PST 107

Project 1

Japan Hostel Dataset

Duong Thanh Bach

Data background

This is a dataset comprising the data of 342 Hostels located in Japan, including 5 cities: **Tokyo, Osaka, Fukuoka City, Hiroshima & Tokyo.**

Table column:

This data is collected from HostelWorld.com

- Hotel Name
- City
- Price
- Summary Score
- Rating Band
- Atmosphere
- Cleanliness
- Facilities
- Location
- Security
- Staff
- Value of Money
- lon
- lat

	hostel.name	City	price.from	Distance	summary.score	rating.band	atmosphere	cleanliness	facilities	location.y	security	staff	valueformoney	lon	lat
1	"Bike & Bed" CharinCo Hostel	Osaka	3300	2.9	9.2	Superb	8.9	9.4	9.3	8.9	9	9.4	9.4	135.5137671	34.682678
2	& And Hostel	Fukuoka-City	2600	0.7	9.5	Superb	9.4	9.7	9.5	9.7	9.2	9.7	9.5	NA	NA
3	&And Hostel Akihabara	Tokyo	3600	7.8	8.7	Fabulous	8	7	9	8	10	10	9	139.7774724	35.6974473
4	&And Hostel Ueno	Tokyo	2600	8.7	7.4	Very Good	8	7.5	7.5	7.5	7	8	6.5	139.783667	35.712716
5	&And Hostel-Asakusa North-	Tokyo	1500	10.5	9.4	Superb	9.5	9.5	9	9	9.5	10	9.5	139.7983712	35.7278979
6	1night1980hostel Tokyo	Tokyo	2100	9.4	7	Very Good	5.5	8	6	6	8.5	8.5	6.5	139.7869499	35.724384
7	328 Hostel & Lounge	Tokyo	3300	16.5	9.3	Superb	8.7	9.7	9.3	9.1	9.3	9.7	8.9	139.7454672	35.5480439
8	36Hostel	Hiroshima	2000	1.6	9.5	Superb	8.8	9.9	9.2	9.6	9.8	9.8	9.5	NA	NA
9	3Q House - Asakusa Smile	Tokyo	2500	10.2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
10	Ace Inn Shinjuku	Tokyo	2200	3	7.7	Very Good	6.7	7.2	6.8	8.5	7.8	8.5	8.1	139.7243036	35.6925119
11	Air Osaka Hostel	Osaka	1600	9.7	9.2	Superb	9.5	9.1	8.7	8.8	8.9	9.8	9.5	135.4769556	34.6222596
12	Aizuya Inn	Tokyo	2000	10.6	8.5	Fabulous	8.1	8.3	8.4	7.8	8.9	9.1	8.9	139.800993	35.727547
13	Akihabara Hotel 3000	Tokyo	2200	8	10	Superb	10	10	10	10	10	10	10	139.779382	35.6974904
14	Almond hostel & cafe Shibuya	Tokyo	2900	2.2	9.3	Superb	9.1	9.5	8.8	9.5	9.4	9.7	9	139.6875352	35.6700092
15	Anne Hostel Asakusabashi	Tokyo	2000	8.9	9.1	Superb	8.8	9.2	8.7	9	9.1	9.5	9.2	139.7893595	35.6989448
16	Anne Hostel Yokozuna	Tokyo	1800	9.5	9.1	Superb	8.8	9.1	9	9.2	9.3	9.3	9.2	139.7967977	35.6954897

Observation

After importing this dataset, I conducted a preliminary analysis to gain a comprehensive understanding of its contents and identify potential features for future exploration.

- The dataset contains user reviews of hostels, scored based on criteria such as Cleanliness, Location, Staff, Worthiness, and Value. The scores range between a minimum of 2 and a maximum of 10. An overarching "summary score" provides an aggregate rating for each hotel based on collective user feedback.
- This information proves invaluable for international travelers planning future visits to Japan.

Summary(hostel_data)

X	Hostel_Name	City	Starting_Price	Distance	Summary_Scor
Min. : 1.00	Length:342	Length:342	Min. : 1000	Min. : 0.000	Min. : 3.1
1st Qu.: 86.25	Class :character	Class :character	1st Qu.: 2000	1st Qu.: 2.300	1st Qu.: 8.6
Median :171.50	Mode :character	Mode :character	Median : 2500	Median : 4.500	Median : 9.0
Mean :171.50			Mean : 8388	Mean : 5.723	Mean : 8.7
3rd Qu.:256.75			3rd Qu.: 2900	3rd Qu.: 8.450	3rd Qu.: 9.4
Max. :342.00			Max. :1003200	Max. :36.600	Max. :10.0
					NA's :15

Observation

When examining this dataset as an international traveler, several pertinent questions can be addressed:

1. **Average Hotel Reviews by City**: What is the average review score for all hotels in each city throughout Japan? This information can help identify cities known for offering the best deals or exceptional service.
2. **Hotel Density by City:** What is the concentration of hotels in each city? This analysis can reveal which city boasts the highest number of hotels and which one has the fewest, aiding in destination selection.
3. **Best Location & Hotel on a Budge**t: If I'm on a limited budget, which hostels provide the best locations for exploring the city? This question helps travelers prioritize convenience and affordability.
4. **Luxury Hotel Selection**: If budget constraints are not an issue, which luxury hotels in Japan stand out as the most opulent and exclusive options? This inquiry assists in identifying top-tier accommodations.

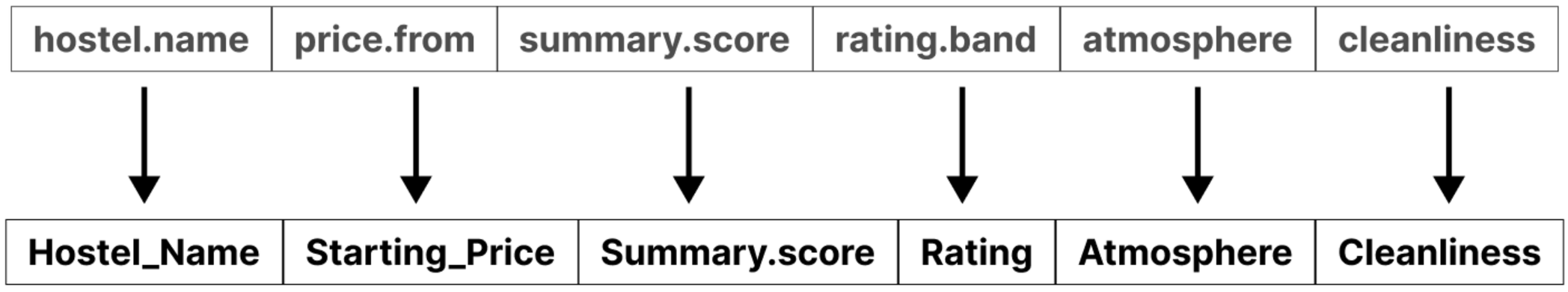
Data Processing - Cleaning Up

To prepare the data for optimal visualization and analysis, you should consider the following data tidying steps:

1. **Remove Unnecessary Columns:** Eliminate the "lon" and "lan" columns, as they are not required to address the listed questions. This can be done using data manipulation tools or code depending on your data format.

~~lon~~	~~lan~~

2. **Rename Header Titles:** Rename the table header titles to adhere to coding best practices. Ensure that the column names are concise and informative.



Data Processing - Converting

Question 1: **Average Hotel Reviews by City**

Methods:

- Group_by()
- Summarise()
- mean()

City	Worthiness	Atmosphere_Ave	Loc_Ave	Staff_Ave	Price_Ave	Clean_Ave	Fac_Ave	Sec_Ave	Acc	Dist_Ave
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>	<dbl>
Fukuoka-City	8.98	8.82	9.02	9.36	2659.	9.38	8.9	9.08	17	2.45
Hiroshima	9.24	8.54	9.23	9.45	2579.	9.54	8.79	9.37	14	3.61
Kyoto	9.01	8.44	8.88	9.17	2238.	9.35	8.79	9.09	73	2.80
Osaka	8.7	8.09	8.56	9.06	12289.	8.75	8.50	8.70	101	5.71
Tokyo	8.81	8.13	8.59	9.10	10905.	8.91	8.50	9	122	8.31

From the table, it's readily apparent that each parameter's maximum and minimum values can be identified for each city. For instance:

- **Pricing in Osaka:** Osaka stands out as the city with the highest hostel prices compared to the other cities in the dataset.
- **Hiroshima's Worthiness Score**: Hiroshima emerges with the highest worthiness score among all cities.

Data Processing - Converting

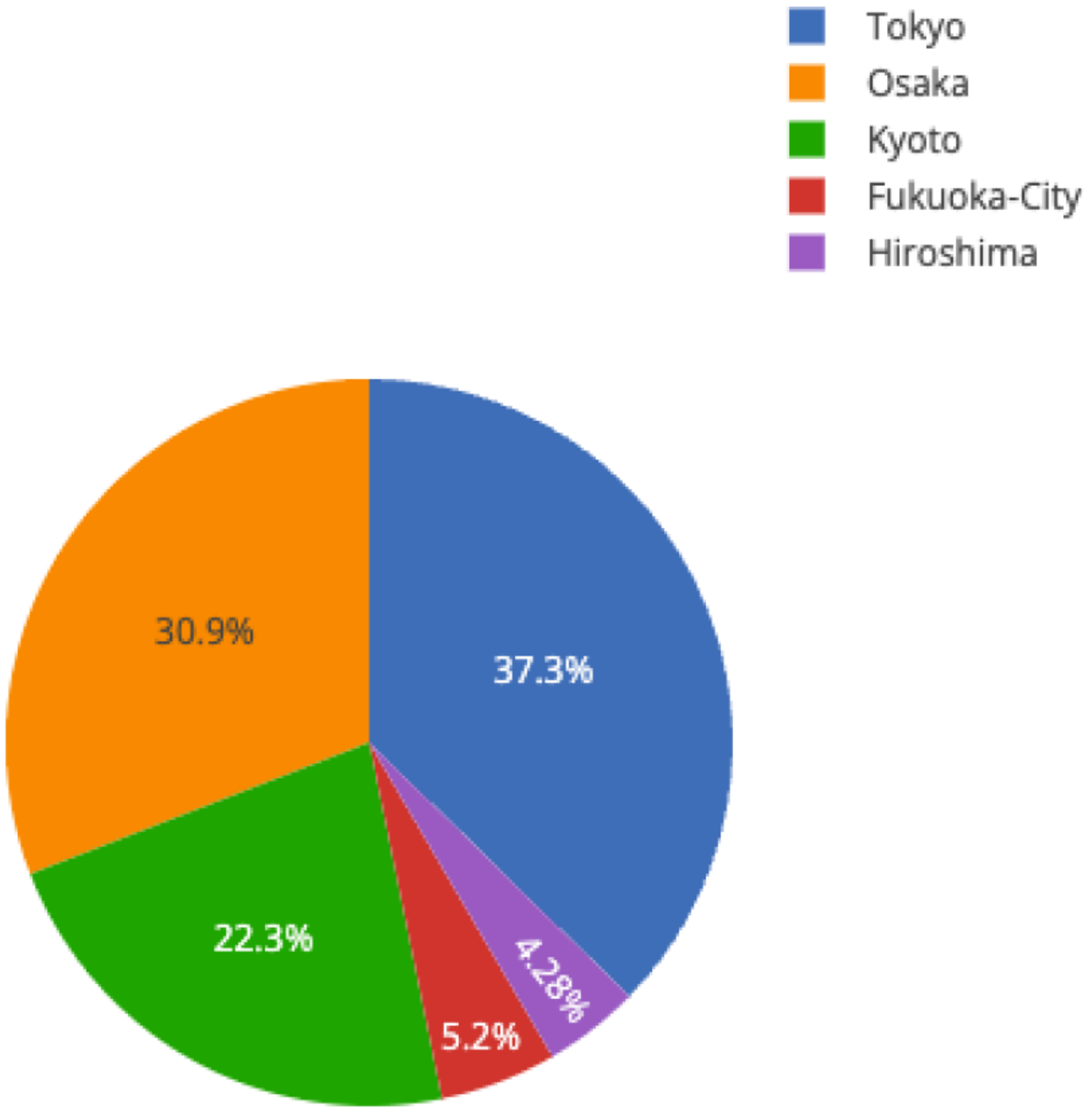
Question 2: **Hotel Density by City**

Methods:

- Plotly()

I utilized the previously transformed dataset and leveraged the Plotly library to construct a Pie Chart. This visualization effectively conveys the distribution of hotels in various cities.

- **Tokyo and Osaka:** Tokyo and Osaka emerge as frontrunners in terms of hotel density. This observation highlights the fact that these two cities offer travelers a more extensive array of hostel options compared to other destinations.



Data Processing - Converting

Question 3: **Best Location on a Budget**

Methods:

- Group_by(City)
- Filter(Starting_Price <= 3000, summary_score > 8, Location >8)
- Summarise(Worthiness, ...)

City	Worthiness	Atmosphere_Ave	Loc	St	Price	Clean	Fac	Sec	Acc	Dist
Fukuoka-City	9.092308	9.053846	9.346154	9.415385	2576.923	9.453846	9.153846	9.076923	13	2.446154
Hiroshima	9.218182	8.854545	9.390909	9.527273	2481.818	9.436364	9.045455	9.272727	11	2.972727
Kyoto	9.344898	8.820408	9.312245	9.493878	2083.673	9.516327	9.236735	9.416327	49	2.469388
Osaka	9.098437	8.581250	9.168750	9.343750	2240.625	9.042188	8.843750	9.060937	64	5.084375
Tokyo	9.025714	8.434286	9.037143	9.265714	2271.429	9.110000	8.761429	9.180000	70	8.390000

Utilizing the cleaned 'Hostel_data_cleaned' dataset, I applied the outlined methods to calculate the overall score for each city across various attributes, utilizing specific filter conditions. These conditions were determined based on the data generated through the summary method.

- Kyoto & Hiroshima emerges as a standout performer, consistently achieving high scores across multiple attributes.

Data Processing - Converting

Question 3: **Best Hotel on a Budget**

Methods:

- Filter(Starting_Price <= 3000, summary_score > 8, Location >8)
- Select()

	Hostel_Name	City	Starting_Price	Distance
1	36Hostel	Hiroshima	2000	1.6
2	Comics & CapsuleHotel COMICAP kyotoshinkyogoku	Kyoto	2000	0.7
3	Downtown Inn Kyoto	Kyoto	1000	1.4
4	Gojo Guest House	Kyoto	2000	1.9
5	Grids Kyoto Shijo Kawaramachi Hotel & Hostel	Kyoto	1800	1.0
6	Guest House Ga-Jyun	Kyoto	2000	1.4
7	Guest House Shiori Ann	Kyoto	1500	2.0
8	Guest House YULULU	Kyoto	1500	0.5
9	Hostel HARUYA Kyoto	Kyoto	1500	1.0
10	Hostel Kyoto Kizuna	Kyoto	1800	2.1
11	Hostel Mundo Chiquito	Kyoto	1400	1.4
12	Jam Hostel Kyoto Gion	Kyoto	1500	0.9
13	Khaosan Kyoto Guest House	Kyoto	2000	1.1
14	Santiago Guesthouse Hiroshima	Hiroshima	2000	0.7
15	The Evergreen Hostel	Hiroshima	2000	1.5
16	Tokyo Hostel Fuji	Tokyo	2000	1.6

Data Processing - Converting

Question 4: **Luxury Hotel Selection**

Methods:

- Filter(Starting_Price <= 3000, summary_score > 8, Location >8)
- Select()

Hostel_Name	Starting_Price	City	Atmosphere	Cleanliness	Worth	Staff	Facilities
Kaneyoshi Ryokan, Namba Dotombori	6500	Osaka	10.0	9.0	9.0	10.0	9.0
Nadeshiko Hotel Shibuya (Female Only)	5500	Tokyo	8.7	9.7	8.7	9.3	9.7
Shell Nell namba	1003200	Osaka	6.0	8.0	8.0	9.0	9.0
Tadaima Japan Shinjuku Ryokan	7600	Tokyo	8.0	9.0	8.0	9.0	9.0
The Millennials Shibuya	6300	Tokyo	8.9	10.0	8.7	8.9	9.3
Tokyo Central Youth Hostel	1003200	Tokyo	6.0	10.0	6.0	8.0	10.0
Turn Table Hostel	5400	Tokyo	9.4	9.9	9.3	9.6	9.7
WIRED HOTEL Asakusa	5200	Tokyo	8.9	10.0	9.5	9.6	9.9

Similar to the previous step, as I aimed to compile a list of hotels with starting prices exceeding the average, there was no necessity to employ the 'Group_by' method. The hotels listed below fall within the higher price range category. Furthermore, this allows for a clear examination of the interplay between pricing and attributes such as cleanliness, atmosphere, worthiness, and staff quality.

Conclusion

From the analysis, several key conclusions can be drawn:

1. Osaka and Tokyo are the cities with the highest concentration of hotels. This density suggests a wide range of options for travelers in terms of budget, location, quality, and other considerations.
2. For budget-conscious travelers prioritizing both affordability and quality, Kyoto stands out as a prime choice given its commendable summary scores and strategic locations of its accommodations.
3. Interestingly, there are hostels that charges premium prices, but their locations do not necessarily correspond to the high costs, suggesting that travelers should exercise caution and conduct thorough research before making bookings.

For future analyses, it might be insightful to examine correlations such as the relationship between distance and price, or how Cleanliness and Worthiness ratings influence the summary score.