

**A Data Analytics Approach: Investigating Factors Influencing Airline Customer Satisfaction**

Aarshi Jain

501178324

Prof. Tamer Abdou

May 20th, 2024

Every year, 8.6 billion passengers fly globally (*Global Passenger Traffic*, n.d.), making it crucial for the airline industry to prioritize customer satisfaction as a competitive advantage. This capstone project delves into the factors influencing customer satisfaction within an undisclosed airline company, utilizing the detailed dataset titled “Airline Customer Satisfaction”. The dataset is constructed of 22 columns and 129,880 rows, providing a strong foundation for predictive analytics aimed at identifying customer satisfaction factors (Huseyn, n.d.).

The primary research question guiding this investigation is: Which factors are the most pivotal in predicting customer satisfaction in the airline industry? Secondary queries include researching demographic variables such as age, travel class, travel type, and customer type to understand their correlations with both high and low satisfaction ratings. Additionally, we explore which aspects of airline service—such as seat comfort, in-flight entertainment, cleanliness, and food and beverage—significantly influence overall customer satisfaction.

To address these questions, the Airline Customer Satisfaction dataset offers comprehensive insights into customer experiences, encompassing factors like customer type, age, purpose of travel, and travel class. Furthermore, it provides specific ratings for a number of services, including seat comfort, convenience of the gate location, departure/arrival time, baggage handling, and more. This information-rich dataset allows for a multifaceted analysis of customer satisfaction factors.

The analytical techniques utilized in this investigation encompass classification and regression models, clustering, and text mining. Classification algorithms such as decision trees, and logistic regression are utilized to predict each customer's satisfaction level based on their characteristics and travel experiences (Logistic Regression, n.d.). Clustering techniques, like K-means, aid in pinpointing different groups of customers, each with their own specific

determinants that contribute to their satisfaction (Webster, n.d.). Text mining of the feedback will provide qualitative insights into customer sentiments (*Text Mining*, n.d.).

Additionally, model evaluation metrics such as precision, accuracy, and F1-score are employed “to assess the performance and effectiveness of a statistical or machine learning model” (Srivastava, 2024). Tools such as R and Python, along with specialized libraries like Scikit-Learn and the “tm” package in R, are utilized for data analysis as well as model building (Choudhary, 2022).

By conducting an investigation into this dataset, airlines can gather critical insights into the determinants that contribute to customer satisfaction, enabling them to tailor their services effectively and enhance the overall customer experience. The research will provide proactive insights for airline companies seeking to strengthen customer loyalty and improve service quality, thereby gaining a competitive edge in the market.

## References

- Choudhary, A. (2022, September 9). *15 Most Important Features of Scikit-Learn!* Analytics Vidhya. Retrieved May 16, 2024, from <https://www.analyticsvidhya.com/blog/2021/07/15-most-important-features-of-scikit-learn/>
- GLOBAL PASSENGER TRAFFIC EXPECTED TO RECOVER BY 2024 AND REACH 9.4 BILLION PASSENGERS.* (n.d.). Airports Council International. Retrieved May 14, 2024, from <https://aci.aero/2023/09/27/global-passenger-traffic-expected-to-recover-by-2024-and-reach-9-4-billion-passengers/#:~:text=Passenger%20traffic%20forecast%202023%E2%80%93932024,94.2%25%20of%20the%202019%20level>
- Huseyn, R. (n.d.). *Airline Customer Satisfaction*. Kaggle. Retrieved May 14, 2024, from <https://www.kaggle.com/datasets/raminhuseyn/airline-customer-satisfaction/data>
- Srivastava, T. (2024, January 8). *12 Essential Evaluation Metrics for Evaluating ML Models*. Analytics Vidhya. Retrieved May 15, 2024, from <https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/>
- Webster, W. (n.d.). *Cluster Analysis: Definition and Examples*. Qualtrics. Retrieved May 14, 2024, from <https://www.qualtrics.com/experience-management/research/cluster-analysis/>
- What is Logistic Regression Used for?* (n.d.). H2O.ai. Retrieved May 14, 2024, from <https://h2o.ai/wiki/logistic-regression/>
- What Is Text Mining?* (n.d.). IBM. Retrieved May 15, 2024, from <https://www.ibm.com/topics/text-mining>