

# **ANALYSIS LUXURY HOTEL REVIEWS**

Pitch Deck

**BY VIPUL BALLUPET & CO.**



# CONTENTS

- 1 Motivation and Problem
- 2 Operations
- 3 Evaluating Results
- 4 Next steps and Conclusion



# MOTIVATION

Data science and Marketing are two of the key driving forces that help companies create value and stay on top in today's fast-paced economy.

Business value of consumer reviews and management responses to hotel performance

# MARKERS OF SUCCESS

1. overall rating
2. location
3. Cleanliness
4. volume of consumer reviews
5. Word-of-mouth (e-WOM)

43% of respondents said that they check consumer ratings and reviews “most of the time” [1]

# OUR SOURCE

Data scrapped from Booking.com

515,000 Customer reviews

1493 Luxury Hotels [2]



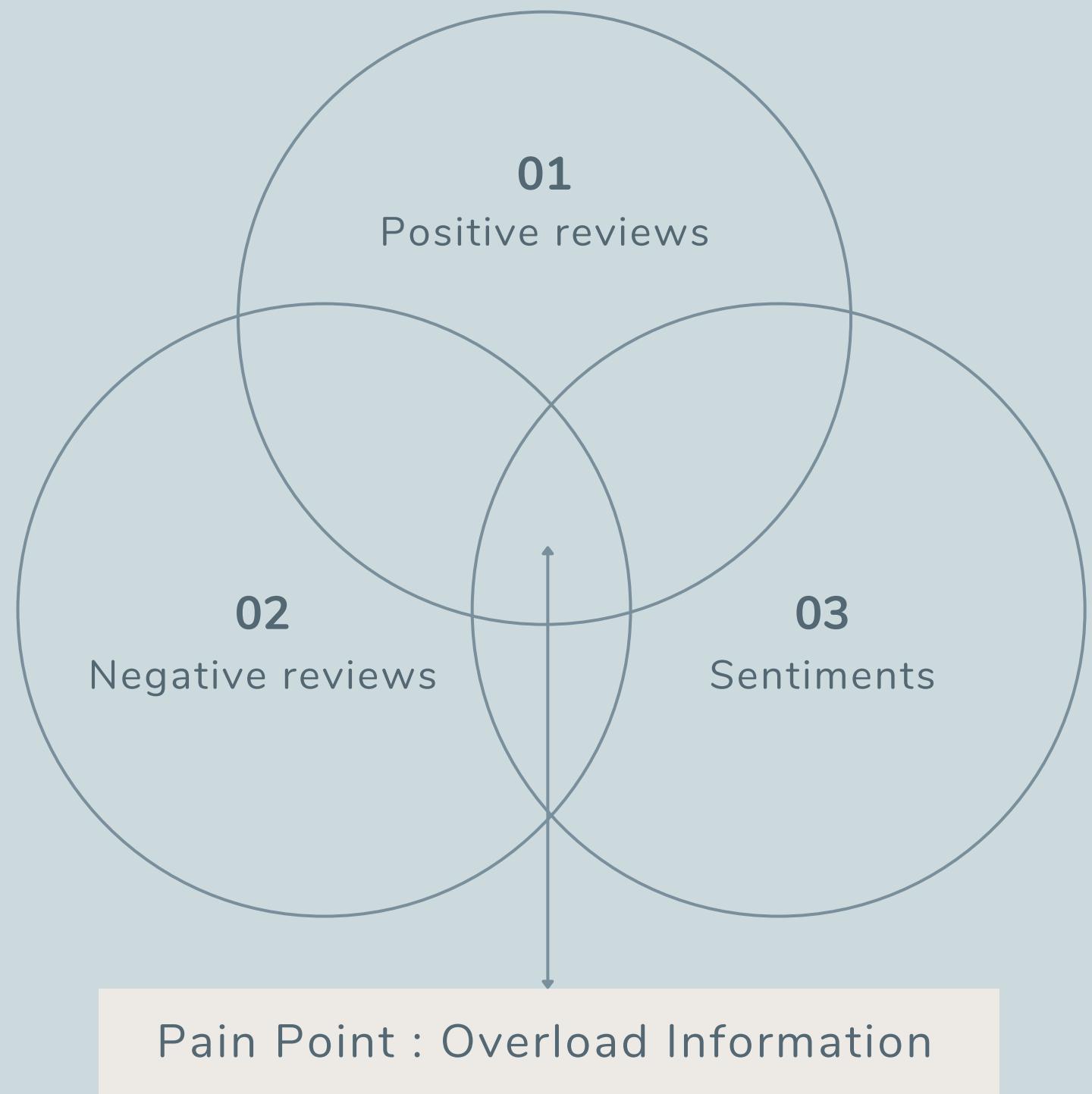
## PROBLEM

**SO MUCH  
INFORMATION SO  
LITTLE TIME**

can we read through all the reviews and ratings and pass a qualitative Judgement?



# PROBLEM ANALYSIS



# PROBLEM ANALYSIS

| hotel_name  | user_review                                       | reviewer_nationality | tags  | average_score | reviewer_score |
|-------------|---|----------------------|---|---------------|----------------|
| Hotel Arena | I am so angry that i made this post available...  | Russia               | [ 'Leisure trip ', 'Couple ', 'Duplex Double...]  | 8             | 3              |
| Hotel Arena | No Negative No real complaints the hotel was g... | Ireland              | [ 'Leisure trip ', 'Couple ', 'Duplex Double...]  | 8             | 8              |
| Hotel Arena | Rooms are nice but for elderly a bit difficul...  | Australia            | [ 'Leisure trip ', 'Family with young childre...] | 8             | 7              |
| Hotel Arena | My room was dirty and I was afraid to walk ba...  | United Kingdom       | [ 'Leisure trip ', 'Solo traveler ', 'Duplex...]  | 8             | 4              |
| Hotel Arena | You When I booked with your company on line y...  | New Zealand          | [ 'Leisure trip ', 'Couple ', 'Suite ', 'St...]   | 8             | 7              |

# MODUS OPERANDI

- 1 Planning and Managing
- 2 Data Exploration
- 3 Natural Language Processing
- 4 Evaluating Algorithms



# PROJECT PLANNING AND MANAGEMENT

**To-do [Examples]**

- DOD: Models successfully predicts ratings with corresponding review words
- DOD: apply these prediction models on the 3(to 5) random Hotels
- EPIC: analyze first 3 hotels with tags, and reviews
- EPIC: find out what key words match what average rating, for which hotels?

**Doing [Example]**

- presentation Draft
- save and update python Module-day1
- Distribution of raw and scaled data

**Done**

- dropping the coloumns ['Reviewer\_Nationality','days\_since\_review','lat','lng']
- apply standard scalar to scale before fitting the regression
- text cleaning modules
- Tableaux Draft
- create documents on Confluece
- text processing functions
- dimensionality reduction with PCA to deal with Multi-collinearity
- data cleaning modules

**Upcoming Events**

- + Add a card

# DATA CLANING AND EXPLORING

Do we have ..

1. missing Values,
2. Null
3. White spaces,
4. noises..?



# **DATA CLANING AND EXPLORING**

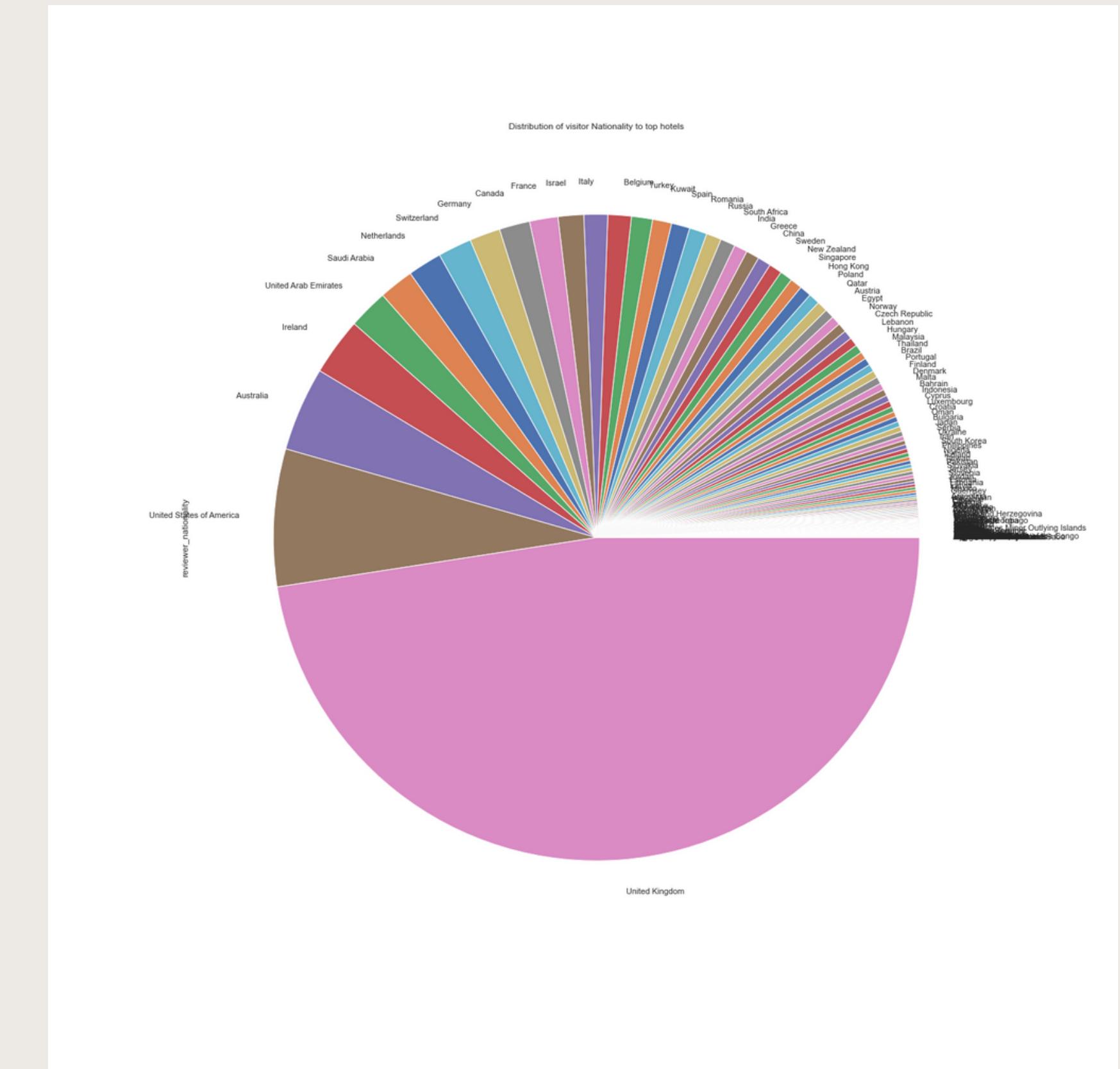
“Give me six hours to chop down a tree and I will spend the first four sharpening the axe.” – Abraham Lincoln

**“Data exploration .... the act of examining data to translate these various sources of Information into actionable insights”**

# NATIONALITY OF REVIEWERS

Highest Number of tourists hail from:

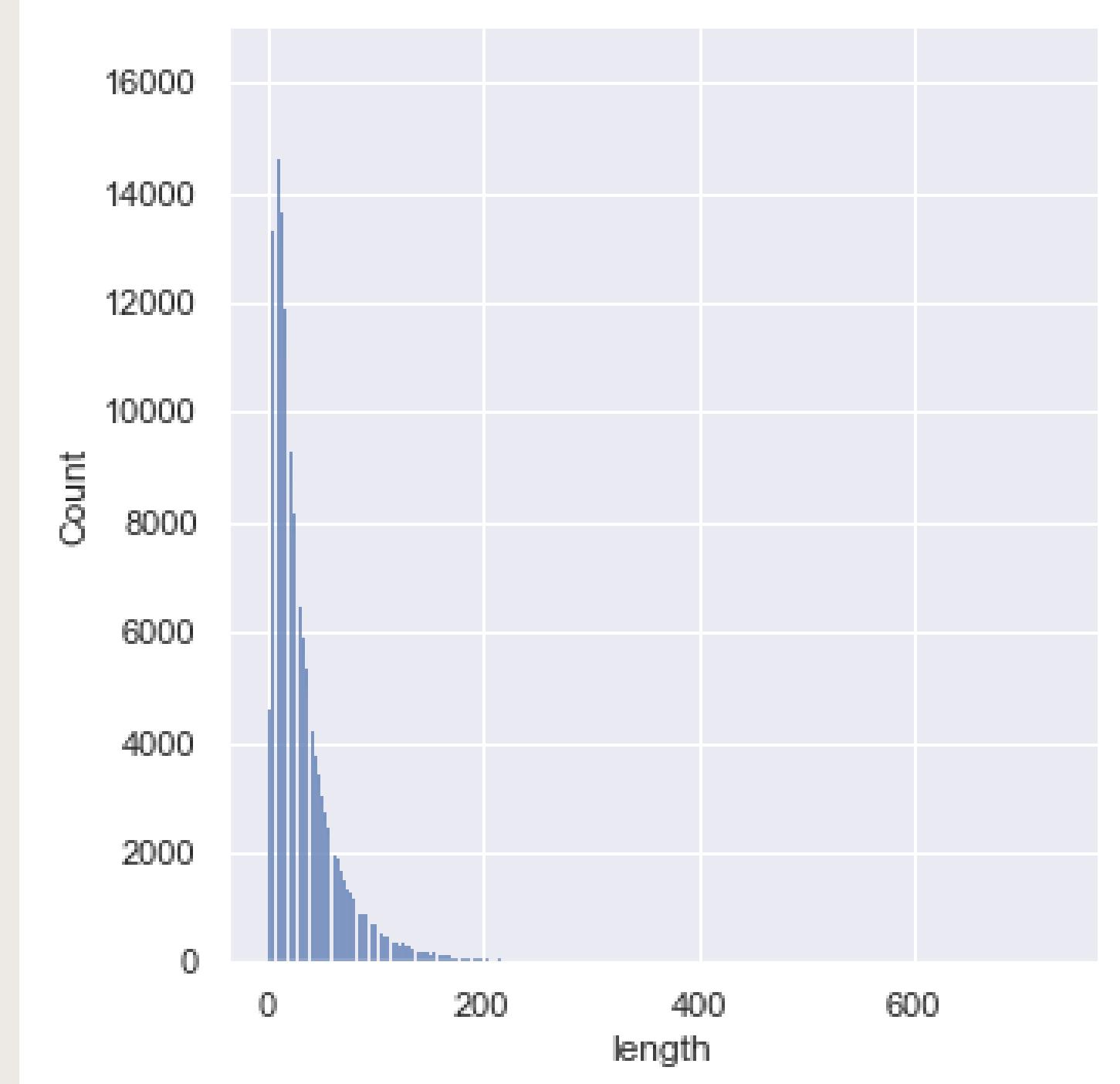
1. United Kingdom
2. United state of America
3. Australia
4. Ireland
5. United Arab Emirates



# WORD COUNT OF EACH REVIEWS

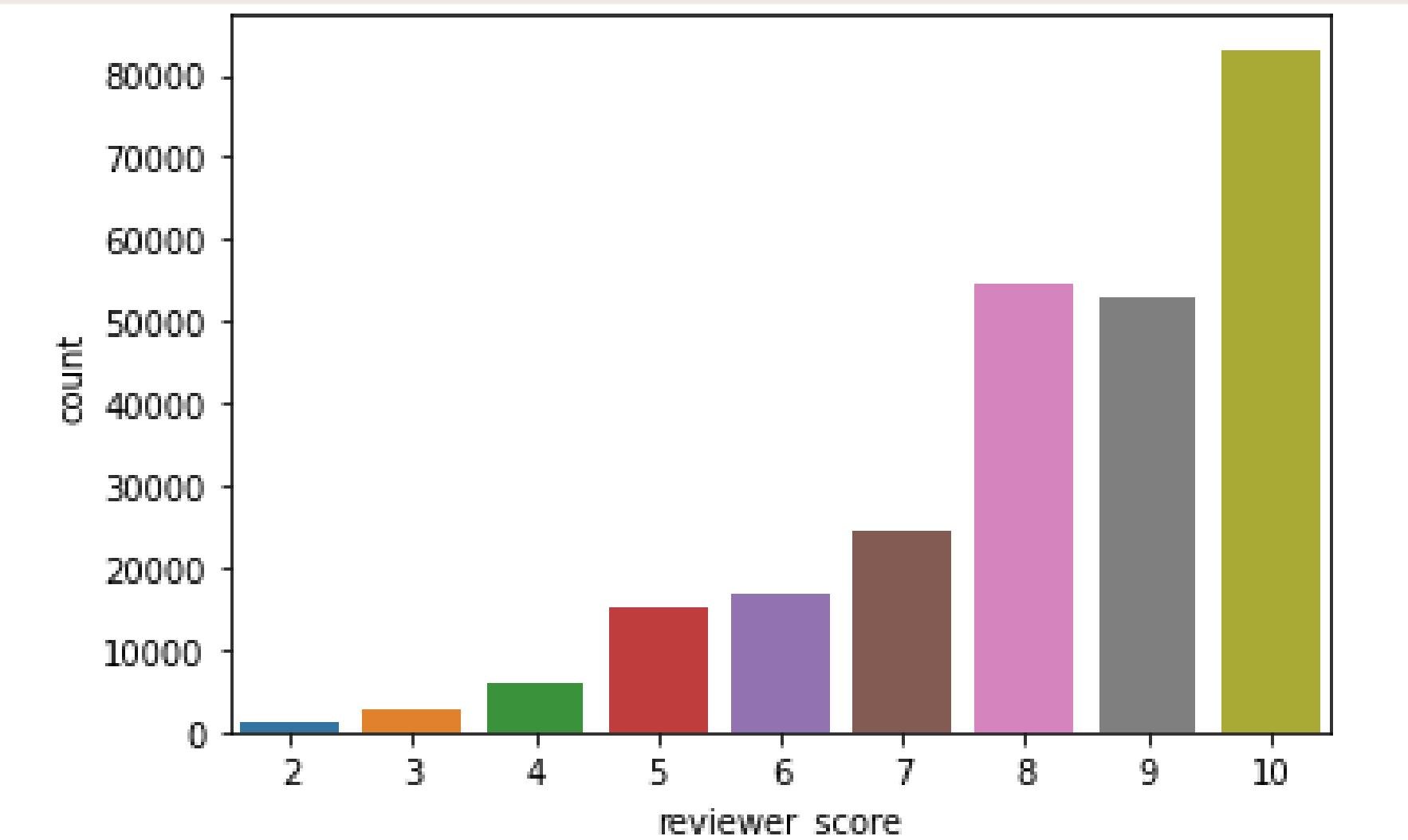
close to 15,000 reviewers used sentences that use 50 words

around 2000 users used upto 200 words to give the review



# WORD COUNT OF EACH REVIEWS

Overall distribution of the reviews for these Hotels



# NATURAL LANGUAGE PROCESSING

|   | hotel_name  | user_review                                       | reviewer_nationality | tags   | average_score | reviewer_score : |
|---|-------------|---|----------------------|--|---------------|------------------|
| 0 | Hotel Arena | I am so angry that i made this post available...  | Russia               | [ 'Leisure trip ', 'Couple ', 'Duplex Double...]   | 8             | 3                |
| 1 | Hotel Arena | No Negative No real complaints the hotel was g... | Ireland              | [ 'Leisure trip ', 'Couple ', 'Duplex Double...]   | 8             | 8                |
| 2 | Hotel Arena | Rooms are nice but for elderly a bit difficul...  | Australia            | [ 'Leisure trip ', 'Family with young children...] | 8             | 7                |
| 3 | Hotel Arena | My room was dirty and I was afraid to walk ba...  | United Kingdom       | [ 'Leisure trip ', 'Solo traveler ', 'Duplex ...]  | 8             | 4                |
| 4 | Hotel Arena | You When I booked with your company on line y...  | New Zealand          | [ 'Leisure trip ', 'Couple ', 'Suite ', 'St...]    | 8             | 7                |
| 5 | Hotel Arena | Backyard of the hotel is total mess shouldn t...  | Poland               | [ 'Leisure trip ', 'Group ', 'Duplex Double...]    | 8             | 7                |
| 6 | Hotel Arena | Cleaner did not change our sheet and duvet ev...  | United Kingdom       | [ 'Leisure trip ', 'Group ', 'Duplex Twin Ro...]   | 8             | 5                |

# NATURAL LANGUAGE PROCESSING

## Text Cleaning

- use lowercase
- drop punctuation marks

## Tokenising

- how many times a word appears in the reviews

## Stemming

- the "root" of each word
- Ex: loved, love and lovers have the same root: love

## Reducing the Stop words

- drop articles, prepositions,...etc

```
[('hotel', 6099),  
 ('staff', 5996),  
 ('location', 3969),  
 ('breakfast', 3421),  
 ('good', 3057),  
 ('great', 2714),  
 ('bed', 2382),  
 ('small', 2199),  
 ('friendly', 2153),  
 ('rooms', 2031),  
 ('helpful', 2001),  
 ('nice', 1867),  
 ('would', 1836),  
 ('comfortable', 1731),  
 ('clean', 1731),  
 ('stay', 1669),  
 ('us', 1553),  
 ('bathroom', 1525),  
 ('excellent', 1482),  
 ('could', 1448),  
 ('bar', 1435),  
 ('negative', 1433),  
 ('one', 1416),  
 ('really', 1364)]
```

# EVALUATING ML MODELS

we used ML to predict and/or classify the results ie user reviews



## PREDICTIONS

How successfully can our model predict the reviews:

5 OR 7

## CLASSIFICATION

How successfully can our model Classify the reviews:

GOOD or BAD

# EVALUATION OF PREDICTION MODEL

```
Results obtained for the TRAIN SET  
=====  
r2_train      = 0.23645877600779475  
r2_test       = 0.2322797868929648  
=====  
Other metrics  
MSE = 2.4087239651989782  
RMSE = 1.5520064320739713  
MAE = 1.209927874996469
```

Our model can predict the user reviews based on the keywords with an accuracy of 23%

Improvement needed!!

# EVALUATION OF CLASSIFICATION MODEL

| =====                             |           |        |          |         |
|-----------------------------------|-----------|--------|----------|---------|
| Results obtained for the TEST SET |           |        |          |         |
| The Cohen's Kappa is: 0.09        |           |        |          |         |
|                                   | precision | recall | f1-score | support |
| 10                                | 0.57      | 0.38   | 0.45     | 24878   |
| 2                                 | 0.02      | 0.36   | 0.04     | 346     |
| 3                                 | 0.03      | 0.21   | 0.06     | 867     |
| 4                                 | 0.05      | 0.14   | 0.08     | 1863    |
| 5                                 | 0.09      | 0.11   | 0.10     | 4538    |
| 6                                 | 0.10      | 0.14   | 0.12     | 5067    |
| 7                                 | 0.12      | 0.13   | 0.13     | 7418    |
| 8                                 | 0.25      | 0.17   | 0.20     | 16488   |
| 9                                 | 0.25      | 0.20   | 0.23     | 15764   |
| accuracy                          |           |        | 0.23     | 77229   |
| macro avg                         | 0.17      | 0.21   | 0.16     | 77229   |
| weighted avg                      | 0.32      | 0.23   | 0.26     | 77229   |

Our model can **classify** the user reviews based on the keywords with an accuracy of 23%

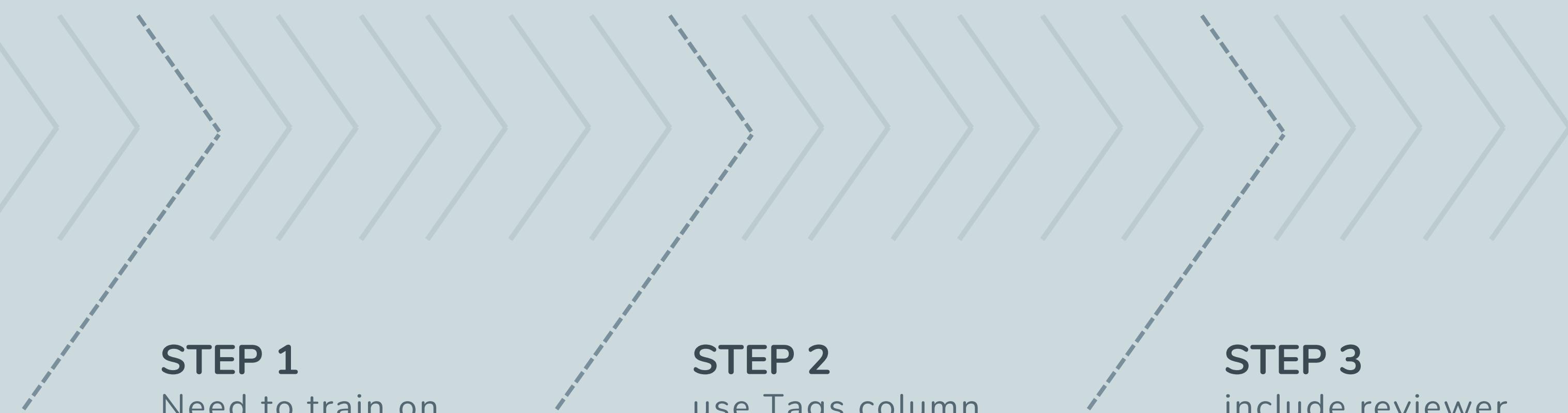
```
print("Accuracy:", nltk.classify.accuracy(classifier, test)*100)
```

Accuracy: 34.22

with NLTK classifier we get accuracy of 34.22%

# NEXT STEPS

Improve the performance of the algorithms with more than 60 % accuracy



## STEP 1

Need to train on more frequent words

## STEP 2

use Tags column that is already provided to save time

## STEP 3

include reviewer Nationality to the equation

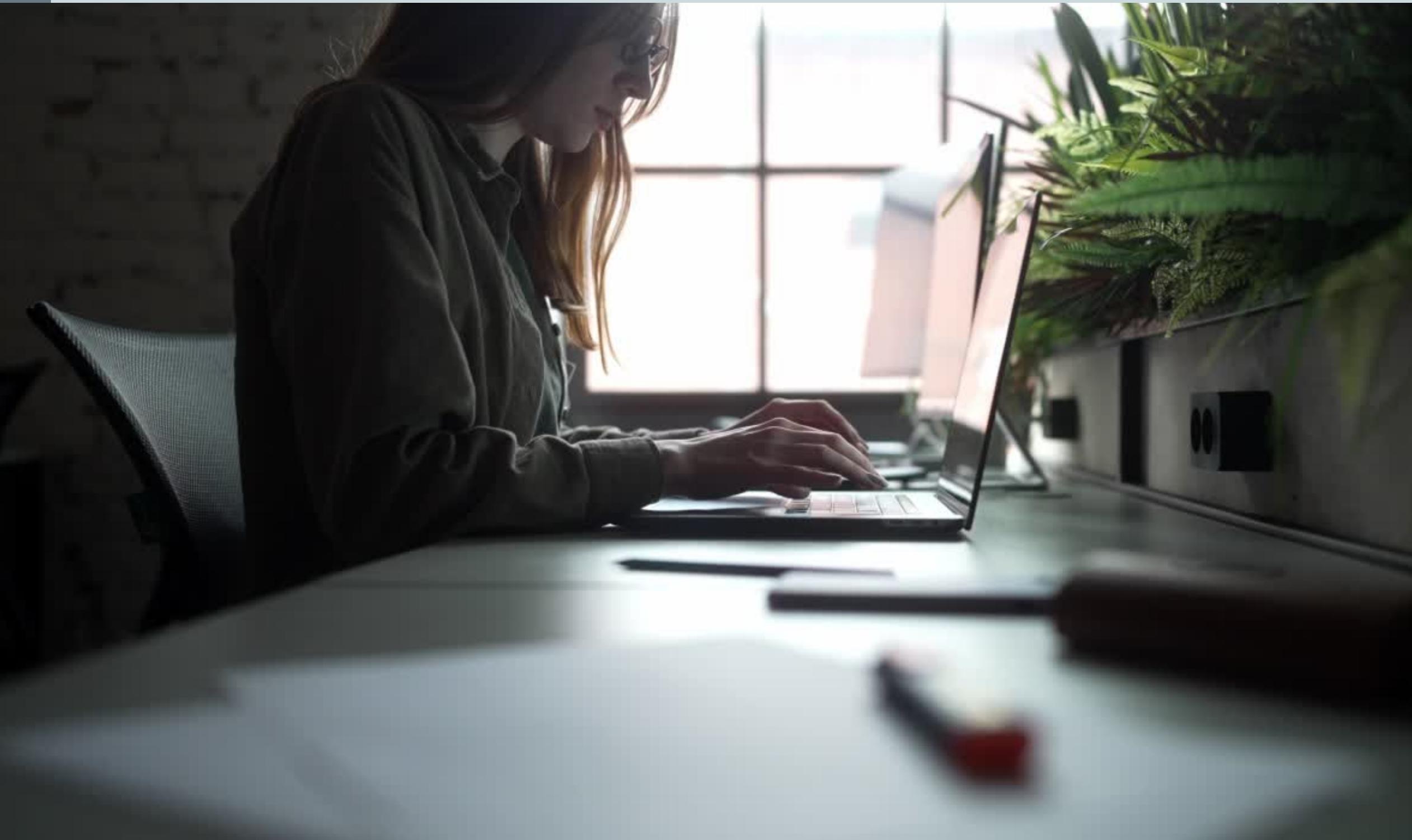
# CONCLUSION

The relationships among consumer reviews, hotel performance management responses were examined and understood in this project.



## CONTACT ME

[vipulballupet@gmail.com](mailto:vipulballupet@gmail.com)  
+49 152 1516 7273



**THANK YOU FOR YOUR ATTENTION**