# Exploring new business opportunities in Wake County, NC

# Srinivas Tatiraju

# 7/17/2020

Table of Contents

[1. Introduction 2](#_Toc45971263)

[1.1 Background 2](#_Toc45971264)

[1.2 Problem 2](#_Toc45971265)

[1.3 Interest 2](#_Toc45971266)

[2. Methodology 2](#_Toc45971267)

[2.1. Data acquisition and cleaning 2](#_Toc45971268)

[2.1.1 Data Sources 2](#_Toc45971269)

[2.1.2 Data Cleansing 3](#_Toc45971270)

[2.2 Exploratory Data Analysis 3](#_Toc45971271)

[2.3 Statistical Data Analysis 8](#_Toc45971272)

[2.4 Clustering 9](#_Toc45971273)

[3. Results 10](#_Toc45971274)

[4. Discussion 10](#_Toc45971275)

[5. Conclusion 11](#_Toc45971276)

# Introduction

## 1.1 Background

A businessman wants to start a new business in North Carolina, USA. With RDU International Airport in the middle and multiple Fortune-100 Tech companies around, Wake county is his first choice for setting up the business in NC. So, he wants to explore various business opportunities in Wake county.

## 1.2 Problem

For him to decide in which city he shall make investment to start the business, he wants to explore the following data:

1) What are the various cities in Wake county in NC state?

2) What are the various businesses in these cities?

3) What is the count of each these businesses in each of these cities?

4) What are the most common businesses & least common food businesses in these cities?

## 1.3 Interest

Apart from exploring the above-mentioned data, he also wants to focus particularly on Restaurants & other food businesses because of his earlier experience in restaurant business.

Fortunately for the businessman, also for all other potential investors/businesses, we can provide this information using Data Science methodologies & Python language.

# Methodology

## 2.1. Data acquisition and cleaning

### 2.1.1 Data Sources

Here are the data points needed for this project:

1. Identify the cities in Wake county, NC: List of cities in NC by county is available on Internet at the following website:

<https://www.sog.unc.edu/resources/microsites/knapp-library/cities-north-carolina>

2) Address info (i.e. longitude & latitude) for these cities: This is obtained using geopy.geocoders.

3) Identify & explore various businesses, venues in these cities: Using the longitude & latitude values of the cities, information about various businesses, venues in these cities can be obtained using Foursquare API.

4) Identify the no. of businesses, their frequency, top-10 most common businesses & least common food businesses in these cities and calculate their frequency.

5) Identify the business densities in these cities using above data points.

### 2.1.2 Data Cleansing

The UNC website (<https://www.sog.unc.edu/resources/microsites/knapp-library/cities-north-carolina>) contains information about all the cities in NC state. From that list, cities in Wake county are filtered out first. Then, additional info column is removed from the list as it is not required for the analysis.

The second step is to obtain the latitude, longitude information for the cities in Wake county using geopy.geocoders and appending this information to the city list. This cleansed final table is the source for all subsequent analysis in the project.

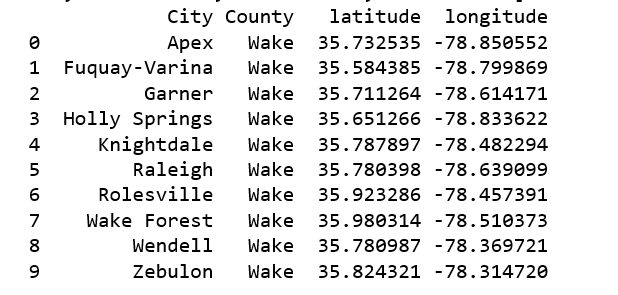
The third step is identifying various food businesses & venues in each of these cities in Wake county. This data is obtained from Foursquare API. But the data from API is in JSON format and needs to be flattened. For this the data is extracted first from Foursquare API, then the venues & businesses information is extracted using Python functions & Pandas Dataframes.

## 2.2 Exploratory Data Analysis

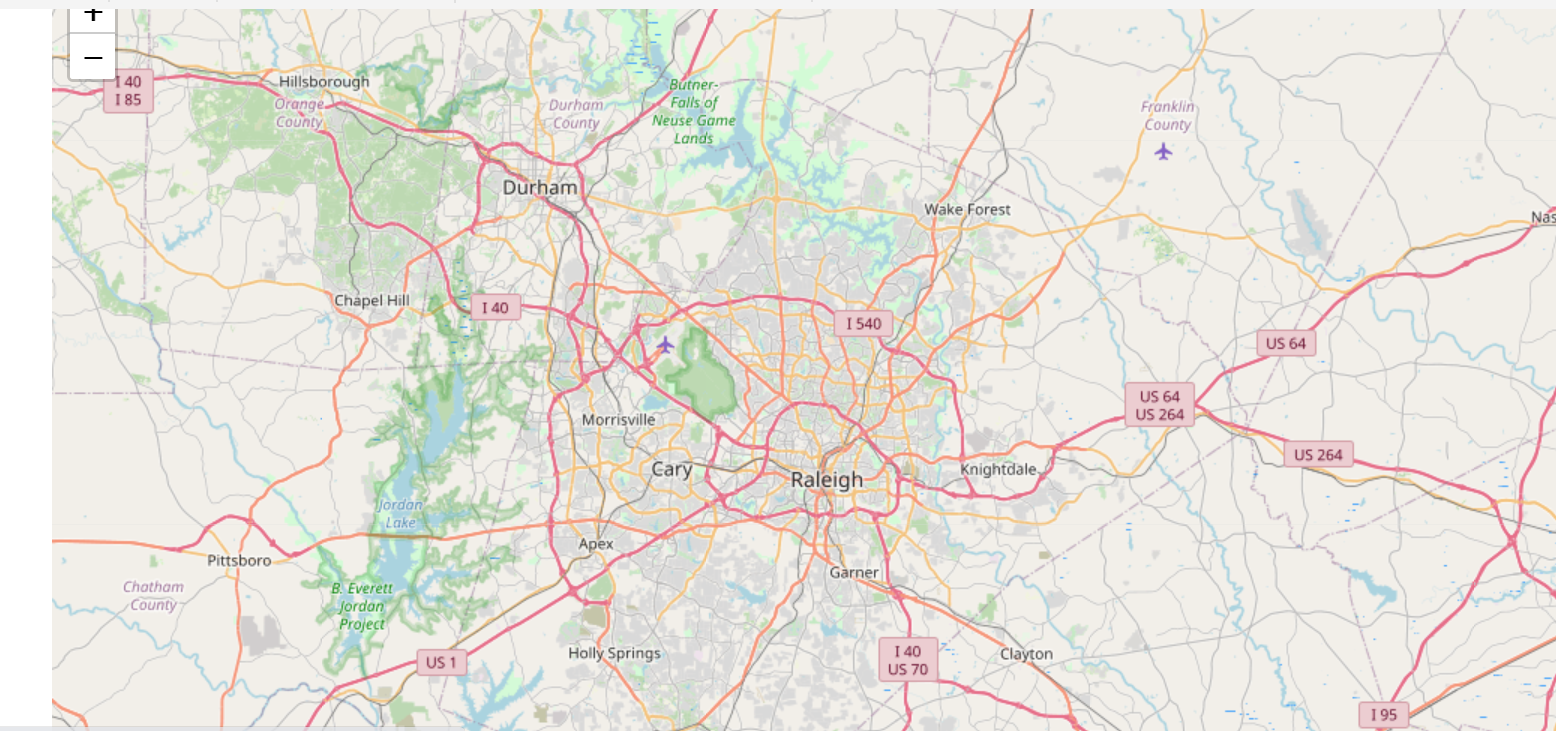
### 2.2.1. Analysis of Wake county city data

After the City data is extracted from UNC website and the corresponding latitude, longitude data is appended from geopy.geocoders, that data can be stored in a dataframe and looked at to make sure the correct data is captured.

Here is the snapshot of the Wake county City data:

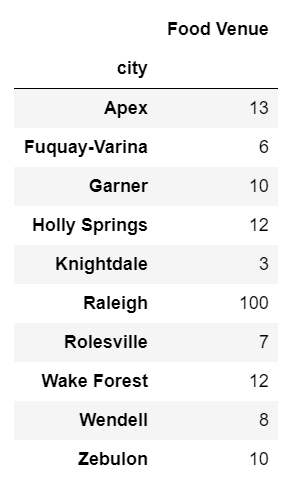


A map of Wake county can be displayed using Folium with this city data.

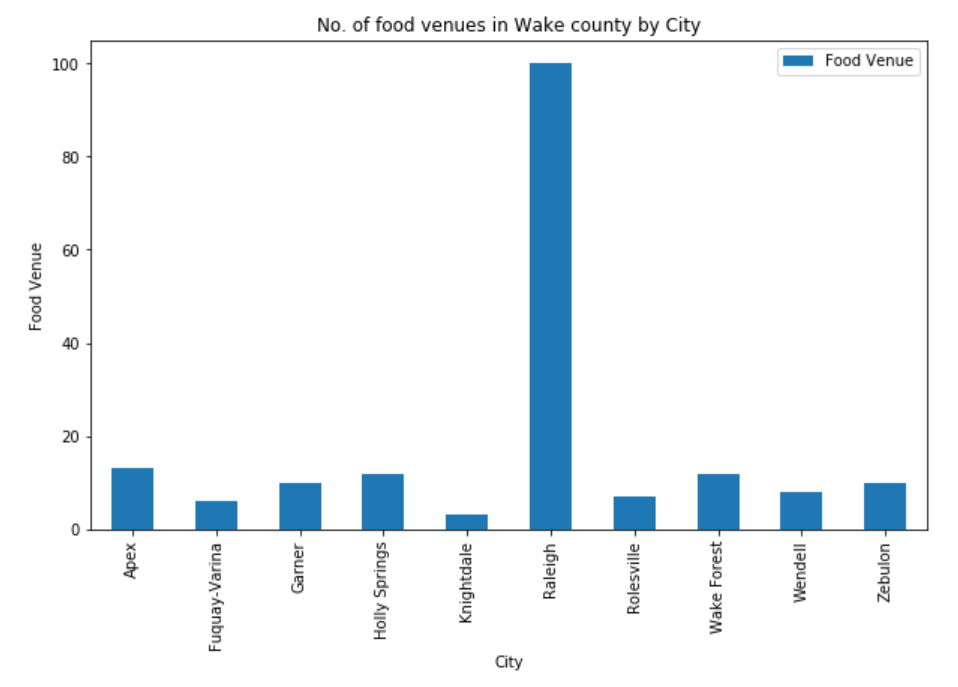


#### 2.2.1.1 Food venue data by city

Once the City data and the corresponding latitude, longitude information is captured, the list of Food venues/businesses is extracted from Foursquare API. Snapshot of Food venues by City is displayed below.



The same information can be displayed in a bar chart using matplot library as shown below:

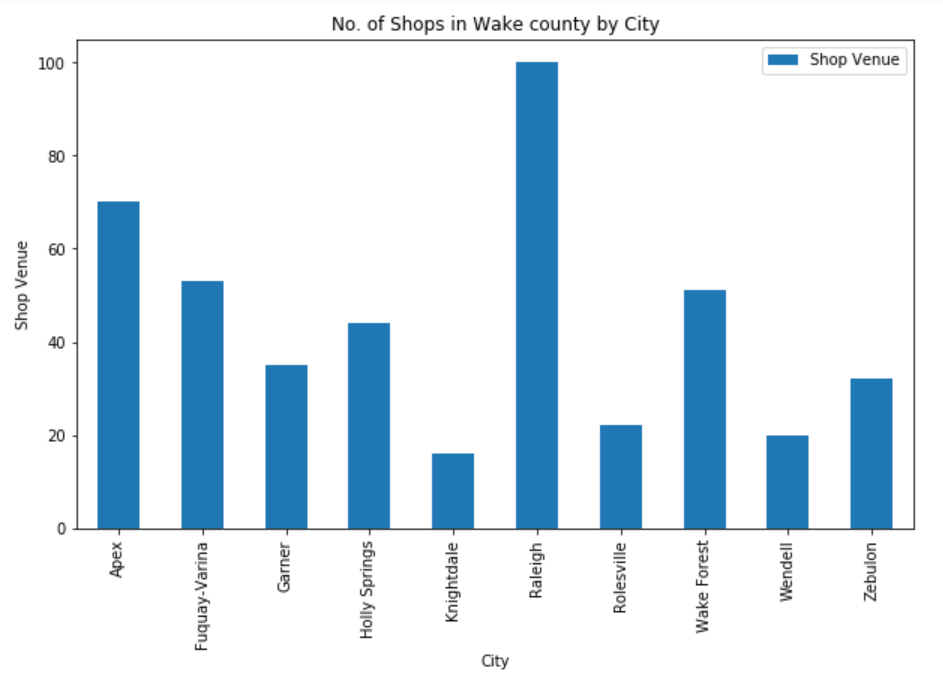


#### 2.2.1.2 Shop Venue data by city

Next the list of Shops/venues is extracted from Foursquare API. Snapshot of this information by city is displayed below:

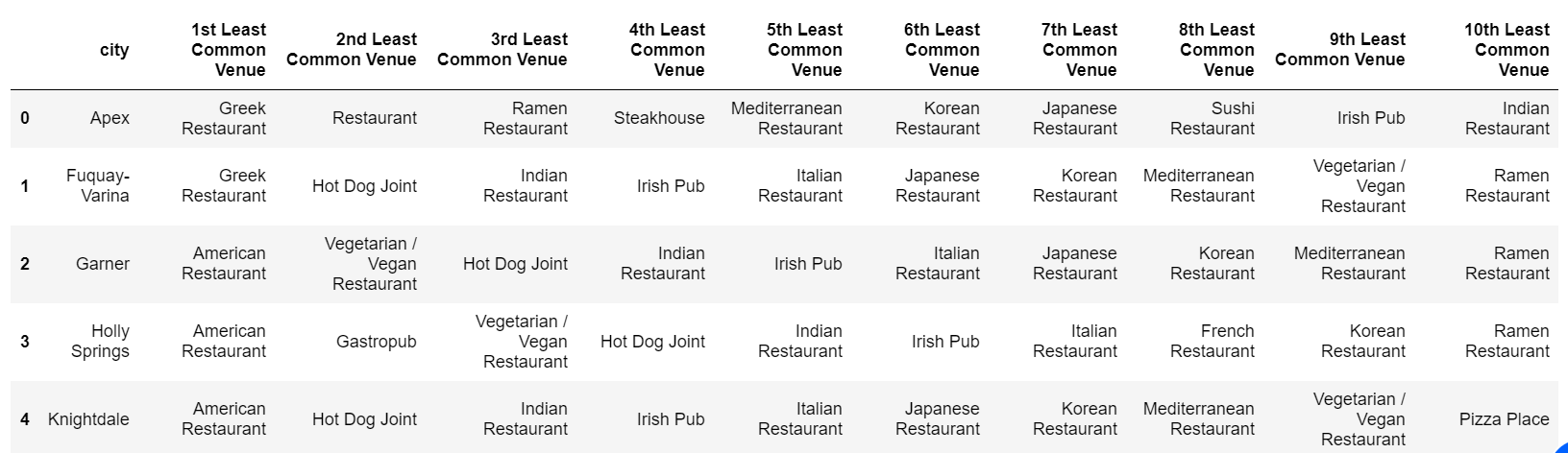


This can be displayed in a bar chart using matplot library as shown below:

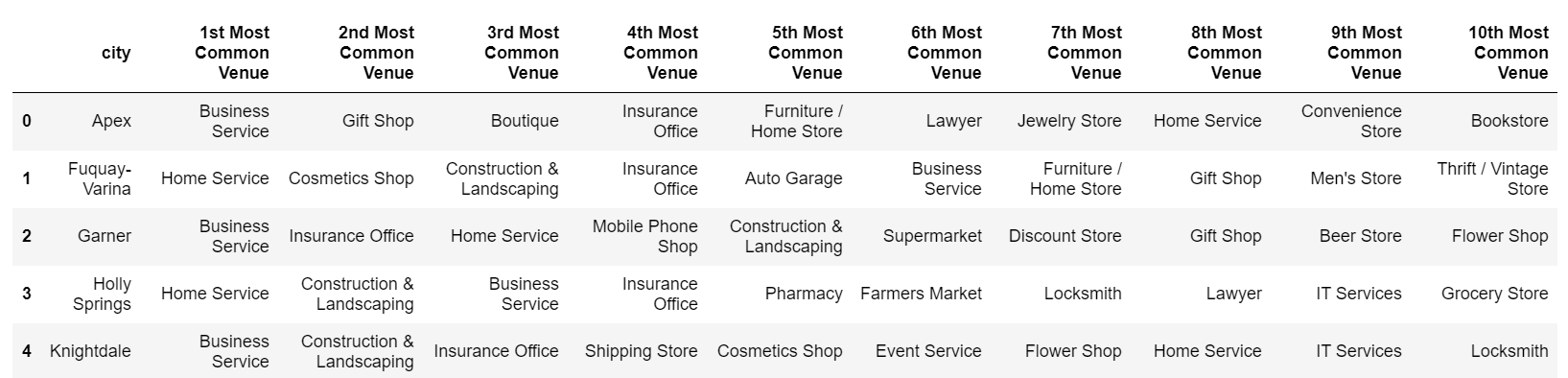


## Statistical Data Analysis

With the data analyzed using visualization technique, statistical analysis of the data is performed to identify least common food venues by city. A snapshot of least common food venues by city is shown below:



Similarly, information of most common venues for each city is compiled:



## Clustering

To further analyze how the restaurants are distributed, the cities can be divided into clusters to find out locations where maximum restaurants are clustered. This is done using Kmeans clustering unsupervised algorithm.

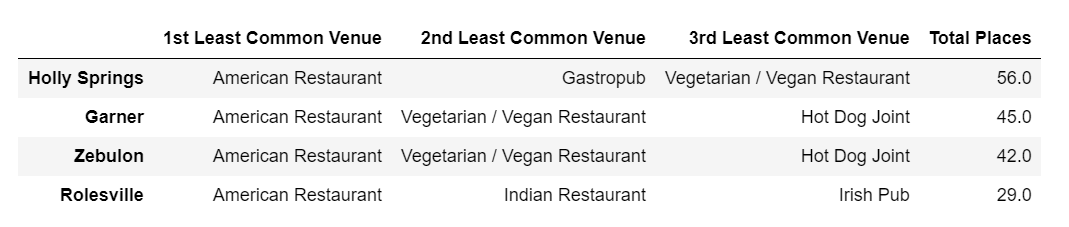


Similarly, to analyze how the shops are distributed, the cities can be divided into clusters to find out locations where maximum businesses are clustered using Kmeans clustering unsupervised algorithm.



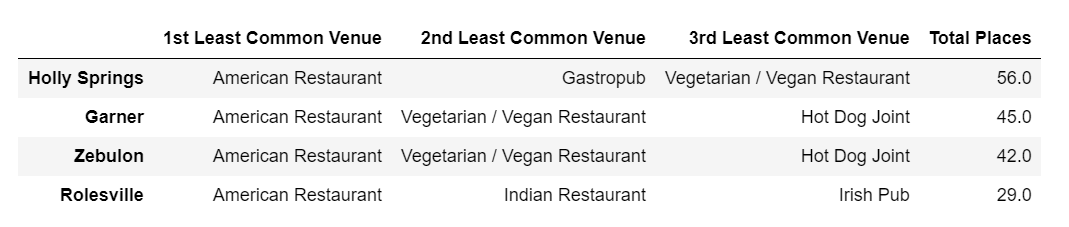
# Results

From the analysis performed in the previous sections and based on the clusters that show the cities with highest number of food venues and other businesses, identify the least common food businesses in cities with relatively higher number of businesses. This is achieved by combining the data sets created in earlier sets. The result is shown below.



# Discussion

Using the city data in UNC website & the businesses information from Foursquare API, applying Visualization techniques & Data Sciences methodologies and using Machine learning techniques, the least common food businesses in cities with relatively higher number of businesses are identified. These results & findings will help the businessman in deciding upon a location and the type of restaurant he can start for achieving maximum profits. He can select one of the Cities with higher number of food, other businesses and one of the least common businesses mentioned in the results section.



This data will be useful for any other investors/firms interested in starting a restaurant in Wake county.

# Conclusion

***Based on the analysis of the cities in Wake county and of various businesses in those cities, the recommendation for the businessman is to start an American Restaurant in the city of Holly Springs, NC.***