What is the pandemic traveller thinking? Insights from online hotel reviews in India

Abstract:

The pandemic has profoundly impacted the mind-set of travellers' about hotel bookings and travel in general. The main goal of this paper is to study this impact using online hotel reviews and bring out its implications for stakeholders in the hospitality space. The amount of user generated content based research done in the intersection of COVID and tourism has been rudimentary, given the nascent nature of the topic itself and is mostly limited to baseline research of the review text alone. This study uses data from March 2020 to April 2021 collected from TripAdvisor pages of hotels across India. A keyword filtration technique is employed to shortlist pandemic specific reviews. The underlying dimensions of the topics of the reviews are unearthed using topic modelling. Data other than reviews like the precautions taken by hotels, type of traveller and the price are also analysed to present a holistic image. The study finds that hotels that take fewer precautions are more prone to having more complaints and concerns raised in online reviews. A temporal analysis of topics shows that the top topic of discussion, when a pandemic wave subsides and free movement begins gradually, is the safety precautions taken by the hotel. We list managerial implications to help hotels maintain a better online profile and cater to the varied needs of each type of traveller and theoretical implications to analyse the new nature of data available to researchers post the pandemic.

Keywords: COVID-19, Hospitality Management, TripAdvisor, Data Mining, User Generated Content, Topic Modelling, LDA

1. Introduction

India has always been a tourism hotspot owing to its rich cultural and historical heritage, variety in ecology, and places of natural beauty. The Indian tourism and hospitality industry has emerged as one of the critical drivers of growth among the service sector in India. Tourism is also a large employment generator besides being a significant source of foreign exchange for the country. It makes up roughly one-tenth of the GDP and generates handsome Foreign Exchange Earnings (Roy et. al., 2020). India’s rising middle class have supported the growth of domestic tourism as well. International hotel chains are increasing their presence in the country and are expected to account for 50% of the tourism and hospitality sector by 2022.

India is one of the most digitally advanced traveller nations regarding digital tools used for planning, booking, and experiencing a journey. According to the Indian Ministry of

Tourism, the country has seen steady growth in the number of e-Tourist Visas being approved, with 171 countries having access to the same. In 2019, 2.9 million foreigners availed the e-Tourist Visa (23.6% year-on-year growth). With Foreign Tourist Arrivals standing at close to 11 million (more than double what it was a decade ago) and international tourist arrivals at approximately 18 million in 2019, the sector was all set to accelerate growth further in 2020 (Ministry of Tourism, 2019).

Kumar (2020) has thrown light on the hardships faced by the Indian tourism and hospitality industry due to the pandemic. A nationwide lockdown was imposed for 21 days starting from the 24th of March 2020. Since then, the country has seen a series of lockdowns (some region-specific) and subsequent unlocks carried out. The sector, still being in its developing stage, needs a swift response to overcome the losses caused by the pandemic.

Starting with the Unlock 1.0, hotels were allowed to open up from the 8th of June 2020. Later in the year, the country had an almost complete return to normalcy, with the last of the unlock (5.0 and 6.0) coming to effect by November (Government Guidelines Unlock 1.0, 2020). However, this does not mean that one can neglect the effect of the pandemic any time soon. In a global survey conducted by UNWTO among its Panel of Tourism Experts on the impact of COVID-19 on tourism and the expected time of recovery, as many as 56% of experts said that international tourism would not return to pre-pandemic levels in the Asia-Pacific region by 2023. (UNWTO, 2021). In the same survey, it was established that almost all the recovery seems to be coming from domestic tourism, especially in the Asia-Pacific region (hence collecting reviews that primarily contain domestic travellers' views is reasonable). India also ranks highly in the "Domestic Visitor Spending" category in 2020 (WTTC, 2020). The Government of India, too, has come up with initiatives to boost the recovery of the sector. On the 28th of June 2021, adding to the existing policies, handsome loans were approved to travel agencies and tourist guides, and a free 1-month tourist visa was announced for the first 500,000 international tourists who visit the country once borders open up. With all these developments, it is of utmost importance to thoroughly study the current situation from the travellers' perspective and make sure that their requirements are satisfied to avoid any more unwarranted outbreaks of the virus and rejuvenate the sector.

In recent years, the analysis of User Generated Content (UGC) has become widespread. In the context of tourism, TripAdvisor is one of the most credible platforms to turn to when the goal is to extract insights from UGC, which in this case, is online customer reviews Ayeh et.al. (2013). Research and analysis of online reviews have numerous advantages, including but not limited to data availability, swift data collection and non-intrusiveness with human subjects Lu et. al. (2015). Adding to this, the analysis of online hotel reviews also helps bring out travellers' needs and preferences Guo et. al. (2017). Scrutinising these analyses can help managers formulate strategies to improve their overall service Lu et.al. (2015). Taecharungroj et. al. (2019) have explained the enormous importance of online reviews in the context of tourism research and how TripAdvisor has grown over the years to be arguably the best online review platform in terms of reliability which goes hand in hand with users considering the platform to be unbiased and trustworthy. The platform has also strictly policed the system to avoid fake reviews despite past controversies (Filieri et. al., 2015; Schuckert et. al., 2015).

As data analysis, in general, becomes more prevalent across all areas of study, recent research often uses a combination of web crawling, data mining, machine learning, and other statistical methods to collect, analyse, and interpret data Xiang et.al. (2017).

Henceforth, the goal of this study will be to unearth insights from the data at hand and bring out underlying trends and patterns. For textual data (reviews), topic modelling (uses advanced software and mathematical techniques developed in the fields of natural language processing) is an effective way to achieve the same. One of the most widely used topic modelling tools is the Latent Dirichlet Allocation, more popularly known as LDA (Blei et. al., 2003). It uses an unsupervised learning algorithm based on the principle of Gibbs sampling and aims to cluster text data into "topics". An analysis of the spread of these topics with respect to various parameters can prove to be insightful.

The remainder of the paper is structured as follows. Section 2 contains an overview of previous research on online reviews, web-scraping and LDA. Section 3 starts with an introduction to the methodology of the conducted research, followed by data collection, cleaning and pre-processing, splitting the datasets, LDA and Exploratory Data Analysis (EDA). We present the analysis and results of our study in Section 4, and discuss the findings and insights for key stakeholders in section 5. The practical and research implications are available in Section 6, followed by the conclusions, limitations, and scope for future work in Section 7.

2. Literature review

2.1 The Analysis of User Generated Content

Research on online reviews started with manual approaches with datasets containing not more than 1000 reviews (Au et al., 2014; Manickas et al., 1997; Levy et al., 2013). Although manual approaches have their advantages, such as solid interpretation, the reliability of the results depends highly on the analysts' knowledge, and the task is laborious. This also makes the results hard to replicate (Hu et al., 2019). As the volume of data to be analysed started to increase, courtesy of developments in the automation of data collection methods, research on tourism and hospitality applications using computer-assisted analysis of UGC (Kirilenko et al., 2021) has become common.

2.2 Data collection methods

The first step in the analysis of online reviews is data collection. There are several approaches when it comes to the collection of publicly available online hotel data. As discussed earlier, doing this manually is not feasible as the current study demands data from multiple hotels spread all across the country. Hence, this calls for an automated data collection approach. One of the most accepted and efficient processes to automate data collection is web scraping (Broucke et. al., 2018). Alternative approaches such as the use of Application Programming Interfaces (APIs) and surveys have also been employed by researchers, but these studies are vulnerable to certain pitfalls that outweigh their pros. Survey-based studies are often on the receiving end of criticism of not fully capturing real consumer behaviour arising from biases of the participating subjects. On the other hand, APIs are often difficult to access, and even if access is obtained, they require an upfront fee, are accessible only for a limited time frame and do not generally expose all the required variables (Han et. al., 2021).

Web scraping for research purposes is also often outsourced to third-party firms (Wu et. al., 2016). This has its cons, as improper communication with the firm can lead to more money and effort being spent to complete the research (Han et. al., 2021). As we had the required expertise and user-friendly software at our disposal, we scraped all our data using a python script.

2.3 Latent Dirichlet Allocation and Polarity Analysis

Han et. al. (2021) have argued that in current literature, all studies that involve text analytics fall under two broad umbrellas: (1) Polarity analysis that deals with the analysis of consumers' sentiment extracted through their reviews on a particular attraction, hotel, or destination, and (2) Topic modelling which aims to extract the review's meaning (generally by assigning a topic label to the review). The textual analysis in this study mainly focuses on the LDA and subsequent topical analysis. Nevertheless, the polarities of the reviews have been explored in brief to obtain a rough estimate of their spread by topic.

The LDA has since found its way into countless studies involving text analysis. In the tourism context, it has been used to examine destination images (Luo et. al., 2020; Wang et.

al., 2020), uncover dimensions of customer satisfaction (Guo et, al., 2017; Kirilenko et. al., 2021), and also to get a feel of the corpus and subsequent topical analyses (Mankad et. al., 2016; Garcia et. al., 2020; Taecharungroj et. al., 2019).

2.4 Covid and Tourism

Some work done so far on analysing online customer reviews during the pandemic years deals with a global perspective (Yu et. al., 2020; Uğur et. al., 2020; Mehta et. al., 2021; Hu et. al., 2021) while there are also few studies focused on a particular country or region (Chatibura, 2020). Though covid based research is picking up pace, studies analysing UGC have been limited to exploring just the text and some supporting analytics. Moreover, there has been no study yet that has taken a pan-India approach to understanding how travellers perceive the situation. We fill these research gaps with a topical analysis on reviews about covid and analyse the interactions of parameters such as precautions (listed on TripAdvisor) taken by hotels, type of traveller(s) that stayed at the hotels, and the price bracket the hotels fall under.

3. Methodology

Our framework contains three stages: (i) data collection, (ii) data pre-processing, (iii) LDA and further analysis. In the first stage, we scrape all relevant data from the TripAdvisor platform. The second stage comprises text cleaning for LDA and appropriate pre-processing, feature engineering, and transformations for the general analytics. In the third stage, we apply LDA to the text data and perform exploratory analysis on the rest of the data to uncover insights and patterns present in the subsequent section.

3.1 Data collection

We have taken a top-down approach to data collection. TripAdvisor has a section where several famous tourist locations are listed, and by clicking on each of these location hyperlinks, we obtain a comprehensive list of hotels in those locations. From here, we proceed to click on the hyperlinks of each hotel (with more than 150 reviews) and extract the hotel's name, location, property amenities, room features, room types, listed covid precautions, reviews (only English reviews from January 2019 to April 2021), date of stay (if mentioned by the reviewer), trip type (if mentioned by reviewer), rating (1-5 bubbles), and price range (if listed on TripAdvisor). After removing duplicate reviews, we finally ended up with 521,402 reviews from 2418 hotels across.

We have an initial filter of 150 reviews as we were observing that hotels with fewer reviews were harder to extract data from and were also missing a hefty amount of data, thus contributing very little to the current study.

We performed all the data collection activities with selenium in python. It is a software package that relies on the HTML tags of various elements on a web page to locate, identify, and extract content from the specified tag. Some data may be bound to be missing as we collect data from both high end and low-end hotels all across the country.

3.2 Cleaning and pre-processing

The next step is to clean all reviews for further analysis using standard text pre-processing techniques. We expanded contractions in the text (for instance, "didn't" to "did not"), applied lemmatisation along with POS tagging, removed punctuations, stop words, extra spaces, converted all non-ASCII characters to ASCII characters using python's Unicode package, and finally, removed all words that do not exist in the English Dictionary. We performed these using the Natural Language Toolkit (NLTK) (https://www.nltk.org/).

After observing the spread of COVID-19, subsequent lockdowns and unlocks in the country, and some manual inspection, we trimmed down the data to range from March 2020 - April 2020 to reflect the pandemic situation. Following this, we extracted a filtered sample (Dataset 2) from the current data frame (Dataset 1), which contained reviews that had mentions of one or more of the keywords: "corona", "covid", "virus", "quarantine" and "pandemic" (Chatibura, 2020; Uğur et al., 2020).

The motivation behind the keyword filter also stems from the fact that TripAdvisor provides a "search reviews" option where users can search for some keywords that they want reviews to have. We then use Dataset 1 for LDA and supporting topical analysis and Dataset 2 for a general exploratory analysis of post-February 2020 data.

3.3 LDA and further analysis

Latent Dirichlet Allocation (LDA) was first introduced by Blei et al. (2003) as a generative probabilistic model of a corpus, in which each item of a collection (each review in this case) is modelled as a finite mixture over an underlying set of topics and a distribution over words characterises each topic. Each review is assigned a set of topic probabilities, and each topic is given as a mixture of the most dominant words in that topic. The labelling of these topics is a manual process.

With the cleaned reviews in Dataset 2, we proceeded with the LDA. We have implemented the LDA MALLET offered as a part of python’s gensim library. Further, we analysed and plotted the coherence scores as a function of the number of topics and applied the “elbow” method (Annisa et. al., 2019), coupled with some manual inspection to ensure general interpretability and topic separation to finally arrive at the optimal number of topics. We then named the topics based on the relevance.

We performed polarity analysis of the reviews using NLTK’s VADER (Valence Aware Dictionary for Sentiment Reasoning) framework. For each review, VADER returns a neutrality score, positivity score, negativity score and an aggregated compound score, all ranging from -1 to 1 (Wang et. al., 2021). We focus only on the compound score as the aim was to get a brief idea of the spread of the polarities across the topics.

Following this, we labelled each review with its corresponding dominant topic, i.e., the topic under which it was classified by the LDA, and carried out further topical analysis by comparing the frequencies of the topics throughout the corpus against various parameters. We analyse the precautions (only those listed on TripAdvisor by hotels) to obtain a reasonably clear idea of how hotels across the country are responding and adapting to the current situation. Further, we also explore subtle interactions between parameters like precautions taken, price, and type of trip.

4. Analysis and results

Figure 1, which is in the shape of the country India, shows a word cloud that displays the most frequently occurring words in the corpus with reference to the other words (Dataset 2). The bigger the word, the more frequently it occurs in the reviews.

Figure 1: Word cloud of the most used words

4.1 LDA

We perform the LDA on 18,644 pre-processed reviews present in Dataset 2. With the analysis of the coherence scores with the number of topics, and some manual inspection (manually checking if the topics were decipherable with the given collection of words), we represent the corpus as an amalgamation of 9 topics. We name each topic by investigating the unique words primarily to the topic under study and manually perusing reviews in some topics. The LDA also returns the topics with their underlying words and the weight of each word for the respective topic. The description of the topics along with their top words and weights are in Table 1. Table 2 shows some summary statistics of the polarity scores obtained from the reviews from different traveller types.

Table 1: Key topics identified and their top words

Topic ID	Topic	Description	Top words and their weights
0	Food quality and compliance with COVID protocols	Reviews belonging to this topic focus mainly on the quality of food and the covid norms employed by hotels while serving meals	0.166 * "good" + 0.071 * "food" + 0.063 * "service" + 0.058 * "excellent" + 0.055 * "room" + 0.045 * "nice" + 0.033 * "clean" + 0.017 * "staff" + 0.016 * "maintain" + 0.013 * "awesome"
1	Staff service during the COVID pandemic	Various protocols and covid appropriate behaviour that the hotel staff followed to ensure safe service	0.121 * "staff" + 0.066 * "service" + 0.049 * "covid" + 0.049 * "great" + 0.033 * "time" + 0.029 * "food" + 0.028 * "pandemic" + 0.021 * "helpful" + 0.016 * "restaurant" + 0.015 * "specially"
2	Concerns and complaints	General complaints raised by customers, including but not limited to: delayed check-ins, inefficient communication regarding bookings, orders and other services, food having insipid taste, sub-par room features and poor maintenance	0.062 * "room" + 0.015 * "check" + 0.011 * "breakfast" + 0.010 * "book" + 0.009 * "guest" + 0.009 * "call" + 0.008 * "order" + 0.007 * "restaurant" + 0.006 * "request" + 0.006 * "serve"
3	Safety precautions during the hotel stay	Covid safety guidelines adopted all across hotels and the stringent means with which they were enforced	0.204 * "hotel" + 0.145 * "stay" + 0.045 * "staff" + 0.024 * "covid" + 0.019 * "comfortable" + 0.017 * "safe" + 0.016 * "room" + 0.015 * "provide" + 0.012 * "pleasant" + 0.009 * "recommend"
4	Hospitality and stay experience	Warmth with which travellers were welcomed and the homely atmosphere created by the hotel staff	0.099 * "stay" + 0.050 * "experience" + 0.047 * "hospitality" + 0.044 * "make" + 0.036 * "amazing" + 0.025 * "great" + 0.023 * "wonderful" + 0.015 * "memorable" + 0.015 * "comfortable" + 0.014 * "take care"
5	Ambience and amenities	Utilities provided by the hotel	0.033 * "property" + 0.015 * "room" + 0.010 * "restaurant" + 0.009 * "pool" + 0.008 * "enjoy" + 0.008 * "offer" + 0.007 * "area" + 0.006 * "view" + 0.005 * "include" + 0.005 * "open"
6	Overall experience	An all-encompassing topic which touches upon all aspects of the travel with the stay at the hotel at its heart	0.045 * "time" + 0.031 * "day" + 0.029 * "give" + 0.020 * "experience" + 0.015 * "people" + 0.012 * "travel" + 0.012 * "work" + 0.012 * "make" + 0.010 * "thing" + 0.009 * "home"
7	Property location	Mainly focuses on leisure travel experiences in various resorts throughout the country with a special emphasis on the quality of the property and location perks like the scenery, shopping, etc.	0.081 * "place" + 0.054 * "visit" + 0.052 * "resort" + 0.043 * "food" + 0.032 * "family" + 0.031 * "property" + 0.022 * "beautiful" + 0.015 * "enjoy" + 0.013 * "time" + 0.013 * "perfect"
8	Compliments to staff	Overall satisfaction with the stay with special mentions to the various hotel team members like chefs and managers	0.043 * "team" + 0.019 * "make" + 0.016 * "chef" + 0.015 * "ensure" + 0.015 * "guest" + 0.011 * "mr" + 0.011 * "entire" + 0.008 * "serve" + 0.007 * "work" + 0.007 * "manager"

Table 2: Statistics on topics, polarities and traveller types

Topic ID	Topic	Trip type	Reviews with polarity scores ranging from				Polarity Score				Remarks based on the polarity spread
			-1 to -0.5	-0.5 to 0	0 to +0.5	+0.5 to +1	25th Percentile	Median	75th Percentile	Highest	
0	Food quality and compliance with COVID protocols	Business	0	1	2	681	0.91530	0.94795	0.96750	0.99630	Travellers with friends are the most pleased whereas solos and couples have a good share of reviews with lesser polarity scores
		Couple	0	0	2	260	0.91493	0.95825	0.97283	0.99740	
		Family	1	0	4	690	0.92160	0.95430	0.97410	0.99700	
		Friends	0	0	0	238	0.93000	0.95710	0.97365	0.99470	
		Solo	0	0	1	157	0.90958	0.94715	0.96625	0.99630	
1	Staff service during the COVID pandemic	Business	0	1	5	402	0.91818	0.95380	0.97388	0.99780	Business, family and solo travellers have lesser polarity scores than couples and friends
		Couple	0	0	1	135	0.94073	0.96660	0.98050	0.99590	
		Family	0	1	5	513	0.92310	0.96070	0.97860	0.99810	
		Friends	0	0	1	143	0.94128	0.96420	0.97718	0.99650	
		Solo	0	0	0	98	0.92563	0.95780	0.97605	0.99430	
2	Concerns and complaints	Business	24	13	17	157	0.49390	0.91080	0.96780	0.99850	Business travellers leave complaint dominant reviews with the most negative polarities. Couples, the complete opposite
		Couple	16	7	4	153	0.80730	0.95950	0.98700	0.99980	
		Family	40	15	27	272	0.62728	0.95215	0.98190	0.99910	
		Friends	10	4	8	78	0.68080	0.93955	0.98338	0.99870	
		Solo	5	2	4	57	0.74468	0.95015	0.97480	0.99840	
3	Safety precautions during the hotel stay	Business	2	6	7	599	0.90620	0.94930	0.97290	0.99830	The polarity scores in the first quartile corresponding to business and solo travellers appears to be relatively lower
		Couple	0	0	3	153	0.91973	0.95900	0.97573	0.99710	
		Family	2	0	4	382	0.91690	0.95415	0.97535	0.99790	
		Friends	0	2	0	128	0.92243	0.95345	0.97513	0.99590	
		Solo	0	0	2	163	0.90220	0.95450	0.97630	0.99610	
4	Hospitality and stay experience	Business	0	0	0	253	0.93490	0.96650	0.98130	0.99700	All types of travellers except the business travellers seem to be very satisfied
		Couple	0	0	0	267	0.95230	0.97450	0.98750	0.99900	
		Family	0	1	4	575	0.94930	0.97360	0.98560	0.99930	
		Friends	0	0	0	147	0.95380	0.97470	0.98635	0.99650	
		Solo	0	0	1	58	0.94895	0.97230	0.98500	0.99750	

5	Ambience and amenities	Business	0	0	1	48	0.94680	0.97190	0.99420	0.99920	No major differences are observed between the traveller types
		Couple	0	0	1	176	0.95990	0.98530	0.99360	0.99910	
		Family	0	0	3	411	0.96153	0.98430	0.99320	0.99960	
		Friends	0	0	0	79	0.96520	0.98460	0.99330	0.99840	
		Solo	0	1	0	16	0.95310	0.98230	0.98940	0.99860	
6	Overall experience	Business	0	1	3	183	0.90810	0.96230	0.98100	0.99870	On a whole, business and solo travellers were the most dissatisfied
		Couple	0	1	3	94	0.94030	0.96705	0.98778	0.99860	
		Family	1	0	6	212	0.93475	0.97560	0.98920	0.99930	
		Friends	0	0	1	74	0.94340	0.97230	0.98585	0.99920	
		Solo	0	1	0	71	0.92565	0.96885	0.98585	0.99920	
7	Property location	Business	0	0	0	75	0.91775	0.95710	0.97880	0.99460	Solo and business travellers had the least enjoyable experience, families had the best experience
		Couple	0	0	1	159	0.94298	0.96940	0.98263	0.99770	
		Family	1	0	2	609	0.94930	0.97410	0.98595	0.99900	
		Friends	1	0	1	174	0.93898	0.96960	0.98440	0.99790	
		Solo	0	0	2	24	0.90715	0.94225	0.95998	0.99610	
8	Compliments to staff	Business	0	4	3	387	0.94300	0.97320	0.98850	0.99940	No major differences are observed between the traveller types
		Couple	0	0	0	162	0.95928	0.98280	0.99230	0.99960	
		Family	0	0	0	486	0.95110	0.97845	0.99010	0.99940	
		Friends	0	1	0	91	0.94575	0.97710	0.98828	0.99900	
		Solo	0	0	0	108	0.94698	0.97565	0.98958	0.99940	

4.2 Topic vs polarity

We explore the interaction between the topics and their corresponding polarity measures (compound score returned by VADER). Figure 2 is a detailed depiction of the spread of the polarity scores of different subgroups that are formed based on the dominant topic and the trip type. The classic boxplot covers the data in the upper and lower bounds of the 75th and 25th percentile respectively, up to 1.5 times the interquartile range. The individual black dots signify outliers. The 25th percentile mark of the polarity scores for almost all the topics is more than 0.75, except for the topic ‘Concerns and complaints’ (Topic ID: 2).

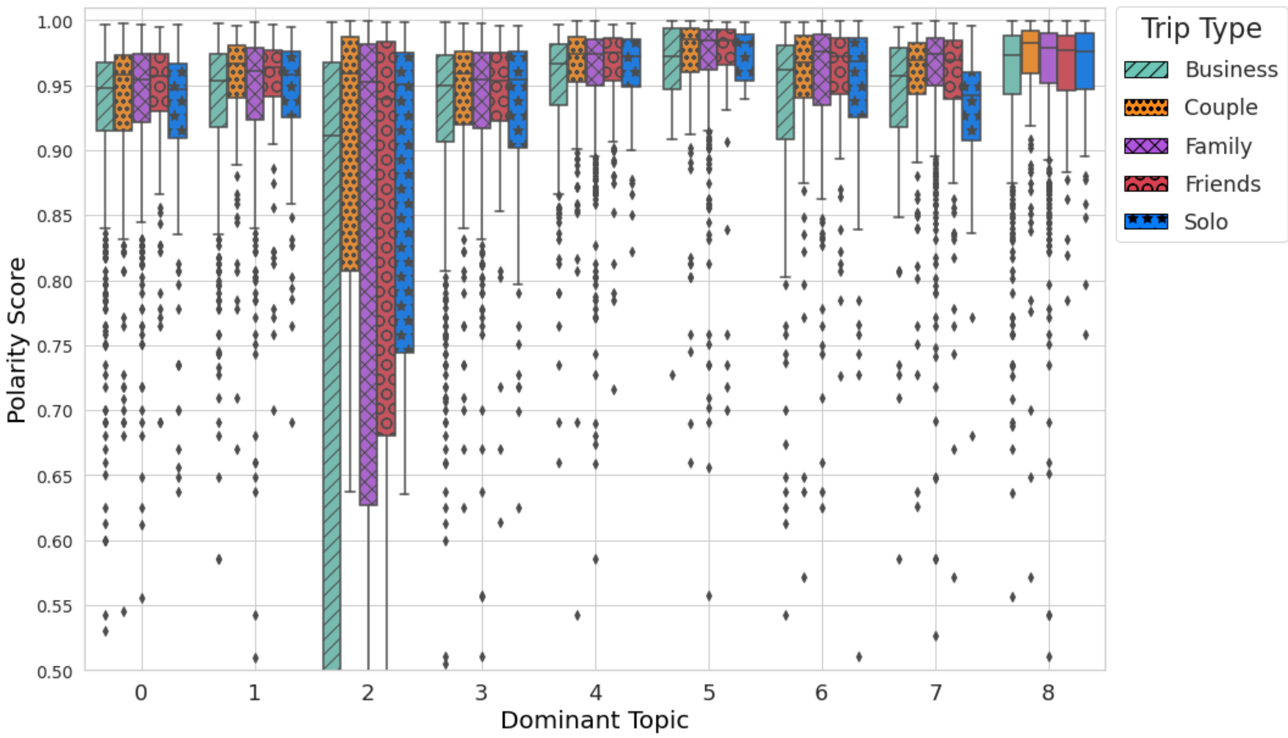

Figure 2: Polarity boxplots per topic per traveller type

Out of all COVID related topics, the polarity scores of “Safety precautions during the hotel stay” stretch out the lowest with particularly negative reviews coming from business and solo travellers. Business travellers and families took the most issue with the quality of staff service. Friends consistently left behind the most positive reviews for all COVID related topics. “Ambience and amenities”, “Hospitality and stay experience” and “Compliments to staff” are the three most positively polarised topics. Solo travellers had the lowest polarity scores while talking about resorts.

We felt the need to visualise the topic ‘Concerns and complaints’ separately as the polarity scores were very widespread for each traveller type and demanded more scrutiny. The violin plot (Figure 3) illustrates the kernel density estimate along with the 25th percentile, median and 75th percentile of the polarity scores pertaining to the topic ‘Concerns and complaints’ for each traveller type.

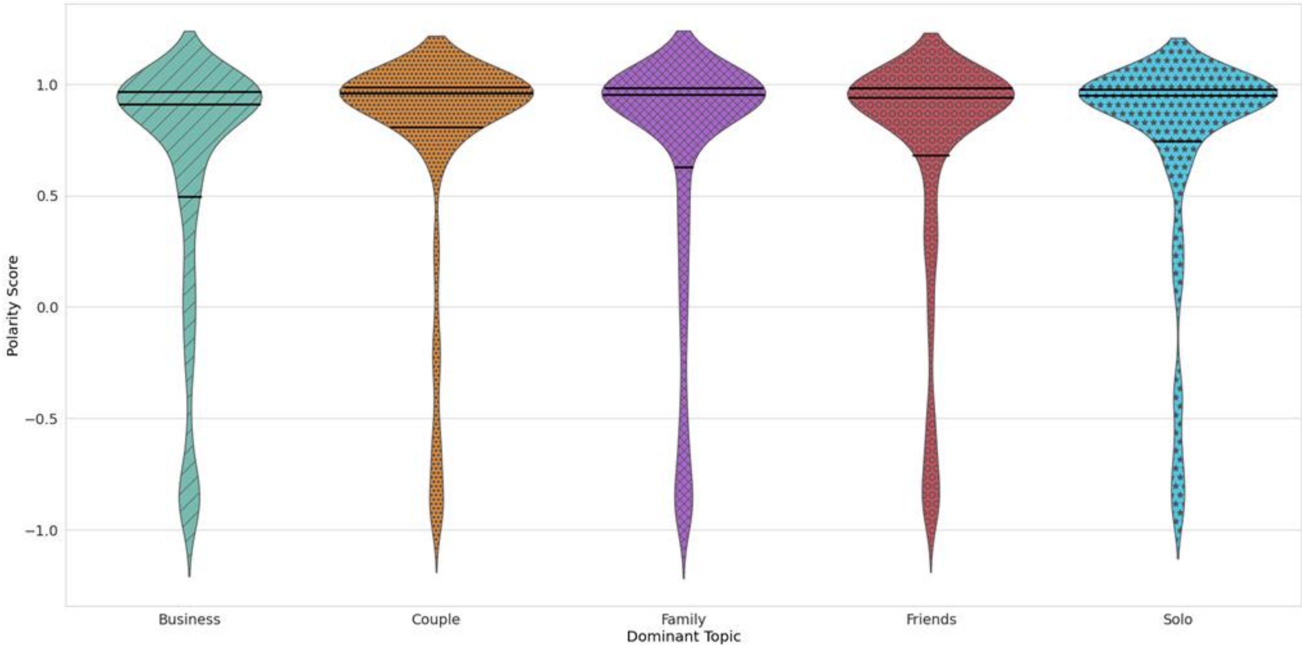

Figure 3: Violin plot of the polarity for the topic ‘Concerns and complaints’

The complaints from Business travellers are the most negatively polarised, followed by the complaints from families and friends. Though couples tend to leave behind several reviews containing concerns and complaints (discussed later in the paper), they do not use as many negative words while describing their concerns.

4.3 Dominant topic vs count of precautions

To explore the potential relationship between the topics and precautions, we explore the spread of topic frequencies in each type of hotel (categorised by the number of precautions taken). To start, we adopt a binary categorisation approach for the precautions variable. We coin two buckets, namely (i) “Listed” and (ii) “Not Listed”, signifying hotels that have listed one or more precautions on TripAdvisor and the ones that have not listed any precautions on TripAdvisor.

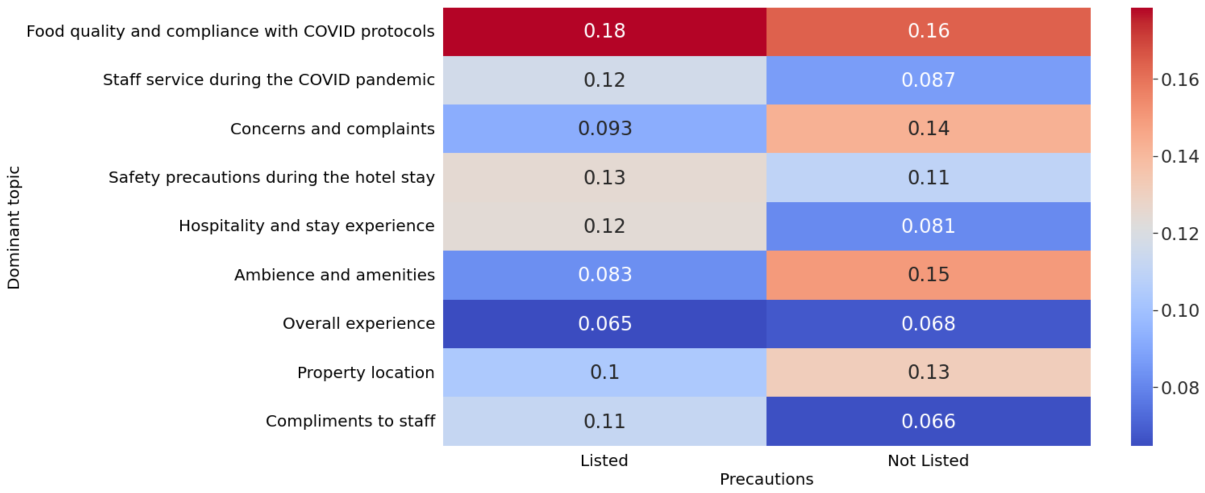

Figure 4: Heat map of topic distribution in hotels that have listed and not listed precautions

The numbers in Figure 4 signify the percentage of reviews that belong to a particular topic in a given category of precautions taken by hotels. We see a clear distinction in the topic distribution above, with COVID-related topics being discussed more in reviews of hotels that have listed precautions. In addition to this, for the other class of hotels, the focus tends to shift to topics such as, “Ambience and amenities”, “Property Location” and “Concerns and Complaints”, and a proportional decrease in topics such as, “Compliments to staff” and “Hospitality and stay experience”. Given these results, we investigate within the “Listed” hotels category by binning the “Number of precautions” into four buckets, [1, 10], (10, 12], (12, 13] and (13, 14] using quantile splitting. The corresponding frequencies are 5332, 4424, 2776 and 3951 respectively.

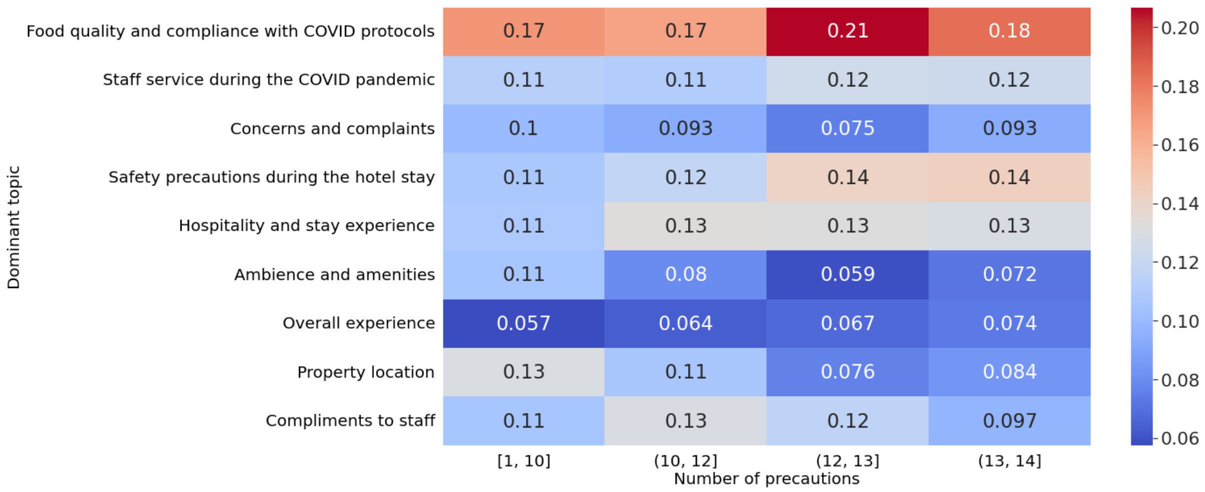

Figure 5: Heat map showcasing the relation between the dominant topic and the binned precautions count

The colour-coded heat map's annotations in Figure 5 signify the percentage of reviews that belong to a particular topic in a given bin of the precautions count. Property location is most discussed in the first two bins and also most resorts across the country take less than 10 precautions. People raise complaints more frequently in hotels taking the least number of precautions. The level of concern for COVID related matters (share of 1st, 2nd and 4th topic) is almost the same across all bins.

We further challenged the thought of there being a relation between the number of precautions taken and the topics people talk about, with a statistical test by framing the following initial hypotheses:

H0: *The number of precautions taken do not have any effect on the spread of the reviews over topics*

H1: *The number of precautions taken may influence the spread of reviews over topics*

We use the chi-square test of independence as both the variables in question are nominal and satisfy all the prerequisite conditions (McHugh, 2013; Ozdemir et. al., 2012; Rahman et. al., 2020; Song et. al., 2019). The results suggest to reject the null hypothesis leading one to believe that the two variables have some dependence.

4.4 Topics vs traveller type

We observe interesting results when we draw a comparison between the topic frequencies against the type of trip. Figure 6 signifies the percentage of reviews that belong to a particular topic, given a traveller type.

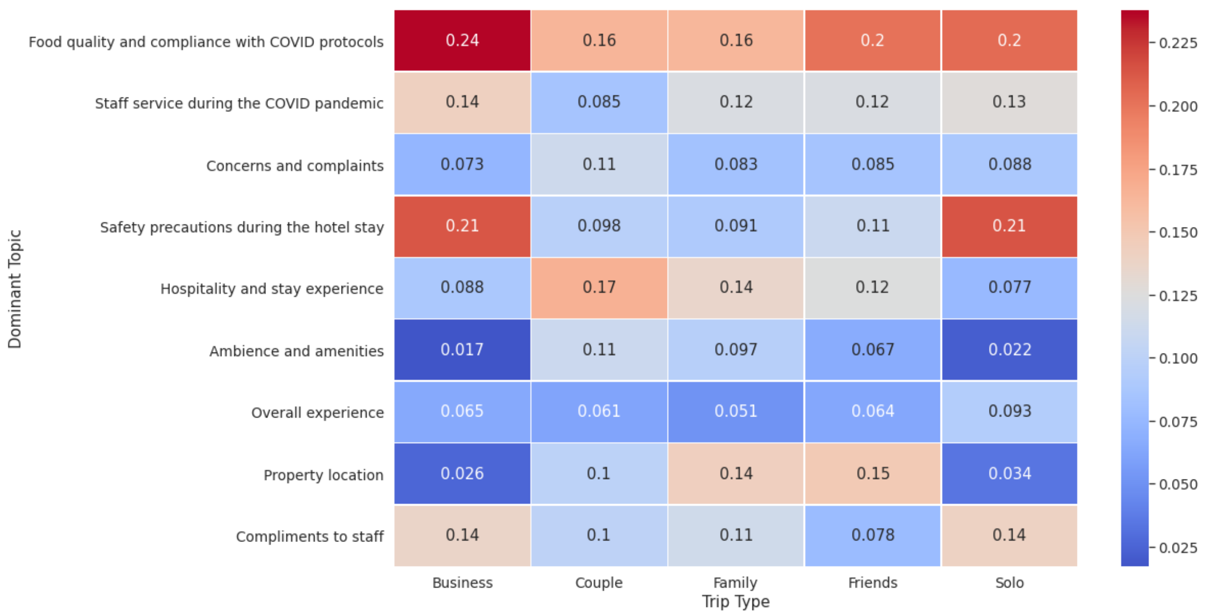

Figure 6: Heat map highlighting the percentage shares of each topic by traveller type

Business travellers, followed by solo travellers, focus the most on covid related topics. Couples, families, and friends emphasise the hospitality, ambience, and property location. It can also be observed that couples are very likely to complain. Table 3 presents the three most discussed topics by the different types of travellers. For all traveller types except couples, at least 2 out of their top 3 topics of discussion are related to COVID.

Table 3: Top 3 topics of discussion for each traveller type

	Business	**Couple**	**Family**	**Friends**	**Solo**
1	Food quality and compliance with COVID protocols	Hospitality and stay experience	Food quality and compliance with COVID protocols	Food quality and compliance with COVID protocols	Safety precautions during the hotel stay
2	Safety precautions during the hotel stay	Food quality and compliance with COVID protocols	Hospitality and stay experience, Property location	Property location	Food quality and compliance with COVID protocols
3	Staff service during COVID pandemic, Compliments to staff	Concerns and complaints, Ambience and amenities	Staff service during COVID pandemic	Staff service during COVID pandemic, Hospitality and stay experience	Compliments to staff

4.5 Topic trend with time

Figure 7 conveys the trend in the volume of the reviews observed for each topic on a monthly basis. The value of the Y axis for a particular month signifies the number of reviews observed during the course of that month. The drop in the volume of the reviews is in line with the lockdown restrictions imposed on travel during the COVID pandemic.

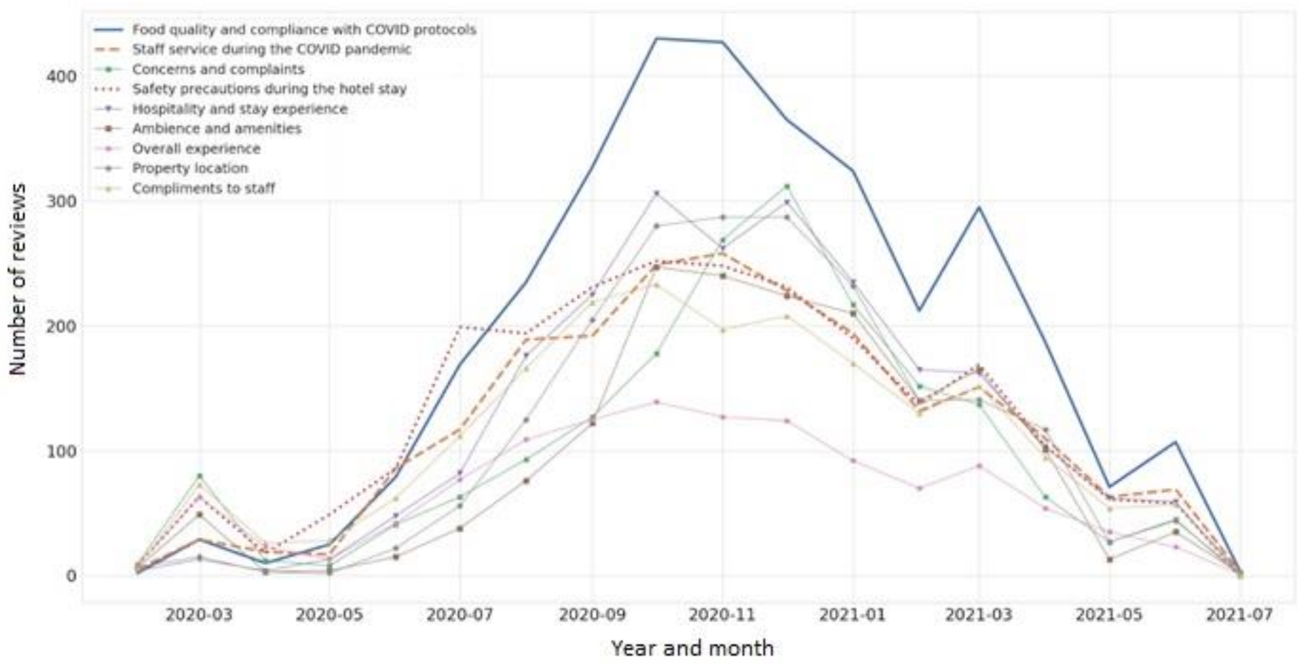

Figure 7: Trend in the number of reviews by topic over time (monthly frequency)

4.6 Price vs precautions

4.6.1 Based on review

We consider all entries with prices above ₹15059.8125 (upper whisker limit) to be outliers and removed them for price-related analysis. Figure 8 compares the mean price of the hotels with the number of precautions taken, faceted on the traveller type. The estimated mean of the mean price for each of the number of precautions taken along with the confidence interval surrounding it is also available. With this, we aim to understand the price perspective for each type of traveller and if they fancy spending more for a hotel that has listed more precautions. We observe little to no deviation from the line for business travellers, similar patterns of deviation for friends, families and couples, and some deviations with large confidence intervals (95%) for solos. Overall, we observe a positive correlation between the number of precautions taken and the mean price for all traveller types.

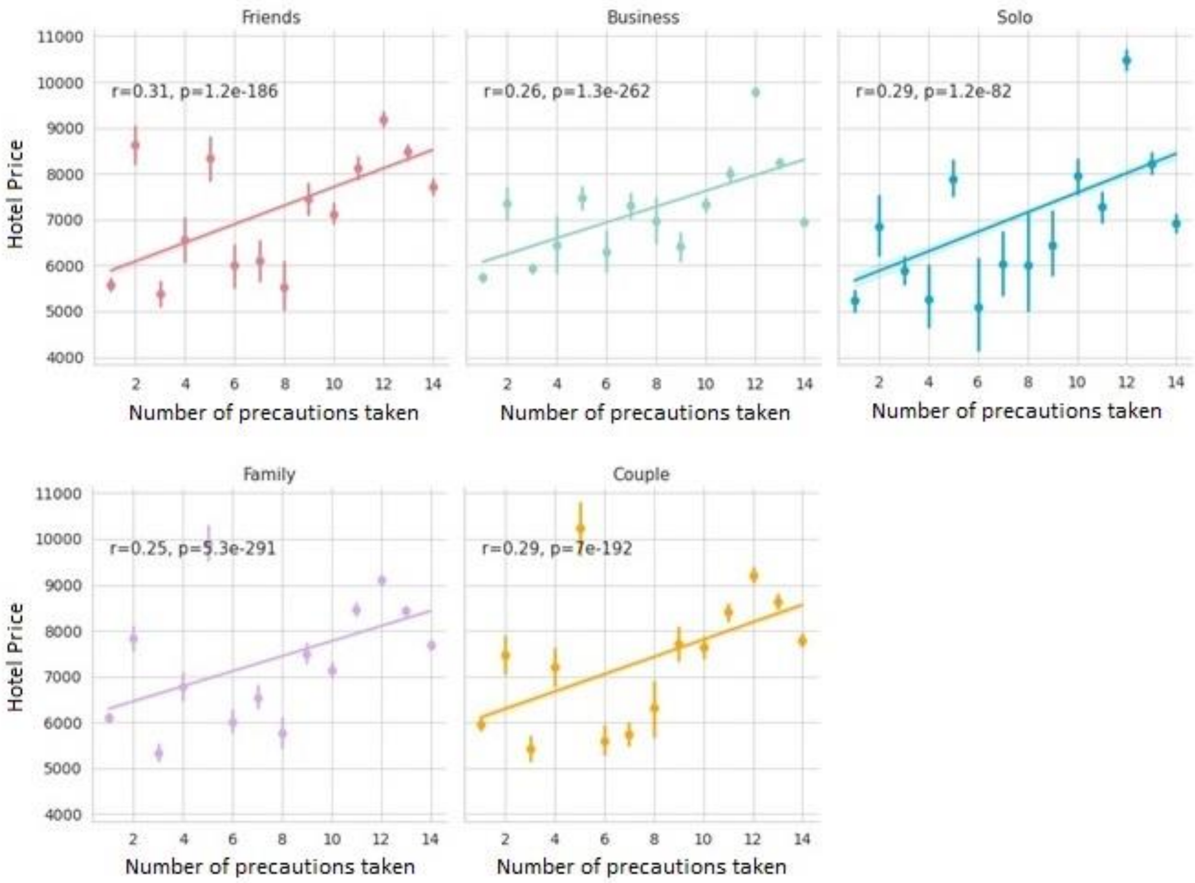

Figure 8: Plot of the price of the hotels (per night) vs the number of precautions taken

4.6.2 Based on hotel

Table 4 presents a summary of the average hotel price. It indicates a large spread in the price with the standard deviation being more than half of the mean and the median. We find that out of 1174 hotels that we have considered, 612 have not listed the precautions they are taking on TripAdvisor. Hence, to ensure the legibility of these figures, all entries that have no precautions listed are removed. Similarly, all entries that did not have a price listed on TripAdvisor are removed. To better observe the price variations, hotels with prices above the upper whisker limit (in this case, 75% price + 1.5 * inter-quartile range of the mean price = 15059.8125) are also removed.

Table 4: Summary statistics of the mean price of hotels (in Indian Rupees)

Mean	Standard Deviation	Minimum	25th percentile	Median	75th percentile	Maximum
7209.77	4682.64	1100	4357.00	5971.25	8638.125	42572.50

Figure 9 shows a bar chart of the frequency of hotels that have listed a certain number of precautions in their TripAdvisor page and also a box plot of the mean hotel price with respect to the number of precautions listed. It helps to visually compare the number of precautions taken by the hotels and the mean price of the hotels.

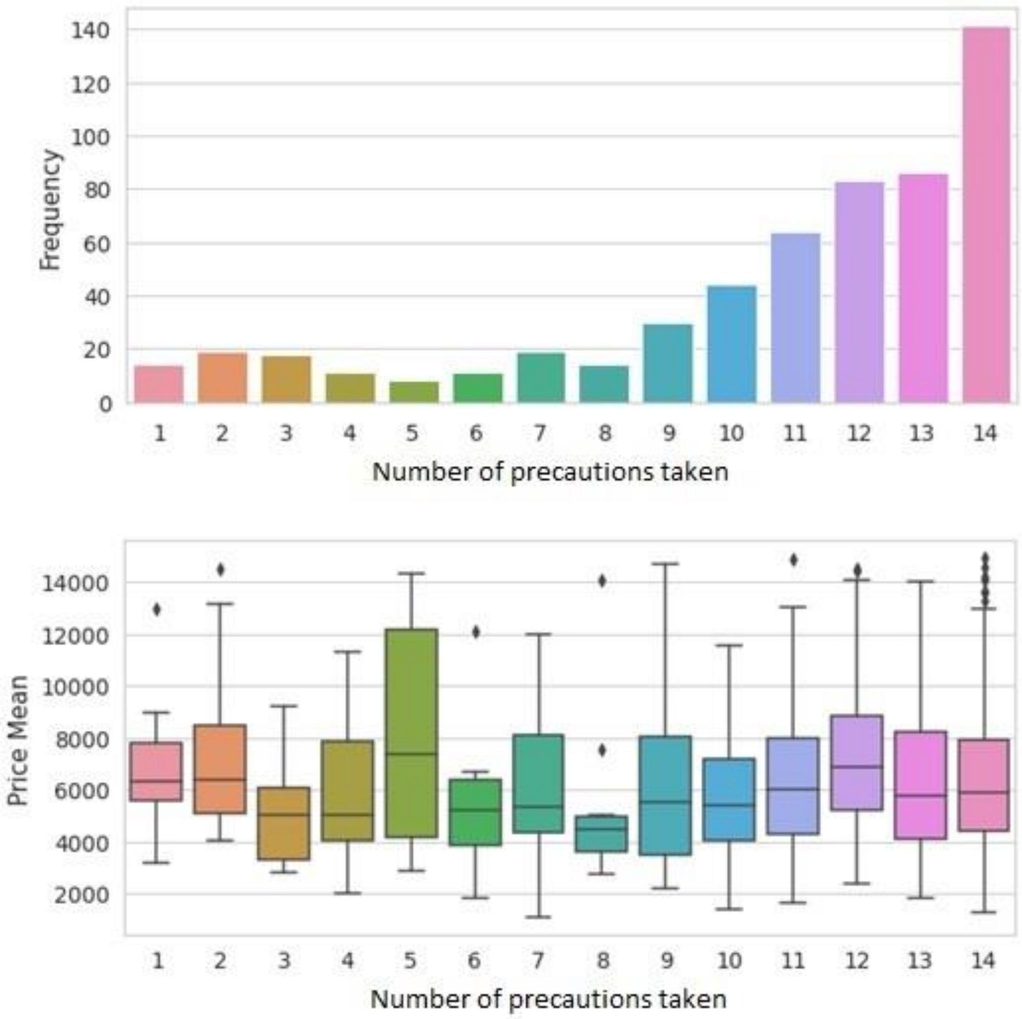

Figure 9: Bar chart of the frequency of hotels and box plot of the mean price

Out of the hotels that have listed their precautions, most of them belong to the 12-14 bracket. The price does not seem to follow any particular trend with the number of precautions. Table 5 presents the summary statistics of the mean price of the hotels according to the number of precautions taken.

Table 5: Descriptive statistics of the mean price and number of precautions taken

Precaution count		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Frequency		14	19	18	11	8	11	19	14	30	44	64	83	86	141
Price	**Average**	6836	8062	7560	6004	8046	8904	6611	5065	7184	6594	7407	7597	7324	7106
	25th Percentile	5570	5109	3375	4025	4165	4086	4420	3628	3543	4195	4310	5249	4271	4435
	Median	6338	6709	5689	5036	7344	5707	5400	4507	6015	5512	6082	7043	5931	5988
	75th Percentile	7808	9266	7018	7873	12190	6577	8883	4948	8602	8021	8516	9027	9162	8431
Number of outliers		0	1	2	0	0	1	1	0	2	3	3	1	4	8

4.7 Types of precautions taken

Figure 10 presents the frequency of the different precautions listed by the hotels on TripAdvisor. As expected, the most adopted precaution is making hand sanitisers available to everyone at the hotel. The least taken measure is a paid stay-at-home policy for staff with symptoms. Many hotels listed a similar set of precautions, which is evident from the frequency of the precautions.

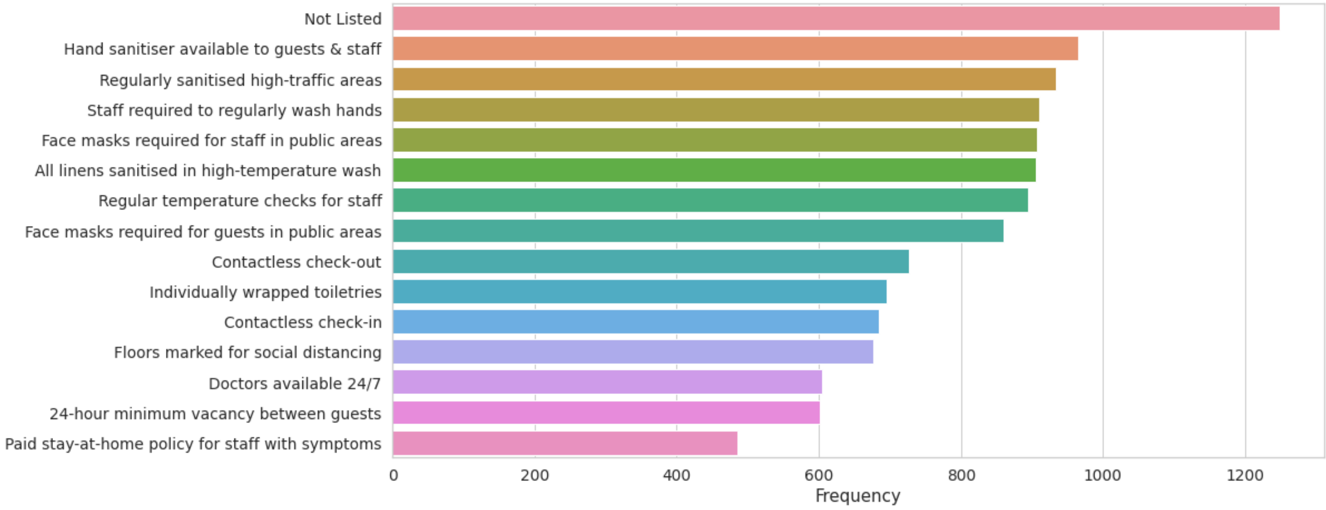

Figure 10: Types of precautions listed on TripAdvisor vs the corresponding number of hotels

Figure 11 shows the polarity distribution for each type of precaution taken, considering only the reviews present in the covid topics. This helps to get an idea of how people talk about the protocols taken in the presence and absence of each type of precaution. On the x-axis, 0 indicates that the precaution has not been listed and 1 indicates otherwise. For each of the first five precautions (Figure 11), not too much of a difference is observed in the spread of the polarities. The lower whisker and 75th percentile of the box corresponding to 1 (the precaution being listed) is consistently above that corresponding to 0 (the precaution not being listed).

Except regular temperature checks for staff, almost every other precaution seems to follow a similar polarity score distribution irrespective of whether the precaution is listed or not. A counter-intuitive result is observed for Doctors available 24/7 with the polarity distribution corresponding to 1 stretching out a bit lower.

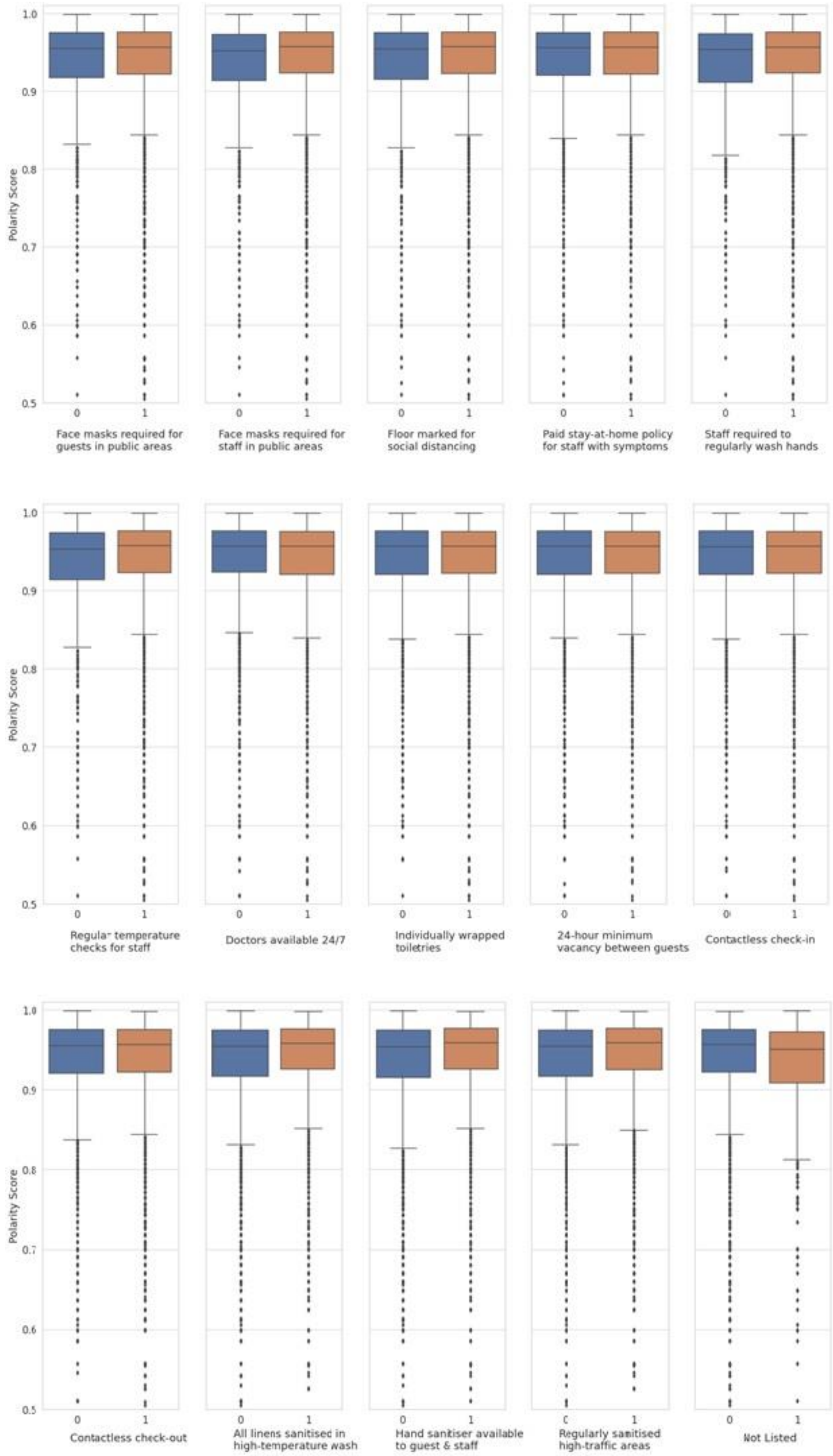

Figure 11: Box plots of the polarity score for each type of precautions listed

For the last four precautions listed in Figure 11, the lower whisker and 75th percentile of the box corresponding to 1 is consistently above that corresponding to 0. Lastly, for the Not Listed category, we observe a relatively large dip in the polarity with the 75th percentile and lower whisker falling by a large amount for 1 while the median and 25th percentile also form a bit lower. Having established the above observations, it is, in fact, counter-intuitive that the median polarity scores are very similar for the "Not Listed" category and all the other categories in Figure 11. This potentially hints towards the weak positive contribution of precautions listed to the polarity of the review.

4.8 Topics vs precautions

The precautions listed may have an impact on the dominant topics that emerged from the LDA. Figure 12 is a heat map comprising of the proportion of reviews belonging to each topic when analysed by the different precautions taken. While there is almost negligible difference in the distribution of topics with the types of precautions taken, there is a stark difference in the same when it comes to hotels that have not listed their precautions compared to those that have.

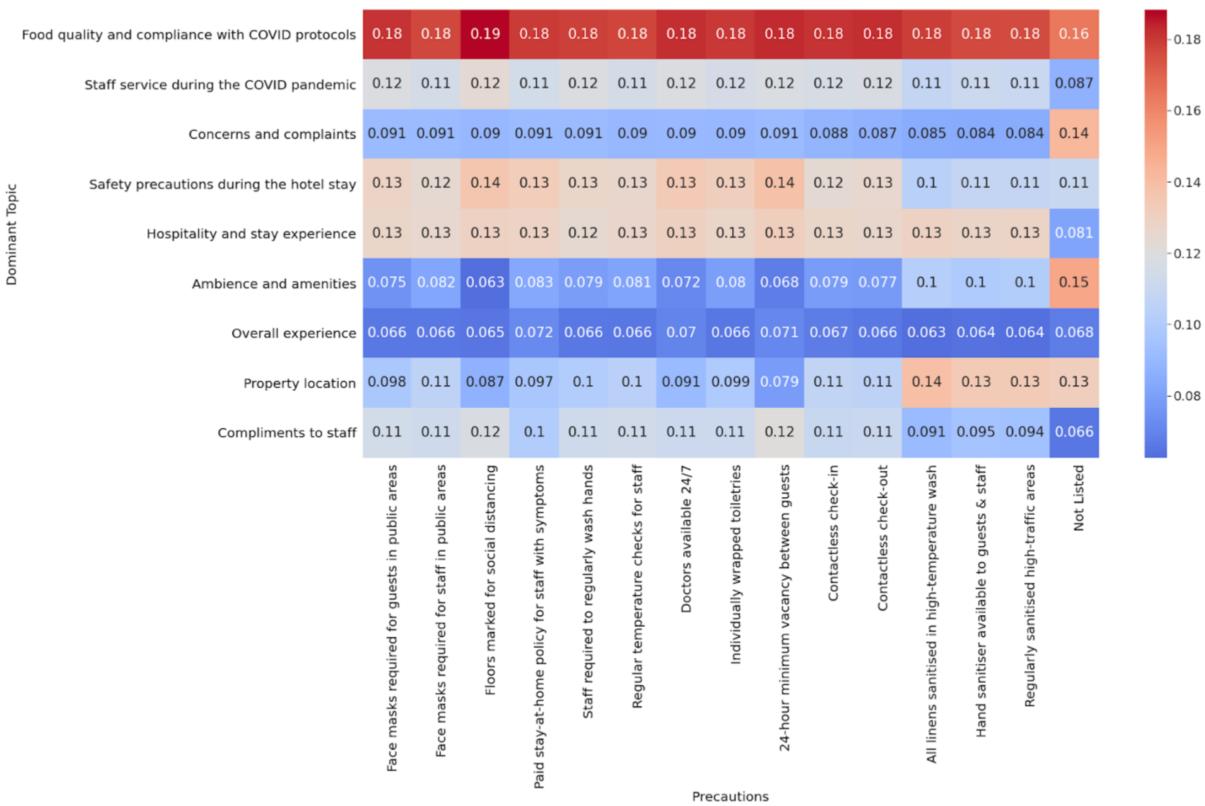

Figure 12: Spread of topics for each type of precaution taken

5. Discussion and insights

5.1 Dominant topics from LDA

As discussed in section 4, we identified nine dominant topics based on our topic modelling results. We looked at the reviews under each topic to understand about the intended nature of the reviews. The topic "Food quality and compliance with COVID protocols" comments on the food and restaurants in the hotel, general hygiene, cleanliness and sanitation protocols. "Staff service during the COVID pandemic" topic mentions specific staff members who assisted guests, about service quality, protocols followed, management's efficiency in enforcing protocols and other regular operations. "Concerns and complaints" highlights delayed check-in complaints, poor food and room quality, subpar amenities and service, shortcomings in sanitisation and breach/ineffectiveness of enforcement of COVID protocols.

"Safety precautions during the hotel stay" comments about the array of precautions the hotels took and their effectiveness, common mentions of sanitisation practices followed by hotels and most reviews confirm the enforcement of COVID protocols. "Hospitality and stay experience" describes how the team handled anxious travellers (as they travelled during the pandemic) in terms of assuring them of safety, making them feel at home, and traits like friendliness, politeness and courtesy showed. The "Ambience and amenities" topic includes description of in-room amenities (like bathroom, spaciousness, ACs, etc.), description of swimming pools (mentions if the pools were closed as part of safety measure; if they were open, descriptions about their surroundings and how well they are maintained), outdoor games for recreation, ambience and decor (both in-room and outside like common areas, lobbies, restaurants, bars, etc.). Some reviews have descriptions of buildings and architecture as well.

"Overall experience" includes reviews related to the time and duration of stay, the purpose of their visit (most travellers mention that they wanted to relieve COVID-induced anxiety. Other reasons for stay include relative's marriage, typical staycation, work, etc.) and if they got what they expected. Some reviews mention experiences with staff, food and overall hygiene. The reviews in the "Property location" topic are exclusive to resorts. They mostly talk about the location, view, scenery, beach (if there was one), etc. Apart from this, we observe general mentions of hospitality, hygiene, protocols, etc. Almost all reviews under "Compliments to staff" refer directly to the staff with their names to thank them. This includes front desk receptionists, chefs, waiters, managers, and everyone else who promptly responded to requests and ensured quality service.

5.2 Polarity analysis

We observe that the reviews written by business and solo travellers across all topics consistently have lower polarities than that of the other travellers (Figure 2). Chang et al. (2019) put forward that business travellers generally tend to rate the lowest (overall and for individual aspects) and more frequently use negative words in their reviews to express their dissatisfaction. They analysed hotels in the Hilton chain in the U.S. According to our results, this finding by Chang et. al. (2019) holds true in the Indian context as well.

We see that solo travellers tend to leave behind negative reviews for the “Property location” topic in particular, which means that leisure stays at resorts have not been too pleasurable for the solo travellers. This could be because solo travellers try to explore nearby places by walk in general, for which the location of the property plays a major role.

5.3 Precautions count vs dominant topic

People notice and talk more about COVID related topics in their reviews in hotels that take a higher number of precautions. These hotels also have a much lesser percentage of complaints raised by travellers. The focus on the ambience and other amenities appears to decrease as the number of precautions increase.

Herrero et al. (2015) argue that TripAdvisor contains rich e-word of mouth (eWOM) information, which affects consumer behaviour. They conclude that a user’s decision-making process is related to the perceived value of the given information, credibility of the source, and similarity between users and review generators. Hotel managers should increase the number of precautions that they are taking and listing on TripAdvisor. This will help more customers to leave behind positive reviews about the quality and variety of precautions, ensuring that potential customers are assured of safety.

5.4 Travel type vs topic

Business and solo travellers focus on covid related topics the most, while couples, families, and friends emphasise the hospitality, ambience, and property location. Couples are most likely to complain with 11% (Figure 5) of their reviews belonging to the “Concerns and complaints” topic, while the same number stands at around 8% for the others. The above contradicts what Banerjee et al. (2016) observed, who say that couples often leave behind the best ratings and feedback. It is also worth noting that the reviews corresponding to the concerns raised by couples were actually in the higher end of the polarity spectrum (Figure 3).

Hotels should ensure that the COVID protocols followed are satisfactory, especially for business and solo travellers. For hotels/resorts that are more popular among leisure travellers, we recommend focusing on ensuring quality hospitality, good amenities and aesthetic ambience.

5.5 Topic trend with time

Figure 6 shows a rise in reviews starting from May, which is steeper in June (Hotel Opening Unlock 1.0, 2020) and continues until October. We observe a little saturation from October to November. A familiar fall starts from November (Jeffrey et al., 1999; Jeffrey et al., 1994) and crashes post-March with the advent of the 2nd wave (Kar et al., 2021). A slight spike occurs during February. The fall during January might have been due to new surge scares.

During June and July, the topic "Safety precautions during the hotel stay" dominates with the maximum volume of reviews. Hence, while opening up, i.e., when a wave subsides, and movement begins gradually, travellers expect the precautions to be at their most stringent. "Concerns and complaints" saw a sharp increase starting September and had the second most number of reviews by December. One possible explanation is that the number of guests the hotels accommodated was at the year's high, and short-staffed hotels could not efficiently enforce COVID protocols and other routine operations (Job losses in hotels, 2020).

5.6 Precautions not listed

Most hotels that have listed their precautions take 12-14 precautions (Figure 8). There is still massive room for improvement on this front as more than half of the Indian hotels on TripAdvisor have not listed any precautions. We recommend the concerned professionals and hotel managers to take swift steps for listing all precautions they take on their TripAdvisor page. As shown by Herrero et. al. (2015) this will give customers much higher confidence about the safety of their stay.

We also show how the polarity score of the reviews drop (though the median remains counter intuitively similar) for hotels that have not listed precautions compared to the ones that have. The share of reviews belonging to the "Concerns and Complaints" topic increases sharply for hotels that have not listed precautions. These facts solidify our case further.

5.7 Price vs precautions count

While comparing the number of precautions count parameter with the price (Figure 9), we conclude that two things are playing out - some hotels provide a good number of precautions at a reasonable price, while some hotels charge a lot but do not provide enough precautions.

Given this, there is also a huge variability in price when compared against the number of precautions taken and hence, nothing conclusive can be said about the exact trend.

Figure 8 reveals that there is a positive Pearson's coefficient R between the price and number of precautions for all traveller types. In a sense, irrespective of the type of trip they plan, all travellers are willing to spend extra if they feel the hotel is diligent in its COVID-related protocols.

5.8 Precautions vs topics

A fall in the share of COVID-related topics and a significant rise in the share of complaints raised (Figure 12) show that people complain more during the pandemic when they stay in hotels that have not listed their precautions. For these hotels, the focus shifts to the ambience and the amenities offered by the hotel. We also observe a proportional decrease in hospitality and compliments to the staff.

6 Practical and research implications

This study provides a robust framework to understand and analyse the effects of the pandemic on travellers and their experience during travel and hotel stays. With the countries going through subsequent waves of the virus due to its mutative capabilities, any UGC based study in the context of tourism in the coming years must strongly consider the role of the pandemic.

We find that our observations on the polarities of reviews left by business travellers are consistently and significantly lower than those left by other travellers. This result closely mimicked that of a study carried out in the context of hotels belonging to the Hilton chain in the U.S. (Chang et. al., 2019). Further, researchers working with Indian UGC data can also expect similar observations.

When analysed against the number of precautions taken, the topic "Property location" came up the most in reviews corresponding to hotels taking less than 10 precautions. In the above types of hotels, the share of complaints was also much higher. These changes and their significance is tested and affirmed with the chi-square test. All COVID related topics are seen to have almost equal weightage across the precautions spectrum.

The study explores the intricacies of the types of precautions taken by hotels, their effects on guests' experience and what they tend to talk about in their reviews. We find that listing the precautions taken in the TripAdvisor page alone is not useful, until and unless the precautions are actually implemented in spirit. "Paid stay-at-home policy for staff with symptoms" is the least listed precaution. This could be due to the fact that most hotels were completely shut-down without any revenue. They might not have got the money or the additional staff to implement this precaution as they had to cut corners during the lockdowns.

The patterns we point out concerning how the volume of reviews and the areas of concern change as the waves begin to grow and subside can also be analysed by studies and are guaranteed to give quality insights. With the slow revival of travel and tourism, we posit that UGC researchers prioritise analysing content generated by domestic tourists to bring out their needs better.

7. Conclusions, limitations and further research

Our study unearths the multiple underlying dimensions in customer reviews written during the pandemic which will be useful for researchers, practitioners and hotel managers. We bring out the spread of the polarity within these dimensions for each type of traveller, where we find that people tend to emphasise COVID-related dimensions when they stay in hotels that take a higher number of precautions. With our analysis and existing literature, we show that hotel managers must take active steps to list their hotel's precautions on TripAdvisor.

Our study provides a good research framework to account for several COVID related aspects and can be used to describe how things change compared to a typical UGC analysis study in the context of tourism. While our work has valuable contributions, it does not come without limitations. We have not considered the finer details like the profile of the reviewers which could be useful to further segment travellers beyond the trip type. Setting aside TripAdvisor's merits, there have been undeniable cases of fake/paid reviews (Filieri et al., 2015; Chang et al., 2019). Our study considers content only from one source, and previous research (Xiang et al., 2017) has shown that this may lead to bias in observations.

Multimedia is also available at the disposal of a researcher in the form of pictures that some travellers share along with their reviews. Other factors such as the aspect ratings, location, room features, amenities and review heading can also be analysed. Their effects on COVID related parameters can be an interesting study for researchers and practitioners.

References:

Aggarwal, S., & Gour, A. (2020). Peeking inside the minds of tourists using a novel web analytics approach. Journal of Hospitality and Tourism Management, 45, 580-591.

Annisa, R., & Surjandari, I. (2019). Opinion Mining on Mandalika Hotel Reviews Using Latent Dirichlet Allocation. Procedia Computer Science, 161, 739-746.

Au, N., Buhalis, D., & Law, R. (2014). Online complaining behavior in mainland China hotels: The perception of Chinese and non-Chinese customers. International Journal of Hospitality & Tourism Administration, 15(3), 248-274.

Ayeh, J. K., Au, N., & Law, R. (2013). "Do we believe in TripAdvisor?" Examining credibility perceptions and online travelers' attitude toward using user-generated content. Journal of Travel Research, 52(4), 437-452.

Banerjee, S., & Chua, A. Y. (2016). In search of patterns among travellers' hotel ratings in TripAdvisor. Tourism Management, 53, 125-131.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. the Journal of machine Learning research, 3, 993-1022.

Chang, Y. C., Ku, C. H., & Chen, C. H. (2019). Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor. International Journal of Information Management, 48, 263-279.

Delly M. Chatibura (2020) Travellers' top comments during the COVID-19 pandemic in Botswana, Research in Hospitality Management, 10:2, 123-130, DOI: 10.1080/22243534.2020.1869462

Filieri, R., Alguezaui, S., & McLeay, F. (2015). Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. Tourism management, 51, 174-185.

Geetha, M., Singha, P., & Sinha, S. (2017). Relationship between customer sentiment and online customer ratings for hotels-An empirical analysis. Tourism Management, 61, 43-54.

Government Guidelines Unlock 1.0 (2020) https://www.news18.com/news/india/unlock-1-0-govt-issues-detailed-guidelines-for-hotels-hospitality-units-ahead-of-reopening-2653635.html

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. Tourism Management, 59, 467-483.

Han, S., & Anderson, C. K. (2021). Web scraping for hospitality research: Overview, opportunities, and implications. Cornell Hospitality Quarterly, 62(1), 89-104.

Herrero, Á., San Martín, H., & Hernández, J. M. (2015). How online search behavior is influenced by user-generated content on review websites and hotel interactive websites. International Journal of Contemporary Hospitality Management.

Hotel Opening Unlock 1.0 (2020) https://timesofindia.indiatimes.com/india/unlock-1-0-govt-releases-guidelines-for-opening-of-restaurants-offices-and-religious-places/articleshow/76201154.cms

Hu, N., Zhang, T., Gao, B., & Bose, I. (2019). What do hotel customers complain about? Text analysis using structural topic model. Tourism Management, 72, 417-426.

Jeffrey, D., & Barden, R. R. (1999). An analysis of the nature, causes and marketing implications of seasonality in the occupancy performance of English hotels. Tourism Economics, 5(1), 69-91.

Jeffrey, D., & Hubbard, N. J. (1994). A model of hotel occupancy performance for monitoring and marketing in the hotel industry. International Journal of Hospitality Management, 13(1), 57-71.

Job losses in hotels (2020) https://timesofindia.indiatimes.com/city/jaipur/job-losses-in-hotels-rise-up-to-80-employees-face-over-50-pay-cut/articleshow/76414720.cms

Kar, S. K., Ransing, R., Arafat, S. Y., & Menon, V. (2021). Second wave of COVID-19 pandemic in India: Barriers to effective governmental response. EClinicalMedicine, 36.

Kirilenko, A. P., Stepchenkova, S. O., & Dai, X. (2021). Automated topic modeling of tourist reviews: does the Anna Karenina principle apply?. Tourism Management, 83, 104241

Kumar, V. (2020). Indian tourism industry and COVID-19: present scenario. Journal of Tourism and Hospitality Education, 10, 179-185.

Levy, S. E., Duan, W., & Boo, S. (2013). An analysis of one-star online reviews and responses in the Washington, DC, lodging market. Cornell Hospitality Quarterly, 54(1), 49-63.

Lu, W., & Stepchenkova, S. (2015). User-generated content as a research mode in tourism and hospitality applications: Topics, methods, and software. Journal of Hospitality Marketing & Management, 24(2), 119-154.

Luo, J. M., Vu, H. Q., Li, G., & Law, R. (2020). Topic modelling for theme park online reviews: analysis of Disneyland. Journal of Travel & Tourism Marketing, 37(2), 272-285.

Manickas, P. A., & Shea, L. J. (1997). Hotel complaint behavior and resolution: A content analysis. Journal of Travel Research, 36(2), 68-73.

Mankad, S., Han, H. S., Goh, J., & Gavirneni, S. (2016). Understanding online hotel reviews through automated text analysis. Service Science, 8(2), 124-138.

McHugh, M. L. (2013). The chi-square test of independence. Biochemia medica, 23(2), 143-149.

Mehta, M. P., Kumar, G., & Ramkumar, M. (2021). Customer expectations in the hotel industry during the COVID-19 pandemic: a global perspective using sentiment analysis. Tourism Recreation Research, 1-18.

Ministry of Tourism (2019) https://tourism.gov.in/sites/default/files/2020-09/ITS%20at%20a%20glance_Book%20%282%29.pdf

Nieto García, M., Muñoz-Gallego, P. A., Viglia, G., & Gonzalez-Benito, O. (2020). Be social! The impact of self-presentation on peer-to-peer accommodation revenue. Journal of Travel Research, 59(7), 1268-1281.

Ozdemir, B., Aksu, A., Ehtiyar, R., Çizel, B., Çizel, R. B., & İçigen, E. T. (2012). Relationships among tourist profile, satisfaction and destination loyalty: Examining empirical evidences in Antalya region of Turkey. Journal of Hospitality Marketing & Management, 21(5), 506-540.

Rahman, S. A. A., Yusof, M. A., Nakamura, H., & Nong, R. A. (2020). Challenges of smart tourism in Malaysia eco-tourism destinations. Planning Malaysia, 18(14).

Roy, A., Suresh, P., & Nirmal, S. S. (2020). Chapter Impact of the Covd-19 Pandemic on Travel and Tourism Industry and Future Revival Strategy with Special Reference to Indian Tourism Sector. Hospitality and Tourism Industry, 18.

Schuckert, M., Liu, X., & Law, R. (2015). Hospitality and tourism online reviews: Recent trends and future directions. Journal of Travel & Tourism Marketing, 32(5), 608-621.

Song, S., Kawamura, H., Uchida, J., & Saito, H. (2019). Determining tourist satisfaction from travel reviews. Information Technology & Tourism, 21(3), 337-367.

Taecharungroj, V., & Mathayomchan, B. (2019). Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. Tourism Management, 75, 550-568.

Uğur, N. G., & Akbıyık, A. (2020). Impacts of COVID-19 on global tourism industry: A cross-regional comparison. Tourism Management Perspectives, 36, 100744.

UNWTO (2021) https://www.unwto.org/impact-assessment-of-the-covid-19-outbreak-on-international-tourism

vanden Broucke, S., & Baesens, B. (2018). Practical Web Scraping for Data Science: Best Practices and Examples with Python. Apress.

Wang, J., Li, Y., Wu, B., & Wang, Y. (2020). Tourism destination image based on tourism user generated content on internet. Tourism Review.

Wang, L., & Kirilenko, A. P. (2021). Do tourists from different countries interpret travel experience with the same feeling? Sentiment analysis of TripAdvisor reviews. In Information and Communication Technologies in Tourism 2021 (pp. 294-301). Springer, Cham.

WTTC (2020) https://wttc.org/Portals/0/Documents/Reports/2021/Global%20Economic%20Impact%20and%20Trends%202021.pdf?ver=2021-07-01-114957-177

Wu, L., Mattila, A. S., Wang, C. Y., & Hanks, L. (2016). The impact of power on service customers' willingness to post online reviews. Journal of Service Research, 19(2), 224-238.

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. Tourism Management, 58, 51-65.

Yu, M., Li, Z., Yu, Z., He, J., & Zhou, J. (2020). Communication related health crisis on social media: a case of COVID-19 outbreak. Current issues in tourism, 1-7.