Data Analysis on Airline Passenger Satisfaction

Christian Appiah
Department Of Information Sciences And Technology
George Mason University
Fairfax, VA.
Email: cappiah@gmu.edu

Nikita Chawla
Department Of Information Sciences And Technology
George Mason University
Fairfax, VA.
Email: nchawla5@gmu.edu

Suyog Suresh Patagave
Department Of Computer Science and Engineering
George Mason University
Fairfax, VA.
Email: spatagav@gmu.edu

***Abstract*—The primary focus of this paper is on addressing the imperative need to enhance passenger satisfaction in the airline industry. This is crucial for the industry's financial performance, customer retention, and competitive positioning. In a highly competitive market with numerous choices for consumers, airlines must attract new business and retain existing customers to succeed. The achievement of these goals is heavily dependent on maintaining a high level of passenger satisfaction. Low satisfaction can lead to detrimental consequences, including a decline in the airline's reputation, diminished customer loyalty, and negative word-of-mouth publicity. Unhappy travelers may explore alternative airlines, sharing their negative experiences on social media and review platforms, further damaging the airline's standing.**

Keywords—Airline industry, passenger satisfaction, tidyverse, ggplot2, boxplots.

I. Introduction

A crucial component of the aviation sector, airline passenger happiness has a significant impact on the financial performance, customer loyalty, and general market standing of airlines. In this milestone, we explore the fundamental issue and the reasons behind the necessity of a thorough data analysis of passenger happiness on airlines. There is intense competition in the aviation industry as several airlines compete for passengers in a highly commoditized market. While pricing, destinations, and travel times are important considerations, the Caliber of the passenger experience is a key differentiator. In order to gain repeat business from satisfied passengers and to draw in new ones through positive word-of-mouth and reviews, airline companies work hard to meet and surpass passenger expectations.

Thus, airlines must comprehend passenger satisfaction above anything else. They can use it to identify areas that need improvement, streamline services, and improve the flying experience as a whole. Airlines can obtain valuable information for their marketing campaigns, operational enhancements, and strategic decision-making by performing a data study on passenger satisfaction. Satisfied passengers are more likely to become devoted customers, which increases revenue and fosters longterm growth. This directly affects profitability.

By demonstrating the vital significance of researching airline passenger pleasure, we established the foundation for our data analysis journey in this milestone. We will go more deeply into the information, approaches, and insights that will help us comprehend how to enhance the flying experience and, consequently, support airlines' performance in this fast-paced and cutthroat sector.

A. *Problem Statement*

The problem that we aim to address is the need to improve passenger satisfaction levels in the airline industry, which has become a pressing concern due to its direct impact on airlines' financial performance, customer retention, and competitive standing. In a market where consumers have many options and competition is fierce, airlines must find a way to both draw in new business and hold onto their current clientele. Reaching these two goals is contingent upon having a high level of passenger satisfaction.

Low passenger satisfaction can have several negative effects, such as a drop in an airline's reputation, a reduction in customer loyalty, and unfavorable word-of-mouth publicity. Disgruntled travelers are more inclined to investigate other airlines for their upcoming trips, and they might also post their negative experiences on social media and review sites, which would further harm the airline's reputation.

Airlines stand to gain a great deal by taking steps to improve passenger satisfaction and solving this issue. These consist of heightened patronage, favorable referrals, greater ticket sales, and a competitive advantage in the industry. This problem statement essentially acts as the cornerstone of our data analysis, directing our efforts to pinpoint the primary factors that influence passenger happiness and offer practical recommendations to the airline sector to improve the quality of the flying experience.

B. *Key Challenges*

Data Availability: Airlines collect vast amounts of data related to passenger satisfaction, including surveys, feedback, and operational data. However, this data is often dispersed and unstructured, making it challenging to gain meaningful insights.

Multi-faceted Nature of Satisfaction: Passenger satisfaction is influenced by various factors, such as in-flight services, punctuality, baggage handling, and customer service. Analyzing these factors and their interactions is complex.

Competitive Pressure: As more airlines enter the market, the competition intensifies. Passengers have more choices than ever before, which means airlines must strive to provide an exceptional experience to retain their customer base. Economic Implications: High passenger satisfaction correlates with increased repeat business and positive word-of-mouth. On the other hand, dissatisfied passengers might choose alternative airlines for their future travels, affecting airlines' profitability.

Regulatory Requirements: Regulatory bodies often require airlines to meet certain passenger satisfaction standards, and failing to do so can lead to legal issues and fines.

C. Motivation

The motivation behind conducting data analysis on airline passenger satisfaction is multi-faceted and extends beyond addressing the problems mentioned above.

- Improving Customer Experience: By identifying the areas of passenger dissatisfaction, airlines can tailor their services to better meet customer expectations. This leads to a more enjoyable and stress-free travel experience for passengers.
- Revenue Enhancement: Satisfied passengers are more likely to book with the same airline for future travel, leading to increased customer loyalty and revenue growth.
- Operational Efficiency: Data analysis can help airlines optimize their operational processes, such as baggage handling, scheduling, and customer service, leading to cost savings and increased efficiency.
- Reputation Management: A positive reputation is an asset for any airline. Analyzing passenger satisfaction data can help airlines identify and address issues that might harm their brand image.
- Compliance with Regulations: Airlines need to meet regulatory requirements related to passenger satisfaction. Data analysis ensures that they remain in compliance and avoid legal issues.

I. Literature Review

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 several reasons, such as an increase in the ticket price, and dissatisfaction with the meals, fees, and other costs. [1]

Some of the key findings of this article depict that overall passenger satisfaction across first/business, premium economy, and economy/basic economy is 798 (on a 1,000-point scale), down more than 20 points from a year ago when the study was conducted. Owing to the pandemic restrictions on serving liquor in the premium economy and business class, a sharp decline in food and beverage satisfaction was recorded. The cost of air travel which has consistently increased since March 2022 has led to customer dissatisfaction. [1]

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. This research can also have practical implications for the decision-making of airlines pertaining to inflight services and by extension their brand image. This leads to the conclusion that the inflight staff must always be well-dressed, well-behaved, and courteous to all passengers. By drawing a direct comparison between how the inflight staff is perceived by passengers and the general sentiment of passengers, the authors of this paper aim to show a correlation between the tangible and intangible aspects of an airline company. [5]

The study is in conformity with the previous studies that the brand image of the airlines has a direct and significant impact on passenger satisfaction and loyalty. An airline company that pays attention to its brand image will produce higher passenger satisfaction and consequently instill higher customer loyalty. The hypothesized theoretical model was a good fit, but the data was generated from a sample of passengers from the New Delhi domestic airport, which is a major limitation of the study. Further, the study focused only on the services provided on-board the airlines and did not consider the airport services. [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]

II. Proposed Approach

In accordance with our objective to understand the key features and predict the satisfaction of the airline service, we have planned to proceed by understanding the two issues, satisfaction, and (neutral or) dissatisfaction.

As of the current stage, we have implemented the use of the R programming language for data analysis. In further stages, we will be implementing various other tools Tableau, for supportive data analysis and visualization.

We followed the data analysis pipeline procedures to explore our dataset.

Data gathering: For this project we were specifically looking for dataset that would have a considerable number of records (rows) and variables/features (columns). This is necessary as it would help us analyze the data extensively and perform machine learning procedures such as creating training and testing datasets as we progress. We were interested in dataset that will have business implications so that we can provide suggestions to improve the business and its growth.

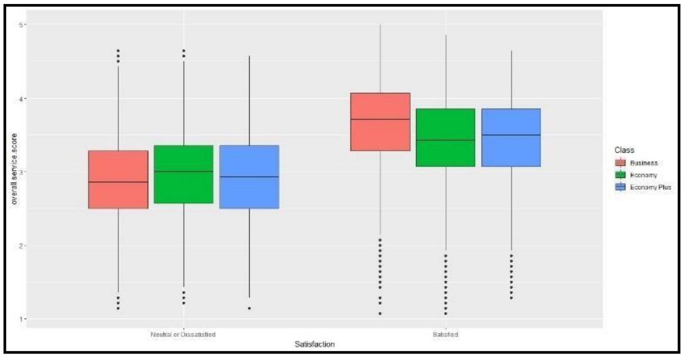


Fig1a: Range of customer satisfaction

Libraries: It is crucial to use libraries (containing various packages) to perform data analysis. Here, we used “tidyverse,” “dplyr,” and “ggplot2.” These were used for data cleaning, data processing, data exploration, and data visualization.

Data Cleaning and Testing: This is one of the essential steps required for data analysis. The data we referred to did not require any extreme data cleaning. We tested the class of every feature and the dataset using the command class (“<column_name>”). This returned values such as “numeric,” “character,” and “dataframe.” We also tested for any null values present in the dataset using the command is.na. The values returned were “False” indicating absence of any null values.

Data analysis: Our approach was to understand the satisfaction levels given by every passer and classify the output based on the class, type of passenger, and travel type. In addition, we created a variable (new column) that shows mean of all the services provided by the airlines called "overall.service.score," which consisted of over 10 distinct types of services, and a mean called "avg.delay," which is the mean of departure and arrival delays for each passenger. We created these new columns to get an overview of the dataset. We considered the possibility for satisfaction or dissatisfaction may be related to the services provided by the airlines or the delays in the flight.

Using the dplyr package, we used the 'mutate' function in creating the two new variables. We created a new sub-dataset that contained only certain variables (columns) from the original dataset. In doing so we were able to focus on a narrower objective which was to understand the satisfaction levels of the passengers considering their type, class, and the travel reason.

Data Visualization: We created boxplots, assigning satisfaction status on the x-axis and the overall.service.scores on the y-axis. Moreover, we classified the boxplots based on the passenger’s class such as economy, economy-plus, and business. Further, we also created tile plot graph. For this we assigned, type of customer (returning or first-time) on x-axis, travel reason on y-axis (business or personal), and faceted (segmented) the graph by the travel class. We filled the tiles with a colour range that depicted average delay the passengers in every combination faced.

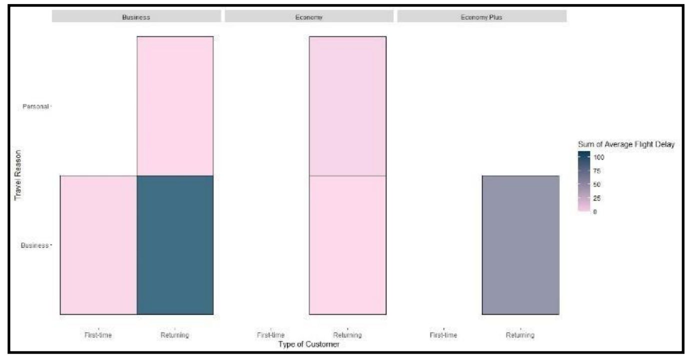
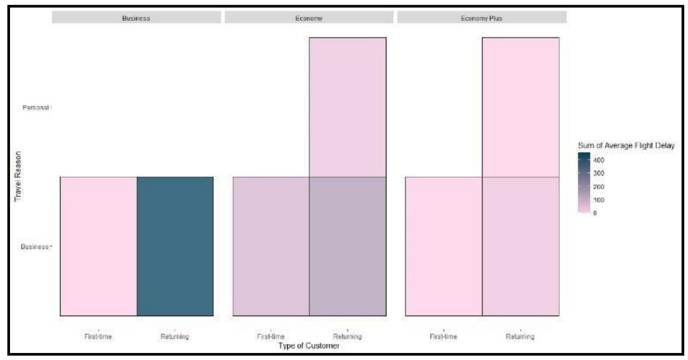


Fig1b: Avergae flight delay



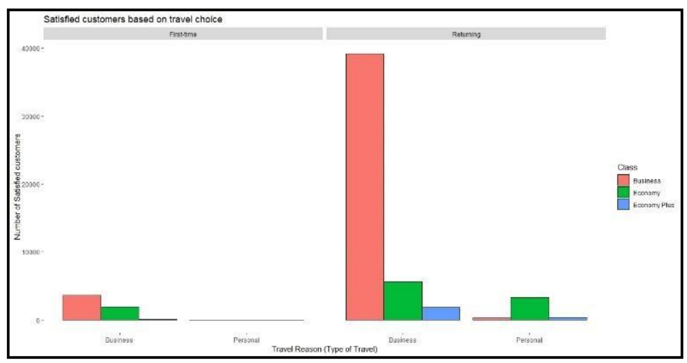


Fig1c: Satisfaction exploration

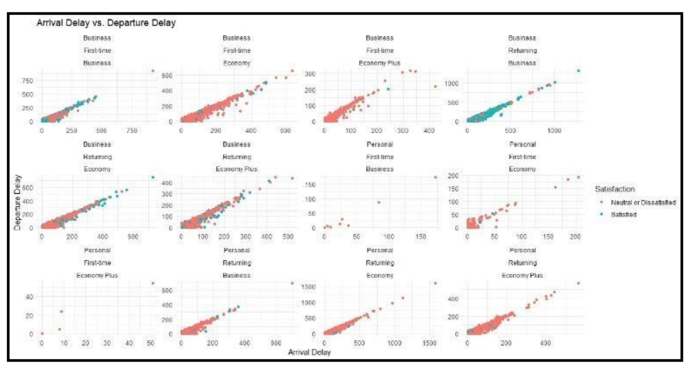


Fig1d: Scatter plot of departure delay vs arrival delay

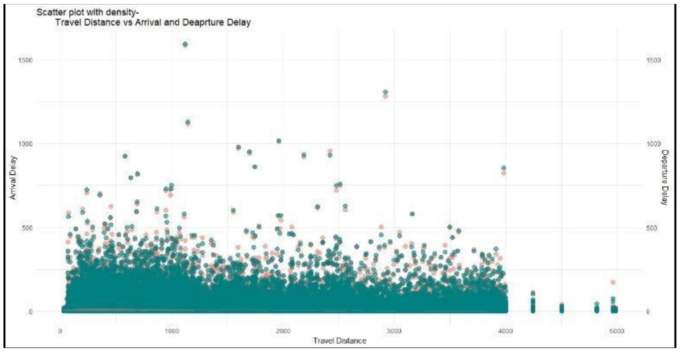


Fig1e: Satisfied and Neutral/Dissatisfied Customers comparing to arrival and departure delay

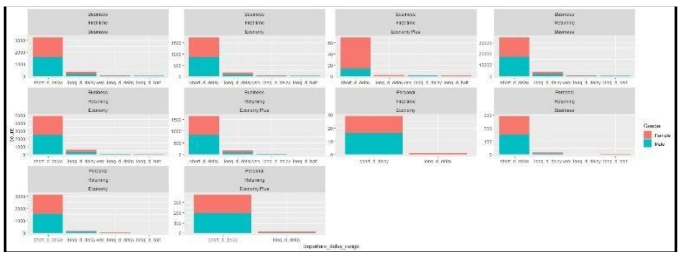


Fig2a: Satisfaction based on range of departure delay

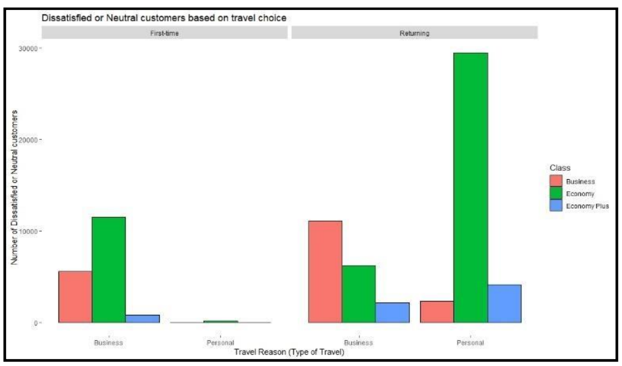


Fig2b: Satisfaction classification based on class and type of travel.

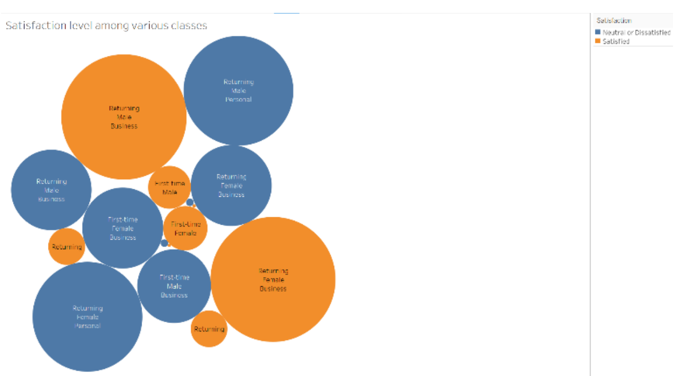


Fig2c: Tableau visualization showing the volume of satisfaction within customers

III. RESULTS AND ANALYSIS

BRIEF OVERVIEW:

We extended our preliminary data exploration and analysis where we used the libraries and packages such as 'dplyr', 'tidyverse', and 'ggplot2'. The requirement of such practice is crucial because it allows an analyst to examine and comfortably understand the important aspects of the dataset. This includes number of features, possible predictors, the most likely output variable, count of columns and rows, and the overall general sense of the dataset.

After gathering the dataset, we planned a series of activities and test related to the dataset that would allow us to recognize the potential features with considerable amount of importance and their function in creating a model to predict the potential outcome based on a series of input variables within the dataset. The data majorly consisted of categorical variables having distinctive values for various features or variables in the dataset. These values were unique and a level of merit in these values served a purpose in the final analysis.

In the preliminary analysis, we have seen the use of data manipulation and data visualization. Although the results generated from these briefly described how the variables were linked to each other, it's also necessary to understand which of these are truly important and have impact on the final output variable.

In accordance with our research quest of understanding which are the most important variables, specifically the type of services we attempted data modeling for classification prediction.

The reason to chose such model mode (classification) is that the outcome variable which is the 'Satisfaction' column of the dataset consist of two unique values 'Satisfied' and 'Neutral or Dissatisfied'. The dataset contained such values in its original form. Hence, the outcome wasn't varied enough, such as range of satisfaction levels.

We planned to use two models to find a reasonable solution to our questions. Being classification problem, we decided to implement and juxtapose two models, namely: Logistic Regression and Decision Tree. The assumption under consideration accounted the ability of logistic regression divide the dataset and the predicted results in two components. Although this may seem to be an ideal solution with respect to the output variable, the use of decision tree model has its benefits. Decision trees have the ability to predict the outcome based on the trained dataset which has same variables and with the presence of values of each variable the decision tree is able to find patterns to predict the outcome on the testing dataset.

Logistic regression:

It is a supervised machine learning algorithm mainly used for binary classification tasks. The goal is to predict the probability that an instance belongs to a given class[1]. It uses a logistic function to map the output of the linear regression function as input and estimates the probability for the given class. The logistic function maps any real value into another value within a range of 0 and 1. The output is a binary or dichotomous outcome limited to two possible outcomes: yes/no, 0/1, or true/false.[12][13]

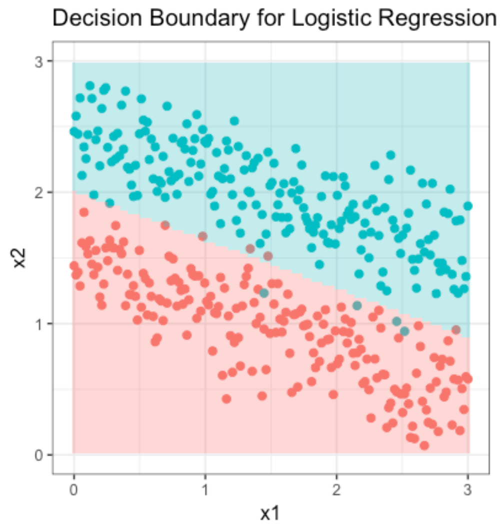


Figure 3.a Logistic regression

Decision Tree: It is a supervised learning algorithm used for both classification and regression tasks. The decision tree starts at the root and branches off to demonstrate various outcomes. Each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label. It is constructed by recursively splitting the training data into subsets based on the values of the attributes until a stopping criterion is met. The decision rules are generally in the form of if-then-else statements.[14][15]

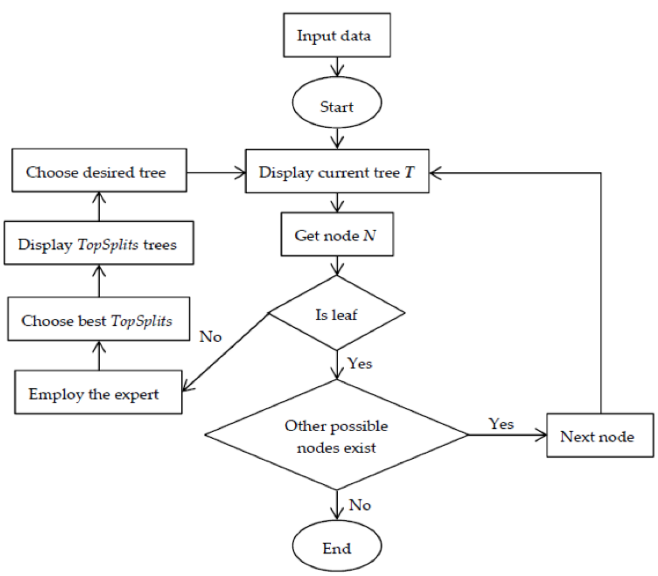


Figure 3.b Decision tree mechanism

The process of initiated with the importing the dataset and followed up with the standard data preprocessing techniques such as calculating ‘NA’ or ‘NULL’ values in the dataset, formatting the variables to retain a constant unit of measurement, counting the number of rows and columns, and summarizing the dataset.

Additionally, we converted the existing character values to factor so that it’s easier and efficient for the model to work with factor and each of the factors in the dataset for certain variables provided unique values.

Once the dataset was cleaned and formatted as required, we downloaded the necessary libraries and applied them for the process of feature engineering, data splitting, and prediction.

The most notable library used in this process is the ‘tidymodels’ library. This library particularly provides a convenient way to set up a regression model and the processing needed for each step.

After initiating our trial model, we learned that certain variables in the dataset possessed multi-collinearity. This factor is undesirable for logistic regression and hence we initially removed one of the variables which were highly correlated.

Following is a correlation plot depicting the collinearity amongst all variables.

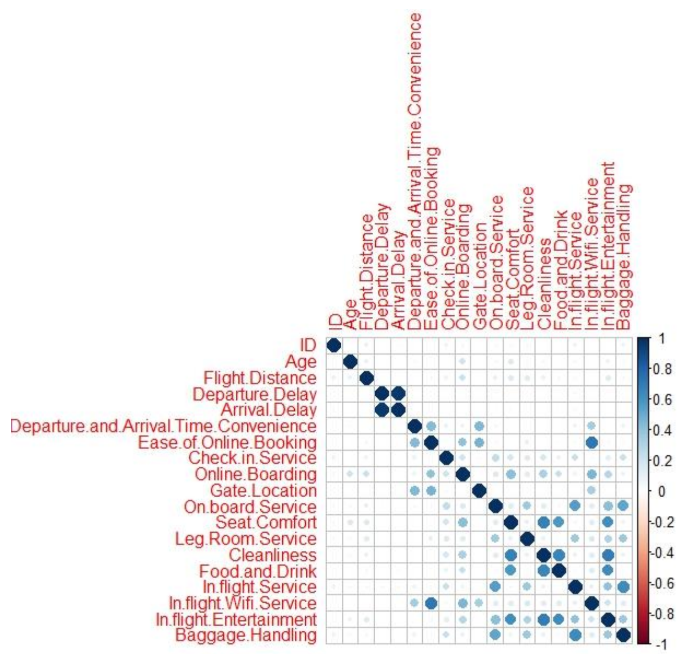


Figure 3.c Correlation plot of the dataset

It is evident from the picture above that the dots and the shade of colours represent the level of collinearity among each variable. We can observe that there’s a high correlation with departure and arrival delay variables in the dataset.

The warning message received in the trail execution allowed us to remove the column ‘Deaprture.Delay’ in hope of mitigating the multicollinearity.

We initiated with the use of ‘recipe’ package in ‘tidymodels’ library. The data was split into training and testing for the model to learn on the trained split. The proportion of the split

was set at 0.75 which signifies the split of 75% for training the model and 25% for testing the model.
Additional preprocessing steps such as use of ‘step_YeoJohnson’ to normalize the continuous numerical data in the dataset such as ‘Arrival.Delay’ and ‘Flight.Distance’. The normalization makes the data uniform, which makes it near ideal for the model to operate. As we have mentioned earlier that the logistic regression unlike decision tree model divides the data into two parts. Hence, it is necessary to convert certain categorical variables such as ‘Class’, ‘Consumer.Type’, and ‘Type.of.Travel’ to one hot coding. This increases the efficiency and eases operation of logistic models.
Then we create a workflow where we set the engine of logistic regression to ‘glm’ and mode to ‘classification’.
The workflow consists of ‘prep’ and ‘bake’ function to set the training model. Once it’s done we train the model and extract the important variables which have impact on the outcome variable. These variables are extracted using ‘parsnip’ and ‘vip’ packages and library, respectively.
Following is the result output of the logistic model’s important feature that affect the satisfaction of the airline customers.

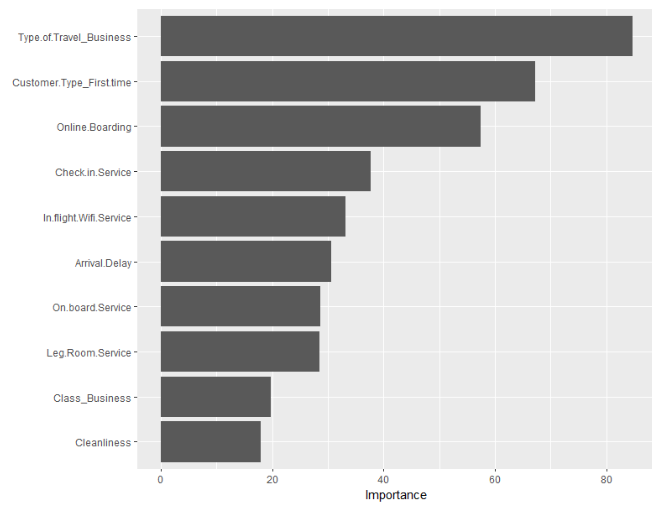


Figure 3.d Important variables in the dataset

The two axis in this picture are Importance on the x-axis quantifying the significance of ten listed features. According to this mode, type of travel is of utmost importance, signifying its impact on the satisfaction of customers, specifically if business purposes following with the customer type. The rest of the features are the combination of services such as online boarding, check-in, and more. This is called feature engineering which help us in generating insight of knowing the important features in the dataset.
Now essentially we have the answer to our curious questions of understanding the airlines satisfaction outcome and its factors.
But moving on we need to know how exactly this model is accurate and knowing if it’s trust worthy. We proceed with fitting the model to split dataset which makes the model to make predictions based on its training information. We received a series of warning messages that alerted stating the predictions may be misleading and less accurate.
We concluded the logistic regression modeling and classification by inspecting a few metric and their results.
Following is the screenshot of metrics that are helpful to judge the model’s performance.

```
# A tibble: 1 × 3
  .metric  .estimator .estimate
  <chr>    <chr>          <dbl>
1 accuracy binary         0.878
>
> #sensitivity
> sens(airlinesf_pred_result,
+      truth= Satisfaction,
+      estimate = .pred_class)
# A tibble: 1 × 3
  .metric .estimator .estimate
  <chr>   <chr>          <dbl>
1 sens    binary         0.905
> #specificity
> specificity(airlinesf_pred_result,
+             truth= Satisfaction,
+             estimate= .pred_class)
# A tibble: 1 × 3
  .metric     .estimator .estimate
  <chr>       <chr>          <dbl>
1 specificity binary         0.842
>
> auc_result <- roc_auc(data = airlines
ed)
> auc_result
# A tibble: 1 × 3
  .metric .estimator .estimate
  <chr>   <chr>          <dbl>
1 roc_auc binary         0.929
>
```

Figure 3.e Performance Metrics of logistic regression

There are four distinctive metrics:
Accuracy: It is the fraction of predictions our model got right, i.e., the number of correct predictions divided by the total number of predictions.
Sensitivity: Also known as the true positive rate, it is the probability of a positive test result, conditioned on the individual truly being positive.
Specificity: Known as the true negative rate, it is the probability of a negative test result, conditioned on the individual truly being negative.
ROC_AUC Score: It measures the entire two-dimensional area underneath the entire ROC curve, providing an aggregate measure of performance across all possible classification thresholds.

Confusion Matrix: A table that summarizes the performance of a classification model on a set of test data for which the true values are known. It displays the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) produced by the model.

```
                          Truth
Prediction                 Neutral_or_Dissatisfied Satisfied
  Neutral_or_Dissatisfied                    16572      2224
  Satisfied                                   1735     11842
>
```

Figure 3.f Confusion Matrix

Thee values essentially show the performance of the logistic regression model.

Although the scores may range within 80-93, it's still not convincing enough to finalize the model. We may see the overall performance of the model using the ROC_AUC plot which shows the comparison of the true positive values correctly predicted.

Considering this fact, the following plot shows the corresponding values of sensitivity to 1-specificity.

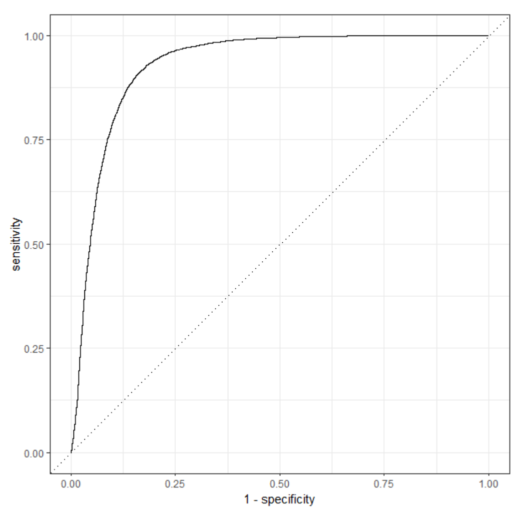


Figure 3.g ROC curve of logistic regression model

Ideally, the model that performs well has the curve extended to the top left corner, depicting higher performance of the model and even better accuracy and accurate prediction. This in-turn justifies that the model is able to predict the values but finding reasonable pattern.

We conducted a similar approach but with a different mode, decision tree. The reason to choose this model is due to its ability to split the data values and come up with a pattern to predict the values of unseen labeled dataset.

The decision tree model has better and enhanced ability to operate on multiple categorical values in the dataset. Unlike logistic regression, the decision tree is able to segregate the values by calculating and applying a threshold.

The initial process to set up the model is similar to the one in the logistic regression. However, the engine here changes to 'rpart'. We also perform cross validation; this makes the model to train on more variation of dataset adding the advantage of adapting to a generalized dataset. We also add a functionality of hyperparameter tuning. This helps in choosing the best values for the decision tree for a better accuracy. We set 'grid_regular' function with the following hyper parameters: cost_complexity, tree depth, and minimum nodes. These essentially create a grid with the best parameters that can be implemented to the model to generate output at higher accuracy.

Conducting the same test of 'vip' for feature selection to choose the best variables we found the following results.

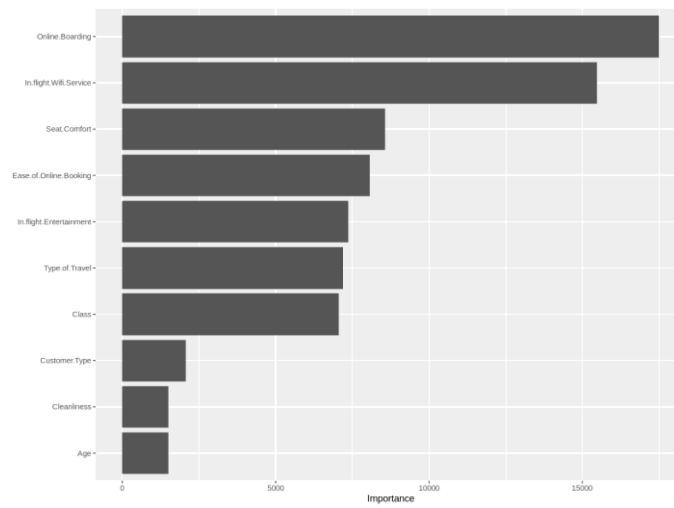


Figure 3.h Important variables in the dataset (decision tree)

According to this model, online boarding, in flight Wi-Fi services, seat comfort and other such service are of highest importance. This list also includes certain other factors such as type of travel, class, and age.

As we progressed through the code execution, we printed the predicted data with the values presenting the chances each record had of being either in the satisfied or neutral or dissatisfied with the satisfaction variable or column of the dataset.

```
tree_predictions<- tree_last_fit%>% collect_predictions()
class(tree_predictions)
head(tree_predictions)
conf_matrix<-conf_mat(tree_predictions, truth= Satisfaction, estimate= .pred_class)
conf_matrix
```

'tbl_df' · 'tbl' · 'data.frame'

A tibble: 6 × 7

id	.pred_Neutral_or_Dissatisfied	.pred_Satisfied	.row	.pred_class	Satisfaction	.config
<chr>	<dbl>	<dbl>	<int>	<fct>	<fct>	<chr>
train/test split	0.945121951	0.05487805	1	Neutral_or_Dissatisfied	Neutral_or_Dissatisfied	Preprocessor1_Model1
train/test split	0.001162115	0.99883788	3	Satisfied	Satisfied	Preprocessor1_Model1
train/test split	1.000000000	0.00000000	16	Neutral_or_Dissatisfied	Neutral_or_Dissatisfied	Preprocessor1_Model1
train/test split	1.000000000	0.00000000	20	Neutral_or_Dissatisfied	Neutral_or_Dissatisfied	Preprocessor1_Model1
train/test split	1.000000000	0.00000000	21	Neutral_or_Dissatisfied	Neutral_or_Dissatisfied	Preprocessor1_Model1
train/test split	0.475609756	0.52439024	23	Satisfied	Neutral_or_Dissatisfied	Preprocessor1_Model1

Fig 3.i Tree predictions showcasing the predicted results and estimates

Similarly, we will assess this model on the similar metrics as that of logistic regression.
Following are the scores of decision tree model.

```
metrics <- tree_last_fit %>% collect_metrics()

# Extracting specific metrics
accuracy <- metrics %>% filter(.metric == "accuracy") %>% pull(.estimate)
sensitivity<- sens(tree_predictions, truth = Satisfaction, estimate = .pred_class)
specificity<-spec(tree_predictions, truth = Satisfaction, estimate = .pred_class)
roc_auc <- metrics %>% filter(.metric == "roc_auc") %>% pull(.estimate)

cat("Accuracy:", accuracy, "\n")
cat("ROC AUC:", roc_auc, "\n")

Accuracy: 0.9532944
ROC AUC: 0.9907322
```

```
sens(tree_predictions, truth = Satisfaction, estimate = .pred_class)
spec(tree_predictions, truth = Satisfaction, estimate = .pred_class)
```

A tibble: 1 × 3

.metric	.estimator	.estimate
<chr>	<chr>	<dbl>
sens	binary	0.9665155

A tibble: 1 × 3

.metric	.estimator	.estimate
<chr>	<chr>	<dbl>
spec	binary	0.936087

Figure 3.j Performance metrics of decision tree model

Contarctisng to the results obtained from logistic regression, all the metrics in this model seem to have higher score. This give it an edge of being superior in classification as compared to logistic regression.
All the values lie above 93% with 99.07% being the best score for the overall model.
Moreover, we can also plot the roc_auc curve that shows the overall performance of the model.

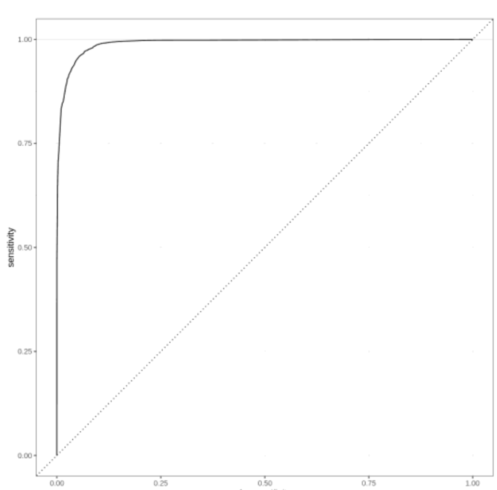


Figure 3.k ROC curve of decision tree model

It is evident that the model has performed well in regards to the task of classification prediction, surpassing the metrcis of logistic regression.

IV. ANALYTICAL CONCLUSION

In brief, we may say that when the task is based on the categorical variables including the output variable, it is ideal to use the classfication mode. With regrads to choosing a model, decision tree has the ability and essential properties that makes it efficient to work on categorically versatile dataset. Feature engineering solved the most of our questions related to research. This solution when applied with the business and domain knowledge and expertise can surely help the airline companies to excel in the service and hospitality. Along with considerable amount of business revenue growth and profitability. From this test of modeling and classsication we may conclude that it's necessary to understand the model and the application for which it is applicable. Judging the model based on metrics may reveal the bets model to choose.

V. LIMITATIONS AND FUTURE SCOPE

The project was indeed a successful attempt in exploring dataset, gathering insights, finding patterns, data visualizzation, data modeling and manipulation, and creating a supervised machine learning workflow for classification prediction.
However, it would be fair to say it wasn't perfect. The lack of additional hyperparameters was observed due to the limitations in the computational power and resources. The absence of such resources have impacted th eefficiency of execution of the code. Thus, occurance of time constraints. Additionally, the code wasn't uniformly adapted to any specific software application. The code for logistic regression was executed on R-Studio sue to the ability of the computational system to undertake such light-weighted task. However, we switched to Google Colab to create an R environment for the execution of decision tree model. This required an efficient system environment that was limited to R-Studio. Google Colab provided us with an efficient resources that were able to execute the code with the heavy-weighted tasks such as hyperparameter tuning and model fitting.
We plan to explore additional datasets with varied and diverse locations and demographics. This would help us in understanding the service demands and performances of passengers across the world. We also aim to undertake social media datasets related to air travel and airline service. This is a crucial aspects such that the information flow in today's world is extremely lucid and this may have serious impacts on the airline business which may result in a series events affecting the business.

VI. 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
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. 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. Stavrinoudis, D., & Orfanou, F. (2019). Airline passenger satisfaction: a structural equation model. Journal of Air Transport Management, 74, 142-148.
9. Han, Y., Karpova, E. G., & Logar, J. (2019). Airline customer satisfaction, trust, and loyalty: An examination of the moderating role of the type of a passenger. Journal of Air Transport Management, 81, 101720.
10. Chang, Y. (2018). Analysis of Factors Affecting Customer Satisfaction in Airline Service Quality. Journal of Air Transport Management, 66, 36-42.
11. Soteriades, M., & Stavrinoudis, D. (2020). Passengers' Expectations, Perceptions, and Satisfaction in Full-Service Airlines: The Case of Athens International Airport. Journal of Air Transport Management, 84, 101735
12. IBM. (2022). What is Logistic regression? Www.ibm.com. https://www.ibm.com/topics/logistic-regression
13. Understanding Logistic Regression. (2017, May 9). GeeksforGeeks. https://www.geeksforgeeks.org/understanding-logistic-regression/
14. GeeksForGeeks. (2017, October 16). Decision Tree - GeeksforGeeks. GeeksforGeeks. https://www.geeksforgeeks.org/decision-tree/
15. Decision Tree Algorithm. (2021, July 9). TowardsMachineLearning. https://towardsmachinelearning.org/decision-tree-algorithm/