Start a new restaurant in Madison, WI

Yaqi Wu
March, 2020

1. Introduction

1.1 Background

Madison, the capital city of Wisconsin, lies west of Milwaukee. It's known for the domed Wisconsin State Capitol, which sits on an isthmus between lakes Mendota and Monona. As of July 1, 2018, Madison's estimated population of 258,054 made it the second-largest city in Wisconsin by population, after Milwaukee, and the 81st-largest in the United States. Located on an isthmus and lands surrounding four lakes-Lake Mendota, Lake Monona, Lake Kegonsa and Lake Waubesa-the city is home to the University of Wisconsin–Madison, the Wisconsin State Capitol, Henry Vilas Zoo, lakes, and an extensive network of parks and bike trails. The presence of the University of Wisconsin–Madison (the largest employer in the state) as well as other educational institutions has a significant impact on the economy, culture, and demographics of Madison. Madison is a growing technology economy and the region is home to the headquarters of Epic Systems, American Family Insurance, Exact Sciences, Promega, American Girl, Sub-Zero, Lands' End, a regional office for Google, the University Research Park, as well as many biotech and health systems startups. So many education and science institutes and large companies in Madison, which means a huge demand for food. Then, there is a good business chance to invest in a restaurant here.

1.2 Problem

Before starting a restaurant, we need to define the restaurants and its content, like what kind of restaurants to start. Also, the location of the restaurant is very important. In this project we will explore the restaurants at Madison, WI. area, and using current restaurants information to figure out what kind of restaurants to start and which locations is the good choice for him.

1.3 Interest

As a restaurant investor who wants to start a restaurant in the Madison, WI area, he/she would be interested in what kind of restaurant he wants to start and where to locate his restaurant.

2. Data acquisition and cleaning

2.1 Data Source

Foursquare location data will be used in this project to analyze the style and location of the new restaurant. The surrounding restaurants data can be got from Foursquare API by using request function by passing url which is the path to access the data we need. The url will be like 'https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={},{}&v={}&query={}&radius={}&limit={}'.format(CLIENT_ID, CLIENT_SECRET, latitude, longitude, VERSION, search_query, radius, LIMIT)", in which (latitude, longitude) is the geographical coordinate of city Madison, which can be got through the "arcgisThe" function in geocoder package by passing 'Madison, WI'. And CLIENT_ID, CLIENT_SECRET are the username and passwords to access Foursquare location data. The key word 'restaurant' is passed as the search_query. I would like to set the version, radius, LIMIT as '20200309', '1000' and '100' respectively. To be able to do deep analysis, the rating score of the restaurants were also extracted from Foursquare location data and will be combined into restaurant data.

2.2 Data cleaning

Madison restaurants data will download from Foursquare as the description above. Then, we can get json data of the surrounding restaurants, which contains unreadable data and some codes. To make the data readable, the json data will be transferred into normal data by using the json_normalize() function, then basic information of restaurant will be extract from the dataframe, like names, id, geographical coordinate, category, address, Zip Code, rating, etc., to form a new clean data frame. I removed the restaurants which don't have the ranting score, since rating score is an important criteria to choose the type and location of a new restaurant.

2.3 Feature selection

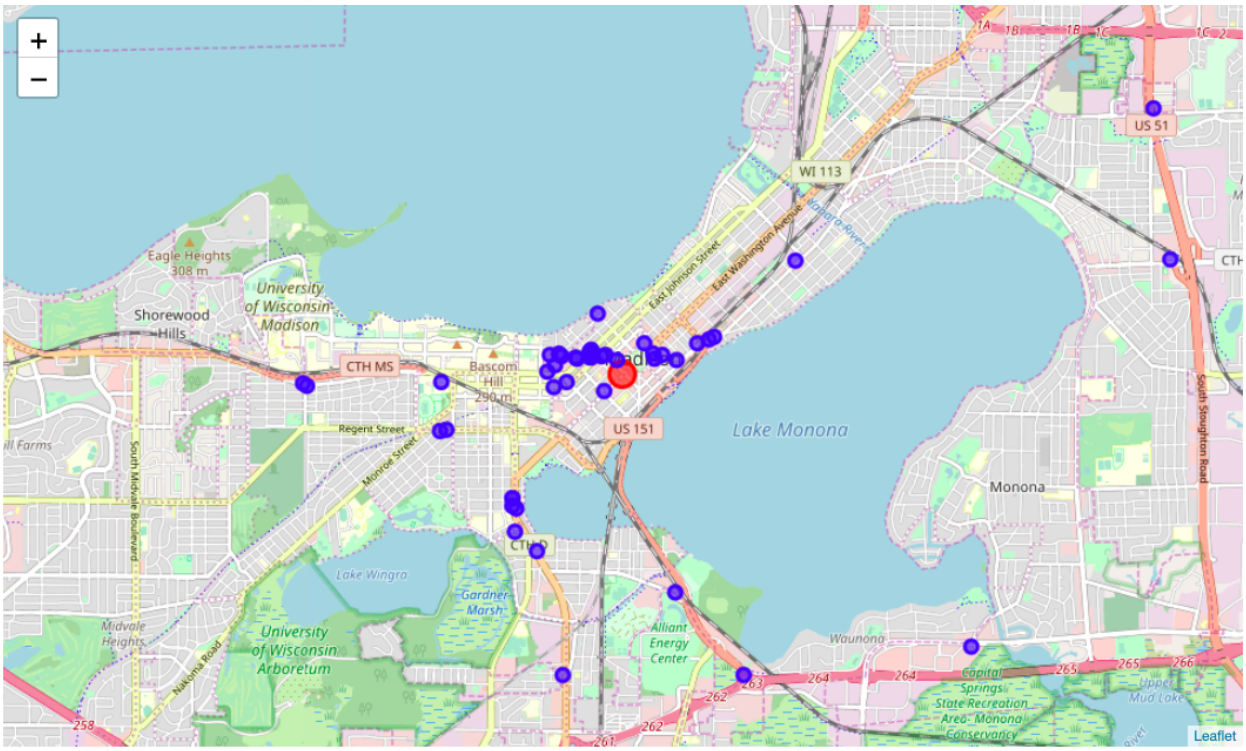
After cleaning the data set, the table of restaurants data contains 'name', 'categories', 'address', 'cc', 'city', 'country', 'crossStreet', 'distance', 'formattedAddress', 'labeledLatLngs', 'latitude', 'longitude', 'neighborhood', 'postalCode', 'state', 'id', 'rating'. But apparently there is some duplicate information of the restaurants, like city, state, country, etc.. To make the data analysis easier and clear, I will choose the restaurant name, category, latitude, longitude and rating as the columns (see the Table 1). The restaurants are all marked on the map, we can see how the restaurants are scattered aroud Madison city (Picture 1).

Table 1 Restaurants Data in the Madison area

Restaurant name	Category	Latitude	Longitude	rating
Nick's Restaurant & Bar	American Restaurant	43.074856	-89.389314	6.8
Restaurant Muramoto	Sushi Restaurant	43.074539	-89.379341	8.5
Crandall's Peruvian Restaurant & Catering	Event Service	43.074973	-89.390995	6.8
The Old Fashioned Tavern & Restaurant	Gastropub	43.076153	-89.383526	9.0
Bandung Indonesian Restaurant	Indonesian Restaurant	43.076788	-89.374243	8.7
Wah Kee Noodle Restaurant	Asian Restaurant	43.076543	-89.374983	7.4
Coliseum Bar and Restaurant	American Restaurant	43.052011	-89.379409	6.7
Maharani Indian Restaurant	Indian Restaurant	43.071510	-89.388882	6.9
The Statehouse Restaurant	American Restaurant	43.079087	-89.389802	7.5
Porta Bella Italian Restaurant	Italian Restaurant	43.073970	-89.395459	6.2
Lombardino's Restaurant	Italian Restaurant	43.072215	-89.428862	8.0
Parkway Restaurant	Diner	43.035972	-89.405006	7.0
Perkins Restaurant & Bakery	American Restaurant	43.035501	-89.408677	6.4
The City Bar & Restaurant	Bar	43.074974	-89.396050	6.0
Kabul Mediterranean Restaurant	Afghan Restaurant	43.074915	-89.394467	5.6
Monona Gardens Family Restaurant	Greek Restaurant	43.046707	-89.340218	7.6
Maharana Restaurant	Indian Restaurant	43.122106	-89.311067	7.4
Edo Japanese Restaurant	Japanese Restaurant	43.061160	-89.401103	5.9
PARTHENON GYROS	Greek Restaurant	43.074822	-89.390627	7.4
Lilliana's Restaurant	Cajun / Creole Restaurant	43.015369	-89.430096	7.8
Nitty Gritty Madison	Bar	43.071960	-89.395554	7.7
Himal Chuli	Asian Restaurant	43.074795	-89.390699	6.5
Dairyland Family Restaurant	Diner	43.084258	-89.313674	7.1

Otto's Restaurant & Bar	American Restaurant	43.059997	-89.490601	8.3
Jamerica Carribean Restaurant	Caribbean Restaurant	43.084210	-89.363428	8.1
The Curve Restaurant	American Restaurant	43.060079	-89.400565	6.2
Vientiane Palace	Thai Restaurant	43.075558	-89.390563	5.4
Essen Haus	German Restaurant	43.076218	-89.376537	7.9
L'etoile	American Restaurant	43.075183	-89.382436	8.6
El Pastor Restaurant And Bar	Mexican Restaurant	43.043956	-89.394384	7.2

Picture 1 Mapping the restaurants' locations in the Madison area



Methodology

In this project, we will use the restaurant's information to analyze the popular restaurant type and top choices for the new restaurant location.

3.1 Average Rating

To figure out what type of restaurant is suggested to start at the Madison area, we want to know what kind restaurant is more popular in the area, then the rating of each restaurant is ranked and the average of each restaurant category is calculated.

3.2 K-Cluster

To decide where to locate the restaurant, we use the K-cluster method to cluster the restaurant by the nearby venues and average the rating of each cluster.

Results and Discussion

4.1 Sort restaurant categories based on the rating score

We can average the rating scores over the restaurant category (Table 2). From the table, we can see the top 3 ranking restaurant categories are Gastropub, Indonesian Restaurant and Sushi Restaurant, the rating scores are 9.0, 8.7 and 8.5 separately. The bottom 3 ranking restaurant categories are Japanese Restaurant, Afghan Restaurant and Thai Restaurant, the rating scores are 5.9, 5.6 and 5.4. It might be because people prefer Gastropub than the other styles, so the rating score is so high. It tells us Gastropub, Indonesian Restaurant and Sushi Restaurant are very popular in the Madison area, which means investors are more likely to succeed if they start a Gastropub compared to the others. On the other hand, starting the Thai Restaurant might make investors lose their money.

Table 2 Restaurant categories sorted by rating scores

Restaurant Category	Rating score
Gastropub	9.000000
Indonesian Restaurant	8.700000
Sushi Restaurant	8.500000
Caribbean Restaurant	8.100000

German Restaurant	7.900000
Cajun / Creole Restaurant	7.800000
Greek Restaurant	7.500000
American Restaurant	7.214286
Mexican Restaurant	7.200000
Indian Restaurant	7.150000
Italian Restaurant	7.100000
Diner	7.050000
Asian Restaurant	6.950000
Bar	6.850000
Event Service	6.800000
Japanese Restaurant	5.900000
Afghan Restaurant	5.600000
Thai Restaurant	5.400000

4.2 Rank restaurant categories by the number of nearby venues

The number of nearby venues will give us an idea about how much activity is around, which means more business or people flow. Usually, people go to bars after meals, or other places which are appropriate for the after-meal activities. Basically, people like to hang around if there are more venues, more places to go, which definitely brings customers into the restaurant. From Table 3, Vientiane Palace, Nick's Restaurant & Bar and The Old Fashioned Tavern & Restaurant have the 100 venues around, which means these areas have more business and customers, we can recommend investors to locate the new restaurant in these areas.

Table 3 Counts of the nearby venues

Restaurant Name	Nearby Venue counts
Vientiane Palace	100
Nick's Restaurant & Bar	100

The Old Fashioned Tavern & Restaurant	100
Restaurant Muramoto	97
Crandall's Peruvian Restaurant & Catering	90
Himal Chuli	87
PARTHENON GYROS	87
L'etoile	83
Kabul Mediterranean Restaurant	75
The City Bar & Restaurant	70
Porta Bella Italian Restaurant	68
Essen Haus	65
Maharani Indian Restaurant	53
Maharana Restaurant	53
Nitty Gritty Madison	51
The Statehouse Restaurant	41
Jamerica Carribean Restaurant	38
Bandung Indonesian Restaurant	37
Wah Kee Noodle Restaurant	36
Otto's Restaurant & Bar	21
Perkins Restaurant & Bakery	19
The Curve Restaurant	19
Edo Japanese Restaurant	18
Lilliana's Restaurant	16
Lombardino's Restaurant	14
Parkway Restaurant	12
Dairyland Family Restaurant	12

Monona Gardens Family Restaurant	11
Coliseum Bar and Restaurant	10
El Pastor Restaurant And Bar	9

4.3 Cluster the restaurants

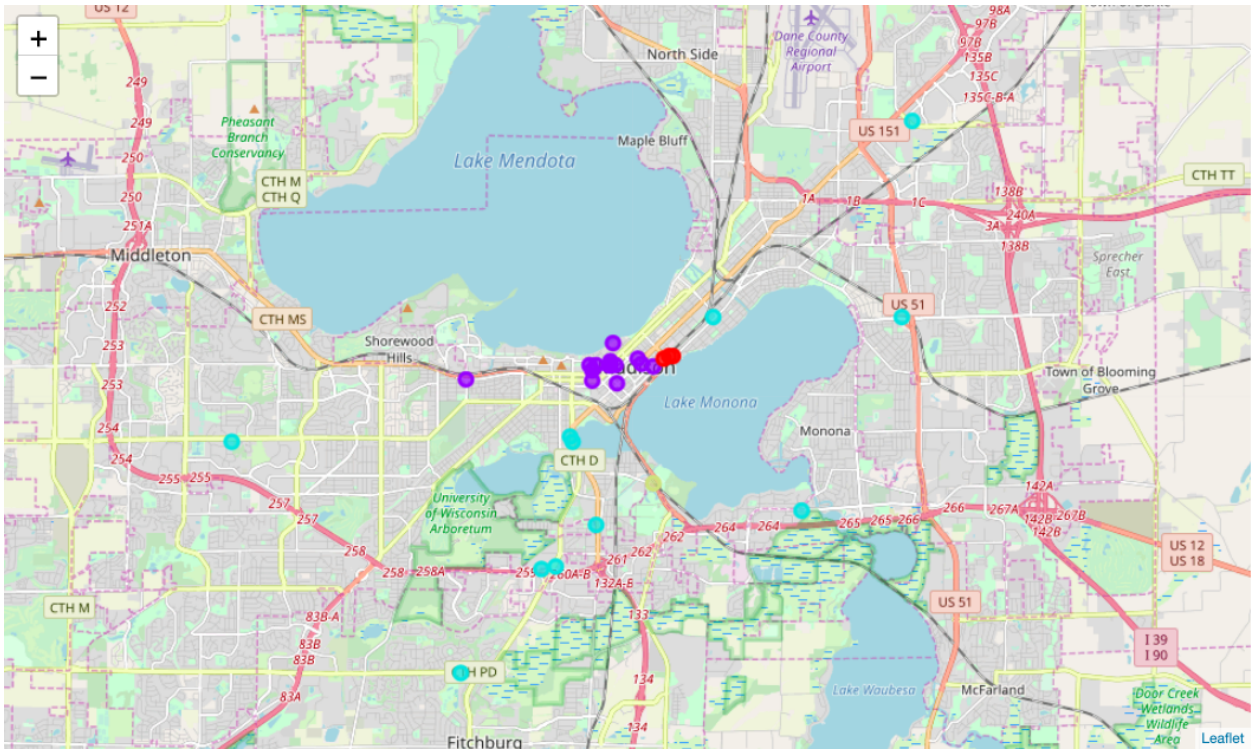
The nearby venues are added as dummy variables into the restaurant table (Picture 2). This dummy data will be used to cluster the restaurants.

Picture 2 The nearby venues of restaurant (partial shown)

	Restaurant Name	ATM	Accessories Store	American Restaurant	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Assisted Living	...	Thrift / Vintage Store	Track	Trail	Video Game Store	Video Store	Vietnamese Restaurant
0	The Old Fashioned Tavern & Restaurant	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
1	The Old Fashioned Tavern & Restaurant	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
2	The Old Fashioned Tavern & Restaurant	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
3	The Old Fashioned Tavern & Restaurant	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
4	The Old Fashioned Tavern & Restaurant	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0

We cluster the restaurants into 4 clusters, and visualize them in the map (Picture 3)

Picture 3 Restaurant Clusters in the Madison area
(Cluster0: red; Cluster1: purple; Cluster2: blue, Cluster3: green)



To compare each cluster, the rating score and nearby venues counts are averaged over the restaurants within each cluster (Table 4). Obviously, Cluster 0 and Cluster rank as the top 2. Cluster 0 (red dots in map, Picture 3) has the highest average rating score, but less average nearby venues counts than Cluster 1. Cluster 1 (purple dots in map, Picture 3) has more average nearby venues count, while lower average rating score than Cluster 0.

Table 4 Average rating score and average nearby venues counts of clusters

Cluster name	Average rating score	Average number of nearby venues
Cluster 0	8.0	46
Cluster 1	7.13	74
Cluster 2	7.18	21
Cluster 3	6.7	10

Conclusion

In conclusion, I will recommend the restaurant investors to start a restaurant from these three types: Gastropub, Indonesian and Sushi, since these restaurants have more high rating scores which

means they are more popular compared with other types of restaurants in the Madison area. Starting with either type gives more potential business to the investors, because people are more willing to go to their restaurants. While on the other hand, I won't recommend starting a Japanese restaurant, Afghan Restaurant or Thai restaurant, because the low average rating score shows the people are not tending to go to these restaurants. For the restaurant location, I would like to recommend the Cluster 0 or Cluster 1, because generally they both have relatively high average rating scores and more nearby venues, which means more potential business and more likely to get investment paid back.