Capstone Project - The Battle of Neighborhoods

Location Analysis for Opening a New Restaurant in Cleveland

Chia-Hung Kuo 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. **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

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

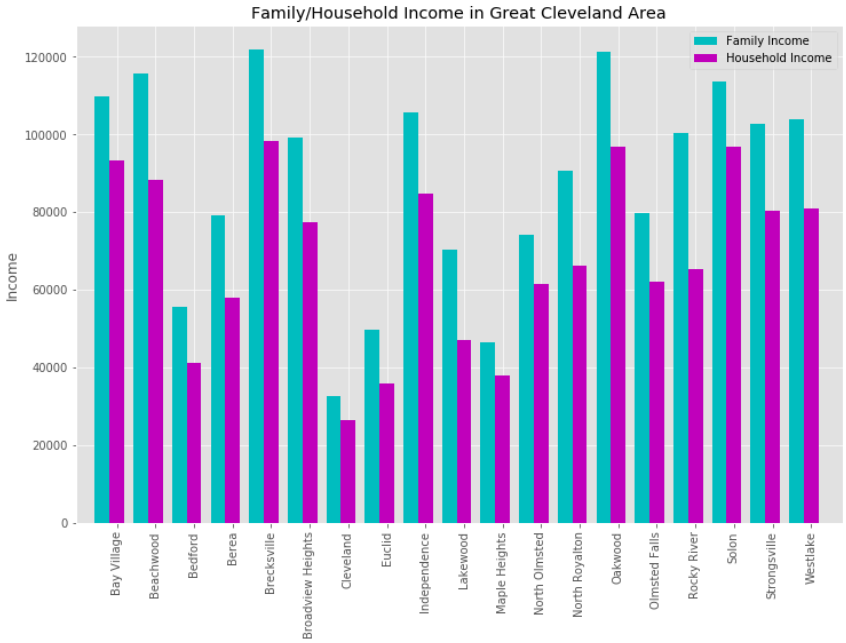

Figure 1. The Family/Household income of greater Cleveland area

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

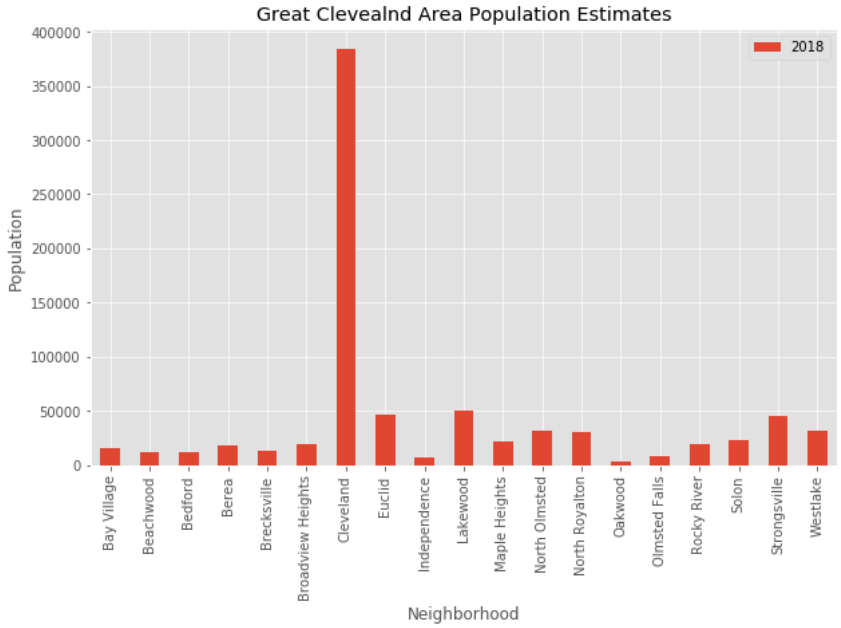

Figure 2. The Greater Cleveland Area Population Estimates of 2018

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