Girl Scouts Cookie Booth Locations Project

Anju Nair

February 26, 2021

1 Business Understanding

1.1 Background

Girl Scouts is a pre-eminent leadership development organization for girls. And with programs from coast to coast and across the globe, Girl Scouts offers every girl a chance to practice a lifetime of leadership, adventure, and success. Their mission is to build girls of courage, confidence, and character, who make the world a better place.

Their extraordinary journey began more than 100 years ago with the original G.I.R.L.(Go-getter, Innovator, Risk-taker, Leader), Juliette Gordon "Daisy" Low. On March 12, 1912, in Savannah, Georgia, she organized the very first Girl Scout troop, and every year since, we've honored her vision and legacy, building girls of courage, confidence, and character who make the world a better place. The organization is 2.5 million strong—more than 1.7 million girls and 750,000 adults who believe in the power of every G.I.R.L. (Go-getter, Innovator, Risk-taker, Leader) ™ to change the world.

Girl Scouts cookies

When you buy Girl Scout Cookies, you feed joy and make a difference with every box. The scout may help the local hospital or animal shelter across town, climb her first mountain or start a nonprofit. It's up to her and her troop. Proceeds stay local to power life-changing programs, experiences, and learning all year long! When girls participate in the Girl Scout Cookie Program, they also develop important life skills—goal setting, decision making, money management, people skills, and business ethics—that will set girls up for success beyond anything they can imagine.

That's why buying the delicious cookies you crave from your local Girl Scout Cookie business pro is important.

Meet the Cookies
- *Thin Mints® - Crisp wafers covered in chocolaty coating. Made with natural oil of peppermint*
- *Samoas® - Crisp cookies, coated in caramel, sprinkled with toasted coconut, and striped with dark chocolaty coating*

- *Tagalongs® - Crispy cookies layered with peanut butter and covered with a chocolaty coating*
- *Trefoils® - Delicate-tasting shortbread that is delightfully simple and satisfying*
- *Girl Scout S'mores® - Crunchy graham sandwich cookies with creamy chocolate and marshmallowy filling*

Most people have a special place in their heart for Girl Scout Cookies. Not only because they're spectacularly delicious but also because the Girl Scout Cookie Program helps girls fulfill their dreams, follow their passions, and change the world!

The program, which is the largest girl-led entrepreneurial program for girls in the world, helps Girl Scouts fund unique adventures for themselves and their troops all year long. It also allows them to give back to the causes they're most passionate about—it's the Girl Scout way!

1.2   Problem

Every year the girls have gone door-to-door to realize their individual and troop goals and contribute with confidence toward their community and troop. As you can gather the sale of girl scouts' cookies are important business and something that each girl scout looks forward to every year.

With COVID-19, door-to-door sales have become next-to-impossible.

On request from my very own fifth grader, a proud girl scout trooper, is a request to find the next possible solution.

Cookie booth sales at the local neighborhood would be an excellent solution to help my daughter's troop and the other troops in town to achieve their goals.

The problem is to find the top 10 prospective locations to set up cookie booths in our town, Grasmere, Staten Island, NY and one or two neighboring towns within a 3-mile radius to maximize sales. And, to make the Girl Scouts troop a happier lot in the process.

1.3   Interest

This analysis will be of great interest to my daughter's Girl Scouts troop and other troops in town to realize their troop's cookie goals. This information may turn out to be helpful while planning any other fund-raising initiatives which might require reaching out to a wider community.

2 Data Understanding

2.1 Data Sources

The data sources used for analyzing and gathering insights are:

- Access to geospatial coordinates (latitude and longitude) will be required. Data is available as a JSON file.

- Neighborhoods to target and compare will be fixed to my town, Grasmere, Staten Island, NY, and the neighboring towns within a 3-mile radius.

- To compare prospective locations in town and neighboring towns, access to location data is imperative. Location data is data describing places and venues, such as their geographical location, their category, working hours, full address, and so on, such that for a given location given in the form of its geographical coordinates (or latitude and longitude values), one is able to determine what types of venues exist within a defined radius from that location. There are many location data providers like Foursquare, Google Places, and Yelp. Providers like Foursquare gives a very wide geospatial coverage free of cost and has rate limits of 95,000 regular API calls and 500 premium calls in a day. They are known to provide comprehensive data that is both accurate and updated. **Foursquare will the preferred choice for extracting location data.**

- To target venues to achieve maximize sales, knowing the venues that are trending at the time of setting the booths will be crucial. Trending venues means places with the highest foot traffic. The Foursquare API provides this data and the venues with the highest foot traffic is fetched live. The only drawback is since it is real-time the data will depend on when the code is run, and most times don't give results. To take care of this issue, the following method can be used.

- As COVID-19 stay-at-home orders begin to lift, the venues people will start to visit more often, and their order of preference will be worth taking into account. Foursquare Recovery Index shows the places consumers go to first as the restrictions are lifted in near real-time. The venues listed by the Foursquare Recovery Index will be considered.

2.2 Data Cleaning and Data Preparation

The geospatial coordinates of New York (NY) was downloaded and loaded.

The dataframe has 5 boroughs and 306 neighborhoods.

Out[5]:

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

The above data with borough, neighborhood and respective latitude and longitude was extracted from the geospatial JSON file.

To restrict this analysis to my town of Grasmere, Staten Island, NY and the neighboring towns, the data was filtered on the neighborhoods of interest.

The neighborhoods of interest are 8 neighborhoods.

Out[19]:

	Borough	Neighborhood	Latitude	Longitude
2	Staten Island	Stapleton	40.626928	-74.077902
7	Staten Island	South Beach	40.580247	-74.079553
24	Staten Island	Park Hill	40.609190	-74.080157
28	Staten Island	Arrochar	40.596313	-74.067124
29	Staten Island	Grasmere	40.598268	-74.076674
30	Staten Island	Old Town	40.596329	-74.087511
48	Staten Island	Shore Acres	40.609719	-74.066678
50	Staten Island	Concord	40.604473	-74.084024

Using the Foursquare API, I was able to extract the location data for each of the neighborhoods of interest. There were 3150 venues in Staten Island Borough and 400 venues in the neighborhoods of interest.

A sample of the data is shown below.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Stapleton	40.626928	-74.077902	Lakruwana	40.625654	-74.075174	Sri Lankan Restaurant
1	Stapleton	40.626928	-74.077902	Defonte's Sandwich Shop	40.627202	-74.077257	Sandwich Place
2	Stapleton	40.626928	-74.077902	The Hop Shoppe	40.629034	-74.079758	Beer Bar
3	Stapleton	40.626928	-74.077902	Every Thing Goes Cafe and Bookstore	40.636920	-74.076661	Café
4	Stapleton	40.626928	-74.077902	Dunkin'	40.631129	-74.076879	Donut Shop

There were 65 venues with no venue category which is crucial for our analysis. Since the number of missing data was not significant, the rows of missing data were deleted. The number of venues in the neighborhoods of interest were 335 and unique venue categories were 135.

A sample of the categories are shown below,

```
Unique categories in the neighborhoods of interest:
['Coffee Shop' 'Department Store' 'Building' 'Discount Store'
 'Sandwich Place' 'Pizza Place' 'Asian Restaurant' 'Medical Center'
 'Miscellaneous Shop' 'Park' 'Salon / Barbershop' 'Bus Stop' 'Church'
 'Bus Line' 'Government Building' 'Bank' 'Mexican Restaurant'
 'Cocktail Bar' 'Diner' 'Library' 'Rock Club' 'Train Station'
 'African Restaurant' 'Seafood Restaurant' 'Recording Studio'
 'Mobile Phone Shop' 'Restaurant' 'Sri Lankan Restaurant'
 'Automotive Shop' 'New American Restaurant' 'Tattoo Parlor' 'Lounge'
 'Bar' 'Deli / Bodega' 'Food' 'Hospital' 'Athletics & Sports'
 "Dentist's Office" 'Art Gallery' 'Beach' 'Theme Park' 'Stables'
 'BBQ Joint' 'Pier' 'Gym' 'Skate Park' 'Arts & Crafts Store'
 'Cosmetics Shop' 'Cafeteria' 'Soccer Field'
 'Residential Building (Apartment / Condo)' 'School' "Doctor's Office"
 'College Communications Building' 'Parking' 'Field' 'Playground'
 'Food Truck' 'Salad Place' 'Toll Plaza' 'Office' 'Bridge'
```

To determine the neighborhoods which are similar to mine, more out of curiosity, the neighborhoods were clustered and segmented. The towns to target were then limited to the towns which belong to the same cluster as my town.

Foursquare Recovery Index data, available as an excel file, was loaded to view the top 20 venue categories for each day. This shows the foot traffic data. The ranking data was available for the last 14 days.

	Date	Category_1	Category_2	Category_3	Category_4	Category_5	Category_6	Category_7	Category_8	Category_9	...	Category_11	Category_12	Category_13	Category_14	Category_15
0	2/8/21	Beaches	Outdoors	Bank	Skiing	Metro	Medical	Auto Shops	Bars	Transit	...	School	Liquor, Beer, and Wine Stores	Warehouse Store	Sports	Convenience Stores
1	2/9/21	Beaches	Outdoors	Bank	Skiing	Metro	Medical	Auto Shops	Bars	Transit	...	School	Liquor, Beer, and Wine Stores	Warehouse Store	Sports	Convenience Stores
2	2/10/21	Beaches	Outdoors	Bank	Skiing	Metro	Medical	Auto Shops	Bars	Transit	...	School	Liquor, Beer, and Wine Stores	Warehouse Store	Sports	Convenience Stores
3	2/11/21	Beaches	Outdoors	Bank	Skiing	Metro	Medical	Auto Shops	Bars	Transit	...	School	Liquor, Beer, and Wine Stores	Warehouse Store	Sports	Convenience Stores
4	2/12/21	Airports	School	Bus	Pharmacies and Drug Stores	Transit	Metro	Apparel Retail	Office	Discount	...	Fitness	Big Box Store	Warehouse Store	Restaurant	Grocery

5 rows × 21 columns

Considering we are setting up cookie booths for girl scouts in March, some of the irrelevant venue categories will need to be removed. For example, beaches, skiing, medical, airports and bars.

Also, looking closely at the above data and comparing it to the venue categories in the neighborhoods of interest, there were some problems. For example, the category, Restaurant, is in the foot traffic data from Foursquare Recovery Index but the venue categories from neighborhoods, obtained from Foursquare API, is more drilled down to each type of restaurant, for example, 'Sri Lankan Restaurant', 'Tapas Restaurant' and 'Seafood Restaurant'. The data was fixed to reflect 'Restaurant' for every type of restaurant. The neighborhood had venue categories, 'Park', 'Playground', 'Sculpture Garden' and so on. All these were consolidated and categorized as 'Outdoors', the category available in foot traffic data. Similar data cleaning was required for each of the categories. The original category of the neighborhood venues was saved for future reference.

The features that would be of interest would be Neighborhood, Venue name, Venue category, Venue address and Venue location (latitude and longitude).

3 Exploratory Data Analysis

While searching for the top 10 locations where the cookie booths can be set up, the following data analysis was performed.

3.1 Count of venues by neighborhood

Filtering out neighborhoods of Staten Island borough, resulted in 336 venues and the following is the count by neighborhood.
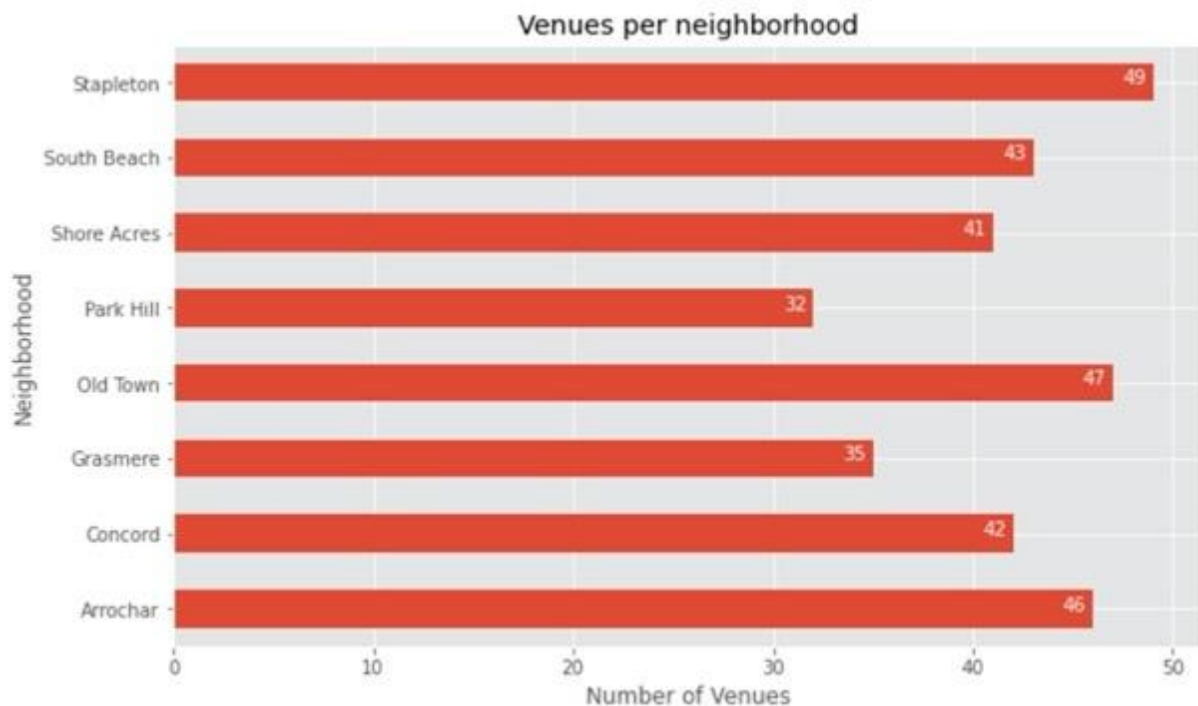
3.2   Top 5 venues by frequency in each neighborhood

Curious to learn what kind of venues occur most in each of the neighborhoods, I found the top 5 by frequency and have listed below a sample of what was found.

```
----Arrochar----
             venue  freq
0    Deli / Bodega  0.07
1         Bus Stop  0.07
2       Bagel Shop  0.04
3     Liquor Store  0.04
4      Pizza Place  0.04

----Concord----
                      venue  freq
0            Medical Center  0.05
1        Martial Arts School  0.05
2      Thrift / Vintage Store  0.05
3            Automotive Shop  0.05
4                     Office  0.05

----Grasmere----
              venue  freq
0    Doctor's Office  0.11
1          Bus Stop  0.09
2              Bank  0.09
3   Dentist's Office  0.06
4            Office  0.06

----Old Town----
                 venue  freq
0    Italian Restaurant  0.09
1        Medical Center  0.09
2       Auto Dealership  0.06
3               Bakery  0.06
4                 Bank  0.06
```

The top 5 venues by frequency are listed below by neighborhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Arrochar	Deli / Bodega	Bus Stop	Bagel Shop	Beach	Italian Restaurant
1	Concord	Peruvian Restaurant	Martial Arts School	Office	Automotive Shop	Doctor's Office
2	Grasmere	Doctor's Office	Bus Stop	Bank	Office	Dentist's Office
3	Old Town	Medical Center	Italian Restaurant	Auto Dealership	Bank	Bakery
4	Park Hill	Doctor's Office	Other Great Outdoors	Office	Thrift / Vintage Store	Medical Center
5	Shore Acres	Deli / Bodega	Dentist's Office	Chinese Restaurant	Building	Office
6	South Beach	Doctor's Office	Hospital	Pier	Beach	Dentist's Office
7	Stapleton	Salon / Barbershop	Bank	Pizza Place	Bus Line	Sandwich Place

3.3 Foursquare Recovery Index categories

Taking the new pattern of how people are returning to shop as the COVID-19 stay-at-home orders begin to lift, Foursquare Recovery Index shows the venue categories that customers would visit by rank. The data available was the top 20 venues visited in New York over 14 days.

There were 34 unique categories from the index.

```
Number of unique categories in the Recovery Index: 34

Unique categories in the Recovery Index:
['Airports' 'Apparel Retail' 'Auto DealerShips' 'Auto Shops' 'Bank' 'Bars'
 'Beaches' 'Big Box Store' 'Bus' 'Convenience Stores' 'Discount'
 'Fast Food' 'Fitness' 'Gas' 'Government' 'Grocery' 'Hardware'
 'Home Stores and Services' 'Hotels' 'Liquor, Beer, and Wine Stores'
 'Medical' 'Metro' 'Office' 'Outdoors' 'Pet Stores'
 'Pharmacies and Drug Stores' 'Restaurant' 'Salons and Services' 'School'
 'Skiing' 'Spiritual Centers' 'Sports' 'Transit' 'Warehouse Store']
```

3.4 Top 20 categories as by per the Foursquare Recovery Index

Based on the foot traffic data of 14 days, the mean ranking was calculated, and categories ranked. Here are the top 10.

Venue Category	Ranking
Airports	7.705882
Skiing	9.941176
Bars	10.000000
Apparel Retail	10.529412
Beaches	11.882353
Discount	12.294118
Hotels	12.882353
Bank	13.058824
Metro	13.529412
Medical	13.588235
School	13.588235

3.5 Clean-up of venue categories to match the generic Foursquare Recovery Index categories

The categories from exploring the neighborhoods of interest were consolidated into the categories found in the Recovery Index. In the picture below, the 'Venue Category Orig' lists the original category of the venue and the 'Venue Category' shows the Recovery Index category.

|	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Category	Venue Address	Venue Latitude	Venue Longitude	Venue Category Orig
15	Stapleton	40.626928	-74.077902	Tappen Park	Outdoors	Bay St (btwn Water & Canal St)	40.626872	-74.076857	Park
64	Arrochar	40.596313	-74.067124	Swing Life Away Swings	Outdoors	71 Sand Ln	40.597928	-74.070138	Playground
66	Arrochar	40.596313	-74.067124	Gateway National Recreation Area	Outdoors	Staten Island, NY 10305	40.592428	-74.065373	Outdoors & Recreation
89	Arrochar	40.596313	-74.067124	The monument	Outdoors	115 lily pond ave	40.600412	-74.064524	Sculpture Garden
180	Shore Acres	40.609719	-74.066578	Poolside At The Dawson's	Outdoors	New York	40.610948	-74.066410	Art Gallery

4 Modeling

To set up the cookie booths, I wanted to stick to neighborhoods that I was comfortable with. More so out of curiosity, I wanted to explore and find out the neighborhoods which were similar to mine. The below maps show the neighborhoods of Staten Island that I was interested in.
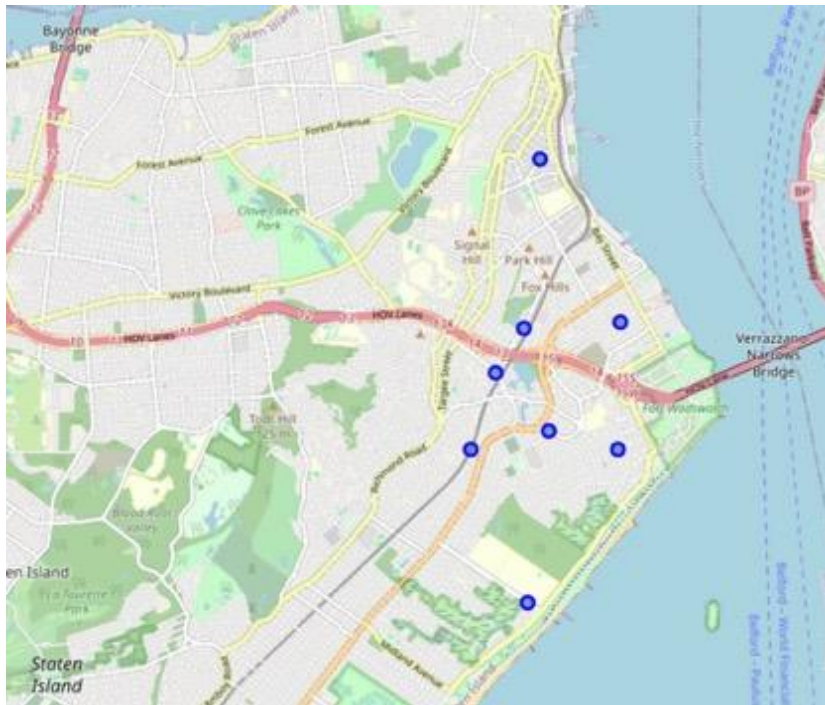
As there were no labeled data and it meant putting the neighborhoods into groups based on similarity of the venues in each neighborhood, I performed k-means clustering of the 8 neighborhoods of Staten Island. The silhouette score was used as a metric to decide on *k* in K-means clustering.

The modeling resulted in the following clusters. The cluster group that my town belongs to was taken as the final target towns.
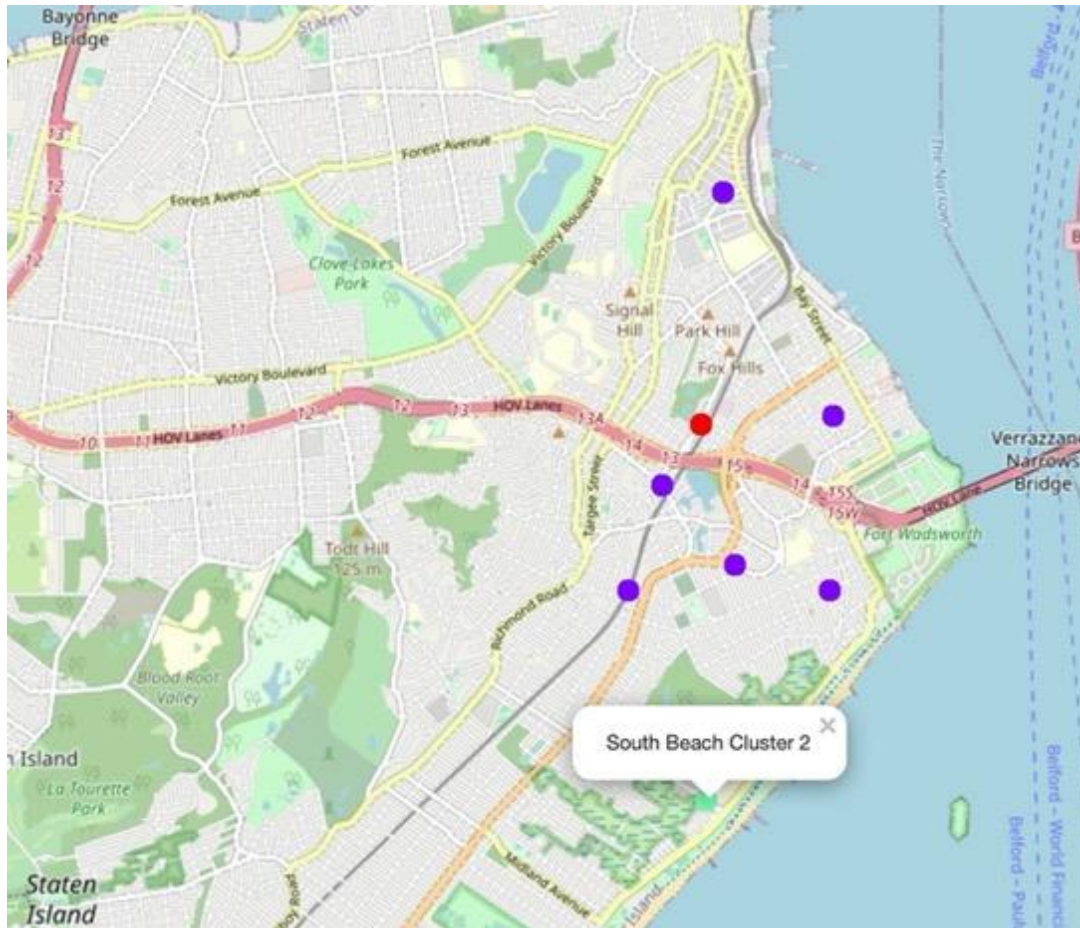
Once the target towns were set, the top locations needed to be determined.

5   Results

I used the Recovery Index ranking to determine the top 10 locations in the resulting target towns.

Venue	Venue Category	Venue Address	Venue Latitude	Venue Longitude	Ranking
The Coupe	Bars	379 Van Duzer St (Beach Street)	40.628754	-74.079730	10.000000
The Phunky Elephant Gastropub	Bars	1271 Bay St (Scarboro Ave)	40.612489	-74.064894	10.000000
Island Chatau	Bars	900 W Fingerboard Rd	40.601069	-74.085779	10.000000
Danny Blaines	Bars	1384 Bay St	40.609820	-74.063232	10.000000
Apple Department Store	Apparel Retail	112 Canal St (Wright Street)	40.626328	-74.077391	10.529412
Dollar General	Apparel Retail	135 Canal St (Water Street)	40.626935	-74.078271	10.529412
Family Dollar	Discount	135 Canal St	40.627018	-74.078461	12.294118
Family Dollar	Discount	40 Prospect St (Prospect Street)	40.628318	-74.075875	12.294118

As seen above, the top locations include bars, and retail. The only reason, bars and other irrelevant locations were kept in the dataset were for my own curiosity to discover what are the locations people are choosing to go once the stay-at-home orders are lifted.

6 Conclusion

The more irrelevant categories were removed and the final top 10 locations for setting up cookie booths were mapped so as to maximize the sales.
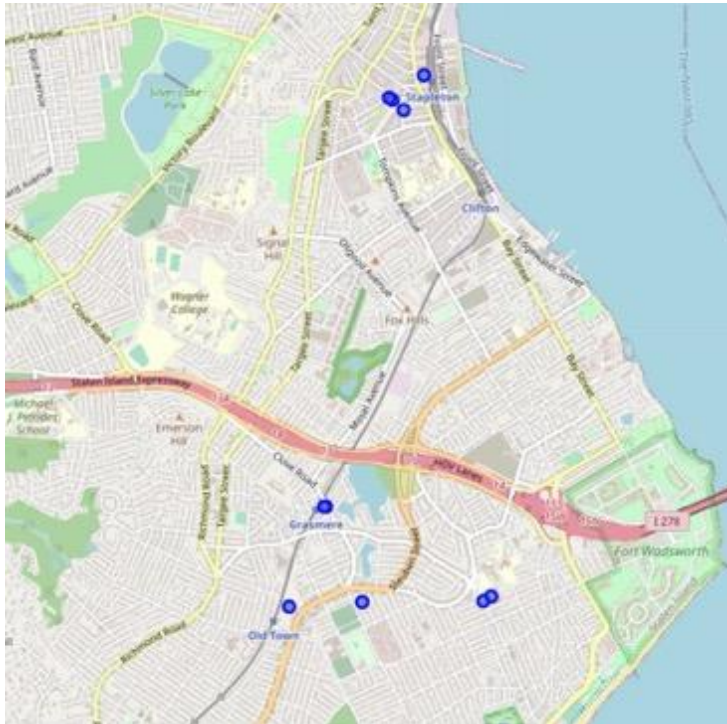
Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Category	Venue Address	Venue Latitude	Venue Longitude	Venue Category Orig	
Stapleton	40.626928	-74.077902	Apple Department Store	Apparel Retail	112 Canal St (Wright Street)	40.626328	-74.077391	Department Store	1
Stapleton	40.626928	-74.077902	Dollar General	Apparel Retail	135 Canal St (Water Street)	40.626935	-74.078271	Department Store	1
Stapleton	40.626928	-74.077902	Family Dollar	Discount	135 Canal St	40.627018	-74.078461	Discount Store	1
Stapleton	40.626928	-74.077902	Family Dollar	Discount	40 Prospect St (Prospect Street)	40.628318	-74.075875	Discount Store	1
Concord	40.604473	-74.084024	2098 clove rd.	School	2098 Clove Rd, Concord, NY 10305	40.603340	-74.083298	Martial Arts School	1
Concord	40.604473	-74.084024	Longview Academy Extreme MMA	School	2098 Clove Rd, Concord, NY 10305	40.603362	-74.083442	Martial Arts School	1
Old Town	40.596329	-74.087511	Academy Of St. Dorothy	School	1305 Hylan Blvd	40.597556	-74.086168	School	1
Grasmere	40.598268	-74.076674	Elite Daycare	School	Staten Island, NY 10305	40.597788	-74.080581	Nursery School	1
Grasmere	40.598268	-74.076674	Saint Joseph Hill	School	850 Hylan Blvd, Staten Island, NY 10305	40.598146	-74.070645	University	1
Arrochar	40.596313	-74.067124	P.S. 39 Francis J. Murphy Jr.	School	71 Sand Ln (btwn Major & MacFarland Ave)	40.597828	-74.071243	Elementary School	1

7 Future Directions

The Foursquare Recovery Index was not accessible real time. If real-time data were available, the prediction of these locations could be more accurate. Also, weather data could also be used in combination with real-time foot traffic data for better results.