

Week 4

Task 1

```
import pandas as pd

data = pd.read_csv("holiday_destinations.csv")

data.shape

(15, 5)
```

From the above output, we have 15 rows and 5 columns in the CSV file.

Task 2

```
print(data.iloc[2:8])
```

	Destination	Feedback_Score	Av_Hotel_Star	All_Inclusive	Popular_City
2	Crete	6.4	3.41	5	Heraklion
3	Mexico	7.6	3.57	8	Cancun
4	Vietnam	4.8	2.92	3	Ho Chi Minh
5	Ghana	3.4	2.95	2	Accra
6	Croatia	5.7	3.24	20	Dubrovnik
7	Iceland	6.3	3.22	6	Reykjavik

Above is printed rows 3-8 of the data since we use zero-indexing.

Task 3

```
data["All_Inclusive"].mean()
```

8.466666666666667

Here we have deduced the mean number of all-inclusive hotels.

Task 4

```
minscore = data["Feedback_Score"].idxmin()

print(data.iloc[minscore])
```

Destination	Ghana
Feedback_Score	3.4
Av_Hotel_Star	2.95
All_Inclusive	2
Popular_City	Accra
Name: 5, dtype: object	

Here we have pulled the data for the destination with the lowest feedback score by first locating its index and using `iloc` to print the data within this index.

Task 5

```
maxscore = data["Feedback_Score"].idxmax()  
print(data.iloc[maxscore])
```

```
Destination      Jamaica  
Feedback_Score    8.7  
Av_Hotel_Star     3.44  
All_Inclusive     6  
Popular_City      Montego Bay  
Name: 1, dtype: object
```

Here we have pulled the data for the highest scoring destination using the same process used in Task 4.

Task 6

```
filter1 = data["All_Inclusive"] > 9  
allinc = data[filter1]  
allinc.head()
```

	Destination	Feedback_Score	Av_Hotel_Star	All_Inclusive	Popular_City
0	Dominican Republic	7.8	3.67	13	Punta Cana
6	Croatia	5.7	3.24	20	Dubrovnik
8	Italy	4.2	3.30	28	Rome
10	Albania	4.3	3.33	11	Tirana
12	Catalonia	7.1	3.25	12	Barcelona

Above is the data for all destinations with more than 9 all-inclusive hotels.

Task 7

```
filter2 = data["Feedback_Score"] > 8  
highestscores = data[filter2]  
highestscores[["Destination", "Feedback_Score"]]
```

	Destination	Feedback_Score
1	Jamaica	8.7
13	Barbados	8.3

Here we have returned destination names and feedback scores for all destinations with a feedback score greater than 8.

Task 8

```
filter3 = data["Feedback_Score"] < 4  
lowestscores = data[filter3]  
lowestscores[["Destination", "Feedback_Score"]]
```

	Destination	Feedback_Score
5	Ghana	3.4
11	Hong Kong	3.8

Here we have filtered the data by destination and feedback scores below 4. To gain some clarity on why these destinations have low feedback scores, I used the same process to return the data for destinations with the lowest average hotel star ratings.

```
filter4 = data["Av_Hotel_Star"] < 3
loweststar = data[filter4]
loweststar[["Destination", "Av_Hotel_Star"]]
```

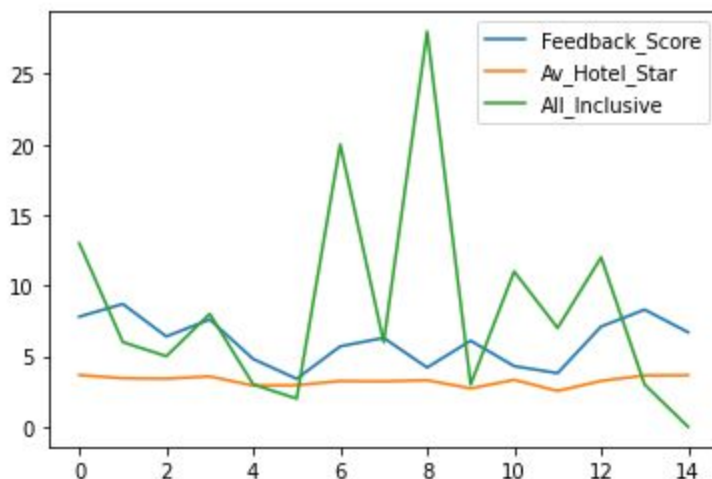
	Destination	Av_Hotel_Star
4	Vietnam	2.92
5	Ghana	2.95
9	Hungary	2.72
11	Hong Kong	2.54

From this, we see that both of the locations with the lowest feedback scores have an average hotel star rating below 3, suggesting a possible correlation in the data between average hotel stars and feedback scores.

Task 9

```
data.plot()
```

<AxesSubplot:>



Though there are some similarities in the data trends of feedback scores and number of all-inclusive hotels, there are too many anomalies to validate a correlation between the two. This is due to the fact that many of the destinations with fewer all-inclusive hotels offer alternative luxury accommodation stays which have not been accounted for within this dataset. We do, however, see that there is a positive correlation between the average hotel stars (though there is less deviation in these scores) and the feedback scores, as the average hotel star of a destination does take into consideration various

different accommodation options. From this, we can conclude that destinations with a higher average hotel star rating receive higher feedback scores.

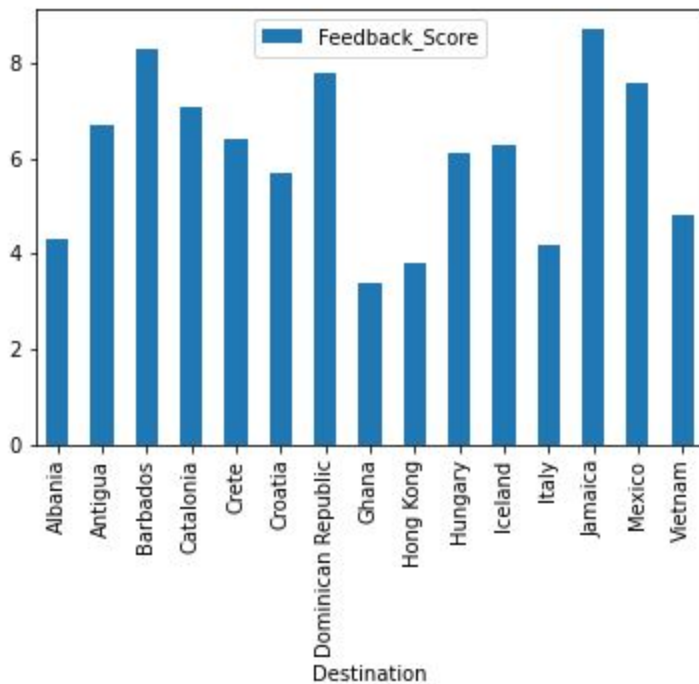
Task 10

```
df = pd.DataFrame(data)

ds = df[["Destination", "Feedback_Score"]]

ds.groupby("Destination").mean().plot.bar()

<AxesSubplot:xlabel='Destination'>
```



Above is a visualisation of feedback scores out of 10 for each destination.