

A Simple Hotel Recommendation System using Foursquare API

Tom Sze

Update: July 15, 2019

1 Introduction

For those who like to travel with more freedom, it is sometimes hard and time-consuming to decide which hotel to stay based on a number of tourist attractions planning to go in a way that would shorten the time between the hotel and attractions. It would be nice to be able to choose a hotel based on the distance to tourist attractions and hotels' rating as a simple recommendation. The further potential application might be a smart trip planning application if local transportation information is easily available.

2 Data

2.1 Data Source

We want the user to input the city, the number of tourist attractions planning to go and the minimum rating. We also want the user to specify hotels' minimum rating so hotels of poor rating will not be recommended. Then, tourist attractions and nearby hotels' geographical locations will be queried using the Foursquare API to calculate the distance from the hotel to tourist attractions. In this case, we set the city to be Toronto, the number of tourist attractions planning to go to be 4, the minimum rating to be 7 (total 10), the hotels' minimum rating to be 8 (total 10).

2.2 Data Cleaning

For data without a rating, geographical information or category, it will be filtered. For data that does not satisfy the user input, it will be filtered also. First five row of data of tourist attractions is shown in **Figure 1** and First 5 row of data of hotels is shown in **Figure 2**.

	name	rating	city	lat	lng	categories
0	Centreville Amusement Park	8.1	Toronto	43.620370	-79.373792	Theme Park
1	Casa Loma	8.8	Toronto	43.677934	-79.409521	Castle
2	The Distillery Historic District	9.3	Toronto	43.650244	-79.359323	Historic Site
3	Royal Ontario Museum	9.1	Toronto	43.668367	-79.394813	Museum
4	Spadina Museum: Historic House & Gardens	7.4	Toronto	43.678883	-79.408858	History Museum

Figure 1: First five row of data of tourist attractions

	name	rating	city	lat	lng	categories
0	Shangri-La Toronto	9.0	Toronto	43.649129	-79.386557	Hotel
1	Delta Hotels by Marriott Toronto	9.3	Toronto	43.642882	-79.383949	Hotel
2	The Fairmont Royal York	8.7	Toronto	43.645449	-79.381508	Hotel
3	The Ritz-Carlton Toronto	8.7	Toronto	43.645330	-79.387089	Hotel
4	The Grand Hotel & Suites Toronto	9.0	Toronto	43.656449	-79.374110	Hotel

Figure 2: First five row of data of hotels

3 Methodology

This system gives recommendations based on the distance between hotels and attractions and hotels' rating. The shorter the distance between the hotel and tourist attraction and the higher the hotel rating, the higher the recommendation will be given.

We find the distance of each hotel to each tourist attraction and we average those distances for each hotel.

3.1 Exploratory Data Analysis

3.1.1 Relation between average hotel distance to chosen attractions and hotel rating

We make a scatter plot to observe the relation between average hotel distance to chosen attractions and hotel rating shown in **Figure 3**.

3.1.2 Distribution of average distance of hotel to chosen attractions

We plot a histogram to observe the distribution of hotel distance to attraction shown in **Figure 4**.

3.1.3 Distribution of hotel rating

We plot a histogram to observe the distribution of hotel rating in **Figure 5**.

Relation between average hotel distance to chosen attractions and hotel rating

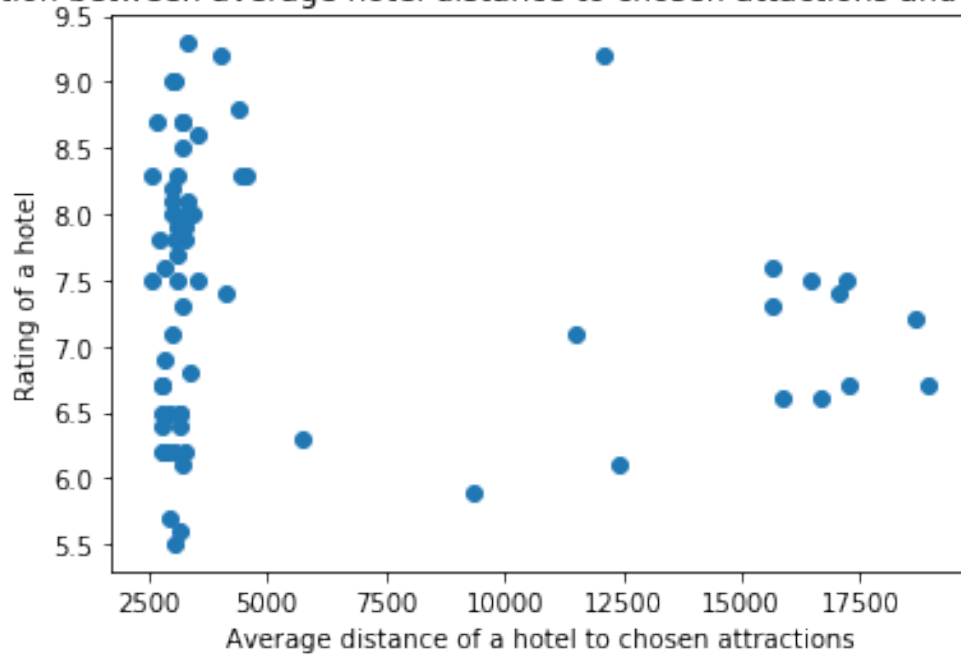


Figure 3: Scatter plot of average hotel distance to chosen attractions and hotel rating

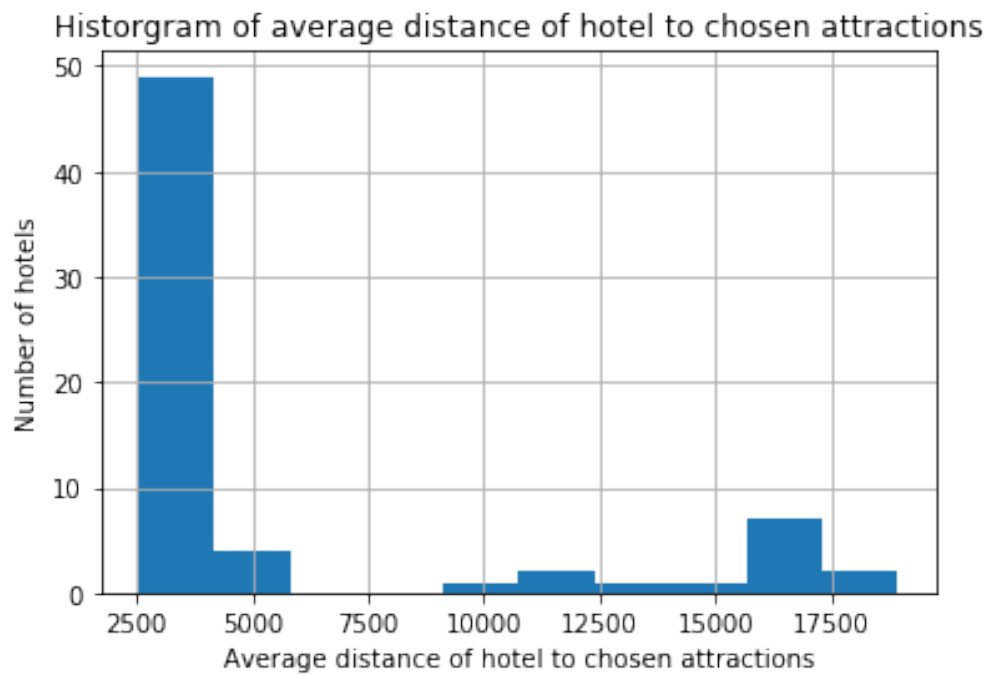


Figure 4: Histogram of average distance of hotel to chosen attractions

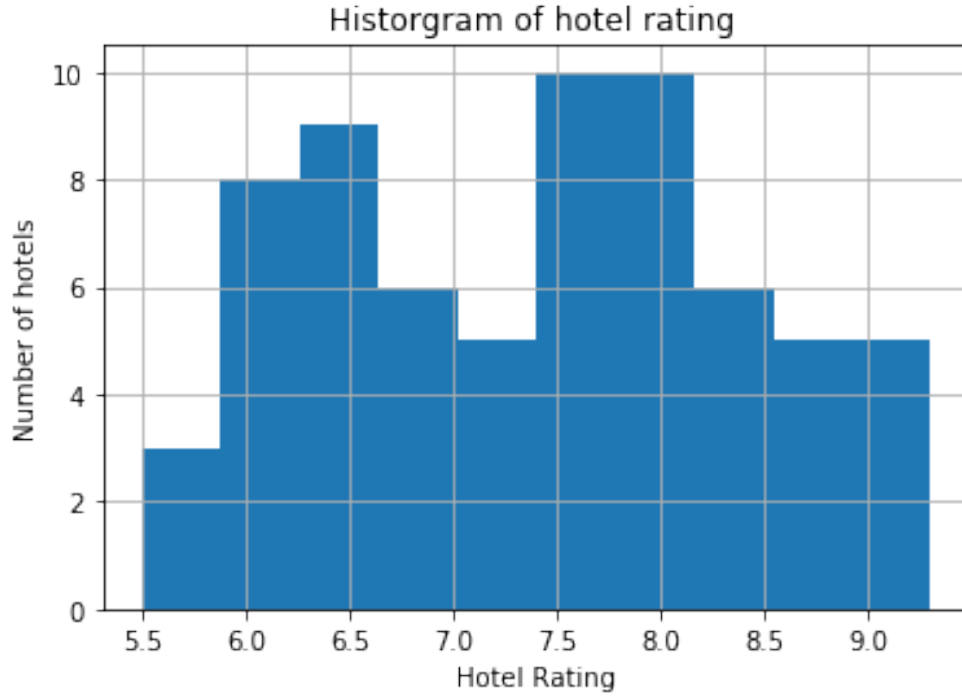


Figure 5: Histogram of hotel rating

	name	rating	city	lat	lng	categories	dist	cluster
0	Shangri-La Toronto	9.0	Toronto	43.649129	-79.386557	Hotel	3050.887840	2
1	Delta Hotels by Marriott Toronto	9.3	Toronto	43.642882	-79.383949	Hotel	3325.326945	2
2	The Fairmont Royal York	8.7	Toronto	43.645449	-79.381508	Hotel	3212.265661	2
3	The Ritz-Carlton Toronto	8.7	Toronto	43.645330	-79.387089	Hotel	3216.308478	2
4	The Grand Hotel & Suites Toronto	9.0	Toronto	43.656449	-79.374110	Hotel	2995.854308	2

Figure 6: First five row of data with k-means clustering result in column 'cluster'

3.2 K-means clustering

We use only two parameters to carry out k-means clustering. They are the average distance from the hotel to chosen attractions and the hotel's rating as shown in **Figure 6**.

3.3 Recommendation strategy

Only with clustering is not enough, we have to decide which cluster to recommend to the user. As we know that the higher the hotel rating and the shorter the distance between the hotel and chosen attractions, the better the recommendation. A simple way to represent this is simply multiply the rating and inverse of the average distance (rating over distance or rod) as shown in **Figure 7**.

	name	rating	city	lat	lng	categories	dist	cluster	rod
0	Four Seasons Hotel Toronto	8.7	Toronto	43.671796	-79.389457	Hotel	2667.946190	2	0.003261
1	The Hazelton Hotel	8.3	Toronto	43.670764	-79.393257	Hotel	2549.653127	2	0.003255
2	The Grand Hotel & Suites Toronto	9.0	Toronto	43.656449	-79.374110	Hotel	2995.854308	2	0.003004
3	Shangri-La Toronto	9.0	Toronto	43.649129	-79.386557	Hotel	3050.887840	2	0.002950
4	Windsor Arms Hotel	7.5	Toronto	43.668781	-79.390850	Hotel	2558.494501	0	0.002931

Figure 7: First five row of data with hotel rating over average distance to attractions in column ‘rod’

4 Results

We show the cluster returned by the k-means algorithm onto the map using different colors. The hotels far away from the attractions have a distinct color from those close to the attractions. The large average distance between the hotel and attractions is making the difference in k-means clustering. In **Figure 8**, we show that the best hotel(pointy marker) to recommend and the cluster it belongs to will also be recommended.

5 Discussion

There is no data for training a model based on a human decision in this case. In fact, this is just a simple recommendation based on two parameters, the average distance between the hotel and chosen attraction, the rating of the hotel.

It is true that it is not accurate simply use the shortest distance between hotel to attractions. If local transportation information is available, more close-to-reality travel time can be calculated. As a result, smart trip planning is also possible with transportation tips.

6 Conclusion

In conclusion, this simple idea based on two parameters, the hotel ratings and the average distance hotel to attractions, to give recommendation on the hotel is possible and is close to expectation.

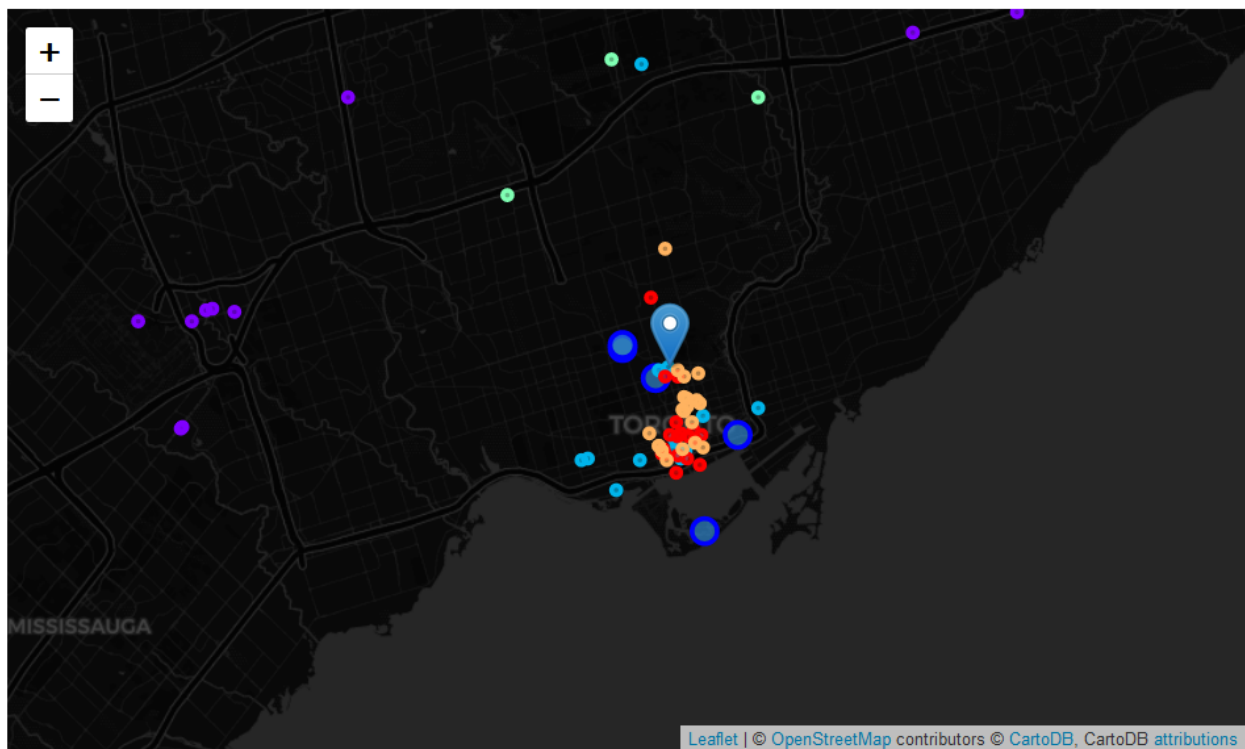


Figure 8: Map of Toronto with markers of tourist attraction(large blue circle), hotels(colored circle) and recommended hotel(pointy marker)