

Capstone Project - The Battle of Neighborhoods
Compare and Analyze Data of Two Cities: Raleigh, NC and
Atlanta, GA
House Sale or Rent Prices and Venues Data Analysis of the
Optimal City

Li Yan Jiang

September 4, 2021

1 Introduction

1.1 Background

Many middle-class people plan to move out New York City due to high tax, high houses' sale price, high houses' rental price, poor housing, the crowded city, education quality declining, and poor security. The stakeholders would like to find a place with a good sense of community, a safe and friendly neighborhood, and less crowded and quiet area.

1.2 Problem

How can we tell which city will be a optimal city to move to based on these criteria? Which location of the optimal city will be the better place to live in? What popular venues are around this location? What are the houses' sale prices or rental prices around this location?

1.3 Interest

In this project, it will compare, and analyze the data about two cities: Atlanta, Georgia and Raleigh, North Carolina that a stakeholder interested in moving to based on the optimal criteria. So people who would like to move to one of these two cities will be interested.

2 Data

2.1 Based on definition of the problem, data will be needed:

- From the AreaVibes websites what locals think about their city, rating of different categories, and top rated schools
- From the Census website for population data
- From the Zillow website for houses' sales and rentals
- From FourSquare for the popular venues, ratings, and tips

2.2 Steps need to take to solve the problem:

1. Web scrape the data about Atlanta, Georgia and Raleigh, North Carolina from AreaVibes and United States Census using Request and BeautifulSoup

	Atlanta_GA	Raleigh_NC
Family friendly	8.19	22.04
Public transit is accessible	19.89	9.86
Walkable to grocery stores	19.11	8.12
Yards are well-kept	8.19	11.89
Lots of parks	17.94	20.01
Walkable to restaurants	19.89	11.89
Friendly neighbors	10.92	17.11
Safe at night	8.19	17.98
Pet friendly	14.82	15.95
Streets have sidewalks	24.96	17.11
A quiet area	5.07	11.02
Sense of community	8.19	13.92

Table 1: What Do Locals Think?

2. Clean, compare and analyze the data to find out which city is optimal to move to using Pandas and Matplotlib for visualization, comparison, and analysis
3. Get the top rated schools' data of the optimal city from AreaVibes and find the optimal areas around these schools using Geopy.geocoders
4. Explore popular spots of each optimal centroid using Request and FourSquare API
5. Find houses' sale and rental data around each optimal area by web scraping data from Zillow using Request and BeautifulSoup
6. Mark these spots/places on a map using Folium
7. Search a specific venue and find rating and tips about the specific venue using Request and FourSquare API

3 Methodology

3.1 Compare and Analyze Data

First, the data were web scraped from [AreaVibes](#) about what locals think about their city. Then the data were combined into one table for compare and analyze. Based on the higher number of people think positive about family friendly, parks, friendly neighbors, safe at night, and sense of community in North Carolina. The data shown that Raleigh, NC is an optimal city to move to. See Table 1 and Figure 1.

Then, the second data were also scraped from [AreaVibes](#) about the overall ratings of each city and combined into one table for comparison and analysis. The ratings are based on the categories: Livability, Amenities, Cost of Living, Crime, Employment, Housing, and Schools. See Table 2. From this table, the data also shown that Raleigh, NC is the optimal city.

Last, the third data were scraped from [the United States Census](#) about the population data of the cities. After filtering and cleaning the data, the data shown that Raleigh, NC is the optimal city based on the number of population, median household income, and poverty. See figure 2, 3, and 4.

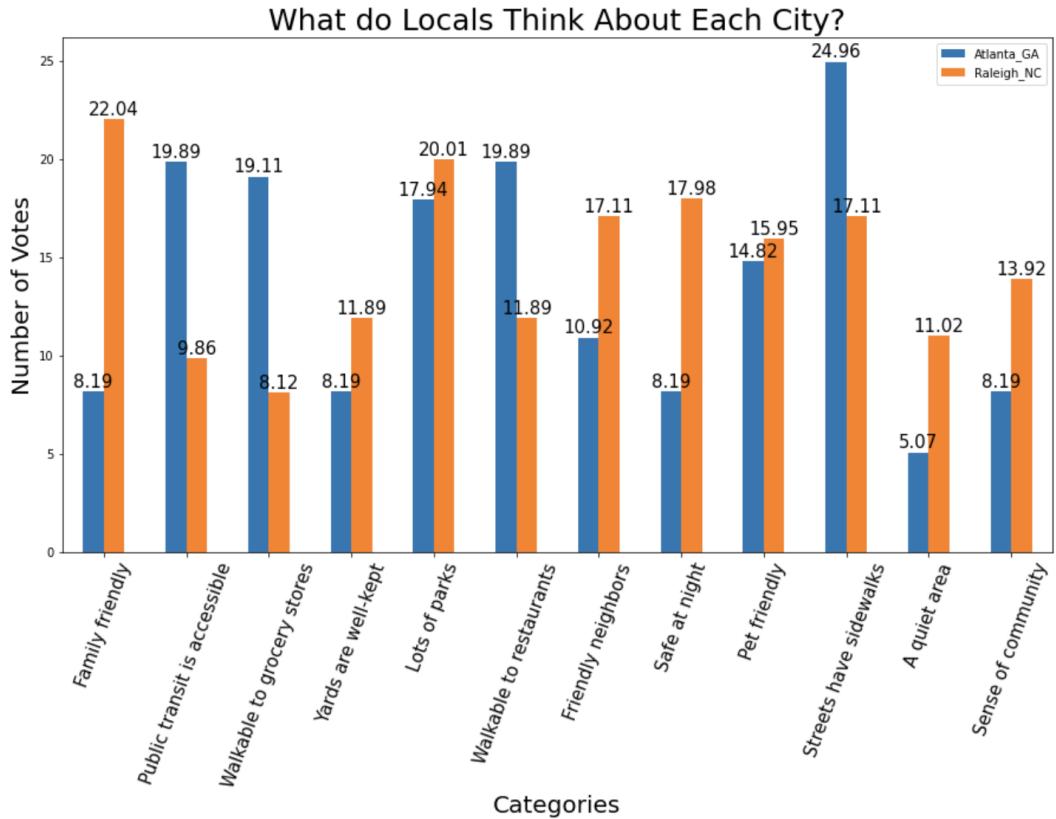


Figure 1: What Do Locals Think?

	Livability	Amenities	Cost of Living	Crime	Employment	Housing	Schools	User Ratings
Atlanta_GA	62	A+	D+	F	D	C+	F	D+
Raleigh_NC	78	A+	D+	B	C	B-	C+	D+

Table 2: The Overall Rating of Each City

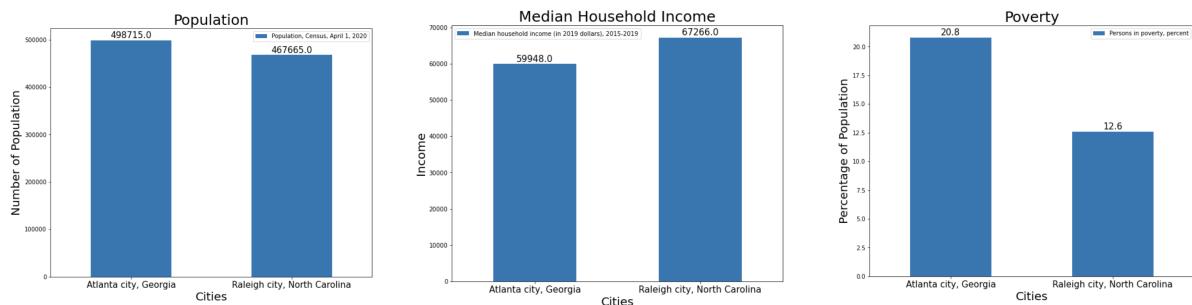


Figure 2:

Figure 3:

Figure 4:

	Name	Grades	Students	Proficiency	Address
0	Raleigh Charter High School	9 - 12	559	94%	1307 Glenwood Avenue, Raleigh NC, 27605
1	Magellan Charter	3 - 8	406	92%	9324 Baileywick Road, Raleigh NC, 27615
2	Quest Academy	KG - 8	144	91%	10908 Strickland Road, Raleigh NC, 27615
3	Wake Early College Of Health And Science	9 - 13	309	80%	2901 Holston Lane, Raleigh NC, 27610
4	Envision Science Academy	KG - 6	252	80%	9400 Forum Dr, Raleigh NC, 27615
5	Wake Stem Early College High School	9 - 12	208	80%	715 Barbour Dr, Raleigh NC, 27603
6	Sycamore Creek Elementary	PK - 5	1,132	78%	10921 Leesville Rd, Raleigh NC, 27613

Table 3: The Top Rated Schools

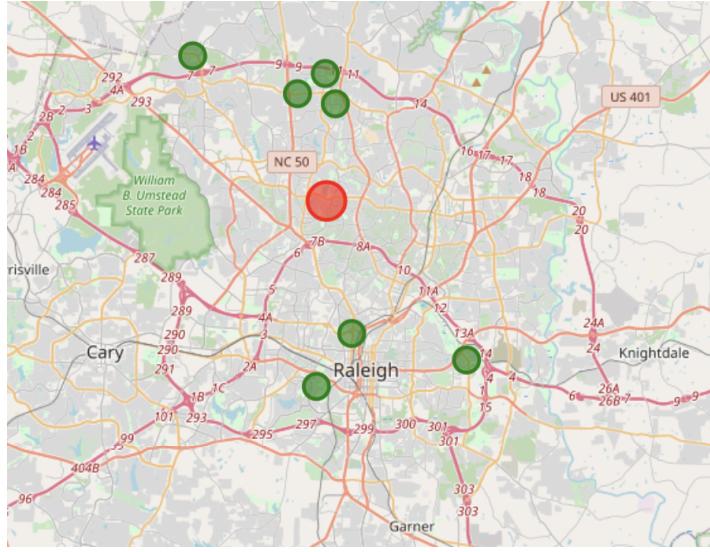


Figure 5: The Centroid and Top-rated Schools

3.2 Find the optimal locations in Raleigh, NC

After finding out the optimal city, the top-rated schools' neighborhoods were used to find the optimal locations in the city of Raleigh, NC. The data were scraped from [Areavibes](#) too. See table 3. Based on these data, Geopy.geocodes library was used to find the coordinate of the center location around these schools. See figure 5.

Moreover, two more centroids were found based the group of K-8 schools and 9-12 schools. After that, continue scraping data from [Areavibes](#) based on the location of zip codes and centroids. These data were combined together into one table for comparison and analysis. See table 4.

Based on the rating of livability, see figure 6, three locations of high rating scores were selected to compare with NY's data. In addition, Cost of Living and Housing (Table 5), Crime (Figure 7), and Employment (Figure 8 and 9) data was scraped deeper from [Areavibes](#) for a closer comparison and analysis. At last, Zip 27612, 27613, and 27615 are shown the most optimal areas in all.

3.3 Explore popular spots around the optimal areas using FourSquare API

The FourSquare API was utilized to explore popular spots around each zip code within radius 2000 meters and the limit was set 50. Here is the head of the list Venues' information and another list of the most common venue. See table 6 and 7 for the example of the zip code 27612.

All popular spots were added to the map for visualization. See figure 10 and 11.

	Livability	Amenities	Cost of Living	Crime	Employment	Housing	Schools
New York, NY	65	A+	F	F	C-	D+	D
Zip_27601	68	A+	D	F	D-	C-	B
Zip_27610	71	C	B-	C+	D-	C+	F
Zip_27612_3	77	B+	F	A+	B	F	A-
Zip_27605	83	A+	D	B+	C	B+	C
Zip_27615	84	A+	D-	A+	B	F	B+

Table 4: Overall Ratings

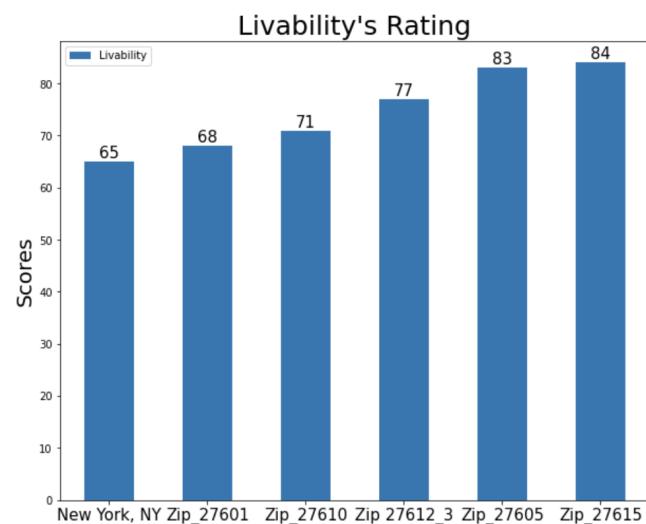


Figure 6: The Rating of Livability

	Cost of Living	Median home value	Median rent price	Home ownership
Zip_27605	104	\$256,791	\$642	20%
Zip_27615	108	\$266,389	\$873	63%
Zip_27612_3	110	\$279,428	\$973	52%
New York, NY	161	\$508,900	\$1,294	32%

Table 5: Cost of Living and Housing

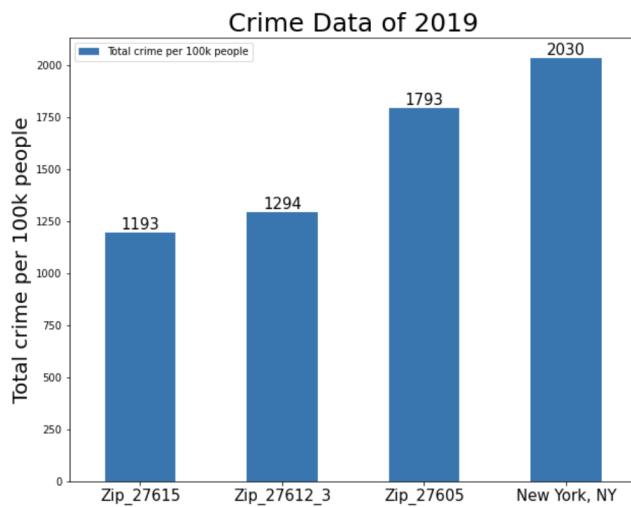


Figure 7: Crime Data

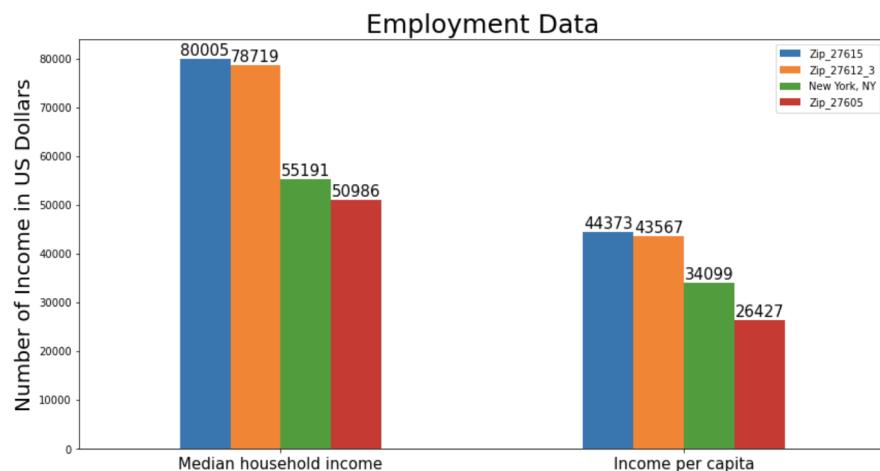


Figure 8: Employment Data

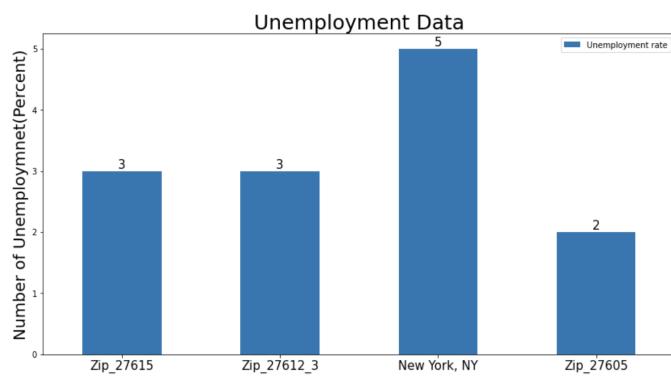


Figure 9: Unemployment Data

	name	categories	address	lat	lng	labeledLatLngs	distance	postalCode	cc	neighborhood	city	state	co
0	Shelley Lake Trail	Trail	Millbrook Road	35.857808	-78.662354	[{"label": "display", "lat": 35.85780832693374, "lng": -78.662354}	390	27612	US	Raleigh, NC	Raleigh	NC	Uni Sta
1	North Hills Club	Social Club	4824 Yadkin Dr	35.847940	-78.651107	[{"label": "display", "lat": 35.84794009711326, "lng": -78.651107}	1147	27609	US	Nan	Raleigh	NC	Uni Sta
2	Harris Teeter	Supermarket	2051 W Millbrook Rd	35.857178	-78.679772	[{"label": "display", "lat": 35.85717816413243, "lng": -78.679772}	1726	27612	US	Nan	Raleigh	NC	Uni Sta
3	Mod Pizza	Pizza Place	2071 West Millbrook Road	35.857870	-78.680248	[{"label": "display", "lat": 35.85787031, "lng": -78.680248}	1783	27612	US	Nan	Raleigh	NC	Uni Sta
4	Relish Craft Kitchen & Bourbon Bar	Café	5625 Creedmoor Rd	35.858261	-78.682087	[{"label": "display", "lat": 35.85826144882770, "lng": -78.682087}	1954	27612	US	Northwest Raleigh	Raleigh	NC	Uni Sta

Table 6: Head List of Venues' Info of Zip Code 27612

categories	Trail	Park	Supermarket	Pizza Place	Cosmetics Shop	Hotel	Track	Social Club	Sandwich Place	Pool	...	Frozen Yogurt Shop	Fried Chicken Joint	Fast Food Restaurant	Coffee Shop	Chinese Restaurant
counts	4	4	2	2	2	1	1	1	1	1	...	1	1	1	1	1

1 rows x 25 columns

Table 7: The Most Common Venues' Ranking of Zip Code 27612

Color Code Chart for Reading the Map

Facts	Colored Markers
Centroids	red
Top Rated Schools	green
Popular Spots	blue

Figure 10:

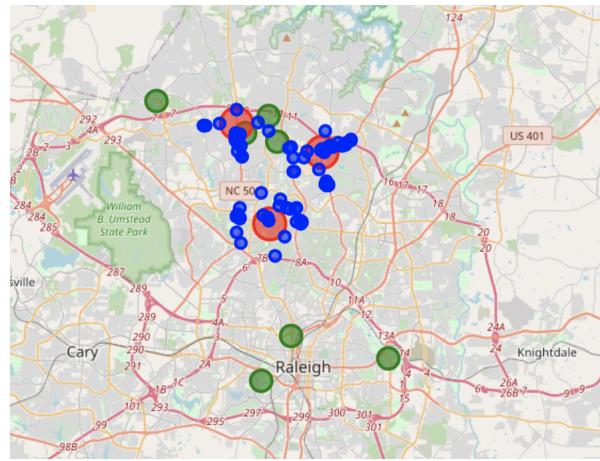


Figure 11:

	SalePrices	Features	Address	List_Link
0	\$360,000	3 bds, 3 ba, 2,213 sqft, Coming soon	530 Weathergreen Dr, Raleigh, NC 27615	https://www.zillow.com/homedetails/530-Weathergreen-Dr-Raleigh-NC-27615/6478013_zpid/
1	\$350,000	4 bds, 3 ba, 2,112 sqft, Coming soon	9320 Carisbrook Ct, Raleigh, NC 27615	https://www.zillow.com/homedetails/9320-Carisbrook-Ct-Raleigh-NC-27615/6531083_zpid/
2	\$265,000	3 bds, 2 ba, 1,522 sqft, House for sale	313 Crown Oaks Dr, Raleigh, NC 27615	https://www.zillow.com/homedetails/313-Crown-Oaks-Dr-Raleigh-NC-27615/6466653_zpid/
3	\$279,900	3 bds, 3 ba, 2,268 sqft, Condo for sale	6529 New Market Way #1, Raleigh, NC 27615	https://www.zillow.com/homedetails/6529-New-Market-Way-1-Raleigh-NC-27615/2068606066_zpid/
4	\$275,000	3 bds, 2 ba, 1,382 sqft, Coming soon	7937 Brown Bark Pl, Raleigh, NC 27615	https://www.zillow.com/homedetails/7937-Brown-Bark-Pl-Raleigh-NC-27615/6470905_zpid/
5	\$250,000	2 bds, 3 ba, 1,543 sqft, Coming soon	900 Bridge Way, Raleigh, NC 27615	https://www.zillow.com/homedetails/900-Bridge-Way-Raleigh-NC-27615/6471478_zpid/
6	\$589,000	4 bds, 3 ba, 3,771 sqft, House for sale	2301 Valley Forge Dr, Raleigh, NC 27615	https://www.zillow.com/homedetails/2301-Valley-Forge-Dr-Raleigh-NC-27615/6460853_zpid/
7	\$425,000	3 bds, 2 ba, 1,314 sqft, House for sale	7405 Valley Run Dr, Raleigh, NC 27615	https://www.zillow.com/homedetails/7405-Valley-Run-Dr-Raleigh-NC-27615/6468550_zpid/
8	\$409,900	3 bds, 2 ba, 1,470 sqft, Coming soon	6729 Van Haven Dr, Raleigh, NC 27615	https://www.zillow.com/homedetails/6729-Van-Haven-Dr-Raleigh-NC-27615/6441056_zpid/

Table 8: The Head List of Houses for Sale of Zip Code 27615

	Features	Address	List_Link
0	\$1,170+/mo 1 bd1 ba, 690 sqft, Apartment for rent	7303 Bryn Athyn Way #5KWCKJ3BB, Raleigh, NC 27615	https://www.zillow.com/homedetails/7303-Bryn-Athyne-Way-5KWCKJ3BB-Raleigh-NC-27615/2090957290_zpid/
1	\$1,800/mo 4 bds, 2.5 ba, 2,250 sqft, Townhouse for rent	747 Weathergreen Dr, Raleigh, NC 27615	https://www.zillow.com/homedetails/747-Weathergreen-Dr-Raleigh-NC-27615/6469824_zpid/
2	\$1,455/mo 3 bds, 2.5 ba, 1,383 sqft, House for rent	7223 Sandy Creek Dr, Raleigh, NC 27615	https://www.zillow.com/homedetails/7223-Sandy-Creek-Dr-Raleigh-NC-27615/132229950_zpid/
3	1,4401bd,1,295+ 2 bds,	2525 Pavilion Pl, Raleigh, NC	https://www.zillow.com/b/legacy-north-pointe-apartments-raleigh-nc-5Xhyfr/
4	\$1,700/mo 3 bds, 3 ba, 1,520 sqft, Townhouse for rent	102 Wythe Cir, Raleigh, NC 27615	https://www.zillow.com/homedetails/102-Wythe-Cir-Raleigh-NC-27615/6469817_zpid/
5	1, 265 + 1bd,1,710+ 2 bds, \$1,755 3 bds,	8810 Autumn Winds Dr, Raleigh, NC	https://www.zillow.com/b/the-village-apartments-raleigh-nc-5XhyjX/
6	1, 438 + 1bd,1,599+ 2 bds, \$1,946 3 bds,	110 Talisman Way, Raleigh, NC	https://www.zillow.com/b/junction-six-forks-raleigh-nc-5g3RJr/
7	\$1,295/mo 2 bds, 1.5 ba, 1,250 sqft, Townhouse for rent	7205 Harps Mill Rd, Raleigh, NC 27615	https://www.zillow.com/homedetails/7205-Harps-Mill-Rd-Raleigh-NC-27615/50098649_zpid/
8	\$1,795/mo 3 bds, 2.5 ba, 1,607 sqft, Townhouse for rent	7004 Racine Way, Raleigh, NC 27615	https://www.zillow.com/homedetails/7004-Racine-Way-Raleigh-NC-27615/82676290_zpid/

Table 9: The Head List of Houses for Rent of Zip Code 27615

3.4 Get Current "House For Sale" and "House For Rent" Data

Showing the stakeholders what the current house sale prices, rental prices are, which will help them be prepared and make a decision for moving. The data were scraped from [Zillow](#) for each zip code. Here is the list of houses for sale and houses for rent of zip 27612. See table 8 and 9.

All houses for sale and houses for rent based on three zip codes were added to the map for visualization. See figure 12 and 13. However, the list of houses for sale and houses for rent may not be same as today because the data on [Zillow](#) will be updated each day.

3.5 Search for a specific venue category

The stakeholder likes craving Italian and Chinese food, who would like to know where these venues are and what the venue id is. The FourSquare API is used to find all these information. Here are the list of Italian Restaurants (Table 10) and the list of Chinese Restaurants (Table 11) around zip code 27615 within the radius as 5000 meters from the center.

All these locations were added to the map for visualization. See figure 14 and 15.

Color Code Chart for Reading the Map

Facts	Colored Markers
Centroids	red
Top Rated Schools	green
Popular Spots	blue
Houses for Sale	purple
Houses for Rent	orange

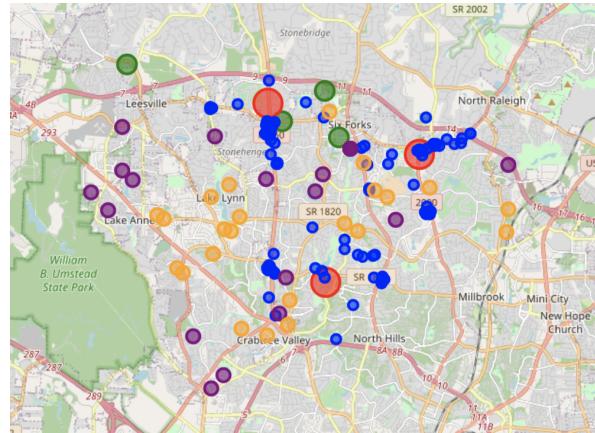


Figure 12:

Figure 13:

id	name	categories	address	lat	lng	labeledLatLngs	distance	postalCode	cc	city	state	country	formattedAddress	crossStreet	
0	4b760c30f964a5200e392ee3	Nina's Ristorante	Italian Restaurant	8801 Lead Mine Rd	35.905173	-78.656968	[{"label": "display", "lat": 35.9051734620298..., "lng": -78.656968}, {"label": "entrance", "lat": 35.8815173, "lng": -78.645271}	3062	27615	US	Raleigh	NC	United States	[8801 Lead Mine Rd, Raleigh, NC 27615]	NaN
1	4f7de2276b7478c17cbc5ad5	Vesuvios Italian Kitchen	Italian Restaurant	7400 Six Forks Rd	35.8815173	-78.645271	[{"label": "display", "lat": 35.8815173, "lng": -78.645271}	2011	27615	US	Raleigh	NC	United States	[7400 Six Forks Rd, Raleigh, NC 27609]	NaN
2	4b326152f964a520120a25e3	Italian Kitchen	Italian Restaurant	4005 Wake Forest Rd	35.835394	-78.615093	[{"label": "display", "lat": 35.835394, "lng": -78.615093}	6400	27609	US	Raleigh	NC	United States	[4005 Wake Forest Rd, Raleigh, NC 27609]	NaN
3	4e9f414e93add232e748d496	Mia Francesca Trattoria	Italian Restaurant	4100 Main At North Hills St	35.836218	-78.643951	[{"label": "display", "lat": 35.836218, "lng": -78.643951}	6399	27609	US	Raleigh	NC	United States	[4100 Main At North Hills St (in North Hills), NC 27609]	in North Hills
4	4c6ab44f3d8bc9b605118d90	Capri Restaurant	Italian Restaurant	6325-17 Falls of Neuse Rd	35.868659	-78.622269	[{"label": "display", "lat": 35.86865910463733..., "lng": -78.622269}	2641	27615	US	Raleigh	NC	United States	[6325-17 Falls of Neuse Rd, Raleigh, NC 27615]	NaN

Table 10: The List of Italian Restaurants around Zip Code 27615

id	name	categories	address	lat	lng	labeledLatLngs	distance	postalCode	cc	city	state	country	formattedAddress	crossStreet	
0	4f328ad219836c91c7e19cf	Eastern Chinese Restaurant	Chinese Restaurant	9650 Strickland Rd	35.903823	-78.656052	[{"label": "display", "lat": 35.90382329999999..., "lng": -78.656052}	2919	27615	US	Raleigh	NC	United States	[9650 Strickland Rd, Raleigh, NC 27615]	NaN
1	4b9iae20af964a5203fd35e3	King's Wok	Chinese Restaurant	8320-160 Litchford Rd	35.892808	-78.600528	[{"label": "display", "lat": 35.892808, "lng": -78.600528}	2399	27615	US	Raleigh	NC	United States	[8320-160 Litchford Rd (Harps Mill, Raleigh, NC 27615)]	Harps Mill
2	4ae095a9f964a5209f8021e3	Imperial Garden	Chinese Restaurant	7713-15 Lead Mine Rd	35.885736	-78.680382	[{"label": "display", "lat": 35.885736, "lng": -78.680382}	3081	27615	US	Raleigh	NC	United States	[7713-15 Lead Mine Rd, Raleigh, NC 27615]	NaN

Table 11: The List of Chinese Restaurants around Zip Code 27615

Color Code Chart for Reading the Map

Facts	Colored Markers
Centroids	red
Top Rated Schools	green
Popular Spots	blue
Houses for Sale	purple
Houses for Rent	orange
Italian Restaurant	yellow
Chinese Restaurant	pink

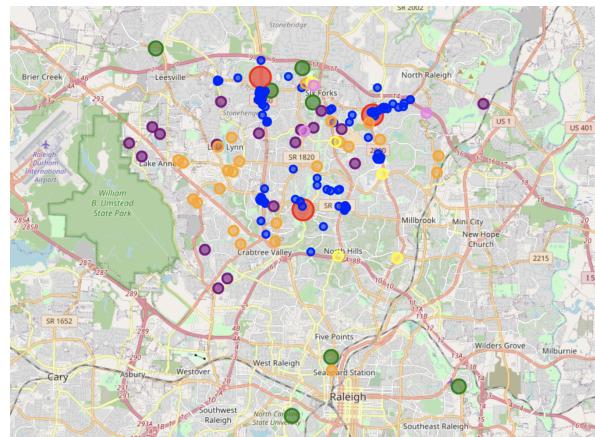


Figure 14:

Figure 15:

	text	agreeCount	disagreeCount		id	user.firstName	user.lastName	user.id
0	Best egg rolls in town!	3	0	524bed26498e0ba58011fee0	Gary	Triplet		299586

Table 12: The tip of Imperial Garden Chinese Restaurant

3.6 Explore a Given Venue

The first and second closest Chinese restaurant - Eastern and King's Wok Chinese Restaurant were explored using FourSquare API, but there was no rating yet. The third closest Chinese restaurant - Imperial Garden Chinese Restaurant was rated 8.3 with 3 count agreed. See table 12.

4 Results

The result of the project has found out that Raleigh city of North Carolina is the optimal city to move to compares to Atlanta of Georgia. In addition, Northwest, Raleigh, NC and North, Raleigh, NC are optimal areas to live in the city. But not all areas in Raleigh, NC are optimal. For example: Central, Raleigh, NC is not a safe neighborhood due to crime rating is bad. The zip code 27612 and 27613 in Northwest and zip 27615 in North all have the popular spots, which are convenient to the residents. The popular spots include the lake trail, parks, pools, supermarkets, fitness centers, shopping mall, bookstores, restaurants, etc. Restaurants include American, Italian, Chinese, Japanese, Mexican, French, and fast-food restaurants. Current houses' sale prices and rental prices are lower than New York City.

5 Discussion

Based on the result, Raleigh, NC is the optimal city. However, when the city's name was used to find the coordinate by geopy.geocoders, it was shown at Central, Raleigh, NC on the map, which is not the optimal area to live in. In addition, when it was reversed to get the zip code, it is 27601, Central, Raleigh, NC. The optimal data from the [Areavibes](#) about Raleigh, NC is the overall rating of the city, which is not the rating for the central. So, some areas of the city are optimal to live in, some of them are not as Central, Raleigh, NC. Another important fact was found that having top rated schools' neighborhoods are not always optimal places to move to, as the example of those top-rated high schools, those neighborhoods aren't. Finding an optimal place is based on many criteria such amenities, crime/safety, cost of living, housing, employment, schools, poverty, and others. Last, the data that scraped from [Zillow](#) were constantly updated every time I ran the script. Therefore, the final list of houses for sale and houses for rent only reflected to the last time I ran the script. It might be different as today's.

6 Conclusion

In this study, only two cities were targeted and only the areas around the top-rated schools of the optimal city were targeted for study. To be better help more people in the future, we should study about more cities. Furthermore, each neighborhood or zip code in the city should be considered for comparison and analysis and find out all the areas in the city that are optimal.