**Opening a Bar in Atlanta: Finding a Location**

Rohit Pappu

June 30, 2020

1. Introduction/Business Problem

1.1 Background

For the last ten years, Atlanta’s economy, particularly its service sector, has been growing at an incredible pace. A number of Fortune 500 companies have headquarters in Atlanta, with many relocating their headquarters from other states to be in Atlanta. Relatively cheap metropolitan property prices and a business-friendly state government makes Atlanta an alluring location to have a headquarters. Atlanta is considered a leader in banking and finance for the Southeast, and is a top five data center market with an enormous amount of growth momentum. All of this means that there are a huge number of young and middle-aged professionals with relatively stressful jobs and a daily disposable income.

1.2 Problem

Atlanta is a large city, and it has many different areas. For this hypothetical, we are considering opening a bar for young to middle-aged professionals with disposable incomes. One of the main questions that we will be focusing on is where to open this bar. Our project aims to find four potential locations to open this bar.

1.3 Interest

Anyone who is considering opening a bar would be interested in this project. The food service industry usually operates on thin margins, so any additional advantage would be welcome.

2. Data acquisition and cleaning

2.1 Data sources

For this project, we used two sources for data. The first source was from the Atlanta Regional Commission and it contained a list of neighborhoods in Atlanta as well as the geographical coordinates and demographic information of each neighborhood. This can be found [here](https://opendata.atlantaregional.com/datasets/d6298dee8938464294d3f49d473bcf15_196?geometry=-85.294%2C33.567%2C-83.547%2C33.967). We focus on the features aspect of the json file, as this is where all the pertinent information can be found. The second source came from using Foursquare data to compile lists of banks, law firms, and colleges in the Atlanta area.

2.2 Data cleaning

For the data that was acquired from the Atlanta Regional Commission, I decided to create a table that contained all the pertinent information. One problem that I ran into was that the file contained a list of multiple geographical coordinates for each neighborhood. This would have made creating a table particularly difficult. For simplicity’s sake, I decided to take the first set of geographical coordinates from the list as the geographical coordinates of the area.

For the data that was acquired from Foursquare, there were no real problems in creating tables from the Foursquare data. For the Bank data table, I decided to exclude ATMs, as the presence of an ATM does not reflect the presence of a flourishing commercial center, and might skew our results.

2.3 Feature Selection

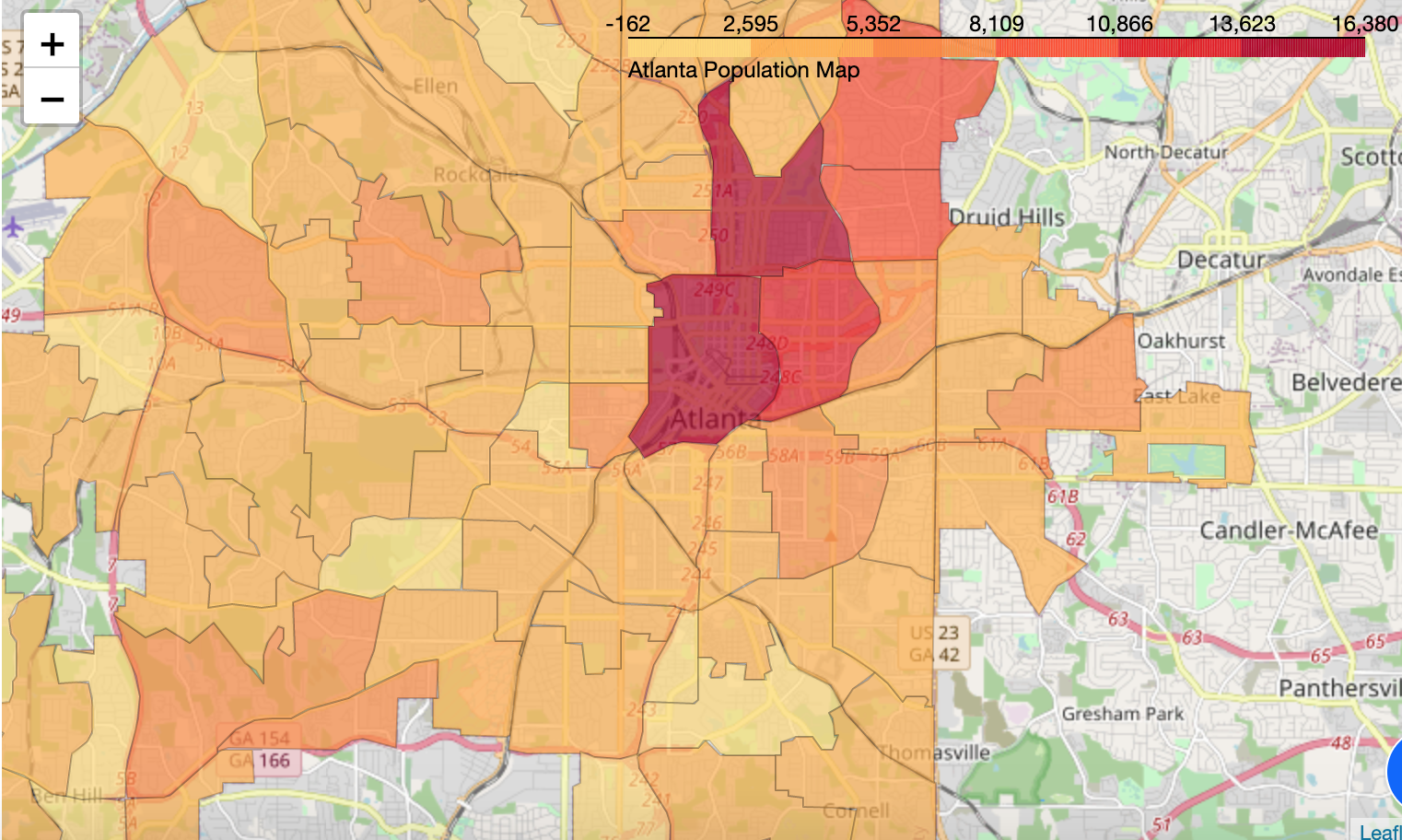
For the data from the Atlanta Regional Commission, I decided to focus on the names of neighborhoods, their population, and their geographical coordinates. I did include a link to the census data in the prepared table, in case there was need for further analysis, however, for the purposes of this project, I will not be needing the census data.

For the data from Foursquare, I focused on the location name, location type, and the geographical coordinates of the neighborhood.

3. Machine Learning Techniques

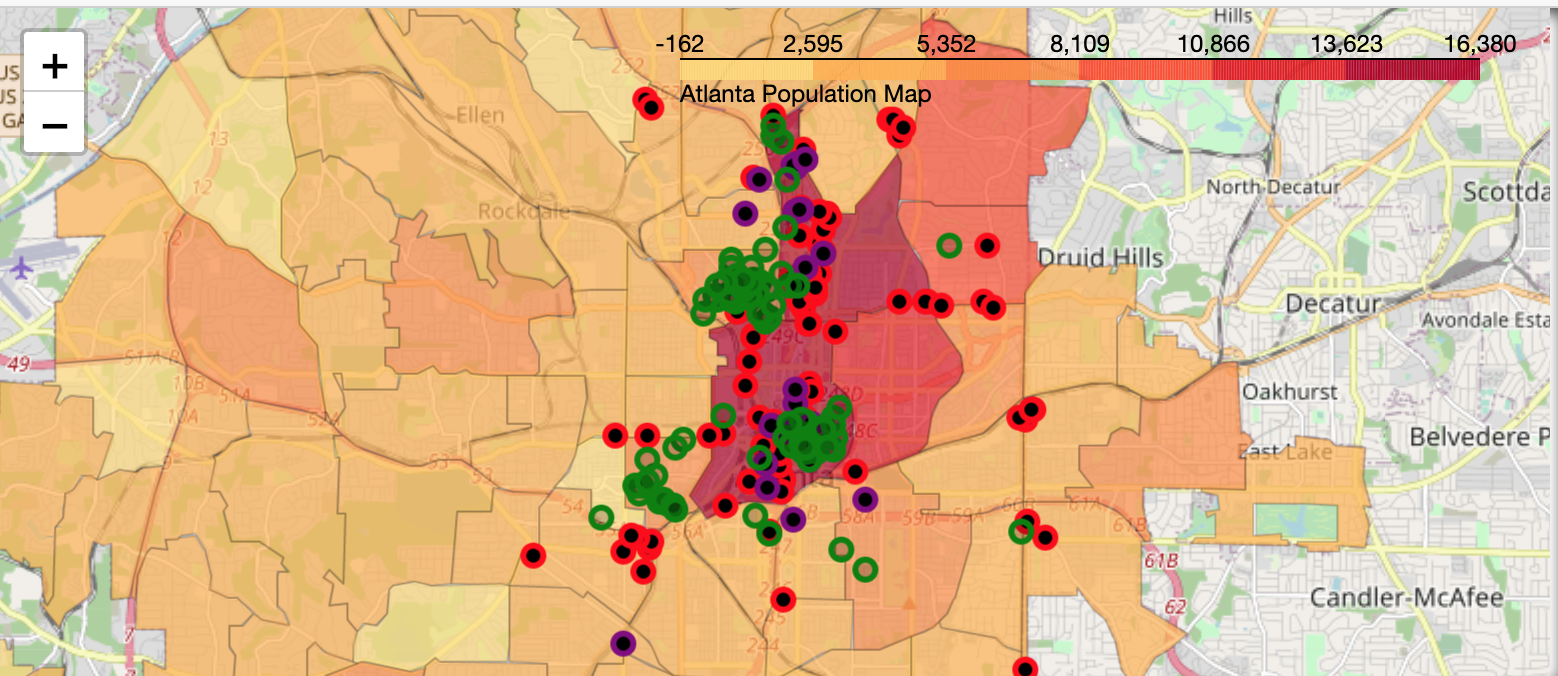
3.1 Choropleth Map

Using the data from the Atlanta Regional Commission, I created a choropleth map that highlighted the location of neighborhoods in Atlanta. The regions are shaded according to population, with darker colors corresponding with a larger population. This is so that a viewer can easily determine the location of neighborhoods in Atlanta, as well as their relative population size. In the case that there are multiple appealing locations for a bar, the population size of each neighborhood can be used to determine the best location.



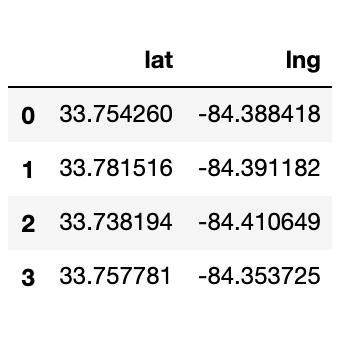
3.2 Folium Circle Marker Map

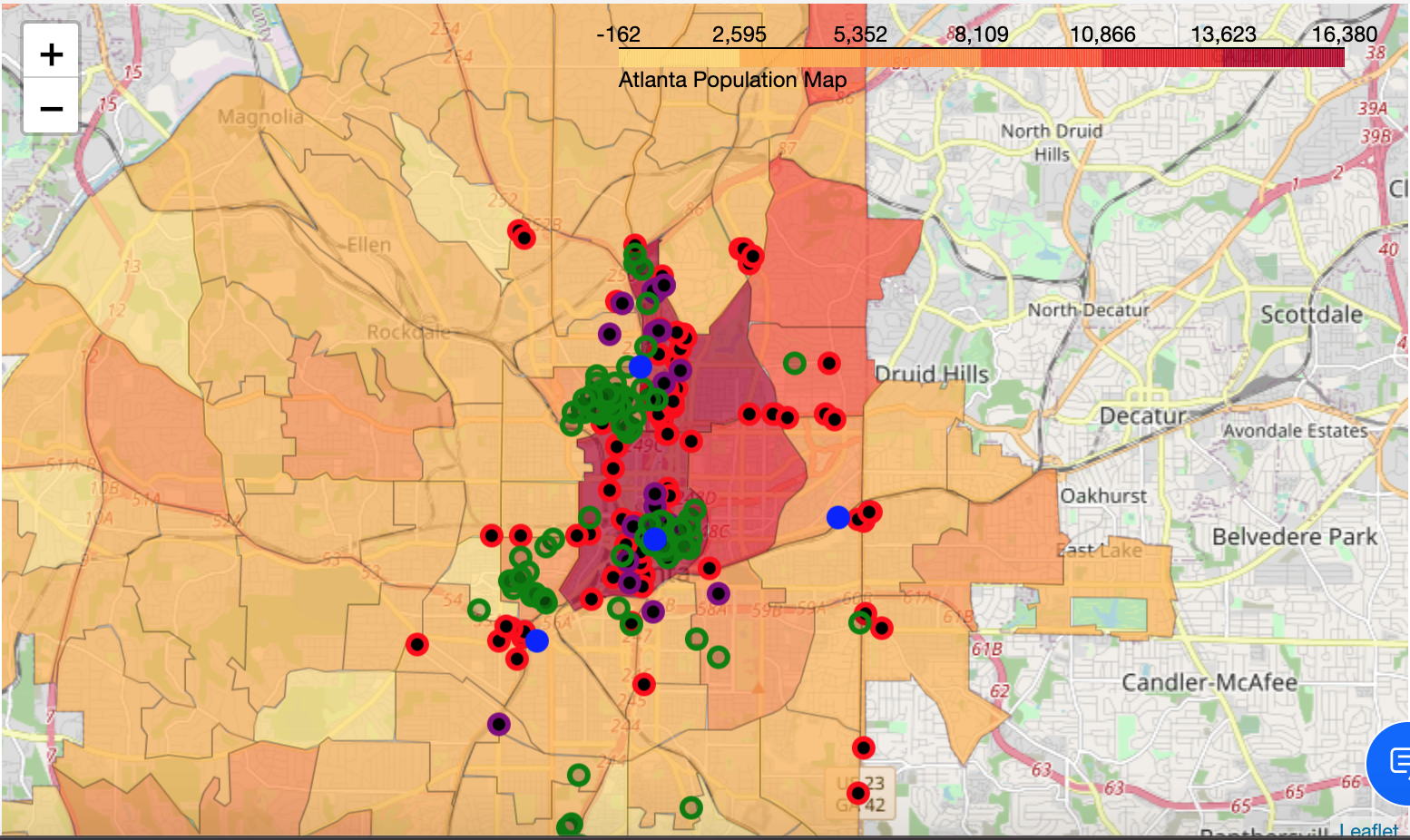
I used the data from Foursquare regarding the location of banks, law firms, and colleges to generate a Folium map consisting of circle markers. I generated circle markers with different colors corresponding to different types of businesses. In my map, red corresponds to banks, purple corresponds to law firms, and green corresponds to colleges. I created the different colors so that a potential bar owner can have a better idea of the type of clientele to expect. For example, if the bar were to be located in an area with a high density of college buildings, the owner might entice students to frequent their bar with discounts.



3.3 K-Means Clustering

I used the geographical coordinates from the Foursquare data and k-means clustering to find three ideal locations for a bar. I did not want to use the elbow method for determining the best k-number, as this might have given me a number of locations, each with a limited customer base. I looked at the distribution of points on the Folium map, and saw that there were two clusters of points. I decided to set the k-number as three as a precaution, in case there was a cluster that was not noticeable to the eye. Using k-means clustering, I was able to obtain three sets of geographical coordinates that correspond to the centers of three clusters. I then plotted these points on the folium map using blue markers.





4. Results

Using k-means clustering, we find that the four optimal locations are located in Downtown, Midtown, and Pittsburg. Using the choropleth map, we find that both Downtown and Midtown are roughly equal in population, so opening a bar at either of these locations would be optimal.

5. Conclusion

Looking at the final map, I would conclude that the best place to open a new bar would be at either the downtown location or the midtown location

[Link to Jupyter Notebook](https://github.com/rohitp1020/github-example/blob/master/Weeks%201%20and%202%20Final%20Project.ipynb)