

EXECUTIVE SUMMARY

The report examines an airline seeking to analyze the current state and issues with its customer satisfaction. The aim is to answer the business question “to what extent does customer satisfaction rating determine the performance of the airline?” and point out recommendations for managerial actions and suggestions to improve the data analysis.

Good customer satisfaction is an important part of any company. Especially airlines where customers have multiple options to choose from when traveling. When an airline is consistently providing positive experience for their passengers, those passengers are more likely to choose that airline for their future travel needs, recommend to their friends and lead to cost savings, as there will be less file complaints or refund demands. Hence, the key outcome variable used in this report is “satisfaction”. In seeking to find the answer to the business question, various other variables were paired with satisfaction such as flight type, connection, business, kids and departure all of which will be a determinant factor in exploring the extent to which customer satisfaction determines the performance of an airline.

After careful analyses, it was revealed that on-time departure positively impact customer satisfaction. However, departure time alone is not a strong predictor of customer satisfaction, indicating that other factors that were not stated in the dataset could also have an impact. It is important to note that while the model's predictors explain a small portion of the variation in customer satisfaction ratings, other factors may have a greater impact. Therefore, it is recommended that the airline continuously gets adequate information from their customers as feedback and address areas that negatively affect customer satisfaction.

DATA PREPARATION

The dataset provided for this analysis contained 29,873 observations and 10 variables; it showed the satisfaction rating of passenger's flight experience of an airline. In preparing the data, outliers were removed, data was restricted. One of the limitations encountered was an extremely high amount of missing observations. 80% of the observation of the outcome variable (satisfaction) were missing. In an attempt to account for the missing values, a thought process that occurred was to make use of the mean or mode to complete the missing values, this however was not plausible as it would have created a bias because a larger percent of our data was missing. Hence, in order to avoid this, we made use of imputation to generate values for the missing data based on existing trend of the original data. This however showed a strong relationship with the original data as shown in figure 1. With the conclusion of the data transformation, a total of six "6" variables and 29,873 observations were selected to be used for the analysis.

VARIABLE DEFINITION

Table 1: Table showing the variable definitions of the dataset

VARIABLE	DEFINITION
Business	This is a categorical variable Indicating if the trip was classified as business or leisure trip. This was transformed to have two responses rather than three "business", "leisure"
Connection	This is a categorical variable indicating if the flight was a connection flight transformed into "connecting flight" and "nonconnectingflight"
Departure	Categorical variable indicating if the flight was on time. Transformed into "ontime", "delayed" "canceled". All flights, which were less than 90 mins or 30 mins, were transformed to Delayed. And disrupted or canceled flights were merged as canceled
Kids	A Categorical variable indicating if the flight ticket was bought together with discounted ticket(s) for kids. This variable was transformed into a dummy variable where "1" are tickets bought at discount price for kids, and "0" are tickets that weren't bought discounted
Satisfaction	A numeric variable "2-10" indicating if the passenger was unhappy with their flight experience or very happy with their flight experience.
Type	A numeric variable indicating if the flight was a city hopper (operated between major cities),

	holiday flight (operated to or from a holiday hotspot) or a long-haul flight (intercontinental flight)
--	--

DESCRIPTIVE STATISTICS

Table 2: Descriptive statistics for continuous variables.

VARIABLE	NO OF OBSERVATION	MEAN	MEDIAN	STANDARD DEVIATION	MINIMUM
Satisfaction	29,873	8.29	8	1.34	2
Kids	29,873	0.204	0	0.403	0

CATEGORICAL VARIABLE		FREQUENCY	PERCENTAGE %
Business	Business	8142	27.18
	leisure	21731	72.86
Connection	Connecting flight	2986	9.94
	Non – connecting flight	26887	90.1
Departure	Canceled	815	2.72
	Delayed	21433	71.6
	On time	7625	25.47
Type	City hopper	6795	22.7
	Holiday flight	12811	42.8
	Intercontinental	10267	34.3

APPENDIX

Appendix 1: The following restrictions were used to create a clean dataset. In the business variable was renamed as “business” in order to create a single subset called “business”.

Outliers were removed in the satisfaction variable.

Under the connection Variable, the space was removed to help R work better with the variable. Transformed into “connecting flight” and “nonconnectingflight

The departure variable was transformed into “on time”, “delayed” “canceled”. All flights, which were less than 90 mins or 30 mins, were transformed to “delayed”. In addition, disrupted or canceled flights were merged as “canceled”

The variable kid was transformed into a dummy variable where “1” are tickets bought at discount price for kids, and “0” are tickets that weren’t bought discounted

DASHBOARD OF SATISFACTION, BUSINESS, DEPARTURE, CONNECTION, TYPE AND KIDS

