COURSERA CAPSTONE PROJECT

The Battle of Neighborhoods

DECEMBER 4, 2019

GAUTAM A

Contents

Discovering the best neighborhood to start a restaurant business in the city of Los Angeles.

1 Introduction

1.1 Description of the Background

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. In this project I will be comparing neighborhoods of Los Angeles to find the best neighborhood to start a restaurant business.

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]

1.2 Problem & Interest

Staring a new business need lot of investments and time, for the investments to yield fruitful results the choices made should be backed by research and not by instincts. By analyzing the city population and current business trends provide an insight on city lifestyle and people's choice.
To exploring the opportunities to start a new restaurant business it is essential to understand the City by its demography, population and the current restaurants businesses located in and around the city.

A data driven presentation to prove if it is worth to start a new restaurant business in L.A. and to better understand the city analysis is done based on following criteria's

- Population of the city and its ethnically-diversity
    - Compare Los Angeles Population by Year.
    - Demonstrating population by Race
    - Compare Los Angeles Population by Age & Sex
- Income of the people
    - Compare Income by Household Type
- Weather conditions
    - Hottest and coldest, wettest and driest averages for a month
- Exploring neighborhoods
    - Number of Hangout places in Top 20 Neighborhoods
    - Number of Restaurants in Top 20 Neighborhoods
- Overall Comparison of the restaurants
    - Conclusion on if it is worth opening a restaurant in L.A. and
    - Best Neighborhoods to start the business

2 Data acquisition and cleaning

2.1 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, L.A. has 15 districts and 272 neighborhoods.

Based on the criteria's defined in Problem & Interest section above. All the data for the project is collected from different sources as discussed in further sections.

2.2 Data sources

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

2.2.1 Population of the city

Population of the city is analyzed by comparing Population by Year trend. This will help in visualizing the growth of the city by population.

Source: Total population 2009-2017 American Community Survey Estimates Population.

2.2.2 Population by Race

Population by race comparison is done to demonstrate the type of people living in the city and the data is segregated into Hispanic and Non-Hispanic population.

Source: US Census 2018 Estimate ACS 1-Year Survey- Hispanic or Latino origin by race (Table B03002)

2.2.3 Compare Los Angeles Population by Age & Sex

This comparison will help in understanding the age group of people living in the city.

Source: US Census 2018 Estimate ACS 1-Year Survey- Age and Sex (Table S0101)

2.2.4 Income of the people

Income data will aid to analyze the spending capacity of the people and to analyze group of people belonging to each income category.

Source: US Census 2018 Estimate ACS 1-Year Survey- Income in the past 12 months (Table S1901)

2.2.5 Weather conditions

Weather conditions will assist in understanding the pattern of the city weather, this information will help in building suitable infrastructure.

Source: Wikimedia Foundation, Inc. I have scraped the wiki table Hottest and coldest, wettest and driest averages for a month From 1895–2019.

2.2.6 Analyzing current restaurants established

Analyzing current restaurants established in each neighborhood and the food choices people are making, will help the business community to get more clarity on the market and its condition.

Source: Wikimedia Foundation, Inc. List of districts and neighborhoods of Los Angeles**.**

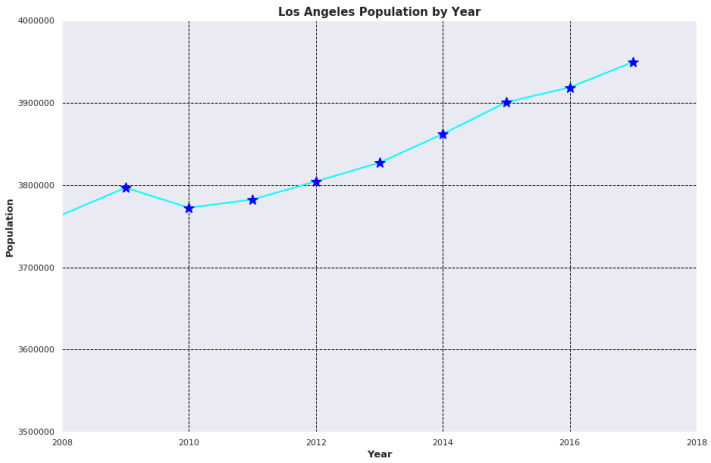
3 Methodology and Results

This project is executed using Python as a programming language, and Pandas libraries are used for data structures and data analysis. Data from US Census Estimate is downloaded and transformed into a Pandas data-frame for all the comparisons made in this report. Matplotlib Python 2D plotting libraries are used for plotting.

3.1 Population of the city

Population of the city data contains Year of the census, Population of the city in that year, the growth in numbers. As it can be seen from the data table and the graph that the city population has steadily increasing over the years. Even today City of Los Angeles is the most populous city in California; the second most populous city in the United States, after New York City; and the third-most populous city in North America, after Mexico City and New York City.[1]

Year	Population	Growth
2009	3796840	-
2010	3772486	-24354
2011	3782544	10058
2012	3804503	21959
2013	3827261	22758
2014	3862210	34949
2015	3900794	38584
2016	3918872	18078
2017	3949776	30904

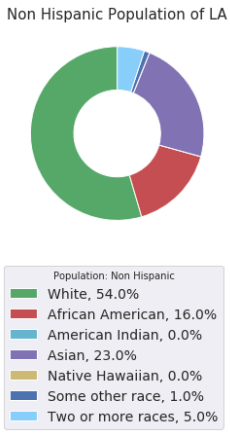


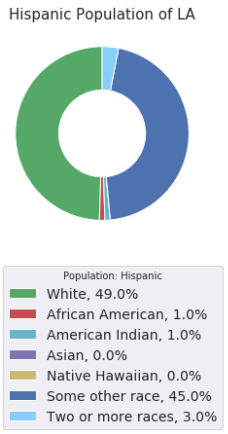
3.2 Population by Race

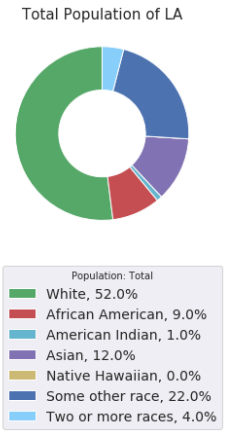
Using the data from US Census estimate 2018 Year Survey to show the comparison of the population by race and it is further segregated as Hispanic or Latino & Non-Hispanic origin.

Population of the city by race data contains total number of people living in each category and the percentage of that category. Total population by race show two major group of people living in the city and the choice of the restaurant type can be based on two groups. But this is just one of the criteria to consider, further analysis are done to understand the city better.

Race	Non Hispanic or Latino	Percentage Non Hispanic	Hispanic or Latino	Percentage Hispanic	Total	Percentage Total
White	1107758	54	965166	49	2072924	52.0
Black or African American	333350	16	12542	1	345892	9.0
American Indian and Alaska Native	6597	0	23969	1	30566	1.0
Asian	475572	23	6391	0	481963	12.0
Native Hawaiian and Other Pacific Islander	5054	0	1963	0	7017	0.0
Some other race	13626	1	883170	45	896796	22.0
Two or more races	96656	5	58655	3	155311	4.0



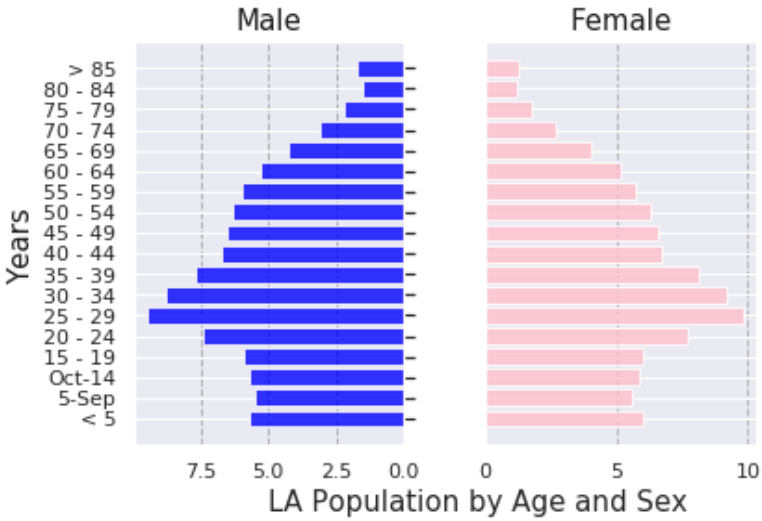




3.3 Compare Los Angeles Population by Age & Sex

From the comparison of Population by Age and Sex it can be noticed that majority of the population both male and female are in the age group of 25 to 39, and if we look into wider results to include next age group we can notice most of the people staying in the city can be grouped between 20 to 49 of age. Which means the bulk of the people living in the city is either young or middle aged.

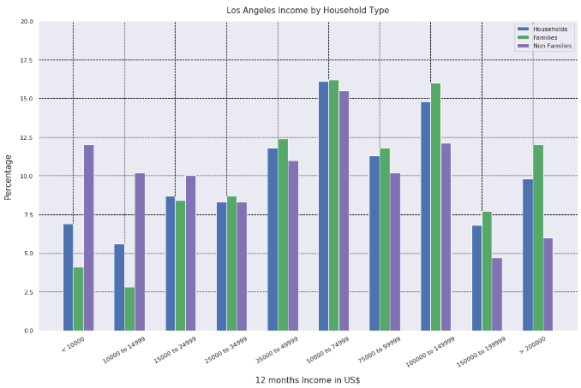
Age	Male	Percent Male	Female	Percent Female	Total	Percent
Under 5 years	228554	5.7	120313	6.0	108241	5.4
5 to 9 years	218804	5.5	111842	5.6	106962	5.4
10 to 14 years	226674	5.7	117800	5.9	108874	5.5
15 to 19 years	235640	5.9	120175	6.0	115465	5.8
20 to 24 years	295953	7.4	153506	7.7	142447	7.1
25 to 29 years	379520	9.5	195535	9.8	183985	9.2
30 to 34 years	353018	8.8	184234	9.2	168784	8.5
35 to 39 years	306004	7.7	160730	8.1	145274	7.3
40 to 44 years	268756	6.7	134630	6.7	134126	6.7
45 to 49 years	260922	6.5	132236	6.6	128686	6.4
50 to 54 years	252251	6.3	125990	6.3	126261	6.3
55 to 59 years	238857	6.0	113489	5.7	125368	6.3
60 to 64 years	213089	5.3	104413	5.2	108676	5.4
65 to 69 years	170179	4.3	79924	4.0	90255	4.5
70 to 74 years	125431	3.1	54309	2.7	71122	3.6
75 to 79 years	85828	2.2	36478	1.8	49350	2.5
80 to 84 years	61234	1.5	24137	1.2	37097	1.9
85 years and over	69755	1.7	25545	1.3	44210	2.2



3.4 Income of the people

US Census estimate 2018 shows that most of the people across all category Households, families, and non-families are in the range of US$ 50,000 to US$ 75,000 and US$ 100,000 to US$ 150,000. Based on the income of the people a casual dining restaurant, Family style restaurants or a Fine dining restaurant can be a good choice to cater.

Income in US$	Percentage Households	Percentage Families	Percentage Non Families
< 10,000	6.9	4.1	12.0
10,000 to 14,999	5.6	2.8	10.2
15,000 to 24,999	8.7	8.4	10.0
25,000 to 34,999	8.3	8.7	8.3
35,000 to 49,999	11.8	12.4	11.0
50,000 to 74,999	16.1	16.2	15.5
75,000 to 99,999	11.3	11.8	10.2
100,000 to 149,999	14.8	16.0	12.1
150,000 to 199,999	6.8	7.7	4.7
> 200,000	9.8	12.0	6.0

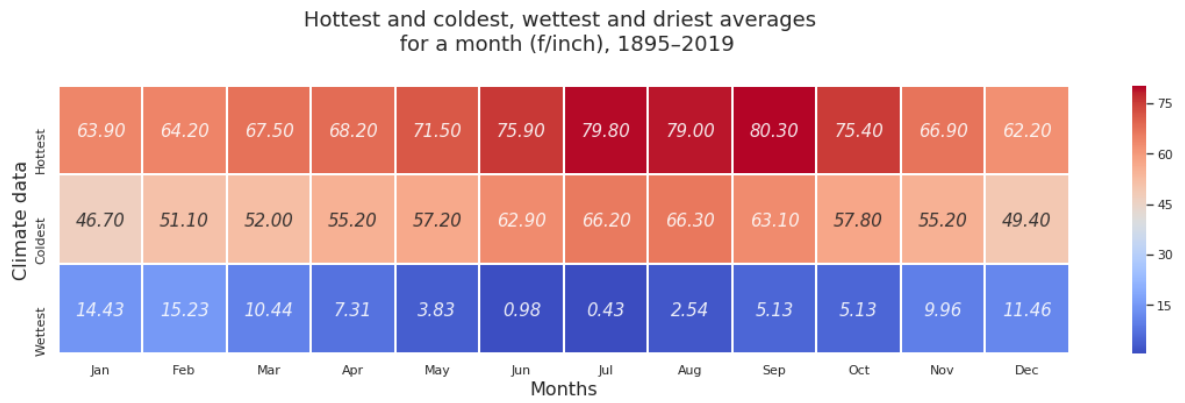


3.5 Weather conditions

BeautifulSoup web scraping (web harvesting) technique is used to pull the data out of wiki table in HTML format and structed into Pandas data-frame.

City weather pattern will help investors to understand the type of investment they may have to make to build infrastructure. Harvested table below show the weather pattern of Hottest and coldest, wettest and driest averages for a month from 1895–2019. From the weather pattern heatmap we can see that the city has very less rainfall throughout the year and it usually dry and cold. Heatmap is generated using Matplotlib based Python data visualization library Seaborn.

	Climate data	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
0	Hottest	63.90	64.20	67.50	68.20	71.50	75.90	79.80	79.00	80.30	75.40	66.90	62.20
1	Coldest	46.70	51.10	52.00	55.20	57.20	62.90	66.20	66.30	63.10	57.80	55.20	49.40
2	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



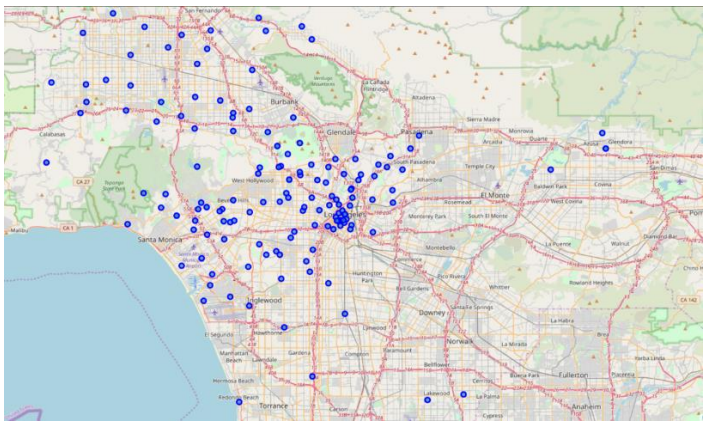
3.6 Analyzing current restaurants established

List of neighborhoods of Los Angeles is harvested from wiki page using the BeautifulSoup web scraping technique in HTML format and structed into Pandas data-frame. Geopy libraries are used to get the latitude and longitude values for each neighborhood.

3.6.1 General search

Using the coordinates of addresses and manipulating the data in Python, data is structured. To analyzing the current restaurants established in each neighborhood within the radius of 500m. Sample of the structured data is shown below. Map of L.A. is plotted on a Leaflet map via folium with neighborhoods superimposed on top.

Neighborhood	LA_Address	Latitude	Longitude
Angelino Heights	Angelino Heights, Los Angeles, California	34.070289	-118.254796
Arleta	Arleta, Los Angeles, California	34.241327	-118.432205
Arlington Heights	Arlington Heights, Los Angeles, California	34.128557	-118.152999
Arts District	Arts District, Los Angeles, California	34.041239	-118.234450
Atwater Village	Atwater Village, Los Angeles, California	34.116398	-118.256464
...	...	...	...
Wilshire Center	Wilshire Center, Los Angeles, California	34.061515	-118.432771
Wilshire Park	Wilshire Park, Los Angeles, California	34.061515	-118.432771
Windsor Square	Windsor Square, Los Angeles, California	34.072593	-118.320810
Winnetka	Winnetka, Los Angeles, California	34.205883	-118.570934
Woodland Hills	Woodland Hills, Los Angeles, California	34.168436	-118.605838



Utilizing the Foursquare API to explore the neighborhoods to get the most common venues of L.A.
General search on each neighborhood of L.A. resulted 3716 venues among those 323 unique categories returned by Foursquare. Sample of the data-frame is shown below.

LA_Address	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Angelino Heights, Los Angeles, California	34.070289	-118.254796	Halliwell Manor	34.069329	-118.254165	Performing Arts Venue
Angelino Heights, Los Angeles, California	34.070289	-118.254796	Guisados	34.070262	-118.250437	Taco Place
Angelino Heights, Los Angeles, California	34.070289	-118.254796	Eightfold Coffee	34.071245	-118.250698	Coffee Shop
Angelino Heights, Los Angeles, California	34.070289	-118.254796	Tsubaki	34.072938	-118.251298	Japanese Restaurant
Angelino Heights, Los Angeles, California	34.070289	-118.254796	Subliminal Projects	34.072290	-118.250737	Art Gallery

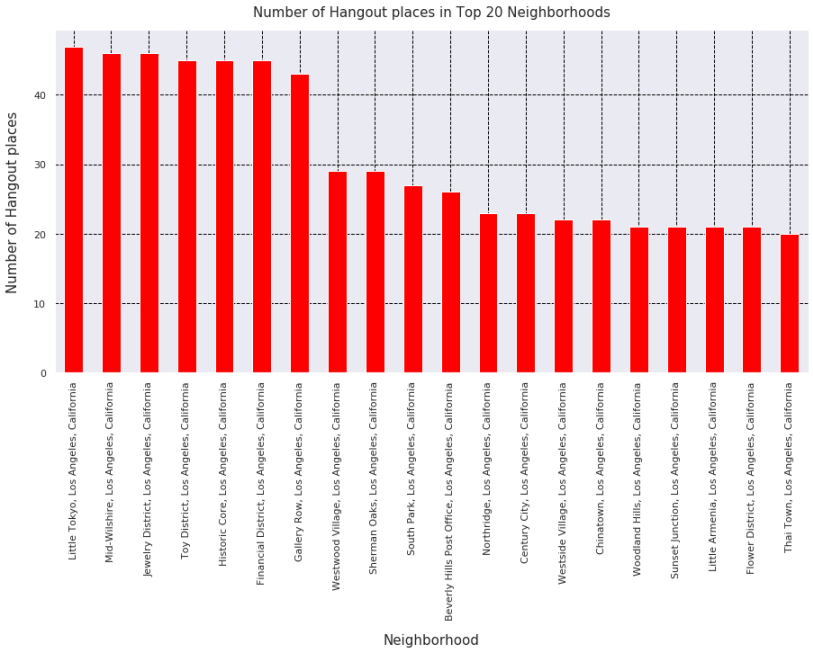
3.6.2 Hangout places search

Searching 'Venue Category' column and looking for specific categories such as Bar, Restaurant, Coffee shop, Pizza, and Food resulted 74 unique categories. Sample of structured data is shown below.

LA_Address	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Angelino Heights, Los Angeles, California	34.070289	-118.254796	Tsubaki	34.072938	-118.251298	Japanese Restaurant
Angelino Heights, Los Angeles, California	34.070289	-118.254796	Ototo	34.074399	-118.254016	Sake Bar
Angelino Heights, Los Angeles, California	34.070289	-118.254796	Bar Henry	34.069062	-118.250465	Cocktail Bar
Angelino Heights, Los Angeles, California	34.070289	-118.254796	Domino's Pizza	34.073528	-118.251187	Pizza Place
Angelino Heights, Los Angeles, California	34.070289	-118.254796	Taqueria Los Amigos	34.073444	-118.251791	Mexican Restaurant
...	...	...	...	...	...	...
Woodland Hills, Los Angeles, California	34.168436	-118.605838	Mazar Mediterranenan Restaurant	34.168358	-118.604793	Mediterranean Restaurant
Woodland Hills, Los Angeles, California	34.168436	-118.605838	Anantra	34.169638	-118.602212	Thai Restaurant
Woodland Hills, Los Angeles, California	34.168436	-118.605838	Lotus Inn China Bistro	34.168605	-118.602055	Chinese Restaurant
Woodland Hills, Los Angeles, California	34.168436	-118.605838	Copper Mine	34.168596	-118.602043	Bar
Woodland Hills, Los Angeles, California	34.168436	-118.605838	Darna Meditaranean Cusine	34.171741	-118.605770	Mediterranean Restaurant

74 unique categories are further grouped to show Top 20 hangout places across all neighborhoods. The current analysis show that Little Tokyo, Mid-Wilshire, Jewelry District, Toy District, Historic Core, Financial District, and Gallery Row neighborhoods are the most favorite hangout places in Los Angeles.

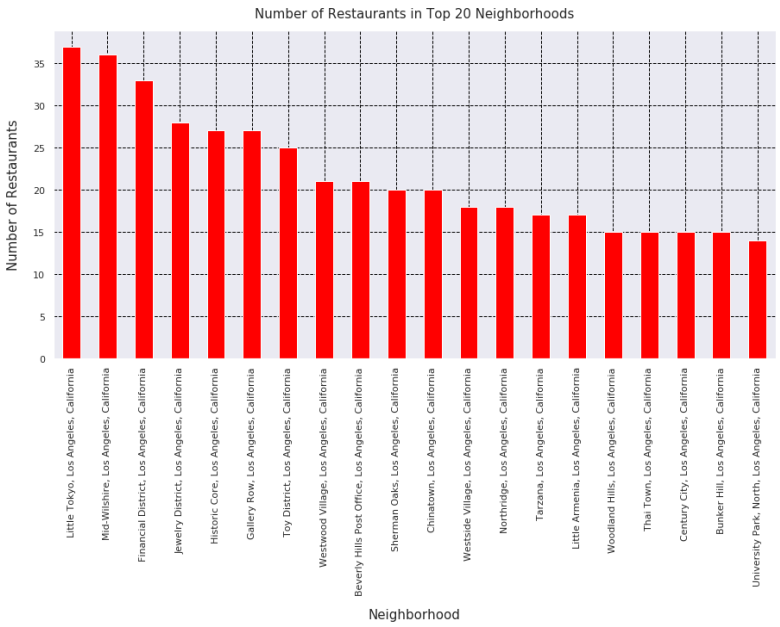
LA_Address	Venue Category
Mid-Wilshire, Los Angeles, California	47
Little Tokyo, Los Angeles, California	47
Historic Core, Los Angeles, California	46
Toy District, Los Angeles, California	45
Jewelry District, Los Angeles, California	44
Gallery Row, Los Angeles, California	44
Financial District, Los Angeles, California	43
Westwood Village, Los Angeles, California	29
Sherman Oaks, Los Angeles, California	28
Century City, Los Angeles, California	26
Beverly Hills Post Office, Los Angeles, California	26
South Park, Los Angeles, California	25
Northridge, Los Angeles, California	23
Westside Village, Los Angeles, California	22
Chinatown, Los Angeles, California	22
Woodland Hills, Los Angeles, California	21
Thai Town, Los Angeles, California	21
Sunset Junction, Los Angeles, California	21
Little Armenia, Los Angeles, California	21
Flower District, Los Angeles, California	21



3.6.3 Restaurants search

Further narrow search restricting to restaurants around each neighborhood yielded 53 unique categories. List of top 20 neighborhood is shown in the below table and graph. Little Tokyo, Mid-Wilshire, Financial District, Jewelry District, Historic Core, Gallery Row, and Toy District where most of the restaurants established in the city.

LA_Address	Venue Category
Little Tokyo, Los Angeles, California	37
Mid-Wilshire, Los Angeles, California	36
Financial District, Los Angeles, California	32
Historic Core, Los Angeles, California	27
Gallery Row, Los Angeles, California	27
Jewelry District, Los Angeles, California	26
Toy District, Los Angeles, California	25
Westwood Village, Los Angeles, California	20
Sherman Oaks, Los Angeles, California	20
Chinatown, Los Angeles, California	20
Beverly Hills Post Office, Los Angeles, California	20
Westside Village, Los Angeles, California	18
Northridge, Los Angeles, California	18
Century City, Los Angeles, California	18
Little Armenia, Los Angeles, California	17
Thai Town, Los Angeles, California	16
Tarzana, Los Angeles, California	16
Woodland Hills, Los Angeles, California	15
Bunker Hill, Los Angeles, California	15
Sunset Junction, Los Angeles, California	14

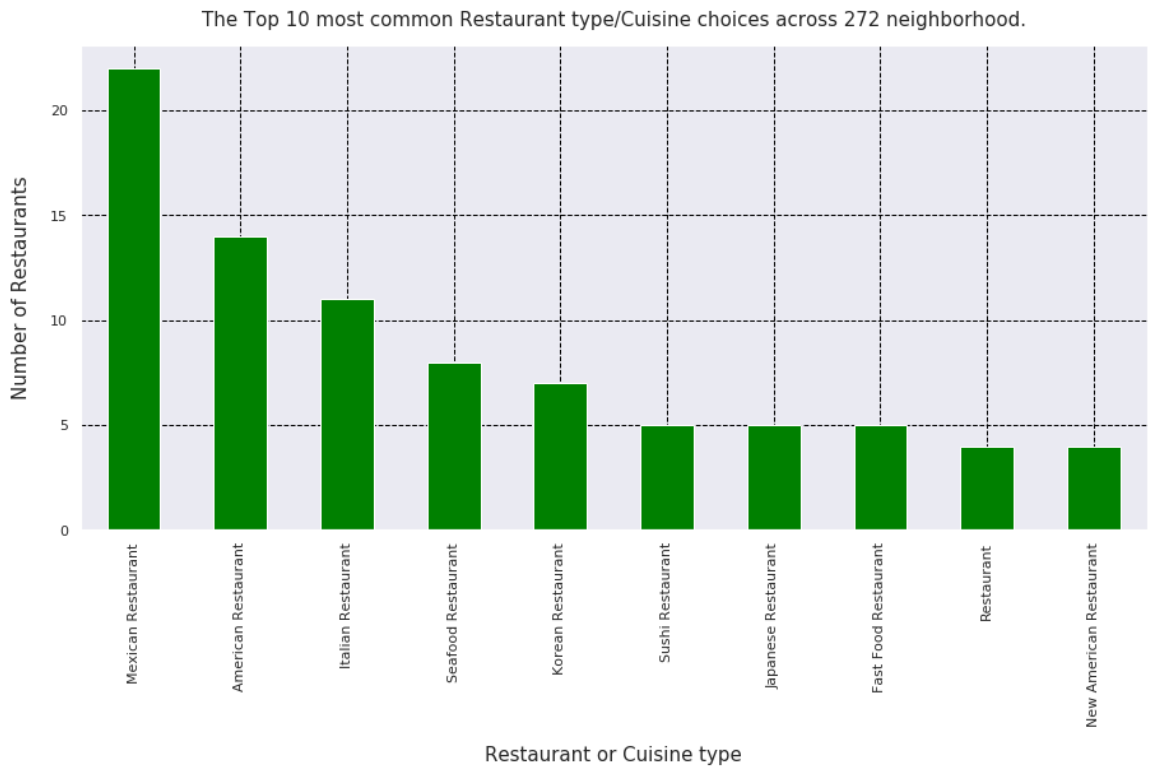


3.6.4 Mean of the frequency of occurrence (cuisine choices)

Below data table show the sample data for each neighborhood along with the top 3 most common restaurant types. This is achieved by grouping rows by Venue category and by taking the mean of the frequency of occurrence of restaurants category.

LA_Address	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
Angelino Heights, Los Angeles, California	Mexican Restaurant	Japanese Restaurant	Sushi Restaurant
Arts District, Los Angeles, California	Italian Restaurant	Asian Restaurant	Mediterranean Restaurant
Atwater Village, Los Angeles, California	Restaurant	Cuban Restaurant	Latin American Restaurant
aldwin Hills/Crenshaw, Los Angeles, California	Fast Food Restaurant	Mexican Restaurant	Southern / Soul Food Restaurant
Beverly Grove, Los Angeles, California	Japanese Restaurant	Yoshoku Restaurant	Italian Restaurant
...	...	...	...
Wilshire Center, Los Angeles, California	Restaurant	Yoshoku Restaurant	Falafel Restaurant
Wilshire Park, Los Angeles, California	Restaurant	Yoshoku Restaurant	Falafel Restaurant
Windsor Square, Los Angeles, California	Italian Restaurant	Sushi Restaurant	Mediterranean Restaurant
Winnetka, Los Angeles, California	Latin American Restaurant	Mexican Restaurant	Filipino Restaurant
Woodland Hills, Los Angeles, California	Mediterranean Restaurant	Sushi Restaurant	Indian Restaurant

First most common restaurant venue plot shows the Top 10 cuisine choices people are making across 272 neighborhoods. From the plot we can see that Mexican and American are most desired type of food, followed by Italian.



4 Observations and recommendations

Observations

Los Angeles is known for its Mediterranean climate the weather pattern of Hottest and coldest, wettest and driest averages for a month from 1895–2019 confirms the climate conditions, from the analysis of the population data by race we can see the ethnic diversity of the city. Bulk of the people living in the city is either young or middle aged with the Households, families, and non-family's income range of US$ 50,000 to US$ 150,000.

Comparing both narrow search and wider search results show that the concentration of restaurants and hangout places around similar neighborhood.

Recommendations:

In this project I have considered all the neighborhoods of Los Angeles across 15 districts. As the scope is very wide this project can be further segmented to focus on one district and looking for a best place to open a restaurant in that district.

We can include trending restaurants in each neighborhood to further understand the trends. Segmentation and clustering the neighborhoods can be done using machine learning models.

5 Conclusion

The city, which covers about 469 square miles (1,210 km2) and with over 4 million in population it would be beneficial to open a restaurant to cater to the Hispanic and Non-Hispanic population.

Based on the income of the people by household type and majority of the population both male and female are between the age group of 25 to 39 it would be a good choice either open a Family style restaurant or a Fine dining restaurant with a good infrastructure to support Mediterranean climate.

Mexican, American, Italian, Japanese cuisine are currently peoples favorite, investor can either opt to go with the current trend or go pick a new variety of cuisines such as Greek, Mediterranean, French or Spanish Cuisine.

With 53 unique categories of restaurants spread across 7 neighborhoods of L.A. Little Tokyo, Mid-Wilshire, Financial District, Jewelry District, Historic Core, Gallery Row, and Toy District. Investors can either chose to establish in these popular neighborhoods or opt for second most popular neighborhoods such as Westwood Village, Beverly Hills, Sherman Oaks, Chinatown, Westside Village, and Northridge.

6 Reference:

1. About Los Angeles, Wikimedia Foundation, Inc. (Retrieved, Nov 2019)
2. Los Angeles Neighborhoods Mapping L.A. Los Angeles Times. (Retrieved, Nov 2019).