

# PROJECT PROPOSAL



STA303 - LEC 5101  
Cristina Su Lam  
1007928566



# INTRODUCTION

## ➤ **RESEARCH QUESTION:**

How do various customer behavior determinants impact the likelihood of completing a booking for passengers traveling on British Airways flights?

The predictors consist of flight hour, trip type, purchase lead time, and sales channel. Length of stay is considered a potential confounder, while booking completion serves as the response variable.

## ➤ **PERSONAL IMPORTANCE:**

Interest in customer behavior and decision-making processes.

## ➤ **BROADER AUDIENCE IMPORTANCE:**

- Insights into booking behavior drive service and marketing improvements.
- Aims to inspire other industries beyond the airline sector to conduct similar studies tailored to their businesses.



# BACKGROUND AND LITERATURE

**PAPER 1:** *"Flight ticket booking app on mobile devices: Examining the determinants of individual intention to use" investigates factors influencing individuals."*

- **Summary:** Investigates factors influencing individuals' inclination to use flight ticket booking apps, particularly in developing nations.
- **Methodology:** Utilizes Structural Equation Modeling (SEM).
- **Results:** Finds perceived value and trust as key factors influencing app usage intention.
- **Relevance:** Include 'sales channel' variable to distinguish between customers who use British Airways' website (internet) and application (mobile) channels.



# BACKGROUND AND LITERATURE

## PAPER 2: *"Passenger booking timing for low-cost airlines: A continuous logit approach."*

- **Summary:** Investigates factors such as fare, trip purpose, and online airfare checking frequency influencing air travelers' booking timing, particularly focusing on low-cost carriers.
- **Methodology:** Utilizes continuous multinomial logit model.
- **Results:** Time-dependent relationships between airfare and booking dates.
- **Relevance:** Include 'purchase lead' variable to capture the time interval between booking and flight date, specifically for completed bookings on British Airways flights.



# BACKGROUND AND LITERATURE

## PAPER 3: *"Consumer choice behaviour and strategies of air transportation service providers."*

- **Summary:** Investigates factors like departure time, travel duration, fare influence European air travelers' itinerary selections.
- **Methodology:** Utilizes models sensitive to fare fluctuations over time and duration of stay, including a continuous departure time function into the choice model.
- **Results:** Morning departures preferred for shorter stays.
- **Relevance:** Inclusion of variables such as 'flight hour', and 'length of stay' in the study.



# DATA DESCRIPTION - VARIABLES CHOSEN

The analysis used a Kaggle dataset with 50,000 observations. Out of 13 variables, 5 were selected as exposure variables and potential confounders based on their relevance and literature background, offering insights into booking completion likelihood.

- **Length of Stay (Confounder):** Trip duration may affect sales channel and booking completion.
- **Flight Hour:** Scheduled departure time, as observed in studies on low-cost airlines' booking timing.
- **Purchase Lead Time:** Duration between booking and flight departure, as highlighted in the study on low-cost airlines.
- **Sales Channel:** Differentiates between booking channels (internet vs. mobile), informed by research on flight ticket booking apps.
- **Trip Type:** Categorizes journeys as RoundTrip or OneWay, revealing passenger preferences.



# DATA DESCRIPTION - EDA

**Table 1. Summary Statistics Table**

	purchase_lead	length_of_stay	flight_hour	booking_complete
Min.	0	0	0.00	0.00
1st Quartile	21	5	5.00	0.00
Median	51	17	9.00	0.00
Mean	85	23	9.07	0.15
3rd Quartile	115	28	13.00	0.00
Max.	867	778	23.00	1.00

## Some notable findings:

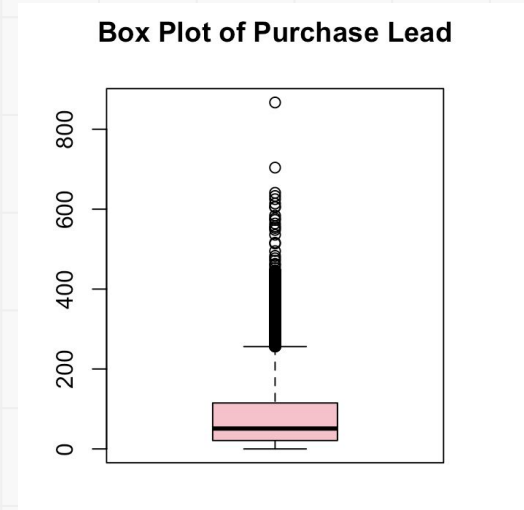
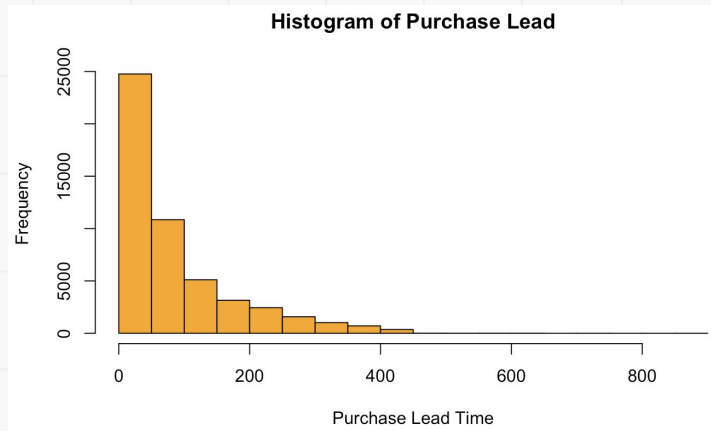
- The average “Purchase Lead” of 85 days suggests bookings made further in advance are more likely to be completed successfully.
- The mean “Length of Stay” of 23 days may influence booking completion, as longer trips require more planning.



# DATA DESCRIPTION - EDA

## Variable “Purchase Lead”:

- Histogram is heavily right-skewed, indicating most bookings are made with a shorter lead time or closer to departure.
- The presence of outliers in the box-plot indicates that some bookings are made well in advance.
- These outliers could impact the model and potentially affect the generalizability of the estimators, making reliable inference challenging.

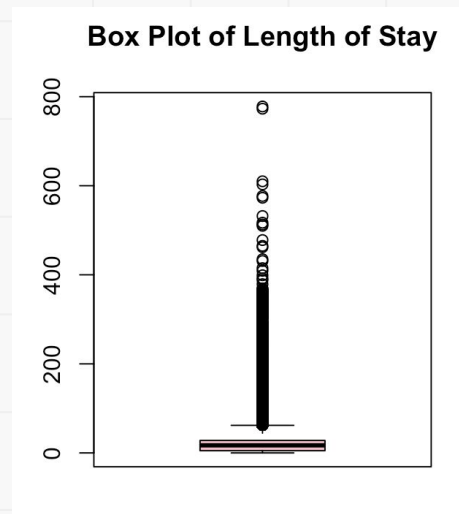
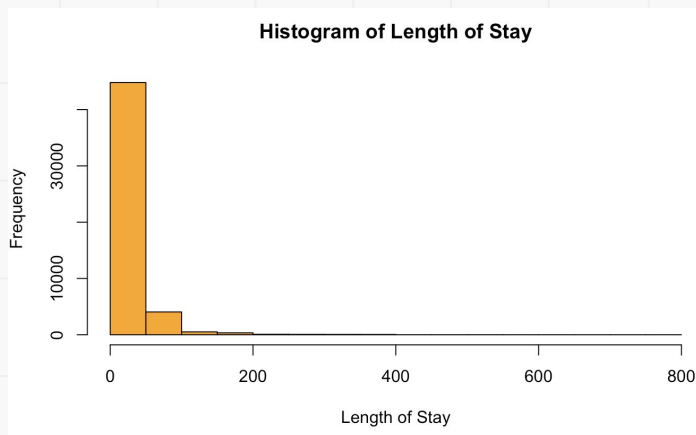




# DATA DESCRIPTION - EDA

## Variable “Length of Stay”:

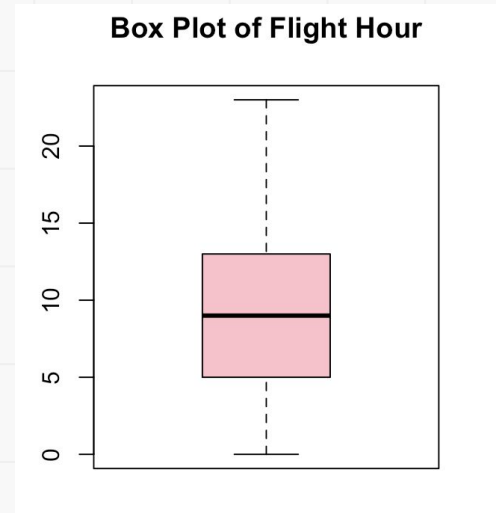
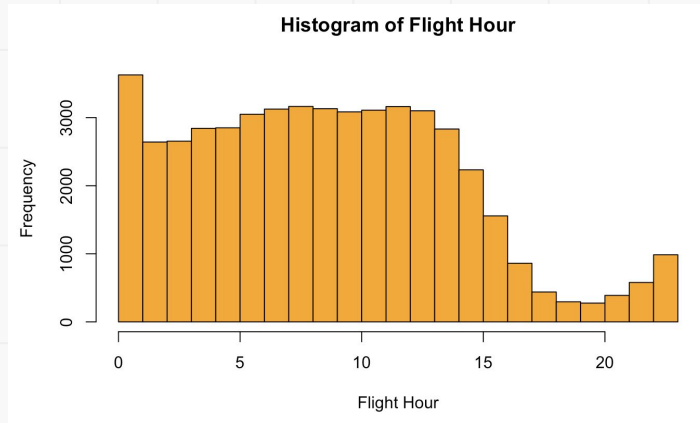
- Histogram is right-skewed, with a long tail extending to the right, suggesting most stays are brief.
- There are outliers indicating that while most stays are short, there are exceptions where stays are significantly longer.
- These outliers could impact the model and potentially affect the generalizability of the estimators, making reliable inference challenging.



# DATA DESCRIPTION - EDA

## Variable “Flight Hour”:

- The histogram exhibits multiple peaks, indicating preferred departure times throughout the day.
- The median is located around 10 hours, suggesting a central tendency for mid-morning to midday departures.
- Minimal outliers beyond the whiskers imply no exceptionally uncommon flight hours.



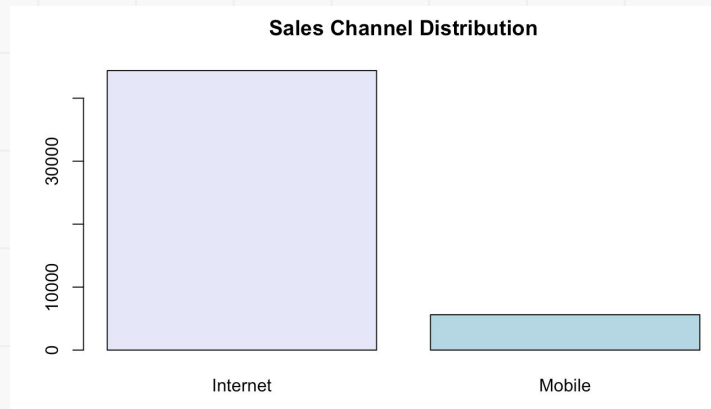
# DATA DESCRIPTION - EDA

## Variable "Sales Channel":

- Internet bookings are substantially higher than Mobile bookings, with 44,382 compared to 5,618 respectively.
- Completed bookings are more prevalent on the Internet (6,869) compared to Mobile (609).

Table 2. Contingency Table (Booking Complete vs. Sales Channel)

	0	1	Total
Internet	37513	6869	44382
Mobile	5009	609	5618
Total	42522	7478	50000



# DATA DESCRIPTION - EDA

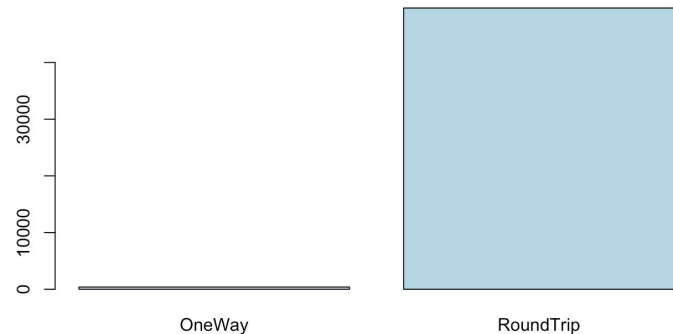
## Variable “Trip Type”:

- Round-trip bookings dominate, with 49,613 compared to 387 one-way bookings.
- The table also shows a comparison between completed and non-completed bookings, with 7,478 completed bookings, of which a majority are round trips.

Table 3. Contingency Table (Booking Complete vs. Trip Type)

	0	1	Total
OneWay	367	20	387
RoundTrip	42155	7458	49613
Total	42522	7478	50000

Trip Type Distribution



# DISCUSSION

## PLAN FOR ANSWERING RESEARCH QUESTION:

The Generalized Linear Model (GLM) is instrumental in addressing the research question by exploring the relationship between various factors and the likelihood of completing a booking. Each chosen variable serves a specific purpose:

- **Sales Channel:** Investigates the impact of booking method on booking completion.
- **Trip Type:** Explores how trip type affects booking completion probability.
- **Purchase Lead Time:** Assesses the influence of lead time on booking completion.
- **Flight Hour:** Examines specific flight times' impact on booking completion.
- **Length of Stay:** Considers how trip duration affects booking completion likelihood.

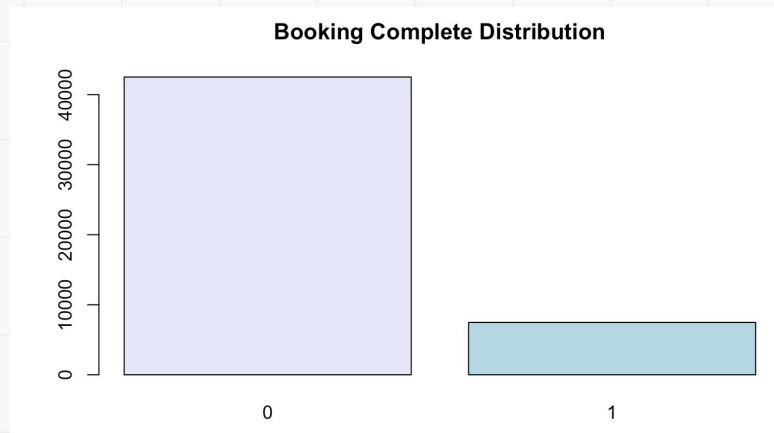
GLM allows estimating variables' impact on booking completion probability, considering confounding effects. Its logistic regression framework, suited for binary outcomes, provides insights into British Airways' booking behavior through regression coefficients, odds ratios, and predictive probabilities.



# DISCUSSION

## ASSUMPTIONS OF GLM:

- **Independence:** Observations are assumed to be independent, unaffected by others.
- **Distribution of Y:** In logistic regression, the binary outcome variable Y follows a binomial distribution.
- **Linearity:** Logistic regression models commonly employ the logit link function, to model the linear relationship between predictors and outcome probability.



# REFERENCES

1. Mohd Suki, N., & Mohd Suki, N. (2017). Flight ticket booking app on mobile devices: Examining the determinants of individual intention to use. *Journal of Air Transport Management*, 62, 146–154. <https://doi.org/10.1016/j.jairtraman.2017.04.003>.
2. Wen, C.-H., & Chen, P.-H. (2017). Passenger booking timing for low-cost airlines: A continuous logit approach. *Journal of Air Transport Management*, 64, 91–99. <https://doi.org/10.1016/j.jairtraman.2017.06.030>.
3. Eggermond, M. van, Schüssler, N., Axhausen, K. W., Eggermond, M. van, Eggermond, M. van, Axhausen, K. W., & Axhausen, K. W. (2007). *Consumer choice behaviour and strategies of Air Transportation Service Providers*. ETH, Eidgenössische Technische Hochschule Zürich, IVT, Institut für Verkehrsplanung und Transportsysteme.
4. Bisht, D. (2024, January 16). *British Airways Customer Bookings*. Kaggle. <https://www.kaggle.com/datasets/deepakb4/british-airways-customer-bookings>.





*Thank you for  
watching!*

