

Quantifying the Effects of COVID-19 on Restaurant Reviews

Ivy Cao, Zizhou Liu, Giannis Karamanolakis, Daniel Hsu, Luis Gravano

Columbia University, New York, NY 10027, USA

{ic2502, z12889}@columbia.edu, {gkaraman, djhsu, gravano}@cs.columbia.edu

Abstract

The COVID-19 pandemic has implications beyond physical health, affecting society and economies. Government efforts to slow down the spread of the virus have had a severe impact on many businesses, including restaurants. Mandatory policies such as restaurant closures, bans on social gatherings, and social distancing restrictions have affected restaurant operations as well as customer preferences (e.g., prompting a demand for stricter hygiene standards). As of now, however, it is not clear how and to what extent the pandemic has affected restaurant reviews, an analysis of which could potentially inform policies for addressing this ongoing situation.

In this work, we present our efforts to understand the effects of COVID-19 on restaurant reviews, with a focus on Yelp reviews produced during the pandemic for New York City and Los Angeles County restaurants. Overall, we make the following contributions. First, we assemble a dataset of 600 reviews with manual annotations of fine-grained COVID-19 aspects related to restaurants (e.g., hygiene practices, service changes, sympathy and support for local businesses). Second, we address COVID-19 aspect detection using supervised classifiers, weakly-supervised approaches based on keywords, and unsupervised topic modeling approaches, and experimentally show that classifiers based on pre-trained BERT representations achieve the best performance (F1=0.79). Third, we analyze the number and evolution of COVID-related aspects over time and show that the resulting time series have substantial correlation (Spearman's $\rho=0.84$) with critical statistics related to the COVID-19 pandemic, including the number of new COVID-19 cases. To our knowledge, this is the first work analyzing the effects of COVID-19 on Yelp restaurant reviews and could potentially inform policies by public health departments, for example, to cover resource utilization.

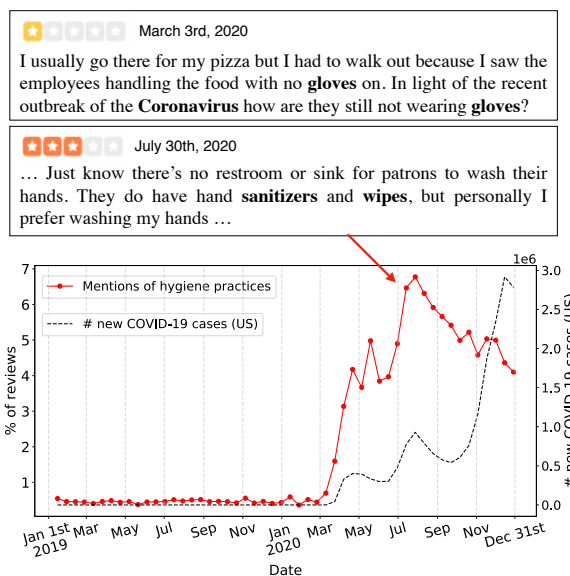


Figure 1: Top: Examples of Yelp restaurant reviews discussing hygiene practices. Bottom: Time series showing the number of reviews discussing hygiene practices and also the number of new COVID-19 cases in the US.

1 Introduction

The outbreak of the SARS-CoV-2 virus in December of 2019 and its evolution to the COVID-19 pandemic have had many devastating consequences in society. Restaurants have been among the hardest-hit businesses during the pandemic.¹ Yelp data (as of September 2020) shows that out of the 32,109 restaurant closures in the U.S., 61% have been permanent, and a greater impact is observed in local businesses in larger metropolitan areas, such as New York City and Los Angeles County, on which we focus in this paper.

Restaurants operate under great uncertainty during this ongoing situation and, therefore, it is critical to understand how the pandemic has affected public attitude towards restaurants. The disruption in daily routines as well as fear and anxiety due

¹<https://www.yelpeconomiccoverage.com/business-closures-update-sep-2020.html>

to the pandemic have been shown to affect eating habits (Naja and Hamadeh, 2020; Di Renzo et al., 2020). The pandemic may have also affected customers’ preferences, such as changes in cuisine types, or higher expectations of hygiene and social distancing practices followed by restaurants.

In this paper, we present our efforts to understand the effects of COVID-19 on restaurant reviews. Reviewers provide ratings and free-form text to express their opinions and experiences about restaurants and we argue that the pandemic has affected such reviews. As an example, Figure 1 shows a Yelp review discussing the hygiene practices of a restaurant, including a mention of “coronavirus” and associated concerns. To understand more broadly the effect of the pandemic on restaurant reviews, we analyze 3 million Yelp reviews published before and during the pandemic, for restaurants in two large metropolitan areas, namely, New York City and Los Angeles County. We measure changes in user activity, ratings, and restaurant type preferences using the corresponding metadata, and quantify changes in written text using relevant extraction and classification techniques.

Overall, we make the following contributions.

Creation of a dataset with fine-grained COVID-19 aspect annotations. To facilitate text analysis, we create a dataset of 600 Yelp restaurant reviews with manual annotations of fine-grained COVID-19 aspects discussed in the reviews, such as hygiene practices, concerns of virus transmission, and sympathy and support messages. Our annotations can support detailed review analyses beyond simple mentions of COVID in text.²

Evaluation of COVID-19 aspect extraction techniques. We use our dataset to evaluate several techniques for COVID-19 aspect extraction from the review text, including unsupervised topic modeling (Blei et al., 2003), weakly-supervised classification based on COVID-related keywords (Karamanolakis et al., 2019), and (fully) supervised classification.

Analysis of the correlation between Yelp reviews and critical COVID-19 statistics over time. We analyze the distribution and evolution of the extracted COVID-19 aspects and other re-

view metadata over time, capturing the period before and during the pandemic. We observe revealing trends, such as increased interest in fast food restaurants compared to traditional American-food restaurants (including brunch restaurants), increased mentions of hygienic practices of restaurants (Figure 1), service changes, racist and xenophobic attacks against the Asian American community, and sympathy and support messages expressed especially for local businesses. Crucially, we show that the resulting time series have substantial correlation (Spearman’s $\rho=0.84$, $p<0.01$) with critical statistics and milestones related to the pandemic, such as the number of COVID-19 cases in the U.S. While our findings do not necessarily imply that the observed trends are caused by the pandemic, they may provide useful insights for restaurant owners, customers, public health officials, and the broad research community.

This paper is organized as follows. Section 2 reviews related work. Section 3 describes the Yelp data collection and annotation procedures. Section 4 outlines the techniques for data analysis. Section 5 summarizes our findings. Finally, Section 6 concludes the paper and suggests future work.

2 Related Work

The natural language processing community has been increasingly pushing efforts towards the better understanding and management of the pandemic. Valuable insight can be extracted from text data, including the COVID-19 scientific literature (Wang et al., 2020; Gutierrez et al., 2020), and web search data (Effenberger et al., 2020; Rovetta and Bhagavathula, 2020). Below, we review related work on the analysis of online user-generated reviews and posts on social media.

Social media reflects public attitudes during the pandemic (Chen et al., 2020). Existing work on sentiment or emotion analysis has considered Twitter (Drias and Drias, 2020; Nemes and Kiss, 2020; Li et al., 2020a; Samuel et al., 2020), Reddit (Biester et al., 2020), Weibo (Li et al., 2020b), and other platforms (Kleinberg et al., 2020). For example, Biester et al. (2020) analyzed how the pandemic has influenced the online behavior of Reddit users and found an increase in posts expressing mental health concerns, including anxiety and concerns for health and family. Beyond sentiment analysis, existing work has considered deep learning techniques for the identification of informative

²Our annotations are available at the following link: <https://drive.google.com/drive/folders/1PwYGO68fdJpJrGkN6rry-P9ji570Ia-r>.

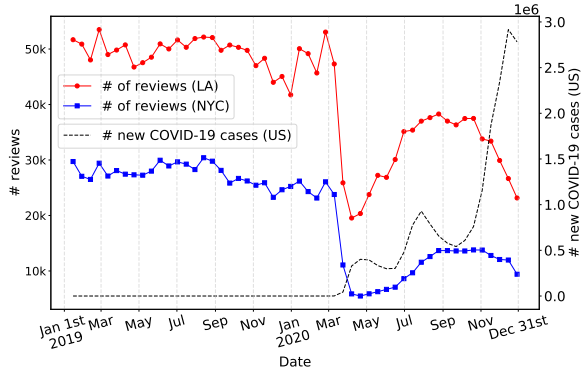


Figure 2: Total number of reviews for New York City and Los Angeles County over January 1, 2019 - December 31, 2020.

tweets that contain information relevant to the pandemic (Nguyen et al., 2020; Laxmi et al., 2020; Verspoor et al., 2020). To our knowledge, our work is the first that analyzes the effects of COVID-19 on restaurant reviews. To perform this analysis, we extract fine-grained COVID-19 aspects related to restaurants (e.g., hygiene practices, sympathy and support, social distancing, etc.).

Other work has studied nutrition during the pandemic by conducting surveys (Di Renzo et al., 2020) or by analyzing Twitter (Van et al., 2020). Van et al. (2020) observe a shift from mentions of healthy to unhealthy foods. Naja and Hamadeh (2020) propose a framework for action to maintain optimal nutrition during the pandemic. As part of our work, we show trends in restaurant preferences, such as increased interest in fast food restaurants compared to traditional American-food restaurants.

While prior work demonstrates changes in public attitude and nutrition during the pandemic, it is not clear how and to what extent restaurant reviews have changed during the pandemic. Yelp has introduced special COVID-19 review guidelines, and subsequently removed more than 4,000 reviews that violated those guidelines.³ Our work demonstrates that many aspects of the review content and metadata have changed during the pandemic.

3 Data

We now describe our procedure for Yelp data collection (Section 3.1) and COVID-19 aspect annotation (Section 3.2).

³<https://blog.yelp.com/2021/01/yelp-will-display-user-feedback-on-health-and-safety-practices>

Aspect	Star Rating					ALL
	1	2	3	4	5	
Hygiene	103	21	16	25	78	243
Non-COVID	39	13	14	28	117	211
Service	21	4	8	9	41	83
Social Distancing	9	2	8	8	40	67
Sympathy & Support	8	1	3	1	28	41
Transmission	26	6	4	1	2	39
Racism	30	1	0	0	2	33
Other	14	1	3	0	3	21

Table 1: Aspect- and rating-related statistics for the 600 labeled Yelp reviews: ALL reports the number of reviews for each COVID aspect; the other columns report the number of reviews for the different star ratings. Out of the 600 reviews, 81 reviews were annotated with more than one aspect.

3.1 Yelp Data Collection

We consider Yelp reviews for New York City (NYC) and Los Angeles County (LA) restaurants uploaded over January 1, 2019 - December 31, 2020. Our dataset overall consists of 1 million reviews for NYC and 2.1 million reviews for LA.

Figure 2 plots the number of reviews across time as well as the number of new COVID-19 cases in the U.S. For both NYC and LA, the number of reviews decreases significantly after January 2020, especially in March and April 2020: shutdowns and more stringent guidelines were put into effect starting in March. Such restrictions were only lifted in July 2020 and a second peak in the number of reviews is observed during September 2020.

3.2 COVID-19 Aspect Annotation

We manually labeled 600 reviews published after March 2020 with annotations relevant to COVID-19. In particular, we aimed to understand what aspects of restaurant operations are discussed in reviews referring to the pandemic. We will use these labels in Section 4.1 to train and evaluate classifiers for COVID aspect detection. For annotation, we considered 600 Yelp reviews posted after March 1, 2020, selected as follows. First, we considered all reviews after March 1, 2020 that contain COVID-related keywords⁴ and selected 400 reviews uniformly at random among them. Second, we selected 200 reviews uniformly at random from

⁴We consider the following COVID-related keywords (case insensitive; adopted from Biester et al. (2020)): “corona,” “outbreak,” “pandemic,” “virus,” “sars-cov-2,” “coronavirus,” “wuhan,” “2019ncov,” “2019-ncov,” “wufu,” “covid-19,” “covid19,” “covid,” “sars,” and “mers.”

all reviews after March 1, 2020 that do not contain such keywords. We considered the following aspects related to COVID-19:

1. Hygiene: hygiene conditions of restaurants and protective equipment (e.g., *“Just know there’s no restroom or sink for patrons to wash their hands. They do have hand sanitizers and wipes, but personally I prefer washing my hands.”*).
2. Transmission: concern of virus transmission (e.g., *“All the whole coughing without covering his mouth”*).
3. Social Distancing: social distancing measures (e.g., *“The tables are set far apart – a more than acceptable social distance”*).
4. Racism: racism experiences (e.g., *“She was the only one waiting at the register but no one came to ring her up. She waited for a while but decided to leave after realizing she was ignored because of her race.”*).
5. Sympathy and Support: messages of solidarity, for example, towards local businesses (e.g., *“Help support your Chinatown restaurants who are deeply hurting from the stigma around corona virus.”*).
6. Service: service changes during the pandemic (e.g., *“Not sure if the restaurant was empty because of the coronavirus scare but the food came out suuuuper fast...”*).
7. Other: aspects that are related to COVID but that do not fall under any of the above categories (e.g., *“Shame on management for taking advantage of people trying to keep safe from coronavirus during a NY state of emergency.”*).

We annotated each review with a COVID-related aspect if at least a sentence of the review discusses such aspect. A single review can be annotated with more than one distinct aspect. In cases where a review did not contain any sentences that were deemed relevant to any of the seven COVID-related aspects, then it received the “Non-COVID” aspect.

Table 1 shows annotation statistics. (We discuss review ratings later.) Most reviews discuss hygiene conditions of restaurants, and many reviews discuss social distance measures as well as changes in the restaurant service related to COVID.

4 Methodology

We now describe the techniques that we apply to the 3.1 million Yelp reviews from Section 3 for

COVID-19 aspect analysis (Section 4.1) and time series analysis (Section 4.2), leveraging the labeled reviews of Section 3.2.

4.1 COVID-19 Aspect Analysis

First, we extract topics from reviews using unsupervised topic modeling. We train Latent Dirichlet Allocation (LDA) topic models (Blei et al., 2003) with different numbers of topics (5, 10, 25, 50, 100). Then, we manually annotate the obtained topics with descriptive labels by examining the highest-probability words for each topic. We noticed that it is hard to align the topics discovered by LDA with the COVID aspects of interest (Section 3.2) and, therefore, we experiment with supervised and weakly-supervised techniques, as discussed next.

We use our annotated dataset from Section 3.2 to train and evaluate review classifiers (via 5-fold cross-validation) for multi-class COVID-19 aspect classification. We consider two alternative training procedures: fully-supervised classification using labeled training data, and weakly-supervised classification using a small number of indicative keywords per class. The fully-supervised approaches are standard and listed at the end of this subsection.

The weakly-supervised approach we use is the co-training method of Karamanolakis et al. (2019), which works as follows. First, we manually define a small number of keywords or key phrases for each COVID-19 aspect.⁵ Then, we employ a teacher-student architecture, where the teacher classifier considers keywords to annotate unlabeled reviews with aspects and the teacher-labeled reviews are used to train a student classifier. The teacher classifier does not require training and instead predicts aspect probabilities proportionally to keyword counts for each aspect. If no keywords appear in a review, then the teacher predicts the “Non-COVID” aspect. The student classifier can be any classifier, and here we consider both stan-

⁵ **Hygiene:** “masks,” “gloves,” “mask,” “glove,” “shield,” “sanitize,” “sanitizer,” “sanitizing,” “disinfect,” “disinfecting,” “face cover,” “covering face,” “face covers,” “wipe,” “wiping,” and “wipes.” **Transmission:** “cough,” “spread,” “infected,” “cautious,” “potential germs,” “concerning,” “worried,” “covid test,” “tested positive,” and “asymptomatic.” **Social Distancing:** “social distance,” “social distancing,” “six feet,” “6 feet,” “spaced out,” “6ft,” and “distanced.” **Racism:** “racist,” “xenophobia,” “racism,” “race,” “xenophobic,” “asian,” and “asians.” **Sympathy and Support:** “small business,” “local business,” “struggling,” “support,” “stress,” “stressful,” “suffer,” “sympathy,” and “stressed.” **Service:** “takeout,” “outdoor dining,” “take out,” “re-stocked,” “restocked,” “curbside pickup,” “on-line order,” “rude,” and “service.” **Other:** “covid,” “pandemic,” “quarantine,” “covid19,” “lockdown,” “shutdown,” and “cdc.”

standard bag-of-words classifiers and classifiers based on pre-trained BERT representations (Devlin et al., 2019). Note that the student is trained using the teacher’s predictions and no manually annotated reviews; in contrast to the teacher, which only considers keywords, the student can identify aspects even if no keywords appear in a review. As labeled data are expensive to obtain, such a weakly-supervised technique is promising to scale classification by leveraging unlabeled reviews (and keywords) for training (Karamanolakis et al., 2019).

Overall, we consider the following approaches:

1. Random: assigns reviews to a random aspect.
2. Majority: assigns all reviews to the “Non-COVID” aspect.
3. Supervised bag-of-words (BoW) classifiers: represents each review as a bag of words, where words can be unigrams and bigrams. We evaluate logistic regression (LogReg) and Support Vector Machines (SVM).
4. Supervised BERT: fine-tunes pre-trained BERT (Devlin et al., 2019) for supervised aspect classification.
5. Weakly-supervised Teacher: classifies a review solely based on keywords (Teacher in Karamanolakis et al. (2019)).
6. Weakly-supervised Student: is trained using Teacher’s predictions on unlabeled data (Student in Karamanolakis et al. (2019)). We evaluate different modeling approaches for Student, namely, BoW-LogReg, BoW-SVM, and BERT.

The above techniques classify Yelp reviews into COVID aspects using either labeled data (supervised approach) or COVID-related keywords (weakly-supervised approach) for training. In addition to COVID aspect classification, we conduct time series analysis to understand how COVID aspects evolve over time, as discussed next.

4.2 Time Series Analysis

To understand how reviews have changed during the pandemic, we extract time series from the text of the reviews. For a given aspect (e.g., Hygiene), the corresponding time series is computed as the percentage of the reviews at each point in time that contain at least one aspect-specific keyword (see Section 4.1). We consider two approaches: time-series cross-correlation and time-series intervention analysis, as discussed next.

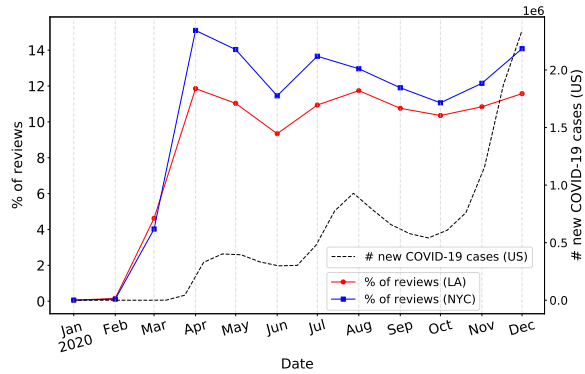


Figure 3: Percentage of reviews containing COVID-related keywords.

As a first approach, we measure the correlation between the Yelp review time series and important statistics related to COVID-19, such as the number of new COVID-19 cases in the U.S or the new COVID-19 cases in NYC and LA individually. As we do not expect Yelp review time series to have a linear relationship with COVID-19 time series, we compute the Spearman’s correlation metric, which only assumes a monotonic but possibly non-linear relationship between the two time series. We also measure the Pearson’s correlation metric as a robustness check.

As a second approach, we consider a time series intervention analysis. First, we train a time-series model on the observations before COVID-19 (i.e., on reviews posted before March 1, 2020) and then we compare the model’s predictions against the observations during COVID-19 (i.e., on reviews posted on March 1, 2020 or later). Similar to Biester et al. (2020), we consider the Prophet time-series forecasting model (Taylor and Letham, 2018), an additive regression model that has been shown to forecast social media time series effectively. After training Prophet on the pre-pandemic data, we check to what degree its forecasts for during COVID-19 differed from the actual values. Specifically, we compute the proportion of observations outside the 95% prediction uncertainty interval produced by Prophet after March 1, 2020.

By constructing Yelp review time series and comparing them to statistics related to COVID-19, we find interesting trends in reviews during the pandemic, as discussed next.

5 Findings

We use the methodology from Section 4 to address various questions on the 3.1 million-review

Topic label (manually assigned)	10 highest-probability words
Protective equipment and social distancing	covid, mask, masks, people, customers, staff, social, wearing, distancing, pandemic
Outdoor seating	outdoor, seating, dining, ramen, good, tables, covid, outside, really, place

Table 2: The two (out of 25 LDA) topics that we identified as relevant to COVID-19, along with manually assigned topic labels and the 10 highest probability words for each topic. (All topics are reported in Table 7 in the Appendix.)

Method	Binary F1	Multi-Class F1
Random	0.459	0.115
Majority	0.346	0.070
<i>Methods below are fully supervised</i>		
BoW-LogReg	0.741	0.481
BoW-SVM	0.739	0.422
BERT	0.786	0.522
<i>Methods below are weakly supervised</i>		
Teacher	0.605	0.270
Student-BERT	0.657	0.407

Table 3: F1 values for binary (left) and multi-class (right) COVID-19 aspect classification. Classifiers based on pre-trained BERT representations outperform simpler bag-of-words classifiers.

dataset from Section 3. First, we analyze the text of the Yelp reviews (Section 5.1), and then we use both the metadata and the text to create time series and evaluate their correlation with the number of COVID-19 cases (Section 5.2).

5.1 COVID-19 Aspect Analysis

In this section, we analyze the text of the Yelp reviews and evaluate the performance of several methods for COVID aspect classification on our manually annotated dataset from Section 3.2.

Number of reviews with COVID-related keywords: Figure 3 shows the percentage of reviews that contain COVID-related keywords. Interestingly, after March 2020, more than 10% of the reviews contain COVID-related keywords: thousands of restaurant reviews per week discuss aspects related to the pandemic.

Topics discussed in reviews: We apply topic modeling on all Yelp reviews after March 1, 2020. Table 2 shows the two (out of the 25) topics that we identified as relevant to COVID-19. The first topic is related to protective equipment and social distancing, while the second topic is related to outdoor seating. The remaining 23 topics did not contain any COVID-related keywords among the 10 highest-probability words: it is hard to align the topics discovered by LDA with the fine-grained

COVID aspects of Section 3.2, so we consider aspect classification approaches, as discussed next.

COVID aspect classification: We evaluate supervised and weakly-supervised approaches for COVID aspect classification via cross-validation using the 600 manually annotated reviews (Section 4.1). Table 3 shows the cross-validation results for binary (COVID vs. Not COVID) and multi-class aspect classification. Table 8 in the Appendix reports additional metrics. The fully supervised BERT-based classifier outperforms BoW-* classifiers on both binary and multi-class classification. The weakly-supervised Teacher that classifies aspects using keywords and no labeled data (Section 4.1) leads to a more accurate Student-BERT classifier: weakly-supervised co-training with keywords leads to substantially better performance than Random. The weakly-supervised Student-BERT has lower F1 score than the fully supervised BERT, which was expected because Student-BERT does not consider labeled reviews for training but instead uses Teacher’s predictions on unlabeled reviews as weak supervision.

5.2 Time-series Construction

In this section, we analyze how reviews have changed during the pandemic by extracting time series from metadata (star ratings, cuisine types) and the text of the reviews (see Section 4.2).

Star ratings: Figure 4 shows the average star rating over time for NYC and LA. For both time series, there is a sharp decrease in average rating starting in March 2020 and an increasing trend after June 2020. Figure 5 shows the number of star ratings across time for NYC. The trends are similar for LA (Figure 9b in the Appendix). For the first time after 2019, the percentage of 1-star ratings in NYC surpassed the percentage of 4-star ratings. Interestingly, for both NYC and LA, a peak in the number of new COVID-19 cases (April 2020 for NYC and July 2020 for LA) coincides with a peak in the percentage of 1-star ratings. Also, after September 2020, there is a decreasing trend in the number of 1-star ratings and an increasing trend in the number

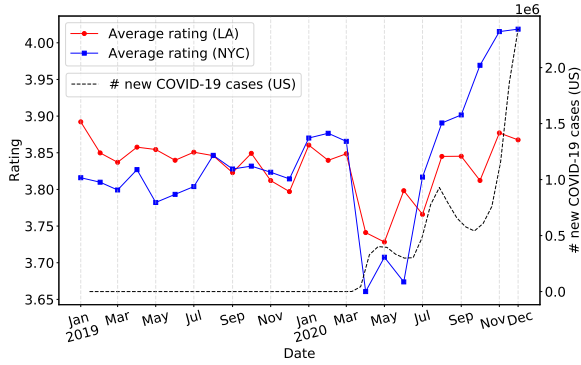


Figure 4: Average star rating across all reviews in NYC and LA over January 1, 2019 - December 31, 2020.

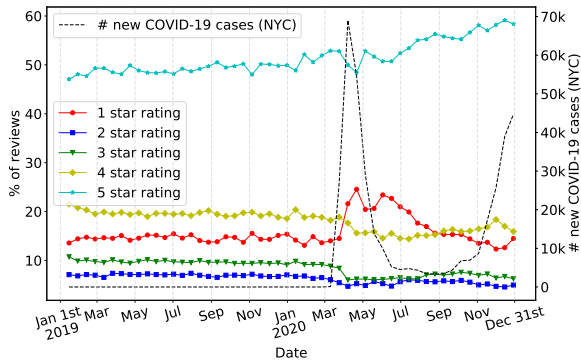


Figure 5: Percentage of reviews for each star rating in NYC over January 1, 2019 - December 31, 2020.

of 5-star ratings. We conclude that, during the first months of the pandemic, users’ ratings shifted to extremely positive (5 star) or extremely negative (1 star) values, but after September 2020, users posted increasingly more 5-star rating reviews, leading to a total increase in average rating.

Types of cuisine: Restaurant metadata include tags that indicate the cuisine types, such as “Italian” and “sandwich.” Figure 6 shows the percentage of reviews for selected groups of cuisine type over time. Such time series are relatively stable during 2019 but change significantly during 2020. “American” substantially dropped at the beginning of the pandemic (March) and rose again after indoor dining re-opened (July). The drop in “American” coincided with the increase of “Fast Food.” “Asian Food” also dropped sharply in March but recovered quickly within 2 weeks. These trends indicate important changes in user activity during the pandemic that affect specific cuisine types, which could be supported by previous observations of nutrition changes (Van et al., 2020).

Time Series (NYC)	NYC Cases	US Cases
Social Distancing	0.768***	0.836***
Hygiene	0.765***	0.822***
Transmission	0.816***	0.804***
Sympathy & Support	0.822***	0.755***
Service	0.772***	0.736***
Racism	0.293**	0.237*
Time Series (LA)	LA Cases	US Cases
Service	0.536***	0.644***
Sympathy & Support	0.490***	0.551***
Hygiene	0.395***	0.538***
Transmission	0.409***	0.522***
Social Distancing	0.347**	0.513***
Racism	-0.006	-0.019

Table 4: Spearman correlation results from comparing COVID aspects and the number of COVID cases in NYC (top) and LA (bottom), sorted in decreasing order by correlation compared with the number of new US cases. Results are marked as statistically significant at the $p < 0.1^*$, $p < 0.05^{**}$, and $p < 0.01^{***}$ levels.

Evolution of restaurant review aspects over time: Figure 7 shows the evolution of aspects over time for NYC. Aspects for LA reviews follow similar trends (see Table 12 in the Appendix). Aspects such as “Hygiene,” and “Social Distancing” have been discussed more frequently after March 2020, covering up to 8% of the restaurant reviews: reviewers discuss such aspects during the pandemic more than before the pandemic. Interestingly, while “Hygiene” peaked during July 2020 (during restaurant re-opening) for both cities and since then keeps decreasing, “Sympathy & Support” peaked during Spring 2020, then decreased, and follows an increasing trend after November 2020.

Correlation of aspects with COVID-19 statistics: We now consider our first approach for time series analysis from Section 4.2 and measure the correlation between Yelp review time series and COVID-19 statistics. Table 4 reports the Spearman correlation between time series constructed from restaurant reviews and the number of new COVID-19 cases. For Pearson correlation results, see Tables 9 and 10 in the Appendix. For both NYC and LA, there is significant correlation between restaurant review aspects and new cases of COVID-19, reaching up to Spearman’s $\rho = 0.84$ for the Hygiene aspect. For LA, COVID aspects have higher absolute correlation to the number of US

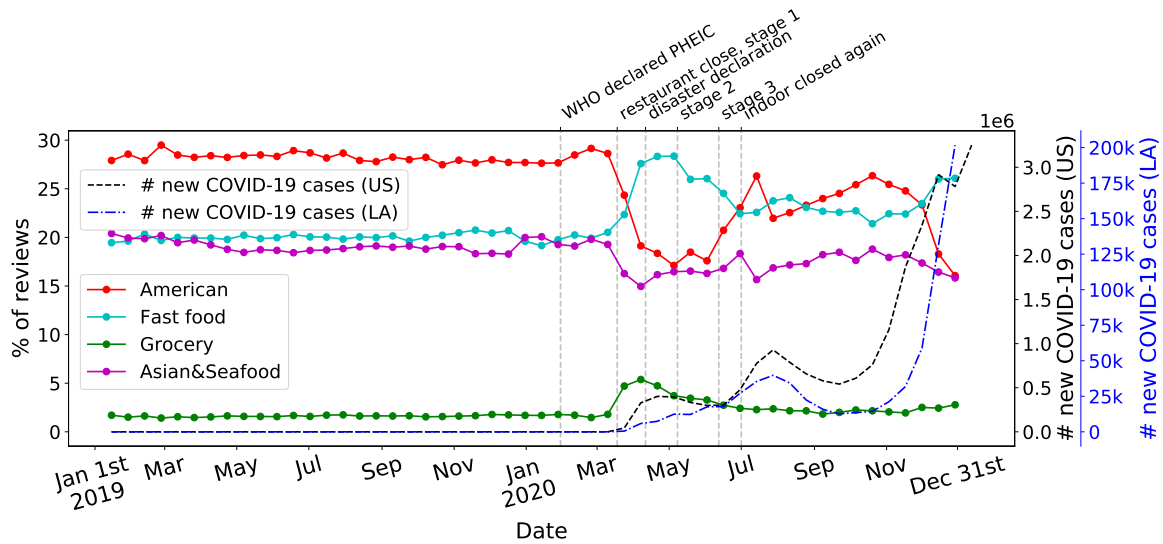


Figure 6: Evolution of cuisine types over time for LA. For each time series, we compute the percentage of reviews that include at least one tag from a predefined tag list: “American”: [“steak”, “cocktailbars”, “bars”, “breakfast brunch”, “newamerican”, “tradamerican”], “Fast Food”: [“sandwiches”, “pizza”, “hotdogs”, “chicken wings”, “thai”], “Groceries”: [“grocery”], “Deserts&Drinks”: [“juicebars,” “bubbletea,” “icecream,” “desserts,” “bakeries”], “Asian&Seafood”: [“sushi”, “japanese”, “seafood”, “asianfusion”, “korean”]. Tags within each category follow similar trends, which we individually report in the Appendix.

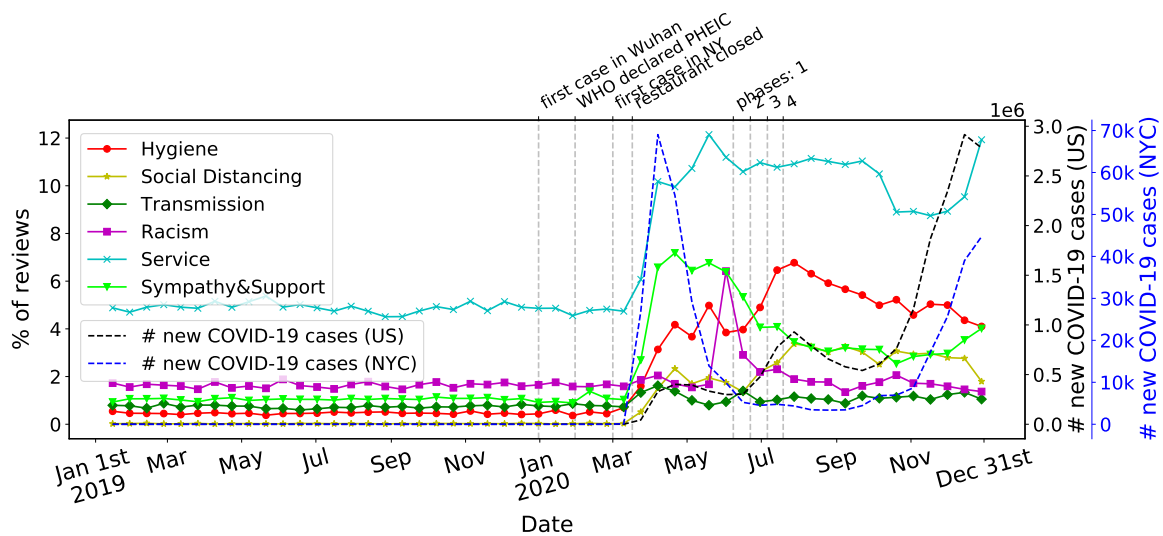


Figure 7: COVID aspects for NYC restaurants over January 1, 2019 - December 31, 2020.

cases compared to the number of LA cases. For NYC, most aspects present higher correlation with the number of NYC cases compared to the number of US cases. Even though we cannot draw causal conclusions from these correlations, our results highlight interesting trends of Yelp reviews during the pandemic.

Time-series intervention analysis: Here, we consider our second approach for time series analysis (Section 4.2) and compare time series constructed from the metadata of Yelp reviews to the

corresponding Prophet forecasts. Figure 8 shows the evolution of the “pizza” tag (left) and “seafood” tag (right) over time and the Prophet forecasts. During COVID-19 (i.e., on March 1, 2020 or later), most true values for “pizza” were higher than forecasts while most true values for “seafood” were lower than Prophet forecasts. The Appendix reports forecasts for more cuisine tags. The difference between Prophet’s forecasts and true values indicates that user activity has shifted towards specific types of businesses, as we further discuss next.

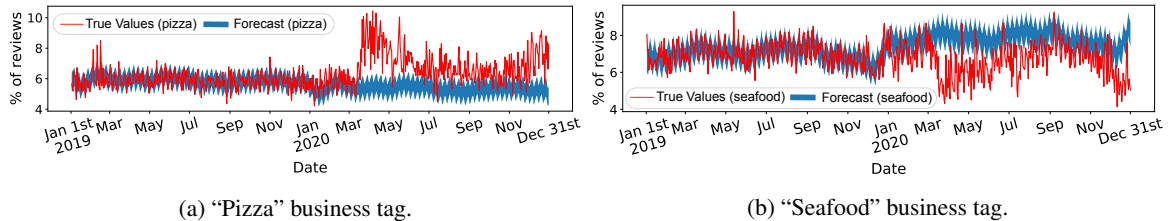


Figure 8: Evolution of business tags and the corresponding Prophet forecasts over time for LA. The red line is the true value and the blue line is the Prophet forecast. After March 1, 2020, most true values for “pizza” restaurants (left) were higher than Prophet forecasts, while most true values for “seafood” restaurants (right) were lower than Prophet forecasts. The Appendix reports all forecasts of Prophet.

Time Series	% of outliers	
	LA	NYC
1 star rating	62.28 ↑	69.55 ↑
2 star rating	55.36 ↓	91.35 ↓
3 star rating	83.04 ↓	47.75 ↓
4 star rating	61.94 ↓	61.24 ↓
5 star rating	88.24 ↑	50.17 ↑
Grocery	82.35 ↑	96.54 ↑
Chicken Wings	64.36 ↑	92.73 ↑
Sandwiches	95.50 ↑	75.78 ↑
Thai	69.20 ↑	68.86 ↑
Bakeries	47.06 ↑	66.78 ↑
Hotdogs	77.85 ↑	65.05 ↑
Pizza	89.96 ↑	56.75 ↑
Ice Cream	71.28 ↑	56.40 ↑
Breakfast&Brunch	81.31 ↓	53.98 ↓
Sushi	41.52 ↓	55.01 ↓
Steak	84.78 ↓	58.48 ↓
Cocktail Bars	99.31 ↓	58.82 ↓
Trad American	93.08 ↓	58.82 ↓
Bars	69.90 ↓	59.86 ↓
Japanese	40.83 ↓	61.24 ↓
Asian Fusion	41.87 ↓	76.47 ↓
New American	89.62 ↓	91.70 ↓

Table 5: Percentage of outliers (observations outside Prophet’s 95% uncertainty interval) for LA and NYC reviews posted after March 1, 2020. Arrows indicate whether the mean value of the outliers is higher (up) or lower (down) than the mean of Prophet’s predictions.

Table 5 reports the percentage of outliers (i.e., true values outside of Prophet’s 95% uncertainty interval) for star ratings (top) and some of the most frequent business tags (bottom). For tags such as “Grocery,” “Chicken Wings,” and “Sandwiches,” upwards pointing arrows indicate that the mean value of outliers is higher than the mean of

Prophet’s predictions. In contrast, for tags such as “New American,” “Asian Fusion,” and “Japanese,” downwards pointing arrows indicate that the mean value of outliers is lower than the mean of Prophet’s predictions. The Appendix reports all forecasts of Prophet. The direction arrows in Table 5 support our previous observations about the corresponding changes of cuisine types and star ratings during the pandemic.

6 Conclusions and Future Work

We presented our effort to understand the effects of COVID-19 on restaurant reviews. We created a dataset with fine-grained COVID-19 aspect annotations, evaluated fully- and weakly-supervised techniques for COVID aspect detection, and showed that BERT-based classifiers outperform bag-of-words classifiers. We observed changes in restaurant reviews (e.g., increased discussions of hygiene practices and messages of solidarity), and showed that they correlate with critical COVID-19 statistics. We found a shift of ratings towards extreme values (1 and 5 stars) and shifts of user activity towards specific types of cuisines. Our insights could potentially be interesting for restaurant owners, customers, and public health officials.

In future work, we plan to expand the regional coverage of our analysis to reveal distinct patterns across cities. It would also be interesting to improve aspect-based sentiment analysis approaches (Pontiki et al., 2016) by considering the new aspects explored in this work.

Acknowledgments

We thank the anonymous reviewers for their constructive feedback. This material is based upon work supported by the National Science Foundation under Grant No. IIS-15-63785.

References

- Laura Biester, Katie Matton, Janarthanan Rajendran, Emily Mower Provost, and Rada Mihalcea. 2020. Quantifying the effects of covid-19 on mental health support forums. In *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020*.
- David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan):993–1022.
- Emily Chen, Kristina Lerman, and Emilio Ferrara. 2020. Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus Twitter data set. *JMIR Public Health and Surveillance*, 6(2):e19273.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT (1)*.
- Laura Di Renzo, Paola Gualtieri, Francesca Pivari, Laura Soldati, Alda Attinà, Giulia Cinelli, Claudia Leggeri, Giovanna Caparello, Luigi Barrea, Francesco Scerbo, et al. 2020. Eating habits and lifestyle changes during covid-19 lockdown: an Italian survey. *Journal of translational medicine*, 18:1–15.
- Habiba H Drias and Yassine Drias. 2020. Mining Twitter data on covid-19 for sentiment analysis and frequent patterns discovery. *medRxiv*.
- Maria Effenberger, Andreas Kronbichler, Jae Il Shin, Gert Mayer, Herbert Tilg, and Paul Perco. 2020. Association of the covid-19 pandemic with internet search volumes: a google trends™ analysis. *International Journal of Infectious Diseases*, 95:192–197.
- Bernal Jimenez Gutierrez, Jucheng Zeng, Dongdong Zhang, Ping Zhang, and Yu Su. 2020. Document classification for covid-19 literature. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, pages 3715–3722.
- Giannis Karamanolakis, Daniel Hsu, and Luis Gra-vano. 2019. Leveraging just a few keywords for fine-grained aspect detection through weakly supervised co-training. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4603–4613.
- Bennett Kleinberg, Isabelle van der Vegt, and Maximilian Mozes. 2020. Measuring emotions in the covid-19 real world worry dataset. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*.
- Sirigireddy Dhana Laxmi, Rohit Agarwal, and Aman Sinha. 2020. DSC-IIT ISM at WNUT-2020 Task 2: Detection of covid-19 informative tweets using RoBERTa. In *Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020)*, pages 409–413.
- Irene Li, Yixin Li, Tianxiao Li, Sergio Alvarez-Napagao, Dario Garcia-Gasulla, and Toyotaro Suzumura. 2020a. What are we depressed about when we talk about covid-19: Mental health analysis on tweets using natural language processing. In *International Conference on Innovative Techniques and Applications of Artificial Intelligence*, pages 358–370. Springer.
- Xiaoya Li, Mingxin Zhou, Jiawei Wu, Arianna Yuan, Fei Wu, and Jiwei Li. 2020b. Analyzing covid-19 on online social media: Trends, sentiments and emotions. *arXiv preprint arXiv:2005.14464*.
- Farah Naja and Rena Hamadeh. 2020. Nutrition amid the covid-19 pandemic: A multi-level framework for action. *European Journal of Clinical Nutrition*, 74(8):1117–1121.
- László Nemes and Attila Kiss. 2020. Social media sentiment analysis based on covid-19. *Journal of Information and Telecommunication*, pages 1–15.
- Dat Quoc Nguyen, Thanh Vu, Afshin Rahimi, Mai Hoang Dao, Long Doan, et al. 2020. Wnut-2020 task 2: Identification of informative covid-19 English tweets. In *Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020)*, pages 314–318.
- Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad Al-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, et al. 2016. Semeval-2016 task 5: Aspect based sentiment analysis. In *International Workshop on Semantic Evaluation*, pages 19–30.
- Alessandro Rovetta and Akshaya Srikanth Bhagavathula. 2020. Global infodemiology of covid-19: analysis of Google web searches and Instagram hashtags. *Journal of medical Internet research*, 22(8):e20673.
- Jim Samuel, GG Ali, Md Rahman, Ek Esawi, Yana Samuel, et al. 2020. Covid-19 public sentiment insights and machine learning for tweets classification. *Information*, 11(6):314.
- Sean J Taylor and Benjamin Letham. 2018. Forecasting at scale. *The American Statistician*, 72(1):37–45.
- Hoang Van, Ahmad Musa, Mihai Surdeanu, and Stephen Kobourov. 2020. The language of food during the pandemic: Hints about the dietary effects of covid-19. *arXiv preprint arXiv:2010.07466*.
- Karin Verspoor, Kevin Bretonnel Cohen, Michael Conway, Berry de Bruijn, Mark Dredze, Rada Mihalcea, and Byron Wallace, editors. 2020. *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at*

EMNLP 2020. Association for Computational Linguistics, Online.

Lucy Lu Wang, Kyle Lo, Yoganand Chandrasekhar, Russell Reas, Jiangjiang Yang, Doug Burdick, Darrin Eide, Kathryn Funk, Yannis Katsis, Rodney Michael Kinney, et al. 2020. Cord-19: The covid-19 open research dataset. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*.

A Appendix

Here, we provide detailed information on our dataset (Section A.1), topic modeling and aspect classification results (Section A.2), time series plots (Section A.3), correlation analysis results A.4, and time-series intervention analysis results (Section A.5).

A.1 Yelp Review Dataset

Table 6 shows more statistics for our dataset. Our COVID aspect annotations for the 600 Yelp reviews are available at the following link: <https://drive.google.com/drive/folders/1PwYG068fDjppjRgKN6rry-P9ji570Ia-r>.

A.2 Topic Modeling and COVID Aspect Classification

Table 7 shows the 25 LDA topics obtained from all reviews posted after March 1, 2020. Table 8 reports detailed evaluation results for COVID aspect classification.

A.3 Time Series Plots

Star ratings: Figure 9 shows the number of star ratings across time for NYC and LA.

Cuisine types: Figure 10 shows the percentage of reviews in NYC (top) and (LA) over time for each selected group of cuisine types. Figure 11 shows the percentage of reviews over time for each individual business tag in our selected groups of cuisine types.

COVID-19 aspects: Figure 12 shows the percentage of reviews over time for each individual business tag in our selected groups of cuisine types.

A.4 Correlation Analysis

Tables 9 and 10 report correlation results between time series constructed from restaurant reviews and the number of new COVID-19 cases for NYC and LA, respectively. Tables 11 and 12 show correlation results between each individual business tag and the number of new COVID-19 cases for NYC and LA, respectively.

A.5 Time-series Intervention Analysis

Table 13 reports the percentage of outliers (according to Prophet’s predictions) for time series constructed from the text and metadata of Yelp reviews. Figures 13-74 plot Prophet’s forecasts for each individual time series.

	NYC	LA County
# Restaurants	55K	65K
# Users	344K	710K
# Reviews	1.0M	2.1M

Table 6: Statistics for our Yelp dataset collected during January 1, 2019 - December 31, 2020.

Topic Label	High Probability Words
Fast food	chicken, fries, burger, wings, sauce, fried, ordered, good, got, food
PPE and social distancing	covid, mask, masks, people, customers, staff, social, wearing, distancing, pandemic
Bakery	cake, chocolate, bakery, cookies, cakes, delicious, flavors, ve, sweet, best
Pizza and pie	pizza, slice, pie, best, good, crust, cheese, sauce, place, new
Service and Environment	great, place, food, staff, friendly, service, love, coffee, amazing, best
Mexican food	tacos, bagel, taco, bagels, mexican, good, chips, burrito, guacamole, delicious
Dinner meals	pasta, steak, good, ordered, sauce, delicious, dish, got, salad, dinner
Savory food	rice, pork, soup, noodles, chicken, thai, good, spicy, fried, beef
Sandwiches and salad	sandwich, cheese, bread, salad, good, egg, sandwiches, bowl, bacon, meat
Ice cream	cream, ice, ordered, like, cold, dessert, tasted, got, came, didn
Sushi	sushi, fish, roll, ordered, shrimp, fresh, good, salmon, rolls, food
Ordering	like, just, don, know, didn, place, people, want, said, order
Delivery service	order, delivery, food, ordered, time, ordering, pick, delicious, great, ve
Brunch	brunch, good, chocolate, eggs, toast, got, really, french, sweet, pancakes
Food quality	food, place, good, love, try, amazing, best, restaurant, really, delicious
Outdoor seating	outdoor, seating, dining, ramen, good, tables, covid, outside, really, place
Business	food, store, don, place, time, ve, years, money, just, business
Milk tea and boba	tea, milk, drink, sugar, drinks, sweet, like, bubble, boba, matcha
Indian and Korean cuisine	food, dishes, restaurant, like, dish, menu, meal, indian, korean, ve
Food and service quality	great, food, service, amazing, place, delicious, definitely, recommend, drinks, restaurant
Service wait time	food, table, came, minutes, wait, time, service, got, drinks, order
Food quality and price	food, good, prices, price, place, great, service, quality, pretty, nice
Service	order, said, told, service, asked, called, customer, manager, restaurant, food
Bars	free, promoter, door, issues, text, entry, drinks, table, vip, nyc
Bars	bar, wine, beer, card, like, bartender, night, drink, credit, place

Table 7: 25 LDA topics with manually assigned topic labels and the 10 highest probability words for each topic.

Method	Binary			Multi-Class		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Random	0.467	0.459	0.459	0.132	0.118	0.112
Majority	0.533	0.267	0.500	0.373	0.048	0.129
<i>Methods below are fully supervised</i>						
BoW-LogReg	0.742	0.741	0.742	0.622	0.501	0.462
BoW-SVM	0.737	0.737	0.740	0.605	0.486	0.419
BERT	0.787	0.785	0.786	0.652	0.542	0.503
<i>Methods below are weakly supervised</i>						
Teacher	0.560	0.610	0.600	0.360	0.334	0.268
Student-LogReg	0.545	0.636	0.511	0.393	0.320	0.389
Student-SVM	0.538	0.568	0.506	0.397	0.321	0.396
Student-BERT	0.603	0.767	0.574	0.472	0.383	0.434

Table 8: Results for binary (left) and multi-class (right) COVID-19 aspect classification.

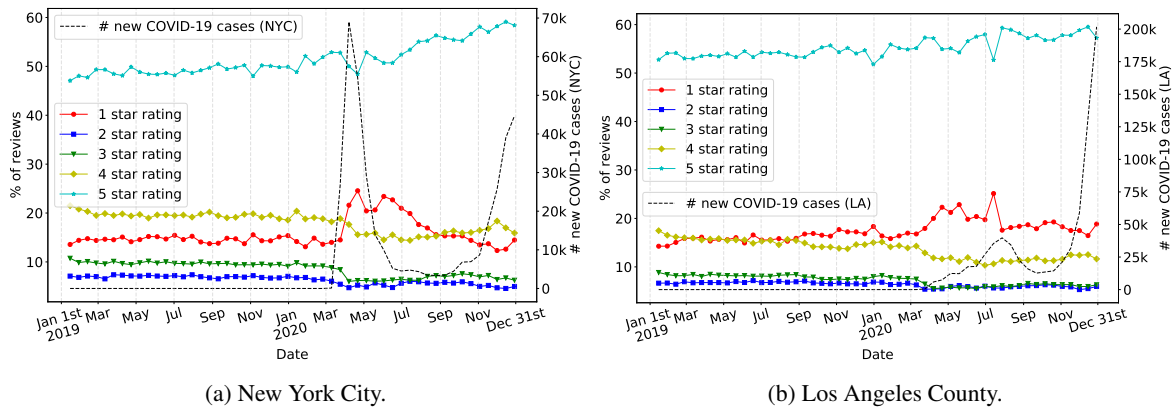


Figure 9: Percentage of reviews for each individual star rating in NYC (top) and LA (bottom) over January 1, 2019 - December 31, 2020.

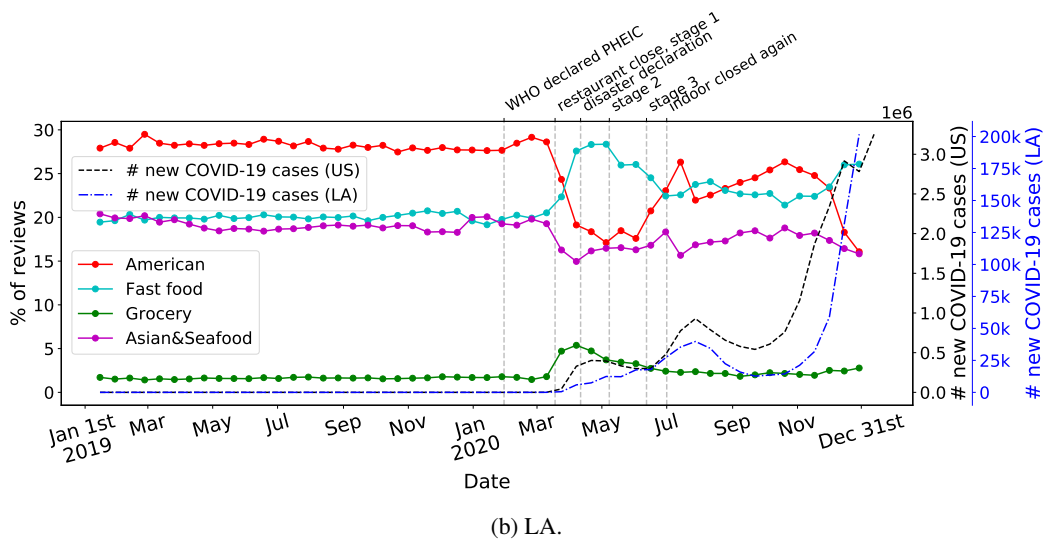
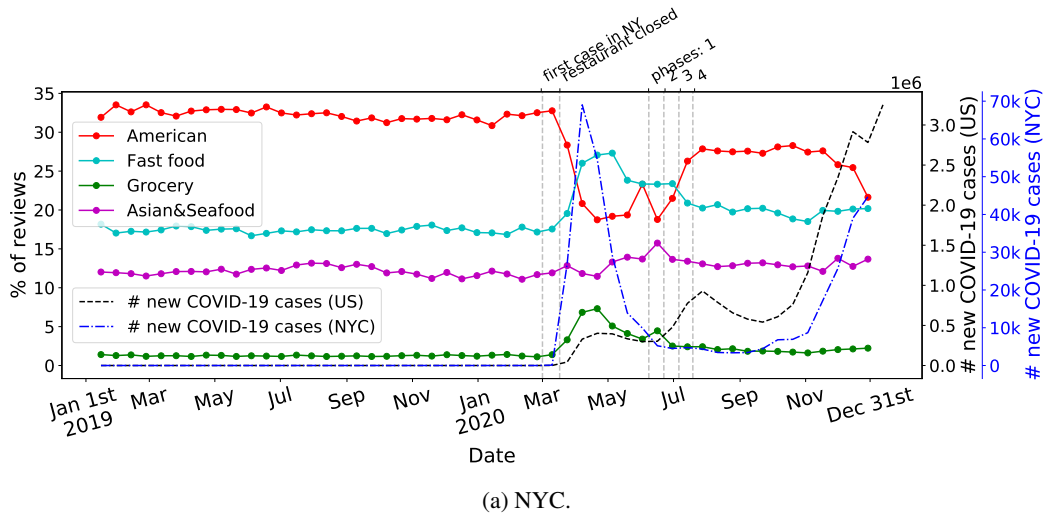
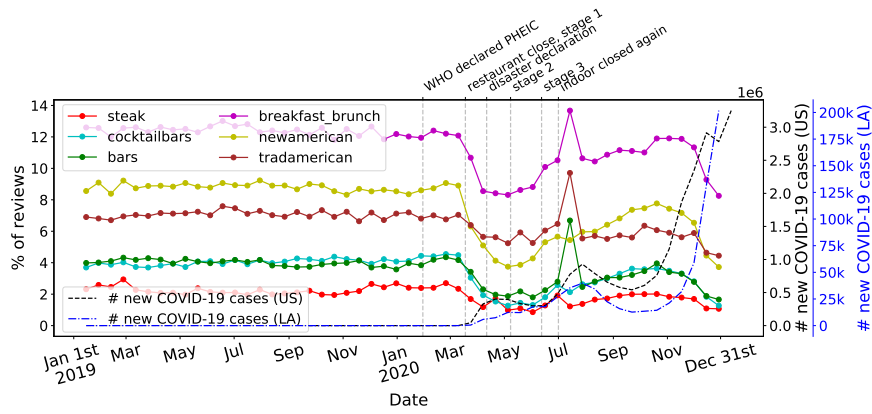
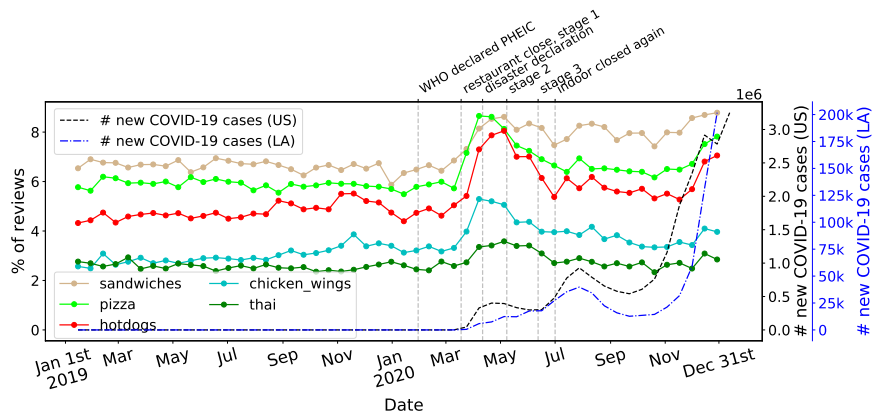


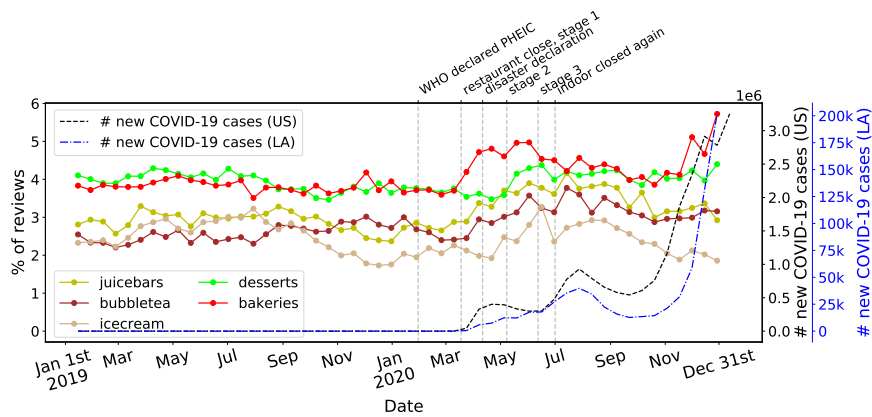
Figure 10: Evolution of cuisine types over time for NYC (top) and LA (bottom).



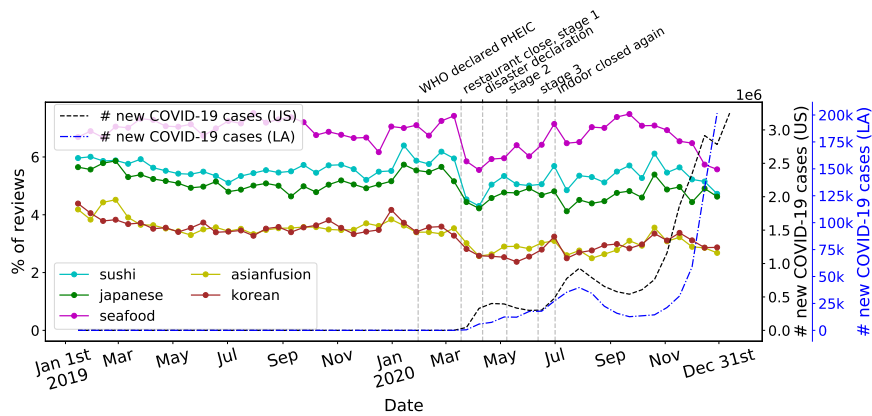
(a) American.



(b) Fast food.

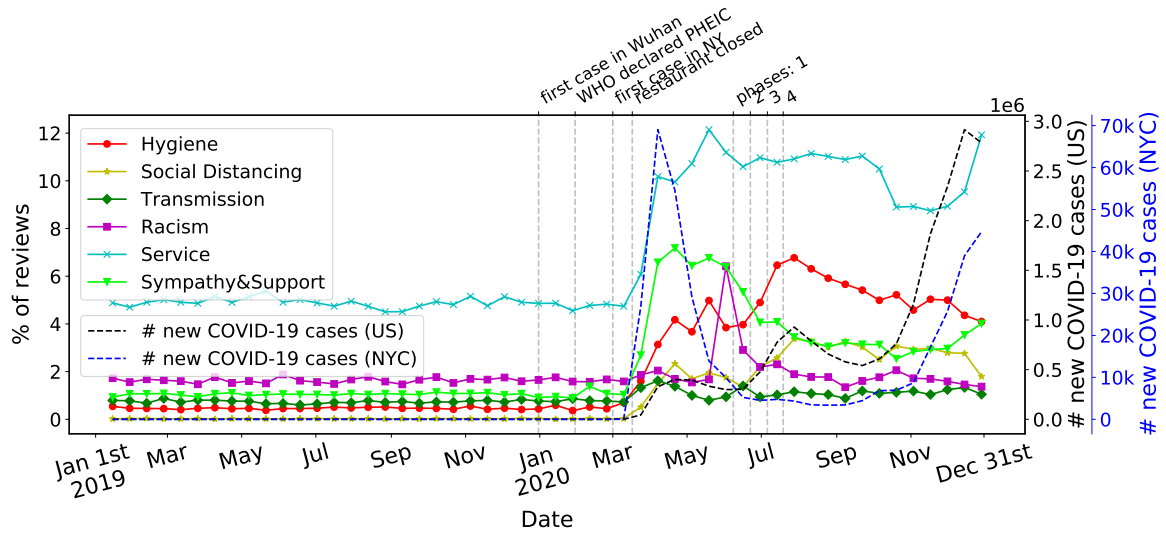


(c) Beverages & Desserts.

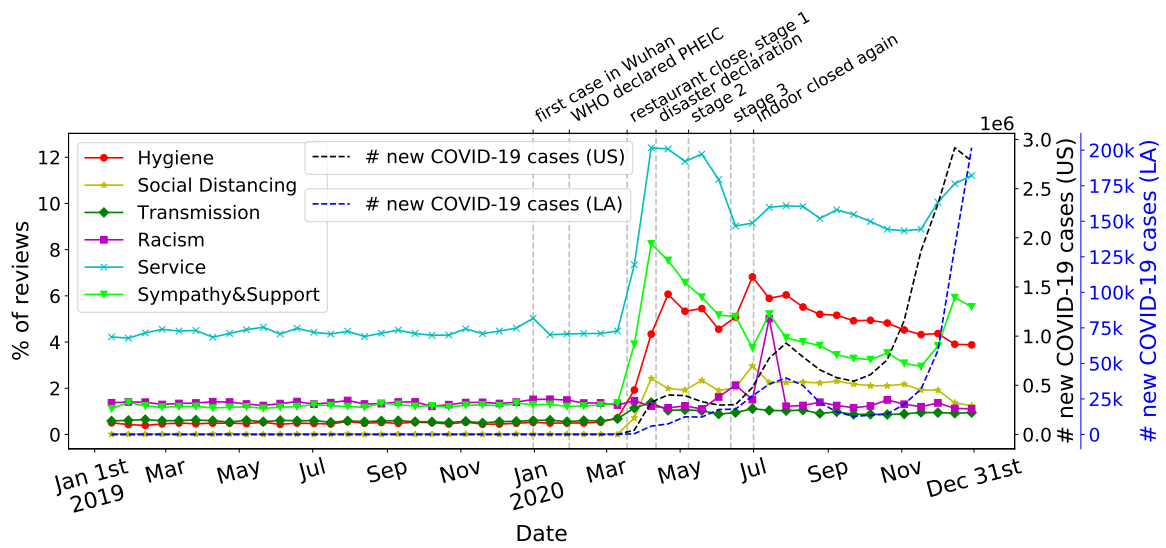


(d) Asian & Seafood.

Figure 11: Time-series of individual tags for each group of cuisine types defined in the main paper.



(a) New York City.



(b) Los Angeles County.

Figure 12: COVID aspects for New York City restaurants (top) and Los Angeles County restaurants (top) over January 1, 2019 - December 31, 2020.

Time Series	NYC Cases	US Cases
<i>Results below are Pearson correlations.</i>		
Hygiene	0.412***	0.612***
Transmission Case	0.708***	0.532***
Social Distancing	0.427***	0.650***
Racism	0.036	-0.025
Sympathy & Support	0.683***	0.407***
Service	0.566***	0.630***
Other	0.590***	0.652***
1 star rating	0.337**	-0.088
2 star rating	-0.724***	-0.712***
3 star rating	-0.624***	-0.619***
4 star rating	-0.384***	-0.476***
5 star rating	0.411***	0.813***
American	-0.731***	-0.541***
Fast Food	0.698***	0.341**
Grocery	0.713***	0.159
Beverages & Desserts	0.221	0.426***
Asian & Seafood	-0.594***	-0.083
<i>Results below are Spearman correlations.</i>		
Hygiene	0.765***	0.822***
Transmission Case	0.816***	0.804***
Social Distancing	0.768***	0.836***
Racism	0.293**	0.237*
Sympathy & Support	0.822***	0.755***
Service	0.772***	0.736***
Other	0.827***	0.808***
1 star rating	0.242***	0.118
2 star rating	-0.882***	-0.859***
3 star rating	-0.855***	-0.833***
4 star rating	-0.770***	-0.825***
5 star rating	0.705***	0.862***
American	-0.830***	-0.750***
Fast Food	0.832***	0.745***
Grocery	0.850***	0.751***
Beverages & Desserts	0.545***	0.524***
Asian & Seafood	-0.463***	-0.238*

Table 9: Correlation results from comparing NYC review statistics and the number of COVID cases in NYC (left) and US (right), sorted in decreasing order by correlation compared with the number of new US cases. Results marked as statistically significant at the $p < 0.1^*$, $p < 0.05^{**}$, and $p < 0.01^{***}$ levels.

Time Series	LA Cases	US Cases
<i>Results below are Pearson correlations.</i>		
Hygiene	0.395***	0.538***
Transmission Case	0.409***	0.522***
Social Distancing	0.347**	0.513***
Racism	-0.006	-0.019
Sympathy & Support	0.490***	0.551***
Service	0.536***	0.644***
Other	0.527***	0.671***
1 star rating	0.244*	0.289**
2 star rating	-0.416***	-0.571***
3 star rating	-0.429***	-0.559***
4 star rating	-0.402***	-0.513***
5 star rating	0.445***	0.608***
American	-0.654***	-0.656***
Fast Food	0.529***	0.577***
Grocery	0.214	0.251*
Beverages & Desserts	0.504***	0.524***
Asian & Seafood	-0.514***	-0.542***
<i>Results below are Spearman correlations.</i>		
Hygiene	0.814***	0.808***
Transmission Case	0.810***	0.779***
Social Distancing	0.773***	0.767***
Racism	-0.235*	-0.291**
Sympathy & Support	0.817***	0.788***
Service	0.812***	0.771***
Other	0.836***	0.811***
1 star rating	0.724***	0.705***
2 star rating	-0.831***	-0.839***
3 star rating	-0.811***	-0.798***
4 star rating	-0.818***	-0.822***
5 star rating	0.730***	0.762***
American	-0.804***	-0.753***
Fast Food	0.822***	0.785***
Grocery	0.801***	0.784***
Beverages & Desserts	0.725***	0.607***
Asian & Seafood	-0.760***	-0.701***

Table 10: Correlation results from comparing LA review statistics and the number of COVID cases in NYC (left) and US (right), sorted in decreasing order by correlation compared with the number of new US cases. Results marked as statistically significant at the $p < 0.1^*$, $p < 0.05^{**}$, and $p < 0.01^{***}$ levels.

Time Series	NYC Cases	US Cases
<i>Results below are Pearson correlations.</i>		
Bakeries	0.684***	0.710***
Thai	0.445***	0.616***
Sandwiches	0.636***	0.466***
Sushi	-0.080	0.450***
Bubble tea	-0.142	0.277**
Grocery	0.713***	0.159***
Pizza	0.698***	0.156
Chicken Wings	0.596***	0.156
Hotdogs	0.518***	0.113
Korean	0.007	0.065
Juice Bars	0.001	0.051
Japanese	-0.427***	0.048
Desserts	-0.164	0.026
Ice Cream	-0.250*	-0.210
Cocktail Bars	-0.635***	-0.229
Asian Fusion	-0.534***	-0.340**
Steak	-0.495***	-0.347**
Seafood	-0.744***	-0.381***
Breakfast & Brunch	-0.667***	-0.469***
New American	-0.614***	-0.528***
Bars	-0.698***	-0.592***
Trad American	-0.639***	-0.658***
<i>Results below are Spearman correlations.</i>		
Thai	0.718***	0.790***
Bakeries	0.864***	0.785***
Grocery	0.850***	0.751***
Chicken Wings	0.798***	0.691***
Pizza	0.761***	0.658***
Hotdogs	0.571***	0.402***
Bubble tea	0.153	0.287**
Sushi	0.029	0.185
Juice Bars	0.203	0.127
Ice Cream	-0.112	-0.114
Seafood	-0.374***	-0.143
Desserts	-0.087	-0.145
Cocktail Bars	-0.438***	-0.179
Korean	-0.290**	-0.252*
Japanese	-0.394***	-0.290**
Asian Fusion	-0.678***	-0.566***
Steak	-0.739***	-0.614***
Bars	-0.778***	-0.730***
New American	-0.816***	-0.732***
Sandwiches	0.844***	-0.736***
Breakfast & Brunch	-0.842***	-0.811***
Trad American	-0.851***	-0.832***

Table 11: Correlation results for individual business tags in NYC. Results marked as statistically significant at the $p < 0.1^*$, $p < 0.05^{**}$, and $p < 0.01^{***}$ levels.

Time Series	LA Cases	US Cases
<i>Results below are Pearson correlations.</i>		
Sandwiches	0.644***	0.747***
Bakeries	0.702***	0.704***
Pizza	0.536***	0.557***
Hotdogs	0.500***	0.523***
Bubble tea	0.465***	0.519***
Desserts	0.414***	0.394***
Chicken Wings	0.340****	0.381***
Sushi	-0.390***	0.357***
Juice Bars	0.202	0.292**
Grocery	0.214***	0.251*
Thai	0.221***	0.225***
Ice Cream	-0.250*	-0.298**
Japanese	-0.309**	-0.378***
Korean	-0.362***	-0.423***
Bars	-0.451***	-0.456***
Breakfast & Brunch	-0.512***	-0.489***
Asian Fusion	-0.432***	-0.506***
Trad American	-0.485***	-0.535***
Seafood	-0.574***	-0.543***
Steak	-0.590***	-0.605***
Cocktail Bars	-0.620***	-0.612***
New American	-0.632***	-0.643***
<i>Results below are Spearman correlations.</i>		
Sandwiches	0.859***	0.807***
Grocery	0.801***	0.784***
Hotdogs	0.799***	0.766***
Pizza	0.779***	0.753***
Chicken Wings	0.743***	0.734***
Bubble tea	0.772***	0.714***
Bakeries	0.795***	0.714***
Juice Bars	0.685***	0.605***
Thai	0.513***	0.454***
Desserts	0.374***	0.267
Ice Cream	-0.118	-0.232*
Seafood	-0.424***	-0.383***
Sushi	-0.522***	-0.427***
Japanese	-0.680***	-0.602***
Bars	-0.699***	-0.645***
Breakfast & Brunch	-0.711***	-0.705***
Cocktail Bars	-0.803***	-0.706***
Steak	-0.793***	-0.708***
Korean	-0.794***	-0.763***
Trad American	-0.746***	-0.765***
New American	-0.810***	-0.767***
Asian Fusion	-0.801***	-0.816***

Table 12: Correlation results for individual business tags in LA. Results marked as statistically significant at the $p < 0.1^*$, $p < 0.05^{**}$, and $p < 0.01^{***}$ levels.

Time Series	% of outliers	
	LA	NYC
1 star rating	62.28 ↑	69.55 ↑
2 star rating	55.36 ↓	91.35 ↓
3 star rating	83.04 ↓	47.75 ↓
4 star rating	61.94 ↓	61.24 ↓
5 star rating	88.24 ↑	50.17 ↑
Fast food	95.85 ↑	96.88 ↑
Beverages&Desserts	51.56 ↑	83.04 ↑
Grocery	96.54 ↑	82.35 ↑
Asian&Seafood	52.94 ↓	87.89 ↓
American	76.47 ↓	92.73 ↓
Grocery	82.35 ↑	96.54↑
Chicken Wings	64.36↑	92.73↑
Sandwiches	95.50↑	75.78 ↑
Thai	69.20 ↑	68.86↑
Bakeries	47.06 ↑	66.78 ↑
Hotdogs	77.85↑	65.05↑
Pizza	89.96↑	56.75↑
Ice Cream	71.28↑	56.40 ↑
Seafood	77.51 ↓	53.28 ↑
Korean	56.40 ↓	32.87↑
Juice Bars	76.12 ↑	50.52 ↓
Desserts	80.62↑	51.90↓
Breakfast&Brunch	81.31↓	53.98 ↓
Sushi	41.52 ↓	55.01 ↓
Bubble tea	56.06 ↑	54.67 ↓
Steak	84.78↓	58.48↓
Cocktail Bars	99.31 ↓	58.82↓
Trad American	93.08 ↓	58.82↓
Bars	69.90 ↓	59.86↓
Japanese	40.83 ↓	61.24 ↓
Asian Fusion	41.87 ↓	76.47↓
New American	89.62 ↓	91.70↓

Table 13: Percentage of outliers (observations outside Prophet’s 95% uncertainty interval) for LA and NYC reviews posted after March 1, 2020. Arrows indicate whether the mean value of the outliers is higher (up) or lower (down) than the mean of Prophet’s predictions.

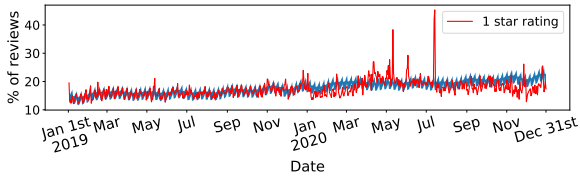


Figure 13: Prophet forecast for 1 star rating (LA)

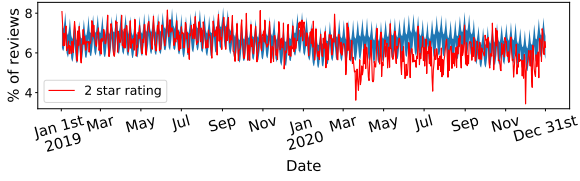


Figure 14: Prophet forecast for 2 star rating (LA)

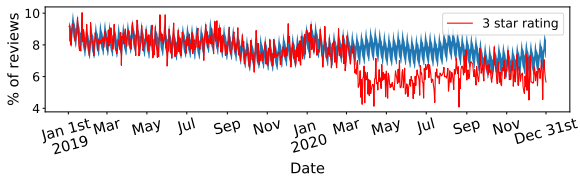


Figure 15: Prophet forecast for 3 star rating (LA)

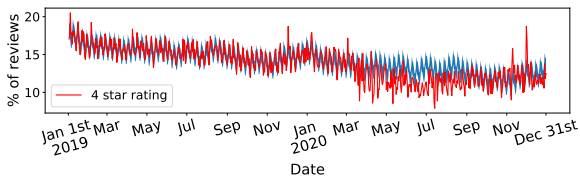


Figure 16: Prophet forecast for 4 star rating (LA)

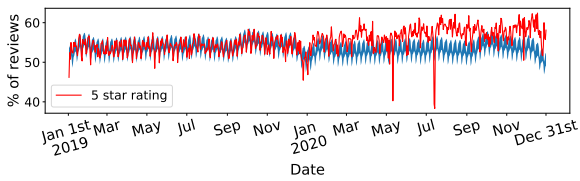


Figure 17: Prophet forecast for 5 star rating (LA)

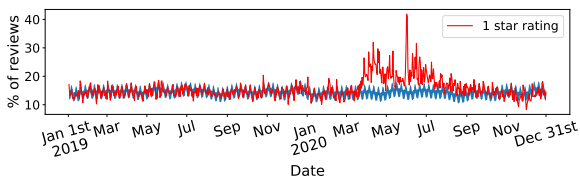


Figure 18: Prophet forecast for 1 star rating (NYC)

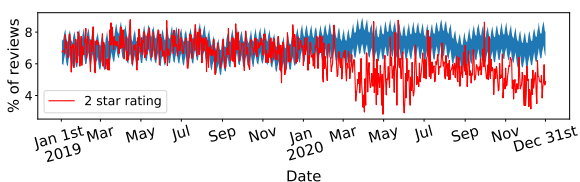


Figure 19: Prophet forecast for 2 star rating (NYC)

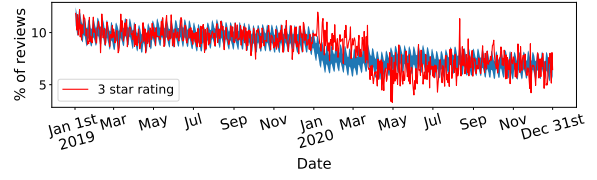


Figure 20: Prophet forecast for 3 star rating (NYC)

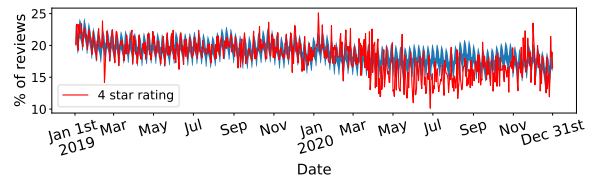


Figure 21: Prophet forecast for 4 star rating (NYC)

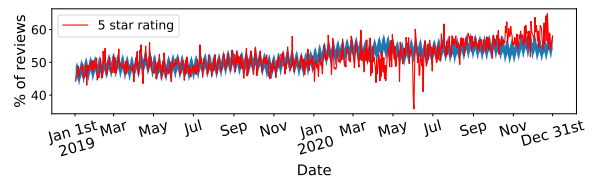


Figure 22: Prophet forecast for 5 star rating (NYC)

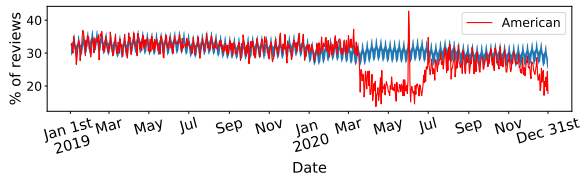


Figure 23: Prophet forecast for "American" group of business tags (LA)

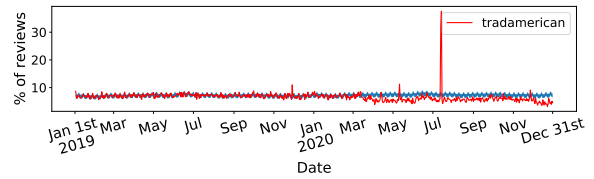


Figure 29: Prophet forecast for "Traditional American" business tag (LA)

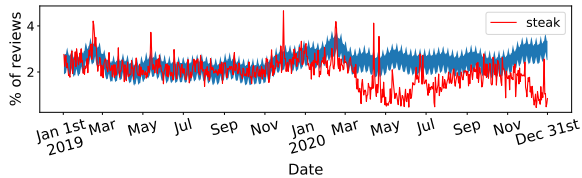


Figure 24: Prophet forecast for "Steak" business tag (LA)

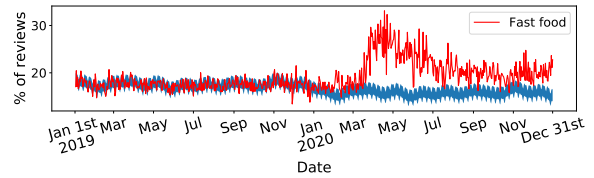


Figure 30: Prophet forecast for "Fast Food" group of business tags (LA)

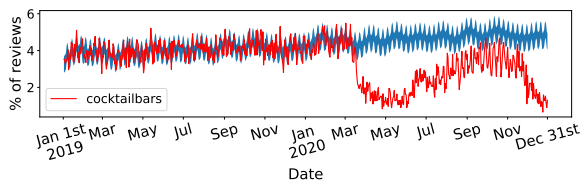


Figure 25: Prophet forecast for "Cocktail Bars" business tag (LA)

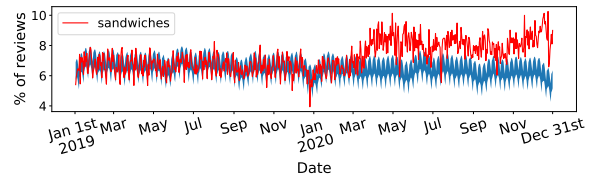


Figure 31: Prophet forecast for "Sandwich" business tag (LA)

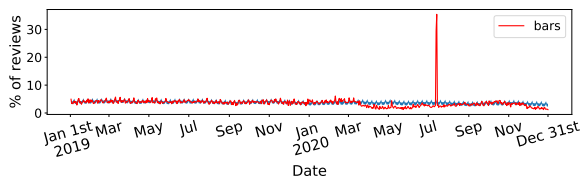


Figure 26: Prophet forecast for "Bars" business tag (LA)

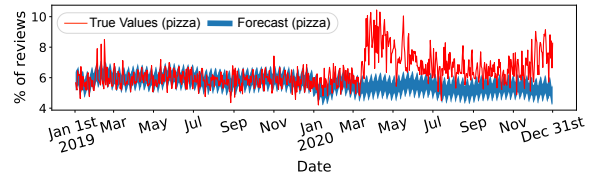


Figure 32: Prophet forecast for "Pizza" business tag (LA)

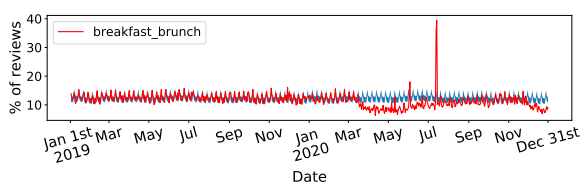


Figure 27: Prophet forecast for "Breakfast&Brunch" business tag (LA)

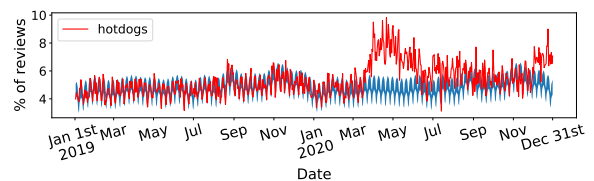


Figure 33: Prophet forecast for "Hot Dog" business tag (LA)

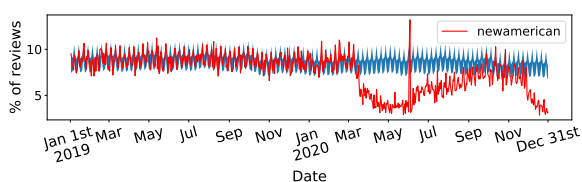


Figure 28: Prophet forecast for "New American" business tag (LA)

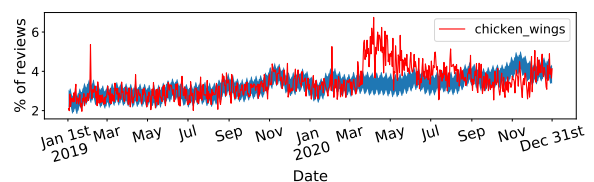


Figure 34: Prophet forecast for "Chicken Wings" business tag (LA)

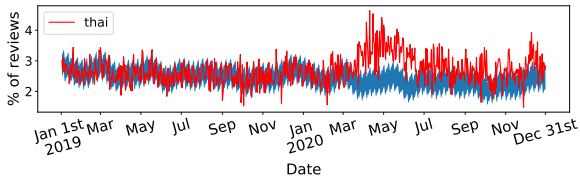


Figure 35: Prophet forecast for "Thai" business tag (LA)

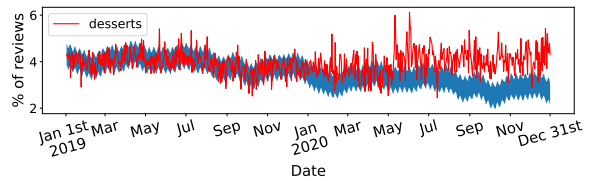


Figure 41: Prophet forecast for "Desserts" business tag (LA)

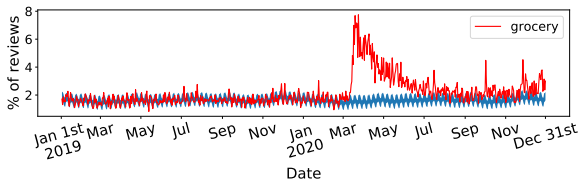


Figure 36: Prophet forecast for "Grocery" business tag (LA)

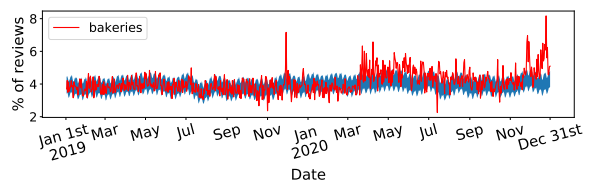


Figure 42: Prophet forecast for "Bakeries" business tag (LA)

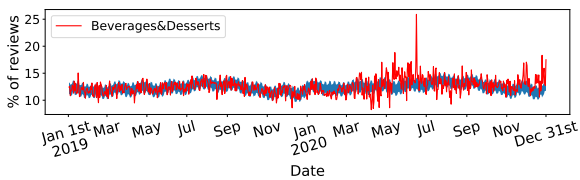


Figure 37: Prophet forecast for "Beverages&Desserts" group of business tags (LA)

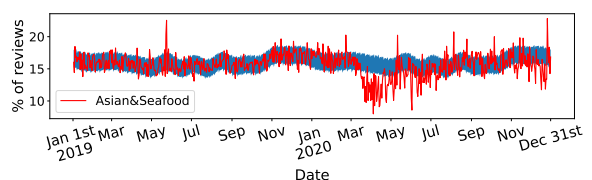


Figure 43: Prophet forecast for "Asian&Seafood" group of business tags (LA)

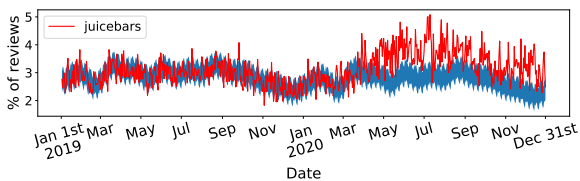


Figure 38: Prophet forecast for "Juice Bars" business tag (LA)

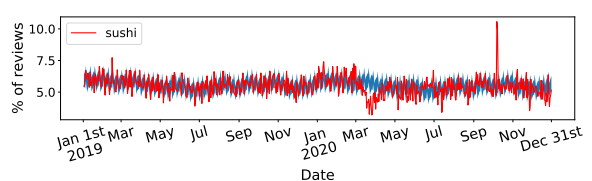


Figure 44: Prophet forecast for "Sushi" business tag (LA)

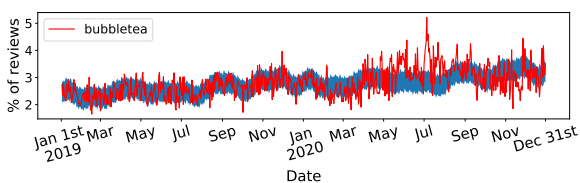


Figure 39: Prophet forecast for "Bubble Tea" business tag (LA)

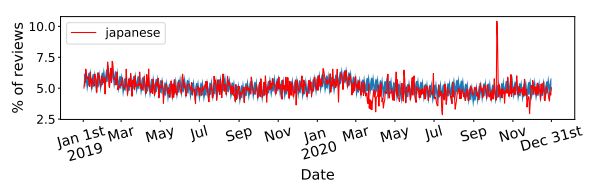


Figure 45: Prophet forecast for "Japanese" business tag (LA)

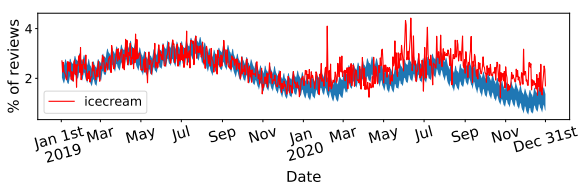


Figure 40: Prophet forecast for "Ice Cream" business tag (LA)

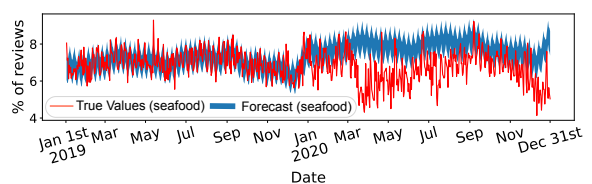


Figure 46: Prophet forecast for "Seafood" business tag (LA)

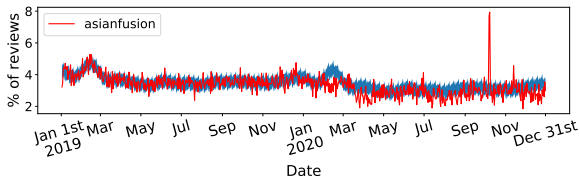


Figure 47: Prophet forecast for "Asian Fusion" business tag (LA)

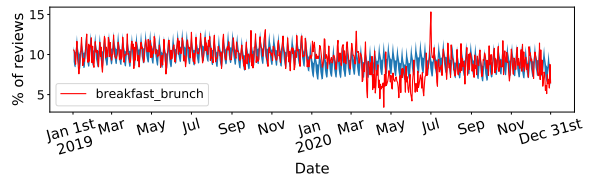


Figure 53: Prophet forecast for "Breakfast&Brunch" business tag (NYC)

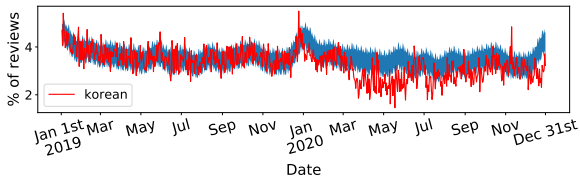


Figure 48: Prophet forecast for "Korean" business tag (LA)

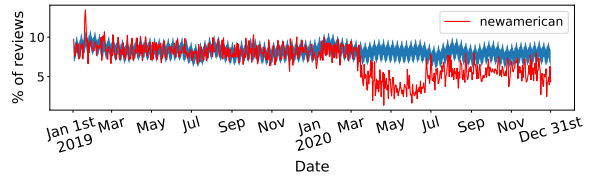


Figure 54: Prophet forecast for "New American" business tag (NYC)

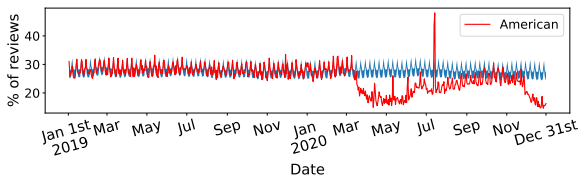


Figure 49: Prophet forecast for "American" group of business tags (NYC)

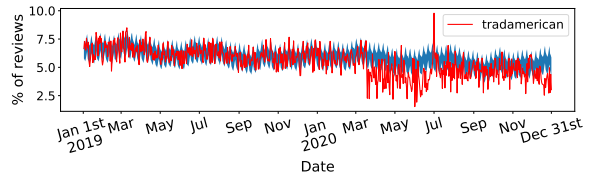


Figure 55: Prophet forecast for "Traditional American" business tag (NYC)

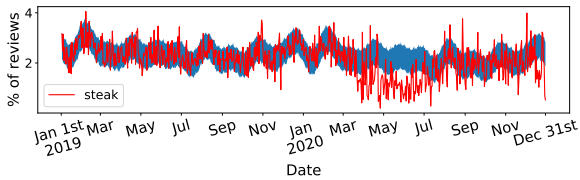


Figure 50: Prophet forecast for "Steak" business tag (NYC)

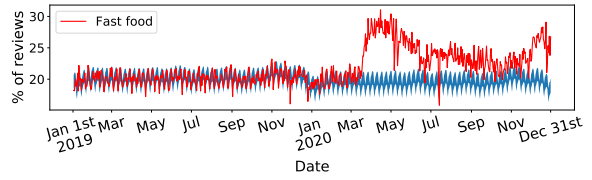


Figure 56: Prophet forecast for "Fast food" group of business tags (NYC)

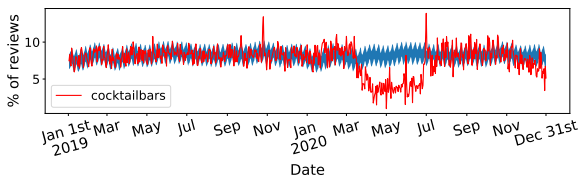


Figure 51: Prophet forecast for "Cocktail Bars" business tag (NYC)

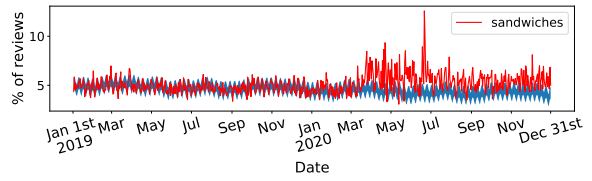


Figure 57: Prophet forecast for "Sandwich" business tag (NYC)

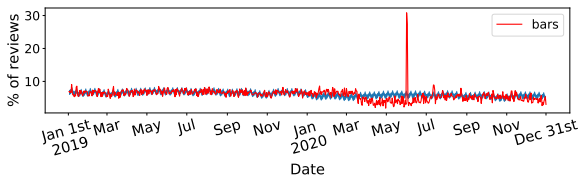


Figure 52: Prophet forecast for "Bars" business tag (NYC)

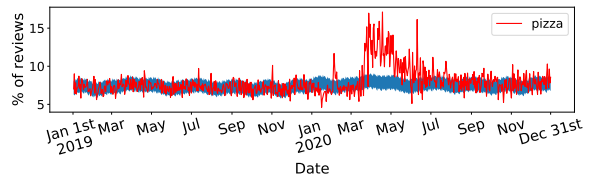


Figure 58: Prophet forecast for "Pizza" business tag (NYC)

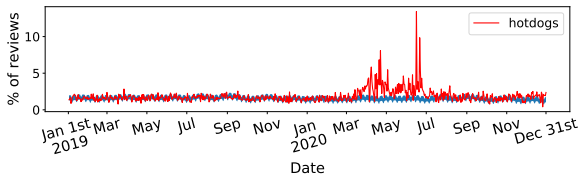


Figure 59: Prophet forecast for "Hot Dog" business tag (NYC)

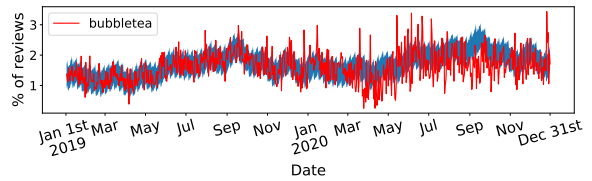


Figure 65: Prophet forecast for "Bubble Tea" business tag (NYC)

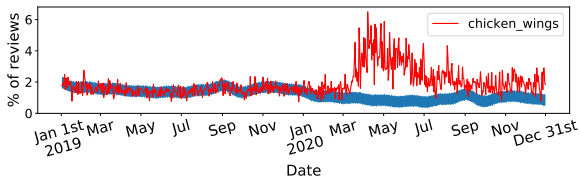


Figure 60: Prophet forecast for "Chicken Wings" business tag (NYC)

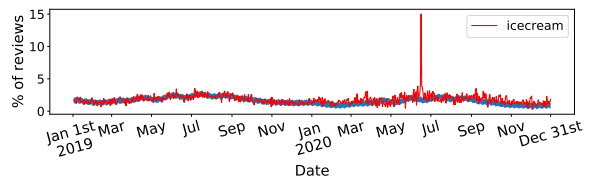


Figure 66: Prophet forecast for "Ice Cream" business tag (NYC)

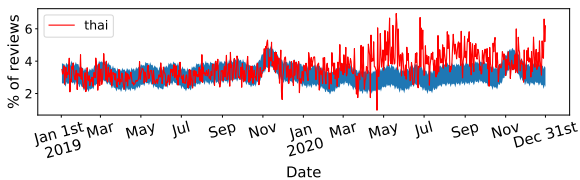


Figure 61: Prophet forecast for "Thai" business tag (NYC)

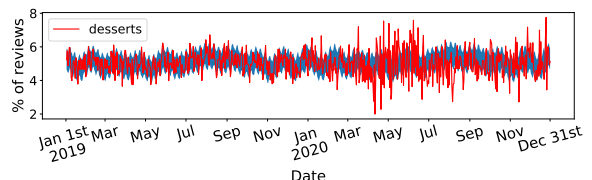


Figure 67: Prophet forecast for "Desserts" business tag (NYC)

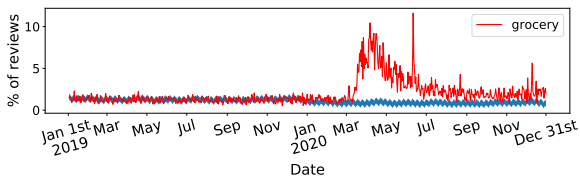


Figure 62: Prophet forecast for "Grocery" business tag (NYC)

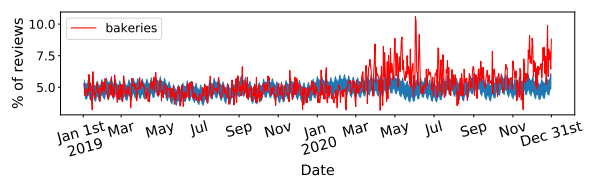


Figure 68: Prophet forecast for "Bakeries" business tag (NYC)

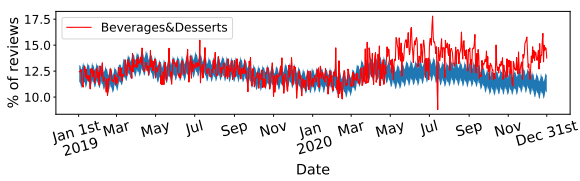


Figure 63: Prophet forecast for Beverages&Desserts group of business tags (NYC)

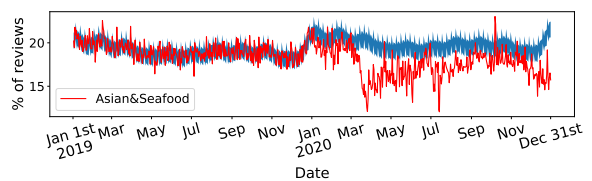


Figure 69: Prophet forecast for "Asian&Seafood" group of business tags (NYC)

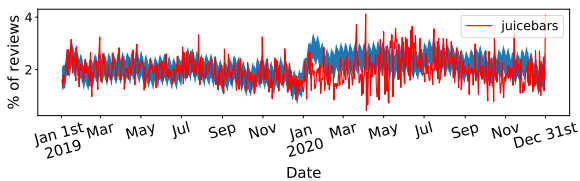


Figure 64: Prophet forecast for "Juice Bars" business tag (NYC)

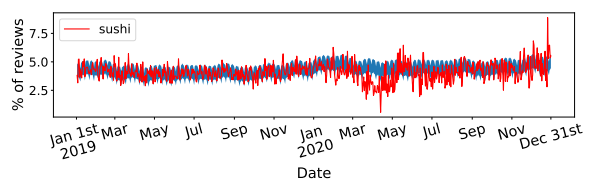


Figure 70: Prophet forecast for "Sushi" business tag (NYC)

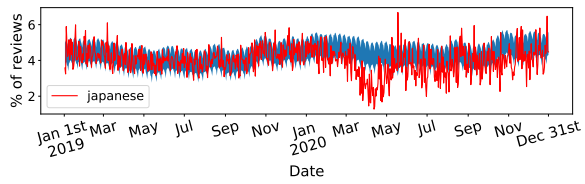


Figure 71: Prophet forecast for "Japanese" business tag (NYC)

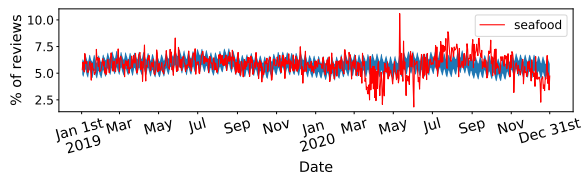


Figure 72: Prophet forecast for "Seafood" business tag (NYC)

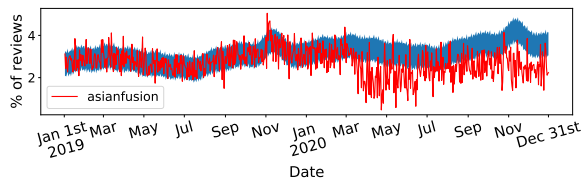


Figure 73: Prophet forecast for "Asian Fusion" business tag (NYC)

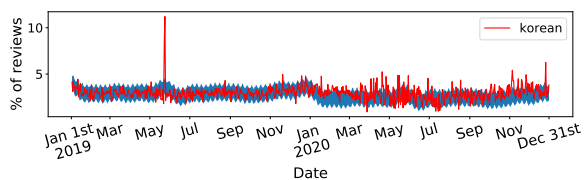


Figure 74: Prophet forecast for "Korean" business tag (NYC)