Prediction and Data Analysis on Airline Passenger Satisfaction

Intermediary Statistical Modelling for Analytics

Supervised by:

Prof. Dr Jin Fang

Group 1

Students: Aashiya Aryal, Rosy Shrestha, Rupesh Kandel, Sachin Kumar, Yashwant Raut

December 13, 2021

Table of Contents

1 Introduction

People use different modes of transportation every day for various purposes, and as a result, transportation has grown to be one of the world's greatest sectors, affecting the economy. Among the different types of transportation the aviation industry plays an important role in the economy of every country. The revenue of airlines can be increased by aircraft utilization rate, i.e. satisfaction level. This research aims to find out the factors that affect customer satisfaction among the customers of airlines.

1.1 The Research Problem

The virtue of liberalization and globalization has brought competition in all business sectors. The airline industry is one of the main industries that has been affected by globalization. Airlines have not only been used for transporting long distance travelers but also used in daily commutes these days. This has brought fierce competition in the airline's industry and resulted in the rise of Economy Budget Airlines. The range of airlines available with the customer and the limited resources available with the airline businesses have brought challenges of how to maintain and retain the customer. Improving customer satisfaction is a useful strategy, particularly in the airline business, where competition is fierce to help organizations survive and grow.

1.2 Research Question

What are the various factors which affect the level of customer satisfaction?

1.3 The Importance of Research Problem

Kotler Philip (Kotler, 2015) has defined Customer Satisfaction as “a person’s feeling of pleasure or disappointment after comparing a product’s perceived performance or outcome against his/her expectations”. Customer satisfaction is an abstract phenomenon that is affected by elements such as product quality and service quality related to the product and services. One of the most significant things that help the airline industry survive in a competitive market is service quality.

Customer satisfaction is linked to an analysis of the service quality offered in tangible elements and online services. Customers that are dissatisfied or disengaged ultimately result in fewer passengers and lower revenue.

Aviation accounts for more than 5% of our Gross Domestic Product, contributes $1.6 trillion in total economic activity and supports nearly 11 million (Bristol,2016). Commercial aviation is a critical economic engine and helps drive more than 10 million American jobs ((A4A), 2021). Covid-19 had a significant impact on the airline sector. The year 2020 was the worst year in airline history. As fewer people traveled, revenue began to decline, producing a big hole in the economy. As the world begins to recover from the pandemic and more people starting to fly, airlines must work harder than ever to attract passengers and meet their requirements. It is no more about profit generation only but has become a matter of survival to the airlines company. "Airline personnel rose to meet the challenges of a drastically altered travel environment. Maintaining that level of flexibility and recognition of individual passenger needs will be a strategic advantage for airlines that want to set themselves apart in passenger satisfaction as travel volumes start to recover," (Slotnick, 2021).

Research on airline service quality and its impact on customer satisfaction and loyalty is a topical issue that has caught the attention of researchers and practitioners because of its potential to influence airline profitability and competitiveness (Saha G. Theingi, 2009) . When airlines focus on their customer service, it will eventually help them in improving their competitiveness and generate a large revenue. Airlines must understand their customers' wants and expectations regarding the numerous services they offer while travelling. Airlines that can provide attractive tangible factors can have an impact on customer satisfaction and can increase their customers. This study seeks to address the relationship of different variables that airlines offer with the customer satisfaction. The findings can help the 1) airline sector better understand their customers' wants, identify the most critical elements, and increase revenue by retaining customers and offering better services 2) customers can lobby the important characteristics and choose the best airlines that caters the most to their needs

1.4 How does it relate to STAT 4600 class?

Different methods that are learnt in STAT 4600 class has been applied to answer the research question. Charts such as Bar graph, pie chart, polygon etc. are used to represent data.

2 Methods

2.1 Description of Dataset

Customer happiness can be influenced by a variety of factors. In this assessment, 25 factors are considered for each passenger to measure their satisfaction.

2.1.1 Contents of Dataset

- Gender: Gender of the passengers (Female, Male)
- Customer Type: The type of customer (Loyal customer, disloyal customer)
- Age: Actual age of the passengers
- Type of Travel: Purpose of the flight of the passengers (Personal Travel, Business Travel)
- Class: Travel class in the plane of the passengers (Business, Eco, Eco Plus)
- Flight distance: The flight distance of this journey
- Inflight Wi-Fi service: Satisfaction level of the inflight Wi-Fi service (0: Not Applicable;1-5)
- Departure/Arrival time convenient: Satisfaction level of Departure/Arrival time convenient
- Ease of Online booking: Satisfaction level of online booking
- Gate location: Satisfaction level of Gate location
- Food and drink: Satisfaction level of Food and drink
- Online boarding: Satisfaction level of online boarding
- Seat comfort: Satisfaction level of Seat comfort
- Inflight entertainment: Satisfaction level of inflight entertainment
- On-board service: Satisfaction level of On-board service
- Legroom service: Satisfaction level of Legroom service
- Baggage handling: Satisfaction level of baggage handling
- Check-in service: Satisfaction level of Check-in service
- Inflight service: Satisfaction level of inflight service

- Cleanliness: Satisfaction level of Cleanliness
- Departure Delay in Minutes: Minutes delayed when departure
- Arrival Delay in Minutes: Minutes delayed when Arrival
- Satisfaction: Airline satisfaction level (Satisfaction, neutral or dissatisfaction)

2.1.2 Variable Types in Dataset

Variable	Type
Gender	**Categorical**
Customer Type	
Type of Travel	
Class	
Inflight Wi-Fi service	**Grade**
Departure Arrival time convenient	
Ease of Online Booking	
Gate Location	
Food and Drink	
Online Boarding	
Seat Comfort	
Inflight Entertainment	
Onboard Service	
Legroom service	
Baggage Handling	
Check-in service	
Inflight service	
Cleanliness	
Age	**Numerical**
Flight Distance	
Departure delayed in minute	
Arrival delayed in minute	

Table 1 Types of variables in the dataset

2.1.3 Source of Dataset

The dataset is obtained from Kaggle.

https://www.kaggle.com/teejmahal20/airline-passenger-satisfaction

2.2 Analysis and Modeling Method

The analysis was carried out to find answers to the research question. During the analysis, different factors and customer level were taken into consideration. The project has charts like bar graph, pie chart, histogram, box plot etc. to analyze the data properly. Prediction is important in any analysis since it allows us to establish strategies. The dataset is used for prediction using catboost method. It will assist the project in identifying the most effective features in historical trends and making accurate predictions for the future.

2.3. Understanding the Dataset

Element: An element or member of a sample or population is a specific subject or object. In the data set, one of the elements is Customer Type.

Variable: An element or member of a sample or population is a specific subject or object. For example,
Age is a variable that varies from 7 to 85.
Flight Distance is a variable that varies from 31 to 4983 miles.

Observations: An element or member of a sample or population is a specific subject or object. For example:
The customer with id 44722 is a Female with the least age of 7 years travelled 67 miles.
The customer with id 69074 is a Male with the highest age of 85 years travelled 3732 miles.

Data set: An element or member of a sample or population is a specific subject or object. For example:

This dataset, airline customer satisfaction has multiple variables like age, flight distance, type of travel, satisfaction and so on with varying values.

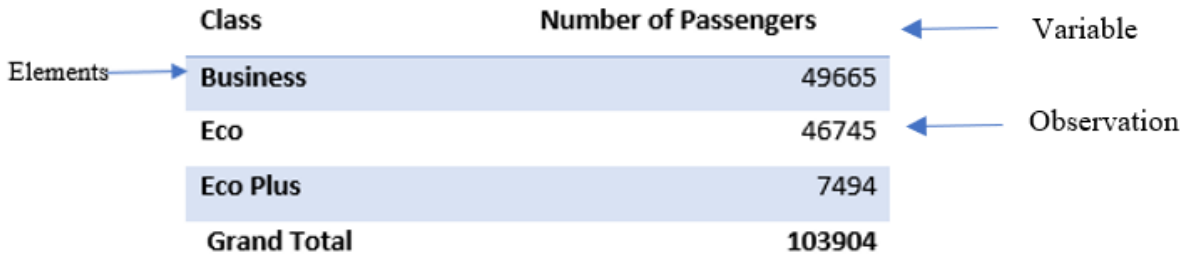


Table 2Understanding element, observation and dataset

Variable and Type of Variable:

- **Variable**: A variable is a characteristic under study that assumes different values for different elements.
- **Quantitative Variable**: A variable that can be measured numerically is called a quantitative variable.
- **Qualitative Variable:** The data collected on a quantitative variable are called quantitative data

Quantitative Variable	Qualitative Variable
Id, Gender, Customer Type	Age, Checkin service, Class, Type of Travel, Flight distance, Inflight Wi-Fi Service, Departure/ arrival, Ease of Online Booking, Gate Location, Food and Drink, Online Boardings, Seat Comfort, Inflight entertainment, Onboard Service, Baggage handling

Discrete and Continuous quantitative variable

- **Discrete Variable -** A variable whose values are countable is called a discrete variable.
- **Continuous Variable -** A variable that can assume any numerical value over a certain interval or intervals is called a continuous variable.

Discrete Variable	Continuous Variable
Id, Gender, Customer Type. Type of Travel, Class, Inflight Wi-Fi service, Departure/ arrival, Ease of Online Booking, Gate Location, Food and Drink, Online Boardings, Seat Comfort, Inflight entertainment, Onboard Service, Baggage handling	Age, Flight distance, Inflight Wi-Fi Service,

Cross Sectional and Time Series Data

- **Cross Sectional Data -** Data collected on different elements at the same point in time or for the same period are called cross-section data. The dataset of flight satisfaction is provided by an airlines over an year so the data set is cross section data set.
- **Time Series Data** - Data collected on the same element for the same variable at different points in time or for different periods of time are called time-series data

Population and Sample

- **Population** - A population consists of all elements – individuals, items, or objects – whose characteristics are being studied. In the dataset, the data in train is the population
- **Sample** - A portion of the population selected for study is referred to as a sample. In the dataset, the data in test is sample

Census

- **Census** - A survey that includes every member of the population is called a **census**.

A census in the dataset is the information about the customers with all the age groups whereas the sample survey is the information about the customers with the age 50 only.

Representative Sample

- Representative Sample - A sample that represents the characteristics of the population as closely as possible is called a representative sample.

Representative Sample is determining which age group of people travels most on Business class by conducting a sample survey or which age group of people are satisfied by their flight by conducting a sample survey.

Sampling with replacement

- In sampling with replacement, each time we select an element from the population, we put it back in the population before we select the next element.

Selecting the customers with the highest age from the highest age group is sampling with replacement.

Sampling without replacement

- Sampling without replacement occurs when the selected element is not replaced in the population.

Selecting the customers with customer id from the highest age group is sampling without replacement.

Random Sample

- A random sample is a sample drawn in such a way that each member of the population has some chance of being selected in the sample.

In this dataset, giving food and drink voucher for 10 passengers randomly is an example of Random Sample.

Non-Random Sample

- In a nonrandom sample, some members of the population may not have any chances of being selected in the sample.

In this dataset, giving 10 % discount for the first 10 bookings is an example of Non-random sample.

Sampling Error and Non-Sampling Error

- Sampling error: The sampling error is the difference between the result obtained from a sample survey and the result that would have been obtained if the whole population had been included in the survey.

For determining average number of Business class tickets booked, If we take only one particular age group as a sample, It might differs when we take in all the age groups.

Non-sampling error

- The errors that occur in the collection, recording, and tabulation of data are called non sampling errors or biases.

A non-sampling error occurs when the age values are null as this is a human error.

Selection Error

- The error that occurs because the sampling frame is not representative of the population is called the selection error or bias.

If the respondents are chosen based on specific class or age group, it will be a selection error.

Non-Response Error

- The error that occurs because many of the people included in the sample do not respond to a survey is called the nonresponse error or bias.

Few passengers did not respond to the ease of online booking and gate location which means that they are not interested in responding to these questions. This is a nonresponse error in data.

Response Error

- The response error or bias occurs when people included in the survey do not provide correct answers.

When passengers give incorrect info on their loyalty this leads to a response error.

Voluntary Response Error

- Voluntary response error or bias occurs when a survey is not conducted on a randomly selected sample but on a questionnaire published in a magazine or newspaper and people are invited to respond to that questionnaire.

If we determine average overall airline satisfaction based on the passenger satisfaction levels who is over 50 years age will lead to response error.

Simple random sampling technique

- Simple Random Sampling: Each sample of the same size has the same probability of being selected.

We divide our dataset into classes based on the age and as there are ages varying from 7 to 85 we select one customer from each age group randomly

Systematic random sampling technique

We divided the data set into 300 samples with each having around 340 rows, after that we select the first row of every sample.

Stratified random sampling technique

We divide the dataset into subpopulations using the age, then one sample is selected from each of the strata. All these collections of samples from strata give the stratified random sample.

Cluster sampling technique

As our data set does not have any regions within, we divide the dataset into clusters based on the type of class and a random sample is selected from each of these clusters.

Observational study and Designed experiment

- Observational Study – In an observational study the assignment of elements to different treatments is voluntary, and the experimenter simply observes the results of the study.
- Designed Experiment - When the experimenter controls the (random) assignment of elements to different treatment groups, the study is said to be a designed experiment.

Our dataset is an observational study as there is no control over factors.

2.4 Organizing and Graphing Data

There are 5 qualitative variables and 18 quantitative data in the data set.

Frequency Distribution - A frequency distribution of a qualitative variable lists all categories and the number of elements that belong to each of the categories.

Relative Frequency – The ratio of Frequency of the category to the sum of all frequencies is called relative frequency.

a. Frequency Distribution of Gender

Gender	Frequency	Relative Frequency	Percentage
Female	52727	0.507	50.745
Male	51177	0.492	49.254
Total	103904	1	100

Table 3 Frequency Distribution of Gender

b. Frequency Distribution of Customer Type

Customer Type	Frequency	Relative Frequency	Percentage
Disloyal Customer	18981	0.183	18.268
Loyal Customer	84923	0.817	81.732
Total	103904	1	100

Table 4 Frequency Distribution of Customer Type

2.4.1 Bar Graph and Pareto Chart for the qualitative variables.

- Bar Graph: A graph made of bars whose heights represent the frequencies of respective categories is called a bar graph

- Pareto Chart : A Pareto chart is a bar graph with bars arranged by their heights in descending order.

a. Bar Graph of Gender

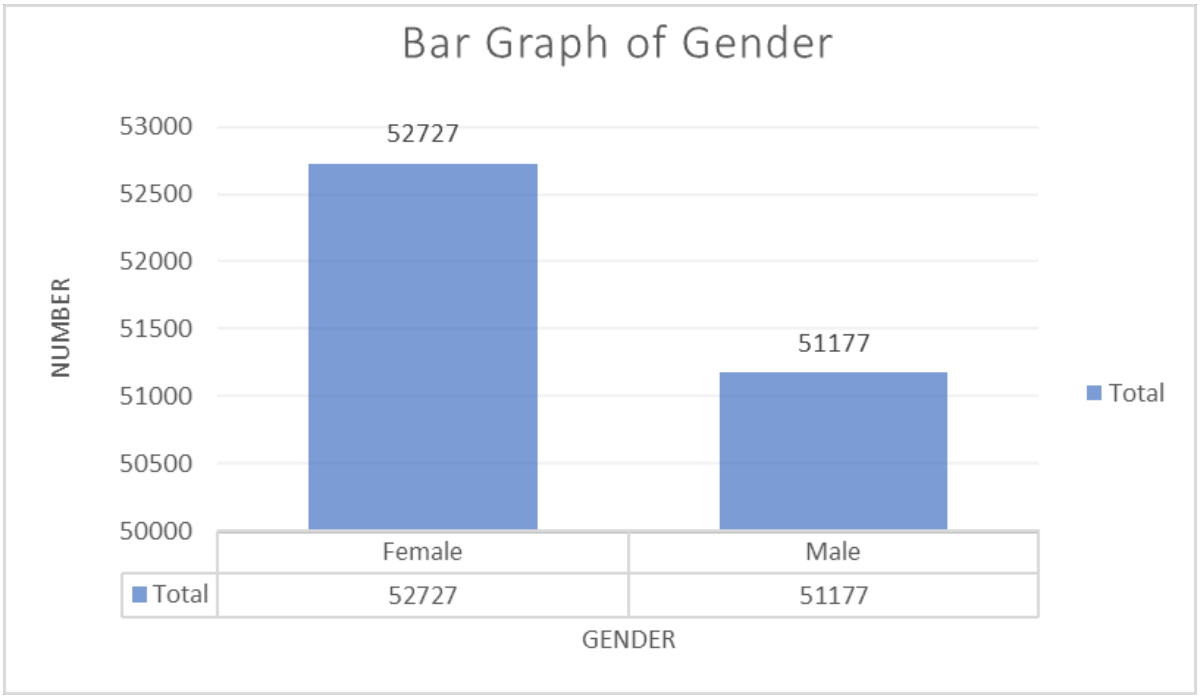


Figure 1 Bar graph representing gender

Analysis – The number of Female passengers is 52,727 and the number of male passengers is 511,177. There are more female passengers travelling by airplane than male.

b. Bar Graph and Pareto Chart of Customer Type

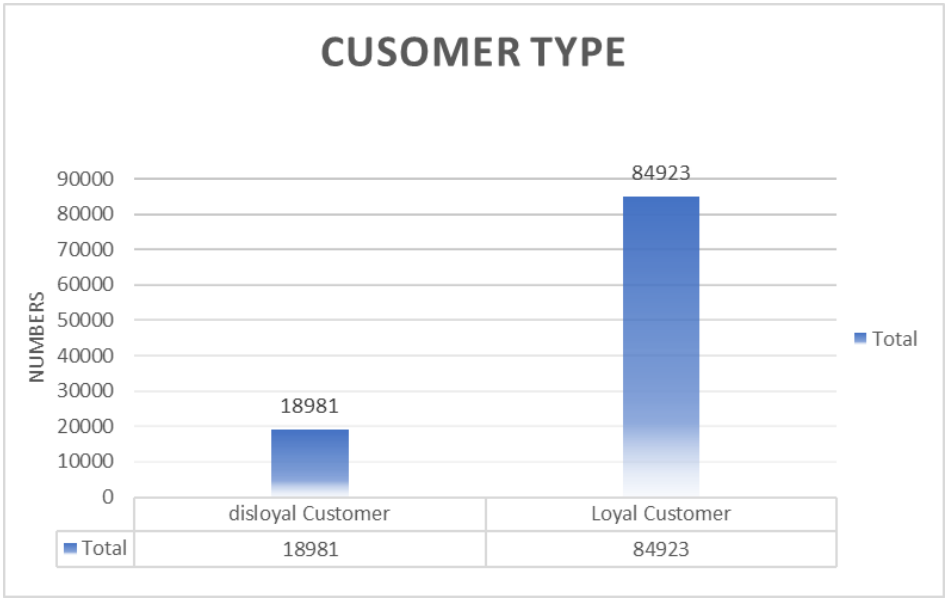


Figure 2 Bar graph of Customer Type

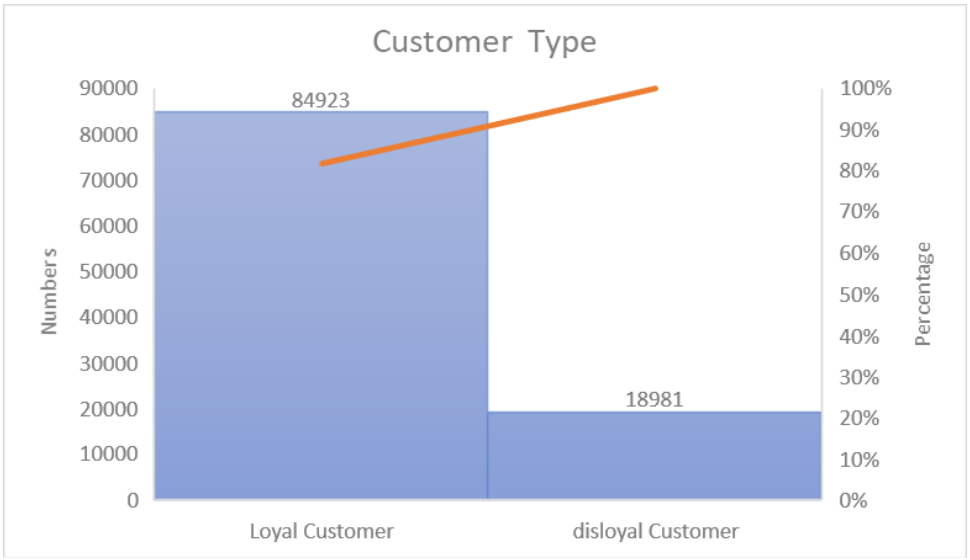


Figure 3 Pareto Chart of Customer Type

Analysis – The number of passengers who are loyal customer is 84,923 and disloyal customer is 18981. We can clearly see that there is a greater number of loyal customer than disloyal customer.

2.4.2 Frequency distribution table for the quantitative variables

Step 1: Find the width of class using Less than Method

To find the width of class using less than method,

Minimum age: 7

Maximum age: 85

The minimum value in the data set is 7, and the maximum value is

85. Suppose we decide to group these data using five classes of equal width.

Then,

Approximate class width = = 15.6

We round this number to a more convenient number 16 and take 16 as the width of each class.

Age	Frequency	Relative Frequency	Percentage
5 to less than 21	11333	0.11	10.91
21 to less than 37	32447	0.31	31.23
37 to less than 53	37546	0.36	36.14
53 to less than 69	20559	0.20	19.79
69 to less than 85	2019	0.02	1.94
	103904		

2.4.3 Histogram

- Histogram : A histogram is a graph in which classes are marked on the horizontal axis and the frequencies, relative frequencies, or percentages are marked on the vertical axis.

a. Histogram of Inflight Wi-Fi Service

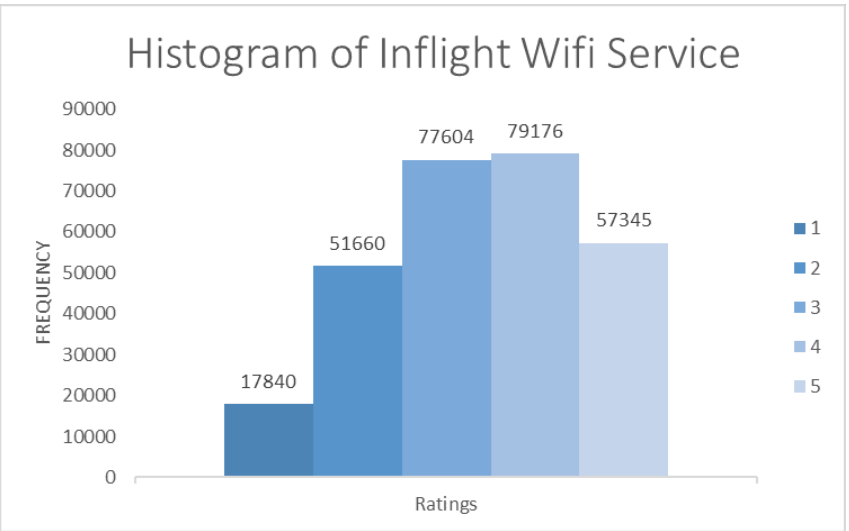


Figure 4 Histogram of Inflight Wi-Fi Service

Analysis – This histogram is symmetric and is almost identical on both sides of its central point.

b. Bar Graph of Online Booking

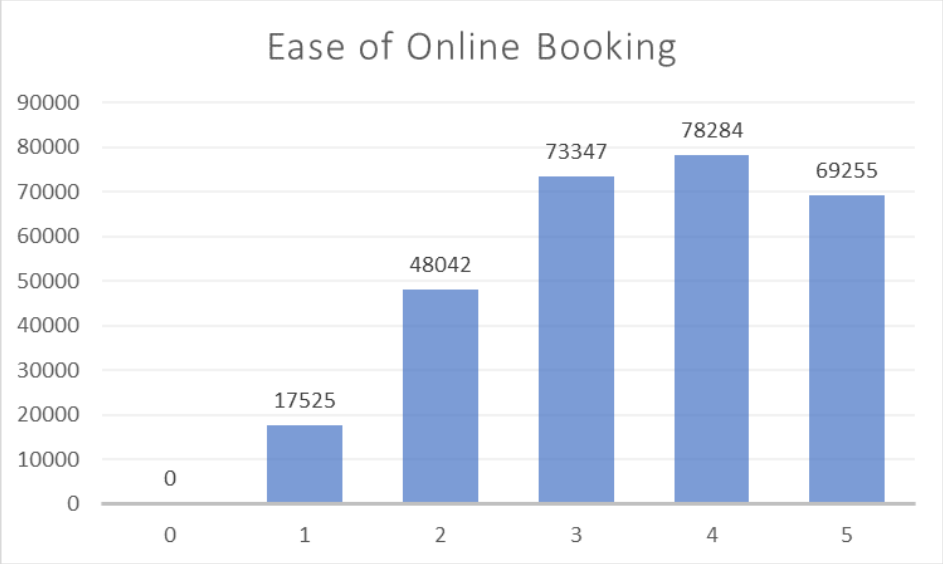


Table 5 Bar graph of Ease of Online Booking

Analysis – The graph shows that maximum customer has rated 4 for online booking. Online Booking service has high rating where maximum has rated it as 3,4 and 5.

2.3.4 Polygon of Gate Location

Polygon - A graph formed by joining the midpoints of the tops of successive bars in a histogram with straight lines is called a polygon.

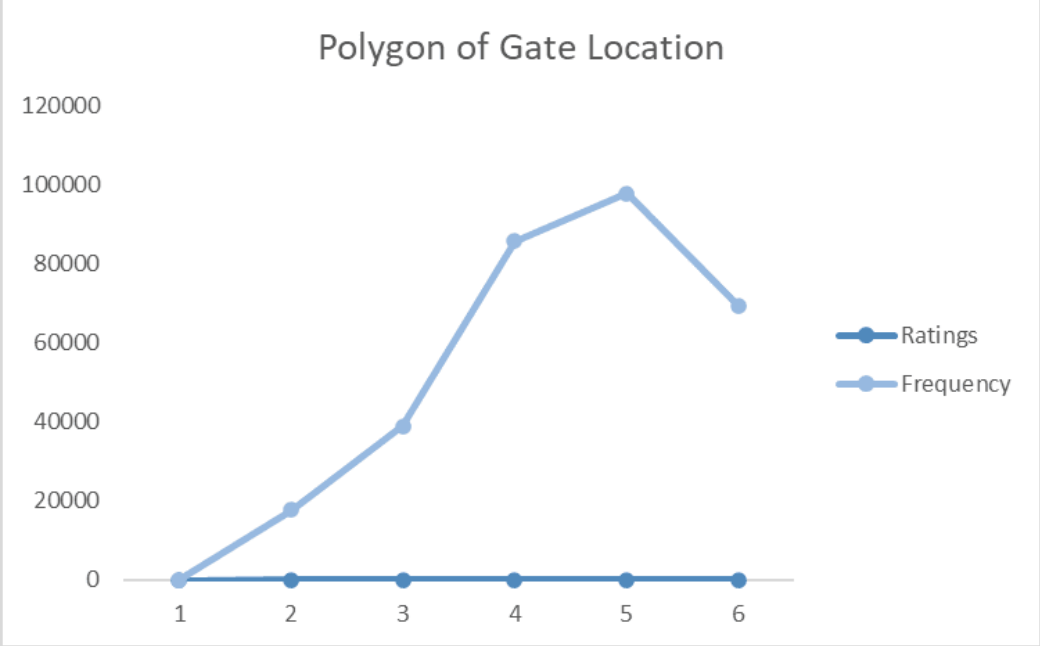
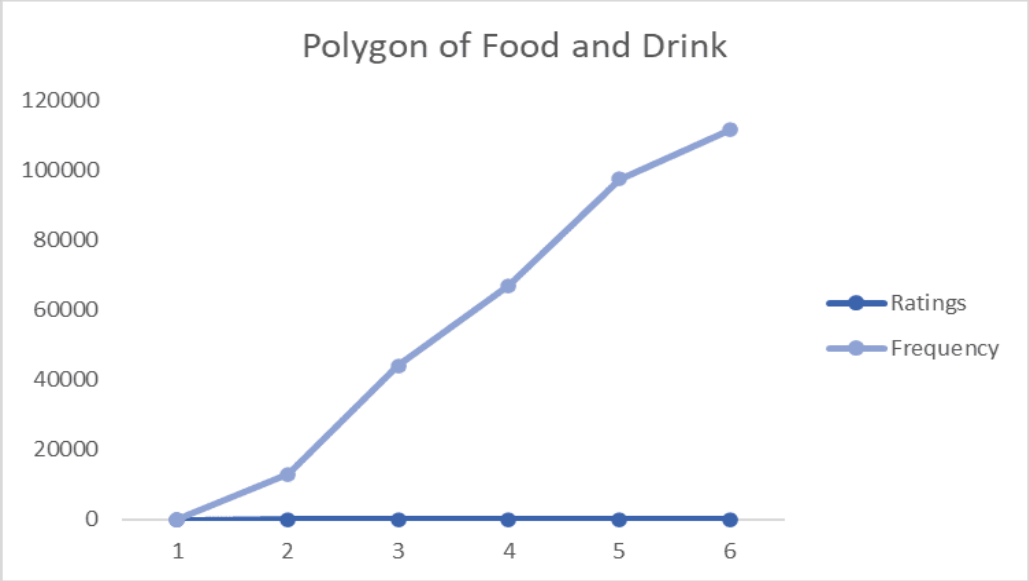


Table 6 Polygon of Gate Location

Analysis – The graphs shows that the midpoint of rating 5 has the highest height.

e. Polygon of Food and Drink



Analysis : The graphs shows that food and drink has more higher ratings than lower.

2.3.5 Stem-and-leaf display

- Stem and Leaf Display - In a *stem-and-leaf display* of quantitative data, each value is divided into two portions—a stem and a leaf. The leaves for each stem are shown separately in a display.

Finding the Stem and Leaf Display of 10 randomly selected Age

13 5 26 25 61 26 47 52 41 20

Step 1 : Split each age into two parts

The first part contains the first digit of a score, which is called the *stem*.

The second part contains the second digit of a score, which is called the *leaf*.

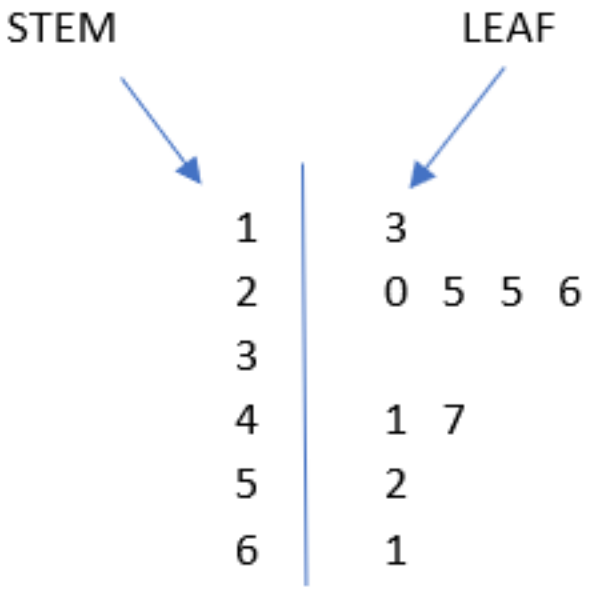


Fig: Stem and Leaf Display

Analysis – It is clearly observed that maximum number of customers in this sample is from age 20-29.

2.4.1 Numerical Descriptive Measures

	count	mean	std	min	25%	50%	75%	max
Age	103594.0	39.380466	15.113125	7.0	27.0	40.0	51.0	85.0
Flight Distance	103594.0	1189.325202	997.297235	31.0	414.0	842.0	1743.0	4983.0
Inflight wifi service	103594.0	2.729753	1.327866	0.0	2.0	3.0	4.0	5.0
Departure/Arrival time convenient	103594.0	3.060081	1.525233	0.0	2.0	3.0	4.0	5.0
Ease of Online booking	103594.0	2.756984	1.398934	0.0	2.0	3.0	4.0	5.0
Gate location	103594.0	2.977026	1.277723	0.0	2.0	3.0	4.0	5.0
Food and drink	103594.0	3.202126	1.329401	0.0	2.0	3.0	4.0	5.0
Online boarding	103594.0	3.250497	1.349433	0.0	2.0	3.0	4.0	5.0
Seat comfort	103594.0	3.439765	1.318896	0.0	2.0	4.0	5.0	5.0
Inflight entertainment	103594.0	3.358341	1.333030	0.0	2.0	4.0	4.0	5.0
On-board service	103594.0	3.382609	1.288284	0.0	2.0	4.0	4.0	5.0
Leg room service	103594.0	3.351401	1.315409	0.0	2.0	4.0	4.0	5.0
Baggage handling	103594.0	3.631687	1.181051	1.0	3.0	4.0	5.0	5.0
Checkin service	103594.0	3.304323	1.265396	0.0	3.0	3.0	4.0	5.0
Inflight service	103594.0	3.640761	1.175603	0.0	3.0	4.0	5.0	5.0
Cleanliness	103594.0	3.286397	1.312194	0.0	2.0	3.0	4.0	5.0
Departure Delay in Minutes	103594.0	14.747939	38.116737	0.0	0.0	0.0	12.0	1592.0
Arrival Delay in Minutes	103594.0	15.178678	38.698682	0.0	0.0	0.0	13.0	1584.0

The given chart shows the mean, median.

- Mean is the average of the numbers: a calculated "central" value of a set of numbers. For example, the average age of the respondent is 39 years.

- Median is the "middle" value of a sorted list of numbers. For example, the median value of age is 40 years. Mode is simply the number which appears **most often**. For example, mode for age is 39.

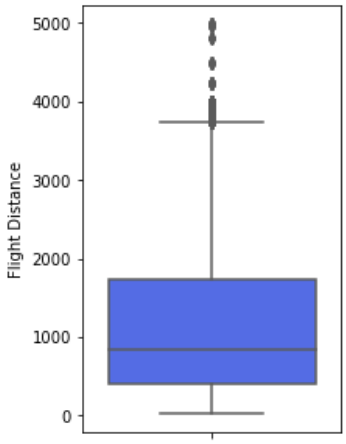
- Mode can be of three types: unimodal, bimodal, and multimodal. Unimodal have single mode, bimodal has two modes and multimodal has more than two modes. The dataset has unimodal variables.

- The chart also shows the standard deviation, first quartile, second quartile, third quartile. Standard deviation is a measure of how spread-out numbers are. The deviation of age in the dataset is 15.11. Similarly, the researcher can find the range and interquartile range. For example, the range for age is 85 - 7 = 78. Similarly, interquartile range for age is Q3 - Q1 i.e. 51 - 27 = 24.

- Chebyshev's Theorem estimates the minimum proportion of observations that fall within a specified number of standard deviations from the mean. For example, the average age was found to be 39 years with a standard deviation of 15. Using Chebyshev's theorem, find the minimum percentage of age which falls between 20 to 50 years.

  = 50-20 = 30/15 = 2

  Therefore, at least 75% of passengers are in the range of 20 to 50



The box plot shows that there are some outliers in flight distance variable.

- **Experiment:** A specific subject or object included in a sample or population about which the information is collected is an element.
- **Outcome:** The result of the performance of an experiment.
- **Sample space:** The collection of all sample points or outcomes of an experiment.

Experiment	Outcomes	Sample Spaces
Gender	Male (M), Female(F)	{M,F}
Class	Business(B), Eco(E), EcoPlus(Ep)	{B,E,Ep}

- **Event:** A collection of one or more outcomes of an experiment.
- **Simple event:** An event that contains one and only one outcome of an experiment. It is also called an elementary event.

Selecting two customers who are loyal (L) and disloyal(D) and observing whether the customer selected each time is a loyal or disloyal. Each of the final four outcomes (*LL*, *LD*, *DL*, and *DD*) for this experiment is a simple event. These four events can be denoted by E_1, E_2, E_3, and E_4, respectively.

Thus,

$E_1 = \{LL\}$,
$E_2 = \{LD\}$,
$E_3 = \{DL\}$, and
$E_4 = \{DD\}$.

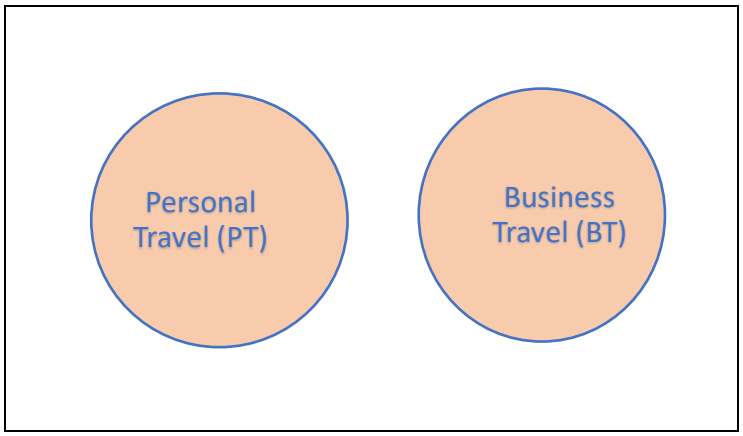
- Compound Event: A **compound event** is a collection of more than one outcome for an experiment. It is denoted by *A, B, C, D*, … or by *A1*, *A2*, *A3*, … or by *B1*, *B2*, *B3*…

For example, selecting two customers and observing whether the customer selected each time is loyal or disloyal. Let *A* be the event that at most one loyal customer is selected.
A = at most one man is selected = {*LD*, *DL*, *DD*}

- **Mutually exclusive event**

Two or more events that do not contain any common outcome and, hence, cannot occur together. The dataset has type of travel i.e. personal travel and business travel. One cannot travel for both. Either he/she travels for personal reasons or for business purpose.

Let the probability of personal travel be PT and Business Travel be BT
Hence, $P(PT \cap BT)$



- **Independent Event**

Those events which do not affect or is not dependent on the occurrence of another event is known as independent events. For example, if a customer travels for a business purpose for the first trip and then again if the same customer travels for a personal reason, the event is known to be independent of each other.

If Bp be the business purpose and Pp be the personal purpose then,
$P(Bp \mid Pp) = P(Bp)$ or
$P(Pp \mid Bp) = P(Pp)$

- **Dependent Event**

Two events for which the occurrence of one change the probability of the occurrence of the other.

From the dataset, the satisfaction is dependent upon the cleanliness variable. The probability of cleanliness is calculated.

Let P(S) and P(C) be the probability of satisfaction and probability of cleanliness respectively.

$P(S) = 45025/103904 = 0.433$

$P(S|C) = 28363/45025 = 0.629$

$P(S|C) \neq P(S)$

Since the probability of cleanliness is not equal to the probability of satisfaction with cleanliness, it can be concluded that satisfaction is dependent of cleanliness.

- **Complementary Event**

Complementary event is the event that includes all the outcomes for an experiment that are not in *A.*

- **Union of Events**

Given by the outcomes that belong either to one or to both events.

- **Intersection of Events**

The intersection of events is given by the outcomes that are common to two (or more) events.

Let A and B be two events which defines cleanliness and food and drinks variable respectively. The intersection of A and B represents the collection of all outcomes that are common to both A and B and denoted by $A \cap B$.

Variables	Total
P (cleanliness)	7660
P (Food & Drinks)	7284
P (Cleanliness ∩ Food&Drinks)	15029

Sampling distributions

The mean distance travelled by 103,904 passengers who travelled in Japan Airlines is 1189 miles and the standard deviation is 964. Let $\bar{x}$ be the mean distance travelled by certain group of passengers from this company. Find the mean and standard deviation of mean x for a sample size of:

- 4,966 (some portion of Business Class).
- 4,674 (some portion of Economy Class).
- 4,494 (some portion of Economy Plus).

= Solution:

Given:

Population mean=1189 miles

Population standard deviation=964

The mean of sampling distribution of mean x, $\mu_{\bar{X}}$, is

$\mu_{\bar{X}} = \mu = 1189$.

a. In this case,

n=4,966,N=103,904. Because n/N=0.047 that is less than 0.05, $\sigma_{\bar{x}}$ is obtained using the formula = $\frac{\sigma}{\sqrt{n}}$and that is:

$\frac{3.7}{\sqrt{4966}} =$ =0.052.

b. In this case, n=4674, N=103,904. Because n/N=0.044 that is less than 0.05, $\sigma_{\bar{x}}$ is obtained using the formula =$\frac{\sigma}{\sqrt{n}}$and that is:

$\frac{3.7}{\sqrt{4674}}$ =0.054

c. In this case, n=4494, N=103,904. Because n/N=0.043 that is less than 0.05, $\sigma_{\bar{x}}$ is obtained using the formula $\frac{\sigma}{\sqrt{n}}$that is

$\frac{3.7}{\sqrt{4494}}$ =0.052

Question no. 2:

Assume that the departure delay in minutes of the Japan Airlines is approximately normally distributed with a mean of 14.80 minutes and a standard deviation of 38.23 minutes. Find the probability that the mean the departure delay in minutes, x, of random sample of 20 business class passengers of Japan Airlines of those passengers will be between 14.10 and 14.50 minutes.

Solution:

Although the sample size is small (n < 30), the shape of the sampling distribution of x is approximately normal because the population is approximately normally distributed.

- The mean and standard deviation of *x* are, respectively,
- Sample mean=Population mean=14.80 minutes.
- Sample standard deviation=population standard deviation/square root of n=38.23/square root of 20=8.54 minutes.

- We are to compute the probability that the value of *x* calculated for one randomly drawn sample of 20 packages is between 14.10 and 14.50 delay in minutes; that is
- P (14.10<mean x<14.50)

$$z = \frac{\bar{p} - p}{\sigma_{\bar{p}}}$$

- The z values for mean x=14.10 and 14.50 are computed below:
- For mean 14.10, z= (14.10-14.80)/8.54= -0.0820.
- For mean 14.50, z= (14.50-14.80)/8.54= -0.0351

The probability that means x is between 14.10 and 14.50 is given by the area under standard normal curve between z=0.0820 and z=0.0351 which is obtained by subtracting the area to the left of the z=-0.0820 from the area to the left of z=0.0351. Thus, the required probability is:

P (14.10<$\bar{x}$<14.50) =P(-0820<z<-0.0351) =0.486-0.467=0.019

Therefore, the probability that the mean departure of sample of 20 business class passengers will be between 14.10 and 14.50 is 0.019.

Question no. 3:

According to survey of passengers of Japan Airlines, 43% of the passengers were satisfied with the services provided by this airline.

Assume that this result is true for the current population of this airlines. Let $\hat{p}$ be the proportion of Japan Airlines passengers in a random sample of 5000 who are satisfied with the services provided by the airlines. Find the mean and standard deviation of $\hat{p}$ and describe the shape of its sampling distribution.

Solution:

Let p be the passengers of Japan Airlines that were satisfied with the services provided by this airline.

p=0.43, q=0.57, n=5000.

The mean of the sampling distribution of $\hat{p}$ is

$\mu_{\hat{p}} = p$ =0.43.

The standard deviation of $\hat{p}$ is =

$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.43 * 0.57}{5000}}$ =0.007

np=5000*0.43= 2150

nq=5000*0.57= 2850

Because *np* and *nq* are both greater than 5, we can apply the central limit theorem to make an inference about the shape of the sampling distribution of $\hat{p}$.

Thus, the sampling distribution of $\hat{p}$ is approximately normal with a mean of 0.43 and a standard deviation of 0.007.

Question no. 4:

In research of Japan Airlines about 20% of the passengers said that they were fully satisfied with online boarding facility. Suppose that this result is true for the current population of all airlines. Let $\hat{p}$ be the proportion in a random sample of 1500 passengers who will hold the said opinion. Find the probability that 76.5% to 78% of passenger in this sample will hold this opinion

Solution:

n=1500, p=0.2, q=0.8

The mean of the sample proportion $\hat{p}$ is $\mu_{\hat{p}} = p$ =0.2

The standard deviation of $\hat{p}$ is $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.2 * 0.8}{1500}}$=0.0103.

- *np* = 1500 (0.20) = 1200, *nq* = 1500 (0.80) = 300.

Because *np* and *nq* are both greater than 5, we can infer from the central limit theorem that the sampling distribution of $\hat{p}$ is approximately normal.

The probability that $\hat{p}$ is between .765 and .78 is given by the area under the normal curve for $\hat{p}$ between $\hat{p}$ = 0.765 and $\hat{p}$ = 0.78 is calculated as below:

The first step in finding the area under the normal distribution curve between $\hat{p}$ = 0.765 and $\hat{p}$ = 0.78 is to convert these two values to their respective *z* values. The *z* value for $\hat{p}$ is computed using the following formula:

$$z = \frac{\bar{p} - p}{\sigma_{\bar{p}}}$$

$$\text{For } \hat{p} = .765: \quad z = \frac{.765 - .75}{.01157275} = 1.30$$

$$\text{For } \hat{p} = .78: \quad z = \frac{.78 - .75}{.01157275} = 2.59$$

P (.765<$\hat{p}$<.78) =.9952-.9032=.0920.

Thus, the probability that 76.5% to 78% of passenger in this sample fully satisfied with this online boarding facility is 9.20%.

Estimations of the mean and proportion

Question no. 1:

A new airlines company has just started its business. Before the company operate its business, it wants to know the average delay in minutes of other airlines in the market. The research

department at the company took a random of 25 airlines company and collected information on average delay in flights. This information produced a mean delay in minutes of 21.90 minutes for this sample. It is known that the standard deviation of all such average in delay is 46.09 minutes and the population distribution of such delay in minutes is approximately normal.

- What is the point estimate of mean delay in minutes of all other airlines?
- Construct a 90% confidence interval for the mean delay in minutes of all other airlines.

Solution:

Here, σ is known and, although $n < 30$, the population is approximately normally distributed.

Hence, we can use the normal distribution.

- From the given information, n = 25, $\bar{x}$ = 21.90, and $\sigma = 46.09$.
- The standard deviation of $\bar{x}$ is

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{46.09}{\sqrt{25}} = 9.218$$

a) The point estimate of the mean price of all such airlines is 21.90. Point estimate of $\mu = \bar{x} = 21.90$

b) The confidence level is 90% or 0.9.

- First, we find the z value for a 90% confidence level.
- In Table IV of Appendix B, look for the areas 0.05 and 0.95, and find the corresponding values of z.

These values are approximately z = -1.65 and z = 1.65.

c) We substitute all the values in the confidence interval formula for μ. The 90% confidence interval for μ is

$$\bar{x} \pm z\sigma_{\bar{x}} = 21.90 \pm 1.65 * 9.21 = \textbf{6.70 to 37.09}$$

Thus, we are 90% confident that the mean delay of such other airlines is between 6.70 and 37.09

Question no. 3:

Policy Interactive of Eugene, Oregon, conducted a study in April 2014 for a Japanese Airlines that included a sample of "Loyal Customer-Personal"-Type traveler 2500 passengers

- 33% of the passengers included in this study said that the food and drinks of the airlines is very good (5 rating).

a. What is the point estimate of corresponding population proportion?
b. Find, with a 99% confidence level, the percentage of all Loyal Customer-Personal traveler who will say that food and drinks of the airlines is very good (5 rating). What is the margin of error of this estimate?

Solution:

Let p be the proportion of all Japanese Airlines of "Loyal Customer-Personal" type traveler who will say that food and drinks of the airlines is very good (5 rating) and p-cap be the corresponding sample proportion. From the given information:

n=2500, $\hat{p}$=0.33 and $\hat{q}$=0.67.

First, we calculate the value of standard deviation of the sample proportion as follows.

$s\hat{p} = \sqrt{\hat{p}\,\hat{q}}/n = 0.009404$.

Note that $n\hat{p}$ and $n\hat{q}$ are both greater than 5. Consequently, the sampling distribution of p cap is approximately normal, and we will use the normal distribution to make a confidence interval about p.

a. The point estimate of the proportion of all Japanese Airlines of "Loyal Customer-Personal" type traveler who will say that food and drinks of the airlines is very good (5 rating) is equal to 0.33; that is,

   Point estimate of p= $\hat{p}$ =0.33.

b. The confidence level is 99% or .99. To find z for a 99% confidence level, first we find the area in each of the two tails of the normal distribution curve which is (1-.99)/2=0.0050. Then, we look for 0.0050+.99=.9950 areas in the normal distribution table to find the two values of z. These two z values are (approximately) -2.58 and 2.58. Thus, we will use z=2.58 in the confidence interval formula. Substituting all the values in the confidence interval formula for p, we obtain:

$\hat{p}$ + z $S_{\hat{p}}$ =0.33+-2.58(0.009404) =0.33+-0.025=0.305 to 0.355=30.5% to 35.5%

Thus, we can state with 99% confidence that 30.5% to 35.5% of Japanese Airlines of "Loyal Customer-Personal" type traveler who will say that food and drinks of the airlines is very good (5 rating).

The margin of error associated with this estimate of p is 0.025 or 2.5%.

2.5 Analysis of dataset

2.5.1 distribution of age

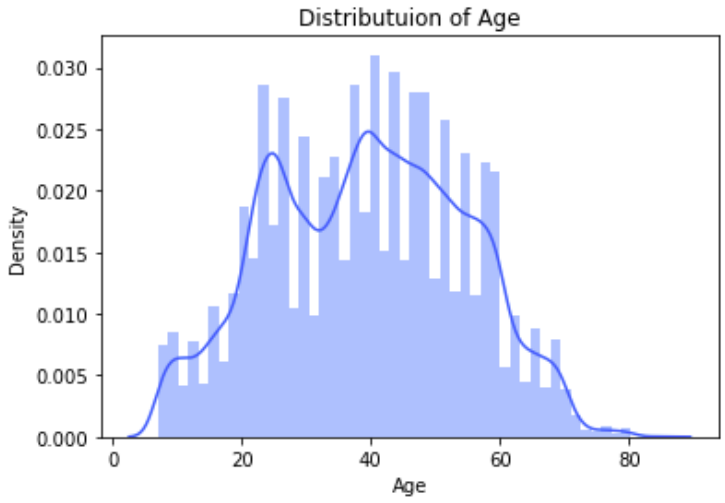


Figure 5 Analysis of age distribution

The above graph shows the distribution of the passenger by age. The densely populated number of passengers was in the age group from 20 to 60.

2.5.1 Analysis of class type

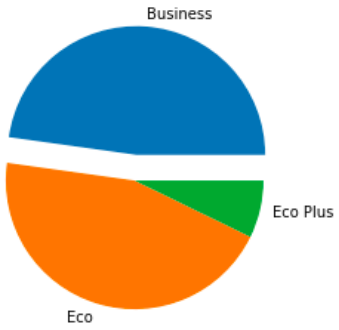


Figure 6 Analysis of class type

The majority of passengers took "Eco" and "Eco Plus" aircraft types. Airlines can generate more revenue if they pay more attention to these passengers.

2.5.2 Analysis of types of customers

Figure 7 Analysis of types of customer

Most of the passengers were loyal which means they have used the airlines more than once. About 81% were Loyal Customer and 18% were disloyal customers.

2.5.3 Analysis of satisfaction

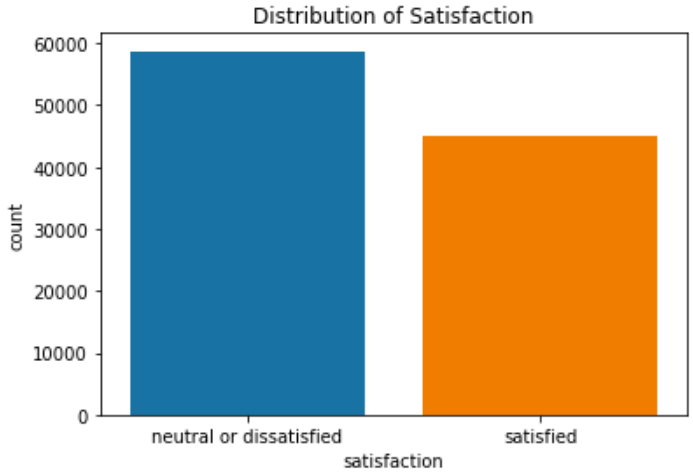


Figure 8 Analysis of Passenger satisfaction

There were more unsatisfied passengers than satisfied passengers, as shown in the bar graph. To be more specific, the number of unsatisfied customers was about 20000 higher than the number of passengers who had a positive experience with the airline.

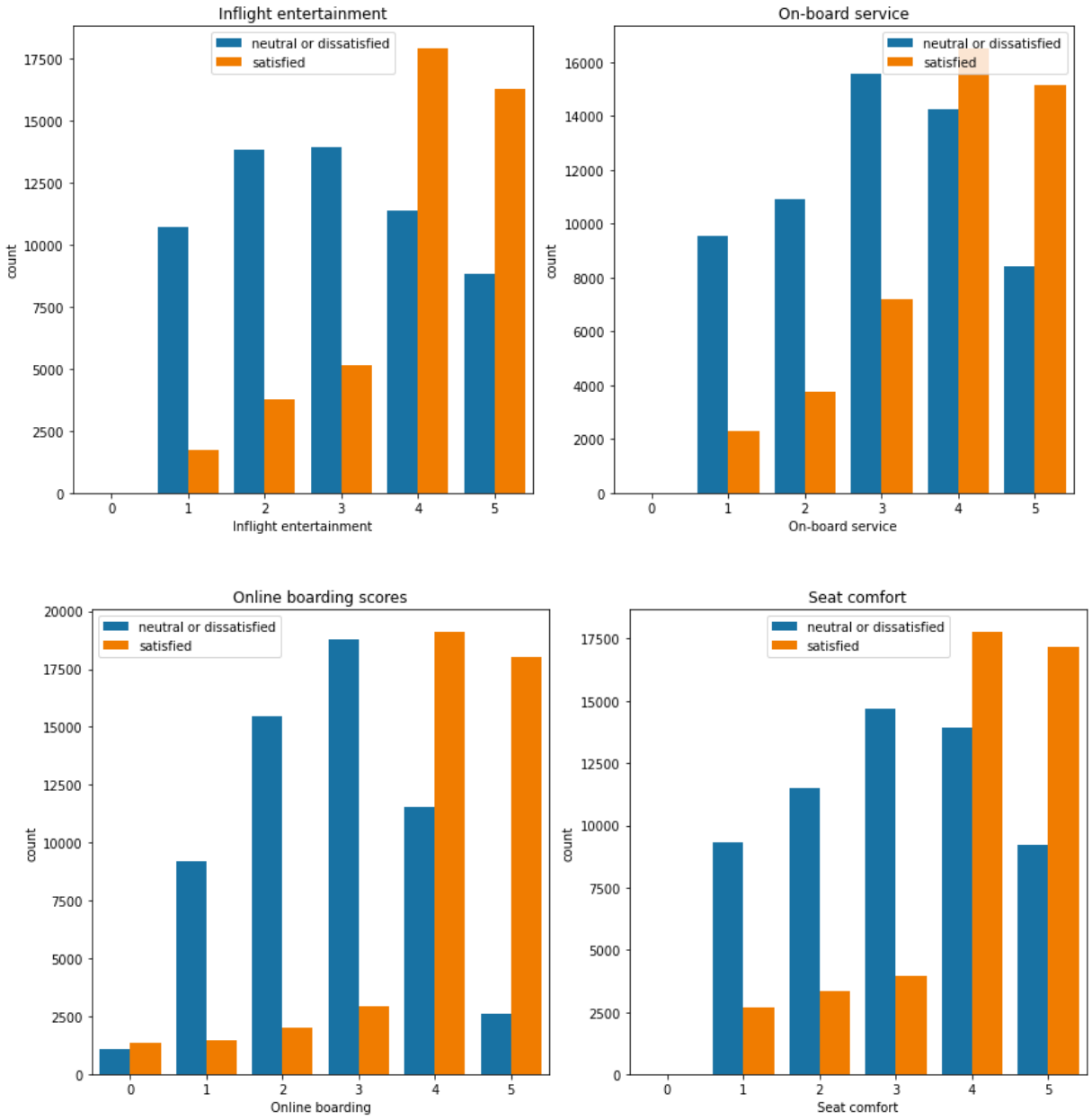


Figure 9 Bivariate analysis of inflight entertainment, onboard service, online boarding, seat comfort

The graphs above show the passengers' satisfaction levels in relation to the many criteria considered to promote customer satisfaction. Airlines should take in consideration of inflight entertainment because many passengers were either neutral or disappointed with the services. Similarly, the majority of passengers who were disappointed with on-board service held a neutral viewpoint. The passengers rated online boarding experience indifferent or dissatisfactory, however some passengers gave it a positive rating. Almost 9,000 satisfied passengers chose a

score of less than four, but none chose zero. Even the unsatisfied passengers were content to give the seat comfort a higher rating.

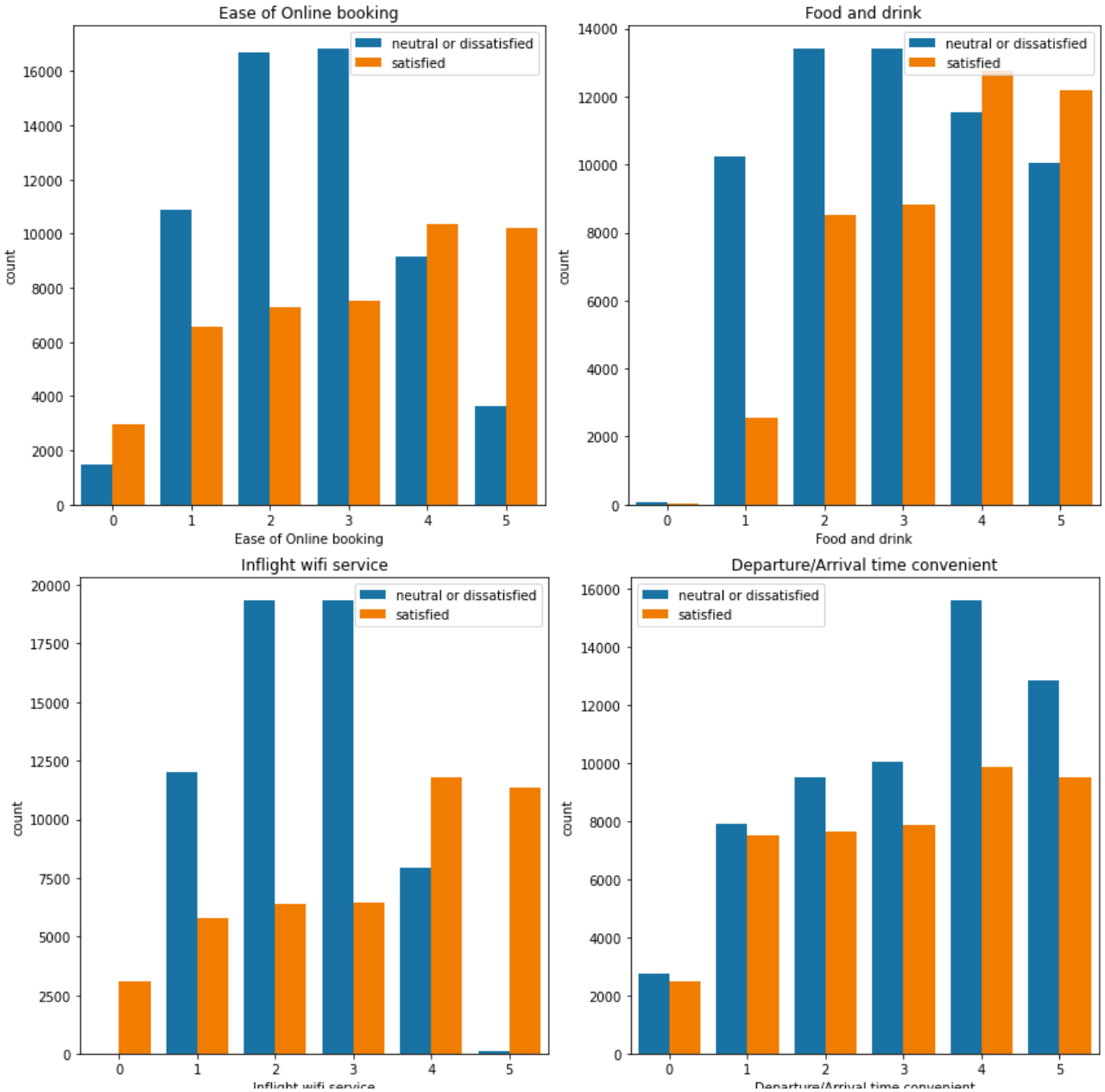


Figure 10 Bivariate analysis of online booking, food and drink, inflight wifi service and time convenient

Passengers were dissatisfied with the online booking facility. Passengers were mostly dissatisfied with the food and drink facility and also with the inflight WI-FI service. However, passengers rated good for the time convenient.

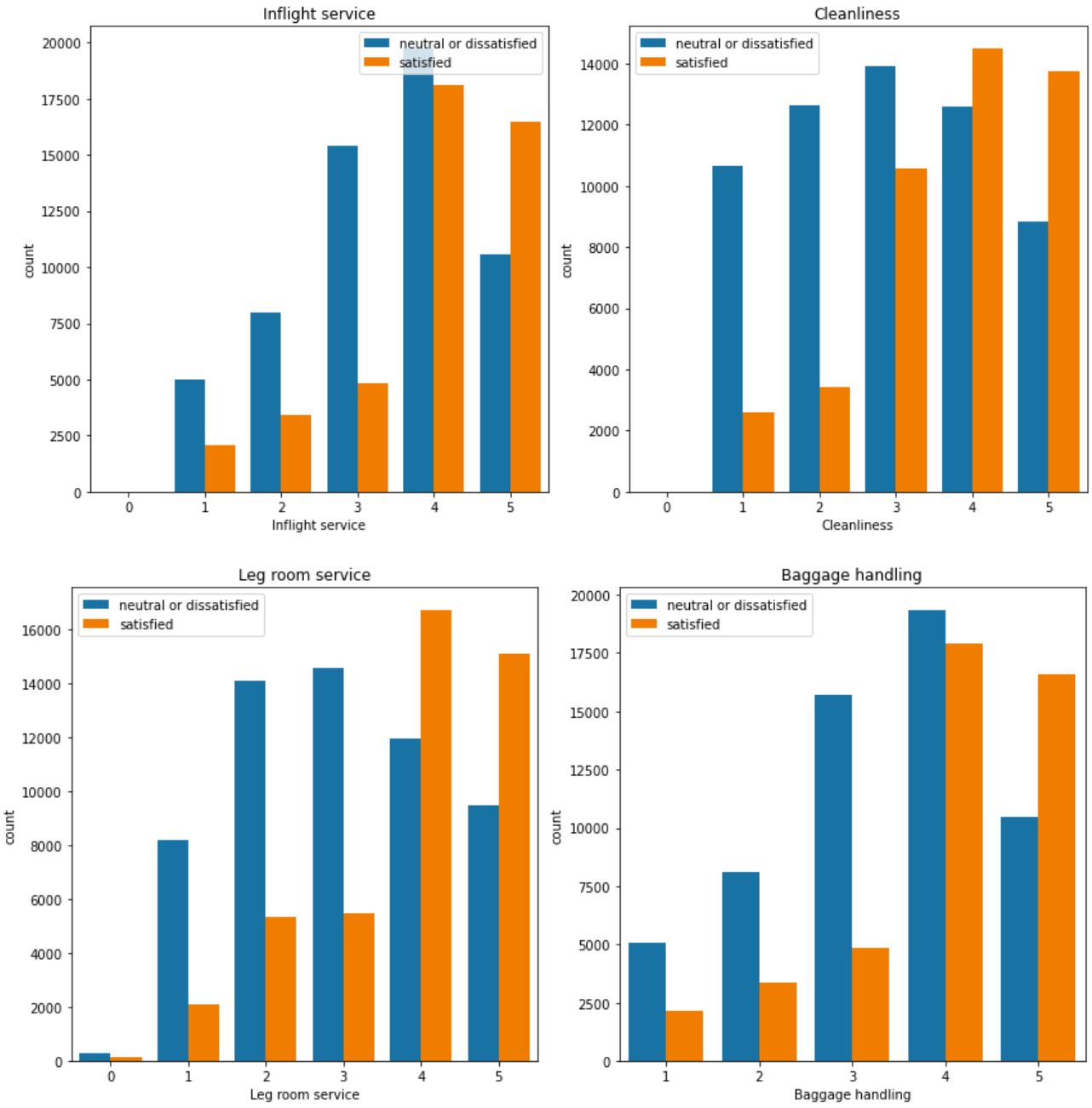


Figure 11 Bivariate analysis of inflight service, cleanliness, legroom service and baggage handling

The graph demonstrates that passengers gave the in-flight service good marks. The dissatisfied passengers praised the inflight service. There were more passengers who were neutral or displeased with the cleanliness. Similarly, the leg room has received a lot of positive feedback. Finally, more unsatisfied passengers rated baggage handling as satisfactory.

2.5.6 Exploring distributions with test v/s train data

The value counts and distributions in all features are explored in this subsection. In both train and test data, the function below counts items in categorical variables.

a. Analysis of Inflight wifi service

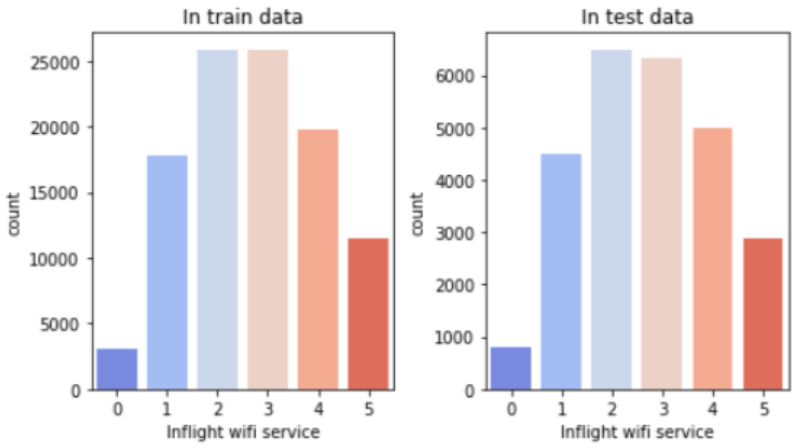


Figure 12 Analysis of inflight service train v/s test data

In both datasets, the majority of passengers were satisfied with the internet service on board.

b. Analysis of Dep/Arr time convenient

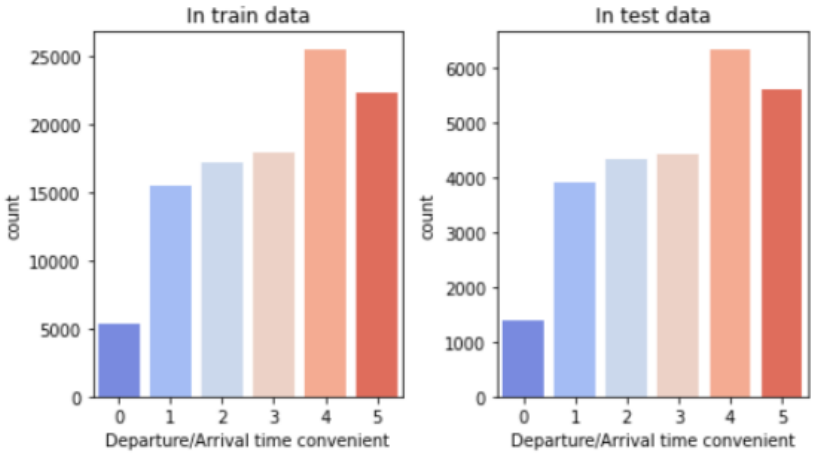


Figure 13 Analysis of dep/arr time convenient train v/s test data

Most passengers considered departure and arrival times convenient for them in both train and test data.

c. Analysis of Ease of online booking

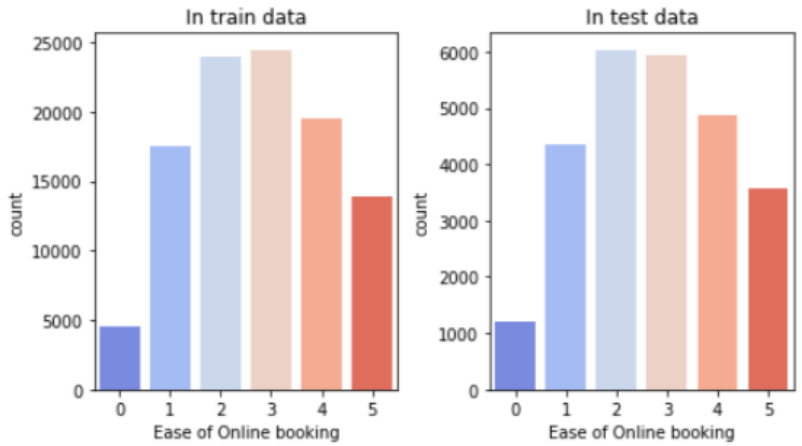


Figure 14 Analysis of ease online booking train v/s test data

Both in train and test data – most passenger admitted that online booking was not so easy for them.

d. Analysis of Food and Drink

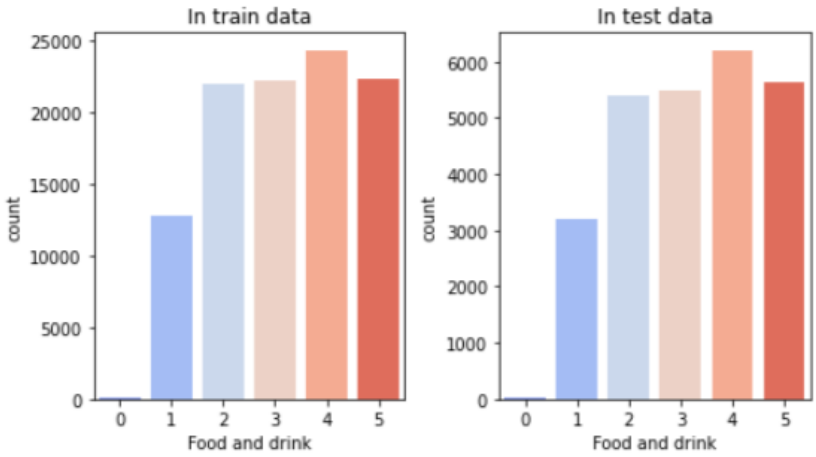
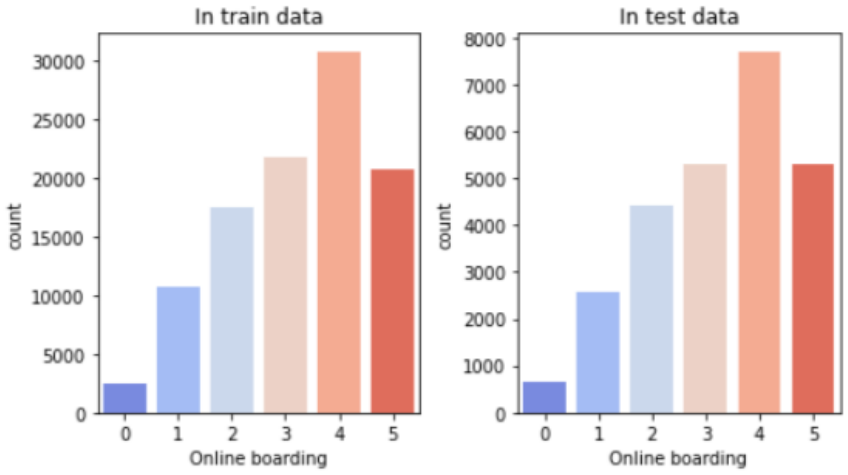


Figure 15 Analysis of food and drink train v/s test data

Equal share of passengers liked and disliked food and drinks onboard both in train and test data.

e. Analysis of Online Boarding



In terms of passenger attitudes about online boarding, there were no differences between the both train and test data.

Catboost Model for Prediction

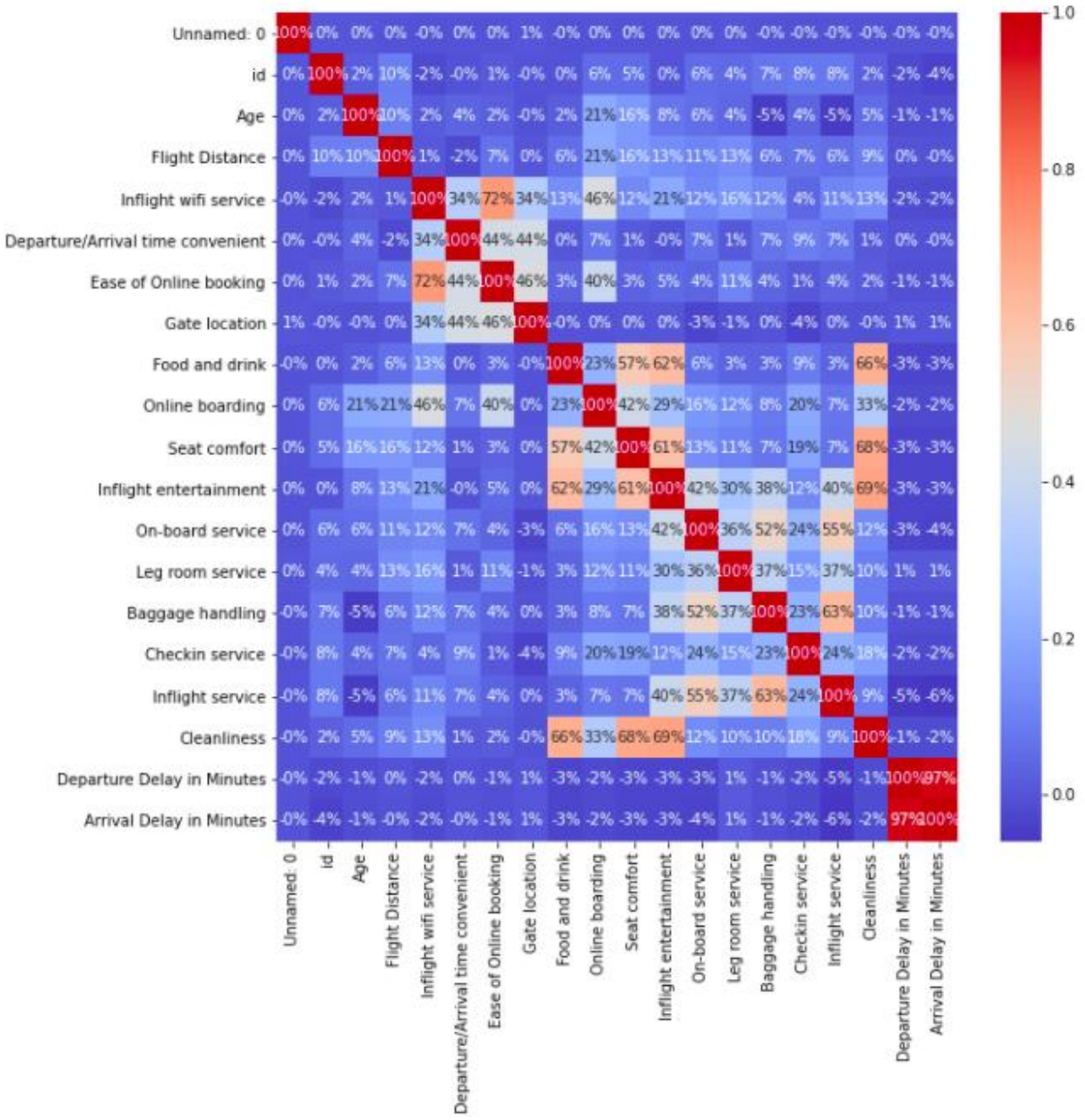
Catboost is a fast, scalable, high performance Gradient Boosting on Decision Trees library, used for ranking, classification, regression and other machine learning tasks for Python, R, Java, C++. Supports computation on CPU and GPU. This model helped in finding the most important feature and the factors correlating with satisfaction.

	satisfaction
Gender	0.012356
Customer Type	-0.187558
Age	0.137080
Type of Travel	-0.448995
Class	-0.449466
Flight Distance	0.298903
Inflight wifi service	0.284163
Departure/Arrival time convenient	-0.051718
Ease of Online booking	0.171507
Gate location	0.000449
Food and drink	0.209659
Online boarding	0.503447
Seat comfort	0.349112
Inflight entertainment	0.398203
On-board service	0.322450
Leg room service	0.313182
Baggage handling	0.247819
Checkin service	0.235914
Inflight service	0.244852
Cleanliness	0.305050
Departure Delay in Minutes	-0.052867
Arrival Delay in Minutes	-0.060050
satisfaction	1.000000

The most important features were online boarding, inflight entertainment and seat comfort.

Correlation between features

The correlation between features was drawn to have better analysis and efficient prediction about the dataset. Due to this heatmap is used because it shows the relation between features clearly.



The heatmap shows that there is significantly strong correlation between cleanliness and food and drink; cleanliness and inflight entertainment; leg room and inflight service. Hence, the project can help to conclude that the airline has to concentrate more on cleanliness, inflight service and food and drink service majorly to increase the number of satisfied passengers.

Conclusion

Findings

- Age also affects the passengers satisfaction level. The project shows that different age group have different expectations with airlines services.
- When the quality of check-in services improves, the percentage of satisfied consumers rises.
- The majority of customers were pleased with the cleanliness.
- When it comes to in-flight wifi, travellers' satisfaction levels 2 and 3 are higher than the other levels.
- The majority of passengers rated their food and beverage satisfaction on a scale of 3 to 5. It indicates that the available food and beverages are well-satisfied. In addition, while looking at the satisfaction and food and drink graphs, it is evident that the majority of 4,5 level selected passengers are satisfied with the airline, whereas the majority of 0 level picked passengers are neutral or unsatisfied.
- The majority of passengers chose 3 to 5 levels of seat comfort pleasure. The majority of passengers who choose seat comfort levels 4 and 5 are pleased with the airline.
- Travellers in business class have more legroom than passengers in economy and eco-plus classes.
- The majority of passengers who chose levels 4 and 5 for onboard service satisfaction are happy with the airline.

The project helps to the conclusion that perceived service quality has a beneficial impact on passenger satisfaction. The airline industry operates in extremely difficult conditions. As a result, company manager must understand that enhancing service quality is critical to achieving and maintaining corporate development and competitiveness. According to the findings of this study, high service quality boosts passenger happiness, which encourages them to use the same service provider again in the future.Hence, business managers should recognize that improving service quality is crucial for gaining and sustaining business growth and competitiveness.

Recommendations

The project finds that the level of customer satisfaction can depend on many factors in airline. Evaluating satisfaction in each of them can lead to efficient overview about customer opinion and total satisfaction level about the airline. Ease of online booking, online boarding, seat comfort, food and drink, inflight entertainment, leg room service and baggage handling plays the vital role in customer satisfaction. The airline should must offer quality services and improve the factors which affects the satisfaction that meets and exceeds passenger expectations.

Limitations

The data comprises of categorical data. Thus, linear regression is hard to conduct.

- The data does not have “safety” feature as a variable. This could be one of the main factor leading to feeling of satisfaction in customers.
- Most of the passengers voted for 3. This can be the regarded as response error.

Appendixes

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from matplotlib.ticker import StrMethodFormatter
```

```
sns.set_palette("rainbow")
sns.distplot(df['Age'])
plt.title("Distributuion of Age")
plt.show()
```

```
#Analysis of Satisfaction
sns.countplot(x ='satisfaction', data = df)
plt.title("Distribution of Satisfaction")
plt.show()
```

```
plt.figure(0)
# Create first chart.
y = np.array([49665, 46745, 7494])
mylabels = ["Business", "Eco", "Eco Plus"]
myexplode = [0.2, 0, 0]
plt.pie(y, labels = mylabels, explode = myexplode)
```

```
#Outliers

fig, axs = plt.subplots(ncols = 4, nrows = 1, figsize = (12, 4))
idx = 0
axs = axs.flatten()
for k, v in numeric_features.items():
    sns.boxplot(y = k, data = numeric_features, ax = axs[idx]) #color = "Blues"
    idx += 1
plt.tight_layout(pad = 0.5, w_pad = 0.5, h_pad = 5.0)
```

```
plt.figure(figsize=(12,12))
plt.subplot(2,2,1)
sns.countplot(data=df, x="Online boarding" , hue="satisfaction", palette="tab10")
plt.title("Online boarding scores")
plt.legend()

plt.subplot(2,2,2)
sns.countplot(data=df, x="Seat comfort" , hue="satisfaction", palette="tab10")
plt.title("Seat comfort")
plt.legend()

plt.subplot(2,2,3)
sns.countplot(data=df,x="Inflight entertainment",hue="satisfaction",palette="tab10")
plt.title("Inflight entertainment")
plt.legend()

plt.subplot(2,2,4)
sns.countplot (data=df, x="On-board service", hue='satisfaction', palette="tab10")
plt.title("On-board service")
plt.legend()
plt.tight_layout()
plt.show()

plt.show()
```

```
plt.figure(figsize=(12,12))
plt.subplot(2,2,1)
sns.countplot(data=df, x="Class" , hue="satisfaction", palette="tab10")
plt.title("Class")
plt.legend()

plt.subplot(2,2,2)
sns.countplot(data=df, x="Gate location" , hue="satisfaction", palette="tab10")
plt.title("Gate location")
plt.legend()

plt.subplot(2,2,3)
sns.countplot(data=df, x="Customer Type" , hue="satisfaction", palette="tab10")
plt.title("Customer Type")
plt.legend()

plt.subplot(2,2,4)
sns.countplot(data=df, x="Checkin service" , hue="satisfaction", palette="tab10")
plt.title("Checkin service")
plt.legend()
plt.tight_layout()
plt.show()
```

```
plt.figure(figsize=(12,12))
plt.subplot(2,2,1)
sns.countplot(data=df, x="Inflight wifi service" , hue="satisfaction", palette="tab10")
plt.title("Inflight wifi service")
plt.legend()

plt.subplot(2,2,2)
sns.countplot(data=df, x="Departure/Arrival time convenient" , hue="satisfaction", palette="tab10")
plt.title("Departure/Arrival time convenient")
plt.legend()

plt.subplot(2,2,3)
sns.countplot(data=df, x="Ease of Online booking" , hue="satisfaction", palette="tab10")
plt.title("Ease of Online booking")
plt.legend()

plt.subplot(2,2,4)
sns.countplot(data=df, x="Food and drink" , hue="satisfaction", palette="tab10")
plt.title("Food and drink")
plt.legend()
plt.tight_layout()
plt.show()
```

```
plt.figure(figsize=(12,12))
plt.subplot(2,2,1)
sns.countplot(data=df, x="Leg room service" , hue="satisfaction", palette="tab10")
plt.title("Leg room service")
plt.legend()

plt.subplot(2,2,2)
sns.countplot(data=df, x="Baggage handling" , hue="satisfaction", palette="tab10")
plt.title("Baggage handling")
plt.legend()

plt.subplot(2,2,3)
sns.countplot(data=df, x="Inflight service" , hue="satisfaction", palette="tab10")
plt.title("Inflight service")
plt.legend()

plt.subplot(2,2,4)
sns.countplot(data=df, x="Cleanliness" , hue="satisfaction", palette="tab10")
plt.title("Cleanliness")
plt.legend()
plt.tight_layout()
plt.show()
```

```
plt.figure(figsize=(10,10))
sns.heatmap(df.corr(),annot=True,cmap="coolwarm",fmt='.0%')
```

```
plot_counts("Departure/Arrival time convenient", [0, 1, 2, 3, 4, 5])
```

```
plot_counts("Ease of Online booking", [0, 1, 2, 3, 4, 5])
```

```
plot_counts("Food and drink", [0, 1, 2, 3, 4, 5])
```

```
plot_counts("Online boarding", [0, 1, 2, 3, 4, 5])
```

```
#Analysis of Customer Type

y=np.array([84923,18981])
mylabels=["Loyal Customer","disloyal Customer"]
myexplode = [0.2,0]
plt.pie(y, labels = mylabels, explode = myexplode, )
plt.show()
```

```
plot_counts("Inflight wifi service", [0, 1, 2, 3, 4, 5])
```

```
from sklearn.preprocessing import LabelEncoder
df = df.apply(LabelEncoder().fit_transform)
```

executed in 453ms, finished 17:06:12 2021-11-30

```
df.corr()[['satisfaction']]
```

References

TJKlein. (n.d.). *Kaggle*. Retrieved from https://www.kaggle.com/teejmahal20/airline-passenger-satisfaction![image.png](attachment:image.png).

Academia. (n.d.). Retrieved from https://www.academia.edu/36990824/RESEARCH_ON_STRATEGIC_ANALYSIS_and_CUSTOMER_SATISFICATION_TOWARDS_SERVICE_PROVIDED_BY_LOW_COST_CARRIER_AIRLINES_IN_INDIA.

Sandu, S. (n.d.). *Kaggle*. Retrieved from https://www.kaggle.com/siumisandu/passenger-satisfaction-descriptive-analysis.

Patil, V. (n.d.). *Kaggle*. Retrieved from https://www.kaggle.com/vaishnavipatil4848/airline-passenger-satisfaction.

Kaggle. (n.d.). Retrieved from https://www.kaggle.com/erntkn/satisfaction.

Kaggle. (n.d.). Retrieved from https://www.kaggle.com/nalkrolu/satisfaction-analysis-and-classification-airline.

Geeks for geeks. (n.d.). Retrieved from https://www.geeksforgeeks.org/seaborn-color-palette/.

Seaborn Pydata. (n.d.). Retrieved from https://seaborn.pydata.org/tutorial/color_palettes.html.

researchgate. (n.d.). Retrieved from https://www.researchgate.net/publication/312889742_Perceived_Service_Quality_and_Customer_Loyalty_The_Mediating_Effect_of_Passenger_Satisfaction_in_the_Nigerian_Airline_Industry.

merriam-webster. (n.d.). Retrieved from https://www.merriam-webster.com/dictionary/unimodal.

Statisticsbyjim. (n.d.). Retrieved from https://statisticsbyjim.com/basics/chebyshevs-theorem-in-statistics/.

mathsisfun. (n.d.). Retrieved from https://www.mathsisfun.com/mode.html.

mathsisfun. (n.d.). Retrieved from https://www.mathsisfun.com/definitions/median.html.

mathsisfun. (n.d.). Retrieved from https://www.mathsisfun.com/definitions/mean.html.

Kotler, P. (2015). *Principles of Marketing.* Pearson Education Inc.

Mann, P. S. (2016). *Introductory Statistics.* John Wiley & Sons, Inc.

(htt1)

Bibliography

Kotler, P. (2015). *Principles of Marketing.* Pearson Education Inc.

Academia. (n.d.). From Academia.edu: 1. https://www.academia.edu/36990824/RESEARCH_ON_STRATEGIC_ANALYSIS_and_CUSTOMER_SATISFICATION_TOWARDS_SERVICE_PROVIDED_BY_LOW_COST_CARRIER_AIRLINES_IN_INDIA

(A4A), A. f. (2021). From https://www.airlines.org/impact/

Slotnick, D. (2021). From THEPOINTSGUY: https://thepointsguy.com/news/airline-rankings-2021-jd-power/

Saha G. Theingi, S. (2009). Service quality, satisfaction, and behavioural intentions: A□study of low-cost airline carriers in Thailand.

TJKlein. (n.d.). *Kaggle*. From https://www.kaggle.com/teejmahal20/airline-passenger-satisfaction![image.png](attachment:image.png).

Academia. (n.d.). From https://www.academia.edu/36990824/RESEARCH_ON_STRATEGIC_ANALYSIS_and_CUSTOMER_SATISFICATION_TOWARDS_SERVICE_PROVIDED_BY_LOW_COST_CARRIER_AIRLINES_IN_INDIA.

Sandu, S. (n.d.). *Kaggle*. From https://www.kaggle.com/siumisandu/passenger-satisfaction-descriptive-analysis.

Patil, V. (n.d.). *Kaggle*. From https://www.kaggle.com/vaishnavipatil4848/airline-passenger-satisfaction.

Kaggle. (n.d.). From https://www.kaggle.com/erntkn/satisfaction.

Kaggle. (n.d.). From https://www.kaggle.com/nalkrolu/satisfaction-analysis-and-classification-airline.

Geeks for geeks. (n.d.). From https://www.geeksforgeeks.org/seaborn-color-palette/.

Seaborn Pydata. (n.d.). From https://seaborn.pydata.org/tutorial/color_palettes.html.

researchgate. (n.d.). From https://www.researchgate.net/publication/312889742_Perceived_Service_Quality_and_Customer_Loyalty_The_Mediating_Effect_of_Passenger_Satisfaction_in_the_Nigerian_Airline_Industry.

merriam-webster. (n.d.). From https://www.merriam-webster.com/dictionary/unimodal.

Statisticsbyjim. (n.d.). From https://statisticsbyjim.com/basics/chebyshevs-theorem-in-statistics/.

mathsisfun. (n.d.). From https://www.mathsisfun.com/mode.html.

mathsisfun. (n.d.). From https://www.mathsisfun.com/definitions/median.html.

mathsisfun. (n.d.). From https://www.mathsisfun.com/definitions/mean.html.

Kotler, P. (2015). *Principles of Marketing.* Pearson Education Inc.

Mann, P. S. (2016). *Introductory Statistics.* John Wiley & Sons, Inc.

(n.d.). From https://blogs.perficient.com/2018/05/14/customer-satisfaction-in-the-airline-industry/