

Where in Raleigh, NC Should I Open an Italian Restaurant?

Scott Josephson | Applied Data Science Capstone | 1/7/19

# Introduction

My friend is looking to move out from New York City as he gets closer to retirement. . He has been in the restaurant business for 20 years and has owned several Italian Restaurants in New York. He wants to move to a quieter town, but he still needs income. He has asked me to research other cities to see which towns are need of a good Italian restaurant. I am currently looking at Raleigh North Carolina. I decided to research Italian restaurant competitors, Cafes, Shopping, and Nightlife. I am most interested in finding places from the different categories and seeing where in Raleigh are the greatest concentration of other Italian Restaurants located in conjunction with other popular cafes, shopping, and Nightlife venues.

# Data

The data that will be using are from Italian Restaurant Venues, Cafes, Shopping, and the Nightlife in the city of Raleigh. We will utilize the data from Foursquare. We will look at features such as name, categories, address, cross street, latitude, longitude.

# Methodology

The goal was to locate Raleigh’s largest concentration of venues from the 4 categories in order to suggest to my friend where his Restaurant would be most populated. Raleigh has 15 zip codes which represents many neighborhoods. I first obtained geographical latitude and longitude coordinates using the zip codes and plugging them into Google’s Geocode API. I then created a dataframe with the Raleigh zip codes and their respective geocoordinates. I then created a Folium Map of Raleigh and its zip codes shown here:

A close up of a map

Description automatically generated

Using the Foursquare API I was able to obtain a listing of the venues I was interested in. To accomplish this I filtered the search by zip code and category name.

Then I created another Folium Map of Raleigh Venues from the four categories shown here:

A picture containing text, map

Description automatically generated

I then prepared the venues for clustering. I used the LabelEncoder class from the scikit-learn library to encode the zip code and category into numerical codes for easier processing. K-Means Clustering was used to group together the venues

# Results

Four clusters were determined to be the best fit. So I color coded them by category:

Cluster 1: Cafes – red

Cluster2: Italian Restaurants – purple

Cluster 3: Nightlife – green

Cluster 4: Shopping - yellow

I counted the totals from the venues and grouped them by zip code. The results are shown by zip code and the quantity of four category venues in them.

27601 23

27612 15

27609 13

27603 10

27617 10

27615 9

27605 5

27616 5

27607 4

27613 4

27614 4

27606 3

27610 1

# Discussion

As a result of this analysis, I will recommend to my friend that his first choice to open his restaurant should be in the 27601 zip code. This neighborhood in downtown Raleigh which is a growing and becoming a very popular travel spot.

# Conclusion

Here is the final map of the Raleigh areas with its color coded categories placed in the specific zip code areas.

A picture containing text, map

Description automatically generated