Florida City Data Analysis

Optimal location for new Coffee Shop Business

Daniela Mosco

March 15th, 2020

1. The Business Problem

There are many reasons for a business failing to turn a profit such as the control of business owners.

The business location is an area of concern that a business has a lot of control over, because the location of a business in the service industry has such a strong impact on whether that business succeeds or fails.

This means that it is an area that demands research and analysis.

A particularly difficult area of business is the restaurant and grocery industry because it is oversaturated and highly competitive which results in new business owners hoping to open a restaurant or shop in an already saturated market.

They should intent on finding a location that gives them the best possible chance for success.

2. Audience

The targeted audience for this project is business people or future investors who plan to start a new coffee shop store business, which they will be specifically looking to locate it in the Florida city area.

The research may also be of interest to business owners looking to expand their coffee shops stores in the Florida city area.

3. Data Sources

- Mapbox API will be used to get geographical latitude and longitude coordinates.
- Geographical location data will be scraped from https://en.wikipedia.org/wiki/List_of_counties_in_Florida.
- Foursquare will be used to gather information on venues in the locations that will be researched. The API returns a JSON list of venues that contains coordinate city, state, country, locations, category.

Example JSON data returned from Foursquare:

```
"url": "http://www.centralparknyc.org",
"likes": {
  "count": 17370,
  "summary": "17370 Likes"
},
"rating": 9.8,
"ratingColor": "008551",
"ratingSignals": 18854,
"beenHere": {
  "count": 0,
  "unconfirmedCount": 0,
  "marked": false,
  "lastCheckinExpiredAt": 0
},
"photos": {
  "count": 26681,
  "groups": [
    {
      "type": "venue",
      "name": "Venue photos",
      "count": 26681,
      "items": [
        {
          "id": "513bd223e4b0e8ef8292ee54",
          "createdAt": 1362874915,
          "source": {
            "name": "Instagram",
            "url": "http://instagram.com"
          },
          "prefix": "https://igx.4sqi.net/img/general/",
          "suffix": "/655018_Zp3vA90Sy4IIDApvfAo5KnDItoV0uEDZeST7bWT-qzk.jpg",
          "width": 612,
          "height": 612,
          "user": {
            "id": "123456",
```


4. Proposed solution

This project will analyze the postal code areas of Florida, USA. The data from 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.

This data will also be used to make recommendations to business owners about the location of optimal area for setting up a new business.

5. Foursquare sample

In this project we will be using the latitude and longitude coordinates of coffee shops across Florida city in order to get an idea of how saturated the coffee shop market is within each postal code.

Then we can use this information to look at areas of lower saturation as possible locations for a new business.

6. 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.

	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]

Figure 1. Resulting dataframe

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

The mapbox API was used to generate accurate latitude and longitude variables for each area code, using geocoding. Then these coordinates data were then built into a dataframe, using Pandas.

	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

Figure 2. Florida's Neighborhoods Coordinates

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

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.

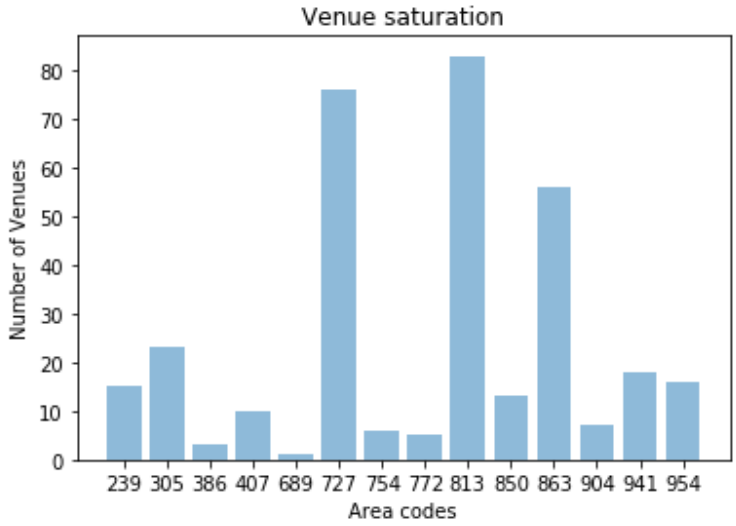

Figure 3. List of venues in each location

Folium was used to visualize the Florida 863 area on the map, as it is shown below:

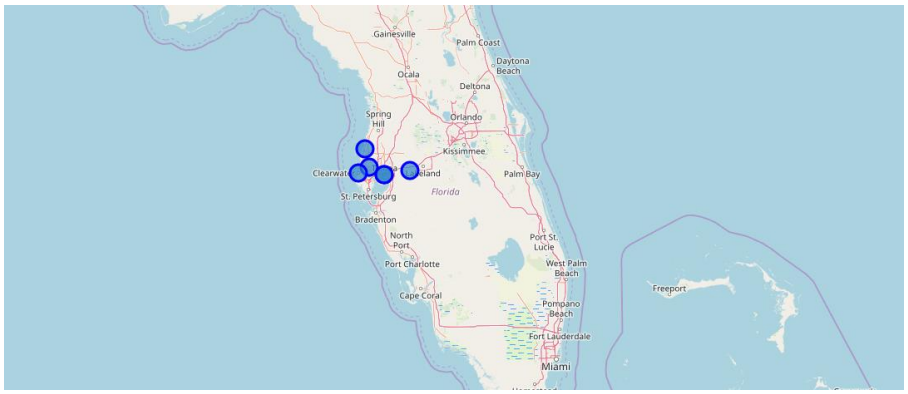

Figure 4. Florida 863 area on the map

The process shown above was repeated on the smaller data set generated by refining our search to Florida 863. Then. it was analyzed with the k-means algorithm which permitted a decision to be made on the best location for a new coffee shop.

7. Results

The results of our 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

Figure 5. K-means clustering results

8. 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.