



# GEO ANALYSIS

---

CHARLOTTE, NORTH CAROLINA

# BACKGROUND

Charlotte was always a vibrant metropolitan area supported by a family oriented suburban cities.  
The growth boom is wonderful to watch.

While looking for a new home upon relocating back to Charlotte, we looked at the various suburban cities in the Charlotte Metropolitan Area for a home and we found that while you can buy a house in any nearby city, the amenities were not always the best in the cities with newer neighborhoods.

This projects aims to analyze the various suburban cities in the Charlotte Metropolitan Area.





# PROBLEM STATEMENT

So my goal is to analyze which cities surrounding the Charlotte Metropolitan and solve for 2 problem from 2 different perspective:

1. **Home Seekers**: Assist people in evaluating various surrounding cities to enable them to find the location which has the amenities that they care about.
2. **Business Owners**: Assist businesses in evaluating under served market to analyze what business opportunity may exist in a particular location.



# DATA

4 data sources were reviewed as follows:

1. [Will not be used] <https://www.gps-latitude-longitude.com/> This site requires location of each city to be fed in one at a time. This is very time consuming and therefore not very efficient.
2. [Will not be used] <https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/export/> This is an excellent source for city information and this dataset gave me a list of over 1,000 cities in North Carolina. However, the data was not very rich and I had no way to know which cities were around the Charlotte Metropolitan Area. Hence, I do not intend to use this dataset.
3. [Will not be used] [https://en.wikipedia.org/wiki/Charlotte%2C\\_North\\_Carolina](https://en.wikipedia.org/wiki/Charlotte%2C_North_Carolina) This link has some very rich information but did not have the structured data that I am looking for. Therefore, I will not be using this data source.
4. [Will be used] <https://simplemaps.com/data/us-cities> Simple map offers a download of most large cities in the world and I was able to download the list of US cities. I was further able to filter out just the cities in North Carolina. This consists of 736 cities. Since our target is cities surrounding the Charlotte Metropolitan Area, which fall in Mecklenburg county, I am going to further filter on the cities within the Mecklenburg county domain.

A copy of the city level data in North Carolina is available in my github repository. Link is provided below.

[https://raw.githubusercontent.com/amitarya76/Coursera\\_Capstone/master/NC\\_CITY\\_DATA.csv](https://raw.githubusercontent.com/amitarya76/Coursera_Capstone/master/NC_CITY_DATA.csv)





# DATA ELEMENTS

City Name allows for identification of cities. From the above dataset, I intend to use the county as the filter to get the cities in the “MECKLENBURG” county and combine that dataset with Foursquare API data to get venues and places in each city. The population and density data will inform our analysis on high, moderate and low growth areas.



## SOURCE = SIMPLE MAPS

#	Field Name	Data Value Example	Data Type
1	city	Trenton	City Name
2	city_ascii	Trenton	City Name
3	state_id	NC	State Abbreviation
4	state_name	North Carolina	State Name
5	county_fips	37103	FIPS Code
6	county_name	Jones	County Name
7	county_fips_all	37103	FIPS Code
8	lat	35.064	Latitude Data
9	lng	-77.3553	Longitude Data
10	population	285	Population in 'ooo
11	density	484	Population Density
12	timezone	America/New_York	Time Zone
13	zips	28585	Zip Code



## SOURCE = FOURSQUARE

#	Field Name	Data Value Example	Data Type
1	Latitude	35.064	Latitude Data
2	Longitude	-77.3553	Longitude Data
3	Category	Juice Bar	Type of Venue
4	Venue	Carolina Smoothies	Venue Name
5	Location	525 Tryon St	Location Address
6	Postalcode	27282	Zip Code
7	distance	125	Distance from Target
8	State	NC	State Abbreviation
9	City	Charlotte	City Name

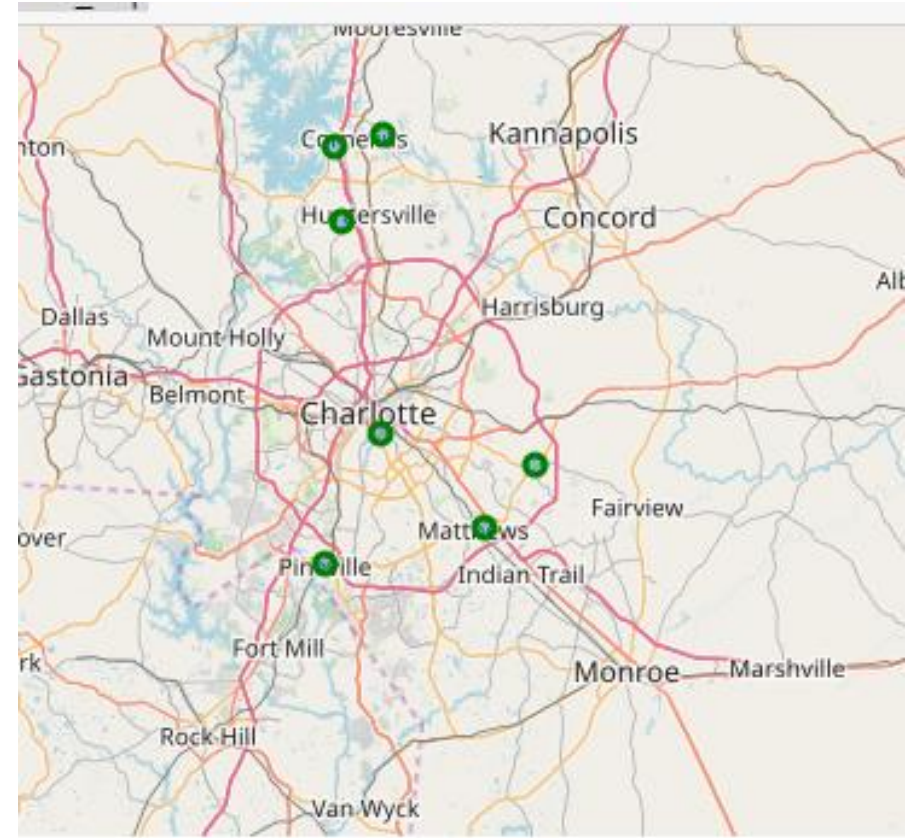
I will use the above dataset from simple maps and use the Foursquare API data to look at venues and locations in the various cities surrounding the Charlotte Metropolitan Area to solve for the 2 problems outlined in the prior section.

# METHODOLOGY

## DATA TRANSFORMATION STEPS



1. Data Import
2. Data Cleaning
3. Data Merging
4. Data Enrichment
5. Visualization Analysis



## VISUALIZATION OF CHARLOTTE AREA CITIES

# EXPLORATORY ANALYSIS

Out[27]:

	name	categories	lat	lng
0	Trader Joe's	Grocery Store	35.211608	-80.835468
1	Mama Ricotta's	Italian Restaurant	35.210081	-80.836012
2	Viva Chicken Elizabeth Avenue	Peruvian Restaurant	35.213407	-80.825966
3	Little Sugar Creek Greenway	Park	35.210849	-80.835981
4	Earl's Grocery	Deli / Bodega	35.213588	-80.826343
5	Duke Mansion	Bed & Breakfast	35.201550	-80.827493
6	Customshop	New American Restaurant	35.213828	-80.826544
7	Tropical Smoothie Cafe	Café	35.209687	-80.835677
8	Laurel Market	Wine Bar	35.202201	-80.824301
9	Sabor Latin Street Grill	Latin American Restaurant	35.216258	-80.821734

In [33]: `charlotte_venues.groupby('City').count()`

Out[33]:

	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
City						
Charlotte	15	15	15	15	15	15
Cornelius	1	1	1	1	1	1
Davidson	2	2	2	2	2	2
Huntersville	5	5	5	5	5	5
Matthews	12	12	12	12	12	12
Mint Hill	9	9	9	9	9	9
Pineville	13	13	13	13	13	13

I then created a data frame with top 100 venues by city and looked at the count of top 100 venues by city. Here we noticed that Cornelius and Davidson did not have many venues to chose from and they were both located in the north of the Charlotte Metropolitan Area. We then analyzed each city for the 46 different types of venues.

Out[35]:

	City	American Restaurant	Arts & Crafts Store	Bakery	Bar	Baseball Field	Beer Bar	Beer Garden	Big Box Store	Brewery	Café	Camera Store	Construction & Landscaping	Cosmetics Shop	Deli / Bodega	Diner	Donut Shop	Dry Cleaner	Fondue Restaurant
0	Pineville	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
1	Pineville	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
2	Pineville	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Pineville	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Pineville	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0

In [36]: `# Charlotte OneHot Dataframe Shape`  
`charlotte_onehot.shape`

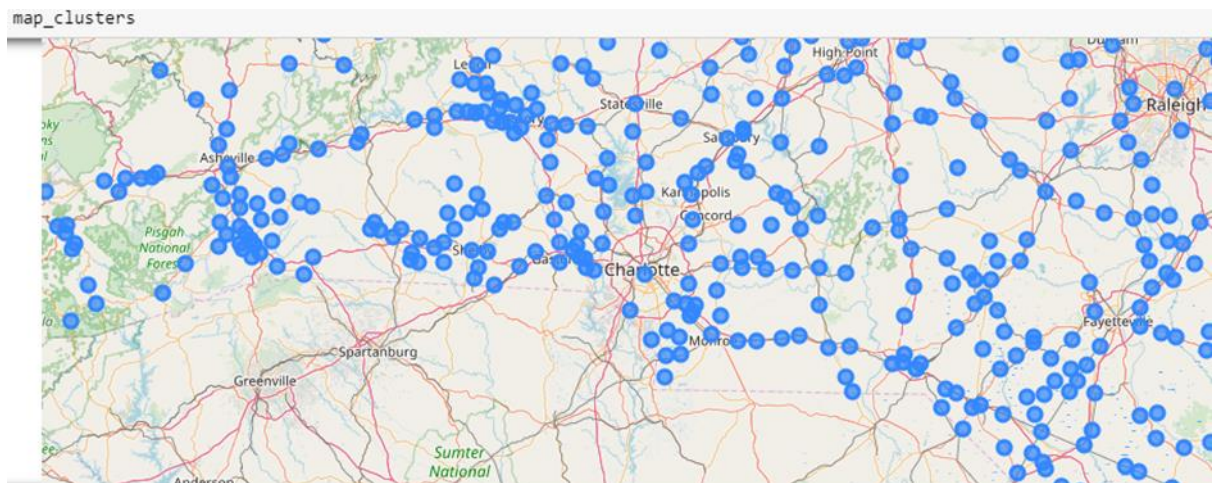
Out[36]: (57, 47)

# DEEPER DIVE

Then we looked at the top 5 venues in each city. As expected, Cornelius and Davidson did not have much to offer.

```
----Cornelius----
      venue  freq
0      Wine Shop  1.0
1 American Restaurant  0.0
2      Pool Hall  0.0
3   Indian Restaurant  0.0
4      Kids Store  0.0
```

```
----Davidson----
      venue  freq
0  Cosmetics Shop  0.5
1 Construction & Landscaping  0.5
2      Record Shop  0.0
3      Kids Store  0.0
4  Martial Arts Dojo  0.0
```



We then looked at the cluster of cities and having run the K means analysis, segmented the cities around Charlotte into 3 clusters.



# FINDINGS



## CLUSTER #1

### 5 Cities

1. Pineville
2. Huntersville
3. Matthews
4. Charlotte
5. Mint Hill

All of these mature cities are good candidate for families with ample amenities and population density.



## CLUSTER #2

### 1 City

1. Davidson

Blue Collar city with low population density. Blue collar establishments.



## CLUSTER #3

### 1 City

1. Cornelius

Up and coming city with good population density and more venues and categories. Good potential for home buyers and business alike.

## RESULTS & DISCUSSION

- Based on the above observations the following 5 suburban areas in the Charlotte Metropolitan Area are good for people looking to buy a house to get a more mature infrastructure and entertainment locations:
  - 
  - 1. Pineville
  - 2. Huntersville
  - 3. Matthews
  - 4. Charlotte
  - 5. Mint Hill
  -
- From a business perspective, the following two cities offer good business potential.
  - 
  - 1. Cornelius
  - 2. Davidson
  -
- Of the two above, **Cornelius** has a larger population base and higher density so the potential for business to succeed will be higher in Cornelius.







THANK YOU !