

# Bert's Best stays

An AI powered search engine

# Purpose

- The purpose of this code is to give a list of the best hostel based on the user's preferences.
- This uses random forest to generate an overall score of each hostel and then compiles a list of the 10 best hostel.
- This is only for selected hostel in Japan.
- I chose to only focus on hostel since this tool is for students.



# Data

- I acquired a dataset from kaggle of Hostel`s in Japan
- Link:<https://www.kaggle.com/datasets/koki25ando/hostel-world-dataset>
- I then used Pandas to remove NA values useless categories like Lat, long and ID
- I also changed the Rating band from Good, Very Good, Fabulous, Superb to a numerical value out of 4, (I later changed this to out of 10)
- Also I removed the “km from city centre” to obtain an integer
- I saved this to a “clean” data file to ensure the original dataset is not changed in case of mistakes

Original

D	hostel_name	City	price.from	Distance	summary.scr	rating.band	atmosphere	cleanliness	facilities	location.y	security	staff	valueform	lon	lat
1	Bike & Bed C	Osaka	3300	2.9km from c	9.2	Superb	8.9	9.4	9.3	8.9	9	9.4	9.4	135.513767	34.682678
2	& And Hostel	Fukuoka-City	2600	0.7km from r	9.5	Superb	9.4	9.7	9.5	9.7	9.2	9.7	9.5	NA	NA
3	& And Hostel	Tokyo	3600	7.8km from r	8.7	Fabulous	8	7	9	8	10	10	9	139.777472	35.6974473
4	& And Hostel	Tokyo	2600	8.7km from r	7.4	Very Good	8	7.5	7.5	7.5	7	8	6.5	139.783667	35.712716
5	& And Hostel	Tokyo	1500	10.5km from	9.4	Superb	9.5	9.5	9	9	9.5	10	9.5	139.798371	35.7278979
6	1night1980h	Tokyo	2100	9.4km from r	7	Very Good	5.5	8	6	6	8.5	8.5	6.5	139.78695	35.724384
7	328 Hostel & Tokyo		3300	16.5km from	9.3	Superb	8.7	9.7	9.3	9.1	9.3	9.7	8.9	139.745467	35.5480439
8	36Hostel	Hiroshima	2000	1.6km from r	9.5	Superb	8.8	9.9	9.2	9.6	9.8	9.8	9.5	NA	NA
9	3Q House - A	Tokyo	2500	10.2km from NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
10	Ace Inn Shinj	Tokyo	2200	3km from cit	7.7	Very Good	6.7	7.2	6.8	8.5	7.8	8.5	8.1	139.724304	35.6925119
11	Air Osaka Hc	Osaka	1600	9.7km from r	9.2	Superb	9.5	9.1	8.7	8.8	8.9	9.8	9.5	135.476956	34.6222596
12	Aizuya Inn	Tokyo	2000	10.6km from	8.5	Fabulous	8.1	8.3	8.4	7.8	8.9	9.1	8.9	139.800993	35.727547
13	Akihbara Hc	Tokyo	2200	8km from cit	10	Superb	10	10	10	10	10	10	10	139.779382	35.6974904
14	Almond host	Tokyo	2900	2.2km from r	9.3	Superb	9.1	9.5	8.8	9.5	9.4	9.7	9	139.687535	35.6700092
15	Anne Hostel	Tokyo	2000	8.9km from r	9.1	Superb	8.8	9.2	8.7	9	9.1	9.5	9.2	139.78936	35.6989448
16	Anne Hostel	Tokyo	1800	9.5km from r	9.1	Superb	8.8	9.1	9	9.2	9.3	9.3	9.2	139.796798	35.6954897
17	Ark Hostel	Osaka	2700	1.9km from r	9	Superb	8.7	9.2	8.9	8.8	9.1	9.2	9	103.873135	1.3108203
18	ARTsHELTE	Tokyo	2400	10.5km from	8.3	Fabulous	8.2	8.1	7.7	7.9	9.1	9	8.3	139.742108	35.604242
19	AS House (A; Tokyo		1600	10km from c	7.8	Very Good	7.9	6.2	7	9	8.2	8	8.3	139.800616	35.707363
20	Asakusa Hos	Tokyo	1300	9.3km from r	8.9	Fabulous	7.9	9.4	8.3	9.5	9.2	8.9	9.2	139.79081	35.713177

Clean

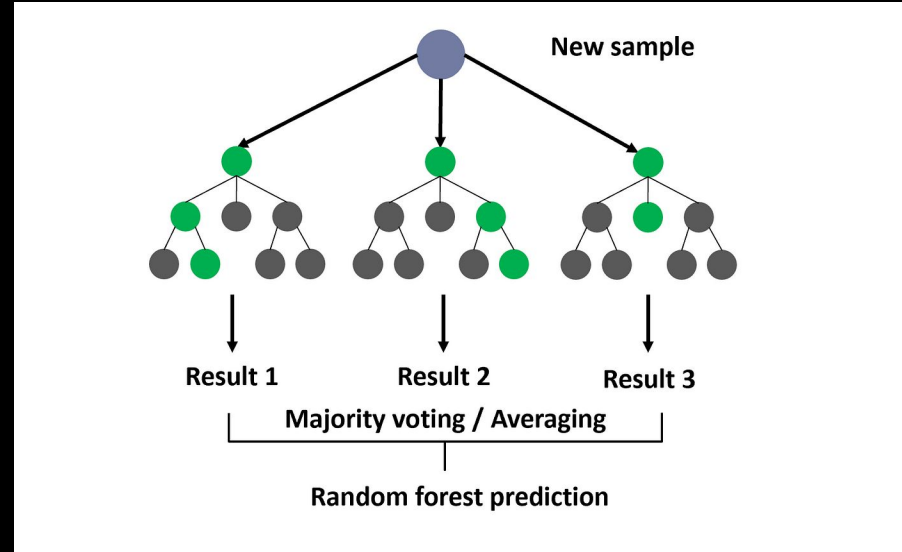
hostel_name	City	price.from	Distance from	summary.scr	rating.band	atmosphere	cleanliness	facilities	location.y	security	staff	valueform	User rating
Bike & Bed C	Osaka	3300	2.9	9.2	3	8.9	9.4	9.3	8.9	9	9.4	9.4	8
& And Hostel	Fukuoka-City	2600	0.7	9.5	3	9.4	9.7	9.5	9.7	9.2	9.7	9.5	8
& And Hostel	Tokyo	3600	7.8	8.7	0	8	7	9	8	10	10	9	2
& And Hostel	Tokyo	2600	8.7	7.4	4	8	7.5	7.5	7.5	7	8	6.5	10
& And Hostel	Tokyo	1500	10.5	9.4	3	9.5	9.5	9	9	9.5	10	9.5	8
1night1980h	Tokyo	2100	9.4	7	4	5.5	8	6	6	8.5	8.5	6.5	10
328 Hostel & Tokyo		3300	16.5	9.3	3	8.7	9.7	9.3	9.1	9.3	9.7	8.9	8
36Hostel	Hiroshima	2000	1.6	9.5	3	8.8	9.9	9.2	9.6	9.8	9.8	9.5	8
Ace Inn Shinj	Tokyo	2200	3	7.7	4	6.7	7.2	6.8	8.5	7.8	8.5	8.1	10
Air Osaka Hc	Osaka	1600	9.7	9.2	3	9.5	9.1	8.7	8.8	8.9	9.8	9.5	8
Aizuya Inn	Tokyo	2000	10.6	8.5	0	8.1	8.3	8.4	7.8	8.9	9.1	8.9	2
Akihbara Hc	Tokyo	2200	8	10	3	10	10	10	10	10	10	10	8
Almond host	Tokyo	2900	2.2	9.3	3	9.1	9.5	8.8	9.5	9.4	9.7	9	8
Anne Hostel	Tokyo	2000	8.9	9.1	3	8.8	9.2	8.7	9	9.1	9.5	9.2	8
Anne Hostel	Tokyo	1800	9.5	9.1	3	8.8	9.1	9	9.2	9.3	9.3	9.2	8
Ark Hostel	Osaka	2700	1.9	9	3	8.7	9.2	8.9	8.8	9.1	9.2	9	8
ARTsHELTE	Tokyo	2400	10.5	8.3	0	8.2	8.1	7.7	7.9	9.1	9	8.3	2
AS House (A; Tokyo		1600	10	7.8	4	7.9	6.2	7	9	8.2	8	8.3	10
Asakusa Hos	Tokyo	1300	9.3	8.9	0	7.9	9.4	8.3	9.5	9.2	8.9	9.2	2

## Filter based on input

- Originally I used a simple code using to create a new dataset using the user`s inputs to based on the city, distance from city centre and price. I kept it simple at first to make it easier from me to work on the random forest code.
- For the User Interface version I used ipywidgets to create interactive widgets to make it more user friendly, for the max and min price and distance I just used excel functions in the csv to find these and then integrated them into the slider.

# Random forest

- I decided to use Random Forest as my Machine learning algorithm since I will use it in another project and is useful for this size dataset.
- I used this youtube tutorial: <https://www.youtube.com/watch?v=YYjvkSJoui4&t=1514s> (however I used random forest regressor not classifier since my rating is numerical)



# Random Forest code explanation

- As my test data (Y) I used the summary score which was provided by the original dataset.
- As my Train data (X) I used the rating.band, atmosphere score, cleanliness score, facilities score, location score, security score, staff score, value for money score and the user rating. Random Forest allowed me to easily use multiple inputs.
- The random state is set to 30 I found this as the best value, this just ensures that the same randomness is kept for each decision tree.
- I used model.predict to then create a new predicted score based on the inputs.
- I then calculated the error which was relatively high but tried to minimise as much as possible by changing the test size.

# Final result

- I finally used `IPython.display` to display the results of the 10 best hotels in a nice table.

Top 10 Hostels based on RandomForestClassifier Analysis:														
	hostel.name	City	price.from	Distance	summary.score	rating.band	atmosphere	cleanliness	facilities	location.y	security	staff	valueformoney	predi
110	Guesthouse Morizou Female only	Osaka	2000	3.7	10.0	8	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0
304	The Blend Inn	Osaka	2300	4.5	9.6	8	9.8	9.9	9.5	8.5	10.0	9.9	9.9	9.6
46	Capsule Hotel Asahi Plaza Shinsaibashi	Osaka	3000	4.0	9.9	8	10.0	10.0	10.0	10.0	10.0	10.0	10.0	9.0
270	Poly Hostel Osaka	Osaka	3800	5.9	9.4	8	8.0	10.0	10.0	8.0	10.0	10.0	10.0	10.0
145	Hostel Jin	Osaka	2500	4.3	9.6	8	8.8	10.0	9.8	8.8	9.8	9.8	9.8	10.0
256	Osaka Hana Hostel	Osaka	2600	3.7	9.5	8	8.5	9.6	9.6	9.8	9.5	9.8	9.8	9.6
156	Hostel Rakutsuki	Osaka	1900	4.7	9.5	8	9.3	9.5	8.7	9.8	9.8	9.8	9.8	9.8
189	J-Hoppers Osaka Guesthouse	Osaka	2300	1.9	9.5	8	9.5	9.6	9.2	9.5	9.6	9.7	9.7	9.6

# Remarks

- In the youtube tutorial RandomForestClassifier is used but I used RandomForestRegressor as my prediction is numerical and it is better for ranking, since I found with classifier I would get a few top results with a score of 10 and many results with 9, which meant that the ranking would show it based on the ID not the score.
- I used ChatGPT mostly to aid me make the User interface. I used it to create the widgets and display the table in a nice manner instead of just printing the csv file. I also used it to arrange my code in a nice manner so it is easier to read and to add description to each line.