

# Japan Hostels: CPI Analysis

## 1. Introduction

### 1.1 Background

Finding the best low cost hostels can save your budget tremendously. Besides saving on flights much of travel costs come from accommodations. People who choose hostels instead of 5-star hotels can allocate their money towards better experience when travelling.

A hostel is a low-budget accommodation shared among travelers but some also come with your own private space. Staying at hostels is an extremely affordable option for travelers. Hostels gives you a chance to experience something you would never at a hotel. You get to make new friends and join activities that you would otherwise have to explore on your own. Therefore, in this study we would explore the ways in which we can find a good budget hostel in Tokyo.

### 1.2 Business Understanding/Problem Description

The accommodation industry is a booming industry with the latest rise in tourism worldwide and Airbnb. This means the demand for hostels is only going to grow further and more people would want to get in the game.

- How should an investor choose a location to open a hostel?
- What are the factors that influence decision making?
- Which neighborhood venues affect a hostel's ratings?

Choosing a hostel from the many options out there can be a daunting task for first time travelers or investors. 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 change with location?
- Where are the *CPI* hostels located?
- How does proximity to transportation affect hostels?
- Which hostels are most secure and where are they located?
- Suggest similar hostel but which cheaper price

## 1.3 Target Audience

This study intends to serve two groups of audience:

1. **Tourists:** Help make better informed decision when choosing a hostel by providing an in-depth analysis of hostels and their CPI.
2. **Investors:** Provide useful information and models which can help in decision making

## 2. Data

I analyzed 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 hostel dataset has the following values:

1. hostel.name: Hostel Name
2. City: City name where hostel is located in
3. price.from: Minimum Price for 1 night stay
4. Distance: Distance from city centre (km)
5. summary.score: Summary score of ratings
6. rating.band: Rating band
7. atmosphere: Rating score of atmosphere
8. cleanliness: Rating score of cleanliness
9. facilities: Rating score of facilities
10. location: Rating score of location
11. security: Rating score of security
12. staff: Rating score of staff
13. valueformoney: Rating score of value for money
14. lon: Longitude
15. 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:

1. HostelName: Name of the hostel
2. VenueName: Name of the venue
3. Category: It is the primary category of the venue, for example, Café, Train Station, Restaurant.
4. 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
- Used Foursquare API to get the venues around the hostel.
- Scraped <https://utinokati.com> to get land prices of various neighborhoods in Tokyo.

#### 3.2 Analytic Approach:

Two analytic approaches was used in this study

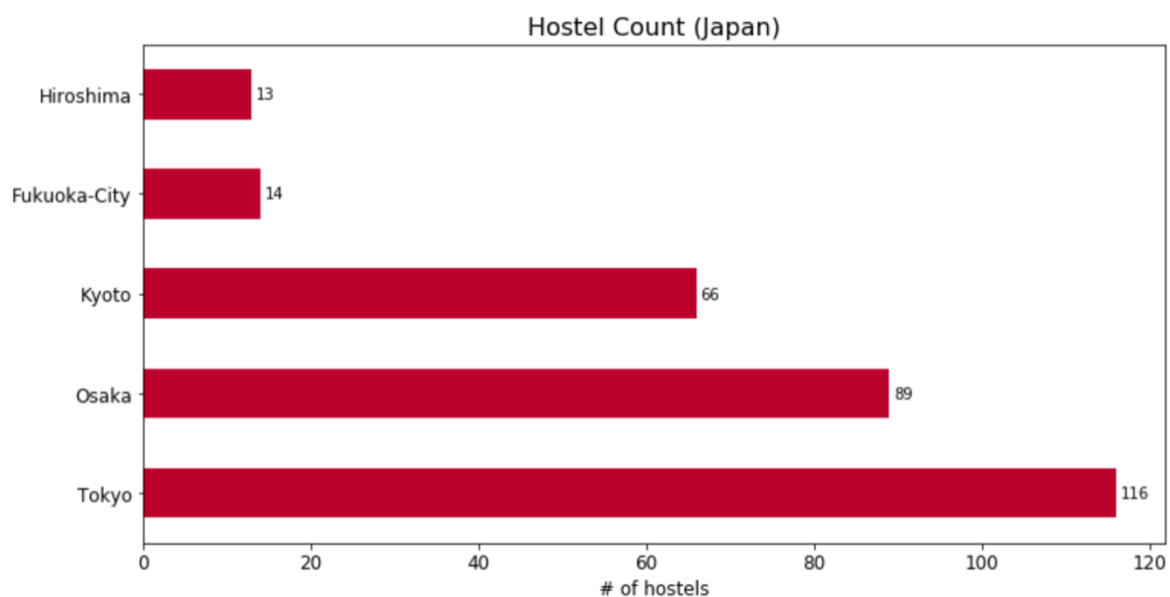
First, exploratory data analysis (EDA) was used to provide useful insights to the reader, travelers and investors. I used the list of hostels from Hostel dataset and use Foursquare API to get venues around the Hostel. I then used EDA to explore the neighborhood and examine its effect on hostel prices. I will also use the combined dataset to cluster similar hostels as per pricing and neighborhood.

Second, 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.

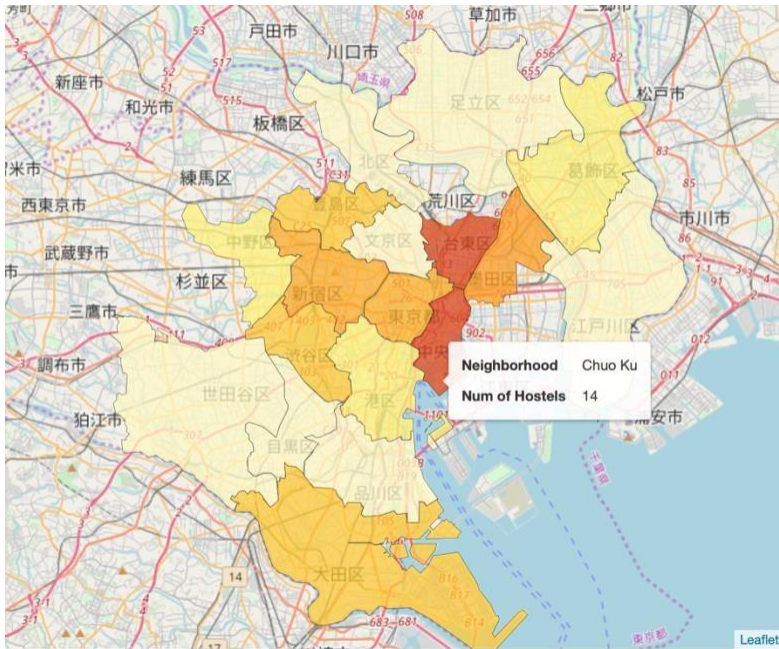


Tokyo has the greatest number of hostels since it is one of the largest cities in the world and attract a lot of visitors. Also, the two airports in Tokyo make it an important hub in the pacific rim.

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

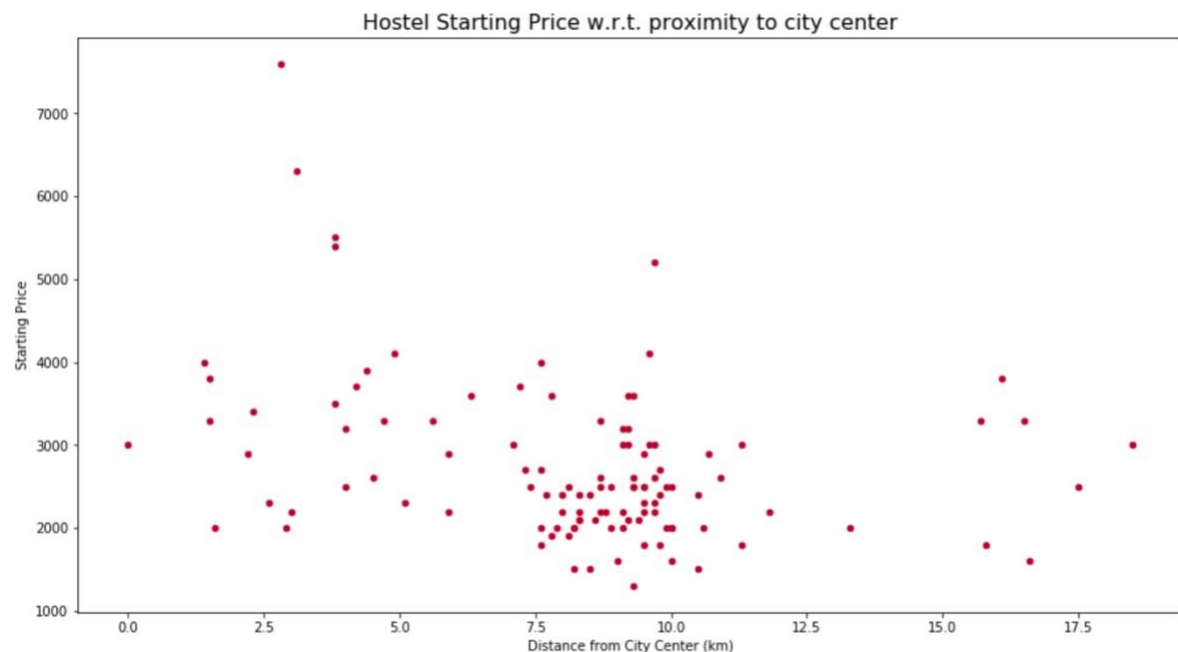
hostels in a region, the higher the demand. It is evident from the above map that **Taito** and **Chuo** are host to many hostels. Hostel density is highest in these neighborhoods. However, if we compare the land prices in these regions, they are amongst the most costly localities to buy property.



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.

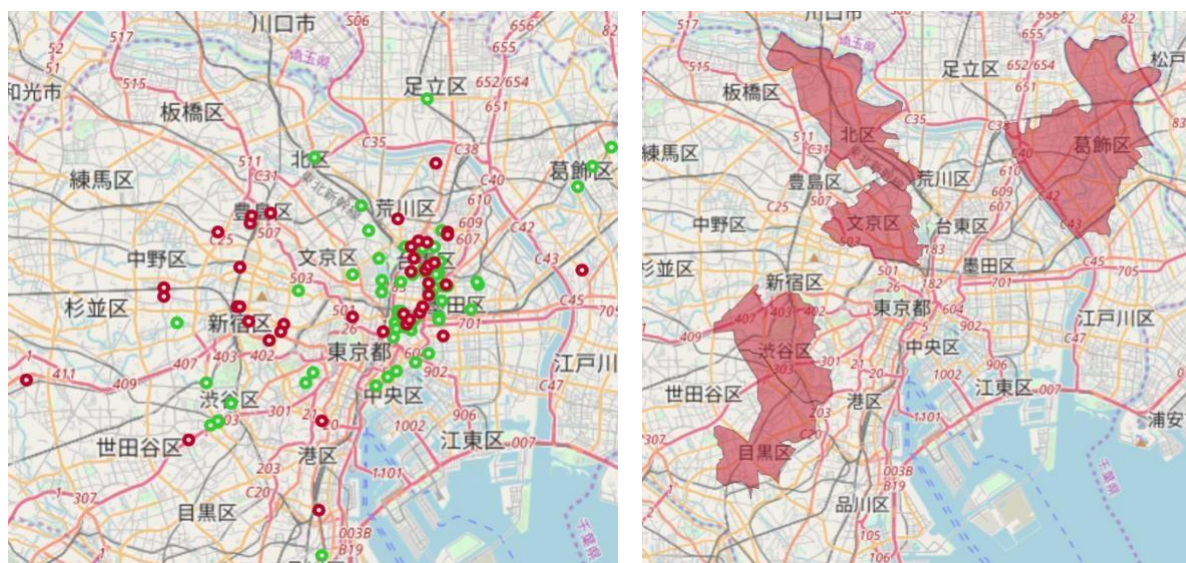
The price of property(per sq. meter) in Sumida is almost 43% less than that in Taito, which combined with the fact that it is close to Taito and has fewer hostels than Taito, makes it an exciting prospect for new investors.

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



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

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

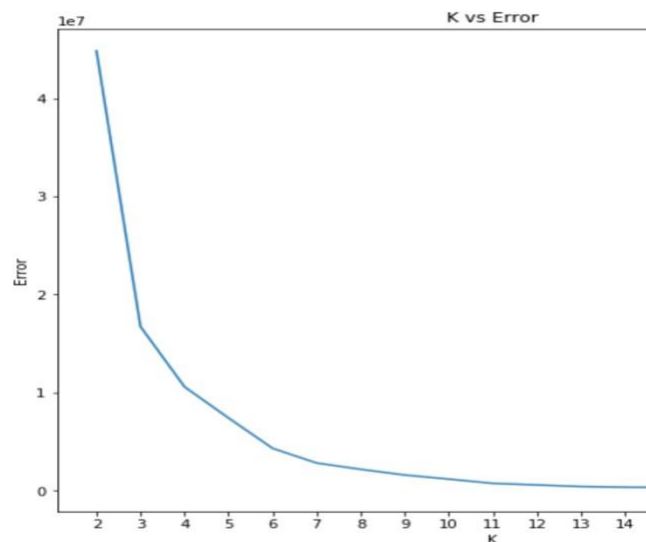






## 4.2 Clustering

We performed clustering twice based on different set of parameters. First, we clustered using the different rating scores, distance from city center, 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.

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

## 5. Result

After analyzing the data we have a summary of findings listed below which may be useful for travelers or investors:

- The highest number of hostels are located in Taito-ku and Chuo-ku.
- Sumida-ku is an outlier since it is close to Taito-ku and 43% cheaper than Taito-ku.
- The price of hostels does depend 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 transportation or historic site positively affects the hostel rating.
- Price and distance from city center influence hostels rating for value greatly.

## 6. Discussion

According to the above summary analysis, in terms of supply and demand, 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. Sumida is a neighbor 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.

Some draw backs of this analysis is its usefulness for investors looking to open a hostel in Tokyo as our prescription to new investor for opening a new hostel is solely based on neighborhood and land price. For future improvement we can include more data that can help in determining the success of a hostel investment by introducing more datasets such as number of bars and restaurants. Therefore, our analysis is more focused on helping a investor identify a region that shows greater potential for opening a hostel

## 7. Conclusion

In conclusion, using existing dataset combined with data collected from Foursquare API and data scraped from the web we have identified potential areas for high CPI hostels. We performed EDA and clustering methodology to arrive at the given solutions. If travelers were to visit Tokyo it would be wise to chose hostels with a high CPI such as those away from city center and those with high security ratings. If travelers are more transit cautious then a hostels in the Chuo-ku area are still viable after weighing in time as a value of money.

