DESIGN LAYOUT

WHY PANDAS?

It clean messy datasets, make them readable and relevant.

Because relevant data is important for Data Science.

The Data I decided to use is based on the black community travelers

Metaplotlib was used to plot the barr chart.

DESTINATION	FEEDBACK	AVERAGE	ALL INCLUSIVE HOTELS	MOST VISTITED CITY
Nigeria	9	4	200	Lagos
Bali	7	5	100	Denpasar
Maldives	10	5	200	Meeru Island
Japan	8	5	250	Tokyo
UAE	9	5	250	Dubai
USA	7	4	250	Las vegas
Spain	8	4	15	Tenerife
Italy	8	4	150	Venice
Greece	9	4	160	Piraeus
France	6	5	140	Paris
Bangkok	7	4	130	Thailand
China	6	4	140	Macau
Istanbul	7	5	130	Turkey
India	6	3	120	Delhi

Task1 How many rows and columns are there in your file?

```python
import pandas as pd

# Create data
data = pd.read_csv('C:/Users/mylaptop/Documents/Data-python/week2/15 holidaydestination.csv')
#Display number of rows, columns, etc.
#df.info()
#Get the number of rows and columns: df.shape
row, col = data.shape
print(row)

print(col)



# Set index
#df = df.set_index('Keyword')

#for dest, group in df.groupby('destination'):
    # group.to_csv(dest + '.csv')

#continents = ['Asia','Europe','Africa','North America','South America','Oceania']
```

✓ 0.3s

14

5

Task2 Print row 3-8 (using iloc/loc).

```python
import pandas as pd
data = pd.read_csv('C:/Users/mylaptop/Documents/Data-python/week2/15 holidaydestination.csv')
data.iloc[3:8]
#df.iloc[0:5] # first five rows of dataframe
#print (type (data.iloc[3:8]))#DataFrame type because list selection used.
```
✓ 0.6s

	DESTINATION	FEEDBACK	AVERAGE HOTEL RATING	ALL INCLUSIVE HOTELS	MOST VISTITED CITY
3	Japan	8	5	250	Tokyo
4	UAE	9	5	250	Dubai
5	USA	7	4	250	Las vegas
6	Spain	8	4	15	Tenerife
7	Italy	8	4	150	Venice

Task3 Find the mean number of all-inclusive hotels across all destinations.

```python
import pandas as pd
data = pd.read_csv('C:/Users/mylaptop/Documents/Data-python/week2/15 holidaydestination.csv')

data['ALL-INCLUSIVE HOTELS'].mean()
```
✓ 0.5s

159.64285714285714

Find the lowest scoring destination.

Markdown

```python
import pandas as pd
data = pd.read_csv('C:/Users/mylaptop/Documents/Data-python/week2/15 holidaydestination.csv')
data = {'DESTINATION':['Nigeria','Bali','Maldives','Japan','UAE','USA','Spain','Italy', 'Greece', 'France', 'Bangkok', 'China', 'Istanbu
        'FEEDBACK':[9,7,10,8,9,7,8,8,9,6,7,6,7,6]}

df = pd.DataFrame(data)

#print(df)
df[ df['FEEDBACK'] == df['FEEDBACK'].min() ]

#df['DESTINATION'].min()
#df['FEEDBACK'].min()
#print(data.DESTINATION.max())
#print(data[data.DESTINATION == data.FEEDBACK.max()])#FEEDBACK 0-10
#data= data.loc[[data.DESTINATION.idxmax(), data.DESTINATION.idxmin()]]
#print (data)
```

✓ 0.4s

	DESTINATION	FEEDBACK
9	France	6
11	China	6
13	India	6

Task5 Find the highest scoring destination

```python
import pandas as pd
data = pd.read_csv('C:/Users/mylaptop/Documents/Data-python/week2/15 holidaydestination.csv')
data = {'DESTINATION':['Nigeria','Bali','Maldives','Japan','UAE','USA','Spain','Italy', 'Greece', 'France', 'Bangkok', 'China', 'Istanbu
        'FEEDBACK':[9,7,10,8,9,7,8,8,9,6,7,6,7,6]}

df = pd.DataFrame(data)

#print(df)
#df['DESTINATION'].min()
#data.max()
df[ df['FEEDBACK'] == df['FEEDBACK'].max() ]

#data['FEEDBACK'].max()
#print(data.DESTINATION.max())
```

✓ 0.4s

Python

	DESTINATION	FEEDBACK
2	Maldives	10

Task6
Find all the destinations where there are more than 9 all-inclusive hotels.

```python
import pandas as pd
data = pd.read_csv('C:/Users/mylaptop/Documents/Data-python/week2/15 holidaydestination.csv')
#df = {'DESTINATION':['Nigeria','Bali','Maldives','Japan','UAE','USA','Spain','Italy', 'Greece', 'France', 'Bangkok', 'China', 'Istanbul
    #'ALL-INCLUSIVE HOTELS':[200,100,200,250,250,15,150,160,140,130,140,130,120]}

#df = pd.DataFrame(data)

#test = df[(df['ALL-INCLUSIVE HOTELS'] > 9)]
test = data[(data['ALL INCLUSIVE HOTELS'] > 9)]
print(test)
```

✓ 0.5s

```
DESTINATION  FEEDBACK  AVERAGE HOTEL RATING  ALL INCLUSIVE HOTELS  \
   Nigeria         9                     4                   200
      Bali         7                     5                   100
  Maldives        10                     5                   200
     Japan         8                     5                   250
       UAE         9                     5                   250
       USA         7                     4                   250
     Spain         8                     4                    15
     Italy         8                     4                   150
    Greece         9                     4                   160
```

Task7
Filter the data by score above 8.

```python
import pandas as pd
data = pd.read_csv('C:/Users/mylaptop/Documents/Data-python/week2/15 holidaydestination.csv')
test = data[(data['FEEDBACK'] > 8)]
print(test)
#data(['data'] > 8)
#print(data > 8)
```

✓ 0.3s

```
   DESTINATION  FEEDBACK  AVERAGE HOTEL RATING  ALL-INCLUSIVE HOTELS  \
0      Nigeria         9                     4                   200
2     Maldives        10                     5                   200
4          UAE         9                     5                   250
8       Greece         9                     4                   160


   MOST VISTITED CITY
0               Lagos
2        Meeru Island
4               Dubai
8             Piraeus
```

Task8 Filter the data score below 2

```python
import pandas as pd
data = pd.read_csv('C:/Users/mylaptop/Documents/Data-python/week2/15 holidaydestination.csv')
test = data[(data['FEEDBACK'] < 2)]
print(test)
```

```
Empty DataFrame
Columns: [DESTINATION, FEEDBACK, AVERAGE HOTEL RATING, ALL-INCLUSIVE HOTELS, MOST VISTITED CITY]
Index: []
```

Task9 Is there a correlation between number of all-inclusive hotels and score?

```python
import pandas as pd
data = pd.read_csv('C:/Users/mylaptop/Documents/Data-python/week2/15 holidaydestination.csv')
#data.get('ALL INCLUSIVE HOTELS')
column_1 = data['ALL INCLUSIVE HOTELS']
column_2 = data['FEEDBACK']
correlation = column_1.corr(column_2)
print(correlation)
```
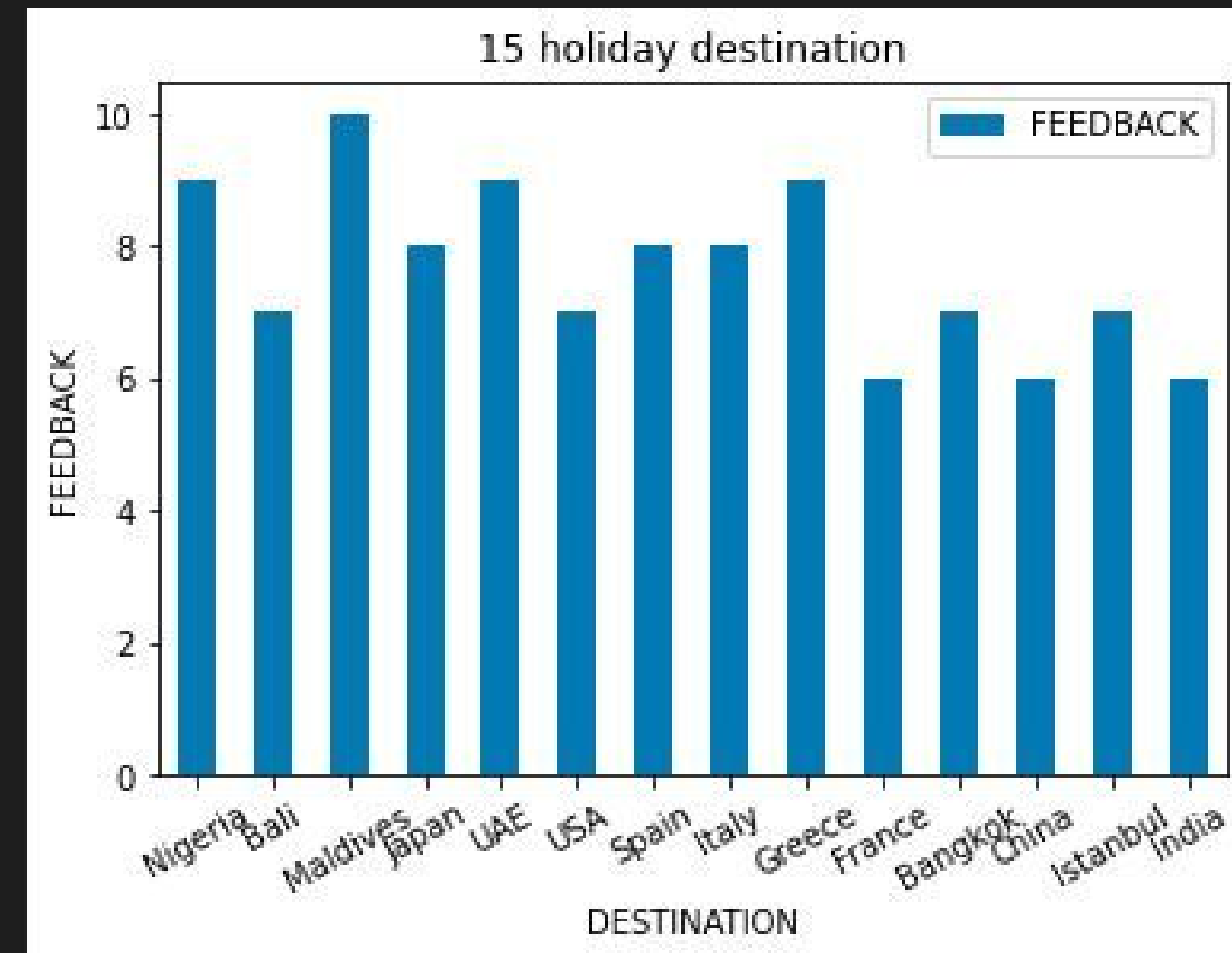
✓ 0.2s

```
0.3639768127222851
```

Create a data visualisation diagram to show destination and hi

Markdown

15 holiday destination

```python
from matplotlib import pyplot as plt
import pandas as pd
#import matplotlib.pyplot as plt
# Create a sample dataframe with an text index
plotdata = pd.DataFrame(
    {"FEEDBACK": [9,7,10,8,9,7,8,8,9,6,7,6,7,6]},
    index=['Nigeria','Bali','Maldives','Japan','UAE','USA','Spain','Italy', 'Greece', 'France', 'Bangkok', 'China', 'Istanbul', 'India']
# Plot a bar chart
plotdata.plot(kind="bar")
#plotdata['pies'].plot(kind="bar", title="test")
# Rotate the x-labels by 30 degrees, and keep the text aligned horizontally
plt.xticks(rotation=30, horizontalalignment="center")
plt.title("15 holiday destination")
plt.xlabel("DESTINATION")
plt.ylabel("FEEDBACK")
```

✓ 1.2s

Python

```
Text(0, 0.5, 'FEEDBACK')
```