# **Battle of the Neighborhoods Next Charlotte Neighborhood COVID 19 Testing Center**

## **Introduction/Business Problem**

As in all communities in the United States and around the world, COVID-19 has had a negative impact to the communities and neighborhoods in Charlotte, North Carolina. In this fall and winter season, Charlotte has seen an influx of COVID-19 cases.

Early on, the city of Charlotte had a shortage of testing availability around the city. Over the pandemic, this has improved; however, the people of Charlotte continue to have difficulty accessing testing around the city.

**Problem Statement**

With the opportunity to improve testing accessibility around the city, city of Charlotte and Mecklenburg County want to provide an additional testing center to members of the community and would like to use data analysis used to aid in their decision.

Criteria of the analysis include the location of the current clusters of COVID-19 cases within the city of Charlotte and the distance of the nearest testing centers in relation to the COVID-19 clusters. In addition, Foursquare API data was also used to aid in determine in population served in the neighborhoods as well as inferring the type of demographics in the neighborhoods (e.g., more affluent vs. less affluent). The neighborhood with the highest clusters and testing centers that are furthest away will be provided the next testing center.

## **Description of the Data Used**

Three sources of data will be used for this project:

* North Carolina Department of Health and Human Services -->COVID-19 Data--> About the Data-->Zip Code Cases and Death Table <https://covid19.ncdhhs.gov/dashboard/about-data>
* North Carolina Department of Health and Human Services COVID-19 --> Test Site Finder <https://covid19.ncdhhs.gov/about-covid-19/testing/find-my-testing-place/test-site-finder>
* FourSquare Locator API Data

As the data from the North Carolina Department of Health and Human Services is web data, Beautiful Soup was used to scrape data from the North Carolina Department of Health and Human Services web pages and import into Python. Venue data from FourSquare imported into Python through API.

After the project was started, it was determined that geo-spatial data was needed to map the Zip Code Cases and Death Table as well as the testing centers. For the testing centers, obtaining the geospatial data was achieved through google search. The data was then appended to the testing center data frame. Adding the geo-coordinates data to the Zip Code and Death Table was achieved through going to the website, <https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/?rows=77>, and importing the csv for Charlotte zip codes. The geo-coordinate data was then merged with the Zip Code Cases and Death data frame.

The Cases and Death by Zip Code Table data, from the North Carolina Department of Health and Human Services, needed a lot of cleaning before it could be used. The table, as given, had the labels “Cases”, “Cases per 10,000 Residents”, “Cases per 100,000 Residents” and Deaths embedded in a column named, “Measured Names”. The data by zip code could not be used in that state. The unstack function in pandas was used to create the data frame “zip\_loc”, which was used to in folium to plot the cases and deaths in a Charlotte map.

The Testing Center data also was not clean and able to be used for analysis in its given state. The address data had to be manually broken apart and then geo-coordinates for the testing centers had to be appended to the data frame. This was achieved by importing the data into github and cleaning it there.

The FourSquare data was used to create the venue data frame and was easily cleaned using the process used while in this capstone course.

## **Methodology**

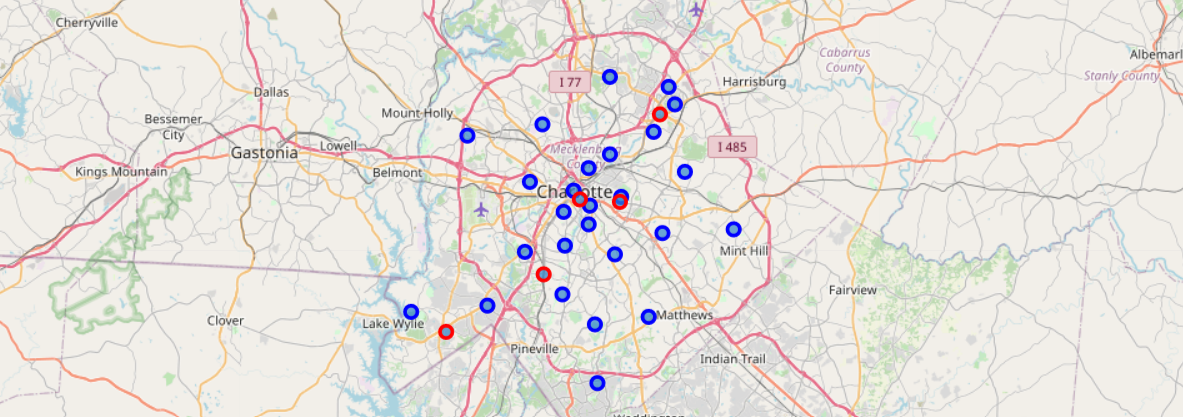
The first step in the analysis was to observe how the data points from the Cases and Deaths by Zip Code data frame in the Charlotte map. Then, the testing center data (red) overlayed the plots from the Cases and Deaths table (blue) so that the testing center locations could be observed in relation to the positive test cases plotted in the map.

The next step in the process was to review the Foursquare API data in relation to the zip codes. The number of venues in the areas could help determine whether or not there is a greater population density in some areas versus others. Anecdotally, it is known that the testing centers currently in place does serve the populations in Charlotte with greater densities as it they are closer to the city center.

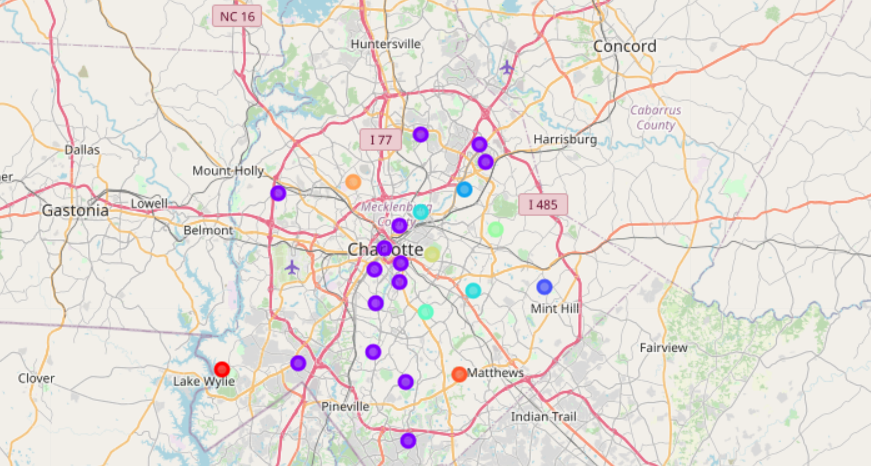
The FourSquare data for Charlotte was used to segment and cluster, using kmeans clustering in s cikit-learnin the neighborhoods by venues and Cases and Deaths by Zip Code. There were 5 clusters created as a result.

## **Observations and Results**

It was observed in these data plots that there are few testing centers in relation to the positive cases in the city. While it is obvious that people well outside the testing center area are somehow getting tested to produce the positive results, this observation does suggest that adding testing centers around the city would make sense for the people of Charlotte (see map below).

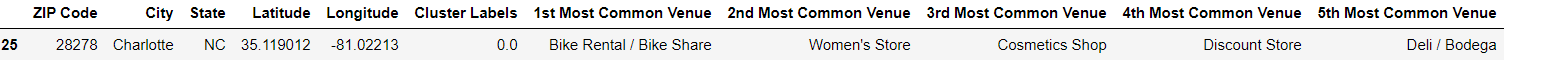


After the FourSquare data was segmented and clustered with the Cases and Deaths by Zip Code Table, the clusters were then mapped as shown below.



Next the clusters were closely examined to make inferences about the neighborhoods and the needs for accessibility to COVID testing. The results are below:

**Southwest Charlotte-Red Cluster**



The lack of restaurants and venues suggest that this is not an area where there are many people gathering. It is also not a very affluent part of the city, but more of a lower middle class area.

**South Charlotte/University City-Purple Cluster**



This area has a high number of restaurants and coffee shops. This cluster also contains very affluent neighborhoods in Charlotte.

**Mint Hill-Blue Cluster**



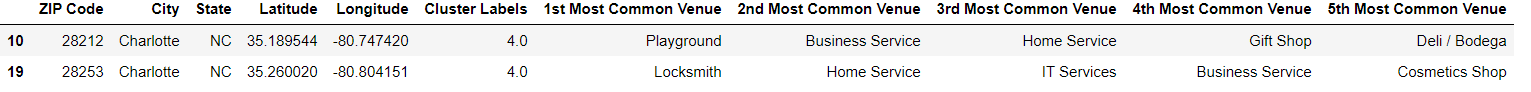
This area does not offer many venues that would suggest a high level of gathering or possibly residing in. This area has some affluence, but it has mainly lower- to middle-class people that reside here.

**North Charlotte-Light Blue Cluster**



This neighborhood also does not offer many venues that suggest a highlevel of gathering. This area is known to have a higher level of unemployment than other areas of the city, while still considered middle class.

**East Charlotte/North Charlotte-Light Green**



This neighborhood also does not suggest a high level of gathering. It also has a higher level of unemployment and contains lower middle class and poverty levels of the city.

## **Recommendation**

As the city of Charlotte is interested in only adding one testing center, my recommendation is that the city focuses on the area of the city with the least accessibility and affluence, as this population is likely the neediest of the entire population. The North Charlotte section of the city does already have a testing center, but the East Charlotte portion of the cluster does not. With that in mind, I recommend that East Charlotte have the next testing center. Those residents would be better served and more insights into the spread of the disease can be gained.

### **Conclusion**

Adding an additional testing center to East Charlotte would allow for the city of Charlotte to serve that community and provide opportunity to the city to engage that community more. COVID-19 is a serious disease and it is important for the Charlotte communities be knowledgeable about the spread in their neighborhoods so they can take steps necessary to prevent it spreading further.