1. Introduction

1.1 Background

If you ask anybody what city has the best bagels, New York City is always going to be at or near the top of the list. Canadians may claim it is Montreal—nonetheless, the status of New York City as top tier in the bagel world is indisputable. From classic bagel joints like Absolute Bagels and Murray's Bagels, to more upstarts like Tompkins Square Bagels, the industry is rich with shops of all kinds.

1.2 Problem

What goes on a bagel, you may ask? Obviously, cream cheese. Many bagel shops have cream cheese of course, but they focus on bagels. Cream cheese is an afterthought. There are no specialized cream cheese stores. Not until now. I propose opening a shop devoted to the art of cream cheese, specifically for bagel applications. And sure, New York City is bustling with bagel shops. However, in order to open my shop I need to find the neighborhood with the highest concentration of bagel shops in order to maximize my profits and please my investors. Though New York City has 5 boroughs, my shop must be exclusive to Brooklyn as I am planning on running my store in a way that will primarily appeal to the young hip folk that have flocked to the East shore of the East River.

1.3 Interest

This project should be of high interest to bagel lovers worldwide. This cream cheese revolution will benefit all that have ever seen look at a bagel. It is of particular interest to those New Yorkers who love grabbing their daily bagel but are disappointed in the lack of diversity in the cream cheese department.

2. Data acquisition

2.1 Data sources

New York City neighborhood data sorted by borough and containing latitude and longitude was obtained from the IBM Skills Lab data file found here. Foursquare API was utilized in order to find all of the names and categories of points of interest in Brooklyn specifically.

2.2 Data cleaning

The New York City data json file was put into a pandas dataframe containing columns for borough, neighborhood, latitude and longitude. These variables were all defined and the data was sorted so the neighborhoods in each borough were grouped together. The data from the

Foursquare API for Brooklyn points of interest was put into a table containing columns for the name of the point of interest, the category it belongs to, as well as latitude and longitude. The tables were then combined so each point of interest is also labeled with its respective neighborhood. The categories were grouped in a way so that for each neighborhood, a table containing all of the most common venues was obtained.

2.3 Feature selection

The feature selected was the frequency of a particular category of points of interest in a particular neighborhood. The ultimate variable to obtain is which neighborhood(s) of Brooklyn have the highest frequency of bagel shops for a suitable business location.

3. Exploratory Data Analysis

Each neighborhood in Brooklyn was sorted and mapped using Folium.
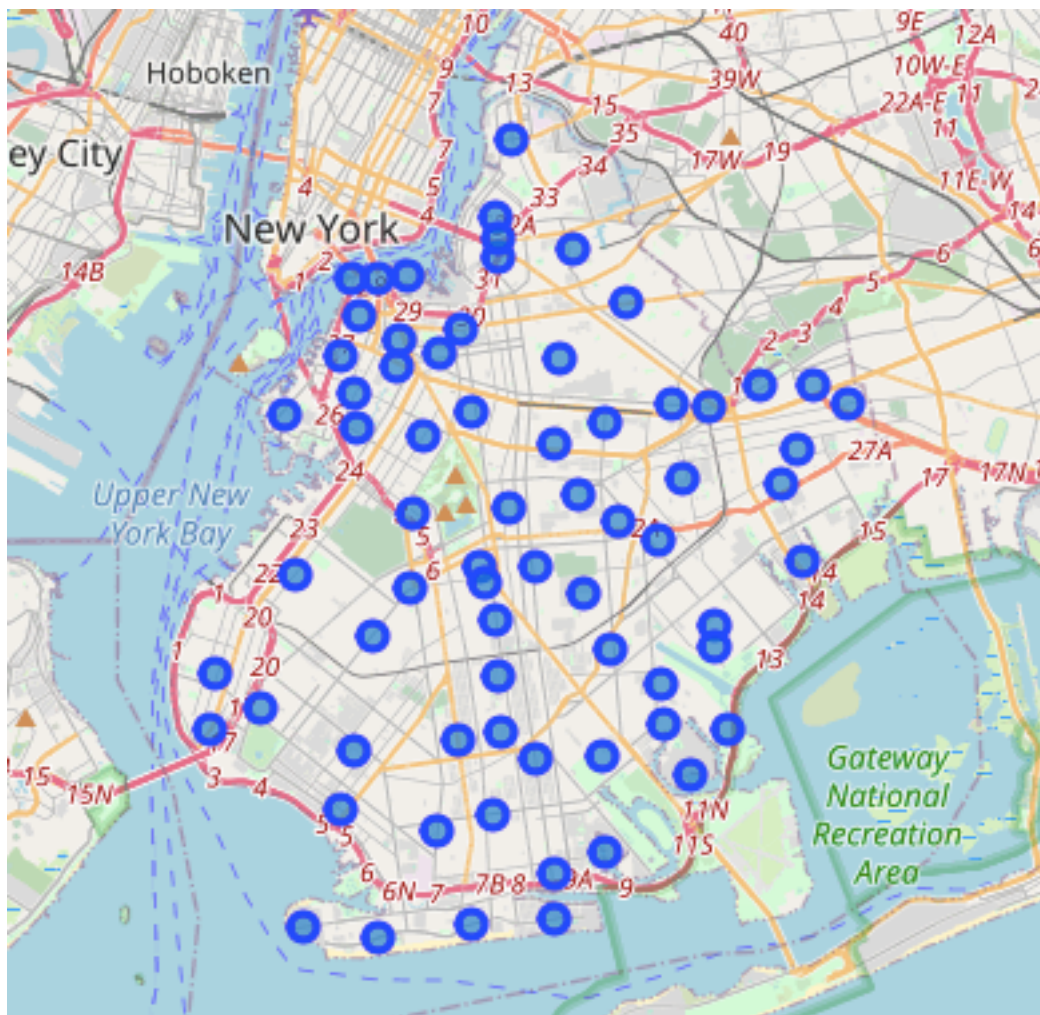
Then, as mentioned, the neighborhood data and the Foursquare API data were combined into a table.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bay Ridge	40.625801	-74.030621	Pilo Arts Day Spa and Salon	40.624748	-74.030591	Spa
1	Bay Ridge	40.625801	-74.030621	Bagel Boy	40.627896	-74.029335	Bagel Shop
2	Bay Ridge	40.625801	-74.030621	Leo's Casa Calamari	40.624200	-74.030931	Pizza Place
3	Bay Ridge	40.625801	-74.030621	Pegasus Cafe	40.623168	-74.031186	Breakfast Spot
4	Bay Ridge	40.625801	-74.030621	Georgian Dream Cafe and Bakery	40.625586	-74.030196	Caucasian Restaurant

All of the data was normalized, processed and quantified to result in a frequency value for a particular category of venue in each neighborhood.

```
----Bath Beach----
                    venue  freq
0       Chinese Restaurant  0.13
1             Pizza Place  0.06
2     Cantonese Restaurant  0.04
3    Fast Food Restaurant  0.04
4         Bubble Tea Shop  0.04


----Bay Ridge----
                    venue  freq
0                     Spa  0.06
1      Italian Restaurant  0.05
2             Pizza Place  0.05
3        Greek Restaurant  0.04
4     American Restaurant  0.04


----Bedford Stuyvesant----
                    venue  freq
0             Coffee Shop  0.12
1                     Bar  0.09
2                    Café  0.06
3             Pizza Place  0.06
4    Japanese Restaurant  0.03
```

These frequencies were then ranked and inserted into the previous table. The data was inserted in a way that the 1st through 10th most common venues were displayed in the table for each neighborhood.

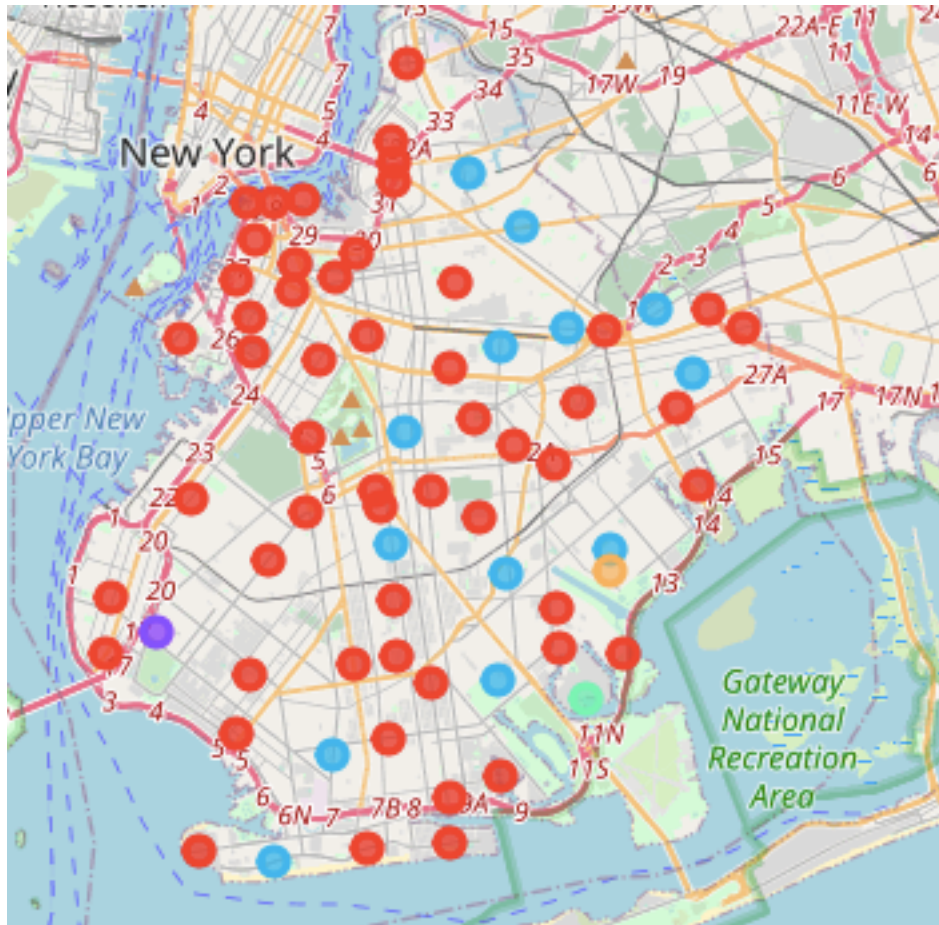|	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
0	Bath Beach	Chinese Restaurant	Pizza Place	Pharmacy	Bubble Tea Shop	Italian Restaurant	Donut Shop	Gas Station	Fast Food Restaurant	Cantonese Restaurant
1	Bay Ridge	Spa	Pizza Place	Italian Restaurant	American Restaurant	Bar	Bagel Shop	Greek Restaurant	Sushi Restaurant	Hookah Bar
2	Bedford Stuyvesant	Coffee Shop	Bar	Café	Pizza Place	Deli / Bodega	Tiki Bar	Fried Chicken Joint	New American Restaurant	Bus Stop
3	Bensonhurst	Chinese Restaurant	Pizza Place	Italian Restaurant	Ice Cream Shop	Donut Shop	Sushi Restaurant	Sporting Goods Shop	Shabu-Shabu Restaurant	Noodle House
4	Bergen Beach	Harbor / Marina	Park	Baseball Field	Playground	Athletics & Sports	Yemeni Restaurant	Event Service	Event Space	Factory

4 Predictive Modeling

4.1 KMeans Clustering

The data was clustered using KMeans clustering. The data was put into 5 clusters based on which neighborhoods had similar most common venues.

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Brooklyn	Bay Ridge	40.625801	-74.030621	0	Spa	Pizza Place	Italian Restaurant	American Restaurant	Bar	Bagel Shop	Greek Restaurant	Sushi Restaurant	Hookah Bar	Cosmetics Shop
1	Brooklyn	Bensonhurst	40.611009	-73.995180	0	Chinese Restaurant	Pizza Place	Italian Restaurant	Ice Cream Shop	Donut Shop	Sushi Restaurant	Sporting Goods Shop	Shabu-Shabu Restaurant	Noodle House	Smoke Shop
2	Brooklyn	Sunset Park	40.645103	-74.010316	0	Pizza Place	Latin American Restaurant	Bank	Bakery	Mexican Restaurant	Deli / Bodega	Fried Chicken Joint	Pharmacy	Mobile Phone Shop	Gym
3	Brooklyn	Greenpoint	40.730201	-73.954241	0	Bar	Pizza Place	Coffee Shop	Cocktail Bar	Yoga Studio	Deli / Bodega	Record Shop	Mexican Restaurant	Grocery Store	Furniture / Home Store
4	Brooklyn	Gravesend	40.595260	-73.973471	2	Bakery	Pizza Place	Lounge	Chinese Restaurant	Deli / Bodega	Bar	Italian Restaurant	Record Shop	Pharmacy	Martial Arts School

These clusters were mapped using Folium to visualize.

Each cluster was defined. Cluster 2 contained only one neighborhood, the only of which had a bagel shop as the most common venue.

```
[ ] brooklyn_merged.loc[brooklyn_merged['Cluster Labels'] == 1, brooklyn_merged.columns[[1] + list(range(5, brooklyn_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
35	Dyker Heights	Bagel Shop	Plaza	Golf Course	Burger Joint	Dance Studio	Flea Market	Fish Market	Fish & Chips Shop	Filipino Restaurant	Field

5. Conclusions

Dyker Heights is the only suitable neighborhood for my bagel shop. It is the only venue that has a bagel shop as its most common venue. I was able to identify this by utilizing KMeans clustering in hopes of identifying a cluster of neighborhood that would be suitable to open my cream cheese shop.

6. Future Directions

The only thing left to do is open the shop.