An Analysis on Hostels in Japan

November 27, 2022

1 An Analysis on Hostel's in Japan

This dataset is taken from Kaggle.com and contains data on over 300 hostels in Japan. This dataset includes things like which city the hostel's are located in, the price per night, the distance from the city center, and many different ratings. I would like to know things like how much the price varies from place to place and figure out what are possible reasons if there are noticeable differences.

I will be using pandas to tidy the dataset and use numpy, matplotlib, and seaborn to visual the data.

[ ]:
```
import pandas as pd
```

[ ]:
```
df = pd.read_csv(r'C:\Users\Brendan␣
↪Sakihara\Desktop\Coding\Python\Projects\Jovian Project\Hostel.csv')
```

[ ]:
```
df
```

[ ]:
```
     Unnamed: 0                 hostel.name         City  price.from  \
0             1  "Bike & Bed" CharinCo Hostel        Osaka        3300
1             2                 & And Hostel  Fukuoka-City        2600
2             3        &And Hostel Akihabara        Tokyo        3600
3             4             &And Hostel Ueno        Tokyo        2600
4             5     &And Hostel-Asakusa North-        Tokyo        1500
..          …                           …            …          …
337         338       YADOYA Guesthouse Green        Tokyo        2300
338         339      YADOYA Guesthouse Orange        Tokyo        2000
339         340             YAWP! backpackers        Tokyo        2500
340         341               You En Me House        Kyoto        2800
341         342               Zabutton Hostel        Tokyo        2900

                 Distance  summary.score rating.band  atmosphere  \
0     2.9km from city centre            9.2      Superb         8.9
1     0.7km from city centre            9.5      Superb         9.4
2     7.8km from city centre            8.7    Fabulous         8.0
3     8.7km from city centre            7.4   Very Good         8.0
4    10.5km from city centre            9.4      Superb         9.5
..                       …             …           …           …
337   2.6km from city centre            8.2    Fabulous         7.9
```

```
338   2.9km from city centre           8.9    Fabulous         8.6
339  17.5km from city centre           9.3      Superb         9.5
340   2.4km from city centre           8.0    Fabulous         7.3
341   5.9km from city centre           8.6    Fabulous         8.1

     cleanliness  facilities  location.y  security  staff  valueformoney  \
0            9.4         9.3         8.9       9.0    9.4            9.4
1            9.7         9.5         9.7       9.2    9.7            9.5
2            7.0         9.0         8.0      10.0   10.0            9.0
3            7.5         7.5         7.5       7.0    8.0            6.5
4            9.5         9.0         9.0       9.5   10.0            9.5
..           …           …           …         …      …              …
337          7.7         6.9         8.9       8.9    8.8            8.3
338          9.0         7.8         9.4       9.0    9.2            9.4
339          9.3         9.4         8.5       9.5    9.2            9.6
340          8.0         6.7         8.0       8.7   10.0            7.3
341          8.5         7.8         9.0       8.9    9.5            8.5

            lon        lat
0    135.513767  34.682678
1           NaN        NaN
2    139.777472  35.697447
3    139.783667  35.712716
4    139.798371  35.727898
..          …          …
337  139.668125  35.702908
338  139.667695  35.706513
339  139.869197  35.752885
340  135.749063  34.997376
341  139.742116  35.655470

[342 rows x 16 columns]
```

[ ]:
```
print(df)
```

```
     Unnamed: 0                   hostel.name         City  price.from  \
0             1  "Bike & Bed" CharinCo Hostel        Osaka        3300
1             2                 & And Hostel  Fukuoka-City        2600
2             3         &And Hostel Akihabara        Tokyo        3600
3             4              &And Hostel Ueno        Tokyo        2600
4             5    &And Hostel-Asakusa North-        Tokyo        1500
..          …                           …          …         …
337         338      YADOYA Guesthouse Green        Tokyo        2300
338         339     YADOYA Guesthouse Orange        Tokyo        2000
339         340            YAWP! backpackers        Tokyo        2500
340         341              You En Me House        Kyoto        2800
341         342               Zabutton Hostel        Tokyo        2900
```

```
                  Distance  summary.score rating.band  atmosphere  \
0     2.9km from city centre            9.2      Superb         8.9
1     0.7km from city centre            9.5      Superb         9.4
2     7.8km from city centre            8.7    Fabulous         8.0
3     8.7km from city centre            7.4   Very Good         8.0
4    10.5km from city centre            9.4      Superb         9.5
..                       …             …         …          …
337   2.6km from city centre            8.2    Fabulous         7.9
338   2.9km from city centre            8.9    Fabulous         8.6
339  17.5km from city centre            9.3      Superb         9.5
340   2.4km from city centre            8.0    Fabulous         7.3
341   5.9km from city centre            8.6    Fabulous         8.1

     cleanliness  facilities  location.y  security  staff  valueformoney  \
0            9.4         9.3         8.9       9.0    9.4            9.4
1            9.7         9.5         9.7       9.2    9.7            9.5
2            7.0         9.0         8.0      10.0   10.0            9.0
3            7.5         7.5         7.5       7.0    8.0            6.5
4            9.5         9.0         9.0       9.5   10.0            9.5
..           …          …          …        …     …            …
337          7.7         6.9         8.9       8.9    8.8            8.3
338          9.0         7.8         9.4       9.0    9.2            9.4
339          9.3         9.4         8.5       9.5    9.2            9.6
340          8.0         6.7         8.0       8.7   10.0            7.3
341          8.5         7.8         9.0       8.9    9.5            8.5

            lon        lat
0    135.513767  34.682678
1           NaN        NaN
2    139.777472  35.697447
3    139.783667  35.712716
4    139.798371  35.727898
..          …          …
337  139.668125  35.702908
338  139.667695  35.706513
339  139.869197  35.752885
340  135.749063  34.997376
341  139.742116  35.655470

[342 rows x 16 columns]
```

[ ]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 342 entries, 0 to 341
Data columns (total 16 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
```

```
 0   Unnamed: 0     342 non-null    int64
 1   hostel.name    342 non-null    object
 2   City           342 non-null    object
 3   price.from     342 non-null    int64
 4   Distance       342 non-null    object
 5   summary.score  327 non-null    float64
 6   rating.band    327 non-null    object
 7   atmosphere     327 non-null    float64
 8   cleanliness    327 non-null    float64
 9   facilities     327 non-null    float64
 10  location.y     327 non-null    float64
 11  security       327 non-null    float64
 12  staff          327 non-null    float64
 13  valueformoney  327 non-null    float64
 14  lon            298 non-null    float64
 15  lat            298 non-null    float64
dtypes: float64(10), int64(2), object(4)
memory usage: 42.9+ KB
```

Key:
hostel.name = name of the hostel
City = the city of the hostel
price.from = the price per one night stay
Distance = the distance from the center of the city in km
summary.score = the average rating based on all the ratings
rating.band = the rating in word form
atmosphere = the rating for the atmosphere of the hostel
cleanliness = the rating for how clean the hostel is
facilities = the rating of the facilities in the hostel
location.y = the location rating of the hostel
security = the rating of the security of the hostel
staff = the rating of the staff of the hostel
valueformoney = the rating of the value of the hostel based on the price

[ ]:
```
df = pd.read_csv(r'C:\Users\Brendan␣
↪Sakihara\Desktop\Coding\Python\Projects\Jovian Project\Hostel.csv',␣
↪index_col=0) #re-importing the dataset without the first column
```

[ ]:
```
df
```

[ ]:
```
                          hostel.name          City  price.from  \
1     "Bike & Bed" CharinCo Hostel         Osaka        3300
2                        & And Hostel  Fukuoka-City        2600
3               &And Hostel Akihabara         Tokyo        3600
4                    &And Hostel Ueno         Tokyo        2600
5         &And Hostel-Asakusa North-         Tokyo        1500
..                                ...           ...         ...
338          YADOYA Guesthouse Green         Tokyo        2300
```

```
339      YADOYA Guesthouse Orange         Tokyo        2000
340             YAWP! backpackers         Tokyo        2500
341               You En Me House         Kyoto        2800
342               Zabutton Hostel         Tokyo        2900

                     Distance  summary.score rating.band  atmosphere  \
1     2.9km from city centre            9.2      Superb         8.9
2     0.7km from city centre            9.5      Superb         9.4
3     7.8km from city centre            8.7    Fabulous         8.0
4     8.7km from city centre            7.4   Very Good         8.0
5    10.5km from city centre            9.4      Superb         9.5
..                        …             …          …           …
338   2.6km from city centre            8.2    Fabulous         7.9
339   2.9km from city centre            8.9    Fabulous         8.6
340  17.5km from city centre            9.3      Superb         9.5
341   2.4km from city centre            8.0    Fabulous         7.3
342   5.9km from city centre            8.6    Fabulous         8.1

     cleanliness  facilities  location.y  security  staff  valueformoney  \
1            9.4         9.3         8.9       9.0    9.4            9.4
2            9.7         9.5         9.7       9.2    9.7            9.5
3            7.0         9.0         8.0      10.0   10.0            9.0
4            7.5         7.5         7.5       7.0    8.0            6.5
5            9.5         9.0         9.0       9.5   10.0            9.5
..             …          …          …       …      …              …
338          7.7         6.9         8.9       8.9    8.8            8.3
339          9.0         7.8         9.4       9.0    9.2            9.4
340          9.3         9.4         8.5       9.5    9.2            9.6
341          8.0         6.7         8.0       8.7   10.0            7.3
342          8.5         7.8         9.0       8.9    9.5            8.5

            lon        lat
1    135.513767  34.682678
2           NaN        NaN
3    139.777472  35.697447
4    139.783667  35.712716
5    139.798371  35.727898
..          …          …
338  139.668125  35.702908
339  139.667695  35.706513
340  139.869197  35.752885
341  135.749063  34.997376
342  139.742116  35.655470

[342 rows x 15 columns]
```

I won't be using the longitude and latitude information so I will drop it from the dataset.

```
[ ]: df_edit = df.copy() #making copies in case I need to go back
```

```
[ ]: df_nolatlong = df_edit.drop(['lon','lat'], axis = 1)
     df_nolatlong
```

```
[ ]:                       hostel.name          City  price.from  \
     1     "Bike & Bed" CharinCo Hostel         Osaka        3300
     2                   & And Hostel  Fukuoka-City        2600
     3            &And Hostel Akihabara         Tokyo        3600
     4                 &And Hostel Ueno         Tokyo        2600
     5      &And Hostel-Asakusa North-         Tokyo        1500
     ..                             …             …          …
     338       YADOYA Guesthouse Green         Tokyo        2300
     339      YADOYA Guesthouse Orange         Tokyo        2000
     340             YAWP! backpackers         Tokyo        2500
     341               You En Me House         Kyoto        2800
     342               Zabutton Hostel         Tokyo        2900

                         Distance  summary.score rating.band  atmosphere  \
     1     2.9km from city centre            9.2      Superb         8.9
     2     0.7km from city centre            9.5      Superb         9.4
     3     7.8km from city centre            8.7    Fabulous         8.0
     4     8.7km from city centre            7.4   Very Good         8.0
     5    10.5km from city centre            9.4      Superb         9.5
     ..                       …            …           …          …
     338   2.6km from city centre            8.2    Fabulous         7.9
     339   2.9km from city centre            8.9    Fabulous         8.6
     340  17.5km from city centre            9.3      Superb         9.5
     341   2.4km from city centre            8.0    Fabulous         7.3
     342   5.9km from city centre            8.6    Fabulous         8.1

          cleanliness  facilities  location.y  security  staff  valueformoney
     1            9.4         9.3         8.9       9.0    9.4            9.4
     2            9.7         9.5         9.7       9.2    9.7            9.5
     3            7.0         9.0         8.0      10.0   10.0            9.0
     4            7.5         7.5         7.5       7.0    8.0            6.5
     5            9.5         9.0         9.0       9.5   10.0            9.5
     ..           …         …         …       …    …            …
     338          7.7         6.9         8.9       8.9    8.8            8.3
     339          9.0         7.8         9.4       9.0    9.2            9.4
     340          9.3         9.4         8.5       9.5    9.2            9.6
     341          8.0         6.7         8.0       8.7   10.0            7.3
     342          8.5         7.8         9.0       8.9    9.5            8.5

     [342 rows x 13 columns]
```

Some hostel's had values that included NA so I decided to drop them as well.

```
[ ]: df_clean = df_nolatlong.copy()
```

```
[ ]: df_clean = df_clean.dropna()
```

I also noticed that in distance, it has 'km from city centre' after the number so I want to get rid of that so that they are just numbers using the str.replace method.

```
[ ]: df_clean["Distance"] = df_clean["Distance"].str.replace('km from city centre',␣
     ↪'')
```

Since distance is still an object due it having the text previously, I will convert it into a float.

```
[ ]: df_clean["Distance"] = df_clean.Distance.astype(float)
```

```
[ ]: df_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 327 entries, 1 to 342
Data columns (total 13 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   hostel.name     327 non-null    object
 1   City            327 non-null    object
 2   price.from      327 non-null    int64
 3   Distance        327 non-null    float64
 4   summary.score   327 non-null    float64
 5   rating.band     327 non-null    object
 6   atmosphere      327 non-null    float64
 7   cleanliness     327 non-null    float64
 8   facilities      327 non-null    float64
 9   location.y      327 non-null    float64
 10  security        327 non-null    float64
 11  staff           327 non-null    float64
 12  valueformoney   327 non-null    float64
dtypes: float64(9), int64(1), object(3)
memory usage: 35.8+ KB
```

I want to rename some of the columns to a more simpler name using the rename method.

```
[ ]: df_clean = df_clean.rename(columns={'hostel.name': 'hostel', 'City': 'city',␣
     ↪'price.from': 'price', 'Distance': 'distance', 'summary.score': 'summary',␣
     ↪'rating.band': 'rating', 'location.y': 'location_rating', 'valueformoney':␣
     ↪'value'})
```

```
[ ]: df_clean
```

```
[ ]:                               hostel         city  price  distance  summary  \
     1    "Bike & Bed" CharinCo Hostel        Osaka   3300       2.9      9.2
     2                    & And Hostel  Fukuoka-City   2600       0.7      9.5
```

```
3           &And Hostel Akihabara         Tokyo   3600       7.8       8.7
4                &And Hostel Ueno         Tokyo   2600       8.7       7.4
5      &And Hostel-Asakusa North-         Tokyo   1500      10.5       9.4
..                            …           …    …       …       …
338        YADOYA Guesthouse Green        Tokyo   2300       2.6       8.2
339       YADOYA Guesthouse Orange        Tokyo   2000       2.9       8.9
340              YAWP! backpackers        Tokyo   2500      17.5       9.3
341                You En Me House        Kyoto   2800       2.4       8.0
342                Zabutton Hostel        Tokyo   2900       5.9       8.6

        rating  atmosphere  cleanliness  facilities  location_rating  \
1       Superb         8.9          9.4         9.3              8.9
2       Superb         9.4          9.7         9.5              9.7
3     Fabulous         8.0          7.0         9.0              8.0
4    Very Good         8.0          7.5         7.5              7.5
5       Superb         9.5          9.5         9.0              9.0
..         …          …           …          …               …
338   Fabulous         7.9          7.7         6.9              8.9
339   Fabulous         8.6          9.0         7.8              9.4
340     Superb         9.5          9.3         9.4              8.5
341   Fabulous         7.3          8.0         6.7              8.0
342   Fabulous         8.1          8.5         7.8              9.0

     security  staff  value
1         9.0    9.4    9.4
2         9.2    9.7    9.5
3        10.0   10.0    9.0
4         7.0    8.0    6.5
5         9.5   10.0    9.5
..        …      …      …
338       8.9    8.8    8.3
339       9.0    9.2    9.4
340       9.5    9.2    9.6
341       8.7   10.0    7.3
342       8.9    9.5    8.5

[327 rows x 13 columns]
```

I noticed later on that Fukuoka was the only city that has '-City' at the end of it so I want to get rid of that for consistency using the str.replace method.

[ ]:
```
df_clean["city"] = df_clean["city"].str.replace('-City', '')
```

[ ]:
```
df_clean
```

[ ]:
```
                          hostel      city  price  distance  summary  \
1    "Bike & Bed" CharinCo Hostel     Osaka   3300       2.9      9.2
2                   & And Hostel   Fukuoka   2600       0.7      9.5
```

```
3          &And Hostel Akihabara    Tokyo   3600       7.8       8.7
4                &And Hostel Ueno    Tokyo   2600       8.7       7.4
5       &And Hostel-Asakusa North-    Tokyo   1500      10.5       9.4
..                              …        …     …          …         …
338        YADOYA Guesthouse Green    Tokyo   2300       2.6       8.2
339       YADOYA Guesthouse Orange    Tokyo   2000       2.9       8.9
340              YAWP! backpackers    Tokyo   2500      17.5       9.3
341                You En Me House    Kyoto   2800       2.4       8.0
342                Zabutton Hostel    Tokyo   2900       5.9       8.6

        rating  atmosphere  cleanliness  facilities  location_rating  \
1       Superb         8.9          9.4         9.3              8.9
2       Superb         9.4          9.7         9.5              9.7
3     Fabulous         8.0          7.0         9.0              8.0
4    Very Good         8.0          7.5         7.5              7.5
5       Superb         9.5          9.5         9.0              9.0
..         …          …           …          …               …
338   Fabulous         7.9          7.7         6.9              8.9
339   Fabulous         8.6          9.0         7.8              9.4
340     Superb         9.5          9.3         9.4              8.5
341   Fabulous         7.3          8.0         6.7              8.0
342   Fabulous         8.1          8.5         7.8              9.0

     security  staff  value
1         9.0    9.4    9.4
2         9.2    9.7    9.5
3        10.0   10.0    9.0
4         7.0    8.0    6.5
5         9.5   10.0    9.5
..        …      …      …
338       8.9    8.8    8.3
339       9.0    9.2    9.4
340       9.5    9.2    9.6
341       8.7   10.0    7.3
342       8.9    9.5    8.5

[327 rows x 13 columns]
```

1.1 Exploratory Analysis and Visualization

After cleaning up the dataset, there are still many things in the data that we don't know about. Since Japan is somewhere many people want to go for a vacation one day (after they open up again due to COVID-19 of course), I mainly wanted to see things based on location like prices based on location or ratings for example.

[ ]:
```
import seaborn as sns
import matplotlib
```

```
import matplotlib.pyplot as plt
%matplotlib inline

sns.set_style('darkgrid')
matplotlib.rcParams['font.size'] = 14
matplotlib.rcParams['figure.figsize'] = (9, 5)
matplotlib.rcParams['figure.facecolor'] = '#00000000'
```

Seeing how many hostel's were counted for in each city

[ ]:
```
df_clean['city'].value_counts()
```

[ ]:
```
Tokyo        122
Osaka        101
Kyoto         73
Fukuoka       17
Hiroshima     14
Name: city, dtype: int64
```

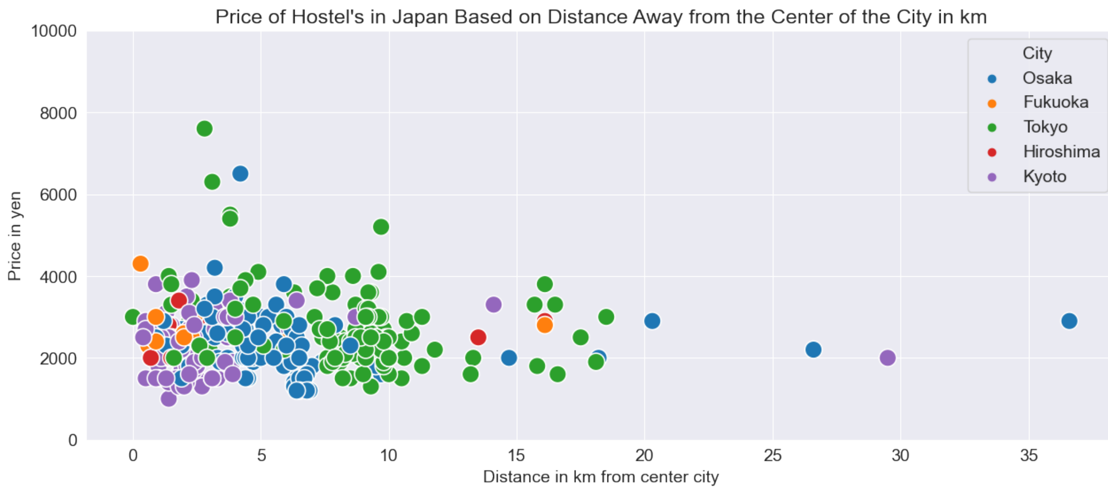
I first wanted to see how many hostel's were counted for in each city.

We first see that the cities that the hostel's are located in are Tokyo, Osaka, Kyoto, Fukuoka, and Hiroshima.

It is no surprised that Tokyo and Osaka have the most hostels listed as they are the two most popular cities in Japan and are where most of the tourists go to. The order more or less goes from most popular to least popular cities for tourists.

Exploring the relationship of the price of the hostels based on the distance away from the center of the city

[ ]:
```
plt.figure(figsize=(15, 6))
plt.title("Price of Hostel's in Japan Based on Distance Away from the Center of␣
↪the City in km")
g = sns.scatterplot(x=df_clean.distance, y=df_clean.price, hue=df_clean.city,␣
↪s=200);
g.set(ylim = (0,10000));
g.set(xlabel = 'Distance in km from center city', ylabel = 'Price in yen');
plt.legend(title='City');
```



I plotted this graph using seaborn's scatterplot method and using matplotlib to label and organize everything.

This graph shows the prices of all the hostel's based on their distance from the center of the city for each city. We see that the majority of the points are all clumped up together around the 1500 yen to 4000 yen area and are mainly within 0 to 10 km of the center of the city. Interestingly enough, there isn't much of a variation in price based on distance so there must be other factors into how the hostel's are priced.

Exploring the ratings of the hostel

```
[ ]: df_clean['rating'].value_counts()
```

```
[ ]: Superb        182
     Fabulous      106
     Very Good      20
     Good           11
     Rating          8
     Name: rating, dtype: int64
```

I used the value_counts method to return the count of each rating.

As we can see, the majority of the hostel's in the dataset have a rating of superb or fabulous. Those two ratings make up around 88% of the total ratings. Japan is known for being clean and having great service so it does not really surprise me seeing the majority of the ratings be at the top. I am not sure what the 'rating' rating is but it makes up only about 2% of the total ratings.

I am now curious about what its like in each city now.

```
[ ]: df_clean.groupby('city')['rating'].value_counts()
```

```
[ ]: city       rating
     Fukuoka    Superb        9
```

```
           Fabulous       8
Hiroshima  Superb        11
           Fabulous       3
Kyoto      Superb        46
           Fabulous      21
           Good           2
           Rating         2
           Very Good      2
Osaka      Superb        50
           Fabulous      35
           Good           7
           Very Good      6
           Rating         3
Tokyo      Superb        66
           Fabulous      39
           Very Good     12
           Rating         3
           Good           2
Name: rating, dtype: int64
```

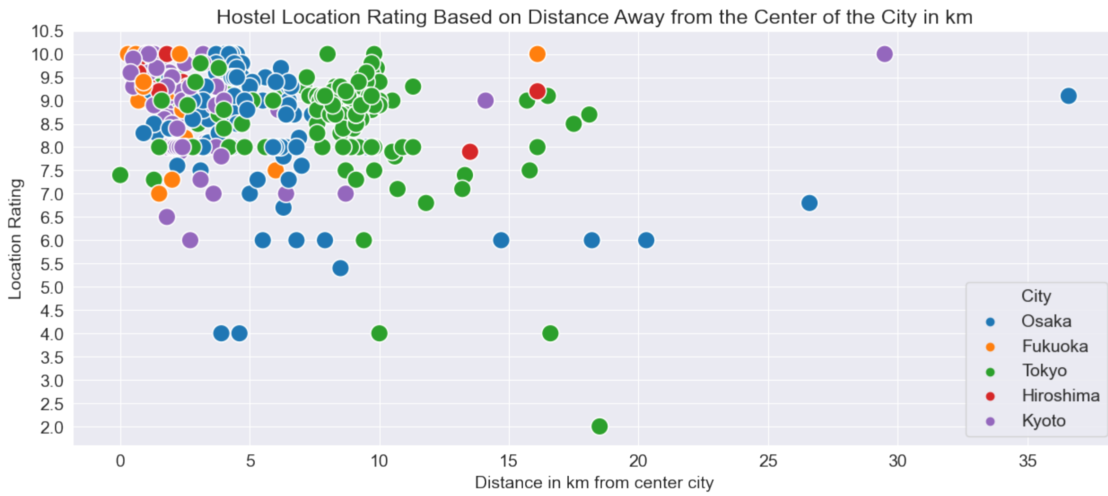
Although there was a small sample size for Fukuoka and Hiroshima hostel's, all of them were rated in either superb or fabulous.

Around 86% of the hostel's in Tokyo are rated superb or fabulous. 84% for Osaka, and 92% for Kyoto.

Exploring if there is a relationship of the location rating and the distance the hostel is from the center of the city

[ ]:
```
import numpy as np
```

[ ]:
```
plt.figure(figsize=(15, 6))
plt.title("Hostel Location Rating Based on Distance Away from the Center of the␣
 ↪City in km")
g = sns.scatterplot(x=df_clean.distance, y=df_clean.location_rating,␣
 ↪hue=df_clean.city, s=200);
plt.yticks(np.arange(min(df_clean.location_rating), max(df_clean.
 ↪location_rating)+1, 0.5))
g.set(xlabel = 'Distance in km from center city', ylabel = 'Location Rating');
plt.legend(title='City');
```



I used seaborn once again to create a scatter plot of the data and used matplotlib as well as numpy this time to label and organize everything.

I wanted to see if there was some kind of correlation based on the location rating and the distance from the center of the city. Though there isn't any correlation, as I expected, the closer the hostel is located from the center of the city, the higher rating it is. Similarly with the first scatter plot between the price of hostel's in Japan based on distance away from the center of the city, there most of the hostels are highly rated when the hostel is located between 0 to 10 km from the center of the city.

Exploring the averages of the other ratings

[ ]:
```
df_clean.
↪groupby(['city'])[['atmosphere','cleanliness','facilities','security','staff','value']].
↪mean()
```

[ ]:

city	atmosphere	cleanliness	facilities	security	staff	value
Fukuoka	8.817647	9.376471	8.900000	9.076471	9.358824	8.982353
Hiroshima	8.535714	9.535714	8.785714	9.371429	9.450000	9.242857
Kyoto	8.441096	9.345205	8.793151	9.094521	9.168493	9.013699
Osaka	8.091089	8.754455	8.501980	8.697030	9.064356	8.700000
Tokyo	8.125410	8.914754	8.495902	9.000000	9.101639	8.808197

I wanted to see the average ratings of all the other ratings that were in the dataset. From looking at all the averages, we see that all the ratings for every category are very close city to city at around 8.0 to mid 9.0. This leads me to believe that Japan holds a high importance on the overall well-being for the visitors to ensure that their stay is as best as possible.

1.2 Asking and Answering Questions

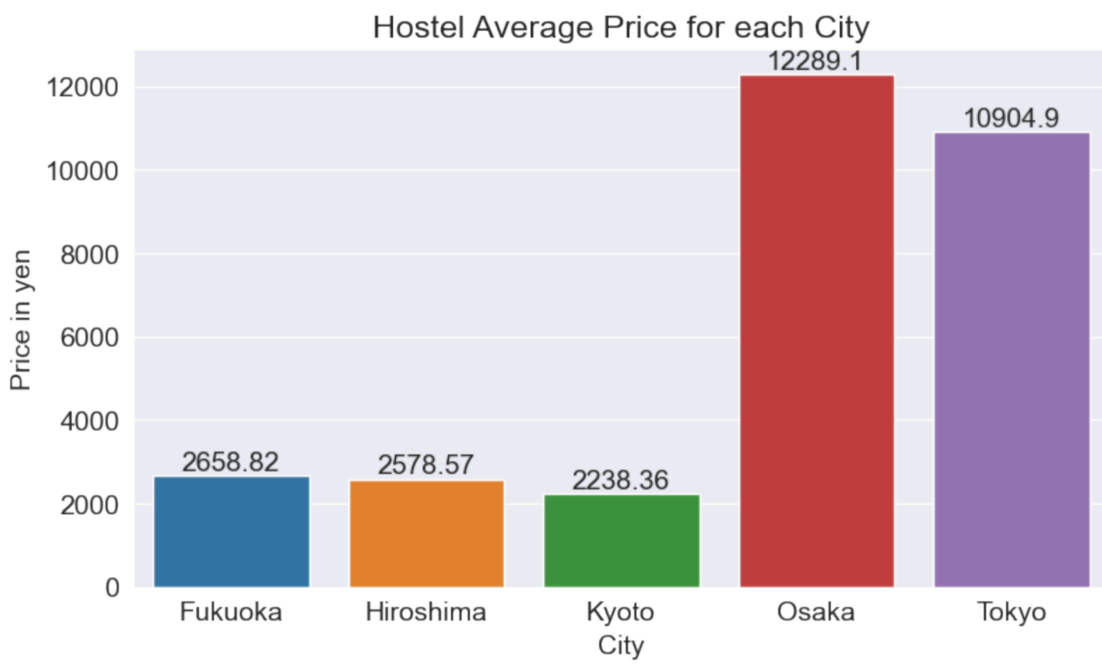
There are still some unanswered question I have so I will be exploring more on the dataset and taking a more in-depth look into it.

Q1: What is the average price of the Hostel's in each city?

[ ]:
```
avg_price_city_df = df_clean.groupby('city')[['price']].mean()
avg_price_city_df
```

[ ]:
```
                  price
city
Fukuoka     2658.823529
Hiroshima   2578.571429
Kyoto       2238.356164
Osaka      12289.108911
Tokyo      10904.918033
```

[ ]:
```
bar0 = sns.barplot(x = avg_price_city_df.index, y = avg_price_city_df.price);
plt.title('Hostel Average Price for each City');
plt.xlabel('City');
plt.ylabel('Price in yen');
bar0.bar_label(bar0.containers[0]);
```



When looking at the scatter plot of all the prices of the hostel across Japan, I assumed that the average prices would be relatively close to each other. Surprisingly we see that Osaka and Tokyo have a much higher average price at around 10000 to 12000 yen (around 100 to 120 usd) while Fukuoka, Hiroshima, Kyoto are relatively close to each other at around 2000 to 3000 yen (around 20 to 30 usd).

This seemed strange based on the information from the scatter plot so I looked into the dataset more in-depth.

```
[ ]: df_clean.groupby(['city'])[['price']].describe()
```

```
[ ]:            price                                                              \
                count          mean           std     min     25%     50%     75%
     city
     Fukuoka     17.0   2658.823529    469.119952  2300.0  2500.0  2500.0  2600.0
     Hiroshima   14.0   2578.571429    428.195806  2000.0  2300.0  2550.0  2900.0
     Kyoto       73.0   2238.356164    682.444572  1000.0  1600.0  2200.0  2700.0
     Osaka      101.0  12289.108911  99588.049384  1200.0  2000.0  2300.0  2800.0
     Tokyo      122.0  10904.918033  90585.744826  1300.0  2100.0  2500.0  3150.0


                      max
     city
     Fukuoka       4300.0
     Hiroshima     3400.0
     Kyoto         3900.0
     Osaka      1003200.0
     Tokyo      1003200.0
```

Using the describe method, I checked the summary statistics of all the prices in each city and saw that Osaka and Tokyo have a much higher max than the other cities at 1003200 yen (around 10,000 usd).

```
[ ]: df_clean.groupby(['city'])[['price']].idxmax()
```

```
[ ]:            price
     city
     Fukuoka       41
     Hiroshima    223
     Kyoto        239
     Osaka        290
     Tokyo        317
```

```
[ ]: df_clean.loc[([41, 223, 239, 290, 317])]
```

```
[ ]:                           hostel       city    price  distance  summary  \
     41   Book And Bed Tokyo Fukuoka    Fukuoka     4300       0.3      8.6
     223            Kyoubashi Ryokan  Hiroshima     3400       1.8      8.6
     239                 O-yado Sato      Kyoto     3900       2.3      9.1
     290            Shell Nell namba      Osaka  1003200       4.8      8.3
     317  Tokyo Central Youth Hostel      Tokyo  1003200       4.8      8.0

            rating  atmosphere  cleanliness  facilities  location_rating  security  \
     41   Fabulous         8.0         10.0         6.0             10.0       8.0
     223  Fabulous         6.0         10.0         6.0             10.0      10.0
     239    Superb         9.0         10.0         9.0              8.0      10.0
     290  Fabulous         6.0          8.0         9.0              9.0       9.0
```

```
317  Fabulous          6.0          10.0         10.0                 8.0        8.0

     staff  value
41    10.0    8.0
223    8.0   10.0
239    9.0    9.0
290    9.0    8.0
317    8.0    6.0
```

Using the idxmax and loc method, I found the index for the highest prices for each city.

Since the hostel in Osaka and Tokyo are clear outliers, I want to compare the average prices of the hostels in each city again but excluding these two hostels.

[ ]:
```
df_no_out = df_clean.copy()
```

[ ]:
```
df_no_out = df_no_out.drop([290, 317])
df_no_out
```

[ ]:
```
                           hostel     city  price  distance  summary  \
1    "Bike & Bed" CharinCo Hostel    Osaka   3300       2.9      9.2
2                    & And Hostel  Fukuoka   2600       0.7      9.5
3           &And Hostel Akihabara    Tokyo   3600       7.8      8.7
4                &And Hostel Ueno    Tokyo   2600       8.7      7.4
5      &And Hostel-Asakusa North-    Tokyo   1500      10.5      9.4
..                            …        …    …       …        …
338       YADOYA Guesthouse Green    Tokyo   2300       2.6      8.2
339      YADOYA Guesthouse Orange    Tokyo   2000       2.9      8.9
340             YAWP! backpackers    Tokyo   2500      17.5      9.3
341               You En Me House    Kyoto   2800       2.4      8.0
342               Zabutton Hostel    Tokyo   2900       5.9      8.6

        rating  atmosphere  cleanliness  facilities  location_rating  \
1       Superb         8.9          9.4         9.3              8.9
2       Superb         9.4          9.7         9.5              9.7
3     Fabulous         8.0          7.0         9.0              8.0
4    Very Good         8.0          7.5         7.5              7.5
5       Superb         9.5          9.5         9.0              9.0
..         …           …            …           …              …
338   Fabulous         7.9          7.7         6.9              8.9
339   Fabulous         8.6          9.0         7.8              9.4
340     Superb         9.5          9.3         9.4              8.5
341   Fabulous         7.3          8.0         6.7              8.0
342   Fabulous         8.1          8.5         7.8              9.0

     security  staff  value
1         9.0    9.4    9.4
2         9.2    9.7    9.5
```

```
3         10.0   10.0    9.0
4          7.0    8.0    6.5
5          9.5   10.0    9.5
..         …    …    …
338        8.9    8.8    8.3
339        9.0    9.2    9.4
340        9.5    9.2    9.6
341        8.7   10.0    7.3
342        8.9    9.5    8.5

[325 rows x 13 columns]
```

I was able to get rid of the two outlier's using the drop method.

[ ]:
```
df_no_out.groupby(['city'])[['price']].describe()
```

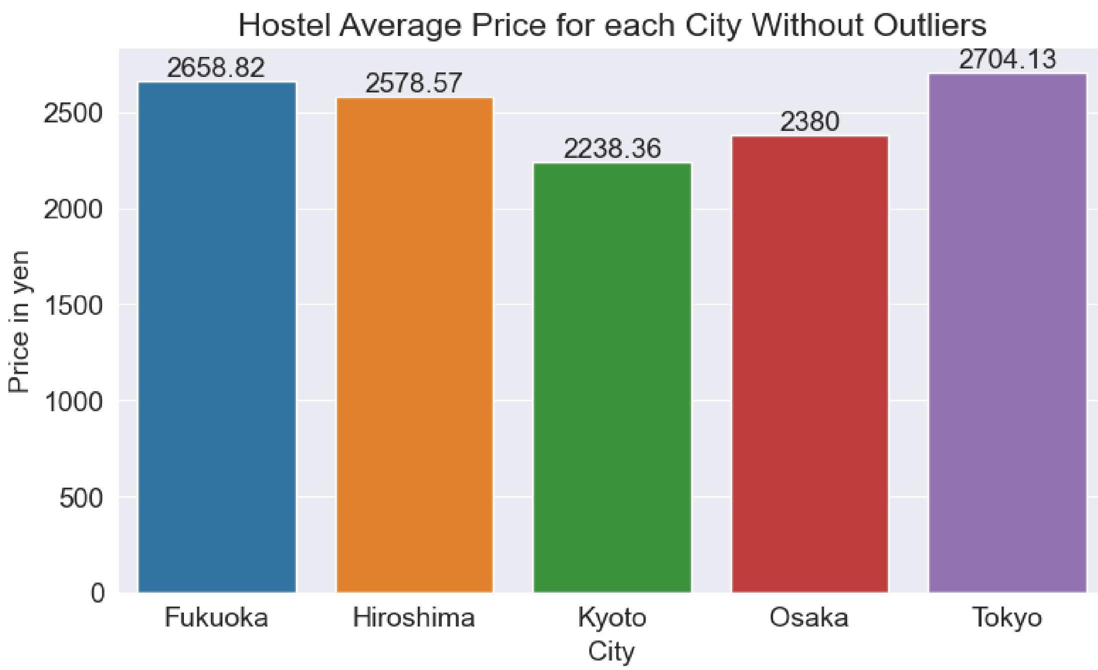
[ ]:
```
           price                                                              \
           count         mean         std     min     25%     50%     75%
city
Fukuoka     17.0  2658.823529  469.119952  2300.0  2500.0  2500.0  2600.0
Hiroshima   14.0  2578.571429  428.195806  2000.0  2300.0  2550.0  2900.0
Kyoto       73.0  2238.356164  682.444572  1000.0  1600.0  2200.0  2700.0
Osaka      100.0  2380.000000  742.096403  1200.0  2000.0  2300.0  2800.0
Tokyo      121.0  2704.132231  965.694629  1300.0  2100.0  2500.0  3000.0


              max
city
Fukuoka    4300.0
Hiroshima  3400.0
Kyoto      3900.0
Osaka      6500.0
Tokyo      7600.0
```

[ ]:
```
avg_price_city_no_out_df = df_no_out.groupby('city')[['price']].mean()
avg_price_city_no_out_df
```

[ ]:
```
                 price
city
Fukuoka    2658.823529
Hiroshima  2578.571429
Kyoto      2238.356164
Osaka      2380.000000
Tokyo      2704.132231
```

[ ]:
```
bar1 = sns.barplot(x = avg_price_city_no_out_df.index, y =␣
 ↪avg_price_city_no_out_df.price);
plt.title('Hostel Average Price for each City Without Outliers');
```

```
plt.xlabel('City');
plt.ylabel('Price in yen');
bar1.bar_label(bar1.containers[0]);
```



Without the outliers, we now see that the average prices are more even with each other at around 2200 to 2700 yen which is overall quite cheap (low 20 usd). Now we can see that Tokyo and Fukuoka have the highest average price but only by few hundred yen (which are about a few usd).

Q2: What kind of scores did the hostel with the 'rating' rating have? The 'rating' rating was something I didn't know what it meant so I wanted to see if the scores of the other variables can give us a possible reason as to why these hostel have that weird rating.

As we found out previously, Fukuoka and Hiroshima had only superb or fabulous ratings so that would mean the hostel's that got the 'rating' rating are in Tokyo, Osaka, and Kyoto.

[ ]:
```
df_clean.loc[df_clean['rating'] == 'Rating']
```

[ ]:

	hostel	city	price	distance	summary	rating	\
35	bnb+ Ninja Dojo Ueno	Tokyo	2500	8.0	5.2	Rating	
36	bnb+Shinjuku Castle	Tokyo	2500	1.3	5.0	Rating	
68	Ezstay Osaka	Osaka	2000	4.6	3.1	Rating	
141	Hostel Ginkakuji	Kyoto	1300	2.7	4.9	Rating	
235	Nagomi-Ryokan Yuu	Kyoto	3100	2.2	5.1	Rating	
273	Qoo Ebisucho	Osaka	2000	5.5	5.7	Rating	
281	Sakura Guest House	Osaka	2700	4.3	4.9	Rating	
291	Shibamata FU-TEN Bed and Local	Tokyo	3000	18.5	4.9	Rating	

```
     atmosphere  cleanliness  facilities  location_rating  security  staff  \
35          4.0          3.3         4.0              8.7       5.3    4.7
36          2.7          4.0         3.3              7.3       5.3    5.3
68          2.0          2.0         2.0              4.0       6.0    2.0
141         2.0         10.0         6.0              6.0       2.0    2.0
235         6.0          6.0         2.0              8.0       6.0    2.0
273         6.0          8.0         6.0              6.0       2.0    8.0
281         4.0          4.0         4.0              8.0       4.0    6.0
291         2.0          4.0         2.0              2.0      10.0   10.0

     value
35     6.7
36     6.7
68     4.0
141    6.0
235    6.0
273    4.0
281    4.0
291    4.0
```

Using the loc method, I was able to see the hostel's that had the 'rating' rating.

We can see that there are 3 hostel's in Tokyo, 3 hostel's in Osaka, and 2 hostel's in Kyoto that have the 'rating' rating.

Looking at the other scores, we can see that there are few scores that are very low (lower that 5.0) for every hostel. We can even see some 2.0 ratings which is a very bad rating.

I want to see the scores of hostel's with the superb rating to see if there are any low scores.

[ ]:
```
df_clean.loc[df_clean['rating'] == 'Superb']
```

[ ]:
```
                          hostel       city  price  distance  summary  \
1    "Bike & Bed" CharinCo Hostel      Osaka   3300       2.9      9.2
2                   & And Hostel    Fukuoka   2600       0.7      9.5
5      &And Hostel-Asakusa North-      Tokyo   1500      10.5      9.4
7              328 Hostel & Lounge      Tokyo   3300      16.5      9.3
8                        36Hostel  Hiroshima   2000       1.6      9.5
..                             …          …      …         …        …
333             WIRED HOTEL Asakusa      Tokyo   5200       9.7      9.5
334        Wise Owl Hostels Shibuya      Tokyo   2500       4.0      9.1
335          Wise Owl Hostels Tokyo      Tokyo   2000       7.9      9.1
336                          With B      Tokyo   4000       8.6      9.1
340               YAWP! backpackers      Tokyo   2500      17.5      9.3

     rating  atmosphere  cleanliness  facilities  location_rating  security  \
1    Superb         8.9          9.4         9.3              8.9       9.0
2    Superb         9.4          9.7         9.5              9.7       9.2
```

```
5    Superb         9.5           9.5          9.0                9.0       9.5
7    Superb         8.7           9.7          9.3                9.1       9.3
8    Superb         8.8           9.9          9.2                9.6       9.8
..      …          …          …          …              …        …
333  Superb         8.9          10.0          9.9                9.1       9.7
334  Superb         8.9           9.6          9.1                8.4       9.6
335  Superb         8.4           9.5          9.1                9.1       9.4
336  Superb         8.5           9.5          9.0                8.0       9.0
340  Superb         9.5           9.3          9.4                8.5       9.5

     staff  value
1      9.4    9.4
2      9.7    9.5
5     10.0    9.5
7      9.7    8.9
8      9.8    9.5
..     …     …
333    9.6    9.5
334    9.5    8.8
335    9.3    9.1
336    9.5   10.0
340    9.2    9.6

[182 rows x 13 columns]
```

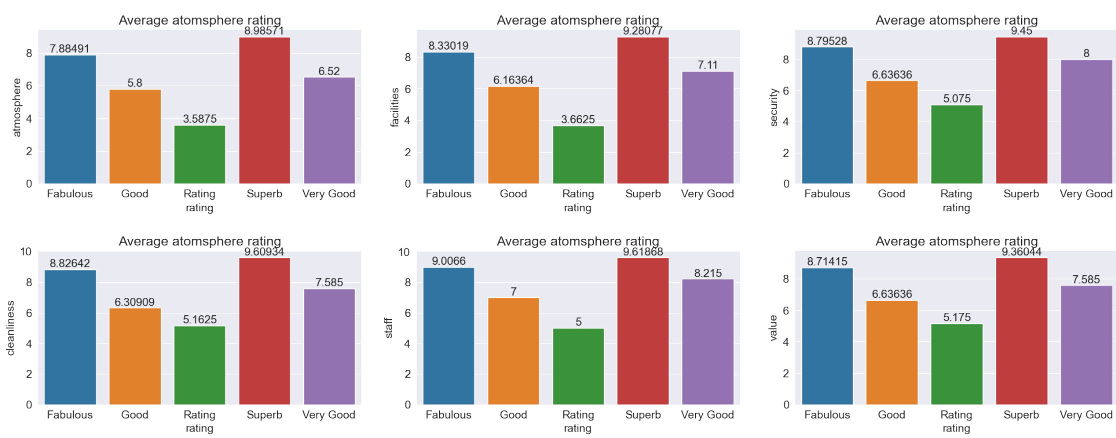
Just as I thought, all the scores (that are visible) have ratings over 8.0 in every category.

[ ]:
```python
avg_rating_df = df_clean.
 ↪groupby(['rating'])[['atmosphere','cleanliness','facilities','security','staff','value']].
 ↪mean()
avg_rating_df
```

[ ]:
```
           atmosphere  cleanliness  facilities  security     staff     value
rating
Fabulous     7.884906     8.826415    8.330189  8.795283  9.006604  8.714151
Good         5.800000     6.309091    6.163636  6.636364  7.000000  6.636364
Rating       3.587500     5.162500    3.662500  5.075000  5.000000  5.175000
Superb       8.985714     9.609341    9.280769  9.450000  9.618681  9.360440
Very Good    6.520000     7.585000    7.110000  8.000000  8.215000  7.585000
```

[ ]:
```python
fig, axes = plt.subplots(2, 3, figsize=(20, 8))

# Use the axes for plotting
axes[0,0].set_title('Average atomsphere rating')
ax0 = sns.barplot(x = avg_rating_df.index, y = avg_rating_df.atmosphere, ax =␣
 ↪axes[0,0])
ax0.bar_label(ax0.containers[0])
```

```
axes[1,0].set_title('Average atomsphere rating')
ax1 = sns.barplot(x = avg_rating_df.index, y = avg_rating_df.cleanliness, ax = axes[1,0])
ax1.bar_label(ax1.containers[0])

axes[0,1].set_title('Average atomsphere rating')
ax2 = sns.barplot(x = avg_rating_df.index, y = avg_rating_df.facilities, ax = axes[0,1])
ax2.bar_label(ax2.containers[0])

axes[0,2].set_title('Average atomsphere rating')
ax3 = sns.barplot(x = avg_rating_df.index, y = avg_rating_df.security, ax = axes[0,2])
ax3.bar_label(ax3.containers[0])

axes[1,1].set_title('Average atomsphere rating')
ax4 = sns.barplot(x = avg_rating_df.index, y = avg_rating_df.staff, ax = axes[1,1])
ax4.bar_label(ax4.containers[0])

axes[1,2].set_title('Average atomsphere rating')
ax5 = sns.barplot(x = avg_rating_df.index, y = avg_rating_df.value, ax = axes[1,2])
ax5.bar_label(ax5.containers[0])



plt.tight_layout(pad=2);
```



I wanted to see the average scores for each rating to see where ‘rating’ lines up in the rating scale. I used the plt.subplots method to create multiple graphs to visual the average scores for every category for each rating.

As we can see 'rating' as the lowest average scores across the board meaning that the 'rating' rating is the lowest rating in the dataset.

Q3: Which hostel has the best value of money rating? Since the value category is a rating for the value for the money, I am curious to see what makes the hostel worth its price. What is the cost? Is the distance very close to the center of city?

[ ]:
```
df_clean.nlargest(n=10, columns = ['value'])
```

[ ]:
```
                              hostel     city  price  distance  summary  \
13              Akihabara Hotel 3000    Tokyo   2200       8.0     10.0
30  Beagle Tokyo Hostel & Apartments    Tokyo   3800      16.1      9.0
48            Capsule Inn Kinshichou    Tokyo   2600      10.9      8.6
55                       Chaganjutei    Kyoto   3300      14.1      9.6
58                           Colours    Kyoto   1600       1.8      8.9
60            Common de Hostel & Bar  Fukuoka   2500       0.8      9.9
71                       Friends Kyo    Kyoto   2200       1.2     10.0
77                  Goen Lounge Stay    Osaka   2000      18.2      8.9
79           Gojo Guesthouse - Annex    Kyoto   2500       1.7      9.6
87                Guest House Denchi    Tokyo   3600       9.2      8.9

      rating  atmosphere  cleanliness  facilities  location_rating  security  \
13    Superb        10.0         10.0        10.0             10.0      10.0
30    Superb         8.0         10.0         9.0              8.0       9.0
48  Fabulous         6.0         10.0         8.0              8.0      10.0
55    Superb        10.0         10.0         9.0              9.0       9.0
58  Fabulous         4.0         10.0        10.0              8.0      10.0
60    Superb        10.0         10.0        10.0              9.0      10.0
71    Superb        10.0         10.0        10.0             10.0      10.0
77  Fabulous         8.0         10.0        10.0              6.0       8.0
79    Superb         8.7         10.0         9.3              9.3      10.0
87  Fabulous         8.0         10.0         6.0              8.0      10.0

    staff  value
13   10.0   10.0
30    9.0   10.0
48    8.0   10.0
55   10.0   10.0
58   10.0   10.0
60   10.0   10.0
71   10.0   10.0
77   10.0   10.0
79   10.0   10.0
87   10.0   10.0
```

Using the nlargest method, I looked at the top 10 hostel's with the highest value for money.

We can see that Akihabara Hotel 3000 is at the top having a 10.0 score for every category. Something I found interesting is the hostel Colours has a 10 value rating but we can see that the atmosphere rating is very low at 4.0. I am guessing that the other scores being high rated while being the cheapest and relatively close to the center of the city brings it back to a 10.0.

I see that Friends Kyo also has 10's across the board so the top 10 here is just by index order if the value is the same which is not exactly what I was looking for.

```
[ ]: len(df_clean[df_clean['value'] == 10.0])
```

```
[ ]: 31
```

Using the len method, we can see that there are 31 hostels with a 10.0 for the value of money score.

Q4: Which hostel has the best summary score? Since it was a bit difficult to see which hostel was considered the "best" by value, I will look at the summary score which is just the average of all the scores in every category including value.

```
[ ]: df_clean.nlargest(n=10, columns = ['summary'])
```

```
[ ]:                              hostel     city  price  distance  \
13                 Akihabara Hotel 3000    Tokyo   2200       8.0
71                          Friends Kyo    Kyoto   2200       1.2
89                     Guest House Hachi    Kyoto   3200       3.2
111        Guesthouse Morizou Female only    Osaka   2000       3.7
135             Hostel Chapter Two Tokyo    Tokyo   2700       9.8
237           Norishico Auto Guesthouse  Fukuoka   2800      16.1
300                 Tabicolle Backpackers  Fukuoka   2500       2.3
302                                Talbot    Kyoto   2000      29.5
47   Capsule Hotel Asahi Plaza Shinsaibashi    Osaka   3000       4.0
60                 Common de Hostel & Bar  Fukuoka   2500       0.8

     summary  rating  atmosphere  cleanliness  facilities  location_rating  \
13      10.0  Superb        10.0         10.0        10.0             10.0
71      10.0  Superb        10.0         10.0        10.0             10.0
89      10.0  Superb        10.0         10.0        10.0             10.0
111     10.0  Superb        10.0         10.0        10.0             10.0
135     10.0  Superb        10.0         10.0        10.0             10.0
237     10.0  Superb        10.0         10.0        10.0             10.0
300     10.0  Superb        10.0         10.0        10.0             10.0
302     10.0  Superb        10.0         10.0        10.0             10.0
47       9.9  Superb        10.0         10.0        10.0             10.0
60       9.9  Superb        10.0         10.0        10.0              9.0

     security  staff  value
13       10.0   10.0   10.0
71       10.0   10.0   10.0
89       10.0   10.0   10.0
111      10.0   10.0   10.0
```

```
135       10.0   10.0   10.0
237       10.0   10.0   10.0
300       10.0   10.0   10.0
302       10.0   10.0   10.0
47        10.0   10.0    9.0
60        10.0   10.0   10.0
```

We can see that 8 hostel's have a perfect 10.0 summary rating. What is surprising to me is that the Capsule Hotel Asahi Plaza Shinsaibashi has a 9.9 summary rating because of the value rating being 9.0. That made me think why the value a 9.0 when every other rating was a 10.0. I then shifted my attention to the price and distance but it isn't much different from hostel's like Guest House Hachi which is located in the same city, very similar in price, and being only about 0.8 km different in the distance from the center of the city.

Q5: Are the cheapest hostel's worth it?

[ ]:
```
df_clean.nsmallest(n=10, columns = ['price'])
```

[ ]:
```
                        hostel   city  price  distance  summary     rating  \
63          Downtown Inn Kyoto  Kyoto   1000       1.4      9.3     Superb
179        Ikidane House Namba  Osaka   1200       6.5      9.3     Superb
262          Peace House Abeno  Osaka   1200       6.9      9.0     Superb
263          Peace House Sachi  Osaka   1200       6.5      8.6   Fabulous
265          Peace House Showa  Osaka   1200       6.8      7.5  Very Good
266       Peace House Suzunami  Osaka   1200       6.4      8.2   Fabulous
20   Asakusa Hostel Toukaisou  Tokyo   1300       9.3      8.9   Fabulous
141            Hostel Ginkakuji  Kyoto   1300       2.7      4.9     Rating
172            Hotel Sun Plaza  Osaka   1300       6.3      6.5       Good
264          Peace House Sakura  Kyoto   1300       1.8      6.3       Good

     atmosphere  cleanliness  facilities  location_rating  security  staff  \
63          8.9          9.7         9.7              9.1       9.1    8.6
179         8.9          9.6         9.7              8.8       8.9    9.6
262         9.0          9.1         9.1              8.2       9.1    9.1
263         8.8          8.2         7.9              8.9       8.4    9.2
265         8.0          6.4         5.6              8.0       6.4    9.6
266         8.4          7.1         7.9              8.7       8.0    8.9
20          7.9          9.4         8.3              9.5       9.2    8.9
141         2.0         10.0         6.0              6.0       2.0    2.0
172         6.0          7.3         6.0              6.7       6.0    7.3
264         6.8          5.8         5.3              6.5       6.3    6.8

     value
63     9.7
179    9.6
262    9.3
263    8.8
265    8.4
```

```
266    8.2
20     9.2
141    6.0
172    6.0
264    7.0
```

I wanted to see if the cheapest hostel's in the dataset are actually worth it.

Using the nsmallest method, we can see the top 10 cheapest hostels in Japan. Surprisingly enough, the cheapest hostel Downtown Inn Kyoto, has an overall high rating across the board, is relatively close to the center of the city and has a superb rating for only 1000 yen (less $10 usd). Something I found interesting is that the 9 out of the 10 cheapest hostel's are all located in Osaka and Kyoto which are both in the same region of Japan (Kansai).

Based on personal experience, staying in Osaka did seem cheaper than Tokyo but I am surprised not to see any Fukuoka and Hiroshima hostel's on here but maybe that is due to the low sample size for those two cities.

[ ]:
```
df_clean.nlargest(n=12, columns = ['price'])
```

[ ]:
```
                                 hostel     city    price  distance  \
290                    Shell Nell namba    Osaka  1003200       4.8
317           Tokyo Central Youth Hostel    Tokyo  1003200       4.8
301          Tadaima Japan Shinjuku Ryokan    Tokyo     7600       2.8
200       Kaneyoshi Ryokan, Namba Dotombori    Osaka     6500       4.2
311              The Millennials Shibuya    Tokyo     6300       3.1
234  Nadeshiko Hotel Shibuya (Female Only)    Tokyo     5500       3.8
327                     Turn Table Hostel    Tokyo     5400       3.8
333                   WIRED HOTEL Asakusa    Tokyo     5200       9.7
41             Book And Bed Tokyo Fukuoka  Fukuoka     4300       0.3
165               Hotel Cargo Shinsaibashi    Osaka     4200       3.2
40             Book And Bed Tokyo Asakusa    Tokyo     4100       9.6
42           Book And Bed Tokyo Ikebukuro    Tokyo     4100       4.9

     summary    rating  atmosphere  cleanliness  facilities  location_rating  \
290      8.3  Fabulous         6.0          8.0         9.0              9.0
317      8.0  Fabulous         6.0         10.0        10.0              8.0
301      8.4  Fabulous         8.0          9.0         9.0              8.0
200      9.4    Superb        10.0          9.0         9.0             10.0
311      9.3    Superb         8.9         10.0         9.3              9.8
234      9.2    Superb         8.7          9.7         9.7              8.7
327      9.6    Superb         9.4          9.9         9.7              9.7
333      9.5    Superb         8.9         10.0         9.9              9.1
41       8.6  Fabulous         8.0         10.0         6.0             10.0
165      9.3    Superb         8.4          9.6         8.8              8.8
40       9.1    Superb         8.7          9.7         8.6              9.4
42       8.2  Fabulous         8.9          8.0         8.0              9.1

     security  staff  value
```

```
290          9.0    9.0    8.0
317          8.0    8.0    6.0
301          8.0    9.0    8.0
200          9.0   10.0    9.0
311          9.5    8.9    8.7
234          9.7    9.3    8.7
327          9.6    9.6    9.3
333          9.7    9.6    9.5
41           8.0   10.0    8.0
165          9.6   10.0    9.6
40           9.5    9.3    8.6
42           7.7    8.6    7.4
```

After looking at the cheapest hostel's, I wanted to also see if the most expensive hostel's are worth it.

Using the nlargest method, we can see the top 12 most expensive hostel's. I decided to look at the top 12 because of the two outliers.

Looking at the two outliers, we can see the rating is only a fabulous despite it's price. It is odd to see that the value is rated relatively high for Shell Nell namba despite these ratings and the price. This leads me to believe that these two prices might be a mistake in the dataset.

Aside from the outliers, it is no surprise that most of the hostel's have a superb rating because it should be expected to have high ratings in all these categories if you are paying a premium for it.

Unlike the cheapest hostel's, the majority of the most expensive hostel's are located in Tokyo which makes sense given the popularity of the city amongst tourists.

1.3 Inferences and Conclusion

I found a lot of interesting information through transforming the dataset using the things I learned in this course. I was able to see that a lot of the hostel's in the dataset had a relatively similar price despite the distance from the center of the city. This is probably due to how efficient the Japanese train system is so distance isn't much of a factor when people choose a place to stay. We found that the average prices of the hostel's were very similar in every city (excluding outliers). Based on my findings, the value for money rating didn't prove to be very useful when trying to figure out what the "best" hostel was and the summary rating proved to be more useful. We also found that the majority of the hostel's had a rating of superb and fabulous and was able uncover what the 'rating' rating was. With all this, we can conclude that hostel's in Japan are overall well maintained and affordable in Tokyo, Osaka, Kyoto, Fukuoka, and Hiroshima.

1.4 References and Future Work

Something I would be interested in is hostels in other cities in Japan especially Hokkaido which is another hot spot for tourists. Another thing that would be interesting to see is an updated version of the dataset as this dataset is a few years old so I would like to see the effects of COVID-19 and see if there are any drastic changes. Something I would improve on is making sure there aren't any errors in the dataset because I am not sold on the two outlier hostel's being accurate in price as they seem too ridiculous to be real.

Link to the dataset: https://www.kaggle.com/datasets/koki25ando/hostel-world-dataset

Pandas Documentation: https://pandas.pydata.org/docs/user_guide/index.html#user-guide

Putting the value over bar plots: https://stackoverflow.com/questions/43214978/how-to-display-custom-values-on-a-bar-plot