

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!

2. 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.

3. The Target Audience:

Any Charlotteans looking for an efficient night on the town!

4. 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.

5. Methodology:

5.1. 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.

	id	lat	lng	name
0	4e2ffb921838f1c552ca5f70	35.226234	-80.844124	Vapiano
1	4bb14a46f964a520fa893ce3	35.226615	-80.842317	Aria Tuscan Grill
2	4b466381f964a5206f1f26e3	35.228656	-80.842479	Dandelion Market
3	55a42c96498e793e0e13b18c	35.229240	-80.839710	The Cellar
4	576eb02b498e5c4439a993cd	35.226891	-80.846126	Not Just Coffee

Figure 1: Example Foursquare Venues

5.2. 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.

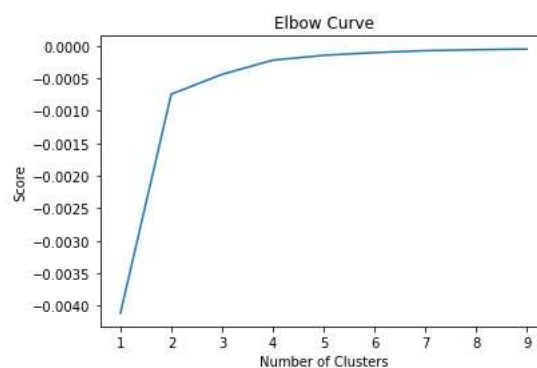


Figure 2: Elbow Curve

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.

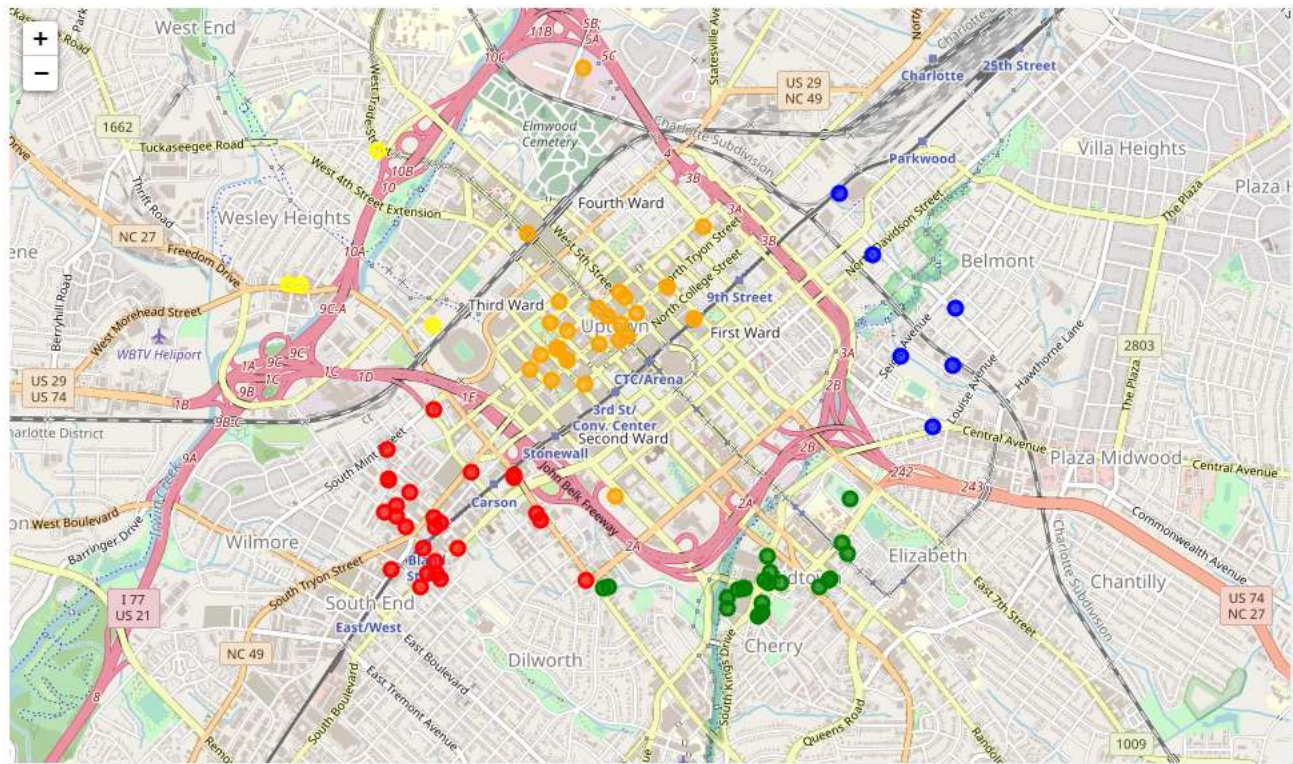


Figure 3: Map of clusters of venues in Charlotte NC

```
cluster centers
[[ 35.21695313 -80.85420957]
 [ 35.21368823 -80.83351362]
 [ 35.22760503 -80.82417577]
 [ 35.23056105 -80.86016398]
 [ 35.2274064 -80.8440707 ]]
```

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.

5.3. 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/{id}?client_id={}&client_secret={}&v={}'
```

	id	name	rating	likes	categories	description	location
16	4b21405cf964a520033924e3	Little Village Grill	9.3	83	greek restaurant	NaN	710 W Trade St Ste G (btwn Graham & Cedar)Char...
5	4cd3543b793d224bff884810	Not Just Coffee	9.1	182	coffee shopcafé	Hand Crafted Espresso and Pour Over Coffee	224 E. 7th St (btw Brevard and College)Charlot...
3	55a42c96498e793e0e13b18c	The Cellar	9.0	44	restaurantcocktail barbeer garden	The Cellar at Duckworth's combines gastropub f...	330 N Tryon St (7th Street)Charlotte, NC 28202...
28	4d890fb1508a143cec9051e	VBGB Beer Hall & Garden/Restaurant	8.6	331	beer garden	GPS Users: 935 N. Graham St then follow signs ...	920 Hamilton StCharlotte, NC 28206United States
0	4e2ffb921838f1c552ca5f70	Vapiano	8.6	207	italian restaurantpizza placebar	PASTA PIZZA BAR	201 S Tryon St (E 4th St)Charlotte, NC 28202Un...
8	56539dc6498e84bcd93141c	Amelle's French Bakery — Uptown	8.6	184	bakery	NaN	301 S Tryon St (Martin Luther King Blvd.)Charl...
1	4bb14a46f964a520fa893ce3	Aria Tuscan Grill	8.6	85	italian restaurantamerican restaurant	NaN	100 N Tryon StCharlotte, NC 28202United States
7	4b50eb53f964a520b73827e3	Courtyard Hooligans	8.6	75	sports bardive barpub	Courtyard Hooligans is an international pub wi...	140 Brevard Ct (3rd and Church)Charlotte, NC 2...
4	576eb02b498e5c4439a993cd	Not Just Coffee	8.6	40	café	NaN	222 S Church StCharlotte, NC 28202United States
2	4b466381f964a5206f1f26e3	Dandelion Market	8.5	234	tapas restaurantamerican restaurantpub	Dandelion Market provides Uptown Charlotte wit...	118 W 5th St (N Church St)Charlotte, NC 28202U...

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.

	name	description	location	id	lat	lng
0	Not Just Coffee	Hand Crafted Espresso and Pour Over Coffee	224 E. 7th St (btw Brevard and College)Charlot...	4cd3543b793d224bff884810	35.227537	-80.838005
1	Which Wich Uptown	NaN	101 S Tryon St (E Trade St)Charlotte, NC 28280...	564cb152498eb31a4de6fd79	35.226412	-80.842841
2	Vapiano	PASTA PIZZA BAR	201 S Tryon St (E 4th St)Charlotte, NC 28202Un...	4e2ffb921838f1c552ca5f70	35.226234	-80.844124
3	The Cellar	The Cellar at Duckworth's combines gastropub f...	330 N Tryon St (7th Street)Charlotte, NC 28202...	55a42c96498e793e0e13b18c	35.229240	-80.839710
4	Vapiano	PASTA PIZZA BAR	201 S Tryon St (E 4th St)Charlotte, NC 28202Un...	4e2ffb921838f1c552ca5f70	35.226234	-80.844124
5	Courtyard Hooligans	Courtyard Hooligans is an international pub wi...	140 Brevard Ct (3rd and Church)Charlotte, NC 2...	4b50eb53f964a520b73827e3	35.226021	-80.846857

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

6. 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.

7. 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.