Coursera Capstone Project

The Battle of Neighborhoods - Final Report (Week 1 and Week 2)

Content

1. Introduction Section

1.1 Discussion of the business problem

Let's consider a scenario when someone who lived for a long time in Russia is moving to New York and looking for a place to live that has Russian restaurants. New York city is the largest city in the US and well-known to be one of the most multi-cultural and diverse cities in the world. Thus, it is not difficult to find any types of restaurants, shops of diverse culture. As a result, a person decided to find a place that has Russian restaurants. Moreover, it is important to have banks and grocery shops in the vicinity.

The following criteria should be taken into account in selecting the place of living:

1. Russian restaurants with high average rating;
2. Banks and grocery shops, which are common venues in the neighborhood.

The challenge is to analyse the neighborhoods that have the venues with a person's priority, if possible, to find the top venues. The data that are required to find a location is described below.

1.2 Audience for this project

This project might be interesting for people who should like to search a place with similar priorities, since the approach and methodology used are applicable for different scenarios.

2. Data Section

2.1 Data required to resolve the problem

The following data are required to find a location:

- New York City data that contain list of Boroughs, Neighborhoods along with their latitude and longitude.
- Venues in each neighborhood of New York city.

2.3 Data sources and how data are used to resolve the problem

The source of the list of New York city Borough, Neighborhood and Geodata information is located here: https://cocl.us/newyork_dataset. These data are used to have an information about New York city's neighborhood.

```
neighborhoods.head(10)
```

Out[2]:

	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
5	Bronx	Kingsbridge	40.881687	-73.902818
6	Manhattan	Marble Hill	40.876551	-73.910660
7	Bronx	Woodlawn	40.898273	-73.867315
8	Bronx	Norwood	40.877224	-73.879391
9	Bronx	Williamsbridge	40.881039	-73.857446

Foursquare API data contain information about the venues with the specified locations. Foursquare and geopy data are used to map top venues for all New York neighborhoods and clustered in groups.

The data will be used as follows: Use Foursquare and geopy data to map top venues for all New York neighborhoods and clustered in groups. A map will be created that shows the neighborhoods with the highest amount of top venues. From the map it will be easily visualized which neighborhood is the most suitable to the person and depicts his/her preferences.

```
In [3]: # Use geopy library to get the latitude and longitude values of New York City
address = 'New York City, NY'

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

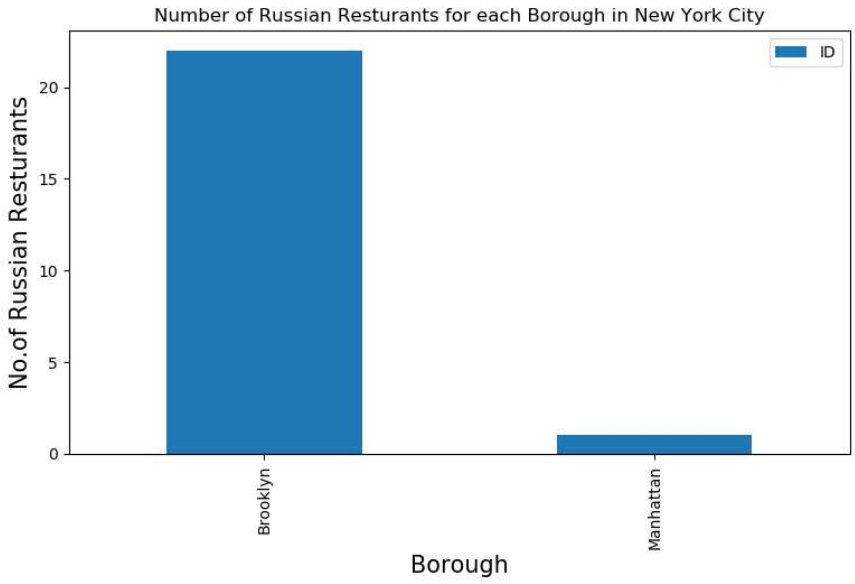
```
The geograpical coordinate of New York City are 40.7127281, -74.0060152.
```

3. Methodology section

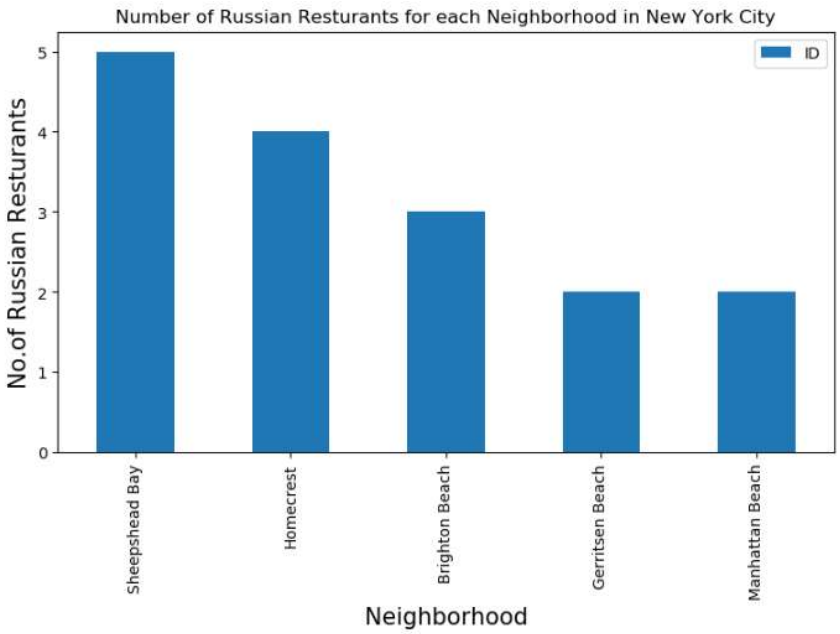
Define a function to interact with Foursquare API and get top 100 venues within a radius of 1000 metres for a given latitude and longitude. Below function will return us the venue id, venue name and category. Define a function to get venue details like count, rating, tip counts for a given venue id. This will be used for ranking. Collect Russian restaurants for each Neighborhood.

	Borough	Neighborhood	ID	Name
0	Brooklyn	Bensonhurst	51f092a2498e9f9edd14b597	Verrazano Grill
1	Brooklyn	Gravesend	4a2937fcf964a52069951fe3	Stolovaya
2	Brooklyn	Brighton Beach	4ac69109f964a52017b520e3	Skovorodka
3	Brooklyn	Brighton Beach	4245f780f964a52070201fe3	Tatiana Restaurant
4	Brooklyn	Brighton Beach	4c1ea63763750f474c19ba67	Volna Café

There are 23 Russian restaurants in New York city returned by Foursquare API.



There are only two boroughs with Russian restaurants. The highest number of Russian restaurants is located in Brooklyn.



Two neighborhoods share the highest number of Russian restaurants - Homecrest and Sheepshead Bay.

Further the ranking of the restaurants is analyzed.

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips
0	Brooklyn	Bensonhurst	0	0	0	0	0
1	Brooklyn	Gravesend	4a2937fcf964a52069951fe3	Stolovaya	41	7.9	27
2	Brooklyn	Brighton Beach	4ac69109f964a52017b520e3	Skovorodka	92	7.9	48
3	Brooklyn	Brighton Beach	4245f780f964a52070201fe3	Tatiana Restaurant	157	7.1	67
4	Brooklyn	Brighton Beach	4c1ea63763750f474c19ba67	Volna Café	32	6.3	20
5	Brooklyn	Sheepshead Bay	5a08f7e2d48ec1369706f57d	Bellarussian Xata	23	9.1	6
6	Brooklyn	Sheepshead Bay	55d6885e498e41b9e7b8b5e8	Cafe Dushanbe	10	7.3	3
7	Brooklyn	Sheepshead Bay	5324ee02498e9a825bb4f689	La Vue Restaurant	23	7.2	3
8	Brooklyn	Sheepshead Bay	4b53cec1f964a520f9ab27e3	Vanka Vstanka	5	6.1	4
9	Brooklyn	Sheepshead Bay	4b2ecbf5f964a520cee624e3	Passage	8	6	5

The neighborhoods with maximum average ratings are selected to prioritize the ones with higher ratings.

	Neighborhood	Average Rating
3	Gramercy	8.70
4	Gravesend	7.90
9	Sheepshead Bay	7.14
1	Brighton Beach	7.10
5	Homecrest	6.85
7	Manhattan Terrace	6.70
8	Ocean Parkway	6.70
2	Gerritsen Beach	6.65
6	Manhattan Beach	6.65
0	Bensonhurst	0.00

The next table shows the maximum average ratings in two boroughs, where Manhattan has higher rating compared to Brooklyn.

	Borough	Average Rating
1	Manhattan	8.700000
0	Brooklyn	6.622727

Only restaurants with average rating above 6 are visualized on the map:

	Neighborhood	Average Rating
1	Brighton Beach	7.10
2	Gerritsen Beach	6.65
3	Gramercy	8.70
4	Gravesend	7.90
5	Homecrest	6.85
6	Manhattan Beach	6.65
7	Manhattan Terrace	6.70
8	Ocean Parkway	6.70
9	Sheepshead Bay	7.14

Average ratings were then included to the list with Borough, Neighborhoods, Geodata information.

	Borough	Neighborhood	Latitude	Longitude	Average Rating
0	Brooklyn	Brighton Beach	40.576825	-73.965094	7.10
1	Brooklyn	Gerritsen Beach	40.590848	-73.930102	6.65
2	Manhattan	Gramercy	40.737210	-73.981376	8.70
3	Brooklyn	Gravesend	40.595260	-73.973471	7.90
4	Brooklyn	Homecrest	40.598525	-73.959185	6.85
5	Brooklyn	Manhattan Beach	40.577914	-73.943537	6.65
6	Brooklyn	Manhattan Terrace	40.614433	-73.957438	6.70
7	Brooklyn	Ocean Parkway	40.613060	-73.968367	6.70
8	Brooklyn	Sheepshead Bay	40.586890	-73.943186	7.14

The dataframe containing the data about borough, neighborhood, geodata and average rating of the restaurants is created. The next step is to create a map.

The map above shows the locations of neighborhoods with maximum average ratings of Russian restaurants. There is only one neighborhood and limited number of Russian restaurants in Manhattan, even if it has the highest maximum rate. It is better to have more options of neighborhoods and restaurants to decide. Thus, Brooklyn is further chosen for the analysis. One of the criteria of the person was to have the banks and grocery stores. So, the ratings and locations of these venues in Brooklyn Borough is analysed.

	Borough	Neighborhood	Latitude	Longitude
0	Brooklyn	Bay Ridge	40.625801	-74.030621
1	Brooklyn	Bensonhurst	40.611009	-73.995180
2	Brooklyn	Sunset Park	40.645103	-74.010316
3	Brooklyn	Greenpoint	40.730201	-73.954241
4	Brooklyn	Gravesend	40.595260	-73.973471
5	Brooklyn	Brighton Beach	40.576825	-73.965094
6	Brooklyn	Sheepshead Bay	40.586890	-73.943186
7	Brooklyn	Manhattan Terrace	40.614433	-73.957438
8	Brooklyn	Flatbush	40.636326	-73.958401
9	Brooklyn	Crown Heights	40.670829	-73.943291

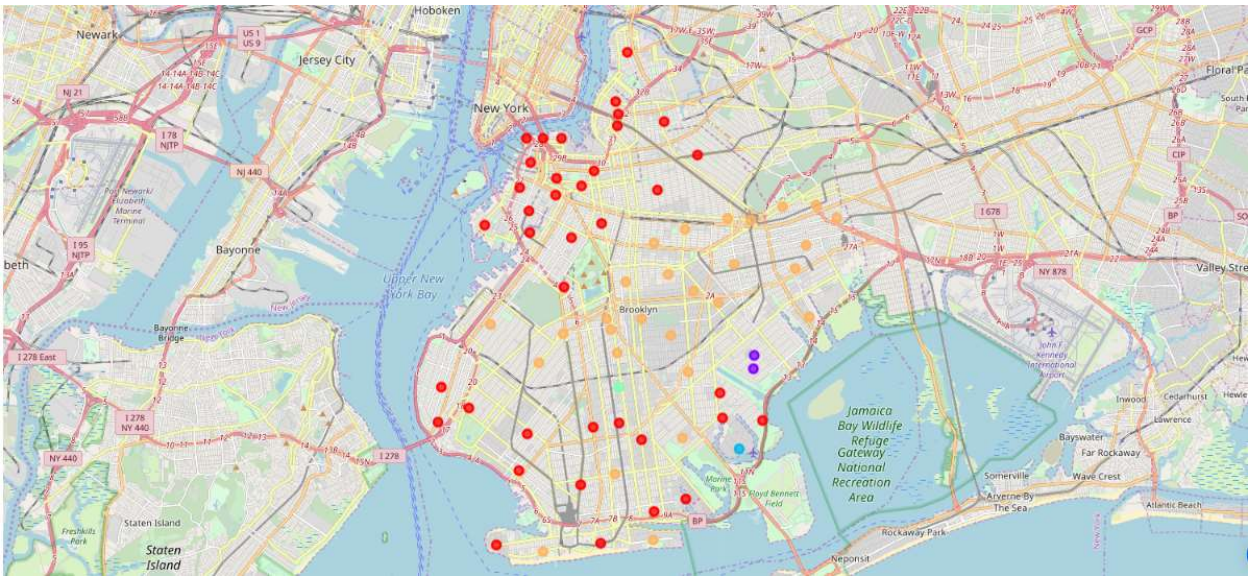
The venues in Brooklyn returned by Foursquare API:

	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	Cocoa Grinder	40.623967	-74.030863	Juice Bar
3	Bay Ridge	40.625801	-74.030621	Pegasus Cafe	40.623168	-74.031186	Breakfast Spot
4	Bay Ridge	40.625801	-74.030621	Ho' Brah Taco Joint	40.622960	-74.031371	Taco Place

Further each neighborhood is analyzed based on the venues. And the list containing the top 10 venues for each neighborhood is created. k-means is run to cluster the neighborhood into 5 clusters. A new dataframe is created that includes the clusters as well as the top 10 venues for each neighborhood.

	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	Italian Restaurant	Pizza Place	Bagel Shop	Greek Restaurant	American Restaurant	Bar	Ice Cream Shop	Sandwich Place	Chinese Restaurant
1	Brooklyn	Bensonhurst	40.611009	-73.995180	0	Chinese Restaurant	Ice Cream Shop	Italian Restaurant	Donut Shop	Grocery Store	Sushi Restaurant	Spa	Food Truck	Pet Store	Smoke Shop
2	Brooklyn	Sunset Park	40.645103	-74.010316	4	Bakery	Bank	Pizza Place	Mexican Restaurant	Latin American Restaurant	Pharmacy	Mobile Phone Shop	Gym	Women's Store	Italian Restaurant
3	Brooklyn	Greenpoint	40.730201	-73.954241	0	Bar	Coffee Shop	Cocktail Bar	Pizza Place	Yoga Studio	French Restaurant	Café	Boutique	Mexican Restaurant	Record Shop
4	Brooklyn	Gravesend	40.595260	-73.973471	0	Chinese Restaurant	Italian Restaurant	Pizza Place	Bakery	Bus Station	Lounge	Gym	Cupcake Shop	Pharmacy	Donut Shop

The resulting clusters can be visualized on the map:



Each cluster is examined and the discriminating venue categories that distinguish each cluster are determined. Based on the defining categories, two clusters (1 & 5) are the most preferred ones that complies the second criteria of having banks and grocery stores in the vicinity.

4. Results section

Only 23 Russian restaurants were found in 306 neighborhoods, and 9 restaurants that have maximum average rating above 6. Most of them are located in Brooklyn borough. Thus, it was decided to select this borough for further analysis as the high number of restaurants give more options to select which is more preferable. All venues in Brooklyn were clustered into 5 clusters based on their most common venues. Homecrest and Manhattan Terrace are the neighborhoods which have all the requirements that have to be complied. They both have Russian restaurants as well as banks and grocery stores. They fall under two different clusters (1 & 5), and have quite similar average ratings of Russian restaurants. Thus, these two neighborhoods both can be considered for living.

5. Discussion section

While doing the analysis some interesting points were found:

- 23 Russian restaurants were provided by Foursquare. It's remarkable that there are many Russian restaurants in New York city, which can show how much the city is diverse.
- The maximum average ratings of Russian restaurants are low compared to other restaurants.

- While analysis of Brooklyn's venues, it was observed that there are lots of other popular restaurants, shops and bars which makes this Borough even more attractive for living.

6. Conclusion section

Although the analysed data are not enough for selecting the place to live, we have come up with a way to solve the problem of finding a place with minimum criteria based on the venues. In order to apply this in real life, it is important to consider rent price, location, neighborhood and lots of other factors which are not discussed in this report. In general, by applying this simple analysis we could decrease the search locations to two neighborhoods as a starting point.