Factors Affect Airline Customers' Satisfaction: Data Mining

Yifei Wang(✉)

Department of Accounting, Dongbei University of Finance and Economics, Dalian 116012, China
wangyifei20021029@163.com

Abstract. This study investigated different factors that affect customer satisfaction in the airline industry. These factors include demographic elements (gender and age), and other objective factors like delay time, travel distance, customer class, type of class, and customers' impressions of different aspects of airline services and facilities. After the exploratory data analysis, statistical approaches were chosen to conduct this research, including comparative analysis and logistics regression. This paper discusses the above-influencing factors and explores the diversity of different categories of customers' concerns about various aspects of flight services. Finally, a logistic regression model was developed to predict flight satisfaction. The main findings are as follows.

- Males and females have different concerns about needs during travel.
- Customers selecting the business class and those taking business travel are almost the same groups of people, who are generally disloyal. In addition, older customers are more likely to be in better cabins.
- Travelers are more tolerant of departure delays than arrival delays.
- Business travelers are more sensitive to delays. Long-distance flights suffer relatively less delay.

Keywords: airline · customer satisfaction · factor analysis · comparative analysis · logistic regression

1 Introduction

Researchers predicted that the airline market will reach $1,048.24 billion by 2027, increasing from $821.96 billion in 2019 at a healthy rate of 16.0%. However, due to the impact of COVID-19, airline companies faced short-term cashflow shortages [1]. Luckily, this shortage has stopped and an upcoming surge in demand will reach. What airline companies should consider right now is how to attract the increasing customer flow. Accordingly, airline companies continually pursue many ways to protect their shares in the market, retain their loyal customers, and attract new customers. Customer satisfaction will directly affect customers' choice of which airline company they are going to choose so improving customers' satisfaction in every aspect is what airline companies should seriously consider right now [2]. This paper aims to find out some

V. Gaikar et al. (Eds.): ECIT 2023, AHE 18, pp. 344–362, 2023.
https://doi.org/10.2991/978-94-6463-210-1_43

major factors that would cause significant affection in customer satisfaction to help airline companies make choices and adjustments to their present services and campaign strategies. The paper is organized as follows: in Sect. 5, the dataset used in the research is described. A series of statistical analyses – factor analysis, competitive analysis, and logistic regression – is processed after the description. In Sect. 6, a discussion on several significant factors that are studied in this research is posted. Speculations and Managerial Implications are given.

2 Literature Review

The factors that affect the satisfaction of airlines have been under discussion. "Airline tangibles" was found to be the most significant to affect both customer satisfaction and repurchase intention. Findings also showed that customer satisfaction is positively related to repurchase and word-of-mouth intentions [3]. Then the study found that the order of importance of the dimensions of service quality tested here is: flight schedules; flight attendants; tangibles; and ground staff. Passenger satisfaction with these service-quality dimensions is very important in explaining behavioral intentions. Satisfied passengers are mostly influenced by the schedule [4]. When focusing on human elements, Babbar, S. and Koufteros, X. found that individual attention, helpfulness, courtesy, and promptness have a significant effect on airline passenger satisfaction [5]. When it comes to recent researches, a study about PIA (Pakistan International Airlines) figured out that customer satisfaction of customers is influenced by all of the five service quality dimensions, including airline tangibles, terminal tangibles, personnel, empathy, and image [6]. In addition, online/mobile boarding, inflight wi-fi service, baggage handling, and inflight entertainment, in succession, is the top 4 crucial service to be improved by the airline to gain passenger satisfaction [7]. Going further, a study presented the application of our MOEOA-based knowledge-discovery business-intelligence technique (fuzzy rule-based classification systems) characterized by genetically optimized interpretability-accuracy trade-off to decision support related to airline passenger satisfaction problems [8]. And scaling out, there are fundamental differences in the drivers of passenger satisfaction depending on the travel class. Friendliness and helpfulness of staff are the key factors for those traveling in Economy Class, product value is key for those in premium cabins [9, 10].

3 Eda

3.1 Shape

3.1.1 Characters

There are 129880 observations in total and 23 characters, which can be divided into the following three categories as Table 1 (except for 'satisfaction').

Table 1. Three Categories from 22 Characters [Owner-draw]

passenger info	passenger score	delay time
Gender	inflight_wifi_service	departure_delay_in_minutes
age	departure_arrival_time_convenient	arrival_delay_in_minutes
Customer_type	ease_of_online_booking	——
type_of_travel	gate_location	
customer_class	food_and_drink	
flight_distance	online_boarding	
——	seat_comfort	
	inflight_entertainment	
	onboard_service	
	leg_room_service	
	baggage_handling	
	checkin_service	
	inflight_service	
	cleanliness	

3.1.2 Distribution

Based on the basic information of the surveyed customers, we had the overall image of the dataset. As illustrated in, the proportion of men and women is almost equal among the surveyed customers. Age was approximately normally distributed, with a mean age of 39. Regarding travel type, respondents on business travel were about twice as likely as those on private travel. As for the cabin type, 48% and 45% of passengers chose business and economy class, respectively. For 'satisfaction', the variable to be predicted, 43% are satisfied and 57% are neutral or dissatisfied (see Fig. 1).

3.2 Correlation

As shown in Fig. 2, Some variables have a strong positive correlation with each other. There is a correlation of 0.96 between the delay time of departure and arrival, which means flights with late departures will arrive late. In addition, a strong correlation exists of satisfaction variables for in-flight services and facilities (see Fig. 3).

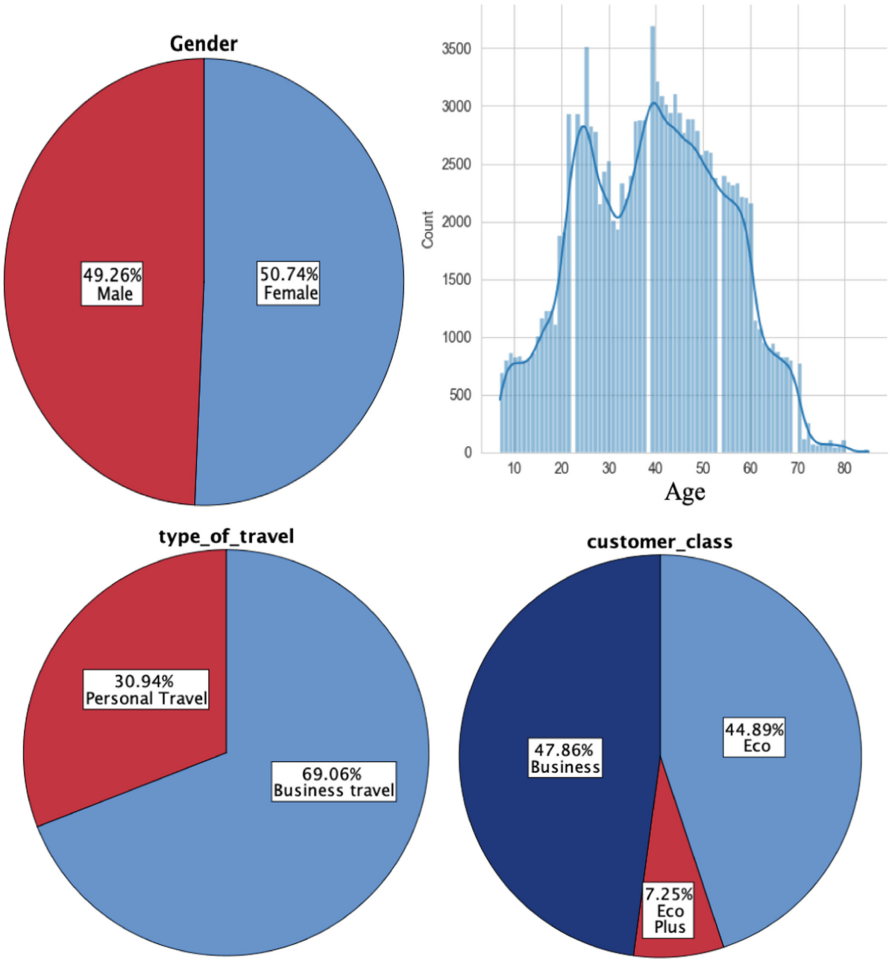

Fig. 1. Customer Distribution of Gender, Age, Type of Travel and Customer Class [Owner-draw]

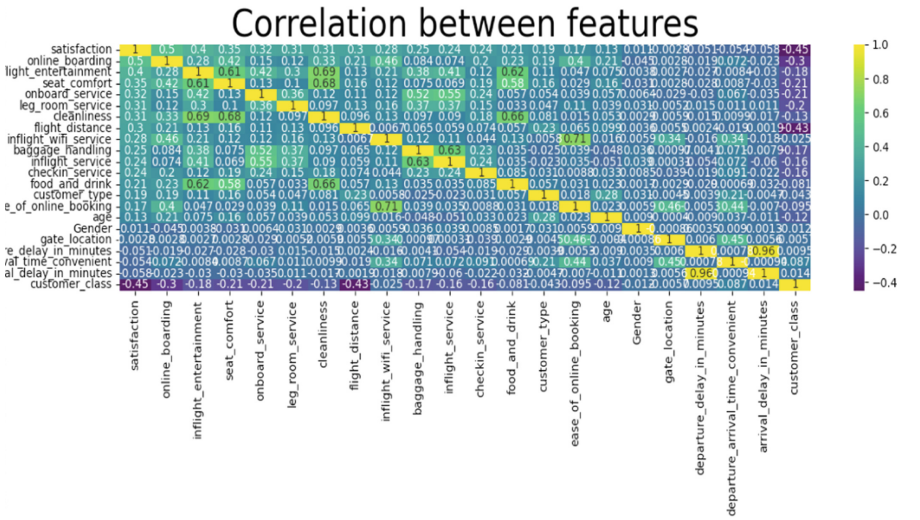

Fig. 2. Correlation between features [Owner-draw]

Fig. 3. Correlation of satisfaction variables for in-flight services and facilities [Owner-draw]

4 Operations and Results

4.1 Basic Data Processing

4.1.1 Missing Value Replacement

There were 393 missing values in the variable “arrival_delay_in_minutes”, which is replaced with the mean value.

4.1.2 Calculate the Variable “TOTAL”

The new variable is formed by summing up all the variables on the five-point scale of satisfaction.

Table 2. KMO and Bartlett's Test [Owner-draw]

Testing Method	Indicator	
Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.274
Bartlett's Test of Sphericity	Approx. Chi-Square	1777150.592
	df	231
	Sig.	.000

4.1.3 Factor Analysis

The first step in the analysis of the dataset is factor analysis. This is to quickly understand the components of airline satisfaction, to explore the principal factors and their components, and to prepare for the subsequent decomposition analysis satisfaction prediction.

4.1.4 Factor Analysis Suitability Test

A major basis of factor analysis is the correlation between each variable, therefore, before factor analysis of information data, data should be tested with Bartlett's Test of Sphericity as well as the KMO test. "Inflight_wifi_service" and "ease_of_online_booking" were found to be extremely correlated, resulting in a non-positive definite matrix, so one of them was deleted to perform the tests. The results in Table 2 show that Bartlett's spherical test results correspond to the $P < 0.05$ confidence level, meaning that variables are correlated and factor analysis can be performed.

4.1.5 Common Factor Extraction and Naming

By observing the screen plot and total variance interpretation (Fig. 4 and Table 3) from principal component analysis, most of the data of the original variables were covered by 10 common factors, which resulted in a total explanatory variance of more than 80%. Among them, the first four variables provided a large portion of the explanation. Additionally, except for 3 variables, the rest of the variables were extracted with more than 70% of their values, indicating that they can all be reasonably expressed by the common factor.

The initial factor loading matrix was rotated orthogonally with maximum variance to obtain 10 new factor loadings, as shown in Table 4. For illustration, we analyzed only the first five factors. The first common factor (Fac1) is mainly related to cleanliness, food and drink, seat comfort, and inflight entertainment, which can be named "Fundamental Items". The second common factor (Fac2) is mainly related to inflight service, baggage handling, onboard service, and legroom service, which can be named "Basic Service". Compared to Fac2, Fac1 is a more significant determinant of satisfaction, probably because Fundamental Items make up the overall environment of the journey. People generally have a deeper experience of it. And the services are a plus or minus for satisfaction based on the environment. The third common factor (Fac3) is mainly related to inflight Wi-Fi service, ease of online booking, and online boarding, which are more

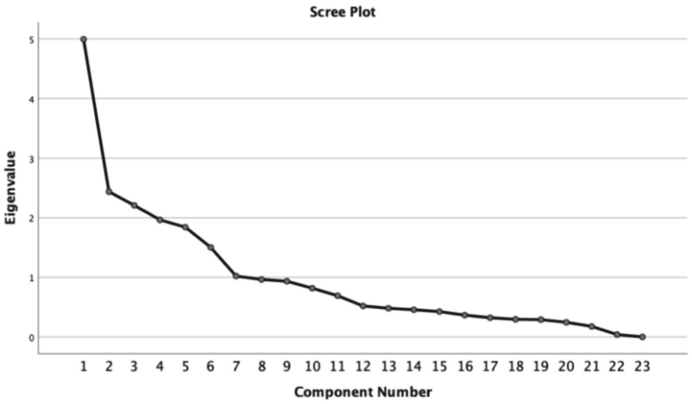

Fig. 4. Screen Plot [Owner-draw]

Table 3. Total Variance Explained [Owner-draw]

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	4.994	21.715	21.715
2	2.439	10.602	32.317
3	2.208	9.601	41.918
4	1.966	8.546	50.464
5	1.843	8.011	58.476
6	1.503	6.537	65.012
7	1.021	4.440	69.452
8	.965	4.197	73.649
9	.934	4.063	77.711
10	.818	3.557	81.268

peripheral services than the basic services in factor 2, named “Secondary Service”. The fourth common factor (Fac4) is mainly related to departure delay and arrival delay, which can be named “Delay Time”. The fifth common factor (Fac5) is mainly related to flight distance, customer class, and type of travel, which are all about flight information of customers, named “Passenger Air Info”. Customer information did not appear until the fifth factor, which indicates that objective factors probably do not determine satisfaction very well.

The implication of the above factors was well defined, and the next step was to use the original variables as well as the factors for comparative analysis to explore more connections in the dataset.

Table 4. Rotated Component Matrix [Owner-draw]

Factor	Name	VAR 1	VAR 2	VAR 3	VAR 4	VAR 5
1	Fundamental Items	cleanliness (0.889)	food_and_drink (0.859)	seat_comfort (0.819)	inflight_ entertainment (0.791)	Total (0.579)
2	Basic Service	inflight_ service (0.832)	baggage_handling (0.815)	onboard_service (0.773)	leg_room_service (0.631)	Total (0.541)
3	Secondary Service	inflight_wifi_service (0.853)	ease_of_online _booking (0.803)	online_boarding (0.714)	N/A	N/A
4	Delay Time	departure_delay _in_minutes (0.989)	arrival_delay _in_minutes (0.989)	N/A	N/A	N/A
5	Passenger Air Info	flight_distance (0.872)	customer_class (-0.785)	type_of_travel (-0.581)	N/A	N/A
6	—	gate_location (0.872)	departure_arrival_ time_convenient (0.732)	N/A	N/A	N/A
7	—	customer_type (0.860)	type_of_travel (0.636)	N/A	N/A	N/A
8	—	checkin_service (0.926)	N/A	N/A	N/A	N/A
9	—	age (0.957)	N/A	N/A	N/A	N/A
10	—	Gender (0.997)	N/A	N/A	N/A	N/A

Extraction Method: Principal Component Analysis.
a. Rotation converged in 8 iterations.

4.2 Comparative Analysis

We were interested in three areas:

1 Association of traveler demographic characteristics (gender & age) with satisfaction as well as these factors. This can help airlines to develop personalized services for passengers.
2 Association between passenger flight information and satisfaction, and the internal association of three types of flight information. This can reflect the group of customers that the airline focuses more on for satisfaction.
3 Objective facts versus subjective evaluation for the same issue (delay time & flight distance). For delay time, we wondered about the patterns of delay times faced by different types of travelers and their subjective ratings of delay. We also wanted to find a threshold, on the two sides of which, satisfaction shows different trends. For flight distance, we wondered about its relationship with delay, customer information, and satisfaction indicators.

4.2.1 Customer Demographic Characteristics

Firstly, about gender. The result in Table 5 showed that there were highly significant differences in satisfaction at gender levels.

Then the differences in specific services were analyzed using the factors. From Table 6 and Table 7, the most significant difference between males and females was observed in Fac2 (Basic Service), on which females give lower scores on average. There is also a significant difference in Fac1&3 (Fundamental Items & Secondary Service), with males scoring lower average.

To further investigate the reasons for the above differences, we conducted independent samples t-tests for the variables in each of the 3 factors. From Fig. 5 and the result of One-way ANOVA, variables with significant differences in Fac2 are legroom service, baggage handling, and inflight service, with women's scores lower. While in Fac1, the

Table 5. Independent Samples T-test [Owner-draw]

	variances assumed	Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig.	Mean Difference	Std. Error Difference	95% Confidence Interval	
									Lower	Upper
satisfaction	Equal	64.996	.000	-4.050	129878	.000	-.011	.003	-.017	-.006
	Not Equal			-4.049	129740.909	.000	-.011	.003	-.017	-.006

Table 6. Cross tabulation- Gender & Fundamental Items, Basic Service, Secondary Service [Owner-draw]

Gender	Fac	Mean	Fac	Mean	Fac	Mean
Female	Fac1	0.006	Fac2	-0.026	Fac3	0.006
Male		-0.006		0.026		-0.006
Total		0.000		0.000		0.000

Table 7. One-way ANOVA [Owner-draw]

	Sum of Squares	df	Mean Square	F	Sig.
FAC1_1	4.396	1	4.396	4.396	.036
FAC2_1	87.648	1	87.648	87.706	.000
FAC3_1	5.083	1	5.083	5.083	.024

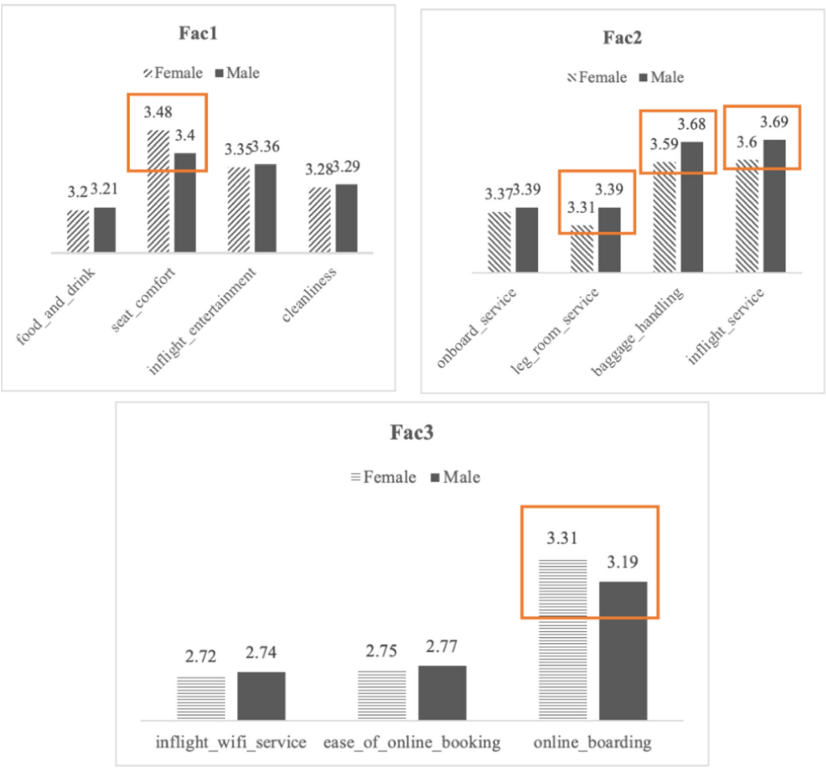

Fig. 5. Differences in Fac1 to 3 by Gender [Owner-draw]

variable with the larger difference is seat comfort and online boarding, with men's scores lower.

Secondly, about age. Figure 6 as well as the result of its One-way ANOVA showed that for different ages, there were significant differences in satisfaction, and the average age of satisfied passengers was 4 years higher than that of neutral or unsatisfied. In addition, by investigating the correlation between age and cabin type, it can be found that the more advanced the cabin type, the older the average age of the passengers.

4.2.2 Flight Information

Explore the connection between airline-related passenger information (customer class, type of travel, loyalty), and which class of people is more important to the airline's sense of experience. Based on the above questions, the following hypotheses are proposed.

H1: Business class passengers are more often on business travel, and most organizations are used to buying tickets from the same airline, so most of them are loyal customers.

H2: Airlines focus more on business class, business trips, and the experience of loyal customers.

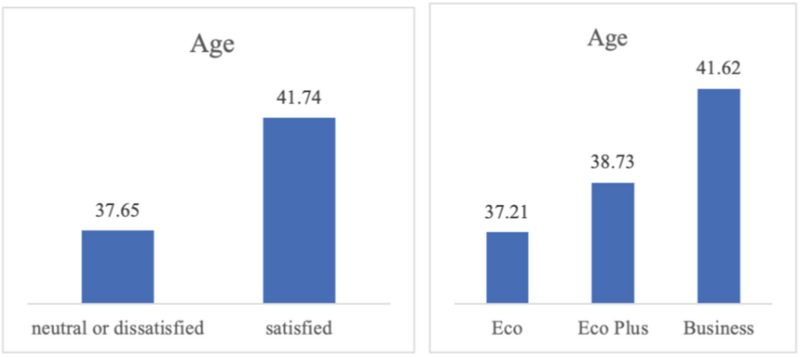

Fig. 6. Age Means for Different Satisfaction Scenarios and Class [Owner-draw]

Fig. 7. Relationship among Business Class, Business Travel and Disloyal Customers [Owner-draw]

From the cross-tabulation between Customer Class and Type of Travel, 95.7% of business class are purchased by travelers of business travel. Then through that of Type of Travel and Customer Type, it is shown that 99.2% of business travel customers are disloyal (as Fig. 7). This means customers selecting the business class and those taking business travel are almost the same group of people, who are generally disloyal.

The H2 conjecture is then confirmed by hypothesis testing. As shown in Table 9, satisfaction was significantly different at different levels of Customer Type, Type of Travel, and Customer Class relatively. And post hoc tests showed that for satisfaction there was a significant difference between each of the two class types. Therefore, it can be seen in Table 8 that the proportion of satisfied customers is highest among business class, business trips, and loyal customers.

4.2.3 Subjective Versus Objective

Firstly, the delay time is an objective value, and we explored its association with the satisfaction of unrelated items —— how could delay time affect the rating for factor 1, factor 2, and factor 3? For using One-way ANOVA, the two kinds of delay time were separately divided into 2 groups (Short Delay & Long Delay) bounded by their mean value. Results were significant, showing that the group with longer delays scored lower for Fundamental Items, Basic Service, and Secondary Service (Table 10 and Table 11).

Table 8. Satisfaction Rate of Different Customer Class, Type of Travel, Customer Type [Owner-draw]

Customer Class	% of satisfied	Type of Travel	% of satisfied	Customer Type	% of satisfied
Business	69%	Business Travel	58%	Loyal Customer	48%
Eco Plus	25%	Personal Travel	10%	disloyal Customer	24%
Eco	19%	—	—	—	—

Table 9. One-way ANOVA for Customer Flight Information [Owner-draw]

Variable	Sum of Squares	df	Mean Square	F	Sig.
Customer Type	1104.237	1	1104.237	4655.170	.000
Type of travel	6458.225	1	6458.225	32952.935	.000
Customer Class	8085.206	2	4042.603	22035.610	.000

Table 10. Fac1 ~ 3 of Different Delay Groups [Owner-draw]

Fac1	Delay Group	Mean of Departure Delay	Mean of Arrival Delay
Fac1	Short Delay	0.0090	0.0139
	Long Delay	-0.0275	-0.0422
Fac2	Short Delay	0.0125	0.0178
	Long Delay	-0.0378	-0.0542
Fac3	Short Delay	0.0105	0.0114
	Long Delay	-0.0319	-0.0346

Then we explored the trend in satisfaction with delay time. Distribution-Skewness of two types of delay time both showed long-tail distribution (see Fig. 8). Therefore, we grouped them from the left side, in groups of 3 min (the last group contains all of the rest), and compared the satisfaction of each group by cross-tabulation.

Figure 9 showed that for departure delay, the satisfaction rate continues to decline from group 0 to 8 (i.e., before the arrival time delay reaches 27 min); beyond 27 min, the satisfaction rate hardly changes. For Arrival Delay, the satisfaction rate continues to decline from group 0 to 2 (i.e., before the arrival time delay reaches 9 min); beyond 9 min, there is virtually no more change in satisfaction.

Next, we examined the differences in subjective perceptions of delays among different groups of passengers. Are business travelers more demanding of schedule? The ANOVA tests were conducted separately for departure-arrival time convenience as well

Table 11. One-way ANOVA for Delay & Fac1 ~ 3 [Owner-draw]

Type of Delay	Variable	Sum of Squares	df	Mean Square	F	Sig.
Departure Delay	Fac1	32.251	1	32.251	32.259	.000
	Fac2	61.262	1	61.262	61.290	.000
	Fac3	43.662	1	43.662	43.676	.000
Arrival Delay	Fac1	76.185	1	76.185	76.229	.000
	Fac2	125.591	1	125.591	125.711	.000
	Fac3	51.462	1	51.462	51.482	.000

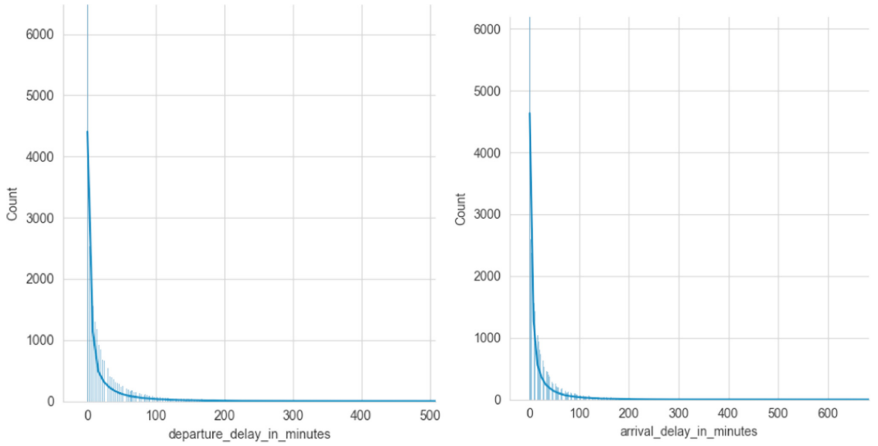

Fig. 8. Distribution-Skewness of Two Types of Delay Time [Owner-draw]

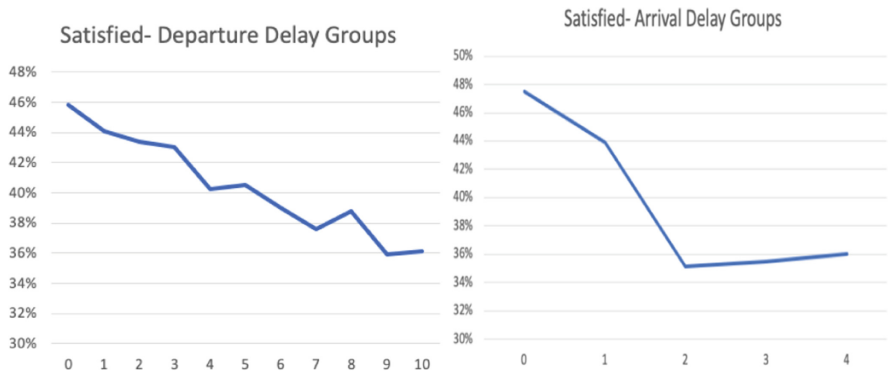

Fig. 9. How Satisfaction Rates Vary with Delay Time [Owner-draw]

Table 12. Subjective Perceptions of Delays among Different Customer Class [Owner-draw]

Customer Class	Mean		
	Time Convenient	Departure Delay	Arrival Delay
Eco	3.19	15.08	15.61
Eco Plus	3.21	15.12	15.80
Business	2.91	14.30	14.49

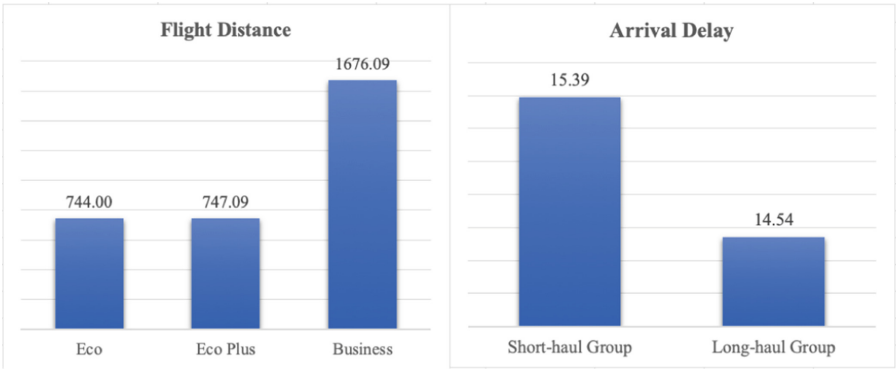

Fig. 10. Relationship of Delay, Flight Distance, and Cabin Type [Owner-draw]

as objective delay time using different cabin types as factors. A post-hoc test revealed that there was no significant difference between Eco and Eco Plus on the rating of this matter, while both were significantly different from business class. Table 12 demonstrates that business class passengers generally experienced shorter delays, but they ranked time convenience lower than other class types.

So why do business class passengers generally experienced shorter delays? We suspected that flight distance, another object data, has a role in it. The comparison test indicates that arrival delays for long-haul and short-haul flights, as well as flight distances for business class and the other classes, differ significantly. As shown in Fig. 10, long-haul flights have shorter average delays, and business travelers have longer average flight distances. Now it is clear that business class passengers have shorter delays because they have more long-distance travel, and long-distance flights suffer relatively less delay.

4.3 Logistics Regression

4.3.1 Purpose

To predict satisfaction and observe the role of each variable in the regression equation, we performed the binary logistic regression using the original data and the downscaled factors respectively, and compared the model effects.

Table 13. Regression with Original Variables - Classification Table [Owner-draw]

Observed		Predicted		
		neutral or dissatisfied	satisfied	Percentage Correct
satisfaction	neutral or dissatisfied	66463	6989	90.5
	satisfied	9295	47133	83.5
Overall Percentage				87.5

4.3.2 Regression with Original Variables

With satisfaction as the dependent variable and all the other original variables (including Total) as covariates, and through the method of Conditional Forward, the binary logistic regression yields the following results. From Table 13, with 21 variables, the equation has an explanatory of 87.5% and relatively balanced predictions for "satisfied" and "neutral or dissatisfied". From Table 14, all coefficients are significant at the 99% confidence interval, which means they all have a strong relationship with the satisfaction of customers. These statistics show that the model has a strong effect.

Among the rating variables on the five-point scale, the evaluation of inflight Wi-Fi service and online boarding have a maximum association with satisfaction without considering the variables excluded from the model, which is consistent with the quality correlation matrix in EDA. Additionally, the coefficients of customer type, class, and type of travel are quite high, indicating that the differences in these features can make a considerable difference in satisfaction, which coincides with the findings in the comparative analysis. However, the number of variables involved in this model is too high, and below we model it with factors to try if we can express the same content with fewer variables.

4.3.3 Regression with Factors Derived by Dimensionality Reduction

With satisfaction as the dependent variable and FAC1 ~ FAC10 as covariates, and through the method of Conditional Forward, the binary logistic regression yields the following results (Table 15 and Table 16). All the coefficients are significant for the model. The variables with the largest coefficient in the first model are in Fac3 and Fac5, which are about secondary service and customer flight information. With only a 2% decrease in expressivity, this new model shrinks by 11 variables, which means this is an efficient and powerful model.

5 Discussion, Speculations, and Managerial Implications

5.1 Gender

The differences in the first three factors (Fundamental Items, Basic Service, and Secondary Service) with males and females may reflect the generally different focus on requirements in a journey between males and females. For Fac2, pregnant women may

Table 14. Regression with Original Variables - Variables in the Equation [Owner-draw]

Variable	B	Sig.	Variable	B	Sig.
Gender	.061	.000	onboard_service	.241	.000
customer_type	2.083	.000	leg_room_service	.189	.000
age	-.008	.000	baggage_handling	.077	.000
type_of_travel	-2.842	.000	checkin_service	.271	.000
customer_class	-.515	.000	inflight_service	.066	.000
inflight_wifi_service	.334	.000	cleanliness	.166	.000
departure_arrival_time_convenient	-.192	.000	departure_delay_in_minutes	.004	.000
ease_of_online_booking	-.211	.000	arrival_delay_in_minutes	-.009	.000
gate_location	-.036	.001	Total	.061	.000
food_and_drink	-.089	.000	Constant	-7.963	.000
online_boarding	.553	.000	—	—	

Table 15. Regression with Factors - Classification Table [Owner-draw]

Observed		Predicted		
		neutral or dissatisfied	satisfied	Percentage Correct
satisfaction	neutral or dissatisfied	64912	8540	88.4
	satisfied	10957	45471	80.6
Overall Percentage				85.0

Table 16. Regression with Factors - Variables in the Equation [Owner-draw]

Variable	B	Sig.	Variable	B	Sig.
Fac1	.982	.000	Fac6	-.413	.000
Fac2	.961	.000	Fac7	-.145	.000
Fac3	1.147	.000	Fac8	.366	.000
Fac4	-.161	.000	Fac9	.494	.000
Fac5	1.391	.000	Fac10	.016	.048
—	—		Constant	-.500	.000

focus more on legroom and baggage handling. And more interaction between people is implied in the inflight service. For Fac1, probably because the male has a larger average body size, they are more concerned about seat comfort, and online boarding can help them choose a seat that suits them better.

5.2 Flight Information & Age

The basic result is that customers selecting the business class and those taking business travel are almost the same group of people, who are generally disloyal. The phenomenon is probably because organizations usually arrange a business class for business travelers. And contrary to the hypothesis, their disloyalty indicates that their organizations' selection of airlines is relatively random. This means airlines have the opportunity to seek partnerships with companies in target areas while developing corporate employees as loyal customers. Additionally, it can be extrapolated that airlines are focusing more on the experience of business class, business trips, and loyal customers. And regarding age, there are two conjectures. Maybe the needs of senior travelers are more easily met and perhaps young people are stricter. And combining the findings on the correlation between age and cabin type, we infer that the reason why passengers with a greater average age are more likely to be satisfied is that older customers are more likely to be business travelers.

5.3 Delay Time

The delay had an impact on passengers' scores for other items. Perhaps this can be attributed to emotional factors. The result of the satisfaction rate with different delay times indicates that travelers are more tolerant of departure delays and more sensitive to arrival delays. This is understandable because departure delays do not necessarily lead to arrival delays. Maybe departures are delayed, but flights arrive on time by reducing the scheduled time en route. Next, about business travelers' stricter scoring in terms of delays, considering that business class travelers are mostly business travelers, it can be assumed that this phenomenon is caused by their tighter schedules compared to personal travelers. Therefore, airlines could provide more care for business class passengers when delays occur. In addition, why do business class passengers generally experience shorter delays? Now it is clear that business class passengers have shorter delays because they have more long-distance travel, and long-distance flights suffer relatively less delay.

6 Conclusion

After a series of data mining, the study mainly yielded the following results.

1) Female passengers pay more attention to legroom service, baggage handling, and inflight service. Differently, male passengers have higher requirements for seating comfort and online boarding.

2) Customers selecting the business class and those taking business travel are almost the same groups of people, who are generally disloyal. In addition, older customers are more likely to be in better cabins.

3) Travelers are more tolerant of departure delays and more sensitive to arrival delays. Business travelers are more sensitive to delays. They generally suffer shorter delays, because they have more long-distance travel, and long-distance flights suffer relatively less delay.

4) The logistic regression model built with the original variables as variables successfully achieved an explanatory power of 87.5% with strong predictive power. All of these factors were significantly associated with satisfaction.

The highlight of this study is the meticulous and in-depth data mining rendering of a web-like approach. For discontinuous variables, firstly, the differences between different categories of passengers at the factor level were compared, and secondly, the specific elements that caused the differences at the factor level were explored. For continuous variables, differences in the explained variables were assessed by grouping them in multiple times. In addition, it was always remembered to probe for correlations between different variables in this study. All these efforts aim to uncover the causes of the different phenomena demonstrated by the airline dataset. Another highlight of the research is the good modeling results. The predictive power of the logistic regression model was excellent. This research hopes to provide possible data mining ideas and help airlines to develop marketing and service strategies in a targeted manner. In the future, the conclusions of the article can be tested in practice with real-life cases. If more comparable data are available at different time points, it would be feasible and meaningful to explore the trends and reasons for the changes in the influence of each variable. Combined with machine learning, such research may play a more powerful and realistic role.

References

1. Research Dive. (2020) *Airline Market: Global Opportunity and Forecast 2020–2027*. Research Dive, www.researchdive.com/8331/airline-market.
2. An, M., & Noh, Y. (2009). *Airline customer satisfaction and loyalty: impact of in-flight service quality*. Service Business, 3(3), 293–307.
3. Nadiri, H., Hussain, K., Haktan Ekiz, E. and Erdoğan, Ş. (2008), *An investigation on the factors influencing passengers' loyalty in the North Cyprus national airline*, The TQM Journal, Vol. 20 No. 3, pp. 265-280.
4. Saha, G.C., and Theingi (2009), *Service quality, satisfaction, and behavioral intentions: A study of low-cost airline carriers in Thailand*, Managing Service Quality: An International Journal, Vol. 19 No. 3, pp. 350-372.
5. Babbar, S. and Koufteros, X. (2008), *The human element in airline service quality: contact personnel and the customer*, International Journal of Operations & Production Management, Vol. 28 No. 9, pp. 804-830.
6. Ali, F., Dey, B.L. and Filieri, R. (2015), *An assessment of service quality and resulting customer satisfaction in Pakistan International Airlines: Findings from foreigners and overseas Pakistani customers*, International Journal of Quality & Reliability Management, Vol. 32 No. 5, pp. 486-502.
7. Tri Noviantoro, Jen-Peng Huang. (2021), *Investigating Airline Passenger Satisfaction: Data Mining Method*. Research in Transportation Business and Management, vol. 43, Elsevier BV, Oct. 2021, p. 100726.
8. Gorzałczany, Marian B., et al. (2021), *Business Intelligence in Airline Passenger Satisfaction Study—A Fuzzy-Genetic Approach with Optimized Interpretability-Accuracy Trade-Off*. Applied Sciences, vol. 11, no. 11, MDPI, May 2021, p. 5098.

9. Sezgen, Eren, et al. (2019), *Voice of Airline Passenger: A Text Mining Approach to Understand Customer Satisfaction.* Journal of Air Transport Management, vol. 77, Elsevier BV, Apr. 2019, pp. 65–74.
10. Tahanisaz, Sahar, and Sajjad Shokuhyar. (2020), *Evaluation of Passenger Satisfaction With Service Quality: A Consecutive Method Applied to the Airline Industry.* Journal of Air Transport Management, vol. 83, Elsevier BV, Mar. 2020, p. 101764.