1. Introduction/Business Understanding

1.1 Problem Description:

Over 20% of the tourists worldwide are aged between 19-25. 80% of them prefer to spend less on accommodation by opting for hostels. This means the demand for hostels is bound to grow everyday (this does not necessarily apply to abnormal years, like 2020). It is, however, interesting to ask these questions:

- How should a new businessperson 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 traveler, 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 decide. 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 attempt to answer the following questions:

- How does price vary with location?
- Where are the *'value of 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

For this project, we will be looking at hostels in Tokyo.

1.2 Target Audience

This project will serve two groups of audience:

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

2. Analytic Approach

I will be taking two approaches in the project:

1. I will use **exploratory data analysis** *(EDA)* to uncover hidden properties of data and provide useful insights to the reader, both future traveller and investor.
2. I will use **prescriptive analytics** to help a business person decide a location for new hostel. I will use *clustering* (K-Means)

3. Data requirements

Following are the datasets used in the project along with the reasons for choosing them:

1. **Japan Hostel Dataset**: This dataset is available on Kaggle. Provides information about hostels in Japan.

2. **Foursquare API**: This API will provide the venues around the hostel which I will use for EDA and clustering.

3. **Tokyo Land Price**: I will scrape this website to get land prices of various neighborhoods in Tokyo.

Firstly, I will use 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.

Second, I will combine the above data with the land price for the area in which the Hostel is situated and them develop clustering and regression models to predict where a new hostel should be opened and how much should it be priced at.