Coursera IBM Capstone Project
The Battle of the (Vegetarian) Neighborhoods

Introduction

This project will examine the viability of opening a new restaurant which targets a **vegetarian / vegan** customer base, specifically in the Jacksonville, FL metropolitan area.

According to a recent Gallup poll conducted in 2018, the percentage of Americans who identify as vegetarians has remained close to 5% since 2012[1]. However, sales of vegetarian food items, also referred to as plant-based foods, has grown steadily. The Plant Based Foods Association states on its web site: "Plant-based foods remain a growth engine, up 29 percent over the last two years." [2]

In order to capitalize on these growing trends, this project will investigate the current restaurant landscape in one of Florida's most populous cities: Jacksonville, and help determine where an investor or business owner should open a new vegetarian-friendly restaurant.

This analysis will attempt to research the current number of restaurants in Jacksonville, FL, including a breakdown of restaurant categories, with an emphasis on the number of restaurants classified as "vegetarian / vegan".

---

[1] https://news.gallup.com/poll/238328/snapshot-few-americans-vegetarian-vegan.aspx
[2] https://plantbasedfoods.org/marketplace/retail-sales-data/

Data

Data sources to be used in the project include:
- Jacksonville neighborhood listing on Wikipedia
  https://en.wikipedia.org/wiki/Neighborhoods_of_Jacksonville

- Latitude and longitude information for each neighborhood.
  This data to be provided by ArcGIS for Developers
  https://developers.arcgis.com/features/geocoding/

- Foursquare API - to be used to search neighborhoods, including restaurant category information.
  https://developer.foursquare.com/docs/resources/categories

One Foursquare category which plays an important part in analysis is the sub-category specific to "Vegetarian / Vegan restaurant.
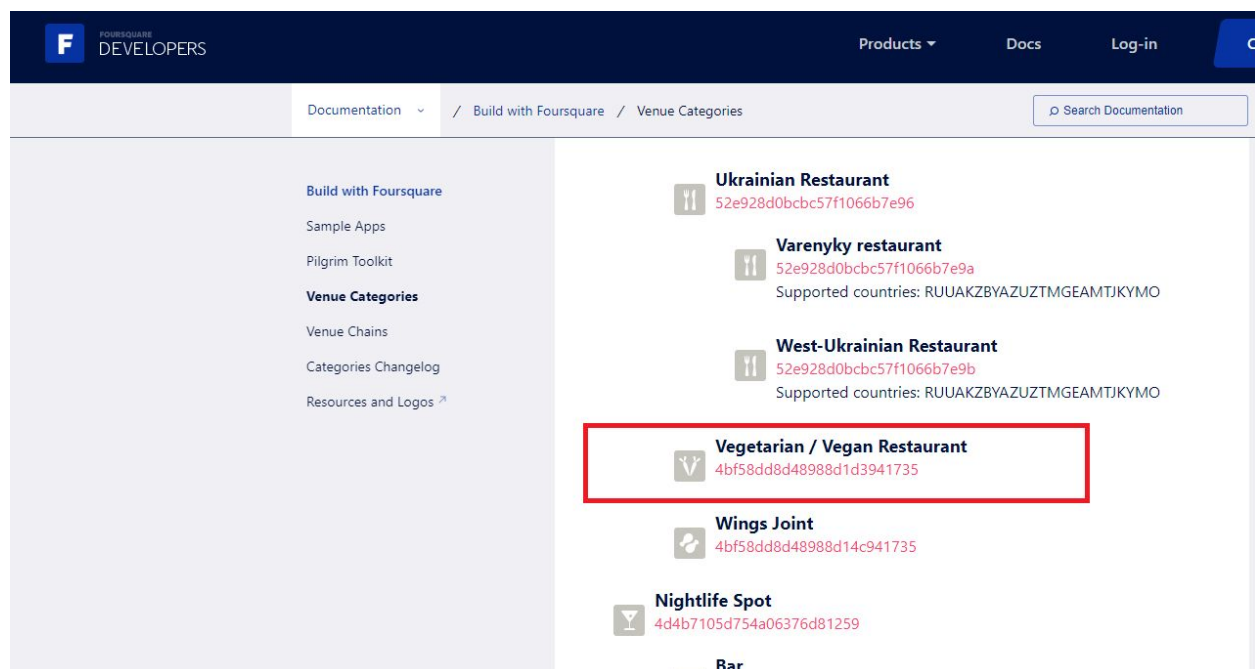
After searching all neighborhoods in the Jacksonville metropolitan area, the results can then be parsed to determine what percentage of existing restaurants are classified as vegetarian in the Foursquare data set.

A map of the area can also help determine which neighborhood has a lower density of vegetarian restaurants, which could then increase the probability of success for a future restaurant..

Methodology

The end result of the analysis should be a simple way to compare a set of all Jacksonville restaurants versus a smaller set of just vegetarian restaurants. The following steps are taken to to accomplish this:

Load all Foursquare API restaurant categories
Foursquare is an online data warehouse which stores location data for cities all around the world. Foursquare data includes many categories of restaurants and other venues, such as Arts and Entertainment, Colleges and Universities, along with other retail and travel establishments.

Each entry in the Foursquare data collection is associated with a category. A top level "Food" category defines a group of more specific categories, such as Asian restaurant or Seafood restaurant.
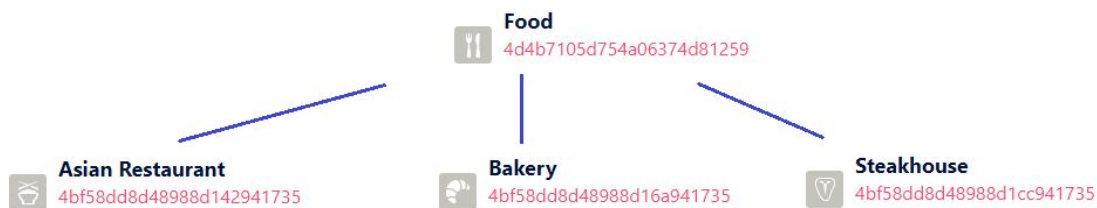
A list of all Foursquare categories is found here:
https://developer.foursquare.com/docs/build-with-foursquare/categories/

After copying the categories, extra formatting was removed to create a simple comma separated values file. This file is then loaded into a Jupyter Notebook for future reference.

Load all Jacksonville neighborhood names
In the United States, Florida is the third most populated state in the country, behind only California and Texas. Jacksonville is the largest city in Florida, in terms of both area (over 870 square miles) and population (over 900,000).[3]

---

[3] https://en.wikipedia.org/wiki/Jacksonville,_Florida

In order to capture the list of approximately 30 Jacksonville neighborhoods, we use Python for web scraping, then the Beautiful Soup library to assist in parsing the captured Wikipedia web page located at https://en.wikipedia.org/wiki/Jacksonville,_Florida

A resulting Python dataframe after parsing (only the first 10 records are shown):

	Neighborhood
0	Downtown Core
1	LaVilla
2	Brooklyn
3	Southbank
4	Riverside
5	Avondale
6	Springfield
7	Eastside
8	Oakland
9	Fairfield
10	San Marco

Find latitude and longitude values for each neighborhood.
Once all neighborhood names have been loaded, the corresponding latitude and longitude coordinates must be populated in order to later execute a Foursquare API query.

A Python geocoder library can take in a simple neighborhood name, city and state name, and return both latitude and longitude values.

geocoder.arcgis is flexible and can accept a specific street address, city, and state, or, in this case, the name of a neighborhood, city and state, and return its corresponding latitude and longitude coordinates.

The end result is a dataframe (again, only the first 10 records are shown):

	Neighborhood	Latitude	Longitude
0	Downtown Core	30.3284	-81.6569
1	LaVilla	30.3308	-81.6696
2	Brooklyn	30.323	-81.6752
3	Southbank	30.1157	-81.5867
4	Riverside	30.3138	-81.6927
5	Avondale	30.3041	-81.7117
6	Springfield	30.3445	-81.6551
7	Eastside	30.34	-81.6403
8	Oakland	30.242	-81.7353
9	Fairfield	30.3498	-81.6287
10	San Marco	30.3117	-81.6605

For each neighborhood, search Foursquare for all available restaurants.
The bulk of the work of the analysis is gathering as much restaurant data as possible using the Foursquare API.

The Foursquare API offers two main options to retrieve venue information, a search command and an explore command:

https://api.foursquare.com/v2/venues/**explore**

https://api.foursquare.com/v2/venues/**search**

In practice, /venues/explore seemed to be missing some restaurants expected to be in the search results. However, the explore option does allow for pagination with its result set using an **offset** parameter.

The /venues/search command returned a more thorough set of venues, but Foursquare imposes a limit of 50 results to any API call using the search method.

To attempt to find as many matching restaurants as possible, our analysis performs a large number of searches using the **/venues/explore** function. Every neighborhood in the data set queries the Foursquare API with each of the 76 sub-level categories.

Since the Foursquare queries use a search radius of 8000 meters (approximately 5 miles), there will be some overlap in the search process. Some neighborhoods are small in size, and so the

searches may find the same restaurant multiple times. After all searches are completed, the duplicates are removed.

Each of the 28 Jacksonville neighborhoods performs a search using each of the 76 Foursquare sub-level categories, totaling 2,128 search commands.

After eliminating duplicates, the total number of all restaurants in all Jacksonville combined neighborhoods was approximately 2700.

Perform a separate search on Foursquare for only vegetarian restaurants.
The Foursquare lower level categories contain a specific category for restaurants focusing a vegetarian or vegan clientele.

For each of the 28 neighborhoods, a single Foursquare API call is made.

Based on some exploratory analysis, there are fewer than fifty vegetarian results in each neighborhood, so only one Foursquare query is necessary.

After eliminating duplicates, the total number of restaurants categorized in Foursquare as Vegetarian / Vegan is 18.

K means Cluster analysis.
K means cluster analysis is performed to create logical groupings of all restaurants in Jacksonville.

Before clustering, the optimal number of clusters is found using the elbow method.
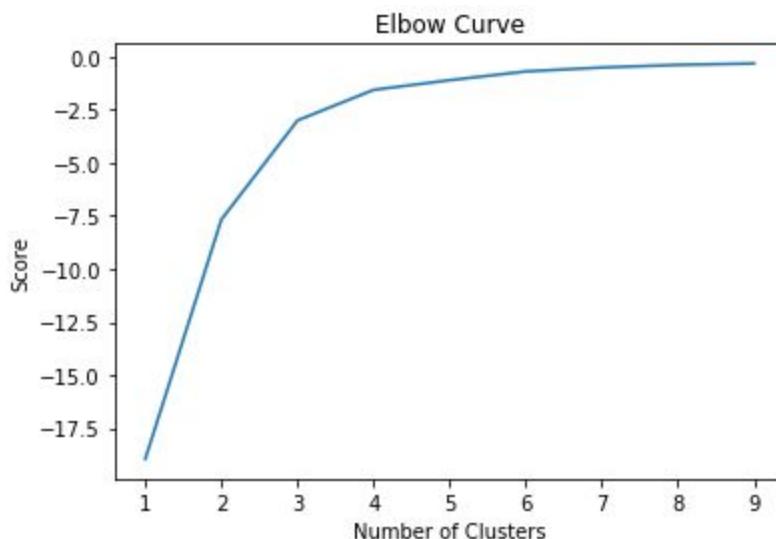
Based on the curve, the number of clusters to be created should be 3.

Using k means clustering analysis, each restaurant in the data set is put into one of three clusters.

Analysis

Map all results.

The Folium Python library provides a simple way to display a map, and programmatically add markers to identify points of interest.

Map #1 displays all 2700 restaurants in the Jacksonville area, broken up into three separate clusters.
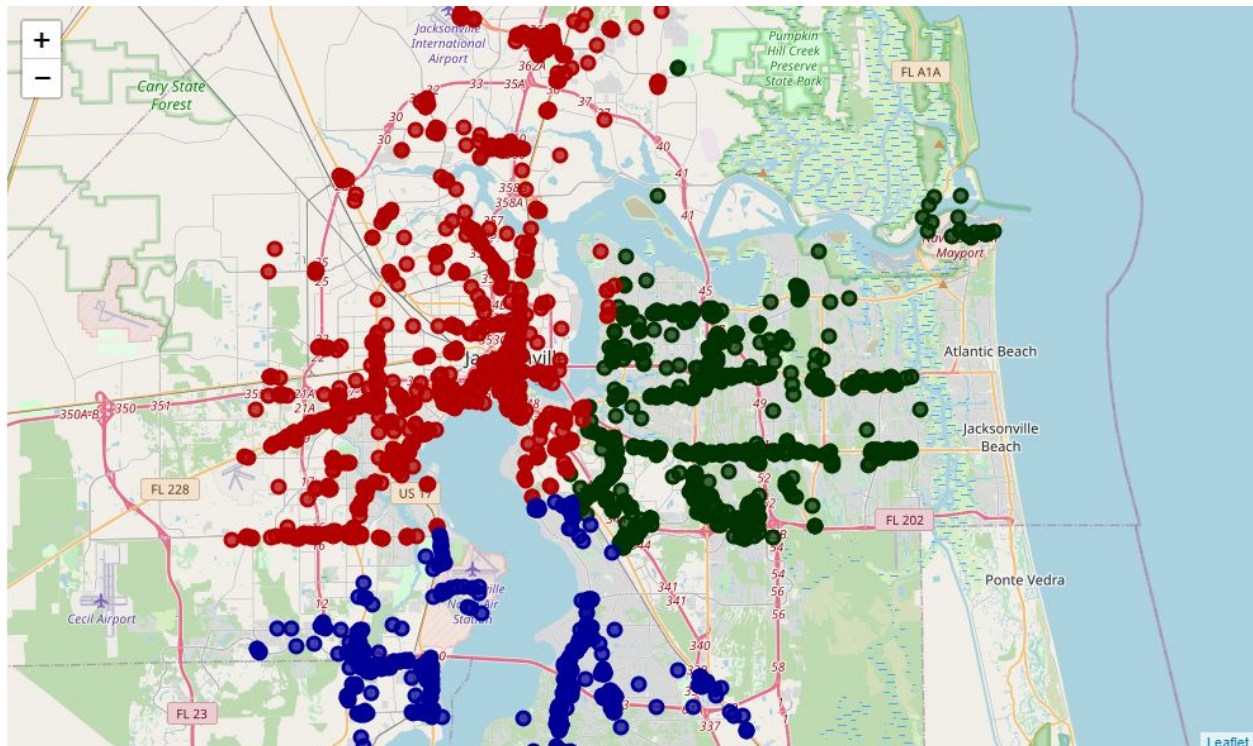
Maps #2 adds the 19 restaurants identified by the vegetarian / vegan Foursquare category, seen below as yellow markers:
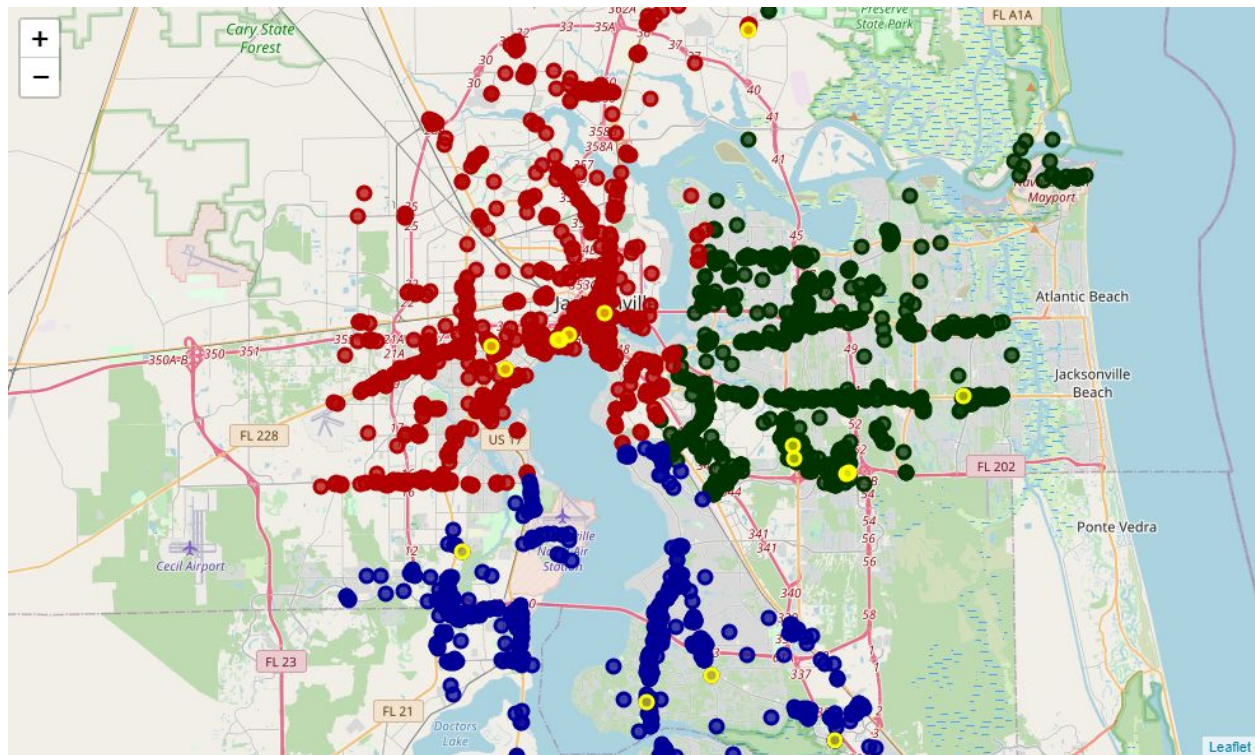
Results

The k means clustering creates the following:

Cluster	Number of restaurants	Number of vegetarian restaurants	Percentage of restaurants marked vegetarian
0 - green	931	5	0.54%
1 - red	1125	7	0.62%
2 - blue	679	7	1.03%
Totals	2,735	19	0.69%

As the clusters indicate, the cluster labeled 0 in the dataframe (the green cluster on the eastern side of Jacksonville) has the lowest number of vegetarian restaurants, as well as having the lowest market penetration. This could present a possible business opportunity for an investor looking to open a new vegetarian restaurant.

Conclusion

Foursquare data is created by its users, so it might not include each and every restaurant or venue in any given city. However for this analysis, Foursquare can present a clear pattern of restaurant types in the neighborhoods searched.

If a restaurant investor was looking to invest in a vegetarian or vegan leaning location, Jacksonville seems to offer many opportunities. As discovered in the current Foursquare data for the Jacksonville metropolitan area, vegetarian restaurants are under-represented as compared to the overall population.

A 2018 Gallup poll claims that 5% of the population of the United States identifies as vegetarian. However the current Foursquare data claims that less than 1% of the existing restaurants in Jacksonville are vegetarian.

More specifically, the eastern section of Jacksonville would be an area to investigate further, as this section of neighborhoods had the fewest ratio of vegetarian restaurants.