Optimizing

Introduction

According to a recent report from RealPage, one of the many impacts of Covid-19 is a decline in rental prices in Los Angeles. In fact, there was a 3.3% decrease to an average of $2,254 for units of all sizes from a year earlier. This might be a good time to relocate if you are looking to save extra money.

But suppose you wish to move into a safer neighborhood but want to spend less money than your current rate. Today, I will show you how you could have your cake and eat it too through the magic of data scraping, a bit of python coding, and data visualization!

I recently took a certificate training in data science from IBM, and this is my capstone project. For this project, I wanted to find out if more expensive rentals or neighborhood locations correlated to lesser crime rates. I also wanted to find out if neighborhood location verse rental prices had better predictor of crime rates.

Data: The Art of Stealing Data..."Legally"

By stealing, I mean leveraging the plethora of available data at your fingertips (and borrowing codes from more seasoned data scientists). I used Zillow to gather the rental data, but Zillow does not provide us with a beautiful spreadsheet of rental info.

To scrape the data from Zillow, we'll need an API, a fancy term for a bot that will retrieve our data for us. Luckily, Maksim Korzh had already created an API for just this occasion. You can check his YouTube and GitHub for more info.

To access crime data for Los Angeles, I used the data from www.data.lacity.org.

Methodology: IF, THEN, RETURN, ELSE #RAGEQUIT!

There is no secret to managing your #RageQUIT during the process of cleaning your data. Data, like life, never behaves like it ought to! I spent 99% of the time cleaning data and trying to figure out why I kept getting errors.

I did find that forcing myself to take a periodic break every 25 minutes helped. I did this project with a friend, so we had a shared a common enemy, errors!

ASK And You Shall Recieve, EXCEPT for ZILLOW…In Which You'll Need To Scrape for It!

To access rental data from Zillow, one must first download the Zillow API created by Maksim Korzh. Then it is just a matter of plug and play after you access the link from the Zillow webpage. I highly recommend watching Korzh's videos on YouTube.

PANDAS for the WIN…And Then More Cleaning!

The output from the Zillow API will return a JSON file. To access the file, you must first convert it to a data frame, which requires importing the Pandas package. Unfortunately, the data will require a bit of cleaning since each cell is storing additional lists.

Out[1]:

	url	details	address	coordinates
0	https://www.zillow.com/b/wren-los-angeles-ca-9...	[1, 875 + *Studio*, 2,045+ 1 bd, $3,105+ 2 bds]	Wren | 1230 S Olive St, Los Angeles, CA	{'latitude': 34.038567, 'longitude': -118.262796}
1	https://www.zillow.com/b/eastown-los-angeles-c...	[1, 705 + *Studio*, 1,965+ 1 bd, $2,960+ 2 bds]	Eastown | 6201 Hollywood Blvd, Los Angeles, CA	{'latitude': 34.10226, 'longitude': -118.3248}
2	https://www.zillow.com/b/cobalt-culver-city-ca...	[2, 446 + *Studio*, 3,186+ 1 bd, $4,637+ 2 bds]	Cobalt | 10601 Washington Blvd, Culver City, CA	{'latitude': 34.017442, 'longitude': -118.405222}
3	https://www.zillow.com/b/the-avenue-hollywood-...	[2, 885 *Studio*, 2,715+ 1 bd, 3, 530 + 2*bds*,...	The Avenue Hollywood | 1619 N La Brea Ave, Los...	{'latitude': 34.100538, 'longitude': -118.345047}
4	https://www.zillow.com/b/alessio-los-angeles-c...	[1, 967 + 1*bd*, 2,538+ 2 bds]	Alessio | 5700 W Centinela Ave, Los Angeles, CA	{'latitude': 33.97548, 'longitude': -118.381095}

First, we need to remove the URL column, then look for the lowest price by separating the details column into separate columns to get the lowest price. We'll need to clean the new price column only to include the money value and remove additional info (like studio or 1bd). After that, we need to separate the latitude and longitude. Our final result should look something like this.

	address	latitude	longitude	price
0	Wren | 1230 S Olive St, Los Angeles, CA	34.038567	-118.262796	$1,875
1	Eastown | 6201 Hollywood Blvd, Los Angeles, CA	34.102260	-118.324800	$1,705
2	Cobalt | 10601 Washington Blvd, Culver City, CA	34.017442	-118.405222	$2,446
3	The Avenue Hollywood | 1619 N La Brea Ave, Los...	34.100538	-118.345047	$2,885
4	Alessio | 5700 W Centinela Ave, Los Angeles, CA	33.975480	-118.381095	$1,967
...	...	...	...	...
182	Del Mor Apartments | 1551 Echo Park Ave, Los A...	34.081258	-118.254915	$1,940
183	7733 Oakwood Ave, Los Angeles, CA	34.078205	-118.358034	$2,850
184	Toscana Apartments | 15736 Vanowen St, Lake Ba...	34.193370	-118.476241	$1,644
185	The Visconti | 1221 W 3rd St, Los Angeles, CA	34.057280	-118.258237	$1,825
186	The Orsini | 550 N Figueroa St, Los Angeles, CA	34.060960	-118.246352	$1,636

Would you like some ZIPCODES to those LATS & LONGS?

The first challenge is figuring out how to integrate our rental data with the crime data. For this, I've decided to use zip code data as the common factor rather than address location. Since we

know our latitude and longitude data, we can use another API called USZIPCODE to search for the relevant data.

	address	latitude	longitude	price	zipcode
0	Wren | 1230 S Olive St, Los Angeles, CA	34.038567	-118.262796	$1,875	90015
1	Eastown | 6201 Hollywood Blvd, Los Angeles, CA	34.102260	-118.324800	$1,705	90028
2	Cobalt | 10601 Washington Blvd, Culver City, CA	34.017442	-118.405222	$2,446	90232
3	The Avenue Hollywood | 1619 N La Brea Ave, Los...	34.100538	-118.345047	$2,885	90028
4	Alessio | 5700 W Centinela Ave, Los Angeles, CA	33.975480	-118.381095	$1,967	90056
...	...	...	...	...	...
182	Del Mor Apartments | 1551 Echo Park Ave, Los A...	34.081258	-118.254915	$1,940	90026
183	7733 Oakwood Ave, Los Angeles, CA	34.078205	-118.358034	$2,850	90036
184	Toscana Apartments | 15736 Vanowen St, Lake Ba...	34.193370	-118.476241	$1,644	91406
185	The Visconti | 1221 W 3rd St, Los Angeles, CA	34.057280	-118.258237	$1,825	90071
186	The Orsini | 550 N Figueroa St, Los Angeles, CA	34.060960	-118.246352	$1,636	90012

L.A. CRIMES: Special Vector Unit (dah dah)

In the criminal justice system, violent crime based offenses are considered especially heinous. In Los Angeles, the dedicated detectives who investigate these vicious crimes are members of an elite squad known as the Special Vector Unit. These are their stories (in CSV vector files).

Now that we have our rental data, we need to interlace that with our crime data. Luckily, you can access the crime data from data.lacity.org. Luckily for us, the files are in CSV format, so we won't need to scrape for the data.

	DR_NO	Date Rptd	DATE OCC	TIME OCC	AREA	AREA NAME	Rpt Dist No	Part 1-2	Crm Cd	Crm Cd Desc	...	Status	Status Desc	Crm Cd 1	Crm Cd 2	Crm Cd 3	Crm Cd 4	LOCATIC
0	10304468	01/08/2020 12:00:00 AM	01/08/2020 12:00:00 AM	2230	3	Southwest	377	2	624	BATTERY - SIMPLE ASSAULT	...	AO	Adult Other	624.0	NaN	NaN	NaN	1100 39TH
1	190101086	01/02/2020 12:00:00 AM	01/01/2020 12:00:00 AM	330	1	Central	163	2	624	BATTERY - SIMPLE ASSAULT	...	IC	Invest Cont	624.0	NaN	NaN	NaN	700 S HI
2	190101087	01/02/2020 12:00:00 AM	01/01/2020 12:00:00 AM	510	1	Central	156	2	626	INTIMATE PARTNER - SIMPLE ASSAULT	...	IC	Invest Cont	626.0	NaN	NaN	NaN	300 E 5
3	191501505	01/01/2020 12:00:00 AM	01/01/2020 12:00:00 AM	1730	15	N Hollywood	1543	2	745	VANDALISM - MISDEAMEANOR ($399 OR UNDER)	...	IC	Invest Cont	745.0	998.0	NaN	NaN	54 CORTE
4	191921269	01/01/2020 12:00:00 AM	01/01/2020 12:00:00 AM	415	19	Mission	1998	2	740	VANDALISM - FELONY ($400 & OVER, ALL CHURCH VA...	...	IC	Invest Cont	740.0	NaN	NaN	NaN	144 TITUS

Next, we need to clean the LA crime data and run the zipcode generator. The final product should look like this.

	Unnamed: 0	lat	lon	zipcode	crimes
0	0	34.0141	-118.2978	90037	VIOLATION OF COURT ORDER
1	1	34.0459	-118.2545	90014	VANDALISM - FELONY ($400 & OVER, ALL CHURCH VA...
2	2	34.0449	-118.2458	90014	OTHER MISCELLANEOUS CRIME
3	3	34.1685	-118.4019	91607	VIOLATION OF COURT ORDER
4	4	34.2198	-118.4468	91402	RAPE, ATTEMPTED
...	...	...	...	...	...
995	995	34.0451	-118.2604	90017	BATTERY - SIMPLE ASSAULT
996	996	34.0480	-118.2577	90017	THEFT-GRAND ($950.01 & OVER)EXCPT,GUNS,FOWL,LI...
997	997	34.0423	-118.2452	90014	BATTERY - SIMPLE ASSAULT
998	998	34.0459	-118.2545	90014	THEFT PLAIN - PETTY ($950 & UNDER)
999	999	34.0388	-118.2411	90021	LEWD CONDUCT

Chances are, multiple crimes are occurring in the same zip code location. So, we'll need to create a separate data frame that calculates the number of crimes per zip code. The results should look something like this:

	zip	crimes
0	90001	1
1	90002	1
2	90003	2
3	90004	1
4	90005	1
...	...	...
74	91602	1
75	91604	1
76	91605	2
77	91606	5
78	91607	3

We also want to know the types of crimes that occur for each zipcode.

	zip	crimes
0	90001	[BATTERY - SIMPLE ASSAULT]
1	90002	[THEFT OF IDENTITY]
2	90003	[OTHER MISCELLANEOUS CRIME, ASSAULT WITH DEADL...
3	90004	[THEFT PLAIN - PETTY ($950 & UNDER)]
4	90005	[INTIMATE PARTNER - AGGRAVATED ASSAULT]
...	...	...
74	91602	[BURGLARY FROM VEHICLE]
75	91604	[BATTERY - SIMPLE ASSAULT]
76	91605	[VANDALISM - FELONY ($400 & OVER, ALL CHURCH V...
77	91606	[VANDALISM - MISDEAMEANOR ($399 OR UNDER), VAN...
78	91607	[VIOLATION OF COURT ORDER, VANDALISM - MISDEAM...

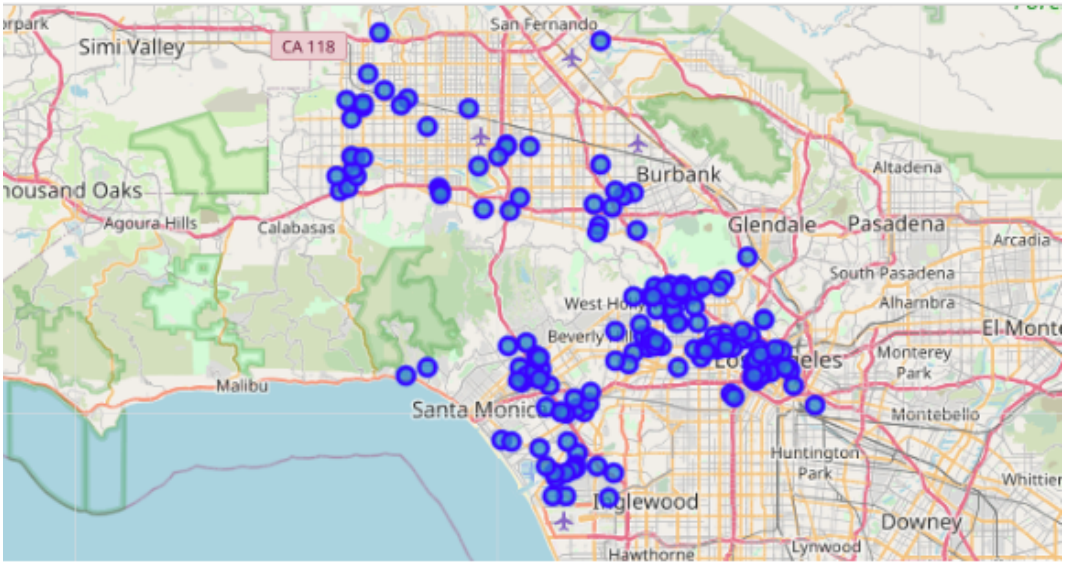
79 rows × 2 columns

The final results should include a data frame with all of the relevant data points such as the address, zip code, frequency of crimes, and the description of the crimes.

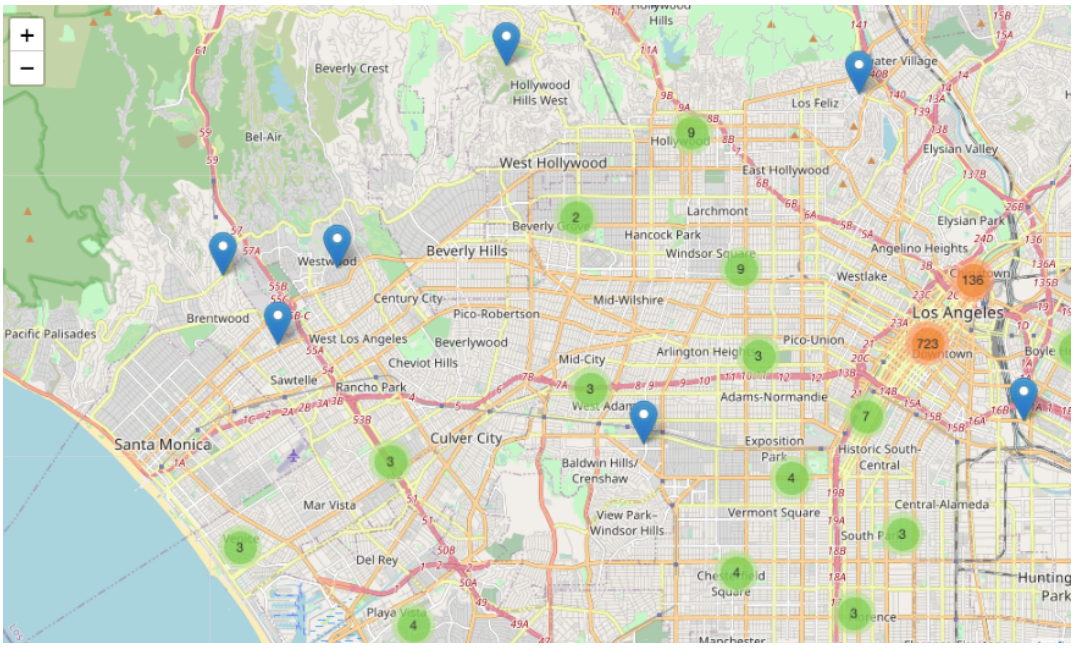
	address	latitude	longitude	price	zipcode	crime_types	crime_freq
0	Wren | 1230 S Olive St, Los Angeles, CA	34.038567	-118.262796	$1,875	90015	[BATTERY - SIMPLE ASSAULT, ATTEMPTED ROBBERY, ...	102.0
1	Eastown | 6201 Hollywood Blvd, Los Angeles, CA	34.102260	-118.324800	$1,705	90028	[OTHER MISCELLANEOUS CRIME, THEFT PLAIN - PETT...	8.0
2	Cobalt | 10601 Washington Blvd, Culver City, CA	34.017442	-118.405222	$2,446	90232	NaN	NaN
3	The Avenue Hollywood | 1619 N La Brea Ave, Los...	34.100538	-118.345047	$2,885	90028	[OTHER MISCELLANEOUS CRIME, THEFT PLAIN - PETT...	8.0
4	Alessio | 5700 W Centinela Ave, Los Angeles, CA	33.975480	-118.381095	$1,967	90056	NaN	NaN
...	...	...	...	...	...	...	...
182	Del Mor Apartments | 1551 Echo Park Ave, Los A...	34.081258	-118.254915	$1,940	90026	[ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAUL...	2.0
183	7733 Oakwood Ave, Los Angeles, CA	34.078205	-118.358034	$2,850	90036	[BATTERY - SIMPLE ASSAULT]	1.0
184	Toscana Apartments | 15736 Vanowen St, Lake Ba...	34.193370	-118.476241	$1,644	91406	NaN	NaN
185	The Visconti | 1221 W 3rd St, Los Angeles, CA	34.057280	-118.258237	$1,825	90071	[OTHER MISCELLANEOUS CRIME, INTIMATE PARTNER -...	35.0
186	The Orsini | 550 N Figueroa St, Los Angeles, CA	34.060960	-118.246352	$1,636	90012	[ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAUL...	130.0

Results: If I MAP It, They Will Come!

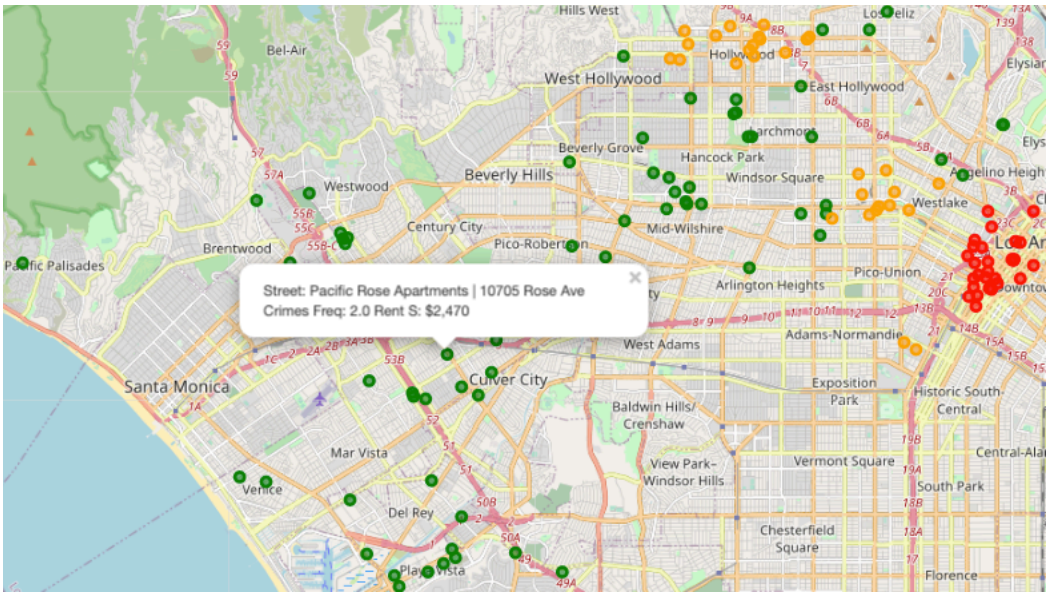
The best method for identifying the best rental location is to show it visually on a map. For this, we use another API called Folium. First, we want to map the apartment data location.



Next, we want to match the crime data.



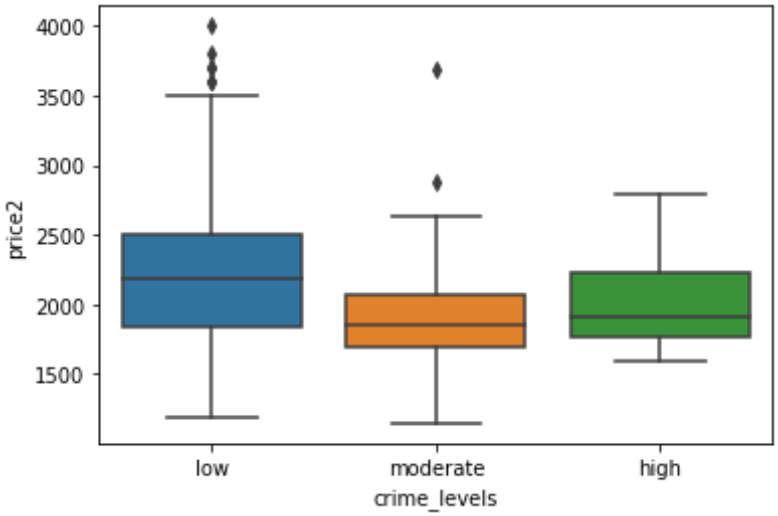
The previous maps help show us our rental data and crime data, but we want to create an interactive map that will 1) show the levels of crimes for the particular rental location, 2) the frequency of crimes around that zip code, and 3) the price. The final graph should look like this (with green, orange, and red being low, moderate, and high levels of crimes).



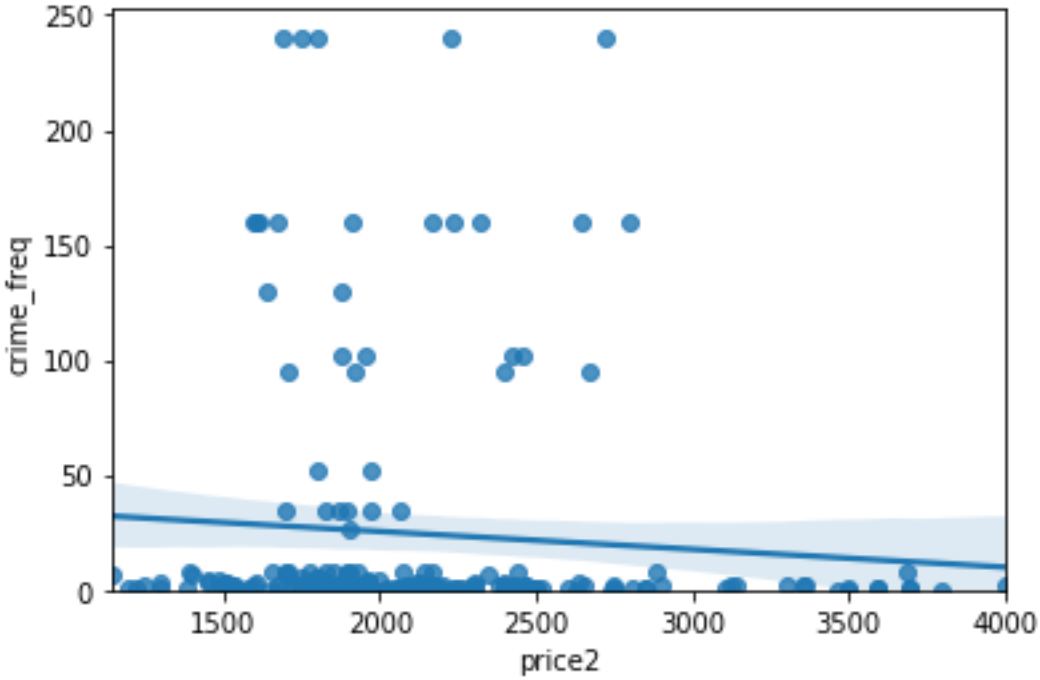
Discussion- The Westside Story

Base on the map, it looks like levels of crimes are more related to neighborhood locations with the fewest being near the West LA, Mid-Wilshire, and Beverly Grove Area. More incidents of crimes seemed to be located around the Downtown area, but it is hard to tell if the price has an impact as well.

To validate this, we could run a simple correlation analysis and boxplot visualization. For instance, while low levels of crime, on average, had slightly higher rental costs, the range of prices was the most diverse. Moderate and high levels of corruption, for the most part, had similar rental fees (which is $200 less, on average, compared to the standards of crime).



Another method for validating our data is running a similar correlation analysis to see if there is a relationship between crime levels and rental prices. That is, do higher rental price predict a safer neighborhood. Results indicate that the correlation between price and crime level was insignificant.



The correlation between crime frequency and price was -0.079, which is insignificant.

	price2	crime_freq
price2	1.000000	-0.079548
crime_freq	-0.079548	1.000000

Conclusion

The relationship between crime levels and price seems negligible, with a majority of low crimes that appeared to be around $1900 and $2400, similar to moderate and high crimes. The range for low crimes seems to range widely from $1100 to $4000, suggesting that neighborhood locations were better predictors of crime levels than price points.

Thus, if you are considering finding a cheaper place that is safe, I would first suggest picking anywhere NOT in Downtown LA and then search for an area with reasonable pricing. The further West you move, the less the level of crimes.