

Opening a Restaurant in Charlotte, North Carolina

Stuart Hunter

May 30th 2021

Introduction

My customer is a serial entrepreneur who is looking for a new investment opportunity.

She believes that after a year of being at home during the pandemic, people are impatient to get back out there and pursue their leisure activities, including dining out.

She is planning to either open a new restaurant or takeover an existing one - perhaps one that has gotten into financial difficulties because of the closures due to the pandemic. She is looking for assistance in two areas:

1. Neighborhood

My customer lives in the SouthPark neighborhood of Charlotte, but her entrepreneurial lifestyle means she travels a lot and has little experience of the food scene outside of her own locale in the city.

She does want to go with what she knows though and outside of SouthPark she wants to consider other 'similar' neighborhoods in Charlotte.

2. Cuisine

No fast food or takeaways like Pizza or Burgers, she is looking to establish or invest in a sit in restaurant, serving good food to groups of people.

She is not tied to a specific cuisine, she is looking for a 'popular' cuisine that may be underrepresented in a desirable neighborhood or where there may be opportunities to grow an already established but non-saturated cuisine.

My deliverable is to provide her with some relevant neighborhoods to choose from and some idea as to what the market is for different cuisines in the city.

Data

The following data sources will be used in this analysis.

1. Wikipedia

Wikipedia provides a page listing the neighborhoods of Charlotte, North Carolina where this analysis will take place.

It is not held in a table so it will need to be scraped using href tags that can be uniquely identified by their common attributes.

https://en.wikipedia.org/wiki/List_of_Charlotte_neighborhoods

```
<li><a href="/wiki/Elizabeth_(Charlotte_neighborhood)" tit'
<li><b>Midtown</b>, located along King's Drive southeast o
<h2><span class="mw-headline" id="East_Charlotte">East Cha
<ul><li><b>Coventry Woods</b> is a neighborhood with matur
<li><b>Easthaven</b> is bounded by W.T. Harris Blvd, Easth
<li><a href="/wiki/Eastland_(Charlotte_neighborhood)" titl
<li><b>Grove Park</b> is a spacious park-like neighborhood
<li><b>Hickory Grove</b> is an area of East Charlotte along
<li><b>Hickory Ridge</b> is an area approximately bordered
<li><b>Idlewild</b> is a neighborhood bordering Eastland a
<li><a href="/wiki/Mint_Hill,_North_Carolina" title="Mint I
<li><b>Oakhurst</b> is the area surrounding Monroe Rd betw
<li><b>Plaza Hills</b> is an area extending from North of I
<li><a href="/wiki/Plaza-Midwood_(Charlotte_neighborhood)"
```

2. Geopy

We need the attitude and longitude information for each of the neighborhoods, and that is provided by the Geopy Python module.

```
for item in neighborhoods:
    loc = item + ", Charlotte, NC"
    geolocator = Nominatim(user_agent="my_request")
    location = geolocator.geocode(loc)
    print(item, location.latitude, location.longitude)
    lat_list.append(location.latitude)
    long_list.append(location.longitude)
```

```
Derita 35.2833128 -80.8192863
Starmount 35.1382582 -80.8683834
Quail Hollow 35.1186739 -80.8399701
Steele Creek 35.14409795 -80.97896658564757
Paw Creek 35.263798 -80.916153
Eastland 35.2088208 -80.7513256
```

3. Foursquare

With the latitude and longitude identified, the neighborhoods are submitted to Foursquare to retrieve the recorded venues within 500 meters of the neighborhood location.

```
# Prepare inputs to retrieve top 100 venues from this first neighborhood from Foursquare
limit = 100
radius = 500
url = 'https://api.foursquare.com/v2/venues/explore?client_id={}&client_secret={}&v={}&l'

# Request URL from Foursquare
results = requests.get(url).json()
results

{'venue': {'id': '4b105607f964a520276e23e3',
            'name': 'Cafe Monte',
            'location': {'address': '6700 Fairview Rd',
                        'crossStreet': 'Allen Tate Building at Phillips Place',
                        'lat': 35.14715397463186,
                        'lng': -80.82872611013036,
                        'labeledLatLngs': [{'label': 'display',
                                           'lat': 35.14715397463186,
                                           'lng': -80.82872611013036}],
                        'distance': 244,
                        'postalCode': '28210',
                        'cc': 'US',
                        'neighborhood': 'SouthPark',
                        'city': 'Charlotte',
                        'state': 'NC',
```

4. KMeans

The KMeans method of the scikit Python module is used to generate data used to cluster the neighborhoods.

```
# Cluster Neighborhoods

# set number of clusters
# Having tried kclusters at 7 and 10 which had most of the neighborhoods in one category
# I found that 15 works well for this exercise, giving a large enough cluster to focus on
# for further processing
kclusters = 15

charlotte_grouped_clustering = charlotte_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(charlotte_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:15]
```

5. Folium

The Folium Python module was used to generate maps showing the location of neighborhoods in the Charlotte area and then to display the neighborhoods after clustering to show which neighborhoods belonged to each cluster.

Methodology

1. Wikipedia Data

Wikipedia provides a page listing the neighborhoods of Charlotte, North Carolina where this analysis will take place.

https://en.wikipedia.org/wiki/List_of_Charlotte_neighborhoods

It is not held in a table so it will need to be scraped using href tags that can be uniquely identified by their common attributes. As shown in the image below neighborhoods can be identified in <a href> tags and are suffixed by *Charlotte_neighborhood* so all text that matched those parameters were retrieved and pulled into a list

```
<li><a href="/wiki/Elizabeth_(Charlotte_neighborhood)" tit'
<li><b>Midtown</b>, located along King's Drive southeast o
<h2><span class="mw-headline" id="East_Charlotte">East Cha
<ul><li><b>Coventry Woods</b> is a neighborhood with matur
<li><b>Easthaven</b> is bounded by W.T. Harris Blvd, Easth
<li><a href="/wiki/Eastland_(Charlotte_neighborhood)" titl
<li><b>Grove Park</b> is a spacious park-like neighborhood
<li><b>Hickory Grove</b> is an area of East Charlotte alon
<li><b>Hickory Ridge</b> is an area approximately bordered
<li><b>Idlewild</b> is a neighborhood bordering Eastland a
<li><a href="/wiki/Mint_Hill,_North_Carolina" title="Mint I
<li><b>Oakhurst</b> is the area surrounding Monroe Rd betw
<li><b>Plaza Hills</b> is an area extending from North of I
<li><a href="/wiki/Plaza-Midwood_(Charlotte_neighborhood)"
```

2. Invalid Data

Some of those tags also describe non-relevant data so they will have to be excluded from the list of neighborhoods. They are mainly navigation links for the bottom of the web page

3. Incorrect Tags

Some of the neighborhoods have been entered with either incorrect tags or with erroneous detail and these will have to be appended to the list of neighborhoods manually.

4. Duplicates

Duplicates appear on the web page so these will be removed programmatically to ensure the uniqueness of each neighborhood.

5. Geolocation

The neighborhood names in the list retrieved from Wikipedia are appended with “, **Charlotte, North Carolina**” so that they can be identified correctly when submitted to Geopy. Without this Geopy finds the first matching neighborhood name which could be in another city and state.

6. Foursquare

With the latitude and longitude identified, the neighborhoods are submitted to Foursquare to retrieve the recorded venues within 500 meters of the Geopy provided neighborhood location.

7. Mapping

Maps were generated first to show the general location and layout of Charlotte, and then with the location of each neighborhood superimposed over the top of the general map.

8. Clustering

All of the venues are used to cluster the data via kmeans rather than just using restaurant venues at this stage to get a rounded picture of the facilities and lifestyles available in each neighborhood.

We are aiming to be able to focus on a prospective neighborhood cluster that is not too large or small but gives an adequate number of related neighborhoods with which we can perform further analysis. After running means with a variety of values for Kclusters, it was found that K=15 gave a reasonable cluster size with which to continue the analysis.

9. Filtering on Restaurants Only

After identifying the cluster containing the neighborhoods of direct relevance to my customer the data was then filtered to exclude all venues other than those which are identified as restaurants (other than eateries that could be classified as takeout such as Pizza places and Chinese Takeout).

10. Unique Cuisines

A list was made containing the unique list of restaurant cuisine types and then sorted by how popular they are. The top 5 will be presented to the customer as popular cruises to consider for her own restaurant.

11. Cuisine Saturation in Neighborhoods

For each neighborhood the number of restaurants was counted to see what neighborhoods may be over or underserved.

Results

25 neighborhoods were identified within the Charlotte area and after clustering we ended up with 10 neighborhoods that were closest to the customer's familiar SouthPark neighborhood - including SouthPark itself.

Foursquare presented us with 150 different venue categories for all of the neighborhoods within Charlotte and 394 individual venues.

KMeans clustering with kcluster=15 gave us the most acceptable results with 10 neighborhoods in a cluster that gave us a food number to compare.

One we filtered the venues to only reflect those in the cluster neighborhoods that were restaurants we had 142 restaurants.

The top5 restaurant cuisines were identified with which the customer will consider what cuisine to pursue. 'Fast Food Restaurant' appeared in the top5 list, so this was removed from the results as the customer is not interested in such cuisine, and it was replaced by the next most popular - 'Italian Restaurant'

2 cross tabs were generated, 1 showed all the cluster neighborhoods and how many of each type of cuisine restaurant they had, and the other shows the top 5 cuisine restaurants and what neighborhoods they were in.

Discussion

The analysis produced two major tools to help my customer make her determination as to the best neighborhoods to consider for opening a restaurant.

1. A table of all of the Neighborhoods in the cluster and how many of each type of restaurant the contained

Neighborhood	Ballantyne	Cotswold	Coulwood	Elizabeth	NoDa	Parkdale	Plaza-Midwood	South End	SouthPark	All
Cuisine										
American Restaurant	2	0	0	0	0	1	1	3	5	12
Asian Restaurant	3	0	0	0	0	0	1	2	0	6
Caribbean Restaurant	0	0	0	0	0	0	1	0	0	1
Chinese Restaurant	1	1	0	1	0	0	1	1	2	7
Eastern European Restaurant	0	0	0	0	1	0	0	0	0	1
French Restaurant	0	0	0	0	0	0	0	1	1	2
Greek Restaurant	1	0	0	1	0	0	0	0	1	3
Indian Restaurant	2	0	0	0	0	0	0	1	0	3
Italian Restaurant	1	0	0	0	0	1	0	1	1	4
Japanese Restaurant	0	0	0	0	0	0	0	0	1	1
Mexican Restaurant	1	1	0	1	0	1	0	1	1	6
New American Restaurant	0	0	0	3	0	0	0	0	1	4
Peruvian Restaurant	0	0	0	1	0	0	0	0	0	1
Ramen Restaurant	0	0	0	0	0	0	0	1	0	1
Restaurant	0	0	0	0	0	0	1	2	0	3
Seafood Restaurant	0	0	0	0	1	0	0	0	0	1
Southern / Soul Food Restaurant	0	0	1	0	1	0	1	0	0	3
Sushi Restaurant	0	1	0	0	0	0	0	0	0	1
Tapas Restaurant	0	0	0	0	0	0	0	1	0	1
Tex-Mex Restaurant	1	0	0	0	0	0	0	0	0	1
Thai Restaurant	0	0	0	0	0	0	0	2	0	2
Vegetarian / Vegan Restaurant	0	0	0	0	0	0	0	1	1	2
All	12	3	1	7	3	3	6	17	14	66

2. A table of the Top 5 cuisines by popularity and which neighborhoods they are in

Neighborhood	Ballantyne	Cotswold	Elizabeth	Parkdale	Plaza-Midwood	South End	SouthPark	All
Cuisine								
American Restaurant	2	0	0	1	1	3	5	12
Asian Restaurant	3	0	0	0	1	2	0	6
Chinese Restaurant	1	1	1	0	1	1	2	7
Italian Restaurant	1	0	0	1	0	1	1	4
Mexican Restaurant	1	1	1	1	0	1	1	6
All	8	2	2	3	3	8	9	35

One interesting observation is that when I noticed that the number of cluster neighborhoods dropped from 10 neighborhoods to 9 after the filter was applied to just show Restaurant categories. This showed that 'University City' was a cluster neighborhood that had no Restaurant category venues listed.

As University City has many things in common with the other cluster neighborhoods we may assume that there is a demand for such sit-in restaurants that we are considering, but that the residents of University City must be traveling to other neighborhoods for such amenities or are 'making do' with FastFood Restaurants.

Conclusion

In advising my customer in how to interpret the results, I would suggest the following.

1. Choose a popular cuisine type
2. Choose a neighborhood that has a large number of restaurants (indicating there is demand), but little or no representation of that popular cuisine type

As an example of this way of interpreting the results I would consider an Asian Restaurant (the 2nd most popular cuisine type) and looking to situate that in SouthPark - which has 9 restaurants, so there is demand there, but none of them are Asian restaurants.