Finding The Best Station To Live Nearby

Sumit Chhabra

March 30th, 2019

1. Introduction

1.1. Background

I live in Charlotte city in North Carolina in USA for last years. Charlotte is one of the fastest growing city where migration rate is very high in the recent years. Though living in Charlotte city is cheaper than Silicon Valley but a car is needed to commute in different parts of the city. City transportation is improved in last few years but its not at par with the big cities like New York or Los Angeles. All buses go thru Transportation Center in Charlotte uptown. So going from one part of a city to another part can sometimes take upto 2 buses and exchanging buses at the Transportation center. On top of that buses run on different frequencies during weekdays peak hours and non-peak hours as well as weekends.

City has Lynx Blue Light rail which opened in 2007 and operated between south end of the city to uptown in the center of the city. In 2018, there was addition of Lynx Blue Light rail in the north part of the city. This light rail connects different parts of north part of the city with the Charlotte Uptown.

1.2. Problem

One of my friend is moving from India to Charlotte City for work. He will be working in Uptown and will be living here for few years. In the beginning, he won't have a car and will not be easy for him to commute. Either he will have to do car pool or take Uber or Lyft which can add expenses to the monthly budget. It was easier for him to go to work if he lives near any train stations. Below is the stations map of Lynx Blue Line Light Rail (source: https://lightrail.uncc.edu/using-light-rail).

Lynx Blue Line Light Rail Station Map

Source: https://lightrail.uncc.edu/using-light-rail

1.3. Interest

The goal of my project is to find the best station where nearby will be lot of venues like grocery stores, convenience stores, etc in the quarter mile walkable distance. Use this station address to find the apartments nearby and manage to commute to work in Uptown and walk to nearby stores for daily needs and grocery without the need for a car.

2. Data Preparation

2.1. Data Sources

Data for this project was collected from the various online sources such as wikipedia (link), Mecklenburg county websites (link)and Foursquare API. No data was stored locally

2.2. Data Cleaning

First step was to find the list of the stations which I was able from the wikipedia page. Table consisting of **26** stations and their location addresses was scraped. Wikipedia covered special bus route called Lynx Gold line. I had to clean up the data to filter out only station data which were part of Lynx Blue line and I needed for this project. After collecting Station Name and Address and had to rearrange them in the same order the way light rail travels from north end of the city to the south end of the city.

Second step was to find latitude and longitude of each station's address. Initially I used Nominatim library from Geopy Geocoder package to get latitude and latitude but I ran into a issue. Station which had North Tryon in their address (see 2-4) Nominatim returned same coordinates for them which created data discrepancies. I signed up trial for Google Maps Platform API and used GoogleV3 which is also part of Geopy Geocoder, and it provided more accurate latitude and longitude of each station address as shown in the figure below.

	Station	Location	Longitude	Latitude
0	UNC Charlotte–Main	9025 Cameron Boulevard	-80.733757	35.311705
1	JW Clay Blvd/UNC Charlotte	9048 North Tryon Street	-80.745372	35.310881
2	McCullough	8312 North Tryon Street	-80.752444	35.300553
3	University City Blvd	7205 North Tryon Street	-80.761479	35.286898
4	Tom Hunter	6505 North Tryon Street	-80.766868	35.277949
5	Old Concord Road	5442 North Tryon Street	-80.774760	35.261053
6	Sugar Creek	644 East Sugar Creek Road	-80.794170	35.252075
7	36th Street	434 East 36th Street	-80.805399	35.248484
8	25th Street	2227 North Brevard Street	-80.817233	35.241397
9	Parkwood	327 Parkwood Avenue	-80.825144	35.235313
10	9th Street	237 East Ninth Street	-80.835324	35.229433
11	7th Street	260 East Seventh Street	-80.838126	35.227316
12	Charlotte Transportation Center/CTC Arena	310 East Trade Street	-80.840633	35.224688
13	3rd Street/Convention Center	305 East Third Street	-80.845430	35.222255
14	Stonewall	260 East Stonewall Street	-80.846994	35.221287
15	Carson	218 East Carson Boulevard	-80.851835	35.218905
16	Bland Street	1511 Camden Road	-80.855333	35.215676
17	East/West Boulevard	1821 Camden Road	-80.858862	35.212430
18	New Bern	129 New Bern Street	-80.868959	35.199771
19	Scaleybark	3750 South Boulevard	-80.875522	35.192032
20	Woodlawn	4756 Old Pineville Road	-80.879288	35.176200
21	Tyvola	5703 Old Pineville Road	-80.877346	35.162835
22	Archdale	6230 South Boulevard	-80.876791	35.150680
23	Arrowood	7717 England Street	-80.876641	35.134202
24	Sharon Road West	8815 Crump Road	-80.882237	35.119992
25	I-485/South Boulevard	9508 South Boulevard	-80.882237	35.106879

Third step was I used latitude and longitude of each station address and collected list of all venues (with limit of 100) in 500 meters of the radius, using FourSquare API. Total of 863 venues were collected.

Fourth step was I used Transit Station Quarter Mile Walk website to collect dataset which included each station and shape of walkable area around the station. This step was carried after the dataset was clustered and segmented to provide good visualization. This dataset was in ArcGIS esri cluster points json. I had to use ogr2ogr web client (website) to convert esri json to geojson which is compatible with Folium library, used to display maps and overlays. Unfortunately this geojson only covers the northeast extension of the light rail, still it was good enough to do the data science project.

2.3. Feature Selection

New dataframe was prepared which contained list of all station, their address and venues data. Each venue data included Venue name, Venue Latitude, Venue Longitude and Venue Category. Here's how data collected looked like before proceeding the the exploratory data analysis

	Station	Station Latitude	Station Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	UNC Charlotte–Main	35.311705	-80.733757	Barnes & Noble	35.308382	-80.733892	College Bookstore
1	UNC Charlotte–Main	35.311705	-80.733757	Einstein Bros Bagels	35.308576	-80.733696	Bagel Shop
2	UNC Charlotte–Main	35.311705	-80.733757	Bojangles' Famous Chicken 'n Biscuits	35.308533	-80.733671	Fast Food Restaurant
3	UNC Charlotte–Main	35.311705	-80.733757	Wendy's	35.308611	-80.733693	Fast Food Restaurant
4	UNC Charlotte–Main	35.311705	-80.733757	Starbucks	35.308353	-80.733830	Coffee Shop

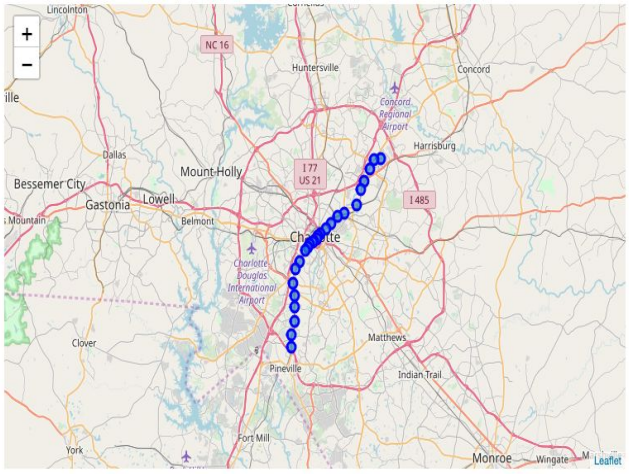
3. Exploratory Data Analysis

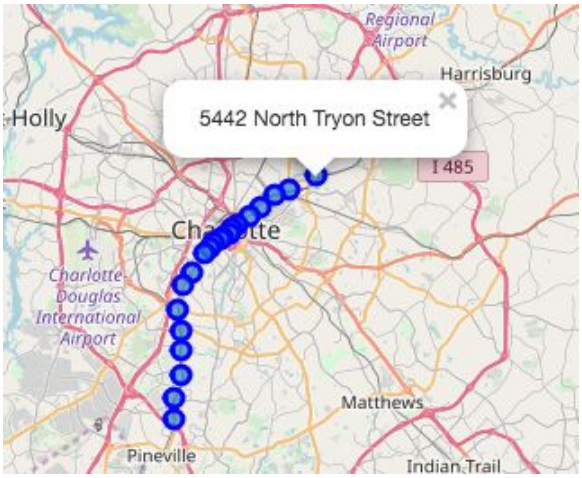
Since the goal of this project is to find a station with lots of venues in the walkable distance and then use that information to find an apartment or house to rent in the same vicinity. This project will only focus on finding the best station based on the data collected.

The analysis was done in two folds. In first part, venues around each train station were explored and in second part, quarter mile walkable distance around the stations in the northeast extension of the light rail were explored.

3.1. Map of Stations

We will start with the basic data visualization which will require gathering latitude and longitude of Charlotte city. Using Folium library, I displayed the city map and plotted all stations as a blue circle markers as shown below. Each station when clicked displayed the station name.

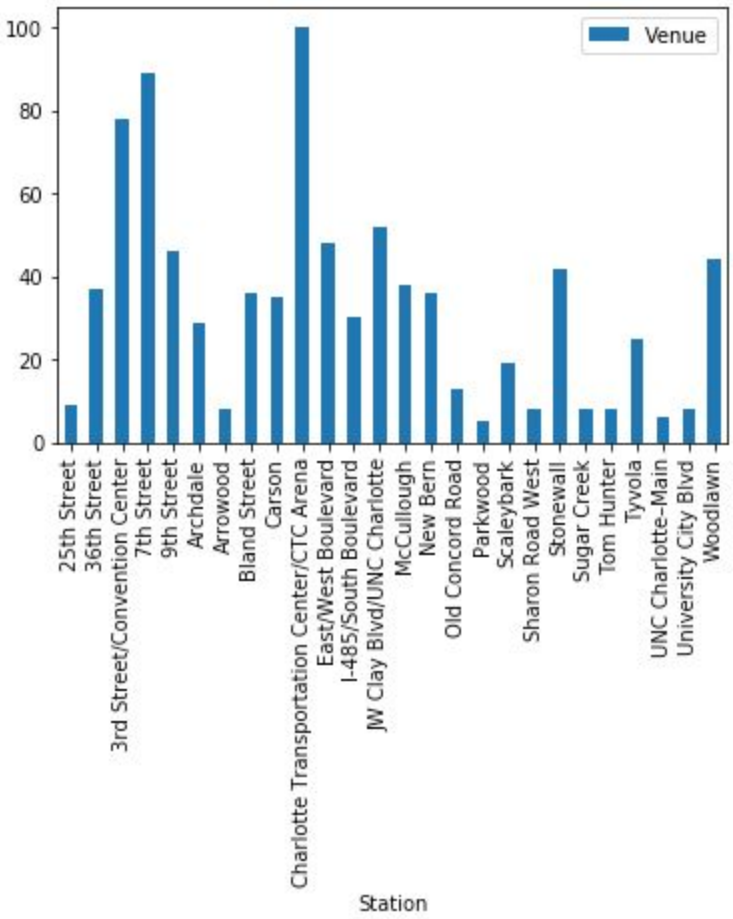




3.2. Grouping

With our given dataset which contains station name, address and venue information, we will do grouping on station name to get the count of venues for each station. Number of venues varied from four to one hundred as per the limit. Result is shown as head of the dataframe as well as best represented by the bar chart (see page 9).

	Station	Station Latitude	Station Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	25th Street	12	12	12	12	12	12
1	36th Street	38	38	38	38	38	38
2	3rd Street/Convention Center	78	78	78	78	78	78
3	7th Street	92	92	92	92	92	92
4	9th Street	45	45	45	45	45	45
5	Archdale	27	27	27	27	27	27
6	Arrowood	8	8	8	8	8	8
7	Bland Street	38	38	38	38	38	38
8	Carson	37	37	37	37	37	37
9	Charlotte Transportation Center/CTC Arena	100	100	100	100	100	100



Bar Chart showing Station names in X -axis and Venues count in Y-axis

3.3. Categories

Using this venues dataset, 176 uniques categories were identified. These categories were converted into columns for next dataset. One-hot coded method was used to identify which category is present (shown as 1) or not (shown as 0). Head of that dataset is shown below

	Station	American Restaurant	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Dealership	...	Trail	Veg Res
0	UNC Charlotte–Main	0	0	0	0	0	0	0	0	0	...	0	
1	UNC Charlotte–Main	0	0	0	0	0	0	0	0	0	...	0	
2	UNC Charlotte–Main	0	0	0	0	0	0	0	0	0	...	0	
3	UNC Charlotte–Main	0	0	0	0	0	0	0	0	0	...	0	
4	UNC Charlotte–Main	0	0	0	0	0	0	0	0	0	...	0	

5 rows × 177 columns

This one-hot coded dataset was grouped by Station name and mean frequency of occurrence of each category was calculated and the result dataset is shown as below. This dataset will be later used for clustering algorithm also.

	Station	American Restaurant	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Dealership
0	25th Street	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000
1	36th Street	0.000000	0.00	0.000000	0.052632	0.000000	0.026316	0.000000	0.000000	0.000
2	3rd Street/Convention Center	0.051282	0.00	0.000000	0.012821	0.025641	0.000000	0.012821	0.000000	0.000
3	7th Street	0.043478	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000
4	9th Street	0.022222	0.00	0.000000	0.022222	0.000000	0.000000	0.000000	0.000000	0.000
5	Archdale	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000
6	Arrowood	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.125000	0.000000	0.125
7	Bland Street	0.026316	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000
8	Carson	0.000000	0.00	0.000000	0.027027	0.000000	0.000000	0.000000	0.000000	0.000
9	Charlotte Transportation Center/CTC Arena	0.040000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000
10	East/West Boulevard	0.060000	0.00	0.000000	0.000000	0.000000	0.000000	0.020000	0.000000	0.000
11	I-485/South Boulevard	0.066667	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000
12	JW Clay Blvd/UNC Charlotte	0.039216	0.00	0.000000	0.000000	0.000000	0.000000	0.058824	0.000000	0.000
13	McCullough	0.025641	0.00	0.000000	0.000000	0.000000	0.000000	0.025641	0.025641	0.000
14	New Bern	0.027027	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000
15	Old Concord Road	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000

This dataset was further analyzed to determine top 5 venues at each station. Here's a sample output of the result.

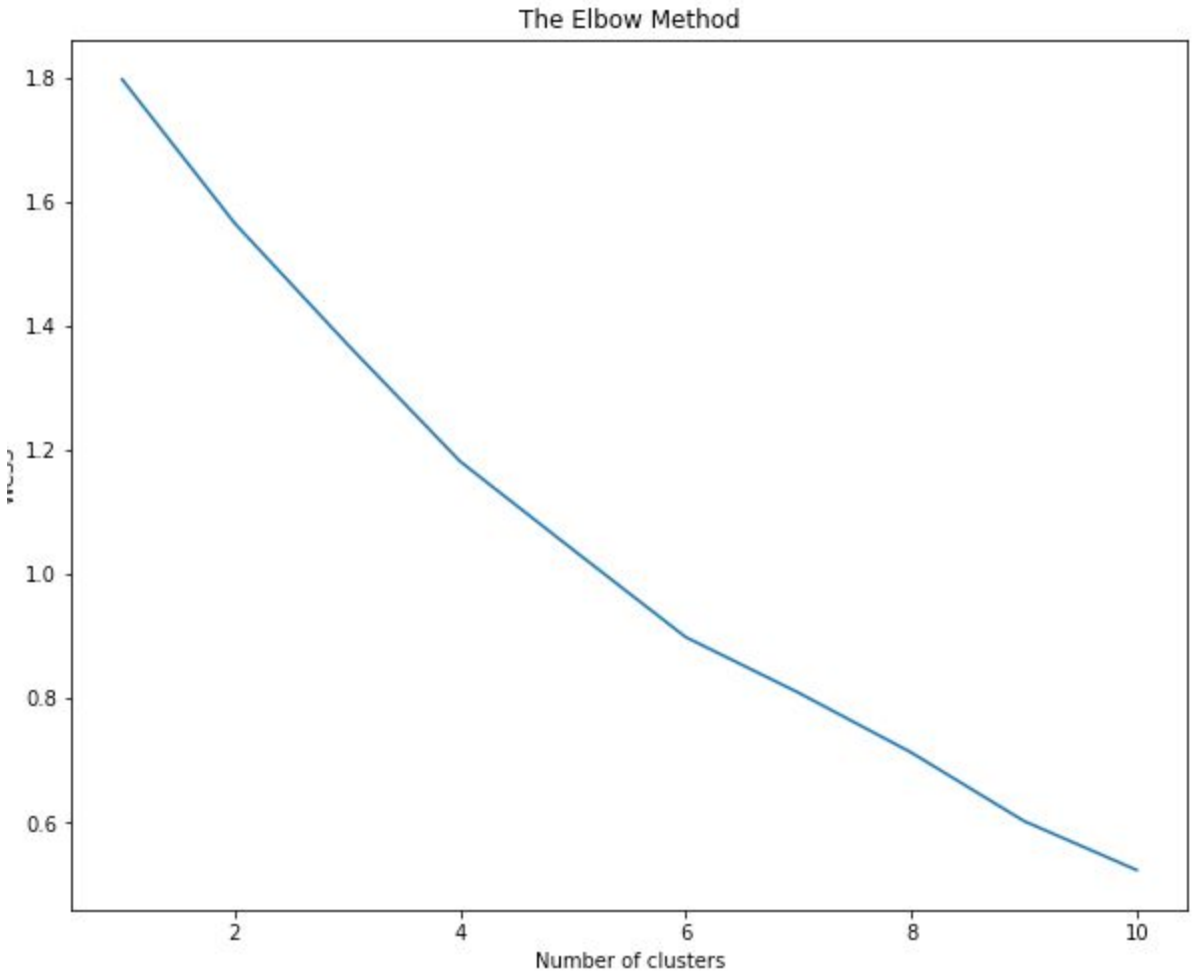
```
----JW Clay Blvd/UNC Charlotte----
                 venue  freq
0          Pizza Place  0.06
1     Asian Restaurant  0.06
2  American Restaurant  0.04
3              Bakery  0.04
4   Mexican Restaurant  0.04


----McCullough----
                  venue  freq
0                 Hotel  0.13
1  Fast Food Restaurant  0.08
2        Sandwich Place  0.08
3                   Spa  0.05
4     Mobile Phone Shop  0.05
```

We used the same dataset to determine the top 10 categories at each station and the result was shown as

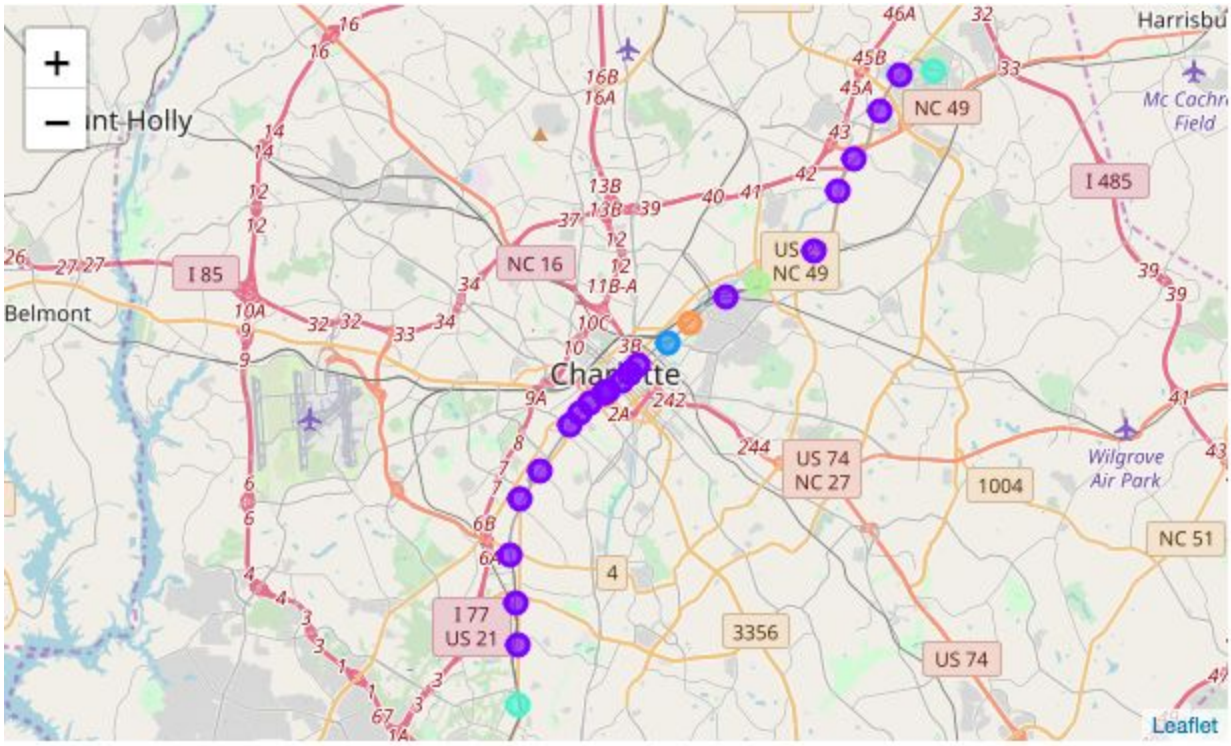
	Station	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Mos Commo Venu
0	25th Street	Brewery	Dog Run	Park	Bakery	Light Rail Station	Theater	Thrift / Vintage Store	Gym	Foo Truc
1	36th Street	Bar	Gastropub	Sports Bar	Art Gallery	Gym	Snack Place	Eastern European Restaurant	Market	Southern Soul Foo Restaurar
2	3rd Street/Convention Center	Hotel	Pizza Place	American Restaurant	Deli / Bodega	Steakhouse	Pub	New American Restaurant	Hot Dog Joint	Smoothi Sho
3	7th Street	Restaurant	American Restaurant	Theater	Pizza Place	Coffee Shop	Hotel	Sandwich Place	Performing Arts Venue	Nightclu
4	9th Street	Coffee Shop	Pizza Place	Lounge	Performing Arts Venue	Restaurant	Theater	American Restaurant	Pub	Light Ra Statio

3.4. K-means Clustering

Elbow method was used to determine best number of cluster using grouped one-hot coded dataset. Station name was dropped from the dataset prior to running for clustering. Elbow method depicted as 4 being the best option for number of clusters, as shown below



So running the K Means clustering algorithm using grouped mean frequency resulted in 6 clusters and clusters were divided based on venue categories around each station. Each station was assigned cluster label and map of Charlotte city was displayed showed each station with their cluster label's color.



3.5. Cluster Analysis

Each of the four cluster was analyzed.

- **Cluster 1** consisted only station and Gas Station was the popular venue at this station, followed by convenience store. Ironically 6th most common venue is station itself.

	Station	Location	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
24	Sharon Road West	8815 Crump Road	-80.882237	Gas Station	Convenience Store	Pizza Place	Storage Facility	Grocery Store	Light Rail Station

- **Cluster 2** clearly collected most of stations either in the Uptown, in the south extension of the light rail and few stations in the northeast of the light rail. All these station had mostly high number eating places like American and Asian restaurants, pizzeria, fast food restaurants in common. Partial dataset is shown below

	Station	Location	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	
1	JW Clay Blvd/UNC Charlotte	9048 North Tryon Street	-80.745372	Pizza Place	Bakery	Restaurant	Asian Restaurant	Me
2	McCullough	8312 North Tryon Street	-80.752444	Hotel	Fast Food Restaurant	Sandwich Place	Mobile Phone Shop	
3	University City Blvd	7205 North Tryon Street	-80.761479	Hotel	Intersection	Breakfast Spot	Steakhouse	
4	Tom Hunter	6505 North Tryon Street	-80.766868	Convenience Store	Light Rail Station	Pizza Place	Department Store	
5	Old Concord Road	5442 North Tryon Street	-80.774760	Cosmetics Shop	Convenience Store	Restaurant	Food Truck	
7	36th Street	434 East 36th Street	-80.805399	Bar	Gastropub	Sports Bar	Art Gallery	A
10	9th Street	237 East Ninth Street	-80.835324	Coffee Shop	Pizza Place	American Restaurant	Restaurant	
11	7th Street	260 East Seventh Street	-80.838126	Restaurant	Hotel	American Restaurant	Pizza Place	
12	Charlotte Transportation Center/CTC Arena	310 East Trade Street	-80.840633	Hotel	American Restaurant	Steakhouse	Italian Restaurant	

- **Cluster 3** consisted of only one train station and Tattoo Parlor and Arcade are the popular venues here. Clearly this cluster isn't ideal for finding apartment.

	Station	Location	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
9	Parkwood	327 Parkwood Avenue	-80.825144	Tattoo Parlor	Arcade	Dessert Shop	Athletics & Sports	Light Rail Station	Yoga Studio

- **Cluster 4** consists of 2 station and they both Fast Food restaurants as the most popular venues. First station is inside the University campus and second station is surrounded by business and corporate offices, so clearly not ideal for the residency.

	Station	Location	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6 C
0	UNC Charlotte–Main	9025 Cameron Boulevard	-80.733757	Fast Food Restaurant	Bagel Shop	Pharmacy	College Bookstore	Coffee Shop	
23	Arrowood	7717 England Street	-80.876641	Fast Food Restaurant	Convenience Store	Latin American Restaurant	Grocery Store	Asian Restaurant	De

- **Cluster 5** consists of only one station and Breweries is the most popular venue, followed by different ethnic restaurants. This station is surrounded by single family houses mostly.

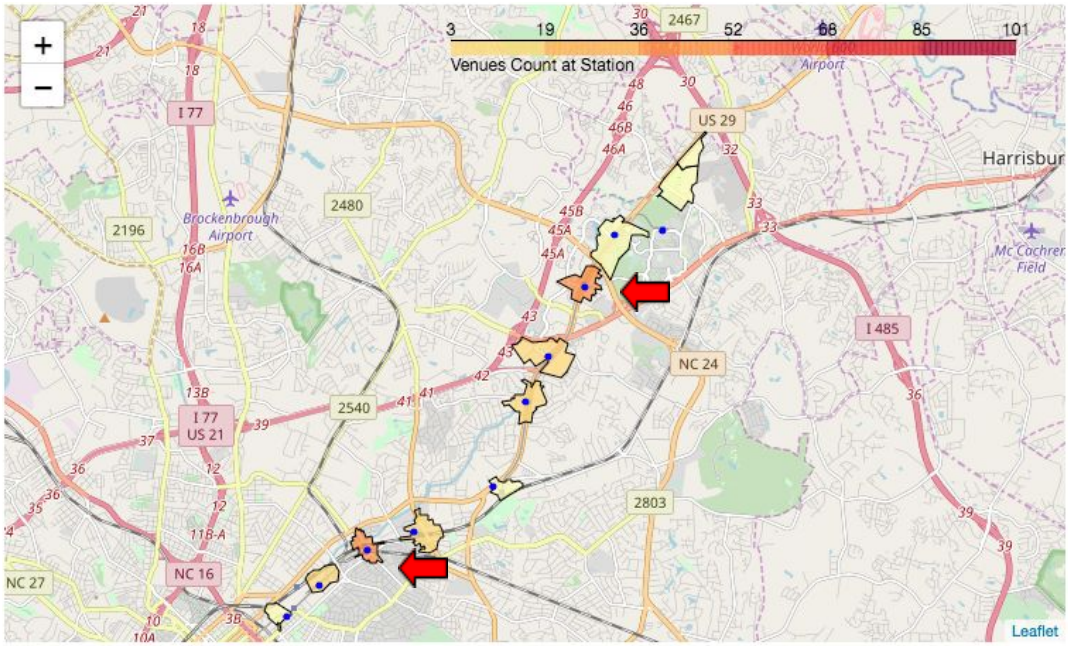
	Station	Location	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
6	Sugar Creek	644 East Sugar Creek Road	-80.79417	Brewery	Thai Restaurant	Breakfast Spot	Vietnamese Restaurant	Shopping Mall	Dim Sum Restaurant

- **Cluster 6** consists of only one station and its one of popular part of the city for its urban neighborhood. This station is also potential candidate for finding the apartment.

	Station	Location	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7t Cc
8	25th Street	2227 North Brevard Street	-80.817233	Brewery	Bakery	Theater	Gym	Coffee Shop	Park	D

3.6. Quarter Mile Walkable Distance Area Map

In this part, geojson data was prepared which was collected Charlotte Open Data Portal which included boundary shape of walkable area and it merged with the dataset in which each station was grouped along with the number of venues at each station. Since geojson was limited to Northeast extension only, Folium map was used to show each station and the walkable area around it. Venue Count was used as threshold. Stations with lower number of venues were shown with yellow color and stations with higher number of venues were shown with red color. Be;ow map shows only north east extension of the light rail.



4. Conclusion

In this study, after analyzing the venue data of each cluster and combining with the quarter mile walkable area map, it was clear two station were picked in the northeast extension of the light rail. Station are McCullough in Cluster 2 and 25th Street in Cluster 5 as shown by two red arrows in the quarter mile walkable area. Since I live this city, this study accurately predicted the stations. In addition to grocery stores, restaurants, shops, they are lot of apartments, hotels, and motels near these two stations. So it will be convenient for anyone relocating to Charlotte city and working in uptown and dont have car or want to save time in traffic congestion in and out of uptown during peak hours to find an apartment or house near these station in a walkable distance range. Not to mention, time spent in train ride can be used for listening to music or audiobook, read a book or do any office work.

5. Future Possibilities

The data collected from this study can be further to assist with the housing needs. I explored the possibility of using websites like Zillow, Trulia and Redfin as they provide the Developer API (note signup is required) and can be use to find apartment or house as well as look at the average price of them. If my county can provide walkable map data for the south extension of the light rail, it can give complete study of the entire city of Charlotte.