

# What factors lead to customer satisfaction for an Airline?

Group 4 - Hyunwoo Jung ,Dongchan Kim, Serim Ahn

# Introduction

1. Analyze differences in customer satisfaction between short-haul ( $<500$  miles) and medium & long-haul ( $\geq 500$  miles) flights
2. Identify how customer characteristics and service factors influence satisfaction
3. Build a predictive model to classify satisfaction (satisfied vs. dissatisfied)

# Dataset

## Data source

Kaggle - Airline Passenger Satisfaction dataset

## Key Variables

Features - Customer Attributes ( Gender, Age, Customer Type),

- Flight Experience (Flight Distance, Seat Comfort)

Target - satisfaction ( Dissatisfied / Satisfied)

## Data processing

Removed flight data with distances less than 100 miles

Dropped duplicate or low-impact features (e.g., Arrival Delay, Gate Location, Inflight Service)

## Data Split Criteria

Data was split based on flight distance.

 TJ KLEIN · UPDATED 5 YEARS AGO

▲ 950 | ◀ Code |  Download | ⋮

## Airline Passenger Satisfaction

What factors lead to customer satisfaction for an Airline?



Data Card | Code (426) | Discussion (11) | Suggestions (1)

---

### About Dataset

Usability ⓘ  
9.41

### Context

This dataset contains an airline passenger satisfaction survey. What factors are highly correlated to a satisfied (or dissatisfied) passenger? Can you predict passenger satisfaction?

### Content

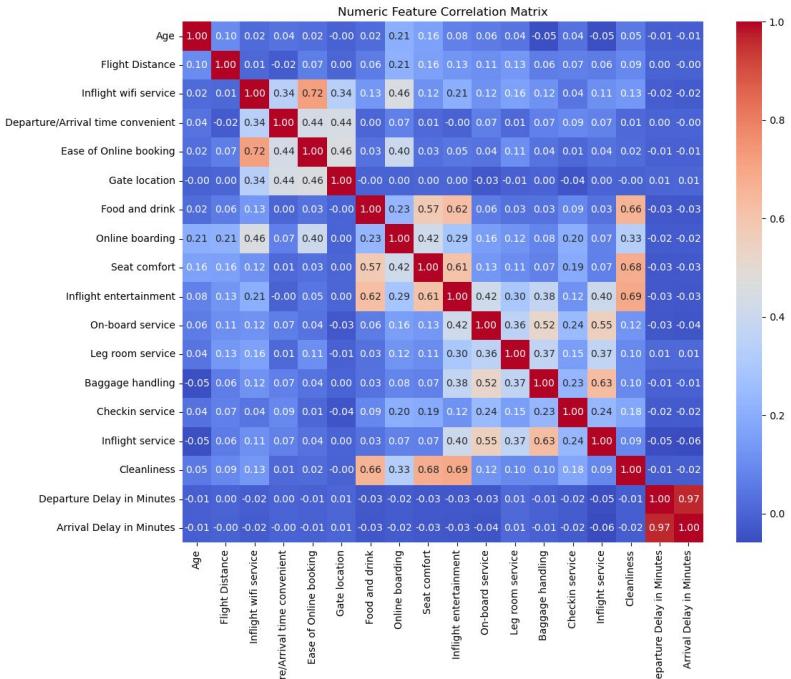
Gender: Gender of the passengers (Female, Male)  
Customer Type: The customer type (Loyal customer, disloyal customer)

License  
Other (specified in description)

Expected update frequency  
Never

Tags  
 Tabular |  Classification  
 Binary Classification

# correlation analysis

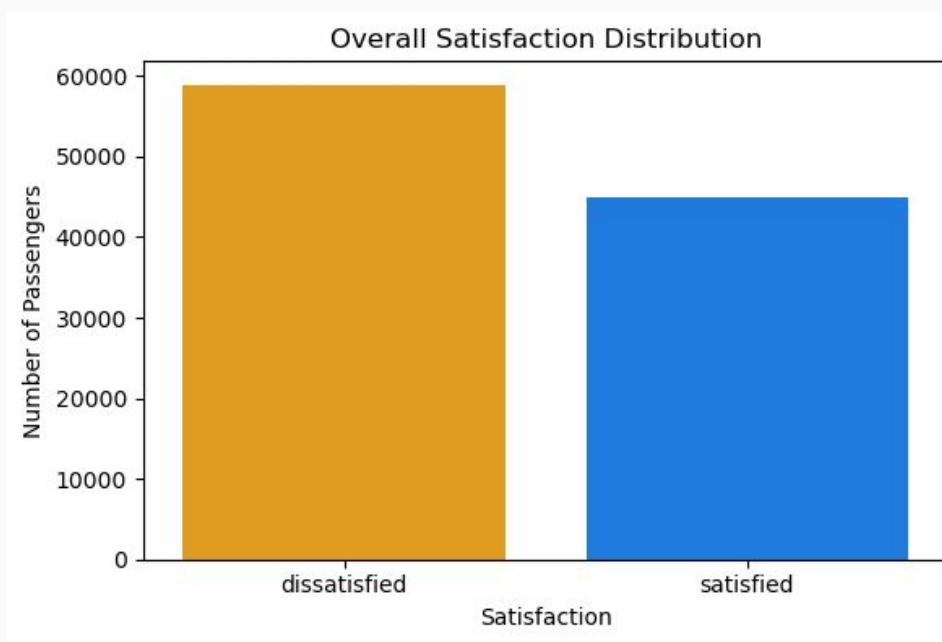


A correlation heatmap showed that seat comfort, inflight entertainment, cleanliness, leg room, and food and drink all correlated positively.

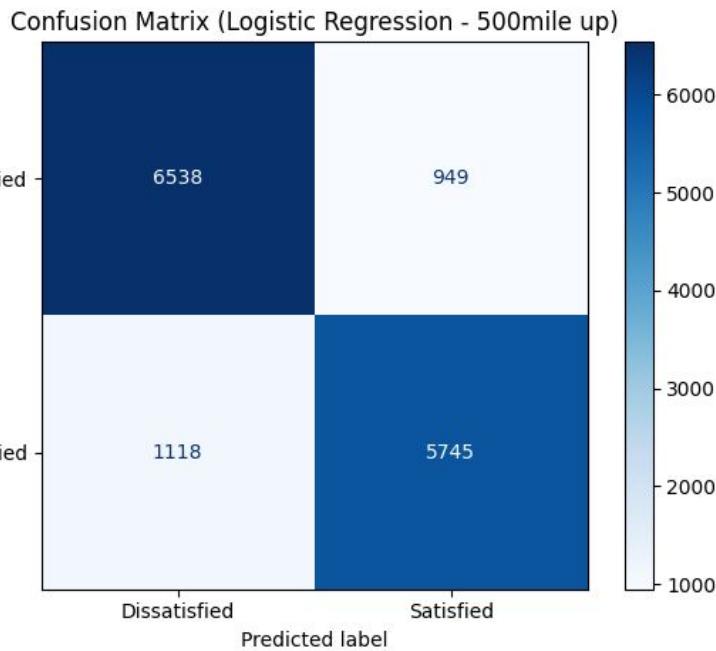
This suggests overall service quality influences satisfaction.

```
# List of unnecessary features
drop_cols = [
    'Arrival Delay in Minutes', # Duplicated with Departure Delay
    'Gate location', # Very low correlation
    'Inflight service', # Potential redundancy
    'Customer Type', 'Type of Travel' # Highly correlated with satisfaction
]
```

# Satisfaction Distribution



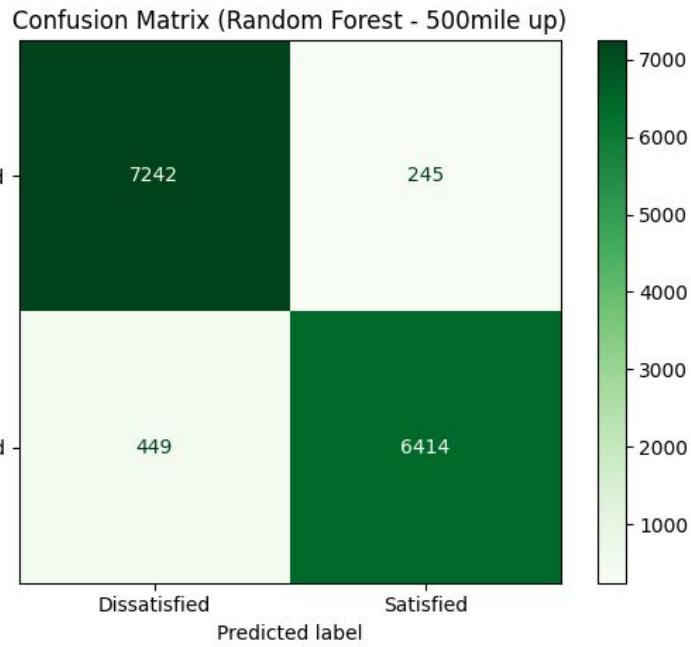
# Logistic Regression - medium&long-haul



Classification Report (500mile up):

	precision	recall	f1-score	support
0	0.85	0.87	0.86	7487
1	0.86	0.84	0.85	6863
accuracy			0.86	14350
macro avg	0.86	0.86	0.86	14350
weighted avg	0.86	0.86	0.86	14350

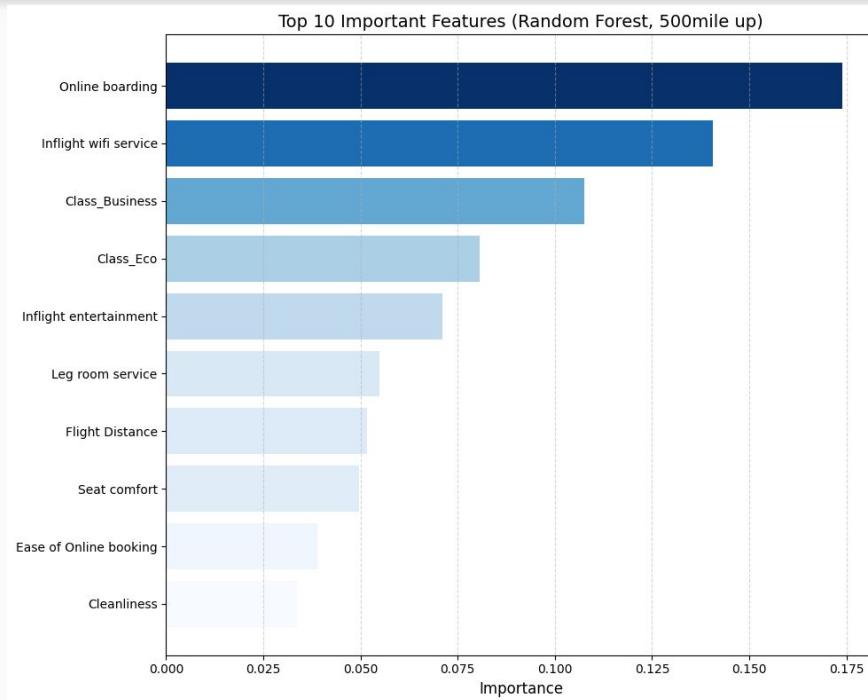
# Random Forest - medium&long-haul



[Random Forest] Classification Report (500mile up):

	precision	recall	f1-score	support
0	0.94	0.97	0.95	7487
1	0.96	0.93	0.95	6863
accuracy			0.95	14350
macro avg	0.95	0.95	0.95	14350
weighted avg	0.95	0.95	0.95	14350

# Analysis of important features -medium&long-haul

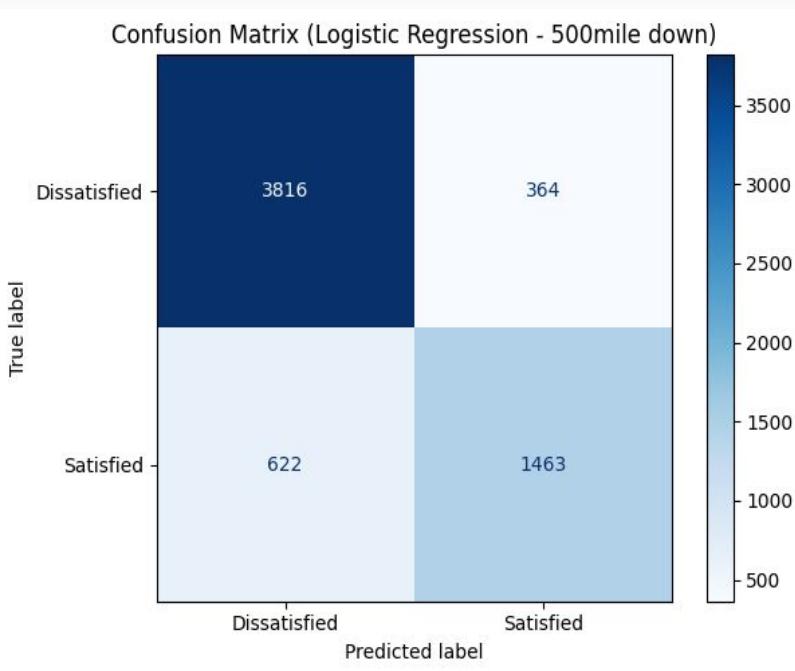


Top features for medium & long-haul flights were online boarding, in-flight Wi-Fi, and business class.

These had the strongest impact on satisfaction.

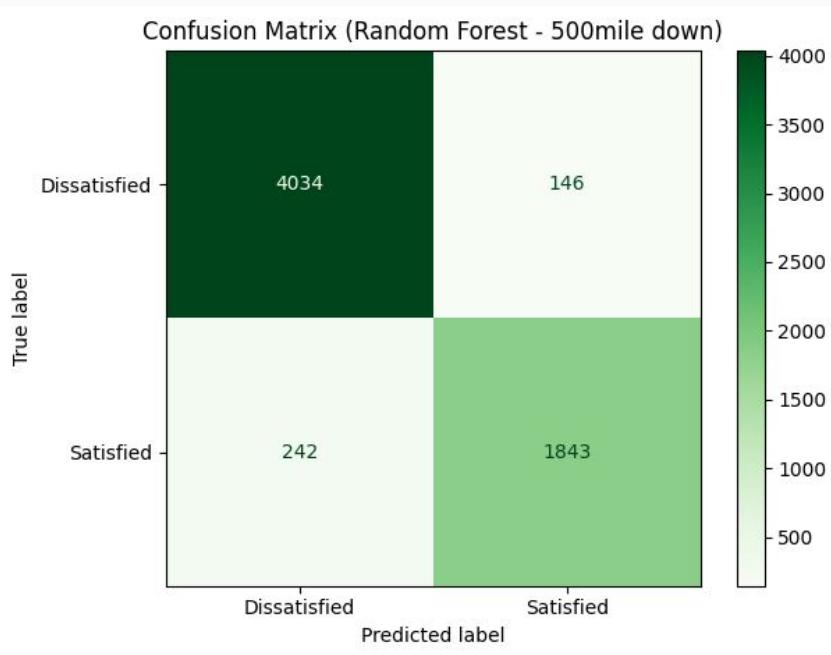
# Logistic Regression - short haul

Confusion Matrix (Logistic Regression - 500mile down)



Classification Report (500mile down):				
	precision	recall	f1-score	support
0	0.86	0.91	0.89	4180
1	0.80	0.70	0.75	2085
accuracy			0.84	6265
macro avg	0.83	0.81	0.82	6265
weighted avg	0.84	0.84	0.84	6265

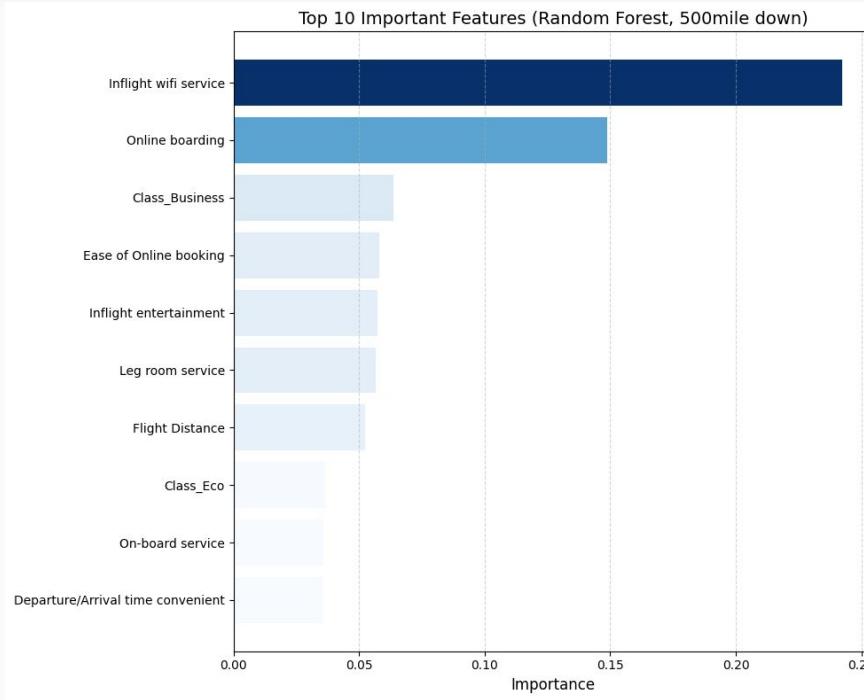
# Random Forest - short haul



[Random Forest] Classification Report (500mile down):

	precision	recall	f1-score	support
0	0.94	0.97	0.95	4180
1	0.93	0.88	0.90	2085
accuracy			0.94	6265
macro avg	0.94	0.92	0.93	6265
weighted avg	0.94	0.94	0.94	6265

# Analysis of important features - short haul



Important features differed for short-haul flights. In-flight Wi-Fi satisfaction was the most important, followed by online boarding and business class. This suggests short-haul passengers prioritize quick, reliable connectivity.

# Conclusion

- In conclusion, airlines should tailor services based on flight distance.
- Online check-in(online boarding) is important for all flights.  
In-flight Wi-Fi matters more for short-haul flights, while seat selection becomes more important on longer flights.
- In the future, customer satisfaction prediction using classification models—such as logistic regression or random forest-based analysis—is also being considered.

Thank you