

Airline Satisfaction Analysis

Yoyo Lu (yl3397), Xinhao He (xh245), Bonju Koo (bk368)

Course: ORIE 4741

Github Link: <https://github.com/xinhaohe245/4741-project>

March 15, 2024

1. Dataset Introduction

The dataset we have chosen to address our research questions is a set of airline reviews spanning from 2013 to 2024. This dataset was compiled by Kaggle user Sujal Suthar and includes reviews for the world's top 10 rated airlines in 2023, including Singapore Airlines, Qatar Airways, ANA (All Nippon Airways), Emirates, Japan Airlines, Turkish Airlines, Air France, Cathay Pacific Airways, EVA Air, and Korean Air according to Skytrax's World Airline Award. It offers perspectives on multiple facets of passenger contentment and service excellence, covering aspects of flight experience, including seat comfort, staff service, food and beverages, inflight entertainment, value for money, and overall rating.

2. Research Questions

Using the dataset, we will examine the following questions for insight into customer satisfaction:

- 1) Which flight aspects have the most significant impact on a passenger's experience? Do the factors that strongly correlate with positive reviews also strongly correlate with negative reviews? How have the emphasis on each flight aspect evolved over time and how would passengers value each aspect in the future?
- 2) What sentiment does each review convey to the airline? How do reviews differ between different demographics? What traits are most prominently found amongst the top airlines and distinguish them from each other? Are negative experiences more likely to result in reviews than positive ones?

3. Discussion

Both of our research questions are essential because they address the core of customer satisfaction, which is a key driver of business success in the airline industry. Understanding the factors that influence passenger experience can help airlines improve their services, improve customer retention, and ultimately increase their market share. By identifying factors that contribute the most to customer satisfaction, airlines can strategically focus their resources on the appropriate, high-impact areas and optimize the return from these investments.

We plan to use passenger satisfaction features from the dataset to address question 1. The rich features in this dataset will help us successfully answer question 1. The airline being reviewed can help us understand if there are specific airlines that consistently receive positive or

negative reviews. The type of traveling and month flown can provide context to the passenger's experience. Ratings for staff, seat, food, entertainment, and value for money can be considered as continuous features and be used to predict flight experience. We aim to build a predictive model that can accurately determine whether a reviewer would recommend an airline, thereby providing a robust tool for assessing and improving customer satisfaction.

Question 2 will be more challenging than question 1; however, by analyzing the unique text associated with each review, we believe it is indeed possible to address question 2 as well. We plan to extract sentiments and identify key themes that passengers care about. Our prediction is the sentiment behind each text, and our features are the texts in each review. This would require textual feature engineering, which we could apply multiple techniques to achieve. A preliminary technique we plan to use is the Bag-of-Words method, which uses the number of times each word appears in a document as feature vectors, then uses the Term Frequency-Inverse Document Frequency method to reduce the weight of meaningless high-frequency words. We can also apply more complex embedding models, such as fine-tuning pre-trained large language models to specifically analyze airline reviews. Using the available data and our feature engineering techniques, we aim to provide insights into passenger sentiment and understand how reviews affect the airlines' future outlook.

References

<https://www.kaggle.com/datasets/sujalsuthar/airlines-reviews>