

For purposes of Forage Virtual Work Experience program

# BRITISH AIRWAYS



Forage - Task 1: Web scraping to gain insights from customer reviews

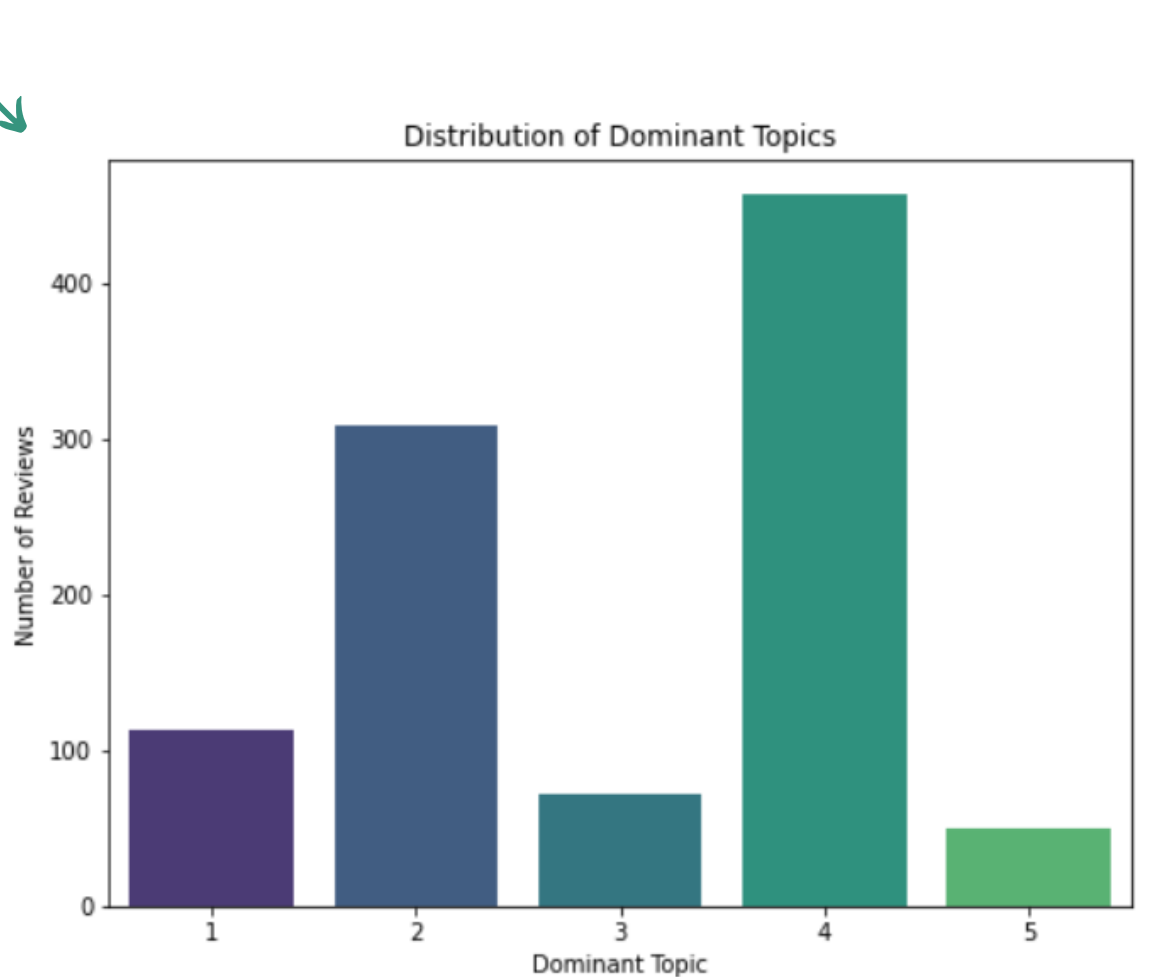
**RHEA MALL**

43.4% of the reviews are of a negative sentiment, and 30.4% are positive.

43.4% of the reviews are of a negative sentiment, and 30.4% are positive.

45.7% of the reviews are about **topic 4.**

45.7% of the reviews are about **topic 4.**



## INSIGHTS FROM WORDCLOUDS

Frequent mentions of the words "service" and "time" in reviews of all 3 sentiments types indicates a focus on the airline's **customer service** and **timeliness**.

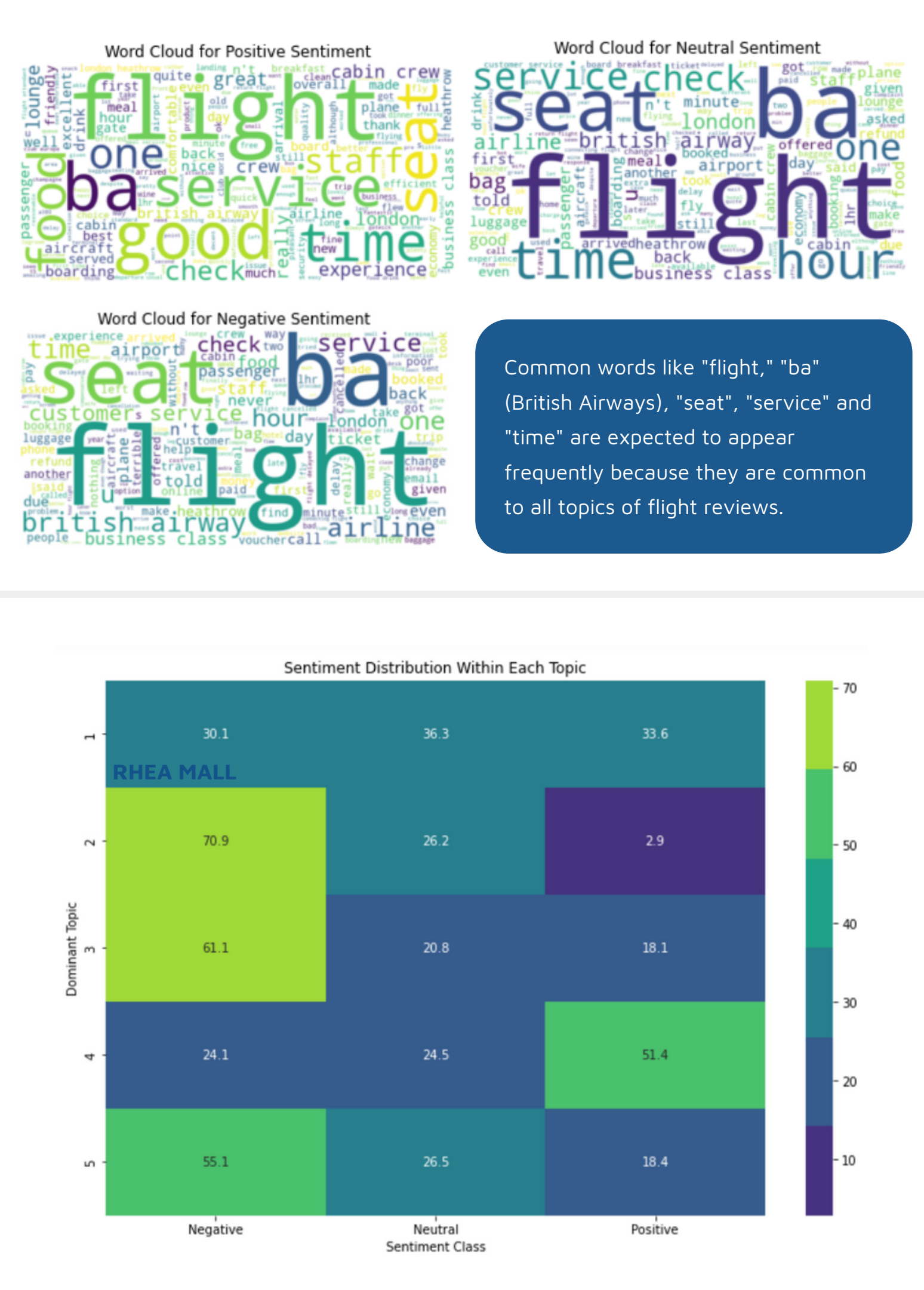
## INSIGHTS FROM WORDCLOUDS

Frequent mentions of the words "service" and "time" in reviews of all 3 sentiments types indicates a focus on the airline's **customer service** and **timeliness**.

## SENTIMENT DISTRIBUTION WITHIN TOPICS

- Reviews about topic 4 (the most common topic) are majorly positive.
- A vast majority (**70.9%**) of the reviews about topic 2 are negative.
- A majority of reviews about topics 3 and 5 are negative, too.

- ## SENTIMENT DISTRIBUTION WITHIN TOPICS
- Reviews about topic 4 (the most common topic) are majorly positive.
  - A vast majority (**70.9%**) of the reviews about topic 2 are negative.
  - A majority of reviews about topics 3 and 5 are negative, too.



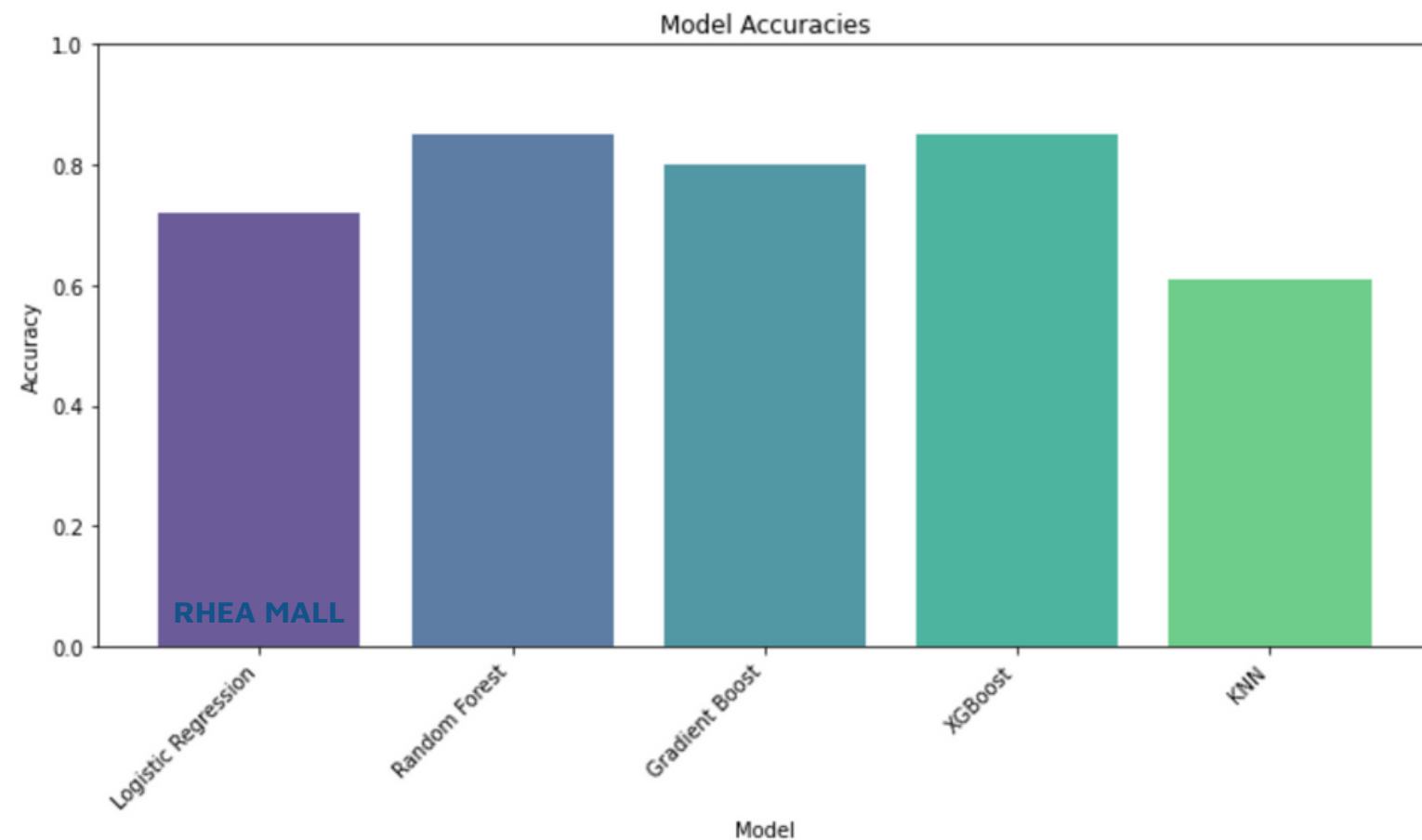
For purposes of Forage Virtual Work Experience program

# BRITISH AIRWAYS



Forage - Task 2: Web scraping to gain insights from customer reviews

**RHEA MALL**



**Model Choice:**  
RF Classifier  
has the highest  
accuracy on our  
data from these  
5 models.

## RANDOM FOREST CLASSIFIER

**85.1%**

Mean cross validation accuracy;  
CV scores were fairly consistent  
across different folds.

**Class 0**

Precision: 0.86  
Recall: 0.97  
F1 Score: 0.91

**Class 1**

Precision: 0.5  
Recall: 0.13  
F1 Score: 0.21

These results indicate that the model performs well for class 0 but struggles to predict for class 1. Thus, potential improvements may involve the employment of other class balancing techniques.

### Feature Importance Analysis:

**purchase\_lead** was the most important feature in the model (number of days between travel date and booking date)

