A Geo-Analysis

of Charlotte, North Carolina

**Peer-graded Assignment:**

**Capstone Project - The Battle of Neighborhoods**

**(Week 1)**

**Updated for**

**Week 2 of the assignment.**

Author: Amit Arya  
Publish Date: April 2, 2020

Contents

[1. Background & Problem Statement 2](#_Toc36726060)

[2. Data Description 3](#_Toc36726061)

[3. Methodology and Exploratory Data Analysis 5](#_Toc36726062)

[4. Results and Discussion 12](#_Toc36726063)

[5. Conclusion 13](#_Toc36726064)

A Geo-Analysis

of Charlotte, North Carolina

Author: Amit Arya  
Publish Date: March 31, 2020

# Background & Problem Statement

I reside in the Charlotte, North Carolina. After living here for 10 years, I moved to Pittsburgh, Pennsylvania for 5 years and have recently moved back to Charlotte about 2 years back.

Charlotte was always a vibrant metropolitan area supported by a family oriented suburban cities. Having come back after 5 years I see that the city has growth exponentially with many families preferring to move to warmer climate with good economic prospects. Charlotte fits this description.

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.

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 Description

The basic data that we need is a list of cities in the Charlotte Metropolitan Area. I performed a quick search on google and found several sites listed below:

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> 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>

The data contains the following fields that I intend to uuse.:

|  |  |  |  |
| --- | --- | --- | --- |
| # | 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 '000 |
| 11 | density | 484 | Population Density |
| 12 | timezone | America/New\_York | Time Zone |
| 13 | zips | 28585 | Zip Code |

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.

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 and Exploratory Data Analysis

(Code in purple text)

Step 1. Data Import

DataURL = 'https://raw.githubusercontent.com/amitarya76/Coursera\_Capstone/master/NC\_CITY\_DATA.csv'

Charlotte\_df = pd.read\_csv(DataURL)

Charlotte\_df.head()

The first step in the process is to import the dataset. My dataset contains all the cities in the state of North Carolina so I am going to clean it up a bit.

A screenshot of a computer

Description automatically generated

Step 2. Data Cleaning

Charlotte\_NC = Charlotte\_df[Charlotte\_df['county\_name'].str.contains('Mecklenburg')].reset\_index(drop = True)

Charlotte\_NC

In this step, I filtered the data for cities in the Mecklenburg county which contains the 7 cities adjoining the Charlotte Metropolitan Area. I then added this data to a data frame called Charlotte\_NC.

A screenshot of a computer

Description automatically generated

We then visualize the data points

charlotte\_map = folium.Map(location = [CLT\_LAT, CLT\_LON], zoom\_start = 6)

for city, lat, lng in zip(Charlotte\_NC['city'],

Charlotte\_NC['lat'],

Charlotte\_NC['lng']):

label = '{}'.format(city)

label = folium.Popup(label, parse\_html = True)

folium.CircleMarker(

[lat, lng],

radius = 4,

popup = label,

color = 'green',

fill = True,

fill\_color = '#3186cc',

fill\_opacity = 0.5,

parse\_html = False).add\_to(charlotte\_map)

charlotte\_map

A picture containing text, map

Description automatically generated

Step 3. Data Merging

url = ('https://api.foursquare.com/v2/venues/explore?client\_id={}'

'&client\_secret={}&v={}&ll={},{}&radius={}&limit={}').format(CLIENT\_ID,

CLIENT\_SECRET,

VERSION,

Explore\_Lat,

Explore\_Long,

RADIUS,

LIMIT)

Top100 = requests.get(url).json()

Top100

We then extracted the Top 100 venues for each location.

A screenshot of a social media post

Description automatically generated

And then extracted the category for each venue.

def get\_category\_type(row):

try:

categories\_list = row['categories']

except:

categories\_list = row['venue.categories']

if len(categories\_list) == 0:

return None

else:

return categories\_list[0]['name']

Step 4. Exploratory Analysis

In the exploratory analysis phase, we filtered for venue name and types and cleaned up the column names as our first step.

venue\_results = Top100['response']['groups'][0]['items']

top\_venues = json\_normalize(venue\_results)

# Filter the columns

filtered\_columns = ['venue.name', 'venue.categories', 'venue.location.lat', 'venue.location.lng']

top\_venues = top\_venues.loc[:, filtered\_columns]

# Filter the category for each row

top\_venues['venue.categories'] = top\_venues.apply(get\_category\_type, axis = 1)

# Clean all column names

top\_venues.columns = [col.split(".")[-1] for col in top\_venues.columns]

top\_venues.head(10)

**A screenshot of a cell phone

Description automatically generated**

We then created a dataframe with top 100 venues by city

**A screenshot of a cell phone

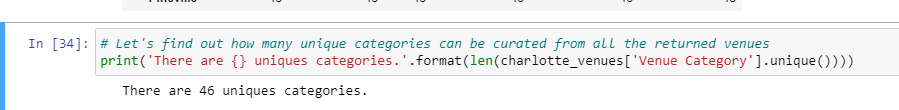
Description automatically generated**

We then looked at the count of top 100 venues by city.

**A screenshot of a cell phone

Description automatically generated**

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 found that there were 46 unique categories in our data and both the city of Cornelius and Davidson were lacking in venues.

We then analyzed each city for the 46 different types of venues.

A screenshot of a cell phone

Description automatically generated

And then checked our data frame to make sure it contained the right data.

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

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

A close up of text on a white background

Description automatically generated

We then looked at the cluster of cities.

A close up of a map

Description automatically generated

And having run the K means analysis, segmented the cities around Charlotte into 3 clusters.

Cluster #1 consisted of 5 cities that had a wide variety of location venues and had large population in them.

A screenshot of a cell phone

Description automatically generated

Cluster #2 was the city of Davidson itself which was more blue collar based on the venues in it.

A screenshot of a cell phone

Description automatically generated

Cluster #3 was the city of Cornelius, which seems more up and coming with larger population base and higher density.

A screenshot of a cell phone

Description automatically generated

# Results and Discussion

Based on the above observations the following 5 suburban areas in the Charlotte Metropolitan Ares 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.

# Conclusion

Charlotte seems to blossoming into a new Mega Metropolis. However, not every area is experiencing the same level of growth. The town of Cornelius seems to be on the cusp of breaking into a good growth momentum and may be a good area for first time home buyers but definitely a good place top open a business. The city of Davidson also seems like a good business potential opportunity given the low number of venues there. Since these two cities are next to each other, there is a good potential for Davidson to benefit from the growth in Cornelius and possible upside from trickle down economic effect to sway its fortunes in the near future.

However for now the following cities are good places for home buyers due to the abundance of population and services:

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