Restaurants in New York

Introduction:

In this project we try to find an optimal location for a restaurant. This report is targeted to stakeholders interested in opening a restaurant in New York.

Since there are lots of restaurants in New York we detected locations that are mostly crowded with restaurants and what category of restaurants are the most.

We have used our data science powers to explore every neighborhood based on this criteria. Advantages of each area is clearly expressed so that best possible final location can be chosen by stakeholders.

Data:

New York has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and longitude coordinates of each neighborhood.

Luckily, this dataset exists for free on the web. Here is the link to the dataset: https://geo.nyu.edu/catalog/nyu_2451_34572

We have used Foursquare API to get the most common venues of New York

We have used geocoder to get the coordinates of each borough.

Methodology:

In this project we will direct our efforts on detecting areas of New York that have most no of restaurant, particularly what category of restaurants are the most.

In first step we have collected the required data: location and type (category) of every restaurant in New York. We have also identified that there is only one Pakistani Restaurant in New York (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration of restaurant across different areas of New York — we will see which areas of New York have most no of restaurant, particularly what category of restaurants are the most.

In third and final step we will focus on areas and create clusters (using k-means clustering) of locations that meet some basic requirements established in discussion with stakeholders and present map of all locations to identify optimal venue location by stakeholders.

Analysis:

First, we will extract the relevant data from our database and store it in a data frame.

```
In [8]: neighborhoods.head()
```

Out[8]:

	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

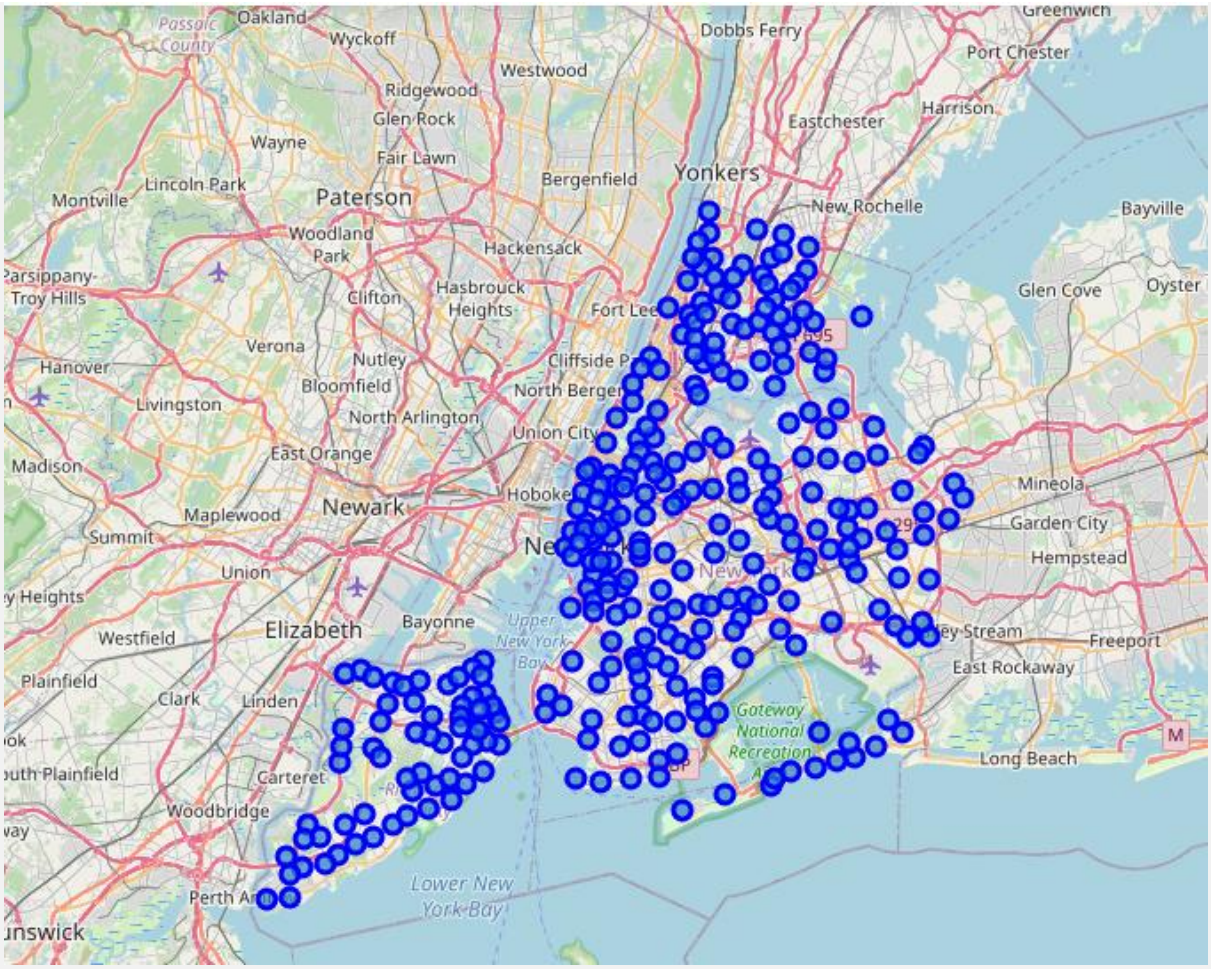
Then we find out latitude and longitude of each borough

```
In [12]: boroughs = getCordinates(addresses = neighborhoods['Borough'].unique())
boroughs
```

Out[12]:

	Borough	Latitude	Longitude
0	Bronx	40.846651	-73.878594
1	Manhattan	40.789624	-73.959894
2	Brooklyn	40.650104	-73.949582
3	Queens	40.749824	-73.797634
4	Staten Island	40.583456	-74.149605

We used python folium library to visualize geographic details of New York and its boroughs and created a map of New York with its neighborhood superimposed on top.



We then utilized the Foursquare API to explore the neighborhoods and segment them.

```
In [17]: newyork_venues.head()
```

Out[17]:

	Borough	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bronx	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Bronx	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
2	Bronx	Wakefield	40.894705	-73.847201	Walgreens	40.896528	-73.844700	Pharmacy
3	Bronx	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
4	Bronx	Wakefield	40.894705	-73.847201	Shell	40.894187	-73.845862	Gas Station

Then we extract restaurant data from venues

```
In [240]: newyork_venues_restaurant = newyork_venues[newyork_venues['Venue Category'].str.contains('Restaurant')].reset_index(drop=True)

newyork_venues_restaurant.rename(columns={'Venue': 'Restaurant', 'Venue Latitude': 'Restaurant Latitude',  'Venue Longitude': 'R
newyork_venues_restaurant.head()
```

Out[240]:

	Borough	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Restaurant	Restaurant Latitude	Restaurant Longitude	Restaurant Category
0	Bronx	Co-op City	40.874294	-73.829939	Arby's	40.870411	-73.828606	Fast Food Restaurant
1	Bronx	Co-op City	40.874294	-73.829939	Townhouse Restaurant	40.876086	-73.828868	Restaurant
2	Bronx	Eastchester	40.887556	-73.827806	Fish & Ting	40.885656	-73.829197	Caribbean Restaurant
3	Bronx	Eastchester	40.887556	-73.827806	Dyre Fish Market	40.889318	-73.831453	Seafood Restaurant
4	Bronx	Eastchester	40.887556	-73.827806	Royal Caribbean Bakery	40.888249	-73.831661	Caribbean Restaurant

2502 restaurants were returned by Foursquare with 88 unique categories.

Then we explore each neighborhood along with the top most common category of restaurant.

```
neighborhoods_restaurant_sorted.head()
```

Out[241]:

	Neighborhood	1st Most number of Restaurant	2nd Most number of Restaurant	3rd Most number of Restaurant	4th Most number of Restaurant	5th Most number of Restaurant	6th Most number of Restaurant	7th Most number of Restaurant	8th Most number of Restaurant	9th Most number of Restaurant	10th Most number of Restaurant
0	Allerton	Chinese Restaurant	Fast Food Restaurant	Spanish Restaurant	Vietnamese Restaurant	Greek Restaurant	Ethiopian Restaurant	Falafel Restaurant	Filipino Restaurant	French Restaurant	German Restaurant
1	Annadale	Restaurant	Vietnamese Restaurant	Kebab Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	Filipino Restaurant	French Restaurant	German Restaurant	Gluten-free Restaurant
2	Arlington	American Restaurant	Vietnamese Restaurant	Hawaiian Restaurant	Falafel Restaurant	Fast Food Restaurant	Filipino Restaurant	French Restaurant	German Restaurant	Gluten-free Restaurant	Greek Restaurant
3	Arrochar	Italian Restaurant	Mediterranean Restaurant	Polish Restaurant	Middle Eastern Restaurant	Vietnamese Restaurant	Greek Restaurant	Falafel Restaurant	Fast Food Restaurant	Filipino Restaurant	French Restaurant
4	Arverne	Restaurant	Thai Restaurant	Vietnamese Restaurant	Halal Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	Filipino Restaurant	French Restaurant	German Restaurant

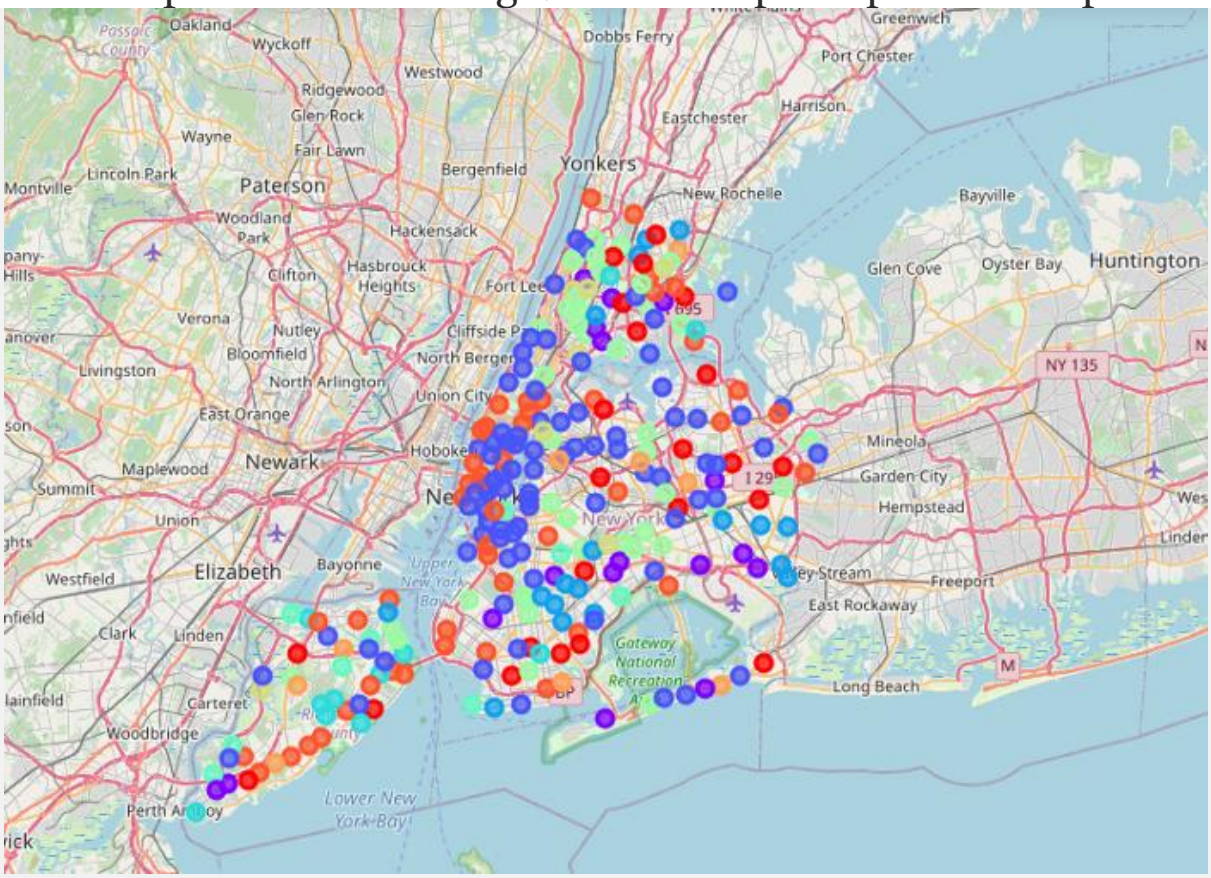
As we have clear indication of neighborhood with the greatest number of restaurants and what category of restaurants. We then cluster those locations to create zones containing similar category restaurants locations. Those zones will be the final result of our analysis.

```
newyork_restaurant_neighborhood_merged.head()
```

Out[189]:

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most number of Restaurant	2nd Most number of Restaurant	3rd Most number of Restaurant	4th Most number of Restaurant	5th Most number of Restaurant	6th Most number of Restaurant	7th Most number of Restaurant	8th Most number of Restaurant
1	Bronx	Co-op City	40.874294	-73.829939	8	Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Kebab Restaurant	Ethiopian Restaurant	Falafel Restaurant	Filipino Restaurant	French Restaurant
2	Bronx	Eastchester	40.887556	-73.827806	3	Caribbean Restaurant	Seafood Restaurant	Fast Food Restaurant	Chinese Restaurant	Ethiopian Restaurant	Falafel Restaurant	Filipino Restaurant	French Restaurant
5	Bronx	Kingsbridge	40.881687	-73.902818	6	Mexican Restaurant	Latin American Restaurant	Spanish Restaurant	Chinese Restaurant	Caribbean Restaurant	Seafood Restaurant	Fast Food Restaurant	Restaurant
6	Manhattan	Marble Hill	40.876551	-73.910660	2	Seafood Restaurant	Kebab Restaurant	Empanada Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	Filipino Restaurant	French Restaurant
7	Bronx	Woodlawn	40.898273	-73.867315	9	American Restaurant	Italian Restaurant	Indian Restaurant	Vietnamese Restaurant	Halal Restaurant	Falafel Restaurant	Fast Food Restaurant	Filipino Restaurant

Now we put clusters of neighborhood superimposed on top.



Discussion:

Our analysis shows that there is a great number of restaurants in New York, but there is only one Pakistani Restaurant in Downtown, Brooklyn. There are 88 unique categories of

restaurant. Most number of restaurants are in Manhattan and in neighborhoods the greatest number of restaurants are in Murray Hill

After directing our attention, we first created a dense grid of location; then we identify areas as which type of restaurant are there most. These locations were then clustered to create zones of interest which contain greatest number of similar locations.

Results:

Result of all these zones containing category of most numbers of restaurants. This, of course, does not imply that those zones are actually optimal locations for a new restaurant! Purpose of this analysis was to only provide info on areas with most number of restaurants in New York — it is entirely possible that there is a very good reason for small number of restaurants in any of those areas, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area. Zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

Conclusion:

Purpose of this project was to explore New York areas with most number of restaurant and areas with most number of

restaurants (particularly Pakistani restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new restaurant. By collecting restaurant data from Foursquare data we have first classified neighborhood and then boroughs and then generated extensive collection of locations which satisfy our requirements. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) were created to be used as starting points for final exploration by stakeholders.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.