<u>Opening up a Japanese Restaurant in the City of Central Los Angeles</u>

Introduction

Shinobi Holdings is Planning to open their newest branch of Japanese restaurant Called Toshi Delights. This will be their 10th restaurant and they have decided to setup the restaurant in the Central Los Angeles City. They have been in the business for more than 20 years and has focused solely on contemporary Japanese cuisine. All of the restaurants they have setup up to now was in the east coast and they were highly successful with the business there. Since this is their first time setting up a restaurant in west coast, there are quite skeptic of the growth of business and wanted to do a feasibility study on how easy it is to do business in LA.

1.1 Business Problem

In order to find out the best place or the neighborhood to setup their restaurant, the director of the company Mr. Tatsumi has entrusted me in finding the most suitable location. Further delving into the history of the restaurant Toshi Delights, I have uncovered the following.

- The restaurant is more popular with the younger crowd, mostly university and high school students.
- Mr. Tatsumi is not in favor with setting up close to their competitors. He is of the view that it would be detrimental to the growth of his business.
- Mr. Tatsumi is of the view that they need to give more attention to/target office crowd.
- Since the Central Los Angeles is one the expensive places to rent out, Mr. Tatsumi is not willing spend more money on rents unless it's absolutely necessary.

Based on the description provided by Mr. Tatsumi the most ideal place they search for will need to be easily accessible by both younger crowd and office professionals. Also, they are quite worried about the already saturated market for Asian restaurants and want to avoid setting up near such similar restaurants especially other Japanese restaurants.

1.2 Target Audience

I believe the report can be used by any person who is venturing into setting up a restaurant business more specifically Asian restaurant in the city of Los Angeles. Also, anyone wants to overcome the saturated restaurant business in LA by locating in the neighborhoods with most prospective clients. This is also beneficial to people who wants to enhance their data science skills.

Data

2.1 Data Required to Resolve the Problem.

There are approximately 17 neighborhoods in central los angels with 4 million residents residing with diverse ethnic background. To find a solution for the problem, we require data sets of the following nature.

1. Data of Distribution of Schools and Universities to locate the Younger Crowds.
2. Data of distribution of Japanese Restaurants in Central LA to locate competitors.
3. Data of Most Popular Places People Visit by Neighborhood to get an idea of the nature of the neighborhood.
4. The Renting Cost for Business based on Neighborhoods in central LA.

All the above data gathered from several palaces. The list of neighborhoods in LA is retrieved from Wikipedia Page using BeautifulSoup to extract it. The Renting cost is also retrieved from a rentcafe website using beautifulsoup. The locations and distributions of Schools, Universities and Other licensed Japanese Restaurants through the use of Foursquare service. The location/coordinates of each place is obtained through OpenStreetMap.

List of LA neighborhoods:
https://en.wikipedia.org/wiki/List_of_districts_and_neighborhoods_of_Los_Angeles

Location of Places (Retrieving coordinates):
http://nominatim.openstreetmap.org/search.php?

Rent Prices for All Neighborhoods
https://www.rentcafe.com/average-rent-market-trends/us/ca/los-angeles/

2.2 How Data Will be used to resolve the problem

- Use OpenStreetMap to retrieve coordinates of the 17 neighborhoods and store these coordinates in a table. Later on the coordinates will be passed from this table onto the folium api to draw a map of all the locations of the neighborhoods.
- Use Foursquare service and geopy data to map top 10 venues for all Central LA neighborhoods and clustered in groups that are related to each other. The explorer api call of foursquare will be used to get all the nearby top 80 places. And then run k-means algorithm to analyze 10 most common venues in each cluster.
- Use foursquare service and geopy data to map schools, universities, office spaces that are essential for analysis of the customer base for the restaurant. The search api call of foursquare will be used get the locations. All of the places will be visualized using folium api on the map to get a better understanding of customer distribution.

- Developing a score system that takes into account the number of schools, universities, hotels and restaurants that are there in each neighborhood and giving a weightage to each of those parameter (positive/negative) so as to calculate a final score of suitability of the Japanese restaurant on each neighborhood.
- Retrieving rent cost from rentcafe.com and storing on a pandas dataframe table. Later the rent values will be compared with final score system already developed to get a better understanding of Central LA neighborhoods and also to uncover any relationship- with close location to those places and the rent cost.

- Maps will be developed to analyses the distribution of places such as schools, offices and restaurants for each neighborhood. The folium api will be used.
- Bar graphs will be used to analyze the distribution of hotels, restaurants among each 17 neighborhoods.

3 Methodology

3.1 Data Mining and Cleaning

Data that was in the Wikipedia page to get list of Neighborhood had some invalid and missing data. So mining was done using beautifulsoup. Items were mined using href link and then all neighbourhoods were copied to pandas dataframe, from there the coordinates were added to each respective neighborhood.

For example some neighborhoods were not able to find coordinates of were renamed or removed, but fortunately it was few and far between.

```
#Get list of neighbourhoods of Los Angeles from Wikipedia
source = requests.get('https://en.wikipedia.org/wiki/List_of_districts_and_neighborhoods_of_Los_Angeles').text

soup = BeautifulSoup(source, 'lxml')
links = soup.find_all("a", href=re.compile("wiki")) #extract <a> tags from html

Neighbourhoods_LA = []
for item in links:
    Neighbourhoods_LA.append(item.get('title'))
```

```
]: df3=df2.copy() #copy to temp dataframe
df3 = df3[~df2['Latitude'].isin(['N'])] #remove neighbourhoods with missing/invalid coordinates
LACity_df=df3.copy()
LACity_df.head(15) #Coordinates for All neighbourhoods in LA
```

|	Neighbourhood	Latitude	Longitude
0	Angelino Heights, Los Angeles	34.0703	-118.255
1	Arleta, Los Angeles	34.2413	-118.432
2	Arlington Heights, Los Angeles	34.1283	-118.156
3	Arts District, Los Angeles	34.0412	-118.234
4	Atwater Village, Los Angeles	34.1164	-118.256
5	Baldwin Hills, Los Angeles	34.0076	-118.351

3.2 Exploratory Data Analysis

Once latitude and longitude were confirmed, the data was used further examine in LA Central region and see how neighborhoods are located. For this geopy library was used to take coordinates. Map of the Entire LA City was generated with markers on the neighborhoods that would be analyzed.

Although we got the coordinates for all neighborhoods in LA city, it was decided to further reduce the concentration are to Central LA city as the data was a large data set. Below is the 17 neighbouhoods of Central LA.
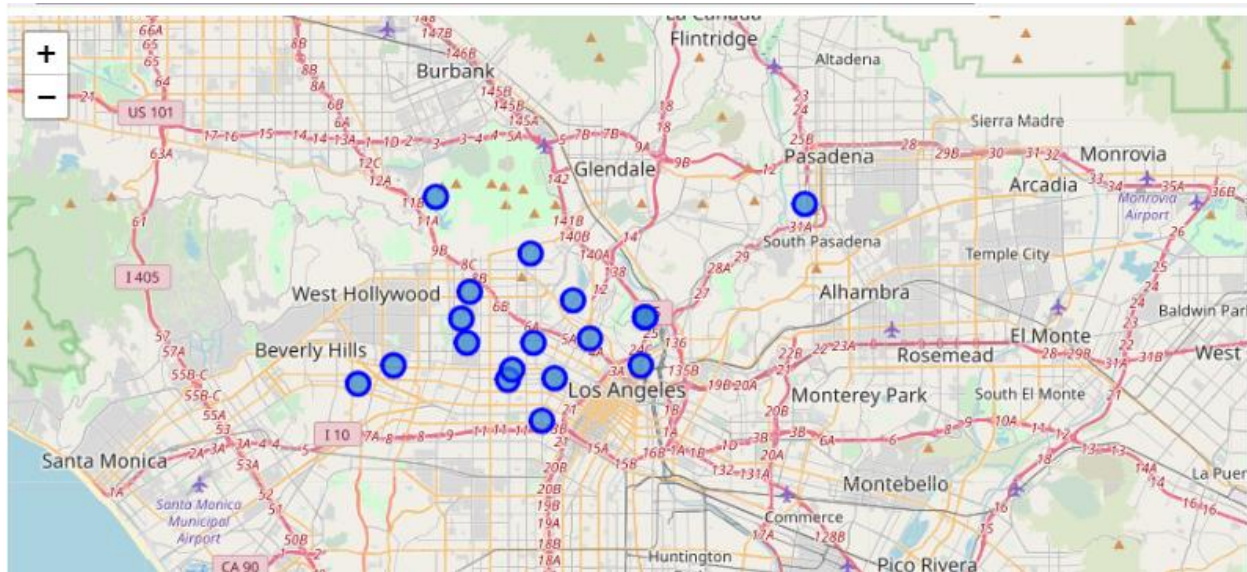

3.3 Relationship of Land Value of Neighborhood with Food Eateries
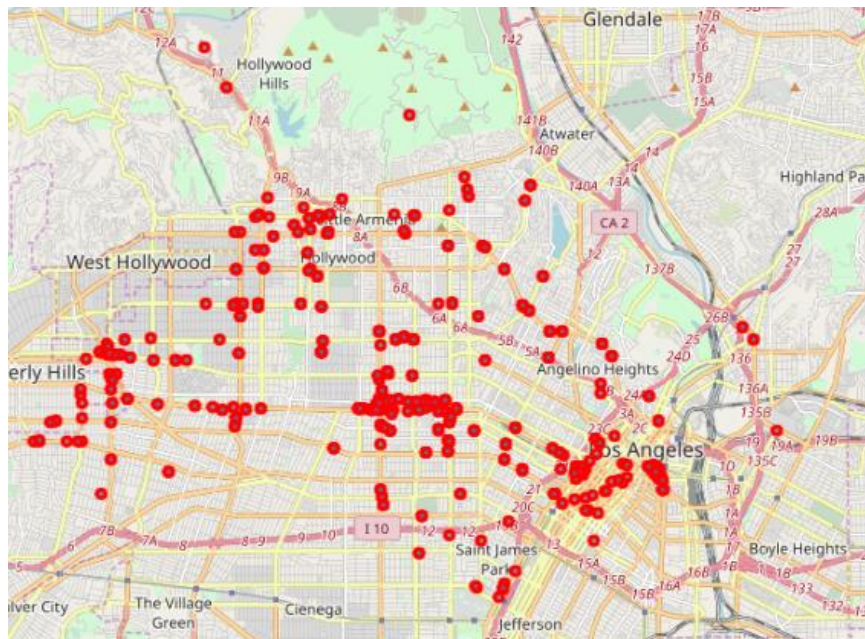
Relationship of Land Value of Neighborhood with Food Eateries including restaurants We don't see land value has an impact with number of food eateries, the central LA remains at higher land value regardless of the Restaurents.
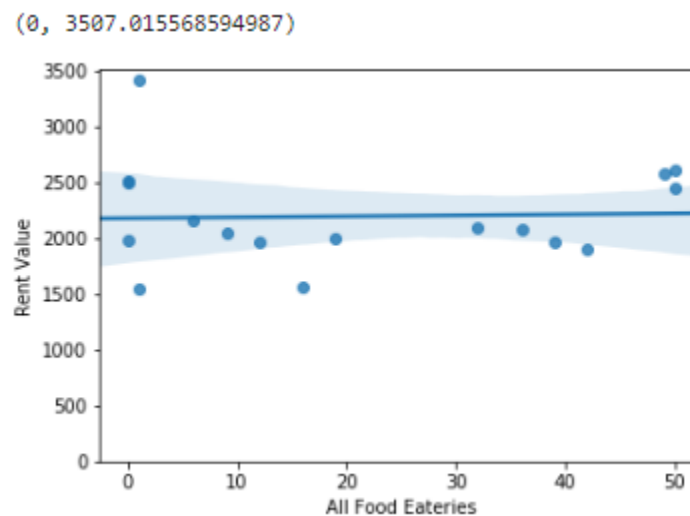
The food eateries and restaurants are evenly distributed as shown below this may be the reason there's no good place of significance concerned with land value.
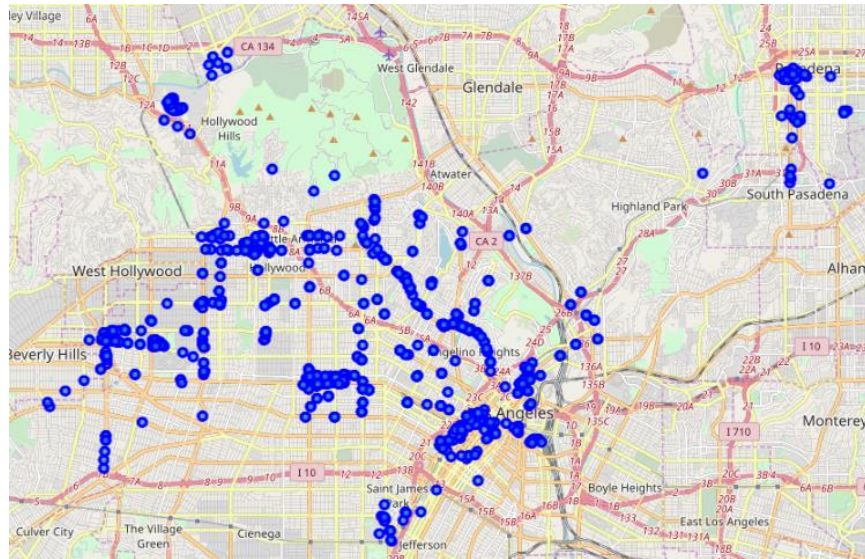
But if drilled down above and only see Japanese restaurants only (marked red), still we can see it has an even distribution throughout the central LA.
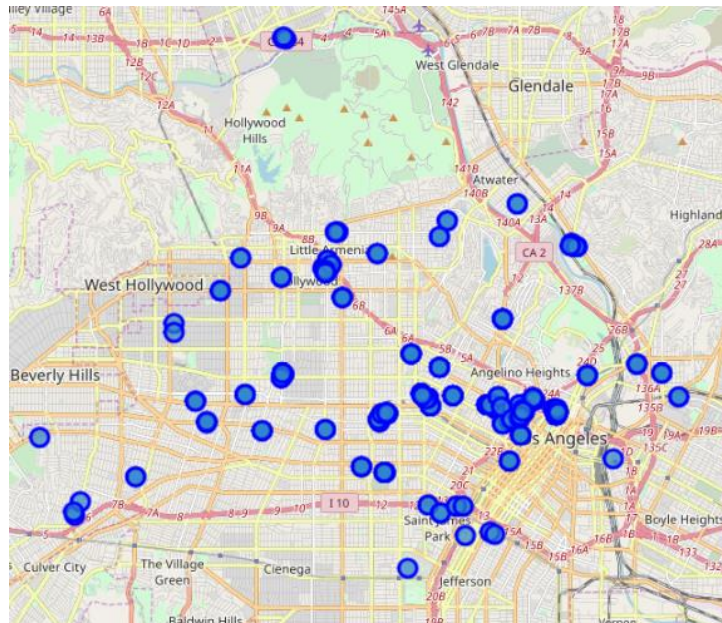

3.4 Relationship of School Distribution with Land Values
Relationship of Land Value with number of schools. There's a slight decrease in Rent Value when number of high school increase in neighborhood. Testing with linear regression proves this.

(0, 3534.4811190010078)
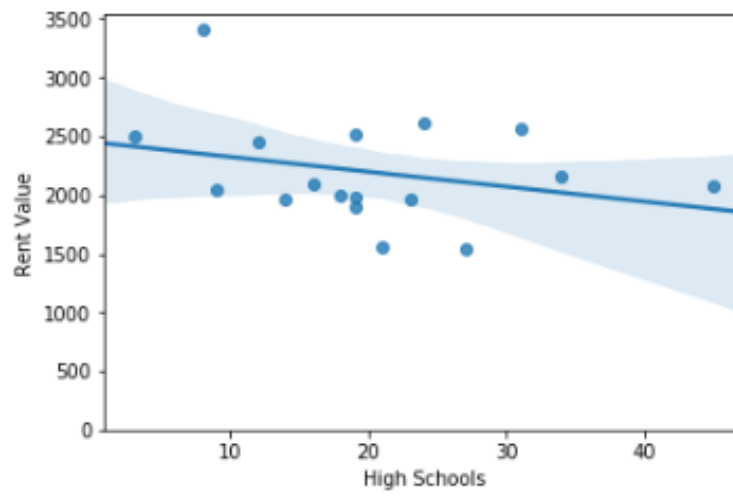
Below shows the school distribution in Central LA.

3.4 Descriptive Analytics

Based on the score system, below 10 were determined as top place for restaurant to setup.

	Neighbourhood	Score	Rent Value
0	Echo Park, Los Angeles	127.0	$2,156
1	Arlington Heights, Los Angeles	124.0	$1,543
2	Windsor Square, Los Angeles	122.0	$1,967
3	Carthay, Los Angeles	113.0	$3,410
4	Hollywood Hills West, Los Angeles	103.0	$2,498
5	Harvard Heights, Los Angeles	98.0	$1,560
6	East Hollywood, Los Angeles	95.0	$1,997
7	Elysian Park, Los Angeles	94.0	$2,509
8	Elysian Valley, Los Angeles	94.0	$1,977
9	Silver Lake, Los Angeles	84.0	$2,039
10	Koreatown, Los Angeles	77.0	$1,903

Above was generated based on a score system that used the number of presence of schools, hotel, restaurants, offices, universities in the neighborhood. **Echo Park, Arlington Heights and Windsor Square** were the top 3 as there were very few eateries or restaurant and large number of office premises and school and universities.

Above scorecard was made using the below table calculations.

|	Neighbourhood	Latitude	Longitude	Office	High Schools	Universities	Japanese Restaurants	Hotels	All Food Eateries	Rent Value
0	Arlington Heights, Los Angeles	34.1283	-118.156	49.0	22.0	4.0	0.0	1.0	1.0	1543.0
1	Carthay, Los Angeles	34.0561	-118.373	50.0	4.0	5.0	0.0	3.0	1.0	3410.0
2	Chinatown, Los Angeles	34.0635	-118.236	50.0	24.0	3.0	0.0	6.0	49.0	2572.0
3	East Hollywood, Los Angeles	34.0932	-118.32	50.0	14.0	3.0	0.0	8.0	19.0	1997.0
4	Echo Park, Los Angeles	34.074	-118.261	50.0	26.0	1.0	0.0	6.0	6.0	2156.0
5	Elysian Park, Los Angeles	34.0829	-118.234	43.0	8.0	0.0	0.0	0.0	0.0	2509.0
6	Elysian Valley, Los Angeles	34.0829	-118.234	43.0	8.0	0.0	0.0	0.0	0.0	1977.0
7	Hancock Park, Los Angeles	34.0637	-118.356	50.0	5.0	4.0	1.0	8.0	50.0	2451.0
8	Harvard Heights, Los Angeles	34.0724	-118.288	50.0	13.0	9.0	0.0	4.0	16.0	1560.0
9	Hollywood Hills West, Los Angeles	34.1312	-118.336	50.0	0.0	3.0	0.0	0.0	0.0	2498.0
10	Koreatown, Los Angeles	34.058	-118.301	50.0	15.0	9.0	1.0	20.0	42.0	1903.0
11	Larchmont, Los Angeles	34.0821	-118.324	50.0	14.0	3.0	1.0	5.0	32.0	2094.0
12	Los Feliz, Los Angeles	34.1082	-118.29	41.0	5.0	1.0	3.0	4.0	39.0	1969.0
13	Mid-Wilshire, Los Angeles	34.0619	-118.299	50.0	16.0	9.0	20.0	24.0	50.0	2611.0
14	Silver Lake, Los Angeles	34.0897	-118.269	45.0	4.0	1.0	1.0	7.0	9.0	2039.0
15	Westlake, Los Angeles	34.0581	-118.278	50.0	31.0	9.0	2.0	22.0	36.0	2071.0
16	Windsor Square, Los Angeles	34.0726	-118.321	50.0	7.0	4.0	3.0	2.0	12.0	1967.0

3.5 Inferential Analytics

In order to further forward out case, Next, we find categories of venue mostly visited by people by neighborhood. In order to normalize data, we perform one hot encoding as well and after that we perform k-mean machine algorithm.

When we list down top 5 categories of places visited by each neighborhood we get the below result

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Arlington Heights, Los Angeles	Gym	Coffee Shop	Sandwich Place	Garden	Gym / Fitness Center
1	Carthay, Los Angeles	Ethiopian Restaurant	Grocery Store	Mobile Phone Shop	Rental Car Location	Middle Eastern Restaurant
2	Chinatown, Los Angeles	Chinese Restaurant	Bakery	Art Gallery	Coffee Shop	Sandwich Place
3	East Hollywood, Los Angeles	Coffee Shop	Bar	Pizza Place	Food Truck	Vegetarian / Vegan Restaurant
4	Echo Park, Los Angeles	Café	Mexican Restaurant	Coffee Shop	Pizza Place	Bar
5	Elysian Park, Los Angeles	Park	Trail	Intersection	Scenic Lookout	Food Truck
6	Elysian Valley, Los Angeles	Park	Trail	Intersection	Scenic Lookout	Food Truck
7	Hancock Park, Los Angeles	Art Museum	Museum	Coffee Shop	Art Gallery	Gym
8	Harvard Heights, Los Angeles	Coffee Shop	Fast Food Restaurant	Grocery Store	Sandwich Place	Mexican Restaurant
9	Hollywood Hills West, Los Angeles	Trail	Park	Convenience Store	Marijuana Dispensary	Mountain
10	Koreatown, Los Angeles	Korean Restaurant	Coffee Shop	Café	Mexican Restaurant	Seafood Restaurant
11	Larchmont, Los Angeles	Coffee Shop	Italian Restaurant	Mexican Restaurant	Bakery	Restaurant
12	Los Feliz, Los Angeles	Coffee Shop	Italian Restaurant	Mediterranean Restaurant	Ice Cream Shop	Breakfast Spot
13	Mid-Wilshire, Los Angeles	Korean Restaurant	Japanese Restaurant	Bakery	Restaurant	Ice Cream Shop
14	Pico-Union, Los Angeles	Mexican Restaurant	Pizza Place	Fast Food Restaurant	Latin American Restaurant	Pharmacy
15	Silver Lake, Los Angeles	Coffee Shop	Ice Cream Shop	American Restaurant	Breakfast Spot	Pizza Place
16	Westlake, Los Angeles	Fast Food Restaurant	Latin American Restaurant	Sandwich Place	Korean Restaurant	Coffee Shop
17	Windsor Square, Los Angeles	Coffee Shop	Spa	Juice Bar	Bakery	Italian Restaurant

As we can see Arlington Heights has no restaurants in Top 5 while Windsor Square and Echo Park only has one restaurant in Top 5 further er confirming they are best places for Setting up restaurants.

Then we perform kmean Clustering and cluster our neighbourhoods,
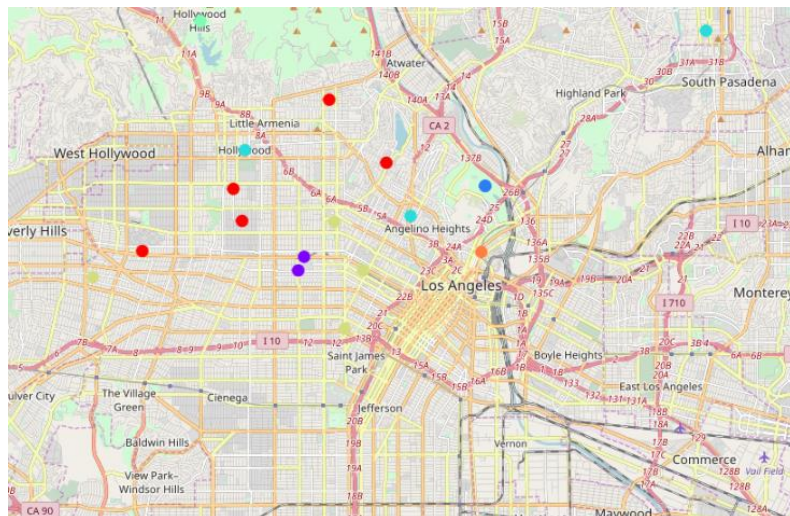
Based on the clustering that I have performed, we can say the neighborhoods that are located north and center or central LA is mostly saturated with food eateries. If we look below clusters Arlington Heights and Echo Park both regions belong to the same cluster. This cluster when looking for any discriminating

category that there are only 2 occurrences of restaurants out of 15 hence it further proves these two neighborhoods are good for setting up restaurant due to lack of them.

```
central_LA_neightbourhoods_Cluster.loc[central_LA_neightbourhoods_Cluster['Cluster Labels'] == 0, central_LA_neightbourhoods_Cluster.colu
```

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
7	Hancock Park, Los Angeles	Art Museum	Museum	Coffee Shop	Art Gallery	Gym
11	Larchmont, Los Angeles	Coffee Shop	Italian Restaurant	Mexican Restaurant	Bakery	Restaurant
12	Los Feliz, Los Angeles	Coffee Shop	Italian Restaurant	Mediterranean Restaurant	Ice Cream Shop	Breakfast Spot
15	Silver Lake, Los Angeles	Coffee Shop	Ice Cream Shop	American Restaurant	Breakfast Spot	Pizza Place
17	Windsor Square, Los Angeles	Coffee Shop	Spa	Juice Bar	Bakery	Italian Restaurant

```
central_LA_neightbourhoods_Cluster.loc[central_LA_neightbourhoods_Cluster['Cluster Labels'] == 1, central_LA_neightbourhoods_Cluster.colu
```

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
10	Koreatown, Los Angeles	Korean Restaurant	Coffee Shop	Café	Mexican Restaurant	Seafood Restaurant
13	Mid-Wilshire, Los Angeles	Korean Restaurant	Japanese Restaurant	Bakery	Restaurant	Ice Cream Shop

```
central_LA_neightbourhoods_Cluster.loc[central_LA_neightbourhoods_Cluster['Cluster Labels'] == 2, central_LA_neightbourhoods_Cluster.colu
```

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
5	Elysian Park, Los Angeles	Park	Trail	Intersection	Scenic Lookout	Food Truck
6	Elysian Valley, Los Angeles	Park	Trail	Intersection	Scenic Lookout	Food Truck

While Echo Park and Arlington heights has low number of food eateries and restaurants, Echo Park may be this reason due to majority of the regional area is covered from nature reserve.

```
central_LA_neightbourhoods_Cluster.loc[central_LA_neightbourhoods_Cluster['Cluster Labels'] == 3, central_LA_neightbourhoods_Cluster.columns[[
```

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Arlington Heights, Los Angeles	Gym	Coffee Shop	Sandwich Place	Garden	Gym / Fitness Center
3	East Hollywood, Los Angeles	Coffee Shop	Bar	Pizza Place	Food Truck	Vegetarian / Vegan Restaurant
4	Echo Park, Los Angeles	Café	Mexican Restaurant	Coffee Shop	Pizza Place	Bar

Discussion

What is evident from the data analyzed is that the Central Los Angels City Region is highly saturated with commercial buildings. From schools to restaurants to office premises is almost evenly distributed. Only slight change that was observed was in south of the city Arlington heights where there as very few to no restaurants. However, when it come the renting prices it was observed that all parts of the region were evenly priced. If there was one thing that could be added here would be the population's ethnicity distribution. If the ethnic distribution can be added to the scorecard system or the data analysis in general we can have a broader view on where a Japanese restaurant can be placed. However, since there was no open source data to read demography on ethnicity of central LA region, it was passed.

Conclusion

Mr. Tatsumi best bet will be to setup restaurant west of central LA. Near Arlington heights. Large number of office premises and the low number of restaurants is the primary reason for this. As far as the renting prices are concerned all regions are equally expensive and hence there is no clear differentiator in that regard.