

Location Options for the Pittsburgh-based, Café Cycle Mobile Fleet

Author: Tom Schott

Date: 26 June 2020

This analysis document is in partial completion of requirements for the [Applied Data Science Capstone Course](#) and the [IBM Data Science Professional Certificate](#). This section of the document completes the first requirement of the “Capstone Project - The Battle of Neighborhoods” to provide an introduction to the problem or opportunity we are trying to solve and discuss the data to be used to solve it.

Table of Contents

INTRODUCTION/BUSINESS PROBLEM:	3
DATA COLLECTION AND PREPARATION	3
METHODOLOGY	4
1. DETERMINE AREA TO ASSESS	4
2. RETRIEVE ALL BIKING RELATED VENUES AND ADD TO DATAFRAME.....	4
3. CREATE LOCATIONS CONSIDERATIONS VIEW MAP	4
4. DETERMINE THE TOP 10 LOCATIONS CENTERS TO CONSIDER.....	5
5. DETERMINE COMPETITOR LOCATIONS	6
RESULTS	7
6. CREATE CLUSTER LOCATION MAPS FOR EACH CLUSTER.....	7
7. CREATE LOCATION OPTION PRIORITY LIST	13
DISCUSSION	14
CONCLUSION	14
REFERENCES	15

Table of Figures

Figure 1: Location Considerations View	5
Figure 2: Color Coded Cluster Circles and Clustered Cycling Venues.....	6
Figure 3: Cluster Locations with Competitors.....	7
Figure 4: Location Option 0.....	8
Figure 5: Location Option 1.....	9
Figure 6: Location Option 2.....	9
Figure 7: Location Option 3.....	10
Figure 8: Location Option 4.....	10
Figure 9: Location Option 5.....	11
Figure 10: Location Option 6.....	11
Figure 11: Location Option 7.....	12
Figure 12: Location Option 8.....	12
Figure 13: Location Option 9.....	13

Table of Tables

Table 1: Mean of each Kmeans Cluster Centers.....	5
Table 2: Location Option Priority List.....	14

INTRODUCTION/BUSINESS PROBLEM:

With the onslaught of the Coronavirus 19 pandemic, there is an increase in cycling activities for both families and individuals as persons change from indoor and group activities to outdoor and socially distanced activities. This increase in cyclists provides a likely increase in demand to provide specialty beverages (coffee) in an outdoor seating situation.

Café Cycle is a custom beverage provider that caters to the cycling community in Pittsburgh, Pennsylvania. While it has a main coffee shop in a fixed location in Pittsburgh, it recently acquired three new coffee trucks that allow the company to deliver the same products and services, along with an outdoor seating for comfort and safe social interaction.

Café Cycle would like to identify possible operating locations for its small fleet of trucks. Since it already has brand awareness within the Pittsburgh cycling community, it is looking for locations where there is a greater likelihood of cyclist to be present. It would also like to know of competitors in the area that might detract from their sales.

DATA COLLECTION AND PREPARATION

Foursquare Developers' Places API allows applications to query its Foursquare global database of venue and user content associated with business locations. These APIs are used to pull data locations of where the Pittsburgh cycling community is more likely to frequent and group these locations into possible vicinities to place their trucks. We will also identify the competitors in the vicinity that might affect their success.

Both the possible locations and the competitors will be displayed on a map using the visualization library, Folium. For the competitors, we will also provide a list with the name and address, location, distance to a group centers, and the overall rating of the business location. We will focus on any business within a 20-mile radius of the center of Pittsburgh (where the three rivers join). The venue categories that we will filter on are below. For reference see <https://developer.foursquare.com/docs/resources/categories>.

Cycling Community Data: Venues associated with cycling

Category	categoryId
Outdoors & Recreation → Bike Trail	56aa371be4b08b9a8d57355e
Shop & Service → Bike Shop	4bf58dd8d48988d115951735
Travel & Transport → Bike Rental / Bike Share	4e4c9077bd41f78e849722f9

Competitor Data: Venues associated with sit down coffee shops

Category	categoryId
Food → Café	4bf58dd8d48988d16d941735
Food → Coffee Shop	4bf58dd8d48988d1e0931735

Shop & Service → Food & Drink Shop → Coffee Roaster	5e18993feee47d000759b256
---	--------------------------

METHODOLOGY

This analysis will use the following approach to develop its recommendations:

1. Determine area to assess. In this case, it is any location within a 20 mile radius of the point where the three rivers of Pittsburgh join ([Point State Park Fountain](#)). Add as center of the Folium map.
2. Retrieve from Foursquare all locations where there are biking related venues and add to a dataframe. Important data is the business location name, location, and venue.
3. Create Folium map of center point with 20-mile radius and add all potential location centers. Add all cycling related venue locations. This map view is the [location considerations view](#).
4. Using the k-mean clustering for the geographic distances between objects, identify the center of 10 potential location centers. Create a dataframe with this list and the center location.
5. For each potential location, identify all Competitor (coffee shops) locations within a three-mile radius.
6. Create a separate Folium map for each location that shows the center point, all cycling related locations and all competitors in the radius.
7. Create a rating for list for each location using a criteria based on number of cycling related locations and number of competitors. Include a Google Maps link to the center location to assist in the site survey.

1. Determine Area to Assess

In this case, it is any location within a 20-mile radius of the point where the three rivers of Pittsburgh join (Point State Park Fountain). This will be the center for the Locations Considerations View Folium map.

2. Retrieve all Biking Related Venues and Add to Dataframe

Retrieve from Foursquare all locations where there are biking related venues and add to a dataframe. Important data is the business location name, location, and venue.

3. Create Locations Considerations View Map

Create Folium map of center point with 20-mile radius and add all potential location centers. Add all cycling related venue locations. This map view is the **Location Considerations View**.

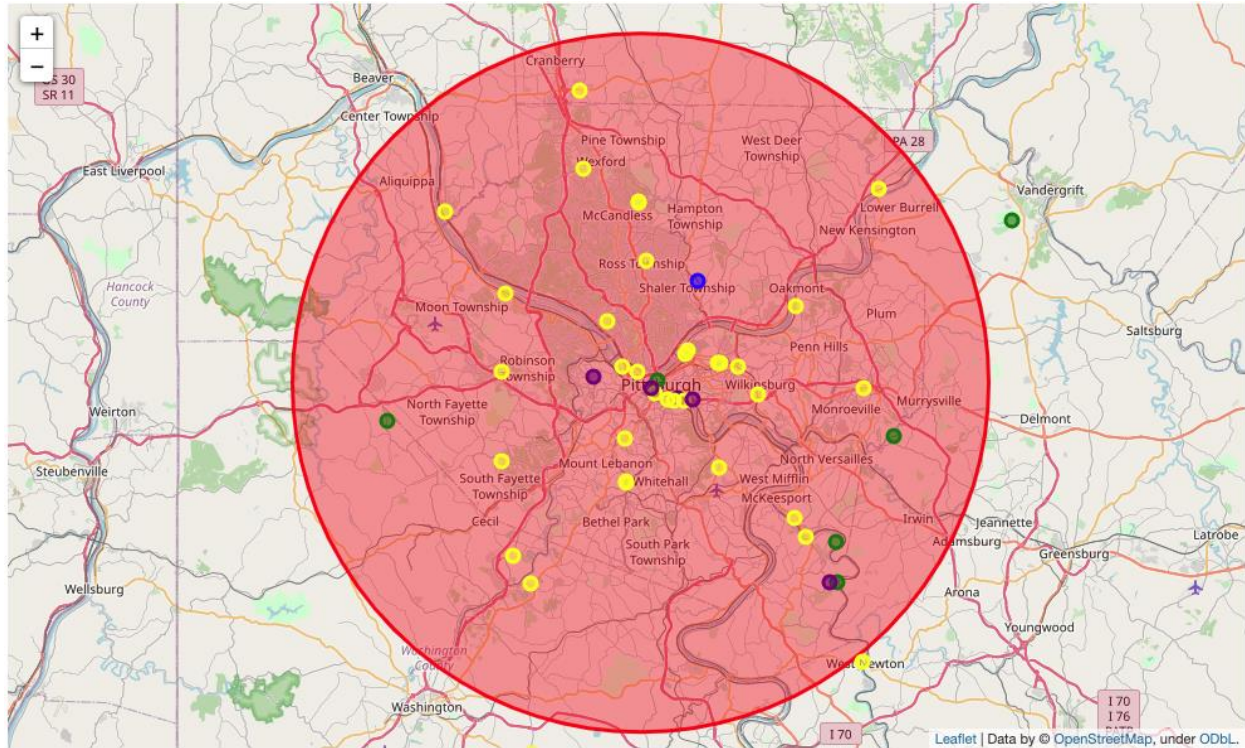


Figure 1: Location Considerations View

4. Determine the top 10 locations centers to consider

Now that we can see where our overall area of consideration around Pittsburgh for placing the cafe shops and the biking related venue locations in this area, we will now identify potential vicinities to locate the food trucks. To do this, we will use K-Means Clustering to find 10 possible locations (10 is an arbitrary number for consideration) that is closest to biking locations. The center of these clusters will be used for further analysis.

Using the k-mean clustering for the geographic distances between objects, identify the center of 10 potential location centers. Create a dataframe with this list and the center location.

Table 1: Mean of each Kmeans Cluster Centers

	Cluster Labels	lat	lng
0	0	40.372650	-79.926689
1	1	40.606710	-80.038295
2	2	40.135228	-80.131575
3	3	40.491360	-80.210466
4	4	40.418810	-79.752237
5	5	40.429238	-80.009142

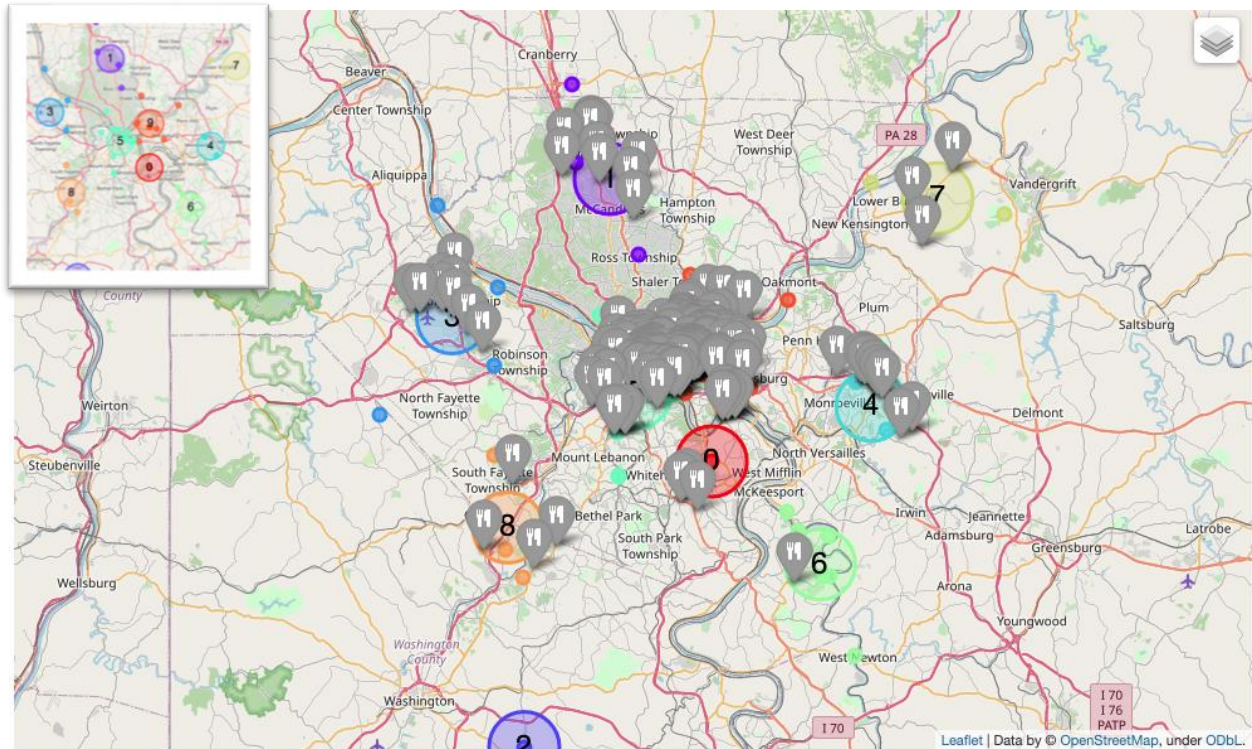


Figure 3: Cluster Locations with Competitors

RESULTS

6. Create Cluster Location Maps for Each Cluster.

The purpose of our analysis was to capture the necessary data to be able to assess each potential area separately. Once a map reconnaissance of the area is completed, individuals will move to the actual ground to perform a site survey. These maps provide the means to identify potential areas to operate the coffee trailers to allow easier and faster on the ground assessments.

There are 47 Pittsburgh area venue locations retrieved from Foursquare that identify either their business as cycling related or are locations where the activity is primarily as a biking trailhead. Using the assumption that cycling related venues are more likely to be established in areas where cycling is popular, or to attract more cyclist to the area, these points provide locations from which Café Cycle will be able to identify target areas to consider locating the coffee truck cafes.

The number of 10 clusters was an arbitrary number chosen to create enough options to consider, but not so many that the staff would not be able to conduct site visits. Of the 10 areas, two were such a distance from Pittsburgh and in relatively rural settings that only one cycling activity was available (Area 7 and 9). These two locations, identified as the least favorable locations, also had no identified coffee shops within its three mile radius.

As expected, Coffee Shops were prevalent in large numbers in the two downtown locations (Areas 5 and 9). Due to retrieval limitations with Foursquare, the results of competitor venues were limited to a maximum 100 returns. This limitation would be addressed in an analysis that would be used for making a financial decision. Areas 1, 6, 3, and 8 had more than three cycling venues and either a small or moderate number of competitor coffee shops. Since these locations all surround the downtown Pittsburgh region where the population and traffic is less dense and cycling activities are more likely to occur, they may prove to be the optimum location to setup the coffee trucks.

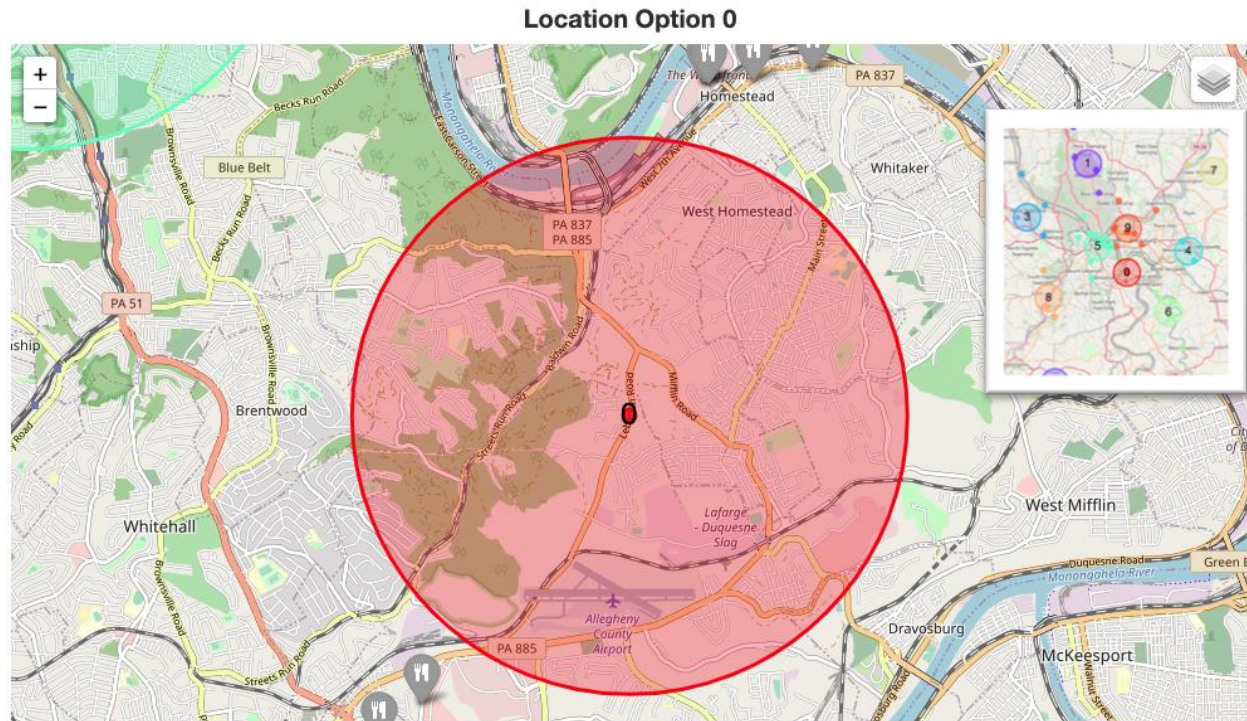


Figure 4: Location Option 0

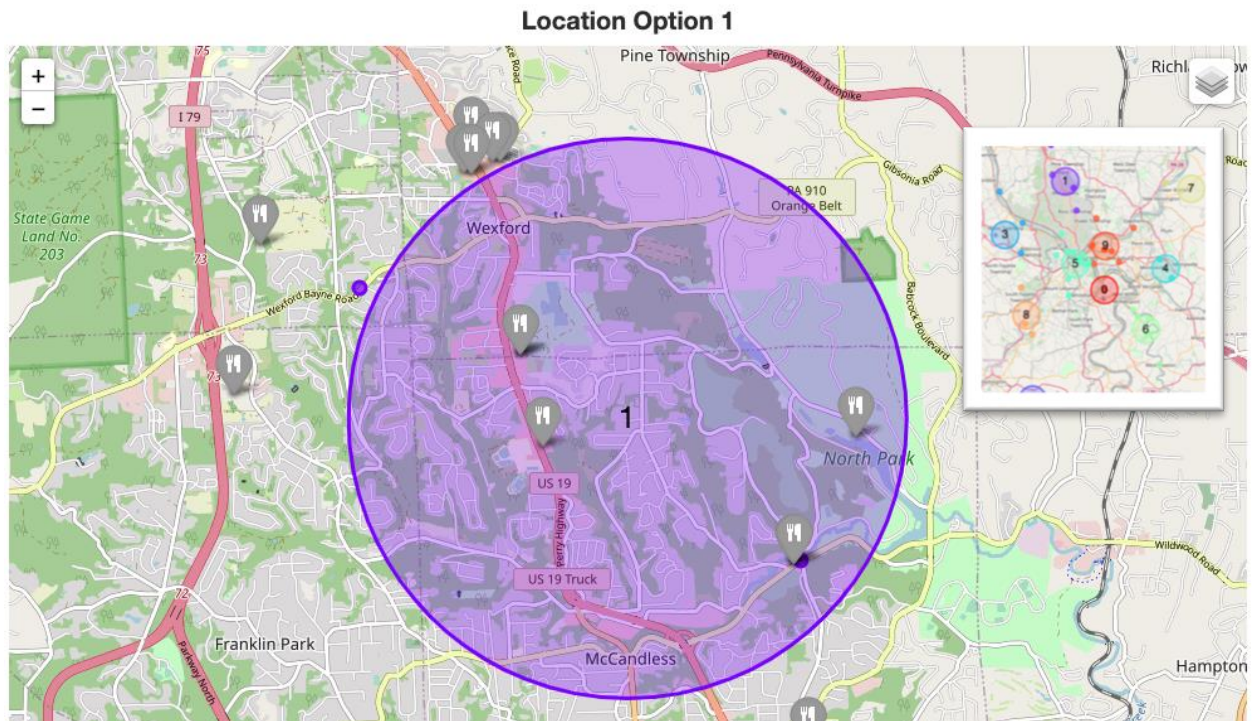


Figure 5: Location Option 1

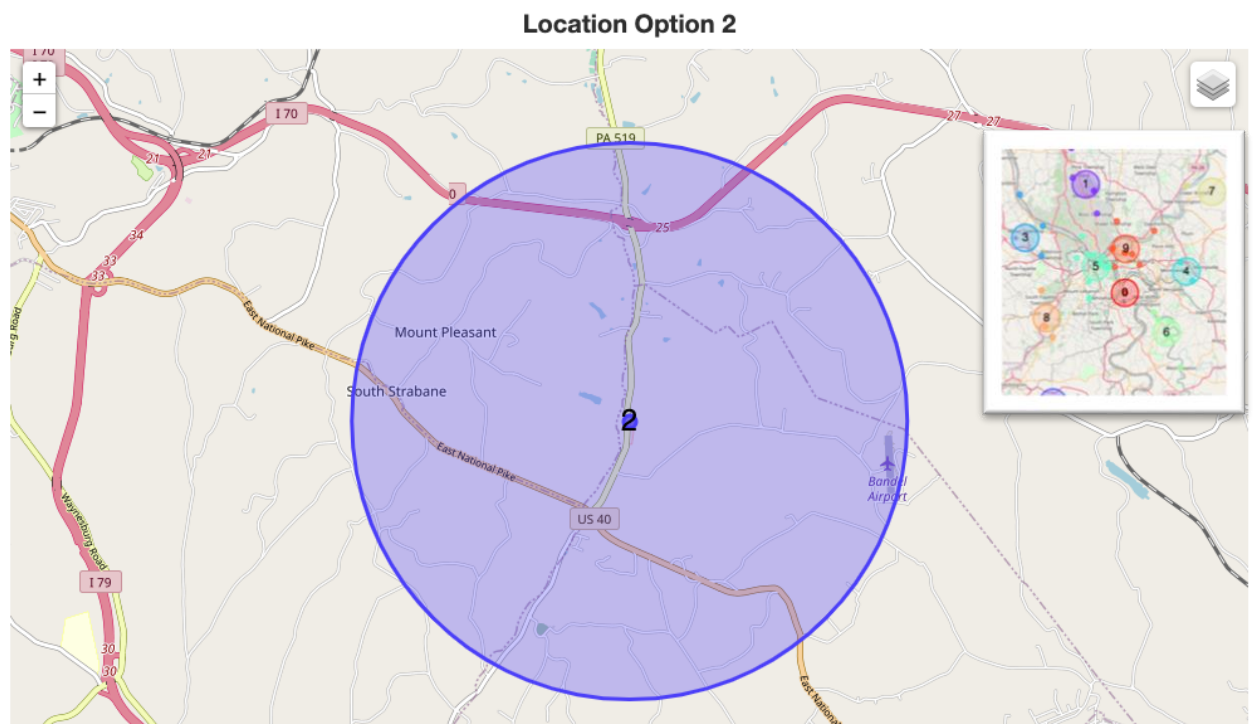


Figure 6: Location Option 2

Location Option 3

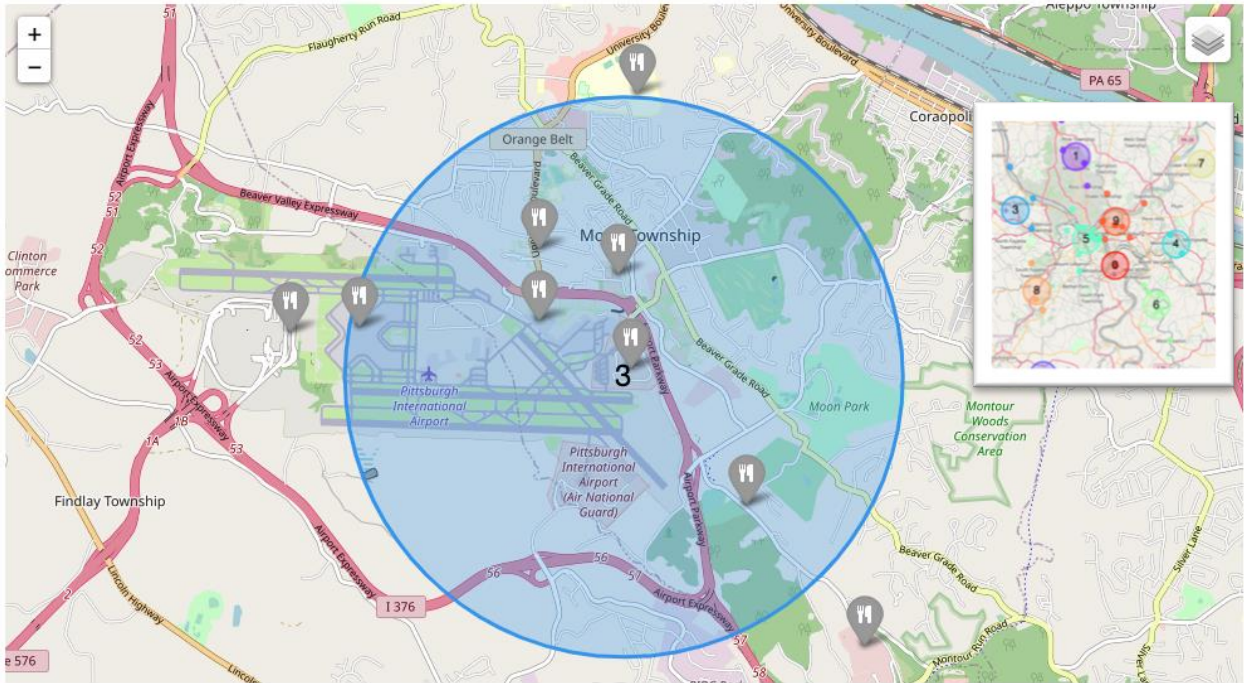


Figure 7: Location Option 3

Location Option 4

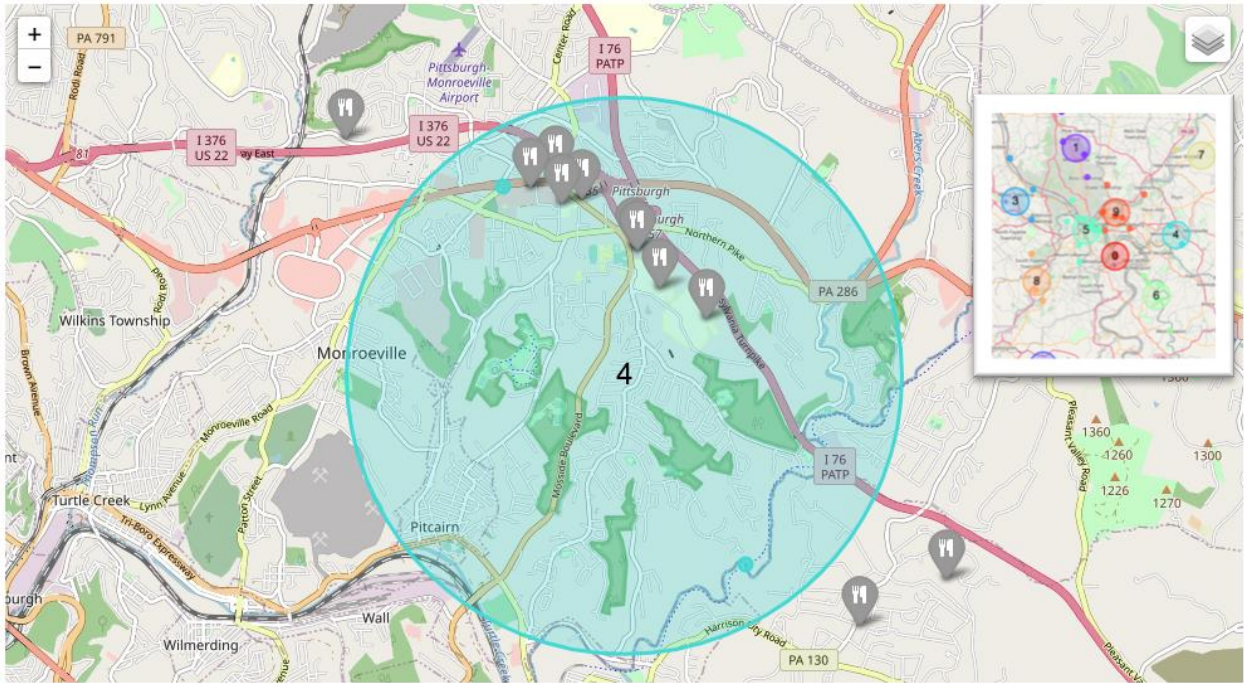


Figure 8: Location Option 4

Location Option 5

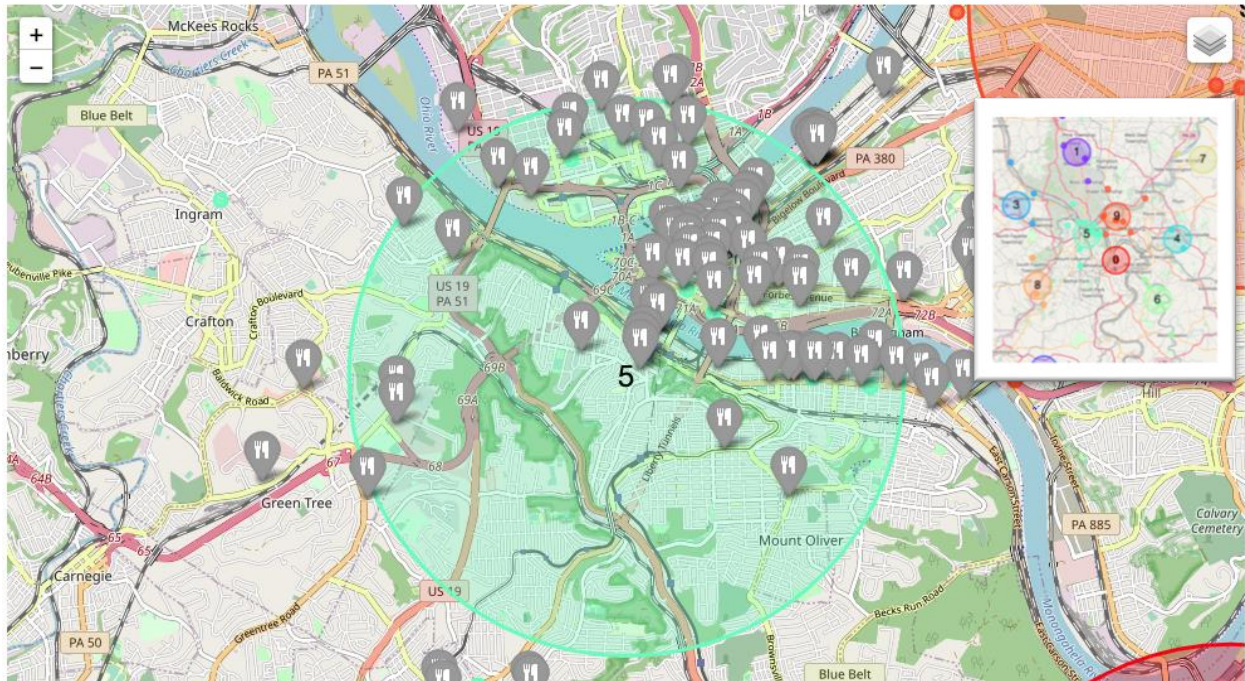


Figure 9: Location Option 5

Location Option 6

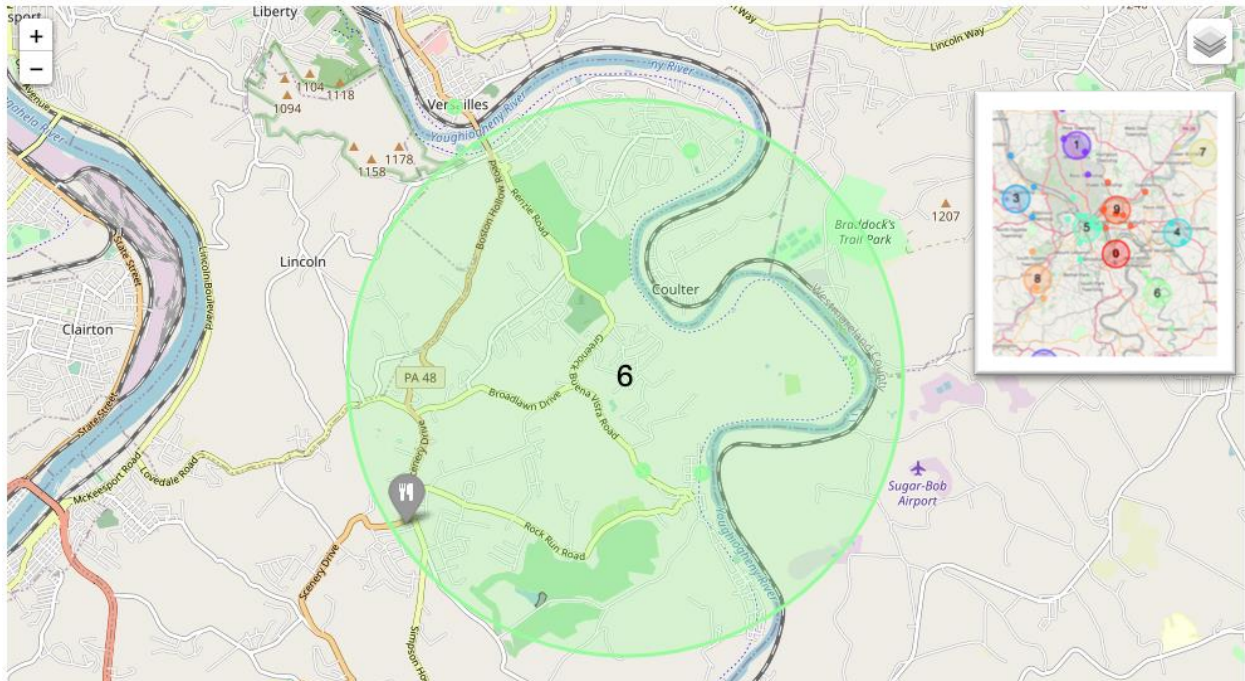


Figure 10: Location Option 6

Location Option 7

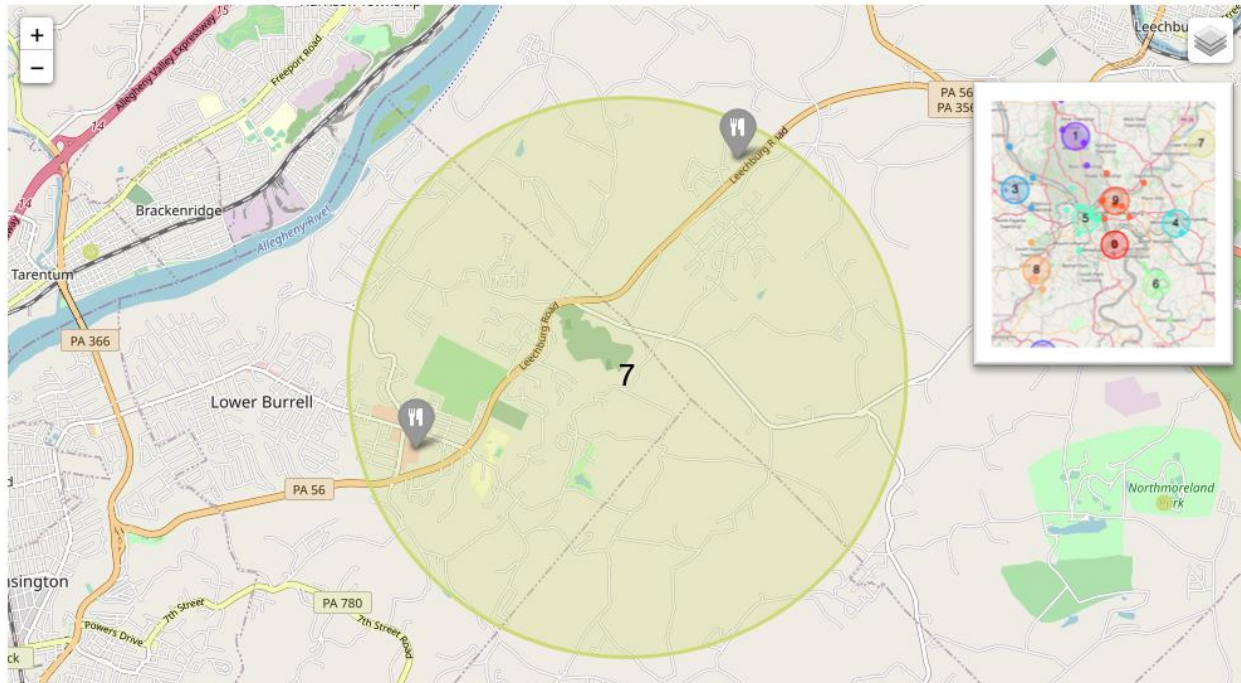


Figure 11: Location Option 7

Location Option 8

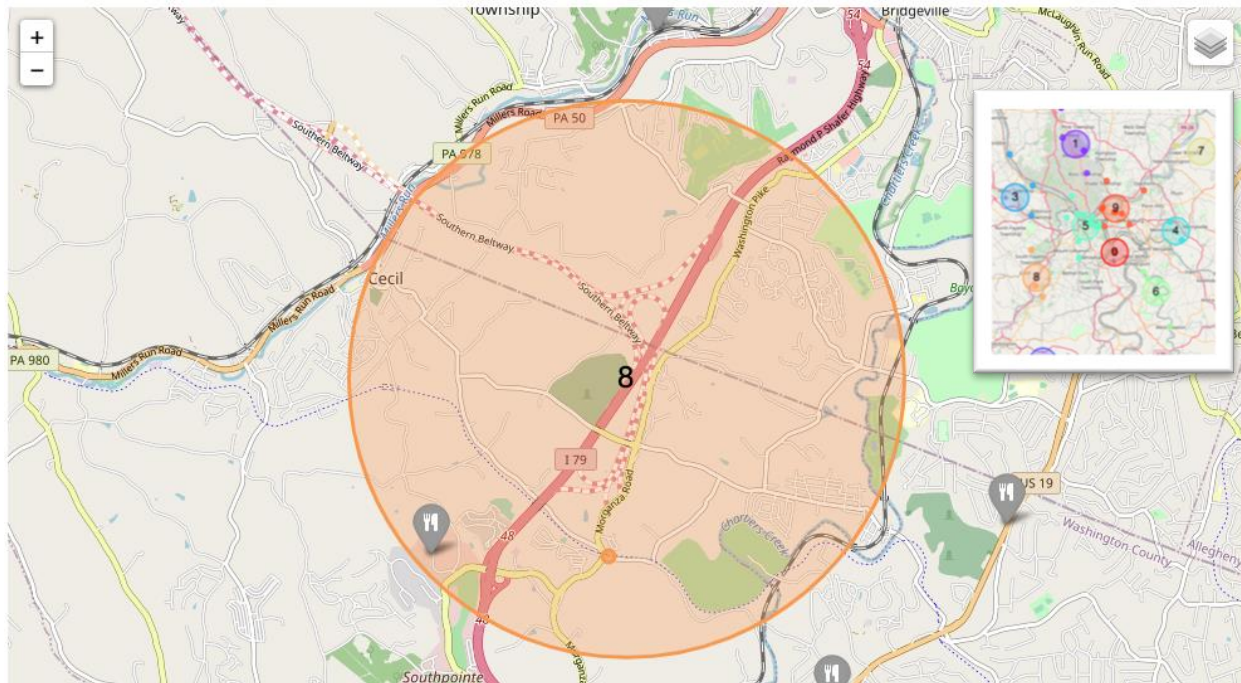


Figure 12: Location Option 8

Location Option 9

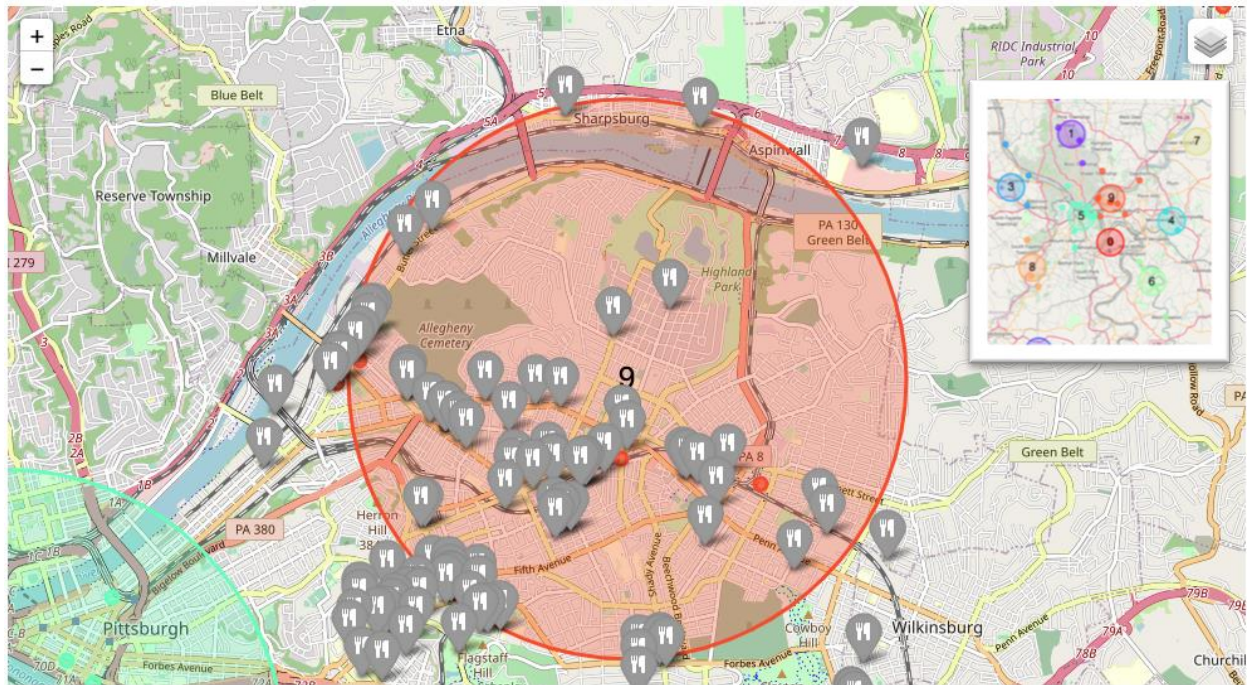


Figure 13: Location Option 9

7. Create Location Option Priority List

Create a rating for list for each location using a user defined criteria based on number of cycling related locations, number of competitors. Include a Google Maps link to the center point.

We will create a dataframe that lists each cluster region along with the following additional values

- Number of Cycling Activities
- Number of Cafes
- Overall Score ($10 * \text{Num of Cycling Venues} - \text{Number of Competitors}$)
- Rank order by Overall Score

Table 2: Location Option Priority List

	Area Name	Cycle Venues Count	Coffee Venues Count	Area_Score	GoogleMapLink
5	Area 5	15	100	249	https://www.google.com/maps/search/?api=1&quer...
9	Area 9	9	100	189	https://www.google.com/maps/search/?api=1&quer...
1	Area 1	5	14	63	https://www.google.com/maps/search/?api=1&quer...
6	Area 6	6	1	60	https://www.google.com/maps/search/?api=1&quer...
3	Area 3	4	9	48	https://www.google.com/maps/search/?api=1&quer...
8	Area 8	3	4	33	https://www.google.com/maps/search/?api=1&quer...
4	Area 4	2	11	30	https://www.google.com/maps/search/?api=1&quer...
7	Area 7	2	3	22	https://www.google.com/maps/search/?api=1&quer...
0	Area 0	1	10	19	https://www.google.com/maps/search/?api=1&quer...
2	Area 2	1	0	9	https://www.google.com/maps/search/?api=1&quer...

DISCUSSION

The results provide recommendations and priorities for target areas to locate the cycling vans. Each location will require a physical site visit and further analysis. The following are some shortcomings with the approach which will require further assessments outside of this assessment.

- Equates each cycling related location as an equal indicator of potential demand. This is not accurate. For example, a bike trail parking lot may provide greater opportunity for coffee drinkers than a location near a bike repair shop.
- Density of traffic that might use your services. Each cycling related location does not have the same number of potential customers frequenting establishment.
- Opportunity for non-cycling coffee drinkers to use your services. While you are targeting the cycling community, you may have more walker/runner traffic near a park with a bike trail that would use your services.
- Availability and cost to locate the mobile truck. This assessment doesn't determine the potential actual locations to host your mobile truck. Other considerations when assessment a location facility include availability of space to host the truck and seating, cost and type of contract needed to use location, or facilities available (i.e. restrooms or water) available at location.

CONCLUSION

The analysis presented in this study does not address all factors necessary for Café Cycle to decide on a location. Other factors such as population density, affluence, and age of the population are also indicators of potential demand and could be added to the model to provide additional value. In addition, the availability of adequate space to locate a mobile truck and the cost for use of these locations are important considerations when making a final decision.

Missing from the data that we used for our analysis are other venues that would attract cyclist. For example, public parks with cycling trails or locations along the Rails to Trails where persons

can enter the trailheads are not listed. Further efforts to map these locations and input them into the model will provide a more reliable indicator of demand.

Interesting enough is that the model returned different results when run multiple times. This is due to Foursquare returning varied number and venues. While some of this variance could be explained by Foursquare returning different venue locations where the maximum value returned of 100 was exceeded. In addition, there is some expected variance with each run of Kmeans to determine the optimum clustering locations when fitting the model due to the random initial location. While these variances affect the results (i.e. different color or cluster label for a cluster location), it had only a minor influence on the actual reported centers or the priority order of Location Option Priority List.

Lastly, it will take Café Cycles leadership to physically assess each area to make the right choice. The advantage of this new mode of delivering Café Cycle's product to its expanded customer base is that if the expected revenues are not fully recognized after operating for a period of time at a location, they have the ability to easily relocate their operations to a new location.

REFERENCES

GitHub Link of Completed Project:

https://github.com/schottr1/Coursera_Capstone/blob/master/Location%20Options%20Pittsburgh-based%20Cafe%20Cycle_hojCm4mBa.ipynb

List of Python Libraries Used

- Pandas: For creating and manipulating dataframes.
- Numpy: library to handle data in a vectorized manner
- Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.
- JSON: Library to handle JSON files.
- Geocoder: To retrieve Location Data.
- Matplotlib: Python Plotting Module.