

Analysis of potential new coffee shop location in Florida:

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.

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.

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_2p3vA90Sy4IIDApvfAo5KnDitoV0uEDZeST7bWT-qzk.jpg",
          "width": 612,
          "height": 612,
          "user": {
            "id": "123456",

```

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.

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.

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.

	Postcode	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront

Figure 1. Example of the 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.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	The Beaches	43.676357	-79.293031	Domino's Pizza	43.679058	-79.297382	Pizza Place
4	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood

Figure 2. Toronto'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.

This dataset was then analyzed with the k-means algorithm which permitted a decision to be made on the best location for a new coffee shop.