

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

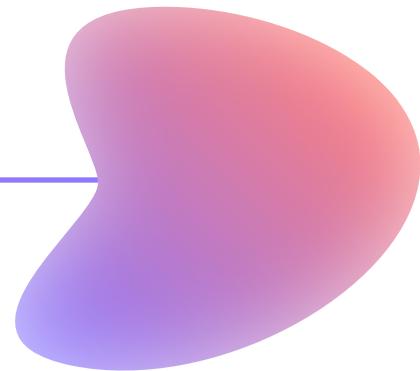


Process

Word Cloud

Sentiment Analysis

Modeling



Data

Size

515,738 reviews

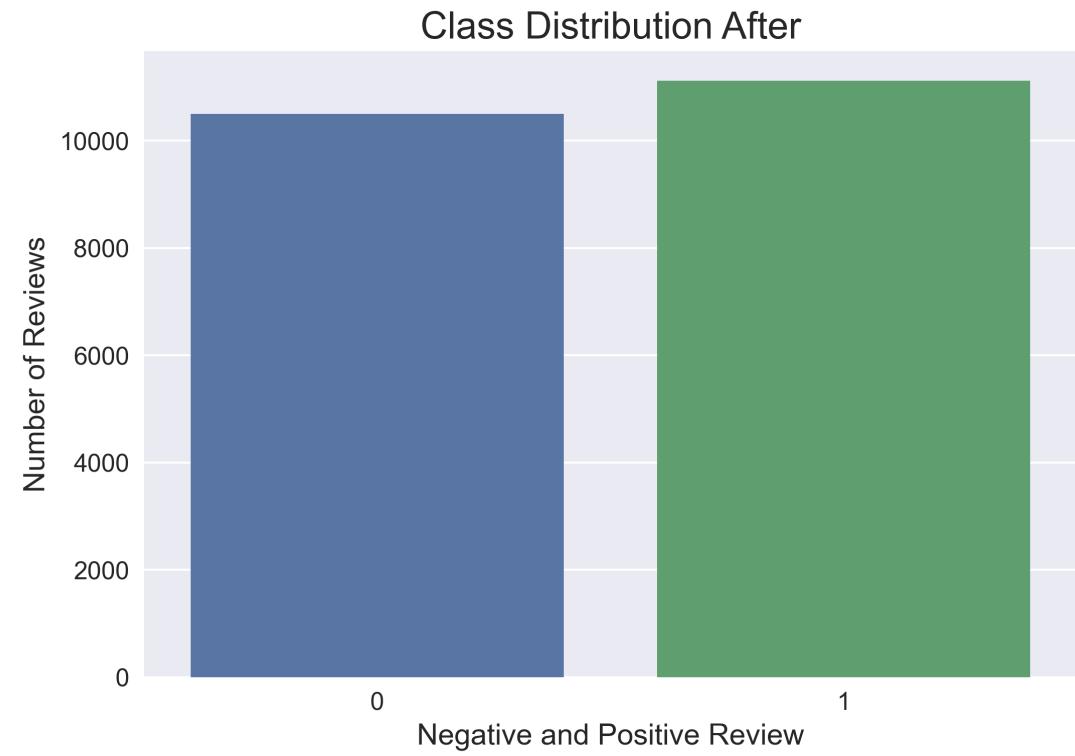
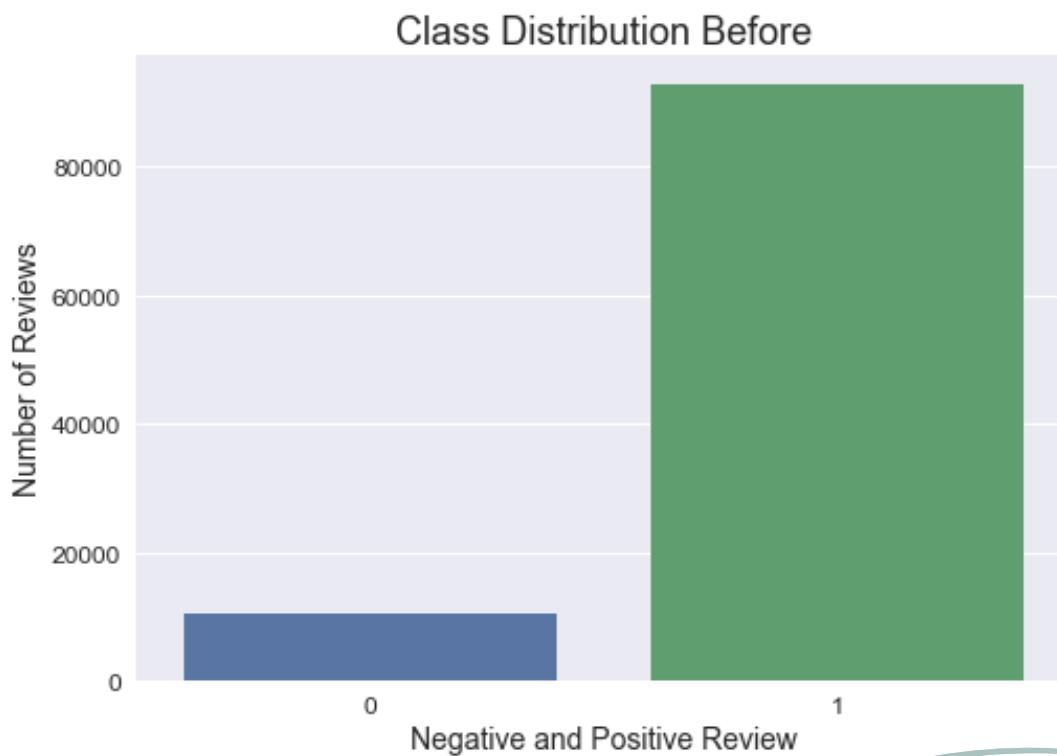
Source

Kaggle

Content

17 columns

Class Imbalance





Italy



Spain



France



United Kingdom



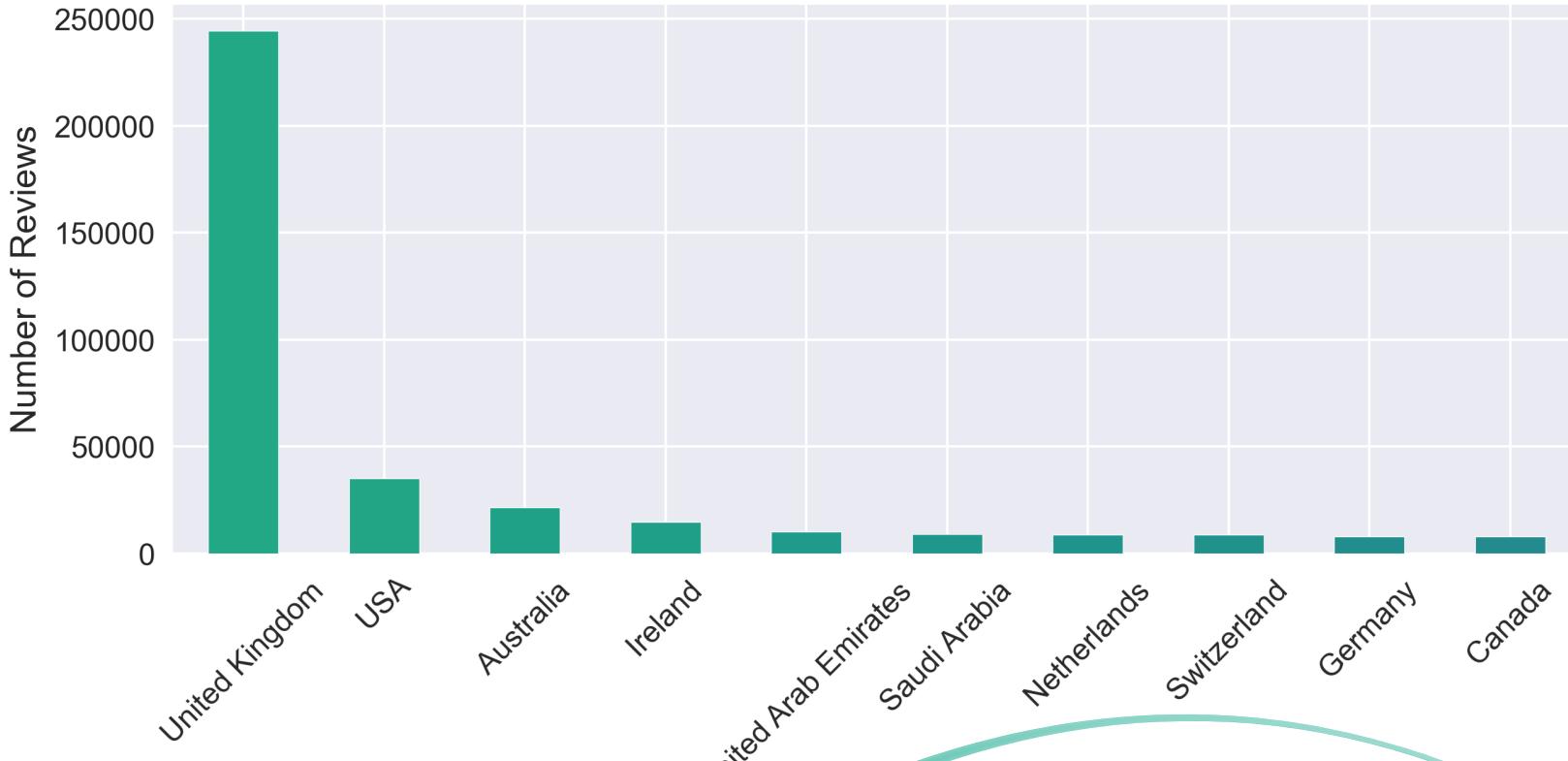
Austria



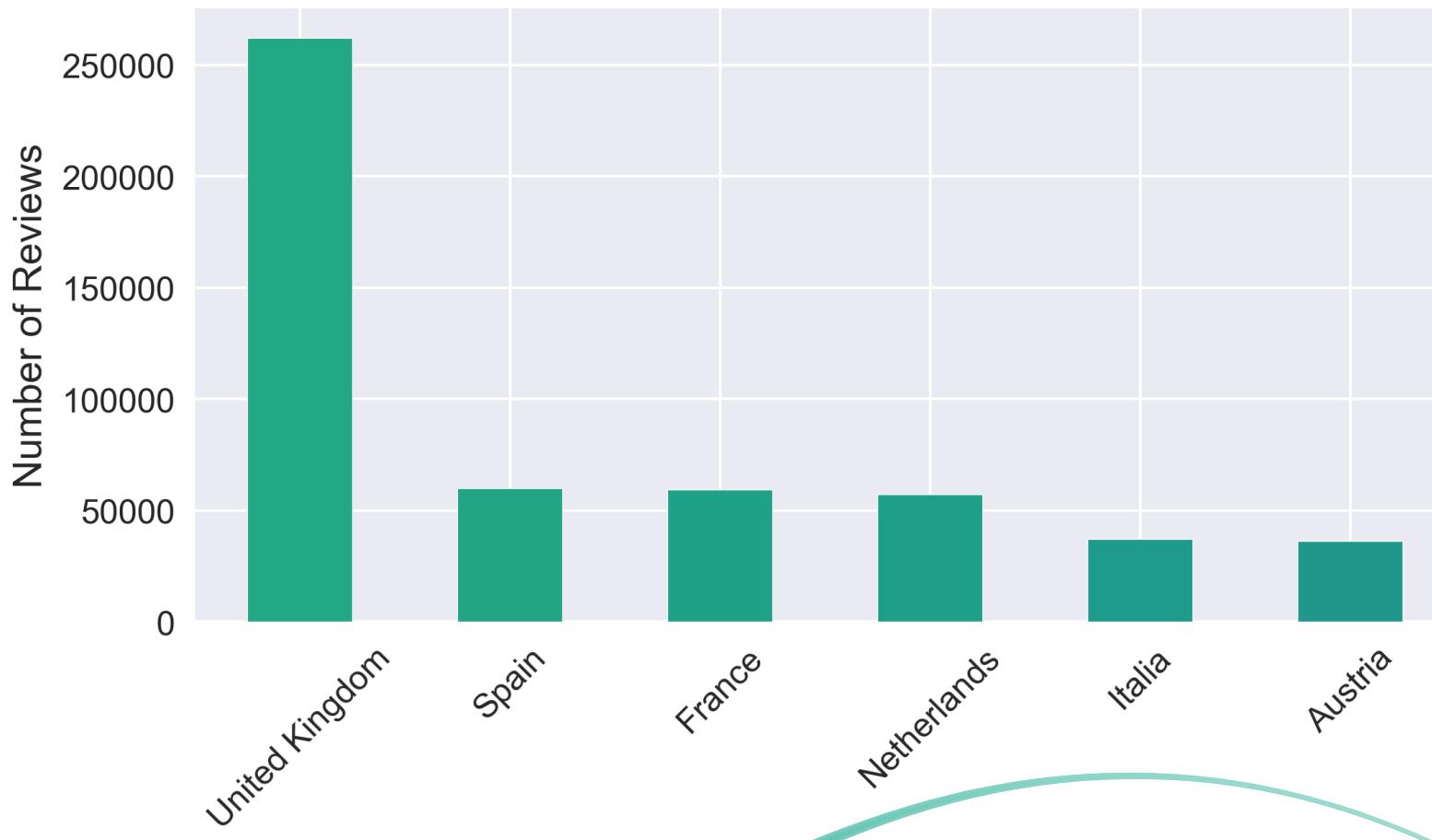
The Netherlands



Most common reviewer nationality



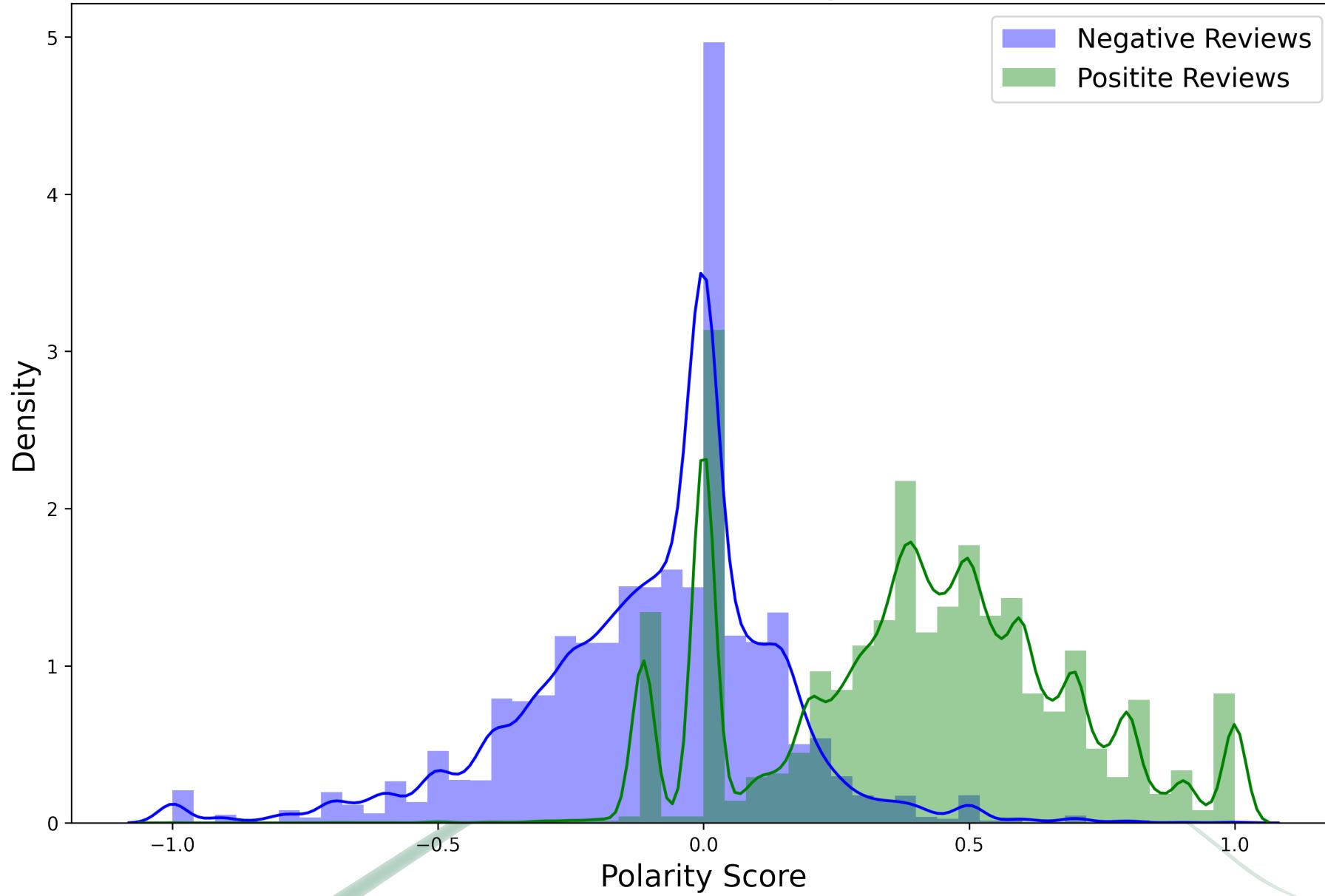
Countries of the Hotels with the Most Reviews



Score Density - Hotel Compared to London



Review Polarity



TripAdvisor Review Section



5,067
Reviews



58
Q+A



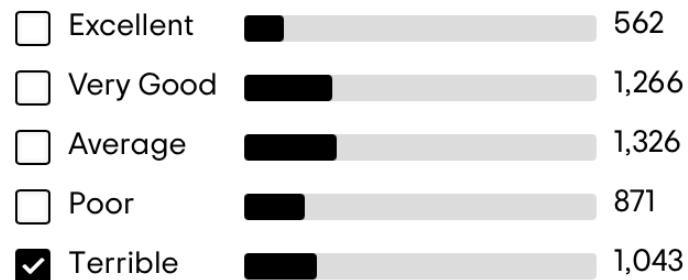
100
Room tips

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

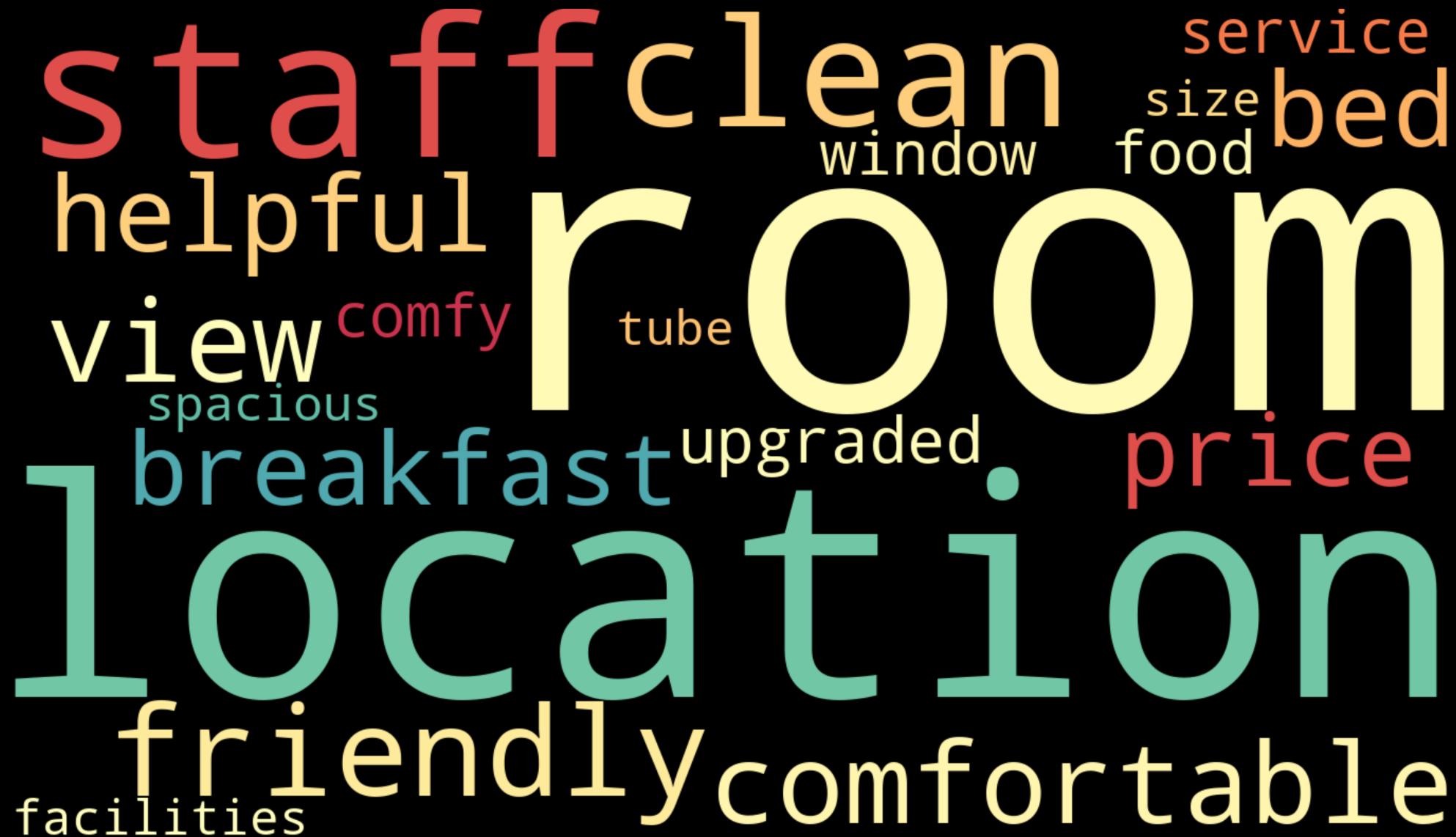
central london

full english

arena

thames

Positive Reviews Word Cloud

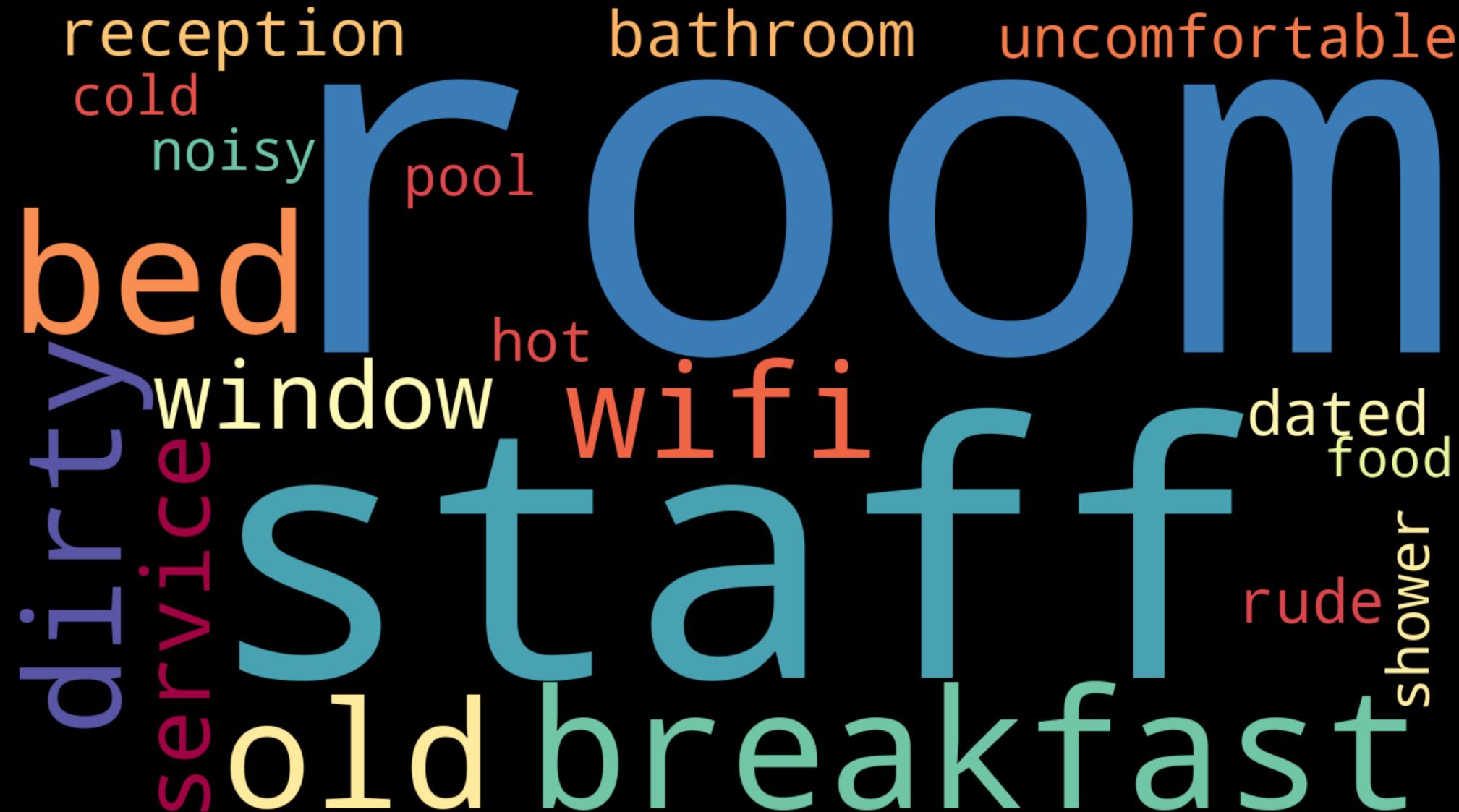


A word cloud visualization showing positive reviews. The words are arranged in a large central cluster with smaller words branching off. The colors of the words vary, including red, orange, yellow, green, blue, and teal.

The most prominent words are:

- room (large, yellow)
- location (large, teal)
- friendly (large, yellow)
- comfortable (large, yellow)
- facilities (medium, yellow)
- breakfast (medium, blue)
- spacious (small, teal)
- comfy (small, red)
- helpful (medium, orange)
- view (medium, yellow)
- clean (large, orange)
- service (small, red)
- size (small, orange)
- bed (small, orange)
- food (small, orange)
- window (small, orange)
- tube (small, yellow)
- upgraded (small, yellow)
- price (medium, red)

Negative Reviews Word Cloud

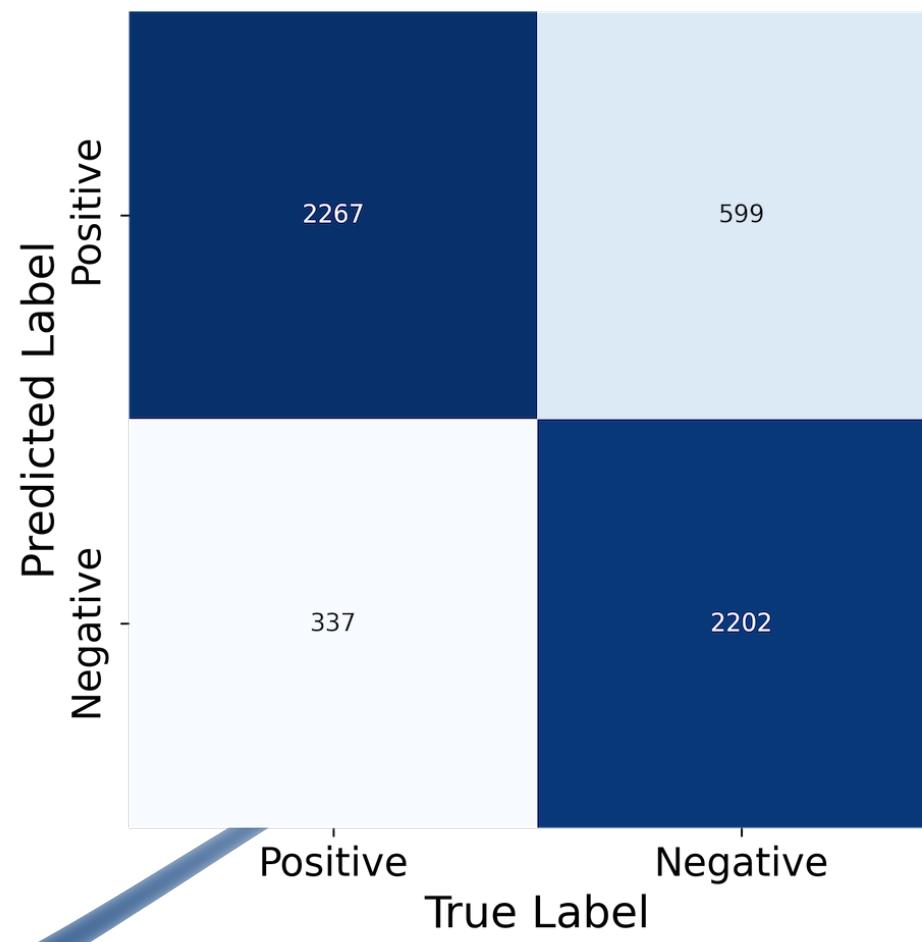


Best Model

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>