Web Scraping to Gain Insights for British Airways

I have analyzed the customer's reviews of British Airways. As we already know, British Airways is the flag carrier airline of the United Kingdom; thousands of flights arrive to and depart from the UK, carrying customers across the world. Meanwhile, British Airways would like to gain deeper insights into customer feelings, needs, and feedback.

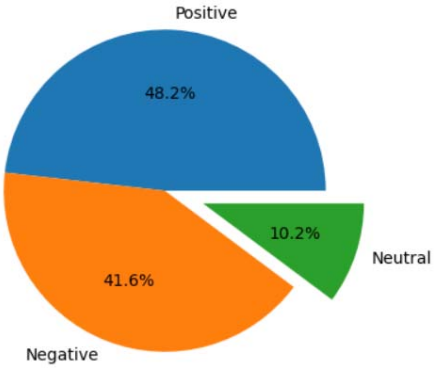
My preliminary analysis indicates:

- **Customers are actively talking about their experience**

  By scraping data from Skytrax, cleaning the data, and processing tokenization, data analysis indicated that customers' top topics are: flight, service, seat, food, crew, time, and late. These top concerns from the customer's review suggested that British Airways needs to improve its overall service.

- **Majority of reviews are polarized as either positive or negative**

  Review analysis indicated:
  48.2% of customers had positive feedback
  41.6% of customers had negative feedback
  10.2% of customers had neutral feedback



To carry on data analysis, I collected the customer reviews of British Airways from the website Skytrax. The data collection process involved the following steps:

1. Web Scraping:
   I used Python along with the Requests library to access the Skytrax website and retrieve the web pages containing customer reviews. This allowed me to gather a substantial amount of text data for analysis.
2. Data Extraction:
   Using the BeautifulSoup library, I parsed the HTML content of the web pages and extracted the relevant information, specifically the customer reviews. These reviews were stored in a structured format for further analysis.
3. Data Quantity:
   I collected a total of 1,000 reviews from the website to ensure a sufficiently large dataset for analysis. The quantity of data is important for obtaining meaningful insights.
4. Data Storage:
   After extracting the reviews, I saved the data to my local computer as a CSV file. This step ensured that the data was readily available for analysis and could be easily imported into data analysis tools.
5. Data Cleaning:
   Before analysis, I performed data cleaning to remove symbols, irrelevant information, and any non-alphabetic items that could potentially introduce noise into the analysis. This step was crucial to ensure the data quality.

By following these steps, I acquired a substantial dataset of customer reviews for British Airways, which was then prepared for analysis to gain insights into customer sentiment and preferences.

The text preprocessing stage was a critical step in transforming the raw customer reviews into a format suitable for analysis. It involved several key processes, each contributing to the refinement of the data:

a. Tokenization:
   Tokenization is the process of breaking down the text into individual units, typically words or phrases, known as tokens. In my analysis, I used Python's NLTK library to tokenize the reviews. Each sentence in a review was split into a list of words or tokens.
b. Stop Word Removal:
   Stop words are common words in a language that do not typically carry significant meaning for analysis. Removing these words is important to focus on the more meaningful content of the reviews.
   To remove stop words, I used NLTK's stop words corpus. Any tokens in the reviews that matched stop words were eliminated from the text.
c. Stemming:

Stemming is the process of reducing words to their root or base form. This helps in consolidating variations of words to their common form. In my analysis, I employed NLTK's stemming tools to achieve this.

By applying these text preprocessing techniques, the data was transformed into a cleaner and more structured format. This made it easier to analyze and extract meaningful information from the reviews. These preprocessing steps ensured that the text data was in a suitable state for subsequent analysis, enhancing the accuracy and efficiency of my insights extraction process.

My analysis has identified the need for British Airways to improve its overall service; several specific recommendations could guide the airline in addressing the concerns raised by customers.

- Punctuality and Timeliness:
  Focus on improving flight punctuality and timeliness. Customers have frequently expressed concerns about delays and late departures. Implement better scheduling, contingency plans for adverse weather conditions, and efficient boarding procedures to minimize delays.
- In-Flight Services:
  Enhance the in-flight experience by upgrading the quality of food, seat comfort, and entertainment options. Consider customer preferences and dietary requirements when designing menus, and invest in more comfortable seating arrangements.
- Data-Driven Decision-Making:
  Continuously monitor and analyze customer feedback to identify trends and areas in need of improvement. Data-driven decision-making can help British Airways adapt to changing customer expectations.

By implementing these specific recommendations, British Airways could proactively address the concerns raised by customers, ultimately leading to an improved customer experience and a more positive brand image. These actions could contribute to increased customer satisfaction, loyalty, and future business growth.