

Auditing Yelp’s Business Ranking and Review Recommendation Through the Lens of Fairness

Mohit Singhal^{1*}, Javier Pacheco², Seyyed Mohammad Sadegh Moosavi Khorzooghi²,
Tanusree Debi², Abolfazl Asudeh³, Gautam Das², Shirin Nilizadeh²

¹Northeastern University

²The University of Texas at Arlington

³University of Illinois Chicago

m.singhal@northeastern.edu, {jxp6862,seyyedmohammads.moosavikhorzoog,txd8597}@mavs.uta.edu, asudeh@uic.edu, gdas@cse.uta.edu, shirin.nilizadeh@uta.edu

Abstract

Auditing is critical to ensuring the fairness and reliability of decision-making systems. However, auditing a black-box system for bias can be challenging due to the lack of transparency in the model’s internal workings. In many web applications, such as Yelp, it is challenging, if not impossible, to manipulate their inputs systematically to identify bias in the output. Yelp connects users and businesses, where users identify new businesses and simultaneously express their experiences through reviews. Yelp recommendation software moderates user-provided content by categorizing it into recommended and not-recommended sections. The recommended reviews, among other attributes, are used by Yelp’s ranking algorithm to rank businesses in a neighborhood. Due to Yelp’s substantial popularity and its high impact on local businesses’ success, understanding the bias of its algorithms is crucial.

This data-driven study, for the first time, investigates the bias of Yelp’s business ranking and review recommendation system. We examine three hypotheses to assess if Yelp’s recommendation software shows bias against reviews of less established users with fewer friends and reviews and if Yelp’s business ranking algorithm shows bias against restaurants located in specific neighborhoods, particularly in hotspot regions, with specific demographic compositions. Our findings show that reviews of less-established users are disproportionately categorized as not-recommended. We also find a positive association between restaurants’ location in hotspot regions and their average exposure. Furthermore, we observed some cases of severe disparity bias in cities where the hotspots are in neighborhoods with less demographic diversity or higher affluence and education levels.

Introduction

The biases of machine learning-based decision-making systems are a major concern. These biases have been identified in various domains, including predicting recidivism against African Americans (Angwin et al. 2016; Dressel and Farid 2018; Flores, Bechtel, and Lowenkamp 2016), and discriminatory practices against women in job screening and online advertisements (Sweeney 2013). These instances underscore

the urgent need for systematic auditing to uncover and address biases, as the consequences of biased decisions can perpetuate societal inequalities and undermine the ethical foundations of these systems.

However, many of these systems are black-box, wherein the training data, algorithms, and model parameters are not publicly available. This lack of transparency makes the task of auditing for bias more challenging (Caton and Haas 2020; Harrison et al. 2020). Prior works proposed methods to audit black-box classifiers and recommendation systems (Adler et al. 2018; Datta, Sen, and Zick 2016; Feldman et al. 2015; Tan et al. 2018). These works either make repeated calls to the black-box model API (Kim, Ghorbani, and Zou 2019) or remove, permute, or obscure some of the features, or mimic the model (Adler et al. 2018; Tan et al. 2018) However, in certain scenarios, ethical and design restrictions may prevent auditors from probing the black-box system with carefully crafted input data. Consequently, the sole available resource for investigating bias becomes the output data already generated by the system. This paper, for the first time, audits one such web application, Yelp, specifically auditing its business ranking and review recommendation system.

On Yelp, users identify businesses in a local city or neighborhood, connect to other users, and share their experiences about the businesses with other Yelp users through their reviews and ratings. Yelp uses an automated content moderation tool called Yelp recommendation software to filter certain reviews based on four key considerations: conflicts of interest, solicited reviews, reliability, and usefulness (Inc. 2022). These filtered reviews are not removed from the site but are moved into a separate, less visible section called *Not Recommended*, shown as a page link at the end of the businesses’ page (Yelp 2023). Reviews are important because other Yelp users rely on them to choose a business. They also affect the business’s overall star rating, which can subsequently affect its ranking in the Yelp search results.

Hypothesis 1 on Yelp Review Recommendation (Filtering) system: Yelp has a mechanism to encourage users to provide “high-quality” reviews. For example, active Yelp users with many reviews, friends, etc., who also constantly provide helpful reviews receive an *Elite* badge (Support 2024). Prior works have shown that *elite* user reviews are

*Work done at The University of Texas at Arlington.

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

given higher weightage (Kamerer 2014; Nilizadeh et al. 2019). However, using these users' account characteristics for filtering their reviews can lead to bias against *less established* users who have a lower no. of reviews and friends. Filtering these users' reviews can be very frustrating for them or even considered unfair to them, as pointed out by Eslamin et al. (Eslami et al. 2019). Quotes from their qualitative study show this frustration and disappointment: "A few friends and a few more reviews will take you out of the filter algorithm and allow your reviews to be posted."

Additionally, prior research has extensively explored the issue of *long-tailed* item recommendation, where less prolific users or items with limited activity receive less attention (Fleder and Hosanagar 2009; Silva et al. 2023). Furthermore, studies have highlighted the adverse effects of unfair content moderation on user disengagement (Duffy and Meisner 2022; Singhal et al. 2023b). Therefore, the first hypothesis examines the bias of Yelp's recommendation system against a new type of sensitive group, i.e., *less-established users* in contrast to *well-established users*: **H1. Reviews written by users with fewer friends and reviews are more likely to be categorized as not recommended.**

Hypotheses 2 and 3 on the Yelp's Business Recommendation System: Yelp uses a black-box ranked-retrieval model (Washcovick 2019) to list businesses per users' search query, looking for a specific type of business, location, services, etc. The search result shows a ranked list of businesses on several pages. Businesses listed among the firsts on the first few pages are considered to have higher visibility and exposure, as some studies have shown users tend only to check the first or a few first pages (Joachims et al. 2017). Studies have also shown that this ranking highly affects businesses because the higher visibility and exposure (Singh and Joachims 2018) means higher foot traffic and increased revenue (Luca 2016). Therefore, a fair ranking is essential for a business. Many factors, including location, can impact a business's ranking on Yelp search results. For example, some neighborhoods are hotspots for restaurants, and several businesses thrive in these locations. This can be due to being in the most touristy and visited parts of the city, which garner more foot traffic. This can, in turn, lead to a higher number of (positive) reviews for the businesses, *giving even more exposure to them*. The higher exposure, in turn, drags even more customers to these businesses, giving them a further boost in fame and popularity, even though they did not provide a better service or higher quality food than other businesses. This cycle of discrimination continuously increases the popularity and revenue gap between these businesses located in different neighborhoods. As a result, it causes "*the rich to get richer and the poor to get poorer*" as it is shown in prior works that there is a correlation between reviewer positiveness and restaurants being in a hotspot or popular location (Kokkodis and Lappas 2020; Huang et al. 2016). Hence, we hypothesize that:

H2. In Yelp's ranking of businesses in a city, businesses located in hotspots tend to have higher exposure.

Furthermore, while the location itself might not be considered sensitive, it can be associated with sensitive attributes, such as neighborhoods' economic status and demographic

and racial composition. Therefore, if Yelp's business ranking algorithm is biased against businesses that are not located in hotspots, and if these neighborhoods are associated with a specific demographic composition, then Yelp is (implicitly) biased against this specific population. Hence, we hypothesize that:

H3. Hotspots are positively associated with less demographic diversity, high education levels, high income, and low unemployment neighborhoods

To examine bias, we relied on popular definitions of fairness and bias, such as *exposure* and used statistical tests, such as quantile linear and logistic regression, to identify associations and relationships between Yelp's outcomes and the sensitive groups. Since the inherent black-box structure and the lack of access to training data and algorithms make it challenging to audit Yelp, we designed two data collection frameworks to minimize bias. Note that in contrast to other auditing frameworks, due to ethical considerations, we do not generate inputs (create accounts or write reviews) to pass them through the system but instead analyze the existing data on the platform.

To audit the algorithmic bias of the Yelp Review Recommendation (Filtering) system, we obtained *recommended* reviews of 15K random businesses from Yelp Dataset Challenge in 2023 (Inc. 2023). We also created a framework to obtain all the *not-recommended* reviews of these businesses from Yelp as Yelp does not provide it in the Dataset Challenge. In total, we obtained 707K recommended and 178K not-recommended reviews. To audit the algorithmic bias of the Yelp Ranking system, we first obtained the ranking of all restaurants for 9 cities in the US using our custom-build framework as Yelp Dataset Challenge does not provide the ranking of businesses. Through the statistical analysis, we found that reviews written by users with a higher number of friends and reviews are more likely to be categorized as *recommended*, hence showing a disparate impact for new or less-active users ($p < 0.001$). Our quantile analysis showed a statistically significant positive correlation between a restaurant's presence in a hotspot and its average exposure ($p < 0.001$). This finding was observed across all nine cities, indicating that the Yelp algorithm disproportionately accords higher exposure to restaurants in the city's hotspots. Furthermore, we noted more severe cases of disparity bias in cities where these hotspots are placed in neighborhoods with less demographic diversity or areas with higher affluence and education levels. For example, in Chicago, we found that hotspot regions are usually in highly educated and wealthy neighborhoods. Hence, we can conclude that there would be bias against the restaurants in neighborhoods with lower wealth and less educated people. The association between hotspots and demographic composition can also be due to existing deep underlying societal biases.

Our results highlight the importance of *responsible data science practices* in data-driven systems such as Yelp. Irresponsible ranking and recommendation models and algorithms cannot only create unfair results but can also create discriminatory feedback loops that keep segregating the cities and marginalizing people.

Related Work

Fairness & Bias in Recommendation Systems. Prior works have found that the recommendation systems incur biases (Zhang and Liu 2021; Ali et al. 2019; Abdollahpouri et al. 2019), some against a gender or race group (Li et al. 2021b; Hannák et al. 2017; Buolamwini and Gebru 2018; Zhao et al. 2017) leading to under-representation of some social groups and denial of economic opportunities for others (Hannák et al. 2017; Abdollahpouri and Burke 2019; Beutel et al. 2019; Koenecke et al. 2023). For example, Ali et al. (Ali et al. 2019) found that skewed delivery of ads occurred on Facebook leading to bias. Ekstrand et al. (Ekstrand and Pera 2017) found that recommender systems suffer of what is known as “sample size bias.” Edizel et al. (Edizel et al. 2020) found that bias that exists in the real world can be modeled and amplified by recommender systems. Marlin et al. (Marlin et al. 2012) found selection bias in rating data collected via user survey. Recommender systems can also affect users decision making process, known as decision bias, and Chen et al. (Chen et al. 2013) show how understanding this bias can improve recommender systems. Dozens of methods to make these ML systems fair are proposed (Li et al. 2021b,a; Fu et al. 2020). **Fairness & Bias in Ranking.** Ranking fairness implies that comparable items or groups of items receive similar visibility. Prior works have extensively studied fairness in ranking (Stoyanovich, Yang, and Jagadish 2018; Patro et al. 2022; Zehlike, Yang, and Stoyanovich 2022a). Scholarships have examined probability-based fairness notions and provided mechanisms to design fair ranking (Asudeh et al. 2019; Zehlike et al. 2017). For example, Asudeh et al. (Asudeh et al. 2019) developed a system that assists individuals in choosing criterion weights that lead to significant fairness. Singh et al. (Singh and Joachims 2018) proposed a framework for formulating fairness constraints on rankings. In fair ranking tasks, a frequently studied problem is how to distribute the exposure opportunity to candidates fairly (Diaz et al. 2020; Singh and Joachims 2018; Zehlike and Castillo 2020). Position bias is very common in recommendation systems (Chen et al. 2023). Popularity bias happens when users to notice or interact with items in certain positions of lists with higher probability (Collins et al. 2018). Maeve et al. (O’Brien and Keane 2006) shows that users often trust the first few results in the lists. Popularity bias is another form of “exposure bias” (Zheng et al. 2021). Our work builds on these works by investigating the bias of Yelp’s ranking system through the lens of exposure bias. **Yelp Recommendation System:** Mukherjee et al. (Mukherjee et al. 2013a) investigated the factors that the Yelp review filtering algorithm uses, reporting that Yelp might filter reviews based on various behavioral features than linguistic ones. In our work, we examine the sensitive user-level features, such as no. of friends and reviews. Kamerer (Kamerer 2014) established that Yelp users who are prolific reviewers and are known and trusted by the community have their reviews in the recommended section. Yao et al. (Yao et al. 2018) found that reviews were most likely to be in the recommended section when conveying an overall positive message. Some works studied the opinions and reactions of users whose re-

views were being removed (Duffy and Meisner 2023; Jhaver et al. 2019). For example, Eslami et al. (Eslami et al. 2019) found that users echoed that the system suppresses their voices. One relevant work to our study (Amos, Maio, and Mittal 2022) empirically studied Yelp’s review recommendation and found that reviews of businesses in lower-density and low-middle-income areas are most likely to be labeled as not recommended. Our study, however, investigates bias toward established and new users.

Bias in Content Moderation. Scholarships have investigated bias and fairness in content moderation and how it affects the users (Duffy and Meisner 2023; Jhaver et al. 2019; Eslami et al. 2019; Singhal et al. 2023b), mostly focusing on its harm on marginalized communities via user studies (Haimson et al. 2021; Seering et al. 2019). Eslami et al. (Eslami et al. 2019) found that elite users on Yelp defend the Yelp review recommendation algorithm because their reviews are rarely filtered.

Auditing Bias of Yelp’s Review Recommendation (Filtering) System

Researchers have proposed dozens of definitions of fairness for ML models (Narayanan 2018; Ekstrand et al. 2022; Lazovich et al. 2022; Patro et al. 2022). However, not all definitions of fairness can satisfy a specific use case. In this particular use case, we have three key stakeholders, i.e., Yelp, Yelp users, and Businesses on Yelp each having their own fairness perspective: **Yelp:** Yelp would want to ensure their review recommendation software has near-perfect accuracy, correctly identifying and highlighting honest and useful reviews. Therefore, fairness to Yelp is to use any features or ML models that can provide such high performance, enhancing the experiences of both people and businesses who use the platform, and maintaining Yelp’s popularity in the highly competitive market over other recommendation applications. **Yelp Review Writers:** Review writers spend time and effort to create and maintain an account on Yelp and participate in the platform by connecting to other Yelp users and sharing their personal experiences with everyone. Therefore, fairness for these users can be acknowledging their opinions, making them available on the platform, and using them to rate the businesses. In contrast, if their reviews are placed in the not recommended section, they echo frustrations and call the system biased (Eslami et al. 2019). **Businesses on Yelp** would want Yelp to remove reviews that are low quality and have a high probability of being fake. Low quality and especially negative fake reviews can cause the star rating of a business to decrease, resulting in less foot traffic (Hussain et al. 2019; Horn et al. 2015).

These three perspectives can contrast each other. For example, to achieve high accuracy, the system might prefer to use the account’s characteristics, such as the number of reviews and friends. This might result in the disproportional removal of reviews by less established users.

Methodology and Bias Metrics

Prior works (Koenecke et al. 2023; Becerril-Arreola 2023) have used statistical association between the systems’ out-

comes and the sensitive attributes to indicate bias. For example, Becerril (Becerril-Arreola 2023) used regression analysis to explain product prices and price differences as functions of consumer demographics. To test **H1**, we used multivariate logistic regression to statistically investigate the relationship between the users’ number of friends and reviews and their reviews being categorized as not recommended. We also controlled for various other confounding metrics for review quality and likelihood of being fake that could contribute to whether the review would be in the *not recommended* section. Specifically, we controlled for the *sentiment* of the review, which is a binary variable for either positive or negative review. Yao et al. (Yao et al. 2018) found that when the review is positive, it is more likely to be in the recommended section. Additionally, we controlled for *review length* and *number of spelling mistakes* as these properties can be indicators of informative vs. low-quality reviews (Kamerer 2014; Mukherjee et al. 2013a).

Fake review consideration. Yelp has two subsections on the Not Recommended page, one for the not recommended reviews and one for those violating its community guidelines (Yelp 2024). We did not crawl the second section, hence our dataset does not have any reviews that are in the violation of Yelp ToS, including fake reviews. Hence, the biases in Yelp’s fake review detection method minimally affect our analysis. The only concern is the false negatives that the method might have and the fake reviews that remain undetected in both sections on Yelp, i.e., recommended and not-recommended sections. We checked recent works such as (Mohawesh et al. 2024; Duma et al. 2024; Ashraf et al. 2024), that are about fake review detection and unfortunately, we could not identify any that open sources their code or data. Also, most of these methods include features extracted from (meta-) data or auxiliary data that we do not have access, making it almost impossible to implement them. To control for such cases, we developed a detection method. We first preprocessed the reviews by removing special characters, emoticons, etc., to minimize the sparsity of data (Singhal et al. 2023a). We extracted bi-grams and uni-grams from all the posts and considered the top 2000 with the highest TF-IDF values. Using the groundtruth dataset provided by (Salminen et al. 2022), we implemented a traditional SVM ML model and found an accuracy of 0.93 and an F1-score of 0.93, which is higher than reported in prior work (Mukherjee et al. 2013b; Harris 2019). Additionally, Figure 1 shows the distribution of the likelihood of reviews being fake and interestingly, we can see that there is a high left skew, i.e., almost all reviews are benign. Hence, compared to this simple model, it seems Yelp is performing well in finding fake reviews. Hence we used this as a control variable in our model.

Since the distribution of these variables is not normal but rather skewed, running the test on all the data cannot provide a full picture. We used quartile regression analysis (Yu, Lu, and Stander 2003), which organizes data into three points—a lower quartile, median, and upper quartile—to form four groups of the dataset, each representing 25% of the observations. Quartiles are examples of quantiles used in research as statistical quantities (Yu, Lu, and Stander

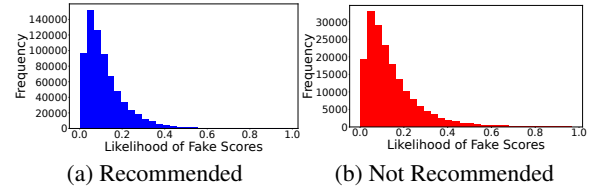


Figure 1: Histogram of likelihood of reviews to be fake

2003; Koenker 2004). After dividing the data into different quartiles, we run our regression model on each quartile separately. This allows us to interpret the results of each quartile separately. Regression models are valuable for analyzing how changes in one variable are associated with changes in another variable. In the context of bias, we used regression to understand how certain variables might be correlated with biased outcomes. However, it’s essential to remember that correlation does not imply causation. A regression model can demonstrate a statistical relationship, but it cannot prove that one variable causes the bias.

Data collection

We developed a framework for collecting the datasets. First, we used the dataset provided by Yelp as part of their Yelp Dataset Challenge in 2023 (Inc. 2023) to identify and collect 707,658 *recommended* reviews of about 144K businesses spanning over about 11 metropolitan cities in the USA. The Yelp Dataset Challenge does not provide the *not recommended* reviews, and therefore, we deployed a custom-built crawler on Yelp to collect them. However, collecting such large-scale data is time-consuming and could generate much traffic, violating Yelp’s Terms of Service. Therefore, we sampled 15K (10%) of the businesses and obtained their not-recommended reviews. To avoid overwhelming Yelp with our requests, we delayed the data collection process in our crawler by 5 to 10 seconds, which was enough, and Yelp never blocked our requests. We were able to obtain 178,747 not-recommended reviews for 13,239 businesses, as 1,761 of them did not have any non-recommended reviews at the time of data collection. Scraping the not recommended reviews, We obtained the full review text, review date, friend and review count, username, location, and the URL of the user profile image. However, it doesn’t provide user IDs for them. Our code and data can be found and tested in this anonymous Dropbox link <https://tinyurl.com/pnt8uudb>. Our dataset does not include paid reviewers.

Variable	Recommended Users (n=707,658)				Not-Recommended Users (n=178,747)			
	Min	Max	Mean	Med.	Min	Max	Mean	Med.
# Friends	1	19,679	157.89	21	0	4,995	23.79	0
# Reviews	0	17,473	121.76	24	0	7,738	7.73	2

Table 1: Descriptive statistics of dataset

Dataset Characterization: Table 1 shows the descriptive statistics of the dataset. We could not find the unique

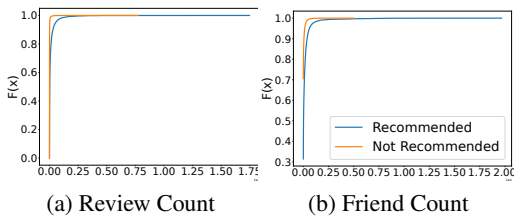


Figure 2: CDF of Reviews and Friends for Recommended and Not Recommended Users

users in the not recommended dataset, as Yelp does not provide user IDs of these users in that section. Therefore, we did not obtain the unique users in the recommended set as well. Figures 2a and 2b show the cumulative density function (CDF) of the number of friends and reviews, respectively. They show that these variables do not follow the normal distribution and on average, users in the recommended dataset have more friends and reviews than those in the non-recommended dataset.

Results

Reviews written by new users with fewer friends and reviews are categorized as not recommended. We ran a multivariate logistic regression model on all the data, with *recommended* as the dependent variable and the number of *friends* and *reviews* as the independent variables. We also controlled for various confounding metrics for review quality such as *review length (in char.)*, *no. of spelling mistakes*, *review sentiment (positive/negative)*, and the likelihood of being a fake review. We used logistic regression because our dependent variable is binary. Our findings show that the reviews of users with more friends and reviews are more likely to be recommended. Additionally, reviews that are less likely to be fake are more likely to be recommended. To check the robustness of our method, we created a binary variable of *Likelihood of Fake*, where a score of less than and equal to 0.5 was labeled as 0, and greater than 0.5 was labeled as 1. Our results were the same as those obtained using the original proportion; hence, our model is robust. However, these findings from the whole dataset might be misleading because the distributions of the number of friends and reviews variables are skewed. Therefore, we also ran two quartile logistic regressions, one for the *friends* and the other for *reviews*, where we could investigate the relationships in each quartile. Table 2 shows the results. The results on the whole data, without quartile analysis (in the second column) still show a statistically significant positive association between the number of friends and reviews and the reviews being categorized as *recommended*. In this paper, we report Incident Rate Ratios (IRR), which are the exponentiated coefficients of Poisson regressions. These ratios show the multiplicative effect on the expected value. Quartile regression on users' number of friends (in Table 2) reveals that a strong positive association exists between the users' number of friends and their reviews being categorized as recommended than those users who have less no. of friends, with

incident rate 1.001 (Q_3) that of users with fewer friends (a 0.1% increase) $p < 0.001$, and 1.000 (Q_4) that of users with fewer friends $p < 0.001$.

Quartile regression on users' number of reviews (in Table 2) reveals a positive association between the users' number of reviews and their reviews being categorized as recommended (IRR = 1.917, $p < 0.001$) in the first quartile, where users have less number of reviews. However, we found that when users' review count starts to increase, i.e., from the second to the third quartile, even though there is still a positive association, the impact decreases as the incidence rate is 1.164 times (16.4% increase) in the second quartile and then 0.016 times (1.6% increase). Interestingly, these results are consistent when we ran the regression model on all the data (See column second of Table 2) as the higher the number of friends and reviews, the more likely they would be in the recommended section, with an increase of 0.3% and 5.1% for friends and reviews respectively. Hence, in summary, we found *support* for our **H1** that new or less-established users with fewer friends and reviews are more likely to have their reviews in the not-recommended section, indicating possible bias against these groups of users and statistically verifying what the authors in (Eslami et al. 2019) found by doing their user study.

Auditing Bias of Yelp's Ranking System

Yelp's Ranking System has three key stakeholders, i.e., Yelp, Businesses on Yelp, & Yelp Users where each has a different fairness perspective. **Yelp** would want to ensure that its business ranking algorithm provides an accurate ranking for businesses so that both users and businesses can benefit from the platform. **Businesses on Yelp** would want Yelp to remove fake reviews from their platform and be ranked fairly compared to their competitors. **Yelp Users** would want Yelp to fairly provide them with the ranking of businesses. These three perspectives are similar: all want a fair ranking. However, the main contrast comes when Yelp uses additional features that explicitly or implicitly consider businesses' location and, subsequently, demographic composition attributes for ranking. This can harm businesses that are in areas of low income or education or are in diverse neighbourhoods (Zehlike et al. 2017).

Methodology & Bias Metrics

Data Collection: Since the Yelp Dataset Challenge does not provide the ranking of businesses, we developed a second data collection framework to collect business rankings in a given city. We only collected organic search results, i.e., *Sponsored Businesses* were not collected. Additionally, the location was set through the form on Yelp's homepage. Hence, making sure that we only get the true ranking of businesses in a city. We limited our analysis to some cities, obtaining a sample of them known for their demographic and social compositions. We first collated a list of cities based on the following criteria: *tourism*, *wealth*, *education*, *diversity*, and *crime*. To do this, we identified reputable and reliable websites, such as HGTV, CNBC, Forbes, USA Today, and WalletHub, which provided up-to-date lists of cities based

	Dependent variable: <i>recommended</i>				
	Logistic	0.25 Qnt.	0.50 Qnt.	0.75 Qnt.	1.00 Qnt.
No. of Friends	0.003(0.000)***	592.9 (0.997)	-0.453(0.000)***	0.001(0.000)***	0.000(0.000)***
No. of Reviews	0.049(0.000)***	0.024(0.000)***	0.024(0.000)***	0.028(0.000)***	0.011(0.000)***
Sentiment	-0.026(0.000)***	-0.528(0.000)***	-0.419(0.000)***	-0.252(0.000)***	-0.300(0.000)***
Spelling Mistakes	0.003(0.000)***	0.031(0.000)***	0.050(0.000)***	0.062(0.000)***	0.062(0.000)***
Length of Review	-0.000(0.000)***	0.000(0.618)	-0.000(0.000)***	-0.000(0.000)***	-0.000(0.834)
Likelihood of Fake	-1.576(0.000)***	-1.651(0.000)***	-1.893(0.000)***	-1.964(0.000)***	-2.073(0.000)***
No. of Friends (IRR)	1.003	3.182e+257	0.635	1.001	1.000
No. of Reviews (IRR)	1.051	1.024	1.025	1.028	1.0117
Sentiment (IRR)	0.769	0.589	0.657	0.776	0.740
Spelling Mistakes (IRR)	1.031	1.033	1.051	1.064	1.064
Length of Review (IRR)	0.999	1.000	1.000	0.999	0.999
Likelihood of Fake (IRR)	0.206	0.191	0.150	0.140	0.125
Observations	886,405	221,602	221,601	221,601	221,601
No. of Reviews		0.651(0.000)***	0.152(0.000)***	0.016 (0.000)***	0.000(0.000)***
No. of Friends		0.004(0.000)***	0.002(0.000)***	0.000(0.000)***	0.000(0.000)***
Sentiment		-0.556(0.000)***	-0.372(0.000)***	-0.230(0.000)***	0.165(0.009)**
Spelling Mistakes		0.002(0.024)*	0.019(0.000)***	0.086(0.000)***	0.080(0.000)***
Length of Review		0.000(0.000)***	0.000(0.000)***	0.001(0.000)***	0.005(0.000)***
Likelihood of Fake		-1.557(0.000)***	-1.950(0.000)***	-2.233(0.000)***	-2.011(0.000)***
No. of Reviews (IRR)		1.917	1.164	1.016	1.000
No. of Friends (IRR)		1.004	1.002	1.000	1.000
Sentiment (IRR)		0.573	0.689	0.791	1.180
Spelling Mistakes (IRR)		1.002	1.019	1.090	1.083
Length of Review (IRR)		1.000	1.000	1.001	1.005
Likelihood of Fake (IRR)		0.210	0.142	0.107	0.133
Observations		221,602	221,601	221,601	221,601

Note: *p<0.05; **p<0.01; ***p<0.001

Table 2: Results of the quartile regression

on specific criteria of our interest. For each criterion, we collated the top 10 cities. Then, we created a 127 by 5 matrix, with each row representing the cities and the column representing the criteria that it fulfilled. In some cases, one city was repeated multiple times for different criteria. For example, one source (HGTV) listed Washington D.C. as a top tourist destination, and the other (USA Today) listed it as a wealthy city; hence, in the matrix, we marked Washington D.C. for both of these categories. To condense this list and find representatives, we followed *The Skyline* approach (Asudeh et al. 2016; Rahman et al. 2017). In the skyline approach, we try to identify the set of non-dominated cities with different attributes, i.e., the cities that fulfill the attributes and are not dominated by any other city for those particular attributes. For example, when creating the skyline for the cities, we found that Anchorage dominated Asheville and Denver in *tourism*, *wealth*, and *crime rate*; hence, we chose Anchorage as it dominated these cities. When multiple cities fulfilled various criteria, such as Tourism, Diversity, etc., and were all part of the skyline, we chose the cities with the highest population count. For example, Chicago, Los Angeles, Louisville, and New York City all satisfy tourism and diversity criteria; however, Louisville is less populated than the other three cities, so we removed it from our list. Using this approach, we obtained ten cities shown in Table 3. We acknowledge the limitation that analyzing all the cities would be ideal. However, considering the limited number of queries enforced by the Yelp server, we had to limit our scope to a subset of cities. Additionally, even though we did not consider all US cities, we selected a representative set based on the considered criteria.

However, we found that many restaurants from Canada

City	Tourism	Wealth	Education	Diversity	Crime Rate
Corpus Christi			X	X	X
Detroit	X		X	X	
Seattle	X	X			
San Jose	X	X	X	X	
San Francisco	X	X	X		
New Orleans	X				X
Anchorage	X	X			X
Chicago	X			X	
Los Angeles	X			X	
New York City	X			X	

Table 3: List of Cities with the criteria. Note that X denotes that the city fulfilled the requirement.

City	# Rests.	City	# Rests.
Anchorage	291	Chicago	329
Corpus Christi	271	Los Angeles	344
New Orleans	281	New York City	348
San Francisco	353	San Jose	346
Seattle	305		

Table 4: Number of Restaurants per City

appeared when searching for restaurants in Detroit. Thus, we removed Detroit from our list of cities. We also limited our analysis to *restaurants* as the category as it is one of the most searched categories on Yelp (Inc. 2019). Using our custom-built crawler, we obtained the list of all the restaurants with their corresponding ranks and addresses. We noticed that the ranking of a restaurant is dynamic and changes at different times, even on a specific day. Therefore, we need more data points on the ranking of businesses. Based on Central Limit Theorem, we decided to collect 30 data points (i.e., rankings for a specific city) so that their means be approximately

normally distributed (Kwak and Kim 2017). Therefore, we extracted the rankings twice a day, once at 11:00 AM and once at 7:00 PM, for 15 days, from March 4th to March 19th, 2023. Table 4 shows the total number of restaurants obtained for each of the nine cities. We obtained the most number of restaurants for the city of San Francisco (353) and the least for New Orleans (281). Please note that Yelp provides a certain number of pages for each city, and we extracted all the restaurants from those pages, and also popular fast food restaurants such as Wendy's, McDonald's, etc., are often not reflected in Yelp's ranking. Since these restaurants are not in the ranking and are not shown to the users, this does not affect our results.

During the data collection, for each restaurant, we also obtained the zip code, where the restaurant is located, and their star rating. We could not obtain the address or the latitude and longitude of these restaurants from Yelp; hence, we used the Google Places API (Google 2023) to extract this information. For each city, we provided the name of the restaurant as well as the latitude and longitude of the city. Moreover, using the zip code, we extracted the *demographic*, *educational*, *income*, and *employment* data from the US Census website of 2020 (Bureau 2023a). The Census provides the demographic information for each zip code <https://tinyurl.com/2w4sbe2k>.

Businesses' exposure by ranking: To identify the disparate impact in the ranking of businesses, we computed the average exposure of each restaurant. We employed the exposure-based fairness metric proposed in (Singh and Joachims 2018). Exposure-based fairness is defined by quantifying the expected attention received by a candidate or a group of candidates, typically by comparing their average *position bias* to that of the other candidates or groups (Zehlike, Yang, and Stoyanovich 2022b; Joachims et al. 2017). This metric allowed us to find the exposure of the businesses to their ranking. Higher exposure means the businesses are better ranked, appear earlier in the list, and have a higher chance of being visited/ checked by them (O'Brien and Keane 2006). The standard exposure drop-off, i.e., the position bias, is given by $\frac{1}{\log(1+j)}$, where j is the position in the ranking. If a restaurant is consistently ranked as #1, then the exposure for each day would be 1, and if the restaurant ranking changes per day then the exposure value is related to its ranking which is between 0 to 1. The closer the value is to 1, more often the restaurant is closer to the top. We used this metric because if a business has a higher exposure, that means that it would be displayed on the earlier pages of the search results and probably get more consumers because of that. We identified some restaurants being present only for one or a few days and not throughout the whole 15 days. Therefore, we assigned 0 exposure on those days and averaged all their data points. Please note that we used the Central Limit Theorem to collect 30 data points, 2 data points per day, at different times to obtain a normal distribution, hence there is no major discontinuity with the exposure of restaurants that are not in top-15.

Hotspot Identification: "City hotspots" typically refer to popular or noteworthy locations within a city that attract a lot of visitors or attention. These spots can include land-

marks, tourist attractions, cultural sites, shopping districts, dining areas, entertainment venues, and vibrant neighborhoods. Restaurants are a key component of popular areas, therefore, to identify hotspots in a city, we tried to detect clusters of restaurants in close proximity.

Using the latitude and longitude information obtained using the Google Places API, we generated various maps using the Folium library in Python. We used the DBSCAN algorithm to perform clustering (Ester et al. 1996) on spatial data to identify city hotspots. In DBSCAN, the user defines the appropriate epsilon (ϵ) value and threshold for the number of neighbors, i.e., *minPts* was set by converting 1KM into radians. We tried multiple epsilon values, including 100 m, 200 m, 500 m, 1000 m, etc., manually cross-checked with their corresponding plot, and found that 200 m was better than others in the number of distinct clusters obtained. When increasing the epsilon, we only obtained one or two clusters with large cluster sizes. We provide the anonymous link to the interactive map of clusters <https://tinyurl.com/4wemktd2>.

We then obtained various clusters for each city containing latitude and longitude information. While in all the cities, many clusters have a size of one, we could obtain hotspots where several restaurants are located close to each other. We found that many clusters were of size 1 or 2 and not hotspots. We called a region a hotspot if more than 6 restaurants were in the vicinity. We chose 6 because we found that not all the cities (8) had clusters of size 7, and while only one city did not have a cluster of size 5, including them would marginally increase the number of clusters for each city. Hence, for each city, we created a binary variable, where we labeled a restaurant to be located in a hotspot as '1' if they were in a cluster of size six or more, else we labeled that as '0'. We obtained the highest number of clusters for San Francisco (13), followed by Chicago (12), Seattle (11), San Jose (10), New York City (9), Los Angeles (8), Anchorage (8), Corpus Christi (7), and the lowest for New Orleans (3). Additionally, we manually checked the obtained hotspots and found that the restaurants in those hotspots were very close, densely populated, and are in the vicinity of popular attractions in the city (for example Figure 4 and Figure 5).

Results of Examining H2

We statistically investigated H2 by using linear regression. We employed this method because our dependent variable, *average exposure*, follows a normal distribution. Additionally, we included the *star* rating of the business as a control variable, based on prior research that demonstrated a one-point increase in star rating leads to improved business ranking and consequently higher revenue (Luca 2016). We also controlled for the number of reviews and available amenities that the restaurant offers as these can also be an indication about how popular the restaurant is.

Businesses located in hotspots have higher exposure. Table 5 shows the result of our regression analysis for hotspots and average exposure, when using aggregated data from all the cities. We see that there is a statistically significant positive association between restaurants' exposure, their location in hotspots, high star ratings, high number of

<i>Dependent variable: average exposure</i>	
Hotspot	0.023 (0.000)***
Stars	0.010 (0.003)**
No. of Reviews	0.000 (0.000)***
Amenities	0.000 (0.001)**
Observations	2,868
Adjusted R-squared	0.0784

Note: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 5: Regression analysis to examine the association between restaurants’ average exposure and hotspots

reviews, and amenities count hence supporting our hypothesis $H2$. We ran the same statistical test per city for all nine cities to investigate if the same pattern is seen everywhere. Interestingly, we observed the same results: restaurants located in hotspots are more likely to have higher exposures, hence corroborating the findings in (Kokkodis and Lappas 2020; Huang et al. 2016). Hence, for each city, we could support our hypothesis $H2$. A complete overview of the results can be found in our Appendix .

Results for Examining $H3$

We used logistic regression models to examine $H3$ and if any relationship exists between the hotspots and demographic composition. Since every city has a unique structure and composition, we run the model for each city separately. While we analyzed all the cities individually (Appendix), due to brevity, we only present the results of Chicago and San Jose as they provided us with different findings.

Statistical Tests: We created binary variables to account for the demographic composition of each zip code in the city, namely White neighborhood (WN), Black neighborhood (BN), American-Indian neighborhood (AIN), Asian neighborhood (AN), highly educated neighborhood (HED), high unemployment neighborhood (HUne), and highly wealthy neighborhood (HWe). This was done due to in most cities, across various zip codes, the percentages of the demographic variables are close to each other, which makes it hard for the model to distinguish between them. This is a common technique used for when the relationship between a continuous predictor and the outcome is not linear. To obtain the binary values, we used the official U.S. Census dataset quickfacts website (Census 2023). We labeled ‘1’ if the percentage of the variable is above what is given on the quickfacts website and ‘0’ if it is not. The U.S. Census Quickfacts website gives the percentage of the white population in the city. Hence, rather than having the harsh threshold of 75.5% which represents the overall percentage of the white population in the US (Bureau 2023b), we used the percentage of the population of that city to give a more nuanced analysis. This is especially because we run the analysis on each city, and compare the demographic composition of zip codes in that city. For example, WN in the city of Chicago was labeled as ‘1’ if the percentage of the White population in the zip code was greater than or equal to 42.4%, which is the total percentage of White population in the city of Chicago. Whereas, WN in the city of San Jose was labeled as ‘1’ if the percentage

<i>Dependent variable: Hotspot</i>	
White Neighborhood (WN)	-16.485 (0.000)***
Black Neighborhood (BN)	-16.287 (0.000)***
American-Indian Neighborhood (AIN)	-0.142 (0.820)
Asian Neighborhood (AN)	-0.457 (0.446)
Highly Educated Neighborhood (HED)	14.957 (0.000)***
High Unemployment Neighborhood (HUne)	-0.600 (0.407)
Highly Wealthy Neighborhood (HWe)	15.706 (0.000)***

Note: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 6: Results of the regression analysis of hotspot with other sensitive attributes for Chicago

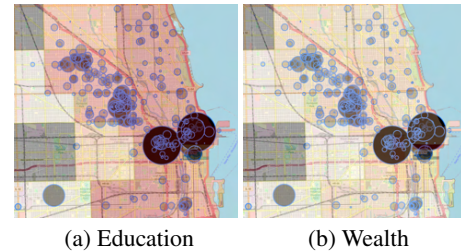


Figure 3: Heatmap for Chicago. The black circles represent individual restaurants’ exposures and colored squares represent the % of sensitive attributes. Darker the color, higher the % of the sensitive attribute and vice-versa.

of the White population in the zip code was greater than or equal to 32%, which is the total percentage of White population in the city of San Jose. We removed the Native Hawaiian (NHN) variable from the models because the sum of all of them would be 100%, making the variables not independent.

We employed clustered multivariate regression clustering by the zip code. This technique combines multivariate regression analysis with clustering methods to account for intra-cluster dependencies or correlations within the data. This approach is useful when dealing with data that exhibit clustering or grouping effects, where observations within the same cluster are likely to be more similar to each other than to observations in other clusters (Price and Sherwood 2018). Here, restaurants located within the same zip code share the demographic composition characteristic of that particular zip code, and therefore, using clustering by zip code helps to account for the dependencies within the data.

Chicago: We obtained 12 hotspots for Chicago. Interestingly, as it is shown in Table 6, White and Black neighborhoods are negatively correlated to hotspots. We suspect this is because there are many White and Black neighborhoods that are far from the touristy parts of Chicago. Still, when only considering the hotspot regions, we find that those are more likely to be in wealthier and higher-educated neighborhoods, hence we find support $H3$. Our statistical results can also visually be seen in Figure 3. We can see that major hotspots are in highly educated (shown in Fig. 3a) and wealthy neighborhoods (shown in Fig. 3b). Combining all the findings, we can conclude that businesses located in Chicago’s hotspots are more likely to experience higher average exposure, particularly in wealthy and highly educated neighborhoods. This finding has significant implications for

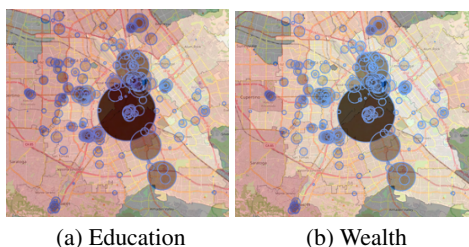


Figure 4: Heatmap for San Jose

businesses situated in less affluent and less educated areas; being outside of a hotspot may result in lower exposure and, consequently, reduced revenue.

San Jose: We obtained 10 hotspots in the city of San Jose. Interestingly, we did not see a statistically significant relationship between any of the sensitive demographic attributes and the hotspots for the city of San Jose. Therefore, while we could support $H2$ that businesses in hotspots have higher exposure, the results do not support $H3$.

This can be explained by the structure of the city as it is shown in Figure 4. This heatmap of San Jose visually shows that many regions of San Jose are characterized by higher levels of education and wealth, and hotspots are all scattered throughout the city (shown in Figure 4a and Figure 4b) and high-wealth neighborhoods. This is in contrast to Chicago, where we observed some segregated regions in terms of wealth and education, with many hotspots being concentrated in specific areas across the city.

Other Cities. For brevity, we do not provide the analysis for other cities. Doing the same analysis, we found support and partial support for $H3$ for Anchorage, New Orleans, San Francisco, and New York City. However, we did not find support for $H3$ for Corpus Christi, Los Angeles, and Seattle. We found that the cities that supported $H3$ tended to be more diverse, having less White population compared to the U.S. average of 75.5%. Moreover, they tended to have more regions with either higher education or wealth than the U.S. average of 34.3% and \$75,149, respectively. These results show that the existing disparities observed in the cities are transferred as bias into Yelp’s search or recommendation system, potentially creating a discriminatory feedback loop against businesses with less exposure.

Implications: Each city has unique demographic characteristics. Hotspots appear in different regions of cities with different demographic distributions. Depending on the associations between groups and hotspots, the ranking algorithm may behave differently for those groups, which explains the implicit bias of the algorithm for different sensitive attributes within each city.

Broader Implication: Our results highlight the importance of *responsible data science practices* in data-driven systems such as Yelp. Yelp and other platforms should open source their models with part of the training data for researchers to audit their systems (Singhal et al. 2023b).

Conclusion

In this work, using a large-scale data-driven approach, we audit Yelp’s business ranking and review recommendation software. We defined and examined three hypotheses to audit both systems. Using statistical and empirical analysis, we found that Yelp’s review recommendation system disproportionately filters reviews by less-established users into the not recommended section. We also discovered that there was a linear association between a restaurant being in a hotspot and having a higher exposure. Additionally, for some cities, we found association between hotspots and some demographic composition. This can create a discriminatory feedback loop against businesses with less exposure.

Limitations: Firstly, we could not obtain the list of unique users. Secondly, we did not analyze the ranking of restaurants for all the cities. Furthermore, we removed Detroit from our ranking analysis. We collected the Yelp reviews in early 2023. Hence, there is a possibility that Yelp might have changed some reviews from recommended to not recommended and vice versa, and our analysis does not capture that. Additionally, our model for finding the likelihood of a review to be fake is simple, as we did not have complex features to build a sophisticated model. Furthermore, we were not able to control for the number of recent reviews and the average star rating of recent reviews because of the limit of calls we could make to Yelp and when we tried using our crawler to get them, Yelp was actively blocking us for one, two or sometimes a week and even blocking requests when using a VPN.

Acknowledgments

This material is based upon work supported by the National Science Foundation under grants No. 2107296 and 2107290. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

- Abdollahpouri, H.; and Burke, R. 2019. Multi-stakeholder Recommendation and its Connection to Multi-sided Fairness. In *Proceedings of the Workshop on Recommendation in Multi-stakeholder Environments co-located with the 13th ACM Conference on Recommender Systems (RecSys 2019)*, Copenhagen, Denmark, September 20, 2019, volume 2440 of *CEUR Workshop Proceedings*. CEUR-WS.org.
- Abdollahpouri, H.; Mansoury, M.; Burke, R.; and Mobasher, B. 2019. The Unfairness of Popularity Bias in Recommendation. In Burke, R.; Abdollahpouri, H.; Malthouse, E. C.; Thai, K. P.; and Zhang, Y., eds., *Proceedings of the Workshop on Recommendation in Multi-stakeholder Environments co-located with the 13th ACM Conference on Recommender Systems (RecSys 2019)*, Copenhagen, Denmark, September 20, 2019, volume 2440 of *CEUR Workshop Proceedings*. CEUR-WS.org.
- Adler, P.; Falk, C.; Friedler, S. A.; Nix, T.; Rybeck, G.; Scheidegger, C.; Smith, B.; and Venkatasubramanian, S.

2018. Auditing black-box models for indirect influence. *Knowledge and Information Systems*, 54: 95–122.
- Ali, M.; Sapiezynski, P.; Bogen, M.; Korolova, A.; Mislove, A.; and Rieke, A. 2019. Discrimination through optimization: How Facebook’s Ad delivery can lead to biased outcomes. *Proceedings of the ACM on human-computer interaction*, 3(CSCW): 1–30.
- Amos, R.; Maio, R.; and Mittal, P. 2022. Reviews in motion: a large scale, longitudinal study of review recommendations on Yelp. *arXiv preprint arXiv:2202.09005*.
- Angwin, J.; Larson, J.; Mattu, S.; and Kirchner, L. 2016. Machine Bias. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
- Ashraf, S. A.; Javed, A. F.; Bellary, S.; Bala, P. K.; and Panigrahi, P. K. 2024. Leveraging Stacking Framework for Fake Review Detection in the Hospitality Sector. *Journal of Theoretical and Applied Electronic Commerce Research*, 19(2): 1517–1558.
- Asudeh, A.; Jagadish, H.; Stoyanovich, J.; and Das, G. 2019. Designing fair ranking schemes. In *Proceedings of the 2019 international conference on management of data*, 1259–1276.
- Asudeh, A.; Thirumuruganathan, S.; Zhang, N.; and Das, G. 2016. Discovering the Skyline of Web Databases. *Proceedings of the VLDB Endowment*, 9(7): 600–611.
- Becerril-Arreola, R. 2023. A Method to Assess and Explain Disparate Impact in Online Retailing. In *Proceedings of the ACM Web Conference 2023*, 3670–3679.
- Beutel, A.; Chen, J.; Doshi, T.; Qian, H.; Wei, L.; Wu, Y.; Heldt, L.; Zhao, Z.; Hong, L.; Chi, E. H.; et al. 2019. Fairness in recommendation ranking through pairwise comparisons. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2212–2220.
- Buolamwini, J.; and Gebru, T. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*, 77–91. PMLR.
- Bureau, U. S. C. 2023a. Census Data. <https://data.census.gov/>.
- Bureau, U. S. C. 2023b. Census Data. <https://www.census.gov/quickfacts/fact/table/US/PST045223>.
- Caton, S.; and Haas, C. 2020. Fairness in machine learning: A survey. *ACM Computing Surveys*.
- Census, U. 2023. QuickFacts. <https://www.census.gov/quickfacts/fact/table/US/RHI225222>.
- Chen, J.; Dong, H.; Wang, X.; Feng, F.; Wang, M.; and He, X. 2023. Bias and debias in recommender system: A survey and future directions. *ACM Transactions on Information Systems*, 41(3): 1–39.
- Chen, L.; De Gemmis, M.; Felfernig, A.; Lops, P.; Ricci, F.; and Semeraro, G. 2013. Human decision making and recommender systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 3(3): 1–7.
- Collins, A.; Tkaczyk, D.; Aizawa, A.; and Beel, J. 2018. A study of position bias in digital library recommender systems. *arXiv preprint arXiv:1802.06565*.
- Datta, A.; Sen, S.; and Zick, Y. 2016. Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems. In *2016 IEEE symposium on security and privacy (SP)*, 598–617. IEEE.
- Diaz, F.; Mitra, B.; Ekstrand, M. D.; Biega, A. J.; and Carterette, B. 2020. Evaluating stochastic rankings with expected exposure. In *Proceedings of the 29th ACM international conference on information & knowledge management*, 275–284.
- Dressel, J.; and Farid, H. 2018. The accuracy, fairness, and limits of predicting recidivism. *Science advances*, 4(1): eaa05580.
- Duffy, B. E.; and Meisner, C. 2022. Platform governance at the margins: Social media creators’ experiences with algorithmic (in) visibility. *Media, Culture & Society*, 01634437221111923.
- Duffy, B. E.; and Meisner, C. 2023. Platform governance at the margins: Social media creators’ experiences with algorithmic (in) visibility. *Media, Culture & Society*, 45(2): 285–304.
- Duma, R. A.; Niu, Z.; Nyamawe, A.; Tchaye-Kondi, J.; Chambua, J.; and Yusuf, A. A. 2024. DHMFRD–TER: a deep hybrid model for fake review detection incorporating review texts, emotions, and ratings. *Multimedia Tools and Applications*, 83(2): 4533–4549.
- Edizel, B.; Bonchi, F.; Hajian, S.; Panisson, A.; and Tassa, T. 2020. FaiRecSys: mitigating algorithmic bias in recommender systems. *International Journal of Data Science and Analytics*, 9: 197–213.
- Ekstrand, M. D.; Das, A.; Burke, R.; Diaz, F.; et al. 2022. Fairness in information access systems. *Foundations and Trends® in Information Retrieval*, 16(1-2): 1–177.
- Ekstrand, M. D.; and Pera, M. S. 2017. The demographics of cool. *Poster Proceedings at ACM RecSys. ACM, Como, Italy*.
- Eslami, M.; Vaccaro, K.; Lee, M. K.; Elazari Bar On, A.; Gilbert, E.; and Karahalios, K. 2019. User attitudes towards algorithmic opacity and transparency in online reviewing platforms. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–14.
- Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X.; et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, 226–231.
- Feldman, M.; Friedler, S. A.; Moeller, J.; Scheidegger, C.; and Venkatasubramanian, S. 2015. Certifying and removing disparate impact. In *proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, 259–268.
- Fleder, D.; and Hosanagar, K. 2009. Blockbuster culture’s next rise or fall: The impact of recommender systems on sales diversity. *Management science*, 55(5): 697–712.

- Flores, A. W.; Bechtel, K.; and Lowenkamp, C. T. 2016. False positives, false negatives, and false analyses: A rejoinder to machine bias: There's software used across the country to predict future criminals. and it's biased against blacks. *Fed. Probation*, 80: 38.
- Fu, Z.; Xian, Y.; Gao, R.; Zhao, J.; Huang, Q.; Ge, Y.; Xu, S.; Geng, S.; Shah, C.; Zhang, Y.; et al. 2020. Fairness-aware explainable recommendation over knowledge graphs. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 69–78.
- Google. 2023. Google Places API. <https://developers.google.com/maps/documentation/places/web-service/overview>.
- Haimson, O. L.; Delmonaco, D.; Nie, P.; and Wegner, A. 2021. Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2): 1–35.
- Hannák, A.; Wagner, C.; Garcia, D.; Mislove, A.; Strohmaier, M.; and Wilson, C. 2017. Bias in online freelance marketplaces: Evidence from taskrabit and fiverr. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, 1914–1933.
- Harris, C. G. 2019. Comparing human computation, machine, and hybrid methods for detecting hotel review spam. In *Digital Transformation for a Sustainable Society in the 21st Century: 18th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2019, Trondheim, Norway, September 18–20, 2019, Proceedings 18*, 75–86. Springer.
- Harrison, G.; Hanson, J.; Jacinto, C.; Ramirez, J.; and Ur, B. 2020. An empirical study on the perceived fairness of realistic, imperfect machine learning models. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 392–402.
- Horn, I. S.; Taros, T.; Dirkes, S.; Hüer, L.; Rose, M.; Tietmeyer, R.; and Constantinides, E. 2015. Business reputation and social media: A primer on threats and responses. *Journal of direct, data and digital marketing practice*, 16: 193–208.
- Huang, N.; Burtch, G.; Hong, Y.; and Polman, E. 2016. Effects of multiple psychological distances on construal and consumer evaluation: A field study of online reviews. *Journal of Consumer Psychology*, 26(4): 474–482.
- Hussain, N.; Turab Mirza, H.; Rasool, G.; Hussain, I.; and Kaleem, M. 2019. Spam review detection techniques: A systematic literature review. *Applied Sciences*, 9(5): 987.
- Inc., Y. 2019. Yelp's 15th Anniversary. <https://www.yelp15.com/>.
- Inc., Y. 2022. Recommendation software. <https://trust.yelp.com/recommendation-software/>.
- Inc., Y. 2023. Yelp Open Dataset. <https://www.yelp.com/dataset>.
- Jhaver, S.; Appling, D. S.; Gilbert, E.; and Bruckman, A. 2019. "Did you suspect the post would be removed?" Understanding user reactions to content removals on Reddit. *Proceedings of the ACM on human-computer interaction*, 3(CSCW): 1–33.
- Joachims, T.; Granka, L.; Pan, B.; Hembrooke, H.; and Gay, G. 2017. Accurately interpreting clickthrough data as implicit feedback. In *Acm Sigir Forum*, volume 51, 4–11. Acm New York, NY, USA.
- Kamerer, D. 2014. Understanding the Yelp review filter: An exploratory study. *First Monday*.
- Kim, M. P.; Ghorbani, A.; and Zou, J. 2019. Multiaccuracy: Black-box post-processing for fairness in classification. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 247–254.
- Koenecke, A.; Giannella, E.; Willer, R.; and Goel, S. 2023. Popular Support for Balancing Equity and Efficiency in Resource Allocation: A Case Study in Online Advertising to Increase Welfare Program Awareness. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 17, 494–506.
- Koenker, R. 2004. Quantile regression for longitudinal data. *Journal of multivariate analysis*, 91(1): 74–89.
- Kokkodis, M.; and Lappas, T. 2020. Your hometown matters: Popularity-difference bias in online reputation platforms. *Information Systems Research*, 31(2): 412–430.
- Kwak, S. G.; and Kim, J. H. 2017. Central limit theorem: the cornerstone of modern statistics. *Korean journal of anesthesiology*, 70(2): 144–156.
- Lazovich, T.; Belli, L.; Gonzales, A.; Bower, A.; Tantipongpipat, U.; Lum, K.; Huszar, F.; and Chowdhury, R. 2022. Measuring disparate outcomes of content recommendation algorithms with distributional inequality metrics. *Patterns*, 3(8): 100568.
- Li, Y.; Chen, H.; Fu, Z.; Ge, Y.; and Zhang, Y. 2021a. User-oriented fairness in recommendation. In *Proceedings of the Web Conference 2021*, 624–632.
- Li, Y.; Chen, H.; Xu, S.; Ge, Y.; and Zhang, Y. 2021b. Towards personalized fairness based on causal notion. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1054–1063.
- Luca, M. 2016. Reviews, reputation, and revenue: The case of Yelp. com. *Com (March 15, 2016). Harvard Business School NOM Unit Working Paper*, (12-016).
- Marlin, B.; Zemel, R. S.; Roweis, S.; and Slaney, M. 2012. Collaborative filtering and the missing at random assumption. *arXiv preprint arXiv:1206.5267*.
- Mohawesh, R.; Salameh, H. B.; Jararweh, Y.; Alkhalaileh, M.; and Maqsood, S. 2024. Fake review detection using transformer-based enhanced LSTM and RoBERTa. *International Journal of Cognitive Computing in Engineering*, 5: 250–258.
- Mukherjee, A.; Venkataraman, V.; Liu, B.; and Glance, N. 2013a. What yelp fake review filter might be doing? In *Proceedings of the international AAAI conference on web and social media*, volume 7, 409–418.
- Mukherjee, A.; Venkataraman, V.; Liu, B.; Glance, N.; et al. 2013b. Fake review detection: Classification and analysis of

- real and pseudo reviews. *UIC-CS-03-2013. Technical Report*.
- Narayanan, A. 2018. Tutorial: 21 fairness definitions and their politics.(2018). <https://www.youtube.com/watch>.
- Nilizadeh, S.; Aghakhani, H.; Gustafson, E.; Kruegel, C.; and Vigna, G. 2019. Think outside the dataset: Finding fraudulent reviews using cross-dataset analysis. In *The World Wide Web Conference*, 3108–3115.
- of Labor Statistics, U. B. 2019. New Orleans had the highest unemployment rate among large metro areas in June 2019. <https://www.bls.gov/opub/ted/2019/new-orleans-had-highest-unemployment-rate-among-large-metro-areas-in-june-2019.htm>.
- O'Brien, M.; and Keane, M. T. 2006. Modeling result-list searching in the World Wide Web: The role of relevance topologies and trust bias. In *Proceedings of the 28th annual conference of the cognitive science society*, volume 28, 1881–1886. Citeseer.
- Patro, G. K.; Porcaro, L.; Mitchell, L.; Zhang, Q.; Zehlike, M.; and Garg, N. 2022. Fair ranking: a critical review, challenges, and future directions. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, 1929–1942.
- Price, B. S.; and Sherwood, B. 2018. A cluster elastic net for multivariate regression. *Journal of Machine Learning Research*, 18(232): 1–39.
- Rahman, M. F.; Asudeh, A.; Koudas, N.; and Das, G. 2017. Efficient computation of subspace skyline over categorical domains. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 407–416.
- Salminen, J.; Kandpal, C.; Kamel, A. M.; Jung, S.-g.; and Jansen, B. J. 2022. Creating and detecting fake reviews of online products. *Journal of Retailing and Consumer Services*, 64: 102771.
- Seering, J.; Wang, T.; Yoon, J.; and Kaufman, G. 2019. Moderator engagement and community development in the age of algorithms. *New Media & Society*, 21(7): 1417–1443.
- Silva, N.; Silva, T.; Werneck, H.; Rocha, L.; and Pereira, A. 2023. User cold-start problem in multi-armed bandits: When the first recommendations guide the user's experience. *ACM Transactions on Recommender Systems*, 1(1): 1–24.
- Singh, A.; and Joachims, T. 2018. Fairness of exposure in rankings. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, 2219–2228.
- Singhal, M.; Kumarswamy, N.; Kinhekar, S.; and Nilizadeh, S. 2023a. Cybersecurity misinformation detection on social media: Case studies on phishing reports and zoom's threat. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 17, 796–807.
- Singhal, M.; Ling, C.; Paudel, P.; Thota, P.; Kumarswamy, N.; Stringhini, G.; and Nilizadeh, S. 2023b. SoK: Content moderation in social media, from guidelines to enforcement, and research to practice. In *2023 IEEE 8th European Symposium on Security and Privacy (EuroS&P)*, 868–895. IEEE.
- Stoyanovich, J.; Yang, K.; and Jagadish, H. 2018. Online set selection with fairness and diversity constraints. In *Proceedings of the EDBT Conference*.
- Support, Y. 2024. What is Yelp's Elite Squad? https://www.yelp-support.com/article/What-is-Yelps-Elite-Squad?l=en_US.
- Sweeney, L. 2013. Discrimination in online ad delivery. *Communications of the ACM*, 56(5): 44–54.
- Tan, S.; Caruana, R.; Hooker, G.; and Lou, Y. 2018. Distill-and-compare: Auditing black-box models using transparent model distillation. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 303–310.
- Washcovich, E. 2019. Understanding Yelp search. <https://blog.yelp.com/businesses/the-guide-to-understanding-yelp-search/>.
- Yao, Y.; Angelov, I.; Rasmus-Vorrath, J.; Lee, M.; and Engels, D. W. 2018. Yelp's Review Filtering Algorithm. *SMU Data Science Review*, 1(3): 3.
- Yelp. 2023. Do reviews that aren't currently recommended impact the business's star rating? <https://www.yelp-support.com/article/Do-reviews-that-arent-currently-recommended-impact-the-businesss-star-rating>.
- Yelp. 2024. Yelp Terms of Service. https://terms.yelp.com/tos/en_us/20240710_en_us/.
- Yu, K.; Lu, Z.; and Stander, J. 2003. Quantile regression: applications and current research areas. *Journal of the Royal Statistical Society Series D: The Statistician*, 52(3): 331–350.
- Zehlike, M.; Bonchi, F.; Castillo, C.; Hajian, S.; Megahed, M.; and Baeza-Yates, R. 2017. Fa* ir: A fair top-k ranking algorithm. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 1569–1578.
- Zehlike, M.; and Castillo, C. 2020. Reducing disparate exposure in ranking: A learning to rank approach. In *Proceedings of the web conference 2020*, 2849–2855.
- Zehlike, M.; Yang, K.; and Stoyanovich, J. 2022a. Fairness in Ranking, Part I: Score-Based Ranking. *ACM Comput. Surv.*, 55(6).
- Zehlike, M.; Yang, K.; and Stoyanovich, J. 2022b. Fairness in ranking, part ii: Learning-to-rank and recommender systems. *ACM Computing Surveys*, 55(6): 1–41.
- Zhang, X.; and Liu, M. 2021. Fairness in learning-based sequential decision algorithms: A survey. In *Handbook of Reinforcement Learning and Control*, 525–555. Springer.
- Zhao, J.; Wang, T.; Yatskar, M.; Ordonez, V.; and Chang, K.-W. 2017. Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2979–2989.
- Zheng, Y.; Gao, C.; Li, X.; He, X.; Li, Y.; and Jin, D. 2021. Disentangling user interest and conformity for recommendation with causal embedding. In *Proceedings of the Web Conference 2021*, 2980–2991.

Ethics Checklist

1. For most authors...

- (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? **Yes**
 - (b) Do your main claims in the abstract and introduction accurately reflect the paper’s contributions and scope? **Yes**
 - (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? **Yes**
 - (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? **NA**
 - (e) Did you describe the limitations of your work? **Yes**
 - (f) Did you discuss any potential negative societal impacts of your work? **No, because we ideate on positive inference from our findings.**
 - (g) Did you discuss any potential misuse of your work? **NA**
 - (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? **Yes**
 - (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? **Yes**
2. Additionally, if your study involves hypotheses testing...
- (a) Did you clearly state the assumptions underlying all theoretical results? **Yes, in the Hypothesis for each of the systems.**
 - (b) Have you provided justifications for all theoretical results? **Yes, in the Hypothesis for each of the systems.**
 - (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? **Yes, in the Hypothesis for each of the systems.**
 - (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? **Yes, in the individual analysis Result section.**
 - (e) Did you address potential biases or limitations in your theoretical framework? **Yes, in the Hypothesis for each of the systems.**
 - (f) Have you related your theoretical results to the existing literature in social science? **Yes, in the Hypothesis for each of the systems.**
 - (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? **Yes, in the Result section for each system.**
3. Additionally, if you are including theoretical proofs... **NA. Our study does not include theoretical proofs.**
4. Additionally, if you ran machine learning experiments... **NA. Our study does not involve any machine learning algorithm.**
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
- (a) If your work uses existing assets, did you cite the creators? **Yes**
 - (b) Did you mention the license of the assets? **NA**
 - (c) Did you include any new assets in the supplemental material or as a URL? **Yes. We include an anonymous Dropbox URL to the dataset in this submission.**
 - (d) Did you discuss whether and how consent was obtained from people whose data you’re using/curating? **NA**
 - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? **Yes**
 - (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR? **Yes, we are committed to abide by FAIR principles when sharing our dataset upon paper acceptance.**
 - (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset **Yes, we commit to creating a Datasheet when sharing our dataset upon paper acceptance.**
6. Additionally, if you used crowdsourcing or conducted research with human subjects... **NA. Our study does not involve research with human subjects/participants.**

Ethical Statement

We only gathered reviews from businesses, and public restaurants, and did not attempt to access or find any accounts on Yelp. We also made no attempts to maliciously gather this data, instead, we used the same backend APIs that a user browser would request data from.

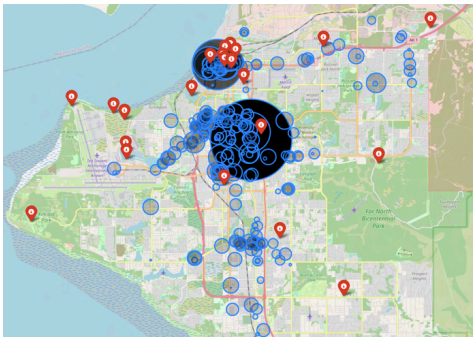
Appendix

Exposure of each city

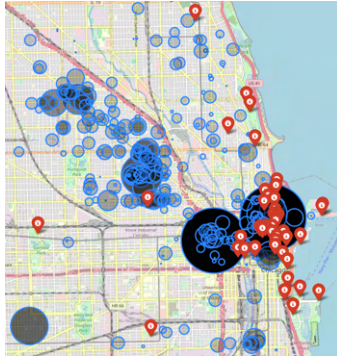
Figure 5 shows the exposure of restaurants per city. The circles shown in the figure show the exposure of each restaurant. The bigger the size, more is the exposure of that restaurant.

Anchorage: Figure 5c shows the exposure of restaurants in the city of Anchorage. By employing the DBSCAN algorithm, we obtained a total of 111 clusters. We found that there were 64 clusters of size 1. We obtained a total of 8 hotspots for the city of Anchorage. We then performed a linear regression to see if statistically there is a correlation between average exposure and hotspots.

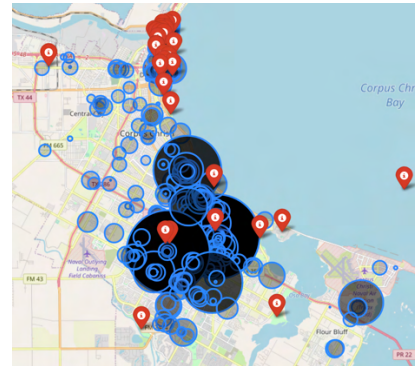
Table 7 shows the results. We observed a consistent result with that of the analysis carried out on the whole dataset, where is a positive correlation between average exposure and the restaurant being in a hotspot, hence we find support for *H2*. We then investigated if a restaurant that is in a hotspot has any relation with demographic features such as racial composition, percentage of educated people, percentage of unemployed people, and percentage of wealthy people i.e., *H3*. Table 8 shows the results from our regression model. We performed the analysis, by clustering them based on zip codes. We can see that restaurants that are in hotspots have a higher Black and American Indian population, they



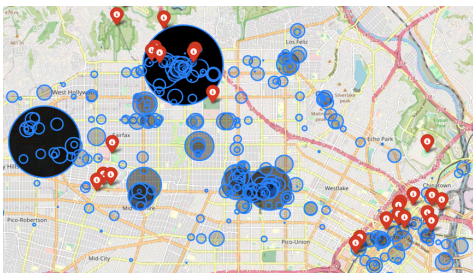
(a) Anchorage



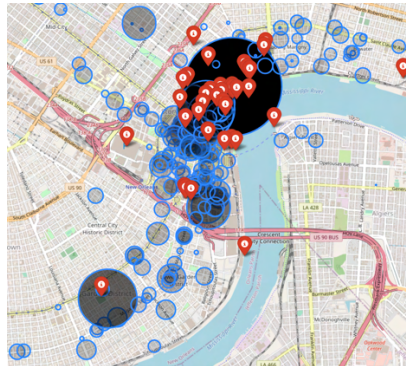
(b) Chicago



(c) Corpus Christi



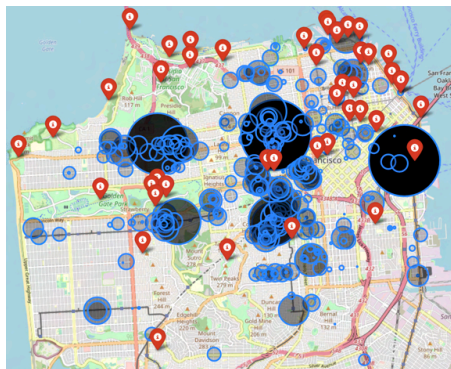
(d) Los Angeles



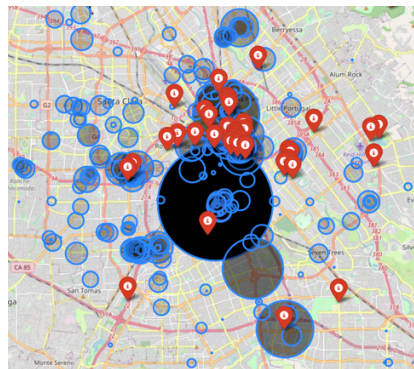
(e) New Orleans



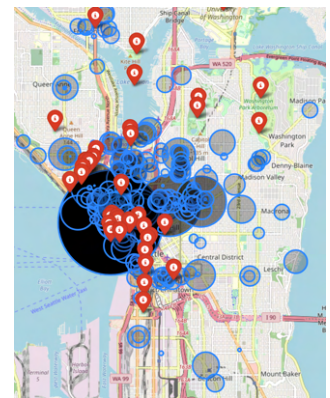
(f) New York City



(g) San Francisco



(h) San Jose



(i) Seattle

Figure 5: Exposure of restaurants in each city. Note that the red marker in the figure shows the popular attraction in each city

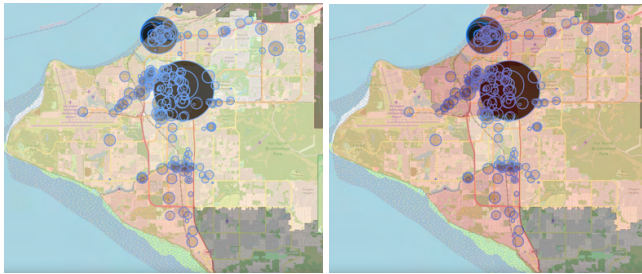
Table 7: Results of the regression analysis of average exposure with hotspot for Anchorage

Dependent variable:	
Average Exposure	
Hotspot	0.033 (0.001)**
Observations	291
Adjusted R-squared	0.03078
Note:	*p<0.05;
	p<0.01; *p<0.001

Table 8: Results of the regression analysis of hotspot with other sensitive attributes for Anchorage

Dependent variable:	
Hotspot	
WN	-0.658 (0.003)***
BN	0.512 (0.000)***
AIN	0.972 (0.000)***
AN	-0.514 (0.000)***
HED	0.830 (0.001)***
HUne	-1.720 (0.000)***
HWe	0.170 (0.595)
Note:	*p<0.05;
p<0.01;	*p<0.001

are also in highly educated and wealthy neighborhoods and hence we find partial support for *H3*



(a) White Population (b) Education

Figure 6: Heatmap for Anchorage

Figure 6 shows the map for educated and white neighborhoods in Anchorage. Visually looking at the map, it confirms our statistical analysis, as restaurants that are in hotspots have higher education (shown in Fig. 6b) and restaurants that are in predominantly white neighborhoods are not in hotspots (shown in Fig. 6a). Therefore, we can draw a conclusion about Yelp in the city of Anchorage, businesses that are in the hotspots, have higher average exposure, and businesses that are in the zip codes with a hotspot, have a higher black and American Indian population, they are also in highly educated and wealthy neighborhoods.

Chicago: Figure 5b shows the exposure of restaurants in the city of Chicago. We obtained 12 hotspots of size 6 or more. Table 9 shows the results of our linear regression model.

We again see a very similar result, where there is a positive correlation between average exposure and hotspot,

Table 9: Results of the regression analysis of average exposure with hotspot for Chicago

Dependent variable:	
average exposure	
Linear	
Hotspot	0.037 (0.000)***
Observations	329
Adjusted R-squared	0.0401
Note:	*p<0.05; **p<0.01; ***p<0.001

hence we find support for *H2*.

Corpus Christi: Figure 5c shows the exposure of restaurants in the city of Corpus Christi. Using the DBSCAN algorithm, we were able to identify 132 clusters. Using our threshold, we found 7 hotspots that were of size six or more. Table 10 shows the results of our linear regression model.

Table 10: Results of the regression analysis of average exposure with hotspot for Corpus Christi

Dependent variable:	
average exposure	
Linear	
Hotspot	0.027 (0.024)*
Observations	271
Adjusted R-squared	0.01502
Note:	*p<0.05; **p<0.01; ***p<0.001

We can see that average exposure and hotspots have a positive correlation. Table 11 shows the results of our logistic regression model. We see that restaurants in zip codes that are in a hotspot are in highly educated neighborhoods and have less number of American Indians and Asian populations. This is because Corpus Christi has a high Hispanic population. We can see the results in Figure. 7. We can visually see that hotspots are in areas that are highly educated neighborhoods. Therefore, we can draw a conclusion about Yelp in the city of Corpus Christi, businesses that are in the hotspots, have higher average exposure, and businesses that are in zip codes that are in hotspots, are in highly educated neighborhoods and have less number of American Indians and Asian population, hence, while we find support for *H2*, we rejected our *H3*.

Los Angeles: Figure 5d shows the exposure of restaurants in the city of Los Angeles. Using the DBSCAN algorithm, we were able to identify 172 clusters. Using our threshold, we obtained 8 hotspots. Table 12 shows the re-

Table 11: Results of the regression analysis of hotspot with other sensitive attributes for Corpus Christi

Dependent variable:	
Hotspot	
Logistic	
WN	0.854 (0.457)
BN	0.730 (0.444)
AIN	-15.037 (0.000)***
AN	-13.889 (0.000)***
HED	13.668 (0.000)***
HUne	-0.241 (0.810)
HWe	-0.504 (0.601)
Note:	*p<0.05; **p<0.01; ***p<0.001

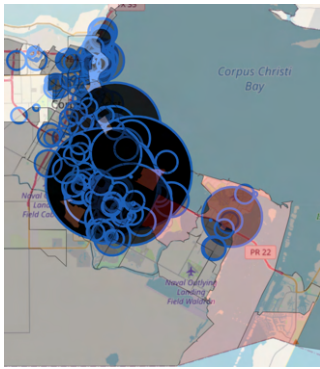


Figure 7: Education heatmap for Corpus Christi

sults of our linear regression model. We can see similar re-

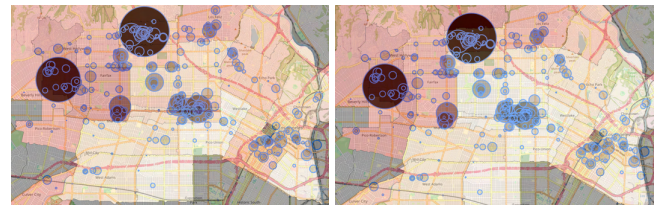
Table 12: Results of the regression analysis of average exposure with hotspot for Los Angeles

<i>Dependent variable:</i>	
<i>average exposure</i>	
<i>Linear</i>	
Hotspot	0.027 (0.009)**
Observations	344
Adjusted R-squared	0.01676
Note:	*p<0.05; **p<0.01; ***p<0.001

sults, that were obtained in the previous cities, where average exposure is positively correlated to hotspots of the city. Table 13 shows the results of our logistic regression model. We see that restaurants in zip codes that are in a hotspot are in highly educated neighborhoods. Interestingly none of the demographic features were significant with hotspot. Figure 8 shows the heatmap for Los Angeles. We can see visually that major hotspots are in highly educated (shown in Figure 8a) and restaurants with higher exposure are more in less white neighborhoods, hence confirming the validity of our regression analysis (shown in Figure 8b). It should also be noted that the areas which are highly educated are Beverly Crest and Westwood where they have about 67.6% and 66.5% adults who have earned a four-year degree or higher respectively. Therefore, we can draw a conclusion about Yelp in the city of Los Angeles, businesses that are in the hotspots, have higher average exposure, supporting our $H2$, and busi-

Table 13: Results of the regression analysis of hotspot with other sensitive attributes for Los Angeles

<i>Dependent variable:</i>	
<i>Hotspot</i>	
<i>Logistic</i>	
WN	- 0.387 (0.648)
BN	0.088 (0.924)
AIN	1.181 (0.186)
AN	0.505 (0.408)
HED	16.627 (0.000)***
HUne	-2.087 (0.007)
HWe	0.680 (0.541)
Note:	*p<0.05; **p<0.01; ***p<0.001



(a) Education (b) White Population

Figure 8: Heatmap for Los Angeles

nesses that are in the zip codes that are in a hotspot, are in highly educated neighborhoods, hence we can reject $H3$.

New Orleans: Figure 5e shows the exposure of restaurants in the city of New Orleans. Using the DBSCAN algorithm, we were able to identify 87 clusters. Using our threshold, we obtained 3 hotspots. Table 14 shows the results of our linear regression model. We can see similar results, that

Table 14: Results of the regression analysis of average exposure with hotspot for New Orleans

<i>Dependent variable:</i>	
<i>average exposure</i>	
<i>Linear</i>	
Hotspot	0.039 (0.000)***
Observations	281
Adjusted R-squared	0.04479
Note:	*p<0.05; **p<0.01; ***p<0.001

were obtained in the previous cities, where average exposure is positively correlated to the hotspots of the city. Table 15 shows the results of our logistic regression model. We see that restaurants in zip codes that are in a hotspot are in highly white, black neighborhoods, they are also in highly educated and highly unemployed areas. This is interesting, however, New Orleans had the highest unemployment rate among large metro areas according to U.S. Bureau of Labor Statistics (of Labor Statistics 2019). Figure 9 shows the heatmap for New Orleans. We can see visually that major hotspots are in highly educated (shown in Figure 9a) and highly unemployed neighborhoods, (shown in Figure 9b). Therefore, we can draw a conclusion about Yelp in the city of New Orleans, businesses that are in hotspots, have higher average exposure, hence finding support for $H2$, and busi-

Table 15: Results of the regression analysis of hotspot with other sensitive attributes for New Orleans

<i>Dependent variable:</i>	
<i>Hotspot</i>	
<i>Logistic</i>	
WN	7.261 (0.000)***
BN	8.379 (0.000)***
AIN	-4.217 (0.000)***
AN	0.754 (0.329)
HED	4.067 (0.000)***
HUne	3.296 (0.000)***
HWe	1.217 (0.140)
Note:	*p<0.05; **p<0.01; ***p<0.001

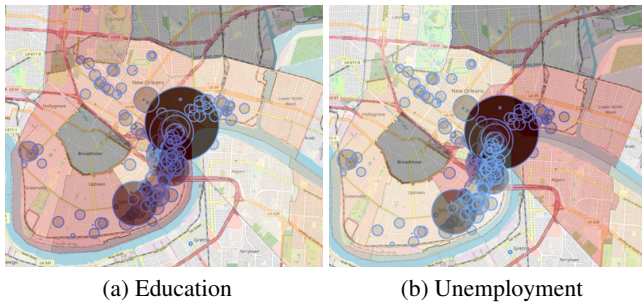


Figure 9: Heatmap for New Orleans

nesses in zip codes where there is a hotspot, are in highly educated and highly unemployed neighborhoods and diverse neighborhoods, hence we find partial support for $H3$.

New York City: Figure 5f shows the exposure of restaurants in the city of New York City. Using the DBSCAN algorithm, we were able to identify 131 clusters. Using our threshold, we obtained 9 hotspots in the city of New York. Table 16 shows the results of our linear regression model.

Table 16: Results of the regression analysis of average exposure with hotspot for New York City

<i>Dependent variable:</i>	
<i>average exposure</i>	
<i>Linear</i>	
Hotspot	0.031 (0.001)**
Observations	348
Adjusted R-squared	0.02542
Note:	*p<0.05; **p<0.01; ***p<0.001

Table 17: Results of the regression analysis of hotspot with other sensitive attributes for New York

<i>Dependent variable:</i>	
<i>Hotspot</i>	
<i>Logistic</i>	
WN	-2.077 (0.018)*
BN	-2.046 (0.027)*
AIN	0.597 (0.359)
AN	1.218 (0.031)*
HED	N/A
HU _{ne}	0.186 (0.749)
HWe	2.197 (0.018)*
Note:	*p<0.05; **p<0.01; ***p<0.001

We can see similar results, that were obtained in the previous cities, where average exposure is positively correlated to hotspots of the city. Table 17 shows the results of our logistic regression model. We see that restaurants in zip codes that are in a hotspot are highly Asian and highly wealthy neighborhoods. We also observed that restaurants in zip codes that are in a hotspot have lower numbers of white and black populations. Interestingly, we didn't find any results for education. Figure 10 shows the heatmap for New York City. We can visually see that major hotspots are all scattered in highly educated neighborhoods (shown in Figure. 10a), hence confirming why we did not obtain any result from

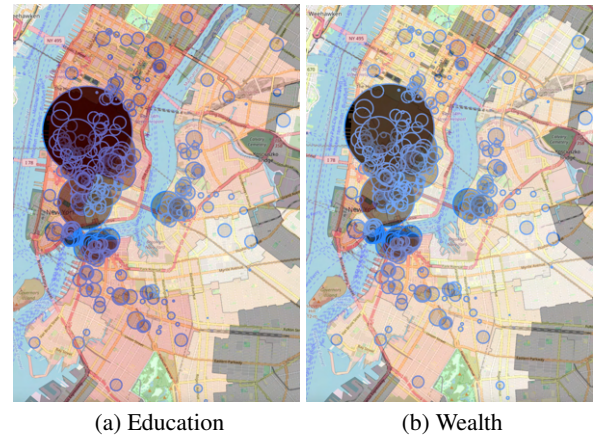


Figure 10: Heatmap for New York City

our regression model and highly wealthy neighborhoods, (shown in Figure 10b). Therefore, we can draw a conclusion about Yelp in the city of New York, businesses that are in the hotspots, have higher average exposure, hence finding support for $H2$, and businesses that are in zip codes that are in a hotspot, are in highly wealthy and Asian neighborhoods and they are also in the neighborhoods that have lower white and black populations, hence we partially find support for $H3$.

San Francisco: Figure 5g shows the exposure of restaurants in the city of San Francisco. Using the DBSCAN algorithm, we were able to identify 120 clusters. Using our threshold, we obtained 13 hotspots in the city of San Francisco. Table 18 shows the results of our linear regression model. We can see similar results, that were obtained in the

Table 18: Results of the regression analysis of average exposure with hotspot for San Francisco

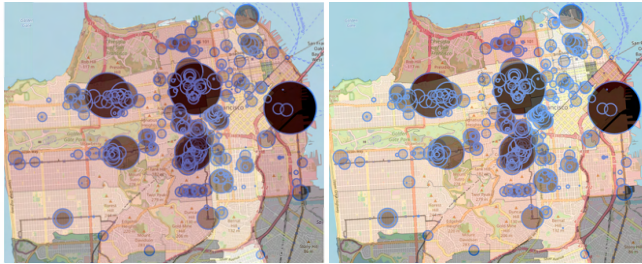
<i>Dependent variable:</i>	
<i>average exposure</i>	
<i>Linear</i>	
Hotspot	0.021 (0.020)*
Observations	353
Adjusted R-squared	0.01239
Note:	*p<0.05; **p<0.01; ***p<0.001

Table 19: Results of the regression analysis of hotspot with other sensitive attributes for San Francisco

<i>Dependent variable:</i>	
<i>Hotspot</i>	
<i>Logistic</i>	
WN	-1.457 (0.144)
BN	0.027 (0.963)
AIN	-0.786 (0.303)
AN	-1.812 (0.225)
HED	13.214 (0.000)***
HU _{ne}	0.826 (0.368)
HWe	1.491 (0.049)*
Note:	*p<0.05; **p<0.01; ***p<0.001

previous cities, where average exposure is positively corre-

lated to the hotspots of the city. Table 19 shows the results of our logistic regression model. We see that restaurants in zip codes that are in a hotspot are in highly educated and are in highly wealth neighbourhoods. Interestingly none of the demographic groups have any correlation with hotspots.



(a) Education (b) Wealth

Figure 11: Heatmap for San Francisco

Figure 11 shows the heatmap for San Francisco. We can visually see that major hotspots are all scattered in highly educated (shown in Figure 11a) and highly wealthy neighborhoods, (shown in Figure 11b). Therefore, we can draw a conclusion about Yelp in the city of San Francisco, businesses that are in the hotspots, have higher average exposure, and businesses that are in zip codes that are in a hotspot, are in highly wealthy and highly educated neighborhoods. While we do find support for *H2*, we were only able to partially find support for *H3*.

San Jose: Figure 5h shows the exposure of restaurants in the city of San Jose. Using the DBSCAN algorithm, we were able to identify 151 clusters. Using our threshold, we obtained 10 hotspots in the city of San Jose. Table 20 shows the results of our linear regression model. We can see sim-

Table 20: Results of the regression analysis of average exposure with hotspot for San Jose

<i>Dependent variable:</i>	
<i>average exposure</i>	
<i>Linear</i>	
Hotspot	0.027 (0.005)**
Observations	346
Adjusted R-squared	0.01239
Note:	*p<0.05; **p<0.01; ***p<0.001

ilar results, that were obtained in the previous cities, where average exposure is positively correlated to hotspots of the city, hence we find support for *H2*.

Seattle: Figure 5i shows the exposure of restaurants in the city of Seattle. Using the DBSCAN algorithm, we were able to identify 103 clusters. Using our threshold, we obtained 11 hotspots in the city of San Jose. Table 21 shows the results of our linear regression model.

We can see similar results, that were obtained in the previous cities, where average exposure is positively correlated to hotspots of the city. Table 22 shows the results of our logistic regression model. We find that restaurants in zip codes that are in a hotspot have a smaller American Indian population and are in less wealthy neighborhoods.

Table 21: Results of the regression analysis of average exposure with hotspot for Seattle

<i>Dependent variable:</i>	
<i>average exposure</i>	
<i>Linear</i>	
Hotspot	0.023 (0.025)*
Observations	305
Adjusted R-squared	0.01312
Note:	*p<0.05; **p<0.01; ***p<0.001

Table 22: Results of the regression analysis of hotspot with other sensitive attributes for Seattle

<i>Dependent variable:</i>	
<i>Hotspot</i>	
<i>Logistic</i>	
WN	-3.145 (0.092)
BN	-0.122 (0.933)
AIN	-2.629 (0.016)*
AN	-0.709 (0.659)
HED	-0.441 (0.795)
HUne	0.059 (0.965)
HWe	-2.066 (0.000)***
Note:	*p<0.05; **p<0.01; ***p<0.001

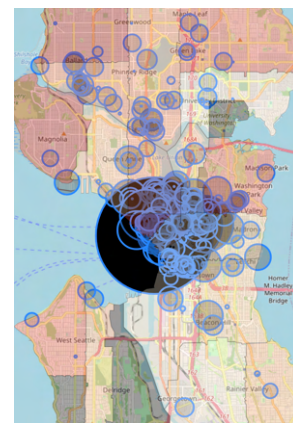


Figure 12: Wealth heatmap for Seattle

Figure 12 shows the heatmap of wealth for the city of Seattle. We can clearly see that the hotspots are in areas where the wealth is less, hence our analysis stands true. Therefore, we can draw a conclusion about Yelp in the city of Seattle, businesses that are in the hotspots, have higher average exposure, and businesses that are in zip codes that are in a hotspot, are in less wealthy neighborhoods and have less number of American Indian population. While we find support for $H2$, we reject our $H3$.