

# Hotel Review Analysis Using NLP

Author: Ismael Araujo



# The Problem

—

Britannia International Hotels have thousands of reviews.  
How they get insights easily from these reviews?

# Process

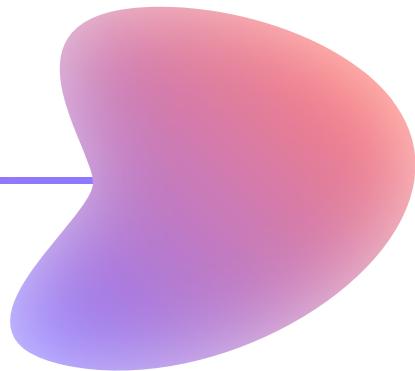


# Process

Word Cloud

Sentiment Analysis

Modeling



# Data

Size

515,738 reviews

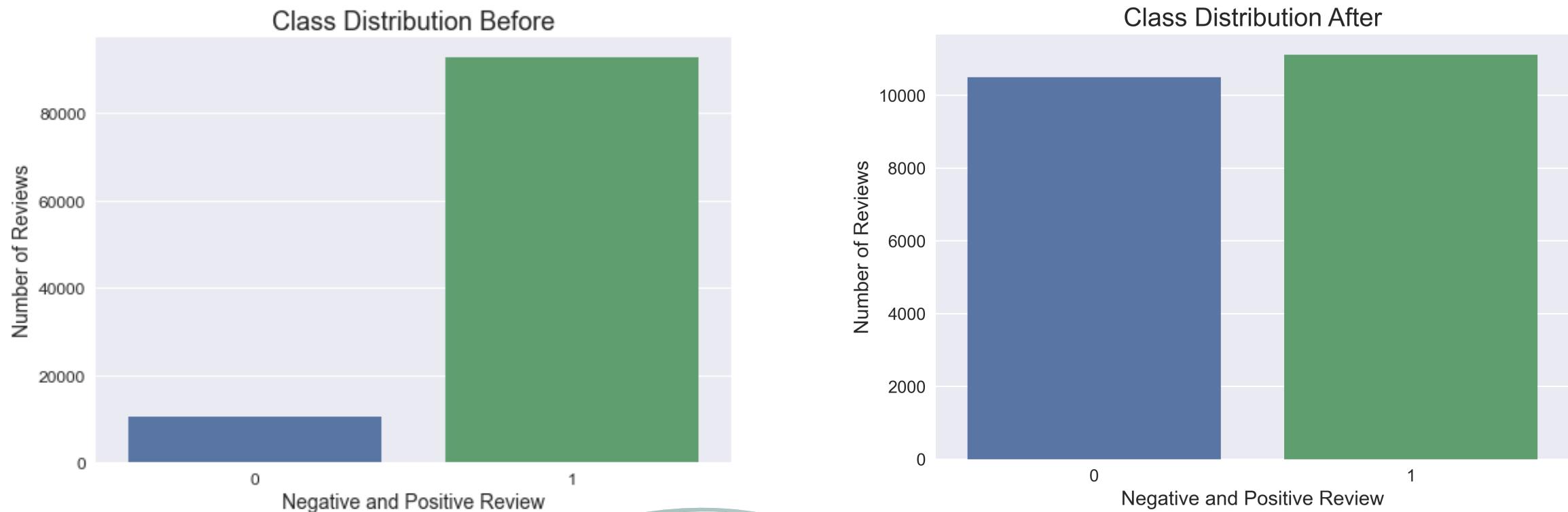
Source

Kaggle

Content

17 columns

# Class Imbalance





Italy



Spain



France



United Kingdom



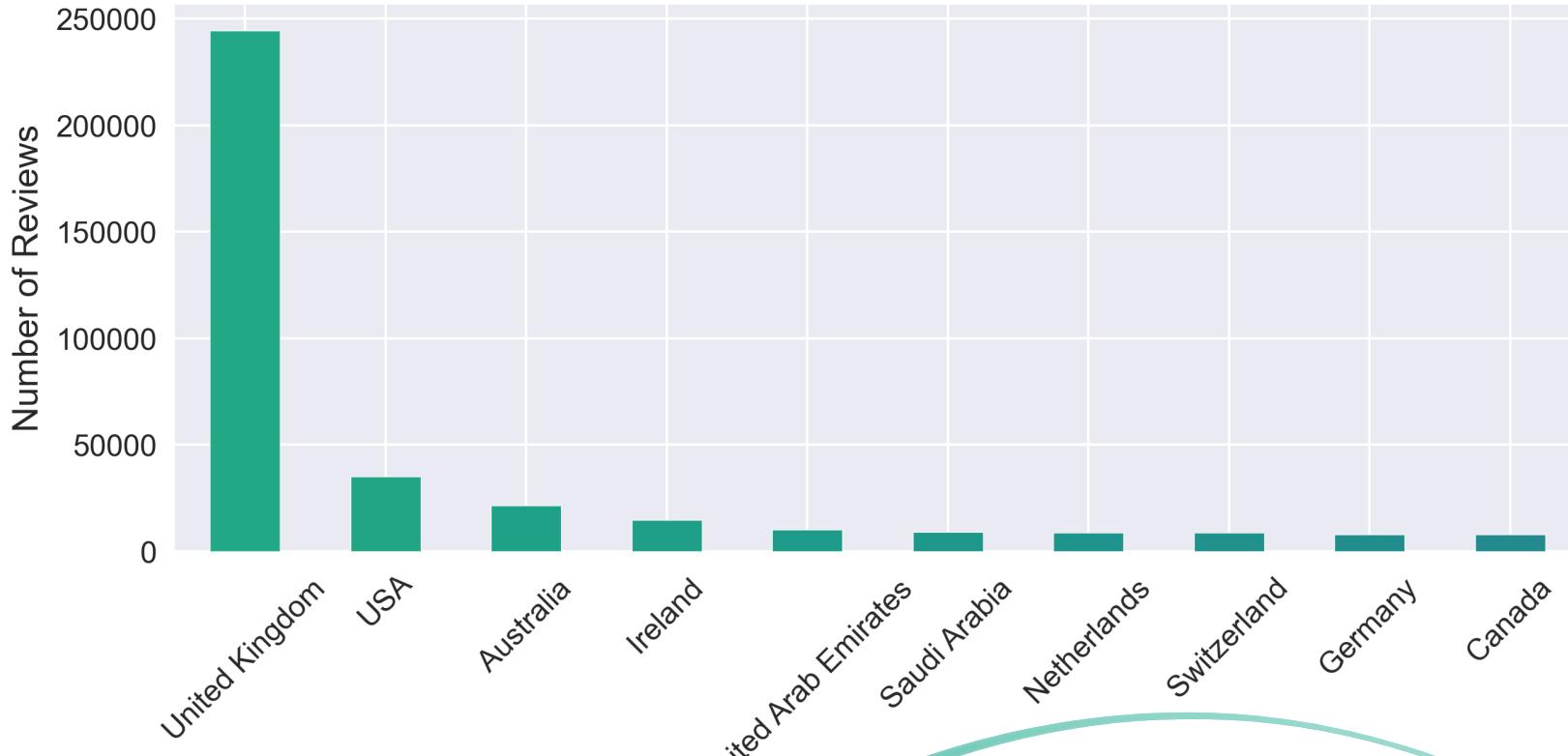
Austria



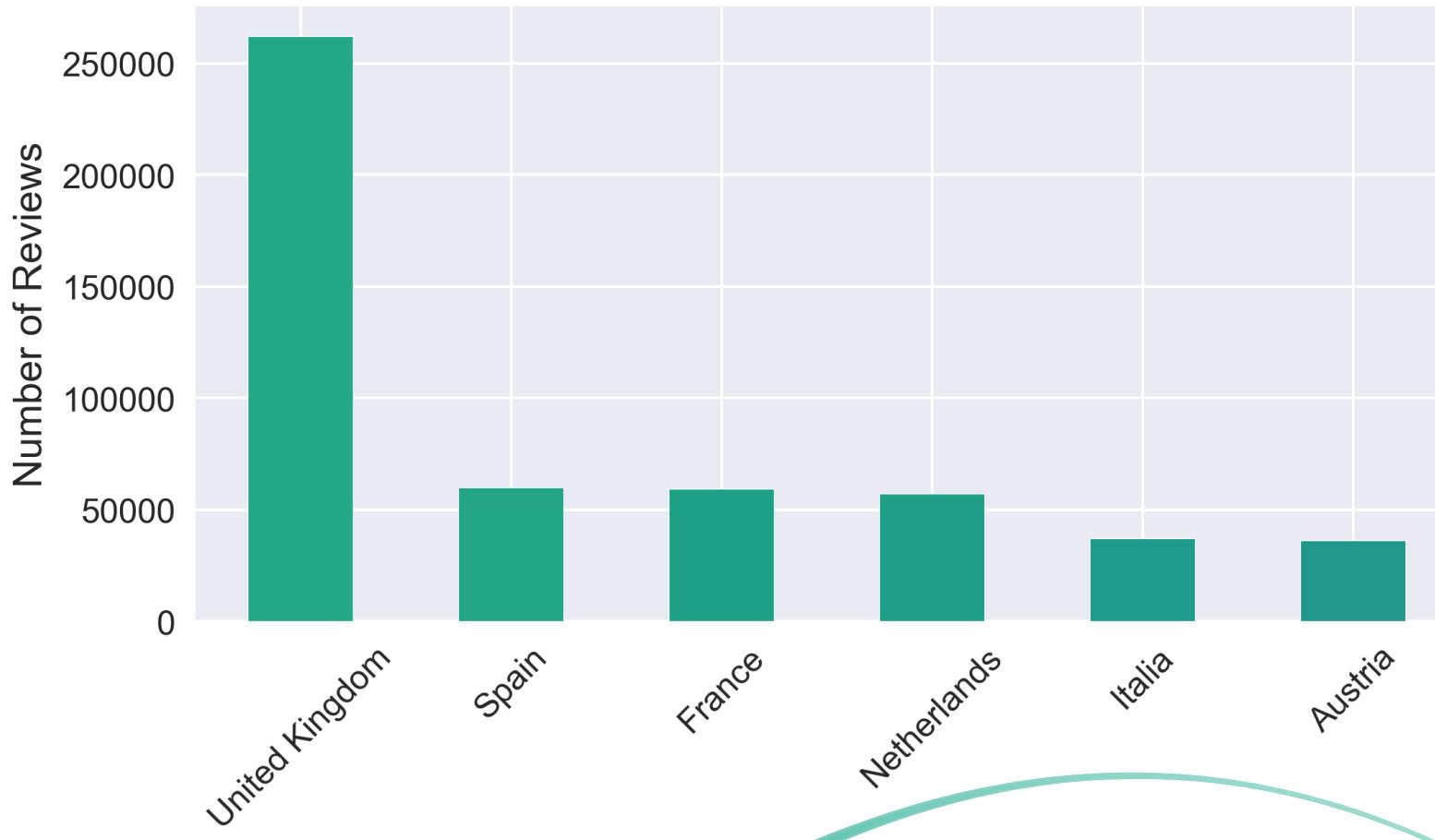
The Netherlands



### Most common reviewer nationality



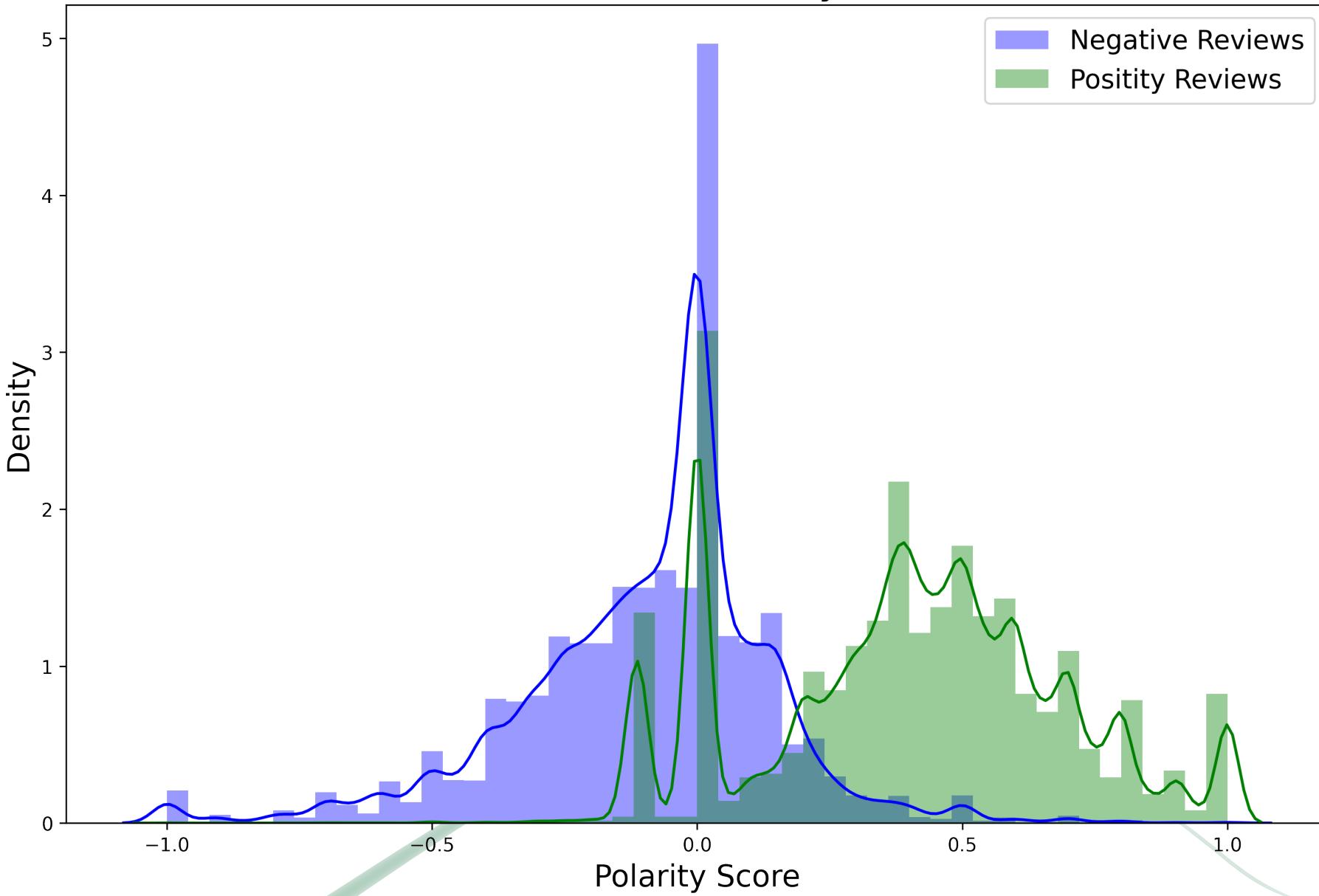
## Countries of the Hotels with the Most Reviews



## Score Density - Hotel Compared to London



## Review Polarity



# TripAdvisor Review Section



5,067  
Reviews



58  
Q+A



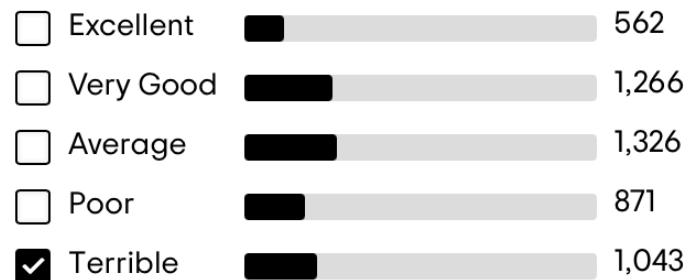
100  
Room tips

## Reviews

[Write a review](#)



### Traveler rating



### Time of year

- Mar-May
- Jun-Aug
- Sep-Nov
- Dec-Feb

### Traveler type

- Families
- Couples
- Solo
- Business
- Friends

### Language

- All languages (5,068)
- English (4,510)
- Chinese (Sim.) (185)
- Chinese (Trad.) (179)

[More](#)

### Popular mentions

All reviews

national holidays

jacuzzi bath

canary wharf tube station

jubilee line

bit dated

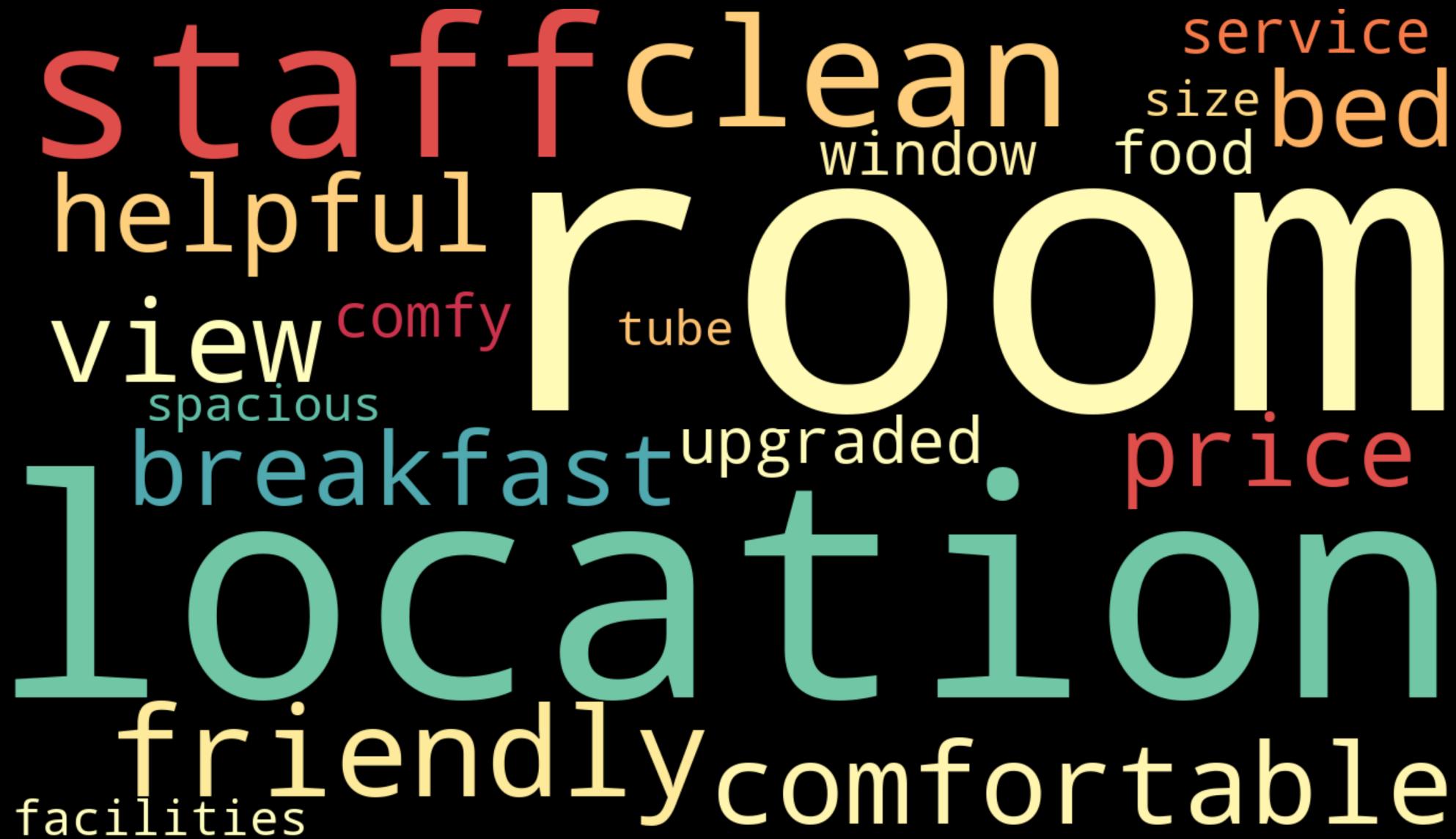
central london

full english

arena

thames

# Positive Reviews Word Cloud



A word cloud visualization showing positive reviews. The words are arranged in a large central cluster with varying sizes and colors. The most prominent word is 'room' in yellow. Other large words include 'location' in teal, 'friendly' and 'comfortable' in yellow, 'breakfast' in blue, 'spacious' in green, 'view' in white, 'helpful' in orange, 'staff' in red, 'clean' in orange, 'upgraded' in light orange, 'tube' in light orange, 'window' in light orange, 'size' in light orange, 'food' in light orange, 'service' in red, 'bed' in orange, 'price' in red, and 'facilities' in light orange.

room

location

friendly comfortable

breakfast

spacious

view

helpful

staff

clean

upgraded

tube

window

size

food

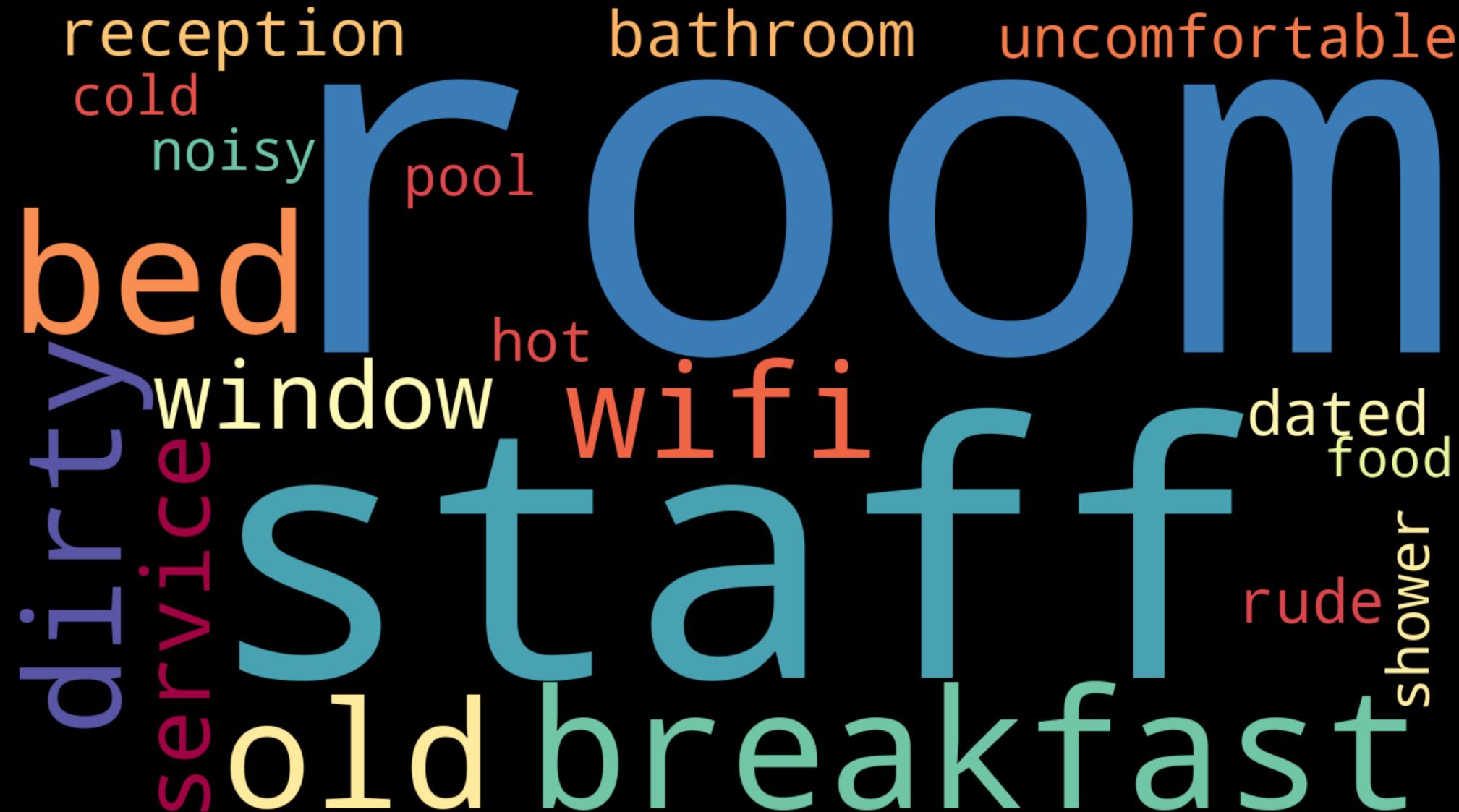
service

bed

price

facilities

# Negative Reviews Word Cloud

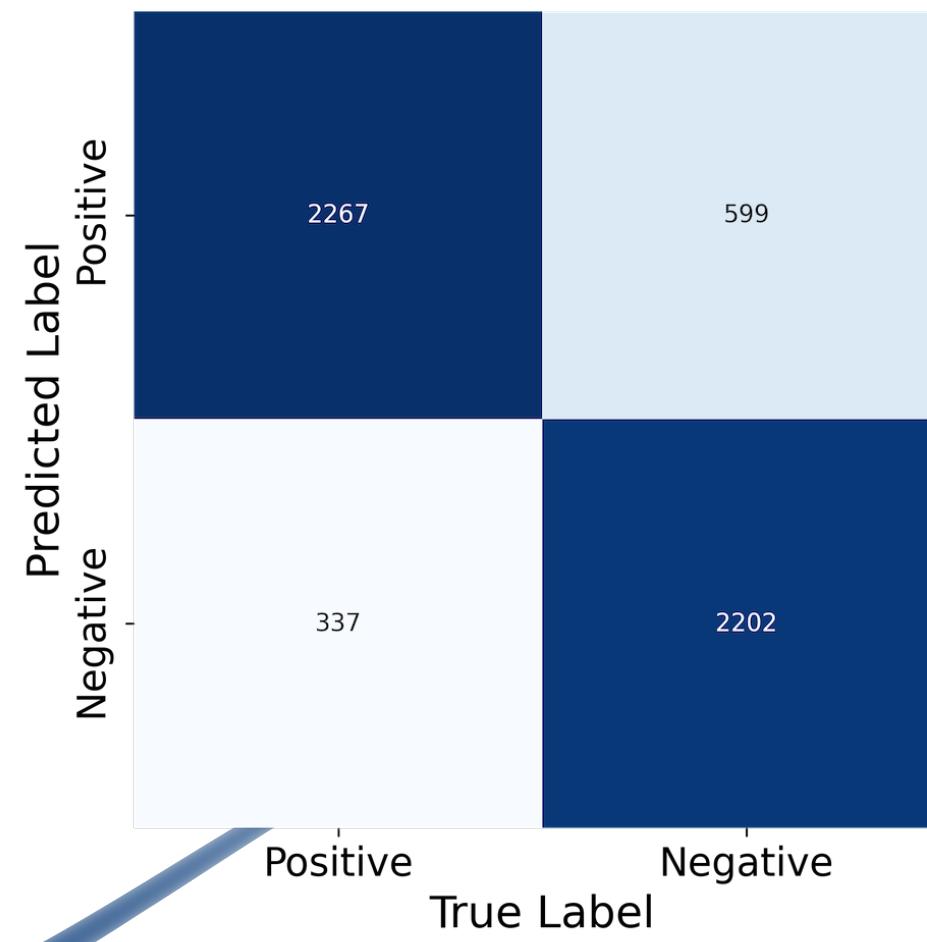


# Best Model

Best Model	Accuracy	Precision	Recall	F1-Score
GridSearch SVC	0.826827	0.867271	0.786148	0.824719

# Final Model

GridSearch SVC Confusion Matrix



# Conclusion

Britannia International Hotel underperforms compared to other hotels

We can use NLP to quickly analyze reviews

Word clouds can give insights

# Next Steps

Test model  
in other datasets

Create a recommendation  
system

Create a dashboard  
for hotels and users

# Thank you!

Email: alves.trevi@gmail.com  
<https://github.com/ismaeltrevi>