

British Airways passenger booking channel
Xiaoyu Fu
Data Science Capstone Project by
Springboard



British Airways Background

British Airways (BA) is the flag carrier of the United Kingdom. It is headquartered in London, England, near its main hub at Heathrow Airport. The airline is the second largest UK-based carrier, based on fleet size and passengers carried, behind easyJet. In January 2011 BA merged with Iberia, creating the International Airlines Group (IAG), a holding company registered in Madrid, Spain.

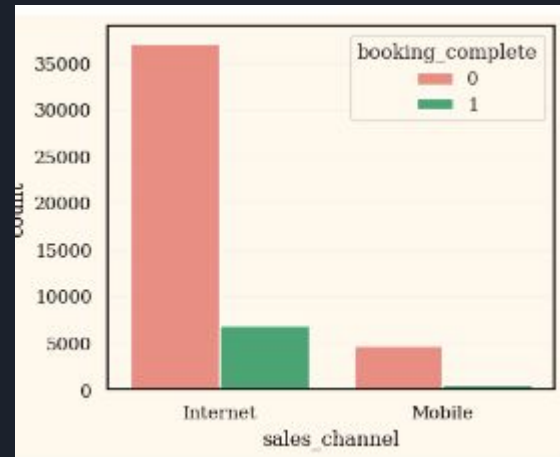


The problem

The gap between online reservations and mobile reservations is very obvious. The nearly seven-fold gap may lead to the loss of some users who focus on mobile reservations, resulting in losses for airlines.

What factor affect the customer booking channel?

Investigating the influence of booking channels on customer behavior.






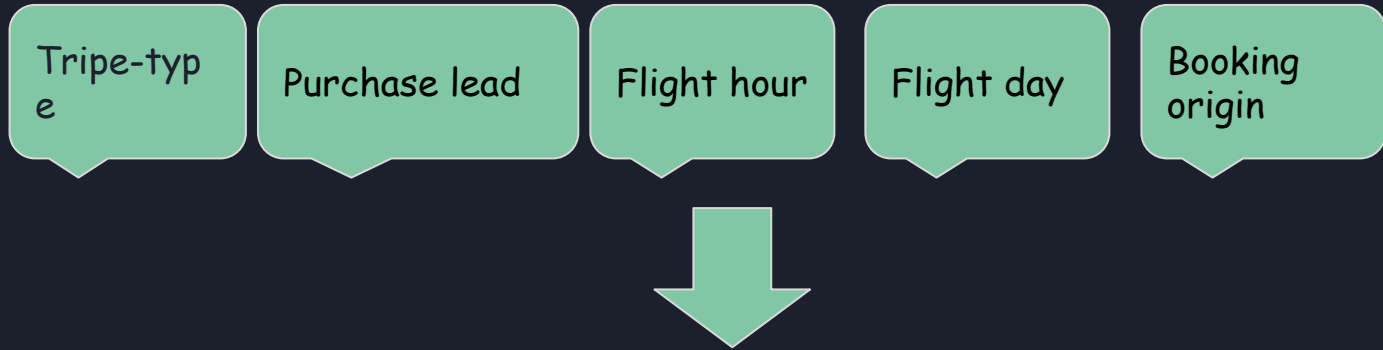
Who might care?



Etc....



What factor might affect the customer's choice of booking channel?



From the data set:
Passanger_booking_data.csv

https://www.kaggle.com/datasets/manishkumar7432698/airline-passangers-booking-data?select=Passanger_booking_data.csv



Data information:

File Name: Passanger_booking_data.csv

File Size: 3.15 MB

Format: CSV (Comma-separated values)

https://www.kaggle.com/datasets/manishkumar7432698/airline-passangers-booking-data?select=Passanger_booking_data.csv

Dataset Statistics

Number of Variables	14
Number of Rows	50002
Missing Cells	0
Missing Cells (%)	0.0%
Duplicate Rows	719
Duplicate Rows (%)	1.4%
Total Size in Memory	28.4 MB
Average Row Size in Memory	596.5 B
Variable Types	Categorical: 9 Numerical: 4 GeoGraphy: 1



Clean the categorical data.

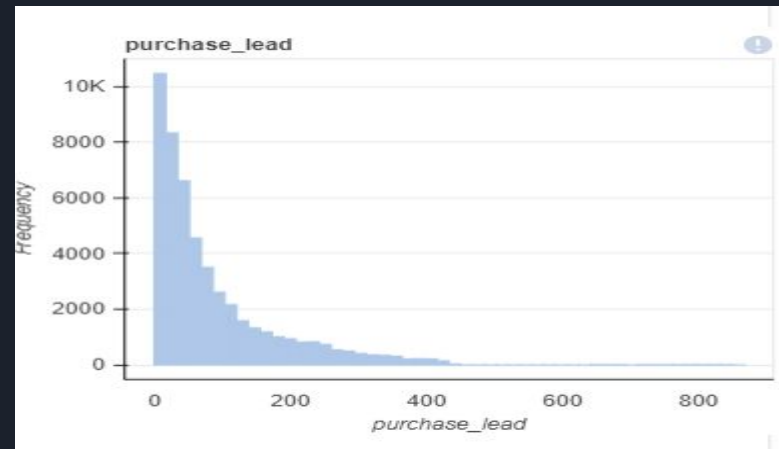
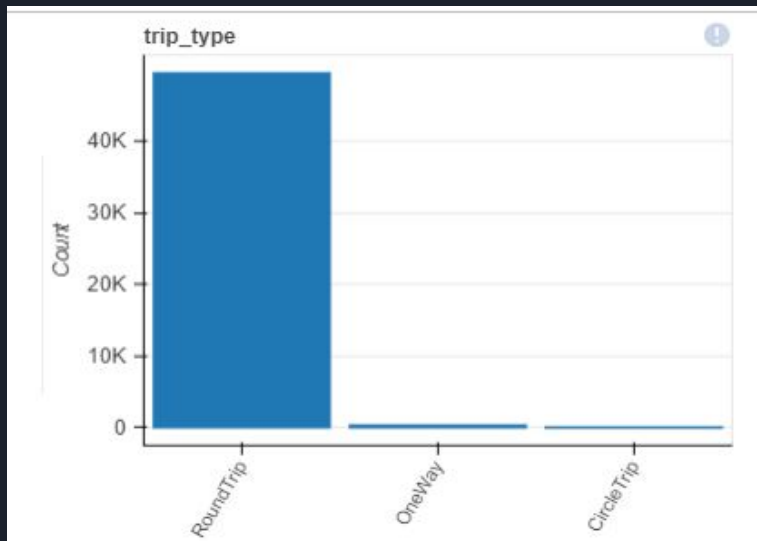
Standardize the magnitude of numeric features using a scaler. And split into testing and training datasets. At this point it is clear to see the "variables" that affect the problem.

```
Index(['num_passengers', 'flight_hour', 'wants_extra_baggage',  
      'wants_preferred_seat', 'wants_in_flight_meals', 'flight_duration',  
      'booking_complete', 'CircleTrip', 'OneWay', 'RoundTrip',  
      ...  
      'Timor-Leste', 'Tonga', 'Tunisia', 'Turkey', 'Ukraine',  
      'United Arab Emirates', 'United Kingdom', 'United States', 'Vanuatu',  
      'Vietnam'],  
      dtype='object', length=922)
```



visualization

- Tripe-type
- Purchase lead
- Flight hour
- Flight day
- Booking origin

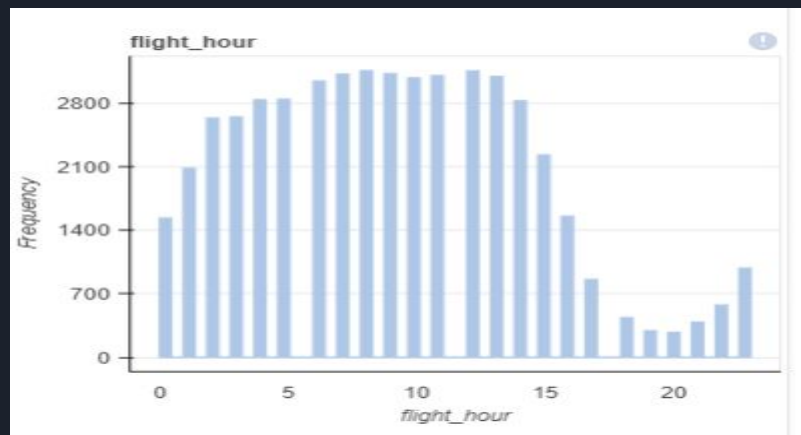


↓

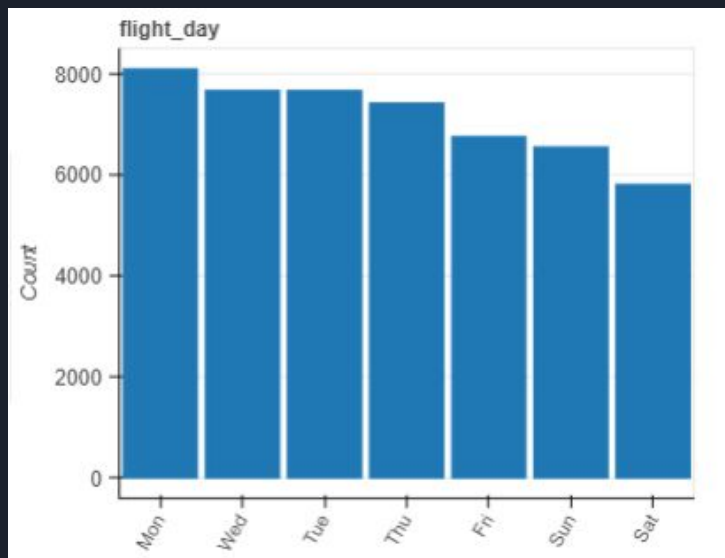
Approximate Distinct Count	3
Approximate Unique (%)	0.0%
Missing	0
Missing (%)	0.0%
Memory Size	3699104

Approximate Distinct Count	470
Approximate Unique (%)	0.9%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Memory Size	800032
Mean	84.9406
Minimum	0
Maximum	867
Zeros	370
Zeros (%)	0.7%
Negatives	0
Negatives (%)	0.0%

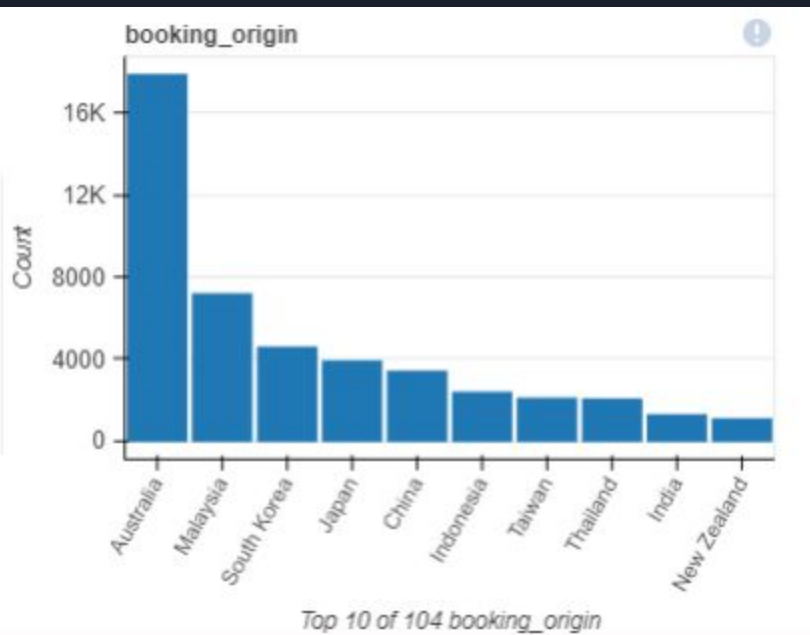




Approximate Distinct Count	24
Approximate Unique (%)	0.0%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Memory Size	800032
Mean	9.0663
Minimum	0
Maximum	23
Zeros	1539
Zeros (%)	3.1%
Negatives	0
Negatives (%)	0.0%



Approximate Distinct Count	7
Approximate Unique (%)	0.0%
Missing	0
Missing (%)	0.0%
Memory Size	3400136



Approximate Distinct Count	104
Approximate Unique (%)	0.2%
Missing	0
Missing (%)	0.0%
Memory Size	3662055

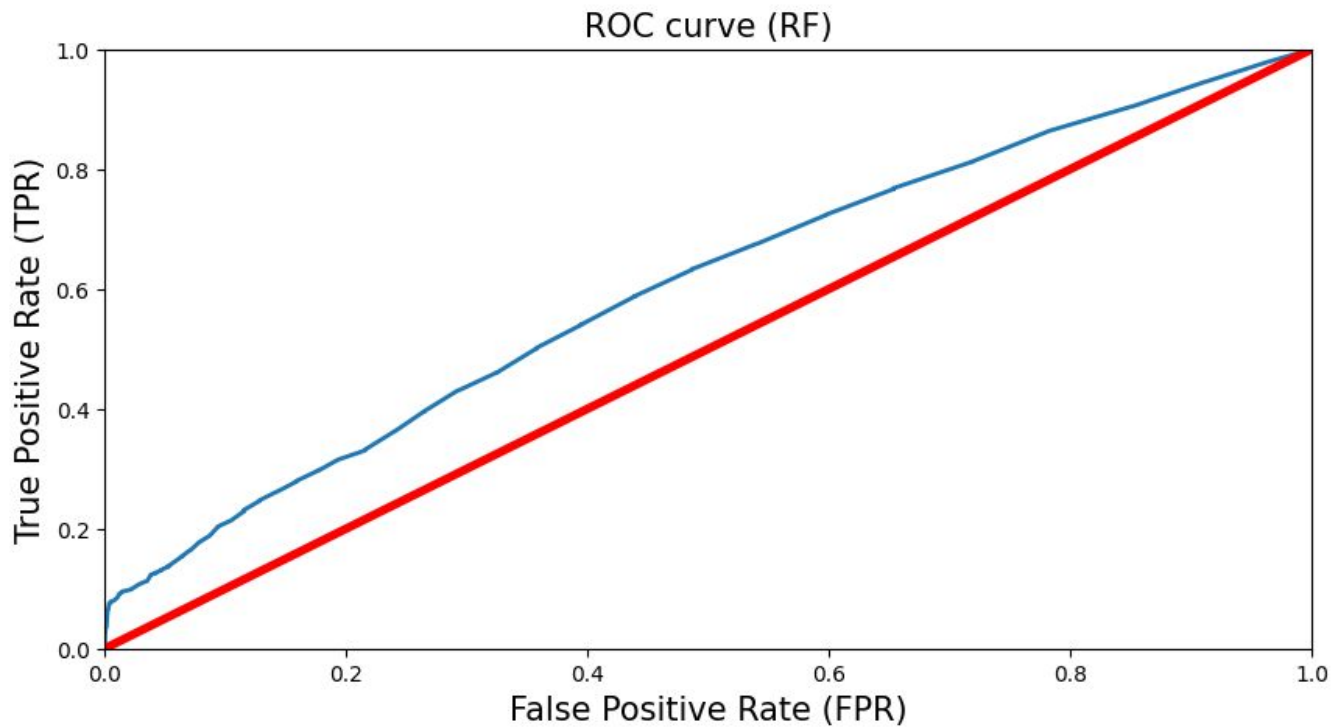


Modeling

Four different models to test the accuracy.

- Logistic Regression (Test Accuracy: 0.8908739417372176).
- K-Nearest neighbor (KNN) (Test Accuracy: 0.896273581761216).
- Random Forest (Test Accuracy: 0.892940470635291).
- Naive Bayes (Test Accuracy: 0.8908739417372176).

Roc Score





Conclusions and Conjectures

- British Airways may need to look at the density of bookings on weekends and analyze the exact times of popular flights.
- Because it can be seen from the specific data that the average number of additional services and the preference for weekend flights change with the change of booking lead time. Booking channels play an important part in this.
- The average values of variables like `total_services` and `is_weekend` vary between different sales channels.
- Based on the above speculation, the completion rate increases in direct proportion to the number of services and booking channels. From 10.68% for 0 services to 18.59% for 3 services.



Two suggestions

- Open more channels for booking air tickets by mobile phone
- Increase the number of services (From 10.68% for 0 services to 18.59% for 3 services.)



**THANK
YOU**