INTRODUCTION

The following data analysis evaluates 15 holiday destinations for a travel website to establish the most popular, and which should be excluded in future recommendations.

The study uses previous holidaymaker feedback data and average hotel star ratings for each location. The number of all-inclusive hotels within each destination is also examined against destination scores to establish if there is any correlation.

Python and Pandas library were utilised to produce this insight.

Yenny Aplin

1. STATE NUMBER OF ROWS AND COLUMNS IN THE FILE

Country	Score	Avg_StartRating	Allinc_Hotels	Top_City
Peru	10	4	20	Cusco
Spain	6	3	15	Barcelona
United Kingdom	7	4	25	London
Italy	9	4	16	Venice
South Africa	3	2	8	Cape Town
Mexico	7	4	14	Cancun
France	8	4	19	Paris
Morocco	1	2	7	Marrakesh
Greece	5	3	13	Athens
Australia	6	3	17	Sidney
Austria	5	5	6	Vienna
Japan	6	3	17	Tokio
USA	8	4	21	New York
Argentine	5	3	19	Buenos Aires
Turkey	4	3	7	Ancara

Holiday.csv

OUTPUT

```
(15, 5)
```

CODE

```
import pandas as pd
holiday_data= pd.read_csv("Holiday.csv")
#shows number of rows and columns
holiday_data.shape
```

✓ 0.7s

2a. PRINT ROW 3-8 (USING ILOC)

OUTPUT

	Country	Score	Avg_StartRating	Allinc_Hotels	Top_City
3	Italy	9	4	16	Venice
4	South Africa	3	2	8	Cape Town
5	Mexico	7	4	14	Cancun
6	France	8	4	19	Paris
7	Morocco	1	2	7	Marrakesh

CODE

```
import pandas as pd
holiday_data= pd.read_csv("Holiday.csv")
print(holiday_data.iloc[3:8])
```

✓ 1.6s

2a. PRINT ROW 3-8 (USING LOC)

OUTPUT

Country	Score	Avg_StartRating	Allinc_Hotels	Top_City
Italy	9	4	16	Venice
South Africa	3	2	8	Cape Town
Mexico	7	4	14	Cancun
France	8	4	19	Paris
Morocco	1	2	7	Marrakesh

CODE

```
import pandas as pd
holiday_data = pd.read_csv("Holiday.csv", index_col = "Country")
print(holiday_data.loc[["Italy", "South Africa", "Mexico", "France", "Morocco"]])
```

✓ 0.5s

3. FIND THE MEAN NUMBER OF ALL-INCLUSIVE HOTELS ACROSS ALL DESTINATIONS

OUTPUT

```
14.933333333333334
```

CODE

```
# Find the mean number of all-inclusive hotels across all destinations
import pandas as pd
holiday_data= pd.read_csv("Holiday.csv")
holiday_data["Allinc_Hotels"].mean()
```

✓ 1.2s

4. FIND THE LOWEST SCORING DESTINATION

OUTPUT

	Country	Score	Avg_StartRating	Allinc_Hotels	Top_City
7	Morocco	1	2	7	Marrakesh

CODE

```
import pandas as pd
holiday_data= pd.read_csv("Holiday.csv")
filter_min = holiday_data["Score"].min()
lowest_country = holiday_data["Score"] == filter_min
holiday_data[lowest_country]
```

✓ 0.2s

5. FIND THE HIGHEST SCORING DESTINATION

OUTPUT

	Country	Score	Avg_StartRating	Allinc_Hotels	Top_City
0	Peru	10	4	20	Cusco

CODE

```
import pandas as pd
holiday_data= pd.read_csv("Holiday.csv")
filter_max = holiday_data["Score"].max()
highest_country = holiday_data["Score"] == filter_max
holiday_data[highest_country]
```

✓ 0.1s

6. FIND ALL THE DESTINATIONS WHERE THERE ARE MORE THAN 9 ALL-INCLUSIVE HOTELS

OUTPUT

	Country	Score	Avg_StartRating	Allinc_Hotels	Top_City
0	Peru	10	4	20	Cusco
1	Spain	6	3	15	Barcelona
2	United Kingdom	7	4	25	London
3	Italy	9	4	16	Venice
5	Mexico	7	4	14	Cancun
6	France	8	4	19	Paris
8	Greece	5	3	13	Athens
9	Australia	6	3	17	Sidney
11	Japan	6	3	17	Tokio
12	USA	8	4	21	New York
13	Argentine	5	3	19	Buenos Aires

CODE

```
import pandas as pd
holiday_data= pd.read_csv("Holiday.csv")
filter_allinc = holiday_data["Allinc_Hotels"] > 9
highest_allinc = holiday_data[filter_allinc]
highest_allinc
```

✓ 0.1s

7. FILTER THE DATA BY SCORE ABOVE 8

OUTPUT

	Country	Score	Avg_StartRating	Allinc_Hotels	Top_City
0	Peru	10	4	20	Cusco
3	Italy	9	4	16	Venice

CODE

```
import pandas as pd
holiday_data= pd.read_csv("Holiday.csv")
filter_score = holiday_data["Score"] > 8
highscore = holiday_data[filter_score]
highscore
```

✓ 0.4s

8. FILTER THE DATA SCORE BELOW 2 (TO IDENTIFY DESTINATIONS TO BE REMOVED)

OUTPUT

	Country	Score	Avg_StartRating	Allinc_Hotels	Top_City
7	Morocco	1	2	7	Marrakesh

CODE

```
import pandas as pd
holiday_data= pd.read_csv("Holiday.csv")
filter_score1 = holiday_data["Score"] < 2
lowest_score = holiday_data[filter_score1]
lowest_score.head()
```

✓ 0.6s

9. CORRELATION BETWEEN NUMBER OF ALL-INCLUSIVE HOTELS AND SCORE

OUTPUT

```
0.7234970592944953
```

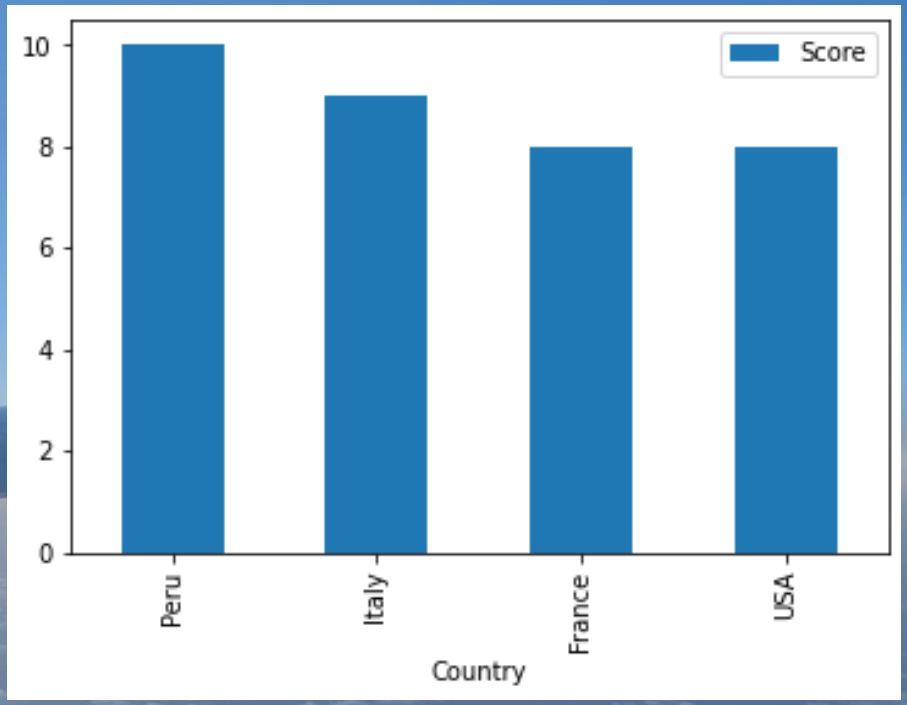
CODE

```
import pandas as pd
holiday_data= pd.read_csv("Holiday.csv")
allinc_hotels  = holiday_data["Allinc_Hotels"]
score_column = holiday_data["Score"]
correlation = allinc_hotels.corr(score_column)
print(correlation)
```

✓ 0.5s

10. DESTINATION AND HIGHEST SCORES

OUTPUT



CODE

```
holiday_data= pd.read_csv("Holiday.csv")
filter_TopScore = holiday_data["Score"] > 7
holiday_highscore = holiday_data[filter_TopScore]
holiday_highscore.plot.bar(x='Country', y='Score')
```

✓ 0.2s

CONCLUSION

- The **mean number** of all-inclusive hotels across all destinations was **14.9** with 11 destinations having more than 9 hotels
- The **lowest** scoring destination was **Morocco** and the **highest** scoring destination was **Peru**
- There is a **moderate to good** correlation **(0.72 coefficient)** between the destination score and number of all-inclusive hotels
- The destination that should be excluded from future considerations is **Morocco** with a score below 2
- **Peru, Italy, France** and **USA** had the highest destination scores (over 7), and are therefore recommended holiday locations