



# Predicting Hotel Direct Booking Behavior

TWITTER SENTIMENT  
ANALYSIS AND BOOKING  
PREDICTION

# Agenda



Background



Objective



Data



Predicting Booking Behavior



Model Performance



Recommendations



Future Work





# Background

---

Online Travel Agencies represent 2/3 of online hotel bookings

- Expedia 28.09%
- Booking.com 19.13%
- Priceline 16.15%

Hotel direct bookings represent only 1/3 of online hotel bookings

- Marriott Hotels 26.21%



## Objective

---

Predict whether a twitter user will book directly with the hotel or through an online travel agency based on class of tweet.

# Data



Over 19,000 tweets gathered



Tweets divided into 6 categories

Luxury Hotels  
Premium Hotels  
Select-Service Hotels  
Expedia  
Priceline  
Booking.com



Tweets categorized by sentiment

Positive  
Negative

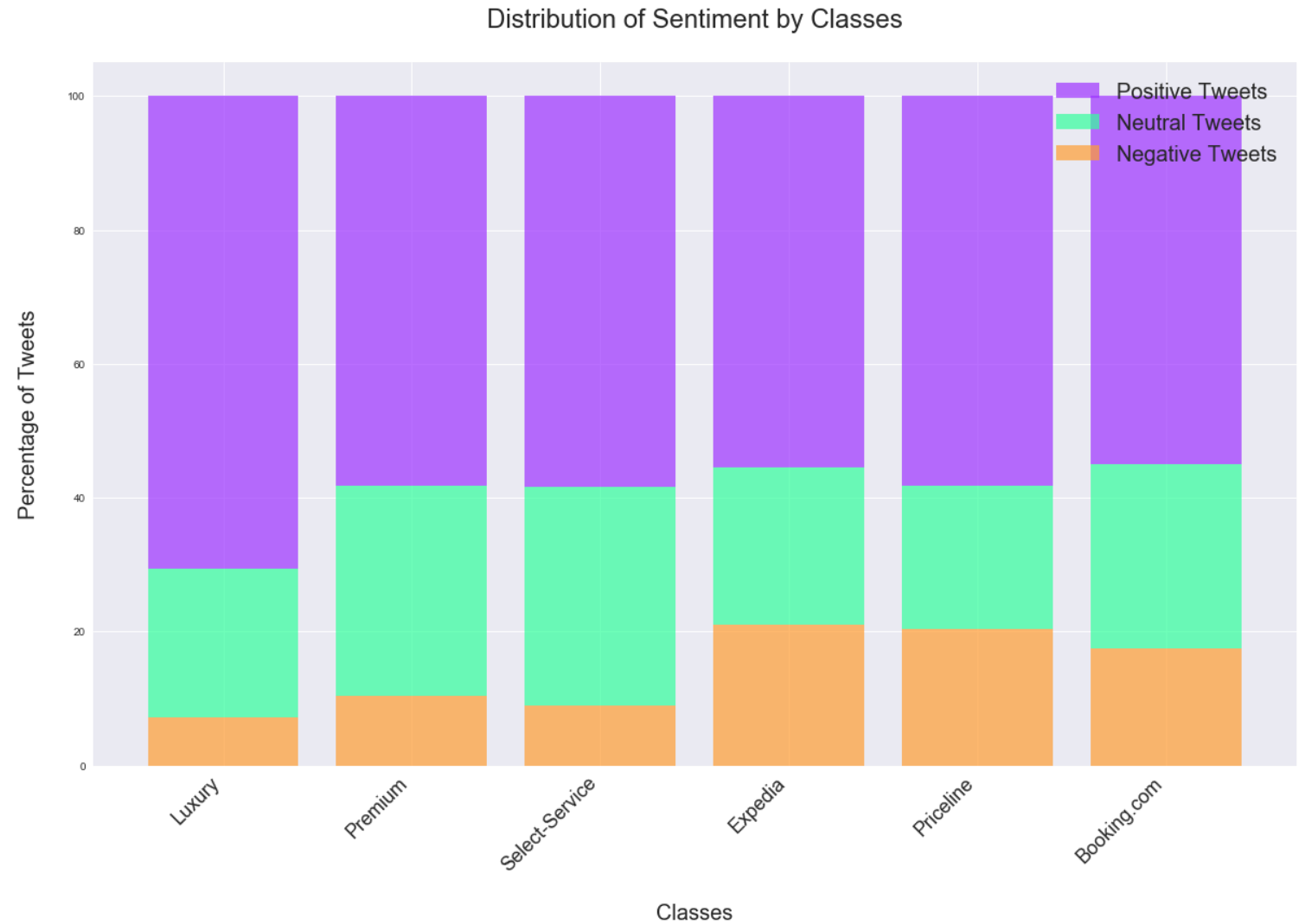
# Distribution of Sentiment by Class

## Most Positive Tweets

- Luxury Hotels

## Most Negative Tweets

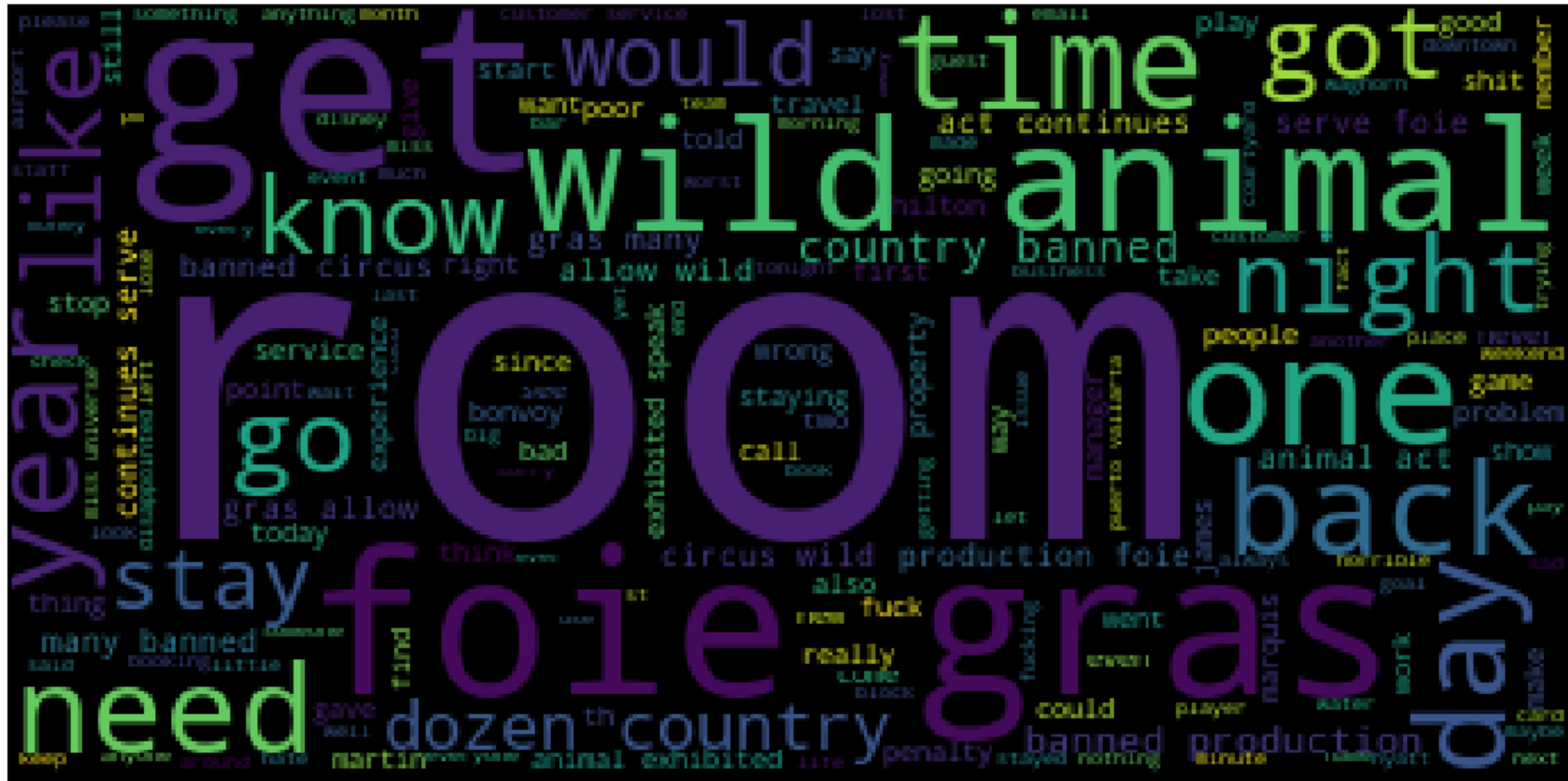
- Expedia



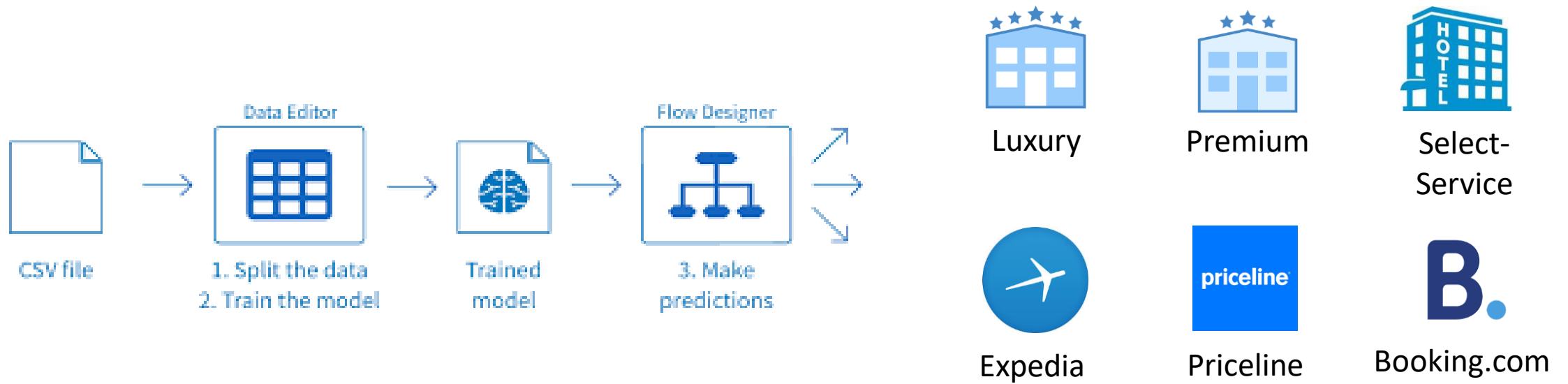
### Most Frequently Used Words in Positive Hotel Tweets



### Most Frequently Used Words in Negative Hotel Tweets







# Predicting Direct Booking Behavior

## Classification Model

- Predict which channel a twitter user will use to book
- 82.3% accuracy

# Model Performance

# Classification of Test Tweets

---

Actual Category	Predicted Category
Premium	Premium
Expedia	Premium
Premium	Premium
Premium	Premium
Expedia	Expedia
Premium	Premium
Premium	Premium
Premium	Premium
Premium	Premium
Luxury	Premium



8 out 10 tweets were placed in the correct category



# Recommendations

---

1. Determine if user will book hotel-direct or with OTA
2. Identify and examine misclassified tweets for themes and brand message opportunities



# Future Work

---



CREATE CLEAR TWEET LABELS TO  
IMPROVE CUSTOMER SERVICE



ANALYZE TWEETS TO DESIGN  
POSITIVE BRAND MESSAGE



CONDUCT ANALYSIS OF  
MEMBERSHIP REWARDS



*thank you*



A microscopic image of a plant tissue section, likely a stem, showing a central vascular bundle. The bundle contains large, circular vessels (xylem) and smaller, more densely stained cells (phloem). The surrounding tissue consists of numerous small, rounded cells, possibly parenchyma or epidermal cells. The image is overlaid with a dark horizontal bar containing the word "Appendix".

# Appendix

---

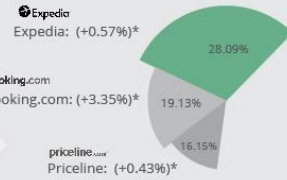


## OTAs vs Hotels: The Age Old Battle Over Online Booking

### Total Share of Online Bookings

OTAs = 2/3 (+0.8%)

Top OTAs market share online bookings



Priceline Group + Expedia predicted to control 94% of online bookings by 2020

Hotels = 1/3 (-0.8%)

Top hotel market share in online booking



\*denotes the % change from 2016

### Causes+Factors

#### Costs

Hotels have limited control over main operational costs

- Labour costs
- Utilities
- Debt services
- Real estate taxes
- Franchise fees
- Distribution costs

### Inventory

OTA sites hold more than

99%

of the worlds hotel inventory

Marketing % of revenue invested in marketing

30-40%

OTAs

6%

Hotels

### Hotels Secret Weapons

#### Artillery

50% of OTA users will visit official hotel website before booking

1

Capture these leads through improved user experience, user driven design + easy booking

85% of travellers say price is the most important factor in deciding where to book

2

Lower costs through digital technology + marketing initiatives. Offer non-monetary incentives

Increasing numbers of occupied rooms

3

Capture + target new travellers through increased investments in marketing

OTA channels reach + traffic

4

Strategically use only during needed periods (ex.off

# Helix Micrometrics

## OTAs vs Hotels: The Age Old Battler Over Online Booking