# Final Capstone Project

"Analysis of Restaurants located in Georgia"

By Daniel Gould

Table of Contents

[Final Capstone Project 1](#_Toc9688449)

[Introduction 2](#_Toc9688450)

[Background: 2](#_Toc9688451)

[Problem Statement: 3](#_Toc9688452)

[Data Sources 4](#_Toc9688453)

[Data Analysis 5](#_Toc9688454)

[Data Analysis - Exploratory, descriptive and statistical 5](#_Toc9688455)

[Feature set 7](#_Toc9688456)

[Methodology 8](#_Toc9688457)

[Results 10](#_Toc9688458)

[Discussion and Conclusion 11](#_Toc9688459)

[What categories of restaurants are popular in different parts of the state? 11](#_Toc9688460)

[Are vegan/vegetarian options available in the state? Where are they located? 11](#_Toc9688461)

[Where can one find a cafe or quick bite? 12](#_Toc9688462)

[Are there specialty southern restaurants in the state? Where are they located? 12](#_Toc9688463)

[Which towns do not have dine-in restaurants, potentially not traveller friendly to stop for dine-ins? 13](#_Toc9688464)

# Introduction

# Background:

While traveling in the southern states of the United States (specifically in Georgia), it has been identified by the client that it can be difficult to locate restaurants that meet a specific person's tastes/needs. This capstone project will focus on the state of Georgia, analysing the various categories of restaurants and food options available for travellers and localities in the different towns and cities.

# Problem Statement:

During this exercise we will be aiming to use the data to address the following key questions:

1. What categories of restaurants are popular in different parts of the state?
2. What categories of restaurants are most popular across the different parts of Georgia?
3. Are there any vegan/vegetarian options available in Georgia? If so, where are they located?
4. Where can someone find a cafe or quick bite? (by this we mean fast-food)
5. Are there any restaurants that are based around southern cuisine? If so, where are they located?
6. Which towns do not have dine-in restaurants?

# Data Sources

To find the answers and get insight into the restaurants landscape in Georgia, United States, the first piece of data that is required is a list of all towns and cities located in Georgia. This is publicly available located in the Wikipedia page. In this project, we will scrape the towns and cities info from Wikipage: <https://en.wikipedia.org/wiki/List_of_municipalities_in_Georgia_(U.S._state>).

The next piece of information we require, is the geolocation (altitude and longitude) of all towns and cities from the first step. To extract this information, we will use the Geocoder Nominatim OpenStreetMap (OSM)API to get the geographical location of each town and city.

Last, we will utilize Foursquare crowdsource data to get all of 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 or and city.

To summarise, we will pull the following data for this project:

List of all town and cities in Georgia, United States Source: <https://en.wikipedia.org/wiki/List_of_municipalities_in_Georgia_(U.S._state>) Pull table data from wikitable sortable

Example: (Name, Type, Country): Alpharetta City Fulton Savannah City Chatham

Geolocation (Latitude & Longitude) of each town and city - using Geocoder Package: geocoder Nomination

Example: Name, Latitude, Longitude 0 Abbeville, GA 31.992122, -83.306824 1 Actworth, GA 34,065933, -84,676880

Restaurants and food venues - from FourSquare crowdsource data, using 'Venues' endpoint Endpoint: FourSquare venue explore endpoint: <https://api.foursquare.com/v2/venues/explore> Parrameters: Radius 5km, limit venues per call = 100, section = Food

Example: Neighbourhood, Neighbourhood Latitude, Neighbourhood 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 and 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 analyse 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

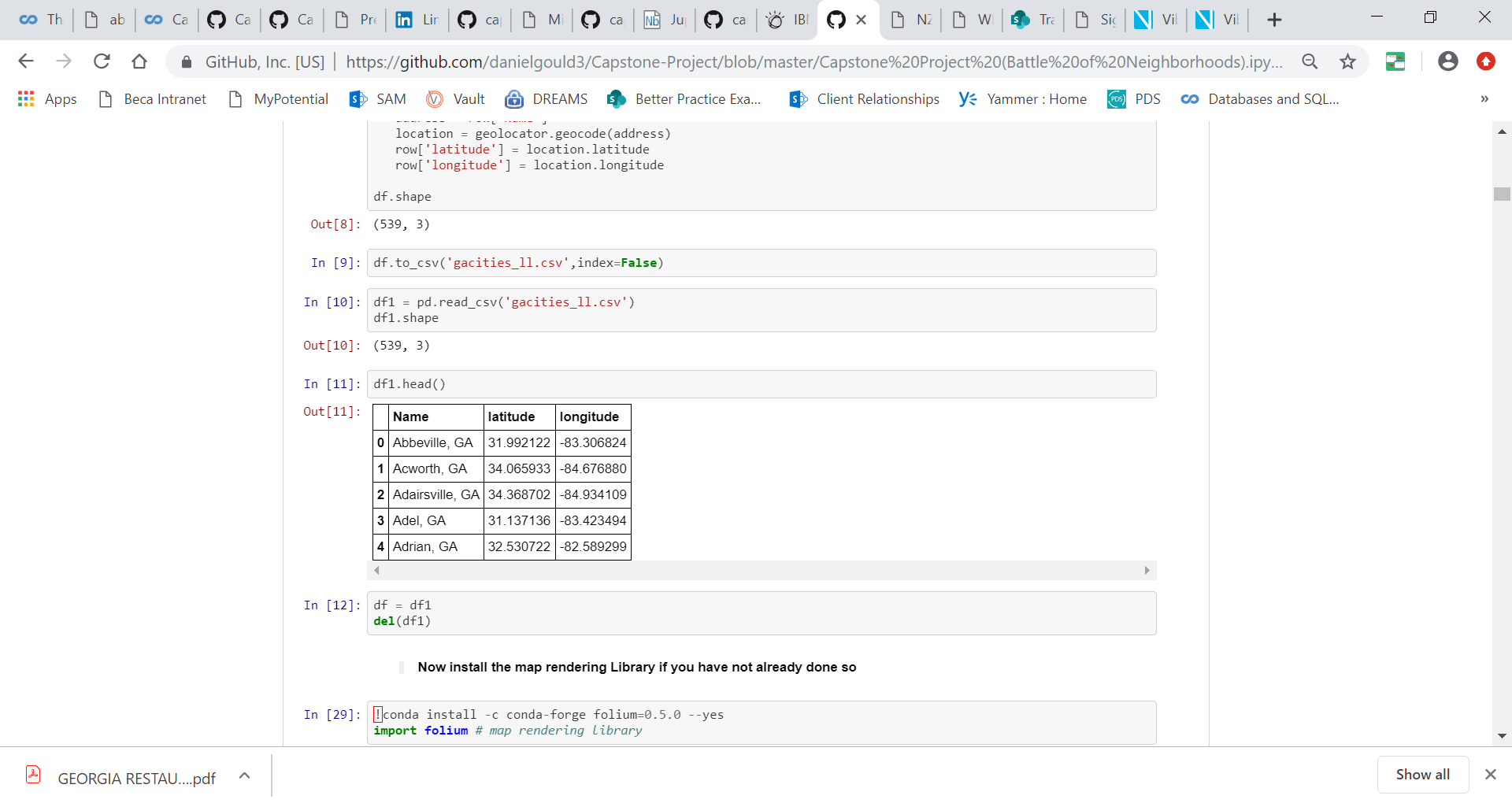
Unified government

6

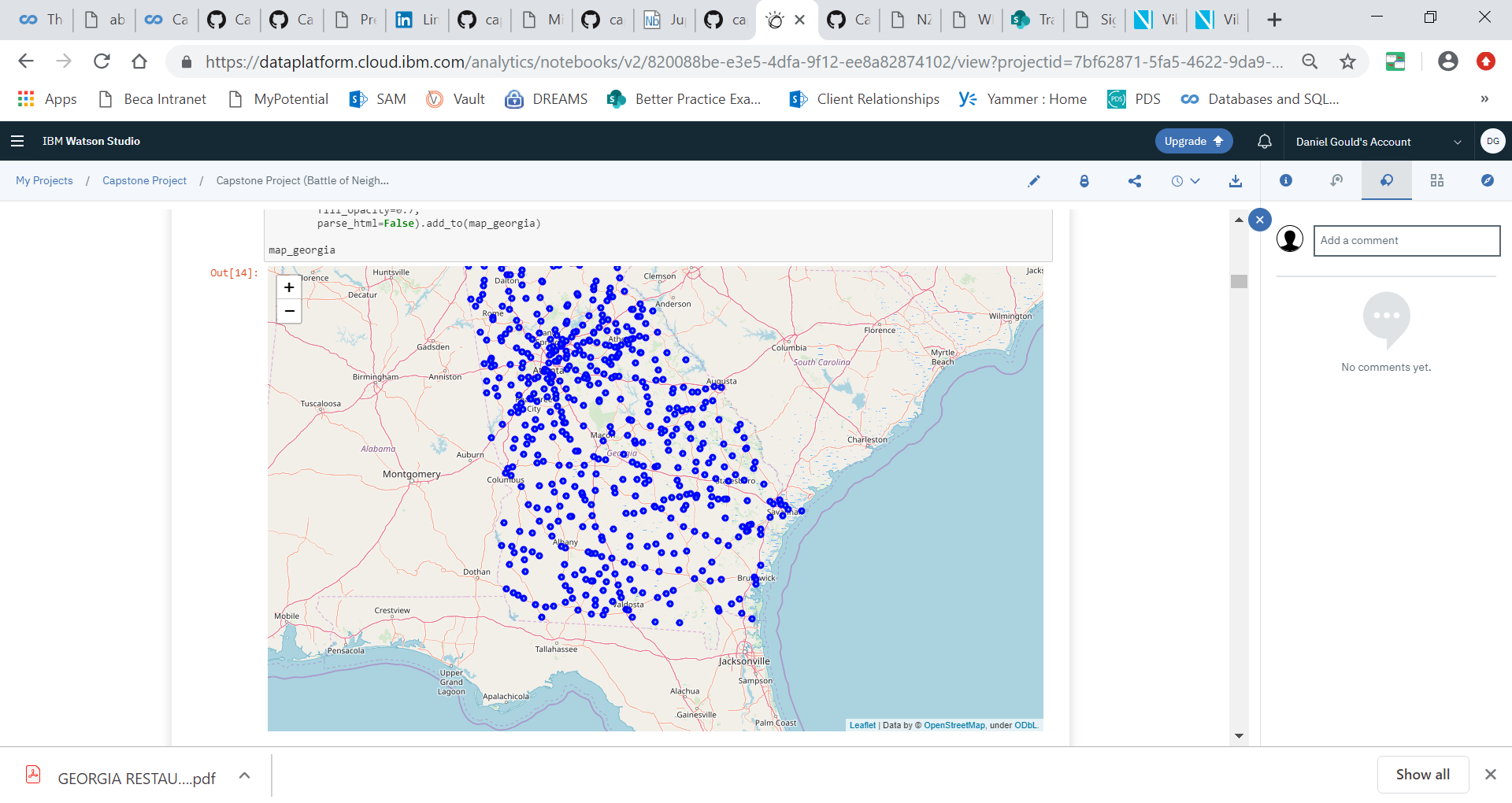
**Grand total**

**539**

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

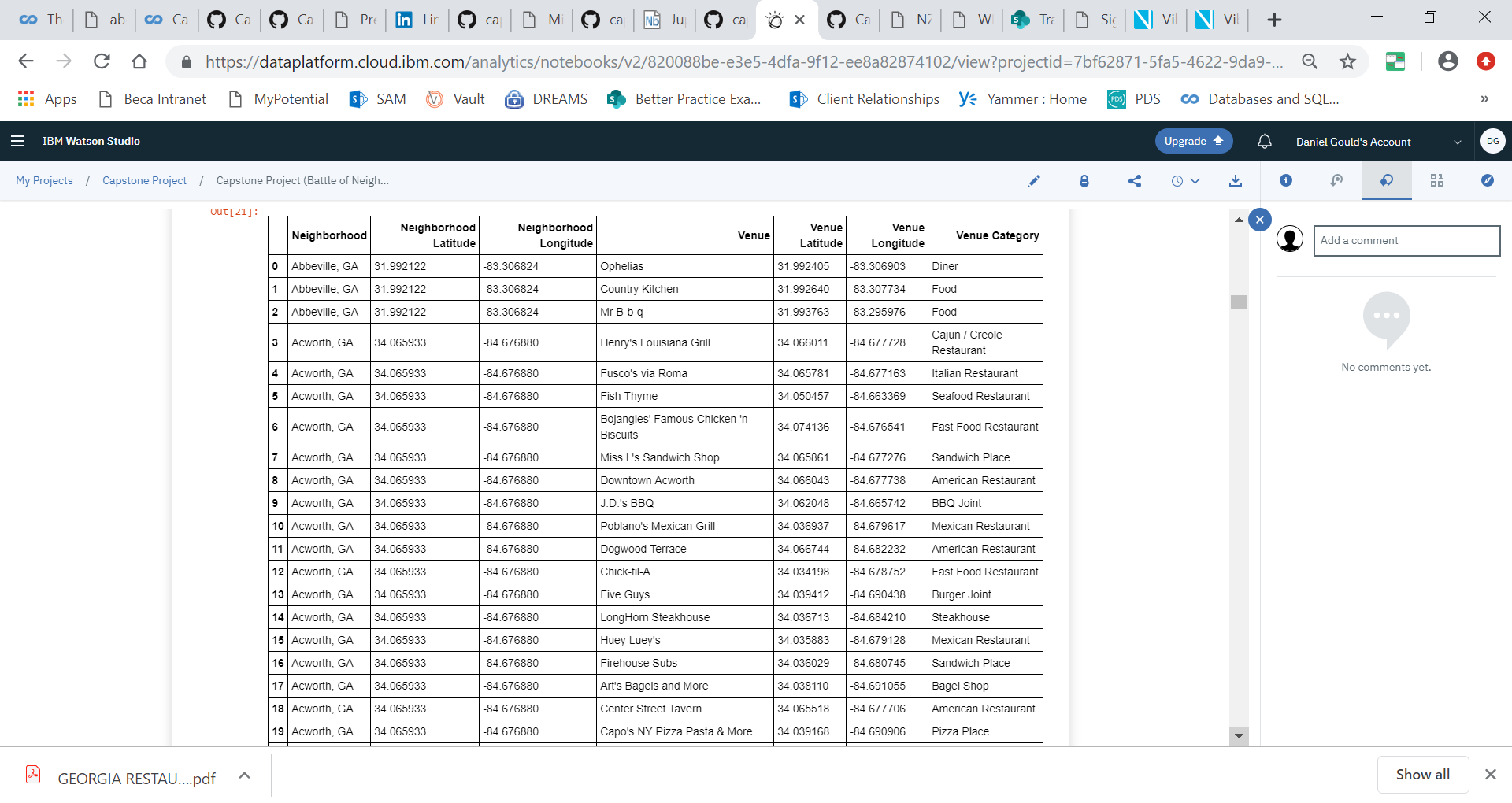


Using folium, visualize all the towns/cities in the state of Georgia



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



**There is a total of 13124 venues, and the total distinct venue categories is 102**

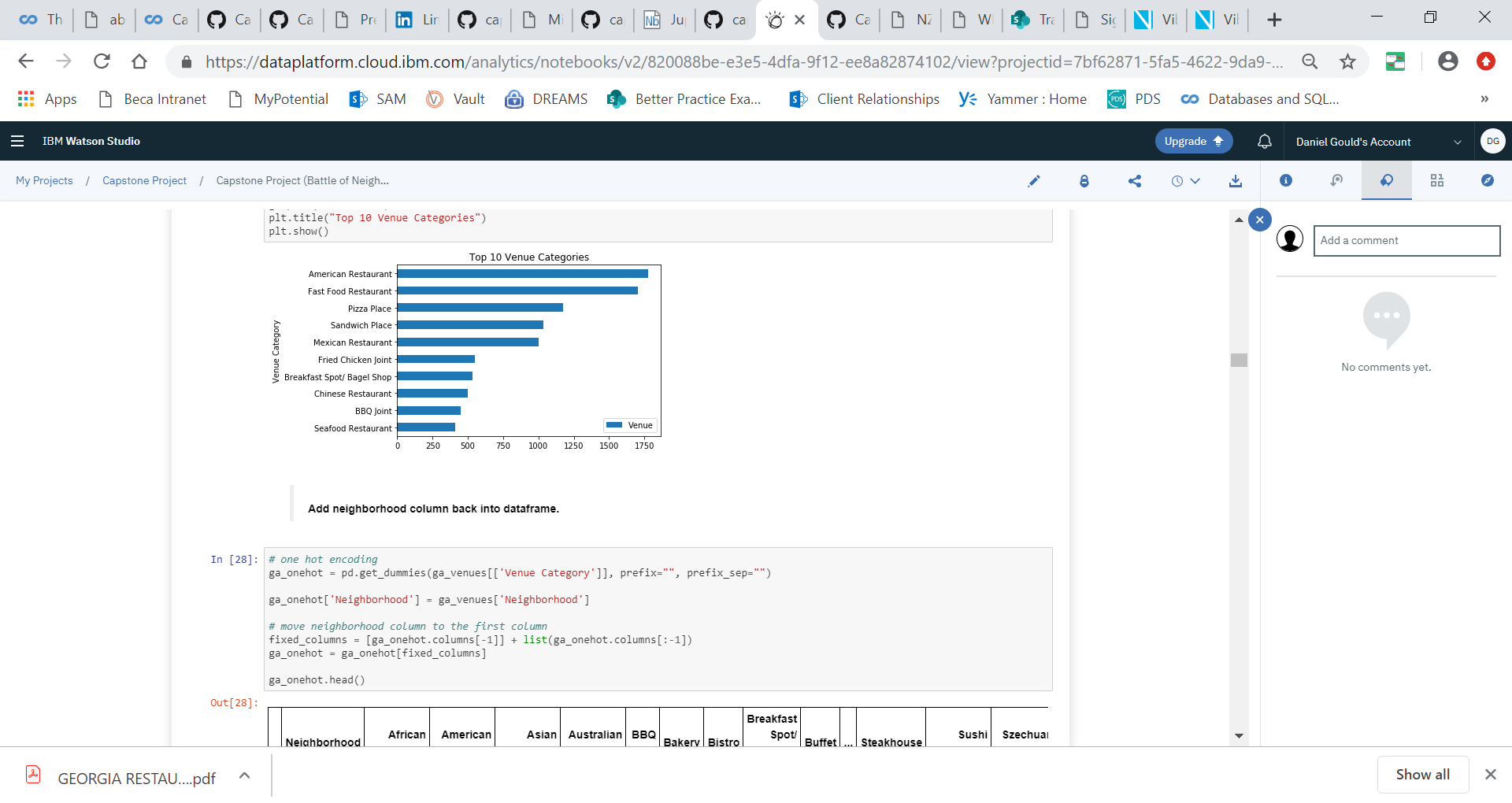
Since the FourSquare data is crowd-sourced, it is prone to have errors and misclassifications. It is important to perform data cleansing including addressing missing values, incorrect category classification, and delete incorrect entries.

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 79**.**

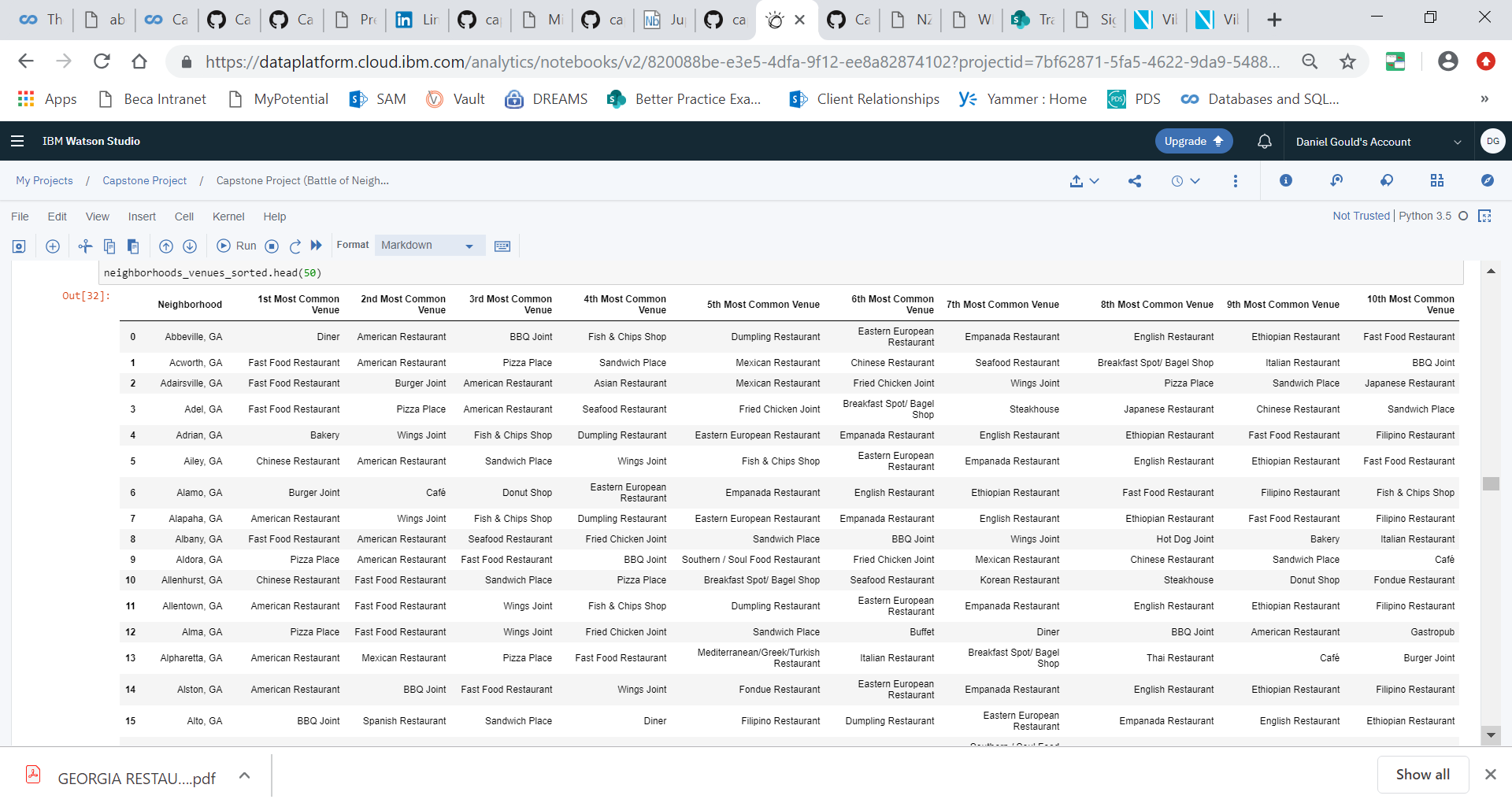
**Below are the top 10 venue categories**



# 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.



At this time, we found an additional 71 town/cities for which Foursquare does not have any reported food venues. We remove these 71 towns/cities from our clustering model.

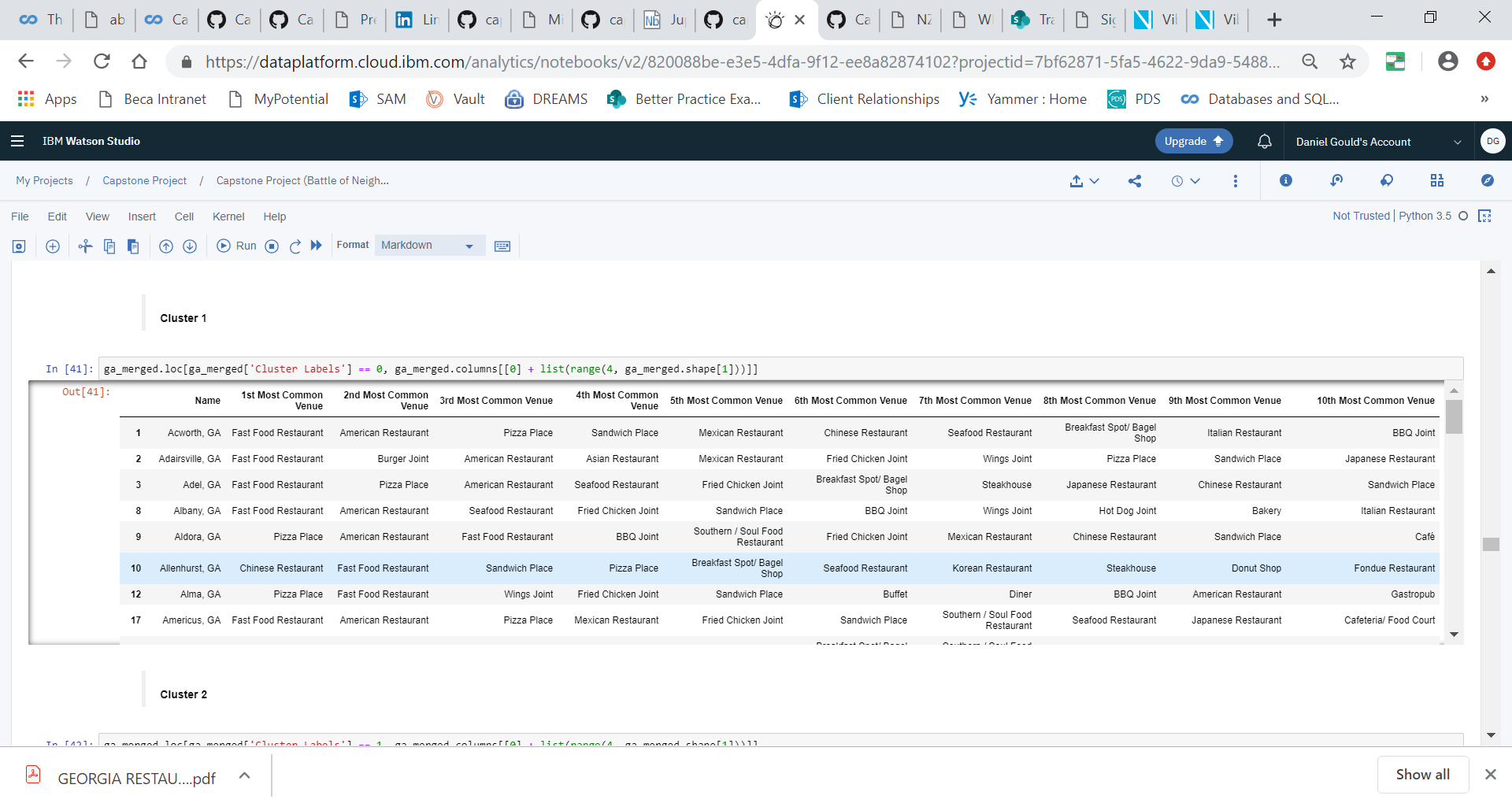
# Methodology

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

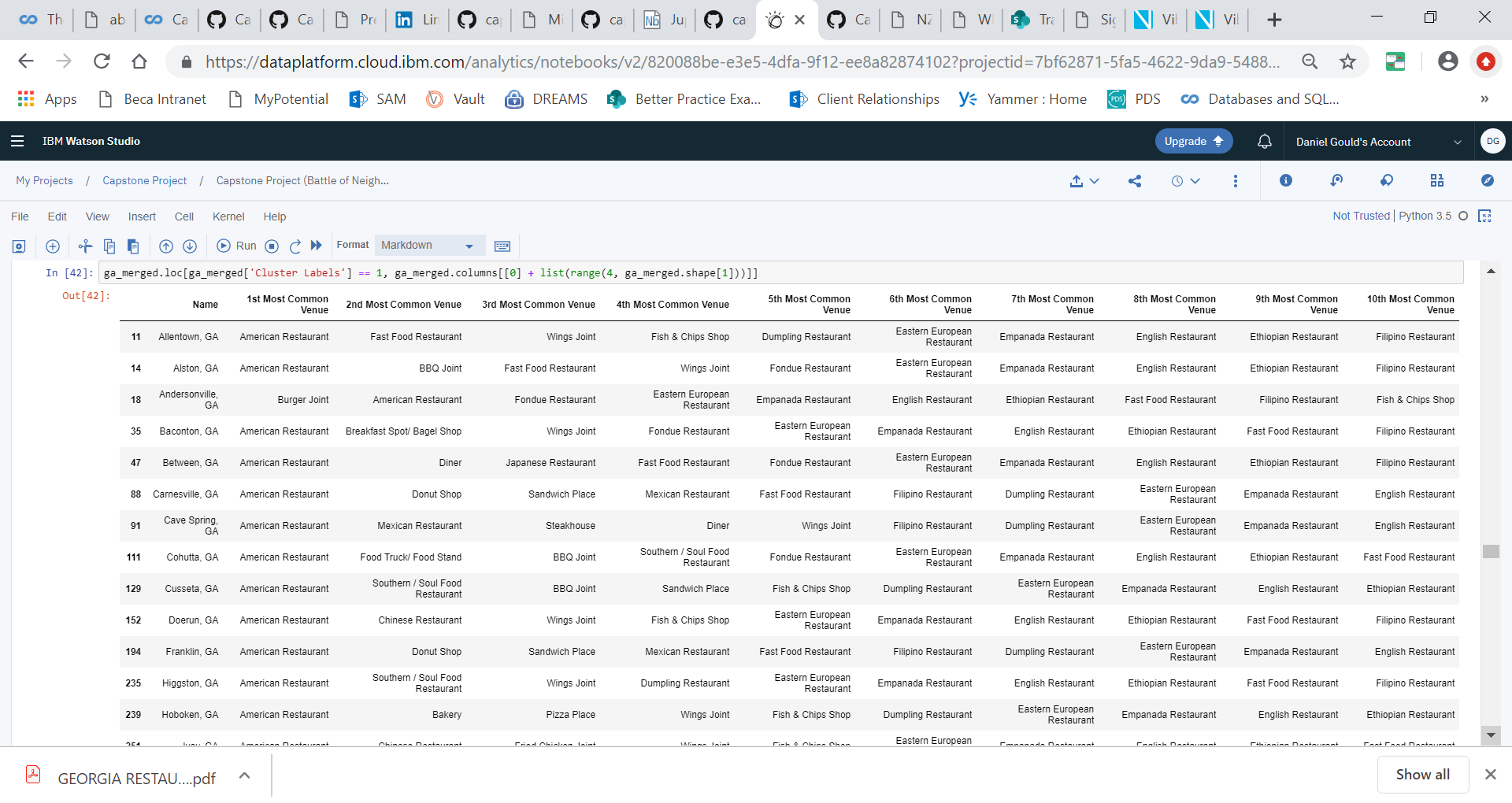
We use the K-means clustering algorithm and label all the towns/cities into **10** clusters.

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

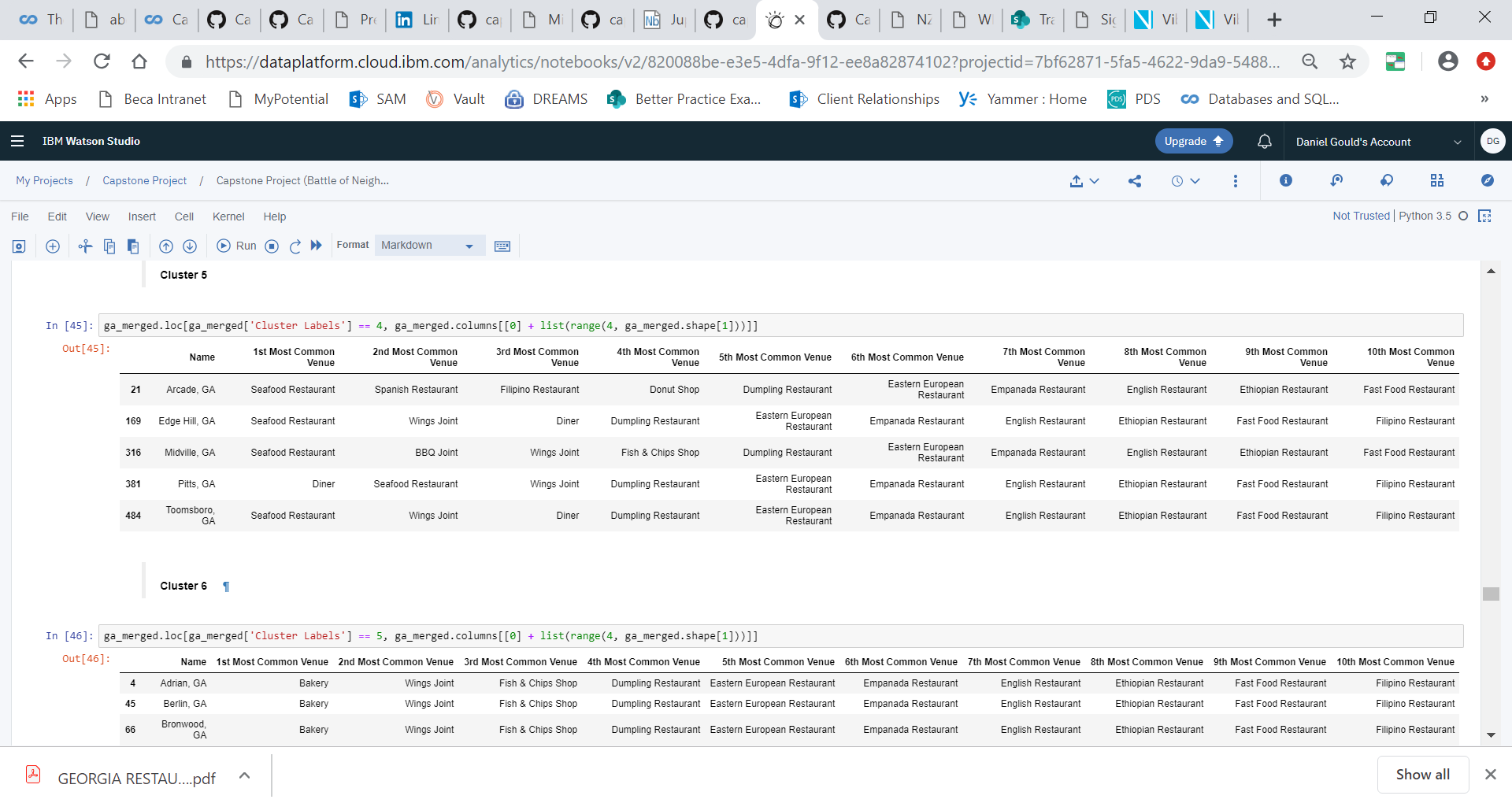
**Cluster 1**



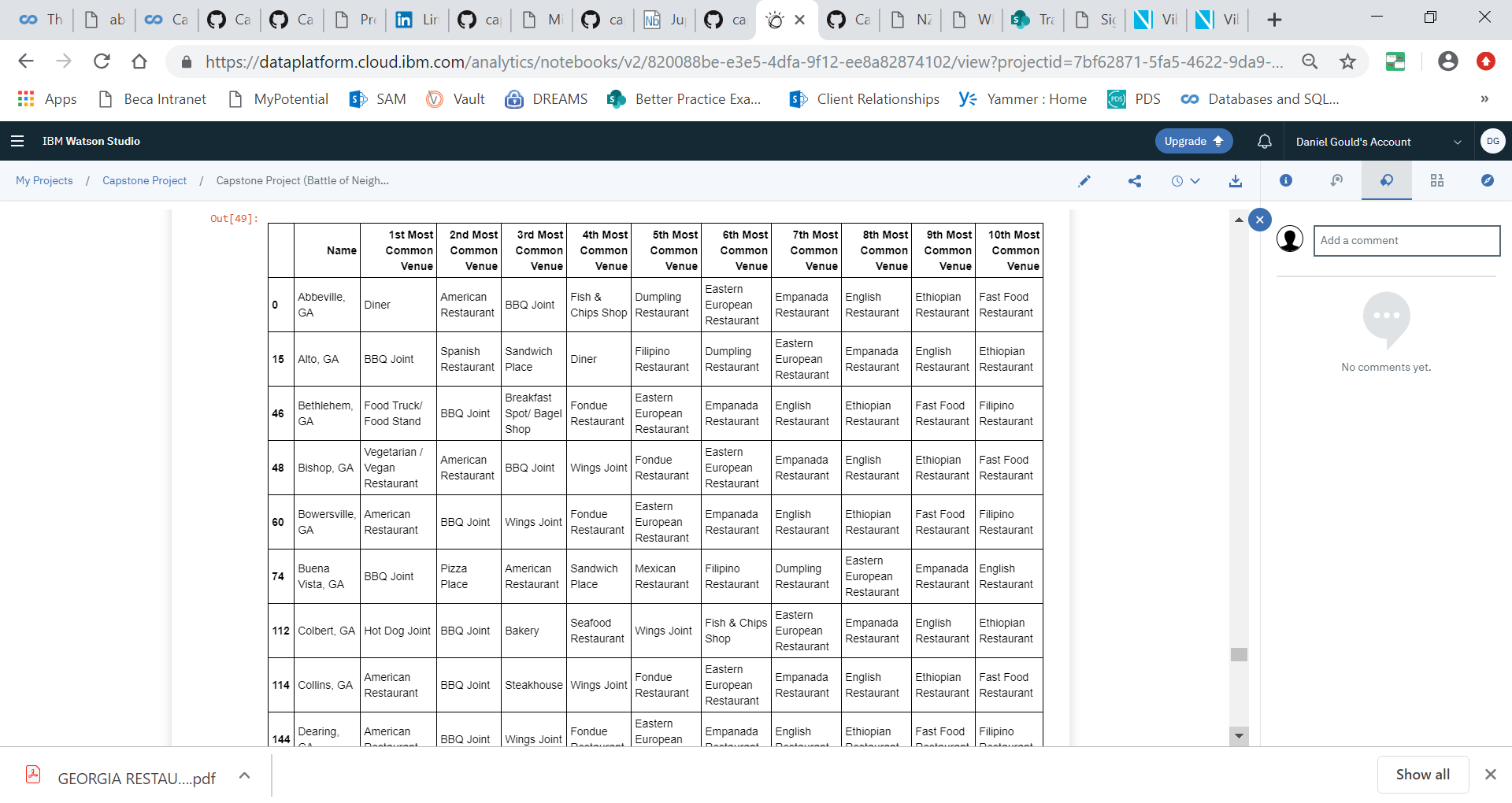
**Cluster 2**



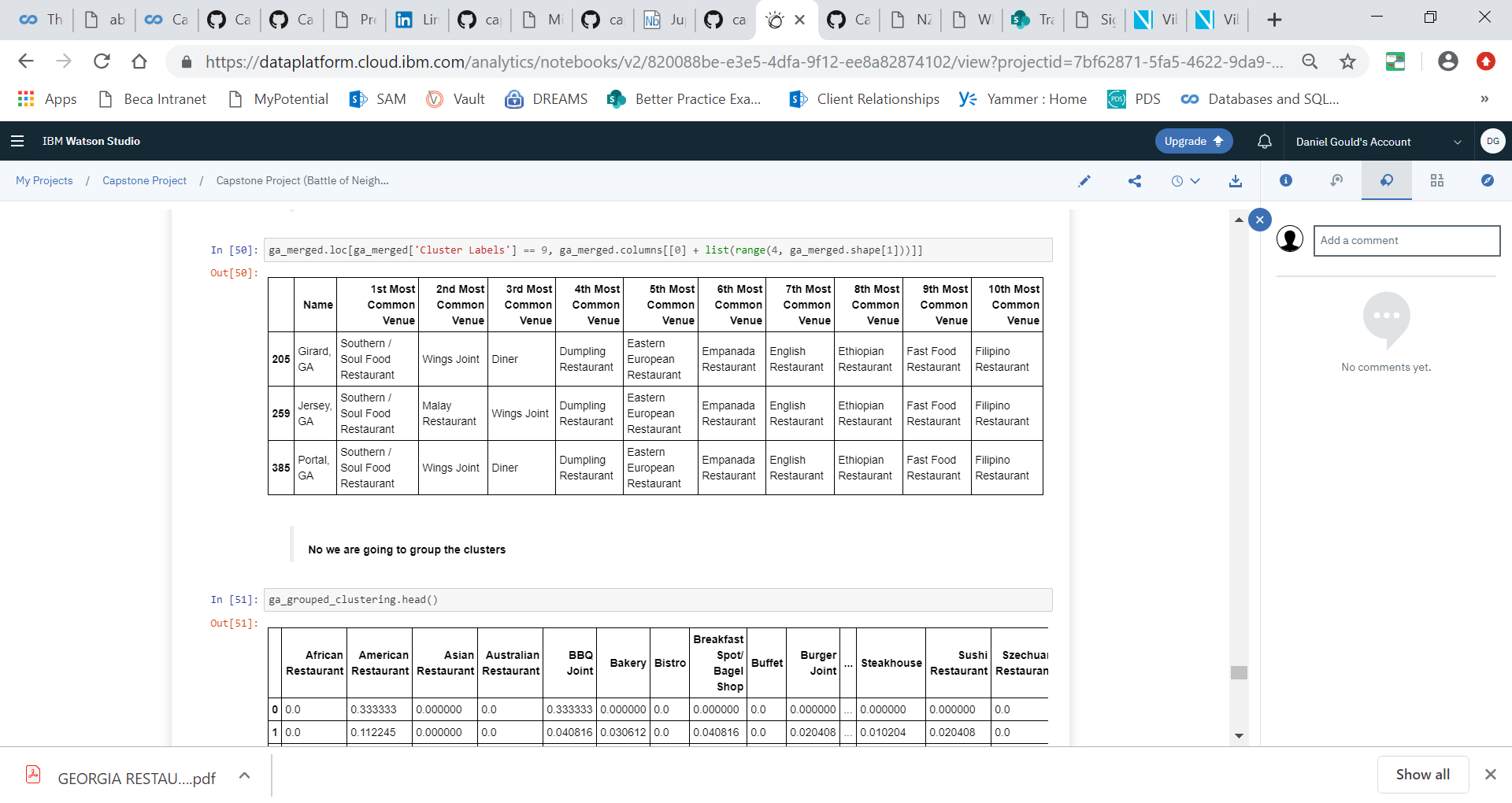
**Cluster 5**



**Cluster 9**



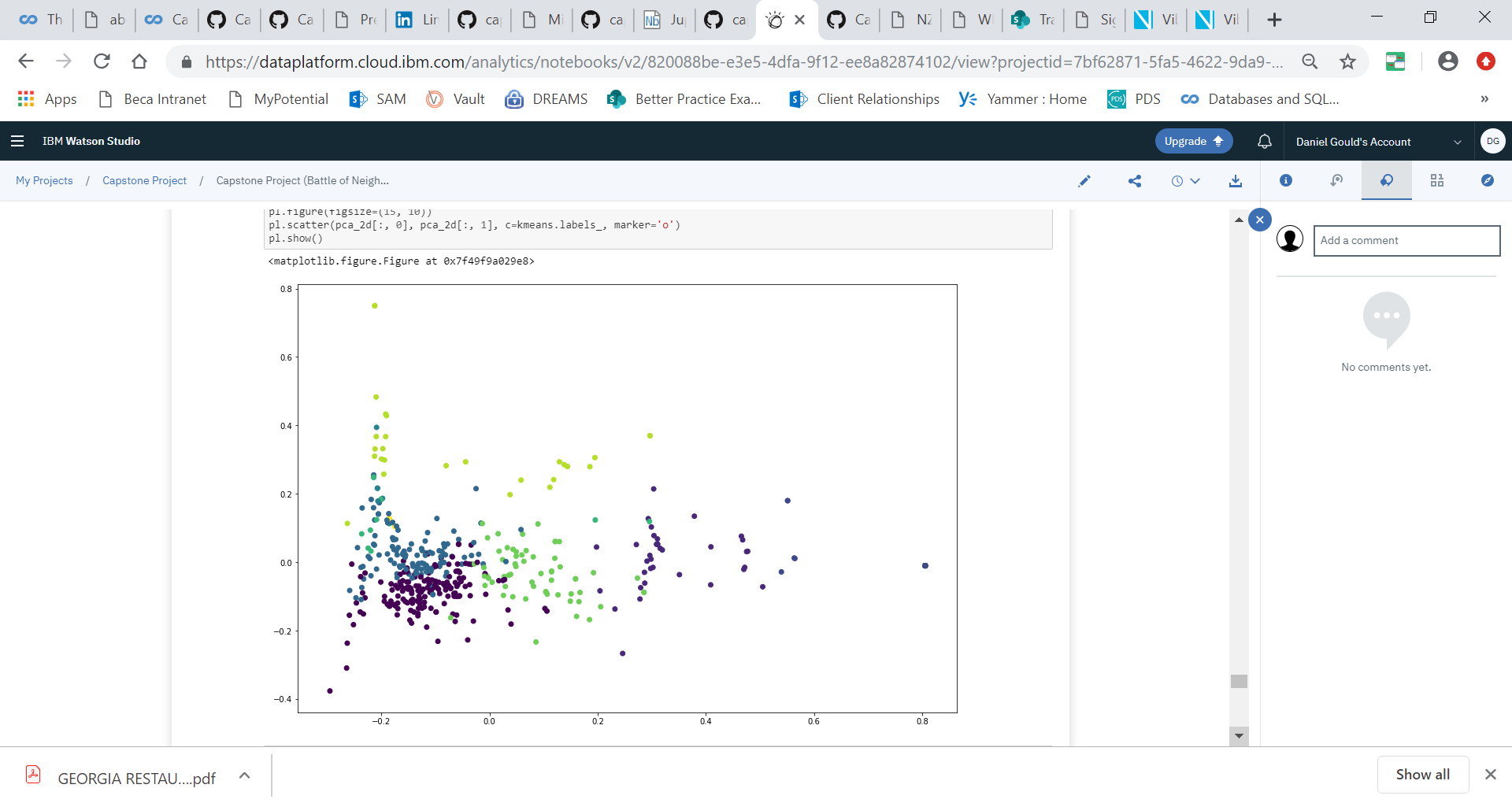
**Cluster 10**



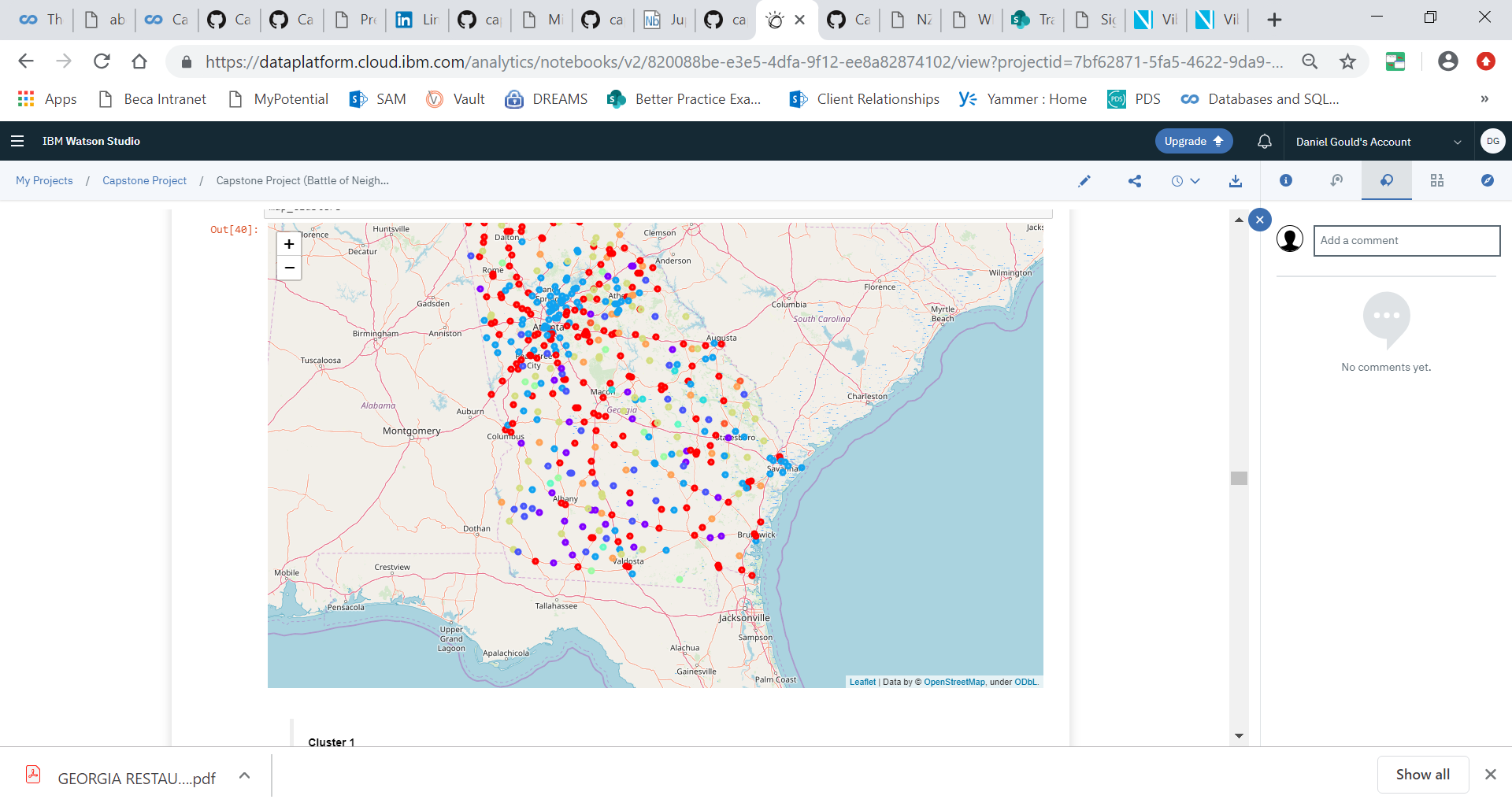
# Results

Once we build the clusters, we visualize the clusters through

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



1. **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.

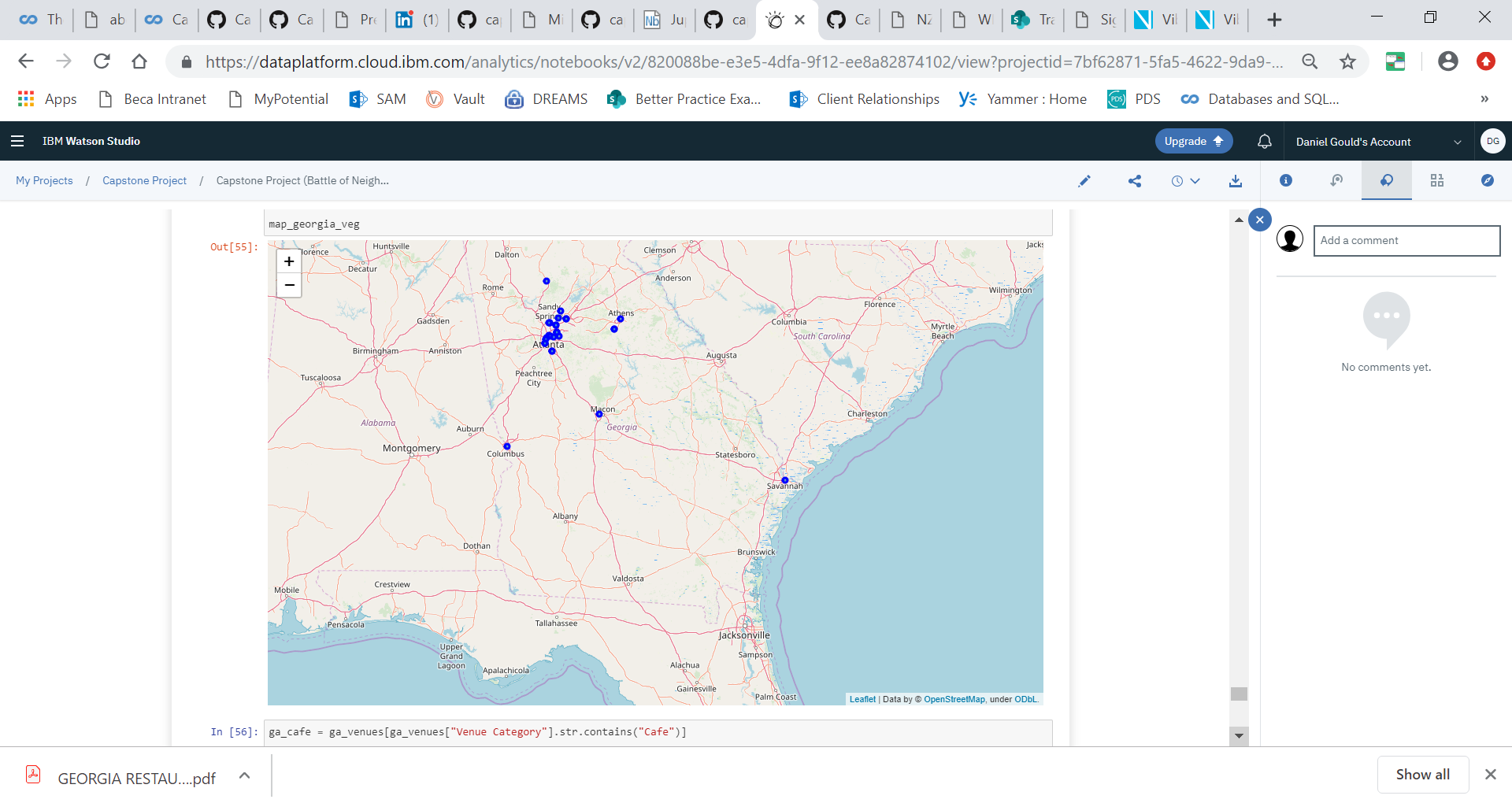
# What categories of restaurants are popular in different parts of the state?

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, Pizza Place, Chinese restaurant |
| Cluster 2 | American, Burger Joint, |
| Cluster 3 | American |
| Cluster 4 | American, Mexican, Pizza Place, Donut, Café, Burger Joint, Sandwich place |
| Cluster 5 | Seafood, Diner |
| Cluster 6 | Bakery |
| Cluster 7 | Burger Joint, Pizza, Café, Sandwich, American, Fried Chicken |
| Cluster 8 | Chinese, Diner, American, Pizza, Southern/Soul Food, Wings, Cafeteria |
| Cluster 9 | BBQ, Diner, Hotdog Joint, American, Vegetarian, Pizza, |
| Cluster 10 | Southern/Soul Food |

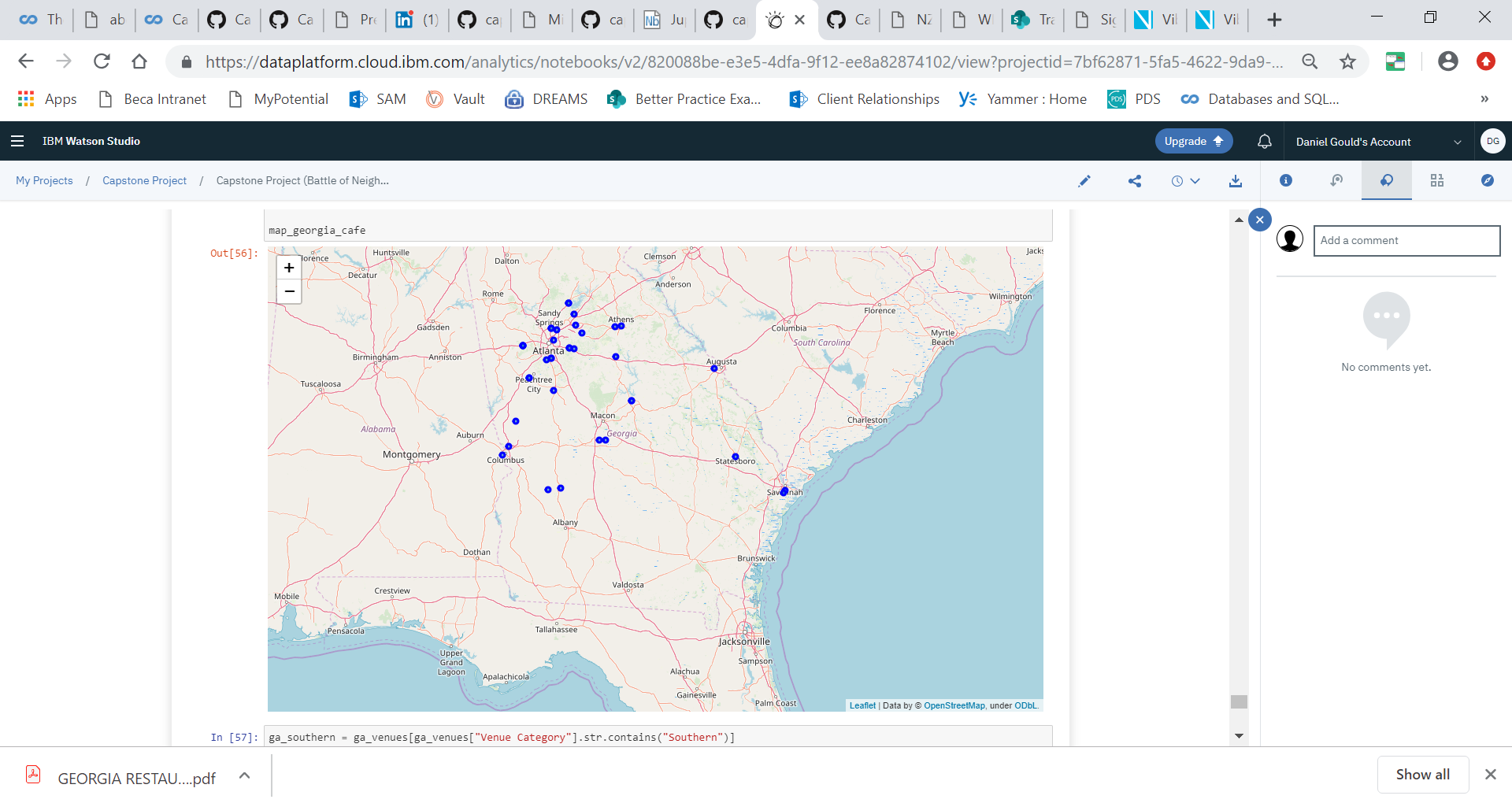
# Are vegan/vegetarian options available in the state? Where are they located?

There is a total of only 34 Vegetarian/Vegan restaurants checked in by FourSquare users in the entire state of Georgia, out of 12,840 venues. This shows that Georgia is not a Vegetarian/Vegan friendly state. Moreover, the few options are centred around a few Cities. Some these cities include Atlanta, Macon, Athens, Savanah and Columbus.



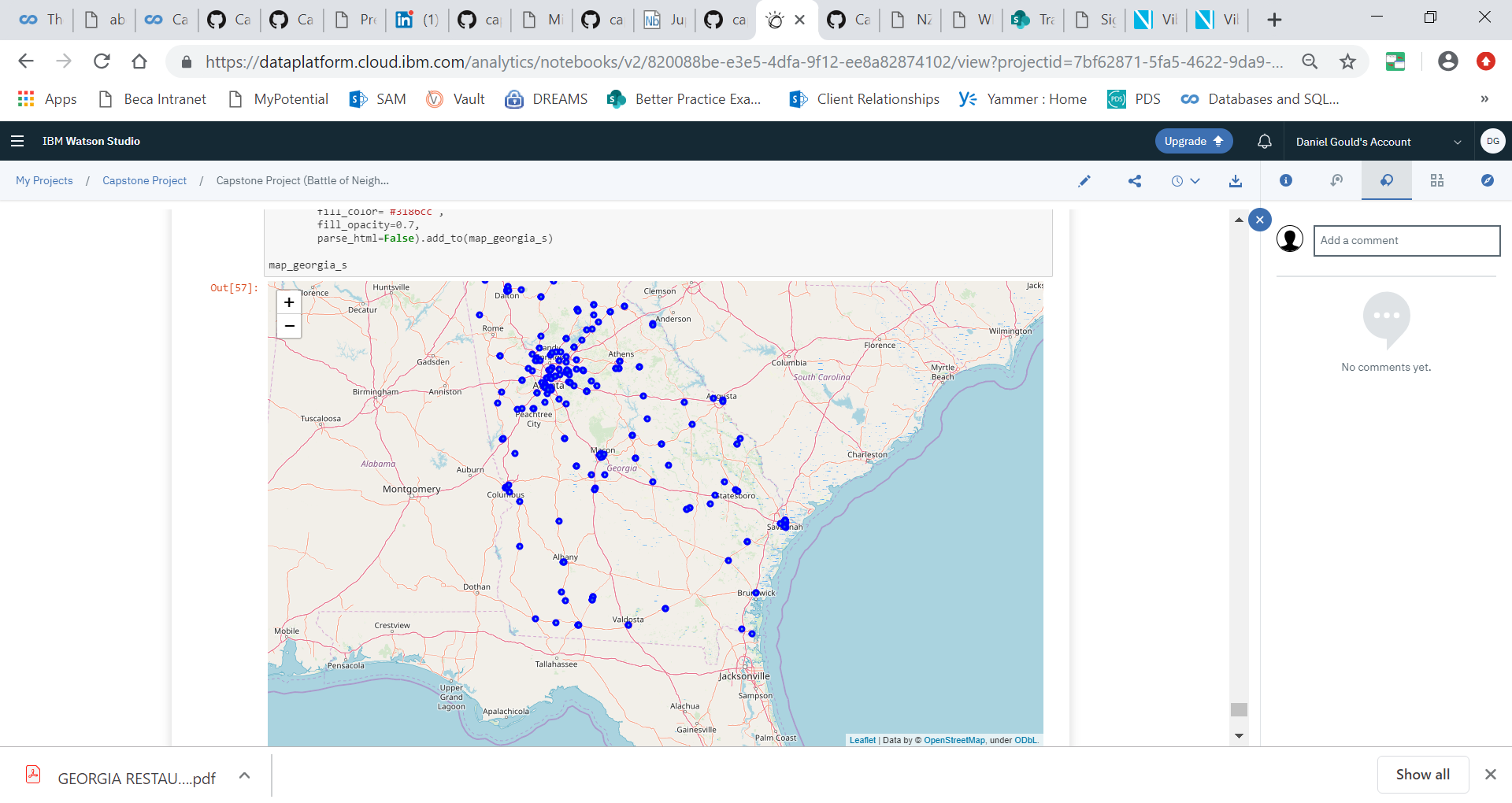
# Where can one find a cafe or quick bite?

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



# 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.



# Which towns do not have dine-in restaurants, potentially not traveller friendly to stop for dine-ins?

There are 71 towns/cities from where FourSquare users not checked in and reported any restaurants. Perhaps, these are cities to avoid dine-in plans.

