

## 1. Introduction

(*Disclaimer:* For this task, the real names of the hotels were used for demonstration purposes only, in order to apply the capabilities of *Foursquare API* more meaningfully. However, we do not claim that the task as stated below was indeed provided by the company mentioned in the report).

**Background:** The international hotel chain *Marriott* recently opened a steakhouse in one of its New York locations, *New York Downtown Marriot*. Because this new establishment was very successful, the owners would be interested to open another steakhouse in one of its Chicago hotels. In order to maximize the chances of success, they would like to find a location which would be similar to the current one. The company provided the following list of nine “candidate” hotels for consideration: *Chicago Marriott Southwest at Burr Ridge*, *Chicago Marriott at Medical District/UIC*, *JW Marriott Chicago*, *Chicago Marriott Downtown Magnificent Mile*, *Marriott Marquis Chicago*, *Chicago Marriott Midway*, *Marriott Chicago O’Hare*, *Chicago Marriott Suites O’Hare*, *Chicago Marriott Oak Brook*. Our task is to analyze the list of candidate hotels for the new location and to make preliminary recommendations based on publically available data.

## 2. Data

Because clientele of the steakhouse consists of the guests who are staying hotel, as well the residents of the city (and possibly tourists residing in other hotels), the similarity of the new location to the current one should somehow address both of these factors.

### 2.1. Data for evaluation of hotel guests similarity

We assume, that guests of the hotel might have similar attitude and level of satisfaction when evaluating both hotel and eating establishments. Thus, to measure the similarity between this part of customer base, we will use publicly available data presented in [www.bookings.com](http://www.bookings.com). More precisely, for each hotel we will use average user scores based on how the guests rated the following characteristics of the hotel:

a) staff; b) facilities; c) cleanliness; d) comfort; e) value for money; f) location.

Example of reviewer scores is presented in the figure below:

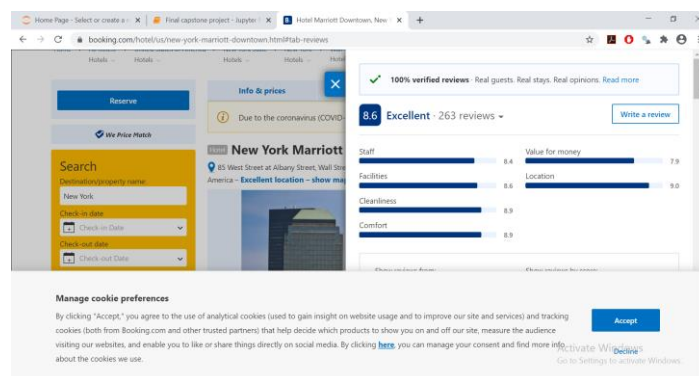


Fig. 1. Example of hotel guests ratings

For our analysis, these scores will be stored in 6-tuple of numerical values, which will serve as our reviews profile of the hotel.

## 2.2. Data for evaluation of neighborhood similarity

To address the similarity between other parts of clientele base, we will analyze neighborhoods around the hotels, based on the data provided by *Foursquare API*. Our assumption is that types of locations residing around the hotel forms or at least reveals a distinct profile, which is connected to foot traffic of possible clients. To form such neighborhood profile, we will extract the categories of the venues located around the hotel.

The Foursquare API presents very distinct list of categories. Here is an example of locations found around the *New York Downtown Marriot*:

```
'Park', 'Memorial Site', 'Gourmet Shop', 'Coffee Shop',  
'Jewelry Store', 'Museum', 'Pet Store', 'Lingerie Store',  
'Food Truck', 'Pub', 'Steakhouse', 'Wine Shop', 'Garden',  
'Auditorium', 'Dog Run', 'Gym / Fitness Center', 'Trail',  
'Grocery Store', 'Ice Cream Shop', 'Gym', 'Tree', 'Hotel', 'Plaza',  
'Gift Shop', 'Mexican Restaurant', 'Salad Place', 'Playground',...
```

We'll group all the categories provided by Foursquare API into six types, which, in our opinion, might be relevant for our task:

a) Various shops and stores; b) Tourists attractions and other walking destinations, such as parks, museums, monuments etc.; c) Restaurants (other than steakhouses); d) Steakhouses; e) Other food and beverage serving establishments, such as pubs, cafes, pizza places, burger joints etc.; f) Other locations.

The extracted number of each type of location will when be divided by the overall number of found locations and stored into another 6-tuple, which will serve us as profile of the neighborhood for a particular hotel.

## 3. Methodology

### 3.1 Evaluating similarity between hotels

The collected review and neighborhood profiles of “candidate” hotels will be compared to the respective profiles of the current location, New York Marriott Downtown. To do so, we will measure Euclidian distances (a particular case of Minkowski distance, with value  $p = 2$ ) between vectors, which will serve as an indicator of dissimilarity between the hotels. To do so, we will apply *minkowski* function which is implemented in *python scipy* package.

To account for both reviews and neighborhood profiles, we will use combined measures, based on the sums of ranks of the distances. The hotels with the lowest values of this joint measure should indicate the most similar hotels to the current location.

### 3.2 Cluster analysis

In addition, we will perform data clustering based on the distances between reviews and neighborhood profiles. This additional analysis will help to better determine the list of best locations, which are similar to each other, so that the hotel owners would have a list of alternative solutions for their development plans.

For cluster analysis, we will apply *k*-nearest neighbors method. More precisely, we will apply *KMeans* method which is implemented in *python sklearn* package. Before the clustering, distances between reviews and neighborhood profiles will be scaled using *StandardScaler* function from *python sklearn* library. This normalized distance measure will provide an additional indicator besides the sum of ranks, mentioned in 3.1 subsection.

Implementations of the code can be seen in the attached link to the *Jupyter Notebook*.

## 4. Results

### 4.1 Evaluation of hotel guests similarity

Customer reviews data, obtained from [www.bookings.com](http://www.bookings.com), is presented in the table below.

Table 1. Customer reviews data:

Name	Staff	Facilities	Cleanliness	Comfort	Value for money	Location
Chicago Marriott Southwest at Burr Ridge	8.4	8.6	8.7	9	8.1	8.6
Chicago Marriott at Medical District/UIC	9.1	8.8	9.3	9.3	8	8.9
JW Marriott Chicago	8.7	8.5	8.7	8.7	7.7	8.9
Chicago Marriott Downtown Magnificent Mile	8.8	8.6	8.9	8.9	8	9.4
Marriott Marquis Chicago	8.9	9.1	9.3	9.3	8	8.9
Chicago Marriott Midway	8.3	8.3	9	8.8	7.6	8.6
Marriott Chicago O'Hare	8.6	8.1	8.5	8.3	7.2	8.2
Chicago Marriott Suites O'Hare	8.9	9	9.2	9.3	8.5	8.9
Chicago Marriott Oak Brook	8.8	8.7	8.9	8.7	8.6	8.6
New York Marriott Downtown	8.4	8.6	8.9	8.9	7.9	9

Euclidian distances, obtained using python code (see attached link to *Jupyter Notebook*) are presented in the table below.

Table 2. Measured Euclidian distances between reviews profiles of the selected Chicago hotels and New York Marriott Downtown:

Name	Distance	Rank
Chicago Marriott Southwest at Burr Ridge	20.58726	4
Chicago Marriott at Medical District/UIC	21.41521	7
JW Marriott Chicago	20.51173	3
Chicago Marriott Downtown Magnificent Mile	21.08837	6
Marriott Marquis Chicago	21.45505	8
Chicago Marriott Midway	20.27652	2
Marriott Chicago O'Hare	19.57895	1
Chicago Marriott Suites O'Hare	21.56384	9
Chicago Marriott Oak Brook	20.94448	5

Thus, based on reviews profiles, the following three hotels are the most similar to the *New York Marriott Downtown*:

- 1) Marriott Chicago O'Hare
- 2) Chicago Marriott Midway
- 3) JW Marriott Chicago.

## 4.2 Evaluation of neighborhood similarity

We used *Foursquare API* and the names of the hotels (see the python code in *Jupyter Notebook*) to obtain geolocation data. The obtained coordinates of the hotels are presented in the Table 3.

Table 3. Geolocation data obtained from Foursquare API:

Name	Latitude	Longitude
Chicago Marriott Southwest at Burr Ridge	41.75605710	-87.90952445
Chicago Marriott at Medical District/UIC	41.87373425	-87.66615671
JW Marriott Chicago	41.87907385	-87.63301085
Chicago Marriott Downtown Magnificent Mile	41.89212200	-87.62485441
Marriott Marquis Chicago	41.85389855	-87.62026699
Chicago Marriott Midway	41.77394745	-87.74461269
Marriott Chicago O'Hare	41.98511205	-87.82987625
Chicago Marriott Suites O'Hare	41.98846610	-87.85891032
Chicago Marriott Oak Brook	41.85389855	-87.62026699
New York Marriott Downtown	40.70942675	-74.01434310

The lists of venues located in 500 meters radius around each hotel were extracted using *Foursquare API* (see the code in *Jupyter Notebook*). An example of some venues found around *New York Marriott Downtown* is presented in table below.

Table 4. The sample of nearest locations around New York Marriott Downtown

Name	Categories	Latitude	Longitude
Liberty Park	Park	40.71038392	-74.01386845
9/11 Memorial South Pool	Memorial Site	40.711039	-74.013082
Eataly Downtown	Gourmet Shop	40.710075	-74.011976
National September 11 Memorial & Museum	Memorial Site	40.711349	-74.013197
Blue Bottle Coffee	Coffee Shop	40.71058901	-74.0123713

All the categories provided by *Foursquare API* there divided into six larger types of categories, as presented in section 2. The extracted data containing the proportions of those types of locations are presented in Table 5. For more details, we refer to the attached link to *Jupyter Notebook*.

Table 5. The neighbourhood profiles of the hotels, based on categories of nearest venues

Name	Sites	Shops	Foods	Restaurants	Steakhouses	Other
Chicago Marriott Southwest at Burr Ridge	0.063	0.563	0.000	0.188	0.000	0.188
Chicago Marriott at Medical District/UIC	0.000	0.214	0.286	0.214	0.000	0.286
JW Marriott Chicago	0.060	0.262	0.298	0.250	0.000	0.131
Chicago Marriott Downtown Magnificent Mile	0.030	0.270	0.160	0.300	0.050	0.190
Marriott Marquis Chicago	0.135	0.077	0.231	0.212	0.000	0.346
Chicago Marriott Midway	0.000	0.130	0.261	0.261	0.000	0.348
Marriott Chicago O'Hare	0.000	0.200	0.200	0.100	0.100	0.400
Chicago Marriott Suites O'Hare	0.125	0.250	0.188	0.125	0.125	0.188
Chicago Marriott Oak Brook	0.135	0.077	0.231	0.212	0.000	0.346
New York Marriott Downtown	0.280	0.290	0.210	0.070	0.010	0.140

Euclidian distances, obtained using python code (see attached link to *Jupyter Notebook*) are presented in the table below.

Table 6. Measured Euclidian distances between neighbourhood profiles of the selected Chicago hotels and New York Marriott Downtown:

Name	Distance	Rank
Chicago Marriott Southwest at Burr Ridge	0.426409428	8
Chicago Marriott at Medical District/UIC	0.36334051	6
JW Marriott Chicago	0.299429541	2
Chicago Marriott Downtown Magnificent Mile	0.349857114	3
Marriott Marquis Chicago	0.360003287	4.5
Chicago Marriott Midway	0.431476123	9
Marriott Chicago O'Hare	0.403980198	7
Chicago Marriott Suites O'Hare	0.211275886	1
Chicago Marriott Oak Brook	0.360003287	4.5

Thus, based on reviews profiles, the following three hotels are the most similar to the *New York Marriott Downtown*:

- 1) Chicago Marriott Suites O'Hare
- 2) JW Marriott Chicago.
- 3) Chicago Marriott Downtown Magnificent Mile

#### 4.3 Evaluation of combined similarity between hotels

We combined the similarity hotels ranks based on reviews and neighborhoods profiles into a single index, which is the sum of these ranks. The lower values of this sum indicate the closer similarity of the particular hotel to *New York Marriott Downtown*. The results are presented in Table 7.

Table 7. Combined similarity index between hotels

Name	Reviews	Neighbourhood	Sum of ranks
Chicago Marriott Southwest at Burr Ridge	4	8	12
Chicago Marriott at Medical District/UIC	7	6	13
JW Marriott Chicago	3	2	5
Chicago Marriott Downtown Magnificent Mile	6	3	9
Marriott Marquis Chicago	8	4.5	12.5
Chicago Marriott Midway	2	9	11
Marriott Chicago O'Hare	1	7	8
Chicago Marriott Suites O'Hare	9	1	10
Chicago Marriott Oak Brook	5	4.5	9.5

The sum of similarity ranks indicates, that the JW Marriot Chicago is most similr to New York Marriott Downtow. Ti was the thrid closest hotel based on costuorem raeveiws data, and the second closes based on the types of venues located in the neighbourd.

#### 4.4 Cluster analysis for evaluation of similarity between hotels

For an additional evaluation of similarity, we performed cluster analysis. The distances between reviews and neighborhood profiles were selected as features for clusterisation. Thus, it also reveals similarity of Chicago hotels among themselves. Fig. 2 shows data points (preprocessed using *StandardScaler* function) for each hotel, before and after clusterisation. In this case, we applied *k*-nearest neighbors method, with value of *k* set to 3.

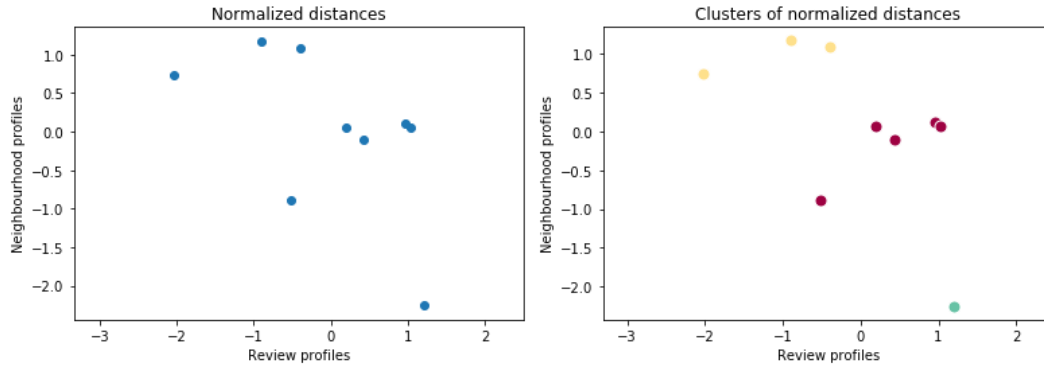


Fig. 2 Clusterisation of hotels similarity data based on distances between reviews and neighborhoods profiles. The left panel shows preprocessed data points, while the right panel shows clustered data (different clusters are denoted by different colors)

Table 8 shows the list of hotels with respective initial values of features, together with assigned cluster labels.

Table 8. Clusterisation results of Chicago hotels based on similarity measures

Name	Distance (reviews)	Distance (neighborhoods)	Cluster label
Chicago Marriott Southwest at Burr Ridge	20.58726	0.426409428	1
Chicago Marriott at Medical District/UIC	21.41521	0.36334051	0
JW Marriott Chicago	20.51173	0.299429541	0
Chicago Marriott Downtown Magnificent Mile	21.08837	0.349857114	0
Marriott Marquis Chicago	21.45505	0.360003287	0
Chicago Marriott Midway	20.27652	0.431476123	1
Marriott Chicago O'Hare	19.57895	0.403980198	1
Chicago Marriott Suites O'Hare	21.56384	0.211275886	2
Chicago Marriott Oak Brook	20.94448	0.360003287	0

Table 9 shows averaged values of (unprocessed) features, which helps us to get a clearer picture on the similarity between the members of the same cluster.

Table 9. Average values of features on the members of each cluster

Cluster	Distance (reviews)	Distance (neighbourhood)
0	21.08	0.3465
1	20.15	0.4206
2	21.56	0.2113

## 5. Discussion

The data shows that Chicago *Marriot* hotels would be ranked somewhat differently according to the highest similarity to the *New York Marriot Downtown* hotel based on reviews and location profiles. This conclusion can be clearly seen from similarity ranks presented in Tables 2 and 6, as well as from our clustering results (see Fig. 2 and Tables 8-9). For example, *Chicago Marriott Suites O'Hare* was most the least similar to *New Yorks Marriot Downtown* based on customer reviews profile, but the most similar according to the distances of neighborhood profiles.

Performed cluster analysis revealed that the considered hotels could be divided in three groups, which could be roughly described as follows:

- Cluster 0: average similarity reviews, average similarity of neighborhoods;
- Cluster 1: high similarity reviews, low similarity of neighborhoods;
- Cluster 2: low similarity reviews, high similarity of neighborhoods.

If both of these factors are of equal importance, the owners should consider the hotels assigned to Cluster 1. Among them, JW Marriot Chicago stands out, because it exhibits the closes similarity among based on both measures among the members of the same cluster. Moreover, its' rank sum was lowest among all hotels under consideration.

## 6. Conclusion

The performed analysis shows that *JW Marriot Chicago* hotel can be viewed as most similar hotel to the *New York Marriott Downtown*. As compared to other candidate hotels, it exhibits rather close similarity based on both reviews of the customers, as well as the types of venues located around the hotel. It was the 3<sup>rd</sup> and 2<sup>nd</sup> best choice based on the respective measures, therefore it can be viewed as very solid compromise combining both of these features.

## 7. Further considerations and recommendations

This preliminary analysis was performed based on publically available data and could serve as a first step for further consideration. For example, using more advanced plan of Foursquare API would allow for more detailed analysis. Customer review data of locations around the hotel, especially regarding all restaurants and steakhouses in particular, their menus, price points etc. could be very valuable. In addition, reviews of the guests of Marriot hotels would be of huge importance. Another publicly unavailable data, collected by owners of the hotel chain would allow to form a more precise view of their customer profiles, because such features as customers age, gender, citizenship, prices rented rooms etc, could also have high predictive value.