Web scraping reviews for British airways

In this assignment, our job is to go to Skytrax website and scrape reviews in an excel sheet.For this purpose the popular pandas package a beautiful soup will be used.

```python
In [1]:    #Getting the required libraries for scraping
           import requests
           from bs4 import BeautifulSoup
           import pandas as pd
```

```python
In [12]:   # Defining the URL for British Airways reviews page on Skytrax
           base_url = 'https://www.airlinequality.com/airline-reviews/british-airways'
```

```python
In [14]:   # Number of pages to scrape
           pages = 20
           # Number of reviews per page
           page_size = 100
```

```python
In [15]:   # Making a request to get the page content
           response = requests.get(url)
```

```python
In [4]:    # Parsing the page content with BeautifulSoup
           soup = BeautifulSoup(response.text, 'html.parser')
```

```python
In [16]:   # List to store reviews
           reviews = []

           # Loop through the desired number of pages
           for i in range(1, pages + 1):

               print(f"Scraping page {i}")

               # Construct the URL for the current page
               url = f"{base_url}/page/{i}/?sortby=post_date%3ADesc&pagesize={page_size}"

               # Send a GET request to the URL
               response = requests.get(url)

               # Check if the request was successful
               if response.status_code == 200:
                   # Parse the content of the page with BeautifulSoup
                   parsed_content = BeautifulSoup(response.content, 'html.parser')

                   # Find all review content
                   for para in parsed_content.find_all("div", class_="text_content"):
                       reviews.append(para.get_text(strip=True))

                   print(f"    ---> {len(reviews)} total reviews collected so far")
               else:
                   print(f"Failed to scrape page {i}. Status code: {response.status_code}")
                   break
```

```
Scraping page 1
    ---> 100 total reviews collected so far
Scraping page 2
    ---> 200 total reviews collected so far
Scraping page 3
    ---> 300 total reviews collected so far
Scraping page 4
```

```
          ---> 400 total reviews collected so far
    Scraping page 5
          ---> 500 total reviews collected so far
    Scraping page 6
          ---> 600 total reviews collected so far
    Scraping page 7
          ---> 700 total reviews collected so far
    Scraping page 8
          ---> 800 total reviews collected so far
    Scraping page 9
          ---> 900 total reviews collected so far
    Scraping page 10
          ---> 1000 total reviews collected so far
    Scraping page 11
          ---> 1100 total reviews collected so far
    Scraping page 12
          ---> 1200 total reviews collected so far
    Scraping page 13
          ---> 1300 total reviews collected so far
    Scraping page 14
          ---> 1400 total reviews collected so far
    Scraping page 15
          ---> 1500 total reviews collected so far
    Scraping page 16
          ---> 1600 total reviews collected so far
    Scraping page 17
          ---> 1700 total reviews collected so far
    Scraping page 18
          ---> 1800 total reviews collected so far
    Scraping page 19
          ---> 1900 total reviews collected so far
    Scraping page 20
          ---> 2000 total reviews collected so far
```

In [17]:
```python
# Convert the list of reviews to a DataFrame
bacr = pd.DataFrame(reviews, columns=['review_text'])
bacr.head()
```

Out[17]:

	review_text
0	✓Trip Verified| British Airways at its best....
1	✓Trip Verified| An excellent flight! Despite t...
2	✓Trip Verified| I recently traveled with Briti...
3	✓Trip Verified| My family and I were booked ...
4	Not Verified| We had to change from AA to BA ...

In [18]:
```python
# Saving the reviews to a CSV file
bacr.to_csv('british_airways_reviews.csv', index=False)
```