



W W F Y L L C

W E H E L P Y O U I N C R E A S E P E R F O R M A N C E

www.weworkforyoullc.com



Capstone: Battle of The Neighborhoods Week 2

Targeting Locations for Hair Salon Owners

ANDREW RANDOLPH – DATA SCIENTIST

NOVEMBER 26, 2019



INTRODUCTION

There are hair salons in my city of Greenville, SC that currently rely on 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 specific demographics 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, 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.



ASSUMPTIONS

- Customers of a particular demographic will generally patronize a salon that generally serves them only.
- These same customers will patronize restaurants that serve cuisines that they like.
- Salons can locate near restaurants to attract the same demographic.



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.





THE DATA

The data that will be used in this effort is a combination of zip codes from the city of reference and Foursquare venue data.

Venues located in a specific zip code are an indication of the cultural demographics in that zip code. In this particular southern-state city, people of similar demographics generally hang out in the same venues together.

As such, I shall utilize a K-means algorithm to segment / cluster those venues as they will be highly correlated to the demographics of the people living in that specific zip code.

This segmentation will help salon owners choose the ideal location of their salons in order to attract specific walk-in customers who fit their demographic – increasing clientele.



THE DATA

Examples of data that will be included in the final dataset:

- ❖ Zip codes in the city
- ❖ Foursquare Location data with a search query against restaurant cuisine types
- ❖ Latitude / Longitude values in order to generate maps
- ❖ Top 10 restaurants per zip code based on cuisine

THE DATA

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	9th Most Common Venue
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	Department Store
29605	gantt	34.799591	-82.394795	0	Hotel	Gym / Fitness Center	Discount Store	Pizza Place	American Restaurant	Miscellaneous Shop	Golf Course	Thrift / Vintage Store	Fast Food Restaurant
29607	greenville 1	34.828555	-82.331346	3	Rental Car Location	Fast Food Restaurant	Bookstore	Discount Store	Department Store	Sandwich Place	ATM	Coffee Shop	Restaurant



METHODOLOGY / ANALYSIS

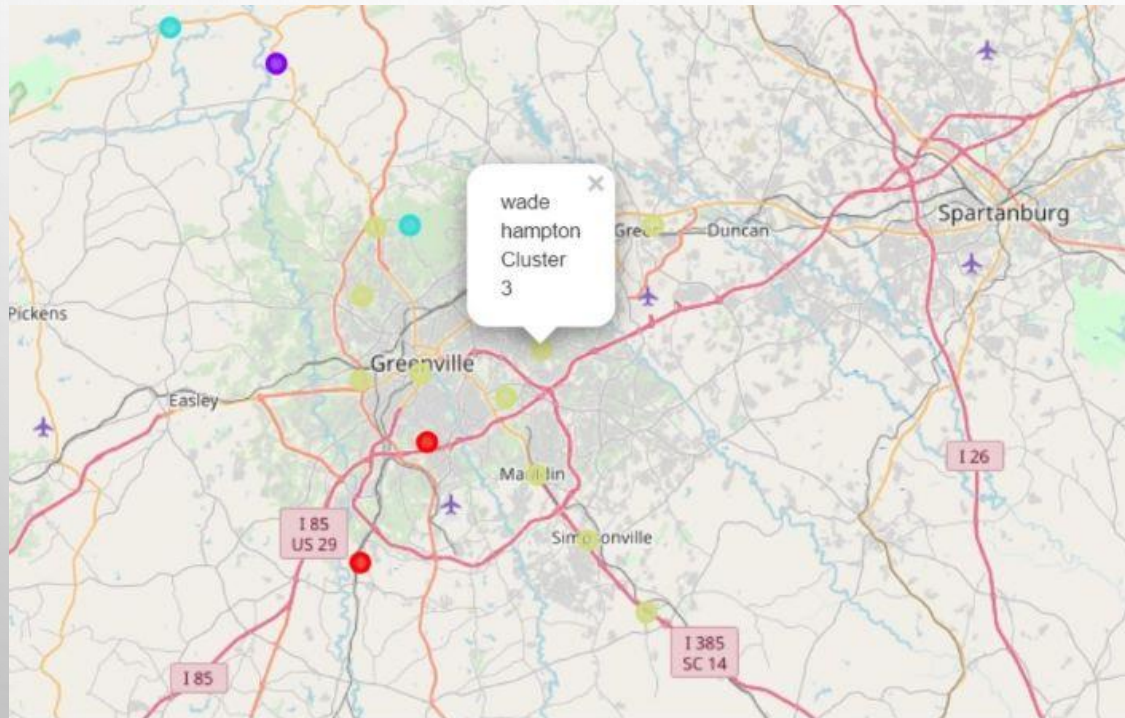
Since the Venue Category is categorical data, I used one-hot encoding to convert to numerical data. Afterwards, I took the mean of the frequency of each venue which shows the most popular restaurant type in each zip code.

Created a new dataframe with only the top 10 venues for that particular zip code for further analysis.

Used K-means algorithm to cluster unlabeled data into several segments of 4 labeled clusters

RESULTS

Map with 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.





RESULTS

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



CONCLUSION

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



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.

