



BATTLE OF NEIGHBORHOODS

Augustin T
Coursera Capstone

Contents

Background	2
Data	2
Methodology	3
Results	3
Discussion	4
Conclusion	5

Background

In my conversations with colleagues and friends about future, job opportunities, environment for raising kids, cost of living, family events and socials, I came to realize there is a very common underlying question that arises in everyone's mind as the move from "single" status to "married" and then to "married with a couple of kids" statuses.

Is there a way to identify the right place for my family to live?

I put together a solution for the above question in the Battle of the Neighborhoods project. My colleagues and friends could also benefit from the solution.

I am very confident that there would be buyers for a custom-built report with all their preferences bells and whistles at reasonable price point.

To make this project as realistic as possible, let create a fictional character Cathy who lives in Hartford CT. She would like to move down south like many others¹. She has narrow down her list to two cities Raleigh NC and Marietta GA but, she is not able to make her decision. Let's put out data science talents to work and help her decide between the two.

Data

- **Foursquare**
Foursquare has a global database of 105 million places with 70+ venue attributes and 900+ categories, which are verified by many users and specialists. It provides us with the ability to perform searches based on location, details about the venues, and customer preferences.
- **Public User Files**
Public user files that provide information about the socio-economic drivers were also be key drivers in determining the appropriate location.
- **Python-Folium**
Folium was used for visualizations as it utilizes the leaflet.js library that has a number of built-in tilesets from OpenStreetMap, Mapbox, and Stamen, and supports custom tile sets with Mapbox or Cloudmade API keys. Folium supports both Image, Video, GeoJSON and TopoJSON overlays.

Data usage in solving the problem

For these kinds of problems, ML algorithms help us in a big way by performing extensive analysis on the Foursquare data points of the two cities in conjunction with the user preferences to uncover the insights and make the decision-making process easier.

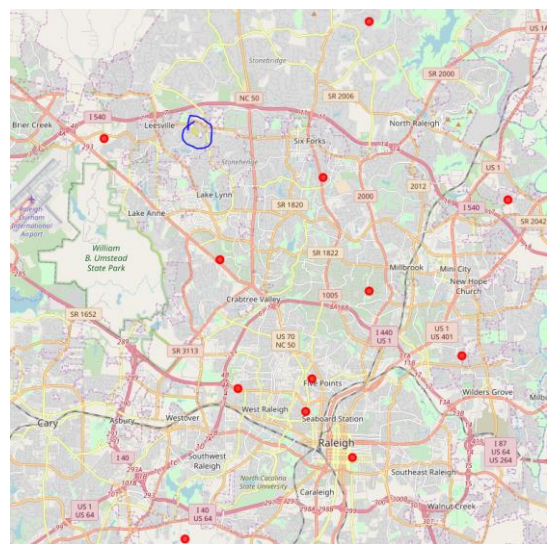
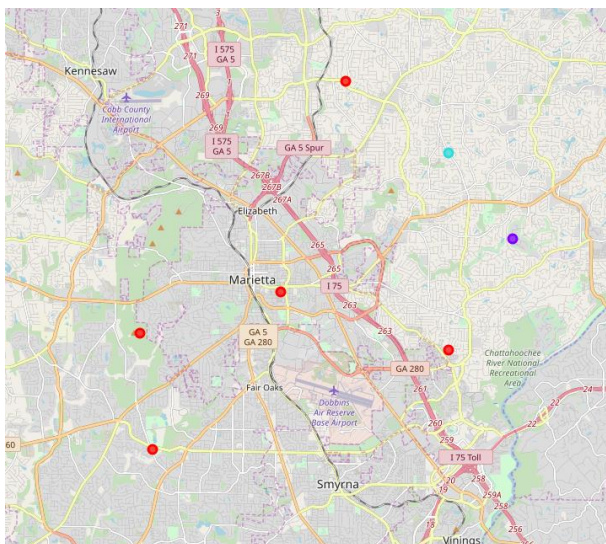
¹ <https://www.move.org/why-americans-are-moving-to-southern-and-western-states/>

K-mean clustering ML algorithm was used to form the clusters of different neighborhoods in and around the cities. These clusters were analyzed further to derive the conclusions.

- Data from different sources were ingested, imputed and normalized to pandas dataframes.
- The goal was to use the FourSquare data and elbow method to derive the number of clusters. However, with the data I had then the elbow method was failing to come up with the number of clusters for this analysis.
- Additional data was added to the then existing data frame to add more context and perspectives. Those additional variables were School ratings, Income levels, Median home value and indicators low income/HPSA neighborhoods.
- With the addition of these variables, I was able to use the elbow method and determine that the four clusters would be appropriate for this analysis.
- Secondary research was also conducted to make reasonable accommodations to personal preferences.

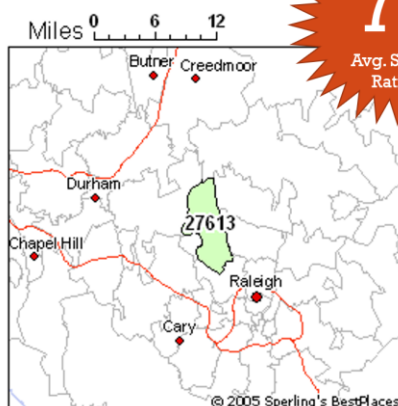
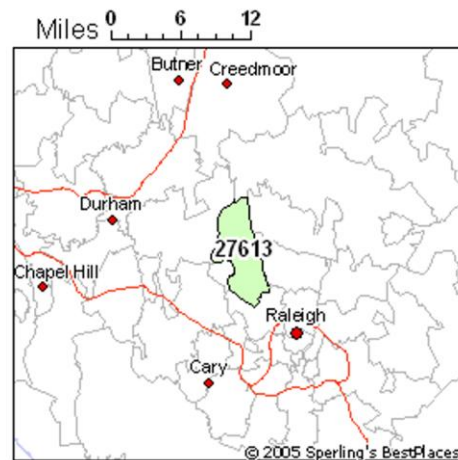
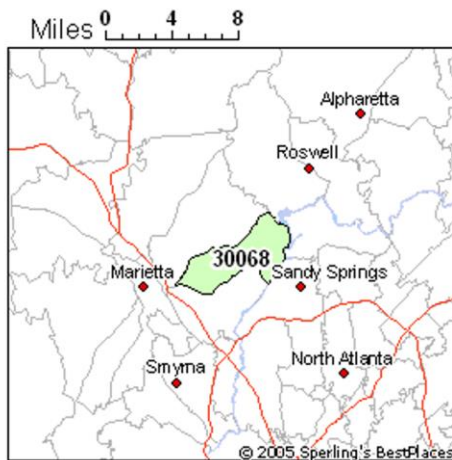
Results

- The results were very interesting. Three clusters had one neighborhood each and the fourth one had all other neighborhoods. The similarities of all the neighborhoods in cluster four was very evident.



Discussion

- Secondary research helped us to further reduce the finalists to two. 30068 in Marietta, GA and 27613 in Raleigh, NC.



75
Avg. School Rating

Option 1:

Neighborhood (27613) in Raleigh, NC

All key factors to consider:

Top 5 Venues:

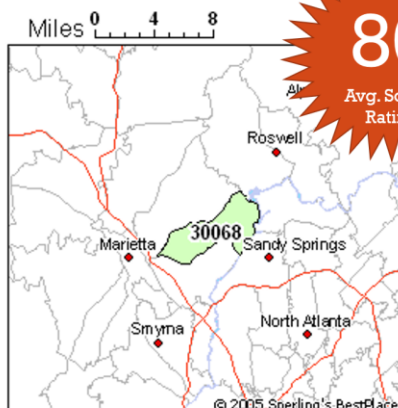
- Deli / Bodega
- Yoga Studio
- Fast Food Restaurant
- Coffee Shop
- Comic Shop



Nearby
Family/Relatives



Appreciating
Housing Market



80
Avg. School Rating

Option 2:

Neighborhood (30068) in Marietta, GA

All key factors to consider:

Top 5 Venues:

- Doctor's Office
- Yoga Studio
- Fast Food Restaurant
- Coffee Shop
- Comic Shop



Nearby
Family/Relatives



Appreciating
Housing Market

Conclusion

- Equipped with this information and taking other personal preferences into consideration, I was able to make a recommendation on the neighborhood that would be appropriate for Cathy.

It was Neighborhood 30068 in Marietta, GA!!

