# Choosing the next train stations in Atlanta

Yung Suk Jeremy Yoo

## 1. Introduction

The public transportation system in the city of Atlanta is Metropolitan Atlanta Rapid Transit Authority (MARTA).

Due to the influx of population into Atlanta and to curb the traffic congestion in the city, MARTA has decided to expand its subway system to service the areas in Fulton county that do not already have any MARTA lines running through them. In other words, it wants to build a new line that connects with one of the existing lines. Due to its limited budget, however, MARTA can only build three new stations during the first phase of the expansion project.

The challenge is to pick three neighborhoods within a reasonable proximity to each other to create a new subway line. To ensure maximum serviceability, MARTA wants to make sure that the neighborhoods already have high foot traffic.

## 2. Data

The data was retrieved from multiple sources to analyze the foot traffic of each neighborhood in Fulton county.

First, a GeoJSON file is needed to map the neighborhoods. A GeoJSON file from "Zillows - US Neighborhoods" from Opendatasoft database was used. The file is too large, so it will be filtered with Georgia state and Fulton county before downloading. A preliminary filtering will be employed by identifying the neighborhoods that already have a MARTA station. The list of MARTA stations was collected from a Wikipedia page.

Second, populations of neighborhoods are an excellent baseline indicator of foot traffic. This data was collected from Wikipedia. The table contains neighborhoods that may not be in the GeoJSON file, and those will be disregarded for sake of simplicity.

Third, each neighborhood's foot traffic was measured by the economic activity in the neighborhood. The restaurant data be collected from Foursquare was used to measure economic activity.

Having a higher population density does not guarantee a higher use of subway systems if the area is residential. For simplicity, number of restaurants will be used as a metric for economic activity.

These data were used to generate choropleth maps that show neighborhoods with a custom foot traffic index to derive the best possible locations for new MARTA stations.

## 3. Methodology

The raw data of the borders of the Fulton county were imported as a GeoJSON file. A table of the population of each neighborhood was scraped from a Wikipedia page using Beautiful Soup. An initial data frame containing the neighborhoods, the latitude and longitude of their geographical centers, and populations was created. Because of the discrepancy in the number of neighborhoods, only the neighborhoods that were on the Wikipedia page and the GeoJSON file were considered.

Before calculating the economic activity of each neighborhood, Foursquare was used to identify the neighborhoods already serviced by MARTA subway. A list of MARTA stations was generated from the Wikipedia page. A general search queries of 'MARTA stations' and 'Metro stations' did not yield a comprehensive result. Therefore, the individual names of each station were used in a for loop to generate a list of results that best match the description of the subway stations. The results were filtered once again to accurately locate the subway stations.

Shapely library was used to calculate how many stations are located within the boundary of each neighborhood. These neighborhoods were excluded from further consideration.

The polygon shapes of the neighborhood boundaries in GeoJSON also allowed for calculation of the neighborhoods' land area. The area allowed for normalization of the population: i.e. population density. The coordinates in GeoJSON were in degrees, but a rough estimate of 111 km/° conversion factor was used to convert the area into meaningful numbers.

Foursquare was used again to estimate the number of restaurants within a *certain distance* of the geographical center of each neighborhood. Assuming that the neighborhoods were circular, the *certain distance* was calculated by solving for the radius. This assumption was thought to be valid since it is a good approximation of the average distance between the borders of the neighborhood to the geographical center. The Foursquare results were counted and tabulated onto a data frame.

The resulting data frames were merged into a single data frame. The population density and the number of restaurants were normalized using the min-max method. The presence of pre-existing MARTA station was considered as well. These three variables were used to generate a *Foot Traffic Index* (FTI) for each neighborhood. The FTI is a rough estimate of the number of people who will be going to each neighborhood, thus a higher FTI indicates a better chance for MARTA to service more people. It is calculated as:

$$FTI = S(0.6 \times P + 0.4 \times R)$$

where $S$ is the station factor, $P$ is the normalized population density, and $R$ is the normalized number of restaurants. S is set to equal zero if there is already a MARTA station in the neighborhood and one otherwise.

An FTI heat map of the Fulton county was generated using folium, along with markers indicating the locations of pre-existing MARTA stations.

## 4. Results

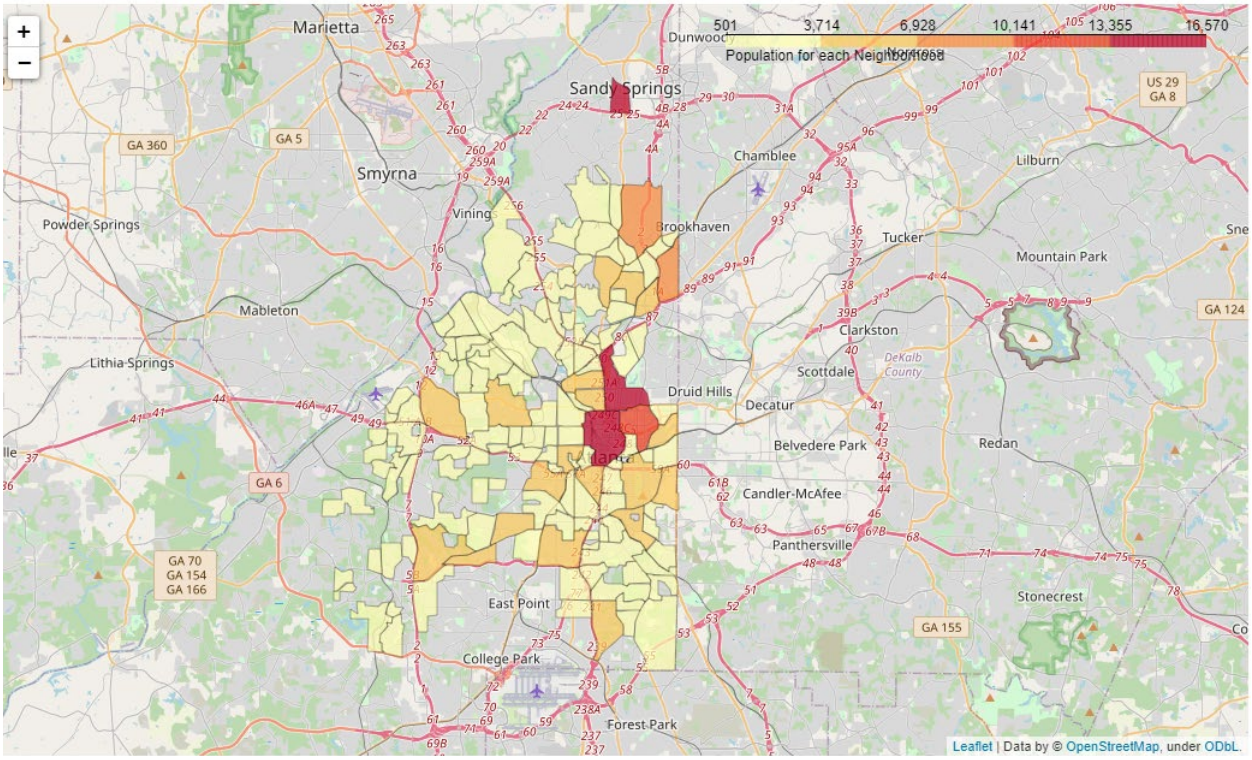


**Figure 1** A heat map showing the population distribution of neighborhoods in Fulton county.

| | Neighborhood | Latitude | Longitude | Population | Area | Pop Den | Stations | Restaurants |
|---|---|---|---|---|---|---|---|---|
| 0 | Adair Park | 33.731149 | -84.410837 | 1331 | 1.401678 | 949.576 | 0 | 4 |
| 1 | Adams Park | 33.713939 | -84.461696 | 1763 | 3.042064 | 579.541 | 0 | 1 |
| 2 | Adamsville | 33.758775 | -84.503071 | 2403 | 3.175911 | 756.633 | 0 | 4 |
| 3 | Almond Park | 33.786533 | -84.462302 | 1020 | 1.627224 | 626.835 | 0 | 1 |
| 4 | Ansley Park | 33.795867 | -84.378577 | 2277 | 1.901030 | 1197.77 | 0 | 1 |
| 5 | Ardmore | 33.805401 | -84.397237 | 756 | 0.400573 | 1887.29 | 0 | 1 |
| 6 | Argonne Forest | 33.841499 | -84.403683 | 590 | 0.833998 | 707.436 | 0 | 0 |
| 7 | Arlington Estates | 33.691336 | -84.540510 | 776 | 1.010329 | 768.067 | 0 | 0 |
| 8 | Ashview Heights | 33.750128 | -84.421897 | 1292 | 0.854518 | 1511.96 | 0 | 0 |
| 9 | Atlanta University Center | 33.749455 | -84.411246 | 5703 | 1.611516 | 3538.9 | 0 | 1 |
| 10 | Atlantic Station | 33.791926 | -84.398118 | 1888 | 0.793080 | 2380.59 | 0 | 5 |
| 11 | Baker Hills | 33.759074 | -84.520024 | 757 | 0.946933 | 799.423 | 0 | 0 |
| 12 | Bankhead | 33.769017 | -84.424535 | 1929 | 2.029152 | 950.644 | 0 | 6 |
| 13 | Beecher Hills | 33.732956 | -84.461730 | 647 | 1.378503 | 469.35 | 0 | 0 |
| 14 | Ben Hill | 33.670380 | -84.522497 | 1725 | 3.334757 | 517.279 | 0 | 1 |
| 15 | Ben Hill Terrace | 33.699987 | -84.521760 | 937 | 1.032244 | 907.731 | 0 | 0 |
| 16 | Benteen Park | 33.715860 | -84.364130 | 893 | 0.868971 | 1027.65 | 0 | 1 |
| 17 | Berkeley Park | 33.796783 | -84.411892 | 1400 | 1.463289 | 956.749 | 0 | 6 |
| 18 | Betmar LaVilla | 33.706664 | -84.398076 | 578 | 0.356112 | 1623.08 | 0 | 1 |
| 19 | Bolton | 33.816551 | -84.456723 | 2996 | 4.649929 | 644.311 | 0 | 3 |
| 20 | Boulevard Heights | 33.723989 | -84.363869 | 675 | 0.678646 | 994.627 | 0 | 2 |

**Figure 2** A data frame showing the first 20 rows of the result of data collection and wrangling. Population, area ($km^2$), population density ($km^{-2}$), number of stations and restaurants are shown.

| | Neighborhood | Latitude | Longitude | Pop Den Norm | Station Filter | Rest Norm | Foot Traffic Index |
|---|---|---|---|---|---|---|---|
| **0** | Adair Park | 33.731149 | -84.410837 | 0.102752 | 1 | 0.08 | 0.0936513 |
| **1** | Adams Park | 33.713939 | -84.461696 | 0.0518948 | 1 | 0.02 | 0.0391369 |
| **2** | Adamsville | 33.758775 | -84.503071 | 0.0762342 | 1 | 0.08 | 0.0777405 |
| **3** | Almond Park | 33.786533 | -84.462302 | 0.0583948 | 1 | 0.02 | 0.0430369 |
| **4** | Ansley Park | 33.795867 | -84.378577 | 0.136864 | 1 | 0.02 | 0.0901184 |
| **5** | Ardmore | 33.805401 | -84.397237 | 0.231632 | 1 | 0.02 | 0.146979 |
| **6** | Argonne Forest | 33.841499 | -84.403683 | 0.0694725 | 1 | 0 | 0.0416835 |
| **7** | Arlington Estates | 33.691336 | -84.540510 | 0.0778056 | 1 | 0 | 0.0466834 |
| **8** | Ashview Heights | 33.750128 | -84.421897 | 0.180046 | 1 | 0 | 0.108028 |
| **9** | Atlanta University Center | 33.749455 | -84.411246 | 0.458628 | 1 | 0.02 | 0.283177 |
| **10** | Atlantic Station | 33.791926 | -84.398118 | 0.29943 | 1 | 0.1 | 0.219658 |
| **11** | Baker Hills | 33.759074 | -84.520024 | 0.0821153 | 1 | 0 | 0.0492692 |
| **12** | Bankhead | 33.769017 | -84.424535 | 0.102899 | 1 | 0.12 | 0.109739 |
| **13** | Beecher Hills | 33.732956 | -84.461730 | 0.0367502 | 1 | 0 | 0.0220501 |
| **14** | Ben Hill | 33.670380 | -84.522497 | 0.0433376 | 1 | 0.02 | 0.0340025 |
| **15** | Ben Hill Terrace | 33.699987 | -84.521760 | 0.097001 | 1 | 0 | 0.0582006 |
| **16** | Benteen Park | 33.715860 | -84.364130 | 0.113483 | 1 | 0.02 | 0.0760898 |
| **17** | Berkeley Park | 33.796783 | -84.411892 | 0.103738 | 1 | 0.12 | 0.110243 |
| **18** | Betmar LaVilla | 33.706664 | -84.398076 | 0.195319 | 1 | 0.02 | 0.125191 |
| **19** | Bolton | 33.816551 | -84.456723 | 0.0607967 | 1 | 0.06 | 0.060478 |
| **20** | Boulevard Heights | 33.723989 | -84.363869 | 0.108944 | 1 | 0.04 | 0.0813664 |

**Figure 3** The final result data frame showing the normalized population density and number of restaurants, along with the station factor. The FTI, indicator of the foot traffic in each neighborhood, is shown on the rightmost column.

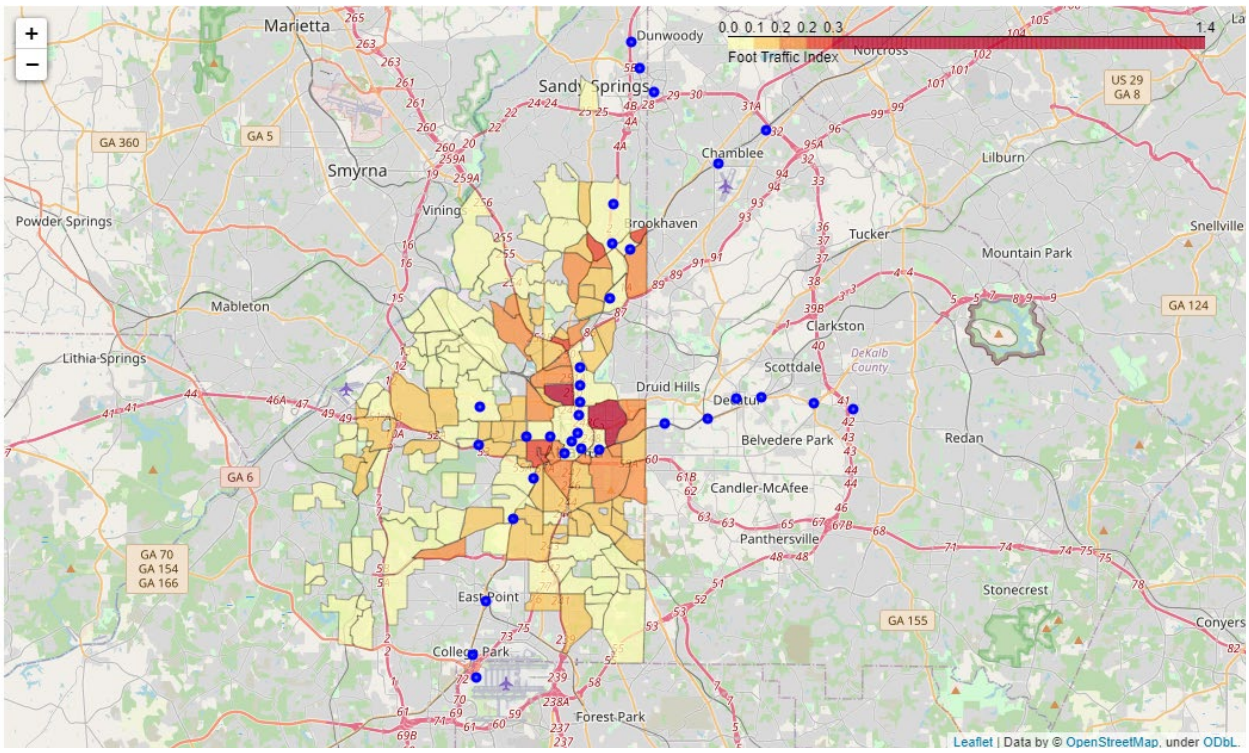


**Figure 4** A heatmap of the FTI overlaid with blue markers showing the locations of MARTA stations. There are two red neighborhoods at the center of the city with the color indicating FTI higher than 1. This is an error that I could not resolve. Therefore, the darker orange will be considered as the “true” high FTI.

## 5. Discussion

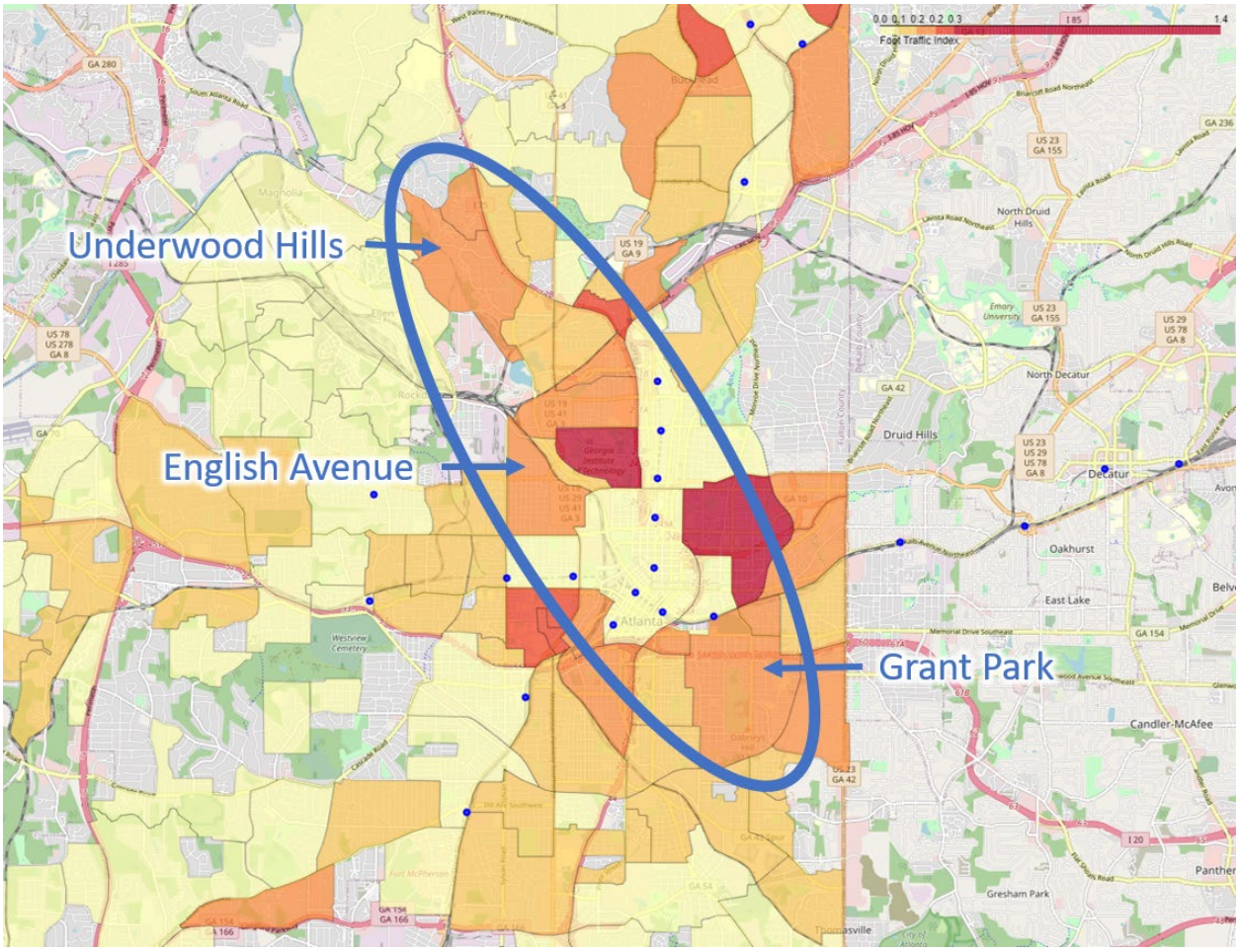


**Figure 5** A magnified choropleth map from **Figure 4** indicating three candidates with the best potential for reaching new MARTA users based on FTI: Underwood Hills, English Avenue, and Grant Park.

Based on FTI, the three best candidate locations for new MARTA stations are Underwood Hills, English Avenue, and Grant Park. There are other neighborhoods with higher FTI, but they usually are adjacent to a neighborhood with a MARTA station. MARTA lacks a line that goes from northwest to southeast, and based on the FTI, MARTA may see the best increase in users if it expands in these neighborhoods.

## 6. Conclusion

Data collected from multiple sources were used to calculate the population density, number of restaurant numbers, and presence of MARTA stations. Number of restaurants was used as a simple metric for economic activity of each neighborhood. A customized index for scoring the foot traffic was used to map the potential serviceability of MARTA subway to new customers. Three likely

candidates (Underwood Hills, English Avenue, and Grant Park) were chosen based on the normalized population density, normalized number of restaurants, and proximity to MARTA stations.

## 7. References


1. Opendatasoft - https://data.opendatasoft.com/explore/dataset/zillow-neighborhoods%40public/export/?refine.state=GA&refine.county=Fulton&dataChart=eyJxdWVyaWVzIjpbeyJjb25maWciOnsiZGF0YXNldCI6InppbGxvdy1uZWlnaGJvcmhvb2RzQHB1YmxpYyIsIm9wdGlvbnMiOnsibG9jYXRpb24iOiIxMiwzMy43NTQzMiwtODQuNDU3OTEiLCJiYXNlbWFwIjoiamF3Zy5zdW5ueSJ9fSwiY2hhcnRzIjpbeyJhbGlnbk1vbnRoIjp0cnVlLCJ0eXBlIjoiY29sdW1uIiwiZnVuYyI6IkNPVU5UIiwic2NpZW50aWZpY0Rpc3BsYXkiOnRydWUsImNvbG9yIjoiIzE0MkU3QiJ9XSwieEF4aXMiOiJzdGF0ZSIsIm1heHBvaW50cyI6NTAsInNvcnQiOiIifV0sInRpbWVzY2FsZSI6IiIsImRpc3BsYXlMZWdlbmQiOnRydWUsImFsaWduTW9udGgiOnRydWV9&location=13,33.79477,-84.35071&basemap=jawg.sunny
2. Wikipedia - https://en.wikipedia.org/wiki/MARTA_rail
3. Wikipedia - https://en.wikipedia.org/wiki/Table_of_Atlanta_neighborhoods_by_population


Jupyter Notebook published on Github:

https://github.com/jeremyyoo/Coursera9_Capstone_Project/blob/master/Calculate_Foot_Traffic_Fulton_County.ipynb