

The background of the slide is an abstract watercolor wash in various shades of blue, ranging from deep navy to light sky blue, with soft, organic edges. A solid dark blue rectangular box is positioned on the left side of the slide, serving as a container for the title and authors.

THE HOTEL INDUSTRY: COVID-19 IMPACTS

ASPEN SCHMIDT, XITONG HU, TRINH
PHAN, AND MARYAM TAHERIRANI

OVERVIEW

1

**Project
Introduction
& Purpose**

2

**Data
Preparation**

3

**Visualization
& Analysis**

4

**Descriptive
Analysis**

5

**Results &
Assessment**

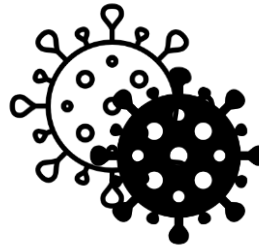
6

Conclusion

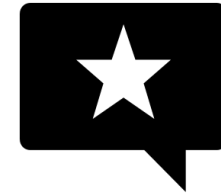
INTRODUCTION



- The hotel industry has been impacted by the COVID-19 Pandemic
- Selected hotels from the city of Chicago for study



- The COVID-19 Pandemic has caused travelers and hotel guests to be more conscious and concerned about where they stay in order to remain safe



- One way hotel guests voice their experiences is through reviewing and rating their hotel stay
- Will explore their reviews to uncover insights

PROJECT PURPOSE

The purpose of this project is to **uncover** the primary hotel guest expectations and the overall feelings guests have towards their hotel stay experience during the COVID-19 Pandemic.

By understanding this, hotel companies can **refine** their current processes and standards.

Therefore, hotels will additionally **improve** their reviews and overall ratings and **increase** occupancy.



DATA PREPARATION



Data Access

Collected from
hotels.com

400 Chicago Hotels

2 Datasets: **Hotel Features** for
Descriptive and
Visualizations
Hotel Reviews for Text
Mining and
Sentiment Analysis



Data Consolidation

CVS Writer was
opened, and the file
was saved on the
directory

Web crawler wrote
data into each row
for each hotel

Two tables:
hotel_listings and
hotel_reviews



Data Cleaning

hotel_listings:

- Special characters removal
- Extract specific values
- Numeric to integer or float
- Removing missing values

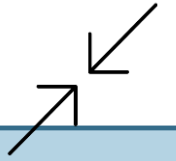
hotel_reviews:

- Remove stop words, numerical values, punctuation, uninformative frequent words
- Convert to lower case
- Stemming



Data Transformation

hotel_listings:
Log transformation
was applied to the
predictor variables
with the goal of
achieving a better
distribution



Data Reduction

COVID-19 Pandemic
timeframe

Recent 100 reviews
for hotels scraped

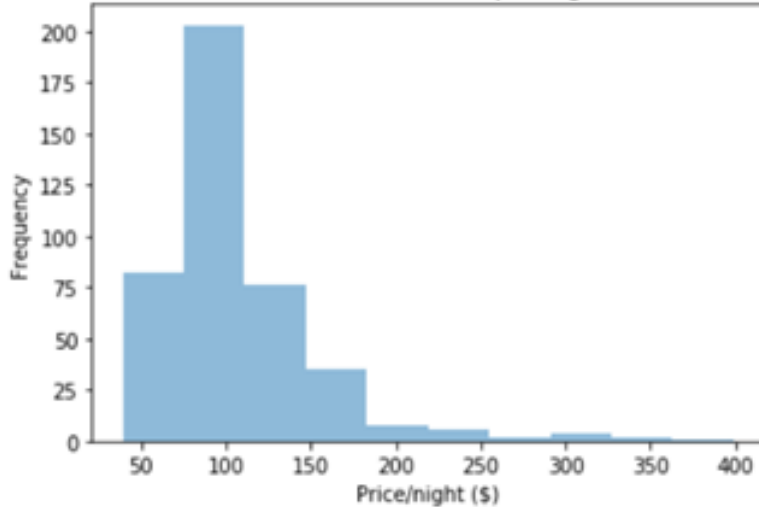
Approximately 10,000
reviews

Hotels that did not
have 100 reviews or
text were removed

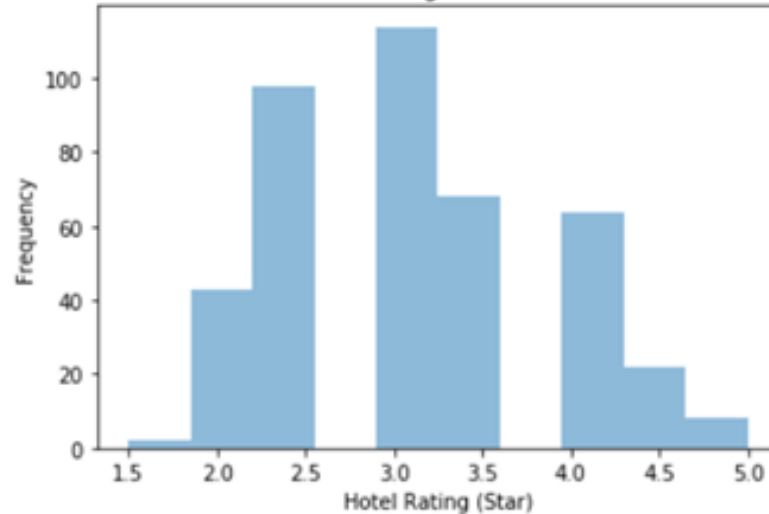
Final dataset had a
total of **9,488**
reviews

VISUALIZATIONS

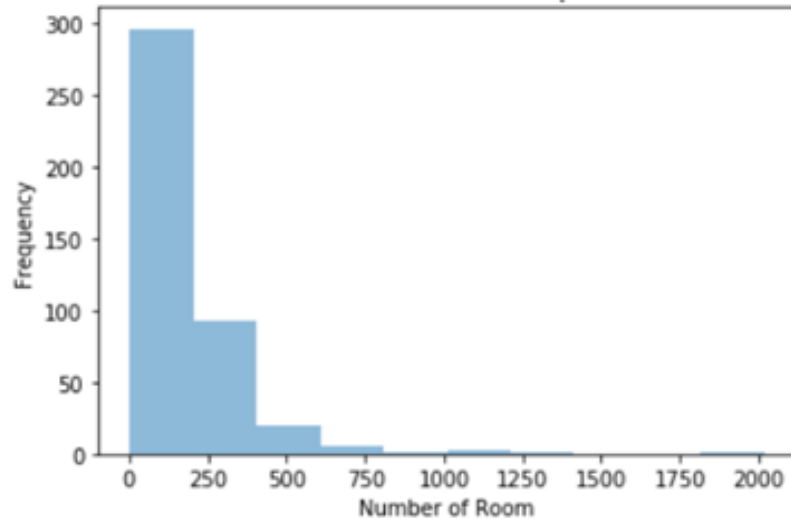
Distribution of Price per Night



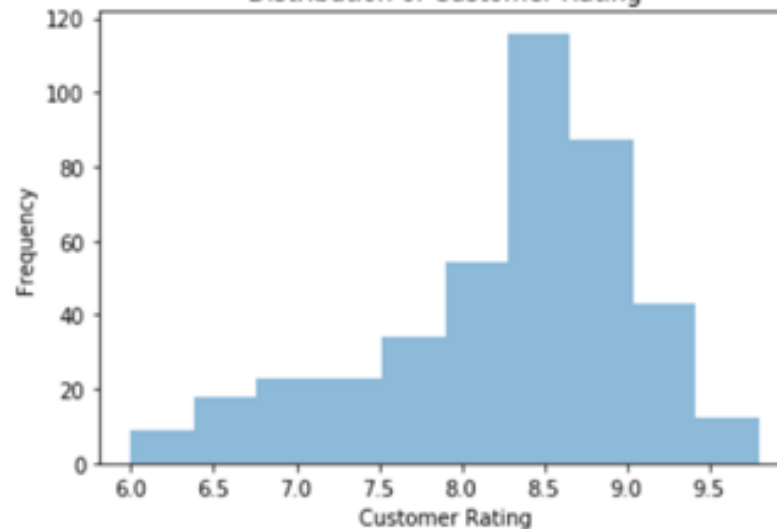
Hotel Rating Distribution



Distribution of Total Room per Hotel



Distribution of Customer Rating

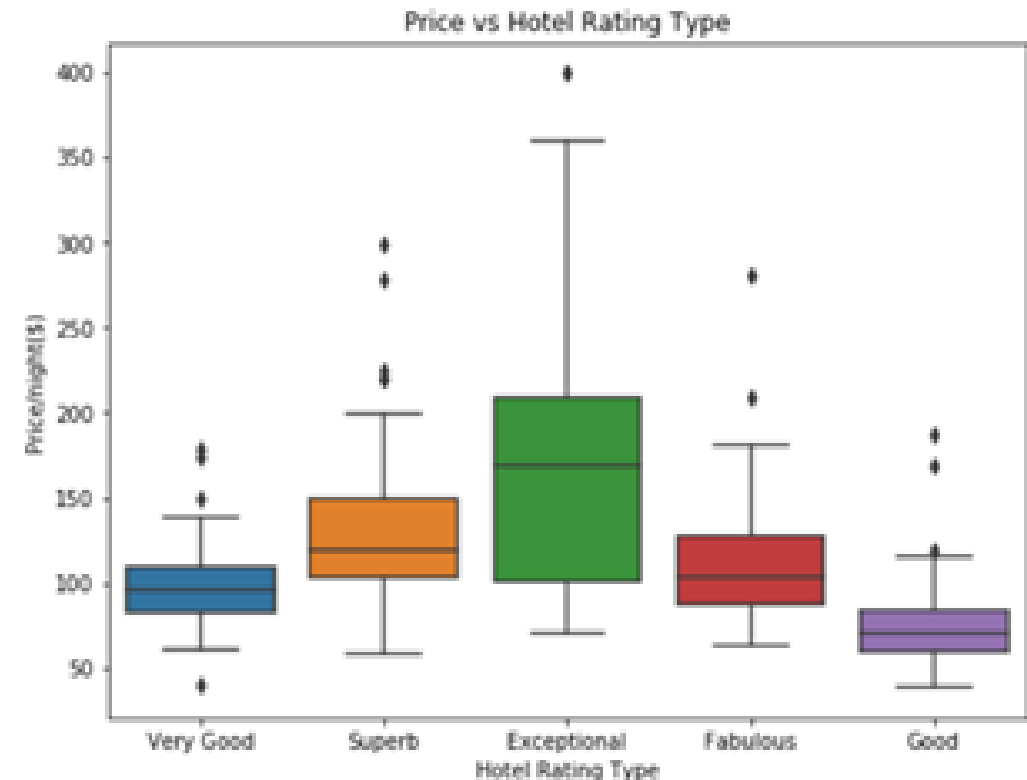
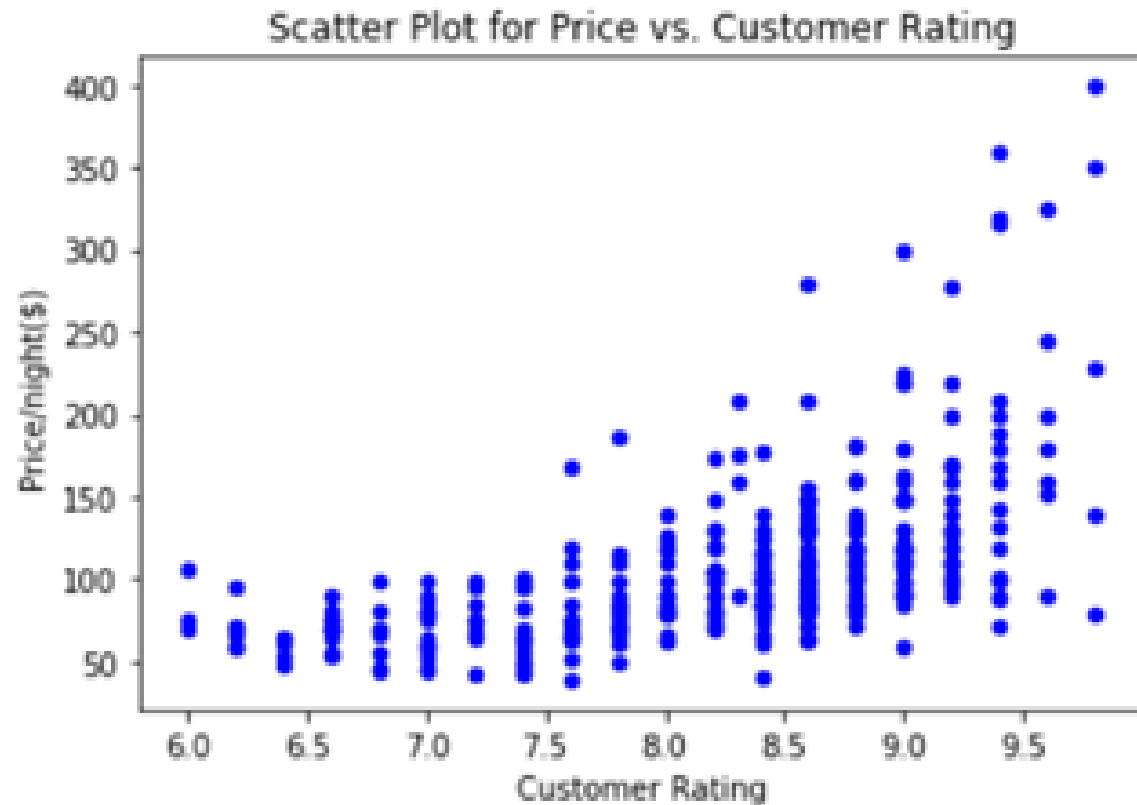


Summary Statistics:

- Avg. Price: \$107/night
- Hotel Star: 3*
- Avg.Total Rooms: 184
- Avg. Rating: 8.3
- Avg.Total Reviews: 591

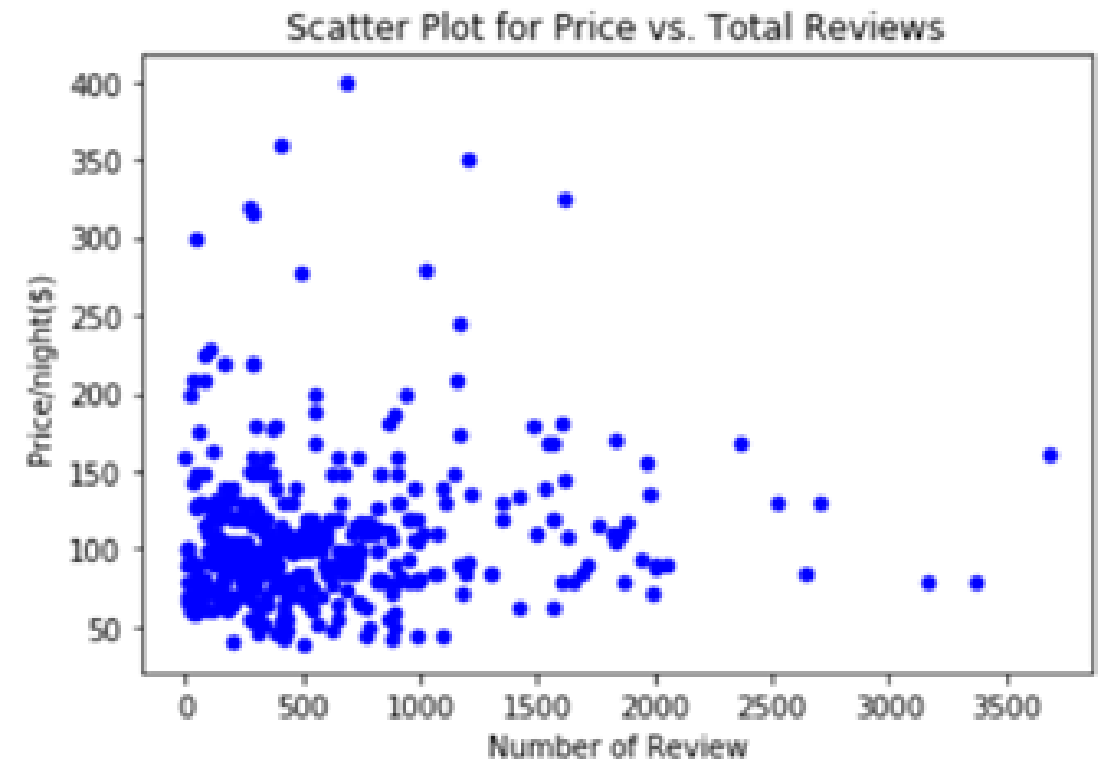
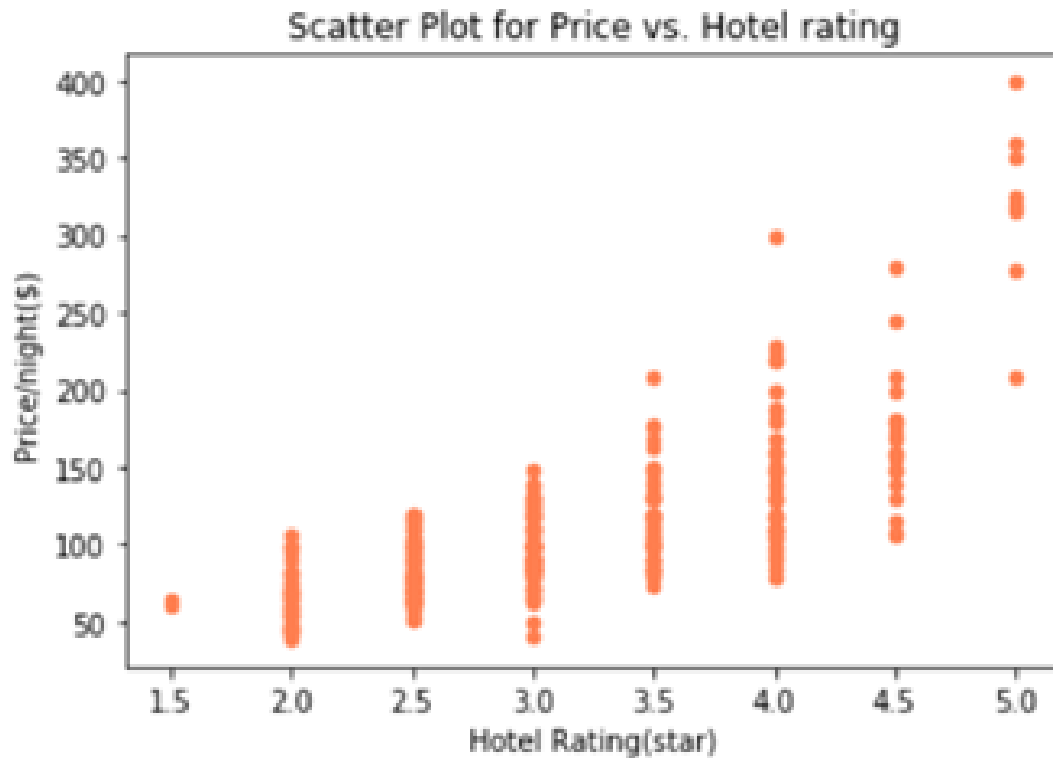
DESCRIPTIVE ANALYSIS

- Strong linear correlation between Customer Rating and Price.
- Exceptional Rating has significantly higher avg. price than others.

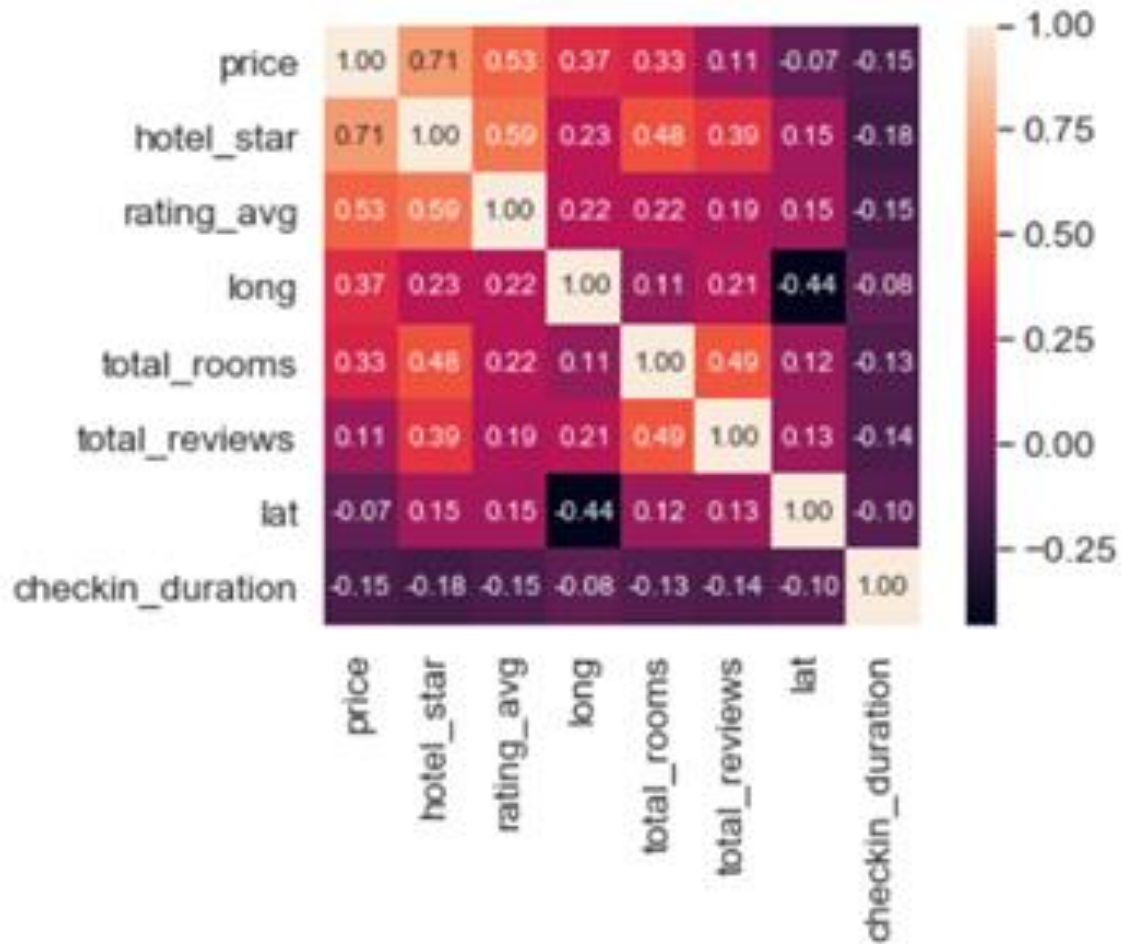


DESCRIPTIVE ANALYSIS

- Hotel Rating and Number of Reviews also affect the Price.

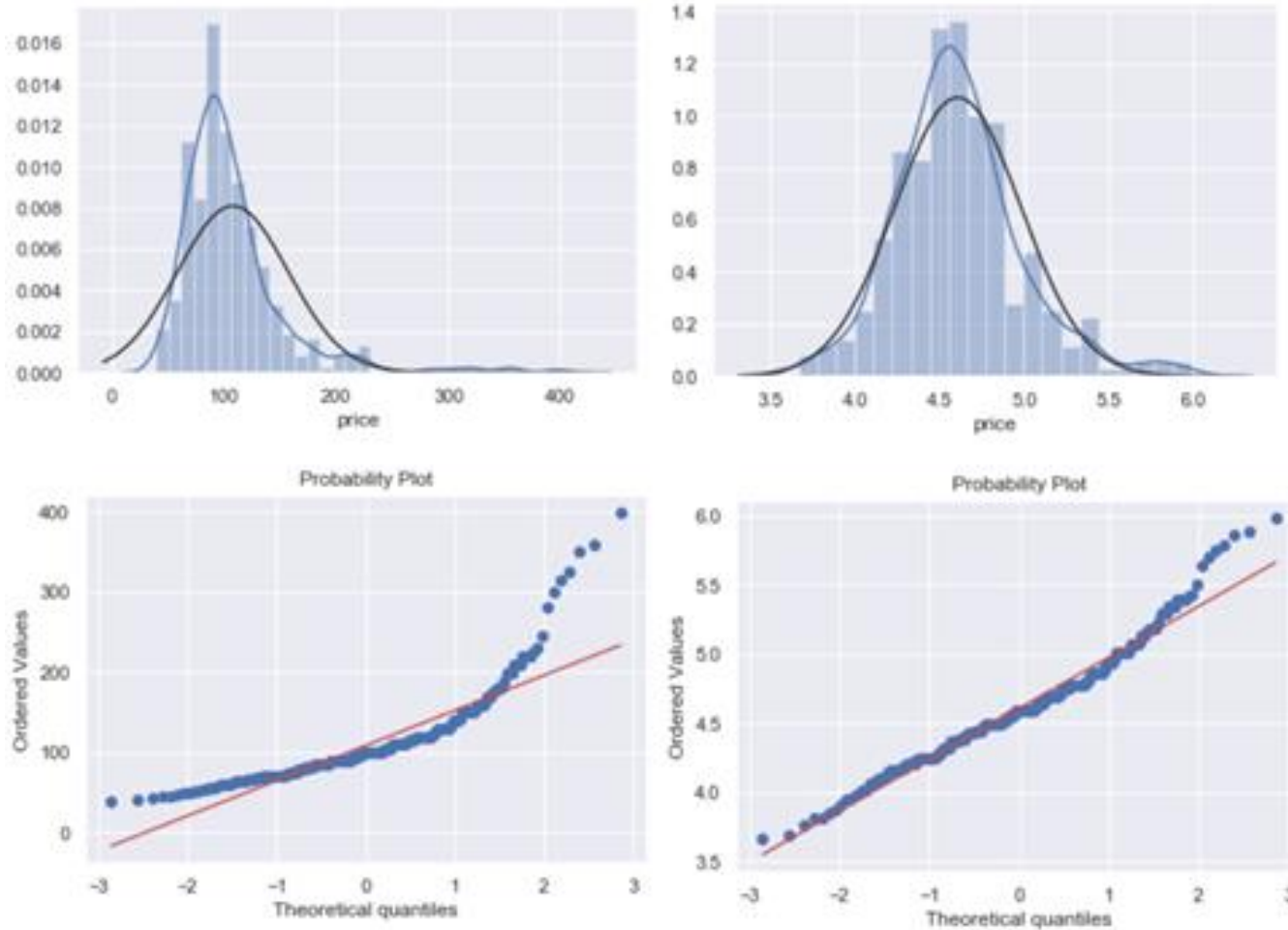


CORRELATION MATRIX



- Hotel Star and Customer Rating are highly correlated.
- Total Reviews has light correlation.

DATA TRANSFORMATION



- Log Transformation:
- Target Variable: Price
- Predictors: Continuous Variables

PRICE PREDICTION

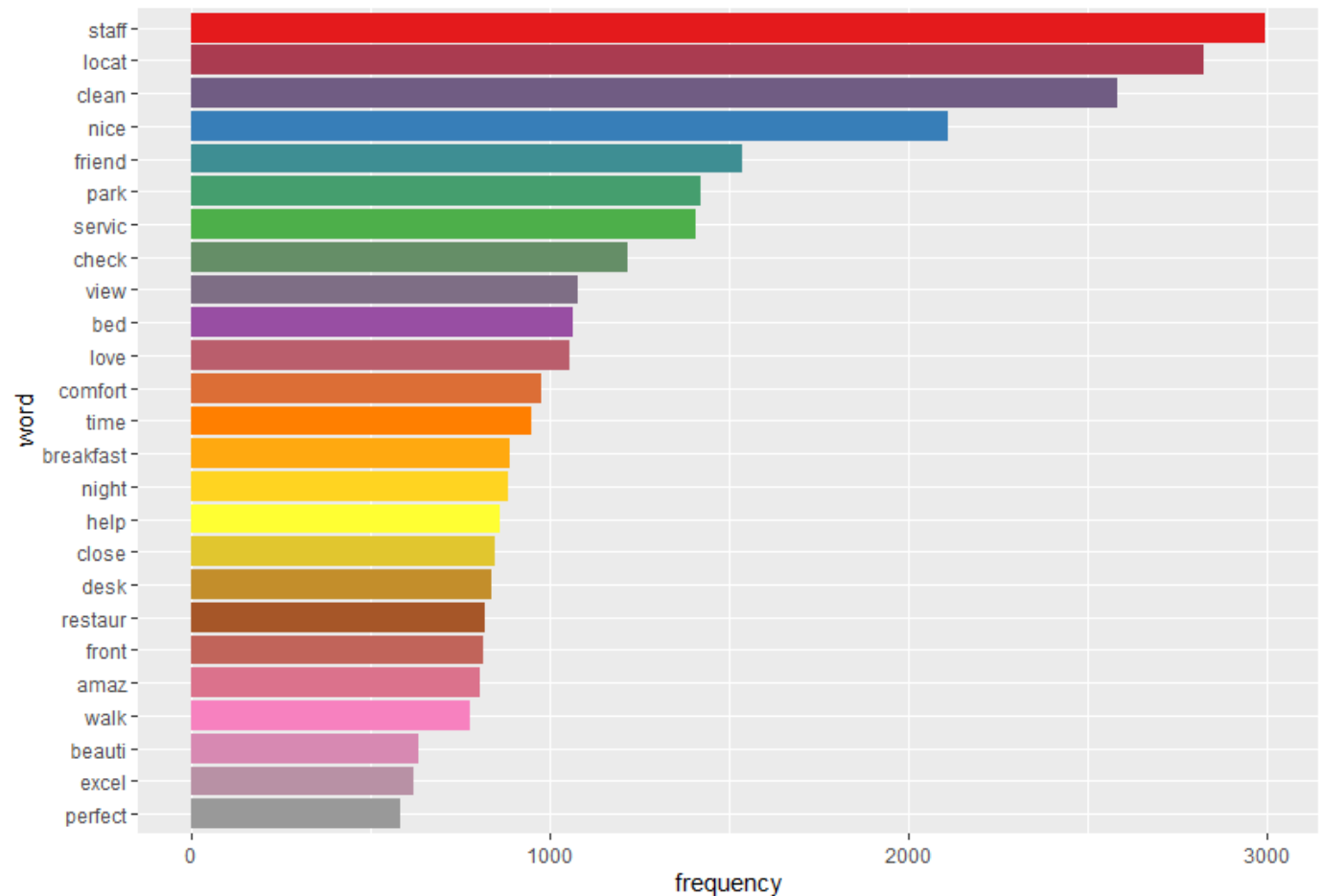
$$\text{Price} = -99.86 + 43.86 * \text{hotel_star} + 9.1 * \text{rating_avg} + 0.02 * \text{total_rooms} - 0.016 * \text{total_reviews}$$

- R-Square: 58%
- Mean Square Error: 803

Index	Actual Price	Predicted Price
324	101	105.97
108	127	100.19
281	80	93.71
242	99	75.01
198	169	121.35
277	139	132.13
209	75	59.40
118	79	71.79
303	99	108.73
244	129	134.24

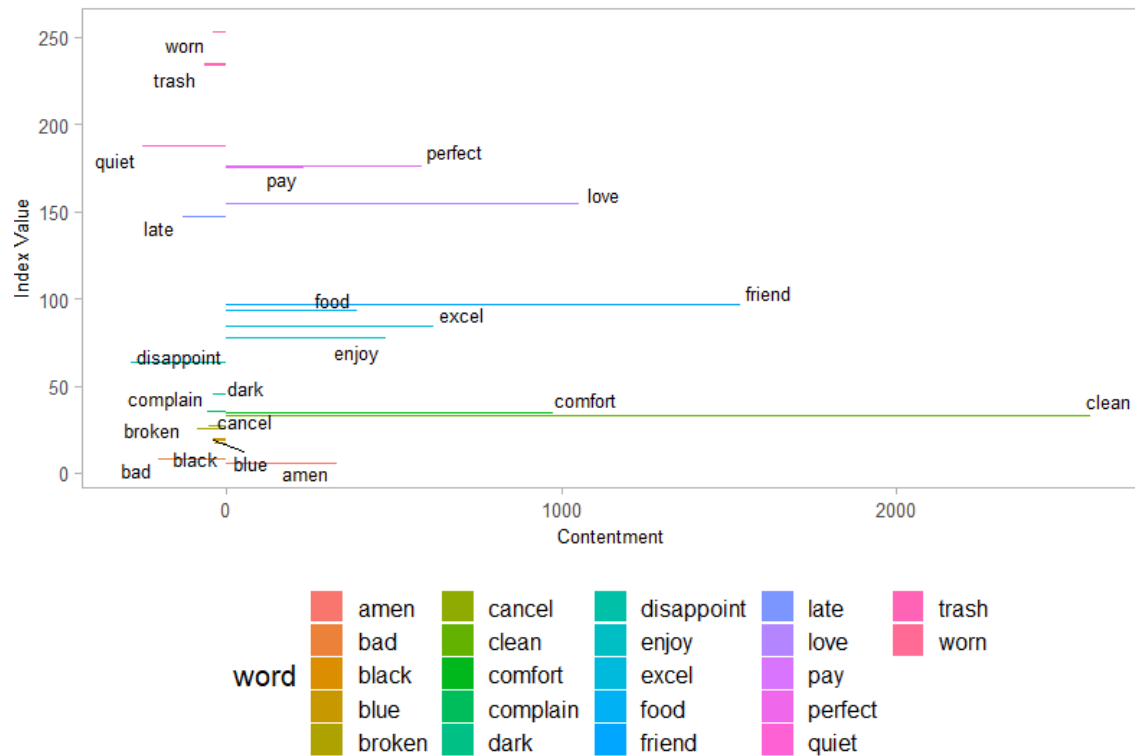
TEXT MINING AND SENTIMENT ANALYSIS

- Staff plays an important role in the customers' feelings about their hotel experience
- Other words such as “locat” and “clean” suggest customers care about the location and the sanitary condition of hotel
- Other import aspects guests evaluate the hotel on such as: parking, breakfast, and restaurant options
- Most of these top words are positive

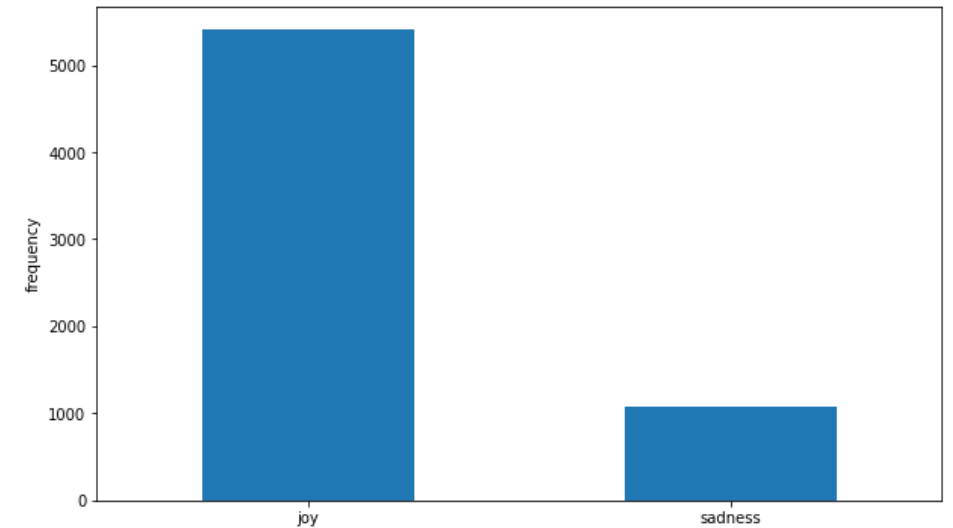


TEXT MINING AND SENTIMENT ANALYSIS

Joy/ Sadness Contentment based on Words



Joy and Sadness Reviews

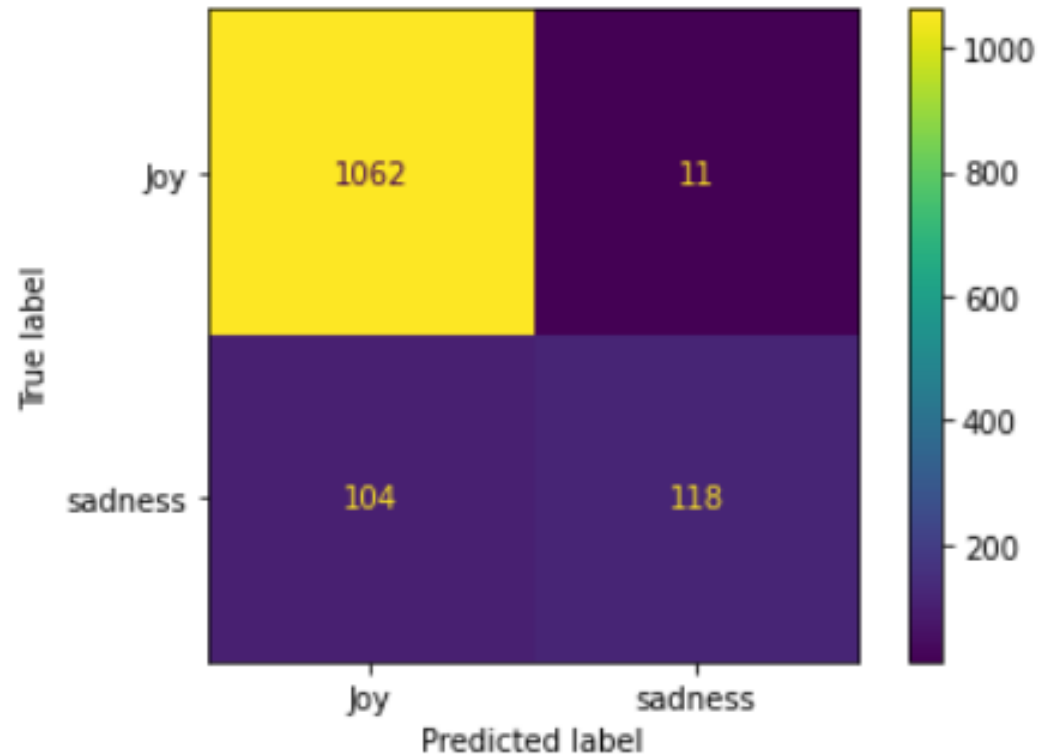


- Joy: Clean, friend, comfortable
- Sadness: Complain, disappointed and bad

- 5,406 joy reviews
- 1,067 sadness reviews

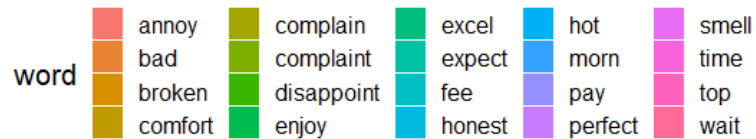
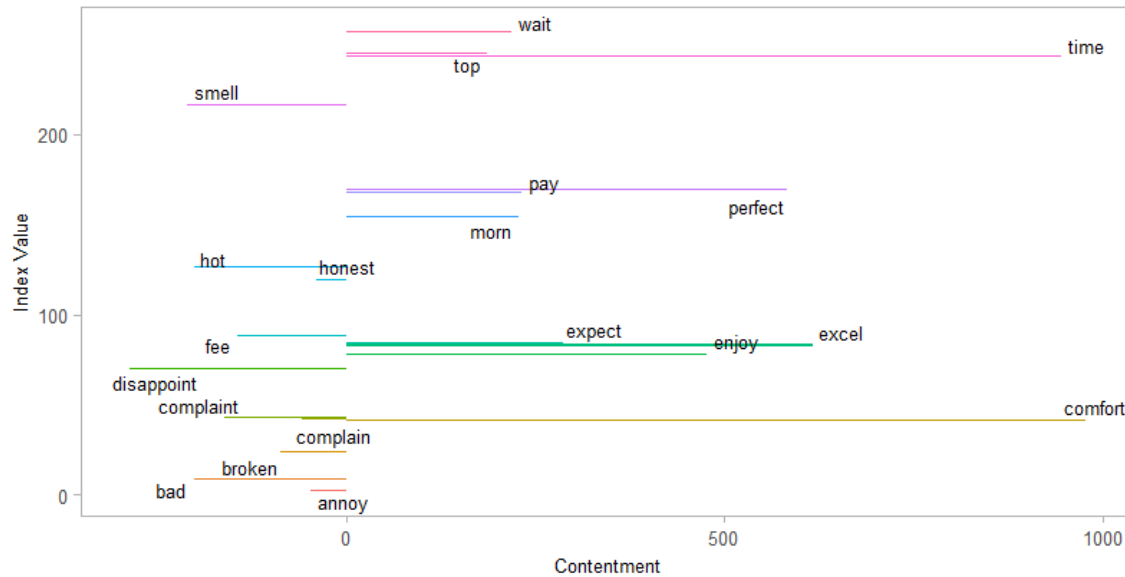
TEXT MINING AND SENTIMENT ANALYSIS

- Random forest classifier
- Tf-idf matrix is used for predict the emotion
- Final model has 91% overall accuracy
- 0.95 F1 score on Joy emotion
- 0.67 F1 score on Sadness emotion



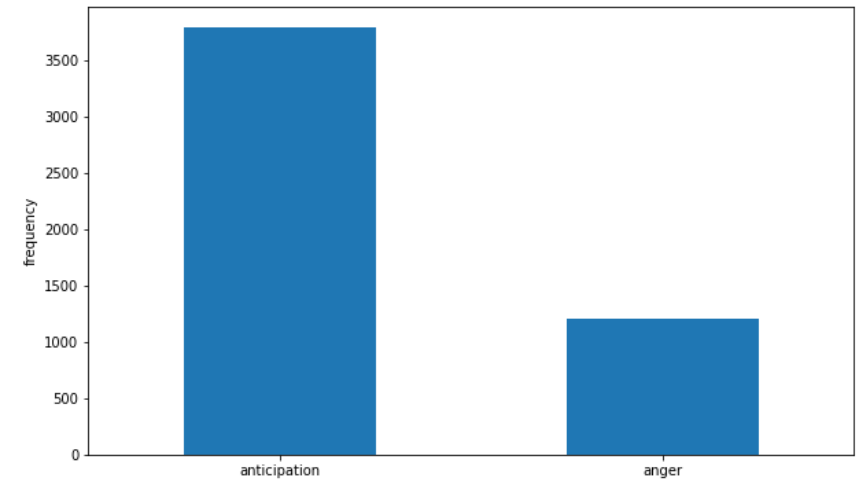
TEXT MINING AND SENTIMENT ANALYSIS

Anticipation/Anger Contentment based on Words



- Anticipation: time, comfort, excel
- Anger: disappointed, complaint, smell

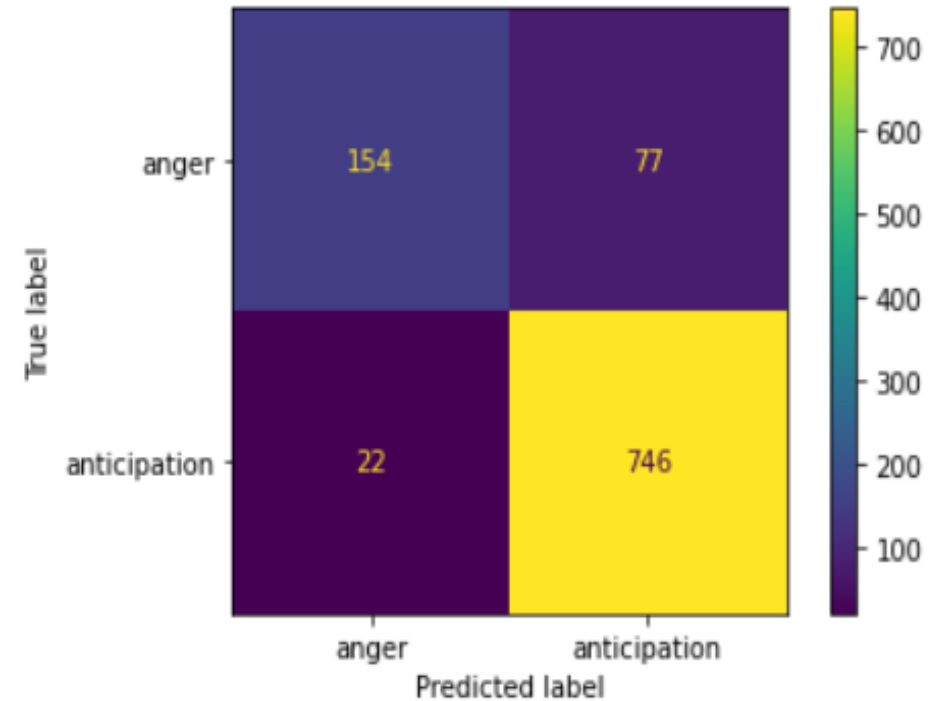
Anticipation and Anger Reviews



- 3,785 anticipation reviews
- 1,209 anger reviews

TEXT MINING AND SENTIMENT ANALYSIS

- Random forest classifier
- Tf-idf matrix is used for predict the emotion
- Final model has 90% overall accuracy
- 0.94 F1 score on Joy emotion
- 0.76 F1 score on Sadness emotion



RESULTS AND ASSESSMENT

Below are some of the most influential factors that affect how a guest feel joyful and anticipated about their hotel experience:



Cleanliness



Positive Staff Attitudes



Comfortable Environment



Time

Below are some of the common factors that make people feel sad and angry.



Complain About the Hotel



Guest Feel Disappointed

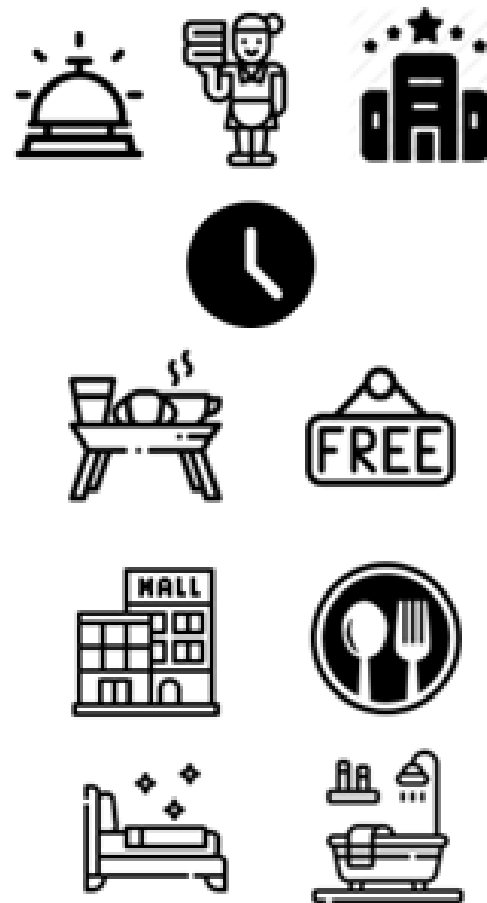


Trash, Broken Facilities
and Bad Environment

TOPIC DETECTION

LDA and NFM

- Service, Staff, and View
- Time Based Services
- Breakfast and Free Services
- Accessibility and Location
- Facilities



TOPIC DETECTION

Negative Reviews

LDA and NFM

- Noisy Atmosphere and Small Rooms
- Parking Not Included
- Staff Delays









ORIGINAL VISION VS FINAL PRODUCT

Original Idea

- Collect data - Airbnb
- Analyze trends - COVID-19
- Customers' opinions



Final Product

- Collected data - hotels.com 
- Analyzed 100 recent reviews (for each hotel) 
- Key considerations for guests 
- Price modeling 
- Guests' feelings 
- Defined standards during the COVID-19 

CONCLUSION

Recommendations

- Cleaning Process



- Staff Training



- Discount



Lessons Learned

- Learned how to collect data through web scraping to solve the project objectives
- Learned how to apply various analytical techniques including: descriptive statistical exploration, visualization development, text mining, and sentiment analysis to uncover insights within the data and tell a story

The background is a vibrant blue watercolor wash. It features a range of blue tones, from deep, dark indigo on the left to lighter, almost white-blue on the right. The texture is organic and fluid, with soft, irregular edges where the colors blend into each other. The overall effect is artistic and calming.

THANK YOU
