

GEORGIA RESTAURANTS – AN ANALYSIS

Introduction and Business Problem

Background:

While traveling in the southern states of the US in general, and in the state of Georgia in specific, it is usually not easy to find the right kind of restaurants to one's need and taste. Especially while traveling through some remote regions in Georgia, it is very handy to know what kind of restaurants one can expect in the neighborhood. Furthermore, there are different towns and cities in Georgia, which are popular for authentic restaurants of specific cuisines. It will be useful to learn more about their specialties before traveling in order not to miss out local specialties. This capstone project will focus on the state of Georgia in USA, analyzing the various categories of restaurants and food options available for travelers and localities in different towns and cities.

Problem statements:

We will attempt to address the following problem statements in this project, all in the geographical region of Georgia, USA

1. What categories of restaurants are popular in different parts of the state?
2. Are vegan/vegetarian options available in the state? Where are they located?
3. Where can one find a cafe or quick bite?
4. Are there specialty southern restaurants in the state? Where are they located?
5. Which towns do not have dine-in restaurants? (potentially not traveler friendly to stop for dine-ins)

Data Sources

In order to find answers and get insight into the restaurants landscape in Georgia, USA, the first piece of data that we require is a list of all towns and cities in Georgia, USA. This is a public data available either in the Georgia government site or in the Wikipedia page. In the project, we will scrape the towns and cities table information available in the Wikipeage: [https://en.wikipedia.org/wiki/List_of_municipalities_in_Georgia_\(U.S._state\)](https://en.wikipedia.org/wiki/List_of_municipalities_in_Georgia_(U.S._state)). The next piece of information we need is to get the geolocation (latitude and longitude) of all the towns and cities we scraped in the first step. In order to get this information, we will use the Geocoder Nominatim OpenStreepMap (OSM) API to get the geographical location of each town/city. Lastly, we will utilize FourSquare crowdsource data in order to get all the restaurant and category details of all the venues of type "food". We will restrict the number of venues by searching for venues within a radius of 5kms around each town/city.

To summarize, we will pull the following data:

1. **List of all towns and cities in Georgia, USA**

Source: [https://en.wikipedia.org/wiki/List_of_municipalities_in_Georgia_\(U.S._state\)](https://en.wikipedia.org/wiki/List_of_municipalities_in_Georgia_(U.S._state))

Pull table data from class witable sortable

Example (Name, Type, County):

Alpharetta City Fulton

Savannah City Chatham

2. **Geolocation (latitude and longitude) of each town and city - Using Geocoder Nominatim OSM API**

Package: geocoder Nominatim

Example:

Name latitude longitude

0 Abbeville, GA 31.992122 -83.306824

1 Acworth, GA 34.065933 -84.676880

3. **Restaurants and food venues - From Foursquare crowdsource data, using 'Venues' endpoint**

Endpoint: Foursquare venue explore

endpoint <https://api.foursquare.com/v2/venues/explore>

Parameters: Radius = 5kms, Limit venues per call = 100, section = 'Food'

Example:

Neighborhood Neighborhood Latitude Neighborhood Longitude Venue Venue

Latitude Venue Longitude Venue Category

Abbeville, GA 31.992122 -83.306824 Ophelias 31.992405 -83.306903 Diner

Abbeville, GA 31.992122 -83.306824 Country Kitchen 31.992640 -83.307734 Food

Abbeville, GA 31.992122 -83.306824 Mr B-b-q 31.993763 -83.295976 Food

Acworth, GA 34.065933 -84.676880 Henry's Louisiana Grill 34.066011 -84.677728

Cajun / Creole Restaurant

Acworth, GA 34.065933 -84.676880 Fusco's via Roma 34.065781 -84.677163 Italian

Restaurant

Combining and blending the data from the 3 data sources stated above, we can get the list of venues and its categorization for each city/town. We will use this as the input to the machine learning model K-means clustering for unsupervised data and build clusters of similar towns/cities offering similar categories of restaurants. Once clusters are formed, we will analyze and label the clusters based on the venue categories within each cluster. Using the cluster and labels, we will attempt to answer all the problem statements.

Data Analysis

Data Analysis - Exploratory, descriptive, and statistical

Using BeautifulSoup package, the wikitable data having all the municipalities (cities and towns) have been loaded into Pandas dataframe. An initial description of this dataframe shows:

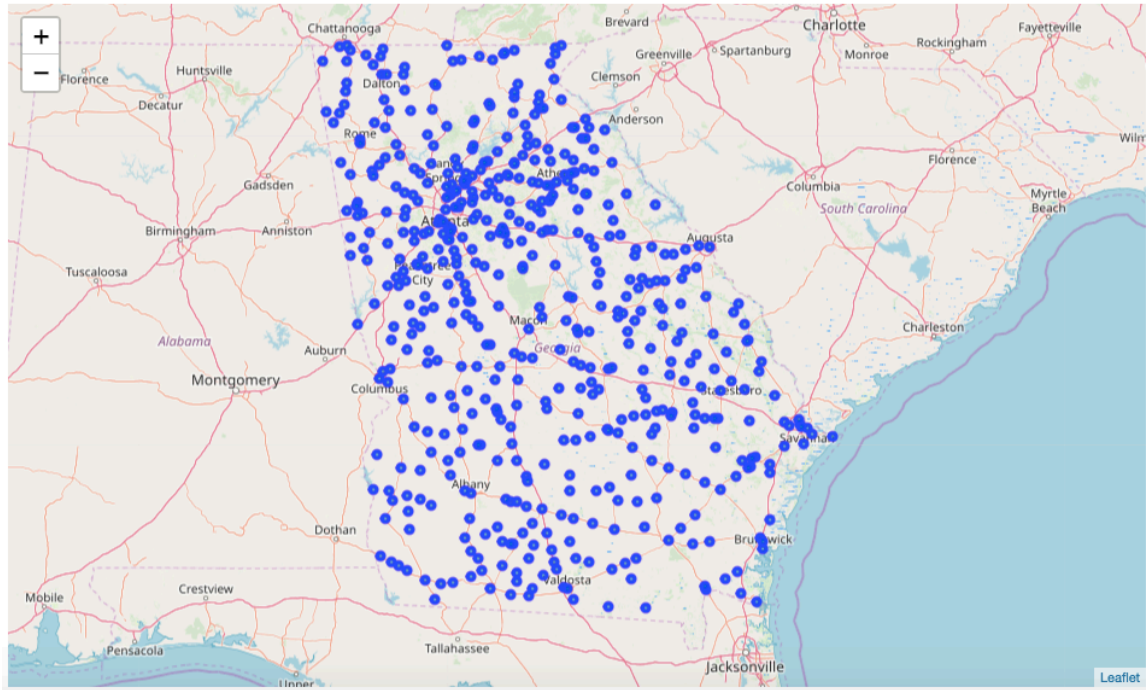
Type	Total
City	409
Consolidated city	2
Town	122
Unified government	6
Grand total	539

Using Geocoder Nominatim package, loop through each town/city in the list, and fetch the geolocation (latitude and longitude).

Name	latitude	longitude
Abbeville, GA	31.992122	-83.306824
Acworth, GA	34.065933	-84.67688

Adairsville, GA	34.368702	-84.934109
Adel, GA	31.137136	-83.423494
Adrian, GA	32.530722	-82.589299

Using folium, visualize all the towns/cities in the state of Georgia



Using FourSquare (venue explore API), all the venues of Food category in the vicinity of each of these towns/cities are obtained.

Parameters used:

- Maximum venues per town/city = 100
- Radius from lat/long = 5kms
- Section (category of venue) = Food

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Abbeville, GA	31.992122	-83.306824	Ophelias	31.992405	-83.306903	Diner
1	Abbeville, GA	31.992122	-83.306824	Country Kitchen	31.992640	-83.307734	Food
2	Abbeville, GA	31.992122	-83.306824	Mr B-b-q	31.993763	-83.295976	Food
3	Acworth, GA	34.065933	-84.676880	Henry's Louisiana Grill	34.066011	-84.677728	Cajun / Creole Restaurant
4	Acworth, GA	34.065933	-84.676880	Fusco's via Roma	34.065781	-84.677163	Italian Restaurant
5	Acworth, GA	34.065933	-84.676880	Miss L's Sandwich Shop	34.065861	-84.677276	Sandwich Place
6	Acworth, GA	34.065933	-84.676880	Downtown Acworth	34.066043	-84.677738	American Restaurant
7	Acworth, GA	34.065933	-84.676880	Bojangles' Famous Chicken 'n Biscuits	34.074136	-84.676541	Fast Food Restaurant
8	Acworth, GA	34.065933	-84.676880	Fish Thyme	34.050457	-84.663369	Seafood Restaurant
9	Acworth, GA	34.065933	-84.676880	J.D.'s BBQ	34.062048	-84.665742	BBQ Joint

There are a total of 13118 venues, and the total distinct venue categories is 102

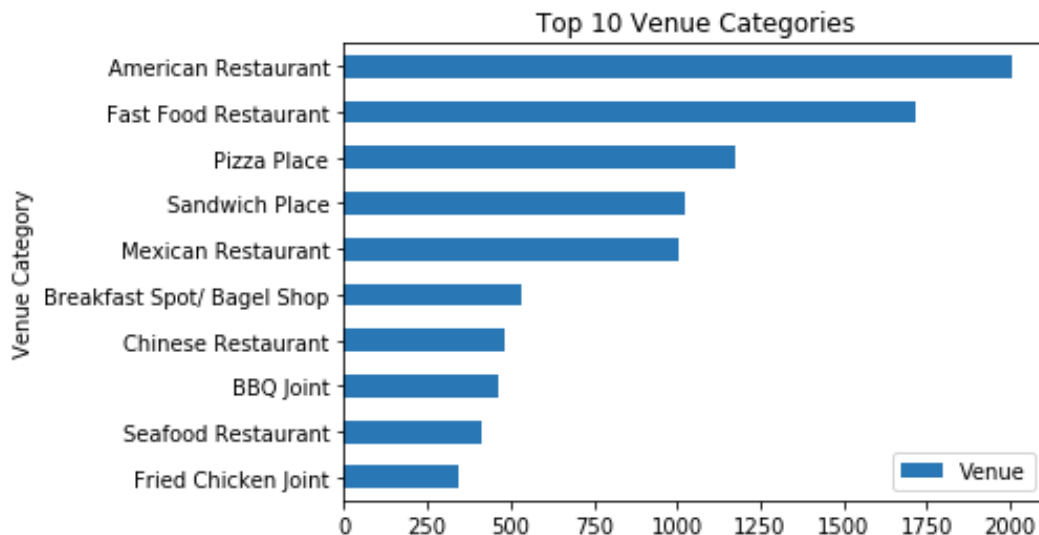
Since the FourSquare data is crowd-sourced, it is prone to have errors and mis-classifications. It is important to perform data cleansing including addressing missing values, incorrect category classification, and delete incorrect entries.

A few examples include:

- Peruvian, Colombian, Argentinian, Brazilian, Latin American, and Venezuelan - need to be reclassified into South American Restaurants
- South Indian, north Indian, Chaat Place - need to be reclassified into Indian Restaurants
- Mediterranean, Greek, Turkish, and Middle Eastern - need to be classified together

After all the cleansing is applied, we have resulting distinct venue categories of **78**

Below are the top 10 venue categories



Feature set

The key to our feature set is applying one-hot encoding technique. We will use the mean of the frequency of occurrence of each category. This will help us arrive at a numerical score for each of the 79 categories, for each town/city.

Since the distribution of the venue category beyond the top 10 categories for each town/city will have very less correlation to what kind of restaurants are offered in the specific town/city, we will generate the top 10 most common venue categories and choose these 10 fields as our feature set.

	Neighborhood	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	10th Most Common Venue
0	Abbeville, GA	Diner	American Restaurant	BBQ Joint	Fish & Chips Shop	Dumpling Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Fast Food Restaurant
1	Acworth, GA	American Restaurant	Fast Food Restaurant	Pizza Place	Sandwich Place	Mexican Restaurant	BBQ Joint	Seafood Restaurant	Chinese Restaurant	Breakfast Spot/ Bagel Shop	Italian Restaurant
2	Adairsville, GA	Fast Food Restaurant	American Restaurant	Asian Restaurant	Mexican Restaurant	Burger Joint	Wings Joint	Fried Chicken Joint	Pizza Place	Sandwich Place	Japanese Restaurant
3	Adel, GA	Fast Food Restaurant	Pizza Place	American Restaurant	Breakfast Spot/ Bagel Shop	Steakhouse	Seafood Restaurant	Japanese Restaurant	Chinese Restaurant	Sandwich Place	Mexican Restaurant
4	Adrian, GA	Bakery	Wings Joint	Fish & Chips Shop	Dumpling Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Fast Food Restaurant	Filipino Restaurant
5	Ailey, GA	Chinese Restaurant	American Restaurant	Sandwich Place	Wings Joint	Filipino Restaurant	Dumpling Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant
6	Alamo, GA	Burger Joint	Café	Donut Shop	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Fast Food Restaurant	Filipino Restaurant	Fish & Chips Shop
7	Alapaha, GA	American Restaurant	Wings Joint	Fish & Chips Shop	Dumpling Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Fast Food Restaurant	Filipino Restaurant

At this time, we found an additional 68 town/cities for which Foursquare does not have any reported food venues. We remove these 68 towns/cities from our clustering model.

Methodology

The dataset for these set of problem statements do not have a predefined label or the primary catogories that a given town/city belongs to. Hence, this dataset is classified as **unsupervised** data. The model we select to solve this particular problem is **K-means clustering for unsupervised dataset**.

We use the K-means clustering algorithm, and label all the towns/cities into **10** clusters.

Here are sample data from some of the clusters generated by our model:

Cluster 1

	Name	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	10th Most Common Venue
1	Acworth, GA	American Restaurant	Fast Food Restaurant	Pizza Place	Sandwich Place	Mexican Restaurant	BBQ Joint	Seafood Restaurant	Chinese Restaurant	Breakfast Spot/ Bagel Shop	Italian Restaurant
2	Adairsville, GA	Fast Food Restaurant	American Restaurant	Asian Restaurant	Mexican Restaurant	Burger Joint	Wings Joint	Fried Chicken Joint	Pizza Place	Sandwich Place	Japanese Restaurant
3	Adel, GA	Fast Food Restaurant	Pizza Place	American Restaurant	Breakfast Spot/ Bagel Shop	Steakhouse	Seafood Restaurant	Japanese Restaurant	Chinese Restaurant	Sandwich Place	Mexican Restaurant
8	Albany, GA	Fast Food Restaurant	American Restaurant	Fried Chicken Joint	Seafood Restaurant	Chinese Restaurant	Sandwich Place	BBQ Joint	Wings Joint	Hot Dog Joint	Italian Restaurant
9	Aldora, GA	American Restaurant	Pizza Place	Fast Food Restaurant	BBQ Joint	African Restaurant	Mexican Restaurant	Breakfast Spot/ Bagel Shop	Steakhouse	Chinese Restaurant	Southern / Soul Food Restaurant
10	Allenhurst, GA	Chinese Restaurant	Fast Food Restaurant	Sandwich Place	Korean Restaurant	Breakfast Spot/ Bagel Shop	Pizza Place	Seafood Restaurant	Steakhouse	Donut Shop	American Restaurant
12	Alma, GA	Pizza Place	Fast Food Restaurant	Wings Joint	Fried Chicken Joint	Sandwich Place	Buffet	Diner	BBQ Joint	American Restaurant	Gastropub

Cluster 5

	Name	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	10th Most Common Venue
21	Arcade, GA	Seafood Restaurant	American Restaurant	Wings Joint	Filipino Restaurant	Dumpling Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Fast Food Restaurant
169	Edge Hill, GA	Seafood Restaurant	Wings Joint	Dim Sum Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Fast Food Restaurant
316	Midville, GA	Seafood Restaurant	BBQ Joint	Wings Joint	Filipino Restaurant	Dumpling Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Fast Food Restaurant
317	Midway, GA	Seafood Restaurant	American Restaurant	Pizza Place	Bakery	Mexican Restaurant	Sandwich Place	Diner	Wings Joint	Ethiopian Restaurant	Dumpling Restaurant
364	Orchard Hill, GA	American Restaurant	Seafood Restaurant	Wings Joint	Filipino Restaurant	Dumpling Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Fast Food Restaurant
381	Pitts, GA	Diner	Seafood Restaurant	Dim Sum Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Fast Food Restaurant
484	Toombsboro, GA	Seafood Restaurant	Wings Joint	Dim Sum Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Fast Food Restaurant

Cluster 6

	Name	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	10th Most Common Venue
4	Adrian, GA	Bakery	Wings Joint	Fish & Chips Shop	Dumpling Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Fast Food Restaurant	Filipino Restaurant
45	Berlin, GA	Bakery	Wings Joint	Fish & Chips Shop	Dumpling Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Fast Food Restaurant	Filipino Restaurant
66	Bronwood, GA	Bakery	Wings Joint	Fish & Chips Shop	Dumpling Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Fast Food Restaurant	Filipino Restaurant

Cluster 8

	Name	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	10th Most Common Venue
15	Alto, GA	BBQ Joint	Spanish Restaurant	Sandwich Place	Diner	Filipino Restaurant	Dumpling Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant
46	Bethlehem, GA	BBQ Joint	Bakery	Breakfast Spot/ Bagel Shop	Wings Joint	Food Truck/ Food Stand	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Fast Food Restaurant
74	Buena Vista, GA	BBQ Joint	Sandwich Place	Mexican Restaurant	Pizza Place	Wings Joint	Ethiopian Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Empanada Restaurant
112	Colbert, GA	Hot Dog Joint	BBQ Joint	Bakery	Seafood Restaurant	Wings Joint	Filipino Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant
191	Fort Gaines, GA	BBQ Joint	Fried Chicken Joint	Wings Joint	Fish & Chips Shop	Dumpling Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Fast Food Restaurant
333	Morven, GA	BBQ Joint	Wings Joint	Fish & Chips Shop	Dumpling Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Fast Food Restaurant	Filipino Restaurant
360	Offerman, GA	BBQ Joint	Fried Chicken Joint	American Restaurant	Breakfast Spot/ Bagel Shop	Food Truck/ Food Stand	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Fast Food Restaurant
368	Patterson, GA	Hot Dog Joint	BBQ Joint	Bakery	Breakfast Spot/ Bagel Shop	Food Truck/ Food Stand	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Fast Food Restaurant

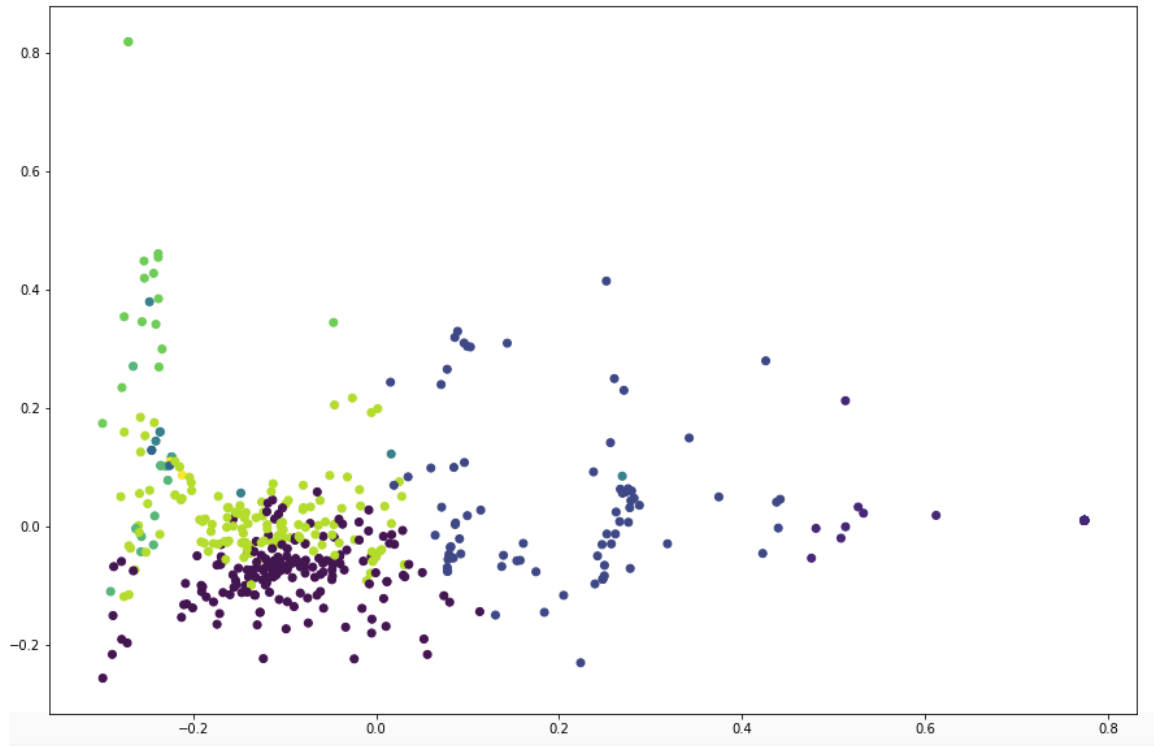
Cluster 10

	Name	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	10th Most Common Venue
205	Girard, GA	Southern / Soul Food Restaurant	Wings Joint	Dim Sum Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Fast Food Restaurant
259	Jersey, GA	Southern / Soul Food Restaurant	Malay Restaurant	Wings Joint	Dumpling Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Fast Food Restaurant	Filipino Restaurant
385	Portal, GA	Southern / Soul Food Restaurant	Wings Joint	Dim Sum Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Fast Food Restaurant

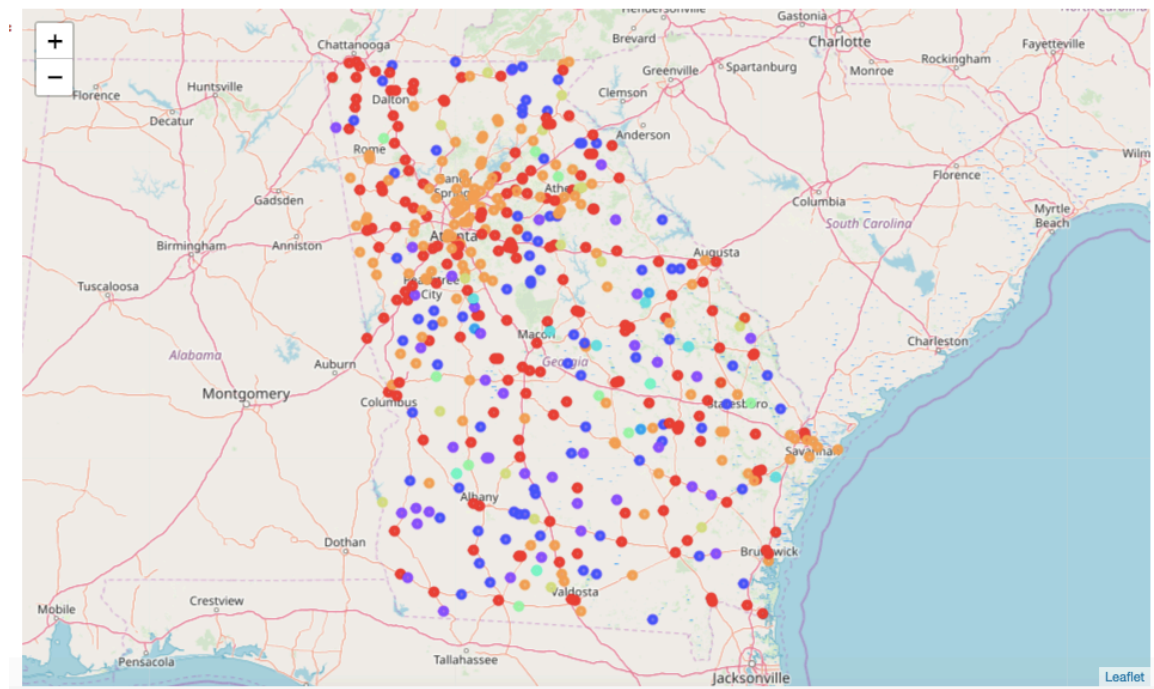
Results

Once we build the clusters, we visualize the clusters through

1) **A scatter plot by converting all features and target into 2-dimensional form**



2) Folium to visualize the clusters and its distribution on the map



Discussion and Conclusion

This section elaborates the answers and findings for all the problem statements defined earlier.

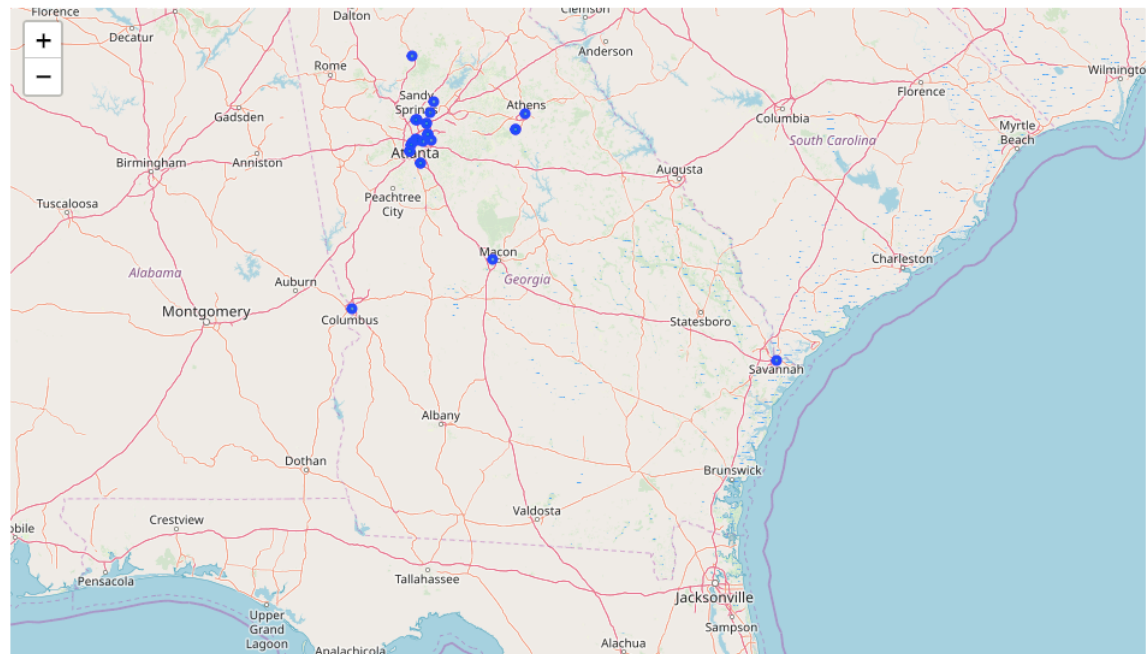
1. What categories of restaurants are popular in different parts of the state?

In the 10 clusters we generated, we were able to find unique restaurant categories. The most popular restaurant categories include American, fast food and Pizza place. They are distributed in clusters 1,2,3, and 9. Below table summarizes the different restaurant categories that are popular in different parts of the state.

Cluster 1	American, Fast Food and Pizza Place
Cluster 2	American, Wings Joint and Fish & Chips Shop
Cluster 3	American Restaurants
Cluster 4	Diner
Cluster 5	Seafood
Cluster 6	Bakery
Cluster 7	Sandwich Place and Café
Cluster 8	BBQ Joint
Cluster 9	American and Mexican Restaurants
Cluster 10	Southern / Soul Food Restaurant

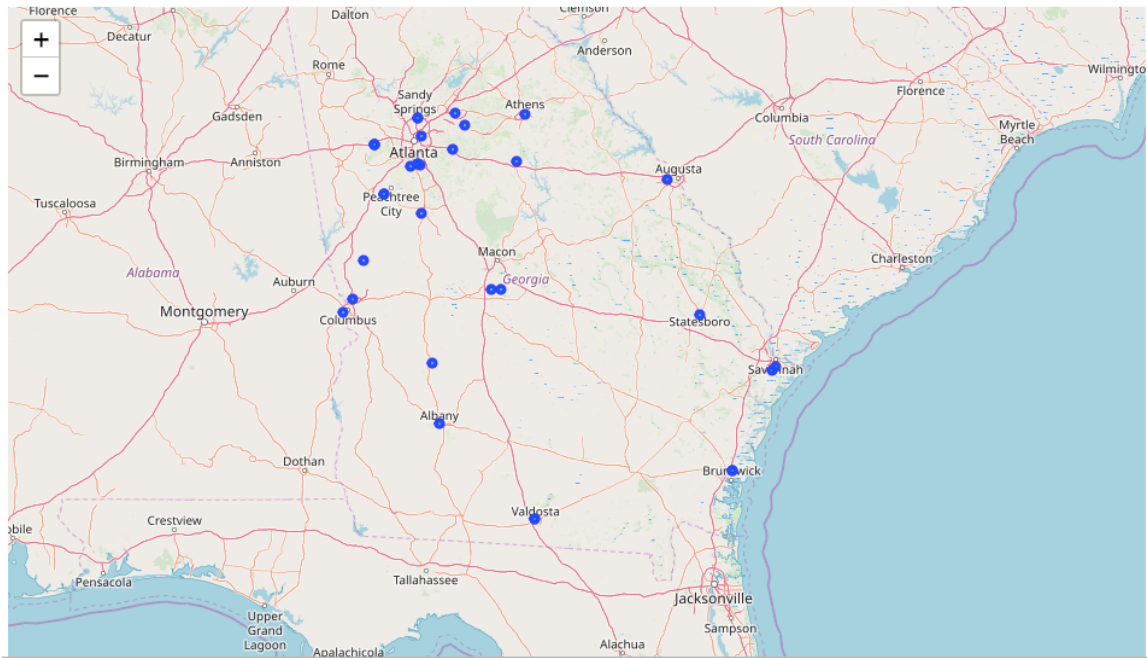
2. Are vegan/vegetarian options available in the state? Where are they located?

There are a total of only 31 vegetarian/vegan restaurants checked in by FourSquare users in the entire state of Georgia, out of 13118 venues. This shows that Georgia is not a veg/vegan friendly state. Moreover, the few options are centered around a few cities including Atlanta, Macon, Athens, Savannah, and Columbus.



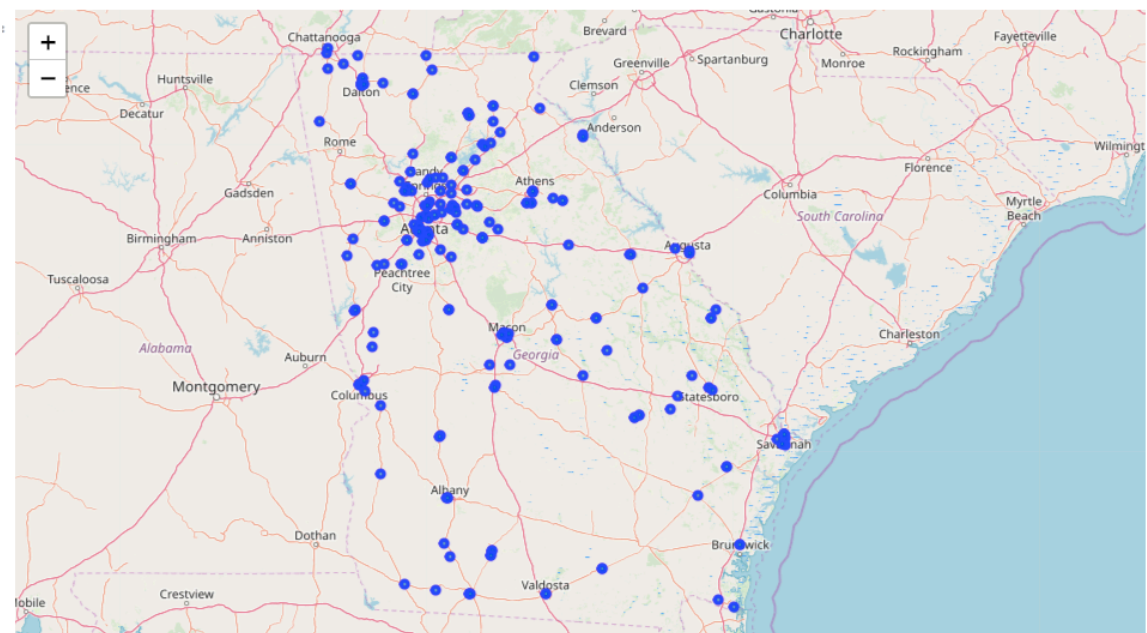
3. **Where can one find a cafe or quick bite?**

Cafe and Sandwich/Quick bite belongs to cluster 7 and are distributed thinly around the entire state



4. **Are there specialty southern restaurants in the state? Where are they located?**

There are 231 restaurants classified as Southern. From the result and distribution, it is evident that Georgia is a typical southern state with many American and southern restaurants well distributed across the entire state.



5. **Which towns do not have dine in restaurants, potentially not traveler friendly to stop for dine-ins?**

There are 68 towns/cities from where FourSquare users not checked in and reported any restaurants. Perhaps, these are cities to avoid dine-in plans.

