

# Capstone Project



## Airline Passenger

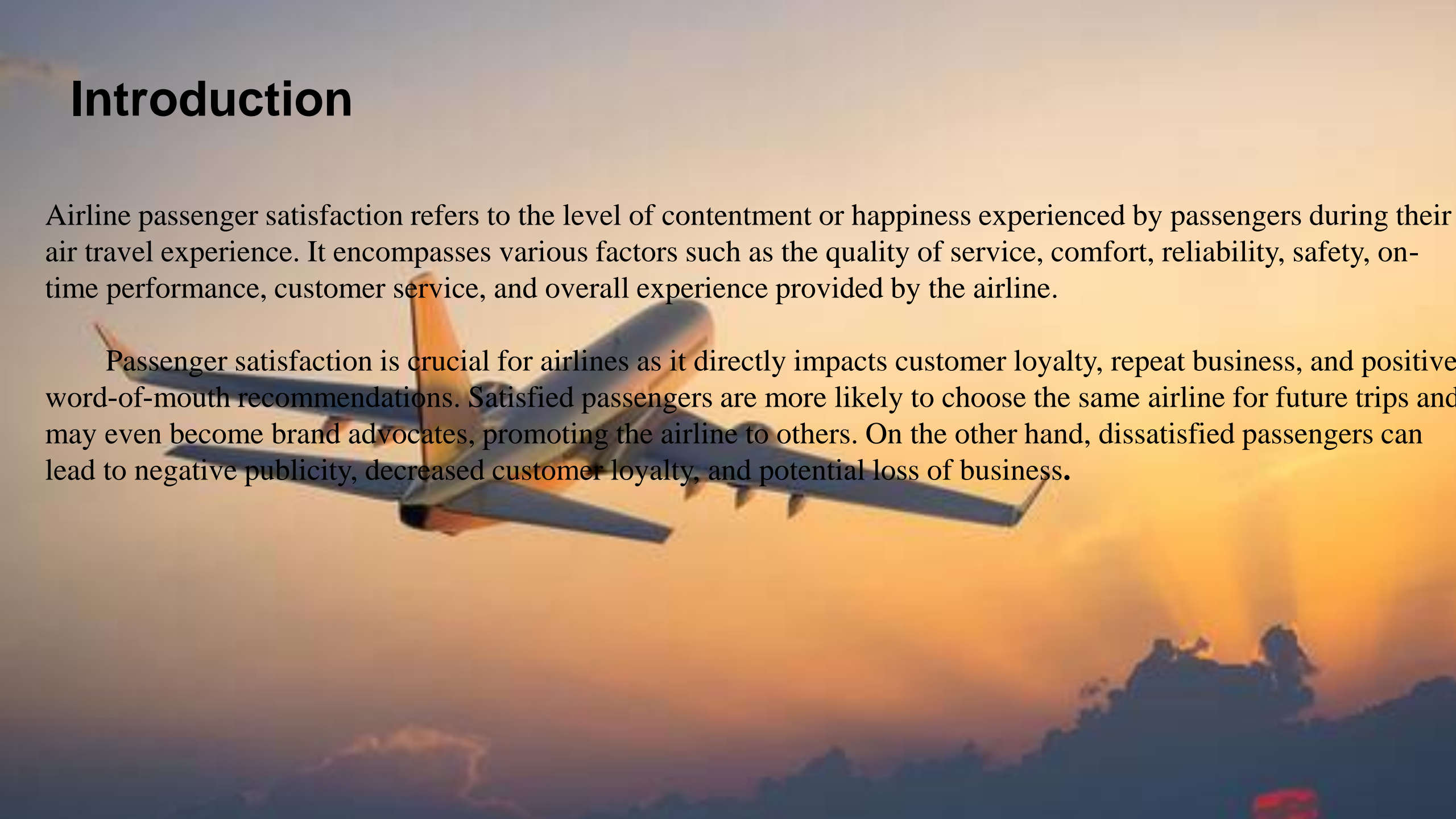
## Satisfaction

Aswini Kumar Motapothula

# Introduction

Airline passenger satisfaction refers to the level of contentment or happiness experienced by passengers during their air travel experience. It encompasses various factors such as the quality of service, comfort, reliability, safety, on-time performance, customer service, and overall experience provided by the airline.

Passenger satisfaction is crucial for airlines as it directly impacts customer loyalty, repeat business, and positive word-of-mouth recommendations. Satisfied passengers are more likely to choose the same airline for future trips and may even become brand advocates, promoting the airline to others. On the other hand, dissatisfied passengers can lead to negative publicity, decreased customer loyalty, and potential loss of business.



# Column Information

**Gender:** Gender of the passengers (Female, Male)

**Customer Type:** The customer type (Loyal customer, disloyal customer)

**Age:** The 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 wifi service:** Satisfaction level of the inflight wifi service (0:Not Applicable;1-5)

**Departure/Arrival time convenient:** Satisfaction level of Departure/Arrival time convenient  
**Of Seat comfort**

**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

**Leg room service:** Satisfaction level of Leg room 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)

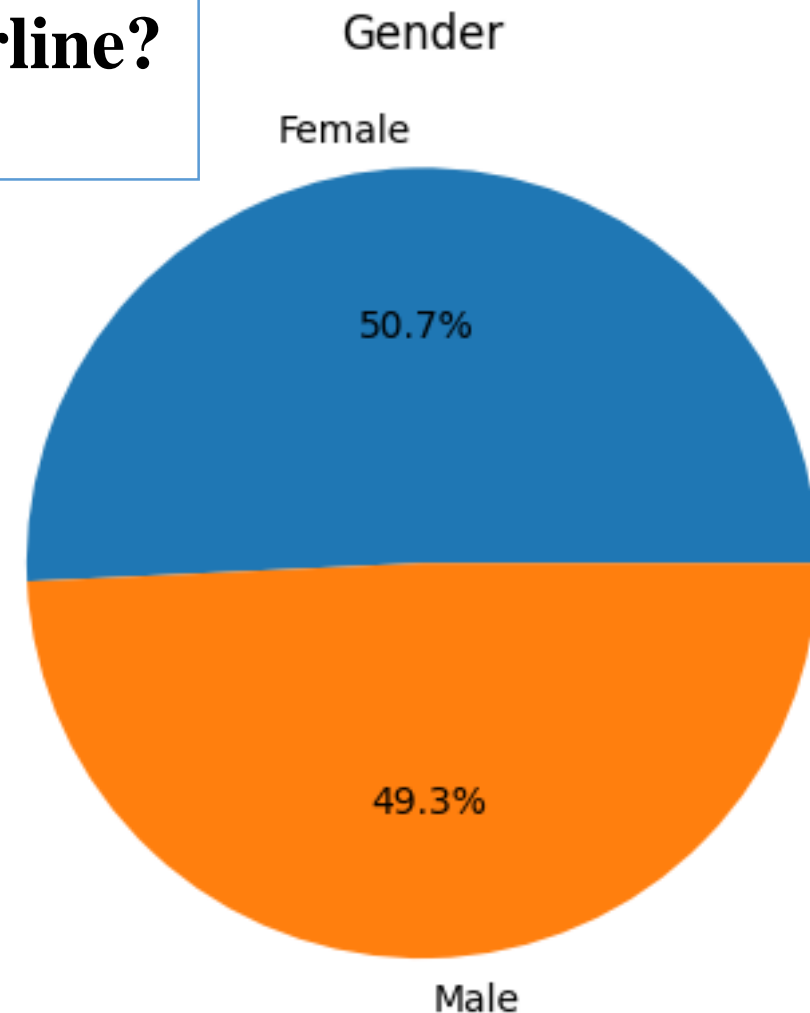
# Business Problem

An international airline company aims to predict the key factors influencing passenger satisfaction and understand the distribution of passenger satisfaction levels (satisfied, neutral or dissatisfied). By addressing these objectives, the airline company seeks to enhance its services and improve overall customer satisfaction



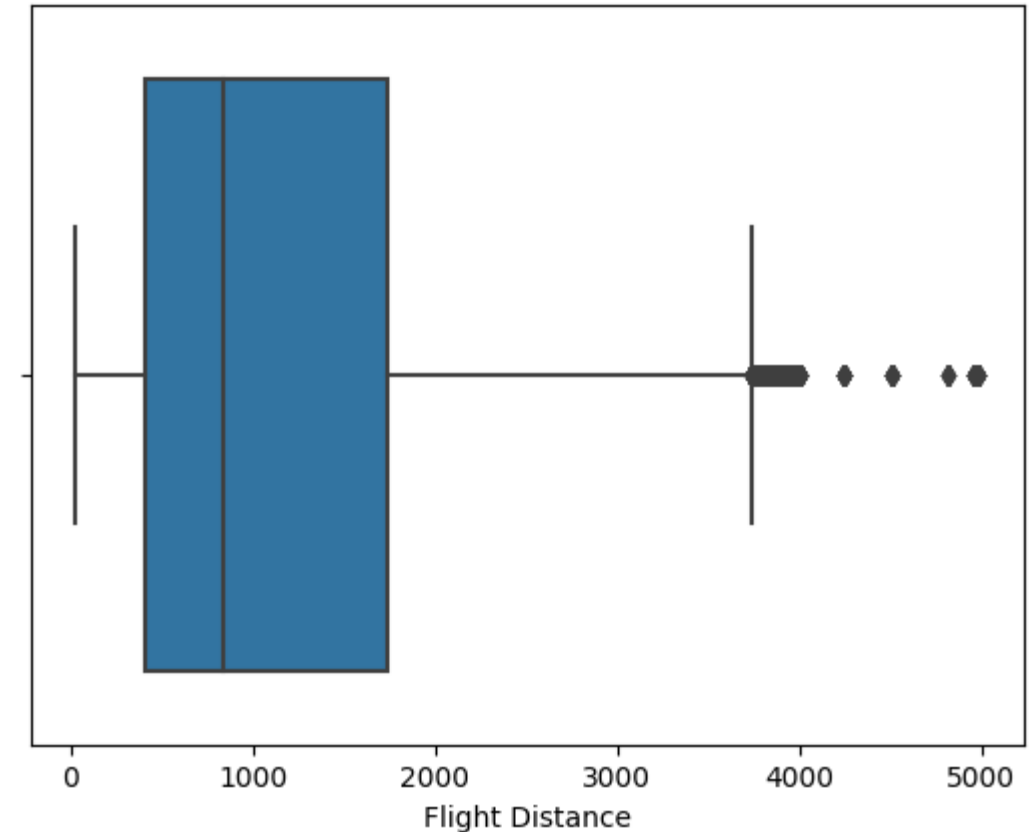
## Which gender travels most frequently on this airline?

- According to our data, the female percentage of travelers on this airline is slightly higher than the male percentage
- .Specifically, females make up approximately 50.7% of the total travelers, while males account for approximately 49.3%.
- It is important to note that these percentages are based on the available data and may vary over time.

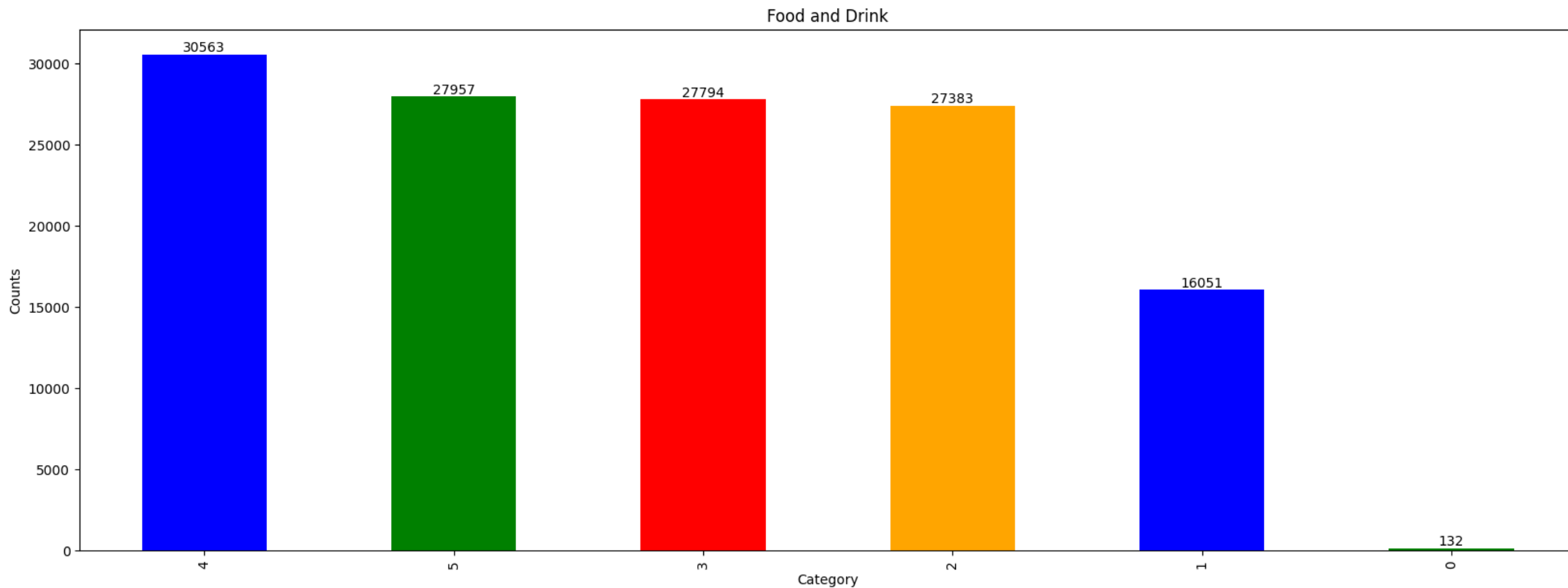


## Box Plot of Flight Travel Distance

- Box Plot Analysis: Flight Travel Distance Distribution (500-1800 km)
- Median Distance is 800 km
- Outliers Ranging from 4000 to 5000 km these distances happened in the rare cases.



# Rating of Food and Drink

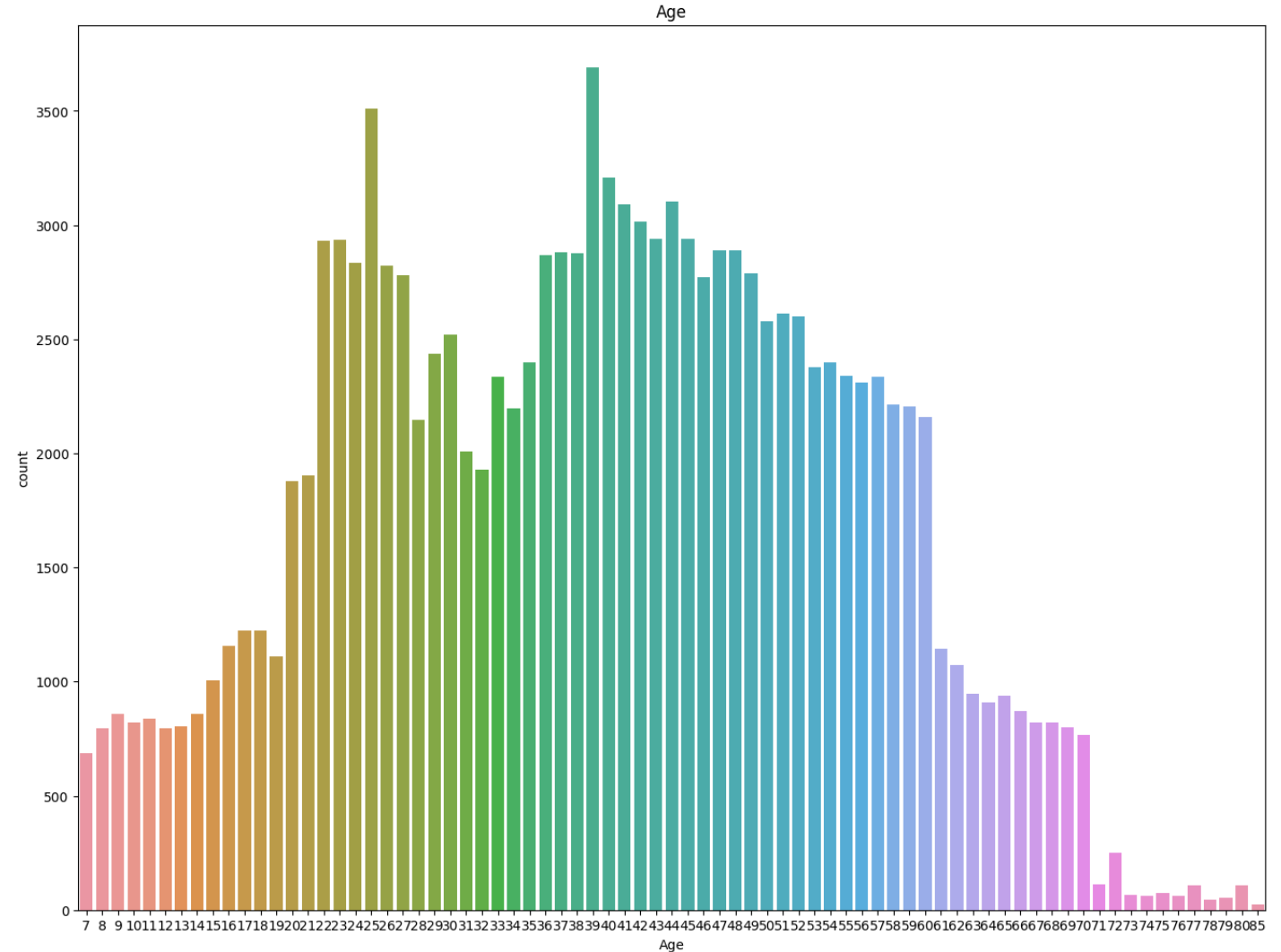


- "The Highest Rating for Food and Drink: 4"
- "30563 Passengers Rated Food and Drink with a Score of 4"
- "132 Passengers Rated Food and Drink with a Score of 0"

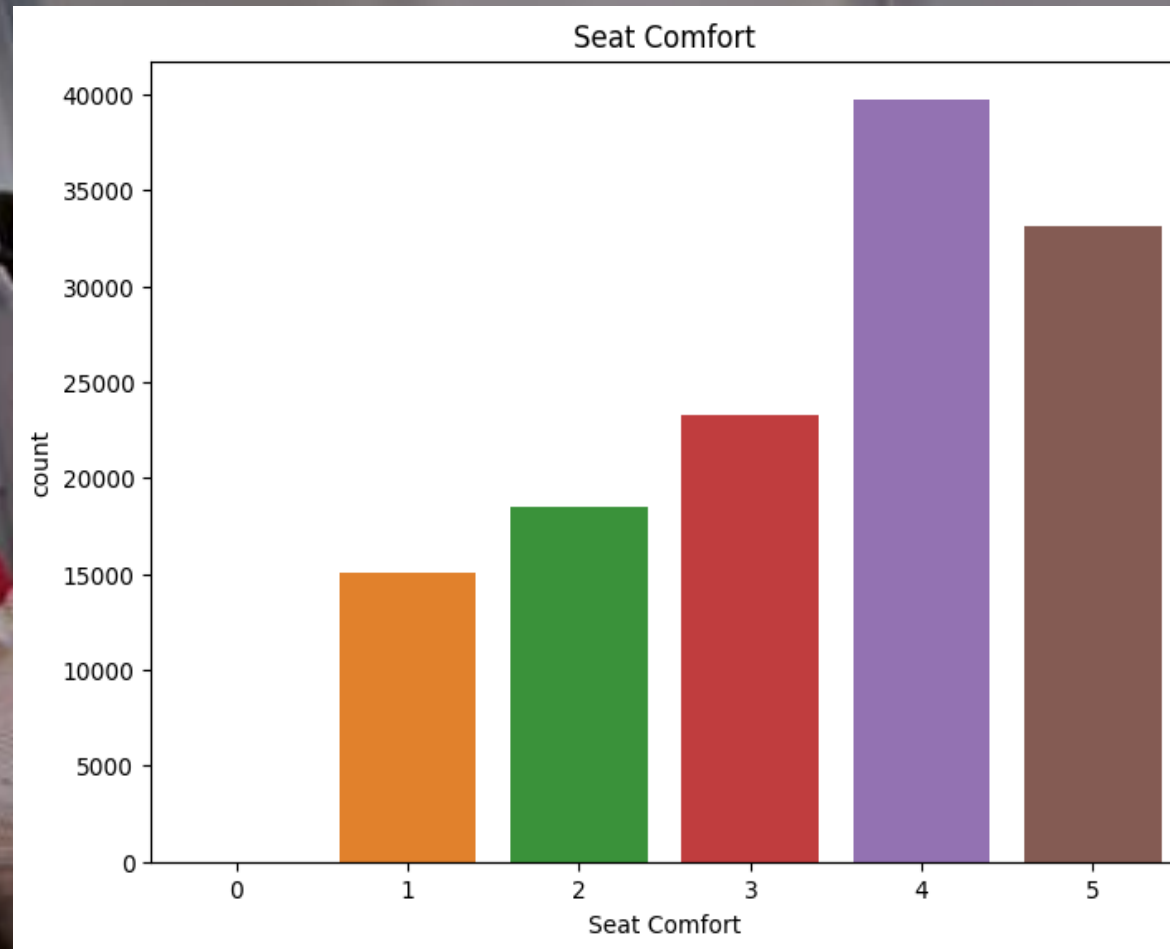


## Distribution of Passenger Age

- ☐ Age Category with Highest Number of Passengers: 39 Years"
- ☐ "The second highest number of passengers traveling on this airline belong to the age group 32
- ☐ The most of the passengers age in between 18 -56
- ☐ Least age for the passenger group is 5



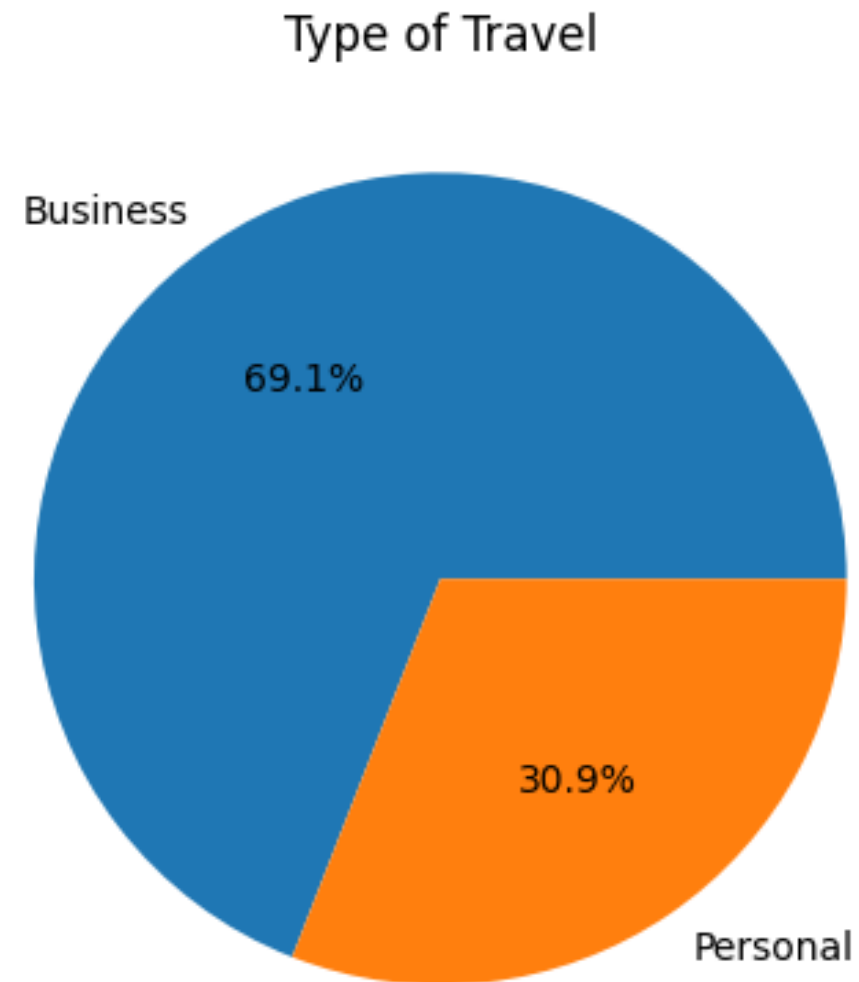
# Passenger Ratings for Seat Comfort



- The Highest Rating for Seat Comfort: 4"
- The Second Highest Rating for Seat Comfort: 5"
- This bar plot illustrates the distribution of passenger ratings for seat comfort."
- Passengers provided ratings for seat comfort, with the majority assigning a score of 4, followed by a significant number of passengers giving a rating of 5.

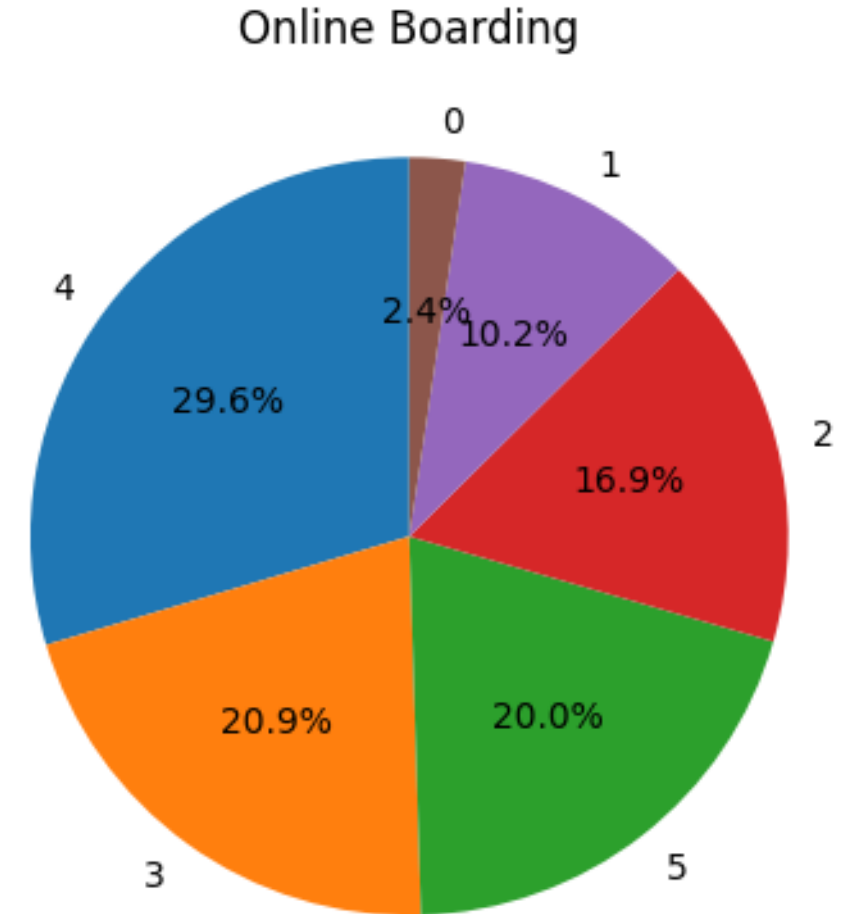
## Percentage Distribution of Travel Types

- The Majority of Passengers, Accounting for 69.1%, Travel for Business Purposes"
- "30.9% of Passengers Travel for Personal Purposes"
- This pie plot visualizes the proportional distribution of travel types among passengers."
- Business travel represents the dominant category, constituting 69.1% of the total, while personal travel accounts for 30.9% of the passengers."

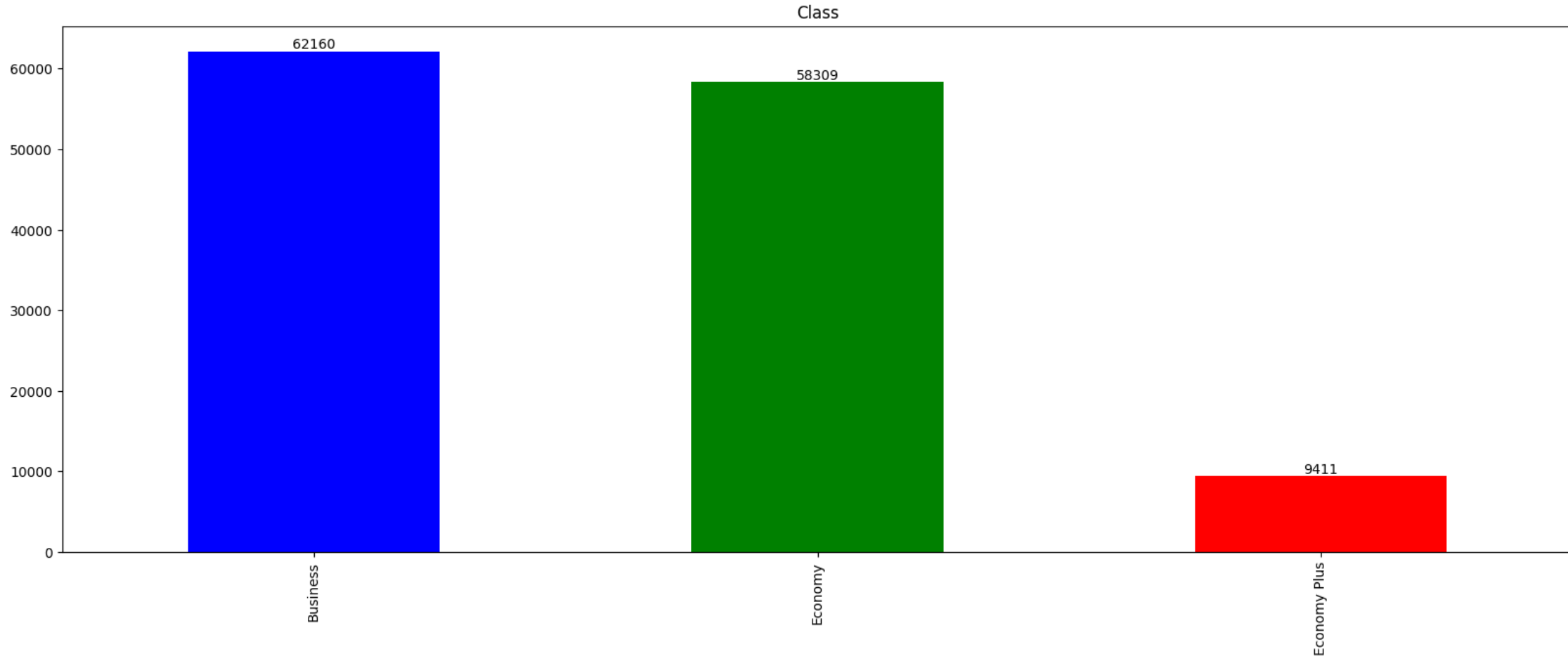


# Ratings for Online Boarding

- 29.6% of Passengers Rated Online Boarding with a Score of 4"
- "4% of Passengers Rated Online Boarding with a Score of 0"
- "This pie plot visualizes the distribution of ratings for the online boarding experience."
- "The majority of passengers, accounting for 29.6%, gave a rating of 4 for online boarding, while a small percentage of 4% rated it with a score of 0."

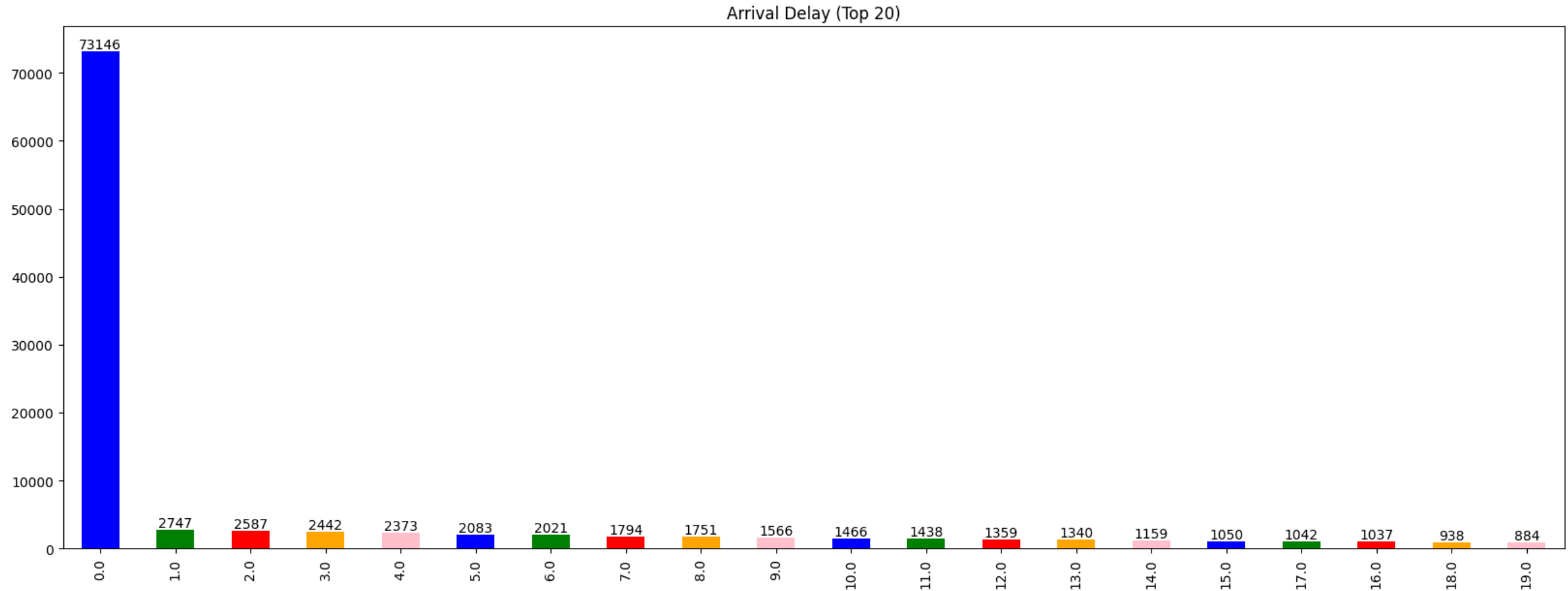


## Distribution of Passenger Travel Classes



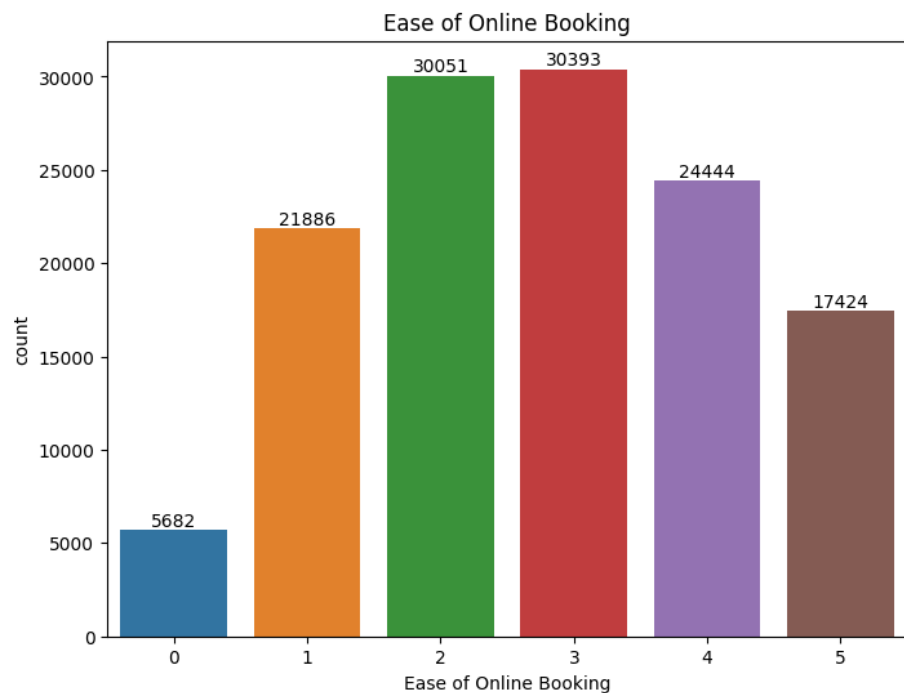
- The Majority of Passengers Traveled in the Business Class
- A Significant Proportion of Passengers Opted for the Business Class
- The Bar plot visually presents the frequency of passengers traveling in different classes.
- Business Class emerged as the preferred choice for a majority of passengers, while fewer passengers traveled in the Economy Plus class.

# Passenger Arrival Delays



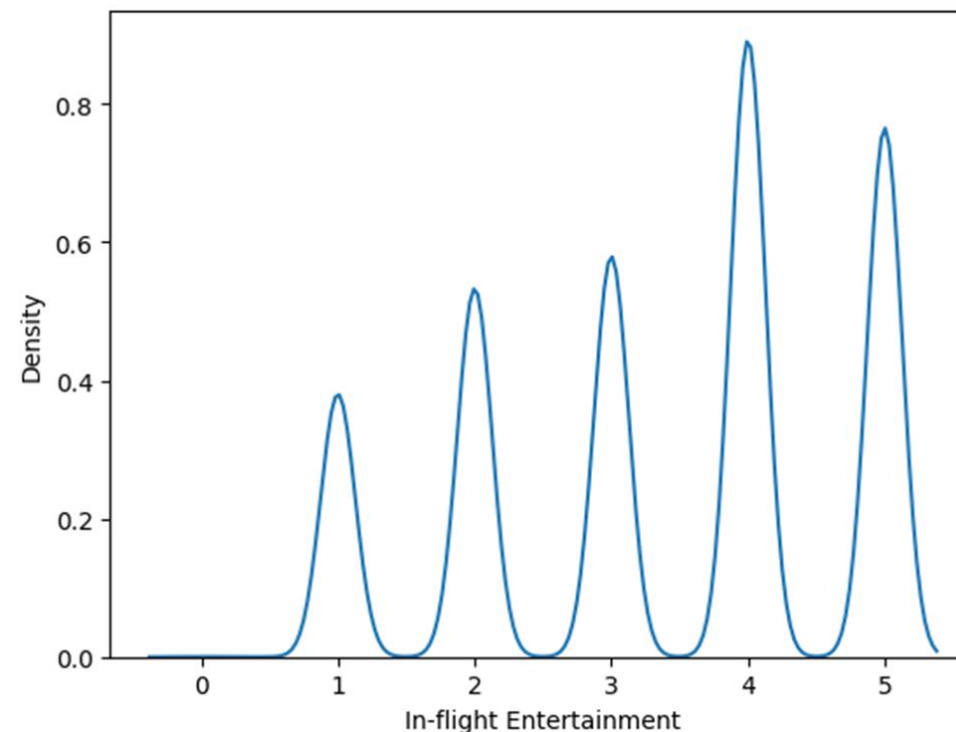
- The Majority of Passengers (73,146) Experienced No Arrival Delay"
- A Significant Number of Passengers Had No Arrival Delay"
- The Bar plot represents the frequency of arrival delays among passengers."
- For a significant portion of passengers (73,146), there were no reported arrival delays

## Ratings for Ease of Online Booking



- A Substantial Number of Passengers (17,424) Rated it with a 5"
- The Bar plot displays the distribution of ratings received for the ease of online booking.
- Most passengers provided a rating of 3 for the ease of online booking, while a significant group of 17,424 passengers gave it a top rating of 5

## Ratings for In-flight Entertainment

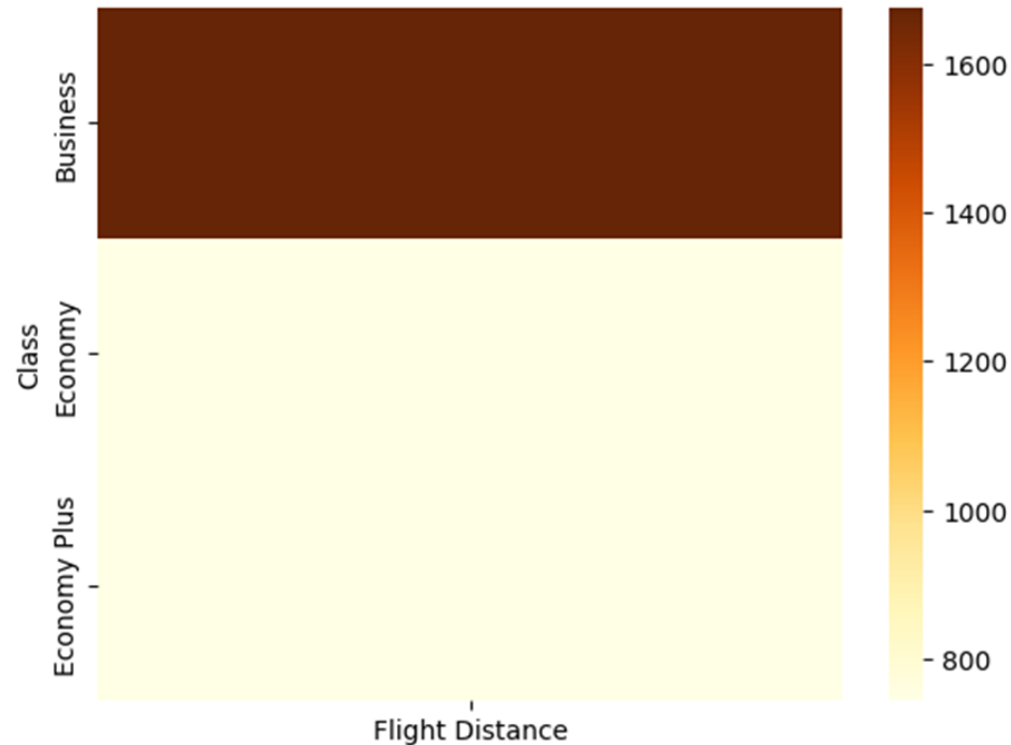


The Majority of Passengers Gave a Rating of 4 for In-flight Entertainment"

"A Substantial Number of Passengers Rated In-flight Entertainment with a Score of 4"

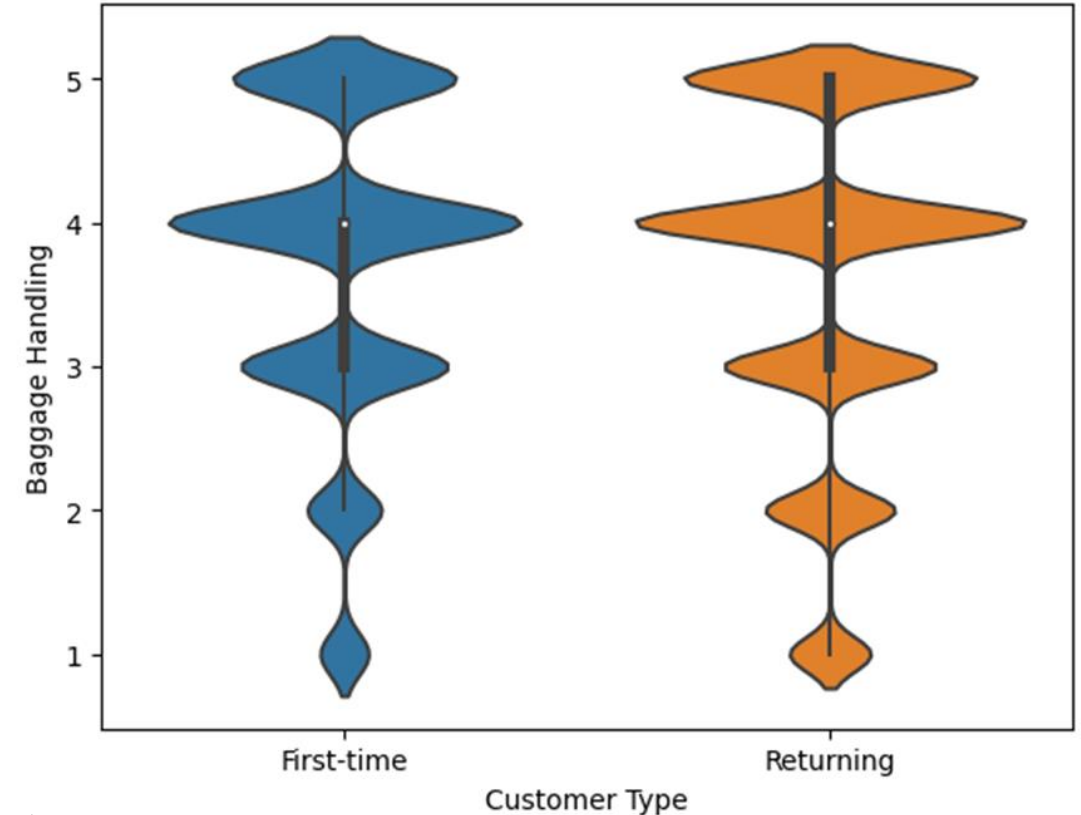


## Travel Distance of Passengers by Class



- "Passengers in the Business Class Travelled Long Distances"
- "Long Distance Travel is Predominantly Observed among Business Class Passengers"
- "The Heatmap visualizes the relationship between passenger class and the distance traveled."

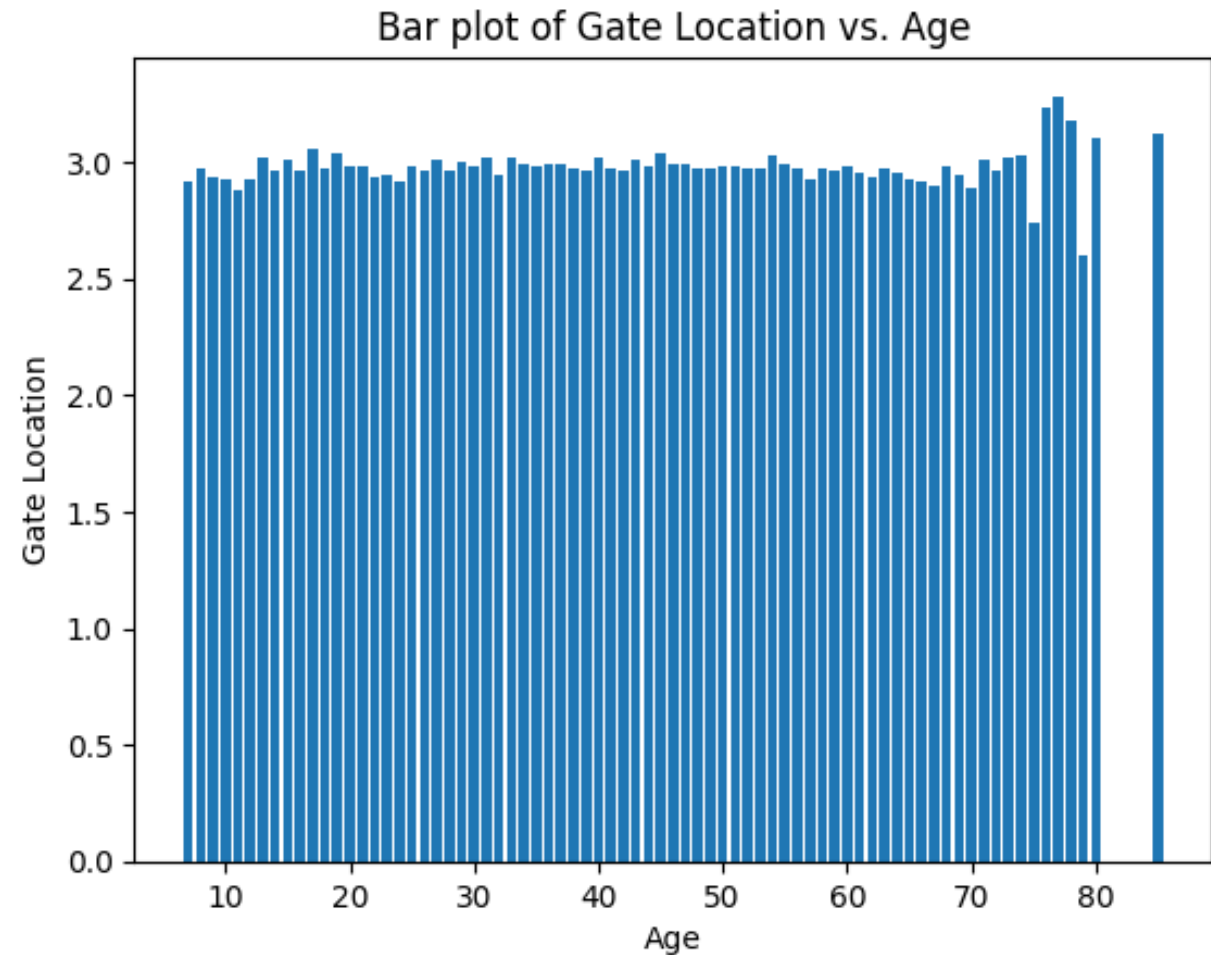
## Ratings of Baggage Handling by Customer Type



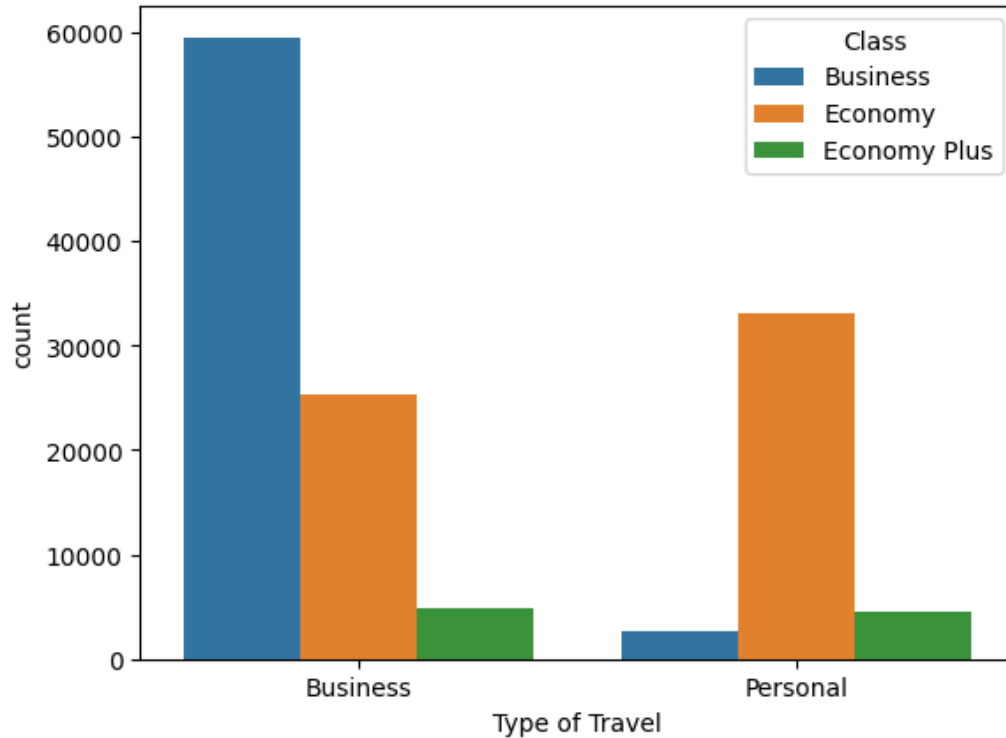
- "Both First-Time and Returning Passengers Gave a Rating of 4 for Baggage Handling"
- "The Violin plot illustrates the distribution and density of ratings for baggage handling across different customer types"

## Ratings of Gate Location by Age Group

- "Passengers in the 70-80 Years Old Age Group Provided the Highest Rating for Gate Location"
- "The 70-80 Years Old Age Group Exhibited the Highest Level of Satisfaction with Gate Location"
- "The Grouped Bar plot showcases the ratings of gate location, segmented by different age groups."
- "Passengers in the 70-80 years old age category rated the gate location the most positively compared to other age groups, indicating their overall satisfaction with this aspect."

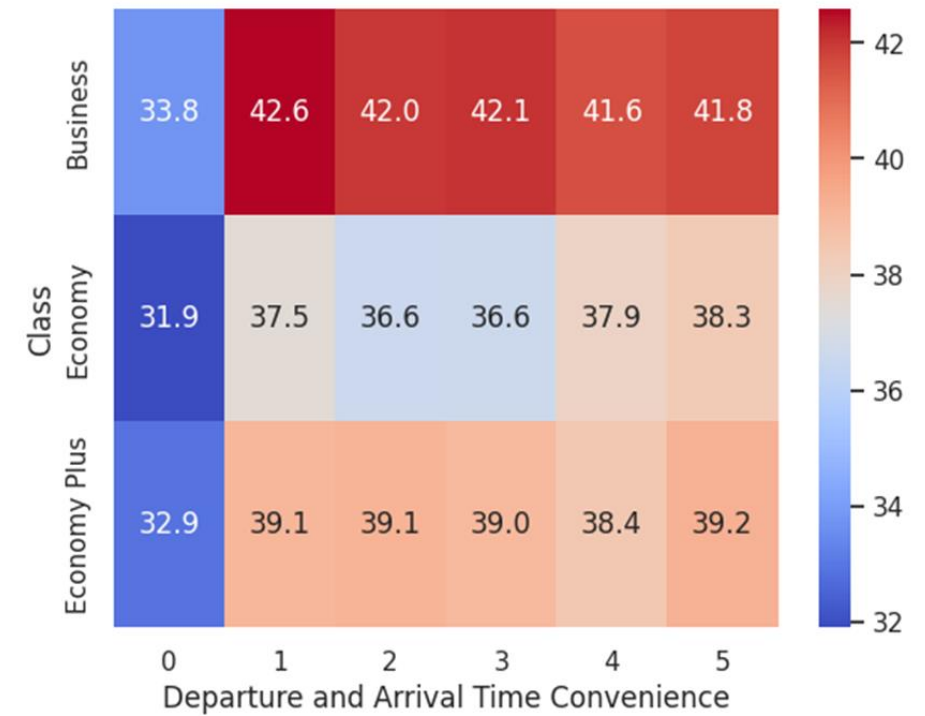


## Distribution of Travel Types by Class



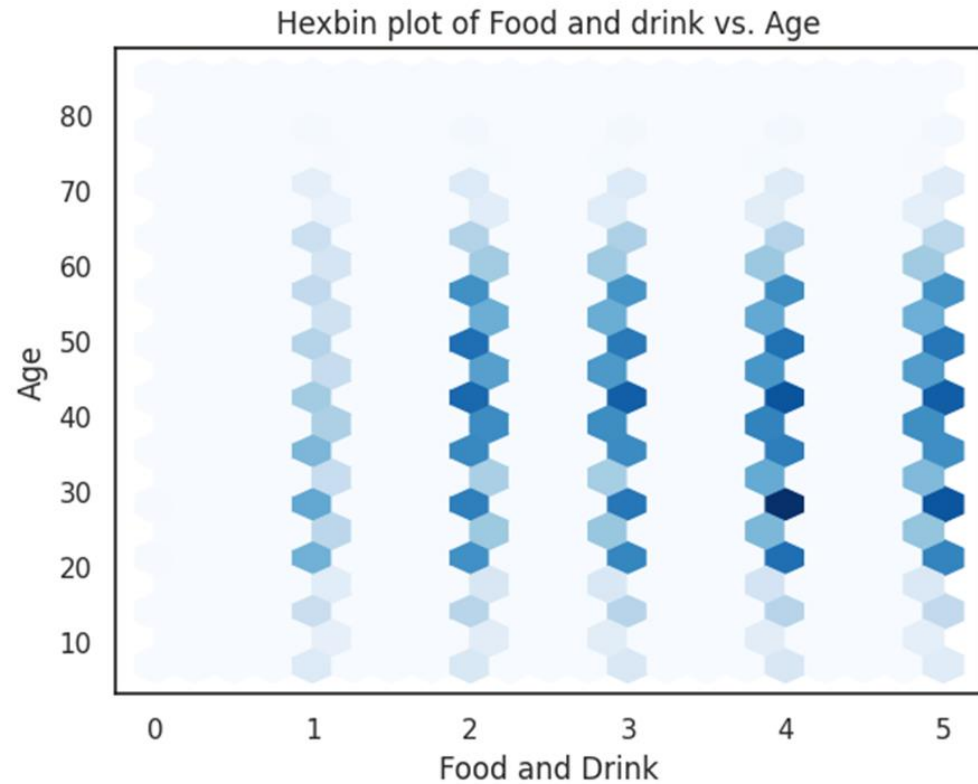
- Business Travel Accounts for the Majority of Travel Types"
- The Count plot visualizes the frequency of travel types categorized by passenger class."
- While a significant number of passengers traveled for business purposes, a smaller proportion traveled for personal reasons, with the majority of personal travelers opting for the Economy Class

## Ratings of Departure and Arrival Time Convenience by Passenger Class



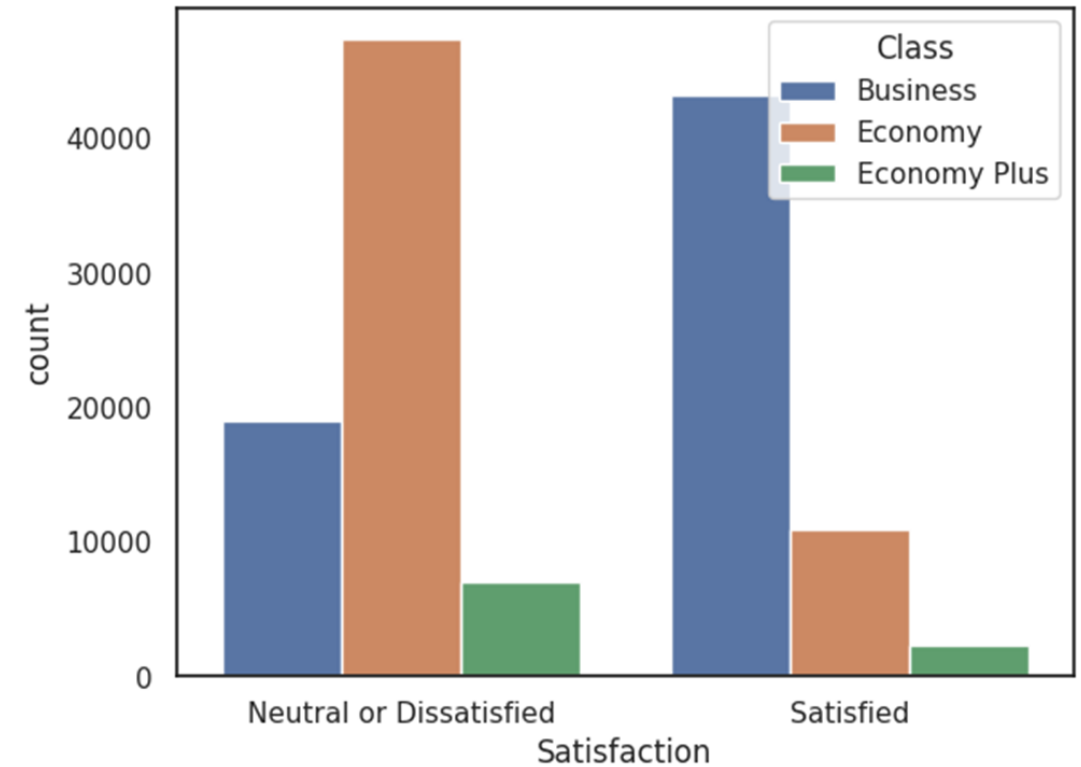
- Passengers in the Business Class Gave a Rating of 5 for Departure and Arrival Time Convenience"
- "Business Class Passengers Highly Rated the Convenience of Departure and Arrival Times"
- "The Heat map visualizes the range of ratings given by passengers of different classes for departure and arrival time convenience

## Ratings of Food and Drink by Age



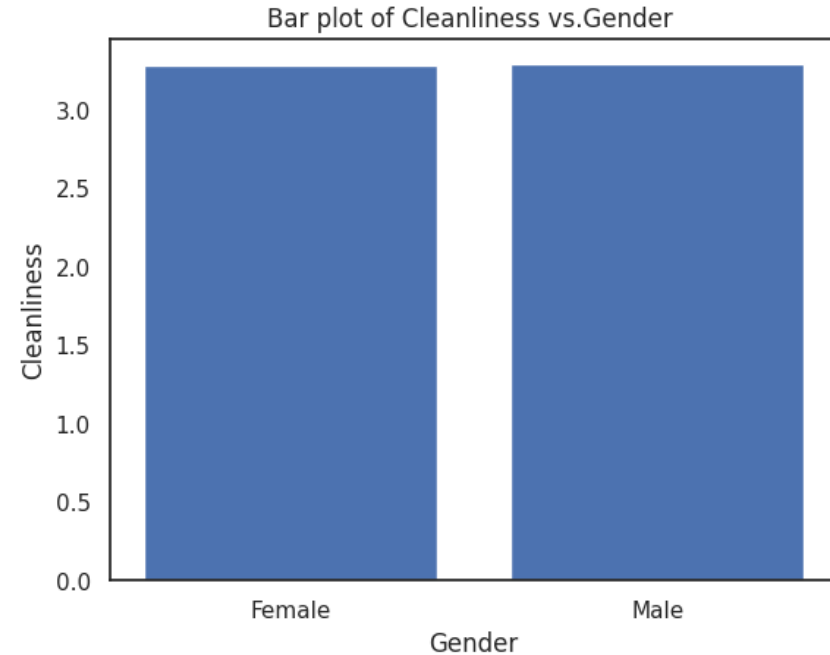
- ❖ Passengers Aged 30 Years Old Gave the Highest Rating for Food and Drink"
- ❖ "The 30 Years Old Age Group Showed the Highest Satisfaction with Food and Drink"
- ❖ The 80 Years old age group gave lowest rating for Food and Drink

## Customer Satisfaction and Dissatisfaction



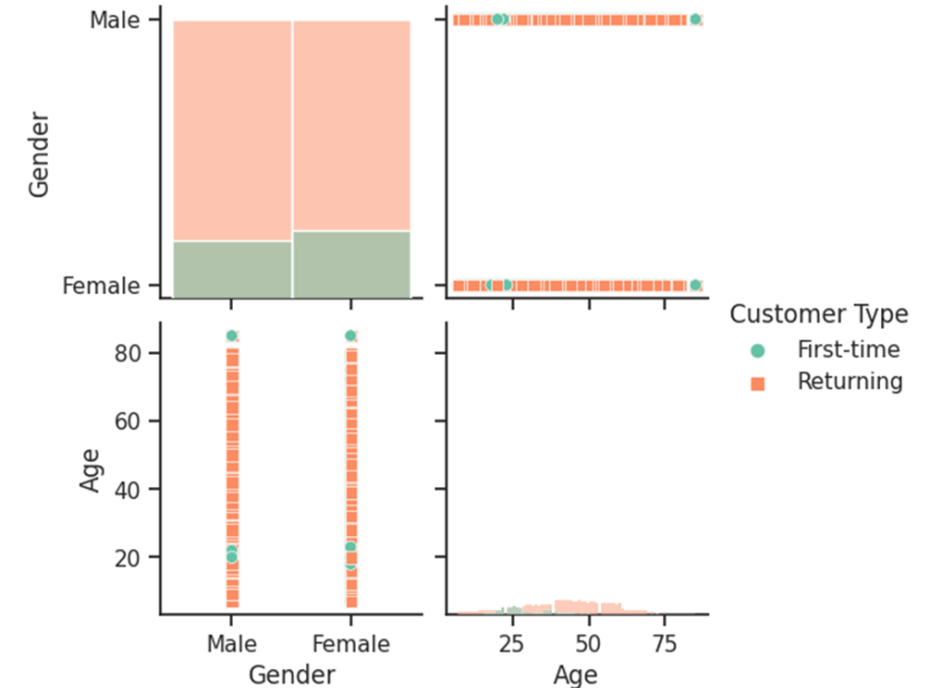
- ✓ Business Class Passengers Display Higher Levels Satisfaction"
- ✓ "Economy Class Passengers Show a Higher Degree of Dissatisfaction"
- ✓ "The Bar plot presents the distribution of customer satisfaction and dissatisfaction."

## Ratings of Cleanliness by Gender



- Both Male and Female Passengers Gave a Rating of 3 for Cleanliness"
- "Male and Female Passengers Display Similar Ratings for Cleanliness"
- "The Bar plot showcases the distribution of cleanliness ratings categorized by gender."
- "Both male and female passengers generally provided a rating of 3 for cleanliness, suggesting a comparable perception of the cleanliness standards onboard"

## Customer Type, Gender, and Age

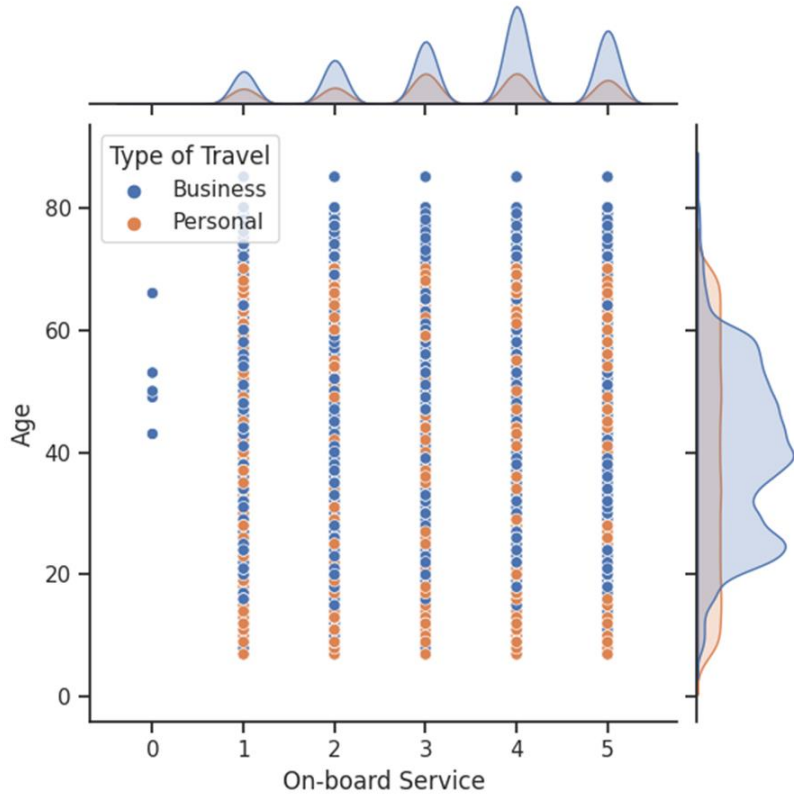


Females Show a Higher Frequency of Travel and Return in Comparison"

"The Pair plot visualizes the relationships among customer type, gender, and age."

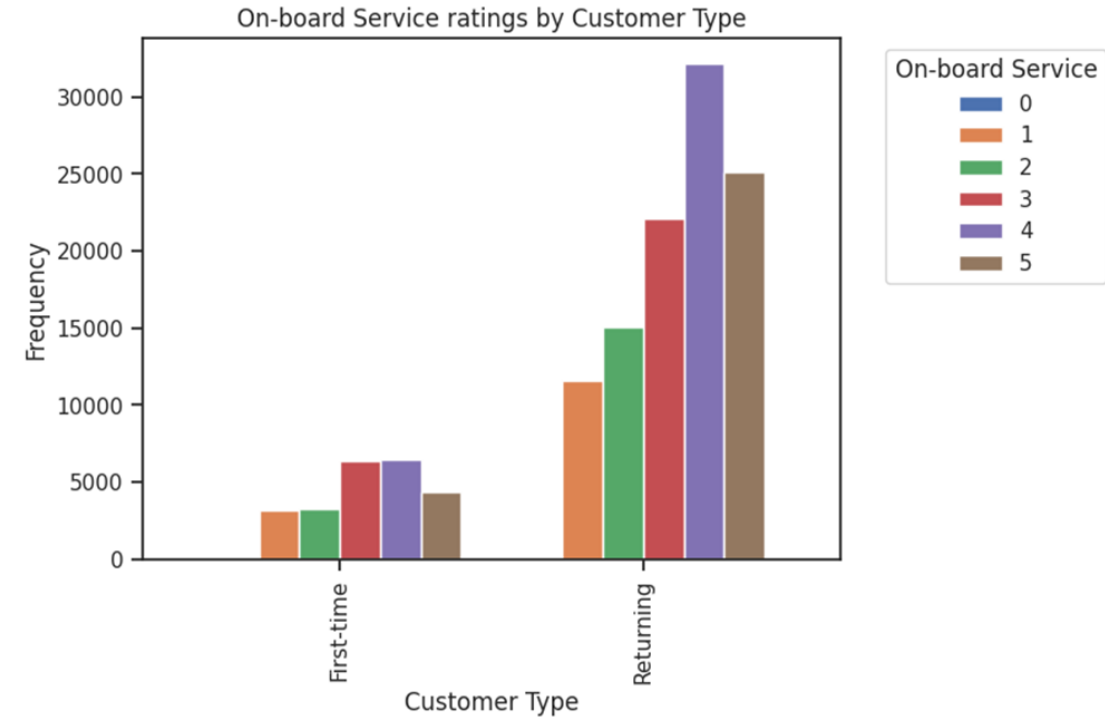
"Based on the plot, it can be observed that females have a higher occurrence of travel and return compared to males."

## On-board Service Ratings by Age and Type of Travel



- "Passengers Traveling for Personal Purposes Gave a Rating of 5 for On-board Service"
- "The 40 Years Old and Above Age Group Showed the Lowest Rating of 0 for On-board Service"
- "The Joint plot visualizes the relationship between age, type of travel, and ratings for onboard service"

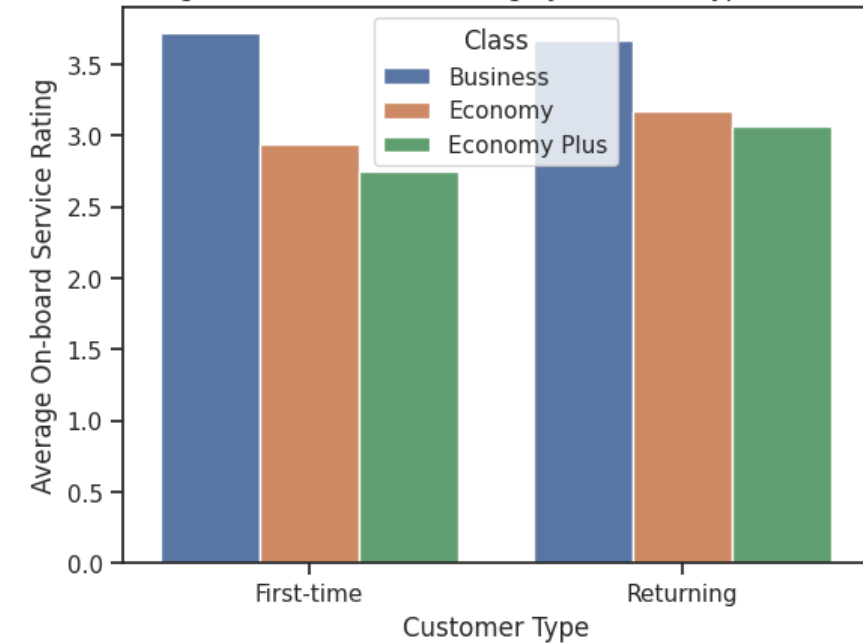
## On-board Services Ratings by Customer Type



- "Returning Passengers Show a Majority Rating of 4 for On-board Services"
- "A Few First-Time Passengers Also Gave a Rating of 4 for On-board Services"
- "The Bar plot visualizes the distribution of ratings for on-board services categorized by customer type"

## Average On-board Services Ratings by First-Time and Returning Passengers

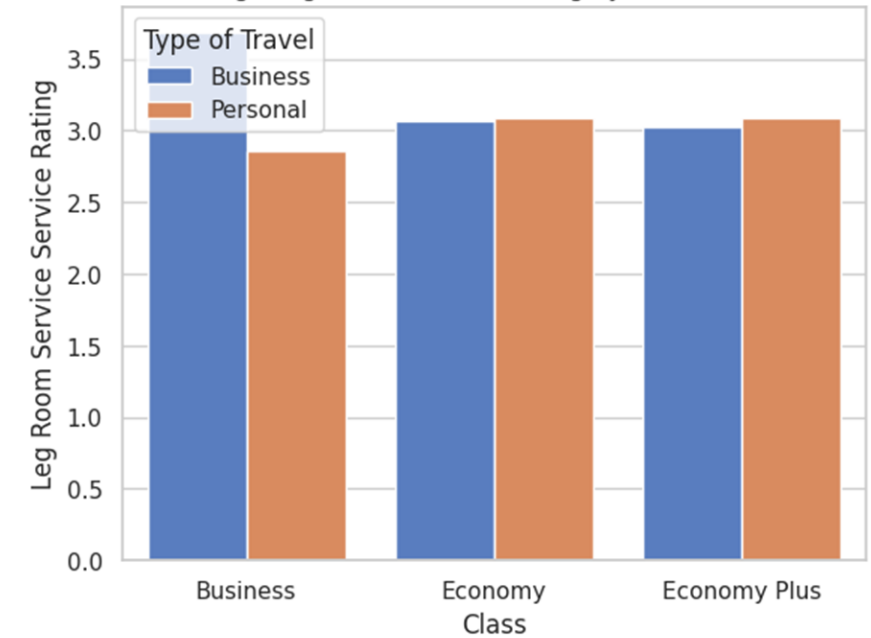
Average On-board Service Rating by Customer Type and Class



- The Average Rating for On-board Services is 3.5 for Both First-Time and Returning Passengers"
- "First-Time and Returning Passengers Display a Similar Average Rating of 3.5 for On-board Services"
- "The Bar plot illustrates the average ratings of on-board services for first-time and returning passengers."

## Ratings of Leg Room Services by Class

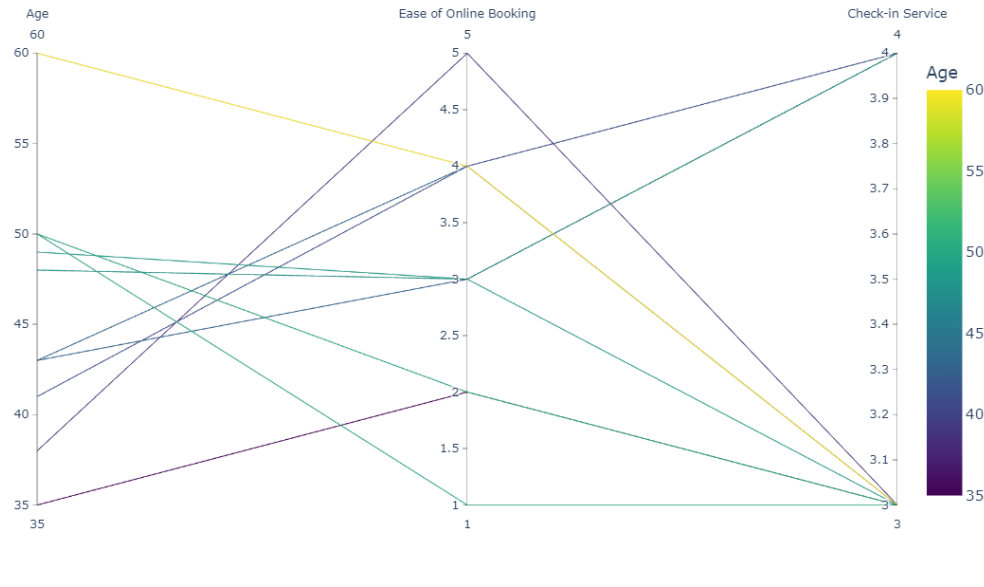
Average Leg Room Service Rating by Customer Class



- Business Class Passengers Gave an Average Rating of 3.5 for Leg Room Services"
- "Economy and Economy Plus Passengers Provided an Average Rating of 3.0 for Leg Room Services"
- "The Grouped Bar plot visualizes the distribution of ratings for leg room services across different passenger classes."

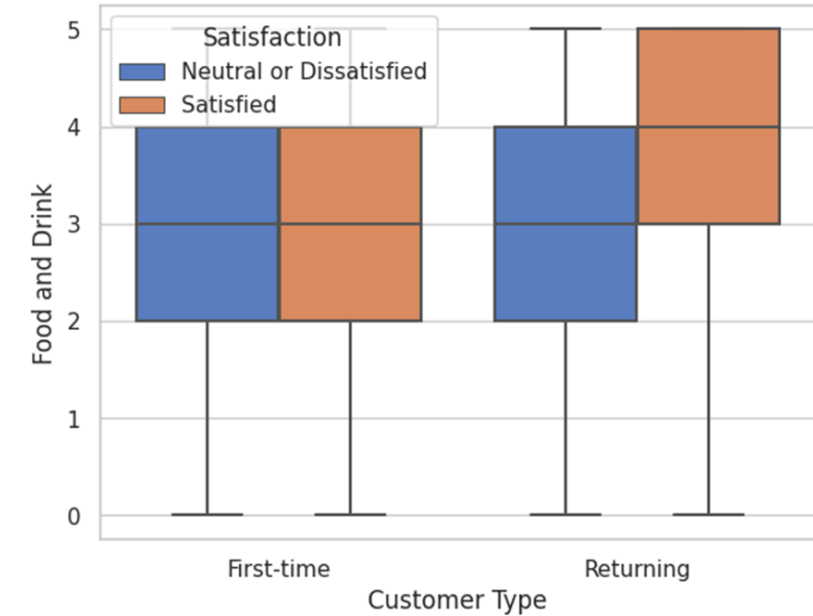


## Ratings for Ease of Online Booking and Check-in Services by Age



- "Passengers in the 60 Age Group Gave a Rating of 4 for Ease of Online Booking and 3 for Check-in Services"
- "Passengers in the 39 Age Group Provided a Rating of 5 for Ease of Online Booking and Check-in Services"
- "The Parallel Coordinate plot visualizes the relationship between age, ratings for ease of online booking, and ratings for check-in services."

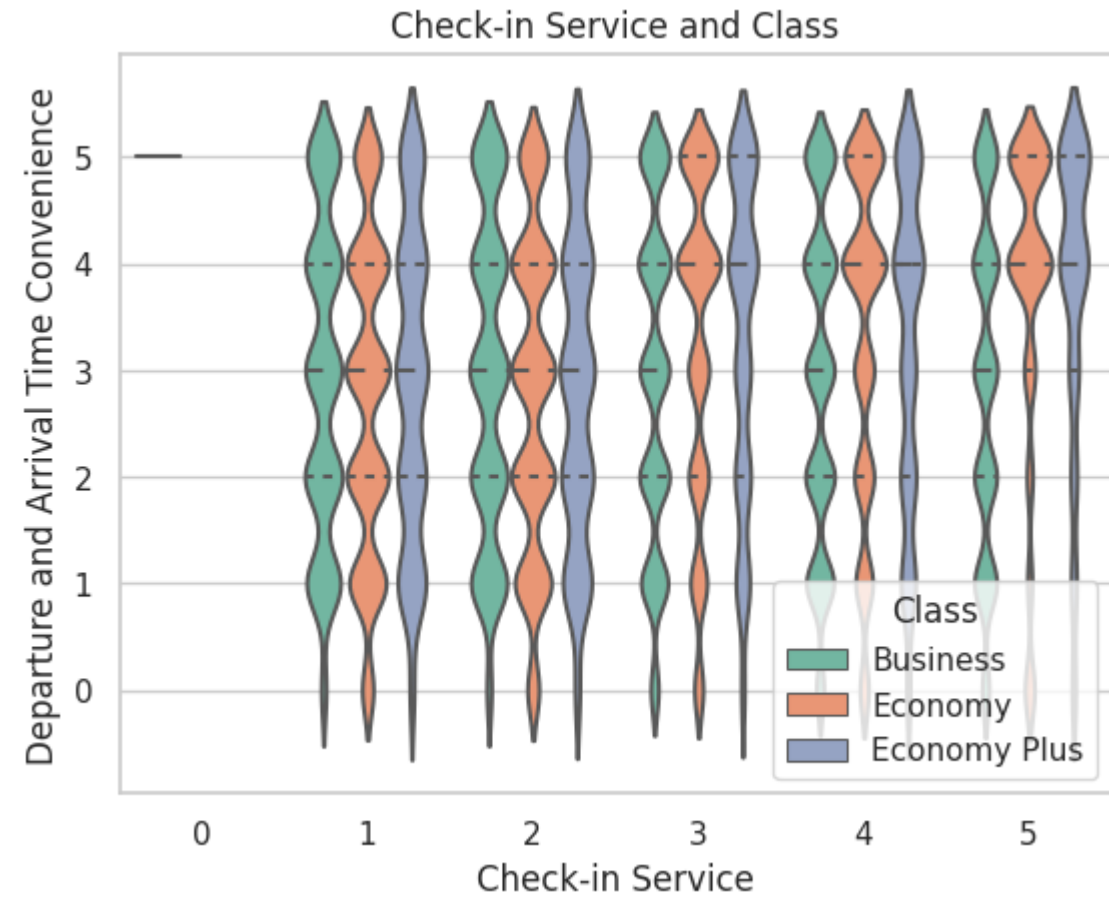
## Satisfaction with Food and Drink by Passenger Type



- First-Time Passengers Gave a Rating of 4 for Food and Drink, Indicating Satisfaction for Some Passengers"
- "Returning Passengers Provided a Rating of 5 for Food and Drink, Demonstrating Higher Satisfaction Overall"
- "The Box plot illustrates the distribution of ratings for food and drink satisfaction among different passenger types"

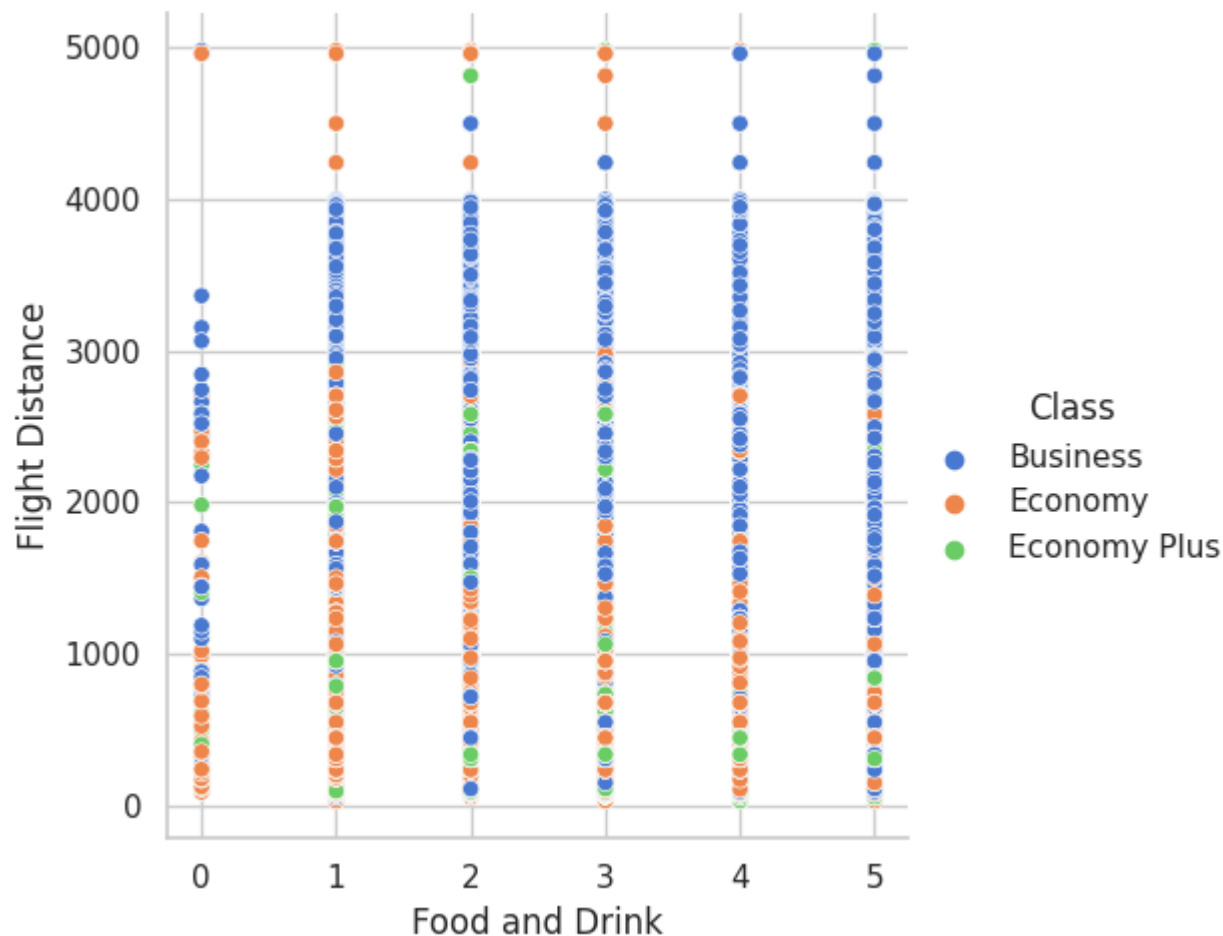
# Ratings of Check-in Service and Departure/Arrival Time Convenience by Class

- Economy Class Passengers Gave a Rating of 5 for Departure and Arrival Time Convenience and a Rating of 4 for Check-in Services"
- "Business Class Passengers Provided a Rating of 4 for Departure and Arrival Time Convenience"
- "The Violin plot visualizes the distribution of ratings for check-in service and departure/arrival time convenience across different passenger classes."
- "Economy class passengers consistently expressed a higher level of satisfaction with a rating of 5 for departure and arrival time convenience and a rating of 4 for check-in services. On the other hand, business class passengers gave a slightly lower rating of 4 for departure and arrival time convenience.



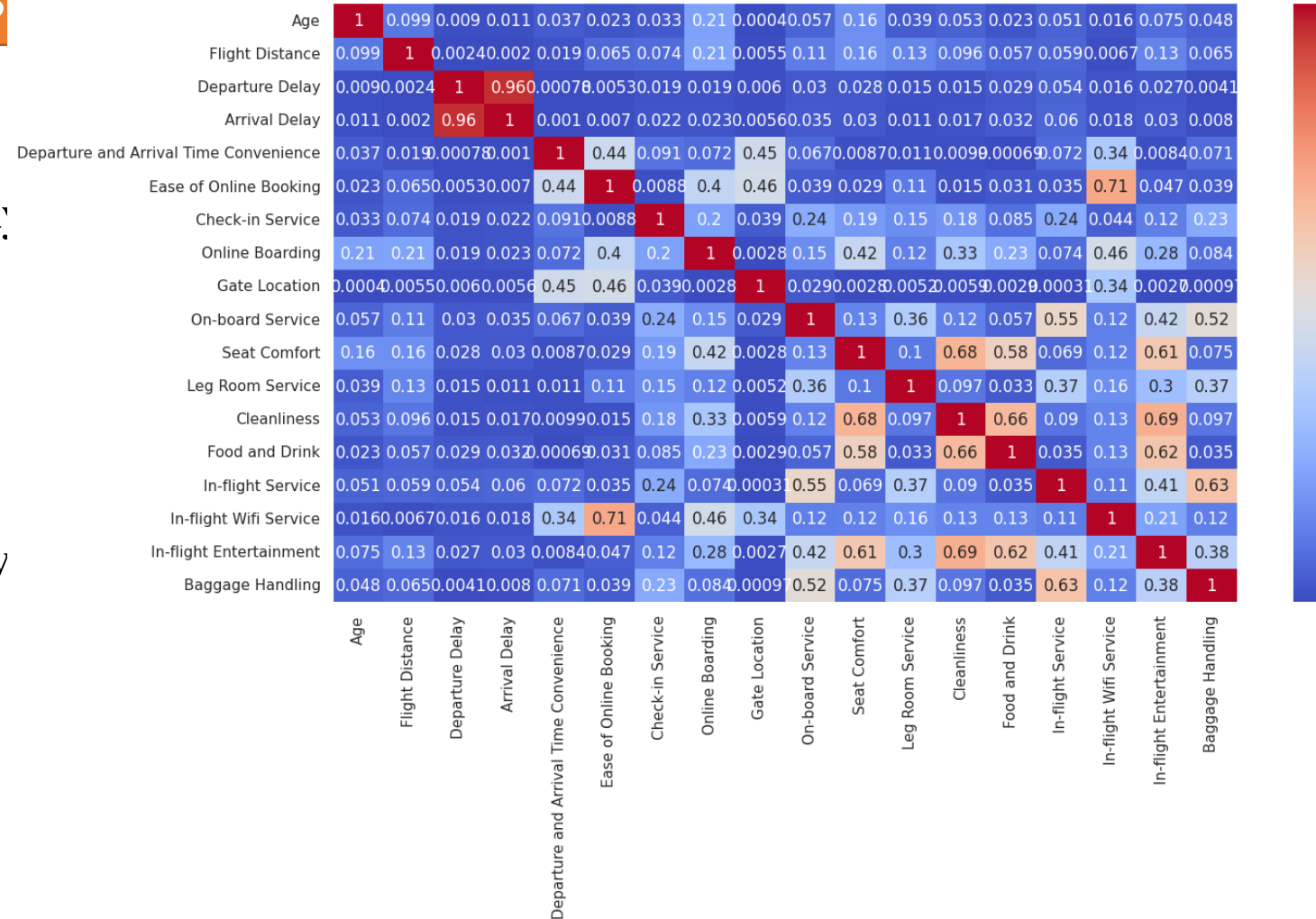
## Flight Distance and Food and Drink Ratings by Class

- Business Class Passengers Mostly Traveled Long Distances and Gave a Rating of 5 for Food and Drink"
- "Economy Class Passengers Who Traveled Long Distances Provided a Rating of 2 for Food and Drink"
- "The Scatter plot visualizes the relationship between flight distance, food and drink ratings, and passenger class."
- "It can be observed that a majority of business class passengers traveled long distances and expressed a high level of satisfaction with a rating of 5 for food and drink. In contrast, economy class passengers who traveled long distances gave a lower rating of 2 for food and drink."

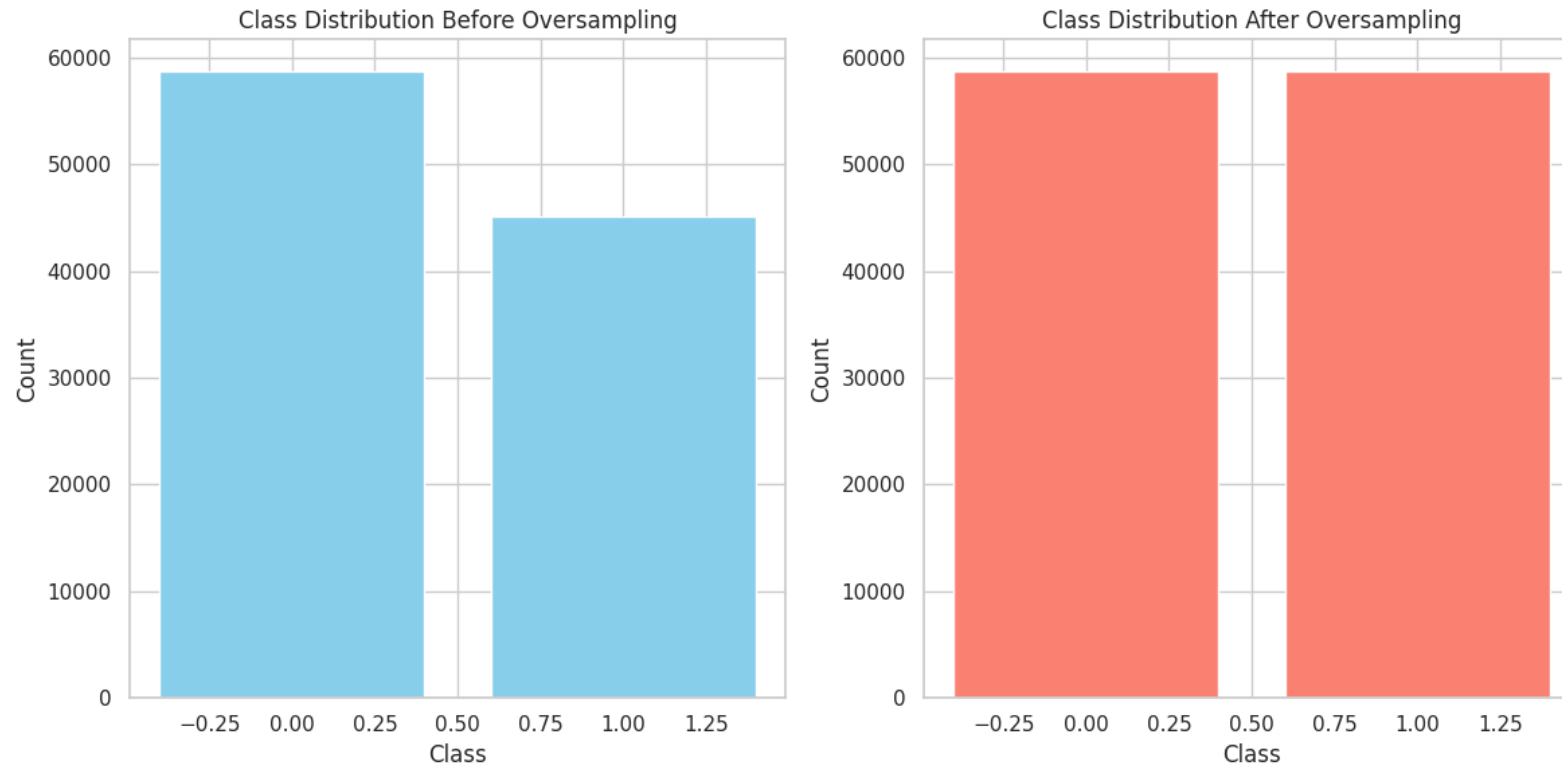


# Correlation between Attributes

- ❑ The highest correlation between arrival delay and departure delay value 0.96
- ❑ The Heatmap Shows a Correlation between Ease of Online Booking and In-flight WiFi Services is 0.7
- ❑ The Remaining Attributes Demonstrate Low Correlation Levels



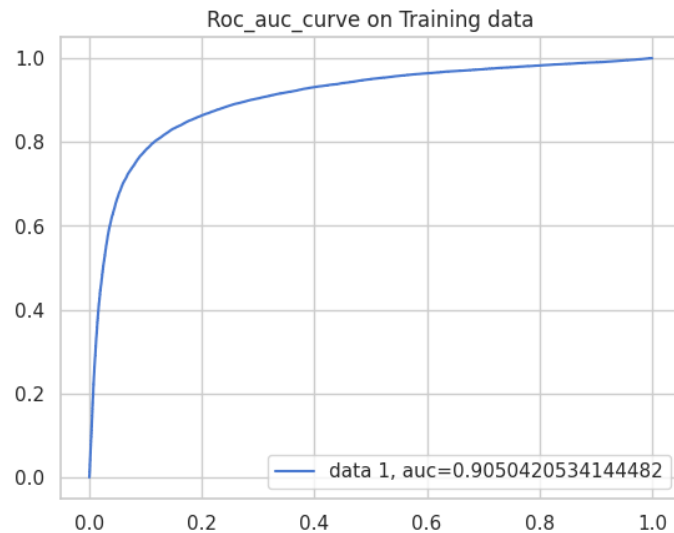
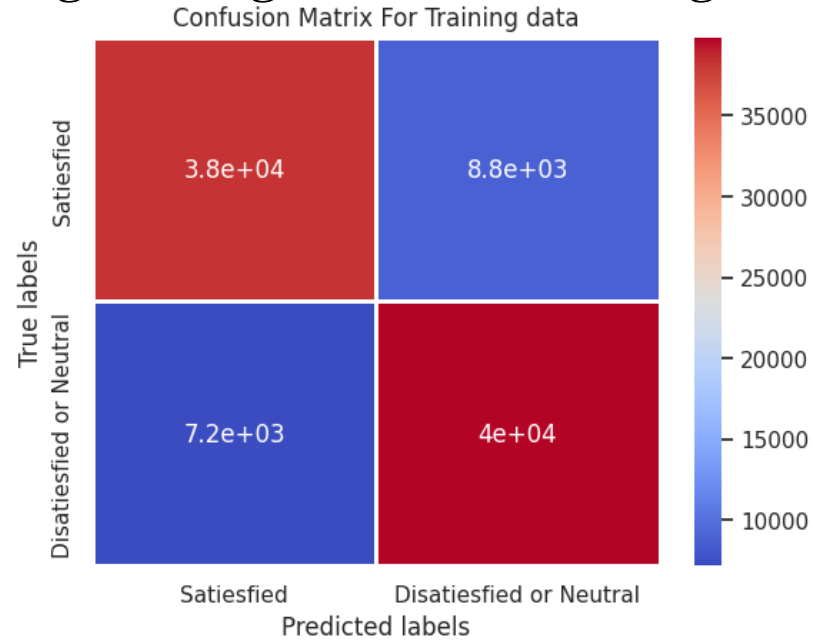
# Dataset Distribution Before and After SMOTE Technique



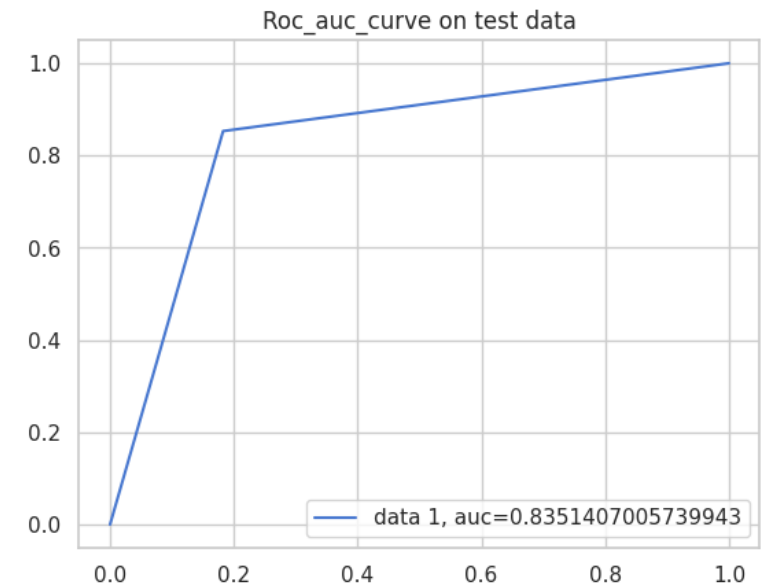
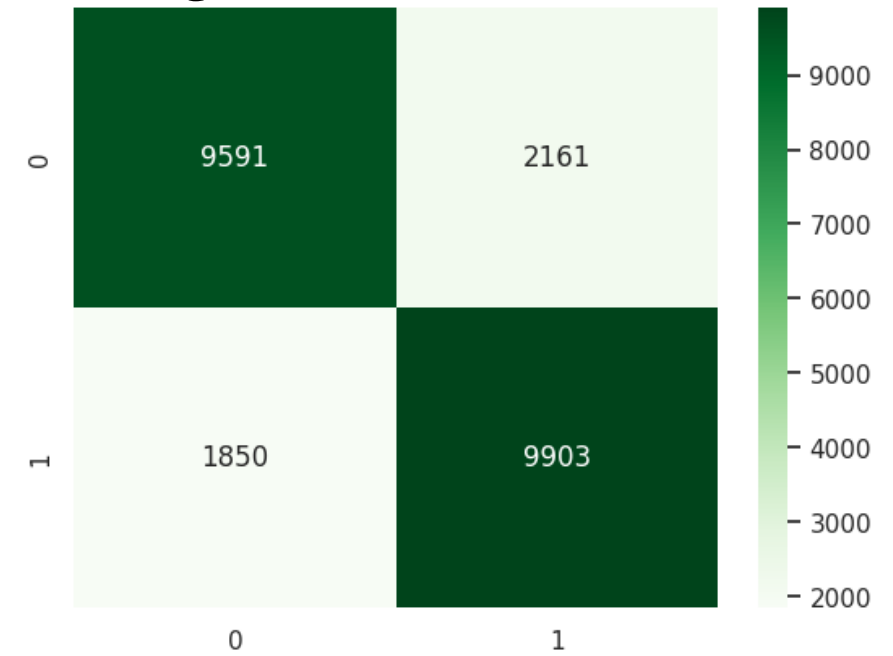
- The Left Bar Plot Represents the Distribution of the Original Dataset, with Counts of 0 and 1 as {0: 58762, 1: 45142}"
- "The Right Bar Plot Illustrates the Distribution of the Dataset After Oversampling, with Counts of 0 and 1 as {0: 73452, 1: 73452}"
- "The Bar plots provide a visual comparison of the dataset distribution before and after oversampling."
- "The original dataset had a class distribution of 0 and 1 as 58762 and 45142, respectively. However, after oversampling, both classes were balanced, resulting in an equal count of 73452 for each class

# Without Optimization

## Logistic Regression on Training Data

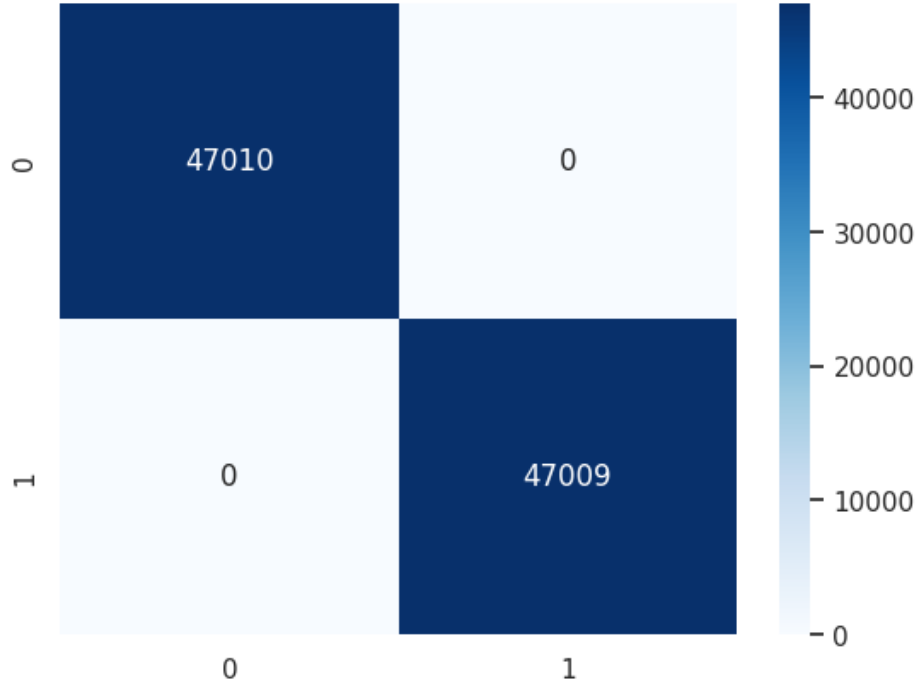


## Logistic Regression on Test Data"



**Accuracy\_score = 0.82**

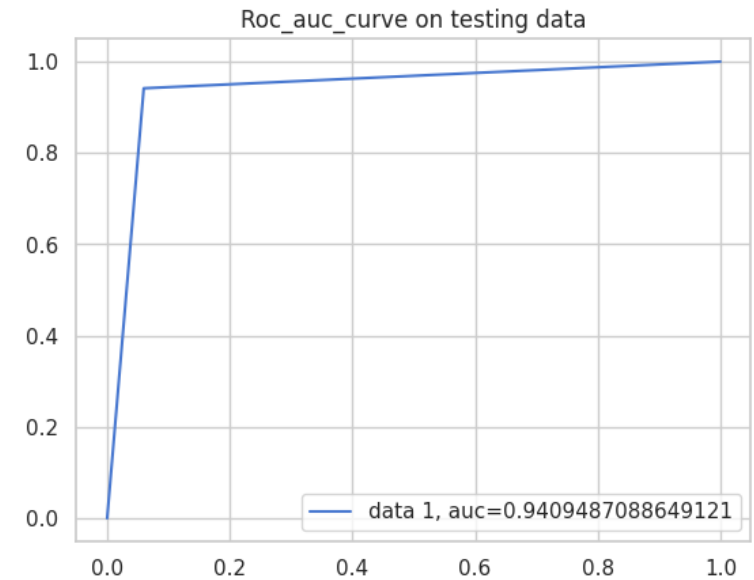
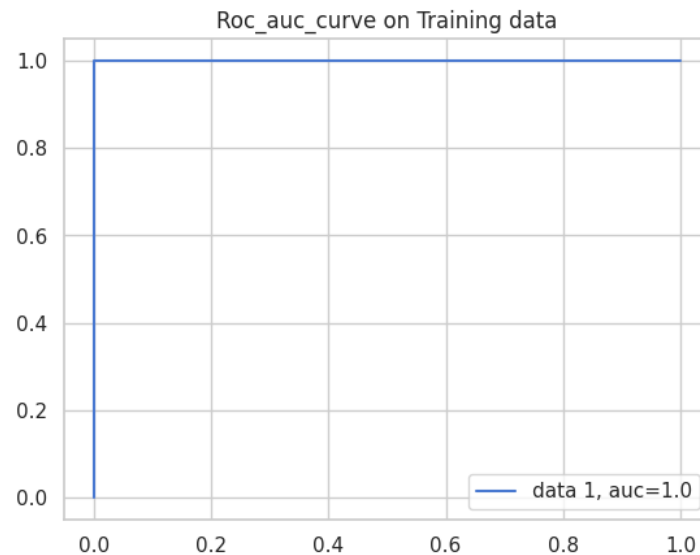
## Decision Tree on Training Data



## Decision Tree

**Accuracy\_score = 0.94**

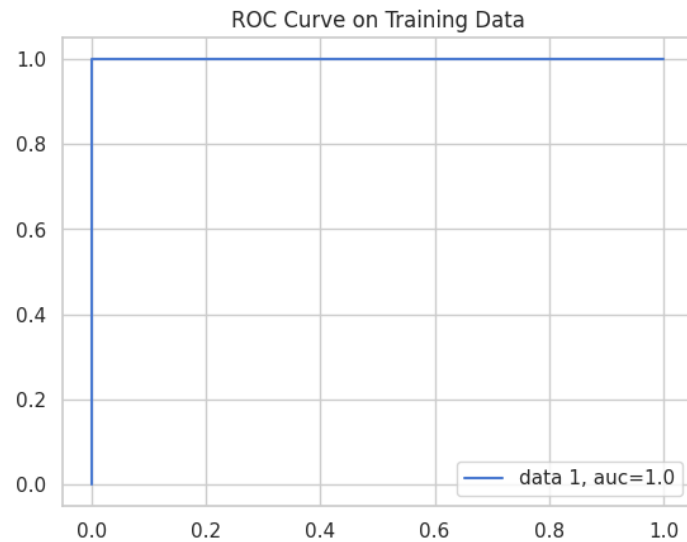
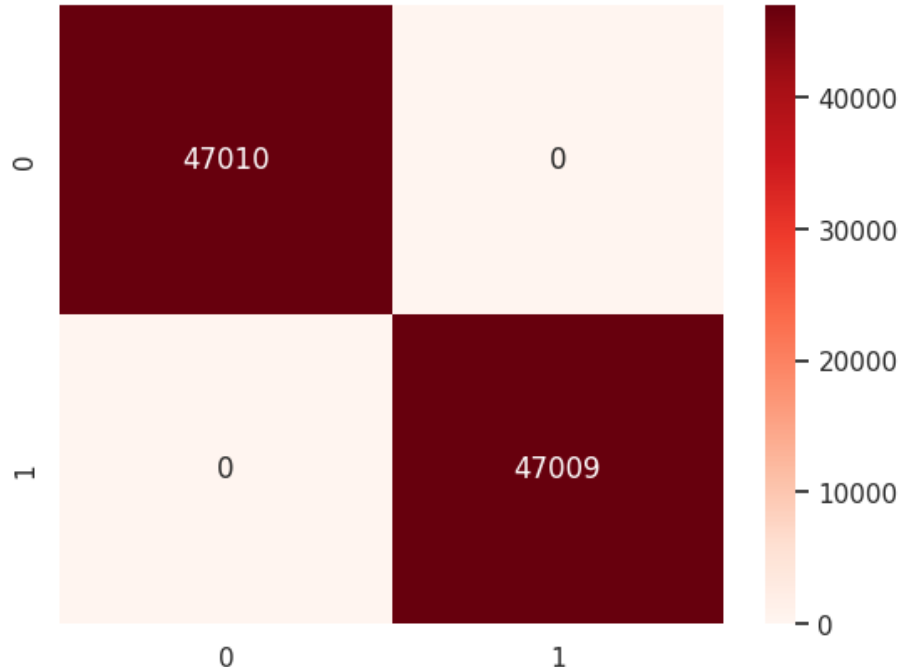
## Decision Tree on Test Data



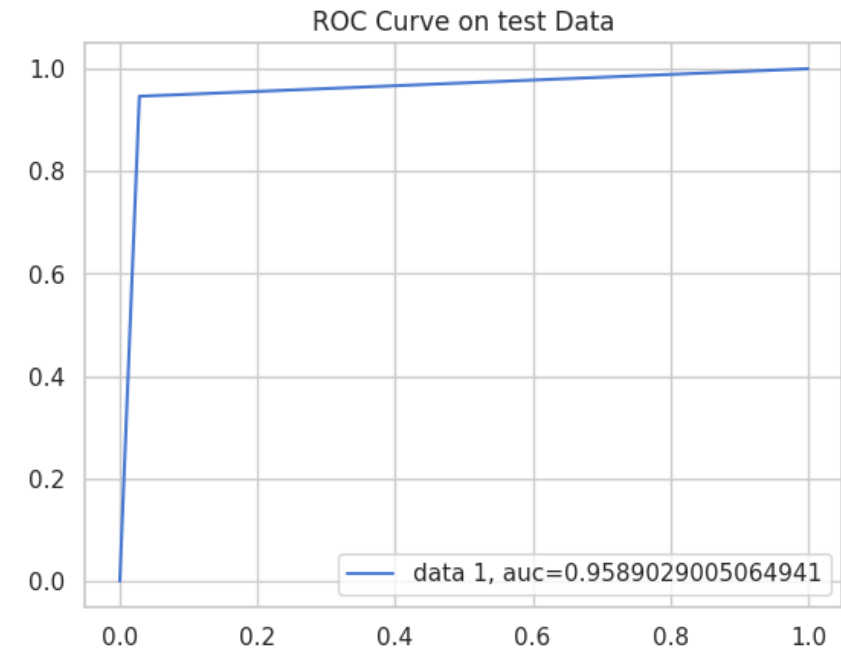
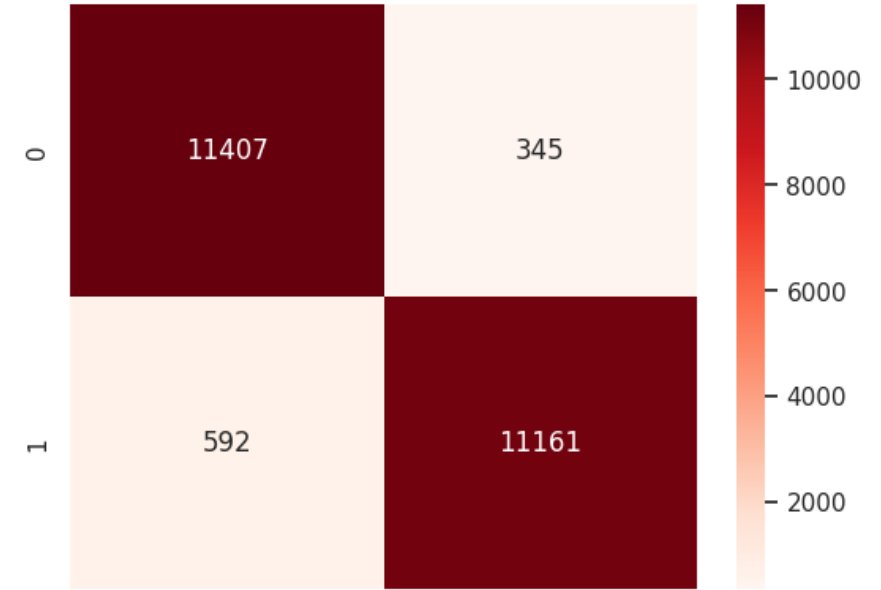


# Random Forest

## Confusion Matrix on Train Data

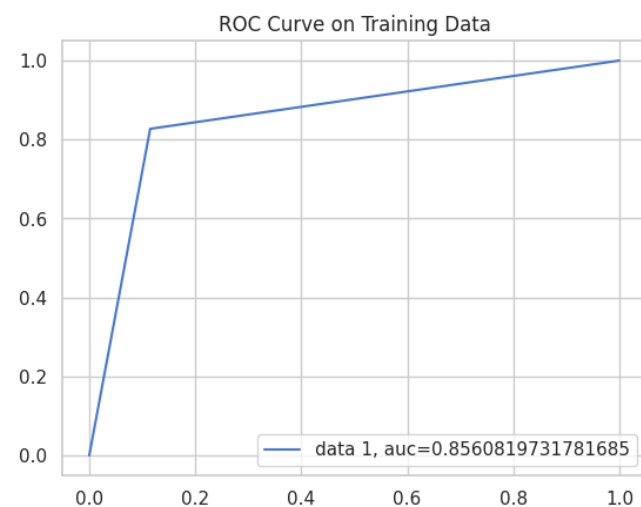
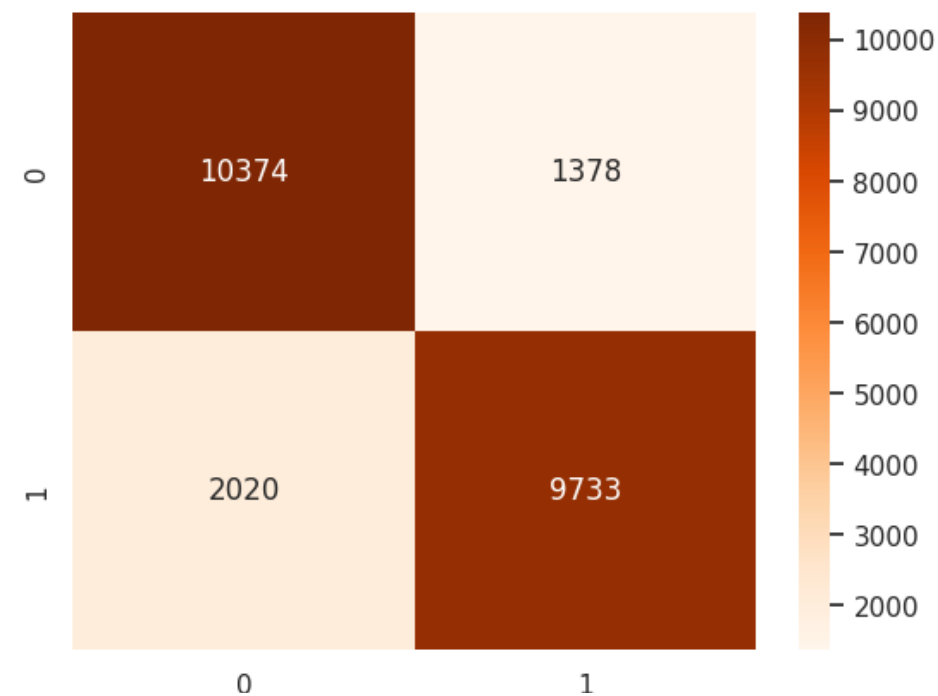
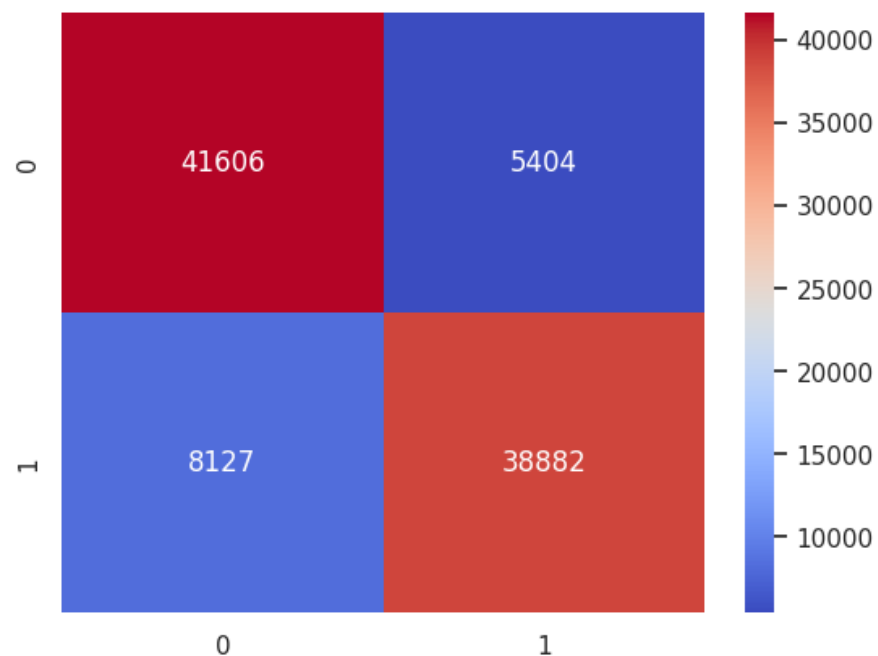


## Confusion Matrix on Test Data

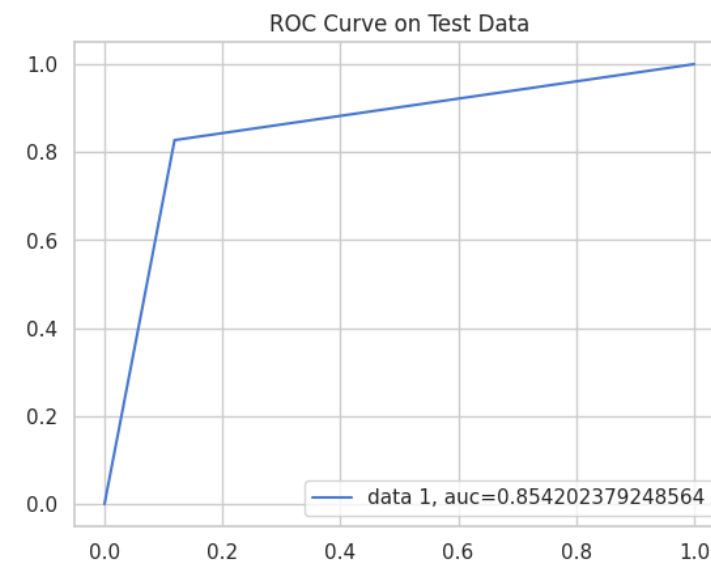


**Accuracy\_score = 0.96**

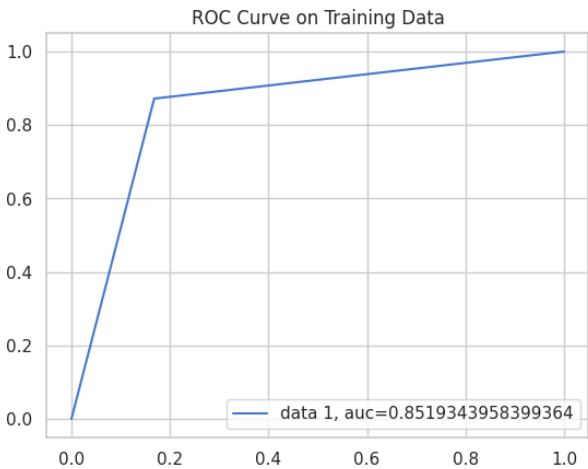
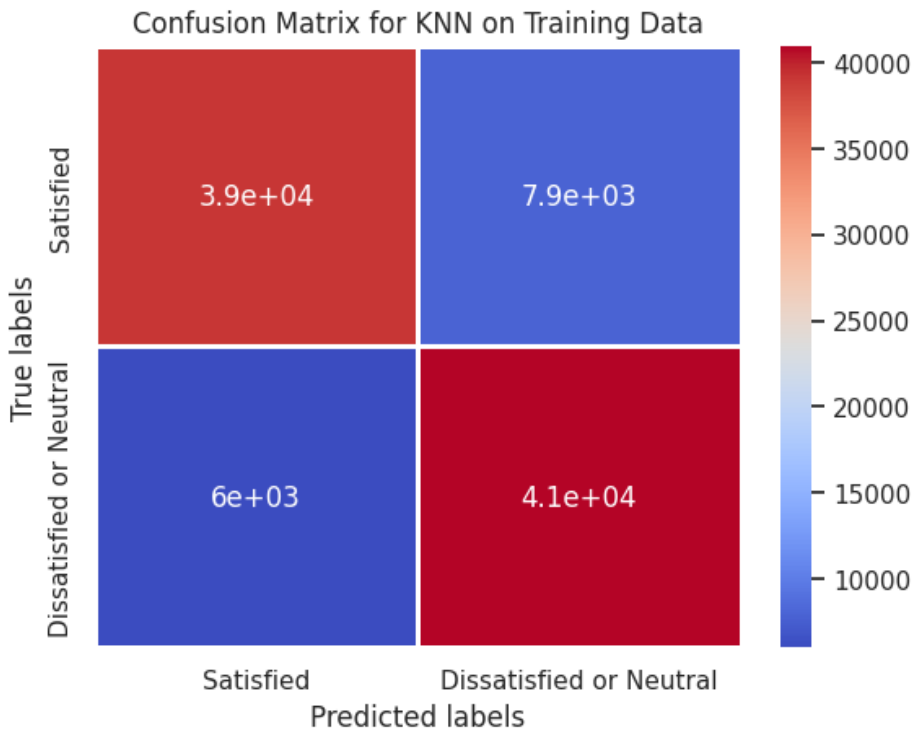
# Gaussian NB



**Accuracy\_score = 0.85**



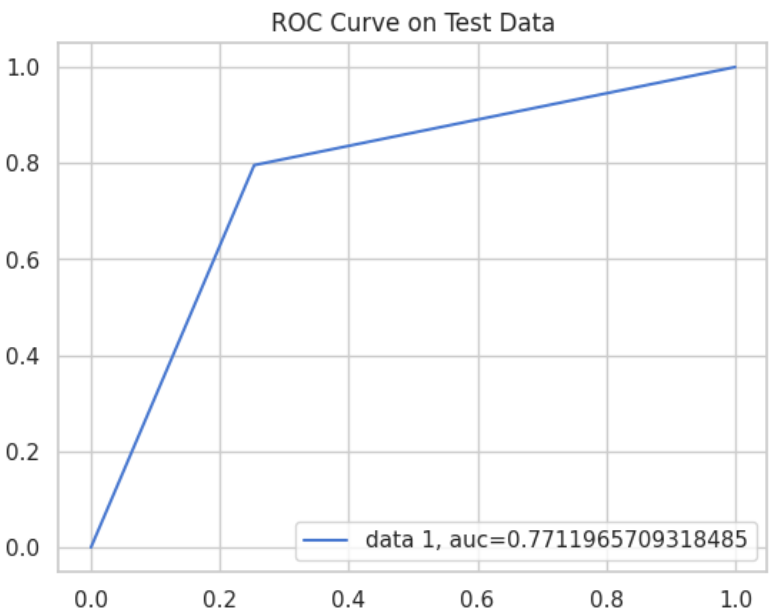
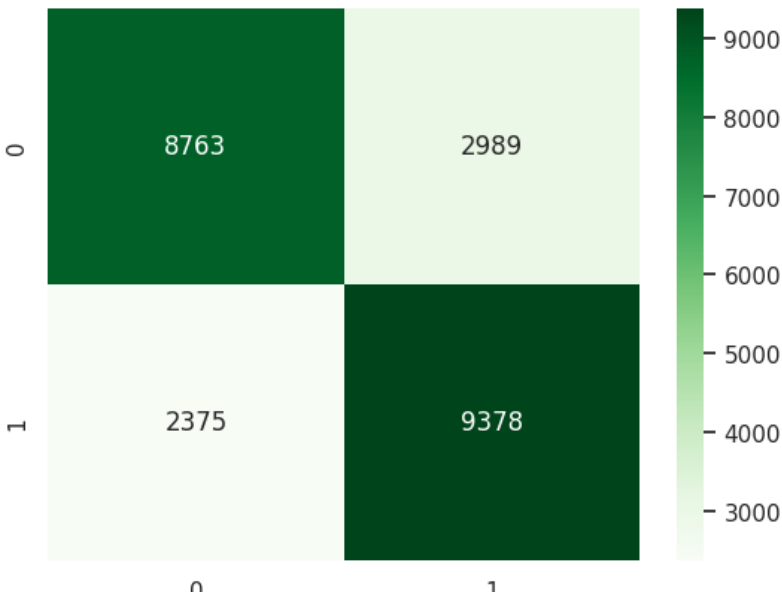
# Confusion Matrix of Training Data



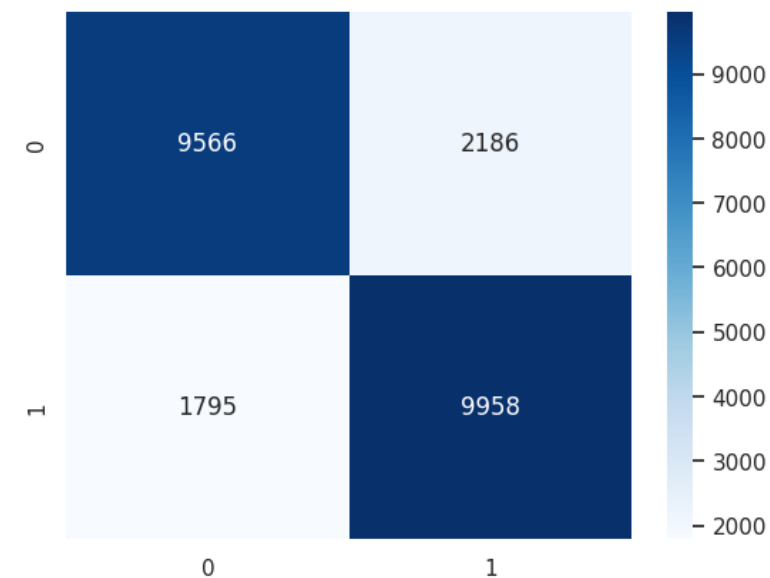
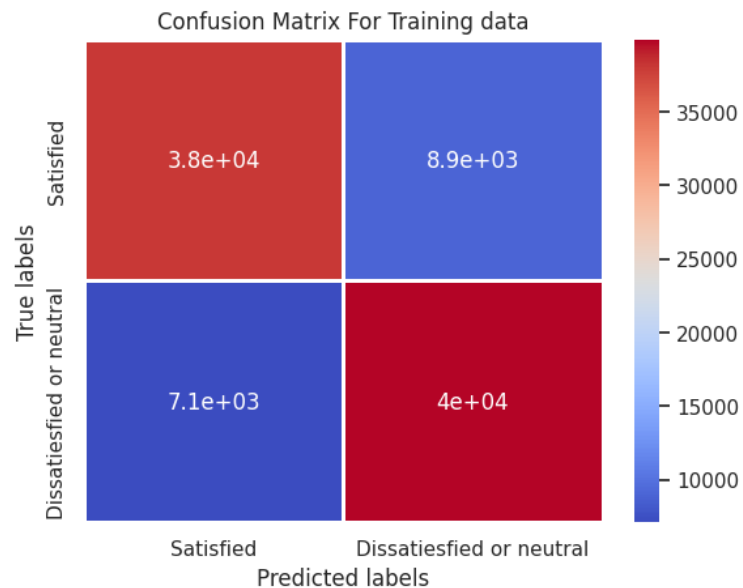
# KNN

Accuracy \_score = 0.77

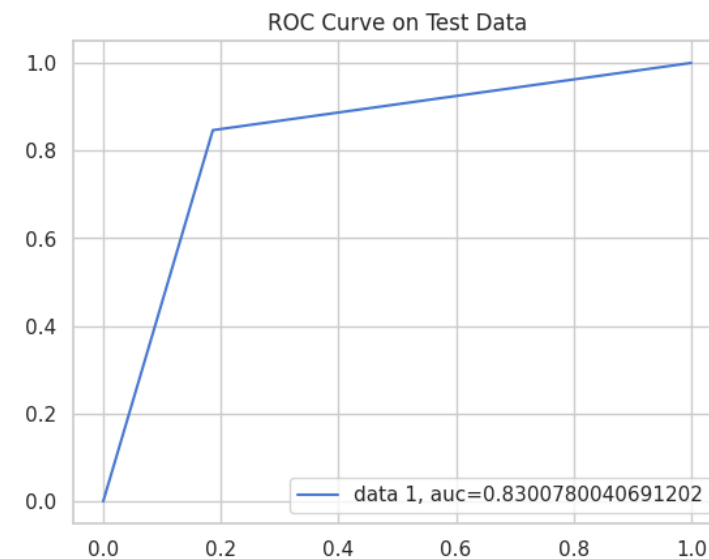
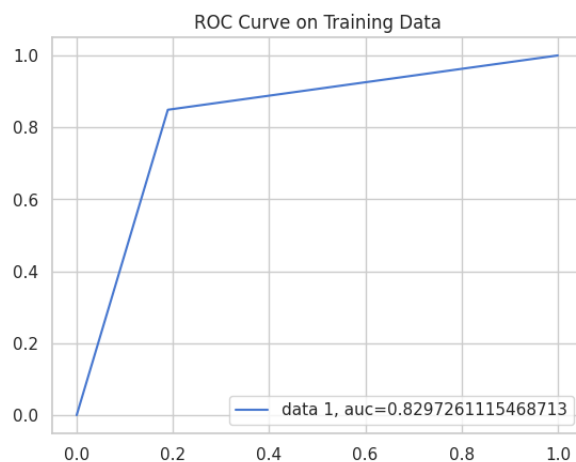
# Confusion Matrix of Test Data



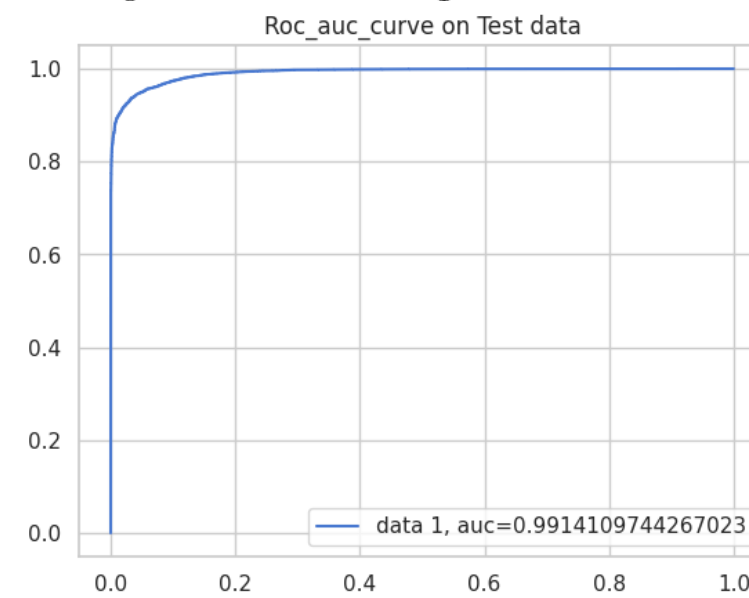
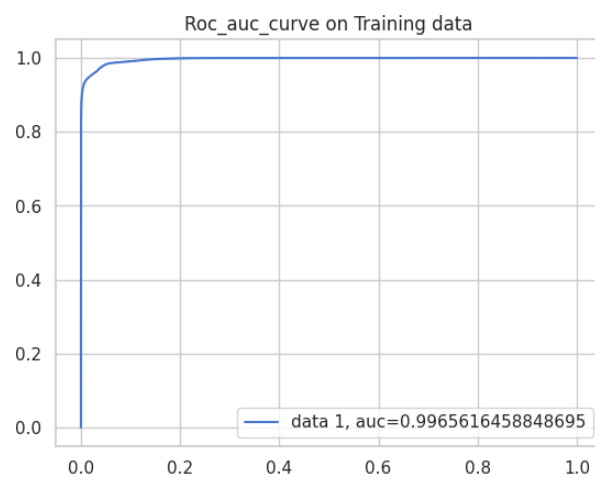
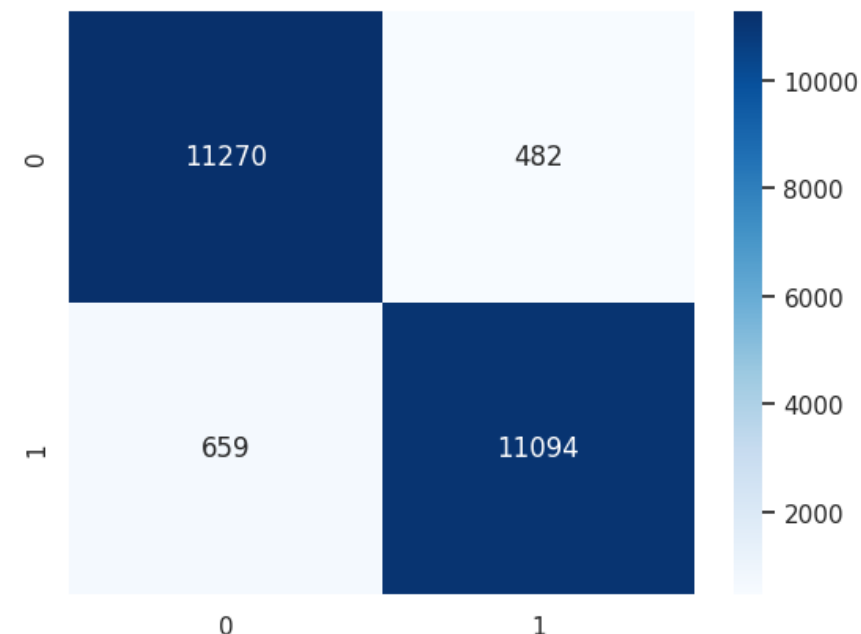
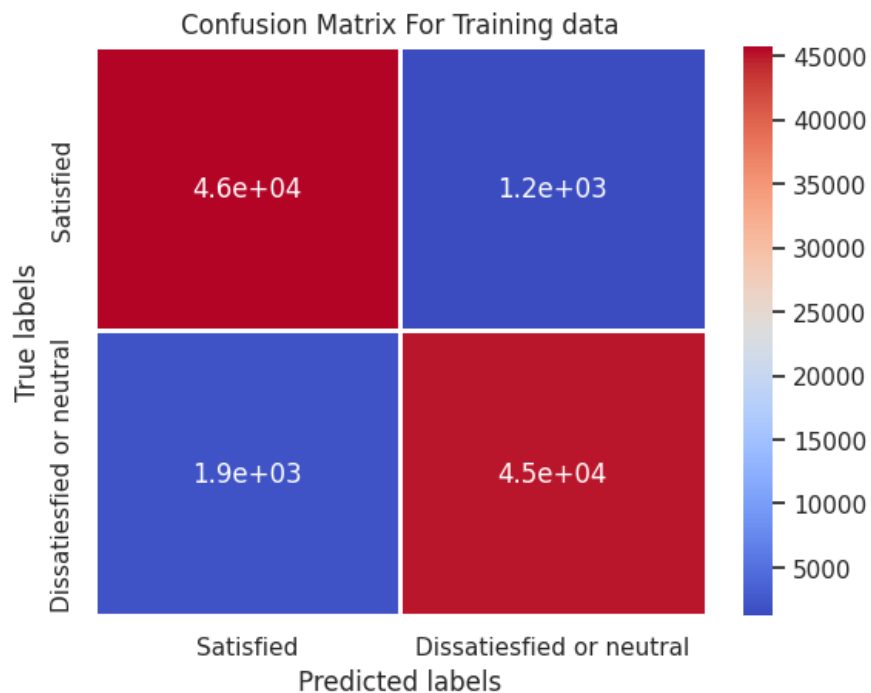
# Logistic Regression with optimization



**Accuracy\_score = 0.8231**

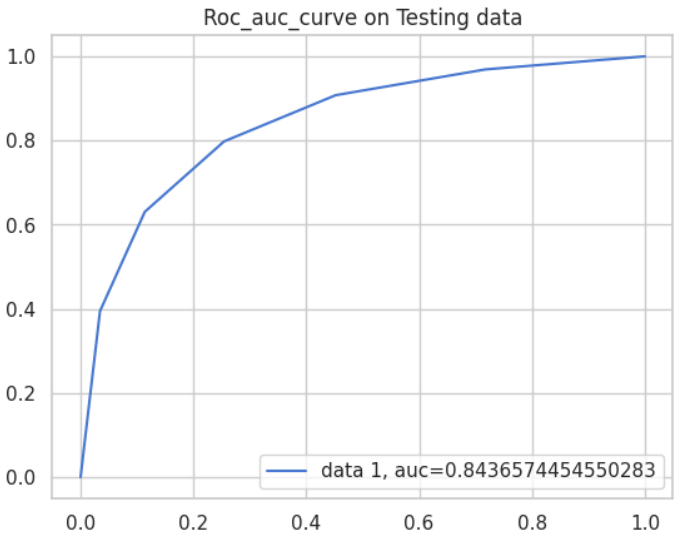
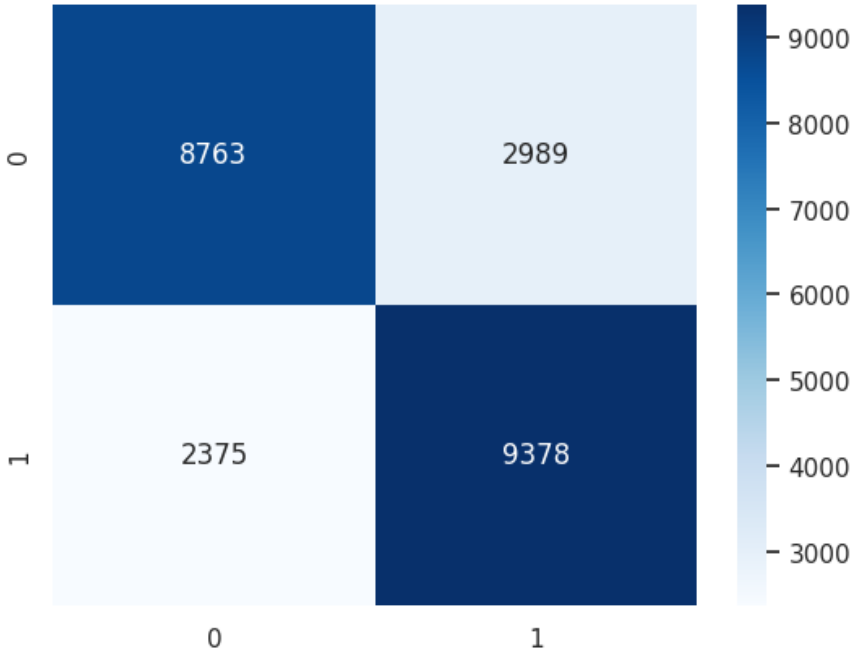
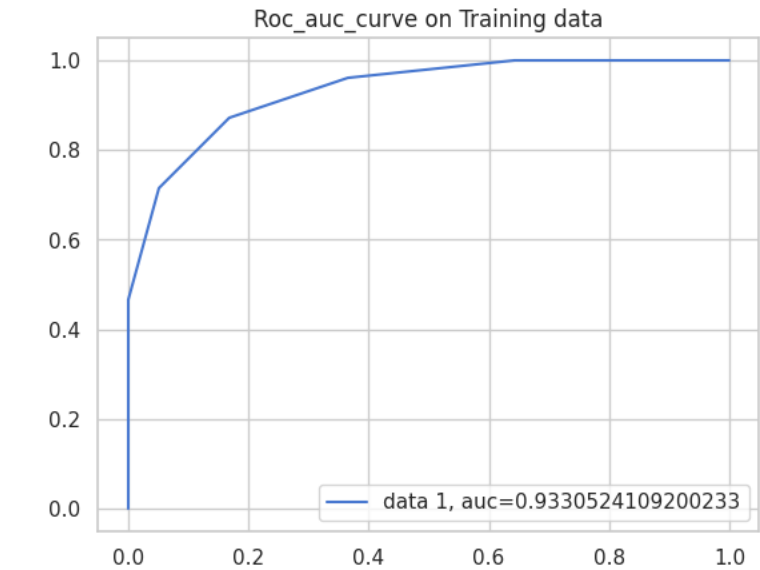
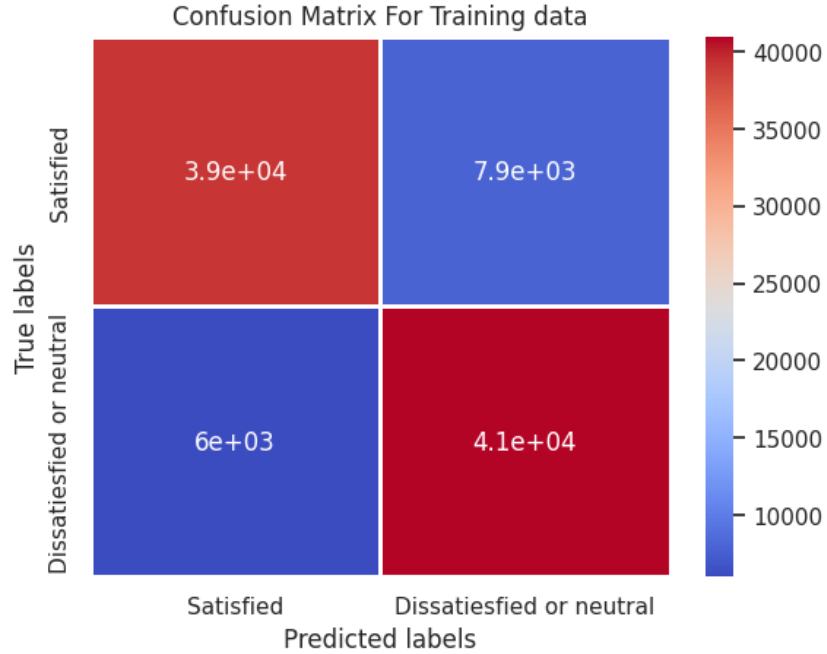


# Random Forest

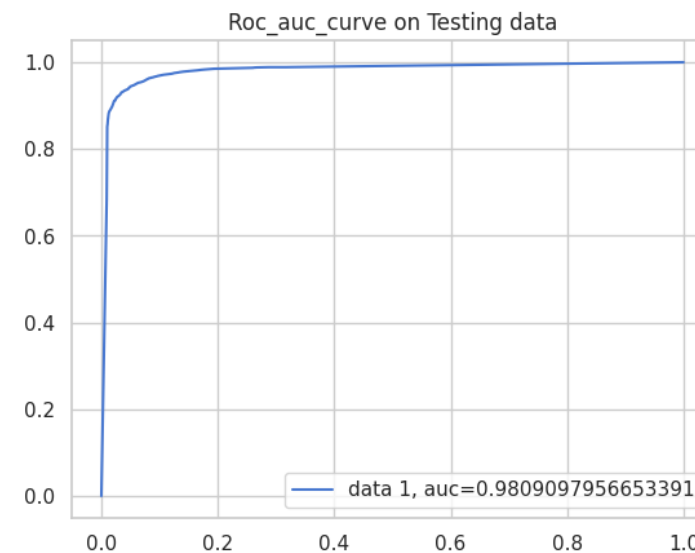
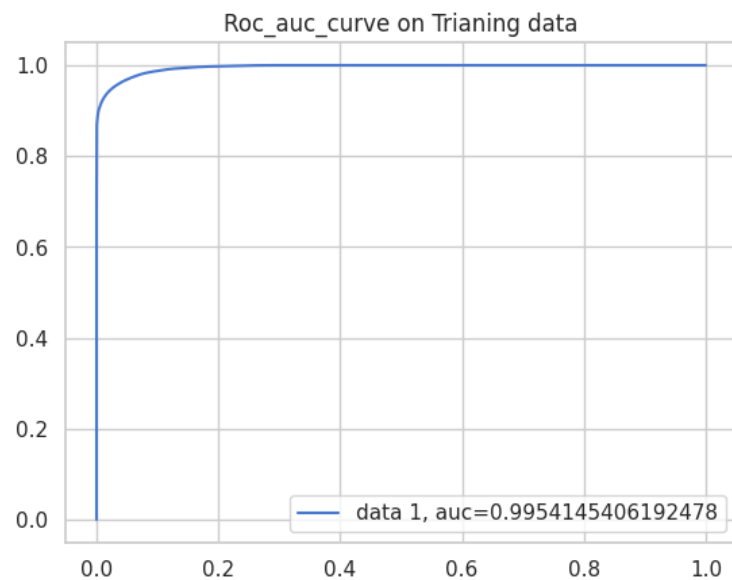
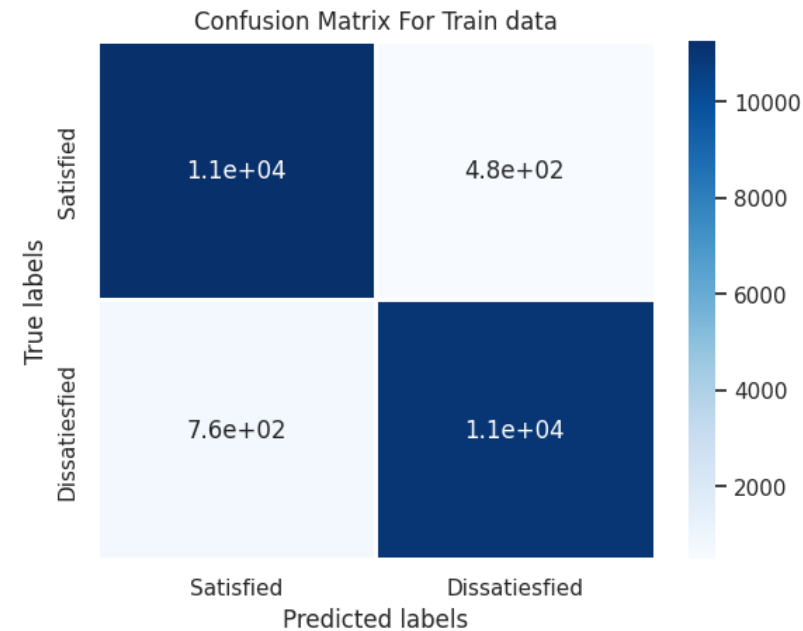
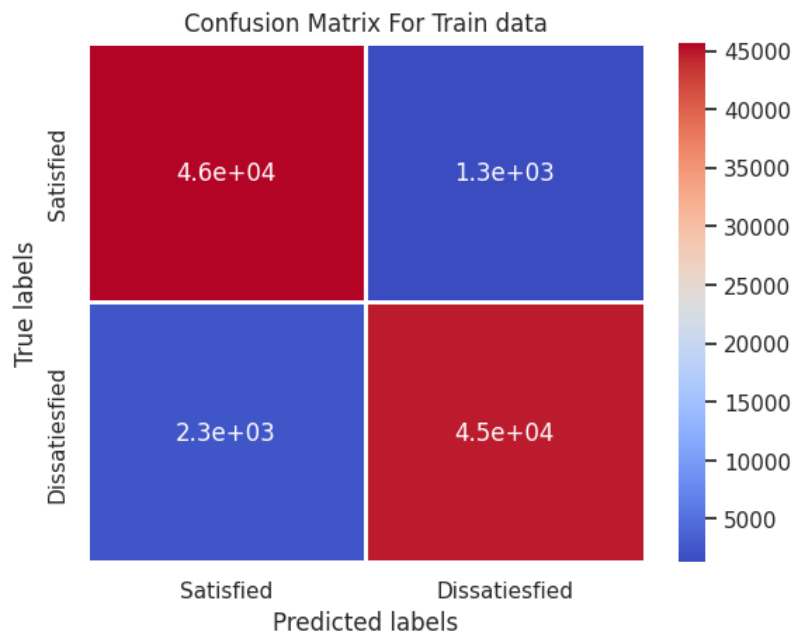


**Accuracy\_score =0.94**

# KNN Classifier



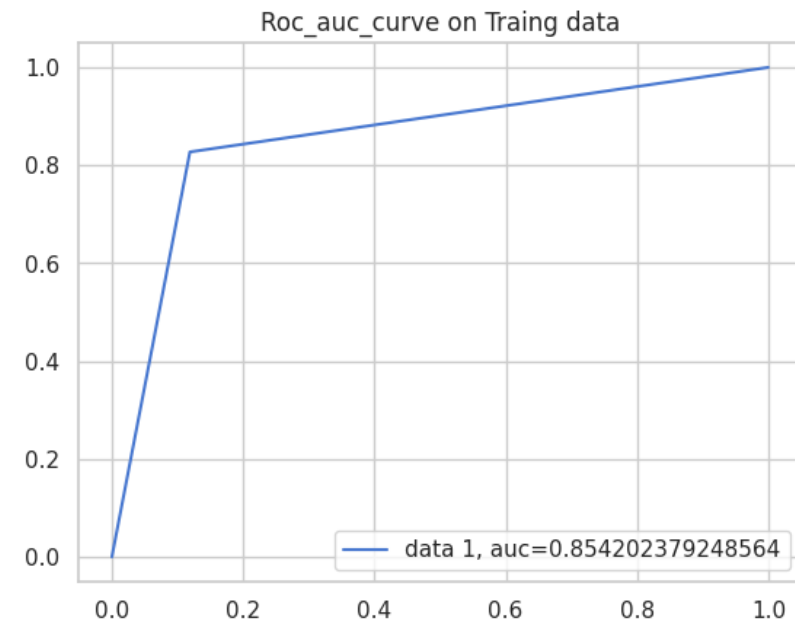
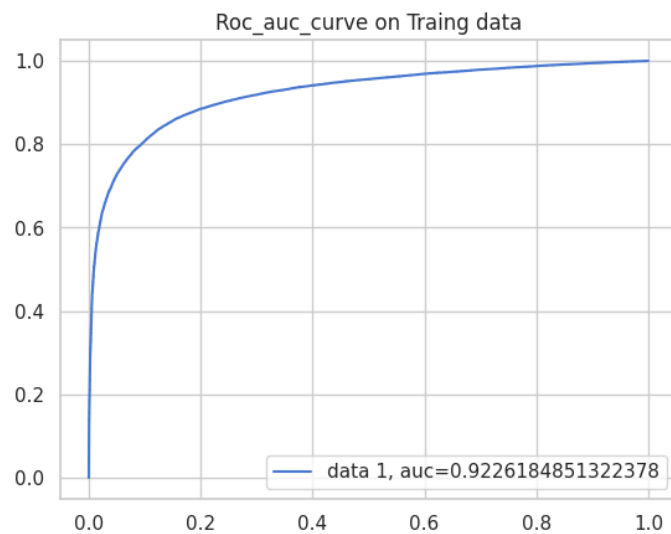
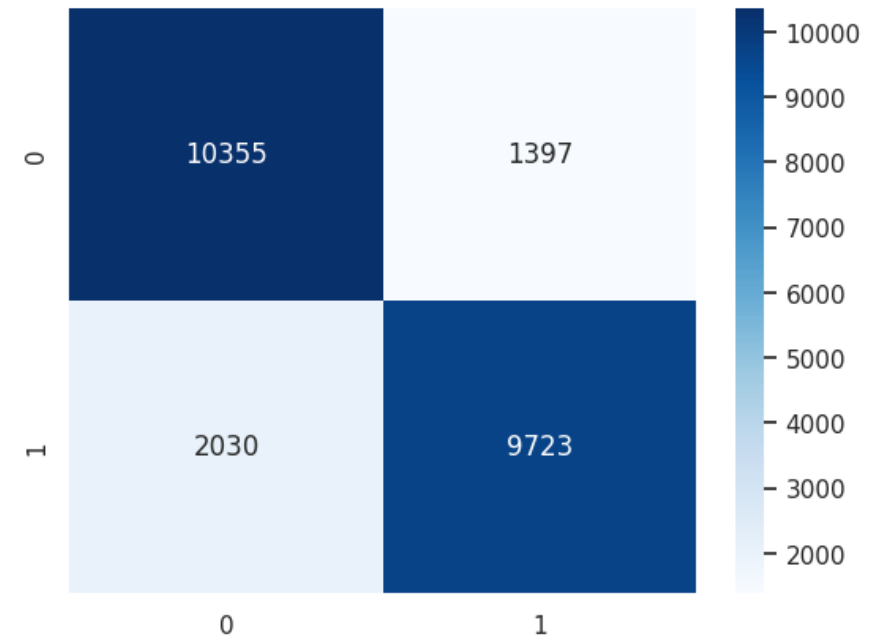
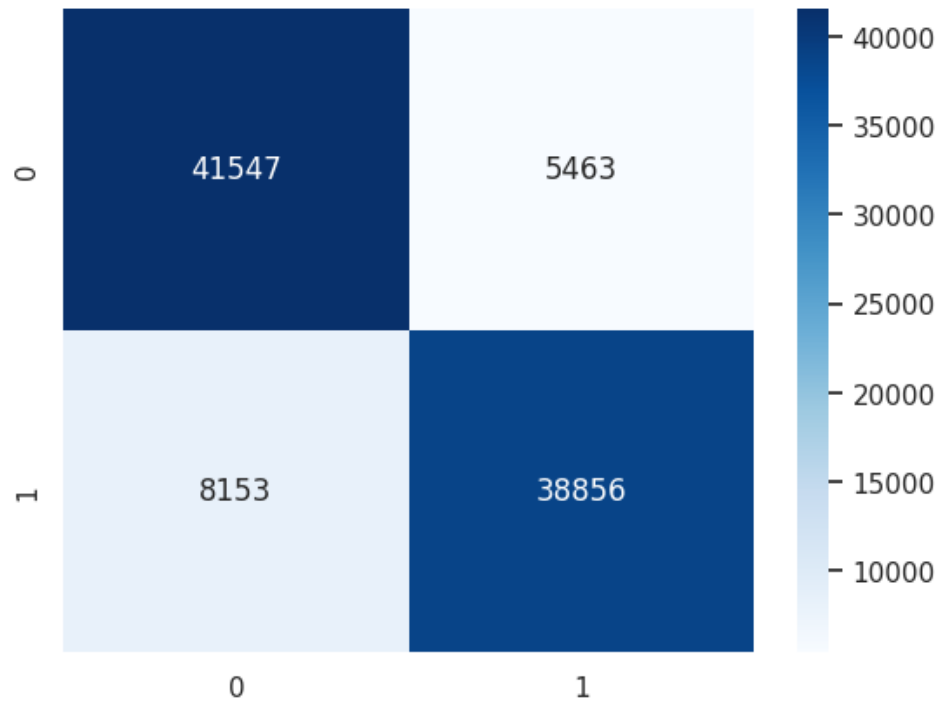
# Decision Tree



**Accuracy score = 0.85**



# Gaussian NB



**Accuracy score = 0.77**

## Model Metrics Without Optimization

Classifier	Accuracy Score	Precision Score	Recall Score	F1 Score	ROC AUC Score
Logistic Regression	0.8231	0.8055	0.8521	0.8281	0.8231
Decision Tree	0.9402	0.9395	0.941	0.9403	0.9402
Random Forest	0.9602	0.9708	0.949	0.9598	0.9602
Gaussian NB	0.8551	0.8758	0.8275	0.851	0.8551
KNN	0.7746	0.7624	0.7981	0.7798	0.7746

## Models Metrics With Optimization

Classifier	Accuracy Score	Precision Score	Recall Score	F1 Score	ROC AUC Score
Logistic Regression	0.8231	0.8055	0.8521	0.8281	0.8231
Decision Tree	0.949	0.9593	0.9378	0.9484	0.949
Random Forest	0.9512	0.9584	0.9433	0.9508	0.9512
Gaussian NB	0.8551	0.8758	0.8275	0.851	0.8551
KNN	0.7746	0.7624	0.7981	0.7798	0.7746

# Conclusion

From the provided results, it appears that the Random Forest classifier achieved the highest accuracy, 0.96 precision, recall, F1 score, and ROC AUC score among the classifiers listed. However, it's important to consider other factors such as the specific problem domain, dataset characteristics, and business requirements when selecting the best classifier

## Potential Benefits

**Improving customer experience:** By analyzing passenger satisfaction data, airlines can gain insights into the factors that contribute to a positive customer experience. This information can help airlines identify areas of improvement and make informed decisions to enhance the quality of their services.

**Identifying trends and patterns:** Analyzing large-scale passenger satisfaction data sets can reveal trends and patterns in customer preferences and behaviors. Airlines can use this information to identify emerging trends, adapt their services accordingly, and stay ahead of the competition.

**Enhancing operational efficiency:** Passenger satisfaction data can provide airlines with insights into the efficiency of their operations. By identifying pain points and bottlenecks, airlines can optimize their processes, reduce delays, and improve overall operational efficiency

## Suggestions

- Increase the facilities in Food and Drink, online-on boarding, Cleanliness
- Provide the best services to Economy and Economy Plus class Passengers .
- Provide the Food and Drinks tasty and Healthy for all age group passengers







# Thank You!