

# The Battle of Neighborhoods: Japan Hostels

## 1. Introduction

### 1.1 Background

As someone who loves travelling, finding ways to save money and travel low-cost is essential. There's little to save while booking flights and I'd rather pay more to traveling comfortably on a 16Hr flight than try to save and wear myself down even before the trip begins! But we can always save a few bucks on accommodation.

A hostel is a low-budget accommodation shared among travellers. Staying at hostels is an extremely affordable option for travellers. Not only that, It also gives you a chance to experience which you would never at a hotel. I've found that experiencing hostel life one of the best ways to travel the world on a budget and here's why:

1. You'll save a lot of money trust me.
2. Meet other like-minded travellers and make new friends around the world!
3. Most hostels organize lot of social activates

### 1.2 Business Understanding/Problem Description

Hostel accommodation is a booming industry with the latest rise in tourism worldwide, especially in Asia. Over 20% of the total tourists worldwide are aged between 19-25. And around 80% of them prefer to spend less on accommodation by opting for hostels. This means the demand for hostels is only going to grow further and more people would want to get their legs in the game.

- How should a new business person decide where to open a hostel?
- What factors should he look at before investing?
- Which neighborhood venues affect a user's rating for location of hostel?

At the same time, it is difficult for a travellers, especially first-timers, to select a hostel from among many options. Hostel reviews are subjective and differ from person-to-person and one cannot solely depend on them to make a decision. It is especially important to consider other aspects like price and neighborhood, which can greatly influence one's experience of the city/country. I will try to answer the following questions

- How does price vary with location?
- Where are the *value for money* hostels located?
- How does proximity to transportation affect hostel rating?
- Which hostels are most secure and where are they located?
- Suggest similar hostel but which cheaper price.

Tourism in Japan is on a rise. It is expected that the number of foreign tourists coming to Japan will be increasing till 2020 when Olympic will be held in Tokyo. Hence, for this project, we will be looking at hostels in Japan, in particular, Tokyo.

### 1.3 Target Audience

This project intends to serve two groups of audience:

1. **Travellers:** Help them make an informed decision while choosing a hostel by providing an in-depth analysis of hostels and their neighborhood.
2. **Business Person:** Provide useful information and models which can help them where to open their first/next hostel.

## 2. Data

I analysed in this project : Tokyo. Following are the datasets used in the project:

1. [Japan Hostel Dataset](#)
2. [Hostel Neighborhood](#)
3. [Tokyo Land Price](#)

### 2.1 Japan Hostel Dataset

The original dataset on Kaggle has the following columns:

- hostel.name: Hostel Name
- City: City name where hostel is located in
- price.from: Minimum Price for 1 night stay
- Distance: Distance from city centre (km)
- summary.score: Summary score of ratings
- rating.band: Rating band
- atmosphere: Rating score of atmosphere
- cleanliness: Rating score of cleanliness
- facilities: Rating score of facilities
- location: Rating score of location
- security: Rating score of security
- staff: Rating score of staff
- valueformoney: Rating score of value for money
- lon: Longitude
- lat: Latitude

Below is a snapshot of the dataset:

Name	City	StartPrice	DistanceFromCityCentre	OverallScore	RatingCategory	Atmosphere	Cleanliness	Facilities	Location	Security	Staff
"Bike & Bed" CharinCo Hostel	Osaka	3300	2.9	9.2	Superb	8.9	9.4	9.3	8.9	9.0	9.4
&And Hostel Akihabara	Tokyo	3600	7.8	8.7	Fabulous	8.0	7.0	9.0	8.0	10.0	10.0
&And Hostel Ueno	Tokyo	2600	8.7	7.4	Very Good	8.0	7.5	7.5	7.5	7.0	8.0
&And Hostel-Asakusa North-	Tokyo	1500	10.5	9.4	Superb	9.5	9.5	9.0	9.0	9.5	10.0
1night1980hostel Tokyo	Tokyo	2100	9.4	7.0	Very Good	5.5	8.0	6.0	6.0	8.5	8.5

## 2.2 Hostel Neighborhood:

This dataset contains all the neighborhoods or venues within 500m radius of a Hostel. It has the following columns:

- HostelName: Name of the hostel
- VenueName: Name of the venue
- Category: It is the primary category of the venue, for example, Café, Train Station, Restaurant.
- VenueLatitude, VenueLongitude: Coordinates of the venue.

Below is a snapshot of the dataset:

	HostelName	VenueName	Category	VenueLatitude	VenueLongitude
0	"Bike & Bed" CharinCo Hostel	Lawson (ローソン S OSL谷町四丁目駅北店)	Convenience Store	34.683282	135.517281
1	"Bike & Bed" CharinCo Hostel	モナコ	Café	34.683210	135.517309
2	"Bike & Bed" CharinCo Hostel	7-Eleven (セブンイレブン 大阪谷町3丁目店)	Convenience Store	34.683509	135.517793
3	"Bike & Bed" CharinCo Hostel	McDonald's (マクドナルド 地下鉄谷町四丁目店)	Fast Food Restaurant	34.683109	135.517443
4	"Bike & Bed" CharinCo Hostel	City Plaza Osaka (シティプラザ大阪)	Hotel	34.684020	135.510206

## 2.3 Tokyo Land Price:

This dataset contains the locality name and the average price of land per square meter. Below is a snapshot:

	Neighborhood	PricePerSqMeter
0	Chiyoda-Ku	1890610
1	Chuo-Ku	3178147
2	Minato-Ku	2339310
3	Shinjuku-Ku	930080
4	Bunkyo-Ku	966787

### 3. Methodology

#### 3.1 Data Collection:

- The Japan Hostels dataset is freely available on Kaggle and was built by scraping Hostelworld.com website.
- We used Foursquare API to get the venues around the hostel.
- We scraped <https://utinokati.com> to get land prices of various neighborhoods in Tokyo.

#### 3.2 Analytic Approach:

I took two approaches in the project.

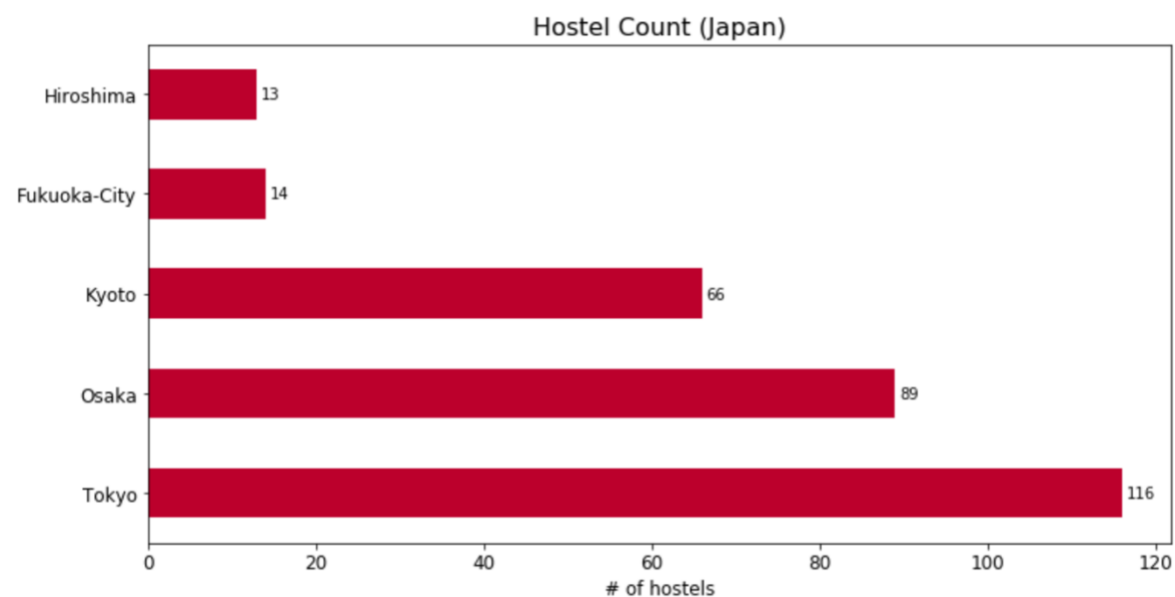
Firstly, I used exploratory data analysis(EDA) to uncover hidden properties of data and provide useful insights to the reader, both future traveller and investor. I used the list of hostels from Hostel dataset and use Foursquare API to get venues around the Hostel. I will then use EDA to explore the neighborhood and how it affects the price of the hostel. I will also use the combined dataset to cluster similar hostels as per pricing and neighborhood.

Secondly, I used prescriptive analytics to help a business person decide a location for new hostel. I will use clustering(K-Means). I combined the above data with the land price for the area in which the Hostel is situated and then develop clustering models to predict where a new hostel should be opened.

### 4. Analysis

#### 4.1 Exploratory Data Analysis

First we tried to see how many hostels were present in our dataset per city.

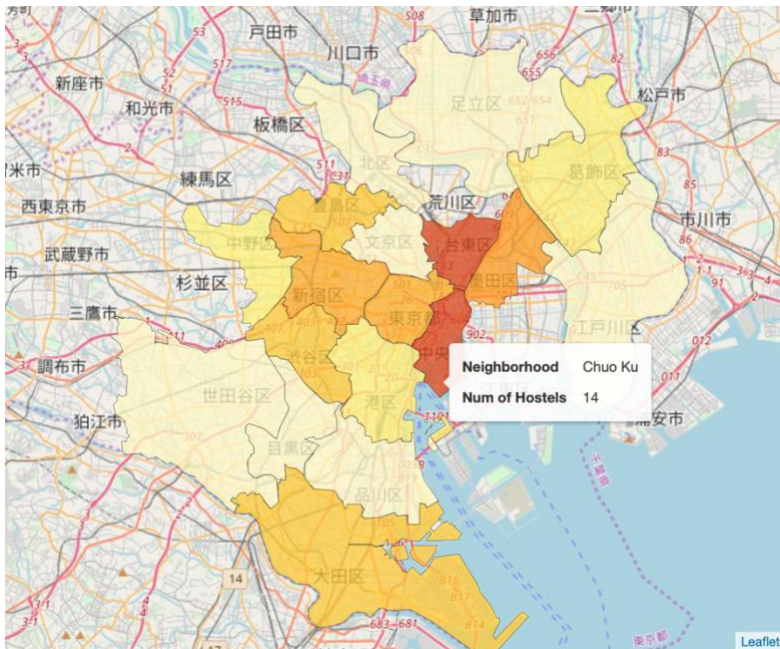


Tokyo has the most number of hostels which is understandable since it is one of the largest cities of the world and hence might attract a lot of visitors. Also, the two airports in Tokyo make it an important hub in the pacific rim.

Next, we visualized all the hostels on map using Folium and Open Street Maps. Below is the geo-visualization of the hostels in Tokyo. As we can see, many hostels are located near the centre with density reducing as we move away from it. In the report we have visualized the same map based on multiple criteria and clusters.



For a person interested in opening a new hostels, it is essential to identify the right area to do so. The main factors to consider while making such a decision is demand, supply and cost. Demand and supply usually go hand-in-hand in such cases i.e. we can say that the more the



Neighborhood	Count	PricePerSqMeter
Edogawa	1	321519
Meguro	1	845562
Adachi	2	291044
Setagaya	2	653804
Shinagawa	2	742731
Bunkyo	2	966787
Nakano	3	599453
Minato	3	2339310
Ota	4	540823
Toshima	4	689177
Shibuya	4	1328586
Chiyoda	8	1890610
Sumida	11	586675
Chuo	14	3178147
Taito	39	1028047

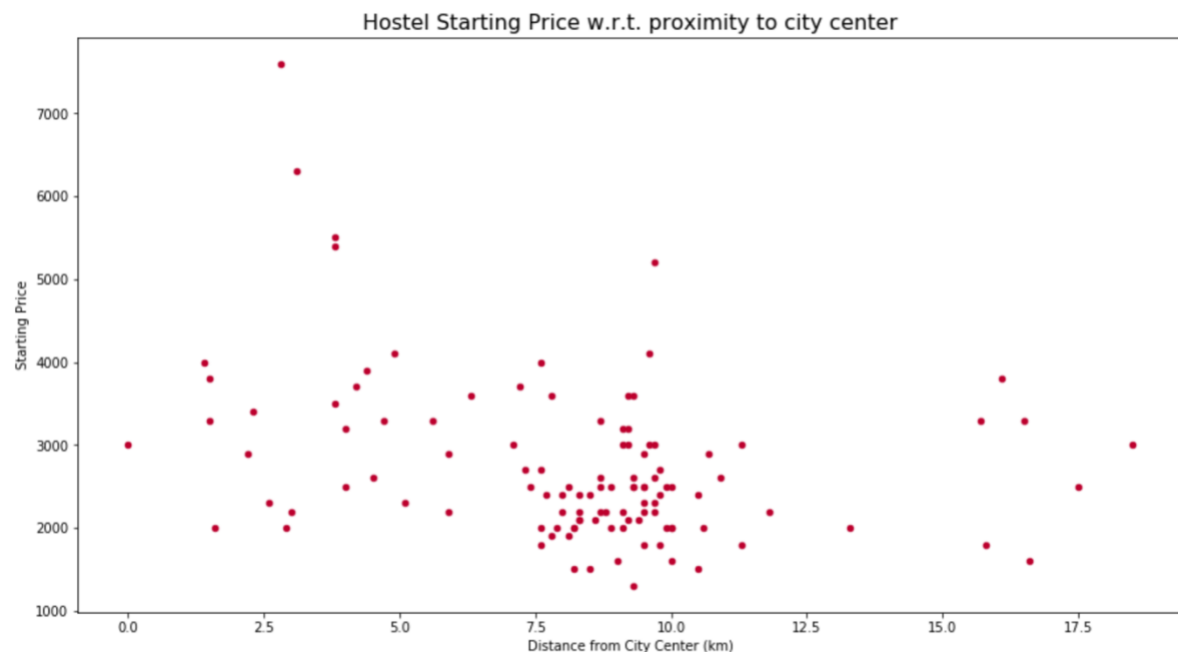
hostels in a region, the higher the demand. It is evident from the above map that **Taito** and **Chuo** are host to many hostels. Hostel density is highest in these neighborhoods. However, if we compare the land prices in these regions, they are amongst the most costly localities to buy property.



Sumida seems to be popular since it ranks third in the list of number of hostels and at the same time, it is the 4th cheapest neighborhood in our list.

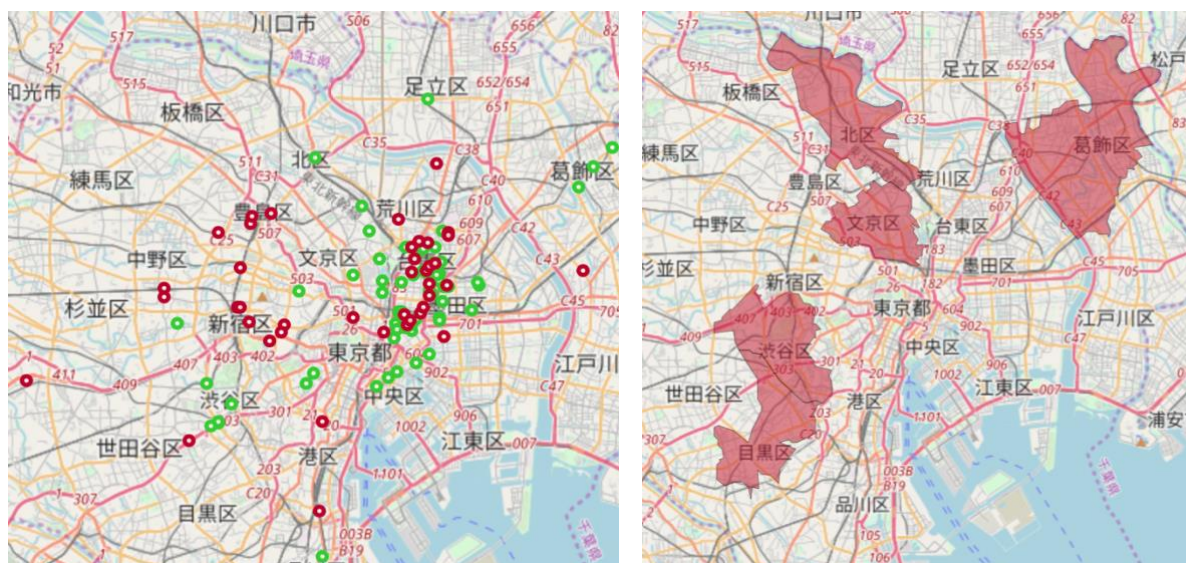
The price of property(per sq. meter) in Sumida is almost 43% less than that in Taito, which combined with the fact that it is close to Taito and has fewer hostels than Taito, makes it an exciting prospect for new investors.

Next, we checked if the distance of a hostel from the city centre has any effect on the price.

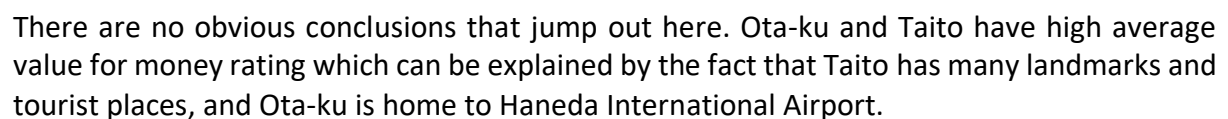


We can see a very weak negative correlation between the price and distance of hostel from city centre. Our intuition is supported mathematically by the Pearson coefficient which turned out to be -0.3. The correlation coefficient is very less to make any strong inferences.

Next we analysed hostels based on their “security” ratings.



Subsequently, we studied similar patterns in “value for money” rating and location of hostels.

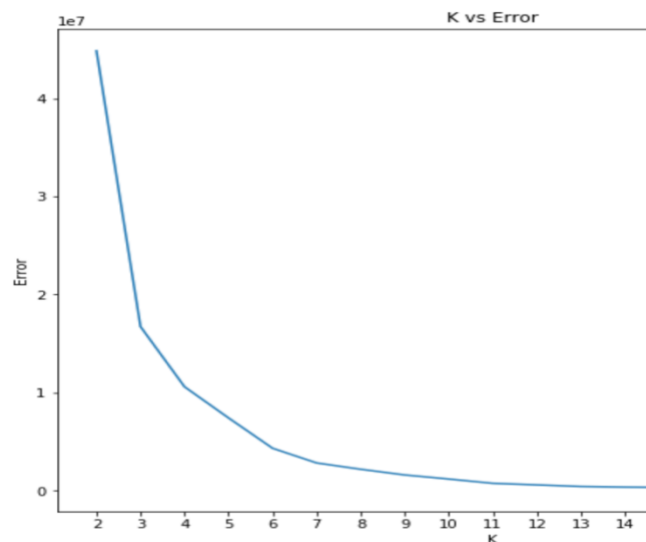


- Proximity to a mode of transportation seemed to greatly affected the ratings, since 50% of hostels near to a bus/metro station were rated Fabulous (highest rating)
- Proximity to convenience store did not have any effect.
- Hostels near museums and malls had a lower rating than the overall average. Only 25% of hostels in such areas were rated high compared to 32% overall.



## 4.2 Clustering

We performed clustering twice based on different set of parameters. First, we clustered using the different rating scores, distance from city centre, and starting price. We used K-Means clustering algorithm and found out the K by using the elbow method. The K on our case is 6, since the error doesn't decrease much after this point.



Let's see a geo-visualization of the clusters and also examine the properties of each cluster.



### Cluster 0:

Very High Cost. Close to City centre.  
Very high cleanliness and security ratings

### Cluster 1:

Medium cost. Far from city centre.  
High Value for Money score.

### Cluster 2:

Low Cost. Average overall score between 8.5 and 9. Low atmosphere rating.

### Cluster 3:

Low Cost. Close to city centre. Low facilities and staff score.

### Cluster 4:

High Cost, On average 7km of City centre. All ratings moderately high

### Cluster 5:

Very low cost. Far away from city centre. High value for money and location ratings.



Second, we clustered the hostels based on the venues in its vicinity. This time, we fixed K in our K-Means algorithm to be 4 since otherwise each hostel would be assigned a unique cluster which defeats the purpose.



We can examine each cluster and determine the discriminating venue categories that distinguish each cluster. Based on the defining categories, we can then assign a name to each cluster.

<i>Cluster Number</i>	<i>Cluster Properties</i>
1	Restaurants and Coffee Shops
2	Convenience Stores
3	Baseball Field and Gym Pool
4	Bar and Restaurants

## 5. Result

We got a glimpse of the hostel scene in Tokyo and were able to find out some interesting insights which might be useful to travellers as well as people with business interests. Let's summarize our findings:

- Most hostels are located in Taito-ku and Chuo-ku.
- Sumida-ku seems to be an interesting locality since it is close to Taito-ku and 43% cheaper than Taito-ku.
- The starting price of hostels does not vary much depending on its distance from the city centre.
- Most of the hostels rated high for their security are in Katsushika, Kita, Meguro, Shibuya and Shinagawa
- Proximity to a mode of transportation or a historic site positively affects the hostel rating.
- Hostels rated highly for being value for money are comparatively cheaper and are located away from the city centre

## 6. Discussion

According to the above analysis, opening a hostel in Sumida-ku is the most appropriate option. Taito and Chuo contain the highest number of hostels from our dataset. This means that these wards must attract lot of tourists, no other reason appropriately justifies why they house so many hostels inspite of being very costly. Sumida is a neighbour of Taito and is almost 41% cheaper than Taito. It also contains decent amount of hostels but not as many as Taito or Chuo, which leaves space for new opportunities.

The clusters will help tourists identify alternate hostels in case their hostel is not available for some reason. For example, I wanted to book a spot at Retrometro Backpackers. However, when I went to the website, it said that all the beds were booked and there was no availability during my dates. From our analysis, Retrometro belongs to Cluster 5. So I checked which other hostels are in that cluster. I found availability at Hostel bedgasm! It was similarly priced and rated as Retrometro Backpackers.

Some drawbacks of the analyses are that our prescription to new business person's for opening a new hostel is solely based on neighborhood and land price and not on other factors like how much the business person is willing to invest, what facilities will he provide, how will he price the hostel. Without this data, it is difficult to predict the success of the investment. However, in our analysis, we have ignored this since we don't have such data and it would be difficult to farm it for a small exploratory study like ours. Hence, our analysis only helps a business person identify a region to open a hostel but doesn't guarantee its success.

Also, it would have been beneficial if we had additional features such as crime rate in the locality, and average number of tourists in the locality. This would give us a more complete picture of the neighborhood of a hostel resulting in better analysis.

## **7. Conclusion**

In the above study, we explored and analysed various aspects of the hostels scene in Tokyo, Japan using data science. We used an existing dataset and combined it with data collected from Foursquare API as well as data scraped from a website. We performed EDA and clustering on these datasets in our pursuit of solutions. We were able to find satisfactory answers to the questions we posed before the study.

The study is based on limited data, but it is nevertheless a significant step in shedding light on the hostel scene in Tokyo. This study can be repeated easily for other cities of the world.