Introduction

Nowadays, the number of people following special or restricted diet has been significantly increased. It is crucial for those people who want to visit or move to new cities to have an overview of the availability of restaurants serving their special diet dishes. As New York City is considered to be one of the most visited cities in USA, this project will compare the following diets' restaurants in New York City; specifically Bronx and Manhattan boroughs:

A. Vegetarian/Vegan
B. Halal
C. Kosher
D. Gluten Free

A travel agency in New York City will use the result of this project to help in creating travel plans for travelers following different diets. This should give them a great idea on which neighborhood to choose for accommodation or sightseeing.

Data

Foursquare API will be used to acquire data about venues and their categories for each diet type. Also, a dataset for NYC neighborhoods geometric coordinates will be used. The following datasets will be used in the project:

1. New York City Neighborhoods dataset. It includes NYC Borough, Neighborhood, Latitude, and Longitude https://geo.nyu.edu/catalog/nyu_2451_34572
   In this solution, we will consider only Bronx and Manhattan boroughs
2. Foursquare API search function will be utilized to populate venues for each neighborhood. This is to get the number of restaurants for each diet in every neighborhood. I will get category IDs to pass to explore function from https://developer.foursquare.com/docs/build-with-foursquare/categories/

Methodology

First, we will get New York data from https://cocl.us/new_york_dataset.
After analysis the data will look like the following:

[13]:
```
neighborhoods.head()
```

[13]:

	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

Since we are considering Bronx and Manhattan only, we created two data frames for each borough.

[13]:
```
Bronx_neigh.head()
```

[13]:

	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

[14]:
```
Manhattan_neigh.head()
```

[14]:

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

Next, we will Use geopy library to get the latitude and longitude values of New York City and its boroughs:

[15]:
```
address = 'New York City, NY'

geolocator = Nominatim(user_agent="foursquare_agent")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of New York City are {}, {}.'.format(latitude, longitude))
```

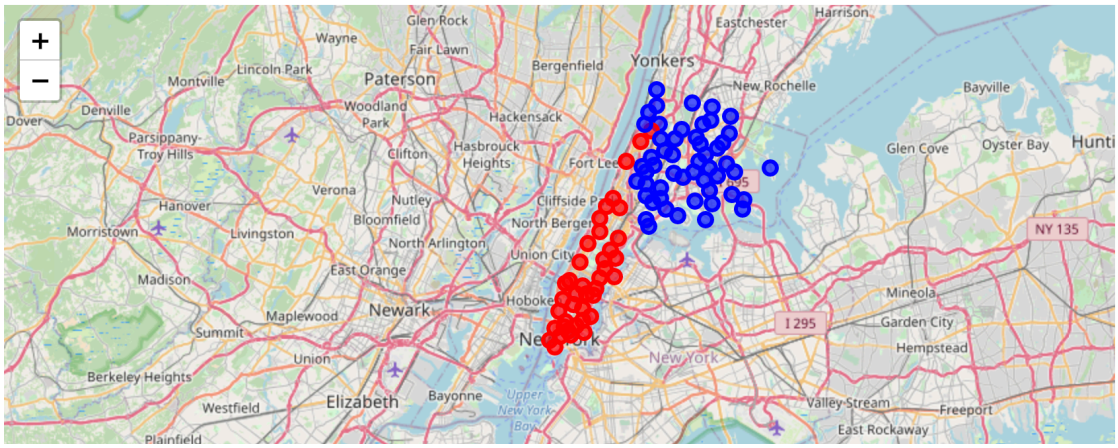
```
The geograpical coordinate of New York City are 40.7127281, -74.0060152.
```

[16]:
```
#Bronx
Bronx_address = 'Bronx, NY'
Bronx_location = geolocator.geocode(Bronx_address)
Bronx_latitude = Bronx_location.latitude
Bronx_longitude = Bronx_location.longitude
print('The geograpical coordinate of Bronx are {}, {}.'.format(Bronx_latitude, Bronx_longitude))

#Manhattan
Manhattan_address = 'Manhattan, NY'
Manhattan_location = geolocator.geocode(Manhattan_address)
Manhattan_latitude = Manhattan_location.latitude
Manhattan_longitude = Manhattan_location.longitude
print('The geograpical coordinate of Manhattan are {}, {}.'.format(Manhattan_latitude, Manhattan_longitude))
```

```
The geograpical coordinate of Bronx are 40.8466508, -73.8785937.
The geograpical coordinate of Manhattan are 40.7896239, -73.9598939.
```

After reading New York data into Pandas data frame, we will use Folium to visualize neighborhoods in both Bronx (in blue) and Manhattan (in red).



Then, we will use Foursquare API to search for food venues within 200 meters of each neighborhood.

3. Define Foursquare Credentials and Version

```
CLIENT_ID = 'YIPE0XDZX0RNL3XZYFBWVP2OCWNCBLMHXXZJWPNXRK5URKTN' # your Foursquare ID
CLIENT_SECRET = 'G4H5YBLNT531U4L2AYAOVD1CAGHUIPHLNASL3OXPKUPZJBQ1' # your Foursquare Secret
VERSION = '20180604'
print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
LIMIT = 50 # limit of number of venues returned by Foursquare API
radius = 200 # define radius of 200 meters
search_query = ''
```

```
Your credentails:
CLIENT_ID: YIPE0XDZX0RNL3XZYFBWVP2OCWNCBLMHXXZJWPNXRK5URKTN
CLIENT_SECRET:G4H5YBLNT531U4L2AYAOVD1CAGHUIPHLNASL3OXPKUPZJBQ1
```

The following Category Id's are used to retrieve the data:

- Vegan/Vegetarian 4bf58dd8d48988d1d3941735
- Halal 52e81612bcbc57f1066b79ff
- Kosher 52e81612bcbc57f1066b79fc
- Gluten Free 4c2cd86ed066bed06c3c5209

```
# create the API request URL
url = 'https://api.foursquare.com/v2/venues/search?&client_id={}&client_secret={}&v={}&ll={},{}&query={}&radius={}&limit={}&categoryId={}'.format(
        CLIENT_ID,
        CLIENT_SECRET,
        VERSION,
        latitude,
        longitude,
        search_query,
        radius,
        LIMIT,
        category)
```

With our extracted information of nearby food venues, we can merge this into new data frame containing our neighborhoods information.

```
print(Bronx_venues.shape)
Bronx_venues.head(10)
```

```
(13, 9)
```

	Borough	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Category	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bronx	Kingsbridge	40.881687	-73.902818	Vegetarian/Vegan	Kingsbridge-Riverdale Farmers' Market	40.879973	-73.907295	Vegetarian / Vegan Restaurant
1	Bronx	Fordham	40.860997	-73.896427	Vegetarian/Vegan	Veggie Mart	40.861740	-73.890566	Vegetarian / Vegan Restaurant
2	Bronx	Fordham	40.860997	-73.896427	Vegetarian/Vegan	C. Kim's Fruit & Vegetable	40.864635	-73.891895	Vegetarian / Vegan Restaurant
3	Bronx	Country Club	40.844246	-73.824099	Vegetarian/Vegan	Vegetarian Joint	40.842828	-73.825764	Vegetarian / Vegan Restaurant
4	Bronx	Parkchester	40.837938	-73.856003	Vegetarian/Vegan	Tandoori Place	40.836010	-73.853482	Indian Restaurant
5	Bronx	Belmont	40.857277	-73.888452	Vegetarian/Vegan	Veggie Mart	40.861740	-73.890566	Vegetarian / Vegan Restaurant

```
print(Manhattan_venues.shape)
Manhattan_venues.head(10)
```

```
(430, 9)
```

	Borough	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Category	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Manhattan	Marble Hill	40.876551	-73.910660	Vegetarian/Vegan	Kingsbridge-Riverdale Farmers' Market	40.879973	-73.907295	Vegetarian / Vegan Restaurant
1	Manhattan	Chinatown	40.715618	-73.994279	Vegetarian/Vegan	Jisu Vegetarian	40.716050	-73.995348	Vegetarian / Vegan Restaurant
2	Manhattan	Chinatown	40.715618	-73.994279	Vegetarian/Vegan	Orchard Grocer	40.717847	-73.990358	Vegetarian / Vegan Restaurant
3	Manhattan	Chinatown	40.715618	-73.994279	Vegetarian/Vegan	Petisco Vegano	40.714040	-73.988815	Vegetarian / Vegan Restaurant
4	Manhattan	Chinatown	40.715618	-73.994279	Vegetarian/Vegan	Dirt Candy	40.717890	-73.991015	Vegetarian / Vegan Restaurant

Now, we have the following breakdown for venue categories

Let's check how many venues were returned for each category

```
Bronx_venues.groupby('Category')['Venue'].count().sort_values(ascending=False)
```

```
Category
Vegetarian/Vegan    8
Halal               2
Gluten Free         2
Kosher              1
Name: Venue, dtype: int64
```

```
Manhattan_venues.groupby('Category')['Venue'].count().sort_values(ascending=False)
```

```
Category
Vegetarian/Vegan    297
Halal                51
Kosher               43
Gluten Free          39
Name: Venue, dtype: int64
```

And the breakdown by neighborhoods

Let's check how many venues were returned for each Neighborhoodh

```
Bronx_venues.groupby('Neighborhood')['Venue'].count().sort_values(ascending=False)
```

```
Neighborhood
Parkchester        2
Fordham            2
Van Nest           1
Unionport          1
Olinville          1
North Riverdale    1
Kingsbridge        1
Country Club       1
Bronxdale          1
Belmont            1
Bedford Park       1
Name: Venue, dtype: int64
```

```
Manhattan_venues.groupby('Neighborhood')['Venue'].count().sort_values(ascending=False)
```

```
Neighborhood
Greenwich Village     36
Noho                  35
Flatiron              31
Soho                  28
East Village          27
Little Italy          26
Midtown               25
Murray Hill           21
Midtown South         20
Chinatown             17
Gramercy              16
West Village          15
Financial District    14
Turtle Bay            13
Upper West Side       12
Sutton Place          10
Chelsea               10
Civic Center          10
Clinton                9
Tudor City             8
Upper East Side        8
Carnegie Hill          6
Lenox Hill             5
East Harlem            4
Central Harlem         4
Tribeca                4
Hudson Yards           3
Yorkville              3
Lower East Side        2
Hamilton Heights       2
Manhattan Valley       2
Marble Hill            1
Roosevelt Island       1
Washington Heights     1
Battery Park City      1
Name: Venue, dtype: int64
```

Next we are using one hot encoding and the sum of the frequency for each one of the four categories.

```
print(Bronx_restaurant_grouped.shape)
Bronx_restaurant_grouped
```

```
(11, 5)
```

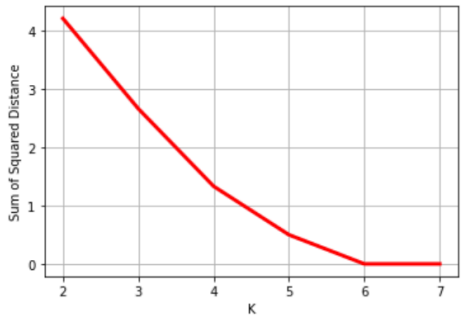
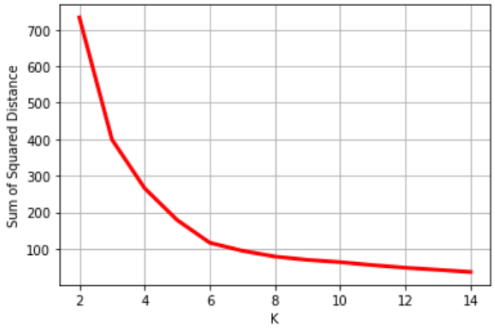
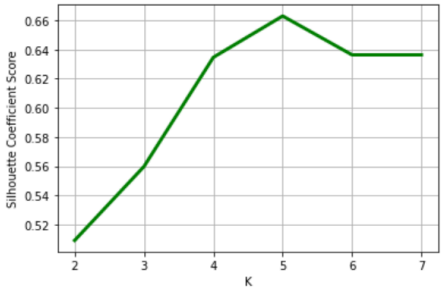
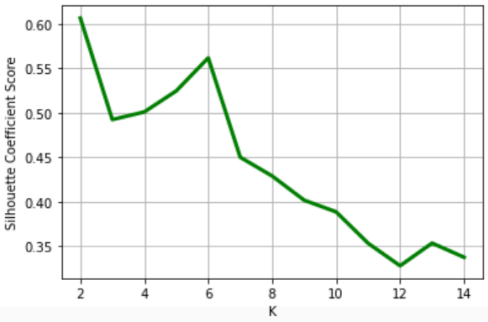
	Neighborhood	Gluten Free	Halal	Kosher	Vegetarian/Vegan
0	Bedford Park	0	1	0	0
1	Belmont	0	0	0	1
2	Bronxdale	0	0	0	1
3	Country Club	0	0	0	1
4	Fordham	0	0	0	2
5	Kingsbridge	0	0	0	1
6	North Riverdale	0	0	1	0
7	Olinville	0	0	0	1
8	Parkchester	1	0	0	1
9	Unionport	0	1	0	0
10	Van Nest	1	0	0	0

```
print(Manhattan_restaurant_grouped.shape)
Manhattan_restaurant_grouped
```

```
(35, 5)
```

	Neighborhood	Gluten Free	Halal	Kosher	Vegetarian/Vegan
0	Battery Park City	0	0	0	1
1	Carnegie Hill	1	0	1	4
2	Central Harlem	0	3	0	1
3	Chelsea	1	2	0	7
4	Chinatown	0	1	0	16
5	Civic Center	1	2	1	6
6	Clinton	2	1	0	6
7	East Harlem	0	2	1	1
8	East Village	2	5	1	19
9	Financial District	1	3	1	9
10	Flatiron	4	2	1	24
11	Gramercy	1	3	3	9
12	Greenwich Village	4	2	2	28
13	Hamilton Heights	0	0	0	2
14	Hudson Yards	0	1	0	2
15	Lenox Hill	2	0	0	3
16	Little Italy	2	4	0	20
17	Lower East Side	0	1	1	0
18	Manhattan Valley	0	1	1	0
19	Marble Hill	0	0	0	1
20	Midtown	4	2	7	12
21	Midtown South	2	1	2	15
22	Murray Hill	3	0	8	10
23	Noho	2	4	3	26

Finally, we will use K-means clustering, an unsupervised learning algorithm used to create K clusters of data points based on feature similarity, to create clusters of the neighborhoods that have similar features.

	Bronx	Manhattan
Elbow Method		
Silhouette score		
Best K value	5	6

By observing the elbow method and silhouette score, we find that the optimal number of clusters to use for Bronx is K=5 and for Manhattan is K = 6.

After applying K-means clustering, we got the following clusters:

Bronx

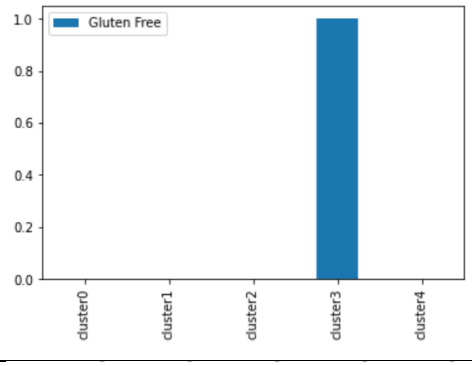
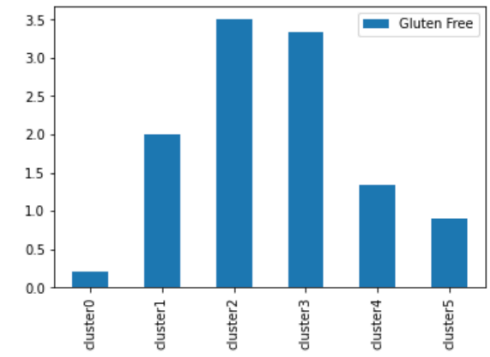
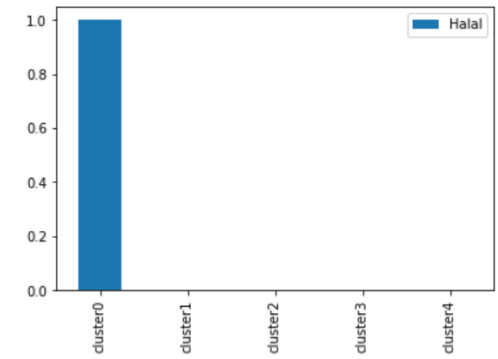
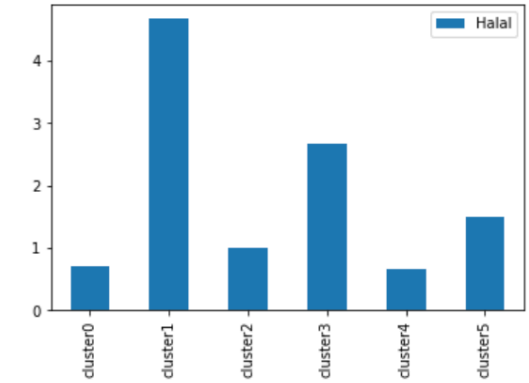
	Gluten Free	Halal	Kosher	Vegetarian/Vegan
cluster0	0.0	1.0	0.0	0.0
cluster1	0.0	0.0	0.0	1.0
cluster2	0.0	0.0	1.0	0.0
cluster3	1.0	0.0	0.0	0.5
cluster4	0.0	0.0	0.0	2.0

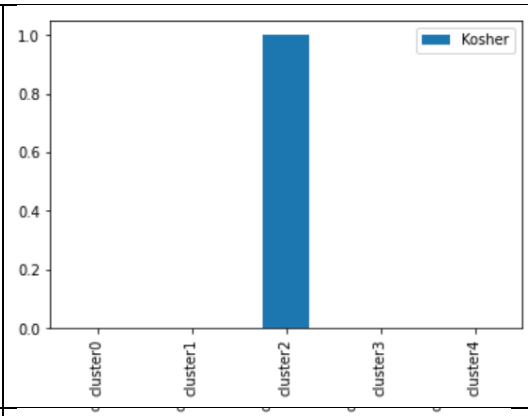
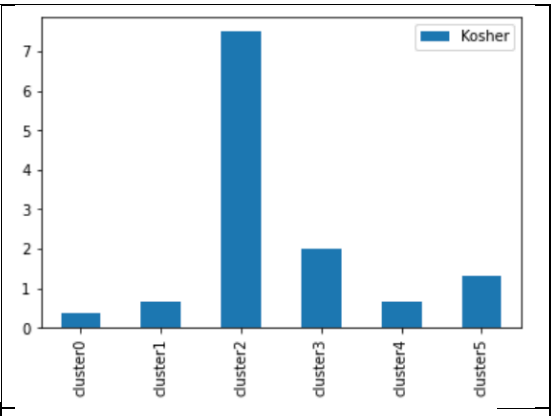
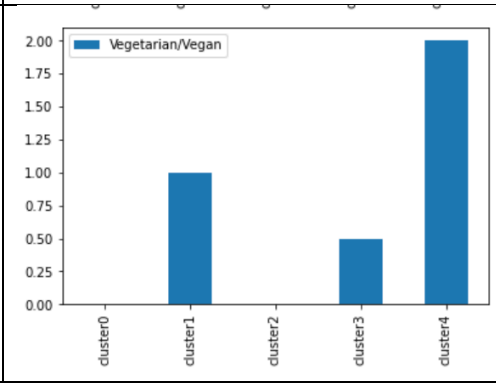
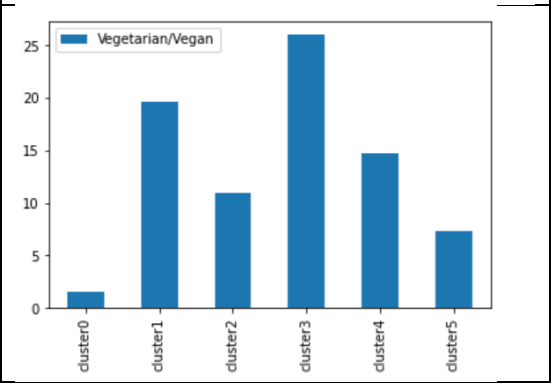
Manhattan

	Gluten Free	Halal	Kosher	Vegetarian/Vegan
cluster0	0.214286	0.714286	0.357143	1.500000
cluster1	2.000000	4.666667	0.666667	19.666667
cluster2	3.500000	1.000000	7.500000	11.000000
cluster3	3.333333	2.666667	2.000000	26.000000
cluster4	1.333333	0.666667	0.666667	14.666667
cluster5	0.900000	1.500000	1.300000	7.300000

Results

To prepare for the results, we need to understand the clustering.

Category	Bronx Clusters	Manhattan Clusters
Gluten Free		
Halal		

Kosher		
Vegetarian		

We can conclude from the charts above the best clusters for each category:

Bronx

- Cluster 0 - best for Halal Restaurants
- Cluster 1 - best for Vegan/Vegetarian Restaurants
- Cluster 2 - best for Kosher
- Cluster 3 - best for Gluten Free
- Cluster 4 - best for Vegan/Vegetarian Restaurants

Manhattan

- Cluster 0 - not recommended for selected diets
- Cluster 1 - best for Vegan/Vegetarian and Halal Restaurants
- Cluster 2 - best for Gluten Free and Kosher
- Cluster 3 - good for all four diets but best for Vegan/Vegetarian
- Cluster 4 - good for Vegan/Vegetarian
- Cluster 5 - low availability of restaurants in all categories

Now we merge the clusters, coordinated and neighborhoods

Bronx

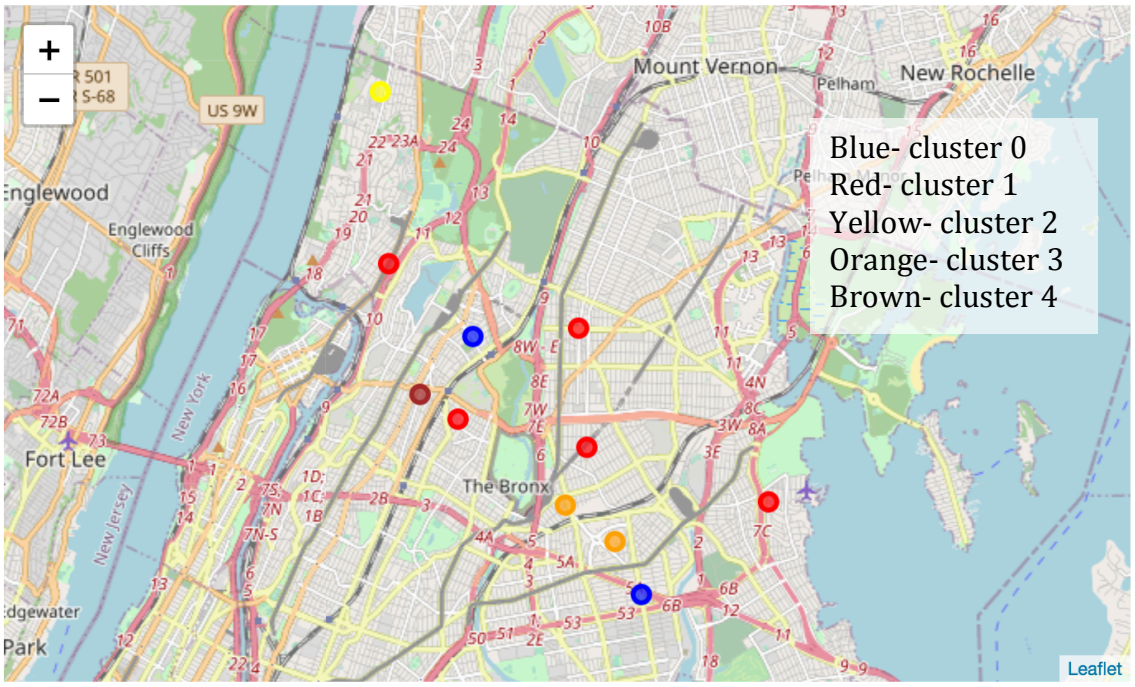
	Neighborhood	Gluten Free	Halal	Kosher	Vegetarian/Vegan	Cluster_Labels	Latitude	Longitude
0	Bedford Park	0	1	0	0	0	40.870185	-73.885512
1	Belmont	0	0	0	1	1	40.857277	-73.888452
2	Bronxdale	0	0	0	1	1	40.852723	-73.861726
3	Country Club	0	0	0	1	1	40.844246	-73.824099
4	Fordham	0	0	0	2	4	40.860997	-73.896427
5	Kingsbridge	0	0	0	1	1	40.881687	-73.902818
6	North Riverdale	0	0	1	0	2	40.908543	-73.904531
7	Olinville	0	0	0	1	1	40.871371	-73.863324
8	Parkchester	1	0	0	1	3	40.837938	-73.856003
9	Unionport	0	1	0	0	0	40.829774	-73.850535
10	Van Nest	1	0	0	0	3	40.843608	-73.866299

Manhattan

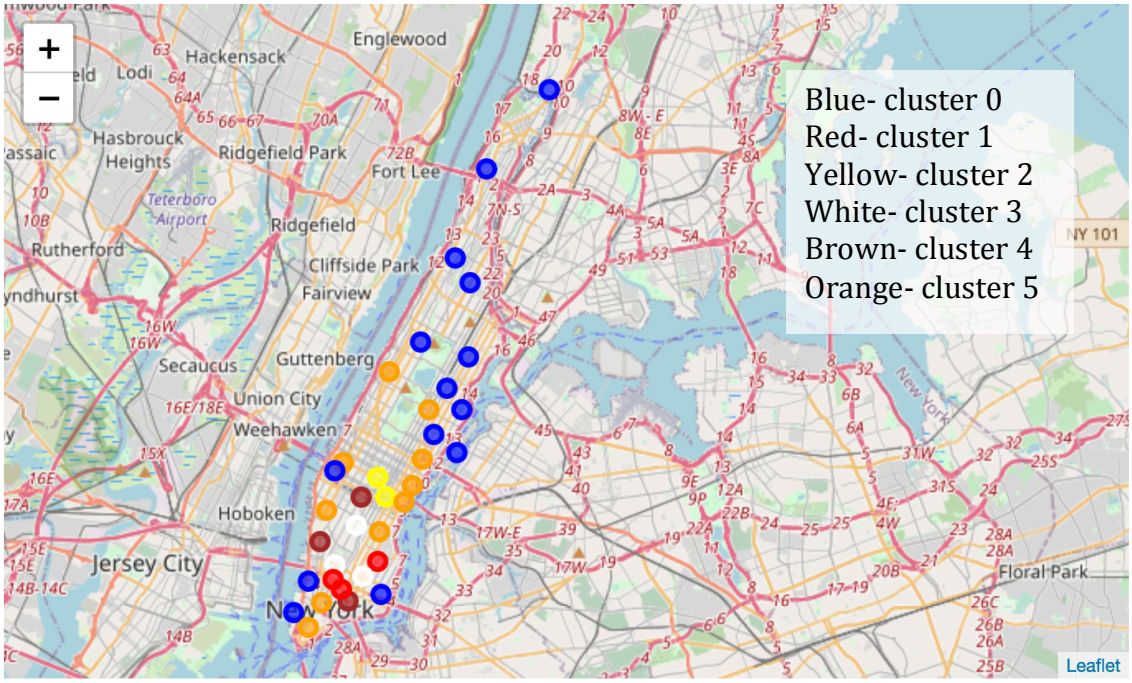
	Neighborhood	Gluten Free	Halal	Kosher	Vegetarian/Vegan	Cluster_Labels	Latitude	Longitude
0	Battery Park City	0	0	0	1	0	40.711932	-74.016869
1	Carnegie Hill	1	0	1	4	0	40.782683	-73.953256
2	Central Harlem	0	3	0	1	0	40.815976	-73.943211
3	Chelsea	1	2	0	7	5	40.744035	-74.003116
4	Chinatown	0	1	0	16	4	40.715618	-73.994279
5	Civic Center	1	2	1	6	5	40.715229	-74.005415
6	Clinton	2	1	0	6	5	40.759101	-73.996119
7	East Harlem	0	2	1	1	0	40.792249	-73.944182
8	East Village	2	5	1	19	1	40.727847	-73.982226
9	Financial District	1	3	1	9	5	40.707107	-74.010665
10	Flatiron	4	2	1	24	3	40.739673	-73.990947
11	Gramercy	1	3	3	9	5	40.737210	-73.981376
12	Greenwich Village	4	2	2	28	3	40.726933	-73.999914
13	Hamilton Heights	0	0	0	2	0	40.823604	-73.949688
14	Hudson Yards	0	1	0	2	0	40.756658	-74.000111
15	Lenox Hill	2	0	0	3	0	40.768113	-73.958860
16	Little Italy	2	4	0	20	1	40.719324	-73.997305
17	Lower East Side	0	1	1	0	0	40.717807	-73.980890
18	Manhattan Valley	0	1	1	0	0	40.797307	-73.964286
19	Marble Hill	0	0	0	1	0	40.876551	-73.910660
20	Midtown	4	2	7	12	2	40.754691	-73.981669
21	Midtown South	2	1	2	15	4	40.748510	-73.988713
22	Murray Hill	3	0	8	10	2	40.748303	-73.978332
23	Noho	2	4	3	26	3	40.723259	-73.988434
24	Roosevelt Island	0	0	1	0	0	40.762160	-73.949168

Now, we are going to use Folium library to visualize the clusters:

Bronx



Manhattan



Discussion

From the results, we can derive that generally there are obvious limitations of restaurants of selected categories in Bronx. However, there are much more varieties in Manhattan.
Here are some points:

- In both Bronx and Manhattan, there are more neighborhoods serving vegetarian and vegan food compared to other categories.
- In Bronx, there is only one neighborhood, which has kosher restaurants.
- In Manhattan, there are 14 neighborhoods with significantly limited selection of restaurants and 10 neighborhoods with low availability of selections.
- Gluten free and kosher restaurants are low in numbers in both boroughs

Conclusion

To conclude, we used K-means clustering which created different clusters for neighborhoods in Bronx and Manhattan boroughs in New York City. The results of the analysis will be used by a travel agency to recommend neighborhoods for visitors following restricted diets.