Data clean project: base on International Hotel Booking

In [1]:
```
import pandas as pd
import numpy as np
```

In [55]:
```
hotel = pd.read_csv('hotels.csv')
hotel.head()
```

Out[55]:

	hotel_id	hotel_name	city	country	star_rating	lat	lon	cleanliness_base
0	1	The Azure Tower	New York	United States	5	40.7580	-73.9855	9.1
1	2	The Royal Compass	London	United Kingdom	5	51.5072	-0.1276	9.0
2	3	L'Étoile Palace	Paris	France	5	48.8566	2.3522	8.8
3	4	Kyo-to Grand	Tokyo	Japan	5	35.6895	139.6917	9.6
4	5	The Golden Oasis	Dubai	United Arab Emirates	5	25.2769	55.2962	9.3

In [26]:
```
hotel.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25 entries, 0 to 24
Data columns (total 13 columns):
 #   Column               Non-Null Count  Dtype  
---  ------               --------------  -----  
 0   hotel_id             25 non-null     int64  
 1   hotel_name           25 non-null     object 
 2   city                 25 non-null     object 
 3   country              25 non-null     object 
 4   star_rating          25 non-null     int64  
 5   lat                  25 non-null     float64
 6   lon                  25 non-null     float64
 7   cleanliness_base     25 non-null     float64
 8   comfort_base         25 non-null     float64
 9   facilities_base      25 non-null     float64
 10  location_base        25 non-null     float64
 11  staff_base           25 non-null     float64
 12  value_for_money_base  25 non-null     float64
dtypes: float64(8), int64(2), object(3)
memory usage: 2.7+ KB
```

dropping columns we dont need

In [58]:
```
hotel = hotel.drop(columns=['lat', 'lon', 'staff_base'], errors='ignore')
hotel.head(30)
```

Out[58]:

	hotel_id	hotel_name	city	country	star_rating	cleanliness_base	comfort_ba
0	1	The Azure Tower	New York	United States	5	9.1	
1	2	The Royal Compass	London	United Kingdom	5	9.0	
2	3	L'Étoile Palace	Paris	France	5	8.8	
3	4	Kyo-to Grand	Tokyo	Japan	5	9.6	
4	5	The Golden Oasis	Dubai	United Arab Emirates	5	9.3	
5	6	Marina Bay Zenith	Singapore	Singapore	5	9.2	
6	7	Sydney Harbour Grand	Sydney	Australia	5	8.9	
7	8	Copacabana Lux	Rio de Janeiro	Brazil	5	9.0	
8	9	Berlin Mitte Elite	Berlin	Germany	5	9.3	
9	10	The Maple Grove	Toronto	Canada	5	9.4	
10	11	The Bund Palace	Shanghai	China	5	9.1	
11	12	Aztec Heights	Mexico City	Mexico	5	8.7	
12	13	The Gateway Royale	Mumbai	India	5	8.9	
13	14	Colosseum Gardens	Rome	Italy	5	9.2	
14	15	Table Mountain View	Cape Town	South Africa	5	9.0	
15	16	Han River Oasis	Seoul	South Korea	5	9.3	
16	17	Kremlin Suites	Moscow	Russia	5	9.1	
17	18	Nile Grandeur	Cairo	Egypt	5	8.8	

	hotel_id	hotel_name	city	country	star_rating	cleanliness_base	comfort_ba
18	19	Gaudi's Retreat	Barcelona	Spain	5	9.2	
19	20	The Orchid Palace	Bangkok	Thailand	5	9.0	
20	21	The Bosphorus Inn	Istanbul	Turkey	5	9.1	
21	22	Canal House Grand	Amsterdam	Netherlands	5	9.4	
22	23	Tango Boutique	Buenos Aires	Argentina	5	8.9	
23	24	The Savannah House	Lagos	Nigeria	5	8.7	
24	25	The Kiwi Grand	Wellington	New Zealand	5	9.3	

In [62]:
```
hotel['city'].value_counts()
```

Out[62]:
```
city
New York          1
London            1
Paris             1
Tokyo             1
Dubai             1
Singapore         1
Sydney            1
Rio de Janeiro    1
Berlin            1
Toronto           1
Shanghai          1
Mexico City       1
Mumbai            1
Rome              1
Cape Town         1
Seoul             1
Moscow            1
Cairo             1
Barcelona         1
Bangkok           1
Istanbul          1
Amsterdam         1
Buenos Aires      1
Lagos             1
Wellington        1
Name: count, dtype: int64
```

average score of each hotel

In [74]:
```
hotel['average_score'] = hotel[['cleanliness_base','comfort_base','facilities_base'
hotel.head()
```

Out[74]:

	hotel_id	hotel_name	city	country	star_rating	cleanliness_base	comfort_base	fac
0	1	The Azure Tower	New York	United States	5	9.1	8.8	
1	2	The Royal Compass	London	United Kingdom	5	9.0	9.2	
2	3	L'Étoile Palace	Paris	France	5	8.8	9.4	
3	4	Kyo-to Grand	Tokyo	Japan	5	9.6	9.0	
4	5	The Golden Oasis	Dubai	United Arab Emirates	5	9.3	9.5	

In [75]:
```
#so now we have the average score of each hotel
```

In [78]:
```
hotel.index = hotel['hotel_name']
hotel.head()
```

Out[78]:

hotel_name	hotel_id	hotel_name	city	country	star_rating	cleanliness_base	comfort
The Azure Tower	1	The Azure Tower	New York	United States	5	9.1	
The Royal Compass	2	The Royal Compass	London	United Kingdom	5	9.0	
L'Étoile Palace	3	L'Étoile Palace	Paris	France	5	8.8	
Kyo-to Grand	4	Kyo-to Grand	Tokyo	Japan	5	9.6	
The Golden Oasis	5	The Golden Oasis	Dubai	United Arab Emirates	5	9.3	

In [80]:
```
hotel.drop(columns=['hotel_name', 'hotel_id'], errors='ignore', inplace=True)
```

In [82]:
```
hotel.head()
```

Out[82]:

hotel_name	city	country	star_rating	cleanliness_base	comfort_base	facilities_base	I
The Azure Tower	New York	United States	5	9.1	8.8	8.9	
The Royal Compass	London	United Kingdom	5	9.0	9.2	8.8	
L'Étoile Palace	Paris	France	5	8.8	9.4	8.7	
Kyo-to Grand	Tokyo	Japan	5	9.6	9.0	9.3	
The Golden Oasis	Dubai	United Arab Emirates	5	9.3	9.5	9.6	

In [88]:
```
hotel['city'].value_counts()
hotel['country'].value_counts()
```

Out[88]:
```
country
United States           1
United Kingdom          1
France                  1
Japan                   1
United Arab Emirates    1
Singapore               1
Australia               1
Brazil                  1
Germany                 1
Canada                  1
China                   1
Mexico                  1
India                   1
Italy                   1
South Africa            1
South Korea             1
Russia                  1
Egypt                   1
Spain                   1
Thailand                1
Turkey                  1
Netherlands             1
Argentina               1
Nigeria                 1
New Zealand             1
Name: count, dtype: int64
```

In [89]:
```
#looks like we dont have to group anything since theres no duplicates cities or cou
```

In [105…
```
hotel_ranked = hotel.sort_values(by='average_score', ascending=False)
```

```
hotel_ranked.head(30)
```

Out[105…

hotel_name	city	country	star_rating	cleanliness_base	comfort_base	facilities_
Canal House Grand	Amsterdam	Netherlands	5	9.4	9.1	
The Bund Palace	Shanghai	China	5	9.1	9.0	
The Golden Oasis	Dubai	United Arab Emirates	5	9.3	9.5	
Colosseum Gardens	Rome	Italy	5	9.2	9.1	
Marina Bay Zenith	Singapore	Singapore	5	9.2	9.1	
Gaudi's Retreat	Barcelona	Spain	5	9.2	9.1	
The Maple Grove	Toronto	Canada	5	9.4	9.3	
The Kiwi Grand	Wellington	New Zealand	5	9.3	9.2	
Berlin Mitte Elite	Berlin	Germany	5	9.3	9.2	
The Orchid Palace	Bangkok	Thailand	5	9.0	9.2	
Han River Oasis	Seoul	South Korea	5	9.3	9.4	
Kremlin Suites	Moscow	Russia	5	9.1	9.3	
Table Mountain View	Cape Town	South Africa	5	9.0	9.2	
The Bosphorus Inn	Istanbul	Turkey	5	9.1	9.0	
Kyo-to Grand	Tokyo	Japan	5	9.6	9.0	
Sydney Harbour Grand	Sydney	Australia	5	8.9	9.0	
L'Étoile Palace	Paris	France	5	8.8	9.4	

hotel_name	city	country	star_rating	cleanliness_base	comfort_base	facilities_
Copacabana Lux	Rio de Janeiro	Brazil	5	9.0	8.7	
The Azure Tower	New York	United States	5	9.1	8.8	
The Royal Compass	London	United Kingdom	5	9.0	9.2	
The Gateway Royale	Mumbai	India	5	8.9	8.8	
Tango Boutique	Buenos Aires	Argentina	5	8.9	8.8	
Aztec Heights	Mexico City	Mexico	5	8.7	8.9	
Nile Grandeur	Cairo	Egypt	5	8.8	8.7	
The Savannah House	Lagos	Nigeria	5	8.7	8.6	

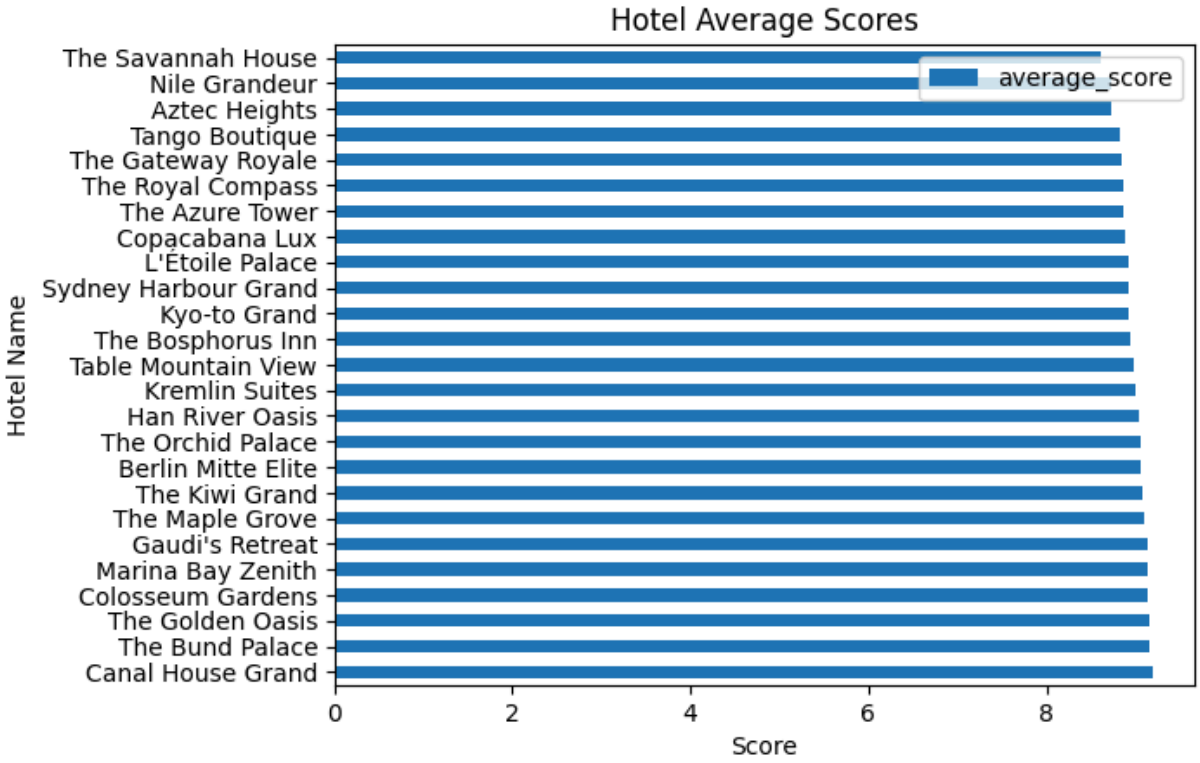
From above we have the final table.

In [96]:
```
#now we have the ranking among those hotels
```

In [97]:
```
import matplotlib.pyplot as plt
```

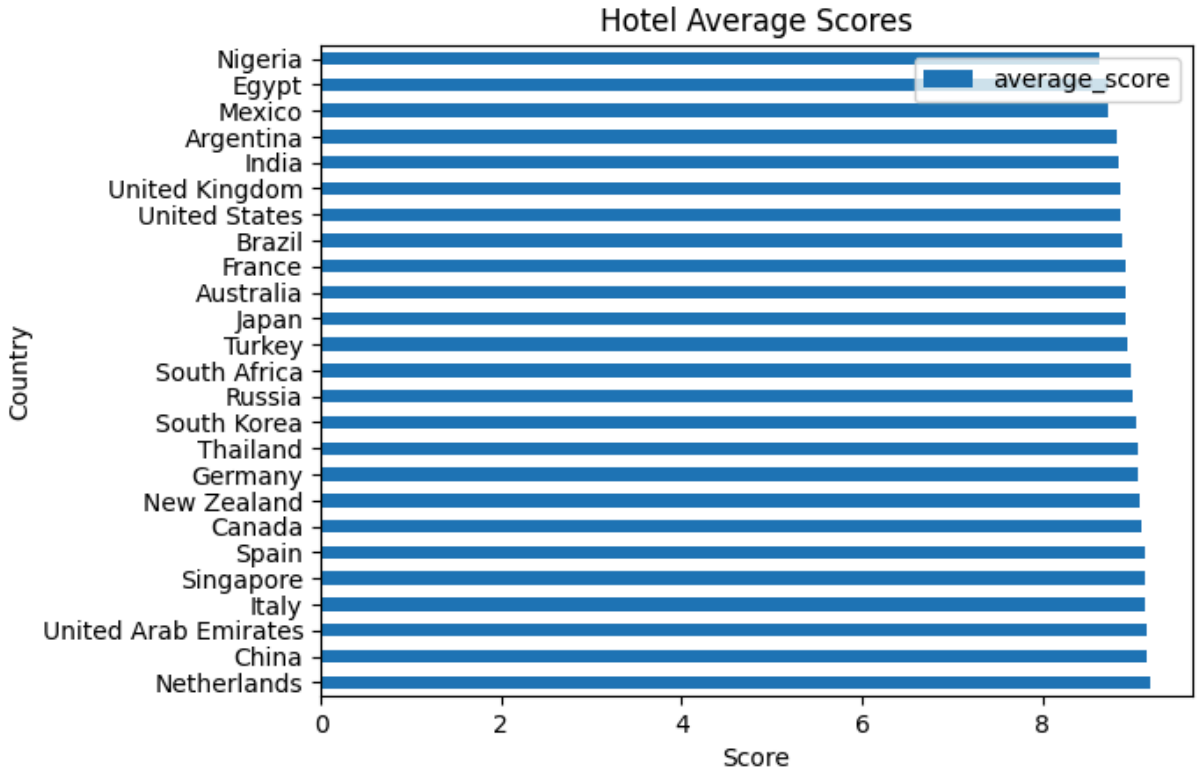
In [102…
```
hotel_ranked.plot(y='average_score', kind='barh', title='Hotel Average Scores')
plt.xlabel('Score')
plt.ylabel('Hotel Name')
```

Out[102…
```
Text(0, 0.5, 'Hotel Name')
```



In [107…
```
hotel_ranked.plot(x='country', y='average_score', kind='barh', title='Hotel Average
plt.xlabel('Score')
plt.ylabel('Country')
```

Out[107…
```
Text(0, 0.5, 'Country')
```



Conclusion: The ranking of 25 luxury hotels shows clear performance differences, even among 5-star properties. Top performers like Canal House Grand (Amsterdam), The Bund Palace (Shanghai), and The Golden Oasis (Dubai) score above 9.15, excelling in cleanliness, comfort, location, and value for money. Lower-ranked hotels, such as The Savannah House (Lagos) and Nile Grandeur (Cairo), are held back mainly by weaker value and facilities. Overall, balanced performance across all dimensions drives high rankings, highlighting areas for strategic improvement.