

MGMT 590 - Web Data Analytics - Group 6



Team



**Sanjana
Santhanakrishnan**



**Sudipta
Singh**



**Surya
Sathiyamurthy**



**Aditya
Gupta**

Agenda



Introduction and Problem Statement



Executive Summary



Case Analysis



Recommendations and Limitations

Current Problems faced by Hotels



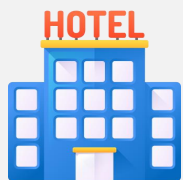
- ☆ **Impact of COVID on tourism, especially hotels** – The impact of COVID on the travel industry has been significant, with many hotels and restaurants closing down in these tough times without proper business
- ☆ **Misleading Information** – People and Hotels both get lost looking for their perfect hotels and guests respectively; TripAdvisor bridges this gap to an extent, so it is important for hotels to rank better in their website in order to get more visibility and in turn more business

Business Problem



- ☆ **Need for Information** - A new hotel coming up needs information on which location to choose, what kind of amenities to have and how to leverage these for a better performance
- ☆ **Customer Perception** – New or underperforming hotels need data on which kind of customers to cater to and expand their business into. They also require information on which customers are least satisfied with their current hotel choices and are looking for new options.

Hotel Profile



Hotels in California



4.35

Average Rating of hotels



550

Average # of Reviews

Hotels with the following Features

Hotels having COVID safety section



36%

Hotels having their own website



32%

Hotels running offers



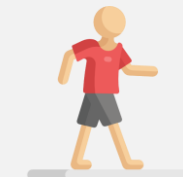
14%

% of Hotels



\$210.75

Average price of Hotels



73

Average Walking Score^



262

Average # of pictures for each hotel

^ Walk Score is dependent on the # of nearby places like restaurants, attractions that are surrounding the hotel

Review Profile



10

Reviews for top 10 Hotels



4.48

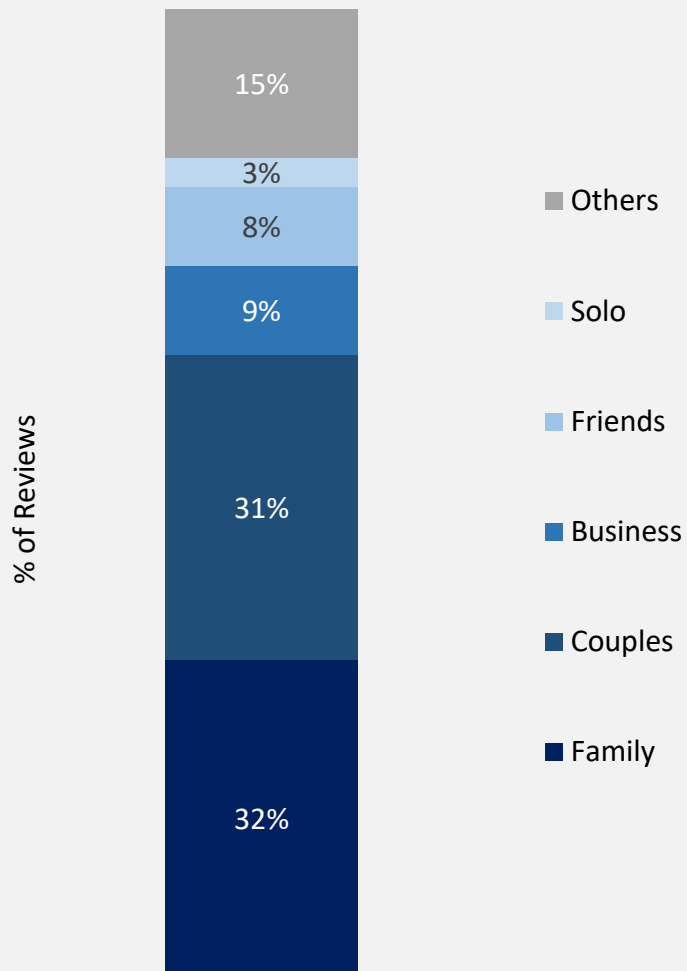
Average Rating of Reviewers



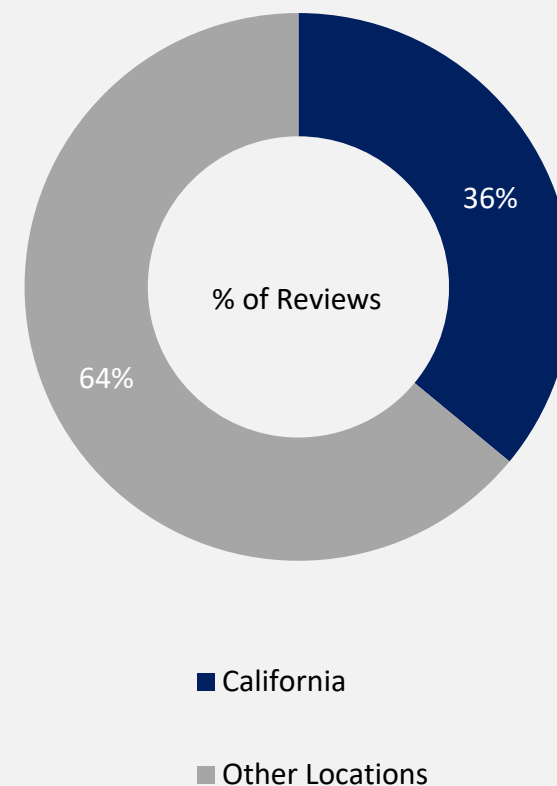
24905

Total # of Reviews

Trip Type of Reviewers



Reviews from California



Agenda



Introduction and Problem Statement



Executive Summary



Case Analysis



Recommendations and Limitations

Executive Summary



- ☆ **Overview** – Help new hotels and underperforming hotels to improve their ranking
- ☆ **Problem** – Impact of COVID, lack of proper information and improvement of customer perception
- ☆ **Recommendations** –
 - ☆ **Information** - Having COVID sections and proper descriptions as well as having non-smoking rooms in hotels
 - ☆ **Location** of a hotel near restaurants rather than tourist attractions
 - ☆ **People** - Focusing on travelers from other location, especially business travelers

Agenda



Introduction and Problem Statement



Executive Summary

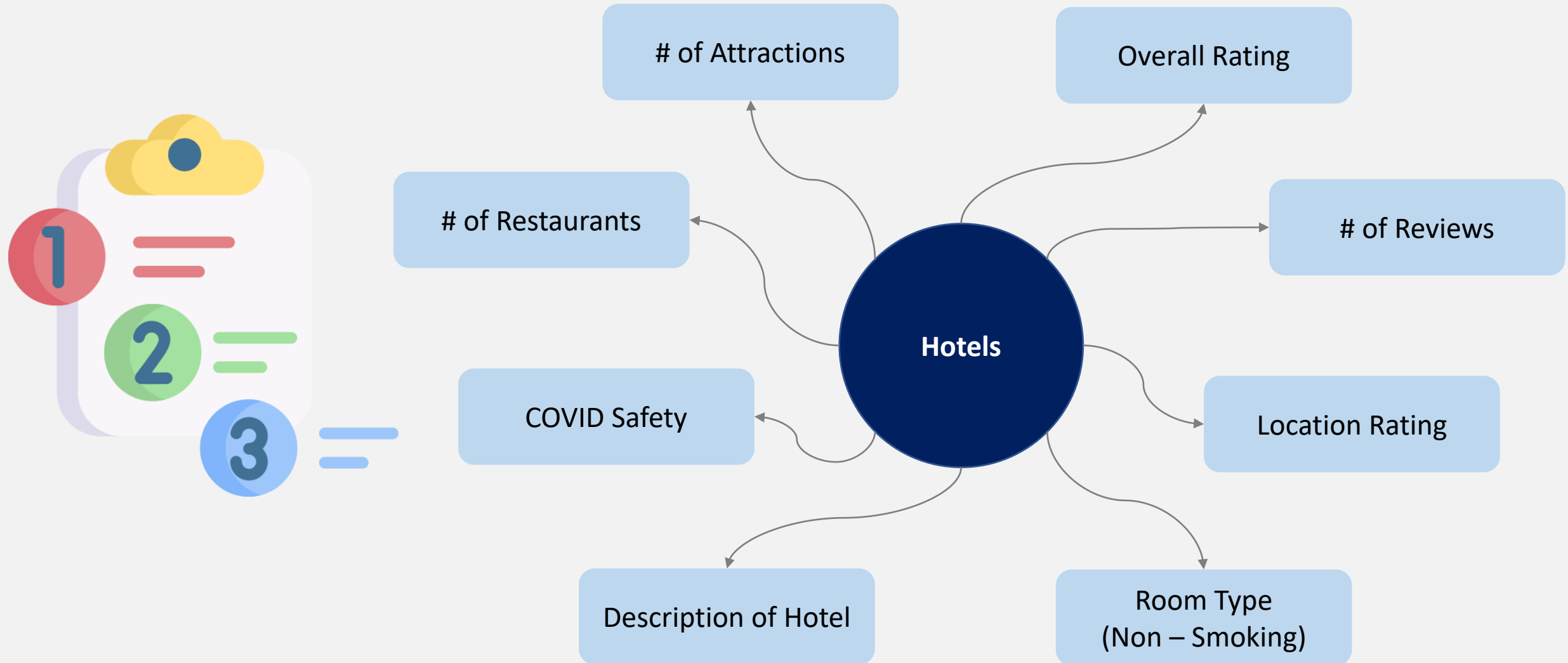


Case Analysis

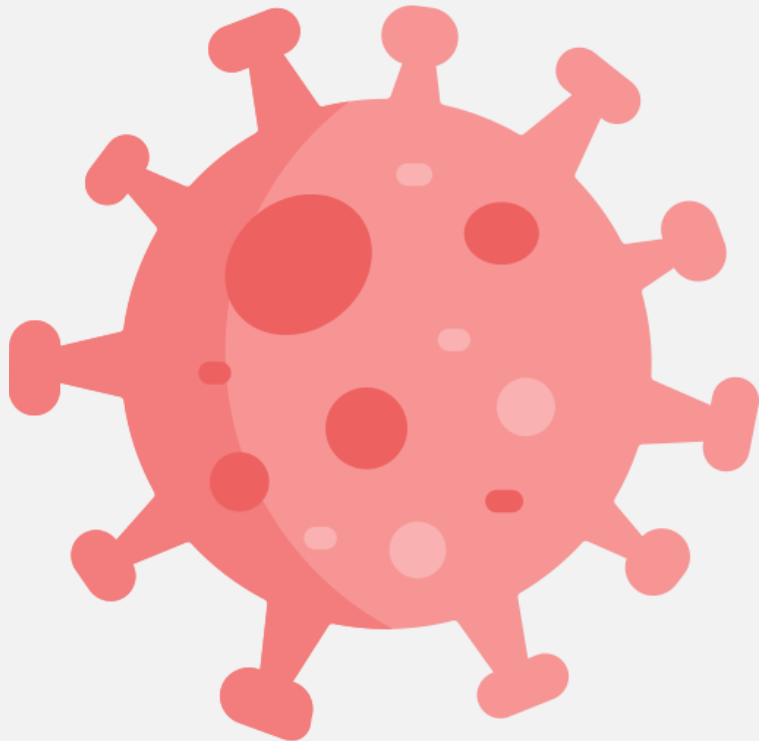


Recommendations and Limitations

Factors Affecting Ranking



Effect of COVID Section on Hotel Rankings



Hotels with
COVID section



838



Hotels without
COVID section



1952

Average Rank of Hotels

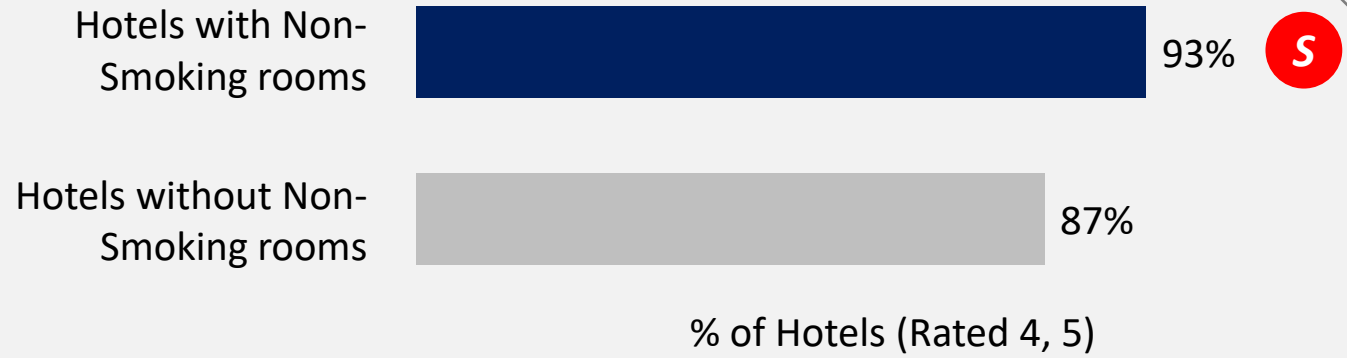


Statistically significant difference (with 95% CL)

Test performed: T- Test

- ✧ Hotels that have COVID safety section have a significantly higher ranking than hotels that do NOT have COVID section
- ✧ It is recommended for the new or existing hotels to have the COVID safety measures implemented and have it listed on their Tripadvisor page

Effect of Non-Smoking Room on Hotel Rating



Statistically significant difference (with 95% CL)
Test performed: Z- Test

- ☆ With the ongoing COVID-19 pandemic, more and more people are preferring hotels with Non-Smoking rooms and this translates into a significantly higher rating of hotels that have non-smoking rooms compared to others
- ☆ It is recommended for the hotels to have or convert some of their rooms to Non-Smoking rooms

Effect of Description on Hotel Rankings



Average Rank of Hotels



Statistically significant difference (with 95% CL)

Test performed: T- Test

- ☆ Hotels that have a brief description of them with more than 100 words have a significantly higher ranking compared to hotels with less than 100 words of description
- ☆ In this world of information, it is advised to provide a more detailed description of hotel for the guests to make an informed decision on their stay

Effect of # of Attractions on Hotel Rankings



Hotels with >11
Attractions near them



1260



Hotels with ≤11
Attractions near them



1303

Average Rank of Hotels



NO Statistically significant difference (with 95% CL)

Test performed: T- Test

- ☆ # of Attractions surrounding a hotel does NOT have a significant impact on the hotel rankings
- ☆ In the future, new upcoming hotels need not spend a fortune paying for a premiere locations that are near multiple attractions

Effect of Restaurants on Hotel Rankings



Average Rank of Hotels



Statistically significant difference (with 95% CL)
Test performed: T- Test

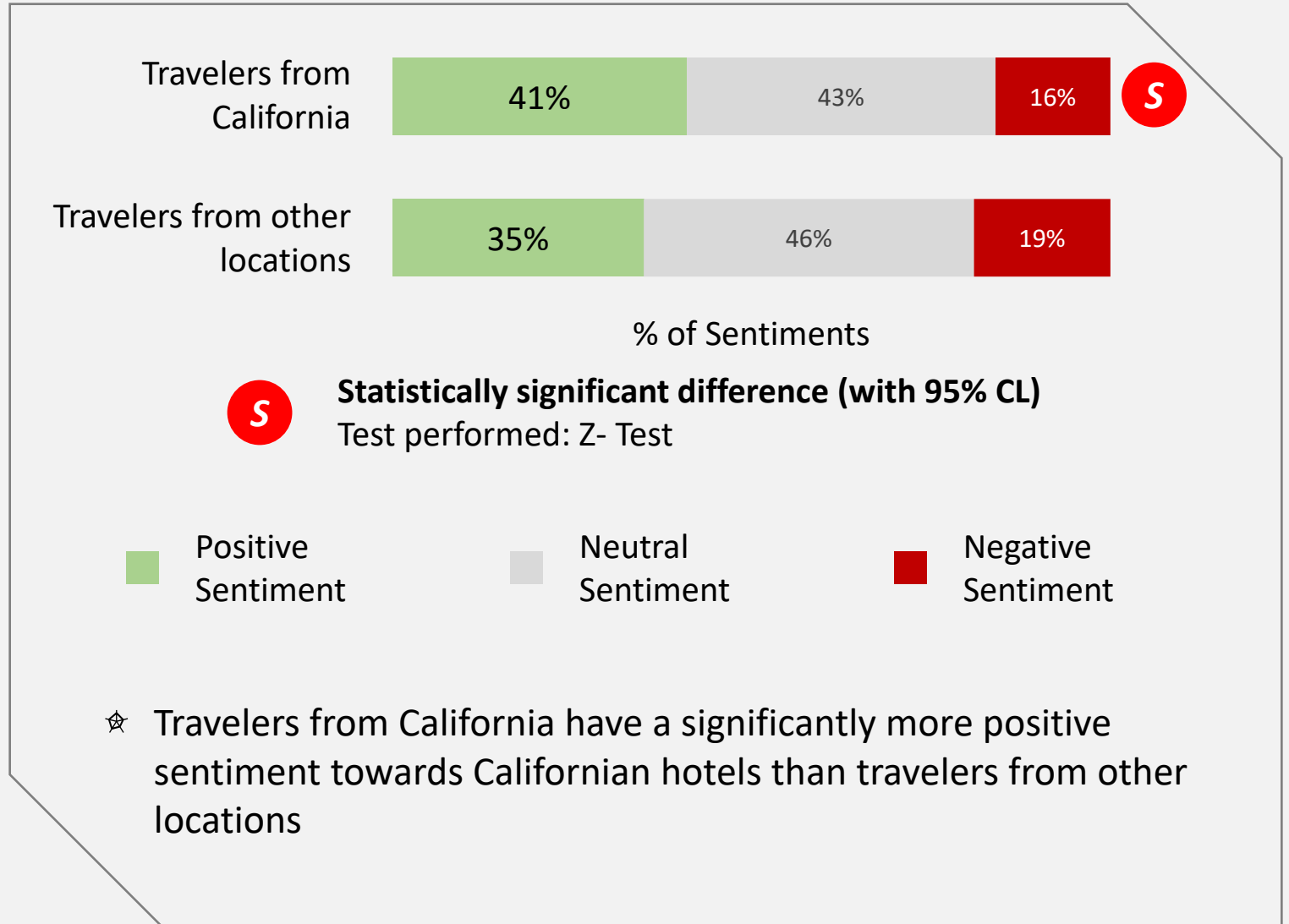
- ☆ Hotels that are in the proximity of >40 restaurants have a significantly better rank than other hotels that are in the proximity of less restaurants

Sentiment of Hotel Reviews

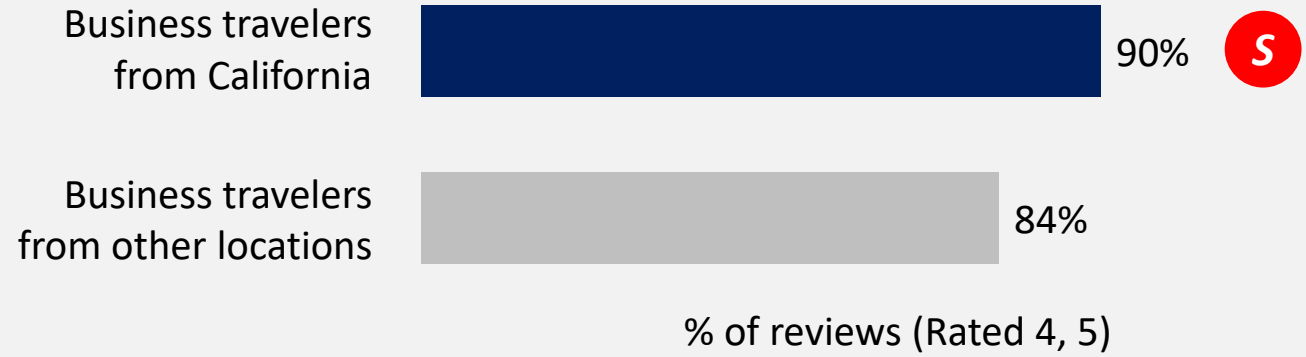


- ☆ Collected reviews and individual rating data of top 10 hotels in California
- ☆ Calculated review sentiment polarity
- ☆ Used Textblob for the sentiment analysis
- ☆ Classified reviews as Positive, Negative and Neutral
- ☆ Analyzed sentiment variation across locations and trip types

Sentiment of Californians on Hotel Reviews



Hotel Ratings from Business Travelers



Statistically significant difference (with 95% CL)
Test performed: Z- Test

- ☆ Even though tourism was affected, businesses were very rapid to adopt to this new normal; Business travelers from CA rated the hotels significantly higher compared to business travelers from other locations
- ☆ This is in line with the sentiment from the previous analysis; focusing on business travelers from other locations can help mitigate this negative sentiment

Agenda



Introduction and Problem Statement



Executive Summary



Case Analysis



Recommendations and Limitations

Recommendations



☆ **Information**

- ☆ **Description** – With a clearer detailed description, the guests will be able to know more about the hotel and make a better choice
- ☆ **COVID** – Having a COVID safety measures implemented and listed makes travelers feel more safe and secure about their stay
 - ☆ As a result, having a non-smoking room types helps with the rating and ranking
- ☆ **Location** – Having the hotels in close proximity to restaurants rather than attractions have a significant effect on ranking and rating
- ☆ **Improvement in the Perception of Travelers from other locations**
 - ☆ Californians vs Travelers from other locations – Focusing more on travelers from other locations helps improve the hotel performance and their sentiment towards these hotels
 - ☆ Especially among business travelers from other location

Limitations of Analysis and Approach



- ☆ **Lack of Proper Metrics** - Hotel rank on TripAdvisor is used as a proxy for the hotel's performance
- ☆ **Missing Data Points—**
 - ☆ Pricing data was not available for some hotels
 - ☆ Location and Trip Type was not mentioned in some reviews
- ☆ **Restriction of Analysis** - Review Analysis is done on reviews of the top 10 hotels only

THANK YOU

APPENDIX

Dep. Variable:	Rank	R-squared (uncentered):	0.811
Model:	OLS	Adj. R-squared (uncentered):	0.808
Method:	Least Squares	F-statistic:	238.6
Date:	Mon, 06 Dec 2021	Prob (F-statistic):	0.00
Time:	17:59:25	Log-Likelihood:	-7394.1
No. Observations:	1016	AIC:	1.482e+04
Df Residuals:	998	BIC:	1.491e+04
Df Model:	18		
Covariance Type:	nonrobust		

Linear Regression Results

	coef	std err	t	P> t	[0.025	0.975]
Rating	-571.8352	36.566	-15.638	0.000	-643.591	-500.079
Reviews	-0.0577	0.010	-5.582	0.000	-0.078	-0.037
Location	210.1995	39.084	5.378	0.000	133.504	286.895
DescLength	0.3994	0.158	2.522	0.012	0.089	0.710
Walkscore	-0.4462	0.591	-0.755	0.450	-1.605	0.713
Attractions	-0.1171	0.249	-0.471	0.638	-0.605	0.371
NonSmokingRoom	1399.7060	147.104	9.515	0.000	1111.037	1688.375
Class_1.5	936.4438	242.792	3.857	0.000	460.002	1412.885
Class_2.0	1042.7917	137.712	7.572	0.000	772.554	1313.029
Class_2.5	982.1879	146.750	6.693	0.000	694.215	1270.161
Class_3.0	935.6351	140.096	6.679	0.000	660.718	1210.552
Class_3.5	983.4215	145.701	6.750	0.000	697.507	1269.336
Class_4.0	956.9870	141.687	6.754	0.000	678.948	1235.026
Class_4.5	1044.3254	203.999	5.119	0.000	644.009	1444.642
Class_5.0	872.2930	156.851	5.561	0.000	564.498	1180.089
Class_NA	1.32e-12	2.55e-13	5.184	0.000	8.2e-13	1.82e-12
Website_1	-26.4799	29.055	-0.911	0.362	-83.496	30.536
Covid_1	-142.0301	25.596	-5.549	0.000	-192.258	-91.802
Offers_1	6.8048	30.616	0.222	0.824	-53.274	66.884

Topic: 0

Words: 0.014*"hotel" + 0.011*"beach" + 0.009*"diego" + 0.009*"san" + 0.008*"resort" + 0.008*"santa" + 0.007*"room" + 0.006*"hollywood" + 0.006*"spa" + 0.006*"california"

Topic: 1

Words: 0.020*"hotel" + 0.010*"room" + 0.010*"guest" + 0.009*"offer" + 0.008*"center" + 0.008*"anaheim" + 0.007*"hollywood" + 0.007*"resort" + 0.006*"pool" + 0.006*"view"

Topic: 2

Words: 0.023*"hotel" + 0.013*"room" + 0.011*"san" + 0.009*"guest" + 0.009*"california" + 0.006*"francisco" + 0.006*"service" + 0.006*"center" + 0.006*"experience" + 0.005*"beach"

Topic: 3

Words: 0.026*"hotel" + 0.013*"san" + 0.010*"room" + 0.008*"guest" + 0.008*"beach" + 0.007*"located" + 0.006*"center" + 0.006*"francisco" + 0.005*"diego" + 0.005*"resort"

Topic Modeling Results