British Airways Virtual Internship Task 1 - Report

Project Overview

This project involves analyzing customer reviews to gain insights into sentiment and key themes. The primary goal is to understand customer satisfaction and areas for improvement.

Data Description

- **Dataset**: Customer reviews dataset with columns such as review text, rating, and date.
- **Variables**: Key variables include review text (for sentiment analysis) and rating (to correlate sentiment with customer satisfaction).

Data Cleaning

- **Processes**:
  - Removed null values and duplicates.
  - Standardized text data by converting to lowercase and removing punctuation.
  - Tokenized and lemmatized text data to prepare for analysis.
- **Libraries Used**: Pandas for data manipulation, NLTK for text preprocessing.

Exploratory Data Analysis (EDA)

- **Visualizations**:
  - Distribution plots of ratings to understand overall customer satisfaction.
  - Word clouds to visualize common words in positive and negative reviews.
  - Correlation heatmaps to explore relationships between variables.
- **Insights**:
  - Most reviews are positive, with a higher concentration of high ratings.
  - Common positive words include "service," "comfortable," and "friendly."
  - Common negative words include "delay," "poor," and "uncomfortable."

Sentiment Analysis

- **Methodology**:
  - Used VADER sentiment analysis to classify reviews into positive, neutral, and negative categories.
  - Created word clouds for both positive and negative reviews to visualize key terms.
- **Tools Used**: NLTK for sentiment analysis, WordCloud for visualization.

Key Findings

- **Positive Sentiment**:

    - Majority of reviews are positive, highlighting excellent service and comfort.

    - Frequent mentions of "friendly staff" and "good experience" indicate strong customer service.

- **Negative Sentiment**:

    - Negative reviews focus on delays, poor customer service, and uncomfortable seating.

    - Common complaints include "long wait times" and "lost luggage," suggesting areas for operational improvement.

Sentiment Distribution

Sentiment Distribution
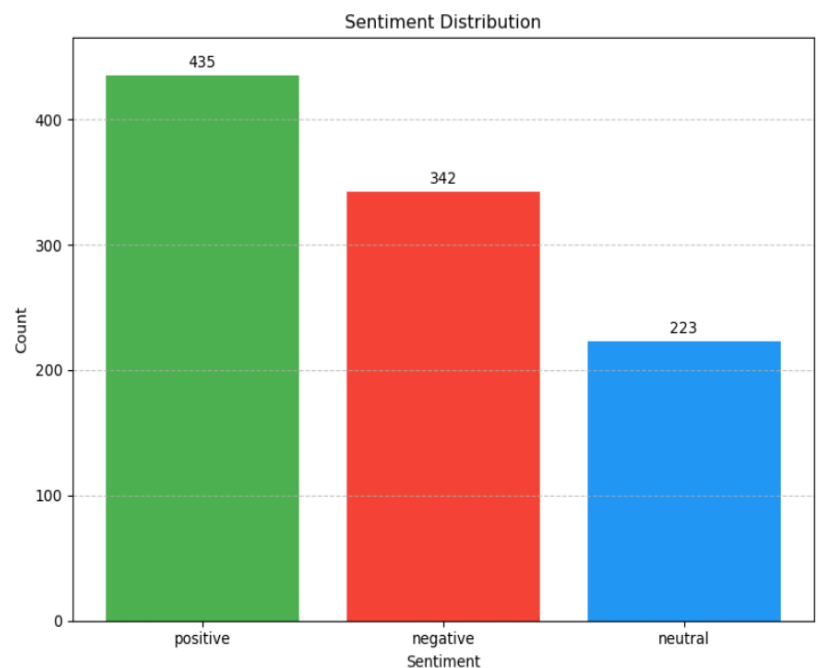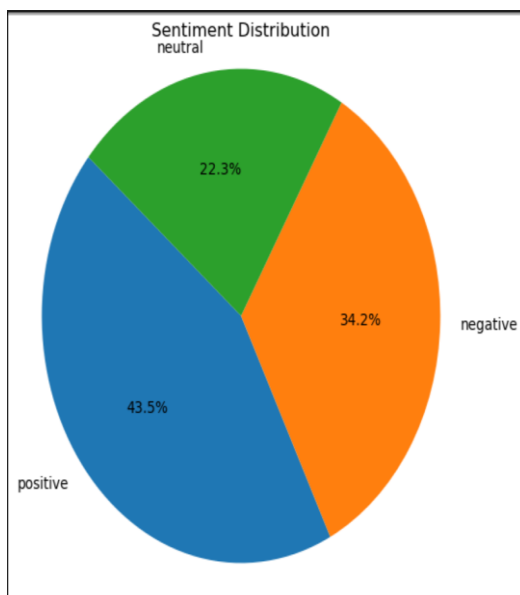
Negative Word Cloud

Overall Word Cloud

Conclusion

- **Summary**:

  - The analysis indicates overall positive customer sentiment, with service quality being a significant strength.

  - Negative feedback primarily revolves around operational inefficiencies such as delays and baggage handling.

- **Future Work**:

  - Conduct deeper analysis into specific routes or times to identify patterns in delays.

  - Implement targeted improvements in operations based on feedback to enhance customer experience.