Los Angeles Crime Data Analysis

Makai Todd
Bsan 360

12-9-25

https://github.com/makaitodd/BSAN360-Project

Data

- Our data comes from the city of Los Angeles and it records each crime reported from 2020 to 2024
- It contains 1,004,991 rows and 28 columns
- The column I utalized for my research and analysis were department, area_name, area_number, part_1-2, crime_code, Crime_code_desc, date_occured, Time_occured, vict_age, vict_sex, vict_descent, and hour.

department	area_name	area_number	part_1-2	crime_code	crime_code_desc
211507896	N Hollywood	15	2	354	THEFT OF IDENTITY
201516622	N Hollywood	15	1	230	ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT
240913563	Van Nuys	9	2	354	THEFT OF IDENTITY
210704711	Wilshire	7	1	331	THEFT FROM MOTOR VEHICLE - GRAND ($950.01 AND ...
201418201	Pacific	14	1	420	THEFT FROM MOTOR VEHICLE - PETTY ($950 & UNDER)

Research Questions

Research Question 1: Which area of LA is the most dangerous and is it statistically more dangerous than the other areas of LA?

Research question 2: What time of day do the most crimes occur and is it statistically more than other times of day?

Research question 3: Which demographics have the most crimes committed on them, how much do certain demographics have crimes committed upon them compared to others?

Research Question 1

	area_name	crime_count	avg_per_day
1	Central	63707	43.605065
0	77th Street	51757	35.425736
12	Pacific	50265	34.404517
15	Southwest	48682	33.321013
6	Hollywood	47375	32.426420
8	N Hollywood	44191	30.247091
11	Olympic	42934	29.386721
20	Wilshire	42726	29.244353
14	Southeast	41299	28.267625
18	West LA	41188	28.191650
13	Rampart	39894	27.305955
9	Newton	39617	27.116359
16	Topanga	36828	25.207392
17	Van Nuys	36725	25.136893
19	West Valley	36590	25.044490
10	Northeast	36297	24.843943
2	Devonshire	36266	24.822724
4	Harbor	34054	23.308693
7	Mission	32822	22.465435
5	Hollenbeck	29712	20.336756
3	Foothill	27404	18.757016

- The most dangerous area (based on average crimes per day) is Central
- Central has an average of 43.6 crimes per day

Research question 1

	area	t_stat	p_value
0	77th Street	15.525173	1.097163e-52
1	Devonshire	43.859626	0.000000e+00
2	Foothill	58.976536	0.000000e+00
3	Harbor	47.258872	0.000000e+00
4	Hollenbeck	54.804281	0.000000e+00
5	Hollywood	23.850107	1.870888e-116
6	Mission	48.849150	0.000000e+00
7	N Hollywood	30.014636	1.354267e-174
8	Newton	35.040521	1.681691e-230
9	Northeast	43.767464	9.881313e-324
10	Olympic	30.957463	2.704030e-185
10	Olympic	30.957463	2.704030e-185
11	Pacific	20.533198	3.914766e-88
12	Rampart	35.893957	1.416137e-238
13	Southeast	31.177531	4.542817e-189
14	Southwest	21.441569	5.888162e-96
15	Topanga	42.460614	3.596475e-310
16	Van Nuys	42.742856	2.556393e-312
17	West LA	35.004235	5.258818e-227
18	West Valley	42.093246	9.918612e-308
19	Wilshire	29.214344	1.647810e-168

- I ran a T test comparing the mean crimes per day from Central to every other area of Los Angeles
- From the P-values (all less than .05) we can see that Central is statistically more dangerous than any other area of Los Angeles

Research question 2

	hour_bin	avg_per_day
0	0–3	55.380470
1	4–7	39.125934
2	8–11	75.608324
3	12–15	101.738527
4	16–19	100.839381
5	20–23	86.396478

- The most crimes are committed between 12pm and 3pm
- There are on average 101.8 crimes committed per day between 12 pm and 3pm
- The safest time of day is between 4am and 7 am

Research question 2

Tukey Test

group1	group2	meandiff	p-adj
12-15	16-19	-0.8991	0.9446
12-15	20-23	-15.342	0.0
12-15	4-7	-62.6126	0.0
12-15	8-11	-26.1302	0.0
0-3	12-15	46.3581	0.0

- After performing a tukey test to compare all of the mean daily crime amounts for each time period we can see the time period of 12pm-3pm has statistically more crimes than 12am-3am, 4am-7am, 8am-11am, and 8pm-11pm.
- We can also see we cannot conclude that more crimes are committed from 12pm 3pm when compared to 4pm-7pm as the pvalue is more than .05

Research question 3

Linear Regression

	coef	std err	t	P>|t|	[0.025	0.975]
const	448.3099	140.688	3.187	0.001	172.451	724.169
vict_age	-7.1508	1.333	-5.364	0.000	-9.765	-4.537
vict_sex_H	-620.5265	203.024	-3.056	0.002	-1018.612	-222.442
vict_sex_Male	36.8036	69.175	0.532	0.595	-98.834	172.441
vict_sex_X	-314.3197	111.703	-2.814	0.005	-533.344	-95.295
vict_descent_B	488.8480	162.800	3.003	0.003	169.633	808.063
vict_descent_C	-68.9683	179.436	-0.384	0.701	-420.804	282.867
vict_descent_D	-215.3540	277.396	-0.776	0.438	-759.268	328.560
vict_descent_F	16.3801	172.893	0.095	0.925	-322.626	355.386
vict_descent_G	-193.8963	270.437	-0.717	0.473	-724.164	336.372
vict_descent_H	1069.8650	160.519	6.665	0.000	755.122	1384.608
vict_descent_I	-134.3335	195.078	-0.689	0.491	-516.839	248.172
vict_descent_J	-90.2668	182.949	-0.493	0.622	-448.990	268.457
vict_descent_K	-41.0977	179.496	-0.229	0.819	-393.050	310.855
vict_descent_L	-188.3307	279.672	-0.673	0.501	-736.707	360.046
vict_descent_O	257.3285	163.166	1.577	0.115	-62.605	577.261
vict_descent_P	-151.0020	206.944	-0.730	0.466	-556.774	254.770
vict_descent_S	-186.9169	292.701	-0.639	0.523	-760.840	387.006
vict_descent_U	-180.1587	227.674	-0.791	0.429	-626.578	266.260
vict_descent_V	-132.4671	199.169	-0.665	0.506	-522.993	258.059
vict_descent_W	738.7011	161.988	4.560	0.000	421.078	1056.324
vict_descent_X	374.0886	161.960	2.310	0.021	56.521	691.656
vict_descent_Z	-136.5109	204.971	-0.666	0.505	-538.414	265.392

- This is a linear regression displaying how many more or less crimes are committed on each demographic when comparing them to the intercept
- In this case we can see the intercept is a victim who is zero years old, of asian decent and a female because those are the variables not displayed in the regression.

Research question 3

From this model we can conclude

- A one year increase in victim age is associated with about 7.15 fewer crimes (As you get older you are less likely to be a victim of a crime
- The most crimes are committed upon people of Hispanic descent (H)
- The least crimes are committed upon people upon people of filipino descent (D)
- More crimes are committed upon males compared to females

Conclusion

The analysis we have gathered can be especially helpful for the police department trying to prevent and fight crime

- Knowing that the most crimes are committed in the Central area can help them know where to send petrol cars and station cops in order to prevent crime and respond to incidents quicker
- Knowing that the most crimes are committed between 12pm and 7pm could also be helpful in deciding when to schedule potrolman and allocate resources during peak crime hours
- Knowing which demographics of people have the most/least crimes committed upon them can be helpful so they can set up crime prevention programs in areas where the most affected demographics are