
APPLIED DATA SCIENCE CAPSTONE

AN INVESTIGATION TO THE IDEAL LOCATION OF A NEW RESTAURANT IN
BRADENTON, FL

April 15, 2019

Michael A Greene, PhD.
IBM Professional Data Scientist Certification
Coursera

Contents

0.1	Introduction/Business Problem	2
0.2	Data Methodology	2
0.3	Results	8
0.4	Discussion	9
0.5	Conclusion	10

0.1 INTRODUCTION/BUSINESS PROBLEM

In a town with a small number of people, it is absolutely crucial to conduct market research to determine the demand for a particular product. Especially if you are trying to sell that product! In a large city, it may be possible to sustain a business on the fact that with such a large and diverse group of people, some of them are bound to stop in and order. This may not be true in smaller towns where profit margins are razor thin.

The purpose of this project is to explore my home town of Bradenton, FL and determine the best place, if any, to open a new restaurant.

This study will take in to account the proximity of other food venues, the categories of venues, and the reviews of existing food venues.

0.2 DATA METHODOLOGY

I will extract the venue information from Foursquare using a developer account and several API calls. Getting the name and location of a venue within a certain radius is completely free up to 500 calls per day with the free developer sandbox account. This is accomplished by following these steps:

1. Connecting to the Foursquare API with a special value of private keys.
2. Creating the query which can be understood by the Foursquare API.
3. Extracting the returned data from the HTML code.

Once the data is in a python-readable format, I will use Folium to visualize the city and plot the various venues. K-means clustering, a machine learning algorithm, will then be used to group similar venues and predict where a new restaurant may be situated in a prime location.

The result of the query returns the following structure:

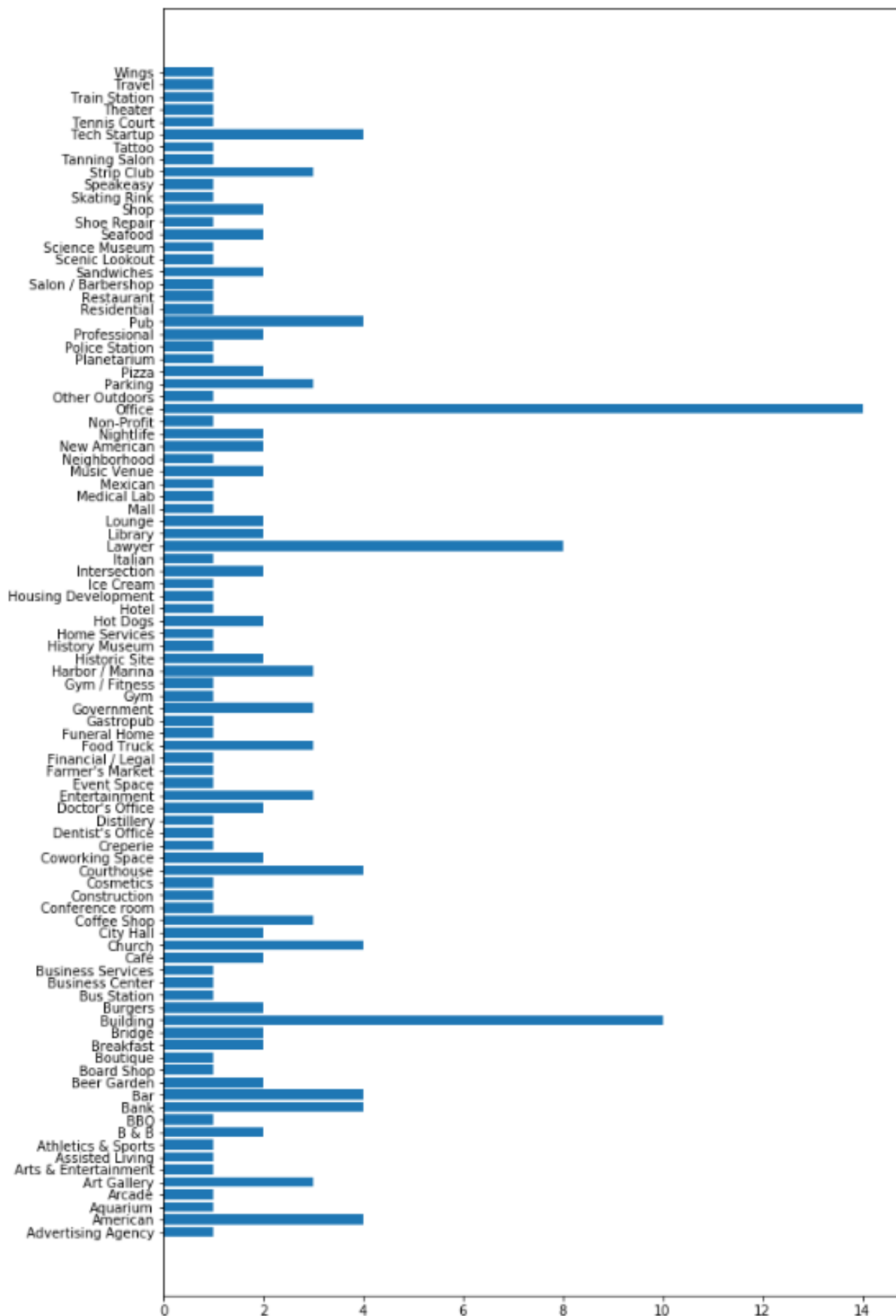
```
{'id': '4b82ea74f964a520fbec30e3', 'name': 'Central Library', 'contact': {},
'location': {'address': '1301 1st Ave W', 'crossStreet': 'at 14 St W',
'lat': 27.49876513403457, 'lng': -82.57469677521266,
'labeledLatLngs': [{'label': 'display', 'lat': 27.49876513403457,
'lng': -82.57469677521266}], 'distance': 21, 'postalCode': '34205',
'cc': 'US', 'city': 'Bradenton', 'state': 'FL',
'country': 'United States',
```

```
'formattedAddress': ['1301 1st Ave W (at 14 St W)',
'Bradenton, FL 34205', 'United States']},
'categories': [{ 'id': '4bf58dd8d48988d12f941735',
'name': 'Library', 'pluralName': 'Libraries',
'shortName': 'Library', 'icon': {'prefix':
'https://ss3.4sqi.net/img/categories_v2/building/library_',
'suffix': '.png'}, 'primary': True}], 'verified': False,
'stats': {'tipCount': 0, 'usersCount': 0, 'checkinsCount': 0,
'visitsCount': 0}, 'beenHere': {'count': 0,
'lastCheckinExpiredAt': 0, 'marked': False, 'unconfirmedCount': 0},
'referralId': 'v-1555100110', 'venueChains': [], 'hasPerk': False}
```

From this, we can see the structure necessary to create the data frame with 190 rows:

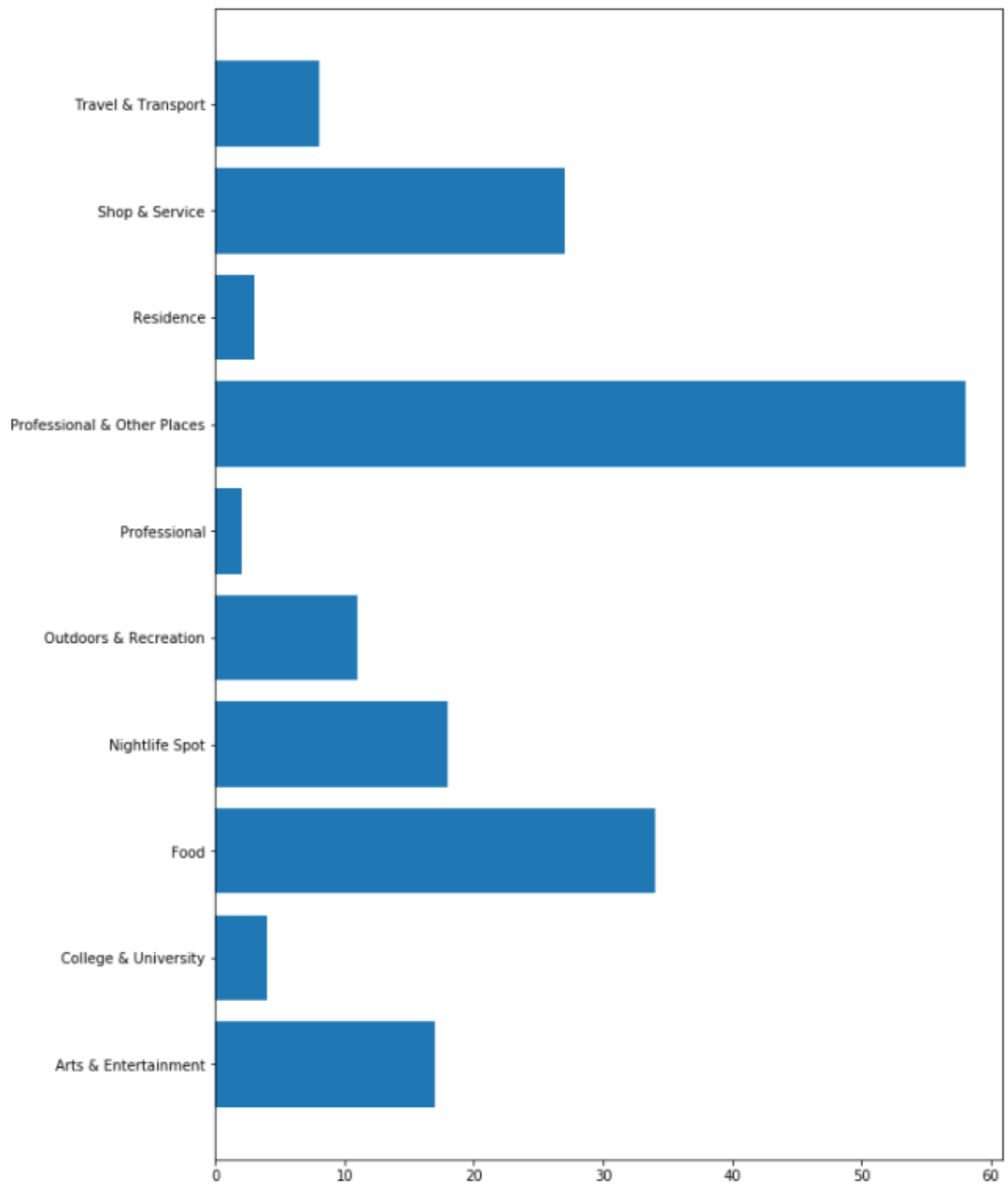
Index	Name	Latitude	Longitude	Category
0	Central Library	27.498765	-82.574697	Library
1	Bradenton Riverwalk	27.499282	-82.567566	Scenic Lookout
2	Downtown Bradenton Farmers Market	27.496181	-82.573458	Farmer's Market
3	The Freckled Fin and Public House	27.498928	-82.574819	Music Venue
4	PIER 22 Restaurant Patio Catering	27.500480	-82.573531	American

There are many different categories of venues:

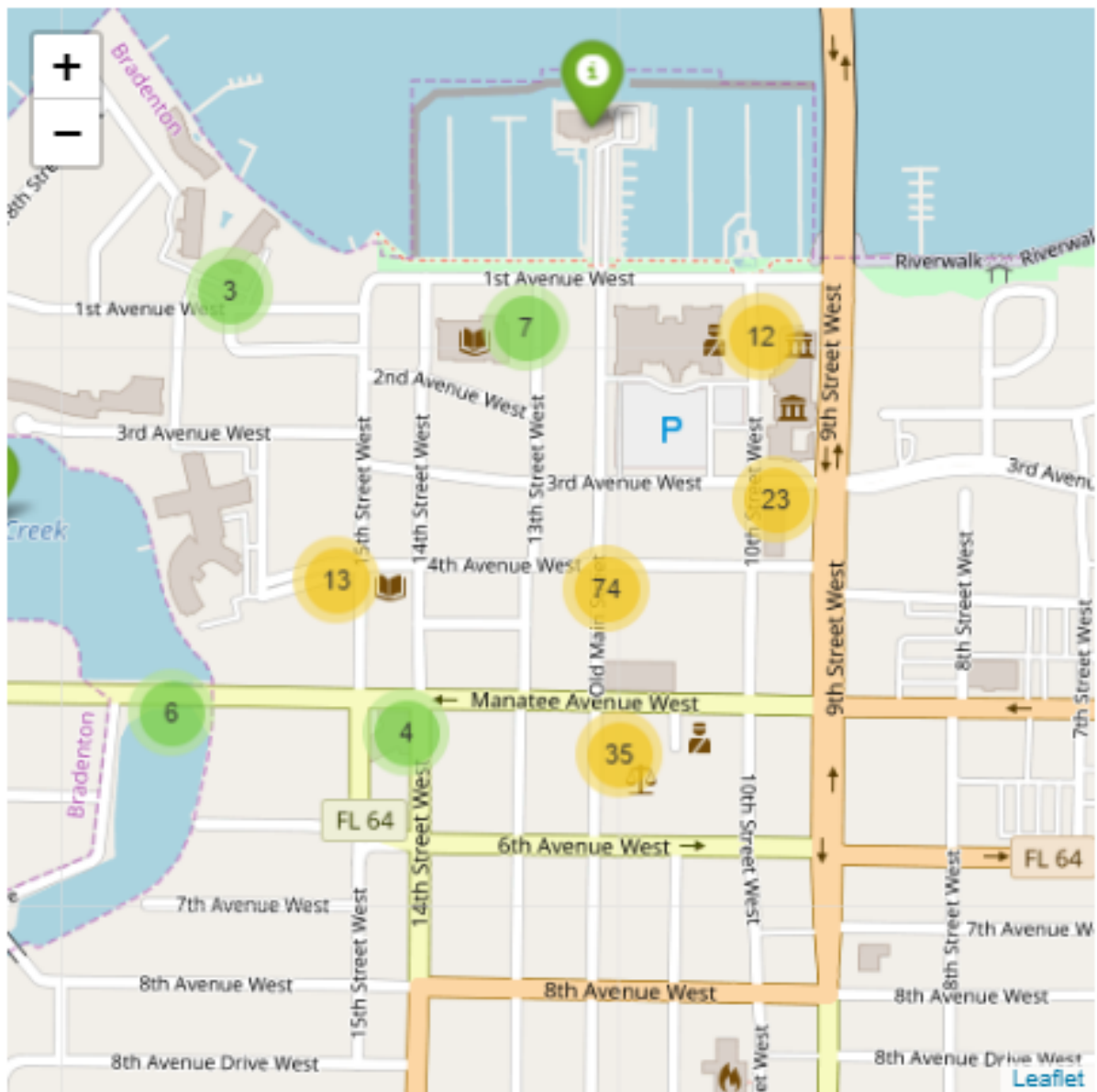


So I will scrape Foursquare's website to get a more general category

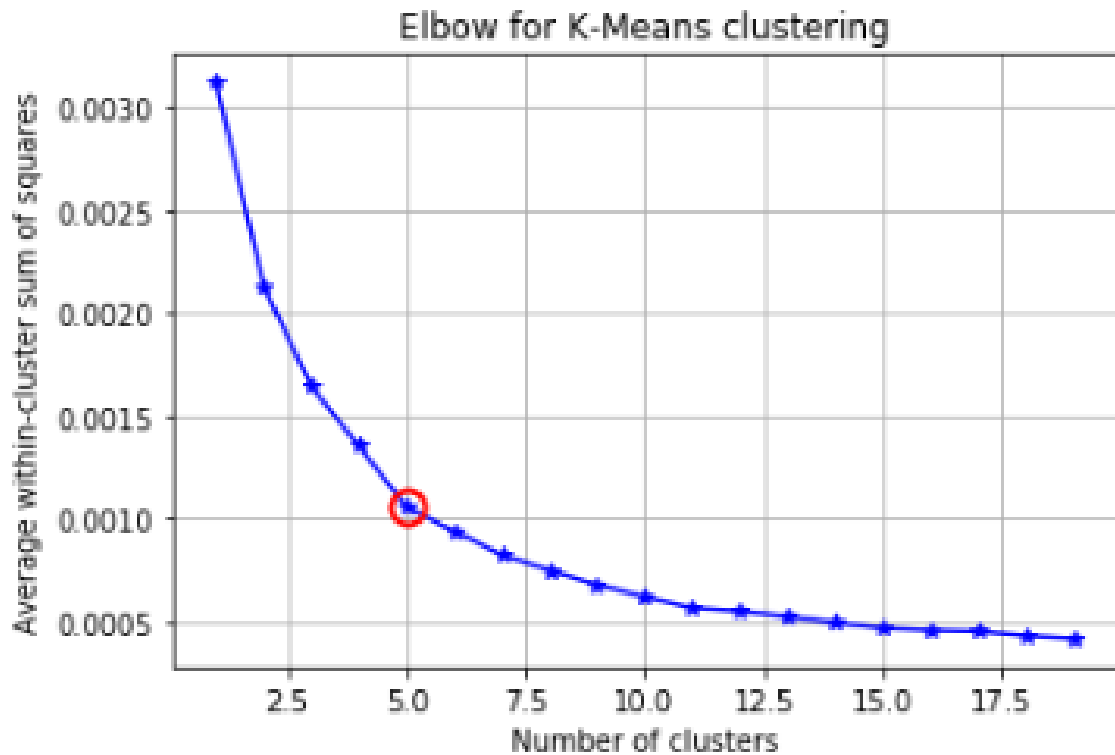
Index	Name	Latitude	Longitude	Category	General Label
0	Central Library	27.498765	-82.574697	Library	College & University
1	Bradenton Riverwalk	27.499282	-82.567566	Scenic Lookout	Outdoors & Recreation
2	Downtown Bradenton Farmers Market	27.496181	-82.573458	Farmer's Market	Shop & Service
3	The Freckled Fin and Public House	27.498928	-82.574819	Music Venue	Arts & Entertainment
4	PIER 22 Restaurant Patio Catering	27.500480	-82.573531	American	Food



Let's now visualize some of the venue and their locations using the Folium map.



Now we can begin to use k-means clustering to determine a cluster which contains the fewest food venues. I will fit the model on the data above, and identify the venues within each cluster. First, let's identify the optimal number of clusters using an elbow plot.



```

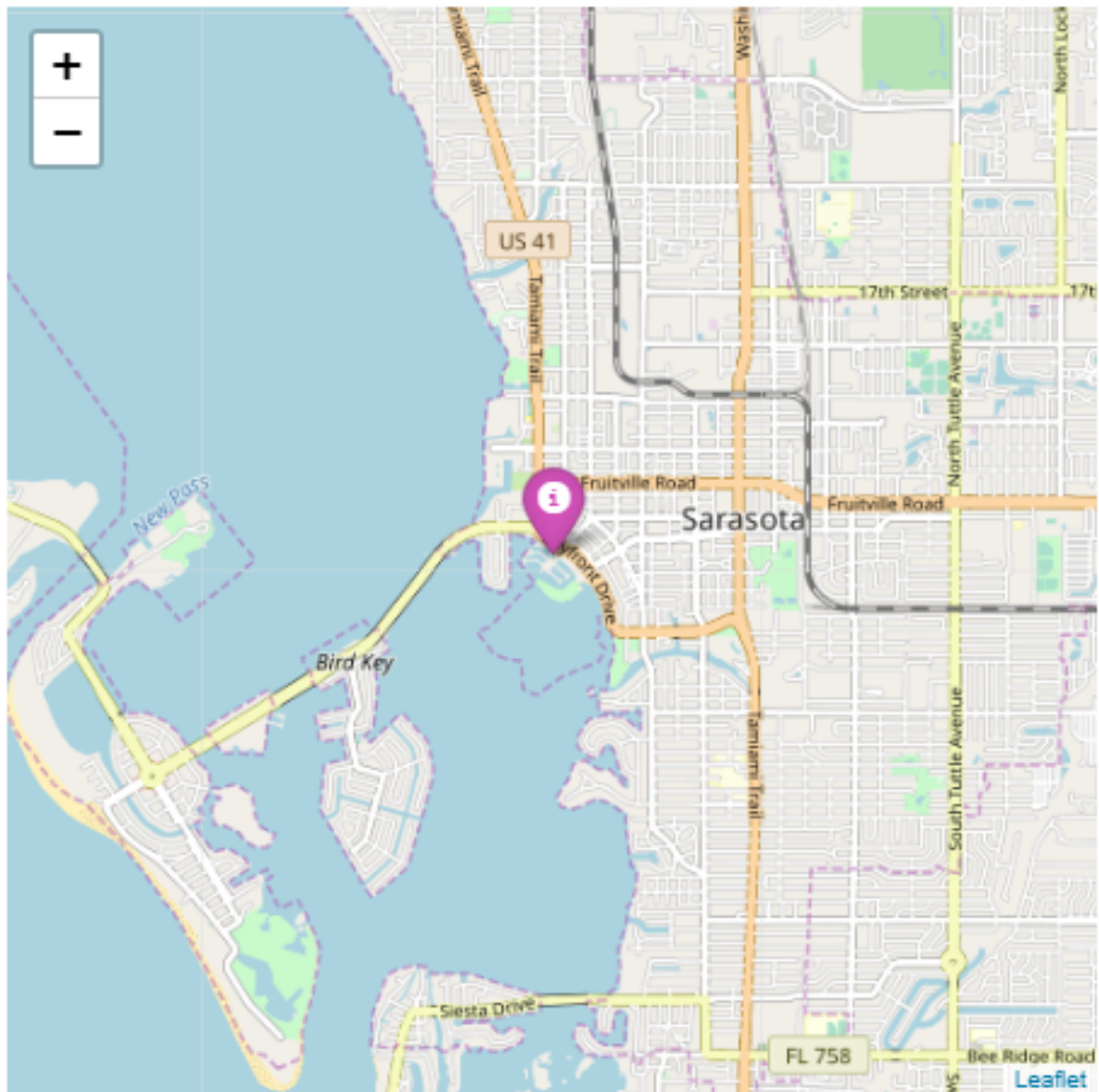
0 --- {'Residence', 'Shop & Service', 'Food', 'Travel & Transport',
'Arts & Entertainment', 'Professional & Other Places', 'Outdoors & Recreation'}
1 --- {'Outdoors & Recreation'}
2 --- {'Food'}
3 --- {'Shop & Service', 'Professional', 'Food', 'Nightlife Spot',
'Travel & Transport', 'Arts & Entertainment', 'Outdoors & Recreation',
'College & University', 'Professional & Other Places'}

```

At first, five clusters resulted in clusters 0 and 4 being identical, so I reduced it to four clusters. Cluster number 1 does not have a "food" category. Incidentally, clusters 1 and 2 have one entry each. This will be discussed later.

0.3 RESULTS

Now that our cluster model is fitted to the data, we can find the centroid of cluster 1. This will be latitude and longitude coordinates of the middle of the cluster.



A prime spot, right on the river!

0.4 DISCUSSION

The biggest concern in this study is that there is only one venue in clusters 1 and 2. This result is likely an issue with the Foursquare return query. I think that Foursquare is only returning venues within a certain radius, but at the same time limiting the results somehow, even after adjusting the search radius. In a future update, I will conduct multiple queries to Foursquare and test my hypothesis by joining the results together for

a larger number of venues.

0.5 CONCLUSION

Thank you for taking the time to read and reflect with me the process of conducting a short data science project for the satisfaction of IBM's Professional Data Scientist certification process through Coursera. In summary, I:

1. Outlined the procedure and background - predicting where to set up a new restaurant in my hometown
2. Defined the required data and how to collect it - Foursquare using a free developer account
3. Described the data once I got it - pandas with matplotlib and Folium
4. Used K-means clustering to group venues by location and category - predicting a restaurant where there aren't any 'food' categories
5. Write up a summary and publish it to a blog or social media post

Please feel free to comment, critique, and leave your reviews however you wish.

Thank you again,

Michael A Greene, PhD