

# The Battle for Neighborhood: A detailed report

## **TABLE OF CONTENT:**

- *Introduction:* Discuss the business problem and get familiar to the project.
- *Data Selection:* Description of valuable data resources to be used to complete the project.
- *Methodology:* Application of data cleaning, sorting, analysis, applying machine learning algorithms, data training & testing in order to establish the purpose to reach the target.
- *Result:* Get familiar to the results or output
- *Discussion:* Explanation of the results and discuss the outcome.
- *Conclusion:* Finalize report

## **1. Introduction:**

This project is regarding the comparison of neighborhoods of two cities namely "Manhattan NY" and "Toronto" in choosing the best place to live provided all the necessary amenities around.

The basic idea behind this comparison is that I have got new job offers as a data scientist in "Manhattan NY" and another in "Toronto". There are some basic evaluation criteria, which I would like to check by means of this project in searching for apartment and living standards in the two cities. Following are the basic queries to be addressed:

1. Two bedrooms or three bedroom apartment.
2. Metro station must be 1 mile from the place of living.
3. Rent not more than \$10000.
4. Top amenities must be included in both places.
5. Restaurants, groceries stores, wine shop are desirable nearby.

## **Problems:**

The difficult part is to locate a suitable apartment for rent in both places "Manhattan NY" and "Toronto" that fulfill the demands of price, location and amenities. In the following section 2 I will explain the process of data collection and importance while addressing the issues.

## 2. Data formulation:

**Following data are required for the project**

1. List of Boroughs and neighborhoods of Manhattan and Toronto with their geodata (Latitude and Longitude).
2. List of subways metro stations with their address location.
3. List of apartment for rent in Manhattan and Toronto.
4. Desired additional information about the price, number of beds etc.
5. Venues for each neighborhood in Manhattan and Toronto.
6. Venues for subway metro stations.

**How to solve problem using above data?**

We will use the data in following way:

1. Use Foursquare and geopy data to sort the top 10 venues in "Toronto" and "Manhattan" neighborhood and cluster them in-group. Then locate the subway stations separately and layer it over the created cluster group.
2. Again use the same geopy and foursquare data to map the location of rental places.
3. Filter the data with the average rental price \$1000 per month and radius of 1 mile near the subway station.
4. Convert the address of the rental location to latitude and longitude.

### METHODOLOGY: Data Mapping

		name	categories	lat	lng
0		Souk Tabule	Mediterranean Restaurant	43.653756	-79.354390
1		Young Centre for the Performing Arts	Performing Arts Venue	43.650825	-79.357593
2		Sukhothai	Thai Restaurant	43.653701	-79.354447
3		SOMA chocolatemaker	Chocolate Shop	43.650622	-79.358127
4	Toronto Cooper Koo Family Cherry St YMCA Centre		Gym / Fitness Center	43.653191	-79.357947
5		Cluny Bistro & Boulangerie	French Restaurant	43.650565	-79.357843
6		The Distillery Historic District	Historic Site	43.650244	-79.359323
7		Underpass Park	Park	43.655764	-79.354806
8		Balzac's Coffee	Coffee Shop	43.649797	-79.359142
9		Corktown Common	Park	43.655618	-79.356211

Table 1: Toronto neighborhood cleaning of dataset according to the latitude and longitude location.

	<b>name</b>	<b>categories</b>	<b>lat</b>	<b>lng</b>
0	The Bar Room at Temple Court	Hotel Bar	40.711448	-74.006802
1	The Beekman - A Thompson Hotel	Hotel	40.711173	-74.006702
2	City Hall Park	Park	40.712415	-74.006724
3	Alba Dry Cleaner & Tailor	Laundry Service	40.711434	-74.006272
4	The Wooly Daily	Coffee Shop	40.712137	-74.008395
5	Gibney Dance Center Downtown	Dance Studio	40.713923	-74.005661
6	Augustine	French Restaurant	40.711310	-74.006660
7	The Class by Taryn Toomey	Gym / Fitness Center	40.712753	-74.008734
8	Takahachi Bakery	Bakery	40.713653	-74.008804
9	Aahar Indian Cuisine	Indian Restaurant	40.713307	-74.007994

Table 2: New York (Manhattan) neighborhood cleaning of dataset according to the latitude and longitude location.

Now once we have sorted out the dataset of both Toronto and Manhattan, NY we can overlay the places onto the geological map as shown in Figure 1 and Figure 2.

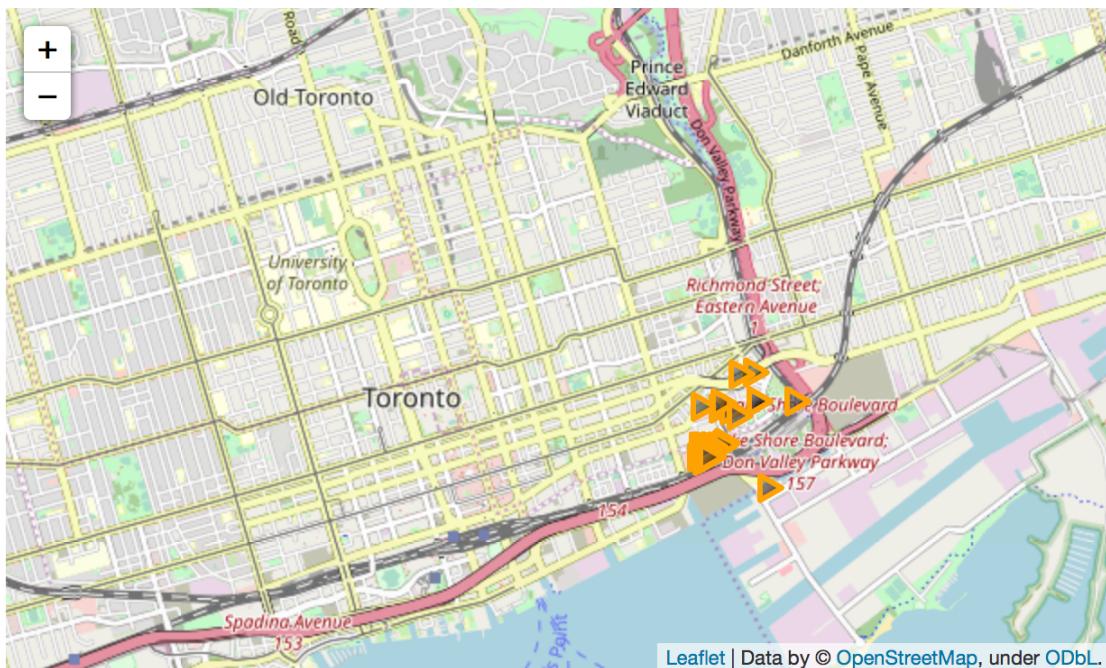


Fig 1: Toronto location map of various places of interests in yellow triangle.

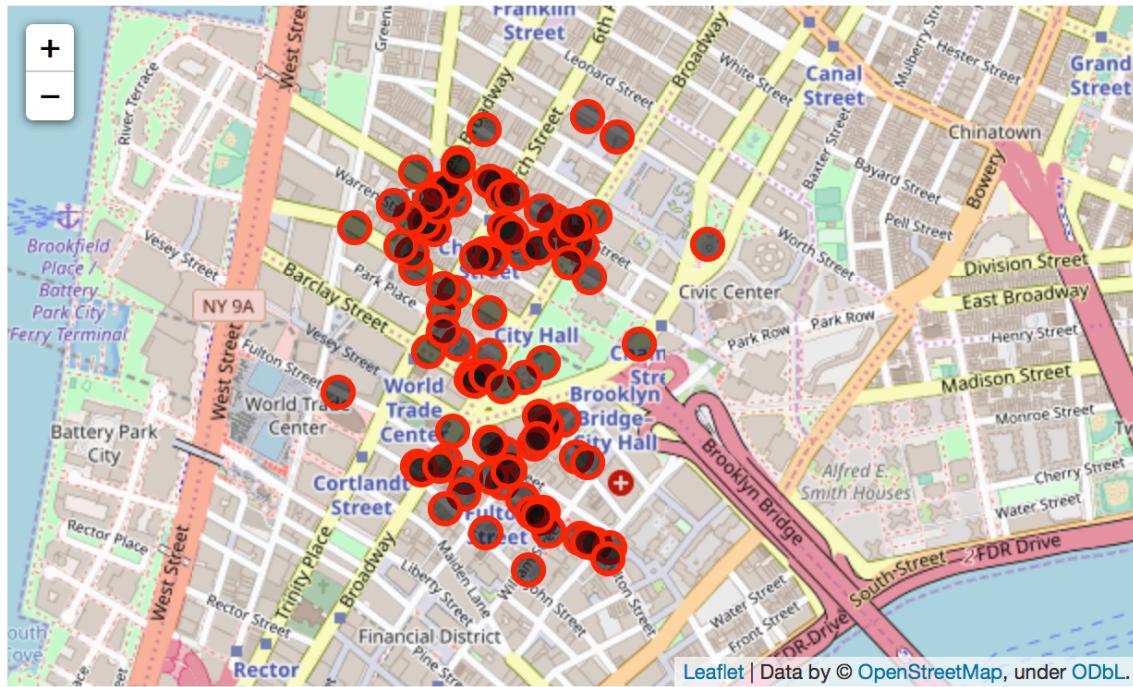


Fig 2: Manhattan, NY location map of various place of interests in red circle.

Next we have sorted out the venues as per most common places in the cities. Below is the Manhattan details about the classified data sorting.

	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
0	Manhattan	Marble Hill	40.876551	-73.910660	2	Coffee Shop	Discount Store	Yoga Studio	Steakhouse	Supplement Shop	Tennis Stadium	Shoe Store	Gym
1	Manhattan	Chinatown	40.715618	-73.994279	2	Chinese Restaurant	Cocktail Bar	Dim Sum Restaurant	American Restaurant	Vietnamese Restaurant	Salon / Barbershop	Noodle House	Bakery
2	Manhattan	Washington Heights	40.851903	-73.936900	4	Café	Bakery	Mobile Phone Shop	Pizza Place	Sandwich Place	Park	Gym	Latin American Restaurant
3	Manhattan	Inwood	40.867684	-73.921210	3	Mexican Restaurant	Lounge	Pizza Place	Café	Wine Bar	Bakery	American Restaurant	Park
4	Manhattan	Hamilton Heights	40.823604	-73.949688	0	Mexican Restaurant	Coffee Shop	Café	Deli / Bodega	Pizza Place	Liquor Store	Indian Restaurant	Sushi Restaurant

Table 3: Clustering of dataset with most common venues around the neighborhood.

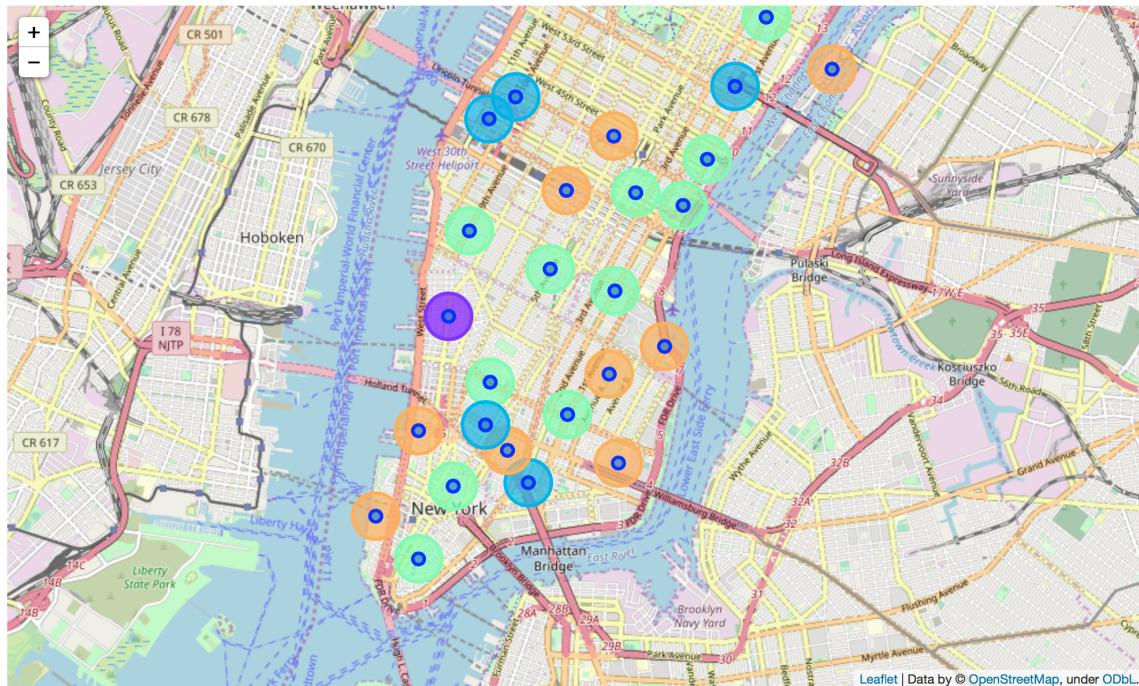


Figure 3: Classification of places with most common venues represented in various colors.

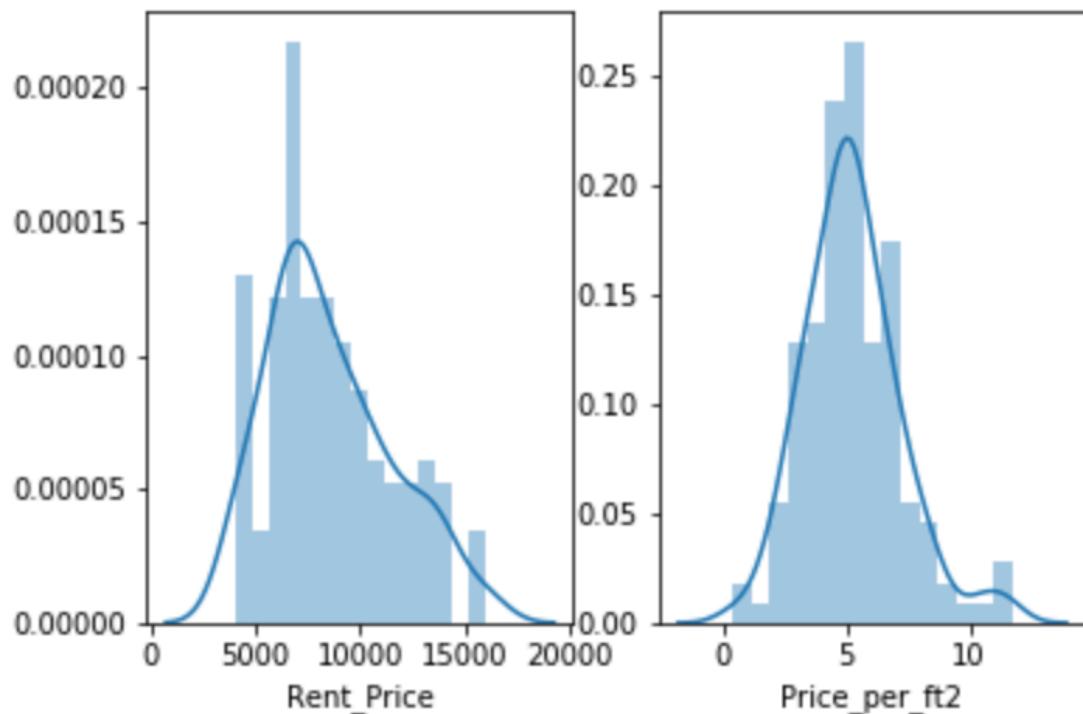


Figure 4: Histogram plot to the apartment rent in Manhattan, NY as per price and price/ $\text{ft}^2$

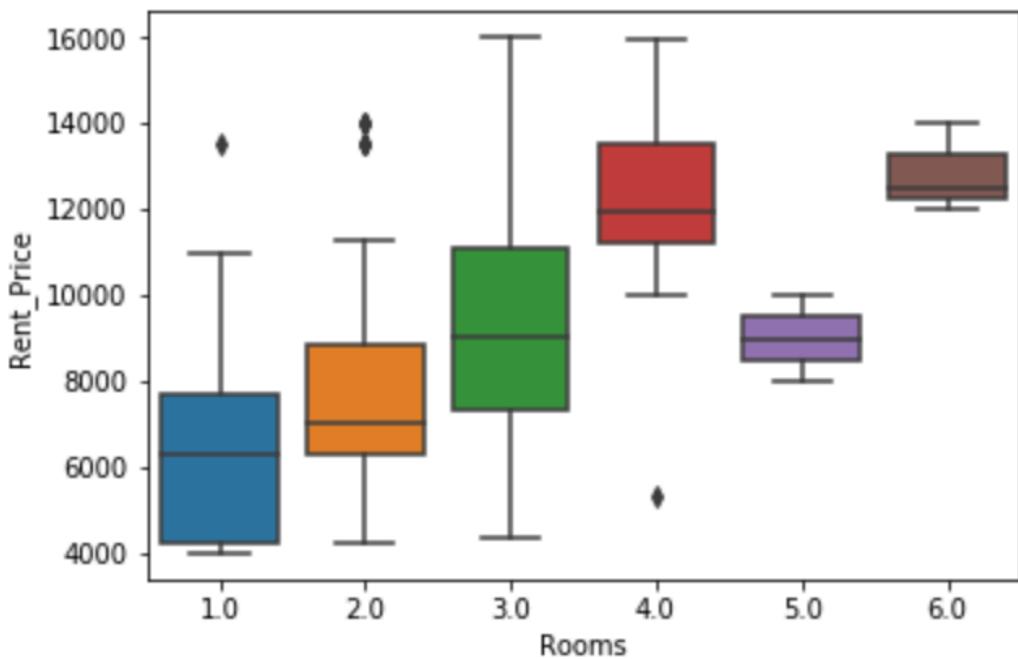


Figure 5: Box plot to the apartment rent in Manhattan, NY as per rooms

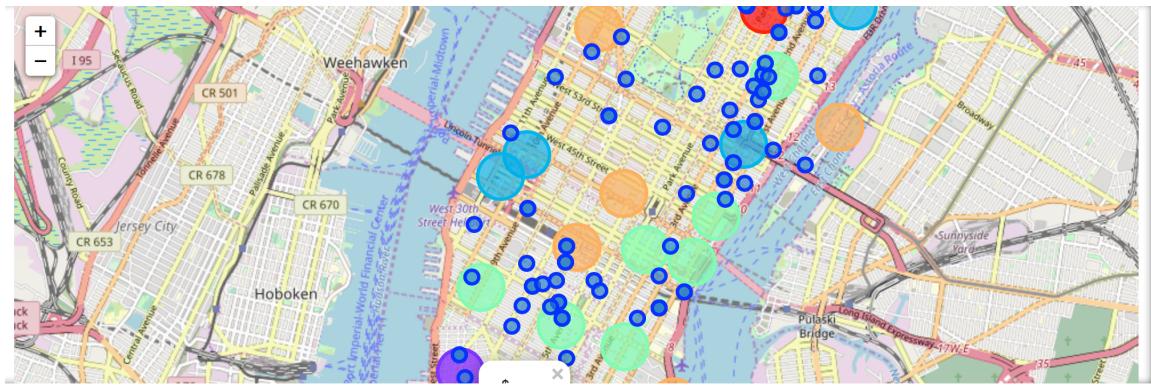


Figure 6: Apartment showing near to the clusters of choices nearby.

Finally in Figure 6 one can easily locate the apartment as per the choice of the amenities around the neighborhood. In conclusion, the target to the specific problem of battle for neighborhood is achieved.

