Hotel Review from Trip Advisor

Group 6's Team Members:

Agnes Calista - 2501980690

Matthew Aaron Sugiyarto - 2501983742

Sharon Zefanya Setiawan - 2501961022

Introduction

This research focus on hotel reviews is crucial due to society's growing reliance on the internet for travel information. Hotel reviews, now a primary resource for travelers, offer insights into accommodation quality. Despite the convenience of access, challenges arise in understanding and interpreting reviews, impacting users' ability to discern reliability. Recognizing and addressing these challenges is essential for improving travel decision-making based on hotel reviews and optimizing hospitality industry services.

About Dataset

This dataset has been sourced from Kaggle:

Tripadvisor Hotel Reviews

Containing 20,491 reviews and ratings, the dataset presents a comprehensive view of hotel reviews on Trip Advisor. Its diverse features offer valuable insights into customer feedback and experiences. This extensive dataset supports in-depth analysis, allowing exploration not only of factors impacting positive, neutral, and negative reviews and also facilitating topic prediction.

Objectives

01

Constructing a predictive model capable of accurately projecting review ratings and providing predictive insights regarding the quality of hotel services and facilities.

02

To discern specific patterns within negative reviews, facilitating targeted corrective actions for continuous improvement in hotel services.

03

Utilizing insights gained from the model and enhance service focus and quality by leveraging the derived information to make informed decisions and strategic improvements.

04

To assess the extent to which LDA topic modeling proves effective in identifying and comprehending the primary topics within hotel reviews on TripAdvisor.

Methodology

01

Exploratory Data Analysis

Project started by importing dataset and libraries into Python using Google Colab. Exploratory Data Analysis includes checking for missing values, duplicates, counting tokens, identifying anomalies, and calculating the average word count based on ratings.

02

Text Preprocessing

Text preprocessing readies raw data for analysis by converting to lowercase, tokenizing, removing punctuation and stopwords, and applying stemming or lemmatization for improved quality.

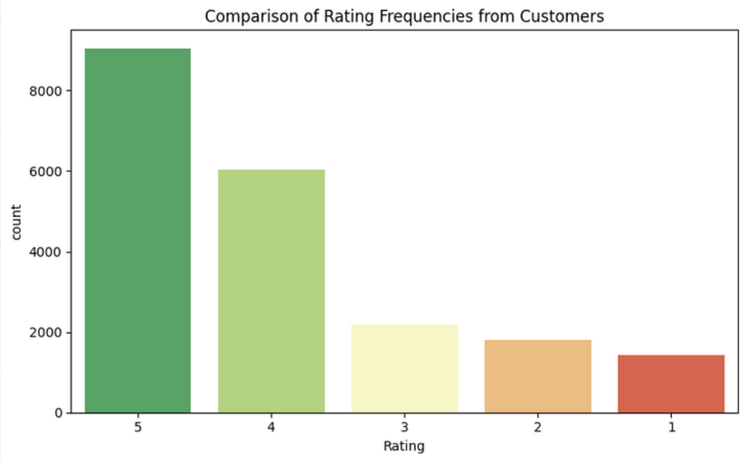
03

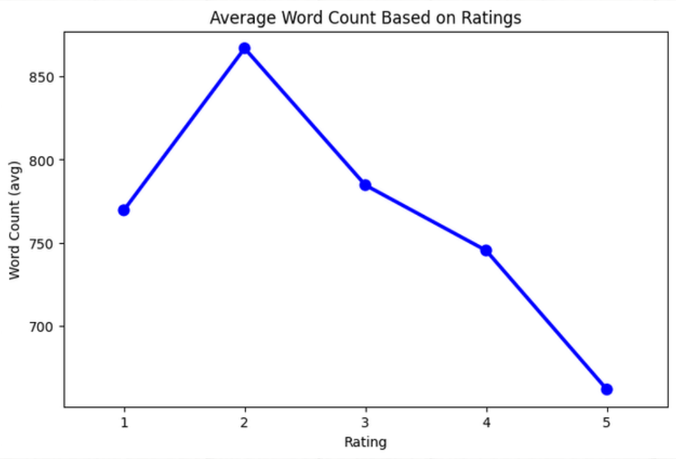
Modeling & Evaluation

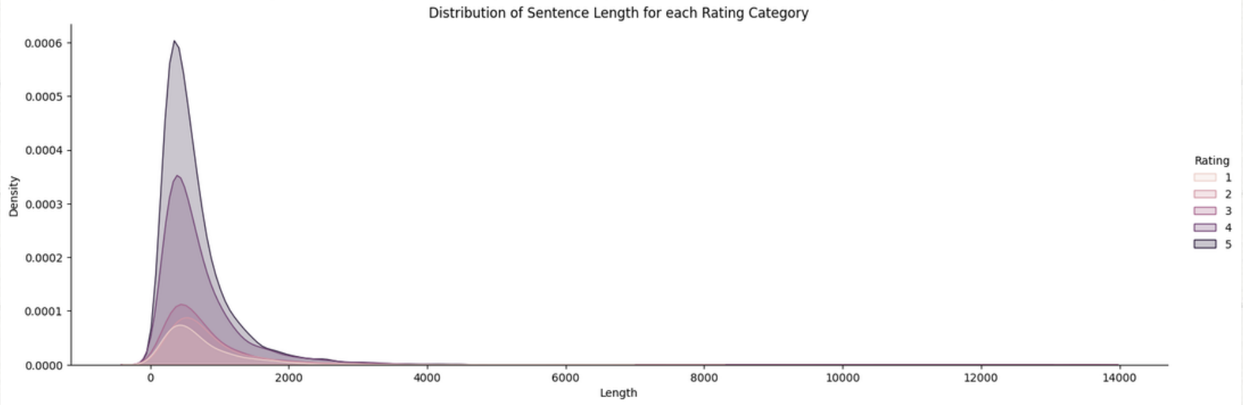
Utilized LDA for modeling with Coherence Model evaluation. Two trials: 5 and 3 topics. Initial performance unsatisfactory, so parameter tuning via grid search conducted (2-11 topics). Best model with 8 topics successfully enhanced the quality of topic analysis.

Exploratory Data Analysis

1. **Load Libraries and Read Files**
2. **Check Missing Value**
3. **Check Duplicated Data**
4. **Comparison of Rating Frequencies from Customers**
5. **Count Token**
6. **Check Anomalies**
7. **Average Length**
8. **Check Distribution**

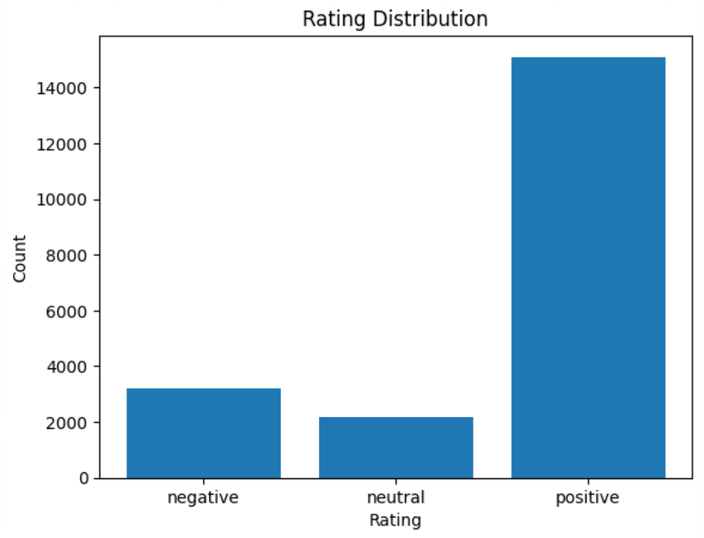






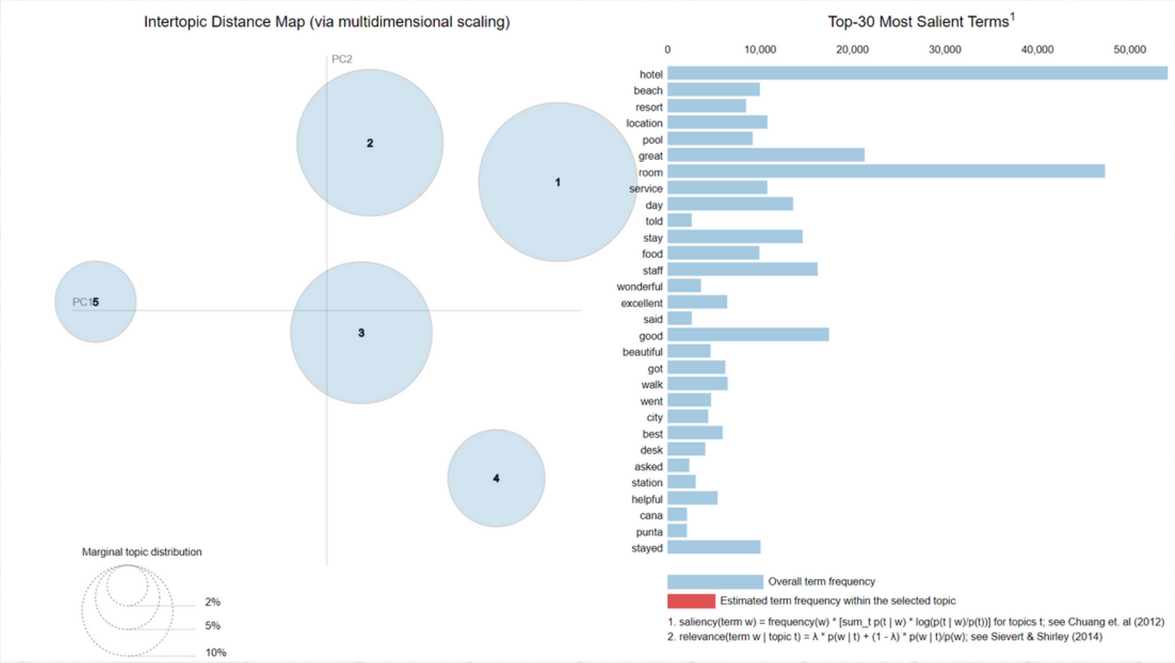
Text Preprocessing

1. **Import Libraries**
2. **Cleaning text**
3. **Labelling Ratings**
4. **Rating Distribution**
5. **Wordcloud**
6. **Tokenization**
   - **using BERT Tokenizer**



Modeling 1

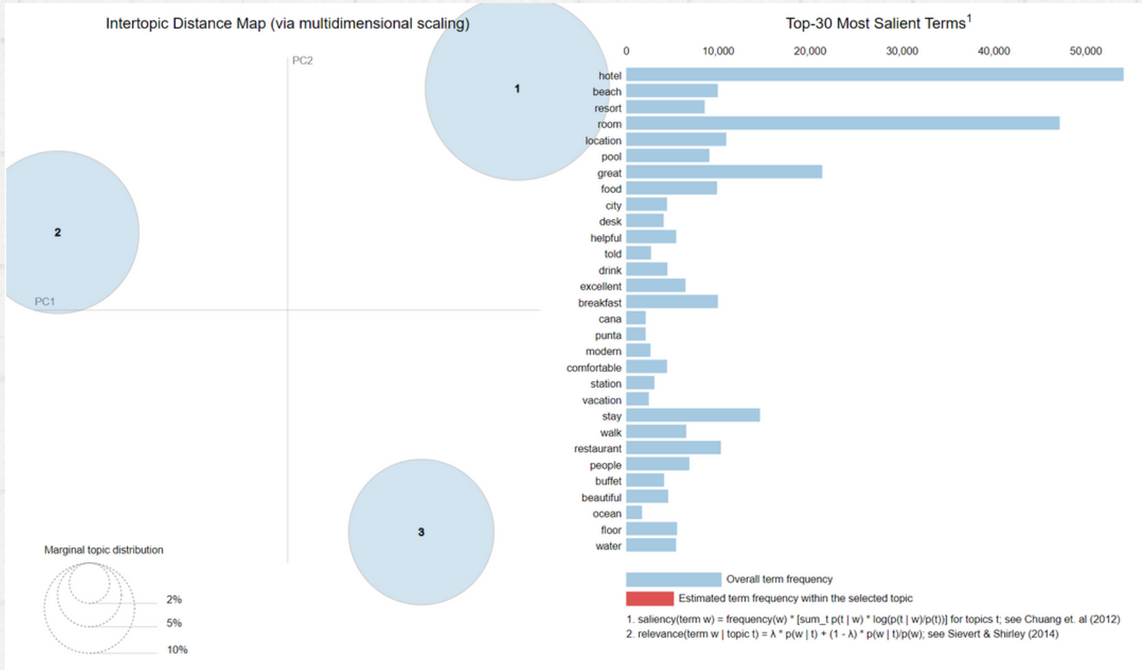
Coherence LDA : 0.36315371515985817 with five Topics.



	Sentence	Topic1	Score1	Topic2	Score2
0	nice hotel expensive parking got good deal sta...	2	0.824494	1	0.153238
1	nothing special charge diamond member hilton d...	2	0.800261	4	0.113130
2	nice room experience hotel monaco seattle good...	2	0.871318	0	0.083890
3	unique great stay wonderful time hotel monaco ...	2	0.487465	1	0.474819
4	great stay great stay went seahawk game awesom...	2	0.356827	4	0.248631
5	love monaco staff husband stayed hotel crazy w...	2	0.375517	0	0.298446
6	cozy stay rainy city husband spent night monac...	1	0.711367	2	0.282257
7	excellent staff housekeeping quality hotel cho...	2	0.396998	1	0.325445
8	hotel stayed hotel monaco cruise room generous...	2	0.657541	1	0.247187
9	excellent stayed hotel monaco past delight rec...	1	0.737545	0	0.243809

Modeling 2

Coherence LDA : 0.3715842573258024 with three Topics.



	Sentence	Topic1	Score1	Topic2	Score2
0	nice hotel expensive parking got good deal sta...	2	0.583649	1	0.411772
1	nothing special charge diamond member hilton d...	2	0.831877	1	0.166372
2	nice room experience hotel monaco seattle good...	2	0.765523	1	0.227847
3	unique great stay wonderful time hotel monaco ...	1	0.687185	2	0.274859
4	great stay great stay went seahawk game awesom...	2	0.537129	1	0.280796
5	love monaco staff husband stayed hotel crazy w...	1	0.466896	2	0.342283
6	cozy stay rainy city husband spent night monac...	1	0.877579	2	0.118522
7	excellent staff housekeeping quality hotel cho...	1	0.497220	2	0.399133
8	hotel stayed hotel monaco cruise room generous...	1	0.615424	2	0.377961
9	excellent stayed hotel monaco past delight rec...	1	0.974169	2	0.014124

Grid Search Tuning

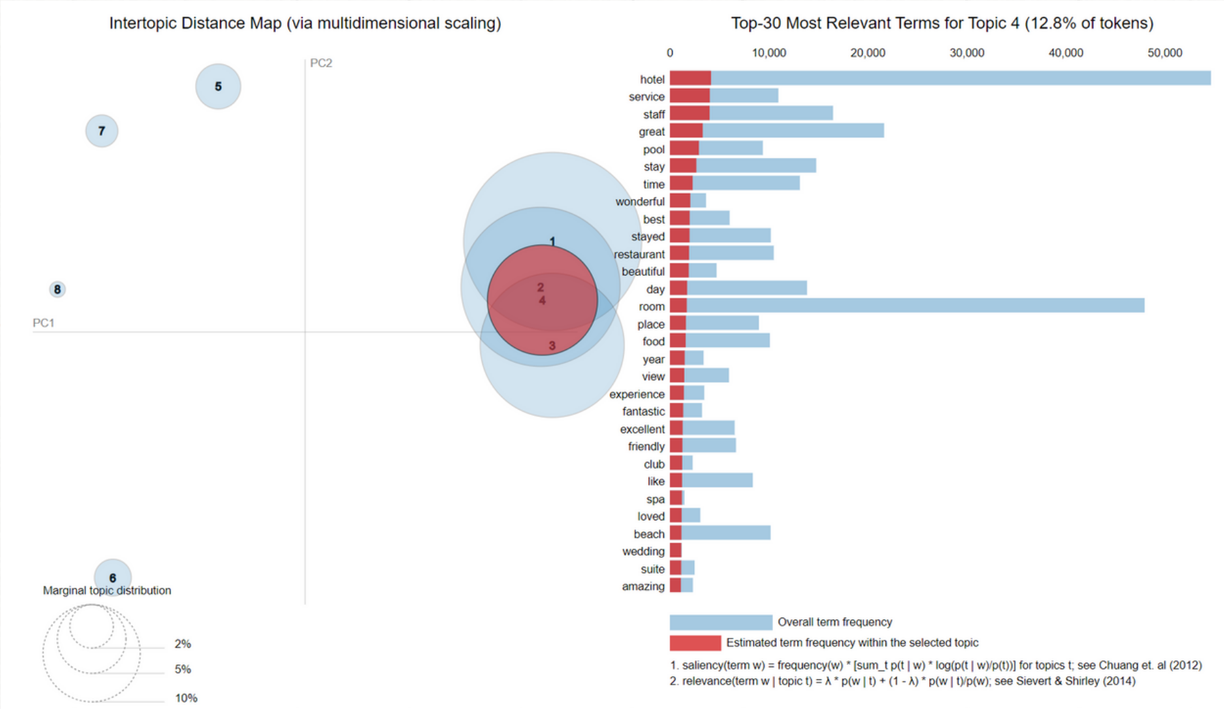
Coherence scores: [0.3455317747523996, 0.3715842573258024, 0.3589270991970778, 0.3631537151598581 7, 0.4545148076729585 4, 0.4800860552988225 3, 0.4902579404352326 3, 0.4805385193538287 4, 0.4883619597790662 7]

Best Coherence scores: 0.4902579404352326 3

Best number of topics: 8

Best Model

Coherence LDA : 0.4902579404352326 with eight Topics.



	Sentence	Topic1	Score1	Topic2	Score2
0	nice hotel expensive parking got good deal sta...	2	0.542683	1	0.380605
1	nothing special charge diamond member hilton d...	2	0.758366	1	0.080195
2	nice room experience hotel monaco seattle good...	2	0.739135	1	0.153156
3	unique great stay wonderful time hotel monaco ...	1	0.579944	2	0.249564
4	great stay great stay went seahawk game awesom...	2	0.497449	1	0.238801
...	...	...	...	...	...
95	excellent stay staff friendly helpful nice hot...	1	0.922277	6	0.050914
96	return going seattle booked hotel knowing budg...	1	0.827227	6	0.122101
97	terrible hotel approximately week ago april re...	2	0.933117	5	0.037636
98	great price okay experience stayed inn queen a...	2	0.495088	1	0.370917
99	old dumpy place problem head start spinning th...	2	0.734411	3	0.126431

100 rows × 5 columns

Topic Name

TOPIC NAME						
	Person 1	Person 2	Person 3		RESULT	
Topic 0	Luxurious	Hospitality	Experience		Topic 0	Facilities
Topic 1	Highlights	Facilities	Hospitalities		Topic 1	Hospitality
Topic 2	Room Service	Room Service	Room service		Topic 2	Room Service
Topic 3	Resort	Resort	Resort		Topic 3	Resort
Topic 4	Hotel Restaurant	Hotel Restaurant	Hotel Restaurant		Topic 4	Hotel Restaurant
Topic 5	Strategic Location	City Exploration	Strategic location		Topic 5	Strategic Location
Topic 6	Parking Area	Parking and Accessibility	Parking Area		Topic 6	Parking Area
Topic 7	Adventures	Unique Experiences	Unique Experiences		Topic 7	Unique Experiences

Result

1. The comparison between human judgement results and topic predictions reveals significant differences.
2. Sentiment analysis based on rating labels shows similarities with sentiment analysis derived from human judgement on review data.
3. Possible bias or the lack of human input in the dataset may impact human judgement results, while model inaccuracy could be a reason for the lack of precision in understanding topic contexts.
4. Further analysis can be conducted to refine and comprehend the factors influencing both the model's performance and human judgement results.

Thank you!