

**Daniela
Mosco**

Florida City Data Analysis

Optimal location for new Coffee Shop Business

The Business Problem

- The location of a business in the service industry has a strong impact on whether that business prospers or fails.
- Look for a location that gives business owners the best possible chance for success.
- The restaurant and grocery industry is:
 - oversaturated
 - highly competitive
- This results in new business owners hoping to open a shop in an already saturated market.

Proposed solution

- Analyze the postal code areas of Florida, USA.
- Foursquare will be used to cluster the areas into groups based on the number of specific amenities in the local areas.
- Data about the given areas will be collected using the Foursquare API and geographical information will be retrieved from the web using the Python library BeautifulSoup as well as the mapbox API.

Methodology

The following libraries were used:

- Matplotlib
- Pandas
- Folium
- BeautifulSoup
- Requests

Geographic data was scraped from Wikipedia using BeautifulSoup to get all the Florida city area codes along with a list of place names within each area code.

The geographic data was made into a dataframe using pandas in order to allow easy manipulation and better comprehension of it.

| Area codes | | Place names |
|------------|-----|---|
| 0 | 239 | [Lee County, Collier County, Monroe County, FL... |
| 1 | 305 | [Miami-Dade County, Florida Keys] |
| 2 | 407 | [Space Coast, Cape Canaveral, Melbourne, Titus... |
| 3 | 689 | [Gainesville, Ocala, Inverness, Spring Hill, D... |
| 4 | 386 | [Pensacola, Tallahassee, Panama City] |
| 5 | 727 | [Pinellas County, Clearwater, St. Petersburg, ... |
| 6 | 754 | [Broward County, Fort Lauderdale, Hollywood, C... |
| 7 | 954 | [Vero Beach, Port Saint Lucie, Fort Pierce, Se... |
| 8 | 772 | [Hillsborough County, Tampa, Plant City, Pasco... |
| 9 | 813 | [Pensacola, Tallahassee, Panama City] |
| 10 | 850 | [Lakeland, Arcadia, Avon Park, Clewiston, Bart... |
| 11 | 863 | [Jacksonville, St. Augustine, Starke, Green Co... |
| 12 | 904 | [Tampa Bay, Manatee County, Sarasota County, C... |
| 13 | 941 | [Broward County, Fort Lauderdale, Hollywood] |

The mapbox API was used to generate accurate latitude and longitude variables for each area code, using geocoding.

Then these coordinates data were built into a dataframe, using Pandas.

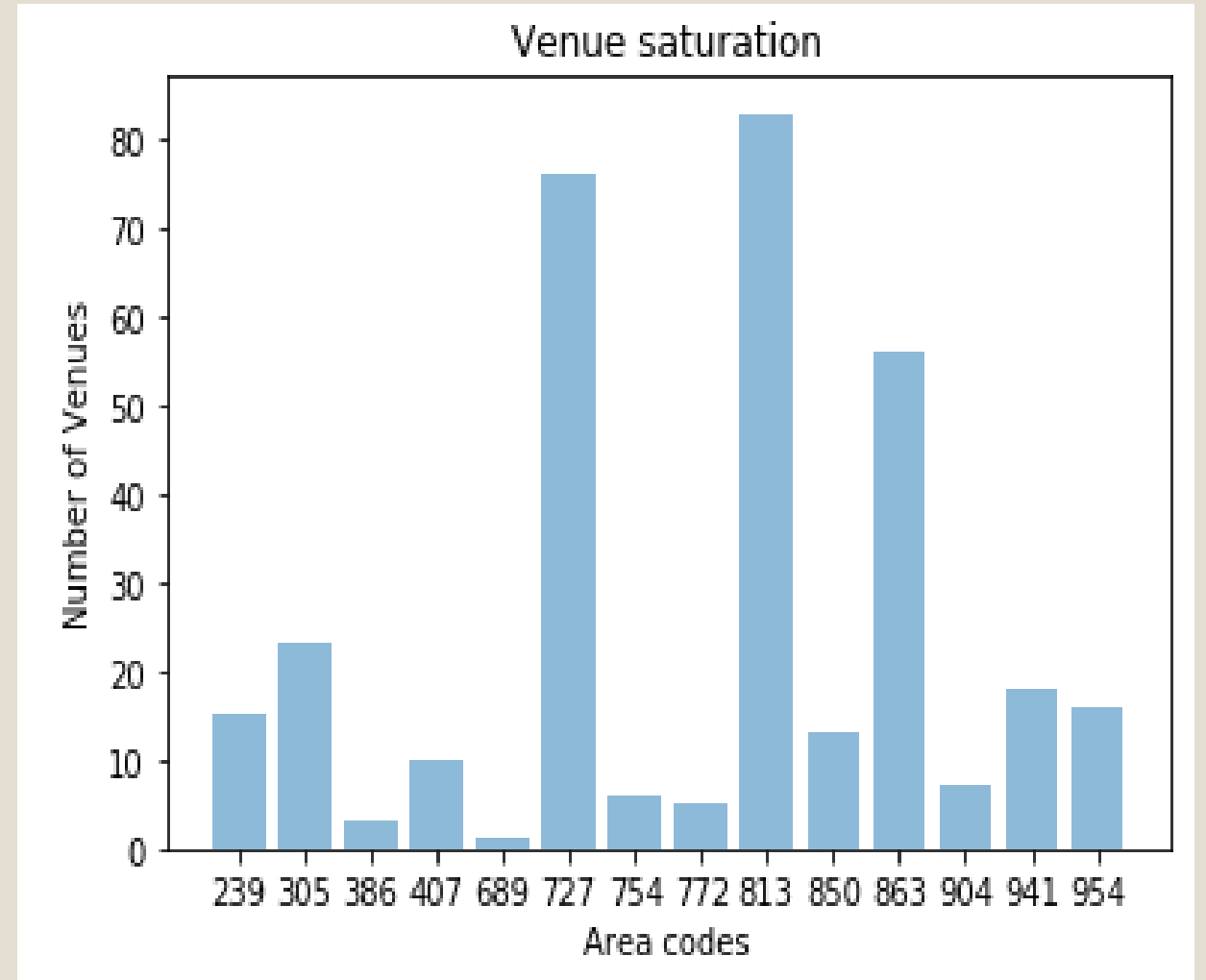
The resulting dataframe was merged with the existing dataframe in order to provide a simply understandable representation of each location and its corresponding coordinates.

| | Area codes | Latitude | Longitude |
|----|------------|----------|-----------|
| 0 | 239 | 28.94703 | -81.29950 |
| 1 | 305 | 28.58214 | -81.75084 |
| 2 | 407 | 27.71523 | -82.43498 |
| 3 | 689 | 29.74793 | -84.85735 |
| 4 | 386 | 30.70996 | -86.76110 |
| 5 | 727 | 26.13284 | -80.13786 |
| 6 | 754 | 27.01878 | -82.17136 |
| 7 | 954 | 29.95433 | -82.10759 |
| 8 | 772 | 30.18817 | -82.61500 |
| 9 | 813 | 26.13401 | -80.13790 |
| 10 | 850 | 26.13323 | -80.20200 |
| 11 | 863 | 28.93470 | -81.93329 |
| 12 | 904 | 29.60769 | -82.81801 |
| 13 | 941 | 26.13502 | -80.20311 |

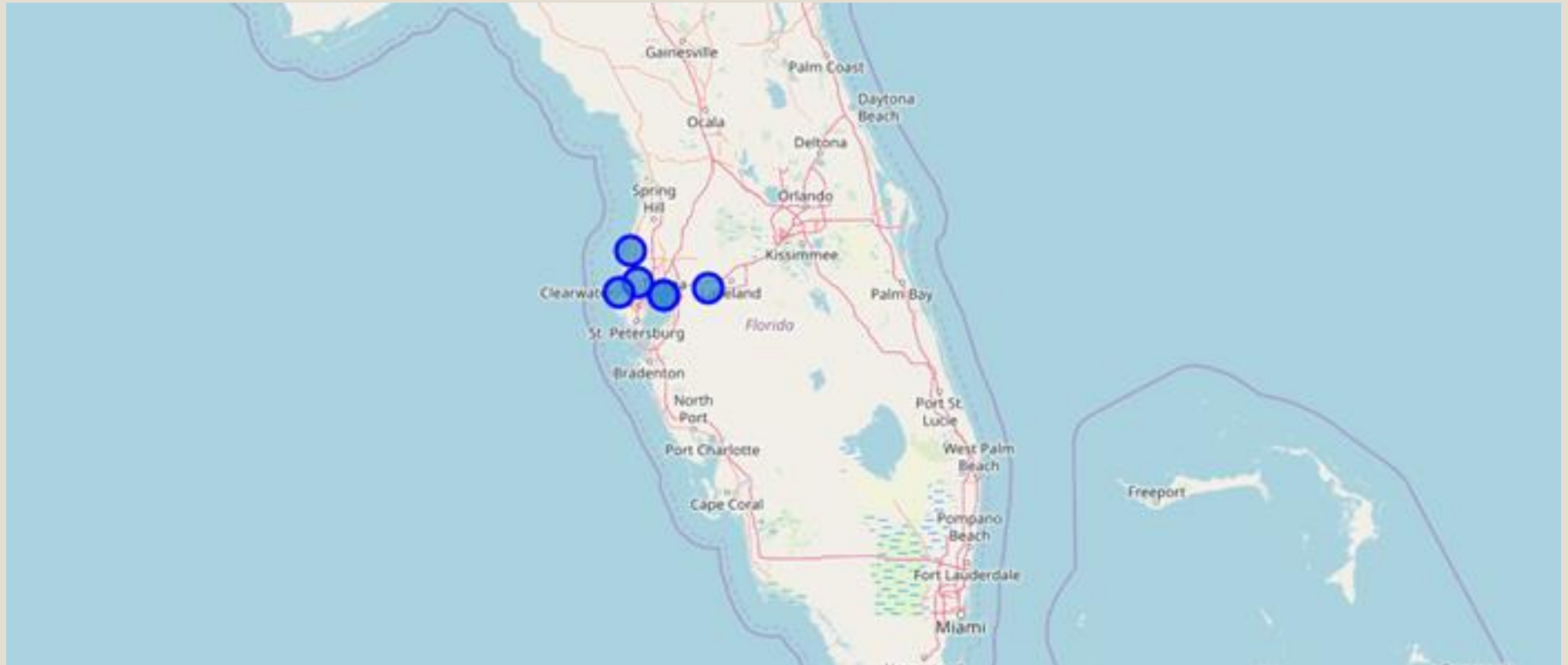
Using the location data gathered, the Foursquare API was queried, and it returned a JSON formatted list of venues in each location.

The data gathered from Foursquare was visualized as a bar chart, using matplotlib.

This gave us an overview of each area code and allowed us to decide on refining our search to areas localized in the Florida 863 area.



Folium was used to visualize the Florida 863 area on the map



The results of k-means clustering showed that:

- The first cluster (cluster 0) had a low to moderate number of coffee shops.
- The second cluster (cluster 1) had the lowest percentage of coffee shops in the clustered areas.
- The third cluster (cluster 2) had the highest concentration of coffee shops.

Regarding these measurements it would be advantageous to open a new coffee shop in the Pinellas County that belong to cluster 2.

| | Area name | Café | Cluster Labels |
|---|---------------------|--------|----------------|
| 0 | Hillsborough County | 0.0200 | 0 |
| 1 | Oldsmar | 0.0000 | 1 |
| 2 | Pasco County | 0.0000 | 1 |
| 3 | Pinellas County | 0.0625 | 2 |
| 4 | Plant City | 0.0000 | 1 |
| 5 | Tampa | 0.0200 | 0 |

Conclusion and Further research

While this project contemplated potential areas for opening a new coffee shop based on preexisting ones, there are many other data sets that should be investigated in order to create a more thorough report.

Some of these may include demographics, affluence, population density, spending trends, and crime rates to name a few.