



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

1.2 Audience for this project

2. Data Section

2.1 Data required to resolve the problem

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

3. Methodology section

3.1 Steps of data analysis

3.2 Data Science Methods, machine learning, mapping tools and exploratory data analysis.

4. Results section

Discussion of the results

5. Discussion section

Discussion on any observations and/or recommendations based on the results.

6. Conclusion section

Report Conclusion



Introduction Section

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:

- ▶ Russian restaurants with high average rating;
- ▶ 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.

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.

Data Section

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.

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.

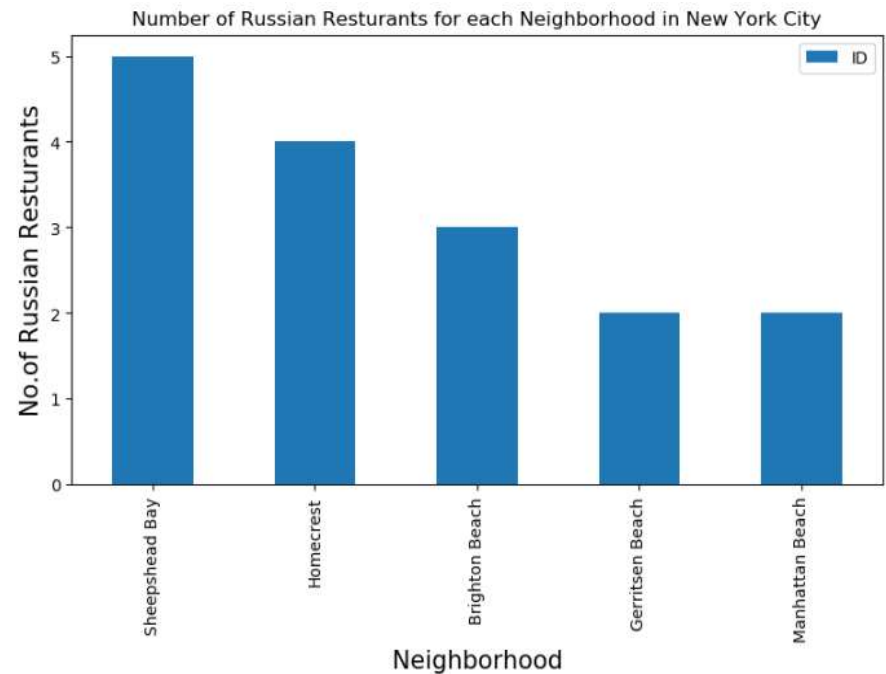
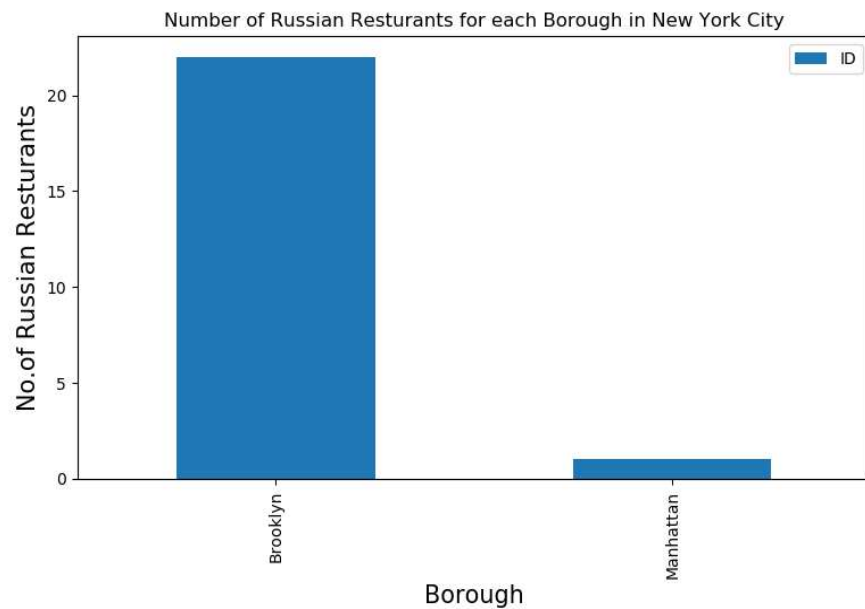
Foursquare API data contain information about the venues with the specified locations. Foursquare and geopy data are used to map top venues for all New York neighborhoods and clustered in groups.

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.
- ▶ 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.
- ▶ 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.

Results

Number of Russian restaurants vs. Neighborhoods and Boroughs



Restaurants with maximum average rating

	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

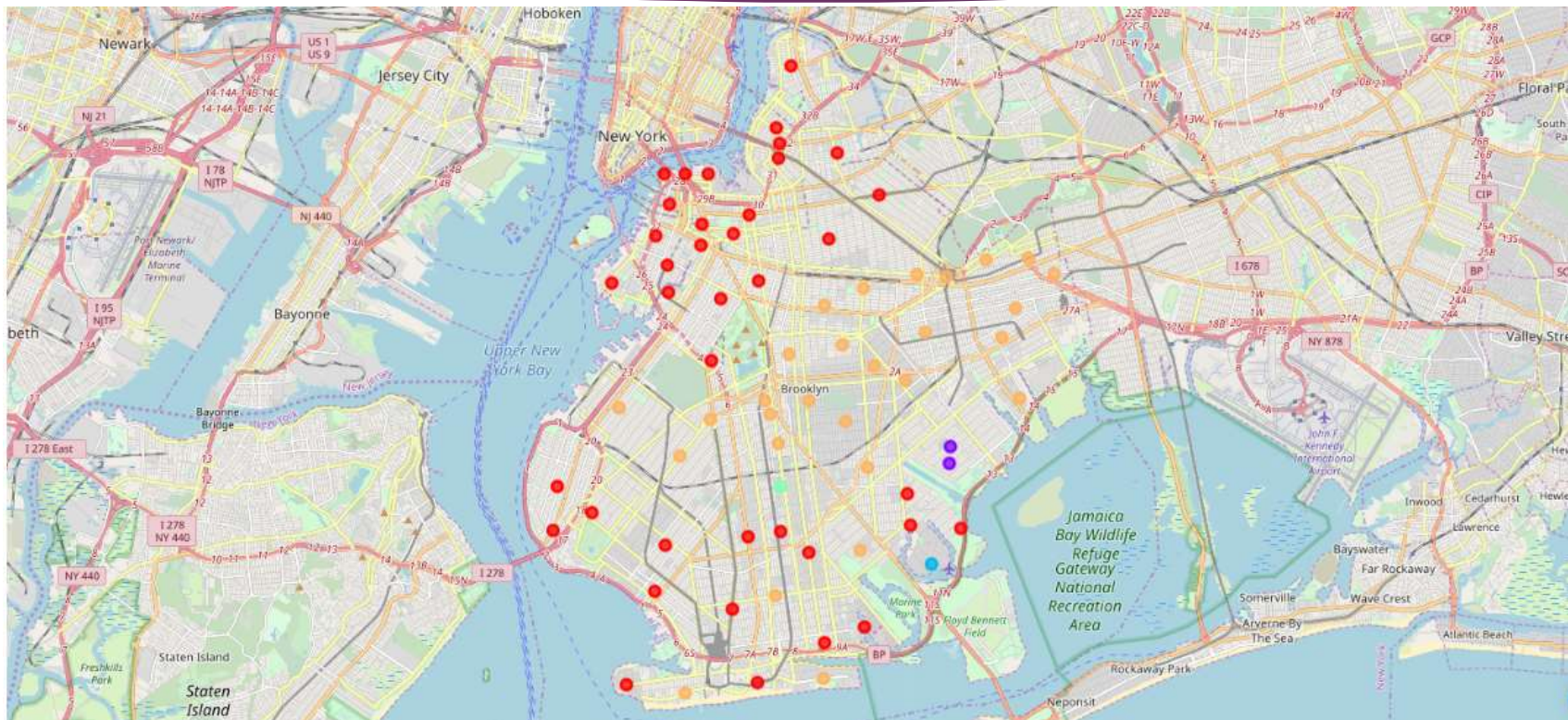
Location of restaurants with maximum average ratings



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

Clusters on the map





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.

Conclusion 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.
- ▶ 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.