000206	0.000154	0.000139	0.000145	0.000153	0.000246	0.000191	0.000139	0.000118	0.000138	0.000183	0.000175
Samoa	0.000038	0.000010	0.000020	0.000010	0.000000	0.000005	0.000022	0.000017	0.000013	0.000009	0.000041	0.000075
Hawaiian	0.000067	0.000100	0.000124	0.000047	0.000092	0.000186	0.000031	0.000095	0.000079	0.000083	0.000091	0.000213
Vietnamese	0.000086	0.000055	0.000060	0.000073	0.000031	0.000046	0.000062	0.000039	0.000157	0.000345	0.000559	0.001065
White	0.257258	0.255021	0.256968	0.251288	0.242773	0.244573	0.232913	0.227582	0.227236	0.223030	0.212977	0.212725
Unknown	0.200425	0.201794	0.199300	0.206645	0.206854	0.220644	0.241032	0.253688	0.255192	0.260055	0.303168	0.304516
Asian Indian	0.000053	0.000020	0.000025	0.000062	0.000026	0.000032	0.000022	0.000030	0.000109	0.000253	0.000198	0.000326

Conclusion

Q1: How does covid-19 impact crime in Los Angeles? What other factors could lead to this impact?

- The population, unemployment rate and locations have an impact on the number of crime significantly
- The average monthly rate of crime decreases after the outbreak of covid-19

Q2: Whether Asian are more likely to become victims after covid-19 in Los Angeles?

- Among all crimes, the proportion of the asian victims increases after the outbreak of covid-19

Reference

1. Kaggle:
https://www.kaggle.com/sumaiaparveenshupti/los-angeles-crime-data-20102020?select=Crime_Data_from_2010_to_2019.csv
2. United States Census Bureau: <https://data.census.gov/cedsci/>
 - a. Household median income:
<https://data.census.gov/cedsci/table?q=income&g=0400000US06%248600000&tid=ACST5Y2019.S1901&hidePreview=true>
 - b. Unemployment:
<https://data.census.gov/cedsci/table?q=unemployment&g=0400000US06%248600000&tid=ACST5Y2019.S2301&hidePreview=true>
 - c. Population:
<https://data.census.gov/cedsci/table?q=demographic&g=0400000US06%248600000&tid=ACSDP5Y2019.DP05&hidePreview=true>
3. Google API Call: <https://maps.googleapis.com/maps/api/geocode/json?latlng=34.0197,-118.2749&key>

Thanks for Listening!

Appendix

Proportion of Victim by Race Over Years

	VICTIM_RACE
Hispanic/Latin/Mexican	33.887%
White	23.798%
Unknown	23.482%
Black	15.665%
Other Asian	2.386%
Korean	0.435%
Filipino	0.141%
Chinese	0.071%
American Indian/Alaskan Native	0.046%
Japanese	0.029%
Vietnamese	0.017%
Pacific Islander	0.017%
Hawaiian	0.009%
Asian Indian	0.008%
Guamanian	0.004%
Samoan	0.002%
Cambodian	0.001%
Laotian	0.001%

CRIME_YEAR	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
VICTIM_RACE												
Other Asian	2.11%	2.19%	2.21%	2.31%	2.40%	2.50%	2.50%	2.53%	2.70%	2.64%	2.23%	1.96%
Black	16.22%	16.22%	16.63%	16.58%	16.83%	15.75%	15.15%	14.97%	15.37%	15.12%	14.26%	14.34%
Chinese	0.04%	0.03%	0.03%	0.04%	0.03%	0.03%	0.06%	0.05%	0.06%	0.13%	0.19%	0.33%
Cambodian	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%
Filipino	0.13%	0.13%	0.13%	0.14%	0.12%	0.10%	0.10%	0.10%	0.10%	0.16%	0.24%	0.42%
Guamanian	0.00%	0.00%	0.00%	0.00%	0.01%	0.01%	0.00%	0.01%	0.00%	0.00%	0.01%	0.01%
Hispanic/Latin/Mexican	35.13%	35.25%	34.82%	34.58%	35.06%	34.48%	34.28%	33.76%	33.03%	33.05%	30.82%	30.25%
American Indian/Alaskan Native	0.04%	0.04%	0.04%	0.05%	0.04%	0.05%	0.05%	0.04%	0.04%	0.07%	0.05%	0.08%
Japanese	0.03%	0.02%	0.01%	0.01%	0.02%	0.01%	0.01%	0.01%	0.02%	0.06%	0.08%	0.13%
Korean	0.50%	0.40%	0.46%	0.46%	0.51%	0.50%	0.42%	0.37%	0.37%	0.37%	0.40%	0.55%
Laotian	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Pacific Islander	0.02%	0.02%	0.01%	0.01%	0.02%	0.02%	0.02%	0.01%	0.01%	0.01%	0.02%	0.02%
Samoan	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%
Hawaiian	0.01%	0.01%	0.01%	0.00%	0.01%	0.02%	0.00%	0.01%	0.01%	0.01%	0.01%	0.02%
Vietnamese	0.01%	0.01%	0.01%	0.01%	0.00%	0.00%	0.01%	0.00%	0.02%	0.03%	0.06%	0.11%
White	25.73%	25.50%	25.70%	25.13%	24.28%	24.46%	23.29%	22.76%	22.72%	22.30%	21.30%	21.27%
Unknown	20.04%	20.18%	19.93%	20.66%	20.69%	22.06%	24.10%	25.37%	25.52%	26.01%	30.32%	30.45%
Asian Indian	0.01%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.01%	0.03%	0.02%	0.03%

→ Total for last 11 years, 34% of victims of crime in LA are Hispanic, followed by 24% White and 16% Black community. Over the years, it may seem like crimes against Black, Hispanic and White community have fallen, but do note the Unknown column. It seems that victim race records have not been kept as well as they were 10 years ago.

Appendix

Population Density by LA Zipcodes

