AIT- 664: Represent, Process & Visualize Applied Information
Prof. Ebrima Ceesay

PROJECT PROPOSAL

ONBOARD INSIGHTS: DATA ANALYSIS OF AIRLINE PASSANGER SATISFACTION

Group 2

Christian Appiah
Nikita Chawla
Suyog Suresh Patagave

Introduction

Air travel holds a crucial role in modern life, connecting people and businesses worldwide. As the airline industry continues to grow and change, airlines must prioritize understanding passenger satisfaction to stay competitive and provide top-notch travel experiences. In this era of data-driven decision-making, leveraging data analysis is not just advantageous but imperative.

This project, titled "Data Analysis on Airline Passenger Satisfaction," aims to explore the intricate landscape of airline passenger satisfaction through data analysis. We will collect and scrutinize data from various sources, including customer feedback, flight records, and social media sentiment analysis, to gain profound insights into the factors driving passenger satisfaction. These insights will enable airlines to refine services, allocate resources effectively, and tailor strategies to consistently enhance passenger satisfaction. In the following chapters, we will outline our research objectives, methodologies, and expected outcomes. Through data analysis, our goal is to not only understand the complex factors influencing passenger satisfaction but also contribute to an industry crucial for global connectivity. By illuminating the path to increased passenger satisfaction, we strive for a future where air travel symbolizes excellence and fosters global connections.

Objective of the Project

The proposed study on airline passenger satisfaction holds significant importance for various stakeholders in the airline industry, including airline operators, customers, airline regulators, and researchers.

For airline operators, understanding passenger satisfaction is crucial for several reasons. It can lead to higher customer loyalty and retention rates, which are vital for the long-term success of airlines. Furthermore, passenger satisfaction data can provide a competitive advantage by attracting more customers and enhancing market shares [7]. Additionally, satisfied passengers are more likely to spend on ancillary services, such as in-flight meals and premium seating, thereby increasing ancillary revenue and overall profitability. Finally, high passenger satisfaction contributes to a positive brand image and reputation, leading to positive word-of-mouth and social media recommendations.

Customers, on the other hand, stand to benefit significantly from this study. Understanding passenger satisfaction can lead to a more enjoyable and comfortable travel experience. Satisfied passengers are likely to perceive better value for their money, as airlines strive to improve services and amenities. Moreover, it assures customers of the airline's commitment to safety, efficiency, and reliability. Additionally, passenger satisfaction surveys provide a platform for timely issue resolution, improving customer relations and loyalty [4]. Passengers can also use satisfaction ratings and reviews to make informed choices when booking flights, leading to more enjoyable journeys [6].

Regulators, including aviation authorities and government agencies, have a vested interest in this study for various reasons. It supports consumer protection, safety, and security. [5] Quality control, regulatory compliance, and policy development benefit from insights gained through passenger satisfaction research. Moreover, it aids in consumer advocacy, public trust, and confidence in the airline industry, promoting data-driven decision-making.

Finally, researchers in the fields of aviation and customer service find value in this study for multiple reasons, including gaining industry insights, analyzing consumer behavior, benchmarking, evaluating policies, exploring innovation and technology adoption, developing service improvement strategies, and assessing the impact of external factors.

Literature Reviews

An article published by J.D. Power, an analytics firm, cites the differences in the satisfaction levels of customers at various airlines. They have ranked various airline companies according to their class, such as business, premium economy, and economy. This shows the preference of customers when they wish to travel in their desired class. The satisfaction levels have also been affected due to various reasons, such as an increase in the ticket price, dissatisfaction with the meals, and fees and other costs. [1]

An IEEE study that relates to our project cites the use of data analytics techniques such as data visualization, exploratory data analysis, and more. The study draws conclusions based on the available dataset variables and uses various machine learning models to compare the best-fitting model. [2]

Hasan and Farooqi (2020) investigated how service quality and brand image influence the satisfaction of airline passengers in India. They found that both service quality and a positive brand image have a direct and positive impact on passenger satisfaction. This highlights the importance of these factors for airlines in India aiming to enhance passenger contentment.[5]

Pabla and Soch (2023) investigate the link between airline passengers' brand experience and brand satisfaction, with brand love as a mediator. They emphasize that a positive brand experience is crucial in fostering brand love, which in turn enhances overall brand satisfaction. The study delves into the multifaceted nature of brand experience, emphasizing various touchpoints, and highlights the importance of emotional attachment in passenger satisfaction. It offers valuable insights for airlines to strategically manage brand experience dimensions to create strong emotional connections and boost passenger satisfaction. [6]

Hannigan, Hamilton, and Mudambi (2015) examined competition and competitiveness in the US airline industry. They emphasized the impact of market deregulation on competition, the role of alliances and partnerships, and the significance of technology and innovation. External factors like economic conditions and regulatory changes were also highlighted as influential in shaping competitiveness within the industry. This study provides valuable insights into the complex dynamics of competition in the US airline sector. [7]

Introduction to Dataset: The dataset provided by Maven Analytics comprises 129,880 records, each representing a passenger who has utilized the services of an airline. It includes a variety of columns capturing key information about these passengers and their flight experiences. These columns encompass demographic data such as Gender and Age, information about the passenger's Customer Type, Type of Travel, and Class of service. Additionally, the dataset incorporates data related to the flight experience, including Flight Distance, Departure Delay, Arrival Delay, and various aspects of onboard services, cleanliness, and satisfaction.

The dataset serves as a valuable resource for conducting comprehensive analyses aimed at understanding passenger satisfaction and the factors that contribute to it. By exploring the

relationships within this dataset, we can gain insights that may inform improvements in airline services and enhance overall customer satisfaction. [3]

Research Questions

The importance of this endeavor lies in the intricate nature of the airline industry, where even the smallest details can significantly impact the customer experience. As the airline business is inherently customer-centric, every piece of feedback from passengers holds immense value in facilitating service improvements and maintaining competitiveness.

To achieve these goals, we have outlined a series of key questions:

1. What are the most influential factors affecting customer satisfaction that can be ranked from highest to lowest?

2. How does customer satisfaction vary when comparing different types of travel and classes?

3. For each customer type, what are the primary reasons for the three levels of satisfaction?

4. When designating "satisfaction" as the output variable, which features and variables in the dataset are most valuable for predicting customer satisfaction? Can we improve predictions through feature engineering and machine learning algorithms, and how do different models perform in terms of metrics?

5. Can we create an overall satisfaction score by summing individual service ratings, and is there a correlation between this score and flight distance, departure date, or arrival delay? Is it possible to predict the overall score based on flight distance?

6. By grouping passengers into age categories, what insights can be gained about different traveler types, their travel behavior, preferred classes, and who tends to take longer trips?

7. Which of the following services—ease of booking, check-in service, or online boarding—is rated the least desirable, and how can they be ranked in order of preference?

8. Among returning passengers, how many travel for personal reasons versus business, and in which class? In comparison to first-time travelers, who is more likely to achieve higher customer satisfaction?

Proposed Approach

1. Data Preprocessing (R/Python): Data Collection: We will download the dataset from Maven Analytics and ensure its completeness and up-to-datedness. Data Cleaning: We will identify and address missing data in columns by employing methods like imputation (mean, median, or mode) or deletion, depending on the significance of the data.We will detect and address outliers that may distort our analysis or modeling results. Feature Engineering: We will generate additional features that have the potential to contribute to

our analysis, such as creating an overall satisfaction score or transforming categorical features. Feature Scaling: To ensure that numerical features have similar scales and prevent the dominance of larger values, we will normalize them. Data Transformation: We will transform categorical variables into numerical formats using techniques like one-hot encoding. The dataset will be split into training, validation, and test sets for the development of machine learning models.

2. Exploratory Data Analysis (R/Python): Data Visualization: We will create various visualizations, including histograms, scatter plots, box plots, etc., to gain insights into the distribution and relationships among variables. Heatmaps and correlation matrices will be generated to identify correlations and potential multicollinearity. Statistical Analysis: To identify significant differences among groups or variables, we will perform statistical tests such as t-tests or ANOVA. If applicable, we will conduct time series analysis to explore trends and seasonality in passenger satisfaction data.

3. Predictive Modeling (R/Python): Data Preparation: We will identify the most relevant features for predicting customer satisfaction using techniques like feature importance. Numerical features will be scaled for modeling, for example, using Min-Max scaling. Model Selection: We will choose appropriate machine learning algorithms for classification tasks, considering options such as Logistic Regression, Decision Trees, Random Forest, Gradient Boosting, and Support Vector Machines. Different models and hyperparameters will be experimented with to find the best-performing model. Model Evaluation: The dataset will be split into training and validation sets to evaluate model performance. Metrics like accuracy, precision, recall, F1-score, and ROC-AUC will be employed to assess model effectiveness. Hyperparameter Tuning: We will optimize model performance by fine-tuning hyperparameters using techniques like grid search or random search. Overfitting or underfitting issues will be addressed by adjusting model complexity.

4. Actionable Insights (R/Python/Tableau): Interpretation of Model Results: Model coefficients (for linear models) or feature importance scores (for tree-based models) will be analyzed to understand the factors driving passenger satisfaction. Interactive Visualizations (Tableau): We will create Tableau dashboards featuring interactive charts and visualizations to present key insights. Dynamic filters and parameters in Tableau will be built, allowing stakeholders to explore data on their own.

R/Python/Tableau [8] [9] [10]

Project Timeline
The project is anticipated to be finished in 8 to 10 weeks, with the following timeline:

Task	Completion Date
Project Proposal	Week 0: Sep 20, 2023
Data Collection and Pre-processing	Week 1-2
Exploratory Data Analysis	Week 3-4
Predictive Modelling	Week 5-6
Analysis and Report Generation	Week 7-8
Final Project Presentation	Week 9-10

Expected Result

Our research project strives to offer valuable insights to the airline industry by identifying and ranking the critical factors that influence customer satisfaction. This ranking will serve as a strategic guide for airlines, enabling them to focus their efforts on enhancing passenger satisfaction more effectively. Furthermore, we expect to uncover variations in satisfaction across different types of travel and service classes, equipping airlines to tailor their services to specific customer segments. Understanding customer types and determining the primary drivers of satisfaction will empower airlines to customize their services for various customer groups. Our predictive models are poised to accurately forecast customer satisfaction, providing airlines with a dependable tool for proactive improvements. Additionally, we aim to create a comprehensive overall satisfaction score and provide demographic insights into traveler types. Moreover, we will rank services based on passenger preferences, assisting airlines in prioritizing areas for enhancement. Lastly, we will identify patterns in travel purposes and class preferences between returning and first-time travelers, allowing airlines to retain loyal customers and enhance personalized experiences.

Conclusion

In conclusion, our research is expected to provide airlines with actionable insights to enhance passenger satisfaction and improve their services. By identifying influential factors, understanding customer segments, and predicting satisfaction levels, airlines can optimize their resources, allocate budgets judiciously, and tailor their strategies effectively. The creation of an overall satisfaction score and the ranking of service preferences will enable airlines to make data-driven decisions to enhance the passenger experience. Furthermore, the insights gained from demographic analysis and the differentiation between returning and first-time travelers will allow airlines to target their marketing efforts more precisely, thereby strengthening customer relationships and loyalty. By actively following our detailed approach, this research project aims to contribute to the airline industry's ongoing efforts to provide exceptional service and maintain competitiveness in the global market.

References

1. Effler, G. (2022, May 11). 2022 North America Airline Satisfaction Study. J.D. Power. https://www.jdpower.com/business/press-releases/2022-north-america-airline-satisfaction-study https://www.jdpower.com/business/press-releases/2022-north-america-airline-satisfaction-study

2. Hibović, L., Smajić, S., & Yaman, E. (2022, October 1). Predicting Satisfaction of Airline Passengers Using Classification. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9932850

3. Maven Analytics. (n.d.). Maven Analytics. https://mavenanalytics.io/data-playground?page=2&pageSize=5

4. Singh, P. P. (n.d.). Ancillary Rev strategies - Airlines. www.linkedin.com. https://www.linkedin.com/pulse/ancillary-rev-strategies-airlines-pramod-pal-singh/

5. Hasan, M., & Farooqi, R. (2020, January 1). Impact of Service Quality and Brand Image on Satisfaction of Airline Passengers in India: A SEM Approach. Amu-in. https://www.academia.edu/46847784/Impact_of_Service_Quality_and_Brand_Image_on_Satisfaction_of_Airline_Passengers_in_India_A_SEM_Approach

6. Pabla, H., & Soch, H. (2023, March 1). Up in the air! Airline passenger's brand experience and its impact on brand satisfaction mediated by brand love. Journal of Air Transport Management; Elsevier BV. https://doi.org/10.1016/j.jairtraman.2022.102345

7. 7.Hannigan, T. J., Hamilton, R. D., & Mudambi, R. (2015, March 16). Competition and competitiveness in the US airline industry. Competitiveness Review; Emerald Publishing Limited. https://doi.org/10.1108/cr-11-2014-0036

8. Python. (2019, May 29). Python. Python.org; Python.org. https://www.python.org/

9. The R Foundation. (2019). What is R? R-Project.org. https://www.r-project.org/about.html

10. Tableau. (2023). Tableau: Business Intelligence and Analytics Software. Tableau Software. https://www.tableau.com/