Battle of the Neighborhoods

Prime Restaurant Locations in Los Angeles, CA

The Problem

1. Introduction to the Business Problem

An eager new chef is looking to establish himself in the Los Angeles culinary scene, but he is not sure which location will be the most successful. Los Angeles has numerous locations with a robust restaurant scene, but data analysis is necessary to chose the correct location. Ensuring the right location could be make it or break it for this this chef and several factors will be considered in the location hunt. According to The Balance Small Business website, there are several pertinent decision making factors.

1. Visibility - Setting up in a high foot traffic area.
2. Size - How big or small is the restaurant going to be.
3. Neighborhood - This is two pronged, in that who else has a restaurant space as well as the crime statistics in the area.
4. Affordability - The cost of the lot.
5. Competitor Analysis - This goes hand in hand with Neighborhood but is important in it's own right.

In this project, we will use data from the Foursqaure API to address the aforementioned data points.

2. Data

We will attempt to address all the factors above, but will likely be limited to lack of availability of some data sets. The most important factors, and the ones we will work with will be *Visibility*, and *Crime Rates*. Competitor analysis will be inherent within the project.

Datasets:

1. Los Angeles Crime data from 2020 to Present - This dataset reflects incidents of crime in the City of Los Angeles dating back to 2020. This data is transcribed from original crime reports that are typed on paper and therefore there may be some inaccuracies within the data. Link
2. Los Angeles Neighborhoods - Official Certified Neighborhood Council boundaries in the City of Los Angeles created and maintained by the Bureau of Engineering / GIS Mapping Division. Link
3. Building Permits by Neighborhood - Link

Search Engines:

Foursqaure. We will use the foursquare API to carry out venue and point of interest search. The results will give us an idea of the neighborhood of the venue's potential location.

3. Methodology

For this project, we will explore Los Angeles Neighborhood and Crime data with the intent to answer the critical factors discussed in the intro. Following that, utilization of the Foursquare API will allow us to explore data within our neighborhoods, providing information on venues and other types of restaurants. The neighborhoods will be clustered using the k-mean algorithm. The culminating results will provide potential restaurant locations for our entrepreneurial chef.

The Data

We started with crime rate. This noted reported incidents from the start of 2020 to present day.


```
[3]: pd.set_option('display.max_columns', 180)
     df_crime.head()
```

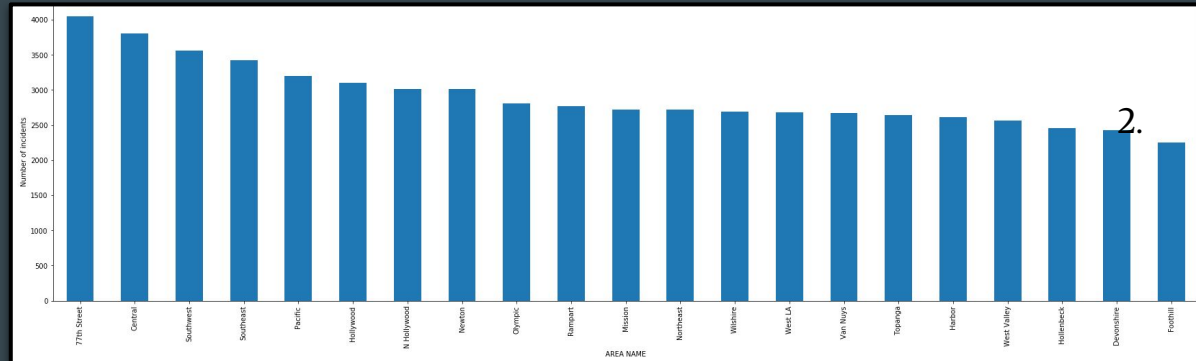|	DR_NO	Date Rptd	DATE OCC	TIME OCC	AREA	AREA NAME	Rpt Dist No	Part 1-2	Crm Cd	Crm Cd Desc	Mocodes	Vict Age	Vict Sex	Vict Descent	Premis Cd	Premis Desc	Weapon Used Cd	Weapon Desc	Status	Status Desc	Crm Cd 1	Crm Cd 2	Crm Cd 3	Crm Cd 4	LOCATION	Cross Street	LAT	LON
0	10304468	01/08/2020 12:00:00 AM	01/08/2020 12:00:00 AM	2230	3	Southwest	377	2	624	BATTERY - SIMPLE ASSAULT	0444 0913	36	F	B	501.0	SINGLE FAMILY DWELLING	400.0	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AO	Adult Other	624	NaN	NaN	NaN	1100 W 39TH PL	NaN	34.0141	-118.2978
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	0416 1822 1414	25	M	H	102.0	SIDEWALK	500.0	UNKNOWN WEAPON/OTHER WEAPON	IC	Invest Cont	624	NaN	NaN	NaN	700 S HILL ST	NaN	34.0459	-118.2545
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	1414 1218 2000 1814 0416 0447	53	F	B	502.0	MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)	400.0	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	IC	Invest Cont	626	NaN	NaN	NaN	300 E 5TH ST	NaN	34.0449	-118.2458
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)	0329 1402	76	F	W	502.0	MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)	NaN	NaN	IC	Invest Cont	745	998.0	NaN	NaN	5400 CORTEEN PL	NaN	34.1685	-118.4019
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...	0329	31	X	X	409.0	BEAUTY SUPPLY STORE	NaN	NaN	IC	Invest Cont	740	NaN	NaN	NaN	14400 TITUS ST	NaN	34.2198	-118.4468

Plots

1. There were several notable columns, but we opted to use But for our purpose, we only need 'Crime Description,' 'Area location,' 'Latitude,' and 'Longitude.' We dropped unwanted columns and cleaning up NaNs, the number of crimes can be plotted by categories:
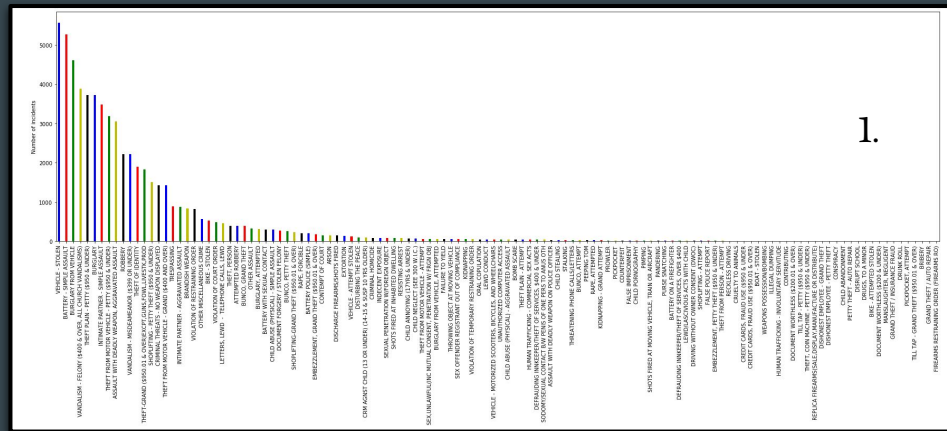
The highest reported crime is Vehicle-Stolen, Battery-Simple Assault, followed by Burglary From Vehicle.

2. Now for the distribution.. We took advantage of 'Area Name' Using the value_counts function, the total number of incidents grouped by Area is plotted as follows:

Crime is prevalent across most the areas in lA, with the 77th Street Area having the highest reported data followed by Central.

Locational Data

We needed to compare the data visually, this was done by combining 'Latitude' and 'Longitude' columns to form a GeoDataFrame.

```
[10]: crime = gpd.read_file('crime.shp')
      crime.head()
```

[10]:	DR_NO	Date_Rptd	DATE_OCC	TIME_OCC	AREA	AREA_NAME	Rpt_Dist_N	Part_1_2	Crm_Cd	Crm_Cd_Des	Mocodes	Vict_Age	Vict_Sex	Vict_Desce	Premis_Cd	Premis_Des	Weapon_Use	Weapon_Des	Status	Status_Des	Crm_Cd_1	Crm_Cd_2	Crm_Cd_3	Crm_Cd_4	LOCATION	Cross_Stre	LAT	LON	geometry
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	0444 0913	36	F	B	501	SINGLE FAMILY DWELLING	400	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	AO	Adult Other	624	0	None	None	1100 W 39TH PL	None	34.0141	-118.2978	POINT (-118.29780 34.01410)
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	0416 1822 1414	25	M	H	102	SIDEWALK	500	UNKNOWN WEAPON/OTHER WEAPON	IC	Invest Cont	624	0	None	None	700 S HILL ST	None	34.0459	-118.2545	POINT (-118.25450 34.04590)
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	1414 1218 2000 1814 0416 0447	53	F	B	502	MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)	400	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	IC	Invest Cont	626	0	None	None	300 E 5TH ST	None	34.0449	-118.2458	POINT (-118.24580 34.04490)
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)	0329 1402	76	F	W	502	MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)	0	None	IC	Invest Cont	745	998	None	None	5400 CORTEEN PL	None	34.1685	-118.4019	POINT (-118.40190 34.16850)
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...	0329	31	X	X	409	BEAUTY SUPPLY STORE	0	None	IC	Invest Cont	740	None	None	None	14400 TITUS ST	None	34.2198	-118.4468	POINT (-118.44680 34.21980)

We also plotted the neighborhood data.
Which will later be combined.

```
[12]: nbrhoods = gpd.read_file('la_neighborhoods.shp')
      nbrhoods.head()
```

[12]:	date_certi	time_certi	dwebsite	name	nc_id	objectid	service_re	waddress	geometry
0	2002-04-27	00:00:00.000	http://empowerla.org/LHNC	LINCOLN HEIGHTS NC	47.0	1.0	REGION 8 - NORTH EAST LA	http://www.lincolnheightsnc.org/	POLYGON ((-118.20503 34.05558, -118.20605 34.0...
1	2008-06-16	00:00:00.000	https://empowerla.org/voicesof90037/	VOICES OF 90037	109.0	2.0	REGION 9 - SOUTH LA 2	https://www.voicesnc.org/	POLYGON ((-118.28273 34.01114, -118.28234 34.0...
2	2002-06-11	00:00:00.000	http://empowerla.org/LFNC	LOS FELIZ NC	36.0	3.0	REGION 7 - EAST	http://www.losfeliznc.org/	POLYGON ((-118.30027 34.15793, -118.29883 34.1...
3	2005-04-19	00:00:00.000	http://empowerla.org/MHNC	MISSION HILLS NC	101.0	4.0	REGION 1 - NORTH EAST VALLEY	http://mhnconline.org/	POLYGON ((-118.46933 34.28050, -118.46931 34.2...
4	2002-09-10	00:00:00.000	http://empowerla.org/GTLNC	GREATER TOLUCA LAKE NC	28.0	5.0	REGION 4 - SOUTH EAST VALLEY	http://www.gtlnc.org/	POLYGON ((-118.35520 34.16491, -118.35529 34.1...

```
[13]: nbrhoods.plot(figsize=(12,14))
      plt.show()
```
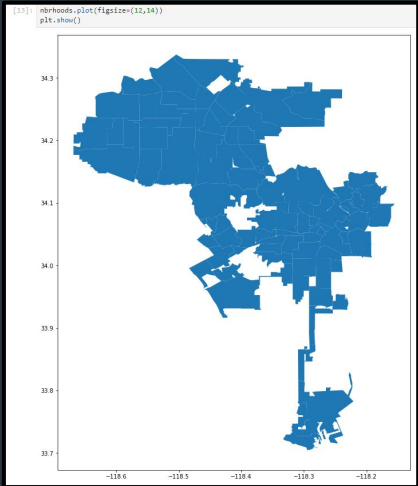
Spatial Join

The crime numbers for each neighborhood were aggregregated with an op=intersects parameter within the Pandas sjoin function.

```
[15]: nbh_crime_counts = gpd.tools.sjoin(crime.to_crs(nbrhoods.crs), nbrhoods, how="inner", op='intersects').groupby('name').size()
      nbh_crime_counts = pd.DataFrame(data=nbh_crime_counts.reset_index())
      nbh_crime_counts.columns=['name', 'incident_counts']
      nbh_crime_counts.head()
```

[15]:

	name	incident_counts
0	ARLETA NC	288
1	ARROYO SECO NC	196
2	ATWATER VILLAGE NC	141
3	BEL AIR-BEVERLY CREST NC	222
4	BOYLE HEIGHTS NC	1295

```
[16]: nbrhoods = nbrhoods.merge(nbh_crime_counts, on='name')
      nbrhoods.head()
```

[16]:

	date_certi	time_certi	dwebsite	name	nc_id	objectid	service_re	waddress	geometry	incident_counts
0	2002-04-27	00:00:00.000	http://empowerla.org/LHNC	LINCOLN HEIGHTS NC	47.0	1.0	REGION 8 - NORTH EAST LA	http://www.lincolnheightsnc.org/	POLYGON ((-118.20503 34.05558, -118.20605 34.0...	455
1	2008-06-16	00:00:00.000	https://empowerla.org/voicesof90037/	VOICES OF 90037	109.0	2.0	REGION 9 - SOUTH LA 2	https://www.voicesnc.org	POLYGON ((-118.28273 34.01114, -118.28234 34.0...	1053
2	2002-06-11	00:00:00.000	http://empowerla.org/LFNC	LOS FELIZ NC	36.0	3.0	REGION 7 - EAST	http://www.losfeliznc.org/	POLYGON ((-118.30027 34.15793, -118.29883 34.1...	548
3	2005-04-19	00:00:00.000	http://empowerla.org/MHNC	MISSION HILLS NC	101.0	4.0	REGION 1 - NORTH EAST VALLEY	http://mhnconline.org/	POLYGON ((-118.46933 34.28050, -118.46931 34.2...	262
4	2002-09-10	00:00:00.000	http://empowerla.org/GTLNC	GREATER TOLUCA LAKE NC	28.0	5.0	REGION 4 - SOUTH EAST VALLEY	http://www.gtlnc.org/	POLYGON ((-118.35520 34.16491, -118.35529 34.1...	161

Crime Map

A choropleth map created using Folium. This contains geospatial and statistical data.
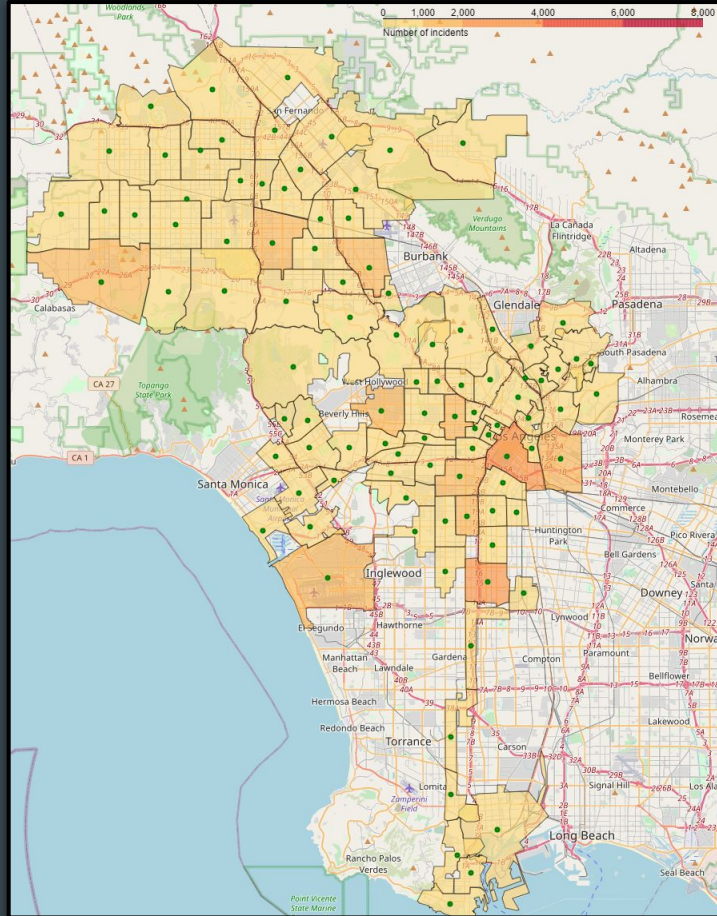
Venue Data

Using the Foursqare API we queried for the top ten most common venue for each neighborhood.

Following that, we plotted the clusters using the k-means algorithm and a representative_point() function from GeoPandas to extract a location that is located in the specific neighborhood


```
[26]:   print(LA_venues.shape)
        LA_venues.head()

        (1287, 7)

[26]:
```

	name	name Latitude	name Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	LINCOLN HEIGHTS NC	34.072527	-118.2099	The Heights Deli & Bottle Shop	34.073909	-118.210758	Pizza Place
1	LINCOLN HEIGHTS NC	34.072527	-118.2099	El Huarachito	34.073638	-118.209958	Mexican Restaurant
2	LINCOLN HEIGHTS NC	34.072527	-118.2099	King Taco Restaurant	34.073551	-118.211480	Mexican Restaurant
3	LINCOLN HEIGHTS NC	34.072527	-118.2099	Tejuino Los Reyes	34.073911	-118.213975	Juice Bar
4	LINCOLN HEIGHTS NC	34.072527	-118.2099	Hm157	34.073419	-118.208700	Music Venue
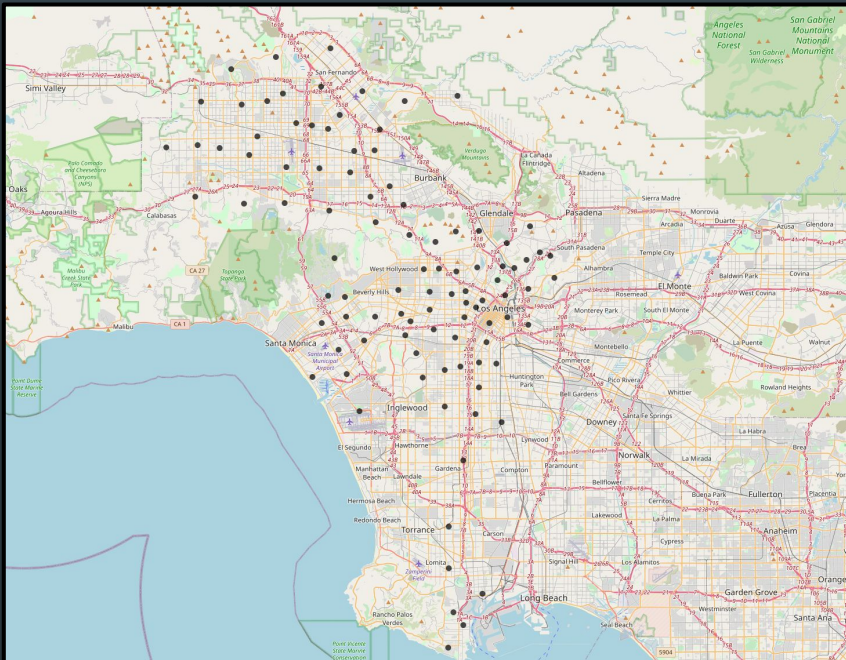
Examine Clusters

By analyzing public datasets obtained from the LA City data portal, we identified Elysian Valley as a possible neighborhood that is ideal for opening a new restaurant. We relied on aggregating crime data for neighborhoods in and around Los Angeles. With the help from the Foursquare API, we further demonstrated that Elysian Valley is a suitable candidate by highlighting a list of already successful restaurants. In addition, we determined that the proximity to Dodger Stadium would ultimately boost revenue due to increased foot traffic before and after sporting events. Lastly, while we have identified an ideal neighborhood, going forward we would like to identify additional information on our competitors and deeper demographic information on our proposed area.

1. Visibility: Our results reveal that **Elysian Valley** should receive notable foot traffic as well as commuter traffic. This is due to the proximity of the Golden State Freeway (I5), the large nearby park, and Dodger Stadium. Blake AVenue would likely serve as the best location to buy a property. Further analysis on the pricing/availability on this street is needed to pinpoint a precise lot.

2. Competitor analysis: In the above sections, we very listed the categories of venues in **Elysian Valley.** This list provides an excellent overview of our competitors. It also guides us in determining which category we should be focusing on or avoiding. An improvement would be to cluster the restaurant categories. The results could let us know the landscape of the restaurant business in this neighborhood.

In this capstone project we identified an eager new chef who was looking to establish himself in the Los Angeles culinary scene, but he was not sure which location will be the most successful. Los Angeles has numerous locations with a robust restaurant scene, but data analysis is necessary to choose the correct location. Ensuring the right location could be make it or break it for this this chef and several factors will be considered in the location hunt. We identified the most important factors that could impact the choice. Using crime and neighborhood datasets, located a possible neighborhood that has a relatively low crime rate and prime location. The Foursquare API recommendation results also seem to support our pick.