# The Impact of Demographics, Service Quality, and Travel Experience on Airline Passenger Satisfaction

Joseph Ogunlade
Adriana Ortega
Amir Khan
Samanth Madhesh

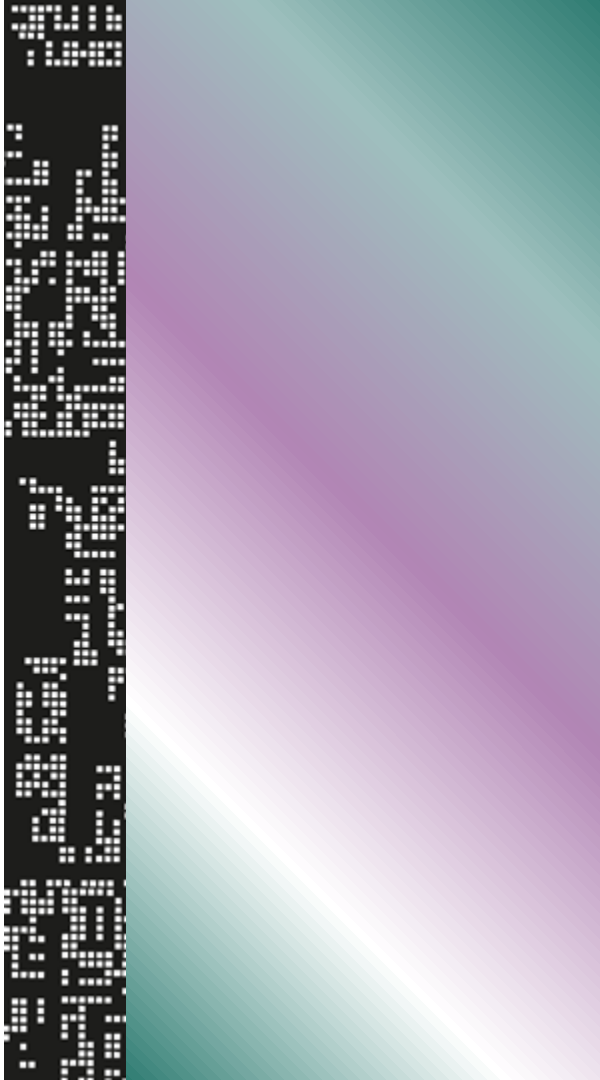
Marketing Analytics Group Project

# Table of contents



"**All** models are wrong, **but some are useful**"

—George E. P. Box



# 01 Problem Statement

# Problem Statement

1. What are the key factors affecting passenger satisfaction?
    1.1 How do these factors vary by demographics, travel purpose, and flight experience?

2. Can we use these factors to predict whether a customer is satisfied or not?

# Hypothesis 02

# Hypothesis

**$H_1$:** Passengers' satisfaction levels significantly vary by age, with younger passengers placing higher importance on flight experience.
**$H_2$:** Passengers' satisfaction levels significantly vary by gender, with potential differences in preferences for service aspects.
**$H_3$:** Business-class passengers report significantly higher satisfaction compared to economy-class passengers due to better service experiences.
**$H_4$:** Business travelers report significantly higher satisfaction compared to personal travelers due to the prioritization of convenience and efficiency.
**$H_5$:** Loyal customers are more likely to report higher satisfaction than disloyal customers.
**$H_6$:** Higher ratings in inflight service factors, such as Wi-Fi quality, ease of online booking, cleanliness, and flight punctuality, significantly predict overall passenger satisfaction.

# Objectives

03

# Objectives

- Conduct **Exploratory Data Analysis (EDA)** to uncover patterns, trends, and insights within the data.
- Perform **Customer Segmentation Analysis** to identify distinct customer groups based on behavioral and demographic factors.
- Utilize **Logistic Regression Analysis** to examine the relationships between independent variables and the target variable, satisfaction.
- Develop and evaluate **Logistic Regression** and **Random Forest Models** for predicting customer satisfaction with a focus on accuracy and interpretability.

# Data 04

# Data

Here is how we grouped our data:
- Demographic Attributes: Gender, Age, Customer Type.
- Flight Experience Factors: Travel purpose, class, flight distance.
- Service Satisfaction Ratings: Wi-Fi, online booking, boarding, seat comfort, food, and cleanliness.
- Logistics: Delays on departure and arrival.

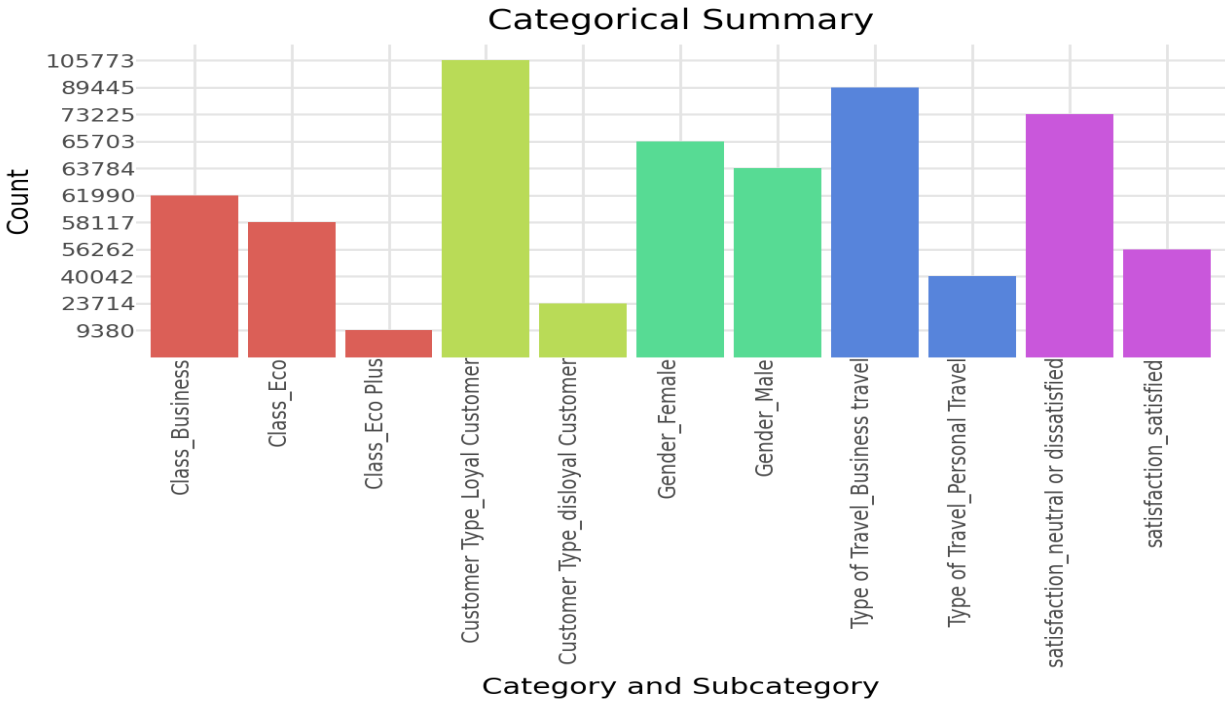
# Data (Summary Statistics)

| | Age | Flight Distance | Inflight wifi service | Departure/Arrival time convenient | Ease of Online booking | Gate location | Food and drink | Online boarding | Seat comfort |
|---|---|---|---|---|---|---|---|---|---|
| count | 129487 | 129487 | 129487 | 129487 | 129487 | 129487 | 129487 | 129487 | 129487 |
| mean | 39.43 | 1190.21 | 2.73 | 3.06 | 2.76 | 2.98 | 3.2 | 3.25 | 3.44 |
| std | 15.12 | 997.56 | 1.33 | 1.53 | 1.4 | 1.28 | 1.33 | 1.35 | 1.32 |
| min | 7 | 31 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 25% | 27 | 414 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 50% | 40 | 844 | 3 | 3 | 3 | 3 | 3 | 3 | 4 |
| 75% | 51 | 1744 | 4 | 4 | 4 | 4 | 4 | 4 | 5 |
| max | 85 | 4983 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |

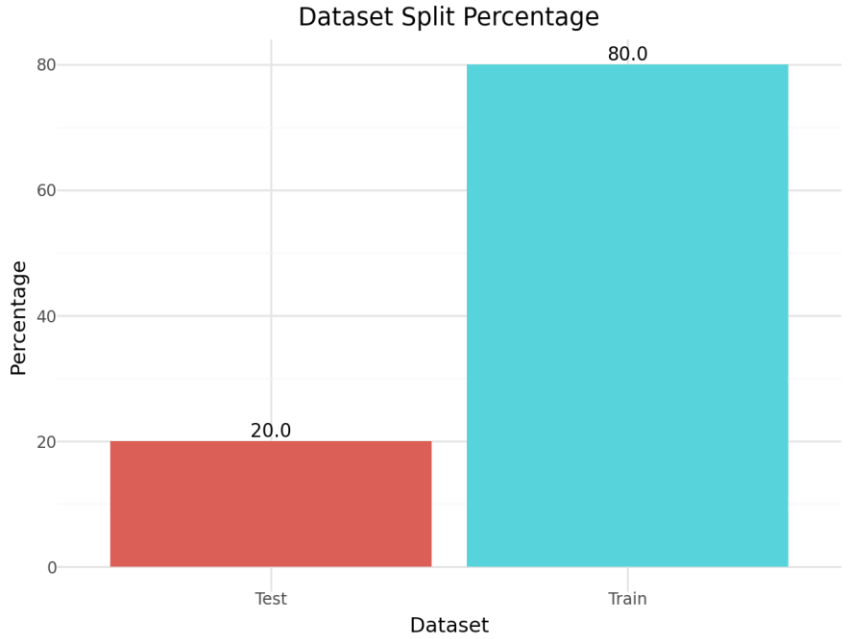
# Data (Summary Statistics)

| | Inflight entertainment | On-board service | Leg room service | Baggage handling | Checkin service | Inflight service | Cleanliness | Departure Delay in Minutes | Arrival Delay in Minutes |
|---|---|---|---|---|---|---|---|---|---|
| count | 129487 | 129487 | 129487 | 129487 | 129487 | 129487 | 129487 | 129487 | 129487 |
| mean | 3.36 | 3.38 | 3.35 | 3.63 | 3.31 | 3.64 | 3.29 | 14.64 | 15.09 |
| std | 1.33 | 1.29 | 1.32 | 1.18 | 1.27 | 1.18 | 1.31 | 37.93 | 38.47 |
| min | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 25% | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 0 | 0 |
| 50% | 4 | 4 | 4 | 4 | 3 | 4 | 3 | 0 | 0 |
| 75% | 4 | 4 | 4 | 5 | 4 | 5 | 4 | 12 | 13 |
| max | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 1592 | 1584 |

# Data (Categorical Summaries)



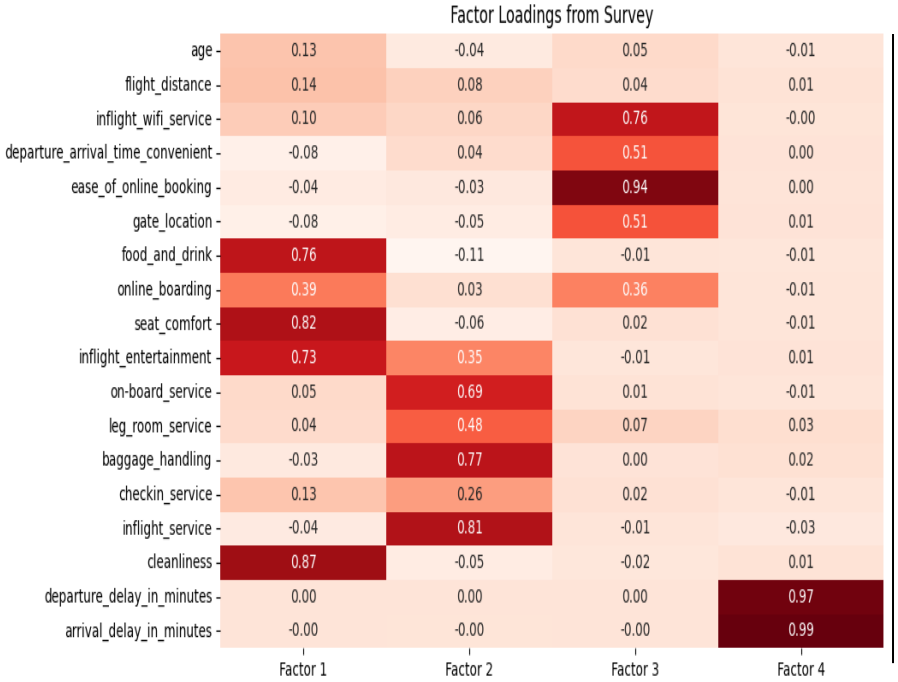
# Data (Dataset Split)



# Exploratory Data Analysis

05

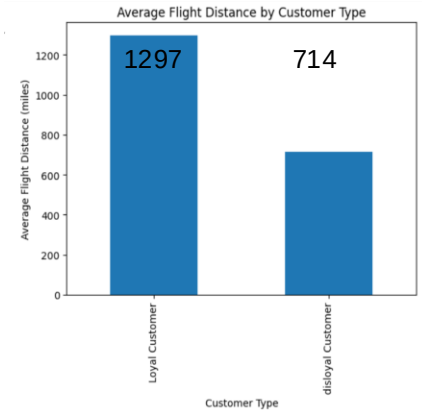
# EDA (Factor analysis)

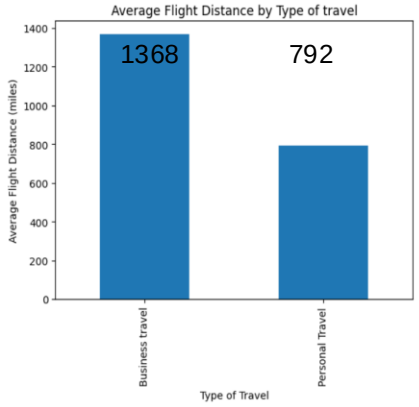



After the factor analysis, this is our new group

- **Flight Experience:** Seat Comfort, Inflight Entertainment, Food and Drink, Cleanliness
- **Service satisfaction:** On-Board Service, Leg Room Service, Baggage Handling
- **Digital and Accessibility Features:** Inflight Wi-Fi Service, Ease of Online Booking, Gate Location, Online Boarding, Departure/Arrival Time Convenience
- **Logistics:** Departure Delay in Minutes, Arrival Delay in Minutes
- **Demography & other factors:** Age, Flight Distance, Gender, Customer Type, Type of travel, Class

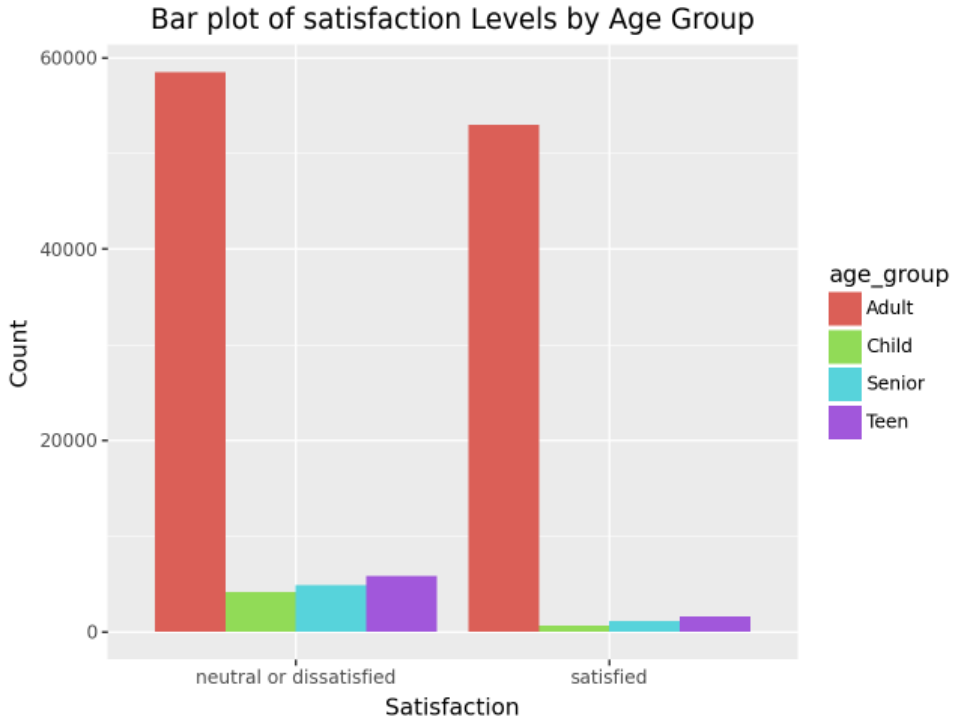
# EDA



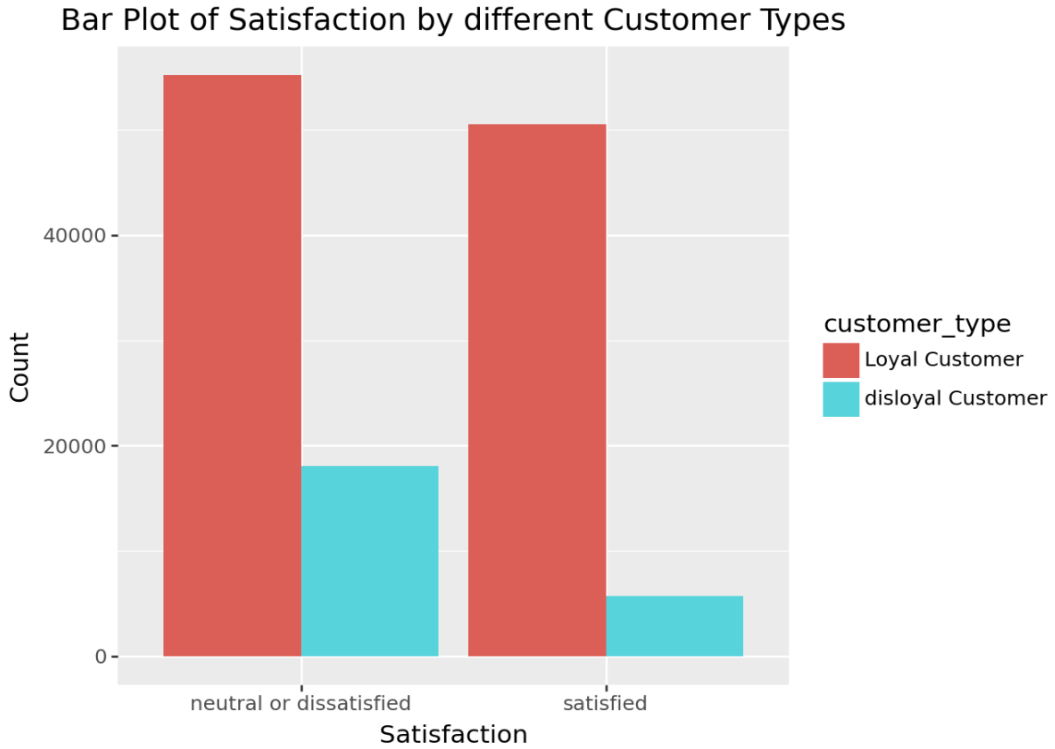




| age group | | Avg flight distance (miles) | Avg Depart. Delay(min) | Avg Arriv. Delay(min) |
|---|---|---|---|---|
| Adult | 20-64 | 1230.683534 | 14.610271 | 15.045807 |
| Child | 0-12 | 905.119262 | 15.287990 | 15.865437 |
| Senior | +65 | 958.301374 | 14.060677 | 14.377305 |
| Teen | 13-19 | 950.844776 | 15.198236 | 15.852646 |

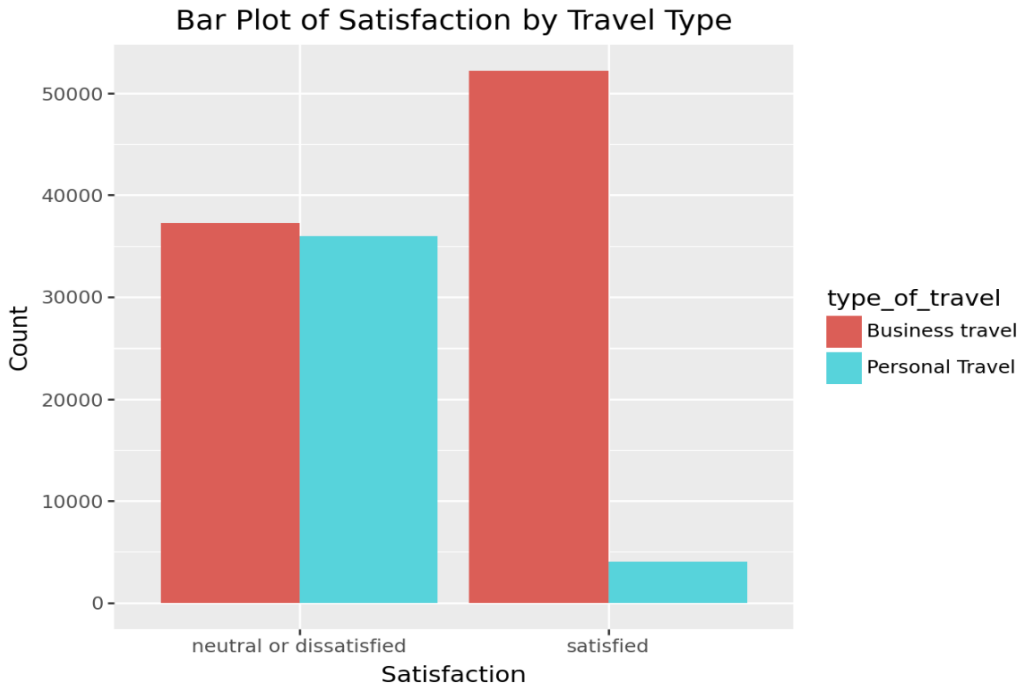
# EDA


Bar plot of satisfaction Levels by Age Group
60000
40000
20000
0
Count
neutral or dissatisfied
satisfied
Satisfaction
age_group
Adult
Child
Senior
Teen

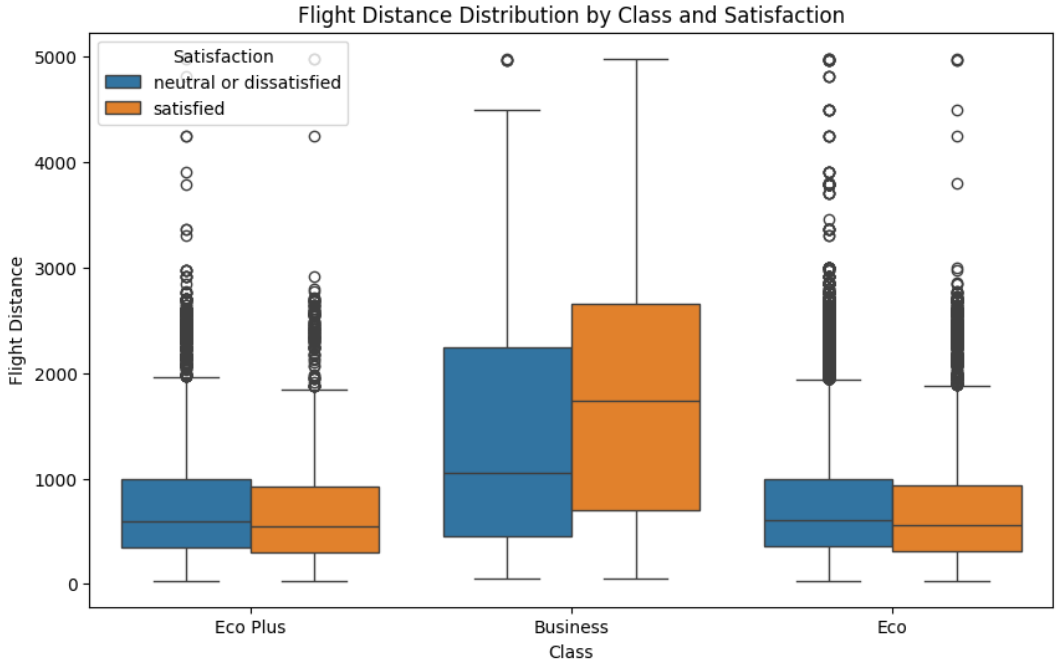
# EDA


Bar Plot of Satisfaction by different Customer Types
Count
40000
20000
0
neutral or dissatisfied
satisfied
Satisfaction
customer_type
Loyal Customer
disloyal Customer

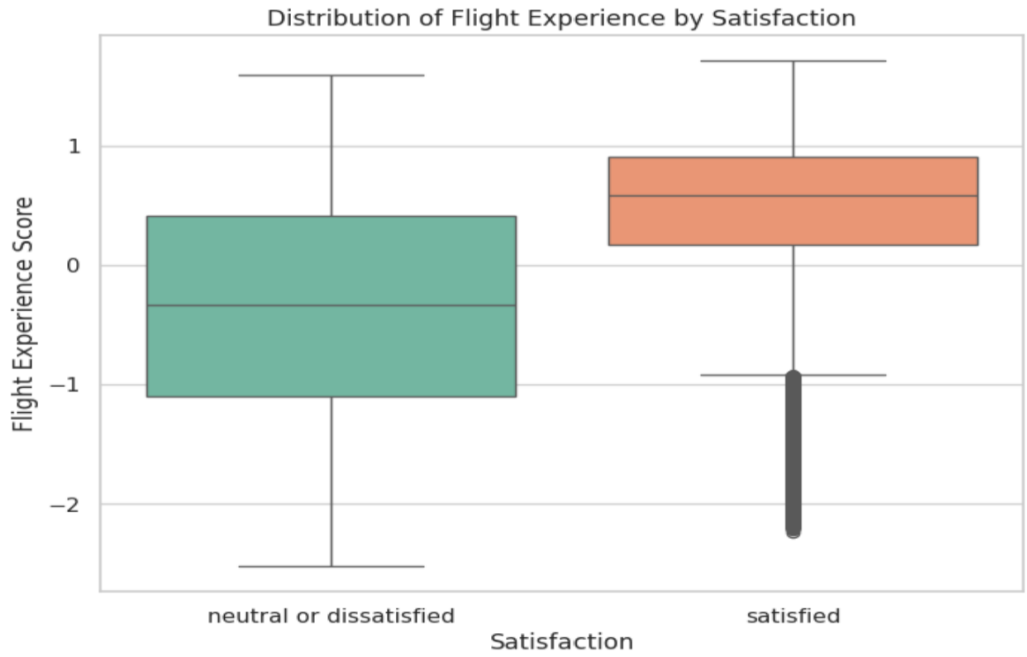
# EDA


Bar Plot of Satisfaction by Travel Type
50000
40000
30000
20000
10000
0
Count
neutral or dissatisfied
satisfied
Satisfaction
type_of_travel
Business travel
Personal Travel

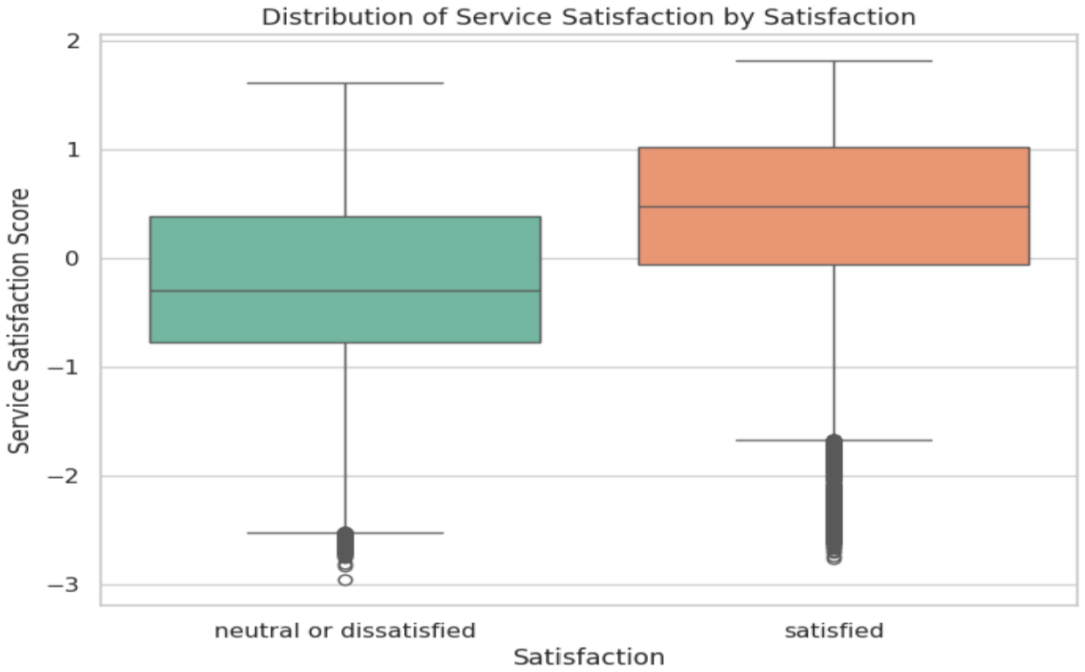
# EDA


Flight Distance Distribution by Class and Satisfaction
Satisfaction
neutral or dissatisfied
satisfied
Flight Distance
5000
4000
3000
2000
1000
0
Eco Plus
Business
Eco
Class

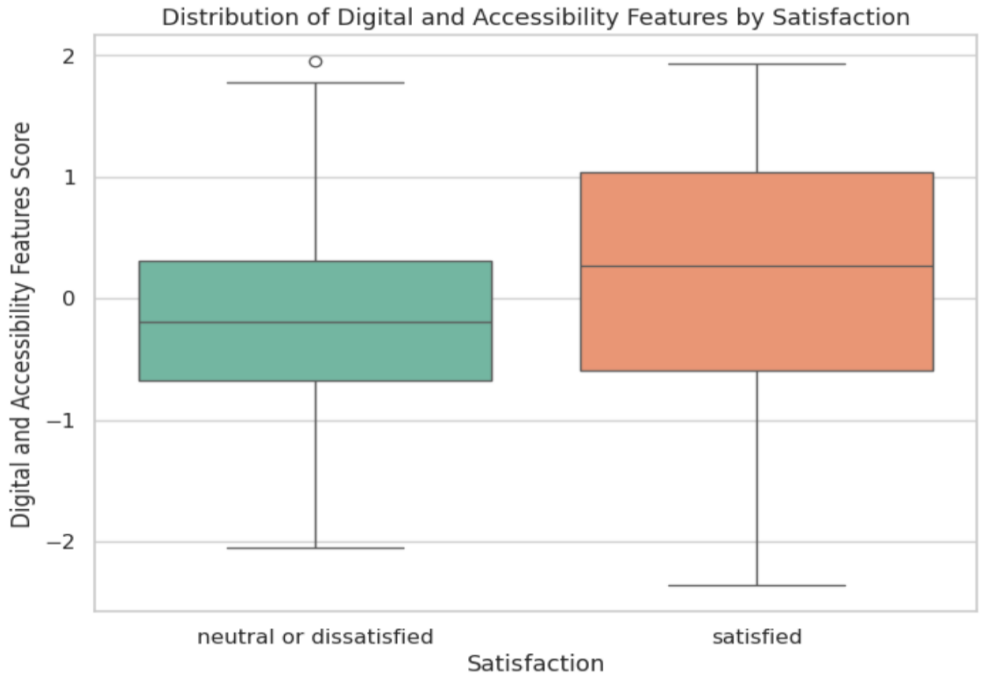
# EDA


Distribution of Flight Experience by Satisfaction
Flight Experience Score
1
0
−1
−2
neutral or dissatisfied
satisfied
Satisfaction

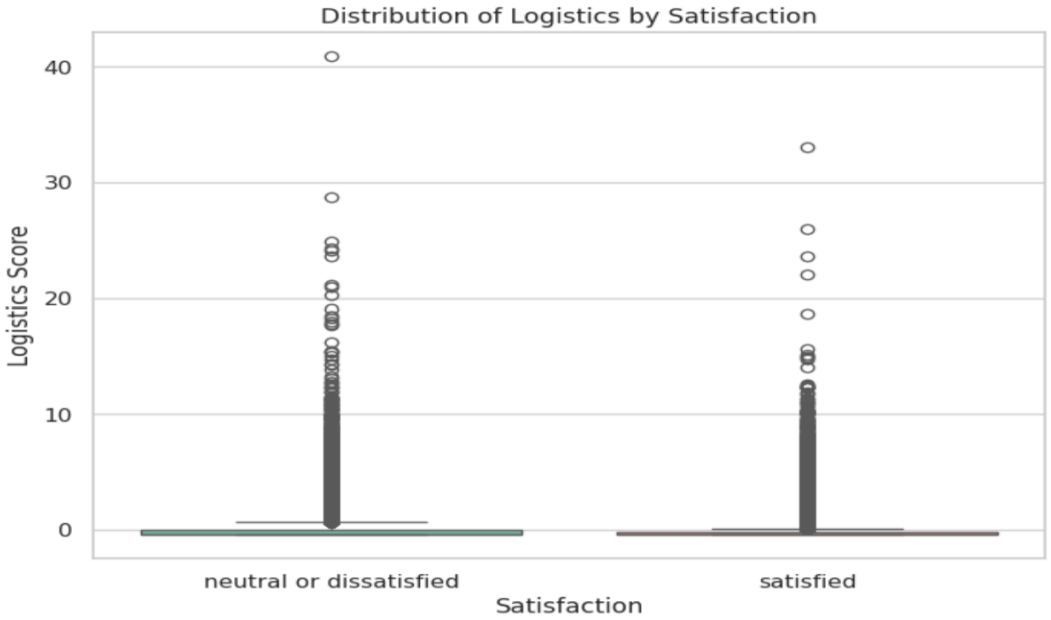
# EDA


Distribution of Service Satisfaction by Satisfaction
Service Satisfaction Score
2
1
0
−1
−2
−3
neutral or dissatisfied
satisfied
Satisfaction

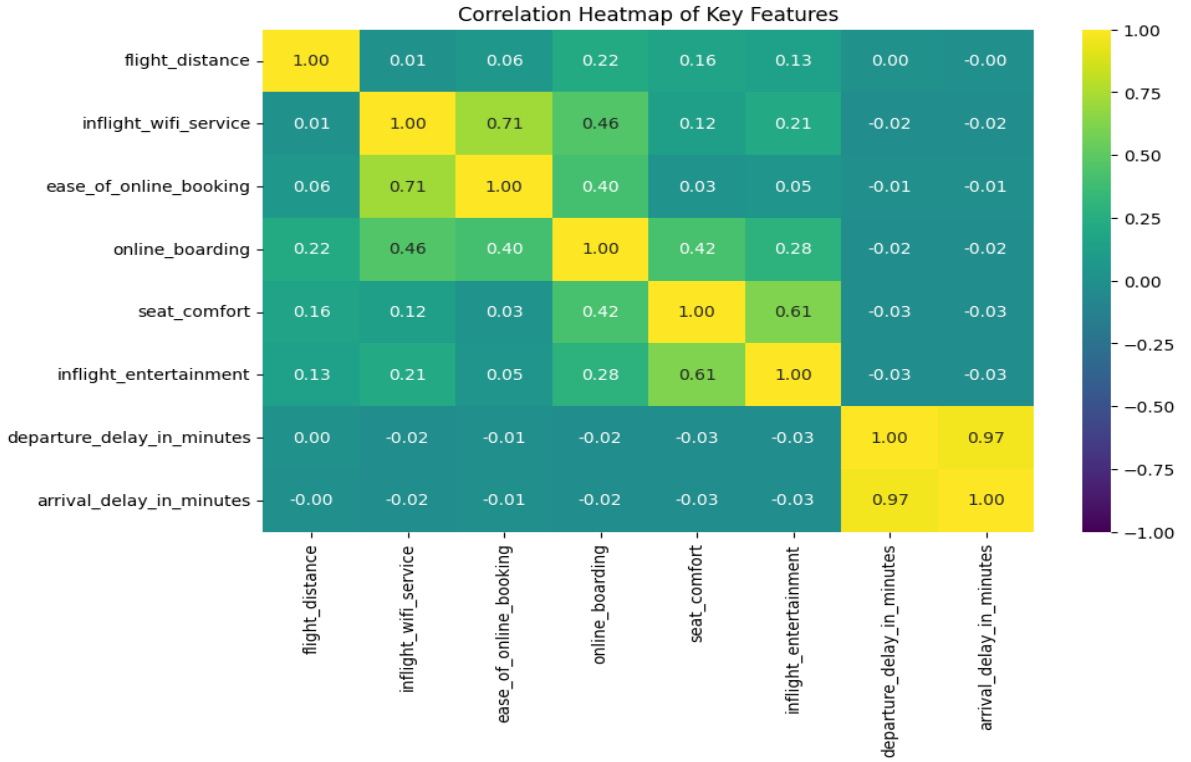
# EDA


Distribution of Digital and Accessibility Features by Satisfaction
Digital and Accessibility Features Score
2
1
0
−1
−2
neutral or dissatisfied
satisfied
Satisfaction

# EDA


Distribution of Logistics by Satisfaction
Logistics Score
40
30
20
10
0
neutral or dissatisfied
satisfied
Satisfaction

# EDA


Correlation Heatmap of Key Features
flight_distance
inflight_wifi_service
ease_of_online_booking
online_boarding
seat_comfort
inflight_entertainment
departure_delay_in_minutes
arrival_delay_in_minutes
1.00 0.01 0.06 0.22 0.16 0.13 0.00 -0.00
0.01 1.00 0.71 0.46 0.12 0.21 -0.02 -0.02
0.06 0.71 1.00 0.40 0.03 0.05 -0.01 -0.01
0.22 0.46 0.40 1.00 0.42 0.28 -0.02 -0.02
0.16 0.12 0.03 0.42 1.00 0.61 -0.03 -0.03
0.13 0.21 0.05 0.28 0.61 1.00 -0.03 -0.03
0.00 -0.02 -0.01 -0.02 -0.03 -0.03 1.00 0.97
-0.00 -0.02 -0.01 -0.02 -0.03 -0.03 0.97 1.00
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

# Statistical Analysis/ML Workflow

06

# Customer Segmentation Analysis

**Objective:**

To segment customers based on flight data (e.g., flight distance, delays, satisfaction) and provide actionable insights for improving customer satisfaction and operational efficiency.

**Methodology Overview:**

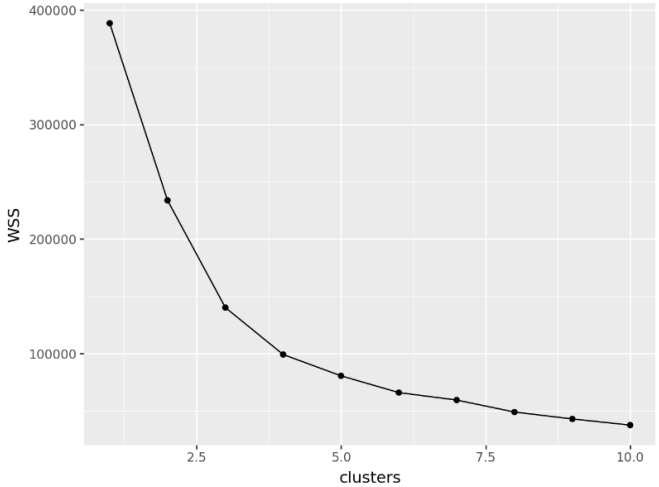
- Data Preprocessing and Feature Selection
- Determining the Optimal Number of Clusters (Elbow Method)
- Customer Segmentation using K-Means
- Validation with Linear Discriminant Analysis (LDA)

# Determining the Optimal Number of Clusters

**Finding the Elbow Point**

- **Method:** Calculated WSS for clusters ranging from 1 to 10.
- **Result:** The elbow was identified at **k=5**, suggesting 5 clusters.

**Visual:**




**Insight:**

Using 5 clusters balances compactness within clusters and simplicity of the model.

# Cluster descriptions

**Characteristics of Each Segment**

**Cluster Statistics:**

| Segment | Mean Flight Distance | Mean Departure Delay (mins) | Mean Arrival Delay (mins) | Size |
|---|---|---|---|---|
| 0 | 664.69 | 3.92 | 4.23 | 81,778 |
| 1 | 1218.76 | 145.57 | 148.06 | 3,671 |
| 2 | 2648.03 | 6.15 | 6.41 | 31,818 |
| 3 | 1323.22 | 347.30 | 350.58 | 551 |
| 4 | 882.85 | 56.03 | 57.19 | 11,669 |

**Key observations:**

- Cluster 0 represents the majority of customers with short flight distances and minimal delays.
- Cluster 3 highlights a smaller group experiencing significant delays.
- Cluster 2 includes long-haul travelers with minimal delays.

# Visualizing segments

**Key differences across clusters.**

| satisfaction | neutral or dissatisfied | satisfied | Total |
|---|---|---|---|
| segment | | | |
| 0 | 52466 | 29312 | 81778 |
| 1 | 2336 | 1335 | 3671 |
| 2 | 10088 | 21730 | 31818 |
| 3 | 341 | 210 | 551 |
| 4 | 7994 | 3675 | 11669 |
| Total | 73225 | 56262 | 129487 |

# Validation with LDA

**Accuracy of Clustering Model**
Used Linear Discriminant Analysis (LDA) to validate cluster assignments.

Result: Achieved an accuracy of **85.52%**, confirming that the segmentation model effectively captures the distinctions between clusters.

# Regression 1 (Satisfaction vs factors)

| | Coef. | Std.Err. | z | P>\|z\| |
|---|---|---|---|---|
| Intercept | -4.3494195 | 0.04533219 | -95.945491 | 0 |
| Age | -0.004486 | 0.00058562 | -7.6602772 | 1.86E-14 |
| Male | 0.03341484 | 0.01612745 | 2.07192343 | 0.03827258 |
| Flight Distance | -4.69E-05 | 9.36E-06 | -5.0117471 | 5.39E-07 |
| Loyal customer | 1.96424532 | 0.02363811 | 83.0965386 | 0 |
| Business Travel | 2.78820846 | 0.02492033 | 111.884886 | 0 |
| Business Class | 1.00592934 | 0.03340027 | 30.117405 | 2.87E-199 |
| Economy Class | 0.11628564 | 0.03270224 | 3.55589256 | 0.0003767 |
| Flight Experience | 0.82616304 | 0.00974665 | 84.7637606 | 0 |
| Service satisfaction | 0.89324244 | 0.01035149 | 86.291157 | 0 |
| Digital and Accessibility Features | 0.3563286 | 0.00860015 | 41.432822 | 0 |
| Logistics | -0.1503217 | 0.0084415 | -17.807461 | 6.19E-71 |

Pseudo R Squared = 0.4448

# Regression 1 (Key insights)

Main Predictors of Satisfaction

- Flight Experience (e.g., seat comfort, food, cleanliness) and Service Satisfaction (e.g., on-board service, baggage handling) are strong positive drivers of passenger satisfaction.

Customer Demographics

- Older passengers and females tend to have slightly lower satisfaction levels.
- Loyal customers and business travelers are much more likely to be satisfied compared to disloyal customers and those traveling for personal reasons.
- Business class passengers are the most satisfied, followed by economy class passengers.
- Economy plus passengers show a slight increase in satisfaction compared to economy plus.

Other Factors

- Delays (departure and arrival) significantly reduce passenger satisfaction, with higher delays leading to lower satisfaction levels.
- Digital and accessibility features (e.g., Wi-Fi, ease of online booking) have a positive but moderate impact on satisfaction.

# Regression 2

- We created a highly detailed logistic regression model, with many predictors as the previous one.
- We included interaction terms between age, gender, class, flight distance, customer type, type of travel and all four factors of Flight Experience, Service satisfaction, Digital and Accessibility Features, and Logistics.
- The intent is to capture nuanced relationships in the data.
- Pseudo R Squared = 0.5477

# Regression 2 (Extra insights)

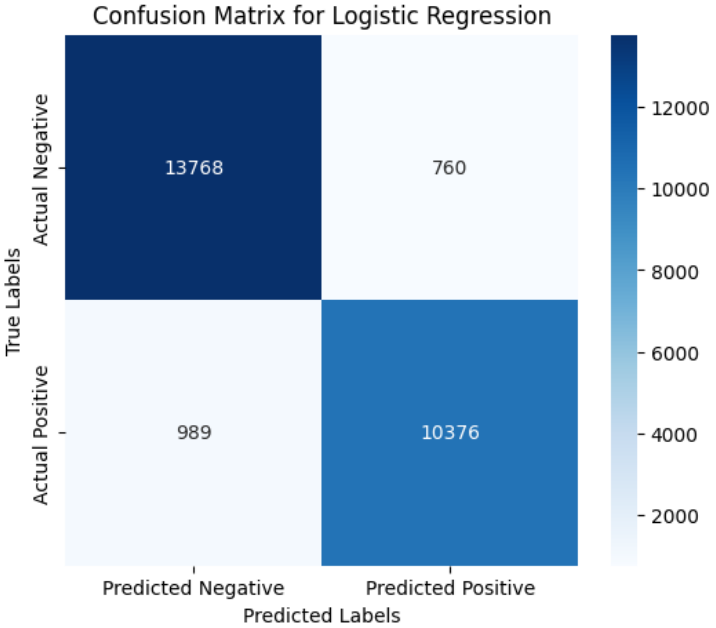
- Older passengers show higher satisfaction with flight experience and lower satisfaction with service satisfaction factor.
- Women report lower satisfaction with service satisfaction factors but higher satisfaction with other service aspects.
- Business class passengers are most satisfied, with service factors related to flight experience and logistics.
- **Economy plus** passengers show slightly higher satisfaction for flight experiences like seating and comfort.
- **Frequent flyers** (loyal customers) report higher satisfaction, especially with **flight experience.**
- **Business travelers** have higher satisfaction compared to personal travelers, especially with **flight experience** and **digital & accessibility features.**
- Longer **flight distances** impact satisfaction, particularly with **flight experience**, making service quality more critical on long flights.

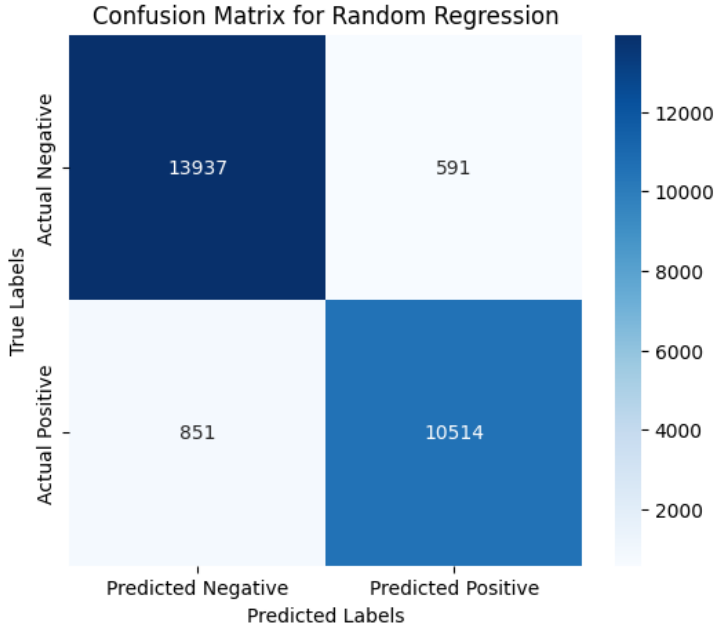
# Regression (Diagnostics)

| feature | VIF |
|---|---|
| Const | 31.2 |
| Age | 1.123 |
| Male | 1.003 |
| Flight distance | 1.362 |
| Loyal customers | 1.478 |
| Business Travel | 1.868 |
| Business Class | 4.682 |
| Economy Class | 4.048 |
| Flight Experience | 1.166 |
| Service Satisfaction | 1.153 |
| Digital & Accessibility Features | 1.034 |
| Logistics | 1.002 |

Overall, the VIFs suggest that your features are well-behaved in terms of multicollinearity and our model seems reliable.

# ML (Logistics Regression vs Random Forest)



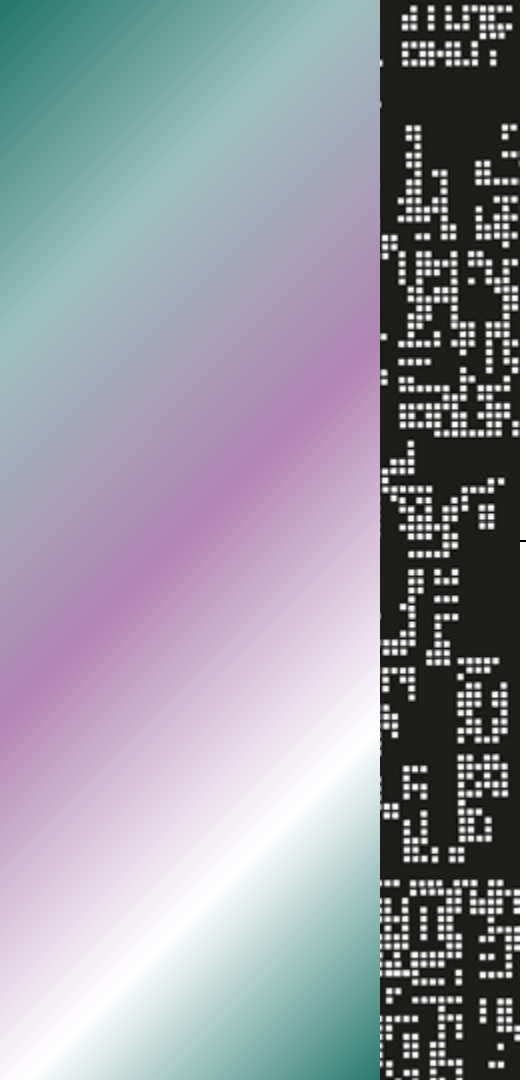

- Random Forest demonstrates higher accuracy (94.4%) compared to Logistic Regression (93.3%).
- Overall, Random Forest performs slightly better because it is capable of capturing non-linear relationships and interactions between features, which Logistic Regression often misses

# Conclusion & Recommendation

07

# Conclusion & Recommendation

- Enhance Flight Experience and Service Quality
- Address Delays Proactively
- Cater to Demographic Preferences
- Leverage Loyalty Programs and Business Travel Insights
- Optimize Service for Long Flights (loyalty programs)
- Strengthen Digital and Accessibility Features
- Focus efforts on reducing significant departure and arrival delays
- Enhance satisfaction through tailored offers or in-flight experiences



# Thank you!

Do you have any questions?