**Web Scraping and Sentiment Analysis Report for British Airways Reviews**

Executive Summary:

This report presents the findings and insights derived from the web scraping and sentiment analysis of customer reviews for British Airways (BA). The reviews were collected from the Skytrax website and analyzed using various text processing techniques, including tokenization, part-of-speech tagging, lemmatization, and sentiment analysis. The sentiment analysis revealed the distribution of positive, negative, and neutral sentiments among the reviews, providing valuable insights into customer perceptions.

Data Collection and Preprocessing:

To collect the data, we used the BeautifulSoup library in Python to scrape customer reviews from the Skytrax website. A total of 1000 reviews were collected from 100 pages. The data was then preprocessed through the following steps:

Cleaning the Text: Punctuation, numbers, and unnecessary characters were removed to clean the text.

Tokenization: The text was tokenized into individual words for further analysis.

Part-of-Speech Tagging: Each token was assigned a part-of-speech tag to understand the grammatical context.

Stopwords Removal: Common stopwords were removed to eliminate noise from the analysis.

Lemmatization: Words were lemmatized to obtain their root forms for accurate sentiment analysis.

Sentiment Analysis:

Sentiment analysis was performed using the VADER (Valence Aware Dictionary and Sentiment Reasoner) sentiment analysis tool. VADER assigns a sentiment polarity score to each review, indicating whether it is positive, negative, or neutral. The sentiment analysis results are as follows:

Positive: [Insert Number of Positive Reviews] reviews (55%)

Negative: [Insert Number of Negative Reviews] reviews (43%)

Neutral: [Insert Number of Neutral Reviews] reviews (2%)

Sentiment Analysis Visual Representation:

A pie chart was created to visually represent the distribution of sentiment in the reviews:

Insights and Implications:

The sentiment analysis provides insights into customer perceptions of British Airways. The majority of reviews are Positive in sentiment, indicating Overall Customer Sentiment. The analysis highlights areas of strength and areas for improvement, which can guide strategic decisions to enhance customer experience and service quality.

Recommendations:

Based on the sentiment analysis, the following recommendations are suggested:

Leverage Positive Aspects: Identify and promote the positive aspects mentioned in positive reviews to reinforce customer satisfaction.

Address Negative Feedback: Analyze negative reviews to identify recurring issues and take prompt corrective actions.

Enhance Customer Interaction: Improve interactions with passengers, cabin crew, and ground staff to positively influence customer sentiment.

Continuous Monitoring: Implement a system for ongoing sentiment analysis to track changes in customer sentiment and respond in real-time.

Conclusion:

In conclusion, the analysis of British Airways customer reviews provides valuable insights into customer sentiments and key topics. By addressing the identified areas of improvement and capitalizing on positive aspects, BA can enhance its customer experience and make informed decisions to drive business growth.

TASK 2

Project Report Summary:

Title: Predictive Modeling of Customer Bookings

Introduction:

The goal of this project is to develop a predictive model that can accurately determine whether a customer's booking will be completed or not based on various features related to the booking process. The dataset contains information such as the number of passengers, sales channel, trip type, purchase lead time, flight details, customer preferences, and more.

Exploratory Data Analysis (EDA):

The initial steps involved data exploration using Python packages such as Pandas, NumPy, Matplotlib, and Seaborn. The dataset was loaded and inspected using the .head() and .info() methods. No missing values were found, and the data types were appropriately adjusted, such as converting the 'flight\_day' column to numerical values representing days of the week.

Descriptive Statistics:

Summary statistics were calculated for the numerical features, revealing insights such as the average number of passengers, purchase lead time, flight duration, and more. The distribution of the target variable 'booking\_complete' was examined, showing that the dataset is imbalanced with a higher number of incomplete bookings (0s) compared to complete bookings (1s).

Feature Importance:

The mutual information (MI) scores were calculated to assess the importance of each feature in relation to the target variable. The top features found to have the highest MI scores were 'route,' 'booking\_origin,' and 'flight\_duration,' suggesting their strong influence on predicting booking completion. On the other hand, 'flight\_day' and 'flight\_hour' had MI scores close to zero, indicating minimal impact.

Data Preprocessing:

Categorical variables were encoded using an ordinal encoder, converting non-numeric categories into numeric values for modeling purposes.

Conclusion:

In this project, exploratory data analysis and feature importance assessment were performed to better understand the dataset and identify relevant features for predicting booking completion. The project serves as a foundation for building predictive models using machine learning algorithms. Further steps would involve feature selection, model training and evaluation, hyperparameter tuning, and performance assessment to create an accurate predictive model for customer bookings.

Note: This project report summary provides a concise overview of the tasks, methodologies, and findings in the project. For a comprehensive report, additional sections such as Methodology, Model Building, Results, Discussion, and Conclusion can be added, along with visualizations and detailed explanations.