The Battle of Neighborhoods - Los Angeles County

By Yaxuan Han

May 2020

Contents

Introduction

Background of the project

Talking about Los Angeles, what comes to your mind? The world-famous Hollywood? Mouth-watering gourmet cheeseburgers? Or the-happiest-place-in-the-world Disneyland? Los Angeles (L.A.) is indeed the ultimate destination of food and entertainment, as well as a dominant financial hub in United States. By July 2019, L.A. County has an estimated population of 10,039,107 according to U.S. Census [1]. L.A. County was the largest county by output, with its $710.9 billion GDP — equivalent to Saudi Arabia — accounting for 3.8% of the U.S. total. L.A. County added $395.2 billion to total U.S. GDP from 2001 to 2018. New York County, home to Manhattan, added $340 billion [2]. Meanwhile, the cost of living at L.A. is very high. By March 2020, L.A has a median home value of $664,495, versus the nation average of $247,084 [3].

Since L.A. is the dream home of many of people, it is interesting to discover the lifestyle and what is around each neighborhood. Therefore, the purpose of this project is to explore cities in L.A County, in terms of median home price and common venues nearby.

Potential Stakeholders

1. People who are relocating to L.A. from a different county/state/country,
2. People who are investing/purchasing a home at L.A.,
3. People who are comparing L.A. with another metropolis.

Data

- Source: **Zillow**, link: https://www.zillow.com/los-angeles-ca/home-values/
    - To get general real estate statistics for Los Angeles County.
- Source: **Zillow**, link: https://www.zillow.com/market-report/05-20/12447/los-angeles-ca.xls?rt=6
    - To export EXCEL that contains names and current median price for L.A. neighborhoods.
- Source: **Neighborhood Data for Social Change by USC**, link: https://usc.data.socrata.com/dataset/Los-Angeles-Neighborhood-Map/r8qd-yxsr
    - To get GeoJSON file of latitude, longitude, and boundary data. They are essential for this project to build location table for all cities, Chloropleth map that reflects different

levels of home price (the darker the color, the higher the median home price), and the clustering of neighborhods based on nearby venue features.

- Source: **Foursquare API**, query: https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}
    - To explore nearby venues, get venue name category, analyze the most common venues of each neighborhood, and cluster neighborhoods to 5 groups based on above criteria.

Methodology

Import Data

Median home price data is imported from Zillow and read as a panda data-frame. The first five rows of the table are shown below:

	Region Name	Region Type	Type	Current	Month Over Month	Quarter Over Quarter	Year Over Year	5 Year Annualized	10 Year Annualized	Current.1	Month Over Month.1	Quarter Over Quarter.1
0	Los Angeles	city	All Homes	752500	0.011249	0.030328	0.058073	0.0628	0.0549	0.1049	0.0386058	-0.0412387
1	Acton	city	All Homes	634700	0.007408	0.018977	0.035056	0.0488	0.0473	---	---	---
2	Agoura Hills	city	All Homes	901400	0.008089	0.020858	0.035714	0.0385	0.0362	0.101	0.0317032	-0.0339105
3	Agua Dulce	city	All Homes	717200	0.007156	0.020719	0.034532	0.0486	0.036	---	---	---
4	Alhambra	city	All Homes	707500	0.013305	0.033073	0.047434	0.0551	0.0465	0.1071	0.0171429	0.00495308

From the Zillow table, only ‘Region Name’ (Neighborhood) and ‘Current’ (Median Price) columns are needed and they turne into a new table – *Price Table* (only shows first five rows):

	Neighborhood	Price
0	Los Angeles	752500
1	Acton	634700
2	Agoura Hills	901400
3	Agua Dulce	717200
4	Alhambra	707500

Since location detail is needed, a second table from USC is imported with latitude and longitude data:

ternal_i	name	display_na	sqmi	type	name_1	slug_1	latitude	longitude	location
acton	Acton	Acton L.A. County Neighborhood (Current)	39.339109	unincorporated-area	NaN	NaN	-118.169810	34.497355	POINT(34.497355239240846 -118.16981019229348)
adams-rmandie	Adams-Normandie	Adams-Normandie L.A. County Neighborhood (Curr...	0.805350	segment-of-a-city	NaN	NaN	-118.300208	34.031461	POINT(34.031461499124156 -118.30020800000011)
agoura-hills	Agoura Hills	Agoura Hills L.A. County Neighborhood (Current)	8.146760	standalone-city	NaN	NaN	-118.759885	34.146736	POINT(34.146736499122795 -118.75988450000015)
agua-dulce	Agua Dulce	Agua Dulce L.A. County Neighborhood (Current)	31.462632	unincorporated-area	NaN	NaN	-118.317104	34.504927	POINT(34.504926999796837 -118.3171036690717)
hambra	Alhambra	Alhambra L.A. County Neighborhood (Current)	7.623814	standalone-city	NaN	NaN	-118.136512	34.085539	POINT(34.085538999123571 -118.13651200000021)

Data Cleaning

Again, the original table from USC is too informative and only name, latitude and longitude data (call it Location Table) will be withdrawn. After merging *Price Table* and *Location Table*, we got a *Result Table*, which has exactly all the information needed – 'Neighborhood', 'Price', 'Latitude' and 'Longitude'. *Result Table* shows the median home price in a descending order. 10 most expensive cities (index 0-9) and 10 lease expensive cities (index 101-110) are presented below:

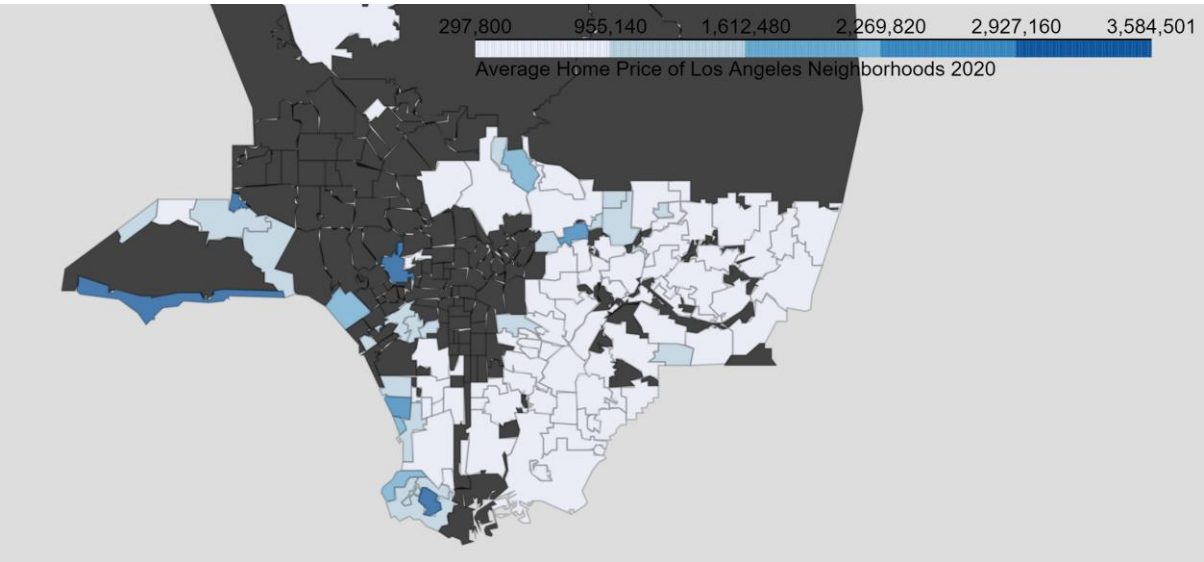
(Note: When merging Location and Price tables, only the intersection part of data is kept based on algorithm rule. Since 'Neighborhood' is the 'shared key' of both tables, cities without Zillow median price information are dropped automatically. Therefore, only 111 out of 158 cities (total number of cities in L.A. County) are represented in the Result Table.)

	Neighborhood	Price	Latitude	Longitude
0	Beverly Hills	3584500	34.082544	-118.399534
1	Hidden Hills	3445300	34.164057	-118.657056
2	Rolling Hills	3173400	33.762683	-118.348127
3	Malibu	3172100	34.033895	-118.754254
4	Manhattan Beach	2641900	33.889508	-118.400939
5	San Marino	2350800	34.121300	-118.115054
6	Palos Verdes Estates	2137800	33.783317	-118.390122
7	Hermosa Beach	1855000	33.864771	-118.397142
8	Santa Monica	1738200	34.021860	-118.480567
9	La Canada Flintridge	1735300	34.210687	-118.200470

101	Paramount	464900	33.899154	-118.165391
102	Willowbrook	462000	33.915711	-118.252312
103	Maywood	461400	33.988339	-118.187303
104	Pomona	461000	34.065723	-117.769942
105	Compton	444200	33.893252	-118.221810
106	Palmdale	340300	34.585422	-118.108094
107	Green Valley	330200	34.620007	-118.414217
108	Lancaster	310300	34.700066	-118.131791
109	Littlerock	298000	34.521322	-117.978276
110	Lake Hughes	297800	34.671409	-118.459334

A Chloropleth Map is generated from *Result Table* using **Folium package**:

(Note: The gray area represents neighborhoods without pricing data. It could be the lack of data, or non-residential nature of the area (most of the gray areas are mountain ranges, valleys, forests, islands, lakes, rivers, and desert).



Since L.A. is on the west coast of the continent, it borders 70 miles (110 km) of coast on the Pacific Ocean. Without surprise, neighborhoods that are closer to the coastline are always more expensive. Homes in Malibu and Beverly Hills (dark-blue) may cost around $ 3 million dollars, while Pomona and Lancaster (white/sky-blue) cost around $30 – $40 thousand dollars.

Foursquare API

In this project, Foursquare API is used to explore venues around each city. The query has following factors: *client_id, client_secret, version, latitude, longitude, radius, and limit*. Radius is set at 490 meters from the city center and limit is set at 90 (due to limitations for free account, API doesn't not allow anything pass 500 meters on my end). Venue information return as a table below: For example, the first row – In Beverly Hills, there is a food truck named Greenz on Wheelz Truck locates at (34.08N, -118.39W).

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Beverly Hills	34.082544	-118.399534	Greenz On Wheelz Truck	34.082529	-118.396762	Food Truck
1	Beverly Hills	34.082544	-118.399534	Harvard's	34.082703	-118.398938	Sake Bar
2	Rolling Hills	33.762683	-118.348127	Hillside Florist	33.760418	-118.347301	Flower Shop
3	Malibu	34.033895	-118.754254	Latigo Beach	34.030577	-118.754597	Surf Spot
4	San Marino	34.121300	-118.115054	San Marino Club	34.118743	-118.113135	Nightlife Spot

In order to find out what is nearby each city, the following table is made to show Top 10 most common venues:

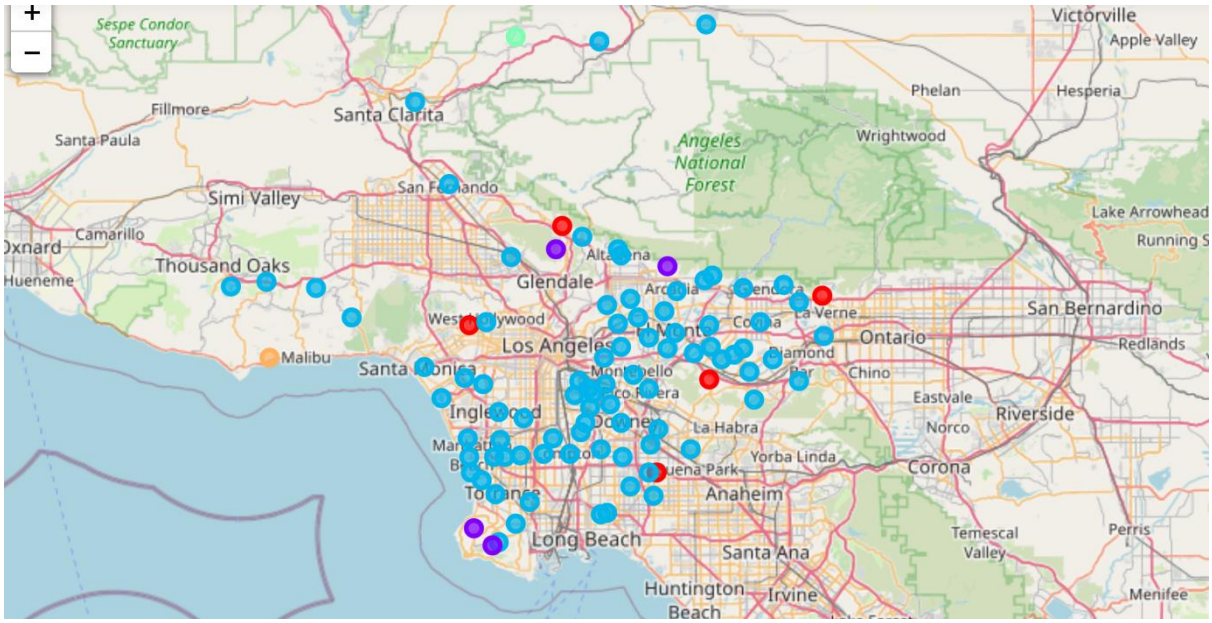
Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Acton	Construction & Landscaping	Yoga Studio	Event Service	French Restaurant	Food Truck	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market
Agoura Hills	Fast Food Restaurant	Chinese Restaurant	Breakfast Spot	Hotel	Bakery	Restaurant	Multiplex	Shipping Store	Café	Mexican Restaurant
Agua Dulce	Airport	Yoga Studio	Event Service	French Restaurant	Food Truck	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market
Alhambra	Convenience Store	Mexican Restaurant	Hardware Store	Sporting Goods Shop	Fast Food Restaurant	Bagel Shop	Construction & Landscaping	Breakfast Spot	Video Store	Pizza Place
Alondra Park	Park	Home Service	Baseball Field	Event Service	Food Truck	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market

Clustering

'Cluster Labels' column is created and each city is assigned to a cluster (0,1,2,3,4) based on the most common venues.

	Neighborhood	Price	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
0	Beverly Hills	3584500	34.082544	-118.399534	0.0	Food Truck	Sake Bar	Yoga Studio	Electronics Store	Food Court	Food & Drink Shop
1	Hidden Hills	3445300	34.164057	-118.657056	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	Rolling Hills	3173400	33.762683	-118.348127	2.0	Flower Shop	Yoga Studio	Event Service	French Restaurant	Food Truck	Food Court
3	Malibu	3172100	34.033895	-118.754254	4.0	Surf Spot	Yoga Studio	Event Service	Food Truck	Food Court	Food & Drink Shop
4	Manhattan Beach	2641900	33.889508	-118.400939	2.0	Hot Dog Joint	Shipping Store	Automotive Shop	Sports Bar	Health & Beauty Service	Bank

Visualizing five clusters on the map:



From the map above, each color represents a cluster/ group. Apparently, most cities are clustered in ‘blue’, and only a few are clustered as ‘purple’ ‘red’ ‘yellow’ and ‘green’.

Please see table below for a list of 5 cities in first cluster (**red**):

	Neighborhood	Price	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
0	Beverly Hills	3584500	Food Truck	Sake Bar	Yoga Studio	Electronics Store	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market
28	La Crescenta-Montrose	959100	Park	Yoga Studio	Food Truck	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market	Fish & Chips Shop
39	Cerritos	753200	Food Truck	Korean Restaurant	Yoga Studio	Electronics Store	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market
51	Hacienda Heights	679000	Park	Yoga Studio	Food Truck	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market	Fish & Chips Shop
52	La Verne	676400	Park	Food Truck	Yoga Studio	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market	Fish & Chips Shop

Please see table below for a list of 4 cities in second cluster (**purple**):

	Neighborhood	Price	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
6	Palos Verdes Estates	2137800	Convenience Store	Trail	Yoga Studio	Electronics Store	Food Truck	Food Court	Food & Drink Shop	Food	Flower Shop
14	Rancho Palos Verdes	1360000	Trail	Yoga Studio	Electronics Store	Food Truck	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market
23	Sierra Madre	1061400	Trail	Coffee Shop	Gourmet Shop	Yoga Studio	Electronics Store	Food Court	Food & Drink Shop	Food	Flower Shop
32	Glendale	886000	Trail	Yoga Studio	Electronics Store	Food Truck	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market

Please see table below for a list of 100 cities in third cluster (**blue**). Only the first five rows are shown:

	Neighborhood	Price	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th M Comm Ver
2	Rolling Hills	3173400	Flower Shop	Yoga Studio	Event Service	French Restaurant	Food Truck	Food Court	Food & Drink Shop	Food	Flea Mar
4	Manhattan Beach	2641900	Hot Dog Joint	Shipping Store	Automotive Shop	Sports Bar	Health & Beauty Service	Bank	Gastropub	Yoga Studio	Fish & Ch Sh
5	San Marino	2350800	Garden	Food Truck	Nightlife Spot	Event Service	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Mar
7	Hermosa Beach	1855000	Mexican Restaurant	Hotel	Coffee Shop	Juice Bar	Sushi Restaurant	American Restaurant	Grocery Store	Gym / Fitness Center	Ital Restaur
8	Santa Monica	1738200	Cycle Studio	Performing Arts Venue	Skate Park	Supermarket	Beach	New American Restaurant	Taco Place	Coffee Shop	P

Please see table below for a list of 1 city in fourth cluster (**yellow**):

	Neighborhood	Price	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	C
43	Agua Dulce	717200	Airport	Yoga Studio	Event Service	French Restaurant	Food Truck	Food Court	Food & Drink Shop	Food	Flower Shop	

Please see table below for a list of 1 city in fifth cluster (**green**):

	Neighborhood	Price	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	C
3	Malibu	3172100	Surf Spot	Yoga Studio	Event Service	Food Truck	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market	

Results

Please find clustering summary in the table below:

Cluster Color	Number of Neighborhoods	Most Common Venues	Price Range
Red	5	Park	$60 k - $3 mil
Blue	100	Variety of restaurants	$20k - $3 mil
Purple	4	Trail	$80k - $2 mil
Yellow	1	Airport	$70 k
Green	1	Surf Spot	$3 mil

Discussion

Scope and Limitation

This project aims to explore the each neighborhood in Los Angeles County in terms of median home price and common venues nearby. Several data science tools were adopted, including: Jupyter Notebook, Pandas, Folium and Foursquare API. These tools are essential for mapping diagrams and data analysis.

Although the data source and tools are robust, this project has certain flaws, and these areas can be improved through:

1. Wider radius and limit for Foursquare API query. In Southern California, everything is distant from each other, and there is not much to get within walking distance. With a radius of 490 meters and 90 limits query, results are limited. This query combination may work better for metropolis that are walking-friendly, such as New York and Sydney. Ideally when there is no query limit, a radius of 3 mile would generate more accurate results.

2. Detailed location data, for example by street name for each neighborhood. The home price in L.A. County is highly dispersed. Home price from one street to another can be significantly different. The result might be more comparable if ‘Zip Code’ or ‘Main Street Address’ column was added.

3. Specify the type of home for median home price. In *Price Table*, the median home price for ‘All types of home’ is picked. This is due to general approach of this project. In fact, Zillow dataset categorized it further into ‘Single-family homes’, ‘condos’ and ‘duplex’ categories for extensive analysis.

Conclusion

During the battle, neighborhoods from L.A. County are highly similar in terms pf common venues nearby. However, the median home prices are widely spread from one to another. Afterall, real estate value is not really about stores and restaurants around, but more based more on location. Generally, a seaside home is far more valuable than a desert house; an old, small downtown apartment may have the value of a suburb single-family home. Living at a particular neighborhood should be less impacted by venues around (since most neighborhoods has similar types of venues), but more by demographical and cultural factors.

References

[1] https://www.census.gov/quickfacts/losangelescountycalifornia

[2] https://www.latimes.com/business/story/2019-12-19/los-angeles-largest-economy

[3] https://www.businessinsider.com/average-home-prices-in-every-state-washington-dc-2019-6#51-west-virginia-108236-1