Investigating Factors Affecting Airline Customer Satisfaction

CIND820: Literature Review

By: Aarshi Jain

Email: [aarshi.jain@torontomu.ca](mailto:aarshi.jain@torontomu.ca)

Student #: 501178324

Supervisor: Prof. Dr. Tamer Abdou

Submission Date: June 17th, 2024

Table of Contents

[Abstract 3](#_Toc109654601)

[Introduction 5](#_Toc109654601)

[Literature Review 7](#_Toc109654602)

[Review 1 7](#_Toc452457338)

[Review 2 8](#_Toc452457338)

[Review 3 9](#_Toc452457338)

[Review 4 9](#_Toc452457338)

[Review 5 10](#_Toc452457338)

[Review 6 11](#_Toc452457338)

[Conclusion 11](#_Toc452457338)

[Methodology 13](#_Toc109654603)

[Dataset & Dictionary 17](#_Toc109654604)

[Link to GitHub Repository & Dataset 17](#_Toc452457338)

[Exploratory Data Analysis 19](#_Toc109654605)

[References 20](#_Toc109654609)

# Abstract

Every year, 8.6 billion passengers fly globally, making it crucial for the airline industry to prioritize customer satisfaction as a competitive advantage (Global Passenger Traffic, n.d.). This capstone project delves into the factors influencing customer satisfaction within an undisclosed airline company, utilizing the detailed dataset titled “Airline Customer Satisfaction.” Comprising of 22 columns and 129,880 rows, this dataset provides a strong foundation for predictive analytics aimed at identifying factors of customer satisfaction (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 examining the influence of demographic variables such as age, travel class, travel type, and customer type on satisfaction ratings. Additionally, we investigate 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 dataset offers comprehensive insights into customer experiences, including customer type, age, purpose of travel, and travel class. It also provides specific ratings for services like seat comfort, gate location convenience, departure/arrival time, and baggage handling. 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 predict customer satisfaction levels based on characteristics and travel experiences (Logistic Regression, n.d.). Clustering techniques, like K-means, identify different groups of customers with specific determinants contributing to their satisfaction (Webster, n.d.). Text mining of feedback provides qualitative insights into customer sentiments (Text Mining, n.d.). Model evaluation metrics such as precision, accuracy, and F1-score assess the performance and effectiveness of the statistical and machine learning models (Srivastava, 2024). Tools such as R and Python, along with specialized libraries like Scikit-Learn and the “tm” package in R, are employed for data analysis and model building (Choudhary, 2022). This research aims to build upon and potentially replicate findings from previous studies using similar datasets, with the goal of providing a detailed analysis of the factors affecting customer satisfaction regarding airlines. This research follows the same techniques used in multiple studies, often combining methods for a deeper analysis, such as the Apriori Algorithm and Logistic Regression. Any differences in research methodology or tools will be highlighted, ensuring clarity on how this study adds to existing knowledge.

By investigating this dataset, airlines can gather critical insights into the determinants of customer satisfaction, enabling them to tailor services effectively and enhance the overall customer experience. This 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.

# Introduction

The airline industry was one of the largest industries to suffer the most from the global pandemic, COVID-19, which began in March 2019. Due to the numerous lockdowns, passenger numbers and revenue dwindled while high fixed costs remained, resulting in financial losses. It was found that “a number of airlines around the world had to file for bankruptcy, while others were kept on financial life support from governments, leading to major layoffs for employees” (Mordvinova, 2022). Ever since the COVID-19 travel restrictions were lifted in 2022, airline companies have been struggling with numerous challenges they were unprepared for such as staffing shortages, a sudden increase in passenger traffic, and the Russian-Ukraine war causing airlines to reroute flights and face fuel availability uncertainty due to this conflict (Singh, 2024).

Another challenge for airline companies is the increase in unruliness from customers due to chaos caused in airports. Since 2021-2022, there has been a 47% increase in customer non-compliance (Sternfield, 2023). The aggravated customers stem from the fact that, compared to 2021, the 2022 mishandling rate of 7.6 bags per thousand passengers represents a 75% increase (*Baggage Mishandling Rates*, 2023). Flight irregularities, mishandled baggage, and long waiting times form the core of passenger dissatisfaction (Mehra, 2023). About 34.4% of aircraft arrived late in 2022 and 37.0% in 2023, with nearly half of the delays being the airlines’ fault since late flights cause a chain reaction leading to more late flights (Thygerson, 2023; Major, 2023). Airfares spiked by 25% compared to 2022, causing more chaos for airlines and creating fussy customers demanding better quality (Holzhauer, 2024).

An estimated 4.7 billion people will be flying in 2024 surpassing the previous record set in 2019, indicating a simultaneous increase in travel demands (Cousins, 2023). It is crucial for airline companies to gain information on how to satisfy customers and understand the needs and expectations of customers flying. To achieve this, aviation companies must investigate passenger satisfaction to identify which aspects of their experience require improvement. The increasing demand for air travel will lead to intense competition in the industry, which is why airline companies must offer high-quality service to exceed customer expectations.

This project report utilizes the Airline Customer Satisfaction dataset, which contains ratings from 1-5 (with 5 being the most satisfied) reflecting passengers’ experiences with the airline along with some general information about the travellers such as their age, class and type of travel. The aim of the report is to identify the most influential predictors of passenger satisfaction using both supervised and unsupervised machine learning techniques. The goal is to help airline companies gain a deeper understanding of how their services impact customer satisfaction and determine which services and passenger types they should invest in to grow and retain loyal customers. Furthermore, the report assesses the efficacy of the models used in the analysis.

# Literature Review

In the aftermath of the global pandemic, aviation companies are facing challenges in accommodating a growing population of air travelers while striving to maintain high levels of customer satisfaction. Understanding how to meet these needs and exceed airline expectations is crucial for airlines to thrive in the competitive aviation industry. This study reviews previous research to explore machine learning techniques that have been utilized to identify the key factors essential for enhancing customer satisfaction and driving business development in aviation companies.

First review:

In Tri Noviantoro and Jen-Peng Huang’s study, they employed data mining techniques to investigate airline customer satisfaction. Initially, they prepared the data by filtering and eliminating entries with missing values. Their study utilized the wrapper method for feature selection, employing forward selection. Using a 10-fold cross-validation strategy to split the data into training and testing sets, the researchers identified "Online Boarding" as the highest-ranked attribute across all validation processes for feature selection. Following feature selection, they retrained the model using the top five selected features and applied eight supervised machine learning algorithms, including decision tree, random forest, gradient boosted tree, k-NN, Naïve Bayes, rule induction, logistic regression, neural net, deep learning, and support vector machine. Model evaluation was conducted using a confusion matrix, revealing that Deep Learning achieved the highest accuracy at 92.08% and the highest F-score at 93.03% in predicting airline passenger satisfaction. Additionally, k-NN demonstrated an accuracy of 91.07% and an F-score of 92.38%. Overall, their study identified the top four factors most crucial for customer satisfaction as: (1) online boarding, (2) inflight Wi-Fi service, (3) baggage handling, and (4) inflight entertainment (Noviantoro & Huang, 2022).

Second Review:

Yifei Wang conducted a comprehensive study on the factors influencing airline customer satisfaction employing structured data mining techniques. The research integrates exploratory data analysis (EDA) and statistical methods, emphasizing the impact of demographic characteristics and service-related factors. Dimensionality reduction techniques were applied to streamline the dataset and enhance model performance, with Principal Component Analysis (PCA) used to reduce feature complexity while retaining data variance. The primary objective was to aid airline companies in optimizing customer satisfaction by identifying key influencing factors. First, Wang conducted EDA to uncover patterns, anomalies, and address missing values by imputing them with the variable averages. Subsequently, a Factor Analysis was performed using Bartlett’s sphericity test, demonstrating significant correlations (P < 0.05) among variables and validating the suitability of factor analysis. Two principal factors emerged: "Fundamental Items," encompassing cleanliness, food and drink, seat comfort, and in-flight entertainment; and "Basic Service," including in-flight service, baggage handling, onboard service, and legroom. Further analysis involved one-way ANOVA to compare satisfaction levels across different demographic groups. Wang also employed Logistic Regression using the factors derived from dimensionality reduction (Wang, 2023). Overall, the study concluded that reducing delays and enhancing the travel class experience are critical steps toward improving customer satisfaction in the airline industry.

Third Review:

Sachin Kumar and Mikhail Zymbler utilized sentiment analysis of tweets to assess customer satisfaction in their study. The primary classification models employed included Support Vector Machines (SVM), Artificial Neural Networks (ANNs), and Convolutional Neural Networks (CNN). Additionally, they conducted Association Analysis using the Apriori algorithm to extract frequent item-sets and generate association rules for both positive and negative sentiment categories. This approach revealed insights into how travel-related issues impact customer sentiments during flights (Kumar & Zymbler, 2019).

Fourth review:

Xuchu Jiang, Ying Zhang, Ying Li, and Biao Zhang conducted a study focusing on predicting aircraft passenger satisfaction using an RF-RFE-LR (Random Forest-Recursive Feature Elimination-Logistic Regression) model. Initially, they employed a Random Forest algorithm with cross-validation to rank feature importance. Subsequently, Recursive Feature Elimination was utilized to select the top 17 features from an initial set of 22. Several machine learning models, including K-Nearest Neighbors, Logistic Regression, Random Forest, Gaussian Naive Bayes, and Backpropagation Neural Network (BNN), were applied to predict satisfaction using both RFE and non-RFE methods. To evaluate the classification models, the study employed accuracy, recall, precision, F-value, and AUC as evaluation metrics across ten models (five with RFE and five without RFE). The findings underscored that feature selection notably enhanced model performance, with the Random Forest-RFE classifier achieving the highest accuracy of 0.963. Additionally, logistic regression was performed to identify the top predictors for customer satisfaction in airlines, highlighting "Type of Travel," "Customer Type," "Customer Class," "In-flight Wi-Fi Service," and "Online Boarding" as the most influential variables based on their coefficient values (Jiang et al., 2022).

Fifth Review:

Chris Bacani conducted research on airline customer satisfaction utilizing Scikit-Learn's K-Best and Chi-Square functions for feature selection. The study began with data cleaning and preparation, including handling null values in Arrival delays, text normalization, aggregating flight delay data, and scaling features. Feature selection techniques K-Best and Chi-Square were applied to identify and extract the 15 most significant features. Subsequently, Bacani trained multiple classifiers including Logistic Regression, Adaptive Boosting, Random Forest, Categorical Naive Bayes, and Extreme Gradient Boosting using the selected features. Model performance was evaluated based on training and test accuracy, precision, recall, and ROC values. The study also included an analysis of model metrics, confusion matrices depicting classification performance, and ROC AUC curves. Among the classifiers tested, the Extreme Gradient Boosting classifier demonstrated the highest accuracy of 95% and 94.6%. Bacani further utilized Python's pickle library to serialize the model, enabling analysis of feature impacts using SHAP (Bacani, 2022).

Sixth Review:

Sangah (Ivy) Kim's study investigates airline customer satisfaction using exploratory data analysis (EDA) and multiple machine learning algorithms. The research begins with thorough data preparation, including imputing missing values with medians for normalization, removing redundant columns, and creating dummy variables for categorical data. Initially, logistic regression with all features identifies the most important variables. Kim then explores interactions between features through additional logistic regression to understand their relationships. To address collinearity, lasso feature selection refines the feature set, identifying significant predictors such as in-flight Wi-Fi and boarding procedures. Classification trees are employed to extract interpretable rules, highlighting factors like online boarding scores and travel type (business vs. personal) that strongly influence satisfaction. A Random Forest model validates the importance of identified features by outperforming the decision tree, leveraging the random selection of predictors during model training. Kim further validates model performance through cross-validation, revealing consistent accuracy across most models, with the exception of the decision tree, which performed less optimally (Kim, 2023).

Conclusion

In reviewing the research reports, it becomes evident that Random Forest consistently demonstrated superior performance across different datasets, with the exception found in Chris Bacani’s research where Extreme Gradient Boosting excelled. Among the studies, "In-Flight Wi-Fi Service" emerged as a consistently influential factor for predicting airline customer satisfaction.

Furthermore, a common practice among researchers was the use of feature selection techniques to reduce dimensionality before applying machine learning models. They then utilized feature importance to identify the most influential predictors after evaluating the models. This approach mirrors methodologies employed by Xuchu Jiang, Ying Zhang, Ying Li, Biao Zhang, Chris Bacani, Tri Noviantoro, and Jen-Peng Huang.

This study intends to integrate these methodologies, drawing particularly from the approaches of Xuchu Jiang et al., Chris Bacani, and Sachin Kumar and Mikhail Zymbler. It will also incorporate techniques such as Association rules and the Apriori algorithm, as explored by Kumar and Zymbler, to uncover relationships and patterns within categorical data.

# Methodology

The following steps outline a tentative, structured methodology for analyzing airline customer satisfaction:

1. Data Collection and Preprocessing

* **Data Cleaning**: Handle missing values.
* Identify and handle all the missing values by replacing them with the mean/median values or removing the row/column. This step was done in all the studies, such as Noviantoro and Huang’s study, where data preparation involved replacing missing values with the medians of the feature and removing columns that are not necessary for the research. This ensures the data is clean for accurate model training which impacts the reliability of the predictions (Noviantoro & Huang, 2022).
* **Data Transformation:** Convert categorical variables into numerical values.
* As used in Tri Noviantoro and Jen-Peng Huang's research, this study will use one-hot encoding for categorical variables such as 'Customer Type', 'Type of Travel', and 'Class' for model compatibility (Noviantoro & Huang, 2022).
* **Normalization/Standardization:**
* Normalizing data makes sure that all features are scaled to be comparable (Bacani, 2022). The standardized data improves model training efficiency and accuracy. This will be done using the Python package “ydata profiling” and “SweetViz”.

2. Exploratory Data Analysis (EDA)

* **Descriptive Statistics**: Summarize the data to gain insights into its basic features and data distribution.
* Sangah Kim used EDA to develop an understanding of the dataset (Kim, 2023). This helps to learn about variances, outliers, the spread of data, and central tendencies which is vital for informed model selection and feature engineering. This will be done through the use of visual tools such as histograms, box plots, and scatter plots to identify patterns and outliers with the help of Python packages like “ydata profiling” and “SweetViz”.
* **Correlation Analysis**: Using a correlation matrix to identify relationships between features.
* By generating a correlation matrix, the study will identify relationships between features. Correlation analysis was performed by Chris Bacani to find key determinants impacting customer satisfaction (Bacani, 2022). The correlation matrix will help select features likely to be significant predictors of customer satisfaction.

3. Feature Selection

* **Recursive Feature Elimination (RFE)**: As used by Xuchu Jiang et al, this study will implement RFE with a Random Forest model to select the most important features by iteratively removing the least significant features (Jiang et al., 2022). RFE helps by reducing dimensionality.
* **K-Best or Chi-Square:** To identify significant features, apple feature selection techniques.
* Initial feature selection was found using K-Best and Chi-Square techniques (Bacani, 2022). This method reduces dimensionality and emphasizes the most important determinants which enhance the model’s performance and interpretability.
* **Lasso Regression:** Refine the feature set by using Lasso Regression.
* To remove collinearity, Sangah Kim used Lasso Regression to increase the robustness of the model (Kim, 2023).

4. Model Building and Evaluation

* **Train-Test Split**: Split the dataset into training and testing sets (20/80).
* **Train Models**: Train machine learning models, including:
  + **Decision Trees**: To understand feature interactions as well as interpretability (Noviantoro & Huang, 2022).
  + **Random Forest**: For its performance and ability to handle large datasets with high-dimensional features. Random Forest was consistently found to be one of the most effective learning models as stated by numerous researchers such as Xuchu Jiang (Jiang et al., 2022).
  + **Logistic Regression**: To provide interpretable coefficients and identify key predictors (Noviantoro & Huang, 2022).
* These models are key to predicting customer satisfaction levels based on a variety of features, directly addressing the primary research question of which factors are the most important in predicting customer service satisfaction.
* **Association Analysis (Apriori Algorithm):** To discover relationships between categorical variables. Sachin Kumar and Mikhail Zymbler’s study was the only one which used this algorithm to find patterns between the variables (Kumar & Zymbler, 2019). This algorithm will answer the secondary questions pertaining to demographic variables corresponding to high and low customer rating values.
* **Model Evaluation**: Evaluate models using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC to assess performance and effectiveness. This ensures that the models are accurate and reliable which will validate the results of the predictions.

5. Cross-Validation

* **Cross-Validation:** Used by numerous researchers such as Noviantoro and Huang, and Xuchu Jiang, cross-validation provides confidence that the model will perform well on unseen data, preventing overfitting and making sure all predictions and satisfactory (Noviantoro & Huang, 2022; Jiang et al., 2022).

6. Reporting and Documentation

* **Report Findings**: Document the methodology, analysis, and results in a comprehensive report.
* **Recommendations**: Provide actionable recommendations based on the analysis to help improve airline customer satisfaction.

# Dataset and Dictionary

This study uses the open-source dataset “Airline Customer Satisfaction” available on the Kaggle website. It contains 22 features and 129,880 rows of data.

**Links**

GitHub Link:

<https://github.com/justatoj/CIND820-.git>

Dataset Link:

<https://www.kaggle.com/datasets/raminhuseyn/airline-customer-satisfaction/data>

Data dictionary:

|  |  |
| --- | --- |
| Feature | Description |
| Satisfaction | Indicates the satisfaction level of the customer. |
| Customer Type | Type of customer ('Loyal’ or 'Disloyal’). |
| Age | Age of the customer. |
| Travel Type | Purpose of the travel ('Business' or 'Personal’). |
| Class | Class of travel ('Business', 'Eco', or 'Eco Plus’). |
| Flight Distance | The distance of the flight (in kilometres). |
| Seat comfort | Rating of seat comfort provided during the flight (1 to 5). |
| Departure/Arrival time convenient | Rating of the convenience of departure/arrival time (1 to 5). |
| Food and Drink | Rating of food and drink quality provided during the flight (1 to 5). |
| Gate location | Rating of gate location convenience (1 to 5). |
| Inflight wifi service | Rating of inflight wifi service satisfaction (1 to 5). |
| Inflight entertainment | Rating of inflight entertainment satisfaction (1 to 5). |
| Online support | Rating of online customer support satisfaction (1 to 5). |
| Ease of Online booking | Rating of ease of online booking satisfaction (1 to 5). |
| On-board service | Rating of on-board service satisfaction (1 to 5). |
| Legroom service | Rating of leg room service satisfaction (1 to 5). |
| Baggage handling | Rating of baggage handling satisfaction (1 to 5). |
| Check-in service | Rating of check-in service satisfaction (1 to 5). |
| Cleanliness | Rating of cleanliness satisfaction (1 to 5). |
| Online boarding | Rating of online boarding satisfaction (1 to 5). |
| Departure Delay in Minutes | Total departure delay in minutes. |
| Arrival Delay in Minutes | Total arrival delay in minutes. |

# Exploratory Data Analysis (EDA)

In the EDA conducted as documented in the repository, comprehensive data cleaning methods were employed, including the reduction of null values, construction of correlation matrices, comparison of categorical and numerical values, and generation of histograms for all numerical features, among other techniques. These steps were crucial in gaining insights into data distributions, identifying patterns, and preparing the dataset for further analysis. For a detailed overview of the EDA procedures followed, please refer to the repository documentation.

# References

**References**

*Air transport industry turns to digitalization as baggage mishandling rates virtually double amid growing passenger volumes*. (2023, May 16). SITA. Retrieved June 6, 2024, from https://www.sita.aero/pressroom/news-releases/air-transport-industry-turns-to-digitalization-as-baggage-mishandling-rates-virtually-double-amid-growing-passenger-volumes/

Bacani, C. (2022, November 21). *Explaining airline passenger satisfaction using interpretable machine learning*. Medium. Retrieved June 7, 2024, from https://medium.com/@chris.bacani7/explaining-airline-passenger-satisfaction-using-interpretable-machine-learning-88d29aa55677

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/

Cousins, B. (2023, December 28). *Will high rates bite into Canadians' travel appetite in 2024?* BNN Bloomberg. Retrieved June 6, 2024, from https://www.bnnbloomberg.ca/economy-won-t-ground-demand-for-air-travel-in-2024-experts-say-1.2016039

*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%932024,94.2%25%20of%20the%202019%20level

Holzhauer, B. (2024, May 2). *Airline Ticket Prices Are Outpacing Inflation. Here's How You Can Save*. CNBC. Retrieved June 6, 2024, from https://www.cnbc.com/select/airline-ticket-prices-are-up-25-percent-why-and-how-to-save/

Huseyn, R. (n.d.). *Airline Customer Satisfaction*. Kaggle. Retrieved May 14, 2024, from https://www.kaggle.com/datasets/raminhuseyn/airline-customer-satisfaction/data

Jiang, X., Zhang, Y., Li, Y., & Zhang, B. (2022, July 1). *Forecast and analysis of aircraft passenger satisfaction based on RF-RFE-LR model*. Nature. Retrieved June 7, 2024, from https://www.nature.com/articles/s41598-022-14566-3

Kim, S. (. (2023, February 17). *Airlines Customer Satisfaction Analysis: EDA & Machine Learning Algorithms*. Medium. Retrieved June 7, 2024, from https://kimsangaha.medium.com/airlines-customer-satisfaction-analysis-eda-machine-learning-algorithms-dd3b883a3a38

Kumar, S., & Zymbler, M. (2019, July 17). *A machine learning approach to analyze customer satisfaction from airline tweets - Journal of Big Data*. SpringerOpen. Retrieved June 7, 2024, from https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0224-1

Major, D. (2023, October 9). *Almost half of all flight delays in 2022 were airlines' responsibility, government data suggests*. CBC. Retrieved June 6, 2024, from https://www.cbc.ca/news/politics/2022-flight-delays-airline-responsibility-compensation-1.6987189

Mehra, S. (2023, December 14). *A new deep dive into passenger frustration with airlines*. TNMT. Retrieved June 5, 2024, from https://tnmt.com/passenger-frustration-with-airlines/

Mordvinova, V. (2022, June 9). *Canadian aviation amid the COVID-19 pandemic: Part 2. Impact on various flying activities*. Statistics Canada. Retrieved June 6, 2024, from https://www150.statcan.gc.ca/n1/pub/45-28-0001/2022001/article/00004-eng.htm

Noviantoro, T., & Huang, J.-P. (2022, June). *Investigating airline passenger satisfaction: Data mining method*. Science Direct. Retrieved June 7, 2024, from https://www.sciencedirect.com/science/article/pii/S2210539521001097

Singh, P. (2024, May 8). *8 Key Challenges and Strategic Imperatives for the Aviation Sector*. Global Market Insights. Retrieved June 4, 2024, from https://www.gminsights.com/blogs/top-challenges-of-aviation-industry

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/

Sternfield, M. (2023, June 14). *Bad behavior among airline passengers soars*. The Hill. Retrieved June 4, 2024, from https://thehill.com/changing-america/well-being/mental-health/4047750-bad-behavior-among-airline-passengers-soars/

Thygerson, J. (2023, December 13). *US Airline Delay Statistics - Analyzed by TravelFreak*. TravelFreak. Retrieved June 5, 2024, from https://travelfreak.com/airline-delay-statistics/

Wang, Y. (2023, July). *Factors Affect Airline Customers' Satisfaction: Data Mining*. ResearchGate. Retrieved June 7, 2024, from https://www.researchgate.net/publication/372608755\_Factors\_Affect\_Airline\_Customers'\_Satisfaction\_Data\_Mining

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/