Capstone Project - The Battle of Neighborhoods

Location Analysis for Opening a New Restaurant in Cleveland

Chia-Hung Kuo 18 May 2020

https://www.timeout.com/cleveland/things-to-do/best-things-to-do-in-cleveland

https://www.facebook.com/ThisisCleveland/photos/a.10150411651779714/10158606223839714/?type=3&theater

https://www.facebook.com/ThisisCleveland/photos/a.10150411651779714/10158054256509714/?type=3&theater

1. Introduction

Business understanding

Introduction Greater Cleveland area is a fast-growing metropolitan area in northeast Ohio. Its 3.5 million residents make it one of the largest Ohio metro in giant Great Lakes Megalopolis. It is fast-growing fast because of the top-ranking universities, major healthcare systems and hospitals, and the balanced work-life lifestyle. The world-class museums and attractions draw approximately 2.5 million visitors annually. Therefore, it is a great place to start a new restaurant. Utilizing the analytic approach to conduct a market analysis is a great start.

Project objectives: Find a neighborhood with the population, household income, and lifestyle for starting a successful restaurant.

Targeted Audience: People who want to start a new restaurant in the greater Cleveland area.

Problem: Needs to be in a go-to neighborhood if people want to eat out and also have regular customers to support the business.

Clear Question: How to find the optimal location for the restaurant?

2. Data

We need the county and city level of family income data and area population data with connected to county and city longitude and latitude information.

a. Data collections

The available data are obtained from Foursquare API, table from websites, and online databases:

i. Foursquare API for the listing of greater Cleveland neighborhoods information with clustered venues through.

ii. Ranking every Ohio city, county for median family income - Census Snapshot by Data Central
https://www.cleveland.com/datacentral/2017/12/ranking_every_ohio_city_county_2.html
Median family income in Ohio cities ranges from a high of $214,850 in the Columbus suburb of New Albany to below $40,000 in 14 separate cities, according to new estimates released this month by the Census Bureau.

iii. Great Cleveand Area Population Estimates by Cuyahoga County Planning Commission
https://www.countyplanning.us/resources/census-data/population-estimates/
The Census Bureau's Population Estimates Program produces population estimates on an annual basis for the nation, states, counties, and communities using components of demographic change.

iv. US Zip Code Latitude and Longitude by Open Data Soft
https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/?refine.state=OH
The data set provides the matching latitude and longitude for the cities and states.

v. ZIP Code Database by United States Zip Codes:
https://www.unitedstateszipcodes.org/zip-code-database/
This dataset provides a more detail insight into the demographics boundaries. It provides the boundary of Cuyahoga County with matched latitudes and longitudes, so that we could narrow down the searching area to only the northeast of Ohio state.

vi. Libraries: Pandas, Urllib, Geopy, Requests, Scikit Learn, and Folium

3. Methodology

We first perform data understanding, visualization, and cleaning the data. Then we analyze the income, population, and similiar business in the borough to obtain the insight of the neighborhood. Since the borough contains many cities, we perform further detailed analyses on the cities from the selected borough.

Using K-Means clustering to find out the groups of neighborhoods that seems to behave similarly to each other in the scale of borough level, and then narrow down to postal code (ZIP Code) areas.

Data understanding/Visualization/Cleaning

To have the data's initial insights, understand the descriptive statistics, and have the visualization, we performed three steps. First, we load the files to the format of datafames and visuall check the datasets' summary information. Second, we performed the data cleaning include removing the missing or invalid values and eliminating duplicate cells. Then we remove the columns that are not needed in this project and then removing the NaN values that might cause problems. Third, we did the visualize the datasets to further under the trend and to see if they are suitable for the project. After cleaning, we combined multiple data sources of the needed data with the matching reference to the city and zip code into a dataframe for use through the project.

Table 1. The Family/Household income data

	Rank	City	2012-2016 Median family income	2007-2011 Median family income	2012-2016 Median household income	2007-2011 Median household income
0	1	New Albany	214850	202243	191375	182321
1	2	Indian Hill	211795	$250,000+	205221	243500
2	3	Pepper Pike	186765	171547	164471	151250
3	4	Powell	146442	151066	132917	142083
4	5	Dublin	144005	143619	125540	121431

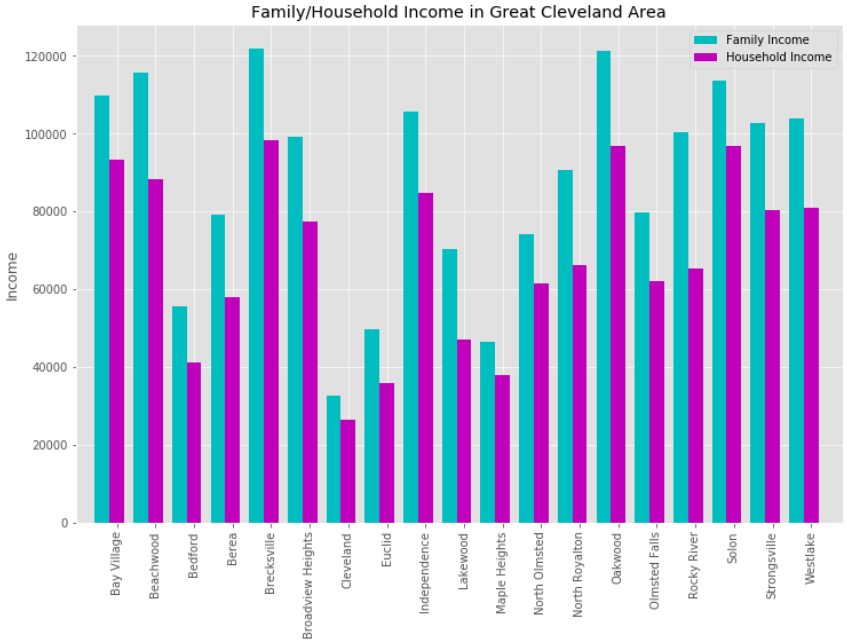

Figure 1. The Family/Household income of greater Cleveland area

Table 2. Greater Cleveland Area Population from 2010 to 2018

[5]:

	Geography	census	Estimate Base	2010	2011	2012	2013	2014	2015	2016	2017	2018
0	Bay Village	15651	15651	15629	15544	15491	15484	15470	15428	15378	15343	15295
1	Beachwood	11953	11927	11909	11844	11814	11803	11785	11743	11703	11677	11658
2	Bedford	13074	13074	13052	12966	12909	12879	12831	12767	12704	12626	12561
3	Bedford Heights	10751	10757	10740	10692	10654	10729	10725	10680	10633	10579	10534
4	Bentleyville	864	864	863	864	859	859	857	855	854	853	851

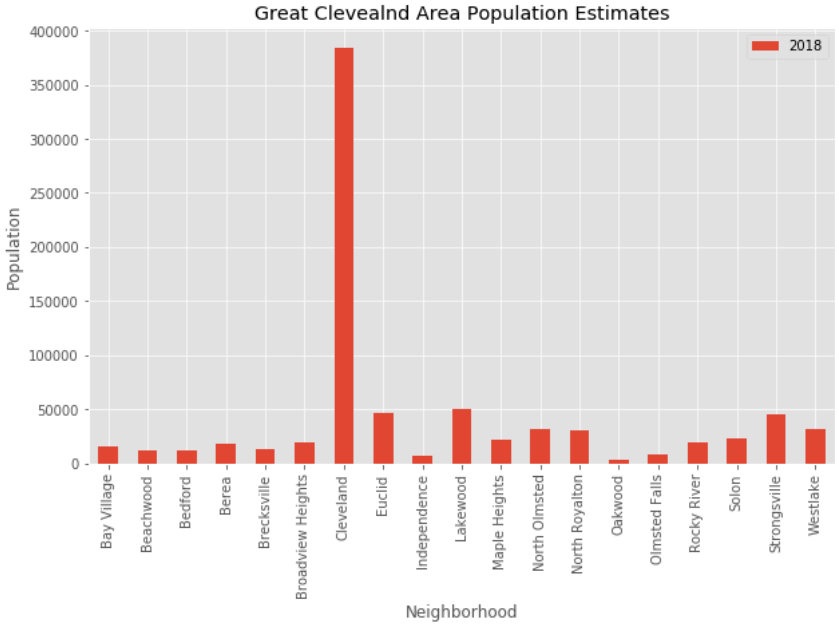

Figure 2. The Greater Cleveland Area Population Estimates of 2018

Although Cleveland has the most population, the income is lower than the surrounded area. The possible reasons could be that Cleveland is a big district contains the most extended list of zip code cities among others, and therefore it has the most residents. For our purpose, we also need to consider that most people commute to work in downtown Cleveland and live in surrounded cities. Cleveland's great feature is that lots of traveling attractions are close to each other, and most of them are in forty minutes' drive from downtown Cleveland.

Table 3. US Zip Code Latitude and Longitude

	Zip	City	State	Latitude	Longitude	Timezone	Daylight savings time flag	geopoint
0	43984	New Rumley	OH	40.296490	-81.102502	-5	1	40.29649, -81.102502
1	43681	Toledo	OH	41.686778	-83.439430	-5	1	41.686778, -83.43943
2	43733	Derwent	OH	39.923616	-81.542965	-5	1	39.923616, -81.542965
3	43334	Marengo	OH	40.399648	-82.807830	-5	1	40.399648, -82.80783
4	45841	Jenera	OH	40.881217	-83.731990	-5	1	40.881217, -83.73199

Table 4. Updated information of selected county and Neighborhood with updated format of zip codes

	county	state	zip	Neighborhood	Latitude	Longitude
18504	Licking County	OH	43001.0	Alexandria	40.08	-82.61
18505	Franklin County	OH	43002.0	Amlin	40.07	-83.18
18506	Delaware County	OH	43003.0	Ashley	40.40	-82.95
18507	Franklin County	OH	43004.0	Blacklick	40.02	-82.80
18508	Knox County	OH	43005.0	Bladensburg	40.29	-82.28

	zip	Neighborhood	Latitude	Longitude
19056	44101	Cleveland	41.49	-81.67
19057	44102	Cleveland	41.48	-81.74
19058	44103	Cleveland	41.52	-81.64
19059	44104	Cleveland	41.48	-81.63
19060	44105	Cleveland	41.45	-81.63

Table 5. The merged data

	Neighborhood	Family Income	Household Income	zip	Latitude	Longitude	2018
0	Bay Village	109673	93220	44140	41.48	-81.92	15295
1	Beachwood	115536	88287	44122	41.47	-81.50	11658
2	Bedford	55684	41285	44146	41.39	-81.53	12561
3	Berea	79306	57896	44017	41.36	-81.86	18655
4	Brecksville	121974	98345	44141	41.30	-81.61	13632
5	Broadview Heights	99286	77480	44147	41.32	-81.68	19202

4. Results

By utilizing the geopy, Folium, and Foursquare libraries, we are able to obtain the location information of each neighborhood in greater Cleveland. The merged demographic data of neighborhoods is then fetched into a location dataframe. Then are fetched into venues near neighborhoods in the selected area through using Foursquare API. OneHotEncoding was also used for creating dummy variables for the venue category as the categorical variables. The K-Means cluster algorithm is used to classifying the neighbors into five groups. Finally, we repeated the previous process and narrow down the scope on a select area to look into the cities identified by zip codes. Finally, we will provide suggestions on choosing a location for our client. The map of the greater Cleveland area is created by using the Folium library with the location information from Nominatim and Foursquare.

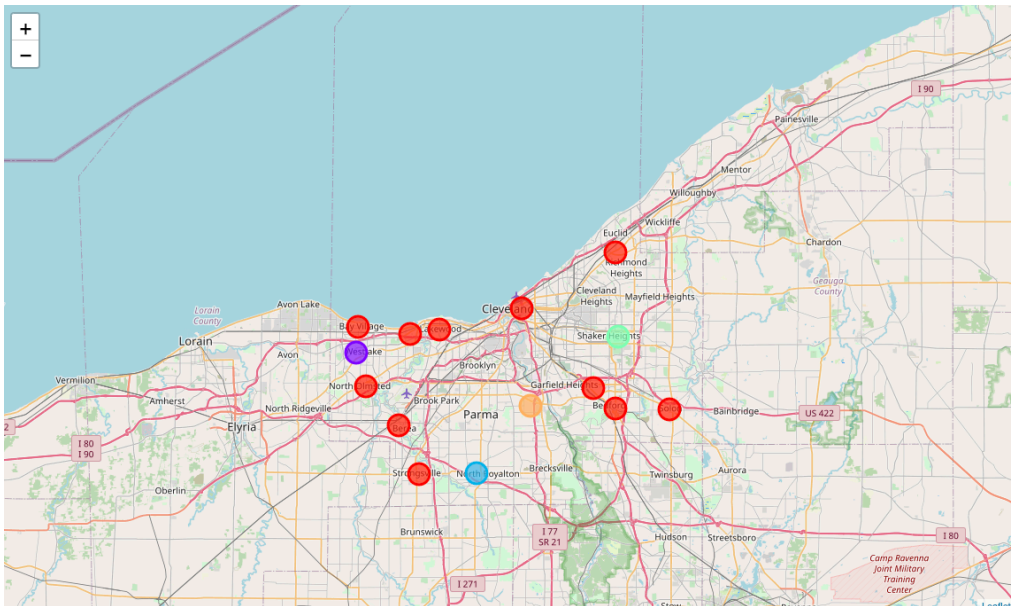

Figure 3. Cluster of neighbothood of Cleveland area

Table 6. The venues of the cluster

	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bay Village	41.484193	-81.926580	2.0	Gym / Fitness Center	Liquor Store	Park	Sandwich Place	Mobile Phone Shop
1	Beachwood	41.472881	-81.522090	4.0	Intersection	Golf Course	Women's Store	Football Stadium	Dim Sum Restaurant
2	Bedford	41.389371	-81.527090	2.0	Bar	Home Service	Pharmacy	Video Store	Supermarket
3	Berea	41.369950	-81.862590	2.0	Pet Store	Fast Food Restaurant	Pharmacy	Vietnamese Restaurant	Discount Store
6	Cleveland	41.505744	-81.672797	2.0	Grocery Store	Gas Station	Bowling Alley	Shopping Plaza	Sandwich Place
7	Euclid	41.570344	-81.527260	2.0	American Restaurant	Discount Store	Sandwich Place	Bar	Fast Food Restaurant
8	Independence	41.391753	-81.658160	2.0	Salad Place	Financial or Legal Service	Hotel	Mobile Phone Shop	Dim Sum Restaurant
9	Lakewood	41.480881	-81.800360	2.0	Sandwich Place	Mobile Phone Shop	Salon / Barbershop	Bakery	Bank
10	Maple Heights	41.412653	-81.560670	2.0	Hardware Store	Sandwich Place	Discount Store	Basketball Court	Farmers Market
11	North Olmsted	41.415097	-81.914360	2.0	Furniture / Home Store	Mobile Phone Shop	American Restaurant	Bank	Shoe Store

Cleveland is one of the clusters with various types of restaurants, coffee shops, and bars. Also, its location is close to the center of the cluster, and therefore, we selected the neighbor, and look into the cities by the index of the zip codes. The results are shown in Figure 4.

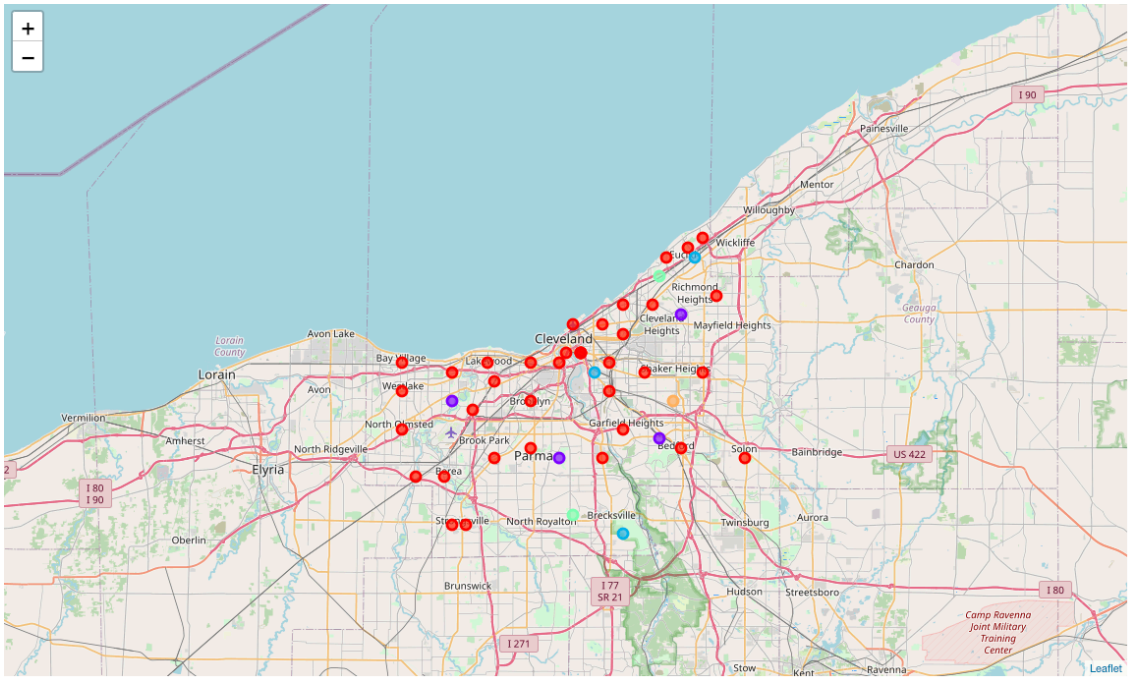

Figure 4. Cluster of zip code cities of Cleveland area

Table 7. The venues of the cluster

	zip	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
0	44140	Coffee Shop	Gym / Fitness Center	Diner	Pizza Place	Pharmacy	New American Restaurant	Italian Restaurant	Ice Cream Shop	Grocery Store	Furniture / Home Store
1	44122	American Restaurant	Cycle Studio	Steakhouse	Park	Cosmetics Shop	Cupcake Shop	Convenience Store	Dance Studio	Falafel Restaurant	Diner
2	44146	Bar	Bowling Alley	Pharmacy	Shipping Store	Lawyer	Café	Supermarket	Pizza Place	Breakfast Spot	Video Store
3	44017	Lake	Scenic Lookout	Snack Place	Discount Store	Falafel Restaurant	Entertainment Service	Dry Cleaner	Donut Shop	Dive Bar	Yoga Studio
6	44129	Coffee Shop	Business Service	Yoga Studio	Farm	Cosmetics Shop	Cupcake Shop	Cycle Studio	Dance Studio	Department Store	Diner

Table 8. The venues of the cluster

	zip	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
8	44134	Park	Construction & Landscaping	Farm	Convenience Store	Cosmetics Shop	Cupcake Shop	Cycle Studio	Dance Studio	Department Store	Diner
39	44126	Park	Beer Garden	Discount Store	Falafel Restaurant	Entertainment Service	Dry Cleaner	Donut Shop	Dive Bar	Yoga Studio	Farm
45	44121	Park	Furniture / Home Store	Construction & Landscaping	Convenience Store	Cosmetics Shop	Cupcake Shop	Cycle Studio	Dance Studio	Farm	Department Store
53	44137	Construction & Landscaping	Concert Hall	Convenience Store	Cosmetics Shop	Cupcake Shop	Cycle Studio	Dance Studio	Department Store	Diner	Discount Store

Table 9. The venues of the cluster

	zip	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
4	44141	Trail	Stables	Yoga Studio	Diner	Entertainment Service	Dry Cleaner	Donut Shop	Dive Bar	Discount Store	Department Store
25	44127	Bus Station	Pizza Place	Intersection	Bike Rental / Bike Share	Clothing Store	Trail	Grocery Store	Cycle Studio	Dance Studio	Department Store
50	44117	Bus Station	Yoga Studio	Farm	Convenience Store	Cosmetics Shop	Cupcake Shop	Cycle Studio	Dance Studio	Department Store	Diner

Table 10. The venues of the cluster

	zip	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
5	44147	Food	Yoga Studio	Farm	Convenience Store	Cosmetics Shop	Cupcake Shop	Cycle Studio	Dance Studio	Department Store	Diner
36	44110	Food	Yoga Studio	Farm	Convenience Store	Cosmetics Shop	Cupcake Shop	Cycle Studio	Dance Studio	Department Store	Diner

Table 11. The venues of the cluster

	zip	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
24	44128	Café	Yoga Studio	Farm	Convenience Store	Cosmetics Shop	Cupcake Shop	Cycle Studio	Dance Studio	Department Store	Diner

The venues in the most significant cluster show that the locations are in the go-to places for people who want to dine out. As long as the new restaurant has a unique feature, there is still a great opportunity to be successful since it is in the area with people's flow.

5. Discussion

Based on the analysis, we recommend to those who plan to open a new restaurant starting from considering the locations in the cluster around the downtown Cleveland area. To find the uniqueness for a new restaurant, we still need to look into the most popular venues from the targeted cluster for further insight to secure the opportunity of a successful business.

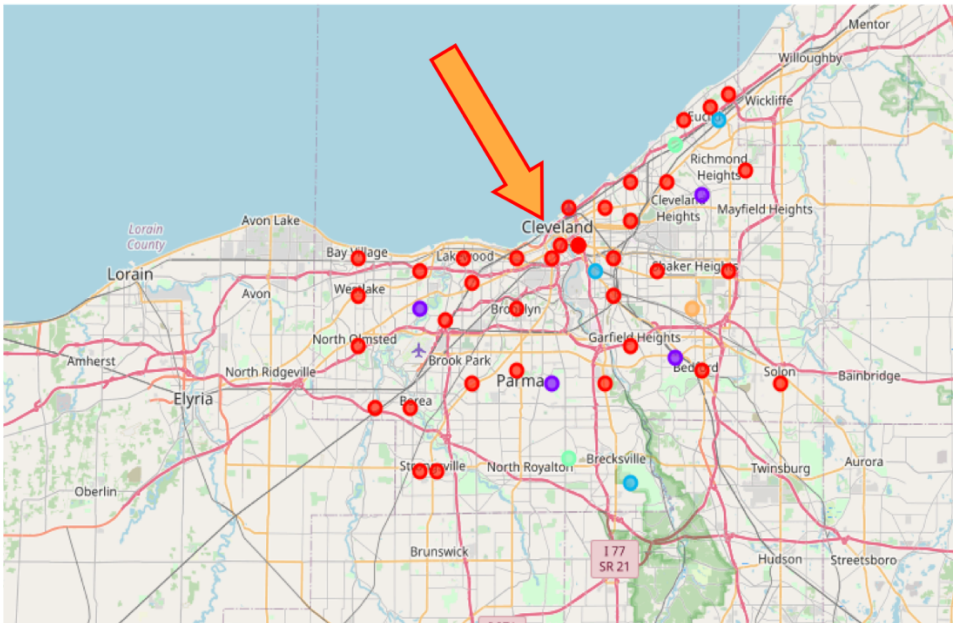

Figure 5. The location suggestion

6. Conclusion

This analysis provides suggestions on the locations for the first step. The next step will be thinking about other factors such as the operating cost, and so on. The demographics, accessibility, and labor costs are also essential factors for opening a new restaurant. The restaurants have been clustered and can help anyone looking forward to start a restaurant business in the greater Cleveland area.