Segmenting and Labeling
MARTA Rail Stations in Atlanta

Nelli Fedorova[1]

July 17, 2019

Abstract

In this report, I explore and visualize data for venues around MARTA rail stations in Atlanta. First, I use Foursquare to get popular venues near each station. Next, I use Foursquare categories hierarchy to assign each venue to a general category, thus significantly reducing data dimensionality. Then I explore primary and general venue categories for each station, and use TF-IDF normalization and LDA algorithm to generate station labels. I also use K-means, DBSCAN and hierarchical clustering algorithms to group stations into category-related clusters. Finally, I visualize all my results in interactive Folium maps.

Introduction

It this report, I am going to explore venues associated with each of 38 MARTA rail stations in Atlanta, Georgia.

This might help us answer the following questions.

- Why people use MARTA?
- What attracts them to each MARTA station and makes them go there?
- Where do they go before/after riding MARTA?
- Where do they spend the most time (and money)?
- Are MARTA stations similar or dissimilar in respect to the venues near them?
- Can we label each MARTA station with venue categories that are specific for this station?

MARTA station names and their coordinates can be easily extracted from Wikipedia. In my analysis, I will use Foursquare API to get the most popular venues near each MARTA station, Foursquare categories hierarchy to assign each venue to a general category, TF-IDF normalization and LDA algorithm to label each station, K-means, DBSCAN and hierarchical clustering algorithms to cluster the stations, and Folium library to visualize the results.

1. MARTA Stations Data

Atlanta MARTA rail network has a total of 38 stations. (Figure 1). 22 stations provide free daily parking for MARTA passengers, 9 of them have long-term parking facilities as well.

To segment, explore and label MARTA stations, I essentially needed a list of MARTA stations with the latitude and longitude coordinates of each station. Luckily, this data exists for free on Wikipedia It can be easily downloaded and converted to Excel file (see first 6 columns in Table 1).

I used station latitude and longitude in the Foursquare queries to get the top 100 most popular venues and their primary categories within a radius of 500 meters from each station. I obtained a list of 1156 venues from 223 different categories. Last column of Table 2 shows number of venues returned by Foursquare for each station.

It should be noted that Foursquare primary categories are rather specific. For example, “Airport Service” and “Airport Lounge” are two separate categories. To reduce data dimensionality and facilitate data analysis I used

[1] LinkedIn https://www.linkedin.com/in/nelli-fedorova-7710b01a/

Foursquare categories hierarchy to assign each primary category to a more general (root parent) category. For example, I assigned "Japanese Restaurant" to "Food" category (Table 2).

Table 1. MARTA rail stations Wikipedia data and Foursquare query results.
Parking: 0 – no parking, 1 – free daily, 2 – free daily & long-term

	Station	Entries/Day	Latitude	Longitude	Parking	Venues
0	Airport	9173	33.640758	-84.446341	0	27
1	Arts Center	6605	33.789705	-84.387789	0	48
2	Ashby	1791	33.756346	-84.417556	1	18
3	Avondale	4327	33.775277	-84.281903	1	21
4	Bankhead	1903	33.771890	-84.428840	0	3
5	Brookhaven	2357	33.860705	-84.340003	2	19
6	Buckhead	2643	33.848410	-84.367018	0	100
7	Chamblee	3785	33.886191	-84.306957	1	17
8	Civic Center	2692	33.766305	-84.387209	0	33
9	College Park	9026	33.651673	-84.448793	2	12
10	Decatur	3821	33.774717	-84.295588	0	61
11	Dome	2107	33.756293	-84.397759	0	58
12	Doraville	5521	33.902079	-84.280389	2	15
13	Dunwoody	3545	33.921200	-84.344400	2	61
14	East Lake	1241	33.765166	-84.312665	1	16
15	East Point	4571	33.677814	-84.440344	1	16
16	Edgewood	1143	33.762001	-84.339579	1	10
17	Five Points	19447	33.753826	-84.391571	0	33
18	Garnett	1516	33.747845	-84.396415	0	30
19	Georgia State	4055	33.750539	-84.386464	0	33
20	H. E. Holmes	6480	33.754638	-84.467940	1	15
21	Indian Creek	5612	33.769794	-84.229656	1	3
22	Inman Park	2525	33.757497	-84.352797	1	40
23	Kensington	5950	33.772664	-84.251937	2	4
24	King Memorial	1517	33.749959	-84.375440	0	44
25	Lakewood	2207	33.700457	-84.428859	1	11
26	Lenox	3284	33.847144	-84.356301	2	25
27	Lindbergh Center	8604	33.821995	-84.367447	2	46
28	Medical Center	1629	33.910689	-84.352684	1	9
29	Midtown	5664	33.781121	-84.386345	0	100
30	North Avenue	5045	33.771712	-84.386699	0	42
31	North Springs	6436	33.944552	-84.356206	2	11
32	Oakland City	4432	33.716848	-84.425200	1	10
33	Peachtree Center	7453	33.759677	-84.387548	0	100
34	Sandy Springs	2322	33.933035	-84.352019	2	23
35	Vine City	821	33.756870	-84.403910	0	16
36	West End	7056	33.735810	-84.412960	1	20
37	West Lake	1378	33.753140	-84.446580	1	6

Figure 1. Map of MARTA Rail System: 38 stations

Table 2. Assigning primary categories to Foursquare categories hierarchy **root nodes**

Food → Asian Restaurant
Food → *Asian Restaurant* → Japanese Restaurant
Residence → Residential Building (Apartment / Condo)
Shop & Service → *Clothing Store* → Women's Store
Shop & Service → Pharmacy
Shop & Service → Gas Station
Nightlife Spot → Nightclub
Nightlife Spot → *Bar* → Whisky Bar
Travel & Transport → *Airport* → Airport Lounge
Travel & Transport→ Light Rail Station
College & University → *College Stadium* → College Basketball Court
Arts & Entertainment → Aquarium
Arts & Entertainment → *Museum* → Art Museum
Outdoors & Recreation → Park
Outdoors & Recreation → *Athletics & Sports* → *Gym / Fitness Center* → Yoga Studio
Professional & Other Places → Event Space
Professional & Other Places → *Medical Center* → Chiropractor

2 MARTA Stations Segmenting

I used **One Hot encoding** to transform primary and general categories for each venue to binary vectors. After that I grouped these vectors by station and computed a sum vector representing categories occurrences for each station. Categories counts reflected categories popularity at each station.

Table 3 shows general categories occurrence vectors for each of 38 stations. Figure 3 provides another view of this data. Obviously, Food and Shop & Service categories dominate areas around all MARTA stations in Atlanta. The same can be revealed from the word cloud chart of primary venue categories (Figure 4).

According to Table 3, there are 13 stations with Professional & Other Places. Georgia State is the only station with College & University venues, and there are only 3 stations with Residence options in Atlanta: Civic Center, Lindbergh and Inman Park/Reynoldstown.

Figure 5 shows **Entries/Day vs Number of Venues** for 16 stations with no parking (blue), 13 stations with free daily parking (yellow) and 9 with free daily & long-term parking (orange). Stations with Professional Services are shown in bold. As we can see from this plot, low business activity around MARTA stations is not correlated with the number of Entries per Day, but it is very strongly **correlated with free parking**. Box plot in Figure 6 illustrates the difference between stations with free parking and with no parking. Evidently, MARTA stations with free parking are mostly used by park-and-ride commuters.

A simple rule "number of venues ≤ 26" surprisingly well separates stations with free parking from those with no parking. There are only 5 exceptions. Dunwoody, Lindbergh Center, Inman Park have parking, but are surrounded with many businesses. Bankhead and Vine City are very small stations with no parking, small number of Entries per Day and practically no venues nearby.

Interestingly, Five Points, which is a transit point between all four MARTA Rail lines, has the highest number of Entries per Day of all MARTA stations, but only 33 venues.

Buckhead, Midtown, Peachtree Center, King Memorial, Dome, Dunwoody, Decatur, North Avenue, Arts Center, Lindbergh Center seem to be located in highly dense areas with many businesses.

Kensington, Indian Creek, College Park and North Springs have many Entries/Day, but a very small number of venues. For example, Kensington has 5950 Entries/Day and only 4 venues: Gas Station, Discount Store, Pharmacy, Chiropractor. One of 4 Kensington's venues, Utley Chiropractic (Utley Chiropractic & Wellness Center, to be precise) happens to fall into Professional & Other Places category.

Table 3. General categories occurrence for 38 stations

	Station	Arts & Entertainment	College & University	Food	Nightlife Spot	Outdoors & Recreation	Professional & Other Places	Residence	Shop & Service	Travel & Transport
0	Airport	2	0	12	0	0	0	0	3	10
1	Arts Center	13	0	21	4	2	0	0	5	3
2	Ashby	1	0	10	1	2	0	0	2	2
3	Avondale	1	0	3	2	6	1	0	7	1
4	Bankhead	0	0	1	0	1	0	0	1	0
5	Brookhaven	0	0	7	2	0	0	0	10	0
6	Buckhead	0	0	34	4	5	2	0	48	7
7	Chamblee	0	0	7	0	1	0	0	9	0
8	Civic Center	1	0	15	4	5	0	1	1	6
9	College Park	1	0	7	0	0	0	0	2	2
10	Decatur	1	0	38	3	4	1	0	13	1
11	Dome	8	0	31	5	6	1	0	5	2
12	Doraville	0	0	8	1	1	0	0	4	1
13	Dunwoody	0	0	14	1	2	0	0	39	5
14	East Lake	0	0	8	0	2	0	0	3	3
15	East Point	1	0	13	0	0	0	0	2	0
16	Edgewood	1	0	2	0	5	0	0	1	1
17	Five Points	4	0	22	3	0	1	0	2	1
18	Garnett	6	0	12	4	1	0	0	4	3
19	Georgia State	2	1	22	0	2	1	0	4	1
20	H. E. Holmes	0	0	4	0	0	0	0	8	3
21	Indian Creek	0	0	0	0	1	0	0	0	2
22	Inman Park	1	0	12	0	2	1	1	21	2
23	Kensington	0	0	0	0	0	1	0	3	0
24	King Memorial	4	0	22	4	9	0	0	3	2
25	Lakewood	1	0	3	0	1	0	0	1	5
26	Lenox	0	0	12	3	1	1	0	4	4
27	Lindbergh Center	0	0	18	6	3	0	1	18	0
28	Medical Center	0	0	4	0	0	0	0	5	0
29	Midtown	3	0	60	6	7	1	0	17	6
30	North Avenue	6	0	25	4	1	0	0	3	3
31	North Springs	0	0	1	0	2	0	0	2	6
32	Oakland City	2	0	4	0	1	0	0	2	1
33	Peachtree Center	5	0	56	14	4	2	0	5	14
34	Sandy Springs	2	0	11	1	1	0	0	8	0
35	Vine City	2	0	9	0	4	0	0	1	0
36	West End	1	0	7	0	0	0	0	11	1
37	West Lake	0	0	2	0	0	0	0	2	2

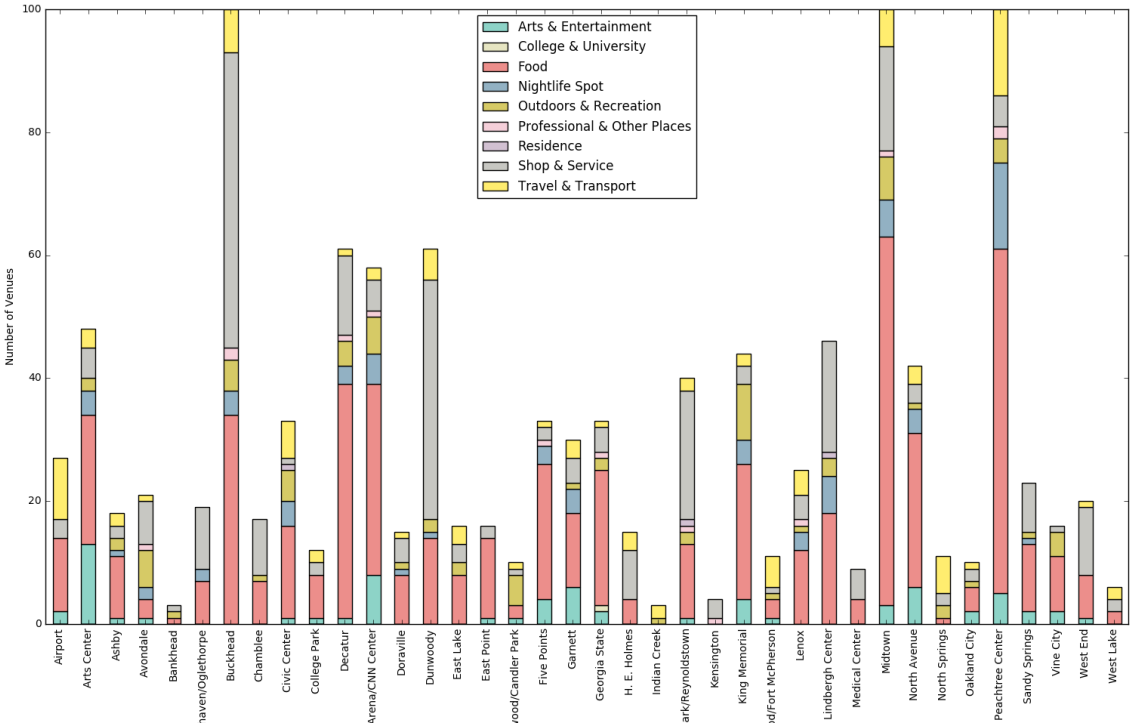

Figure 3. General categories occurrence for 38 MARTA stations

Figure 4. MARTA station venues word cloud

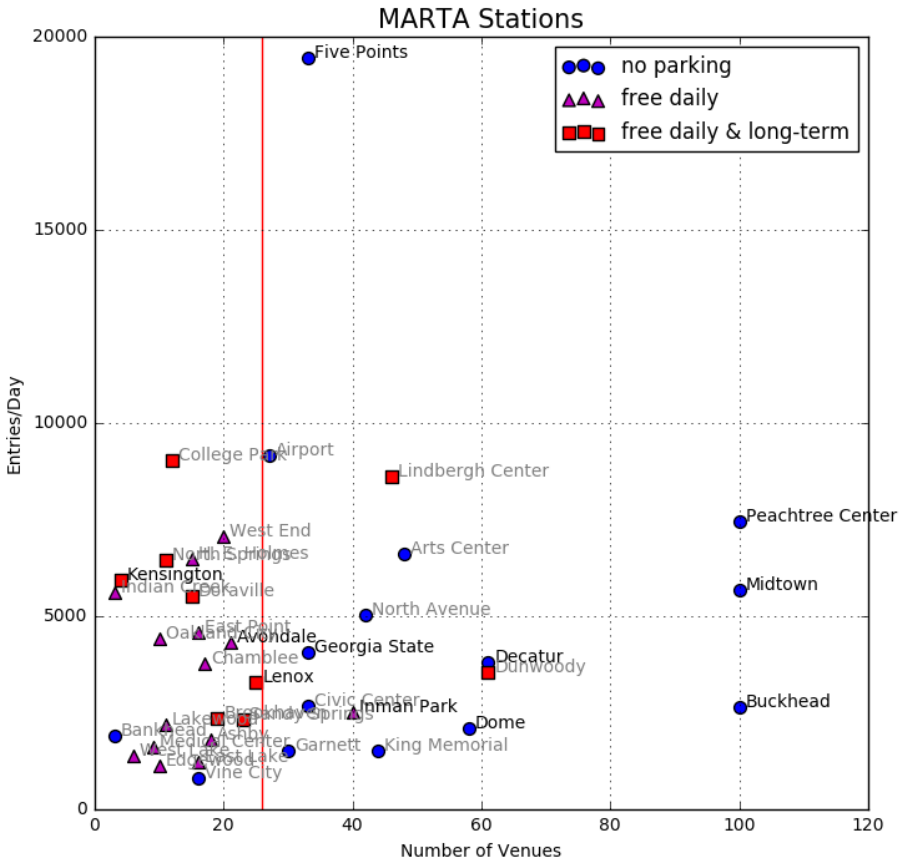

Figure 5. Entries/Day vs Number of Venues for 38 MARTA stations. Stations with Professional & Other Places are shown in bold

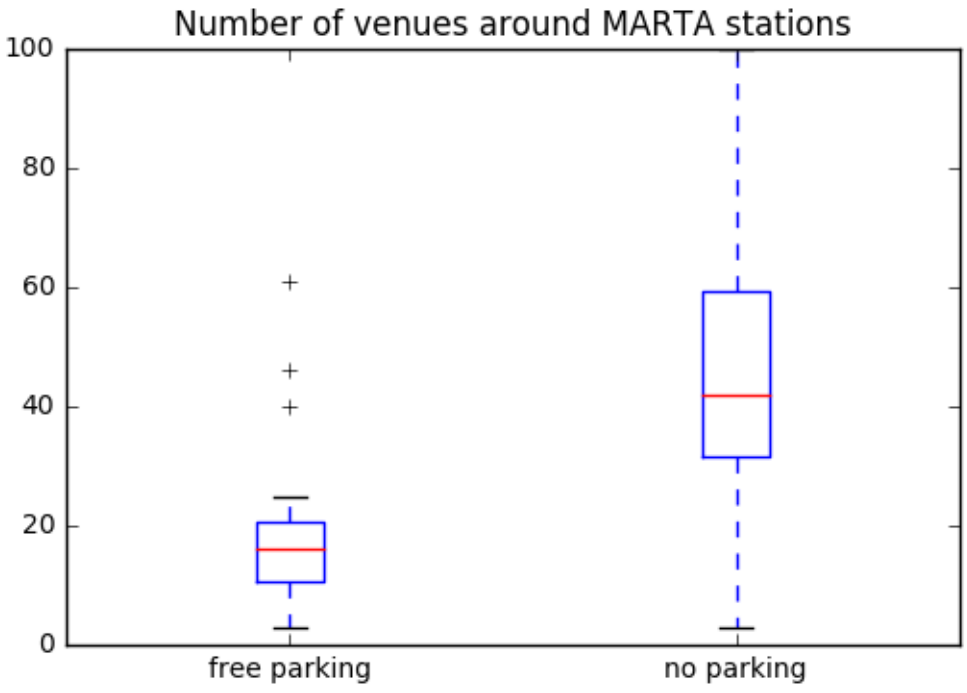

Figure 6. Number of venues around MARTA stations with free parking and with no parking

3. MARTA Stations Labeling

Next, I wanted to go deeper, beyond the first “fast food and coffee shops” level of the data, and see what makes each MARTA station really special and distinct from other stations, i.e. label each station with its discriminative categories.

For this I utilized Term Frequency-Inverse Document Frequency (TF-IDF) and Latent Dirichlet Allocation (LDA) that are widely used in text processing.

In MARTA data each station ("document") can be viewed as a bag of categories ("terms") of all venues near that station. Alternatively, we can view each station as a bag of English words, i.e. for example, consider “Asian” and “Restaurant” as two separate terms instead of a single “Asian Restaurant” category. I tried both approaches in my analysis.

TF-IDF normalization is defined as follows.

TF-IDF value of a term *t* in document *d* = TF * IDF

Term Frequency (TF) = (Number of times a term *t* appears in a document *d*) / (Total number of terms in the document *d*)

Inverse Document Frequency (IDF) = -log(n/N), where, *N* is the total number of documents in the corpus and *n* is the number of documents a term *t* has appeared in.

Due to the IDF term, TF-IDF method penalizes common and frequent terms like "Coffee Shop" but assigns greater weight to more rare and specific terms like "Airport Lounge" or "Aquarium".

Table 4 shows 9 general categories sorted with respect to their TF-IDF values for each MARTA station. Notice that Nightlife Spots now prevail among non-parking stations as opposed to Food and Shop & Service categories in the original categories counts (Table 3).

I applied TF-IDF sorting to the primary categories data as well, and got very detailed labels for each station (Table 5).

Next, I labeled MARTA stations using **Latent Dirichlet Allocation (LDA) with TF-IDF preprocessing**. LDA represents documents as weighted mixtures of latent topics. LDA topics are, in a sense, analogous to Foursquare categories but are automatically extracted from a corpus of documents. With LDA we can get a more detailed representation of data by simply increasing the number of topics.

Tables 6 and 7 present my LDA results for the bag of words approach. Table 6 shows distribution over stemmed English words for each of 10 topics that LDA built for my corpus of 38 documents. I manually gave each LDA topic a short name, for reference purposes. Table 7 shows topic weights for each station. LDA topic 7 (“place”) seem to be analogous to “Nightlife Spots” category in Table 4.

TF-IDF and LDA results can be utilized in many different ways, not just for labeling, but also for searching, clustering and visualization.

I applied TF-IDF normalization to the general categories data from Table 3 and divided each row by the sum of its elements so that all stations were treated based on venue categories and not on the number of venues. I used the result frequencies to group MARTA stations into category-related clusters.

Figure 7 shows Euclidean distance matrix for 38 stations. We see that all distances are very similar and close to 0, except for Bankhead, Indian Creek, Kensington, West Lake stations. With a distance matrix like this K-means and other clustering algorithms are very likely to separate only these 4 stations and consider the other 34 stations as one big cluster.

As expected, **K-means clustering algorithm** showed no elbow results for this data (Figure 8). It suggested that either there were no meaningful clusters or that the algorithm could not separate them, e.g. as in case when K-means is used for concentric circles (vs DBSCAN). Nevertheless, I ran K-means with k=9. The results actually were rather good (Table 8). I also tried DBSCAN and hierarchical clustering with Euclidean, Wasserstein and Hamming distances. They all provided similar clusters.

Finally, I used TF-IDF and LDA results for labeling stations in various **Folium maps** (Figures 9 and 10).

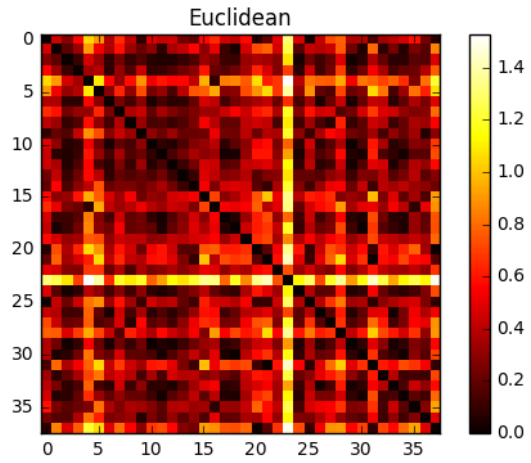

Figure 7. Euclidean similarity matrix for TF-IDF normalized general categories data

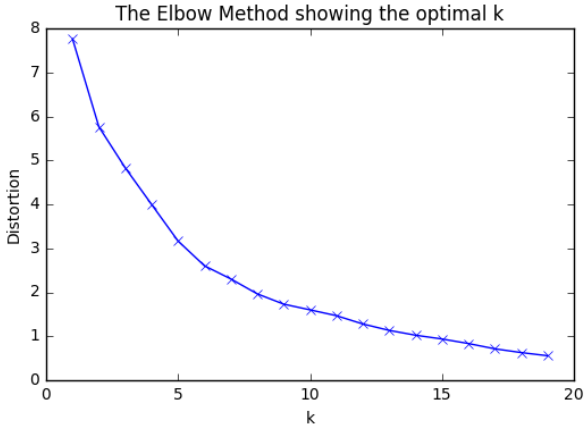

Figure 8. No elbow K-means for TF-IDF normalized general categories data

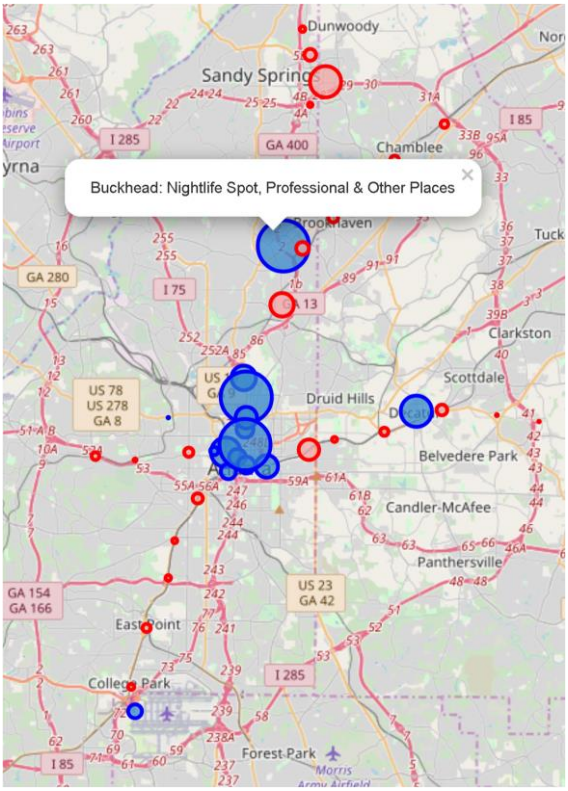

Figure 9. MARTA Map. Marker size reflects the number of venues at each station.
No parking (blue), free parking (red).

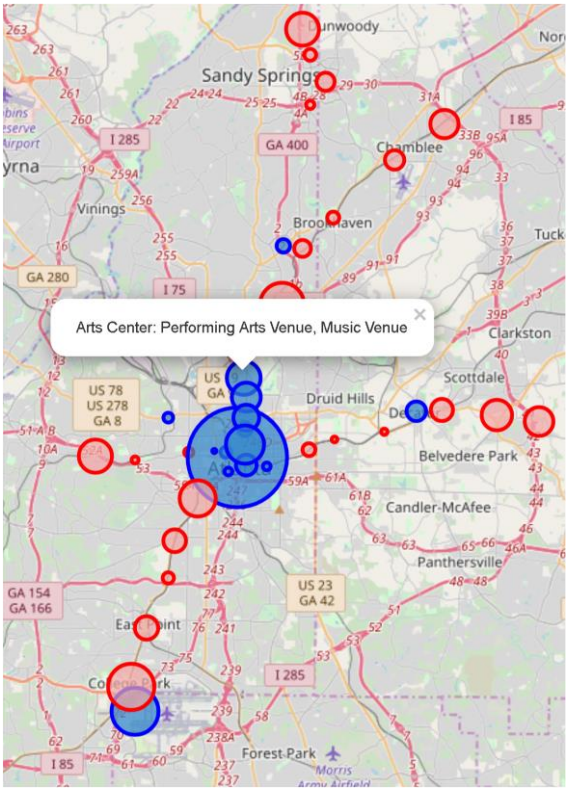

Figure 10. MARTA Map. Marker size reflects the number of Entries per Day at each station.
No parking (blue), free parking (red).

Table 4. MARTA stations: 9 general categories sorted with respect to their TF-IDF values. Stations with no parking are shown in blue

	Station	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category	6th Most Common Category	7th Most Common Category
0	Airport	Travel	Arts	Food	Shop	-	-	-
1	Arts Center	Arts	Nightlife	Food	Travel	Outdoors	Shop	-
2	Ashby	Nightlife	Outdoors	Food	Travel	Arts	Shop	-
3	Avondale	Outdoors	Nightlife	Prof et al	Arts	Travel	Shop	Food
4	Bankhead	Outdoors	Food	Shop	-	-	-	-
5	Brookhaven	Nightlife	Food	Shop	-	-	-	-
6	Buckhead	Nightlife	Prof et al	Travel	Food	Outdoors	Shop	-
7	Chamblee	Food	Outdoors	Shop	-	-	-	-
8	Civic Center	Nightlife	Residence	Travel	Outdoors	Food	Arts	Shop
9	College Park	Travel	Arts	Food	Shop	-	-	-
10	Decatur	Nightlife	Food	Prof et al	Outdoors	Arts	Shop	Travel
11	Dome	Arts	Nightlife	Outdoors	Food	Prof et al	Travel	Shop
12	Doraville	Nightlife	Food	Outdoors	Travel	Shop	-	-
13	Dunwoody	Travel	Shop	Food	Nightlife	Outdoors	-	-
14	East Lake	Travel	Outdoors	Food	Shop	-	-	-
15	East Point	Food	Arts	Shop	-	-	-	-
16	Edgewood	Outdoors	Arts	Travel	Food	Shop	-	-
17	Five Points	Nightlife	Arts	Prof et al	Food	Travel	Shop	-
18	Garnett	Arts	Nightlife	Travel	Food	Outdoors	Shop	-
19	Georgia State	College	Prof et al	Food	Arts	Outdoors	Travel	Shop
20	H.E.Holmes	Travel	Food	Shop	-	-	-	-
21	Indian Creek	Travel	Outdoors	-	-	-	-	-
22	Inman Park	Residence	Prof et al	Food	Outdoors	Shop	Travel	Arts
23	Kensington	Prof et al	Shop	-	-	-	-	-
24	King Memorial	Nightlife	Outdoors	Arts	Food	Travel	Shop	-
25	Lakewood	Travel	Arts	Outdoors	Food	Shop	-	-
26	Lenox	Nightlife	Prof et al	Travel	Food	Outdoors	Shop	-
27	Lindbergh Center	Nightlife	Residence	Food	Outdoors	Shop	-	-
28	Medical Center	Food	Shop	-	-	-	-	-
29	Midtown	Nightlife	Food	Outdoors	Travel	Arts	Prof et al	Shop
30	North Avenue	Arts	Nightlife	Food	Travel	Outdoors	Shop	-
31	North Springs	Travel	Outdoors	Food	Shop	-	-	-
32	Oakland City	Arts	Outdoors	Travel	Food	Shop	-	-
33	Peachtree Center	Nightlife	Travel	Food	Arts	Prof et al	Outdoors	Shop
34	Sandy Springs	Arts	Nightlife	Food	Outdoors	Shop	-	-
35	Vine City	Outdoors	Arts	Food	Shop	-	-	-
36	West End	Arts	Food	Shop	Travel	-	-	-
37	West Lake	Travel	Food	Shop	-	-	-	-

Table 5. MARTA stations: Top 10 primary categories sorted with respect to their TF-IDF values.

	Station	Top10
0	Airport	Airport Service, Airport Lounge, Airport Terminal, Food Stand, Betting Shop, Taxi Stand, Public Art, Caribbean Restaurant, Electronics Store, Bus Station
1	Arts Center	Performing Arts Venue, Music Venue, Theater, Exhibit, Indie Theater, Print Shop, Coffee Shop, Concert Hall, Art Museum, Lawyer
2	Ashby	Pool, Wings Joint, Fried Chicken Joint, Intersection, Light Rail Station, Gas Station, Sports Bar, Seafood Restaurant, Music Venue, Southern / Soul Food Restaurant
3	Avondale	Liquor Store, Martial Arts Dojo, Automotive Shop, Gym / Fitness Center, Cycle Studio, Office, Bakery, Brewery, Wine Shop, Museum
4	Bankhead	Ice Cream Shop, Gas Station, Park
5	Brookhaven	Sports Bar, Pharmacy, Department Store, Diner, Shopping Plaza, Wine Shop, Chinese Restaurant, Shipping Store, Salon / Barbershop, Bank
6	Buckhead	Clothing Store, Department Store, Women's Store, Cosmetics Shop, Steakhouse, Hotel, Furniture / Home Store, Jewelry Store, Shopping Plaza, Gift Shop
7	Chamblee	Vape Store, Health & Beauty Service, Optical Shop, Pet Store, Shopping Plaza, Trail, Wine Shop, Video Store, Sushi Restaurant, Furniture / Home Store
8	Civic Center	Hotel, Sculpture Garden, Steakhouse, Lounge, Dog Run, Ethiopian Restaurant, Residential Building (Apartment / Condo), Bistro, Plaza, Brewery
9	College Park	Pet Store, Concert Hall, Tapas Restaurant, Shoe Store, BBQ Joint, American Restaurant, Rental Car Location, Asian Restaurant, Southern / Soul Food Restaurant, Metro Station
10	Decatur	Gastropub, South American Restaurant, Sushi Restaurant, Plaza, Ice Cream Shop, Farmers Market, Ramen Restaurant, Soup Place, Spanish Restaurant, Candy Store
11	Dome	Basketball Stadium, Outdoor Sculpture, Sports Club, Fast Food Restaurant, Food Court, Dive Bar, Middle Eastern Restaurant, American Restaurant, Cafeteria, Souvenir Shop
12	Doraville	Taiwanese Restaurant, Seafood Restaurant, Karaoke Bar, Latin American Restaurant, Train Station, Korean Restaurant, Convenience Store, Indian Restaurant, Gas Station, Bank
13	Dunwoody	Department Store, Clothing Store, Furniture / Home Store, Lingerie Store, Toy / Game Store, Electronics Store, Cosmetics Shop, Hotel, Shoe Store, Men's Store
14	East Lake	Intersection, Beer Store, Scenic Lookout, Falafel Restaurant, Food, Wine Shop, American Restaurant, Donut Shop, Gym / Fitness Center, Italian Restaurant
15	East Point	Disc Golf, Shop & Service, Taco Place, Cupcake Shop, Vegetarian / Vegan Restaurant, Breakfast Spot, Caribbean Restaurant, Southern / Soul Food Restaurant, Pharmacy, Mexican Restaurant
16	Edgewood	Flower Shop, Platform, Yoga Studio, Park, Art Gallery, New American Restaurant, Food Truck, Gym / Fitness Center, Gym
17	Five Points	Hookah Bar, Event Space, Cuban Restaurant, Deli / Bodega, Sandwich Place, Korean Restaurant, Falafel Restaurant, Bubble Tea Shop, Poke Place, Mediterranean Restaurant
18	Garnett	Art Gallery, Moving Target, Strip Club, Lounge, Neighborhood, Paper / Office Supplies Store, Performing Arts Venue, Boutique, Café, Salon / Barbershop
19	Georgia State	College Basketball Court, Cuban Restaurant, Event Space, Asian Restaurant, Sandwich Place, Tapas Restaurant, Bubble Tea Shop, Movie Theater, Convenience Store, Sporting Goods Shop
20	H. E. Holmes	Lawyer, Cosmetics Shop, Discount Store, Liquor Store, Intersection, Chinese Restaurant, Fast Food Restaurant, Gas Station, Grocery Store, Bus Station
21	Indian Creek	Light Rail Station, Bus Station, Park
22	Inman Park	Mobile Phone Shop, Bookstore, Bank, Playground, Other Great Outdoors, Eye Doctor, Paper / Office Supplies Store, Supplement Shop, Health & Beauty Service, Poke Place
23	Kensington	Chiropractor, Discount Store, Gas Station, Pharmacy
24	King Memorial	Scenic Lookout, Fried Chicken Joint, Bar, Cemetery, Aquarium, Tunnel, Rock Club, Recreation Center, Shop & Service, Greek Restaurant
25	Lakewood	Train Station, Taxi, Movie Theater, Food, Liquor Store, Light Rail Station, Metro Station, Park, Burger Joint, Sandwich Place
26	Lenox	Food Court, Mediterranean Restaurant, Supplement Shop, Clothing Store, Office, Bistro, Japanese Restaurant, Chinese Restaurant, Cocktail Bar, Bus Station
27	Lindbergh Center	Shopping Plaza, Salon / Barbershop, Shipping Store, Nightclub, Garden Center, Big Box Store, Hardware Store, Tex-Mex Restaurant, Residential Building (Apartment / Condo), Video Game Store

28	Medical Center	Pharmacy, Smoke Shop, Diner, Food Court, Spa, Café, Coffee Shop
29	Midtown	Mediterranean Restaurant, Hotel, Irish Pub, Athletics & Sports, Taco Place, Mobile Phone Shop, Yoga Studio, Seafood Restaurant, Vietnamese Restaurant, American Restaurant
30	North Avenue	History Museum, Hotel, New American Restaurant, Theater, Market, Fondue Restaurant, Bath House, Speakeasy, Historic Site, Hot Dog Joint
31	North Springs	Metro Station, Construction & Landscaping, Garden, Bus Line, Train Station, Jewelry Store, Restaurant, Gym
32	Oakland City	Convenience Store, Art Gallery, Food, Vegetarian / Vegan Restaurant, Metro Station, Park, Wings Joint, Breakfast Spot
33	Peachtree Center	Hotel, Hotel Bar, Mediterranean Restaurant, Rental Car Location, Indian Restaurant, Bar, American Restaurant, Southern / Soul Food Restaurant, Theater, Cajun / Creole Restaurant
34	Sandy Springs	Business Service, Bagel Shop, Cafeteria, Art Museum, Hardware Store, Movie Theater, Bakery, Sporting Goods Shop, Discount Store, Japanese Restaurant
35	Vine City	Sports Club, Souvenir Shop, Football Stadium, Outdoor Sculpture, Ice Cream Shop, Park, Fast Food Restaurant, Pizza Place, Steakhouse, Wings Joint
36	West End	Vegetarian / Vegan Restaurant, Thrift / Vintage Store, Music Store, Discount Store, Women's Store, Light Rail Station, Video Store, Chinese Restaurant, Bank, Shoe Store
37	West Lake	Gas Station, Light Rail Station, Bus Station, Wings Joint, Café

Table 6. 10 LDA TF-IDF topics

Topic	Short Name	Stemmed Words
Topic0	store	0.020*hotel + 0.019*store + 0.017*chiropractor + 0.015*furnitur + 0.015*home + 0.014*plaza + 0.013*intersect + 0.013*cosmet + 0.013*shop + 0.013*steakhous
Topic1	restaur	0.024*conveni + 0.014*restaur + 0.014*vegan + 0.014*vegetarian + 0.014*galleri + 0.013*breakfast + 0.013*spot + 0.013*seafood + 0.012*scenic + 0.012*lookout
Topic2	station	0.036*station + 0.032*rail + 0.032*light + 0.026*cream + 0.020*park + 0.018*busi + 0.013*discount + 0.012*vegetarian + 0.012*vegan + 0.011*vintag
Topic3	airport	0.032*airport + 0.015*pharmaci + 0.014*depart + 0.013*sport + 0.013*diner + 0.012*barbershop + 0.012*wine + 0.012*ship + 0.011*servic + 0.011*stand
Topic4	colleg	0.019*basketbal + 0.015*stadium + 0.014*court + 0.012*fast + 0.012*colleg + 0.012*park + 0.011*asian + 0.009*space + 0.009*cuban + 0.009*event
Topic5	metro	0.018*metro + 0.014*gastropub + 0.013*construct + 0.013*line + 0.013*landscap + 0.013*station + 0.012*jewelri + 0.012*sushi + 0.012*south + 0.011*restaur
Topic6	fast	0.019*lawyer + 0.017*cosmet + 0.015*discount + 0.014*station + 0.014*intersect + 0.014*liquor + 0.013*fast + 0.013*chines + 0.013*groceri + 0.011*pharmaci
Topic7	place	0.021*hotel + 0.019*place + 0.016*sandwich + 0.015*theater + 0.015*restaur + 0.015*venu + 0.014*coffe + 0.013*museum + 0.012*shop + 0.012*liquor
Topic8	galleri	0.015*galleri + 0.013*hall + 0.013*concert + 0.013*restaur + 0.013*pool + 0.013*store + 0.012*shoe + 0.012*wing + 0.012*soul + 0.012*southern
Topic9	pharm	0.027*pharmaci + 0.021*flower + 0.021*platform + 0.019*smoke + 0.017*diner + 0.016*court + 0.016*yoga + 0.014*park + 0.013*studio + 0.012*galleri

Table 7. MARTA Stations: 10 LDA topics

	Station	Topic 0 store	Topic 1 restaur	Topic 2 station	Topic 3 airport	Topic 4 colleg	Topic 5 metro	Topic 6 fast	Topic 7 place	Topic 8 galleri	Topic 9 pharm
0	Airport	0.057	0	0	0.334	0	0	0	0.595	0	0
1	Arts Center	0	0	0	0	0	0	0	0.989	0	0
2	Ashby	0	0	0.070	0	0	0	0	0.497	0.414	0
3	Avondale	0.169	0	0	0	0	0	0	0.806	0	0
4	Bankhead	0.020	0.020	0.820	0.020	0.020	0.020	0.020	0.020	0.020	0.020
5	Brookhaven	0.278	0	0	0.236	0	0	0	0.462	0	0
6	Buckhead	0.748	0	0	0	0	0	0	0.213	0.020	0.016
7	Chamblee	0.750	0	0	0	0	0	0	0.227	0	0
8	Civic Center	0.373	0	0	0	0	0	0	0.610	0	0
9	College Park	0	0	0	0	0	0	0	0	0.962	0
10	Decatur	0.207	0	0	0	0	0.168	0	0.618	0	0
11	Dome	0.041	0	0	0	0.117	0	0	0.835	0	0
12	Doraville	0	0.967	0	0	0	0	0	0	0	0
13	Dunwoody	0.702	0	0	0	0	0	0	0.165	0.127	0
14	East Lake	0.542	0	0	0	0	0	0	0.432	0	0
15	East Point	0	0.565	0	0	0	0	0	0.413	0	0
16	Edgewood	0	0	0	0	0	0	0	0.483	0	0.464
17	Five Points	0	0	0	0	0	0	0	0.985	0	0
18	Garnett	0	0	0	0	0	0	0	0.548	0.436	0
19	Georgia State	0	0.057	0	0	0.101	0	0	0.831	0	0
20	H. E. Holmes	0.261	0	0	0	0	0	0.211	0.503	0	0
21	Indian Creek	0.017	0.017	0.850	0.017	0.017	0.017	0.017	0.017	0.017	0.017
22	Inman Park	0.737	0	0	0	0	0	0	0.253	0	0
23	Kensington	0.688	0.017	0.179	0.017	0.017	0.017	0.017	0.017	0.017	0.017
24	King Memorial	0.053	0.490	0	0	0	0	0	0.448	0	0
25	Lakewood	0	0	0.304	0	0	0	0	0.658	0	0
26	Lenox	0	0.520	0	0	0	0	0	0.461	0	0
27	Lindbergh Center	0.395	0	0	0	0	0	0	0.595	0	0
28	Medical Center	0	0	0	0	0	0	0	0.357	0	0.577
29	Midtown	0.091	0.062	0	0	0	0.119	0	0.724	0	0
30	North Avenue	0	0	0	0	0	0	0	0.987	0	0
31	North Springs	0	0	0.256	0	0	0.700	0	0	0	0
32	Oakland City	0	0.950	0	0	0	0	0	0	0	0
33	Peachtree Center	0	0.031	0	0	0	0	0	0.965	0	0
34	Sandy Springs	0.140	0	0.227	0	0	0	0	0.616	0	0
35	Vine City	0	0	0	0	0	0	0	0.972	0	0
36	West End	0.311	0	0.325	0	0	0	0	0.349	0	0
37	West Lake	0.010	0.010	0.910	0.010	0.010	0.010	0.010	0.010	0.010	0.010

Table 8. K-means clusters for TF-IDF normalized general categories data

Cluster	#	Station	1st Most Common Category	2nd Most Common Category
Cluster 0	1	Arts Center	Arts & Entertainment	Nightlife Spot
	11	Dome	Arts & Entertainment	Nightlife Spot
	17	Five Points	Nightlife Spot	Arts & Entertainment
	18	Garnett	Arts & Entertainment	Nightlife Spot
	30	North Avenue	Arts & Entertainment	Nightlife Spot
	34	Sandy Springs	Arts & Entertainment	Nightlife Spot
Cluster 1	0	Airport	Travel & Transport	Arts & Entertainment
	14	East Lake	Travel & Transport	Outdoors & Recreation
	20	H. E. Holmes	Travel & Transport	Food
	21	Indian Creek	Travel & Transport	Outdoors & Recreation
	25	Lakewood	Travel & Transport	Arts & Entertainment
	31	North Springs	Travel & Transport	Outdoors & Recreation
	37	West Lake	Travel & Transport	Food
Cluster 2	4	Bankhead	Outdoors & Recreation	Food
	16	Edgewood	Outdoors & Recreation	Arts & Entertainment
	35	Vine City	Outdoors & Recreation	Arts & Entertainment
Cluster 3	2	Ashby	Nightlife Spot	Outdoors & Recreation
	3	Avondale	Outdoors & Recreation	Nightlife Spot
	5	Brookhaven	Nightlife Spot	Food
	6	Buckhead	Nightlife Spot	Professional & Other Places
	10	Decatur	Nightlife Spot	Food
	12	Doraville	Nightlife Spot	Food
	13	Dunwoody	Travel & Transport	Shop & Service
	24	King Memorial	Nightlife Spot	Outdoors & Recreation
	26	Lenox	Nightlife Spot	Professional & Other Places
	29	Midtown	Nightlife Spot	Food
	33	Peachtree Center	Nightlife Spot	Travel & Transport
Cluster 4	19	Georgia State	College & University	Professional & Other Places
Cluster 5	23	Kensington	Professional & Other Places	Shop & Service
Cluster 6	9	College Park	Travel & Transport	Arts & Entertainment
	15	East Point	Food	Arts & Entertainment
	32	Oakland City	Arts & Entertainment	Outdoors & Recreation
	36	West End	Arts & Entertainment	Food
Cluster 7	8	Civic Center	Nightlife Spot	Residence
	22	Inman Park	Residence	Professional & Other Places
	27	Lindbergh Center	Nightlife Spot	Residence
Cluster 8	7	Chamblee	Food	Outdoors & Recreation
	28	Medical Center	Food	Shop & Service

Results

- I used Foursquare to retrieve 223 primary and 9 general categories for 1156 popular venues near 38 MARTA stations
- I analyzed Entries/Day and number of venues for each MARTA station
- I discovered that number of nearby venues is strongly correlated with free parking
- I discovered that MARTA stations can be roughly divided into two major groups: park-and-ride stations, with < 26 venues, and urban stations with no parking and ≥ 26 venues. Three important exceptions are Dunwoody, Lindbergh Center and Inman Park/Reynoldstown located in dense areas. These stations have free parking and many venues nearby
- I discovered that fast food restaurants, coffee shops and retail shops prevail in areas around all MARTA stations
- I used TF-IDF and LDA algorithms to extracted venue categories that are most specific\discriminative for each station and used these categories to label\annotate each station
- I used K-means and other clustering algorithms to group MARTA stations into ~10 clusters based on their venues categories
- I visualized MARTA data on interactive Folium maps

Discussion

Results for Buckhead, Midtown and Peachtree Center are incomplete, since these stations reached Foursquare API's venue result limit of 100. We can get more Foursquare results by using nearby points. The results may overlap, so we will need to filter the duplicates out and keep changing the point by a tiny margin till the number of unique venues for each station stops growing or reaches some limit.

MARTA venue analysis can be enriched also by using secondary categories, foot traffic and other detailed information for each venue. This require Foursquare primary calls, so I left it beyond the scope of this small project.

Category merging using Foursquare category tree also has much room for improvement. For example, parent category for "Pharmacy" is "Shop & Service". This makes Medical Center look like any other station.

MARTA station labeling is basically a text document labeling/annotation and topic modeling problem. There are many algorithms for this, with many tunable parameters. It would be interesting, for example, to vary number of topics and build LDA models with different detalization.

Conclusion

I analyzed popular venues near 38 MARTA rail stations in Atlanta, Georgia.

My analysis revealed very low business activity (less than 26 venues) around 19 out of 22 MARTA stations with free parking. Three important exceptions are Dunwoody (I-285/SR 400 interchange), Lindbergh Center (I-85/SR 400 interchange) and Inman Park/Reynoldstown. These three stations provide free daily parking for MARTA passengers, and are located in heavily dense areas with many businesses.

Buckhead, Midtown and Peachtree Center have the largest number of venues, they reached Foursquare API's 100 result limit. Civic Center, Lindbergh Center, Inman Park/Reynoldstown, and potentially Buckhead, Midtown and Peachtree Center have Residence venues within walking distance from the train station. Georgia State is the only MARTA station with venues in College & University category.

Fast food restaurants and coffee shops dominate areas around all MARTA stations. Obviously, this is where people spend most of their time and money, which is a good news for the fast food companies.

I used TF-IDF and LDA techniques to label each MARTA station with its discriminative categories/topics. It turned out that there are many Nightlife Spots near MARTA, combined either with Arts & Entertainment, or Food venues.

I created interactive Folium maps of Atlanta with information for each MARTA stations (name, venue categories-based label, Entries/Day, number of venues, free parking) superimposed on top.

I also discovered that there is a good chiropractic at Kensington station in Atlanta.

Acknowledgements

This project was part of a course on Coursera called Applied Data Science Capstone, and I really enjoyed doing it. You can take this course online by clicking here

References

MARTA Rail System Data, Available at https://en.wikipedia.org/wiki/MARTA_rail

Foursquare categories hierarchy, Available at https://developer.foursquare.com/docs/resources/categories

Applied Data Science Capstone Coursera Online Course http://cocl.us/DP0701EN_Coursera_Week3_LAB2