IBM Data Science Capstone

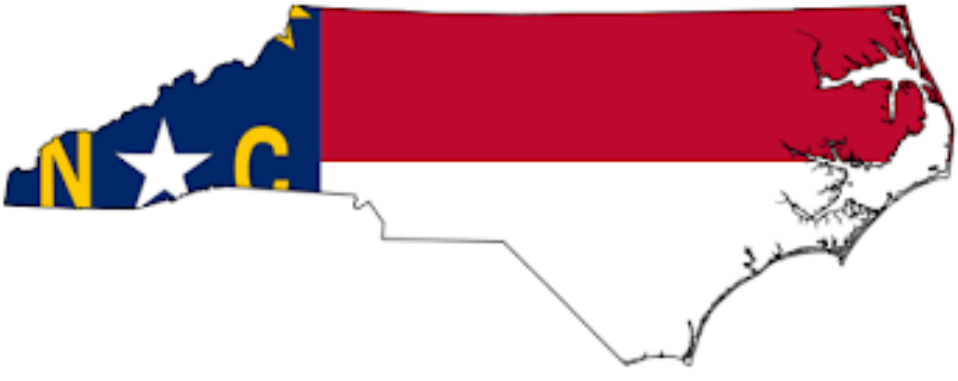
The Battle of Neighborhoods:

Finding the best place to stay.

North Carolina, United States

(Wake County Vs Mecklenburg County)

```
NC_data = df[df.State == 'NC']
NC_data.head()
```



1. INTRODUCTION

Discussion of the business problem and the audience who would be interested in this project.

1.1 Scenario and Background

I am a BI Developer and currently live within walking distance to Downtown Union Station in Denver, Colorado and have an easy commute to work with access to good public transportation. Likewise, I enjoy many amenities in the neighborhood that includes sports bars, restaurants, basketball courts, food and drink shops and entertainment areas.

I have been thinking of moving to North Carolina but I am a bit stressed towards the process to secure a comparable place to live in Raleigh or Charlotte. Therefore, I decided to apply my learned skills in the IBM Cognitive Class for Data Science to explore ways to make sure my decision can be rewarding. Of course, there are alternatives to achieve the answer using available Google and Social media tools, but I thought digging more into the raw data of the state would be rewarding as it would give me a clear picture of what i might face with decision.

On finding the best spot to live, a great deal of things are viewed as when settling on the

choice between urban areas, towns, or neighborhoods. Some of propositions incorporate, but not limited to:

***Overall Comparison*:**

- The correlation of similar components for every city, bringing about having a general diagram of the two urban areas. A portion of the prevalent components incorporate populace, typical cost for basic items, normal lease, wrongdoing rate, charge rates, and air quality.

Crime Rates:

- Here, the correlation is made to realize the crimes of two urban communities, at that point estimates them both against the national statistics.

Cost of Living and Salary:

- Comparison: This considers looking at pay rates and typical cost for basic items in urban areas for a choice to be made. It mostly takes into consideration test scores and teacher and student ratios, including the teacher's experience of the lists schools in the city of your choice.

Neighborhood Comparison:

- This looks at neighborhood comparison and helps one choose the best place to live within any given city. These sites allow you to see some pretty interesting facts about the various communities.

1.2 Problem and Purpose of this Project

The dataset incorporates the coordinates of the cities and neighborhoods in the USA. The datasets does not include the venues within these locations. With venue information, it would be easy to find out more information about the neighborhoods. For example, how many sports bars and restaurants there are, and any basketball courts or playgrounds? We could also need to find out about any banks and food and drink shops? It would better comprehend or settle on good choice about where to move or migrate to if this data was accessible.

Subsequently, the reason for this project is to, algorithmically, find a way to use the location coordinates and tag each data point into a neighborhood in ***two Counties in North Carolina- Wake County and Mecklenburg County***. The algorithm used is k-means clustering. The main idea is to determine neighborhood with venues clustered around each other so that one can make a decision on the right neighborhood to chose based on the proximity of amenities and venues to each other.

A. Clustering the Neighborhoods

The k-means clustering algorithm is an unsupervised clustering technique searches for a pre-determined number of clusters within an unlabeled multidimensional dataset. It accomplishes this using a simple conception of what the optimal clustering looks like:

- The "cluster center" is the arithmetic mean of all the points belonging to the cluster.
- Each point is closer to its own cluster center than to other cluster centers in the dataset.

The two assumptions above are presumably the basis of the k-means model.

1.3 Interested Audience

Information discovered in this project is useful to any person or entity considering moving to a major around the United States or anywhere in the world, since the approach and methodologies used here are applicable in all cases. The use of FourSquare data and mapping techniques combined with data analysis will help resolve the key questions arisen.

NB: *While all of these analyses are useful for comparing the neighborhoods, there is nothing like visiting the actual city, seeing the neighborhoods, and talking to the residents. If it's possible, an in-person visit is highly recommended before making a big move or relocating decision.*

2. DATA

Description of the data and its sources that will be used to solve the problem

The dataset for this project consists of information regarding the cities in the USA obtained from https://simplemaps.com/data/us-cities. Specifically, the data contain: **City Name, County Code, County Name, Density, Id, Latitude, Longitude, Source, State Id, State Name, and Timezone**. I used business intelligence tools for geocoding the data to obtain the correct coordinates. The data was then exported and converted into a .json, read into a pandas dataframe and sliced into Wake and Meclenburg data for use in the project.

In addition to this data, the Foursquare API will be used to collect venues near the neighborhoods for cluster analysis to be performed on the data.

US_cities data

- Since the dataframe contains information of the whole of the United States, North Carolina (NC), the State of interest was segmented from the whole and some of the column names renamed

```
df.rename(columns={'State Id': 'State', 'County Name':'County', 'County Code':'CountyCode', 'City':'Neighborhood', 'Lat': 'Latitude','Lng': 'Longitude'}, inplace=T
df.head()
```

	Number of Records	Neighborhood	County	Density	Id	Latitude	Longitude	Population	Ranking	State	State Name	Timezone	Zip
0	1	South Creek	Pierce	125	1840116412	46.9994	-122.3921	2500	3	WA	Washington	America/Los_Angeles	98580
1	1	Roslyn	Kittitas	84	1840097718	47.2507	-121.0989	947	3	WA	Washington	America/Los_Angeles	98941
2	1	Sprague	Lincoln	163	1840096300	47.3048	-117.9713	441	3	WA	Washington	America/Los_Angeles	99032
3	1	Gig Harbor	Pierce	622	1840097082	47.3352	-122.5968	9507	3	WA	Washington	America/Los_Angeles	98332
4	1	Lake Cassidy	Snohomish	131	1840116371	48.0639	-122.0920	3591	3	WA	Washington	America/Los_Angeles	98223

3. METHODOLOGY

3.1. *Exploratory Analysis*

Exploratory analysis was performed by examining tables and plots of the downloaded data. This was used to:

1. Segment the data into Cities in Wake and Mecklenburg Counties in North Carolina
2. identify missing values, verify the quality of the data
3. determine likely approaches to modelling, which might best yield to good clustering.

3.2. *Segmenting and Slicing, and visualizing the data*

An important part of cluster modelling is the careful selection of the variables of available data. A prerequisite of the study is that the foursquare API is used to collect the venue information. Hence it is very important that the dataset for this work includes the coordinates of the cities to be studied. Segmentation and slicing of the data resulted in Table 1. The subjects included in the data for analysis includes: Neighborhood Name, County Name, Density, Latitude, Longitude, and State Name.

Segmenting North Carolina Data

- Since the dataframe contains information of the whole of the United States, North Carolina (NC), the State of interest was segmented from the whole and some of the column names renamed

Figure.1

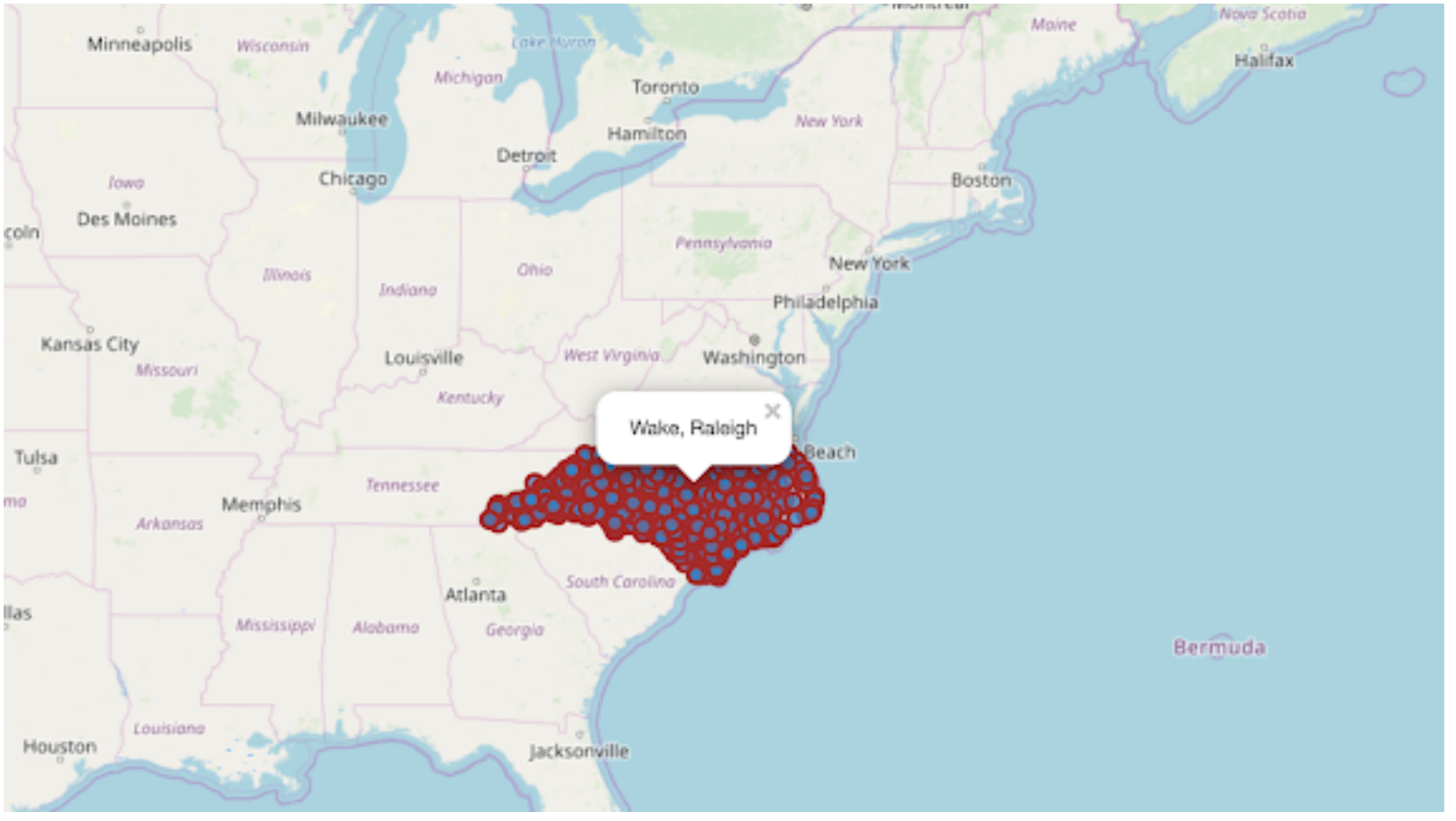


Figure.2

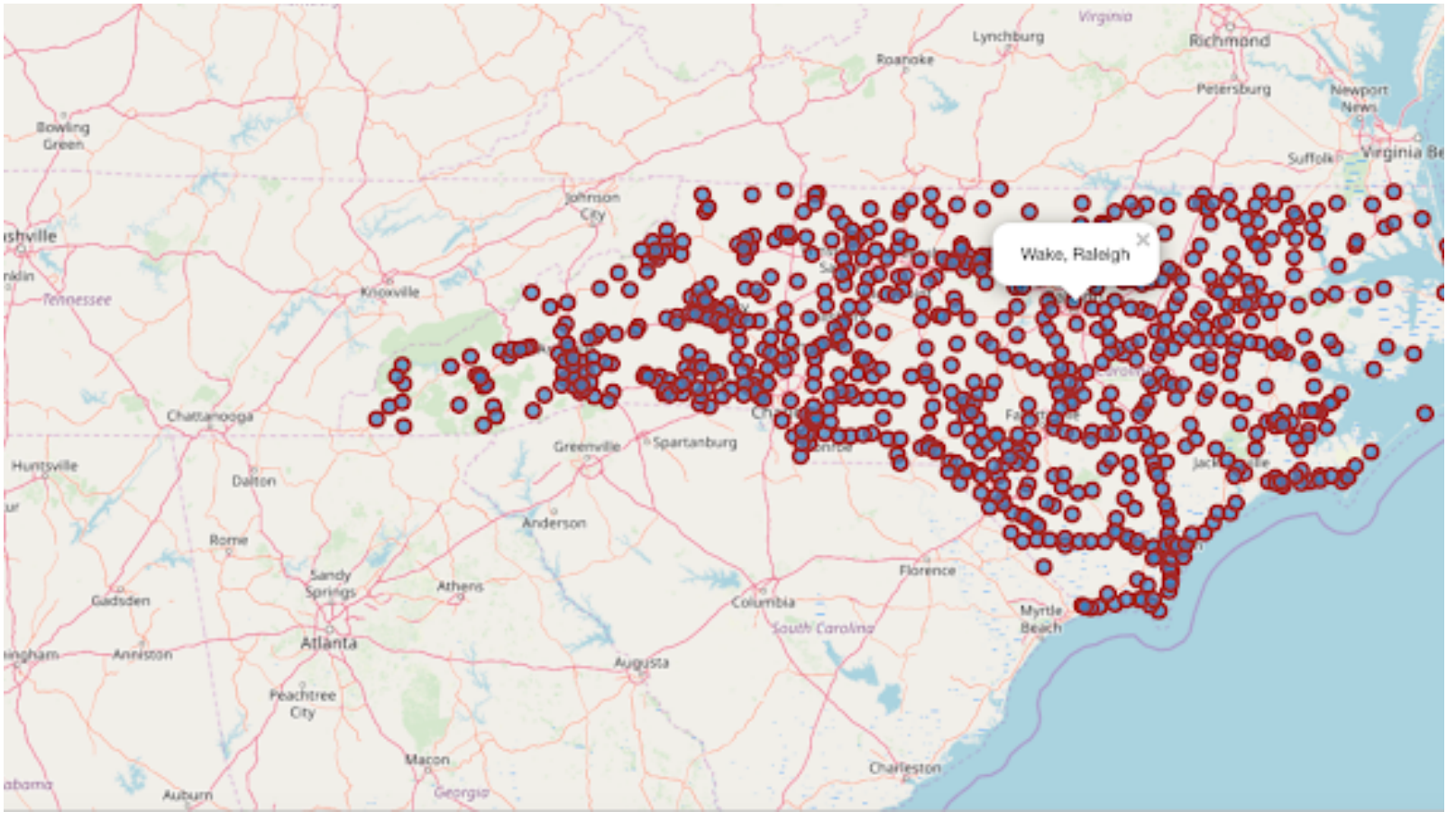


Figure.3

```
NC_data = df[df.State == 'NC']
NC_data.head(20)
```

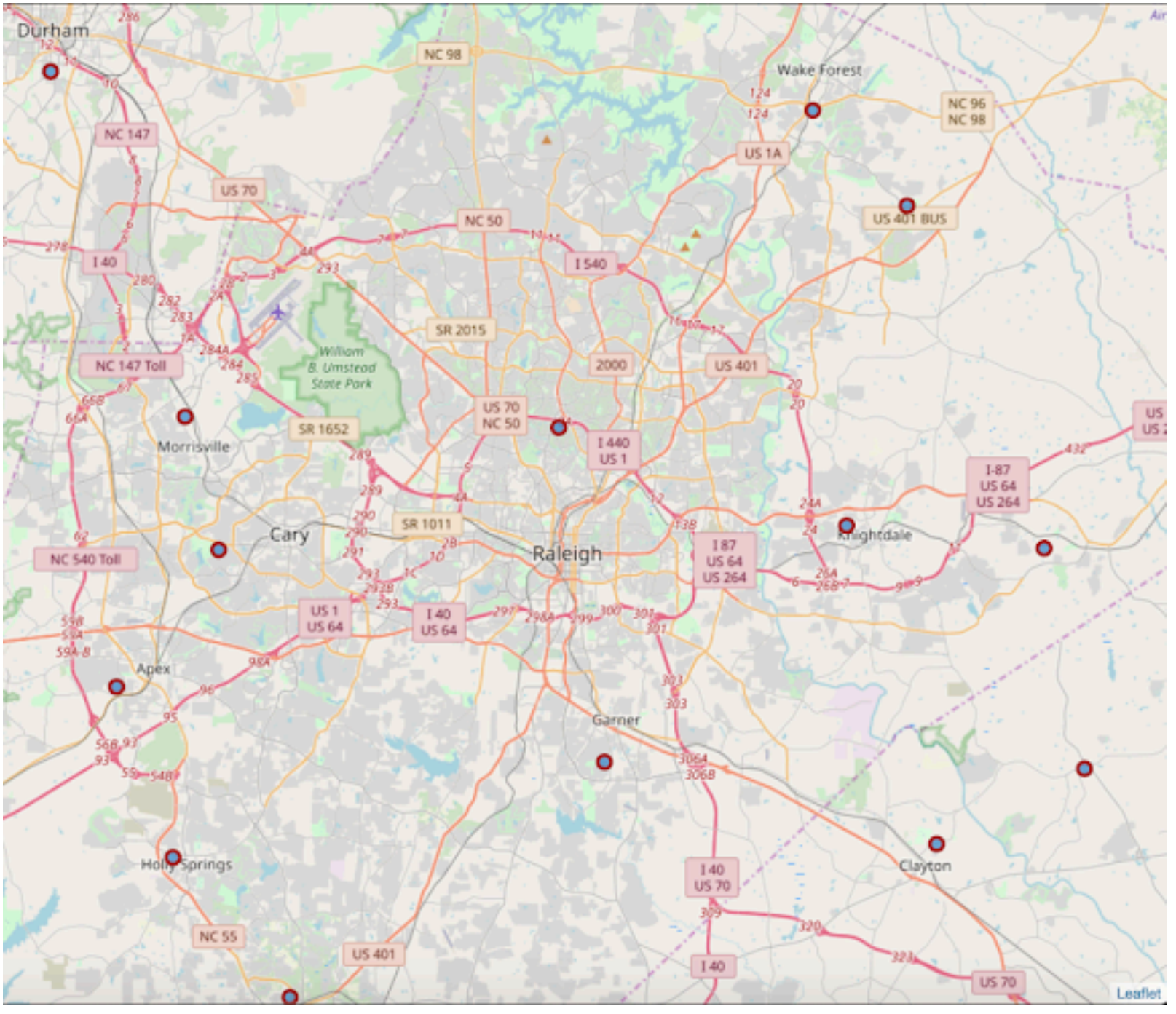
	Number of Records	Neighborhood	County	Density	Id	Latitude	Longitude	Population	Ranking	State	State Name	Timezone	Zip
6677	1	Trenton	Jones	484	1840061975	35.0640	-77.3553	285	3	NC	North Carolina	America/New_York	28585
6678	1	Stanfield	Stanly	131	1840064499	35.2336	-80.4303	1517	3	NC	North Carolina	America/New_York	28163
6679	1	Sanford	Lee	389	1840064038	35.4874	-79.1772	34557	3	NC	North Carolina	America/New_York	27330
6680	1	Cofield	Hertford	46	1840061224	36.3567	-76.9105	382	3	NC	North Carolina	America/New_York	27922
6681	1	Thomasville	Davidson	612	1840064227	35.8813	-80.0807	26615	3	NC	North Carolina	America/New_York	27360
6682	1	Aurora	Beaufort	212	1840062662	35.3030	-76.7889	512	3	NC	North Carolina	America/New_York	27806
6683	1	Casar	Cleveland	64	1840061200	35.5127	-81.6168	293	3	NC	North Carolina	America/New_York	28020
6684	1	Mebane	Alamance	633	1840063605	36.0876	-79.2756	14973	3	NC	North Carolina	America/New_York	27302
6685	1	Vass	Moore	93	1840064299	35.2533	-79.2845	774	3	NC	North Carolina	America/New_York	28394
6686	1	Forest Hills	Jackson	297	1840108305	35.2958	-83.1954	382	3	NC	North Carolina	America/New_York	28723
6687	1	Cape Carteret	Carteret	315	1840062872	34.6973	-77.0572	2062	3	NC	North Carolina	America/New_York	28584
6688	1	Lowgap	Surry	121	1840117457	36.5233	-80.8671	359	3	NC	North Carolina	America/New_York	27024
6689	1	Greenevers	Duplin	157	1840062122	34.8271	-77.9247	638	3	NC	North Carolina	America/New_York	28458
6690	1	Lillington	Harnett	300	1840063521	35.3999	-78.8140	3719	3	NC	North Carolina	America/New_York	27546
6691	1	Gorman	Durham	108	1840110868	36.0418	-78.8091	807	3	NC	North Carolina	America/New_York	27704
6692	1	Colerain	Bertie	268	1840062950	36.2015	-76.7663	183	3	NC	North Carolina	America/New_York	27924
6693	1	Liberty	Randolph	327	1840061559	35.8556	-79.5682	2659	3	NC	North Carolina	America/New_York	27298
6694	1	Bakersville	Mitchell	226	1840062677	36.0154	-82.1582	445	3	NC	North Carolina	America/New_York	28705

Folium was used to view the sliced data for both counties.

What is Folium ?

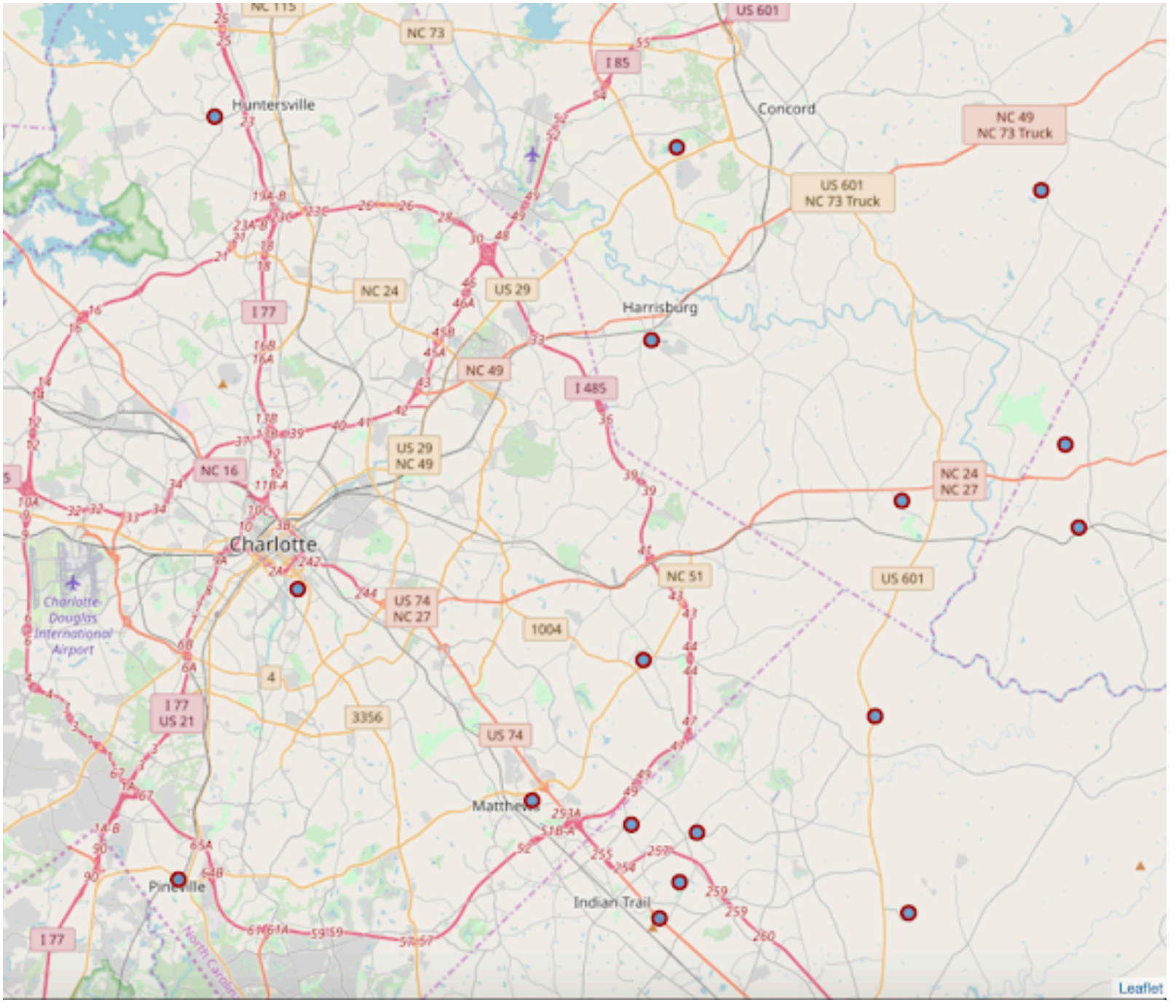
Folium is a powerful python library that builds on the data wrangling strengths of the python ecosystem and the mapping strengths of the Leaflet.js library. Generally, data is manipulating in Python, and then visualize it in on a Leaflet map via Folium.

Figure.4 Raleigh, North Carolina



Latitude and longitude values of Raleigh are 35.8324, -78.6438.

Figure.5 Charlotte, North Carolina



The geograpical coordinate of Charlotte, NC are 35.2356385, -80.8139485.

3.3. *Neighborhood Exploration and Cluster – Wake and Mecklenburg County*

For the neighborhood exploration, the Foursquare API was used. The get request was deployed on the Foursquare URL to get the category types of venues limiting the number of venues to 100 within a 500 radius.Because the aim of the project is to determine the cluster of venues in the neighborhoods, one-hot encoding was performed on the venue categories to get dummies for each venue. That is to say, the venues were coded into 0s and 1s. The result was then grouped by neighborhood by taking the mean of the frequency of occurrence of each category.

Figure.6

We use the get_category_type function to get the category types.

```
# function that extracts the category of the 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']
```

Now json is cleaned and the and structured into a pandas dataframe.

```
venues = results['response']['groups'][0]['items']

nearby_venues = json_normalize(venues) # flatten JSON

# filter columns
filtered_columns = ['venue.name', 'venue.categories', 'venue.location.lat', 'venue.location.lng']
nearby_venues =nearby_venues.loc[:, filtered_columns]

# filter the category for each row
nearby_venues['venue.categories'] = nearby_venues.apply(get_category_type, axis=1)

# clean columns
nearby_venues.columns = [col.split(".")[-1] for col in nearby_venues.columns]

nearby_venues.head(10)
```

```
Using the Foursquare API, the venues within the neighborhoods in both wake and Mecklenburg counties resulted in a vast number of outcomes. The radius defined for the venue returned venues with 81 rows and 7 columns for Wake County and 58 rows and 7 columns for Mecklenburg County. The one-hot encoding produced a total number of 106, 63 and 71, 53 rows and columns for wake and Mecklenburg counties respectively. Tables below show the results of the top 3 venues in each neighborhood venues for both wake and Mecklenburg County was grouped by neighborhood.
```

```
There are (81, 7) rows and columns of venues and neighborhoods in Wake County.
```

```
There are (58, 7) rows and columns of venues and neighborhoods in Macklenburg County
```

3.4. Cluster of Neighborhoods in Wake County.

For the clustering of venues categories in the neighborhoods, the k-means cluster was employed. to cluster the neighborhood into 4 clusters. The k-means clustering machine learning algorithm is an unsupervised clustering technique searches for a pre-determined number of clusters within an unlabeled multidimensional dataset. It accomplishes this using a simple conception of what the optimal clustering looks like:

- The "cluster center" is the arithmetic mean of all the points belonging to the cluster.
- Each point is closer to its own cluster center than to other cluster centers in the dataset.

The two assumptions above are presumably the basis of the k-means model. To be able to produce the clusters and visualize it on a map, the sliced Wake and Mecklenburg county data were merged with the grouped venue data. This was done so that the coordinates form the sliced data can aid in visualizing the clusters on a map.

4.RESULT

```
NC_data.shape
print('The dataframe has', NC_data.shape, 'rows and columns respectively.')
```

```
The dataframe has (736, 6) rows and columns respectively.
```

Figure.7 Wake County data *(wake_data)*

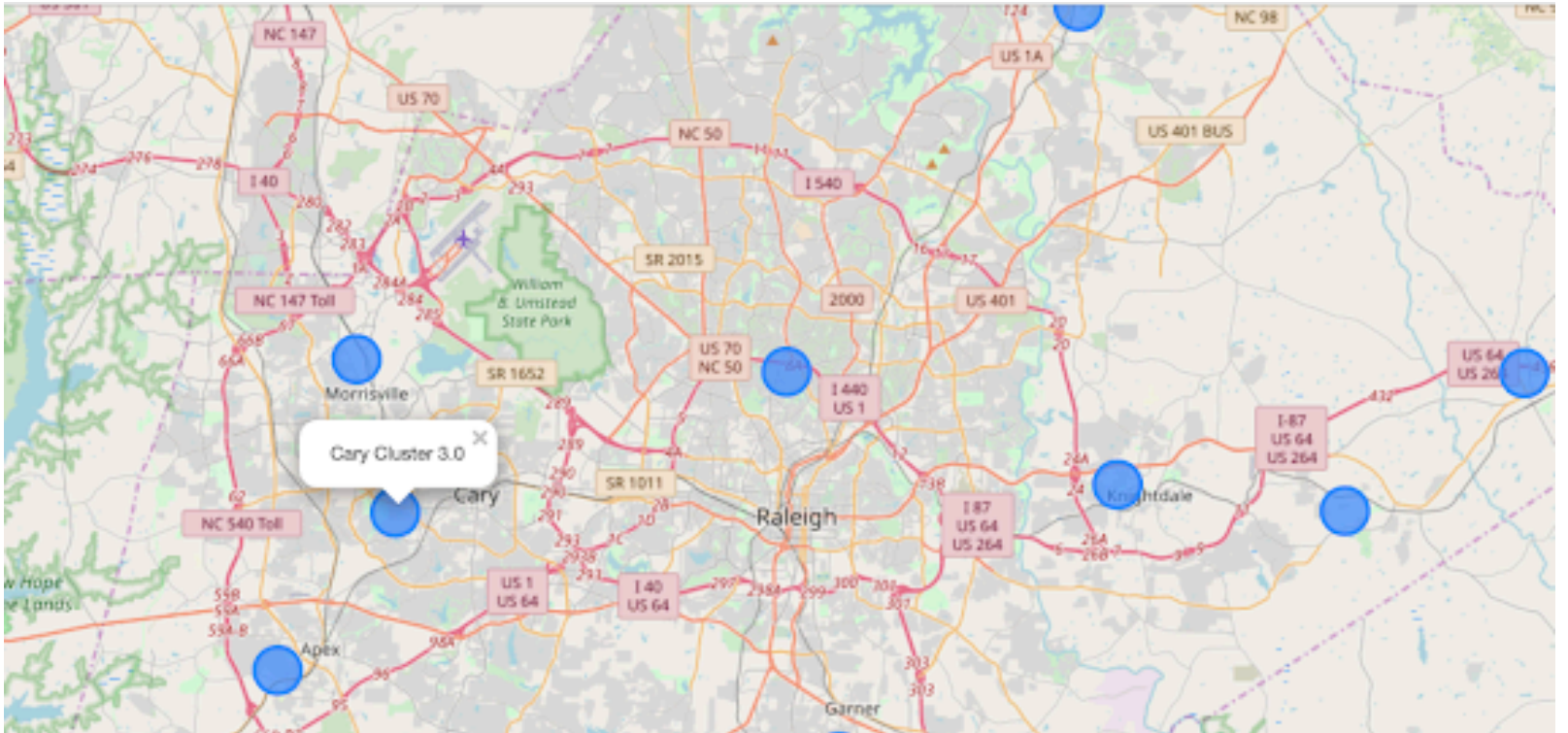
	index	Neighborhood	County	Density	Latitude	Longitude	State
0	6710	Morrisville	Wake	1177	35.8366	-78.8348	NC
1	6722	Knightdale	Wake	892	35.7917	-78.4966	NC
2	6904	Apex	Wake	1020	35.7246	-78.8698	NC
3	6917	Cary	Wake	1118	35.7817	-78.8175	NC
4	6919	Raleigh	Wake	1235	35.8324	-78.6438	NC
5	6929	Garner	Wake	714	35.6938	-78.6199	NC
6	7063	Fuquay-Varina	Wake	781	35.5961	-78.7807	NC

Figure.8 Macklenburg County data *(macklenburg_data)*

	index	Neighborhood	County	Density	Latitude	Longitude	State
0	6697	Pineville	Mecklenburg	508	35.0864	-80.8915	NC
1	6749	Davidson	Mecklenburg	846	35.4862	-80.8271	NC
2	6769	Huntersville	Mecklenburg	535	35.4057	-80.8729	NC
3	6902	Cornelius	Mecklenburg	951	35.4731	-80.8822	NC
4	7114	Matthews	Mecklenburg	724	35.1195	-80.7101	NC
5	7152	Charlotte	Mecklenburg	1081	35.2079	-80.8304	NC
6	7309	Mint Hill	Mecklenburg	426	35.1781	-80.6532	NC

Since the aim of the project is to cluster the neighborhoods, the k-means algorithm is applied to the onehot encoded venue dataset, assuming there are 4 different clusters. The tables below show the neighborhood and the cluster labels assigned to it after the k-means algorithm was applied. Cluster label ‘0’ represents the 1st cluster and ‘3’ the 4th cluster. This series of plots shows the data for each pair of

Figure.9 Neighborhood Clusters for Wake County



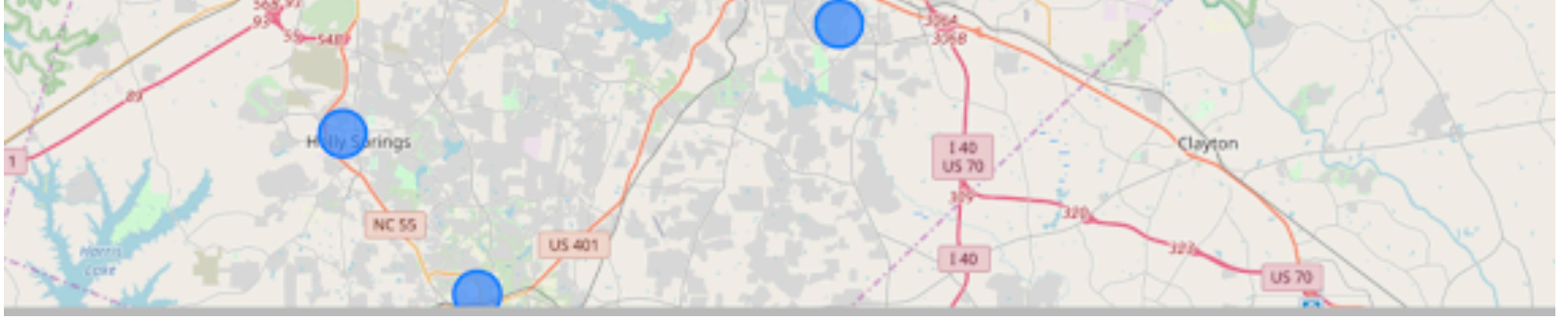


Figure.10 Neighborhood Clusters for Macklenburg Cluster

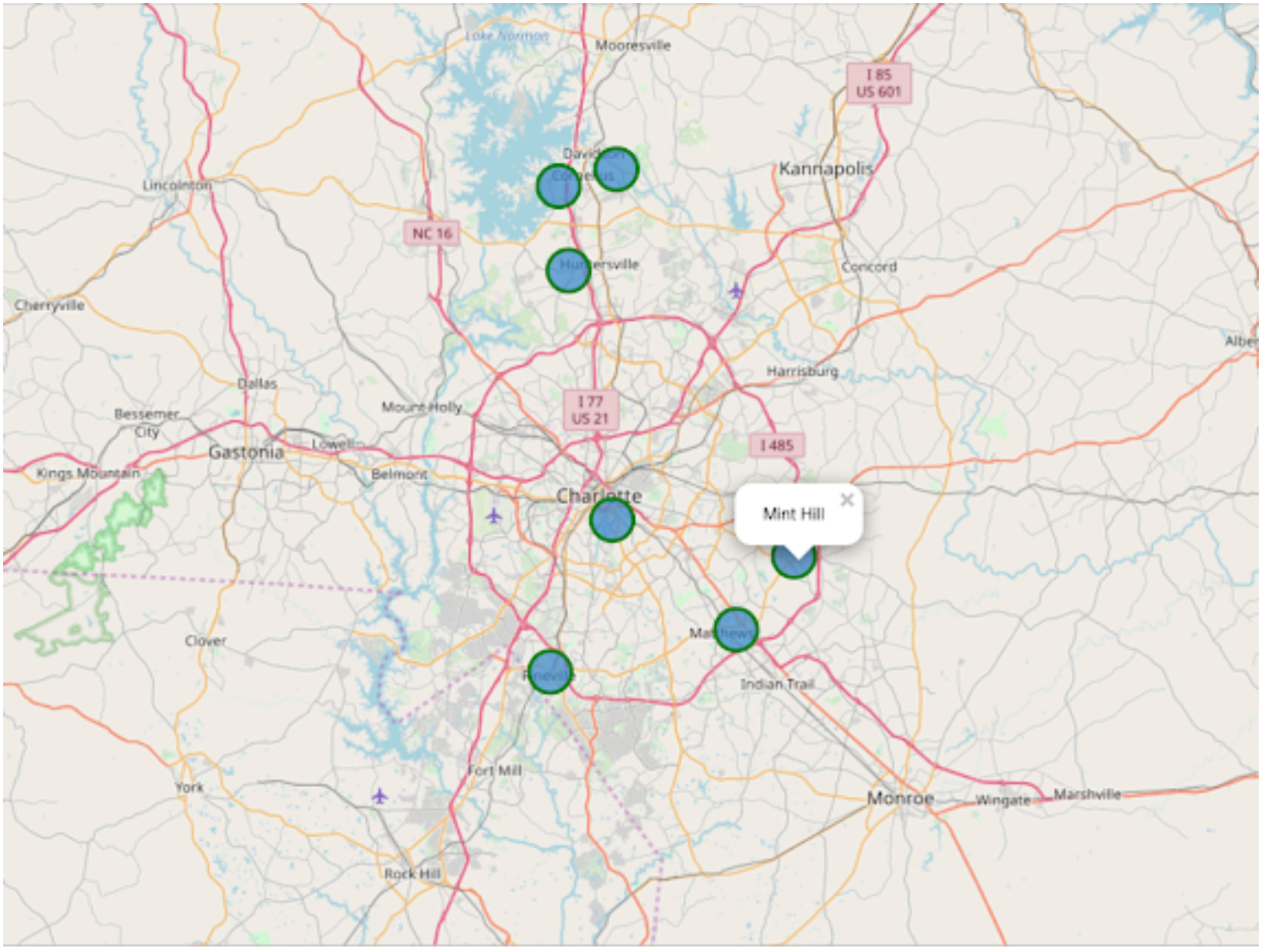


Figure.11

	Neighborhood	County	Density	Latitude	Longitude	State	Cluster Labels
6710	Morrisville	Wake	1177	35.8366	-78.8348	NC	2.0

6722	Knightdale	Wake	892	35.7917	-78.4966	NC	1.0
6904	Apex	Wake	1020	35.7246	-78.8698	NC	0.0
6917	Cary	Wake	1118	35.7817	-78.8175	NC	3.0
6919	Raleigh	Wake	1235	35.8324	-78.6438	NC	0.0
6929	Garner	Wake	714	35.6938	-78.6199	NC	0.0
7063	Fuquay-Varina	Wake	781	35.5961	-78.7807	NC	0.0
7084	Wake Forest	Wake	995	35.9634	-78.5139	NC	0.0
7108	Zebulon	Wake	436	35.8318	-78.3162	NC	0.0

Examining each cluster for the various neighborhoods in the analyzed counties, it was determined that some discriminating venue categories were distinguished each cluster. Based on the defining categories the following names were assigned to each cluster. Since 10 common venues were defined in this work, the assigned names were based only on the 1 st common venues for ease of name assignment.

From the exploratory analysis, it seems a lot of the neighborhoods are in the cluster for both Wake and Mecklenburg County. When we look at the clusters for Wake County, it becomes clear that the first two most common venues in the neighborhoods contain a lot of mixed amenities: Scenic Lookouts, Baseball field, Mexican Restaurant, Park, Beer Store, basketball Court, Italian Restaurant, Fast Food Restaurant, Mobile Phone Shop, electronic store, Yoga Studio, Smoothie Shop. Again, looking at that of Mecklenburg County, we have: Chinese Restaurant, Beer Garden, Gym, Ice Cream Shop, Pool, Bakery, Wine Shop, Women's Store,, Brewery, and Pool Hall.

Figure.12 Wake County most common venues

	Neighborhood	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 Most Common Venue	10th Most Common Venue
0	Apex	Scenic Lookout	Mobile Phone Shop	Basketball Court	Farm	Yoga Studio	Gym	Gift Shop	Gas Station	Furniture / Home Store	Fast Food Restaurant
1	Cary	Baseball Field	Electronics Store	Pool	Yoga Studio	Department Store	Home Service	Gym	Gift Shop	Gas Station	Furniture / Home Store
2	Fuquay-Varina	Mexican Restaurant	American Restaurant	Sandwich Place	Gift Shop	Donut Shop	Indian Restaurant	Discount Store	Mobile Phone Shop	Paper / Office Supplies Store	Cosmetics Shop
3	Garner	Park	Yoga Studio	Sports Bar	Dance Studio	Coffee Shop	Cosmetics Shop	Beer Store	Donut Shop	Gym	Gift Shop
4	Holly Springs	Beer Store	Sporting Goods Shop	Pharmacy	Pizza Place	Yoga Studio	Department Store	Gym	Gift Shop	Gas Station	Furniture / Home Store
5	Knightdale	Park	Yoga Studio	Department Store	Home Service	Gym	Gift Shop	Gas Station	Furniture / Home Store	Fast Food Restaurant	Farm
6	Morrisville	Basketball Court	Convenience Store	Yoga Studio	Department Store	Home Service	Gym	Gift Shop	Gas Station	Furniture / Home Store	Fast Food Restaurant
7	Raleigh	Italian Restaurant	Smoothie Shop	Big Box Store	Coffee Shop	Furniture / Home Store	Gym	Hotel	Women's Store	Mobile Phone Shop	Movie Theater
8	Wake Forest	Italian Restaurant	Supermarket	Pharmacy	Dance Studio	Gym	Gift Shop	Gas Station	Furniture / Home Store	Fast Food Restaurant	Farm
9	Zebulon	Fast Food Restaurant	Pizza Place	Department Store	Sandwich Place	Steakhouse	Optical Shop	Pharmacy	Chinese Restaurant	Home Service	BBQ Joint

Figure13. Macklenburg County most common venues

	Neighborhood	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 Most Common Venue	10th Most Common Venue
0	Charlotte	Bakery	Women's Store	Residential Building (Apartment / Condo)	Donut Shop	Ice Cream Shop	Indian Restaurant	Mobile Phone Shop	Camera Store	Park	Pizza Place
1	Cornelius	Wine Shop	Women's Store	Coffee Shop	Garden Center	Furniture / Home Store	Food Truck	Fondue Restaurant	Dry Cleaner	Donut Shop	Diner
2	Davidson	Cosmetics Shop	Construction & Landscaping	Women's Store	Coffee Shop	Garden Center	Furniture / Home Store	Food Truck	Fondue Restaurant	Dry Cleaner	Donut Shop
3	Huntersville	American Restaurant	Brewery	Kids Store	Optical Shop	Pool	Basketball Court	Cosmetics Shop	Food Truck	Fondue Restaurant	Bakery
4	Matthews	Furniture / Home Store	Gym	Baseball Field	Basketball Court	Big Box Store	Breakfast Spot	Construction & Landscaping	Cosmetics Shop	Dry Cleaner	American Restaurant
5	Mint Hill	Gym	Coffee Shop	Bakery	Bar	Beer Bar	Shipping Store	Sandwich Place	Restaurant	Pizza Place	Museum
6	Pineville	Garden Center	Gas Station	Toy / Game Store	Indian Restaurant	Shoe Store	Beer Bar	Beer Garden	Latin American Restaurant	Diner	Food Truck

Figure.14 Macklenburg County Venue Count Bar graph

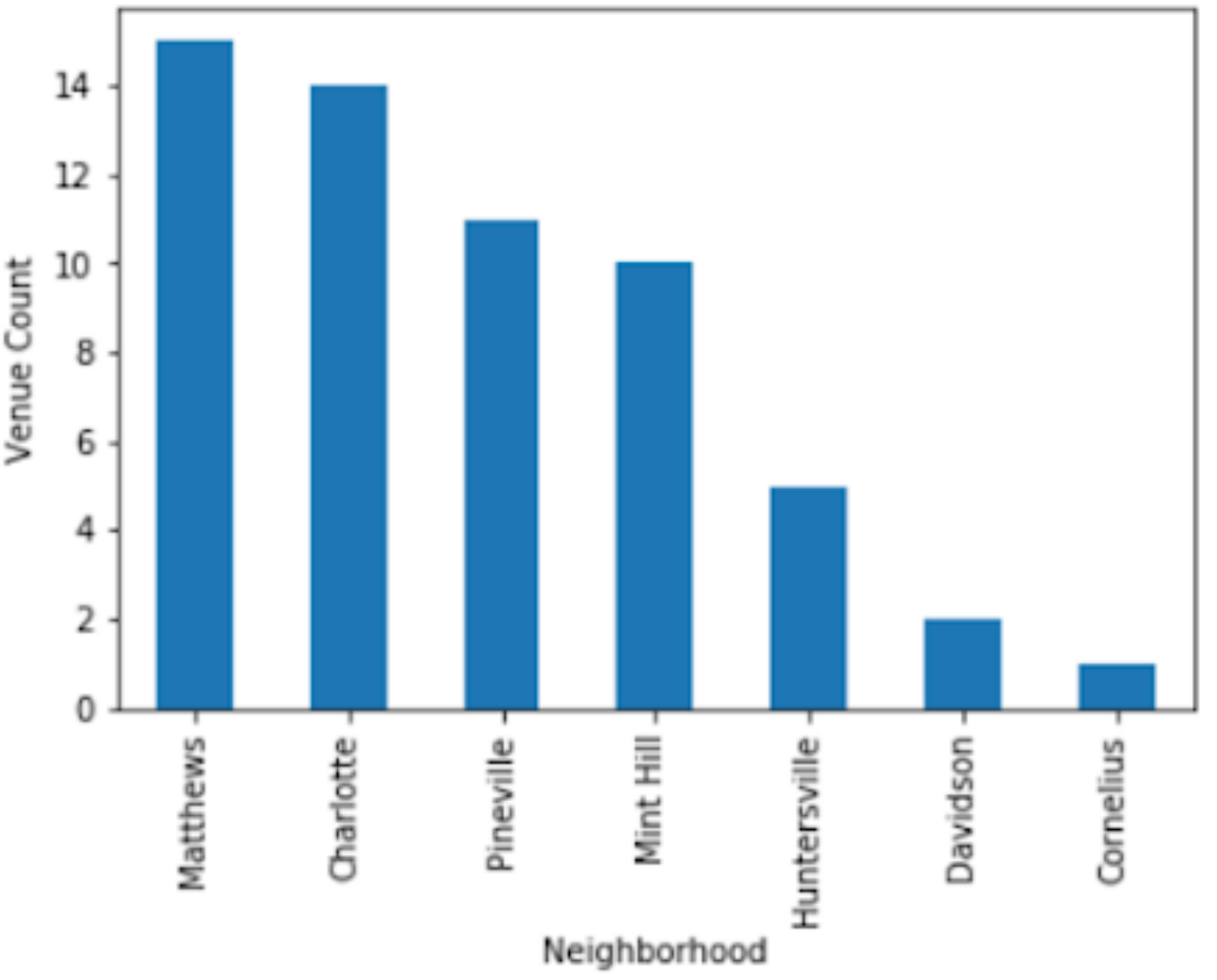


Figure.15 Macklenburg County Venue Count Clusters

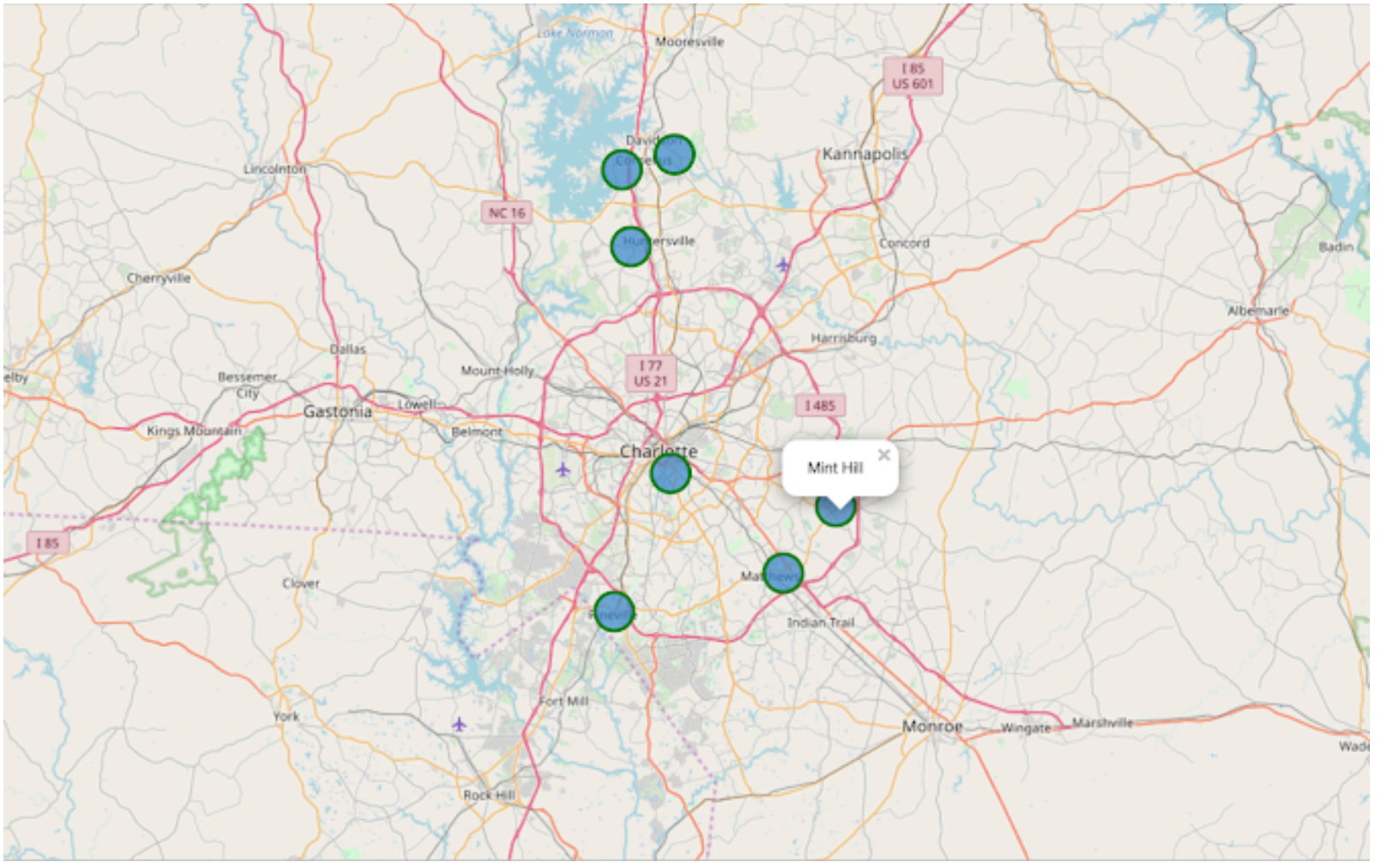


Figure.16 Venues in Macklenburg County

```
print('There are', mecklenburg_venues.shape, 'rows and columns of venues and neighborhoods in Macklenburg County.')

mecklenburg_venues.head(10)
```

```
There are (58, 7) rows and columns of venues and neighborhoods in Wake County.
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Pineville	35.0864	-80.8915	Pintville Craft Beer	35.085629	-80.891111	Beer Garden
1	Pineville	35.0864	-80.8915	Kit's Trackside Crafts	35.085960	-80.891085	Beer Bar
2	Pineville	35.0864	-80.8915	OooWee BBQ	35.084593	-80.890404	Food Truck
3	Pineville	35.0864	-80.8915	Zafran Kabab Palace	35.087350	-80.886476	Indian Restaurant
4	Pineville	35.0864	-80.8915	The Dive-N Diner	35.085959	-80.887051	Diner
5	Pineville	35.0864	-80.8915	Rite Aid	35.085890	-80.886333	Pharmacy
6	Pineville	35.0864	-80.8915	retro reboot	35.085282	-80.890522	Toy / Game Store
7	Pineville	35.0864	-80.8915	Shell	35.084880	-80.886362	Gas Station
8	Pineville	35.0864	-80.8915	Grower's Outlet	35.088002	-80.887929	Garden Center
9	Pineville	35.0864	-80.8915	Delicias Colombianas	35.087483	-80.886384	Latin American Restaurant

Figure.17 Wake County Venue Count

name	categories	lat	lng
bartaco North Hills	Taco Place	35.836575	-78.643819
Mura Japanese Restaurant	Sushi Restaurant	35.836640	-78.643369
Target	Big Box Store	35.836292	-78.642610
Renaissance Raleigh North Hills Hotel	Hotel	35.836195	-78.643695
Mia Francesca Trattoria	Italian Restaurant	35.836218	-78.643951
Starbucks	Coffee Shop	35.836133	-78.642261
Natural Body Spa & Shop	Spa	35.836444	-78.643495
Anthropologie	Women's Store	35.836825	-78.643161
Regal Cinemas North Hills 14	Movie Theater	35.836608	-78.642118
Fitness Connection - North Hills	Gym	35.836327	-78.643142

```
plot1=wake_venues["Neighborhood"].value_counts().plot(kind = "bar")
plot1.set_xlabel('Neighborhood')
plot1.set_ylabel('Venue Count')
```

```
Text(0, 0.5, 'Venue Count')
```

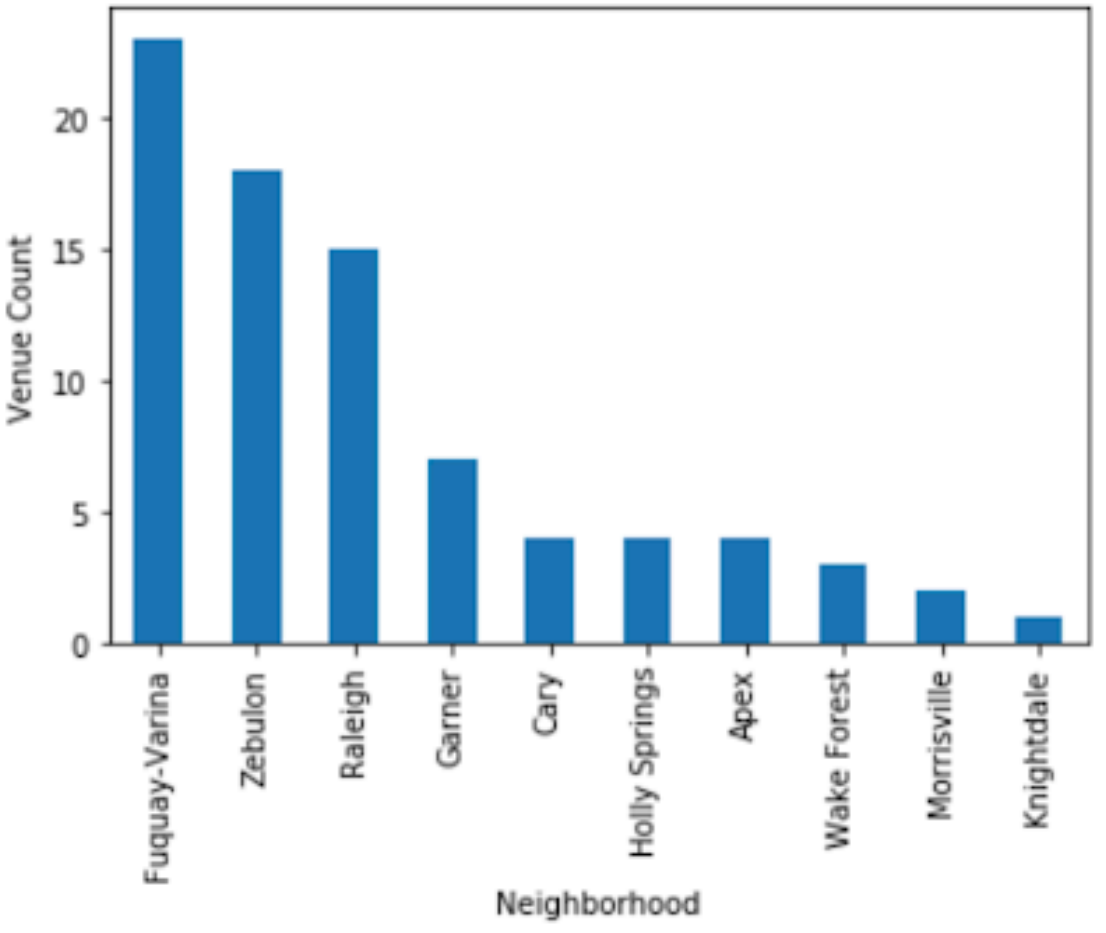


Figure.18 Wake County Venues

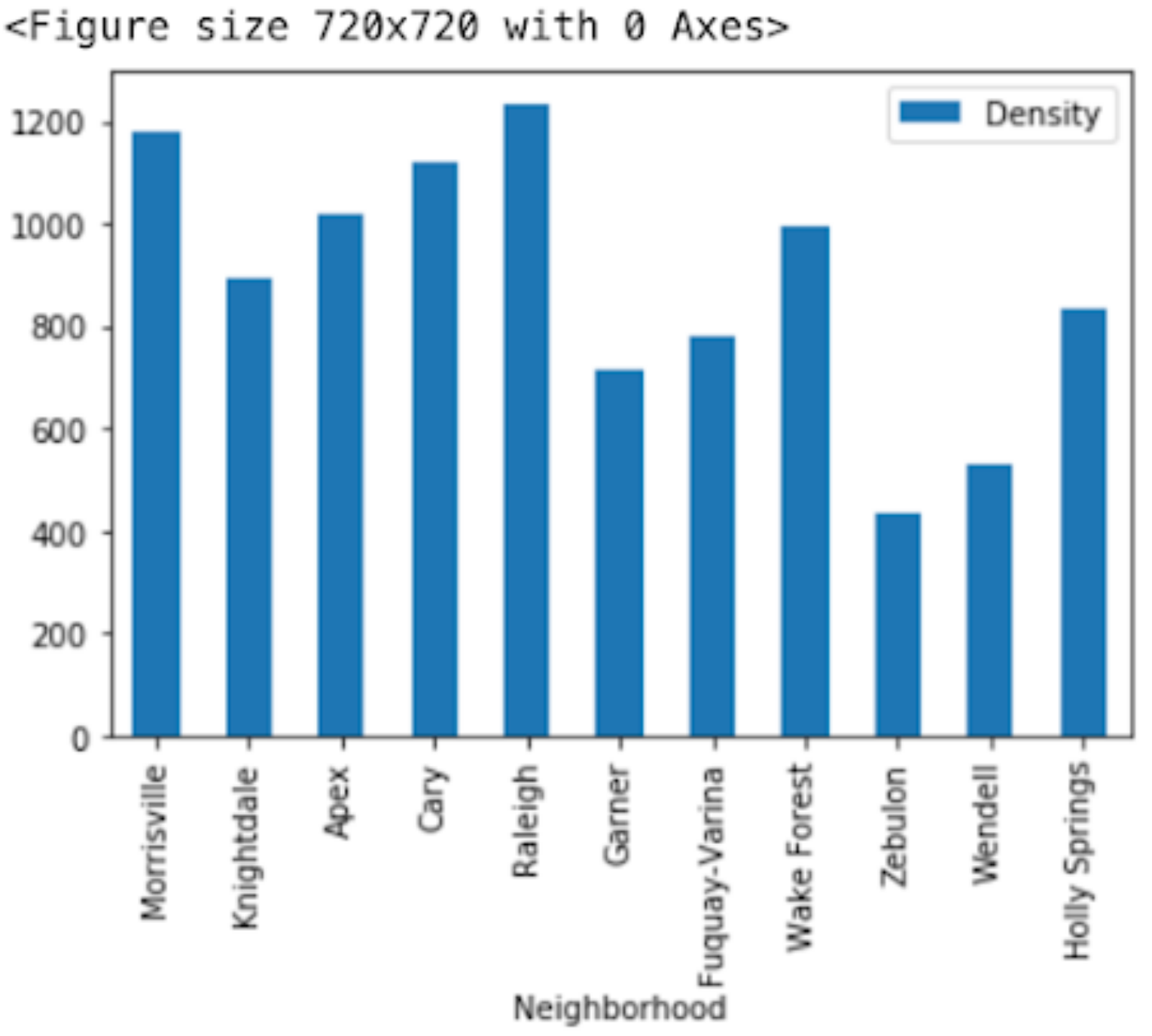
```
print('There are', wake_venues.shape, 'rows and columns of venues and neighborhoods in Wake County.')

wake_venues.head(10)
```

```
There are (81, 7) rows and columns of venues and neighborhoods in Wake County.
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Morrisville	35.8366	-78.8348	Cedar Fork Community Center	35.837397	-78.839706	Basketball Court
1	Morrisville	35.8366	-78.8348	Triangle Mini Mart	35.839855	-78.831110	Convenience Store
2	Knightdale	35.7917	-78.4966	Mingo Creek Park	35.790788	-78.494254	Park
3	Apex	35.7246	-78.8698	STMM Basketball Gym	35.725050	-78.874514	Basketball Court
4	Apex	35.7246	-78.8698	AT&T	35.723800	-78.874710	Mobile Phone Shop
5	Apex	35.7246	-78.8698	The Creek	35.727985	-78.872740	Scenic Lookout
6	Apex	35.7246	-78.8698	Prince Farms	35.720968	-78.867303	Farm
7	Cary	35.7817	-78.8175	Ibm of Highland Oaks	35.784714	-78.816842	Electronics Store
8	Cary	35.7817	-78.8175	Bond Park Field 1	35.781055	-78.821230	Baseball Field

Other factors which may determine the best place to stay may include population density in an area in relation to the availablr venues. One may need to compare the ratio.

Figure.19 Wake County density bar graph and dataframe



	index	Neighborhood	County	Density	Latitude	Longitude	State
0	6710	Morrisville	Wake	1177	35.8366	-78.8348	NC
1	6722	Knightdale	Wake	892	35.7917	-78.4966	NC
2	6904	Apex	Wake	1020	35.7246	-78.8698	NC
3	6917	Cary	Wake	1118	35.7817	-78.8175	NC
4	6919	Raleigh	Wake	1235	35.8324	-78.6438	NC
5	6929	Garner	Wake	714	35.6938	-78.6199	NC
6	7063	Fuquay-Varina	Wake	781	35.5961	-78.7807	NC
7	7084	Wake Forest	Wake	995	35.9634	-78.5139	NC
8	7108	Zebulon	Wake	436	35.8318	-78.3162	NC
9	7289	Wendell	Wake	530	35.7823	-78.3959	NC
10	7395	Holly Springs	Wake	834	35.6540	-78.8403	NC

5. DISCUSSION

So, the question is where should someone considering relocating move to a new neighborhood given the choice between Wake and Mecklenburg County?Well, by looking at the two neighborhood maps, it shows that a foodie would choose to live in the Wake county since there are a lot of restaurant and food outlets in the neighborhoods in Wake County. If you enjoy outdoors like scenic viewing, outdoor sports, choosing Wake county as your place of relocation would be a good idea. A couple breweries in the Mecklenburg county would invite a bibulous person.

However, decision is left to the individual looking at relocating to make. But in general, though all these analyses are useful, there is nothing like visiting the actual city, seeing the neighborhoods, and speaking with residents. If it's possible, an in-person visit is highly recommended before making a big move.

6. CONCLUSION

The aim of this work is to provide the necessary amenities to help people decide on the best to live or relocate to should they think about that. Using public datasets obtained from the web, i was able to, the address a few factors by analyzing the neighborhoods within two major Counties in North Carolina, Wake and Mecklenburg by based on the spatial distribution of venues in the chosen neighborhoods. My analysis has shown that using folium- python library that assists in building a quick interactive data visualization and Foursquare API for neighborhood data collection, it is feasible to cluster neighborhoods cities data based on known and accepted machine learning techniques –K-Means Algorithm.

These results must be considered bounded in scope to the dataset used, since there is no information available as to its provenance. Such results will be of interest to people or citizens whose aim to compare different neighborhoods when thinking about relocation or vacationing in a different environment, considering the ease of accessing numerous venues within a clustered setting. There certainly is lot of room for improvement. For example, obtaining more than the current neighborhood locations to analyze and cluster a wide expanse of geographical setting. We may also use and analyze crime data – which is publicly available for his two counties - to help to provide enough room for decision making with regards to choosing a location to relocate.This information may be extremely useful because we certainly don’t want to live in a crime infested neighborhood. Though the approach used here may not be vigorous enough, it nevertheless showcases the usefulness of neighborhood data analysis.

Git Link - Ognash/TheBattleOfNeighborhoods_FINAL.ipynb