**Mapping Airline Passenger Preferences to Maximize Satisfaction: A Statistical Analysis**

1. INTRODUCTION

In the dynamic aviation industry, where passenger satisfaction holds paramount importance, “airline passenger satisfaction metric” is an ultimate tool for airline carriers to meticulously examine strategies that can elevate the satisfaction levels of their customers.

The objective of our research is to construct a predictive model that furnishes airlines with actionable insights, enabling them to proactively address issues, enhance service quality, and tailor their offerings to cater to the diverse needs of passengers.

By offering a nuanced comprehension of passenger satisfaction, this study not only enriches academic knowledge but also equips airline companies with implementable findings. These findings empower them to elevate the overall travel experience and gain a competitive advantage in an ever-evolving market landscape.

PROBLEM STATEMENTS-

"What is the marginal effect of flight delays, online service, distance, and in-flight amenities on determining customer satisfaction levels in the airline industry?"

"How can airlines leverage this knowledge to substantially improve customer satisfaction?"

1. LITERATURE REVIEW

Various studies in recent years have employed passenger surveys and statistical analysis techniques to assess satisfaction levels and service quality perceptions of airline travelers (Feng & Jeng, 2005; Mikulić & Prebežac, 2011; Kei, 2019). With increasing competition between full-service carriers (FSC) and low-cost carriers (LCC), monitoring attributes that influence customer satisfaction has become imperative for strategic decision making in the air transport sector.

Importance-performance analysis (IPA) is a widely used methodology adopted across the literature to evaluate airline service quality (Choon, 2017). This two-dimensional approach quantifies passenger-rated importance of service attributes along with corresponding satisfaction levels. By mapping the means on a grid divided into four quadrants, IPA identifies strengths and improvement areas for airlines. Kei (2019) used an IPA model to compare Hong Kong’s FSC and LCC passengers. The study found significant gaps in FSC’s on-time performance and seating comfort, while LCC passengers were unsatisfied with amenities and entertainment options. In contrast, Choon's (2017) IPA analysis of a Singapore-based budget airline revealed gaps in check-in, entertainment and convenience dimensions. Thus geographic market variations exist.

In Taiwan, Feng & Jeng (2005) utilized IPA to determine passengers accorded highest importance ratings to safety, punctuality and complaint response efficacy. This highlights how cultural orientations shape priorities and satisfaction criteria. Mikulić & Prebežac (2011) investigated drivers of passenger loyalty by surveying over 1500 travelers across Europe. Using structural equation modeling, their study concluded service quality and perceived value as key antecedents of satisfaction and loyalty - though differing between LCC and FSC segments.

While IPA provides actionable insights for practitioners, scholars have argued it relies solely on respondent ratings without considering determinants of perceived satisfaction. Thanasupsin et al. (2010) thus proposed IPA-Weighted ServPerf approach which integrates importance-performance matrix with multivariate regression weighting of satisfaction attributes. Applying this hybrid model for Thai aviation, the authors found customized in-flight services and responsiveness having maximum effect on satisfaction.

In summary, passenger surveys coupled with IPA emerge as predominant technique for diagnostic assessment of service attributes affecting satisfaction and loyalty. The literature highlights significant gaps between passenger expectations and actual airline performance - necessitating better alignment. As competitiveness intensifies, more advanced analytics approaches could provide nuanced insights to boost customer centricity efforts amongst both full-service and budget airlines globally.

1. METHODOLOGY

The dataset for the US Airline passenger satisfaction survey was sourced from Kaggle through a meticulous data collection process. The dataset contains total of 130,000 observations defined across 22 features that impacts customer satisfaction levels. The primary information was gathered in 2015 from a diverse sample of passengers who had recently participated in domestic flights across the United States. This dataset incorporates a broad range of variables, including demographic details such as age and gender, as well as factors influencing passenger satisfaction, like the purpose of travel (Personal or Business), travel class (Business, Economy, or Economy Plus), and customer loyalty status (Loyal or Disloyal). Flight-specific attributes, such as flight distance, and departure and arrival delays were meticulously recorded. Additionally, passengers provided feedback on satisfaction levels across various aspects of the air travel experience, encompassing in-flight services, online booking, entertainment, food and drink quality, seat comfort, and overall service quality. The dataset aims to offer a comprehensive and insightful view into the factors influencing passenger satisfaction and to aid in the evaluation and enhancement of airline services in the United States.

After finalizing the research question, we started by thoroughly examining the dataset and removing any missing or invalid data points represented by NA (Not Available) or NAN (Not a Number) values. To enhance clarity, we renamed the variables to more descriptive names. Since the outcome variable (customer satisfaction) and most predictor variables were categorical, we grouped them into two broad categories: "good" and "bad." Predictors with ratings below 3 were classified as "bad," indicating customer dissatisfaction, while those with ratings above 3 were classified as "good," signifying customer satisfaction.

Initially, we employed a simple linear regression or probability model to assess the impact and significance of each variable on the outcome variable. After analyzing the results, we eliminated insignificant variables and ran another regression analysis. Although this did not drastically improve the model's performance, it ensured that all remaining variables were statistically significant.

To further delve into the problem, we employed logistic regression models, which are better suited for categorical outcome and predictor variables. In this phase, we iteratively omitted insignificant variables and incorporated interaction terms with each regression model until all remaining variables were statistically significant. This meticulous approach provided valuable insights into the marginal effects of various predictors on the log-odds ratios, enabling us to pinpoint the key factors that substantially influenced customer satisfaction levels. By systematically refining the models and incorporating relevant interactions, we ensured that our final logistic regression analysis captured the most significant determinants of customer satisfaction in a robust and reliable manner.

DATA CLEANING AND PREPROCESSING

The original dataset was obtained from Kaggle that contains close to 130,000 survey entries/data points from a US airline, with a total of 22 feature columns and a binary target column.

14 out of all the features rate the flight experience on a scale of 0 to 5 while others are numerical features.

We followed the below steps to clean our data in Rstudio:

* Renaming the variables,
* Cleaning NA an NAN,
* Dealing with missing data,
* Clustering of variables.

SUMMARY STATISTICS-

Summary statistics of shortlisted predictors

1. **Seat comfort**

**Min. 1st Qu.  Median    Mean 3rd Qu.    Max.**

**0.000   2.000   3.000   2.839   4.000   5.000**

**[1] 2.838597**

**[1] 3**

**[1] 1.392983**

**[1] 6**

**[1] 0 1 4 5 2 3**

1. **In-flight service**

**Min. 1st Qu.  Median    Mean 3rd Qu.    Max.**

**0.000   2.000   4.000   3.383   4.000   5.000**

**[1] 3.383477**

**[1] 4**

**[1] 1.346059**

**[1] 6**

**[1] 4 2 0 3 5 1**

1. **Departure Time Delay**

**Min. 1st Qu.  Median    Mean 3rd Qu.    Max.**

**0.00    0.00    0.00   14.71   12.00 1592.00**

**[1] 14.71371**

**[1] 0**

**[1] 38.07113**

**[1] 466**

1. **Arrival Time Delay**

**Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's**

**0.00    0.00    0.00   15.09   13.00 1584.00     393**

**[1] 15.09113**

**[1] 0**

**[1] 38.46565**

**[1] 473**

1. **Customer Type-**

**Length     Class      Mode**

**129880 character character**

**[1] 2**

**[1] "Loyal Customer"    "disloyal Customer"**

1. **Class Type-**

**Length     Class      Mode**

**129880 character character**

**[1] 3**

**[1] "Eco"      "Business" "Eco Plus"**

1. **Food & drink quality**

**Min. 1st Qu.  Median    Mean 3rd Qu.    Max.**

**0.000   2.000   3.000   2.852   4.000   5.000**

**[1] 2.851994**

**[1] 3**

**[1] 1.443729**

**[1] 6**

**[1] 0 1 2 3 4 5**

1. **Leg Room**

**Min. 1st Qu.  Median    Mean 3rd Qu.    Max.**

**0.000   2.000   4.000   3.486   5.000   5.000**

**[1] 3.485902**

**[1] 4**

**[1] 1.292226**

**[1] 6**

**[1] 0 4 3 2 5 1**

1. **Cleanliness**

**Min. 1st Qu.  Median    Mean 3rd Qu.    Max.**

**0.000   3.000   4.000   3.706   5.000   5.000**

**[1] 3.705759**

**[1] 4**

**[1] 1.151774**

**[1] 6**

**[1] 3 4 1 2 5 0**

1. **Check-in Service**

**Min. 1st Qu.  Median    Mean 3rd Qu.    Max.**

**0.000   3.000   3.000   3.341   4.000   5.000**

**[1] 3.340807**

**[1] 3**

**[1] 1.260582**

**[1] 6**

**[1] 5 2 4 3 1 0**

1. **Baggage Handling**

**Min. 1st Qu.  Median    Mean 3rd Qu.    Max.**

**1.000   3.000   4.000   3.696   5.000   5.000**

**[1] 3.695673**

**[1] 4**

**[1] 1.156483**

**[1] 5**

**[1] 3 4 1 2 5**

RATIONALE FOR DIMENSIONALITY REDUCTION

As the dataset has 22 variables capturing different aspects of customer experience, dimensionality reduction serves two purposes:

1. Remove noise and collinear features: Helps improve model performance
2. Focus on key drivers of satisfaction: Results in more interpretable models

Implications on Analysis

The large sample size (130k observations) ensures even after splitting data for modeling, there is sufficient representation for all satisfaction levels (classes). It does not cover impact over time.

The data characteristics like high dimensionality with a mix of variable types necessitate specialized pre-processing actions like encoding, scaling before applying ML algorithms.

The inferences are focused on core aspects of the travel experience that are likely to shape satisfaction rather than causes unrelated to airline control like external delays.

1. MODEL SELECTION, ANALYSIS & RESULT

Our analysis commenced with the development of a linear regression model, wherein we created dummy variables to represent the binary outcome variable (customer satisfaction). Subsequently, we factorized all categorical variables and constructed a new data frame, selectively incorporating relevant columns. We then executed multiple linear regression iterations, systematically eliminating insignificant variables to refine the model.

During this refinement process, we removed variables such as gate location and departure delay, among others, based on their lack of statistical significance. Counterintuitively, the overall model fit decreased, as evidenced by a drop in the adjusted R-squared value to 0.40, indicating a poorer fit than anticipated.

Recognizing the binary nature of our outcome variable, we employed the 'glm' function to implement a generalized linear model (GLM). We performed multiple regression iterations, analyzing insignificant variables and judiciously incorporating an interaction term to enhance the model's significance. To evaluate the prediction accuracy achieved through this approach, we constructed a confusion matrix for the refined GLM.

Furthermore, we refined the model by excluding the 'departure delay' variable due to the presence of high multicollinearity observed between arrival and departure delays, which could potentially compromise the model's reliability.

EMPIRICAL RESULTS-

In this comprehensive analysis of the airline passenger satisfaction dataset, we developed a series of regression models to uncover the key factors influencing customer satisfaction, with a particular focus on flight delays, class, distance, in-flight amenities etc. The dataset was rigorously cleaned and preprocessed in R, ensuring the integrity and reliability of the data. We created a dummy variable, 'Feedbackdummy', to represent satisfied customers as 1 and neutral or dissatisfied customers as 0, enabling us to perform both linear and logistic regression analyses.

Our journey began with *model\_1*, a linear regression model encompassing a wide range of predictor variables. This initial model provided valuable insights into the significant factors contributing to customer satisfaction, such as customer loyalty, entertainment, service quality, and various aspects of the travel experience. However, we recognized the need to refine our approach and delve deeper into the intricacies of the data.

As we progressed through the modeling process, we made strategic decisions to exclude certain variables, aiming to enhance the model's fit and interpretability. In *model\_2*, we removed variables such as gate location, time convenience, online support, WiFi service, and baggage handling. This iterative process allowed us to focus on the most influential predictors and capture the essence of customer satisfaction.

For *model\_3* and *model\_4,* we further refined our analysis by exploring the impact of specific variable combinations. By carefully selecting and excluding variables based on their significance and relevance, we sought to uncover the true drivers of customer satisfaction. These models provided valuable insights into the complex interplay between various aspects of the airline experience and their effect on passenger sentiment.

In *model\_5*, we introduced a suitable interaction term between seat comfort and leg room. This interaction effect shed light on the crucial role of comfort in shaping passenger satisfaction. The model revealed that the positive impact of good seat comfort on satisfaction was significantly enhanced when accompanied by ample leg room. This finding underscores the importance of a holistic approach to passenger comfort, emphasizing the need for airlines to prioritize both seating arrangements and spaciousness.

Our final model, *model\_6*, provided a robust understanding of the key factors influencing passenger sentiment. The model highlighted the significance of class, with economy and economy plus classes having a negative impact on satisfaction compared to business class. Flight distance also emerged as a significant predictor, with longer distances associated with lower satisfaction levels.

In-flight amenities played a crucial role in shaping passenger satisfaction. The model revealed that good seat comfort, entertainment, and leg room had a positive impact on satisfaction. Notably, the interaction effect between seat comfort and leg room underscored the importance of a holistic approach to passenger comfort, emphasizing the need for airlines to prioritize both seating arrangements and spaciousness.

Flight delays, represented by the 'arrival\_delay' variable, demonstrated a significant negative impact on customer satisfaction. The model quantified the marginal effect of delays, indicating that a one-minute increase in arrival delay decreased the odds of a passenger being satisfied by approximately 0.5%. This finding highlights the critical importance of punctuality and efficient operations in maintaining high levels of customer satisfaction.

Throughout our analysis, we remained vigilant about the potential impact of multicollinearity on our models. The high correlation between 'depart\_delay' and 'arrival\_delay' variables warranted careful consideration. By examining the variance inflation factors (VIF) and making informed decisions about variable inclusion, we ensured the stability and reliability of our models.

From a real-world perspective, our findings offer actionable insights for airline executives and decision-makers. The models suggest that airlines should focus on providing a superior in-flight experience, particularly in terms of seat comfort, entertainment, and leg room. Moreover, the significant impact of flight delays underscores the need for airlines to prioritize operational efficiency and minimize disruptions to enhance customer satisfaction. The models consistently highlight the paramount importance of customer loyalty, emphasizing the need for airlines to foster long-term relationships with their passengers.

The evaluation of model performance through metrics such as the confusion matrix and accuracy score provided a clear indication of the effectiveness of our models. *Model\_6*, our final model, achieved an impressive accuracy of 78.65% in predicting customer satisfaction, showcasing its robustness and predictive power.

1. CONCLUSIONS AND RECOMMENDATIONS

Our comprehensive analysis of the airline passenger satisfaction dataset has unveiled crucial insights that can guide airlines in their pursuit of customer satisfaction and loyalty. The models consistently highlight the significance of factors such as customer loyalty, entertainment, service quality, seamless travel experiences, flight delays, class, distance, and in-flight amenities in shaping passenger sentiment and overall satisfaction.

The interaction effect between seat comfort and leg room emerges as a key finding, emphasizing the need for airlines to prioritize passenger comfort holistically. By investing in superior seating arrangements, ensuring ample leg room, and providing a differentiated in-flight experience across classes, airlines can create a memorable travel experience for their customers.

Our analysis also highlights the critical role of operational efficiency in maintaining high levels of customer satisfaction. Flight delays, represented by the 'arrival\_delay' variable, demonstrated a significant negative impact on satisfaction, with a one-minute increase in arrival delay decreasing the odds of a passenger being satisfied by approximately 0.5%. This finding underscores the importance of prioritizing punctuality and investing in robust systems, processes, and contingency plans to mitigate the negative impact of delays.

Based on our findings, we recommend the following strategies for airlines to maximize customer satisfaction:

1. Foster long-term relationships with passengers through effective loyalty programs and personalized experiences.

2. Invest in high-quality in-flight entertainment systems to keep passengers engaged and satisfied throughout the journey.

3. Prioritize service quality by training and empowering frontline staff to deliver exceptional customer service.

4. Streamline travel processes, such as check-in and boarding, to create a seamless and hassle-free experience for passengers.

5. Adopt a passenger-centric approach, focusing on comfort and well-being at every touchpoint of the journey, with particular emphasis on seating arrangements and leg room.

6. Optimize seat comfort and leg room, particularly on longer flights, by investing in ergonomic seating designs and exploring innovative ways to maximize space and comfort.

7. Continuously monitor and analyze customer feedback and satisfaction levels, leveraging advanced analytics techniques and machine learning algorithms to gain real-time insights and proactively address areas of improvement.

8. Prioritize operational efficiency to minimize flight delays and ensure punctuality, investing in robust systems, processes, and contingency plans to mitigate the negative impact of delays on customer satisfaction.

9. Communicate proactively with passengers in the event of delays or disruptions, providing timely and transparent information, offering assistance and compensation when appropriate, and demonstrating empathy and care for passengers' well-being.

By leveraging the insights from this analysis and implementing these recommendations, airlines can substantially improve customer satisfaction levels. The marginal effects of flight delays, class, distance, and in-flight amenities provide a roadmap for airlines to allocate resources effectively and prioritize initiatives that directly contribute to passenger satisfaction.

In conclusion, our empirical analysis emphasizes the critical role of service attributes in shaping airline passenger preferences and satisfaction. By understanding and quantifying the impact of these factors, airlines can make informed decisions, optimize their offerings, and create a differentiated travel experience that meets and exceeds customer expectations. The insights derived from this analysis serve as a foundation for airlines to build strong customer relationships, enhance loyalty, and thrive in an increasingly competitive market. By embracing a customer-centric approach, investing in key areas of the travel experience, and leveraging advanced analytics, airlines can position themselves for success and foster long-term customer relationships.

**References**

Choon, C. L. (2017). An importance-performance analysis to evaluate airlines service quality: The case study of a budget airline in Asia. Journal of Quality Assurance in Hospitality & Tourism, 8(3), 39-59.

Feng, C. M., & Jeng, K. Y. (2005). Analyzing airline service improvement strategy through importance and performance analysis. Journal of the Eastern Asia Society for Transportation Study. 6, 782–797.

Kei, C. M. (2019). A study of customer satisfaction in airlines. International Journal of Trade, Economics and Finance, 10(6), 144-148.

Mikulić, J., & Prebežac, D. (2011). What drives passenger loyalty to traditional and low-cost airlines? A formative partial least squares approach. Journal of Air Transport Management. 17(4), 237–240.

Thanasupsin, K., Chaichana, S., & Pliankarom, S. (2010). Factors influencing mode selections of low-cost carriers and a full-service airline in Thailand. Transportation Journal, 49(1), 35-47.

1. APPENDIX

**DATA DICTIONARY**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sr. No | Variable Name | Categories | Description | Renamed |
|  | Satisfaction | Satisfaction, neutral or dissatisfaction | Airline satisfaction level | Feedback |
|  | Age | Numerical | The actual age of the passengers | - |
|  | Gender | Female, Male | Gender of the passengers | - |
|  | Type of Travel | Personal Travel, Business Travel | Purpose of the flight of the passengers | travel\_type |
|  | Class | Business, Eco, Eco Plus | Travel class in the plane of the passengers | - |
|  | Customer Type | Loyal customer, disloyal customer | The customer type | cust\_type |
|  | Flight distance | Numerical | The flight distance of this journey | distance |
|  | Inflight wifi service | 0: Not Applicable ;1-5 | Satisfaction level of the inflight wifi service | wifi\_service |
|  | Ease of Online booking | Ordinal scale of 0-5 | Satisfaction level of online booking | online\_booking |
|  | Inflight service | Ordinal scale of 0-5 | Satisfaction level of inflight service | - |
|  | Online boarding | Ordinal scale of 0-5 | Satisfaction level of online boarding | online\_boarding |
|  | Inflight entertainment | Ordinal scale of 0-5 | Satisfaction level of inflight entertainment | entertainment |
|  | Food and drink | Ordinal scale of 0-5 | Satisfaction level of Food and drink | food\_drink |
|  | Seat comfort | Ordinal scale of 0-5 | Satisfaction level of Seat comfort | seat\_comfort |
|  | On-board service | Ordinal scale of 0-5 | Satisfaction level of On-board service | on\_board\_serv |
|  | Leg room service | Ordinal scale of 0-5 | Satisfaction level of Leg room service | leg\_room |
|  | Departure/Arrival time convenient | Ordinal scale of 0-5 | Satisfaction level of Departure/Arrival time convenient | - |
|  | Baggage handling | Ordinal scale of 0-5 | Satisfaction level of baggage handling | baggage\_handling |
|  | Gate location | Ordinal scale of 0-5 | Satisfaction level of Gate location | gate\_location |
|  | Cleanliness | Ordinal scale of 0-5 | Satisfaction level of Cleanliness | cleanliness |
|  | Check-in service | Ordinal scale of 0-5 | Satisfaction level of Check-in service | checkin\_serv |
|  | Departure Delay in Minutes | Numerical | Minutes delayed when departure | depart\_delay |
|  | Arrival Delay in Minutes | Numerical | Minutes delayed when Arrival | arrival\_delay |

**MODEL OUTPUTS**

Model 1

Call:

lm(formula = Feedbackdummy ~ . - Feedback, data = Project)

Residuals:

     Min       1Q   Median       3Q      Max

-1.19088 -0.29420  0.03151  0.25339  1.46404

Coefficients:

                             Estimate Std. Error t value Pr(>|t|)

(Intercept)                 3.763e-02  6.932e-03   5.429 5.69e-08 \*\*\*

GenderMale                 -1.663e-01  2.214e-03 -75.114  < 2e-16 \*\*\*

cust\_typeLoyal Customer     3.335e-01  3.411e-03  97.759  < 2e-16 \*\*\*

Age                        -1.109e-03  7.917e-05 -14.005  < 2e-16 \*\*\*

travel\_typePersonal Travel -1.145e-01  3.252e-03 -35.195  < 2e-16 \*\*\*

ClassEco                   -1.410e-01  3.039e-03 -46.406  < 2e-16 \*\*\*

ClassEco Plus              -1.552e-01  4.646e-03 -33.406  < 2e-16 \*\*\*

distance                   -1.952e-05  1.135e-06 -17.201  < 2e-16 \*\*\*

seat\_comfort2Good           5.921e-02  3.436e-03  17.233  < 2e-16 \*\*\*

time\_convenient2Good       -6.043e-02  3.088e-03 -19.568  < 2e-16 \*\*\*

food\_drink2Good            -3.264e-02  3.619e-03  -9.019  < 2e-16 \*\*\*

gate\_location2Good          5.533e-03  2.940e-03   1.882   0.0599 .

wifi\_service2Good          -4.302e-02  3.232e-03 -13.312  < 2e-16 \*\*\*

entertainment2Good          2.067e-01  3.465e-03  59.653  < 2e-16 \*\*\*

online\_support2Good        -6.164e-02  3.816e-03 -16.154  < 2e-16 \*\*\*

online\_booking2Good         1.475e-01  4.824e-03  30.572  < 2e-16 \*\*\*

on\_board\_serv2Good          1.355e-01  3.202e-03  42.308  < 2e-16 \*\*\*

leg\_room2Good               8.946e-02  2.805e-03  31.893  < 2e-16 \*\*\*

checkin\_serv2Good           1.524e-01  2.789e-03  54.647  < 2e-16 \*\*\*

baggage\_handling2Good      -9.755e-03  3.842e-03  -2.539   0.0111 \*

cleanliness2Good           -3.220e-02  3.964e-03  -8.123 4.57e-16 \*\*\*

online\_boarding2Good        1.397e-01  4.257e-03  32.809  < 2e-16 \*\*\*

depart\_delay                5.086e-04  1.092e-04   4.658 3.19e-06 \*\*\*

arrival\_delay              -1.218e-03  1.077e-04 -11.310  < 2e-16 \*\*\*

---

Signif. codes:  0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3891 on 129463 degrees of freedom

Multiple R-squared:  0.3891, Adjusted R-squared:  0.389

F-statistic:  3585 on 23 and 129463 DF,  p-value: < 2.2e-16

Model 2

Call:

lm(formula = Feedbackdummy ~ . - Feedback - gate\_location2 -

    time\_convenient2 - online\_support2 - wifi\_service2 - baggage\_handling2,

    data = Project)

Residuals:

     Min       1Q   Median       3Q      Max

-1.10490 -0.29763  0.02794  0.25265  1.44563

Coefficients:

                             Estimate Std. Error t value Pr(>|t|)

(Intercept)                -3.737e-03  6.550e-03  -0.571    0.568

GenderMale                 -1.690e-01  2.215e-03 -76.312  < 2e-16 \*\*\*

cust\_typeLoyal Customer     3.252e-01  3.381e-03  96.198  < 2e-16 \*\*\*

Age                        -1.149e-03  7.930e-05 -14.494  < 2e-16 \*\*\*

travel\_typePersonal Travel -1.226e-01  3.242e-03 -37.826  < 2e-16 \*\*\*

ClassEco                   -1.368e-01  3.044e-03 -44.940  < 2e-16 \*\*\*

ClassEco Plus              -1.507e-01  4.657e-03 -32.362  < 2e-16 \*\*\*

distance                   -1.978e-05  1.137e-06 -17.390  < 2e-16 \*\*\*

seat\_comfort2Good           4.940e-02  3.367e-03  14.675  < 2e-16 \*\*\*

food\_drink2Good            -6.009e-02  3.152e-03 -19.064  < 2e-16 \*\*\*

entertainment2Good          2.101e-01  3.185e-03  65.967  < 2e-16 \*\*\*

online\_booking2Good         1.066e-01  4.325e-03  24.659  < 2e-16 \*\*\*

on\_board\_serv2Good          1.402e-01  3.144e-03  44.592  < 2e-16 \*\*\*

leg\_room2Good               9.656e-02  2.782e-03  34.711  < 2e-16 \*\*\*

checkin\_serv2Good           1.444e-01  2.779e-03  51.965  < 2e-16 \*\*\*

cleanliness2Good           -2.064e-02  3.594e-03  -5.744 9.26e-09 \*\*\*

online\_boarding2Good        1.006e-01  3.936e-03  25.558  < 2e-16 \*\*\*

depart\_delay                4.963e-04  1.096e-04   4.529 5.92e-06 \*\*\*

arrival\_delay              -1.199e-03  1.081e-04 -11.091  < 2e-16 \*\*\*

---

Signif. codes:  0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3905 on 129468 degrees of freedom

Multiple R-squared:  0.3847, Adjusted R-squared:  0.3846

F-statistic:  4497 on 18 and 129468 DF,  p-value: < 2.2e-16

Model 3

Call:

lm(formula = Feedbackdummy ~ . - Feedback - gate\_location2 -

    time\_convenient2 - online\_support2 - wifi\_service2 - baggage\_handling2 -

    cleanliness2 - depart\_delay, data = Project)

Residuals:

     Min       1Q   Median       3Q      Max

-1.10636 -0.29741  0.02635  0.25219  1.43758

Coefficients:

                             Estimate Std. Error t value Pr(>|t|)

(Intercept)                -1.485e-02  6.271e-03  -2.369   0.0179 \*

GenderMale                 -1.694e-01  2.214e-03 -76.528   <2e-16 \*\*\*

cust\_typeLoyal Customer     3.267e-01  3.370e-03  96.920   <2e-16 \*\*\*

Age                        -1.124e-03  7.922e-05 -14.190   <2e-16 \*\*\*

travel\_typePersonal Travel -1.236e-01  3.237e-03 -38.174   <2e-16 \*\*\*

ClassEco                   -1.363e-01  3.042e-03 -44.796   <2e-16 \*\*\*

ClassEco Plus              -1.506e-01  4.657e-03 -32.336   <2e-16 \*\*\*

distance                   -1.987e-05  1.137e-06 -17.480   <2e-16 \*\*\*

seat\_comfort2Good           4.909e-02  3.367e-03  14.581   <2e-16 \*\*\*

food\_drink2Good            -6.019e-02  3.153e-03 -19.093   <2e-16 \*\*\*

entertainment2Good          2.111e-01  3.181e-03  66.368   <2e-16 \*\*\*

online\_booking2Good         9.906e-02  4.118e-03  24.053   <2e-16 \*\*\*

on\_board\_serv2Good          1.356e-01  3.038e-03  44.623   <2e-16 \*\*\*

leg\_room2Good               9.420e-02  2.750e-03  34.253   <2e-16 \*\*\*

checkin\_serv2Good           1.429e-01  2.767e-03  51.653   <2e-16 \*\*\*

online\_boarding2Good        1.056e-01  3.840e-03  27.498   <2e-16 \*\*\*

arrival\_delay              -7.197e-04  2.847e-05 -25.277   <2e-16 \*\*\*

---

Signif. codes:  0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3905 on 129470 degrees of freedom

Multiple R-squared:  0.3844, Adjusted R-squared:  0.3844

F-statistic:  5054 on 16 and 129470 DF,  p-value: < 2.2e-16

Model 4

Call:

glm(formula = as.factor(Feedback) ~ . - Feedbackdummy - gate\_location2 -

    time\_convenient2 - online\_support2 - wifi\_service2 - baggage\_handling2,

    family = binomial, data = Project)

Coefficients:

                             Estimate Std. Error z value Pr(>|z|)

(Intercept)                -2.884e+00  4.308e-02 -66.936  < 2e-16 \*\*\*

GenderMale                 -1.042e+00  1.489e-02 -69.975  < 2e-16 \*\*\*

cust\_typeLoyal Customer     1.905e+00  2.252e-02  84.573  < 2e-16 \*\*\*

Age                        -7.319e-03  5.161e-04 -14.181  < 2e-16 \*\*\*

travel\_typePersonal Travel -6.926e-01  2.076e-02 -33.360  < 2e-16 \*\*\*

ClassEco                   -8.630e-01  1.936e-02 -44.571  < 2e-16 \*\*\*

ClassEco Plus              -9.382e-01  2.941e-02 -31.905  < 2e-16 \*\*\*

distance                   -1.209e-04  7.723e-06 -15.649  < 2e-16 \*\*\*

seat\_comfort2Good           2.801e-01  2.196e-02  12.751  < 2e-16 \*\*\*

food\_drink2Good            -3.004e-01  2.065e-02 -14.551  < 2e-16 \*\*\*

entertainment2Good          1.152e+00  2.026e-02  56.874  < 2e-16 \*\*\*

online\_booking2Good         5.390e-01  2.693e-02  20.016  < 2e-16 \*\*\*

on\_board\_serv2Good          8.957e-01  2.121e-02  42.232  < 2e-16 \*\*\*

leg\_room2Good               5.455e-01  1.795e-02  30.392  < 2e-16 \*\*\*

checkin\_serv2Good           9.134e-01  1.811e-02  50.425  < 2e-16 \*\*\*

cleanliness2Good           -2.018e-01  2.377e-02  -8.487  < 2e-16 \*\*\*

online\_boarding2Good        6.534e-01  2.402e-02  27.204  < 2e-16 \*\*\*

depart\_delay                3.086e-03  7.275e-04   4.241 2.22e-05 \*\*\*

arrival\_delay              -8.079e-03  7.159e-04 -11.286  < 2e-16 \*\*\*

---

Signif. codes:  0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 178341  on 129486  degrees of freedom

Residual deviance: 119278  on 129468  degrees of freedom

AIC: 119316

Number of Fisher Scoring iterations: 5

Model 5

Call:

glm(formula = as.factor(Feedback) ~ . - Feedbackdummy - gate\_location2 -

    time\_convenient2 - online\_support2 - wifi\_service2 - baggage\_handling2 +

    seat\_comfort2 \* leg\_room2, family = binomial, data = Project)

Coefficients:

                                  Estimate Std. Error z value Pr(>|z|)

(Intercept)                     -3.227e+00  4.514e-02 -71.495  < 2e-16 \*\*\*

GenderMale                      -1.082e+00  1.504e-02 -71.930  < 2e-16 \*\*\*

cust\_typeLoyal Customer          1.955e+00  2.264e-02  86.387  < 2e-16 \*\*\*

Age                             -6.475e-03  5.175e-04 -12.512  < 2e-16 \*\*\*

travel\_typePersonal Travel      -7.062e-01  2.080e-02 -33.957  < 2e-16 \*\*\*

ClassEco                        -9.071e-01  1.948e-02 -46.558  < 2e-16 \*\*\*

ClassEco Plus                   -9.886e-01  2.950e-02 -33.512  < 2e-16 \*\*\*

distance                        -1.376e-04  7.792e-06 -17.656  < 2e-16 \*\*\*

seat\_comfort2Good                1.014e+00  3.336e-02  30.401  < 2e-16 \*\*\*

food\_drink2Good                 -2.482e-01  2.083e-02 -11.914  < 2e-16 \*\*\*

entertainment2Good               1.156e+00  2.061e-02  56.097  < 2e-16 \*\*\*

online\_booking2Good              4.635e-01  2.717e-02  17.057  < 2e-16 \*\*\*

on\_board\_serv2Good               8.641e-01  2.124e-02  40.692  < 2e-16 \*\*\*

leg\_room2Good                    1.216e+00  2.937e-02  41.404  < 2e-16 \*\*\*

checkin\_serv2Good                9.080e-01  1.817e-02  49.970  < 2e-16 \*\*\*

cleanliness2Good                -2.686e-01  2.395e-02 -11.216  < 2e-16 \*\*\*

online\_boarding2Good             6.725e-01  2.415e-02  27.854  < 2e-16 \*\*\*

depart\_delay                     3.208e-03  7.300e-04   4.395 1.11e-05 \*\*\*

arrival\_delay                   -8.296e-03  7.184e-04 -11.548  < 2e-16 \*\*\*

seat\_comfort2Good:leg\_room2Good -1.057e+00  3.553e-02 -29.755  < 2e-16 \*\*\*

---

Signif. codes:  0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 178341  on 129486  degrees of freedom

Residual deviance: 118369  on 129467  degrees of freedom

AIC: 118409

Number of Fisher Scoring iterations: 5

Model 6

Call:

glm(formula = as.factor(Feedback) ~ . - Feedbackdummy - gate\_location2 -

    time\_convenient2 - online\_support2 - wifi\_service2 - baggage\_handling2 +

    seat\_comfort2 \* leg\_room2 - depart\_delay, family = binomial,

    data = Project)

Coefficients:

                                  Estimate Std. Error z value Pr(>|z|)

(Intercept)                     -3.228e+00  4.513e-02  -71.53   <2e-16 \*\*\*

GenderMale                      -1.081e+00  1.504e-02  -71.91   <2e-16 \*\*\*

cust\_typeLoyal Customer          1.954e+00  2.263e-02   86.36   <2e-16 \*\*\*

Age                             -6.448e-03  5.174e-04  -12.46   <2e-16 \*\*\*

travel\_typePersonal Travel      -7.056e-01  2.080e-02  -33.93   <2e-16 \*\*\*

ClassEco                        -9.073e-01  1.948e-02  -46.57   <2e-16 \*\*\*

ClassEco Plus                   -9.889e-01  2.950e-02  -33.52   <2e-16 \*\*\*

distance                        -1.367e-04  7.788e-06  -17.56   <2e-16 \*\*\*

seat\_comfort2Good                1.014e+00  3.336e-02   30.39   <2e-16 \*\*\*

food\_drink2Good                 -2.482e-01  2.083e-02  -11.91   <2e-16 \*\*\*

entertainment2Good               1.156e+00  2.061e-02   56.11   <2e-16 \*\*\*

online\_booking2Good              4.630e-01  2.717e-02   17.04   <2e-16 \*\*\*

on\_board\_serv2Good               8.642e-01  2.124e-02   40.70   <2e-16 \*\*\*

leg\_room2Good                    1.216e+00  2.937e-02   41.40   <2e-16 \*\*\*

checkin\_serv2Good                9.081e-01  1.817e-02   49.98   <2e-16 \*\*\*

cleanliness2Good                -2.683e-01  2.395e-02  -11.20   <2e-16 \*\*\*

online\_boarding2Good             6.729e-01  2.414e-02   27.87   <2e-16 \*\*\*

arrival\_delay                   -5.266e-03  2.009e-04  -26.21   <2e-16 \*\*\*

seat\_comfort2Good:leg\_room2Good -1.056e+00  3.552e-02  -29.73   <2e-16 \*\*\*

---

Signif. codes:  0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 178341  on 129486  degrees of freedom

Residual deviance: 118388  on 129468  degrees of freedom

AIC: 118426

Number of Fisher Scoring iterations: 5

MULTICOLLINEARITY

Model 1

    GVIF Df GVIF^(1/(2\*Df))

Gender             1.048199  1        1.023816

cust\_type          1.489039  1        1.220262

Age                1.225214  1        1.106894

travel\_type        1.932656  1        1.390200

Class              1.880946  2        1.171100

distance           1.161476  1        1.077718

seat\_comfort2      2.458822  1        1.568063

time\_convenient2   1.935836  1        1.391343

food\_drink2        2.724266  1        1.650535

gate\_location2     1.709094  1        1.307323

wifi\_service2      1.951634  1        1.397009

entertainment2     1.982942  1        1.408170

online\_support2    2.272012  1        1.507320

online\_booking2    3.802440  1        1.949985

on\_board\_serv2     1.573412  1        1.254357

leg\_room2          1.283877  1        1.133083

checkin\_serv2      1.205119  1        1.097779

baggage\_handling2  1.738470  1        1.318511

cleanliness2       1.833780  1        1.354171

online\_boarding2   2.991856  1        1.729698

depart\_delay      14.673759  1        3.830634

arrival\_delay     14.680472  1        3.831510

Model 2

    GVIF Df GVIF^(1/(2\*Df))

Gender            1.041009  1        1.020299

cust\_type         1.452024  1        1.205000

Age               1.220447  1        1.104738

travel\_type       1.906757  1        1.380854

Class             1.871620  2        1.169646

distance          1.158473  1        1.076324

seat\_comfort2     2.344021  1        1.531020

food\_drink2       2.051457  1        1.432291

entertainment2    1.663486  1        1.289762

online\_booking2   3.034420  1        1.741959

on\_board\_serv2    1.505967  1        1.227179

leg\_room2         1.253716  1        1.119695

checkin\_serv2     1.188101  1        1.090001

cleanliness2      1.496884  1        1.223472

online\_boarding2  2.540380  1        1.593857

depart\_delay     14.672683  1        3.830494

arrival\_delay    14.679324  1        3.831361

Model 5

    GVIF Df GVIF^(1/(2\*Df))

Gender                   1.083126  1        1.040733

cust\_type                1.420020  1        1.191646

Age                      1.227105  1        1.107748

travel\_type              1.910963  1        1.382376

Class                    1.704919  2        1.142683

distance                 1.208622  1        1.099374

seat\_comfort2            4.904632  1        2.214640

food\_drink2              1.930759  1        1.389517

entertainment2           1.460619  1        1.208561

online\_booking2          2.660676  1        1.631158

on\_board\_serv2           1.580600  1        1.257219

leg\_room2                3.239018  1        1.799727

checkin\_serv2            1.141029  1        1.068189

cleanliness2             1.696209  1        1.302386

online\_boarding2         2.077992  1        1.441524

depart\_delay            13.197825  1        3.632881

arrival\_delay           13.208558  1        3.634358

seat\_comfort2:leg\_room2  6.042340  1        2.458117

CONFUSION MATRIX

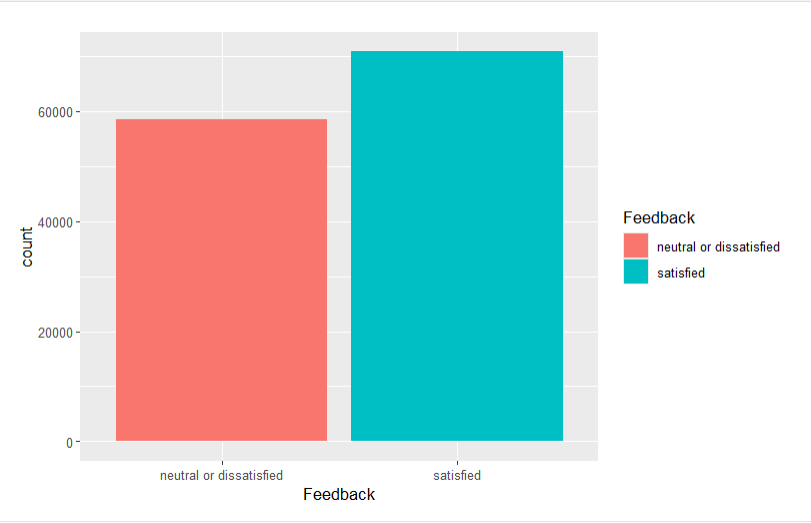
Model 5

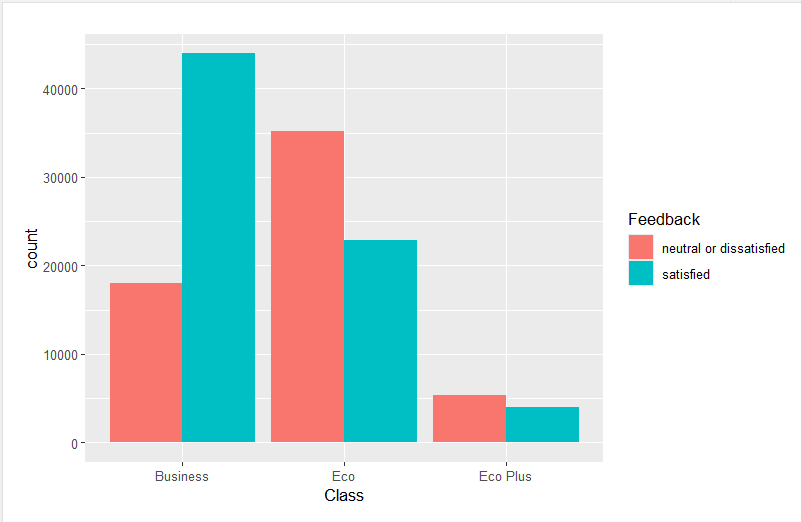
**neutral or dissatisfied   satisfied**

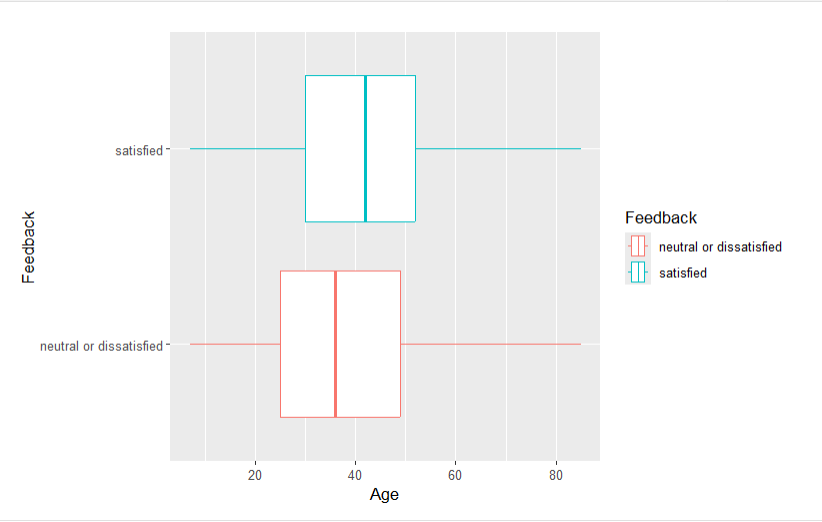
**neutral or dissatisfied                   43137            12173**

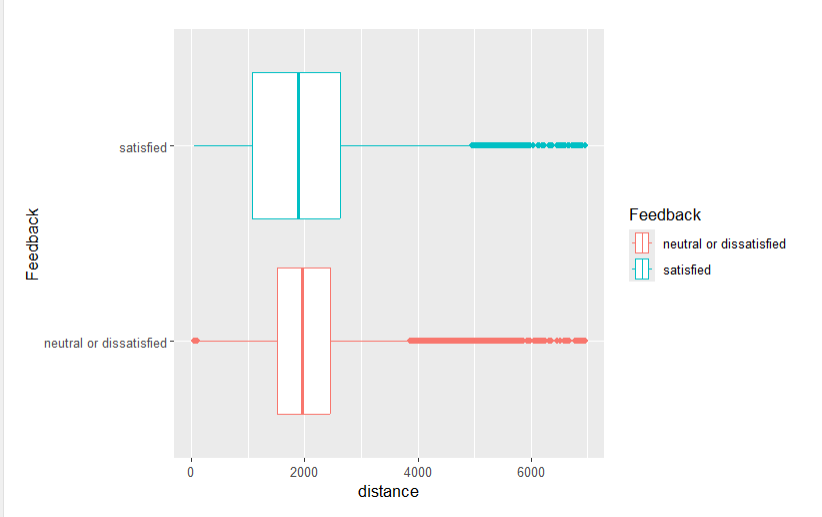
**satisfied                                 15468            58709**

**DATA VISUALIZATION**







****