Targeting Locations for Hair Salon Owners

Andrew Randolph – Data Scientist

November 26, 2019

1. INTRODUCTION

1.1 Background

WWFY, LLC is a Technology firm launched in November 2018. We are dedicated to helping companies increase equity through increased operational performance. We do this by providing business insights through data analytics.

However, there are hair solons in my city of Greenville, SC who currently rely on city location only to attract new walk-in clientele. They open a shop on a busy street corner and hope that footfall traffic will notice their salon and walk in for service. But, when it comes to hair styles and beauty, this city's salons are known for catering to a specific demographic only. For instance, there are salons for Caucasians, African-Americans, Hispanics, etc. It is unfortunate, but true.

Therefore, just because a salon is located on a busy street with high footfall, it doesn't mean that that salon will be successful. Locating a salon in the proper area code that serves its demographic is paramount for success.

1.2 The Problem

There are hair solons in my city that have historically relied on location to attract new walk-in customers in order to build their clientele.

However, the problem with this approach is since most salons in this city cater to specific demographics, simply having a salon in a high footfall area may not necessarily generate a high number of walk-in clients, thus causing the owner to have a money-losing operation.

1.3 Assumptions

One assumption is that in this particular southern-state city, people of a particular demographic will generally patronize a salon that generally serves them only. Additionally, these customers will patronize restaurants that serve them also. Therefore, I can assume that if I cluster restaurants that serve a particular cuisine, that cuisine will be highly correlated to that specific demographic. Salon owners will find optimal locations near restaurants that serve their demographic.

1.4 Who's Interested

Those who would be interested in this analysis are hair salon owners that are either new to the area or locals who are in under-performing locations who desire to relocate their shop(s). This owner group pertains to corporate owned and chain salons.

2. THE DATA

2.1 Data Acquisition

The data that will be used in this effort is a combination of zip codes from the city of reference and Foursquare venue data.
I also used Python's Nominatim to get geospatial data for the zip codes.

Examples of data that will be included in the final dataset:
Zip codes in the city
Foursquare Venue data with a search query against restaurant types.
I will use geopy to get geospatial data in order to generate maps to show segmentation and clustering then segment the data based on certain cuisines.

2.2 Data Wrangling

Since there are only 17 area codes in this city I decided to simply create a .csv file with neighborhood names and load the file into python as a pandas dataframe. See figure 1.

	zip	neighborhood
0	29617	berea
1	29635	cleveland
2	29644	fountain_inn
3	29605	gantt
4	29607	greenville 1
5	29609	greenville 2
6	29613	greenville 3
7	29614	greenville 4
8	29601	greenville 5
9	29650	greer 1
10	29651	greer 2
11	29661	marietta
12	29662	mauldin
13	29611	parker
14	29673	piedmont
15	29680	simpsonville 1
16	29681	simpsonville 2
17	29683	slater-marietta
18	29687	taylors
19	29690	travelers rest
20	29615	wade hampton

Figure 1. Greenville, SC Neighborhoods with Zip Codes

Wrangling the geospatial data was a bit more cumbersome. The data from this source not only has coordinates of the zip codes, but also complete address information. The only features that I needed are the longitude and latitude values. Therefore, I modified the dataframe with a lambda function that not only stripped away the geospatial information, but also created two columns for this information. The first few rows of the resulting dataframe are show in figure 2.

zip	neighborhood	latitude	longitude	
0	29617	berea	34.896035	-82.447021
1	29635	cleveland	35.105984	-82.626093
2	29644	fountain_inn	34.687266	-82.218567
3	29605	gantt	34.799591	-82.394795
4	29607	greenville 1	34.828211	-82.331043

Figure 2. Geospatial dataframe

The features of the FourSquare restaurant data include name, address, cross street, coordinates, distance, country and category. I was only interested in keeping the name and anything associated with its location. The resulting dataframe was grouped by zip code. See figure 3.

zip	neighborhood latitude	neighborhood longitude	venue	venue latitude	venue longitude	venue category	
0	29617	34.896035	-82.447021	Celebrity's Hot Dogs	34.887423	-82.456889	Hot Dog Joint
1	29617	34.896035	-82.447021	Tomato Vine	34.896697	-82.432131	Farmers Market
2	29617	34.896035	-82.447021	CVS pharmacy	34.894302	-82.431767	Pharmacy
3	29617	34.896035	-82.447021	Goodwill	34.892352	-82.450551	Thrift / Vintage Store
4	29617	34.896035	-82.447021	Family Dollar	34.893799	-82.432545	Discount Store

Figure 3. Venue (Restaurants) Dataframe grouped by zip code

3. METHODOLOGY / ANALYSIS

3.1 Analysis of Restaurants in Each Neighborhood

Since the Venue Category column of figure 3 is categorical data, I used a python function called one-hot encoding to convert to numerical data. Afterwards, I took the mean of the frequency of each venue. Doing this gives me a measure of the most popular restaurants in that area.
Then I created a new dataframe with only the top 10 venues for that particular zip code for further analysis. See figure 4.

zip	neighborhood	latitude	longitude	Cluster Labels	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	9 C
29617	berea	34.895414	-82.447059	3	Fast Food Restaurant	Convenience Store	Chinese Restaurant	Pharmacy	Sandwich Place	Discount Store	Diner	Gas Station	
29635	cleveland	35.069949	82.600574	2	Trail	Park	Zoo	Discount Store	Fast Food Restaurant	Farmers Market	Farm	Electronics Store	
29644	fountain_inn	34.687266	-82.218567	3	Convenience Store	Rental Car Location	Pizza Place	Breakfast Spot	Baseball Field	Discount Store	Fast Food Restaurant	Grocery Store	Dep
29605	gantt	34.799591	-82.394795	0	Hotel	Gym / Fitness Center	Discount Store	Pizza Place	American Restaurant	Miscellaneous Shop	Golf Course	Thrift / Vintage Store	F
29607	greenville 1	34.828555	-82.331346	3	Rental Car Location	Fast Food Restaurant	Bookstore	Discount Store	Department Store	Sandwich Place	ATM	Coffee Shop	Re:

Figure 4. Top 10 Restaurants per Zip Code

3.2 Results

I want my results to segregate the most popular cuisine in a particular zip code or neighborhood.

The data has no predefined labels. I do not currently have a method of clumping or clustering the restaurants into specific groups based on cuisines. Therefore, I will need to utilize a machine learning algorithm to create the labels. I created a K-Means algorithm which is a machine learning concept that creates labels and segregates those labels into 3 clusters with.

However, three clusters did not provide enough segregation of the restaurants for analysis. Most of the cuisines of the restaurants were clustered in the same vicinity and did not provide ample segregation to determine the type of demographic that patronized them for that area.

Therefore, I used a k-value of 4 which produced several zones with cluster values of 0 to 3 (red, green, yellow, purple) making it easier to determine the most popular restaurant cuisine in a particular zip code. See figure 5, Neighborhood Clusters map.

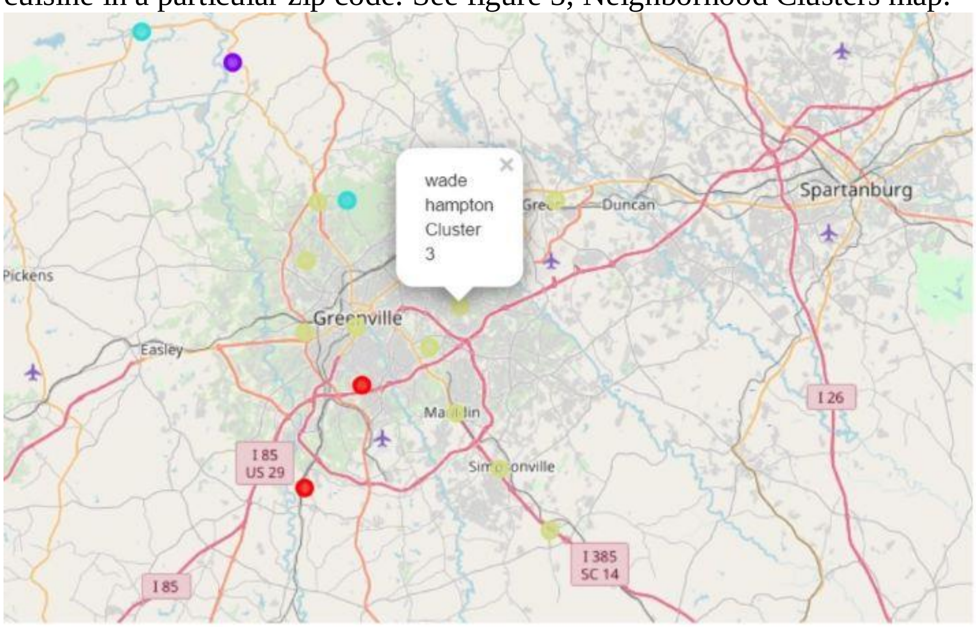

Figure 5. Neighborhood Clusters map

The clusters were selected and assigned to their demographic accordingly.

Cluster 0 = Italian demographic

Cluster 1 = Working class Caucasian demographic

Cluster 2 = Caucasian - Professional demographic

Cluster 3 = African-American / Hispanic demographic

4. CONCLUSION

The clustering algorithm indicated and analysis confidently showed cuisines of restaurants typically enjoyed by African-Americans, upscale Caucasians, working-class Caucasians and Italians. But failed to distinguish Asians or Hispanics which could be important to salon owners.

5. DISCUSSION

It may be possible to dig deeper into the analysis by searching Foursquare on the main menu item served by each restaurant, then correlating this to a specific demographic. Further analysis could possibly give insights as to other cultures in the city of Greenville.