**IBM Data Science Capstone**

**IBM Applied Data Science Coursera**

**Taking a Trip to Charlotte, NC**

By: Adam Herrman

Jan 2021



1. **Introduction:**

Charlotte is the most populous city in the U.S. state of North Carolina. Located in the Piedmont, it is the county seat of Mecklenburg County. Between 2004 and 2014, Charlotte was ranked as the country's fastest-growing metro area, with 888,000 new residents. Based on U.S. Census data from 2005 to 2015, Charlotte tops the U.S. in millennial population growth. It is the second-largest city in the southeastern United States, just behind Jacksonville, Florida. It is the third-fastest-growing major city in the United States with a huge variety of shops and restaurants, thus deciding where to visit is no easy task!

1. **The Problem:**

We are taking a trip to Charlotte, NC. Time is valuable, so we will want to find the optimal spot to visit. Which is the area with the highest density of restaurants within our price range. We will also want to find the venues in this area with the highest ratings and most likes, serving coffee, lunch, dinner, and drinks respectively, so we will know which venues to visit. Finally we will want a map to be able to easily find the locations.

1. **The Target Audience:**

Any Charlotteans looking for an efficient night on the town!

1. **The Data:**

We will use the Foursquare API to collect data for the venues of Charlotte NC. Foursquare

has one of the largest databases of 105+ million places and is used by over 125,000 developers.

Foursquare API will provide many categories of the venue data, we are particularly interested in venues providing: coffee, lunch, dinner, and drinks. This is a project that will make use of many data science skills, from working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). We will need:

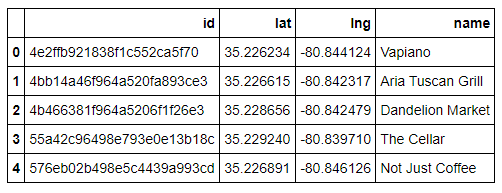
* The geographical coordinates of Charlotte NC
* The geographical coordinates of restaurants and shopping in Charlotte NC. This will be returned by a Foursquare API “explore” query.
* The details, such as, likes, rating, and categories of venues within regions of Charlotte NC. This will be returned by Foursquare API “details” queries for each venue.

1. **Methodology:**
   1. **Exploratory Data Analysis**

We start by getting the coordinates of Charlotte NC (lat:35.2272086, lng:-80.8430827). Using the Foursquare API’s “explore” query we are able to obtain a list of all the restaurants in the area along with their latitudes and longitudes. In order to ensure only restaurants within our price range are returned we add the “price” input to the query.

https://api.foursquare.com/v2/venues/explore?&client\_id=**{}**&client\_secret=**{}**&v=**{}**&ll=**{}**,**{}**&radius=**{}**&limit=**{}**&price=**{}**

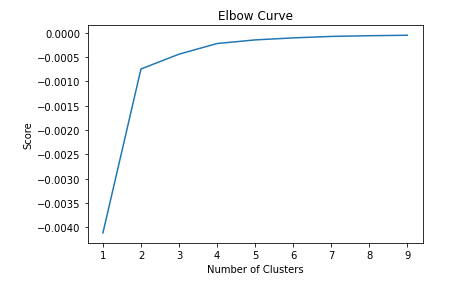
The Foursquare API returns json files with a great deal of information about each venue. After cleaning the data and converting it to a dataframe, the following table is returned. We can now group the venues in Charlotte by location.

  
**Figure 1: Example Foursquare Venues**

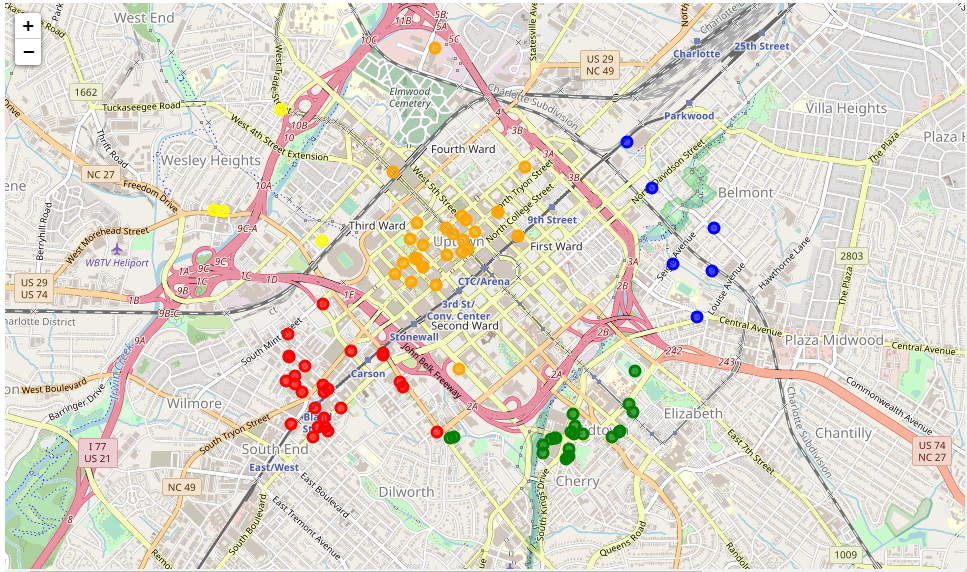
* 1. **Cluster Analysis**

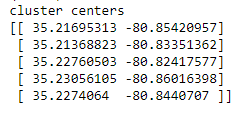
In order to find the optimal spot to visit, we must determine the area with the highest density of restaurants within our price range. To identify groups (clusters) with similar characteristics, we apply the unsupervised learning method, K-Means clustering algorithm to our data.

A fundamental step for any unsupervised algorithm is to determine the optimal number of clusters into which the data may be clustered. The Elbow Curve is one of the most popular methods to determine this optimal value. The method consists of plotting the explained variation as a function of the number of clusters, and picking the elbow of the curve as the number of clusters to use.

  
**Figure 2: Elbow Curve**

Based on the elbow curve we will choose 5 as the number of clusters, which will create very dense clusters. Using the Python Folium library we can visualize our clusters in a map of Charlotte.

  
**Figure 3: Map of clusters of venues in Charlotte NC**

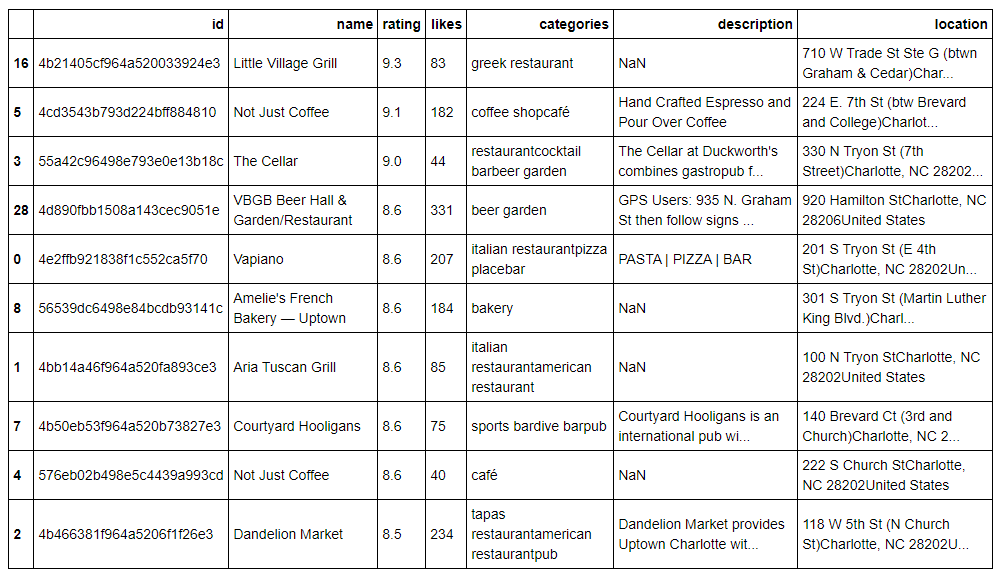
  
**Figure 4: Cluster Data**

We can see that cluster 4, located in Charlotte Uptown has the most and densest venues. We will choose this location to visit.

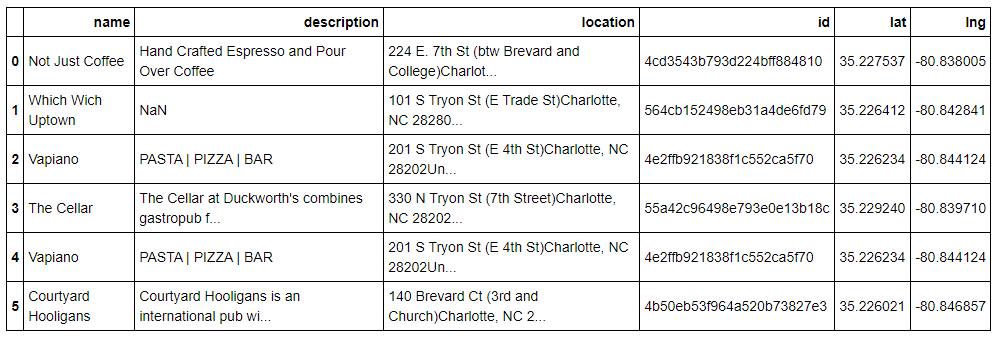
* 1. **Sentiment Analysis**

In order to find the optimal venues in this cluster we will want to isolate the venues with the highest ratings and most likes. We are specifically interested in stopping for coffee, lunch, dinner, and drinks during our day/night on the town. Using the FourSquare API’s “venue details” query to search for each venue’s ID, we can retrieve the ratings, likes, and categories for each venue on our list. After cleaning the data we are able to make a sorted list of the details for each venue in our chosen cluster.

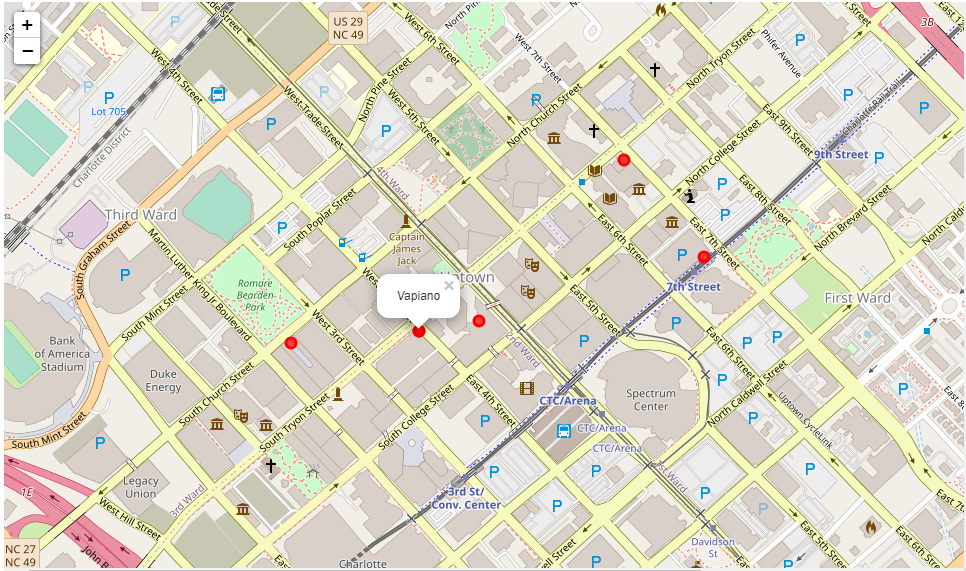
'https://api.foursquare.com/v2/venues/**{}**?&client\_id=**{}**&client\_secret=**{}**&v=**{}**'

 **Figure 5: Venue's Ratings, Likes, Categories, and Addresses**

We filter on each venue for categories containing the words: “coffee”, “sandwich”, “Italian”, and “Bar” respectively. Each time we select the top rated venue in the returned list. Then we append the results into a single dataframe and merger the dataframe with the latitude and longitude columns from the earlier foursquare data, using a left outer join.

  
**Figure 6: The top rated venues for coffee, lunch, dinner, and drinks**

We now have a list of the optimal venues to visit and their locations. Finally we create a map of our destinations using Folium.

 **Figure 7: A map of our destinations**

1. **Results and Discussion:**

During the analysis 5 clusters were identified, with the densest cluster being Charlotte Uptown. However, Midtown and Southend were also very dense clusters with a wide range of venues. These could be good places to explore during subsequent visits to Charlotte. There were 29 venues within the Uptown cluster, however, we selected 4 based on their ratings and the type of food they were serving. We have also produced a printable map and a list of the venue’s addresses so we can find them during our visit.

1. **Conclusion:**

Based on our data analysis we conclude the optimal places to visit while spending the day in Charlotte NC are: Not Just Coffee, Which Wich Uptown, Vapiano’s Italian Restaurant, and Courtyard Hooligans International Pub.