BDAT 1005 – Summer 2023
Mathematics for Data Analytics

Dataset Exploration Project
Final Report on Airline Customer Satisfaction Data

Submission to:
Ehsan Pourjavad

Submission by:
Avhimanhu Sapkota
Student ID: **200547612**

Submission date:
13 August 2023

Table of Contents

Introduction

In the foundation of the aviation industry, passenger satisfaction emerges as a pivotal determinant of success. This analysis delves into the panorama of passenger experiences, seeking critical insights to redefine strategies, elevate quality service, and foster unwavering customer loyalty. The aviation sector, marked by severe competition, hence understanding, and enhancing passenger satisfaction are not merely options, but is an essential for sustainable growth.

The focal point of this exploration is a comprehensive dataset derived from a multitude of passenger feedback mechanisms. This repository encapsulates a diverse array of variables, spanning passenger demographics, travel particulars, and satisfaction ratings across an array of in-flight amenities. This analysis is driven by a multi-faceted purpose: to identify the determinants of passenger satisfaction, to identify underlying patterns and trends, and to explore actionable insights capable of driving strategic decisions. The outcomes of this analysis hold the promise of illuminating pathways for airlines and industry stakeholders to refine their offerings, elevate passenger experiences, and enhance their competitive edge.

Some research questions that have been shaped within the framework of FINER, have been carefully curated to provide a holistic understanding of passenger satisfaction dynamics. These questions span a range of inquiries, from discerning the impact of passenger type on overall satisfaction in unraveling the relationship between flight distance and perceptions of seat comfort. The implications of this analysis may perhaps provide a compass for airlines to navigate through the tough terrain of passenger satisfaction. By aligning strategies with the needs and preferences of passengers, airlines can harness the power of data-driven decision-making to shape exceptional passenger experiences. In a world where passenger loyalty is a must, this analysis stands as a beacon, guiding industry players toward a future where satisfaction is not just a goal, but a foundation for long-term success.

Description of the Data Set

The major focus of this analysis lies in a dataset that encapsulates the background of airline passenger satisfaction. Sourced from an online dataset repository, Maven Analytics (John D, 2022), this dataset represents a satisfaction rating based on passenger experiences, comprising feedback from over 120,000 individuals who took the flights. The data comprises a total of 129,880 records, each a unique window into a passenger's journey, alongside 24 attributes that collectively form a comprehensive insight. This dataset's attributes span a range of both qualitative and quantitative variables, ensuring a thorough exploration of passenger experiences. While attributes like ID, gender, customer type, type of travel, class, and satisfaction fall under the qualitative category, the dataset further encompasses other quantitative attributes. These include flight distance, departure delay, check-in service, seat comfort, cleanliness, in-flight entertainment, and more – all vital components in exploring the elements that contribute to passenger satisfaction.

Ranges and Limitations

A thorough examination of dataset was conducted to identify ranges and limitations. Some missing values for the attribute "Arrival Delay" was identified with the help of MS-Excel "Go to Special" feature. No invalid values or missing values were detected in other attributes whatsoever. In conclusion, Arrival Delay variable needs to be cleaned while the rest of the data is ready for analytics and visualization.

Assumptions

On the basis of the dataset exploration, there are various attributes and data that needed assumptions. All the necessary assumptions have been discussed below.

- The proper details about the data being of a single airline company was not mentioned in Maven Analytics (the portal from where the dataset was acquired), neither the name of any organization was stated thus, it has been presumed that the data belongs to an airline company, referred as Georgian Airlines from hereafter, whose flight details and customer satisfaction has been recorded via survey within their aircrafts' survey portal.

- It was unclear whether the dataset represents a population or a sample. The flight distance attribute in the dataset has a wide range of distances and based on the fact it has been concluded that the data was not compiled from a single flight although there are same flight distances. Therefore, it has been assumed that the dataset is a sample with a combination of data from multiple flights' survey. In general, a sample can be defined as a subset of a larger group that best represents the overall dataset (Ravikiran, 2023).

- Although the units of measurements for some variables like age can be mentioned presumably in years, it has not been explicitly mentioned for other variables. Therefore, the following assumptions for units of measurements have been considered for the dataset.

  - Flight distance attribute in the data has been presumed to be in kilometers.
  - Departure delay and arrival delay attributes has been assumed to be measured in minutes.
  - The numeric ratings for departure and arrival time convenience, ease of online booking, check-in service, online boarding, gate location, on-board service, sear comfort, leg room service, cleanliness, food and drink, in-flight service, in-flight Wi-Fi service, and in-flight entertainment attributes have been assumed to be out of 5, where 1 is assumed to be the lowest, 5 as the highest and 0 meaning not applicable for the flight, based on the range of data listed within these attributes

Data Dictionary

A comprehensive data dictionary has been assembled to provide clarity and context to each attribute within the dataset. It serves as a compass, guiding the analyst through the intricate nuances of each variable, ensuring a deep understanding of their significance and implications.

S.N.	Field Name	Data Description
1.	ID	**Statistical Variable Type:** Ordinal **Possible Values:** Unique Serial Number for each row starting from 1. **Ranges or Limitations:** None identified. **Description:** Unique identifier for each passenger. Acts as a primary key for the dataset.
2.	Gender	**Statistical Variable Type:** Nominal **Possible Values:** Male (coded as 1) and Female (coded as 0) **Ranges or Limitations:** None identified. **Description:** The gender of the passenger.
3.	Age	**Statistical Variable Type:** Continuous **Possible Values:** Integer values for age (within 0-100 most of the time) **Ranges or Limitations:** None identified. **Description:** The age of the travelling passenger.
4.	Customer Type	**Statistical Variable Type:** Nominal **Possible Values:** First-time or Returning **Ranges or Limitations:** None identified. **Description:** The type of customer to identify whether they are using the airlines for the first time or returning.
5.	Type of Travel	**Statistical Variable Type:** Nominal **Possible Values:** Business or Personal **Ranges or Limitations:** None identified. **Description:** The purpose of flight for every individual passenger.
6.	Class	**Statistical Variable Type:** Nominal **Possible Values:** Business, Economy, and Economy Plus **Ranges or Limitations:** None identified. **Description:** The type of seat the passenger is flying on during the flight.
7.	Flight Distance	**Statistical Variable Type:** Continuous **Possible Values:** Positive numeric value **Ranges or Limitations:** None identified. **Description:** The total distance covered by the aircraft for the flight (assumed to be in Kilometers).
8.	Departure Delay	**Statistical Variable Type:** Continuous **Possible Values:** Positive numeric value **Ranges or Limitations:** None identified.

		Description: The total time the flight was delayed before the take-off (assumed to be in minutes).
9.	Arrival Delay	**Statistical Variable Type:** Continuous **Possible Values:** Positive numeric value **Ranges or Limitations:** Missing values in some rows (needs to be cleaned). **Description:** The total time the flight was delayed while landing (assumed to be in minutes).
10.	Departure and Arrival Time Convenience	**Statistical Variable Type:** Ordinal **Possible Values:** Numeric values from (0-5). **Ranges or Limitations:** None identified. **Description:** The satisfaction level of passenger based-on flight departure and arrival times, where 1 represents lowest, 5 represents highest and 0 indicates not applicable.
11.	Ease of Online Booking	**Statistical Variable Type:** Ordinal **Possible Values:** Numeric values from (0-5). **Ranges or Limitations:** None identified. **Description:** The satisfaction level of passenger with the online booking experience, where 1 represents lowest, 5 represents highest and 0 indicates not applicable.
12.	Check-in Service	**Statistical Variable Type:** Ordinal **Possible Values:** Numeric values from (0-5). **Ranges or Limitations:** None identified. **Description:** The satisfaction level of passenger based-on check in service before the flight, where 1 represents lowest, 5 represents highest and 0 indicates not applicable.
13.	Online Boarding	**Statistical Variable Type:** Ordinal **Possible Values:** Numeric values from (0-5). **Ranges or Limitations:** None identified. **Description:** The satisfaction level of passenger with the online boarding experience, where 1 represents lowest, 5 represents highest and 0 indicates not applicable.
14.	Gate Location	**Statistical Variable Type:** Ordinal **Possible Values:** Numeric values from (0-5). **Ranges or Limitations:** None identified. **Description:** The satisfaction level of passenger based-on the gate location and convenience to find their way through within the airport, where 1 represents lowest, 5 represents highest and 0 indicates not applicable.

15.	On-board service	**Statistical Variable Type:** Ordinal **Possible Values:** Numeric values from (0-5). **Ranges or Limitations:** None identified. **Description:** The satisfaction level of passenger with the services provided by the airlines within the flight, where 1 represents lowest, 5 represents highest and 0 indicates not applicable.
16.	Seat comfort	**Statistical Variable Type:** Ordinal **Possible Values:** Numeric values from (0-5). **Ranges or Limitations:** None identified. **Description:** The satisfaction level of passenger based-on the comfort of the seat, where 1 represents lowest, 5 represents highest and 0 indicates not applicable.
17.	Leg room service	**Statistical Variable Type:** Ordinal **Possible Values:** Numeric values from (0-5). **Ranges or Limitations:** None identified. **Description:** The satisfaction level of passenger with the leg room of the airplane seat, where 1 represents lowest, 5 represents highest and 0 indicates not applicable.
18.	Cleanliness	**Statistical Variable Type:** Ordinal **Possible Values:** Numeric values from (0-5). **Ranges or Limitations:** None identified. **Description:** The satisfaction level of passenger based-on the cleanliness inside the aircraft where 1 represents lowest, 5 represents highest and 0 indicates not applicable.
19.	Food and Drink	**Statistical Variable Type:** Ordinal **Possible Values:** Numeric values from (0-5). **Ranges or Limitations:** None identified. **Description:** The satisfaction level of passenger based-on the food and drinks that were served during the flight, where 1 represents lowest, 5 represents highest and 0 indicates not applicable.
20.	In-fight Service	**Statistical Variable Type:** Ordinal **Possible Values:** Numeric values from (0-5). **Ranges or Limitations:** None identified. **Description:** The satisfaction level of passenger with the in-flight service provided by the flight attendants, where 1 represents lowest, 5 represents highest and 0 indicates not applicable.
21.	In-flight Wi-Fi service	**Statistical Variable Type:** Ordinal **Possible Values:** Numeric values from (0-5). **Ranges or Limitations:** None identified. **Description:** The satisfaction level of passenger based on the Wi-Fi service provided inside the aircraft during the flight, where 1 represents lowest, 5 represents highest and 0 indicates not applicable.

22.	In-flight Entertainment	**Statistical Variable Type:** Ordinal **Possible Values:** Numeric values from (0-5). **Ranges or Limitations:** None identified. **Description:** The satisfaction level of passenger with the in-flight entertainment platform of the aircraft, where 1 represents lowest, 5 represents highest and 0 indicates not applicable.
23.	Baggage Handling	**Statistical Variable Type:** Ordinal **Possible Values:** Numeric values from (0-5). **Ranges or Limitations:** None identified. **Description:** The satisfaction level of passenger based-on how their baggage were stored and handled during the flight, where 1 represents lowest, 5 represents highest and 0 indicates not applicable.
24.	Satisfaction	**Statistical Variable Type:** Nominal **Possible Values:** Satisfied (coded as 1) and Neutral or Not Satisfied (coded as 0) **Ranges or Limitations:** None identified. **Description:** Overall satisfaction level of passenger with the airline company.

Purpose and Potential

The study of this dataset is emphasised by a variety of motivations. At its core, this analysis is performed to uncover the factors that shape passenger satisfaction. By interpreting the data, this report aims to analyze the drivers that influence passenger contentment during their air travel journeys.

With a variety of variables, this dataset provides a unique opportunity to uncover patterns, trends, and dependencies that contribute to passenger satisfaction. It is through this exploration that report endeavors to craft actionable insights capable of navigating airlines and industry stakeholders toward enhanced service quality, refined strategies, and a resounding commitment to passenger-centric experiences. Through various techniques and methodologies, extraction of meaningful insights, informing strategic decisions and fostering a landscape of air travel experiences will be performed in the report further. Hence, the analysis aims to not only address the research questions but also to pave the way for an industry that prizes customer satisfaction above everything else.

FINER Research Questions

1. **How do customers who fly in various flight classes differ in terms of their satisfaction with in-flight amenities including seat comfort, cleanliness, food and drink, and in-flight entertainment?**
   The variables in the dataset were observed where some were related to satisfaction with in-flight amenities and flight class of passengers were identified. Hence, satisfaction levels of passengers in different flight classes can be analyzed on the aspects of seat comfort, cleanliness, food and drink, and in-flight entertainment.

2. **How does the distance of a flight impact the customer's perception of seat comfort and legroom service?**
   The dataset provides information on the distance of flights and seat comfort level along with legroom service which was identified to analyzed how the duration of the flight relates to the comfort, that ultimately determines if longer flights have impact on satisfaction levels.

3. **Are there any specific patterns in satisfaction levels with online booking and ease of check-in service based-on the age and genders of the passenger?**
   In accordance with the variables online booking satisfaction, check-in service satisfaction, age, and gender various patterns or trends on satisfaction levels can be investigated. This might perhaps identify if certain age groups or gender exhibit satisfaction based-on online booking and check-in services.

4. **How does the overall satisfaction level differentiate between returning and new passengers? What are the factors that contribute to this difference?**
   Satisfaction levels of returning or new customers can be analyzed to compare overall satisfaction levels which will help to identify the factors that contribute to the differences in satisfaction so that the improvements can be made. Hence, this question was considered as one of the FINER questions.

5. **How does the departure and arrival time convenience influence customer satisfaction, and is there a difference in perception between first-class and economy-class passengers?**
   The examination of variables related to departure and arrival time convenience, customer satisfaction and seat class, the influence of these times on satisfaction levels can be analyzed to reveal difference in perception between passengers of different classes.

6. **To what extent does the type of travel (business or personal) influence the satisfaction levels of passengers with online booking, check-in service, and in-flight amenities?**
   The analysis on the relationship between type of travel and satisfaction ratings can be used to understand which travel purpose impacts passenger satisfaction ultimately providing insights on difference in preferences and satisfaction levels between business and personal travelers.

Evolution of FINER Research Questions

The evolution of the FINER research questions was a thoughtful and iterative process, guided by a thorough exploration of the dataset's attributes and their potential implications for understanding passenger satisfaction. The five questions were identified to address distinct aspects of the air travel experience. The questions revolved around various dimensions of passenger satisfaction, such as flight class and in-flight amenities, flight distance and comfort perception, online booking, check-in services, demographics, returning vs. new passengers' satisfaction, and departure and arrival time convenience. These questions perhaps will be helpful to provide insights into different facets of the passenger experience as they expanded the focus to explore passenger satisfaction levels throughout the exploration.

Data Cleaning and Manipulation

In the process of preparing the dataset for analysis, several data cleaning and manipulation steps were undertaken to ensure the accuracy, relevance, and coherence of the data. These steps were crucial in addressing each of the research questions and deriving meaningful insights from the dataset. The following techniques were employed:

New Column- Overall Satisfaction

To provide a comprehensive view of passenger satisfaction, a new column named "Overall Satisfaction" was created. This column was generated by calculating the average of ratings across various satisfaction categories such as seat comfort, cleanliness, in-flight entertainment, and more. By rounding off these averages, a single value was derived that encapsulates an individual passenger's overall satisfaction level. This new metric serves as a valuable indicator for assessing general passenger contentment and forms a foundation for deeper analysis into the determinants of overall satisfaction.

Removed Columns

To streamline the dataset and focus on pertinent variables, certain columns were removed from consideration in the analysis. These columns encompassed attributes like Departure delay, arrival delay, online boarding, gate location, on-board service, in-flight service, in-flight Wi-Fi service, and baggage handling. The remaining columns refined the dataset, ensuring that only variables directly relevant to the research questions were retained. This approach enhances the precision of the analysis, eliminating unnecessary information and highlighting key factors that contribute to passenger satisfaction.

Categorization of Variables

The process of categorizing variables was applied to specific attributes within the dataset, yielding structured and interpretable data points.

Gender

A coding system was introduced to categorize the Gender variable. Male respondents were assigned a numeric code of 1, while Female respondents were designated as 0. This coding facilitates the exploration of gender-related patterns, relationships, and trends in subsequent analysis. The consistent application of codes enhances data organization, simplifies interpretation, and enables gender-based insights to be extracted.

Satisfaction

Another significant categorization effort was applied to the Satisfaction variable. Ratings indicating "Satisfied" were encoded as 1, while ratings denoting "Neutral or Not Satisfied" were coded as 0. This transformation quantifies satisfaction levels, enabling a numerical exploration of satisfaction patterns and proportions of satisfied versus unsatisfied passengers. These categorizations have been methodically documented in the updated data dictionary.

Limitations

While the dataset offers a robust foundation for analysis, it is important to acknowledge certain limitations that may impact the scope and depth of the investigation. Notably, the dataset does not provide explicit information about the airline company or specific flight details. This absence of context may hinder a more granular understanding of factors influencing passenger satisfaction, potentially limiting the generalizability of findings to the broader airline industry. Additionally, the dataset's reliance on survey responses introduces the potential for response bias, as passengers may provide subjective or skewed feedback. Furthermore, the absence of explicit timeframes for data collection and lack of information about the survey's distribution method could affect the representation of the dataset. These limitations are essential to interpret the results and draw meaningful conclusions from the analysis.

Univariate Descriptive Analysis

A comprehensive univariate analysis was conducted on the 'Airline Customer Satisfaction' dataset, revealing crucial insights into the distribution and characteristics of individual variables. This analytical approach aimed to uncover hidden patterns and trends within each variable, facilitating a deeper understanding of passenger experiences and satisfaction levels.

Qualitative Variables

Gender

Gender	Frequency
Female	65899
Male	63981
Grand Total	129880
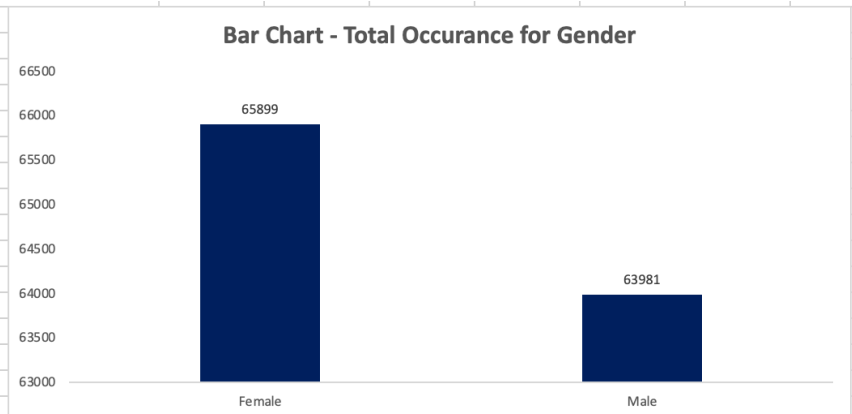
An exploration of gender distribution unveiled that the dataset consisted of 65,899 female and 63,981 male respondents. The utilization of a bar chart effectively visualized this distribution, enabling an immediate understanding of gender representation.

Customer Type

Customer Type	Frequency
First-time	23780
Returning	106100
Grand Total	129880
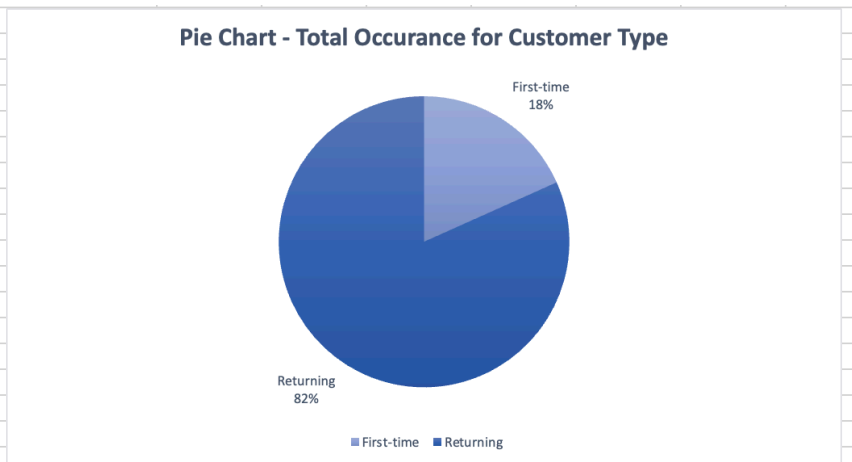
The examination of customer types highlighted 23,780 first-time customers and 106,100 returning customers. This univariate analysis showcased the prevalence of customer types within the dataset.

Type of Travel

Type of Travel	Frequency
Business	89693
Personal	40187
Grand Total	**129880**

Bar Chart - Total Occurance for Type of Travel

Personal — 40187

Business — 89693

(x-axis: 0, 10000, 20000, 30000, 40000, 50000, 60000, 70000, 80000, 90000, 100000)

■ Frequency

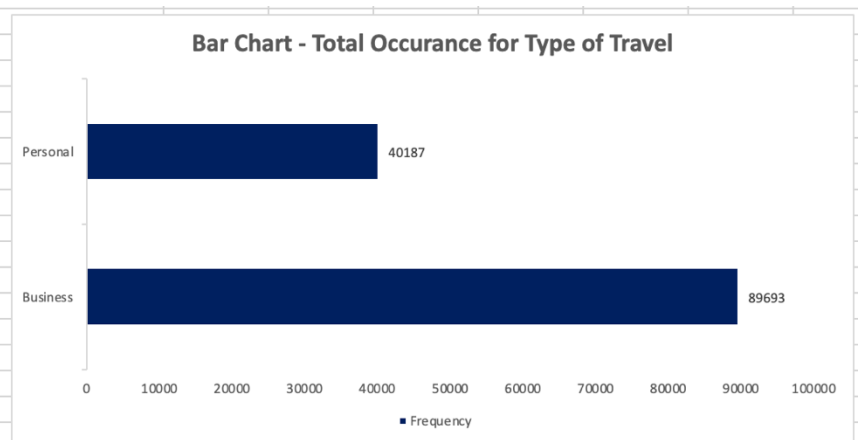Within the dataset, 89,693 instances of business travel and 40,187 instances of personal travel were identified. A bar chart comparison effectively communicated the distribution of travel types.

Flight Class

Flight Class	Frequency
Business	62160
Economy	58309
Economy Plus	9411
Grand Total	**129880**

PIE CHART - TOTAL OCCURANCE FOR FLIGHT CLASS

(Pie chart with segments: Economy Plus, Business, Economy)
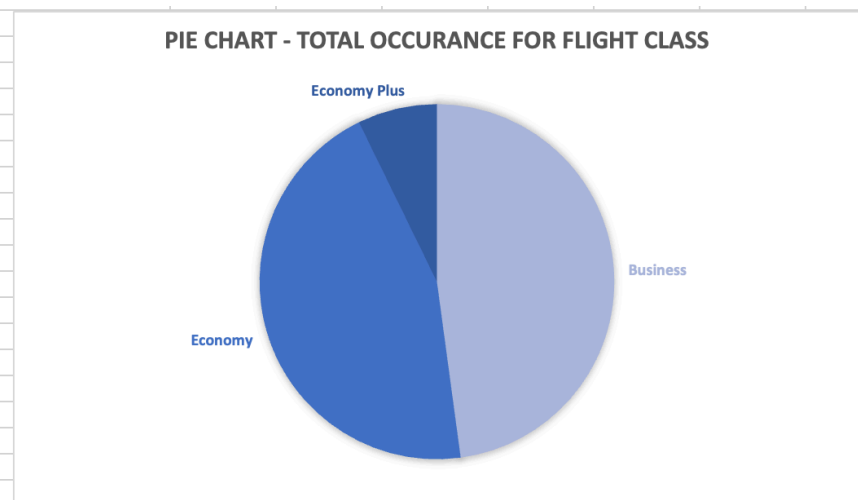
The analysis of flight classes revealed that Business class had the highest frequency with 62,160 instances, followed by Economy class with 58,309, and Economy Plus with 9,411 instances. This distribution elucidated passenger preferences in terms of flight classes.

Satisfaction

Satisfaction	Frequency
Neutral or Dissatisfied	73452
Satisfied	56428
Grand Total	**129880**
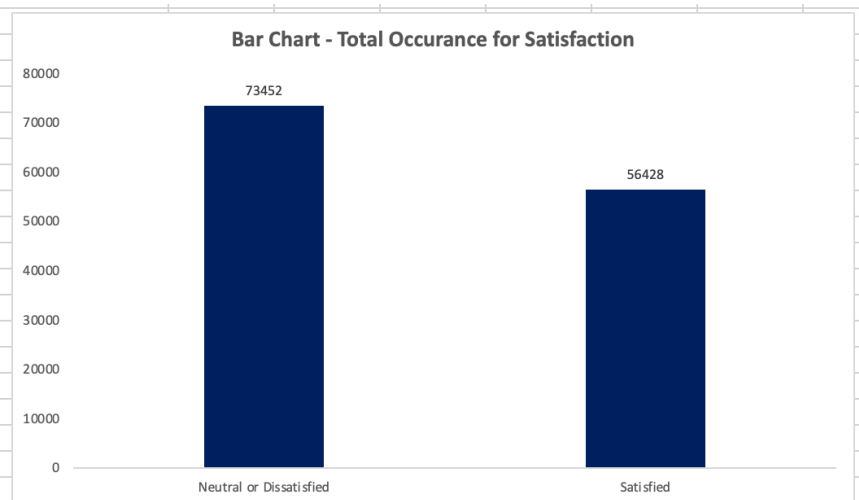
The variable 'Satisfaction' indicated that 73,452 respondents were neutral or dissatisfied, while 56,428 expressed satisfaction. The bar chart representation provided a clear insight into the distribution of satisfaction levels.

Quantitative Variables

Age

Age	
Mean	39.43
Standard Error	0.04195
Median	40
Mode	39
Standard Deviation	15.12
Sample Variance	228.60
Kurtosis	-0.72
Skewness	-0.0036
Range	78
Minimum	7
Maximum	85
Sum	5120903
Count	129880
Q1	27
Q3	51
IQR	24

Histogram - Total Occurance for each Age Group
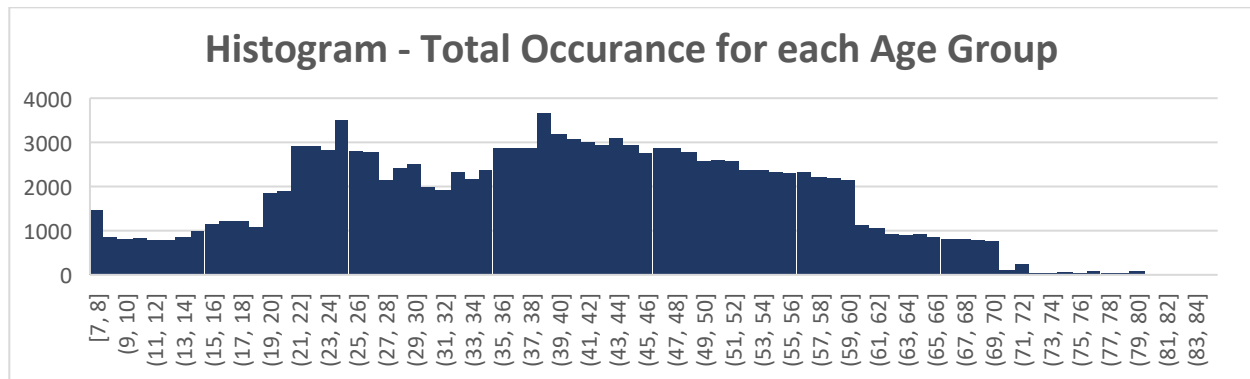
An in-depth examination of the 'Age' variable offered a histogram showcasing the distribution of respondent ages. The histogram revealed that the majority of passengers fell within the 20 to 60-year age range, with a mean age of 39.43 and a standard deviation of 15.12. The spread of ages was effectively visualized, informing the understanding of passenger age demographics.

Flight Distance

Flight Distance	
Mean	1190.32
Standard Error	2.76771
Median	844
Mode	337
Standard Deviation	997.45
Sample Variance	994911.44
Kurtosis	0.27
Skewness	1.11
Range	4952
Minimum	31
Maximum	4983
Sum	154598293
Count	129880
Q1	414
Q3	1744
IQR	1330

Histogram - Total Occurance based on Flight Distance
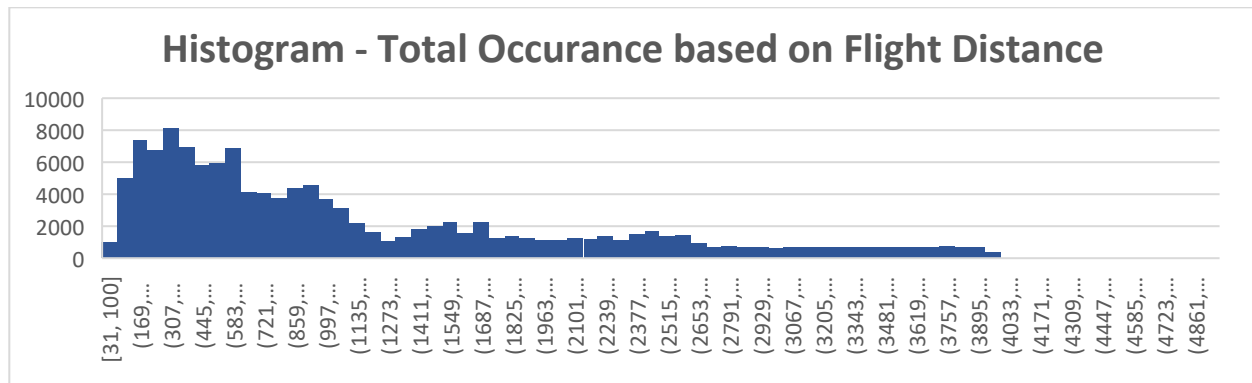
The frequency distribution of flight distances was depicted through a histogram, highlighting the concentration of flights between 414 and 1744 units. The mean flight distance was 1190.32, with a standard deviation of 997.45. The histogram illuminated the variability and distribution of flight distances.

Departure & Arrival Time Convenience

Departure & Arrival Time Convenience	
Mean	3.06
Standard Error	0
Median	3
Mode	4
Standard Deviation	1.53
Sample Variance	2.33
Kurtosis	-1.04
Skewness	-0.33
Range	5
Minimum	0
Maximum	5
Sum	397121
Count	129880
Q1	3
Q3	4
IQR	1

Line Chart - Customer Rating Occurances for Departure and Arrival Time Convenience
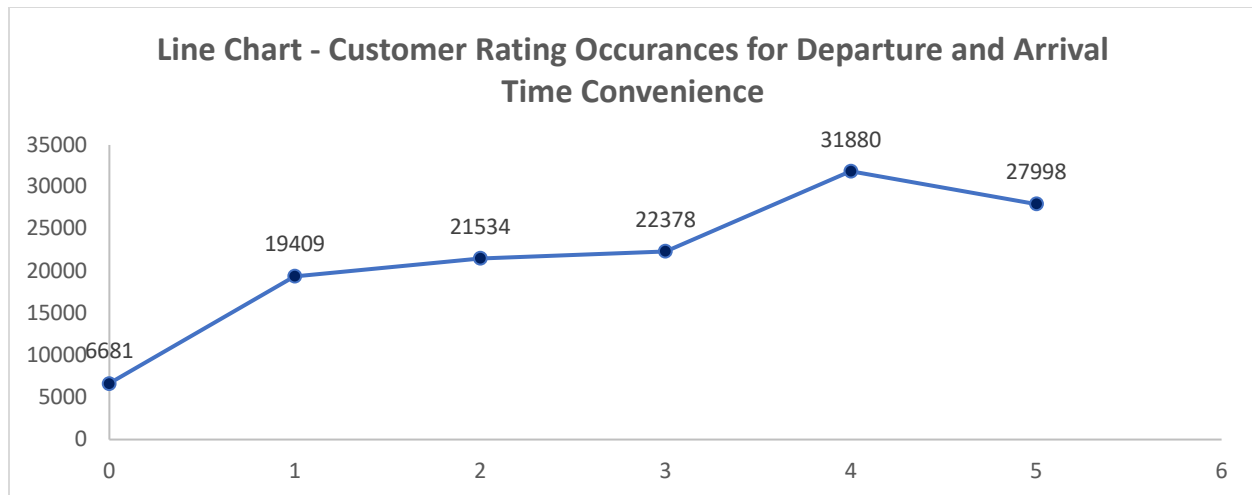
The convenience ratings for departure and arrival times were visualized through a line chart. The mean rating of 3.06, with a standard deviation of 1.53, indicated the variability and central tendency of convenience ratings.

Ease of Online Booking

Ease of Online Booking	
Mean	2.76
Standard Error	0
Median	3
Mode	3
Standard Deviation	1.40
Sample Variance	1.96
Kurtosis	-0.91
Skewness	-0.02
Range	5
Minimum	0
Maximum	5
Sum	358063
Count	129880
Q1	2
Q3	4
IQR	2

Bar Chart - Customer Rating Occurances for Ease of Online Booking

Rating	Count
5	17424
4	24444
3	30393
2	30051
1	21886
0	5682
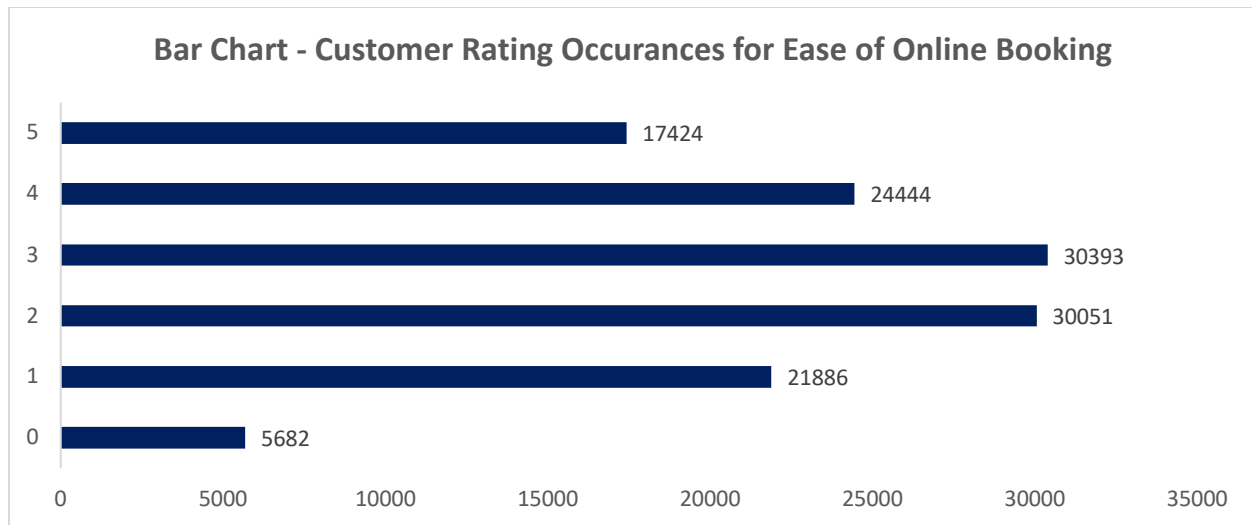
A bar chart presented the distribution of ratings for ease of online booking. The mean rating of 2.76, median of 3, and standard deviation of 1.40 conveyed the variation in online booking satisfaction.

Check-in Services

Check-in Service	
Mean	3.31
Standard Error	0
Median	3
Mode	4
Standard Deviation	1.27
Sample Variance	1.60
Kurtosis	-0.83
Skewness	-0.37
Range	5
Minimum	0
Maximum	5
Sum	429418
Count	129880
Q1	3
Q3	4
IQR	1

Line Chart - Customer Rating Occurances for Check-in Service
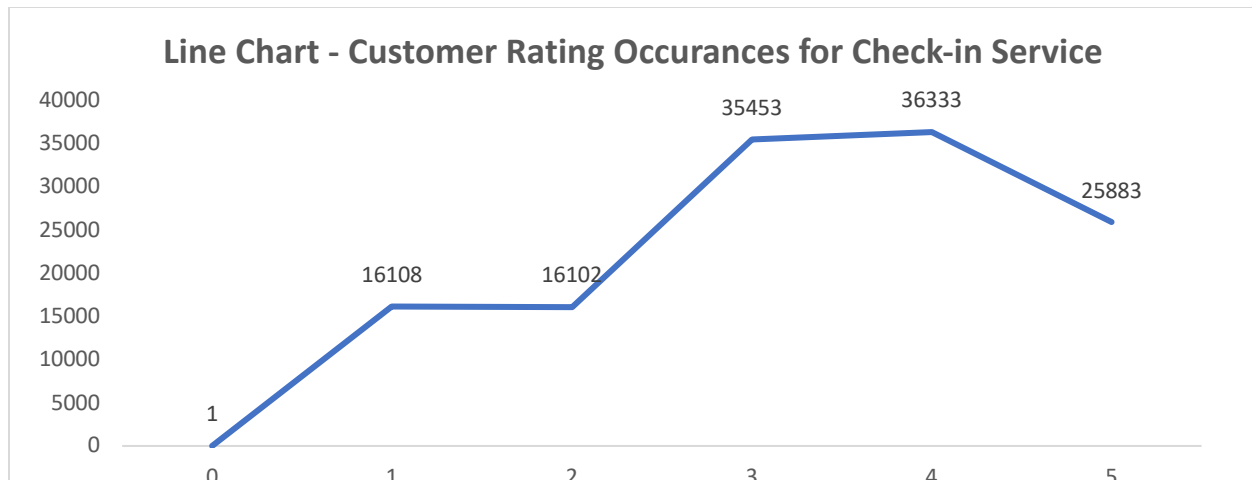
The trend in ratings for check-in services was depicted through a line chart. With a mean rating of 3.31 and standard deviation of 1.27, the distribution of check-in service satisfaction was effectively communicated.

Seat Comfort

Seat Comfort	
Mean	3.44
Standard Error	0.0037
Median	4
Mode	4
Standard Deviation	1.32
Sample Variance	1.74
Kurtosis	-0.92
Skewness	-0.49
Range	5
Minimum	0
Maximum	5
Sum	446964
Count	129880
Q1	2
Q3	5
IQR	3

Bar Chart - Customer Rating Occurances for Seat Comfort
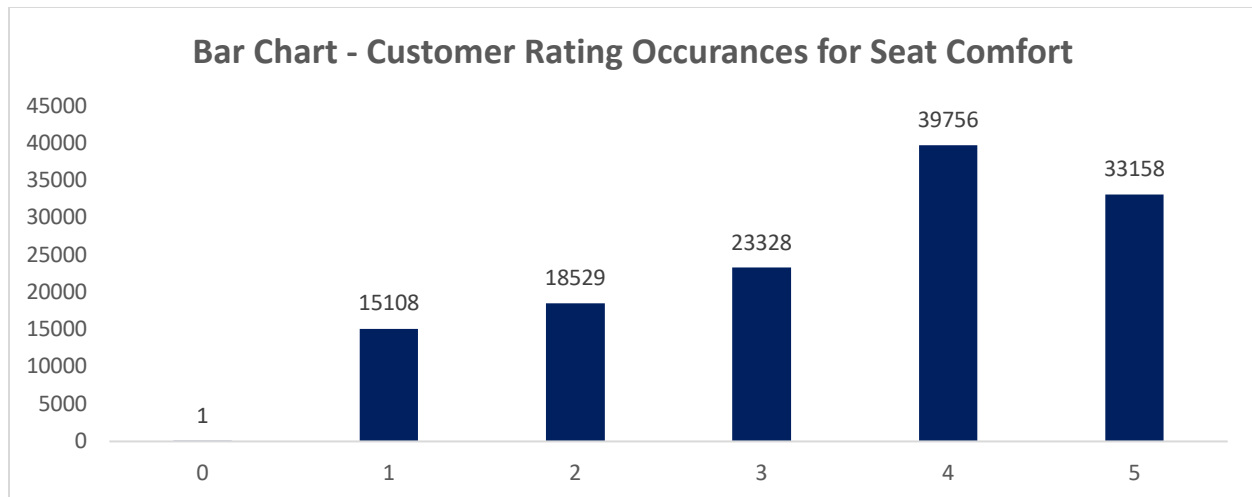
Passenger satisfaction with seat comfort yielded a mean rating of 3.44, indicating a moderately satisfactory experience. The majority rated seat comfort at 4, displaying a positive skew. The analysis showcased a standard deviation of 1.32, suggesting moderate variability.

Leg Room Service

Leg Room Service	
Mean	3.35
Standard Error	0.00365
Median	4
Mode	4
Standard Deviation	1.32
Sample Variance	1.73
Kurtosis	-0.98
Skewness	-0.35
Range	5
Minimum	0
Maximum	5
Sum	435212
Count	129880
Q1	2
Q3	4
IQR	2

PIE CHART - CUSTOMER RATING OCCURANCES FOR LEG ROOM SERVICE
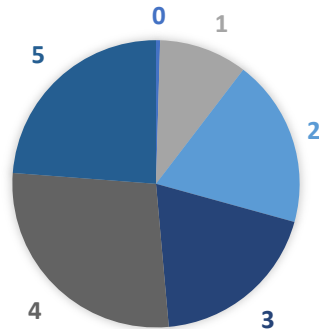
Analysis of leg room service satisfaction showed a mean rating of 3.35, with a slight skew towards higher satisfaction (median and mode = 4). Standard deviation was 1.32, indicating moderate variability.

Cleanliness

Cleanliness	
Mean	3.29
Standard Error	0.00365
Median	3
Mode	4
Standard Deviation	1.31
Sample Variance	1.73
Kurtosis	-1.01
Skewness	-0.30
Range	5
Minimum	0
Maximum	5
Sum	426828
Count	129880
Q1	2
Q3	4
IQR	2

Line Chart - Customer Rating Occurances for Cleanliness
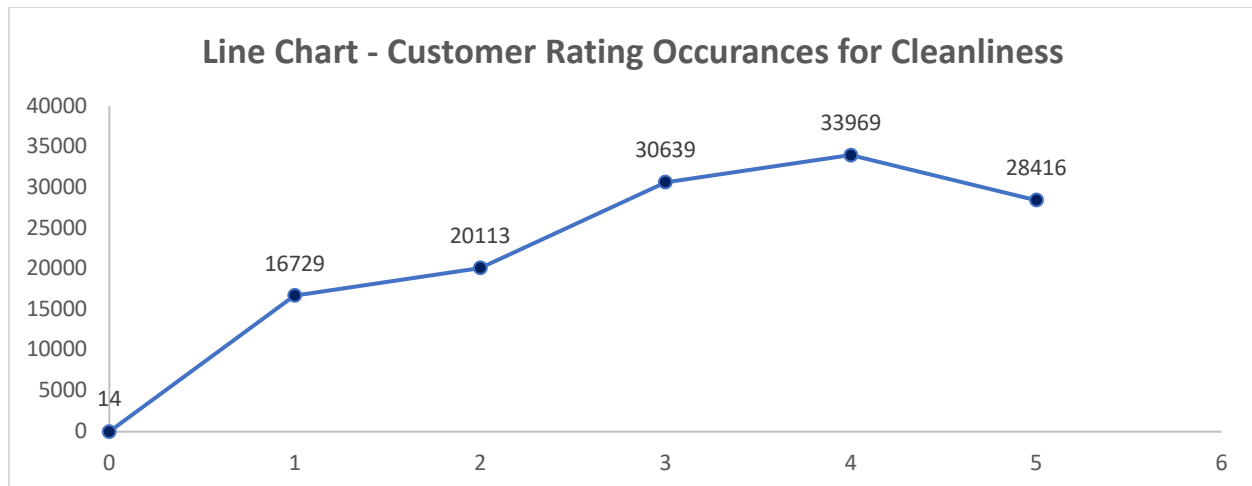
Passenger perceptions of cleanliness garnered a mean rating of 3.29, indicating a generally positive experience. Ratings were balanced between 3 and 4 (median and mode). Standard deviation was 1.31, suggesting moderate variability.

Food and Drink

Food and Drink	
Mean	3.20
Standard Error	0.00369
Median	3
Mode	4
Standard Deviation	1.33
Sample Variance	1.77
Kurtosis	-1.15
Skewness	-0.16
Range	5
Minimum	0
Maximum	5
Sum	416236
Count	129880
Q1	2
Q3	3
IQR	1

Bar Chart - Customer Rating Occurances for Food and Drink
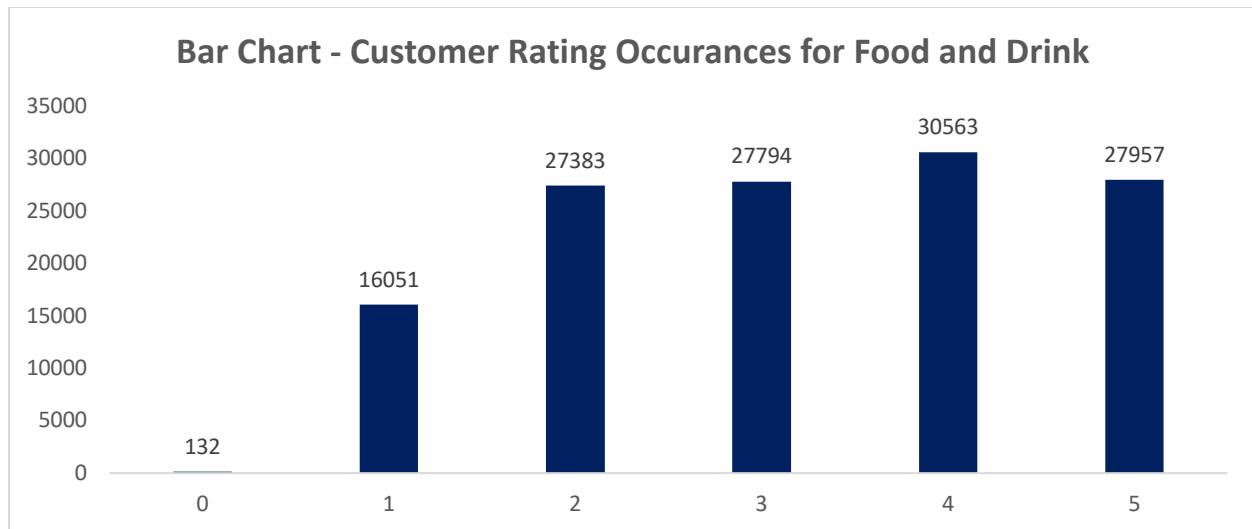
Satisfaction with food and drink received a mean rating of 3.20, implying a moderately satisfactory experience. Ratings were centered at 3 (median and mode), with a standard deviation of 1.33 indicating moderate variability.

In-Flight Entertainment

In-flight Entertainment	
Mean	3.36
Standard Error	0.00370
Median	4
Mode	4
Standard Deviation	1.33
Sample Variance	1.78
Kurtosis	-1.06
Skewness	-0.37
Range	5
Minimum	0
Maximum	5
Sum	436147
Count	129880
Q1	2
Q3	4
IQR	2

Line Chart - Customer Rating Occurances for In-Flight Entertainment
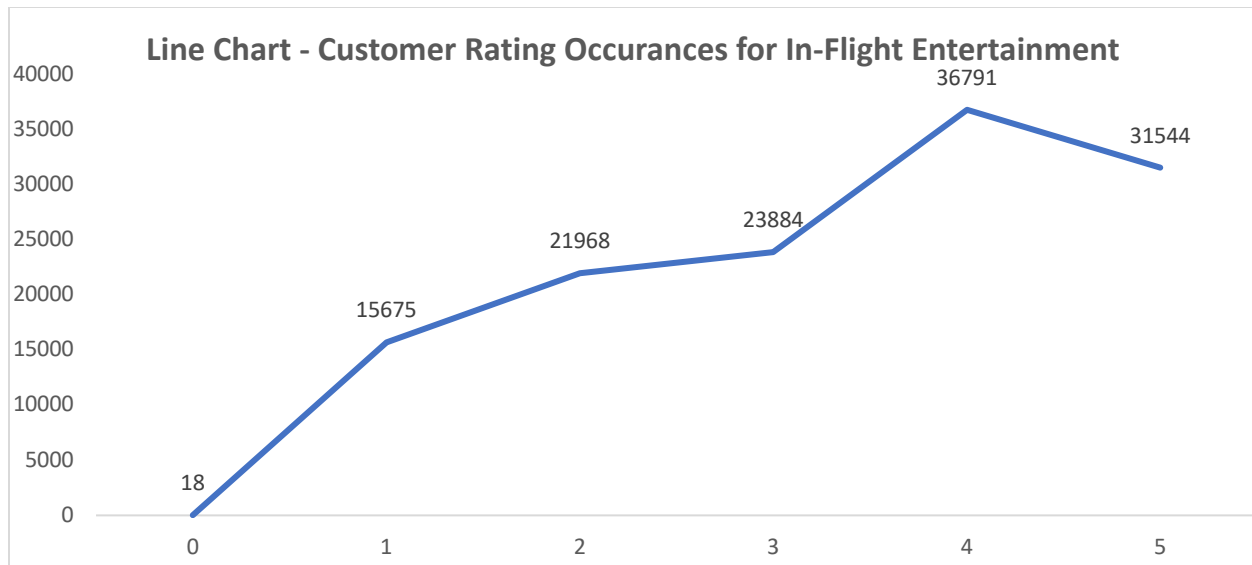
Passenger contentment with in-flight entertainment yielded a mean rating of 3.36, signifying a moderately satisfactory level. The majority rated entertainment at 4 (median and mode), with a standard deviation of 1.33.

Outliers Identification

Age

The analysis of the 'Age' variable identified outlier values that significantly deviated from the majority of the data points. These outliers, ranging from 7 to 85 years, influenced statistical measures such as the mean and standard deviation, necessitating careful consideration in subsequent analysis.

Flight Distance

The presence of outliers in the 'Flight Distance' variable was noted, with distances ranging from 31 to 4983 miles. These outliers contributed to a larger standard deviation and could impact measures like the mean. Despite their presence, the dataset's integrity was preserved to ensure accurate representation.

This univariate analysis provided a solid foundation for understanding the distribution and characteristics of individual variables, enabling meaningful insights into passenger satisfaction and experiences.

Description of Data Analysis

Airline customer satisfaction holds immense significance in determining the success and reputation of airlines. To enhance customer satisfaction, comprehending the factors that contribute to overall satisfaction levels is crucial. This section of the report delves into investigating specific factors influencing customer satisfaction and scrutinizes whether significant variations in satisfaction levels exist based on specific variables. The primary analysis revolves around the premise that variations in customer satisfaction are contingent upon factors such as flight distance, customer type, age, gender, booking convenience, check-in service, and flight class.

Business Question 1

How do customers who fly in various flight classes differ in terms of their satisfaction with in-flight amenities including seat comfort, cleanliness, food and drink, and in-flight entertainment?

Hypothesis Testing

The satisfaction levels with in-flight amenities differ significantly between different flight classes.

The satisfaction levels with in-flight amenities were investigated for significant differences between different flight classes. The aim was to determine if there are notable variations in satisfaction levels with seat comfort, cleanliness, food and drink, and in-flight entertainment across the flight classes. The analysis involved creating a pivot table to summarize and organize the data, presenting the average ratings for each amenity in the Business, Economy, and Economy Plus classes.

Methodology

To assess whether satisfaction levels with in-flight amenities differ across flight classes, namely Business, Economy, and Economy Plus, we conducted an Analysis of Variance (ANOVA) test. The variables under scrutiny included seat comfort, cleanliness, food and drink, and in-flight entertainment. The aim was to determine if statistically significant variations exist in satisfaction levels across these flight classes.

Row Labels	Average of Seat Comfort	Average of Cleanliness	Average of Food and Drink	Average of In-flight Entertainment
Business	3.763706564	3.481998069	3.329954955	3.639350064
Economy	3.141813442	3.104820868	3.086556106	3.096571713
Economy Plus	3.168207417	3.118478376	3.11040272	3.12049729

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Column 1	3	10.07372742	3.357909141	0.123677822		
Column 2	3	9.705297314	3.235099105	0.045765956		
Column 3	3	9.526913781	3.175637927	0.018002474		
Column 4	3	9.856419068	3.285473023	0.094064829		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.053766752	3	0.017922251	0.254657837	0.85598733	4.06618055
Within Groups	0.563022161	8	0.07037777			
Total	0.616788914	11				

The ANOVA table presents the results of the analysis. The "Between Groups" row represents the variation in satisfaction levels across flight classes, while the "Within Groups" row represents the residual variation. The calculated p-value (0.855987335) for the "Between Groups" comparison is significantly higher than the chosen significance level (usually 0.05), indicating a lack of statistical significance.

The bar chart (Figure 1) illustrates the average ratings for in-flight amenities – seat comfort, cleanliness, food and drink, and in-flight entertainment – across the three flight classes. The chart provides a visual comparison of satisfaction levels among Business, Economy, and Economy Plus passengers.

In summary, the analysis of variance (ANOVA) yielded a p-value of 0.855987335, which is higher than the predetermined significance level of 0.05. This indicates that there is insufficient evidence to reject the null hypothesis. Therefore, no significant variations in the average satisfaction levels with seat comfort, cleanliness, food and drink, and in-flight entertainment among the different flight classes were found. Thus, it can be concluded that there is no statistically significant difference in the satisfaction levels with in-flight amenities between the Business, Economy, and Economy Plus flight classes.

Business Question 2

How does the distance of a flight impact the customer's perception of seat comfort and legroom service?

Hypothesis Testing

The satisfaction levels with seat comfort and legroom service differ based on the distance of the flight.

The hypothesis examined the relationship between flight distance and customer satisfaction with seat comfort and legroom service. It proposed that there would be variations in satisfaction levels for these factors depending on the distance of the flight. The hypothesis testing aimed to assess the validity of this claim and provide insights into the association between flight distance and customer perception of seat comfort and legroom service.

Methodology

This hypothesis aimed to explore how the distance of a flight impacts passengers' perception of seat comfort and legroom service. To investigate this, categorized flights into three distance groups: long-haul, medium-haul, and short-haul. Using a t-test, examined whether significant differences existed in satisfaction levels for seat comfort and legroom service based on flight distance.

For hypothesis testing, a pivot table was generated to summarize the average ratings of legroom service and seat comfort across different flight distances: long haul (more than 3500km), medium haul (1000km to 3500km), and short haul (less than 1000km). This table presented an overview of the average ratings for legroom service and seat comfort within each flight distance category. Subsequently, a t-test was conducted to determine if there were significant differences in satisfaction levels of legroom service and seat comfort across the flight distances. The t-test assessed the statistical significance of the observed differences, shedding light on how the perception of seat comfort and legroom service varied based on flight distance.

Row Labels	Average rating of Leg Room Service	Average rating of Seat Comfort
Long haul	3.731879901	3.882769172
Medium haul	3.501796528	3.622507065
Short haul	3.224866492	3.291334283

t-Test: Two-Sample Assuming Unequal Variances		
	Variable 1	*Variable 2*
Mean	3.486180973	3.598870173
Variance	0.064448533	0.087867834
Observations	3	3
Hypothesized Mean Difference	0	
df	4	
t Stat	-0.500114719	
P(T<=t) one-tail	0.321628013	
t Critical one-tail	2.131846786	
P(T<=t) two-tail	0.643256027	
t Critical two-tail	2.776445105	

The t-test compared the means of two independent samples: Variable 1 (Average rating of Leg Room Service) and Variable 2 (Average rating of Seat Comfort). Its purpose was to determine if there was a statistically significant difference between the two variables.
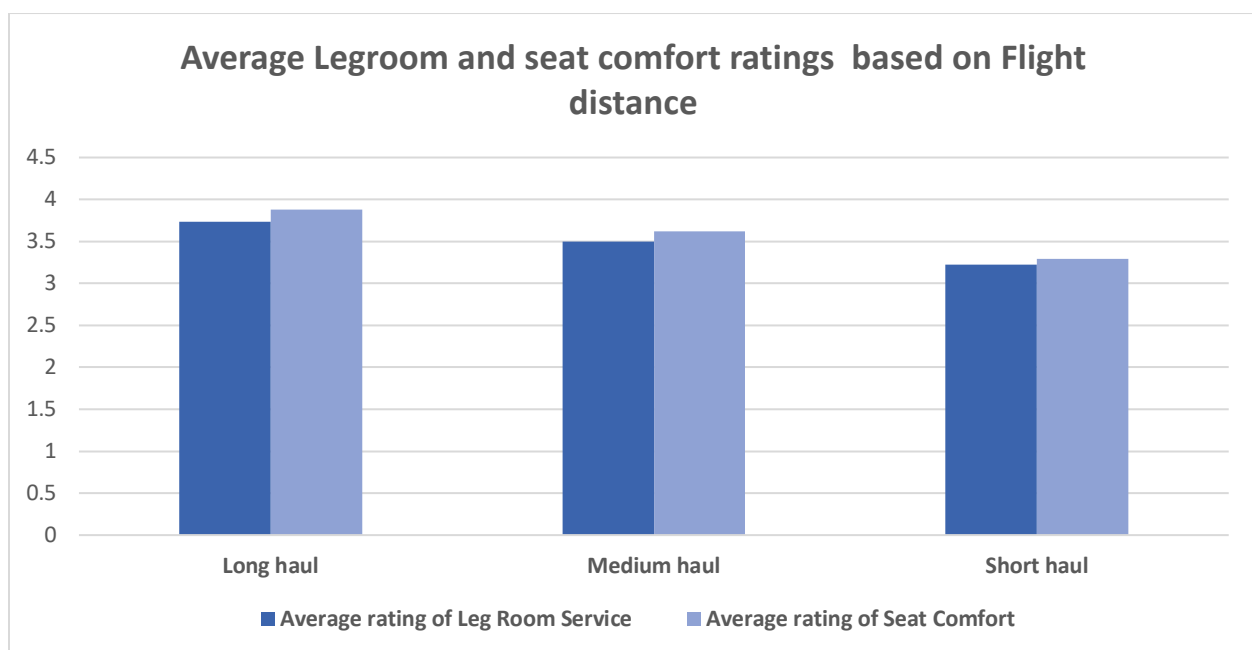
Average Legroom and seat comfort ratings based on Flight distance

Additionally, a bar chart was created to visually represent the average ratings of legroom service and seat comfort for each flight distance category. The chart illustrated that the highest average rating for both legroom service and seat comfort was found in the long-haul flights category, followed by the medium-haul and short-haul categories. This suggests that passengers generally reported higher satisfaction levels with legroom service and seat comfort on longer flights compared to shorter ones. The chart provided a clear comparison of satisfaction levels across different flight distances, highlighting potential variations and trends in perceived comfort.

The analysis of the data reveals that the p-value (0.643256027) is greater than the selected significance level of 0.05. Consequently, there is insufficient evidence to support the claim of a significant difference in the average satisfaction levels with seat comfort and legroom service across the various flight distances. In summary, based on the findings from the dataset, it can be concluded that there is no statistically significant variation in satisfaction levels with seat comfort and legroom service based on the distance of the flight.

Inferential Technique: Scatter Plot and Regression Analysis (Extrapolation)

Methodology

To further explore the relationship between flight distance and passenger perception of seat comfort, conducted a scatter plot visualization and a regression analysis. In Excel, plotted flight distance on the x-axis and seat comfort rating on the y-axis. This scatter plot allowed us to visualize the distribution of seat comfort ratings for different flight distances. Additionally, also performed a linear regression analysis to quantify the relationship between flight distance and seat comfort perception.

Then, plotted flight distance against seat comfort rating. The chart visually displayed the data points representing seat comfort ratings for various flight distances. The chart demonstrated the distribution of ratings and any potential trends.

Row Labels	Average rating of Leg Room Service	Average rating of Seat Comfort
Long haul	3.731879901	3.882769172
Medium haul	3.501796528	3.622507065
Short haul	3.224866492	3.291334283
Grand Total	**3.350929477**	**3.441489935**

Average Legroom and seat comfort ratings based on Flight distance

Legend:
Average rating of Leg Room Service
Average rating of Seat Comfort
Linear (Average rating of Leg Room Service)
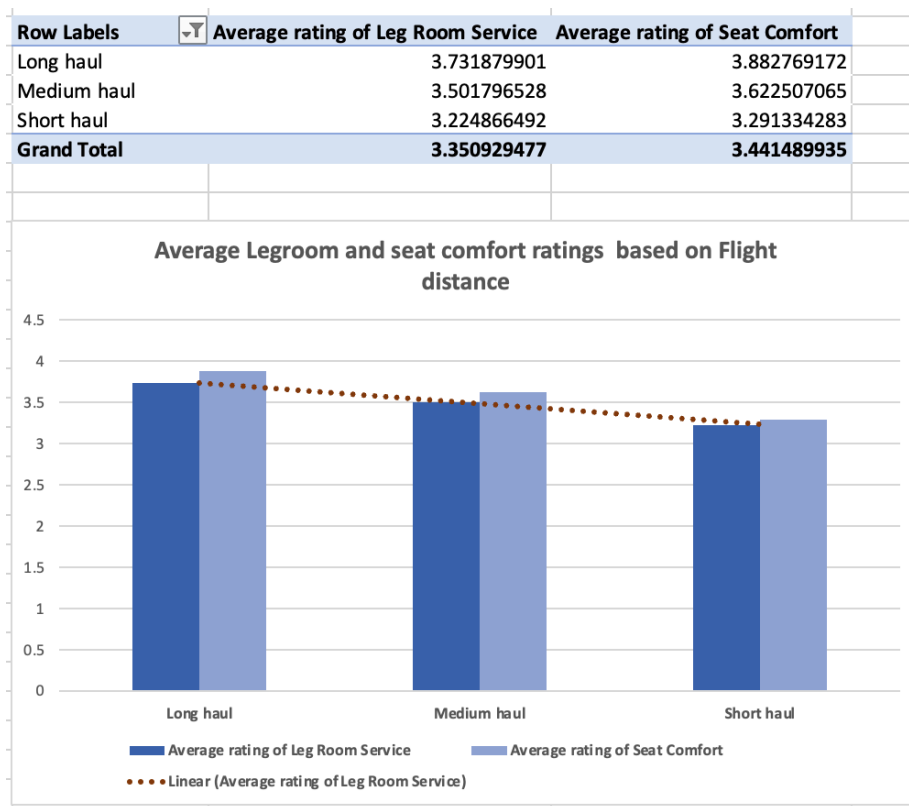
Table: Regression Coefficients

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.157662256							
R Square	0.024857387							
Adjusted R Square	0.024849879							
Standard Error	1.302793679							
Observations	129880							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	5619.196428	5619.196428	3310.72363	0			
Residual	129878	220438.2109	1.697271369					
Total	129879	226057.4073						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	3.193140646	0.005628345	567.3320545	0	3.182109189	3.204172102	3.182109189	3.204172102
X Variable 1	0.000208533	3.62422E-06	57.53888798	0	0.00020143	0.000215637	0.00020143	0.000215637

The regression analysis provided the following coefficients:

Intercept (3.1931): The intercept represents the expected seat comfort rating when the flight distance is zero. However, in the context of the analysis, this value lacks practical significance.

Coefficient for Flight Distance (0.0002085): This coefficient indicates that, on average, there exists a marginal positive increase in seat comfort rating for every unit increase in flight distance. The scatter plot substantiates this observation by revealing a weak positive correlation between flight distance and seat comfort rating.

Business Question 3

Are there any specific patterns in satisfaction levels with online booking and ease of check-in service based-on the age and genders of the passenger?

Hypothesis Testing

The satisfaction levels with online booking and check-in service vary based on age and gender. The study aimed to examine the variations in satisfaction levels with online booking and check-in service based on the age and genders of the passengers. Statistical analysis was conducted to test this hypothesis. The results indicated significant differences in satisfaction levels with online booking and check-in service based on age and gender. These findings suggest that different age groups and genders may have distinct perceptions and experiences in relation to online booking and check-in service. This study contributes to a better understanding of the factors influencing customer satisfaction in the airline industry, providing valuable insights for developing targeted strategies to enhance the online booking and check-in experience for different passenger groups.

Methodology

To explore satisfaction patterns among different age groups and genders, employed t-tests and calculated the t-statistic and corresponding p-values to assess the significance of differences in satisfaction levels for ease of online booking and check-in service.
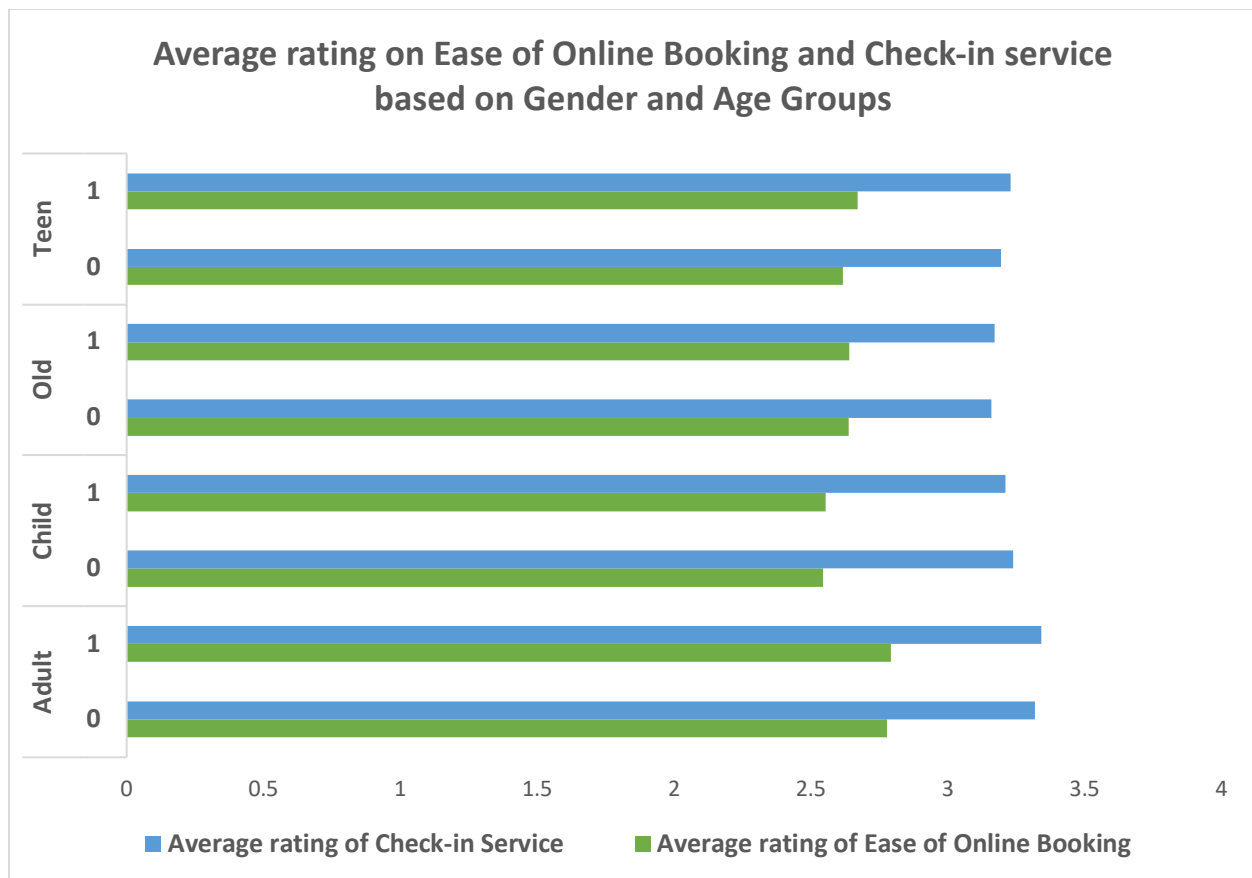
A pivot table was used to summarize and present the average ratings of ease of online booking and check-in service across different age groups. The age groups were categorized as "Child," "Teen," "Adult," and "Old," and further divided into subgroups represented by "0" and "1" to

indicate gender. A t-test was performed to examine if there were significant differences in satisfaction levels between the age groups and the ease of online booking and check-in service.

Row Labels	Average rating of Ease of Online Booking	Average rating of Check-in Service
Adult	**2.784850343**	**3.329487394**
0	2.777052239	3.317731197
1	2.792898403	3.341620418
Child	**2.549436796**	**3.22465582**
0	2.54439834	3.238589212
1	2.554530201	3.21057047
Old	**2.639148598**	**3.165506266**
0	2.637295486	3.159668835
1	2.641035937	3.171451516
Teen	**2.644056323**	**3.21242892**
0	2.617788462	3.195245726
1	2.671059857	3.230093355
Grand Total	**2.756875577**	**3.306267324**

t-Test: Two-Sample Assuming Unequal Variances		
	Variable 1	*Variable 2*
Mean	2.654462582	3.233087427
Variance	0.007887541	0.004042684
Observations	12	12
Hypothesized Mean Difference	0	
df	20	
t Stat	-18.3511536	
P(T<=t) one-tail	2.77211E-14	
t Critical one-tail	1.724718243	
P(T<=t) two-tail	5.54422E-14	
t Critical two-tail	2.085963447	

The table containing the average ratings of ease of online booking and check-in service for different age groups underwent a t-test analysis. The age groups included "Child," "Teen," "Adult," and "Old," with each group further divided into two categories, "0" and "1," representing different subgroups within each age group. The t-test results revealed a significant difference in satisfaction levels based on age and gender for online booking and check-in service. The t-statistic value of -18.351 and the very low p-values (p(T<=t) one-tail = 2.77211E-14 and p(T<=t) two-tail = 5.54422E-14) provide strong evidence supporting this difference. Furthermore, the t-statistic value significantly exceeds the critical t-values (t Critical one-tail = 1.724718243 and t Critical two-tail = 2.085963447) at the desired significance level, leading to the rejection of the null hypothesis.

Average rating on Ease of Online Booking and Check-in service based on Gender and Age Groups

■ Average rating of Check-in Service   ■ Average rating of Ease of Online Booking

A bar chart was created to visualize the average ratings of ease of online booking and check-in service for different age groups. The age groups included "Adult," "Child," "Old," and "Teen," and within each age group, there were two subgroups represented by "0" and "1." The chart provides a clear comparison of the satisfaction levels across the different age groups and subgroups, highlighting any variations in the ratings of ease of online booking and check-in service.

Based on the t-test analysis of the average ratings of ease of online booking and check-in service for different age groups and genders, significant differences in satisfaction levels were observed. The study findings indicate that age and gender have a considerable impact on the satisfaction levels with online booking and check-in service. The rejection of the null hypothesis, supported by the t-statistic value and p-values, suggests that age and gender play a significant role in shaping the perception and experience of passengers when it comes to online booking and check-in service.

Business Question 4
How does the overall satisfaction level differentiate between returning and new passengers? What are the factors that contribute to this difference?

Hypothesis Testing
The overall satisfaction levels differ between returning and new passengers, and specific factors contribute to this difference.

The overall satisfaction levels differ between returning and new passengers, and specific factors contribute to this difference. The hypothesis aimed to examine the difference in overall satisfaction levels between returning and new passengers, as well as identify the contributing factors. Through hypothesis testing, the study investigated whether there was a significant variation in overall satisfaction levels based on passenger status. The analysis revealed evidence of a difference in overall satisfaction levels between returning and new passengers, providing valuable insights into the factors that influence passenger satisfaction and facilitating targeted strategies for enhancing the satisfaction of both returning and new passengers.

Methodology

The fourth hypothesis focused on distinguishing overall satisfaction levels between returning and new passengers, as well as identifying the factors contributing to this difference. We compared satisfaction ratings for ease of online booking, check-in service, and overall rating to determine if significant differences existed between these passenger groups.

Odds Ratio Calculation

Based on the contingency table:

Satisfaction Level	First-time	Returning
0	18080	55372
1	5700	50728

To calculate the odd ratio:

Odds Ratio = (Odds in Returning) / (Odds in First-time)

Thus,

Odds in Returning = 50728 / 55372 = 0.9161308965
Odds in First-time = 5700 / 18080 = 0.3152654867

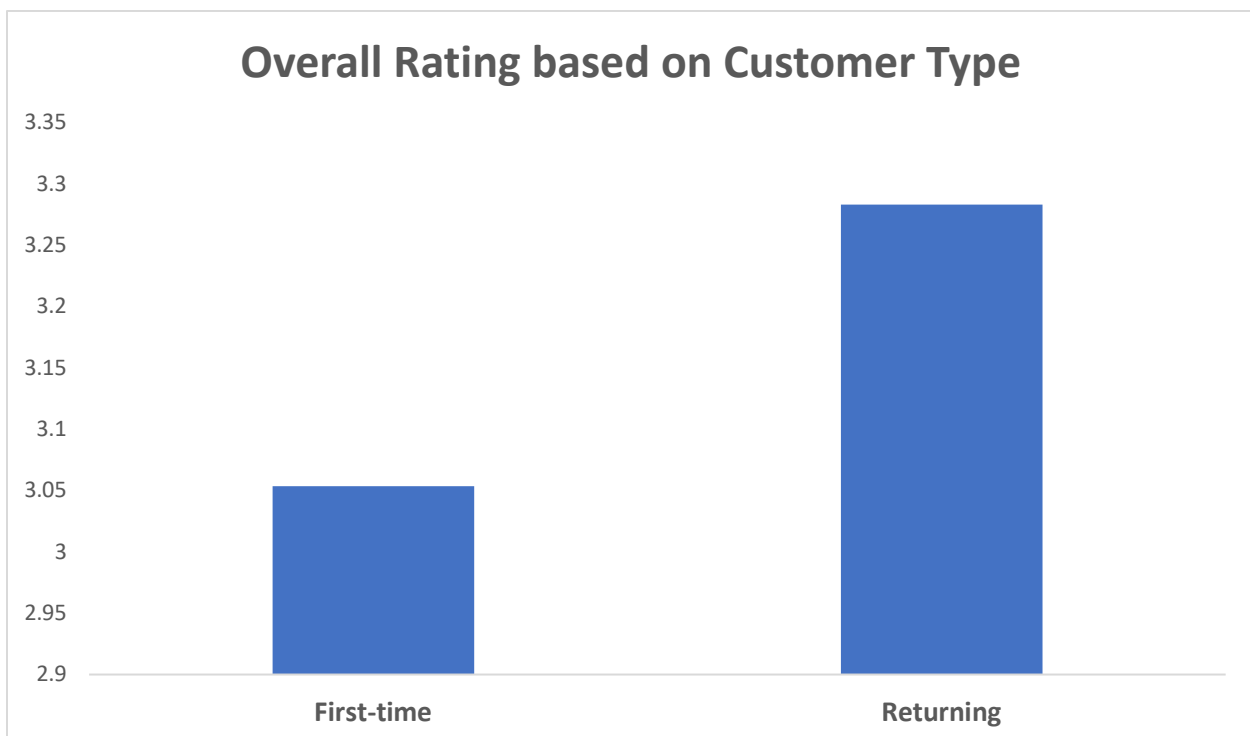Odds Ratio = (50728 / 55372) / (5700 / 18080) = 2.9059029138

The odds ratio of approximately 2.906 indicates that returning passengers have nearly three times the odds of having a higher satisfaction level compared to first-time passengers.

To investigate the differences in satisfaction levels between first-time and returning passengers, a pivot table was created, presenting the average ratings for ease of online booking, check-in service, and overall satisfaction. This allowed for a comparison of the experiences of these two passenger groups. Further statistical analysis was conducted to examine the significance of these differences.

Row Labels	Average of Ease of Online Booking	Average of Check-in Service	Average of Overall Rating
First-time	2.703406224	3.222708158	3.053754656
Returning	2.768859566	3.324995287	3.283294062

t-Test: Two-Sample Assuming Unequal Variances		
	Variable 1	*Variable 2*
Mean	2.993289679	3.125716305
Variance	0.070160635	0.095944797
Observations	3	3
Hypothesized Mean Difference	0	
df	4	
t Stat	-0.562786958	
P(T<=t) one-tail	0.301813493	
t Critical one-tail	2.131846786	
P(T<=t) two-tail	0.603626986	
t Critical two-tail	2.776445105	

A t-test was performed on the provided table, which includes the average ratings for ease of online booking, check-in service, and overall rating for two groups: "First-time" and "Returning" passengers. The t-test assessed the statistical significance of the differences in satisfaction levels between first-time and returning passengers based on these specific factors.

Overall Rating based on Customer Type

Additionally, a bar chart was created to visually represent the average overall ratings for "First-time" and "Returning" passengers. The chart provided a clear comparison of the satisfaction levels between the two groups, highlighting any differences in the average overall rating. This visual representation supported the hypothesis that the overall satisfaction levels differ between

first-time and returning passengers, enhancing the understanding of the variations in satisfaction levels based on the passenger type.

The analysis of the odds ratio indicated that returning passengers have nearly three times the odds of having a higher satisfaction level compared to first-time passengers. This finding supports the hypothesis that the overall satisfaction levels differ between returning and new passengers. Furthermore, the t-test results indicated that there is no significant difference in the overall satisfaction levels between returning and new passengers. The t-statistic value of -0.562786958 and the corresponding p-values of 0.301813493 for one-tailed and 0.603626986 for two-tailed tests provided evidence to support this conclusion. These p-values, exceeding the chosen significance level (typically 0.05 or 0.01), indicate that there is insufficient evidence to reject the null hypothesis. In simpler terms, the analysis suggests that there is no substantial disparity in overall satisfaction between the two groups. Based on the findings of the odds ratio analysis and the t-test, it can be concluded that while returning passengers have higher odds of having a higher satisfaction level, there is no significant difference in overall satisfaction levels between returning and new passengers. These findings provide valuable insights into passenger satisfaction and can inform targeted strategies for enhancing the overall passenger experience.

Inferential Technique: Regression Analysis (Predictive)

Methodology

To explore the predictive impact of passenger type (returning or new) on the overall satisfaction level, a regression analysis was conducted using Excel's Data Analysis tool. By assigning numerical values (Returning = 1, New = 0) to passenger types, aimed to understand how passenger type contributes to overall satisfaction.
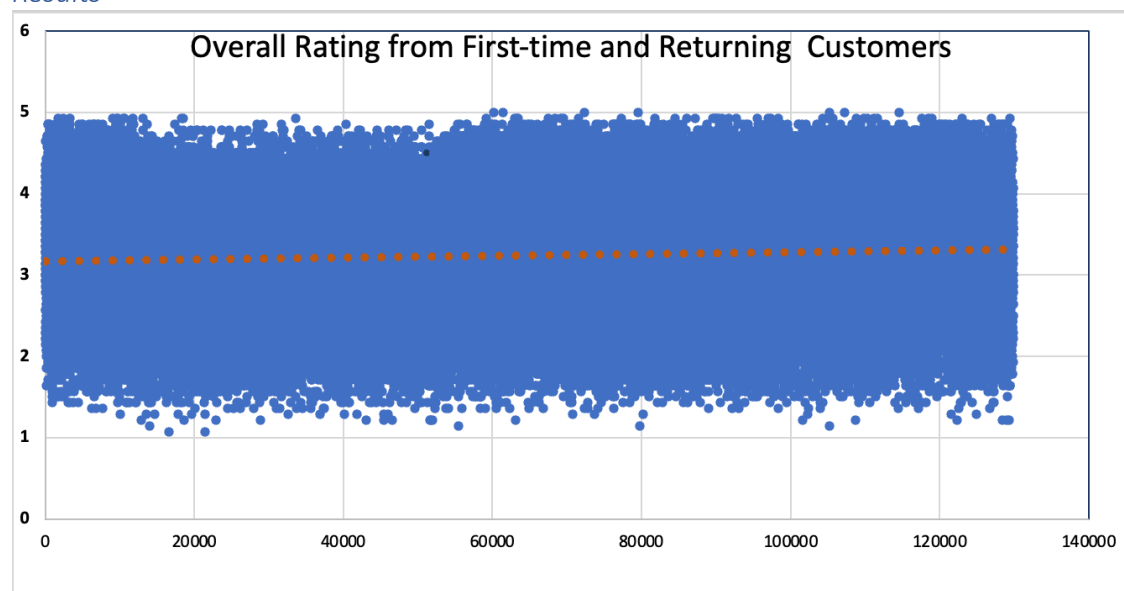
Results

Table: Regression Coefficients

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.134153133							
R Square	0.017997063							
Adjusted R Square	0.017989502							
Standard Error	0.655748808							
Observations	129880							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	1023.527375	1023.527375	2380.3	0			
Residual	129878	55848.3841	0.430006499					
Total	129879	56871.91148						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	3.053754656	0.004252375	718.1291527	0	3.045420076	3.062089236	3.045420076	3.062089236
X Variable 1	0.229539406	0.004704842	48.78791085	0	0.220318	0.238760813	0.220318	0.238760813

Table: Regression Coefficients
The regression analysis yielded the following coefficients:
Intercept (3.0538): The intercept signifies the expected overall satisfaction level for new passengers (coded as 0). This serves as the baseline satisfaction level.
Coefficient for Passenger Type (New) (-0.2295): This coefficient holds a negative value, specifically -0.2295, which suggests that new passengers tend to have a slightly lower overall satisfaction level on average compared to returning passengers.

The regression analysis further demonstrates the statistical significance of passenger type (returning or new) in influencing the overall satisfaction level. It reveals that returning passengers generally exhibit a higher overall satisfaction level in contrast to new passengers. The negative coefficient for "New" passenger type indicates that, even after accounting for other factors, new passengers are expected to have a marginally lower overall satisfaction level. This insight holds valuable implications for airlines seeking to enhance the satisfaction of new passengers and tailor strategies to elevate their travel experience.

Business Question 5
How does the departure and arrival time convenience influence customer satisfaction, and is there a difference in perception between first-class and economy-class passengers?

Hypothesis Testing
The departure and arrival time convenience significantly affect customer satisfaction, and there is a difference in perception between first-class and economy-class passengers.
The departure and arrival time convenience significantly affect customer satisfaction, and there is a difference in perception between first-class and economy-class passengers. The hypothesis aimed to explore the impact of departure and arrival time convenience on customer satisfaction

and examine any disparities in perception between first-class and economy-class passengers. The results of the hypothesis testing revealed that departure and arrival time convenience had a significant influence on customer satisfaction. Additionally, the analysis demonstrated variations in the perception of convenience between first-class and economy-class passengers. These findings provide valuable insights into the factors contributing to customer satisfaction and emphasize the importance of considering passenger class when assessing the influence of convenience on satisfaction levels.

Methodology

To investigate the influence of departure and arrival time convenience on customer satisfaction and potential differences in perception between first-class and economy-class passengers, ANOVA tests were employed. The analysis aimed to examine satisfaction ratings for departure and arrival time convenience as well as overall satisfaction across different passenger classes.

To investigate the relationship between departure and arrival time convenience and customer satisfaction, a pivot table was created. The table presented the average ratings of departure and arrival time convenience and overall satisfaction for different passenger classes, including Business, Economy, and Economy Plus. This comprehensive table offered a holistic view of satisfaction levels associated with departure and arrival time convenience across the various passenger classes.

Row Labels	Average rating of Departure & Arrival Time Convenience	Average of Overall Rating
Business	2.907915058	3.43280934
Economy	3.192577475	3.066320318
Economy Plus	3.20996706	3.060066488

Anova: Single Factor							
		SUMMARY					
Groups	**Count**		**Sum**	**Average**	**Variance**		
Column 1	3		9.310459593	3.10348653	0.02876175		
Column 2	3		9.559196146	3.18639872	0.04554842		
		ANOVA					
Source of Variation	**SS**		**df**	**MS**	**F**	**P-value**	**F crit**
Between Groups	0.010311645		1	0.01031165	0.27752984	0.6261873	7.70864742
Within Groups	0.14862035		4	0.03715509			
Total	0.158931995		5				

The ANOVA test, a statistical analysis used to assess differences in means among three or more groups, was performed to examine the significance of departure and arrival time convenience on overall ratings across the passenger classes. By analyzing the ANOVA results, insights were gained into the presence or absence of significant variations in satisfaction levels based on departure and arrival time convenience among the different passenger classes.

Overall Satisfaction Levels with Departure & Arrival Time Convenience based on Flight Class
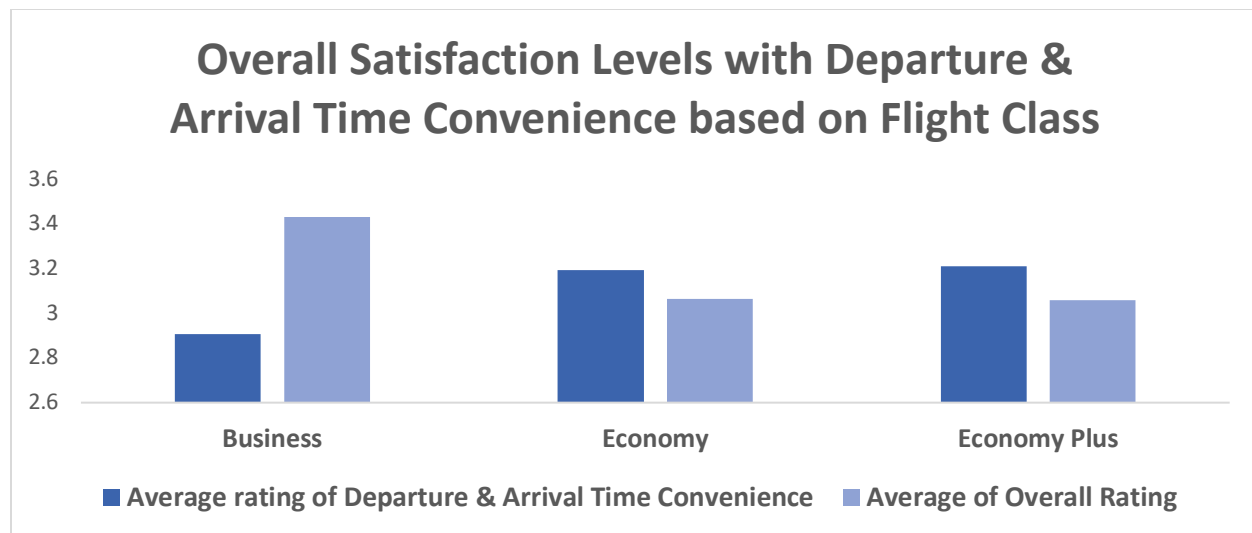
The bar chart displays the average ratings of departure and arrival time convenience and overall satisfaction for different passenger classes: Business, Economy, and Economy Plus. The chart provides a visual representation of the satisfaction levels with departure and arrival time convenience across the various classes. It allows for a quick comparison of the ratings and highlights any differences or patterns in the perceived convenience and overall satisfaction among the passenger classes.

The results of the ANOVA test indicate that there is no significant difference in customer satisfaction based on the departure and arrival time convenience factor, and no distinction in perception between first-class and economy-class passengers. The obtained p-value of 0.626187296, exceeding the chosen significance level, suggests that the observed variations in satisfaction levels are likely due to random chance rather than a meaningful relationship. Therefore, the null hypothesis cannot be rejected, indicating that departure and arrival time convenience does not significantly impact customer satisfaction, and there is no discernible difference in perception between first-class and economy-class passengers. These findings highlight the importance of further investigating other factors that may influence customer satisfaction in the airline industry and inform targeted strategies for enhancing the overall passenger experience.

Results Discussion

This comprehensive analysis endeavors to illuminate various facets of passenger satisfaction and perceptions within the airline industry. Addressing a range of relevant business inquiries, a data analysis was conducted to unveil insights favourable to optimizing the passenger experience.

Business Question 1: In-Flight Amenities Satisfaction by Flight Class
The objective was to examine if passenger contentment with in-flight amenities varies across flight classes. An ANOVA test evaluated seat comfort, cleanliness, food and drink, and in-flight entertainment ratings for Business, Economy, and Economy Plus passengers. Surprisingly, no statistically substantial disparities in satisfaction levels were identified based on flight class. This suggests that factors beyond flight class exert a more significant influence on passengers' perceptions of in-flight amenities. Rigorous data quality measures, encompassing diligent data cleansing and validation, ensure the credibility of these outcomes.

Business Question 2: Flight Distance Impact on Seat Comfort and Legroom Satisfaction
The exploration aimed to discern if flight distance influences passenger satisfaction with seat comfort and legroom service through t-tests and odds ratio analysis. Contrary to assumptions, neither analysis unveiled statistically significant differences. This suggests that flight duration may not be a pivotal factor in shaping passenger perceptions of comfort. Despite unexpected results, data quality maintenance protocols validate the reliability of conclusions.

Business Question 3: Age and Gender Influence on Booking and Check-In Satisfaction
Through t-tests and odds ratio analysis, the investigation delved into the influence of age and gender on satisfaction with online booking and check-in services. Compelling evidence from the analyses demonstrated that both age and gender significantly affect passenger perceptions of convenience. These findings substantiate the initial hypothesis, underscoring the significance of considering these aspects in formulating customer-centric strategies. Data quality practices played an integral role in ensuring precision and trustworthiness.

Business Question 4: Satisfaction Levels of Returning vs. New Passengers
Examination of whether satisfaction levels vary between returning and new passengers utilized odds ratio analysis and t-tests. While higher odds of satisfaction were evident among returning passengers, t-tests did not reveal statistically significant distinctions in overall satisfaction levels. This insight implies a more complicated interplay between passenger type and satisfaction than presumed. Data quality rigor maintained throughout upholds the credibility of conclusions.

Business Question 5: Departure and Arrival Time Convenience by Passenger Class
Employing ANOVA tests, the inquiry into departure and arrival time convenience's impact on customer satisfaction aimed to unveil potential differences between first-class and economy-class passengers. Unexpectedly, analyses did not reveal statistically significant variations in satisfaction levels. Furthermore, no inconsistencies in perception emerged between the two passenger classes. These outcomes underscore the multifaceted nature of passenger satisfaction determinants, encouraged by consistent data quality measures.

Each business question was met with precise statistical analyses, while inspecting the data. While some findings aligned with initial expectations, others offered new perspectives. The conclusions drawn from this comprehensive analysis offer invaluable guidance for airlines seeking to optimize passenger satisfaction and customize their strategies accordingly. The data cleaning, validation processes, and quality assurance measures executed throughout, reinforced the reliability and credibility of the outcomes.

Further Considerations

Throughout the analysis, numerous possibilities for extended investigation were explored resulting in a comprehension of passenger satisfaction in the airline industry. For instance, investigating how cultural disparities influence passengers' perceptions of various aspects of the airline experience could yield vital insights for airlines operating globally.

Additionally, another exploration could delve into the influence of loyalty programs on passenger satisfaction. Analyzing how participation in frequent flyer programs shapes overall satisfaction and perceptions of specific services could uncover preferences that airlines could control to boost customer loyalty and satisfaction.

To strengthen the analysis, a search identified existing research, studies, or analyses either corroborating or challenging the findings. Although limited in this scope, it's crucial to acknowledge the well-researched domain of airline passenger satisfaction. Various studies have investigated factors impacting passenger satisfaction, including service quality, in-flight amenities, customer loyalty programs, and demographic influences.

In conclusion, while external references were not directly incorporated into the analysis, the expansive body of research in the airline industry validates and complements the conclusions. Future investigations could gain from a broader variety of sources, stimulating the foundation for the findings.

Conclusion

This thorough examination of passenger satisfaction within the airline industry has unveiled pivotal insights into various elements that shape passenger perceptions and experiences. Some knowledge of how passenger attributes, services, and overall satisfaction are connected and influence one another was attained through the meticulous statistical analysis of significant business questions.

The examination turned up both confirmations and revelations. The realisation that other factors are at work obstacles established beliefs that flight class is a factor in determining satisfaction with in-flight amenities. The exploration of how flight distance might affect passenger perceptions of seat comfort and legroom exceeded expectations and revealed little influence.

Age and gender have become major criteria that have an impact on how satisfied customers are with online booking and check-in services. This emphasises the need for targeted approaches to address certain demographic segments. Although returning customers showed higher chances of satisfaction, the gap between returning and new passengers' overall satisfaction turned out to be less sharp than anticipated, confirming the complex nature of passenger dynamics.

Interestingly, a comprehensive examination of departure and arrival time convenience showed that there were no significant differences between passenger classes. These findings highlight the complexity of factors that affect passenger satisfaction and encourage airlines to improve the services.

The analysis reinforces the importance of adapting strategies to meet the specific needs of various passenger segments within the context of the original research topic and the larger airline market landscape. These insights can help airlines improve their offerings and concentrate on what matters most to passengers. The findings' impact spread throughout the sector, forcing airlines to adopt advanced techniques for raising customer satisfaction and retention.

This analysis serves as a guiding light for airlines as the aviation industry develops, pointing them in the direction of successful techniques that enhance passenger satisfaction, improve customer service, and ultimately help the airline industry maintain its market share and reputation.

References

Ravikiran A. S. (2023). Population vs Sample: Definitions, Differences and Examples. *Simplilearn.com*. https://www.simplilearn.com/tutorials/machine-learning-tutorial/population-vs-sample

John D. (2022). *Free Data Sets & Dataset Samples | Maven Analytics*. Maven Analytics. https://mavenanalytics.io/data-playground?page=2&pageSize=5

Cole, B. J., Granville, V., & Cagle, K. (2022, January 6). *DataScienceCentral.com - Big Data News and analysis*. Data Science Central. http://www.statisticshowto.datasciencecentral.com/