How the Pandemic Affected Airbnb and Hotel Consumer Sentiment

Team 85

Brendan Bailey; Roman Casillas; Thomas Dahlstrom;
Vida Gholami; Michael Martinez; Omar Rana

Motivation/Introduction

The COVID-19 pandemic has challenged the business community to provide consistent and quality service in the middle of a public health crisis. Nowhere has this challenge been greater than in the travel industry, where factors such as social distancing, travel limitations, and hygienic prioritizations have influenced normal business operations.

Consumer sentiment is critical in determining whether a business survives or fails during such a tumultuous time; using natural language processing techniques, we can analyze reviews of Airbnb and hotels to determine the level of satisfaction and key topics within those reviews both pre- and post-pandemic. As far as we can tell, no study has been performed on consumer sentiment evolution throughout the pandemic as it pertains to the lodging sector of the hospitality industry.

Data

Data for Airbnb reviews was downloaded from Inside Airbnb for 28 cities for multiple years as compressed *.csv* files. Airbnb data required significant deduplication efforts. Data for TripAdvisor reviews was scraped using Apify in *.json* format and contained extraneous data out of which only price level, zip code of hotel, date of review and the review text were kept. Raw data size exceeded 5GB and over 10 million records. Naturally, the data tilted heavily towards pre-pandemic reviews since more people were traveling. More than 85% of the reviews were pre-pandemic.

Our final two output files were sized 1.13GB for the VADER algorithm and 1.3GB for the LDA algorithm. Figure 1 displays total review count per quarter.

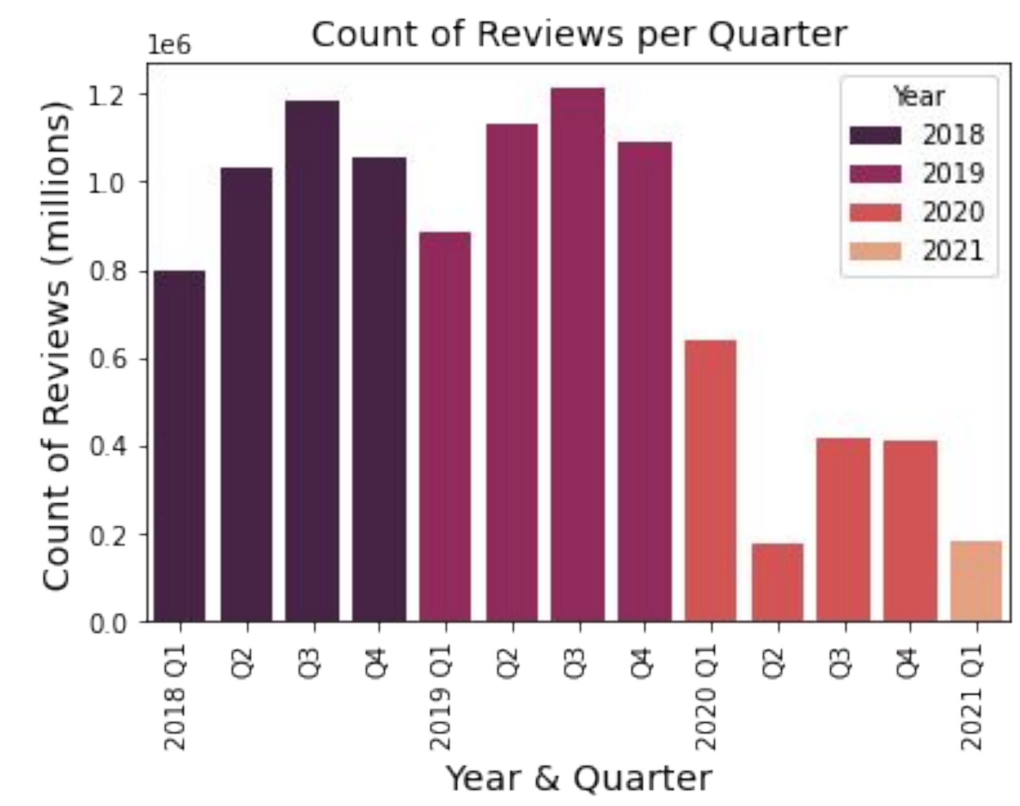

Figure 1. Bar Chart - Count of Reviews per Quarter.

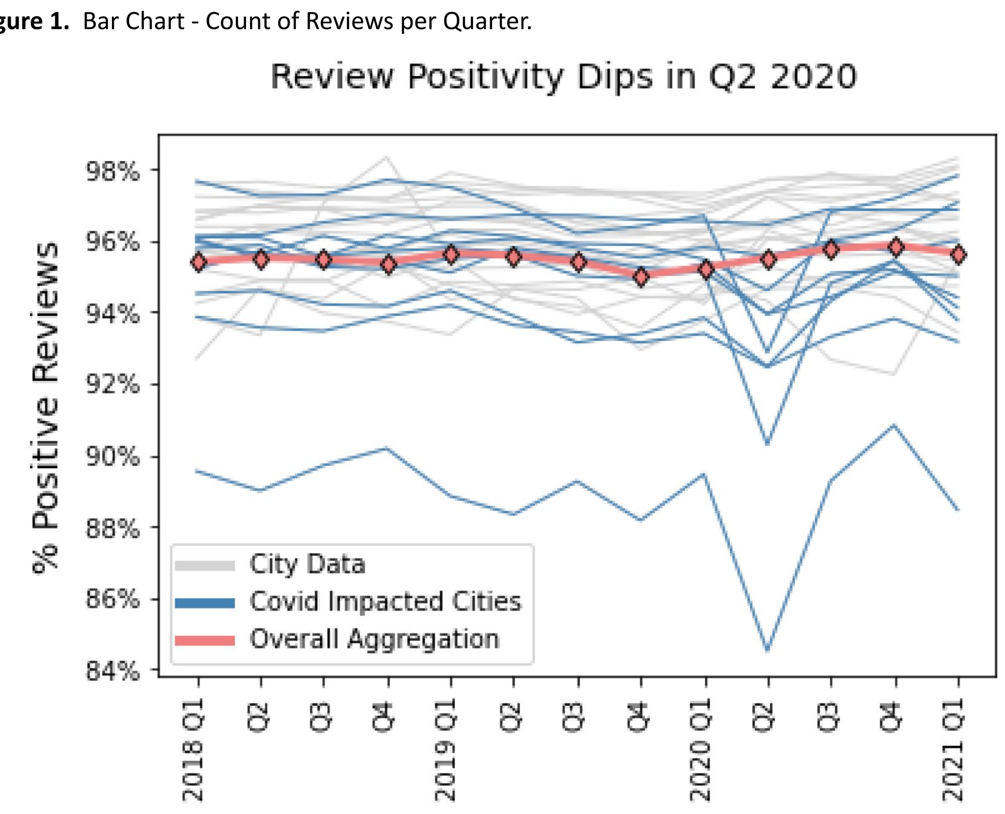

Figure 5. Line Chart - Review Positivity over Time.

Method and Approach

After data had been collected and cleaned, our analytical purpose was to determine both the overall sentiment of the reviews and extract specific topics that are frequently discussed.

To achieve these analytical goals, two different techniques were selected. For sentiment analysis, Valence Aware Dictionary and sEntiment Reasoner (**VADER**) was chosen. Developed in 2014, VADER compares the text to a rule-based lexicon dictionary to calculate a positive, negative, neutral, and a compound (normalized) polarity score [1]. The compound result is generally used to classify the entire text as positive, negative, or neutral, and scores range between -1 and +1. The strength of the score itself also indicates the intensity of the sentiment, so scores may be compared and ordered. VADER is preferred to other techniques because it includes acronyms, emoticons, slang, and punctuation common in online contexts, and it outperformed other baseline methods [2, 3].

For topic extraction, latent Dirichlet allocation (**LDA**) was selected. Developed in 2003, LDA employs a tri-level Bayesian mixture model to the reviews in order to generate a set of underlying topics [4]. Prior to executing the algorithm, LDA requires that the text is lemmatized, where stopwords are removed and word variations are standardized. One parameter set by the user is the number of topics to generate--this value may affect how much clarity and/or overlap occurs between extracted topics. LDA then scores the relevancy of particular words to that topic--so it is left for the researcher to define the topic, based on the words that comprise its distribution [5]. We experimented with different numbers of topics to determine which produced the most meaningful results. One significant aspect to our approach was removing proper nouns during topic extraction to produce more meaningful results.

We consider the following to be our innovations: creating scripts to run both textual analysis algorithms, and the interactive visualization (website) for comparing lodging reviews pre- and post-COVID, by location. This combination of analysis and visualization exceeds the state-of-the-art, and demonstrates a contribution to COVID effects' research. Also, the amount of reviews processed is 10 million (compared to 1 million processed in the following study [6]), and contains reviews from several websites and multiple cities.

The results from both VADER and LDA model were visualized. The visualization depicts three maps where the user can compare review statistics between different locations and time periods (see Figures 2, 3, and 4). For the visualization, D3.js was used. In addition to the D3.js base code, various plugins were included. d3-legend.js simplifies the creation of legends and is primarily used on the first map. d3-tip.js simplifies the creation of tooltips and is utilized for all locations on all three maps. d3.layout.cloud.js generates word clouds easily and is specifically used on the third map.

Table 1. LDA Topics and Key Words.

Assigned Topic Name	Key Descriptive Words
Overall Positive Stay	host place stay great clean beautiful space recommend apartment everything
Issue Related	night room place get door issue bit house time stay
Nightlife Location	great walk restaurant location close downtown walking distance place minute
Accommodations Related	kitchen space room bed comfortable clean water nice bedroom bathroom
Great Value Location	great location place good stay nice clean value price really
Scenic Location	boston hill exploring garden bay ferry federal east rent trail
Communication Related	check question communication respond arrival day response communicate message reservation
Relaxing Vacation	beach perfect weekend cottage house getaway family day night stayed

Experiment and Results

The results from both VADER and LDA models were validated. In the case of the VADER algorithm, 100 review samples were selected and assessed against human-interpreted labels (Table 2). The accuracy and precision were calculated to be 96.25% and 97.96% respectively. For LDA model evaluation, the topic u_mass coherence metric was used; our output had minimal variation, ranging from -3.16 to -3.05. Although not ideal, these scores fall within an acceptable range.

Results for the VADER algorithm show 2.91% of reviews being negative during COVID-19 and 2.92% reviews being negative prior to COVID-19, virtually identical. However, we found a difference between Airbnb (private owners) and TripAdvisor (commercial hotels) when comparing negative review rates: Airbnb registered only 1.74% while TripAdvisor posted 8.29% negative. These results suggest that, while overall positivity towards lodging experiences remained the same in spite of the pandemic, there is a prominent difference –consumers are more likely to criticize hotels with a negative review. We also analyzed positivity trends over time, by quarter: there is a distinguishable drop in positivity during 2019 Q4 that steadily improves through 2020 Q2 (Figure 5). However, the overall positivity percentage for city-wise aggregation across quarters suggests minimal variance from the expected norm during 2020 Q2, which is when we would expect reviews to be the worst due to the beginning of COVID.

Similar to sentiment, there were only minor shifts in review topics before and after the pandemic began. The largest changes were with "Great Value Location" reviews which increased by 2.25 percentage points, "Issue Related" reviews decreased by 2.24 percentage points, and "Overall Positive Stay" reviews, which increased by 2.05 percentage points.

In general, compared to similar studies performed on sentiment analysis, the amount of data we used is significantly larger [6]. Also, this seems to be a unique study specifically performed on the sentiment analysis field using pre- and post-pandemic data, applied to the hospitality and tourism industry.

Table 2. Confusion Matrix VADER.

	Predicted Positive	Predicted Negative
Actual Positive	48	1
Actual Negative	2	29
Total	**50**	**30**

Conclusions

The analysis of the data showed a significant decline in the number of reviews as the pandemic started, which corresponds to the decrease in tourism (Figure 1). There was a notable difference in the polarity of the reviews between Airbnb and hotel industry, which implies that customers of hotel industry are more likely to post negative reviews. The negative to positive ratio in the comments increases starting 2019-Q4, which might not be completely tied to the pandemic, but the trend continued during the pandemic.

Although we were expecting changes in sentiment and topics before and after the pandemic began, the analysis has shown negative results. This may be due to a lack of diversity in review content as compared to other analyses that sentiment analysis and topic modeling is performed on (i.e. news articles, social media posts, etc.). Still we think this methodology can be applied to these other corpora and domains outside of hospitality to gain a broader understanding of how the pandemic shifted discourse in our society.

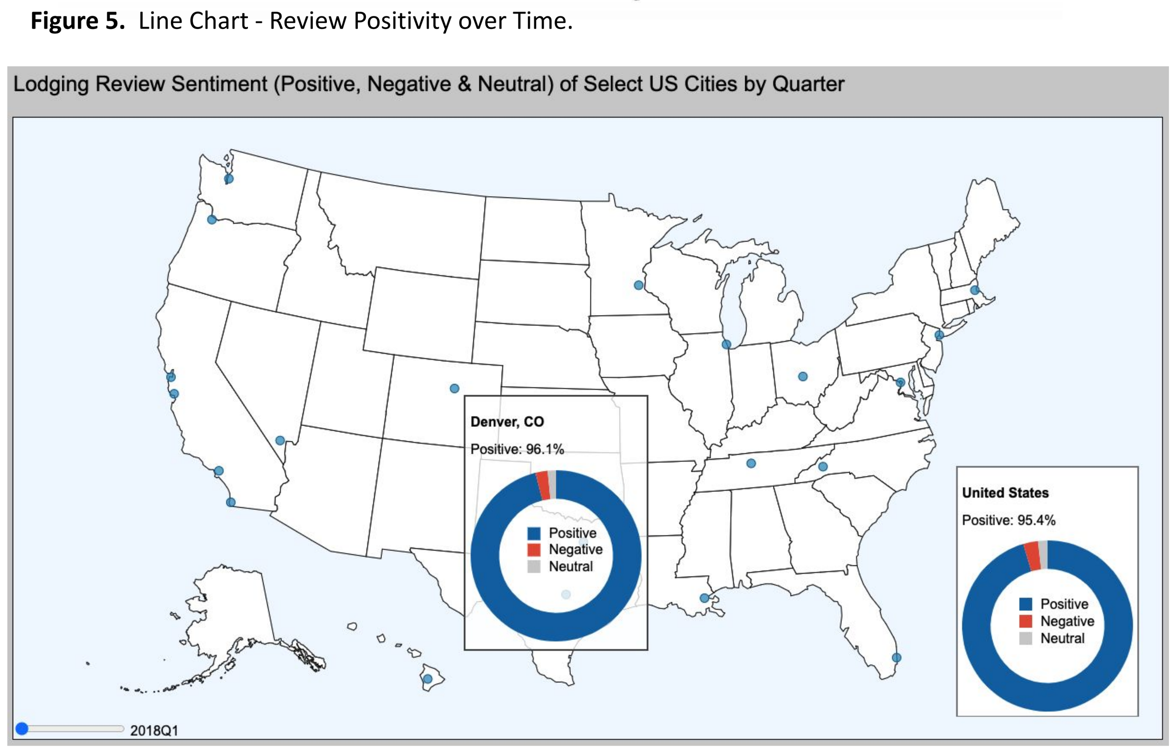

Figure 2. Map 1 - Lodging Review Sentiment.

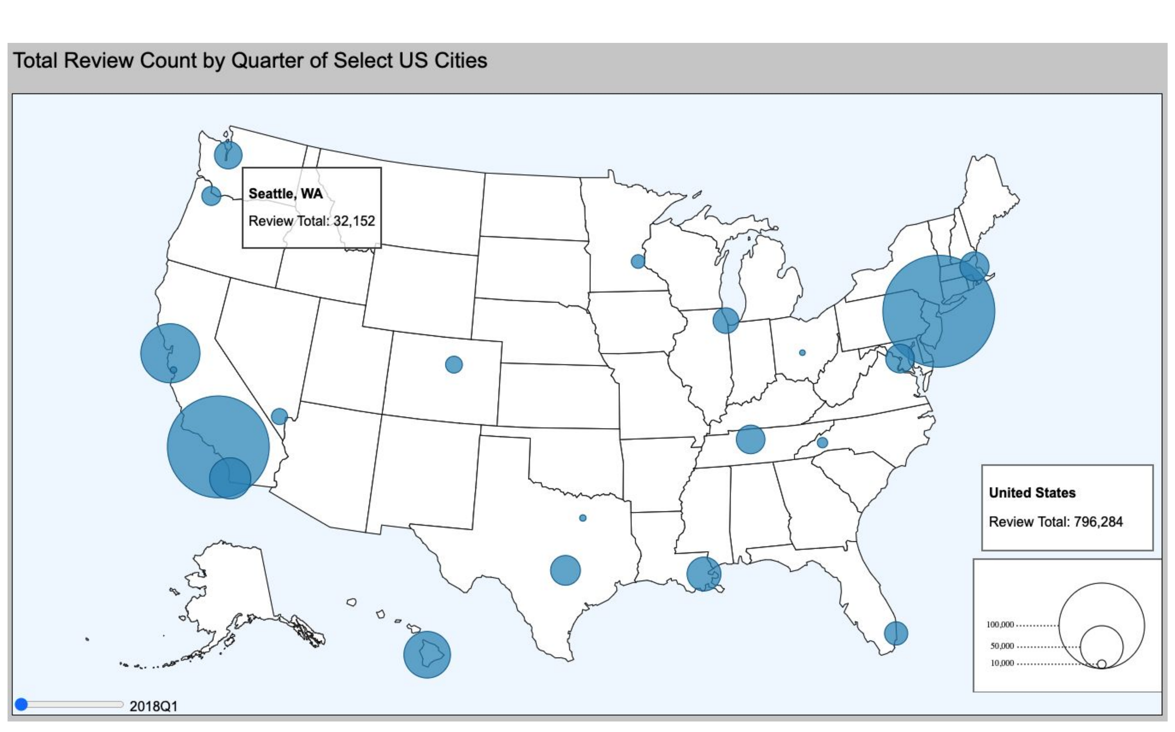

Figure 3. Map 2 - Total Review Count.

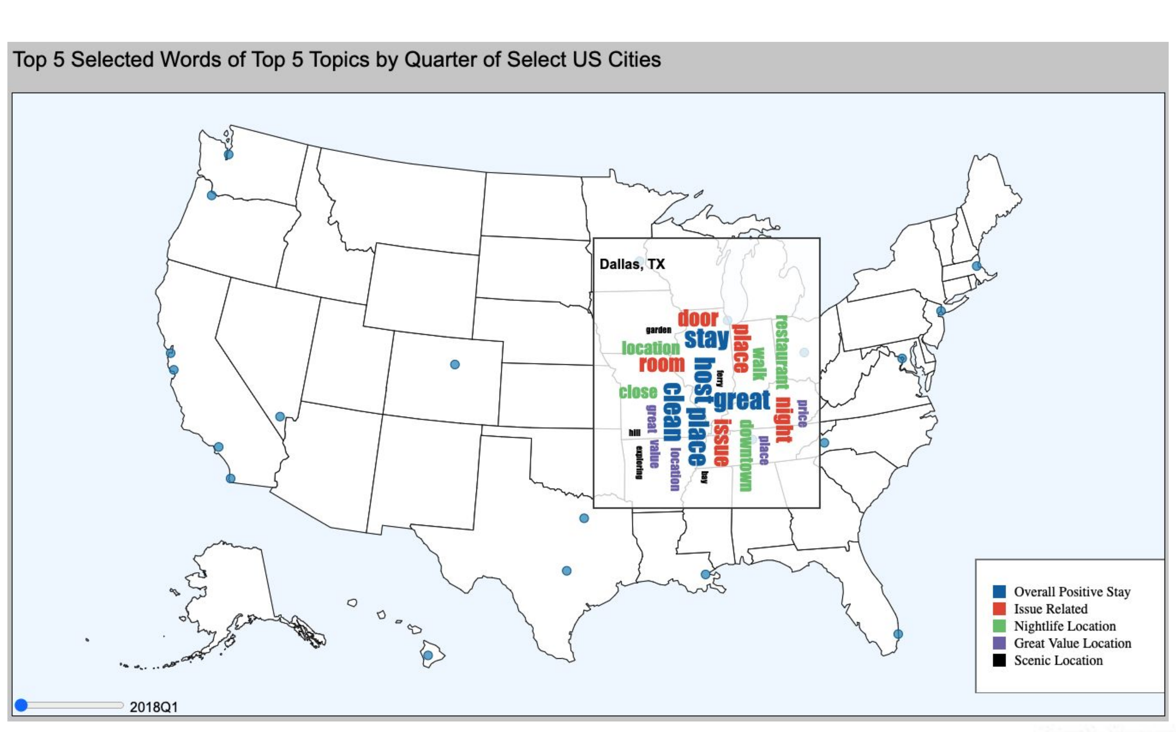

Figure 4. Map 3 - Topic Word Clouds.

CSE 6242 - Data & Visual Analytics
Georgia Tech, College of Computing

Acknowledgements

Inside Airbnb
Adding data to the debate
APIFY
airbnb
tripadvisor

References

1. C. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text," in *Proceedings of the Eighth AAAI Conference on Weblogs and Social Media*, 2014.
2. S. Al-Natour and O. Turetken, "A comparative assessment of sentiment analysis and star ratings for consumer reviews," *International Journal of Information Management*, vol. 54, pp. 102-132, 2020.
3. X. Fang and J. Zhan, "Sentiment analysis using product review data," *Journal of Big Data*, vol. 2, no. 5, 2015.
4. M. D. Hoffman and D. M. B. F. Blei, "Online Learning for Latent Dirichlet Allocation," *NIPS'10: Proceedings of the 23rd International Conference on Neural Information Processing Systems*, vol. 1, 2010.
5. D. M. Blei, A. Y. Ng and M. I. Jordan, "Latent dirichlet allocation," *The Journal of Machine Learning Research*, vol. 3, pp. 993-1022, March 2003.
6. I. Sutherland and K. Kiatkawsin, "Determinants of Guest Experience in Airbnb: A Topic Modeling Approach Using LDA," *Sustainability* , vol. 12, no. 8, 2020.