Allen Li and Sherzod Bobojonov

British Airways Reviews Project Report

**Data Description**

This project analyzes 3700 British Airways reviews data from 2023. The dataset was downloaded from Kaggle, using the following citation:

Anshul Chaudhary, & Muskan Risinghani. (2023). Airline Reviews [Data set]. Kaggle. <https://doi.org/10.34740/KAGGLE/DS/4044107>

The dataset includes the review header and body, information about the traveler, including the type of class (business, economy, etc.), type of traveler (single, couple, family, etc.), and route traveled, the overall rating (1-10 scale), and the rating of different aspects of the traveler experience on a 1-5 scale (comfortability, service, food, etc; summary statistics on each column is in preliminary study).

**Project Questions, Implementation Tools and Techniques, and Analysis of Results**

In addition to the analysis in the preliminary study,

* What is the overall sentiment of customers towards British Airways based on their reviews?

To analyze the overall sentiment based on British Airways, our group took two approaches. First, our group counted the number of each value (“yes” and “no”) in the “Recommended” column using value\_counts. This produced these results:

A close-up of numbers

Description automatically generated

Second, our group used a large language model (LLM) and Hugging Face classifier "jdhr/AirlineTweetAnalysis-RoBERTa" to predict whether the review is positive, neutral, or negative, based on the Review Header. Our group did not use the Review Body because this produced an error, which our group assumed was because the Review Text was too large. This produced these results:

A group of black text

Description automatically generated

Comparing the two results, the number of actual not recommended reviews were about the same as predicted negative by the Hugging Face model. However, the number of predicted positive reviews was a lot lower than the actual recommended reviews. Many of the actual recommended reviews probably were predicted as neutral by the Hugging Face model.

* What are the most frequent terms related to people recommending or not recommending British Airways (extension from Preliminary Study)?

To find the most frequent terms related to people recommending or not recommending British Airways, our group split the initial dataset into two datasets, one with all customers recommending British Airways and one with all customers not recommending British Airways. From there, for each individual dataset, our group removed stop words on each review, but did not do stemming. Also, we used term frequency (Count Vectorizer) and bigrams. Since “British Airways” and “London Heathrow” are at the top of both lists, they were removed.

Based on this code, our group found these to be the top ten terms related to people recommending or not recommending British Airways.

Not Recommended Recommended

A screenshot of a table

Description automatically generatedA screenshot of a number of flight numbers

Description automatically generated

The top ten most frequent bigrams for people recommending/not recommending British Airways are very similar. There are a few small differences, however. Customer service ranks high among people who do not recommend the airline, while customer service is not even mentioned for customers who recommended the airline. Also, “via London” means that they had a connection, which was mentioned for not recommended customers. This means that people who had connections were more likely to not recommend British Airways.

* What are the terms contributing most to people recommending or not recommending British Airways?

To find the terms related to people recommending or not recommending British Airways, our group built two models, a logistic regression cross validation model and a XGBoost model.

First, our group used TFIDF, removed stop words, but did not do stemming on the training and testing (size = 0.3) set for both models. Then, our group built a logistic regression cross validation model with 5 folds and roc\_auc as the performance metric. Based on this model built, these were the accuracy and AUC, as well as the terms relating to people not recommending and recommending the airline.

Train Accuracy:

0.9370656370656371

Test Accuracy:

0.8802880288028803

Train AUC:

0.9840139418020838

Test AUC:

0.9382216459669716

Recommended Not Recommended

A screenshot of a phone

Description automatically generatedA screenshot of a number

Description automatically generated

Comparing the two lists, comfortable/uncomfortable has a large impact on recommending or not recommending the airline. Customer service-related words (ex. attentive, friendly, efficient, professional, rude) and basic positive or negative words (ex. good, excellent, great, happy, terrible, worst, awful, poor) also have a great impact on recommendation.

For the XGBoost model, our group used the same parameters as the LR CV model. Based on candidates of max depth 3, 5, and 10 and candidates for n\_estimators 200, 300, and 500, the best parameters for our XGBoost model was max depth: 5 and n\_estimators: 200. This produced the following accuracy and AUC scores the terms most related to people recommending or not recommending the airline.

Train Accuracy:

1.0

Test Accuracy:

0.873987398739874

Train AUC:

1.0

Test AUC:

0.9365429234338749

A screenshot of a number

Description automatically generated

The same conclusions can be drawn from the XGBoost model as the LR CV model in terms of the most important terms relating to people recommending or not recommending the airline.

* Are certain cabin types or flight routes more inclined to comment about specific aspects of British Airways’ service?

To analyze if certain cabin types of flight routes more inclined to speak about different aspects of British Airways’ service, our group used topic modeling to cluster similar reviews.

First, our group subset only the SeatType, ReviewBody, and Route columns from the initial dataset and split the dataset into training and testing (size = 0.33). Using the review body, our group did stemming, removed stop words, set the max features to 200, and used count frequency.

To find the number of topics for LDA, our group used the elbow method to compute the perplexity score. Based on the graph below, two topics (lowest perplexity score) best suit the number of topics.

A line graph with numbers and a line

Description automatically generated

Based on the LDA model built, these are the words most associated with each topic.

A screenshot of a phone

Description automatically generated

Based on the terms in each topic, topic 1 is more related to customer service (pre-flight) because people are more likely to mention the airline by name when complaining about their customer service (ex. “When I spoke to British Airways customer service” and “I am amazed that British Airways was so insensitive to the problem”). Topic 2 is more related to the in-flight experience and service, with terms “seat”, “crew”, “food”, “cabin” standing out.

A screenshot of a computer

Description automatically generated

These are the reviews most related to topic 0. In the review body, some reviews begin with “I booked”, which can foreshadow them talking about booking problems (i.e. customer service related). More people with economy class are related to customer service, which makes sense, since business class passengers who might face booking problems are prioritized over economy class passengers. Also, many passengers with connecting flights (indicated with the “via” in the route column) are related to customer service, since delayed flights can lead to missed connections, resulting in an interaction with customer service.

A screenshot of a website

Description automatically generated

These are the topics most related to topic 1. Many business/first class are related to topic 1, since they may not face significant issues with customer service, and they expect a high-quality on-board experience, from seat comfortability to the food and cabin service. Also, people without connections are more related to topic 1 because since they may not face significant booking/connections issues, they are more likely to focus and comment about their on-board experience.

* Frequent Bigrams Word Cloud

Finally, for fun, our group created a word cloud in the shape of an airplane depicting the most frequent bigrams (from the preliminary study) from all the reviews, using count vectorizer, without stemming. Also, we removed the top two terms, “London Heathrow” and “British Airways” since they were obviously going to be the most mentioned terms.

A plane made out of words

Description automatically generated

The most frequent bigrams include mentions about its business class, its cabin crew, and its customer service.

**Conclusion and Takeaways**

Overall, customer sentiment towards British Airways is negative. This is evident by more customers not recommending British Airways than recommending the airline. Also, using a large language model, most reviews were predicted to be negative or neutral, with very few being predicted as positive.

Based on the analysis of our questions, the two main contributing factors towards British Airways sentiment is customer service (pre-flight) and on-board experience. To improve customer sentiment, British Airways should focus on improving first-class on-board experience and connecting customer’s customer service experience, since the reviewers of these respective groups tend to comment on those areas of British Airways service.