Capstone Project - The Battle of Neighborhoods

A description of the problem and a discussion of the background.

Compare Neighborhoods in the city of Los Angeles to find the Best Place to start a restaurant business.

Introduction and business problem selection:

City of Los Angeles often known by its initials L.A., being the most popular city in California, and culturally very diverse, starting any kind of a restaurant business would be a great business opportunity.

Los Angeles is the cultural, financial, and commercial center of Southern California. The city is known for its Mediterranean climate, ethnic diversity, Hollywood, the entertainment industry, and its sprawling metropolis. [1]

Los Angeles lies in a basin, adjacent to the Pacific Ocean, with mountains as high as 10,000 feet (3,000 m), and deserts. The city, which covers about 469 square miles (1,210 km2). [1]

In this project I will explore the opportunities to start a new restaurant business and understand the City of Los Angeles, by its demography, population and the current restaurants businesses located in and around the city. Furthermore, I will investigate on current popular or most liked restaurants in each neighborhood.

My stakeholders looking for data to prove if it is worth to start a new restaurant business in L.A. and to help them to better understand the city using following criteria's.

- Population of the city and its ethnically-diversity
    - Compare Los Angeles Population by Year.
    - Demonstrating population by Race and segregating it based on Hispanic and Non-Hispanic population.
    - Compare Los Angeles Population by Age
- Income of the people
    - Visualize Income by Household Type
- Weather conditions
    - Hottest and coldest, wettest and driest averages for a month
- Restaurants
    - Number of restaurants in each neighborhood
    - Number of restaurants based on cuisines in each neighborhood
    - List of top 100 restaurants
    - List of trending or liked restaurants
- Overall Comparison of the restaurants
    - Conclusion on if it is worth opening a restaurant in L.A. and
    - Best Neighborhood(s) to start the business

A description of the data and how it will be used to solve the problem.

Data section:

The city is divided into many different districts and neighborhoods, some of which were incorporated cities that merged with Los Angeles.[2] These neighborhoods were developed piecemeal and are well-defined enough that the city has signage marking nearly all of them.[1] LA city has 15 districts and 272 neighborhoods. Just targeting one district Central LA. More-than 450,000 people living in this district aged between 19 to 49 and these are considered as the spending population who like to visit restaurants.

81.4% of households are renters. The neighborhood with the highest rental rate is Westlake, and the neighborhood with the lowest rental rate is Hollywood Hills West.[3] Renters are considered as moving and relocating population with higher rented population has some advantage on restaurant businesses.

Based on the criteria's I will be using the approaches such as web scraping, segmentation, clustering and creating data visualization to get a better understanding of the city. All the data for the project is collected from different sources list of data source as follows.

Data description and data source that will be used to solve the problem

- Population of the city and its ethnically-diversity are analyzed by comparing Population by Year trend. This will help in visualizing the growth of the city by population.
  Source: Annual Estimates of the Resident Population: April 1, 2010 to July 1, 2018
  Sample:

Geography	April 1, 2010		Population Estimate (as of July 1)								
	Census	Estimates Base	2010	2011	2012	2013	2014	2015	2016	2017	2018
United States	308,745,538	308,758,105	309,326,085	311,580,009	313,874,218	316,057,727	318,386,421	320,742,673	323,071,342	325,147,121	327,167,434
Alabama	4,779,736	4,780,138	4,785,448	4,798,834	4,815,564	4,830,460	4,842,481	4,853,160	4,864,745	4,875,120	4,887,871
Alaska	710,231	710,249	713,906	722,038	730,399	737,045	736,307	737,547	741,504	739,786	737,438
Arizona	6,392,017	6,392,288	6,407,774	6,473,497	6,556,629	6,634,999	6,733,840	6,833,596	6,945,452	7,048,876	7,171,646
Arkansas	2,915,918	2,916,028	2,921,978	2,940,407	2,952,109	2,959,549	2,967,726	2,978,407	2,990,410	3,002,997	3,013,825
California	37,253,956	37,254,523	37,320,903	37,641,823	37,960,782	38,280,824	38,625,139	38,953,142	39,209,127	39,399,349	39,557,045
Colorado	5,029,196	5,029,316	5,048,281	5,121,771	5,193,721	5,270,482	5,351,218	5,452,107	5,540,921	5,615,902	5,695,564
Connecticut	3,574,097	3,574,147	3,579,125	3,588,023	3,594,395	3,594,915	3,594,783	3,587,509	3,578,674	3,573,880	3,572,665
Delaware	897,934	897,934	899,595	907,316	915,188	923,638	932,596	941,413	949,216	957,078	967,171
District of Columbia	601,723	601,766	605,085	619,602	634,725	650,431	662,513	675,254	686,575	695,691	702,455
Florida	18,801,310	18,804,580	18,845,785	19,093,352	19,326,230	19,563,166	19,860,330	20,224,249	20,629,982	20,976,812	21,299,325
Georgia	9,687,653	9,688,709	9,711,810	9,801,578	9,901,496	9,973,326	10,069,001	10,181,111	10,304,763	10,413,055	10,519,475
Hawaii	1,360,301	1,360,307	1,363,963	1,379,252	1,394,905	1,408,453	1,414,862	1,422,484	1,428,105	1,424,203	1,420,491
Idaho	1,567,582	1,567,657	1,570,773	1,583,828	1,595,441	1,611,530	1,631,479	1,651,523	1,682,930	1,718,904	1,754,208
Illinois	12,830,632	12,831,572	12,840,762	12,867,291	12,884,119	12,898,269	12,888,962	12,864,342	12,826,895	12,786,196	12,741,080
Indiana	6,483,802	6,484,061	6,490,436	6,516,045	6,537,640	6,568,367	6,593,533	6,608,296	6,633,344	6,660,082	6,691,878
Iowa	3,046,355	3,046,872	3,050,767	3,066,054	3,076,097	3,093,078	3,109,504	3,121,460	3,131,785	3,143,637	3,156,145

- Demonstrating population by Race and segregating it based on Hispanic and Non-Hispanic population are compared to show what kind of people live in the city and how to cater them and to analyze what kind of food they may prefer.
  Source:
  US Census 2018 Estimate ACS 1-Year Survey- Hispanic or Latino origin by race (Table B03002)
  Sample:

	Los Angeles County, California
	Estimate
Total:	10,105,518
Not Hispanic or Latino:	5,190,231
White alone	2,619,709
Black or African American alone	783,932
American Indian and Alaska Native alone	18,875
Asian alone	1,473,462
Native Hawaiian and Other Pacific Islander alone	25,303
Some other race alone	30,761
Two or more races:	238,189
Two races including Some other race	12,528
Two races excluding Some other race, and three or more races	225,661
Hispanic or Latino:	4,915,287
White alone	2,564,403
Black or African American alone	27,544
American Indian and Alaska Native alone	57,214
Asian alone	19,092
Native Hawaiian and Other Pacific Islander alone	3,073
Some other race alone	2,045,788
Send Feedback	198,173

- Compare Los Angeles Population by Age, this comparison will help to understand the age group of people living in the city and what kind of food choice they may make.
  Source: US Census 2018 Estimate ACS 1-Year Survey- Age and Sex (Table S0101)
  Sample:

	Los Angeles County, California				
	Total	Percent	Male	Percent Male	Female
	Estimate	Estimate	Estimate	Estimate	Estimate
Total population	10,105,518	(X)	4,982,529	(X)	5,122,989
AGE					
Under 5 years	602,507	6.0%	309,399	6.2%	293,108
5 to 9 years	575,136	5.7%	293,989	5.9%	281,147
10 to 14 years	640,817	6.3%	328,518	6.6%	312,299
15 to 19 years	628,872	6.2%	319,566	6.4%	309,306
20 to 24 years	701,010	6.9%	352,945	7.1%	348,065
25 to 29 years	856,123	8.5%	435,807	8.7%	420,316
30 to 34 years	783,915	7.8%	401,667	8.1%	382,248
35 to 39 years	721,419	7.1%	370,758	7.4%	350,661
40 to 44 years	659,703	6.5%	321,754	6.5%	337,949
45 to 49 years	686,881	6.8%	340,062	6.8%	346,819
50 to 54 years	661,688	6.5%	327,277	6.6%	334,411
55 to 59 years	637,656	6.3%	307,249	6.2%	330,407
60 to 64 years	573,832	5.7%	277,115	5.6%	296,717
65 to 69 years	450,513	4.5%	210,929	4.2%	239,584

- Income of the people is one of the most crucial criteria for any business to analyze the spending capacity of the people and to analyze group of people belong to each income category.
  Source:
  US Census 2018 Estimate ACS 1-Year Survey- Income in the past 12 months (Table S1901)
  Sample:

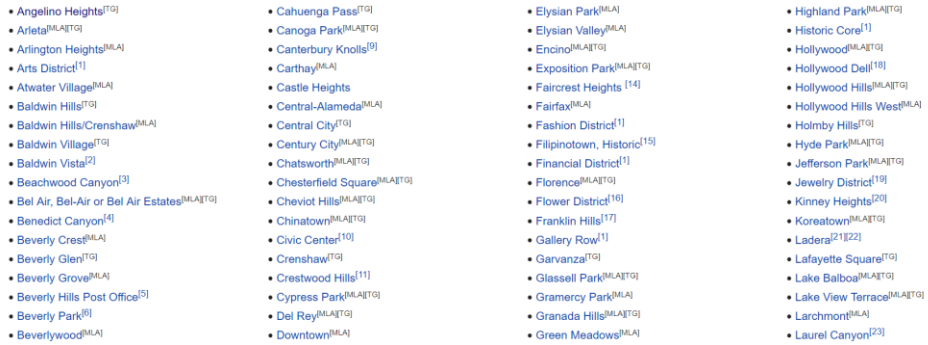
	Los Angeles County, California			
	Households	Families	Married-couple families	Nonfamily households
	Estimate	Estimate	Estimate	Estimate
Total	3,313,908	2,193,349	1,475,430	1,120,559
Less than $10,000	6.0%	3.8%	1.8%	11.6%
$10,000 to $14,999	4.6%	2.3%	1.4%	9.5%
$15,000 to $24,999	8.2%	7.4%	5.6%	10.9%
$25,000 to $34,999	7.8%	7.7%	6.0%	8.8%
$35,000 to $49,999	11.2%	11.3%	9.7%	11.0%
$50,000 to $74,999	16.2%	16.5%	15.2%	15.6%
$75,000 to $99,999	12.3%	12.9%	13.0%	10.5%
$100,000 to $149,999	16.0%	17.5%	20.2%	11.9%
$150,000 to $199,999	7.6%	8.6%	10.8%	4.5%
$200,000 or more	10.2%	12.0%	16.3%	5.6%

- Weather conditions will help the stakeholders to understand the pattern of the city weather such that they can accommodate the infrastructure needed to cater to different weather conditions and investment they may have to make to build such infrastructure.
  Hottest and coldest, wettest and driest averages for a month From 1895–2019 Source: Wikimedia Foundation, Inc.
  Sample:

Hottest and coldest, wettest and driest averages for a month (f/inch), 1895–2019[92]

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Hottest	63.9	64.2	67.5	68.2	71.5	75.9	79.8	79.0	80.3	75.4	66.9	62.2
Coldest	46.7	51.1	52.0	55.2	57.2	62.9	66.2	66.3	63.1	57.8	55.2	49.4
Wettest	14.43	15.23	10.44	7.31	3.83	0.98	0.43	2.54	5.13	5.13	9.96	11.46
Driest	0	0	0	0	0	0	0	0	0	0	0	0

- Analyzing current restaurants established in each neighborhood, the current cuisines they offer and to know the top 100 restaurants and its trends will help the stakeholder to get more clarity of the market and its conditions.
  Source:
  https://en.wikipedia.org/wiki/List_of_districts_and_neighborhoods_of_Los_Angeles#Notes , Wikimedia Foundation, Inc. (Retrieved Nov, 2019) with Foursquare location data
  Sample:



- Angelino Heights[TG]
- Arleta[MLA][TG]
- Arlington Heights[MLA]
- Arts District[1]
- Atwater Village[MLA]
- Baldwin Hills[TG]
- Baldwin Hills/Crenshaw[MLA]
- Baldwin Village[TG]
- Baldwin Vista[2]
- Beachwood Canyon[3]
- Bel Air, Bel-Air or Bel Air Estates[MLA][TG]
- Benedict Canyon[4]
- Beverly Crest[MLA]
- Beverly Glen[TG]
- Beverly Grove[MLA]
- Beverly Hills Post Office[5]
- Beverly Park[6]
- Beverlywood[MLA]
- Cahuenga Pass[TG]
- Canoga Park[MLA][TG]
- Canterbury Knolls[9]
- Carthay[MLA]
- Castle Heights
- Central-Alameda[MLA]
- Central City[TG]
- Century City[MLA][TG]
- Chatsworth[MLA][TG]
- Chesterfield Square[MLA][TG]
- Cheviot Hills[MLA][TG]
- Chinatown[MLA][TG]
- Civic Center[10]
- Crenshaw[TG]
- Crestwood Hills[11]
- Cypress Park[MLA][TG]
- Del Rey[MLA][TG]
- Downtown[MLA]
- Elysian Park[MLA]
- Elysian Valley[MLA]
- Encino[MLA][TG]
- Exposition Park[MLA][TG]
- Faircrest Heights [14]
- Fairfax[MLA]
- Fashion District[1]
- Filipinotown, Historic[15]
- Financial District[1]
- Florence[MLA][TG]
- Flower District[16]
- Franklin Hills[17]
- Gallery Row[1]
- Garvanza[TG]
- Glassell Park[MLA][TG]
- Gramercy Park[MLA]
- Granada Hills[MLA][TG]
- Green Meadows[MLA]
- Highland Park[MLA][TG]
- Historic Core[1]
- Hollywood[MLA][TG]
- Hollywood Dell[18]
- Hollywood Hills[MLA][TG]
- Hollywood Hills West[MLA]
- Holmby Hills[TG]
- Hyde Park[MLA][TG]
- Jefferson Park[MLA][TG]
- Jewelry District[19]
- Kinney Heights[20]
- Koreatown[MLA][TG]
- Ladera[21][22]
- Lafayette Square[TG]
- Lake Balboa[MLA][TG]
- Lake View Terrace[MLA][TG]
- Larchmont[MLA]
- Laurel Canyon[23]

Finally, an overall Comparison of the restaurants will be drawn, and a conclusion will be provided based on findings, if it is a viable option to open a restaurant in L.A. if so which are the best Neighborhood(s) to start the business.

Criteria's are analyzed using following approaches.

- Transform the data into a pandas data-frame
- Using geopy library to get the latitude and longitude values
- Creating a map of LA with neighborhoods superimposed on top.
- Segment and cluster only the neighborhoods
- The script queries the official Foursquare API to search for venues, and then query additional metadata for each unique venue.
- Finding the top 100 venues that are in LA within a radius of 500 meters.
- Extracting the category of the venue
- Exploring neighborhood along with the top 5 most common venues
- Finding the top 10 venues for each neighborhood
- Clustering Neighborhoods by k-means to cluster the neighborhood into 5 clusters and visualizing the resulting clusters

Reference:

1. https://en.wikipedia.org/wiki/Los_Angeles, Wikimedia Foundation, Inc. (Retrieved Nov, 2019)
2. Neighborhoods list: http://maps.latimes.com/neighborhoods/neighborhood/list/ , Mapping L.A. Los Angeles Times and https://en.wikipedia.org/wiki/List_of_districts_and_neighborhoods_of_Los_Angeles#Notes , Wikimedia Foundation, Inc. (Retrieved Nov, 2019)
3. "Central L.A.," Mapping L.A. Los Angeles Times (Retrieved Nov, 2019), http://maps.latimes.com/neighborhoods/region/central-la/, Mapping L.A. Los Angeles Times (Retrieved Nov, 2019).