# Capstone Project

## The Battle of Neighborhoods – Week 2

### Applied Data Science Capstone by IBM / Coursera

Stephenas Ali
2019

# Table of Contents

Introduction: Business Problem....3
Data....4
    Sample of Wikipedia List of Neighborhoods in Miami data....4
    Sample of Foursquare API data for List of Venues....5
    Sample of Foursquare API data for List of Coffee Shops....5
Methodology....6
Results....7
Discussion & Recommendations....12
Conclusion....12

# Introduction: Business Problem

In this project we will try to find an optimal location for a coffee shop. Specifically, this report will be targeted to stakeholders interested in opening an **coffee shop** near to **Downtown Miami**, in Florida.

Since there are lots of cafes in Downtown Miami, we will try to detect **the most suitable location, in or around the surrounding neighborhoods,** where the business can thrive. We are also particularly interested in **areas with popular venues or businesses to attract potential customers**.

We will use our data science skills to evaluate a few of the most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

# Data

Based on definition of our problem, factors that will influence our decision are:

- the distance of neighborhood from the Downtown area
- number of coffee shops in the neighborhood
- number of possible supporting business areas and popular venues in the neighborhood, if any.

This information will be key in developing our analytical model and will be sufficient to obtain our solution. To gather this information the following data sources will be used:

- A Wikipedia entry with a list of neighborhoods in Miami. The web page consists of the neighborhoods and also their geographical coordinate data. This data will be scrapped, formatted and cleaned to be utilized in our analysis. https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Miami

- To investigate the competitive environment in each neighborhood, such as existing coffee shops and other businesses or general places of interest. This data will be retrieved using the **Foursquare API.** Foursquare claims to be the most trusted, independent location data and technology platform for businesses.

### Sample of Wikipedia List of Neighborhoods in Miami data

| | Neighborhood | Population2010 | Population/Km² | Latitude | Longitude |
|---|---|---|---|---|---|
| 0 | Brickell | 31759 | 14541 | 25.758 | -80.193 |
| 1 | Downtown | 71,000 (13,635 CBD only) | 10613 | 25.774 | -80.193 |
| 2 | Little Havana | 76163 | 8423 | 25.773 | -80.215 |
| 3 | Lummus Park | 3027 | 3680 | 25.777 | -80.201 |
| 4 | Overtown | 6736 | 3405 | 25.787 | -80.201 |
| 5 | Park West | 4655 | 3635 | 25.785 | -80.193 |
| 6 | The Roads | 7327 | 4899 | 25.756 | -80.207 |

## Sample of Foursquare API data for List of Venues

| location.address | location.lat | location.lng | location.neighborhood | name |
|---|---|---|---|---|
| 1451 Brickell Ave | 25.758121 | -80.192405 | Brickell | Echo Brickell |
| 1501 Brickell Ave | 25.757828 | -80.192876 | Brickell | St. Jude's Catholic Church |
| 1450 Brickell Ave | 25.758584 | -80.192926 | Brickell | 1450 Brickell Ave |
| NaN | 25.758211 | -80.193154 | Brickell | Google Miami |
| 1450 Brickell Ave | 25.758583 | -80.192805 | Brickell | City National Bank |
| 185 SE 14th Ter | 25.759420 | -80.190957 | Brickell | Fortune House Hotel |
| 1451 Brickell Avenue | 25.758066 | -80.192415 | Brickell | JOE & THE JUICE |
| NaN | 25.761573 | -80.198373 | Brickell | Haitian Consulate |

## Sample of Foursquare API data for List of Coffee Shops

| location.address | location.neighborhood | location.lat | location.lng | name |
|---|---|---|---|---|
| 1541 Brickell Ave | Brickell | 25.757370 | -80.193590 | Mercon Coffee |
| 1200 Brickell Bay Dr | Brickell | 25.761971 | -80.190529 | Finca's Coffee |
| 299 Se 3rd St | Downtown | 25.772535 | -80.190724 | Allegro Coffee Company |
| 117 SE 2nd Ave | Downtown | 25.772994 | -80.190179 | Eternity Coffee Roasters |
| 2 W Flagler St | Downtown | 25.773989 | -80.193687 | Bistro Coffee |
| 145 E Flagler St | Downtown | 25.774293 | -80.190945 | Ever Coffee |
| 444 Brickell Ave | Downtown | 25.769225 | -80.190680 | Coffee To Go |
| 1 SE 3rd Ave | Downtown | 25.774260 | -80.189000 | Q Coffee Club |
| 16 NE 3rd Ave | Downtown | 25.774416 | -80.188576 | Panini Coffee Bar |
| 200 NW 1st Ave | Downtown | 25.777711 | -80.195711 | Parliament Coffee |
| 110 SE 3rd Ave | Downtown | 25.773300 | -80.189060 | Starbucks |

# Methodology

In first step we established our understanding of the business problem. Secondly, we collected and cleaned the required data, inclusive of location information for neighborhoods, coffee shops, and other general venues of interest that would provide customers. During the data cleaning process we also narrowed the list of available neighborhoods to only those within 2.5km of Downtown Miami. This was done using the haversine formula, which is used to determine the great-circle distance between two points on a sphere given their longitude and latitude values.

The third step, will be our analysis of the data acquired. Some basic exploratory statistics will be calculated to get acquainted with the data, such as number of coffee shops found, the average number of coffee shops per neighborhood etc. Then, all the information collected will be sketched onto a heatmap to provide a visual perspective of the data. We would be able to see all of the selected neighborhoods, all venues of interests, and all coffee shops.

As we continue our analysis process, we would then further refine the information by categorizing the data of the neighborhoods by using the average number of coffee shops. The categories would cover those above average, below average, or those with no coffee shops at all within an initial 500 meter radius. This process would be iterated by eliminating neighborhoods with coffee shops above the average, and then by selecting neighborhoods with no coffee shops at all and increasing the scan radius. Folium maps would again be used to visualize the data.

Once a neighborhood is selected, in the final step of the analysis we would seek to pinpoint the most promising location in this preferred neighborhood for the coffee shop by identifying the area with the most dense venues. To accomplish this, we would use the k-means clustering, machine learning algorithm to find the largest groups of areas with potential customers.

# Results

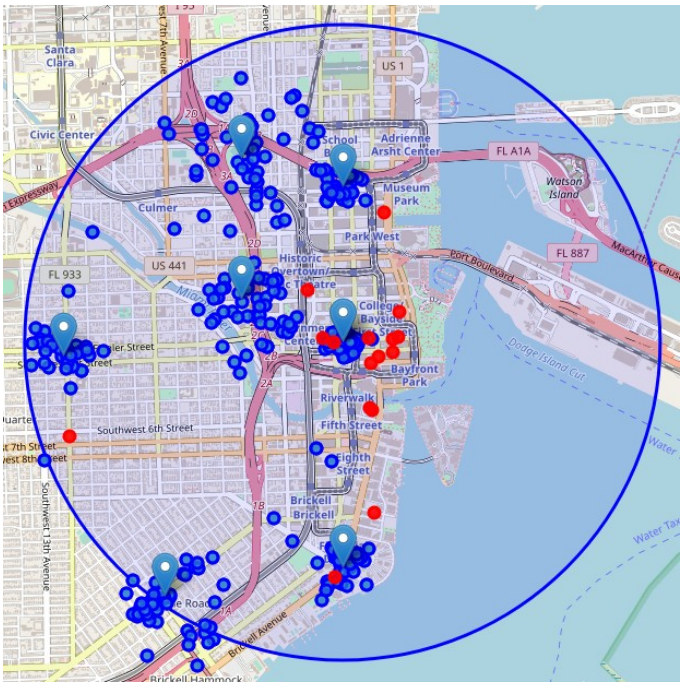
The following is a list of exploratory statistics found on the area.

- Total number of places of interests: 350
- Total number of coffee shops: 16
- Percentage of coffee shops to general places of interests: 4.57%
- Average number of coffee shops per neighborhood: 2.286

The following dataset lists the selected neighborhoods with a count of venues and coffee shops found per area

| Neighborhood | Population2010 | Population/Km² | Latitude | Longitude | Venue_Count | Coffee_Shop_Count |
|---|---|---|---|---|---|---|
| Brickell | 31759 | 14541 | 25.758 | -80.193 | 50.0 | 2.0 |
| Downtown | 71,000 (13,635 CBD only) | 10613 | 25.774 | -80.193 | 50.0 | 12.0 |
| Little Havana | 76163 | 8423 | 25.773 | -80.215 | 50.0 | 1.0 |
| Lummus Park | 3027 | 3680 | 25.777 | -80.201 | 50.0 | 0.0 |
| Overtown | 6736 | 3405 | 25.787 | -80.201 | 50.0 | 0.0 |
| Park West | 4655 | 3635 | 25.785 | -80.193 | 50.0 | 1.0 |
| The Roads | 7327 | 4899 | 25.756 | -80.207 | 50.0 | 0.0 |

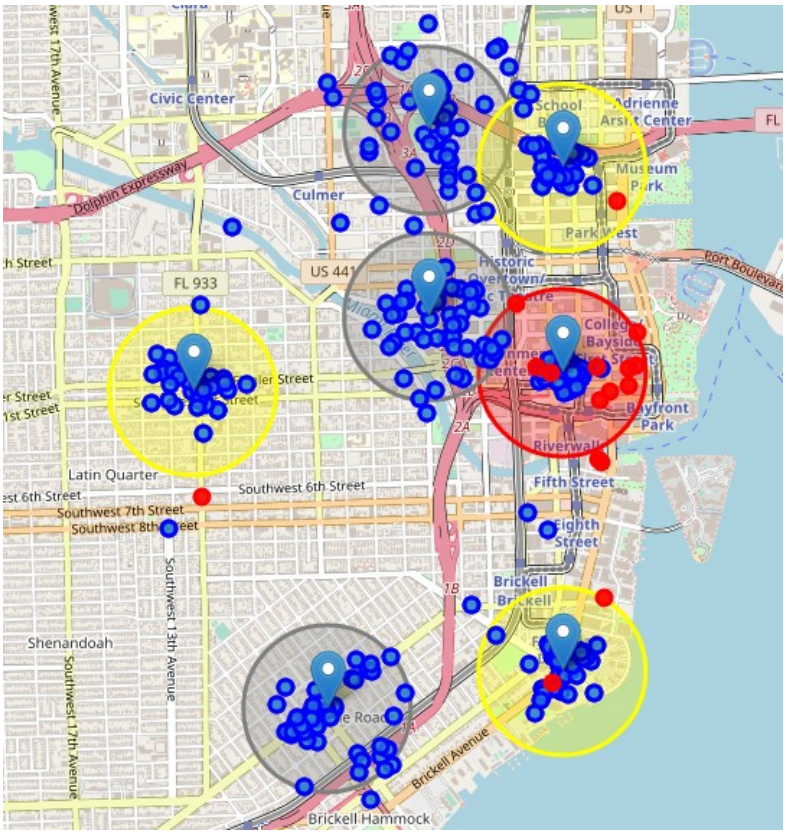
The information when charted on the Folium map looks as follows, with markers representing neighborhoods, red dots representing coffee shops and blue dots representing venues of interest:



Next, the neighborhoods were categorized according to the average number of coffee shops. The column **Category** was added and given a color encoding to be easily visible on the map.

| | Neighborhood | Latitude | Longitude | Category |
|---|---|---|---|---|
| 0 | Brickell | 25.758 | -80.193 | yellow |
| 1 | Downtown | 25.774 | -80.193 | red |
| 2 | Little Havana | 25.773 | -80.215 | yellow |
| 3 | Lummus Park | 25.777 | -80.201 | gray |
| 4 | Overtown | 25.787 | -80.201 | gray |
| 5 | Park West | 25.785 | -80.193 | yellow |
| 6 | The Roads | 25.756 | -80.207 | gray |

The resultant map looks as follows:



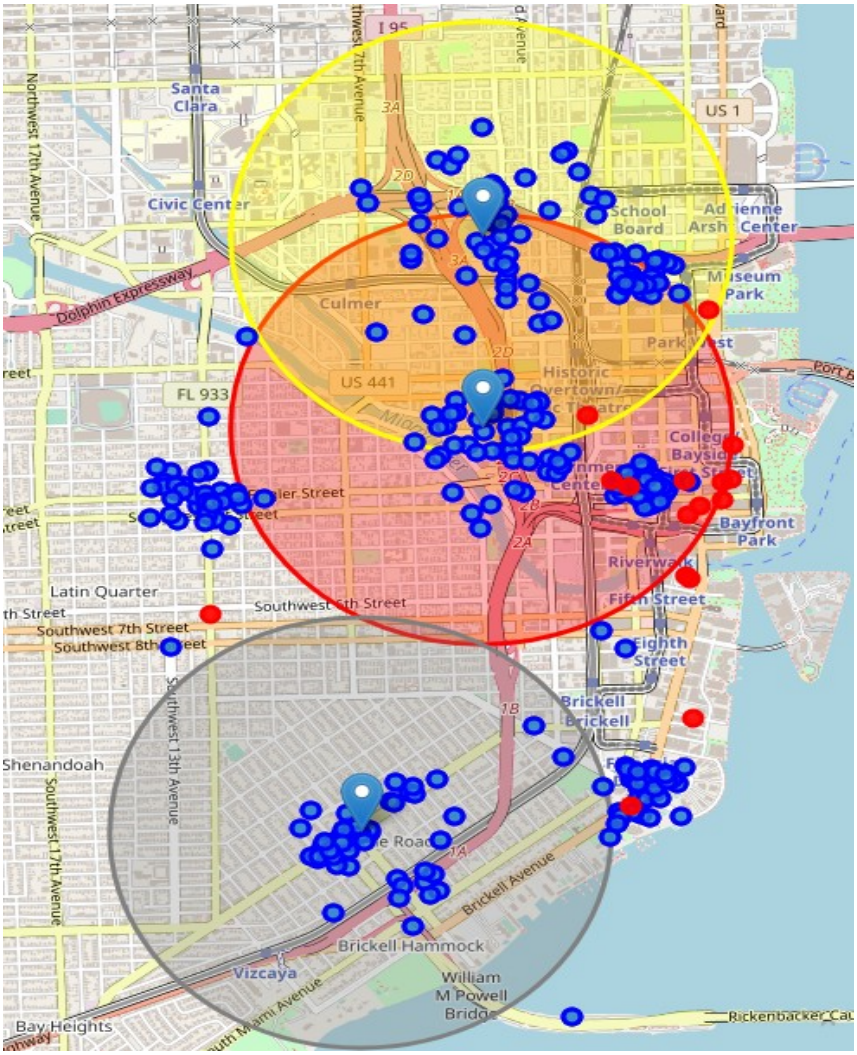

With this visualization we can clearly see that 3 areas have no coffee shops at all within the initial 500 meter radius.

The process was then repeated, the neighborhoods categorized in gray were focused on and the others were removed.

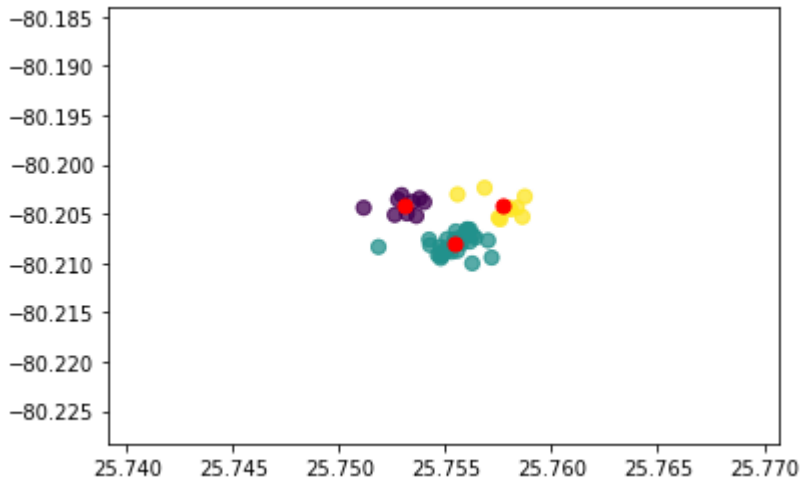
The radius distance was increased from 500 meters to 1.25km, and the data was re-categorized. The dataset now looked as follows:

| | Neighborhood | Latitude | Longitude | Category |
|---|---|---|---|---|
| 0 | Lummus Park | 25.777 | -80.201 | red |
| 1 | Overtown | 25.787 | -80.201 | yellow |
| 2 | The Roads | 25.756 | -80.207 | gray |

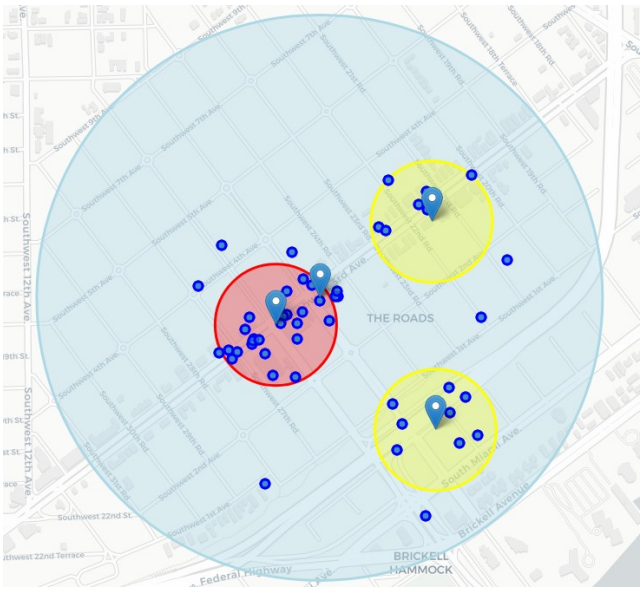
The re-plotted map looked as follows:



As we can see within the new range only one neighborhood still has no coffee shops listed. This neighborhood was identified as ***The Roads*** and will be used specifically in further analysis. K-means clustering was performed to determine where had the best cluster of venues of interest to establish the business. A cluster of 3 was applied on the venues dataset and a simple scatter plot was created.




To get a better perspective, each cluster point was taken and plotted onto the map and categorized to produce the following:




| | Latitude | Longitude | Centroid | Venues | Category |
|---|---|---|---|---|---|
| **0** | 25.753058 | -80.204135 | Centroid 1 | 16 | yellow |
| **1** | 25.755396 | -80.208083 | Centroid 2 | 48 | red |
| **2** | 25.757696 | -80.204231 | Centroid 3 | 12 | yellow |

As we can see, the location at Centroid 2 seems to be the most popular location, densely populated with business and potential customers.

A final dataset was created to confirm the listed venues and locations. The following is a sample of the data:

| location.address | location.lat | location.lng | location.neighborhood | name |
|---|---|---|---|---|
| 2525 SW 3rd Ave, unit Cu2 | 25.755430 | -80.207555 | The Roads | Bocas Grill Brickell |
| 2498 SW 3rd Ave | 25.756415 | -80.207406 | The Roads | Dr. Idalia Lastra, D.M.D. |
| 2525 SW 3rd Ave | 25.755683 | -80.207425 | The Roads | Nordica Condo |
| 2525 SW 3rd Ave | 25.755627 | -80.207822 | The Roads | Nordica Apartments |
| NaN | 25.755930 | -80.206997 | The Roads | Kiki's mind |
| 2525 SW 3rd Ave | 25.755074 | -80.207552 | The Roads | Pool @ Nordica Condominium |
| NaN | 25.755758 | -80.208011 | The Roads | Roads Pediatrics |
| NaN | 25.756279 | -80.207183 | The Roads | Saint Sophia |
| 2525 SW 3rd Ave | 25.755590 | -80.207900 | The Roads | Dr Cordero, Perez-Silva Pediatrics |
| NaN | 25.754230 | -80.207594 | The Roads | Beth David Congregation |
| SW 3rd Ave & SW Oak St | 25.755489 | -80.206770 | The Roads | Taqueria Villanueva - SW 3rd |

# Discussion & Recommendations

The analysis and results surfaced that the neighborhood which would be most recommended to establish a new coffee shop would be The Roads. There are no current coffee shops there within a 1.25km radius and there is a dense population of businesses for potential customers.

One key observation to be pointed out would be the accuracy and comprehensiveness of the data available particularly with the Foursquare API. Note that not all coffee shops may have been listed and note that not all businesses may be listed accurately. Also there are limitations to the the amount of data that the API is able to return at a time, therefore it was noticed that the randomized information could skew results.

# Conclusion

Purpose of this project was to aid stakeholders in narrowing down the search for optimal location for a new coffee shop in the vicinity of Downtown Miami. We identified prospective neighborhoods, and then identified existing coffee shop distributions and supporting businesses using Foursquare data.

Clustering analysis of those locations was then evaluated to determine the most viable option for a new coffee shop.

The recommended solution satisfied low competition from other coffee shops, with supporting business establishments around whilst also being close to Downtown Miami.