APPENDIX

A COMPLIMENTARY EXPERIMENT RESULTS

The following report presents the remaining results from the paper.

A.1 Exposure of each city

Figure 6 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 6c 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 9: Results of the regression analysis of average exposure with hotspot for Anchronage

	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 10: Results of the regression analysis of hotspot with other sensitive attributes for Anchronage

	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

Table 9 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 *H3*. 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., *H4*. To run our logistic regression model, we created some new binary variables, namely white neighborhood (WN), black neighborhood (BN), American-Indian neighborhood (AIN), Asian neighborhood (AN), high educated neighborhood (HED), high unemployment neighborhood (HUne), and high wealth neighborhood (HWe). Table 10 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 are also in highly educated and wealthy neighborhoods and hence we find partial support for *H4*

Figure 7 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. 7b) and restaurants that are in predominantly white neighborhoods are not in hotspots (shown in Fig. 7a). 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 6b shows the exposure of restaurants in the city of Chicago. We obtained 12 hotspots of size 6 or more. Table 11 shows the results of our linear regression model.

Table 11: 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

We again see a very similar result, where there is a positive correlation between average exposure and hotspot, hence we find support for *H3*.

Corpus Christi: Figure 6c 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 12 shows the results of our linear regression model. We can see that average exposure and hotspots have a positive correlation. 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 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. 8. 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 *H3*, we rejected our *H4*.

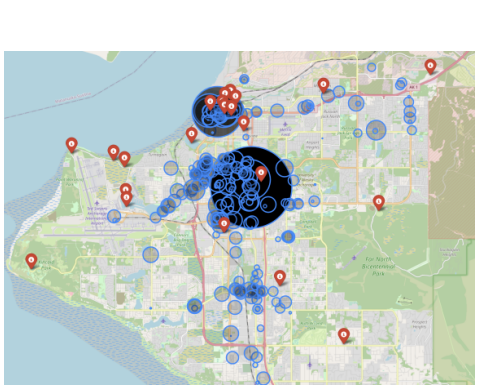
Table 12: Results of the regression analysis of average exposure with hotspot for Corpus Christi

	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

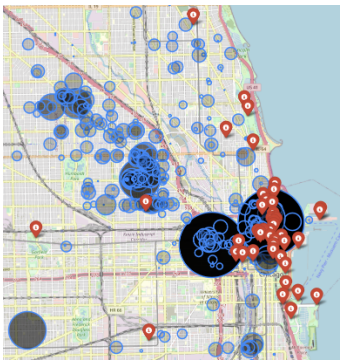
Table 13: Results of the regression analysis of hotspot with other sensitive attributes for Corpus Christi

	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

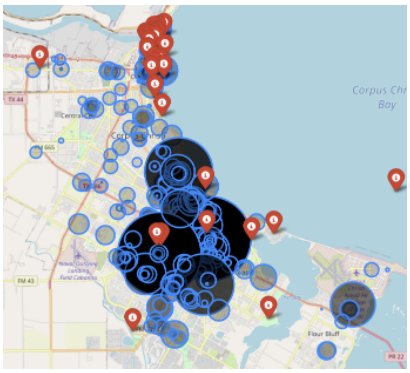
Los Angeles: Figure 6d shows the exposure of restaurants in the city of Los Angeles. Using the DBSCAN algorithm, we were able to identify 172 clusters. Using our threshold, we obtained 8 hotspots. Table 14 shows the results of our linear regression model. We can see similar results, that were obtained in the previous cities, where average exposure is positively correlated to hotspots of the



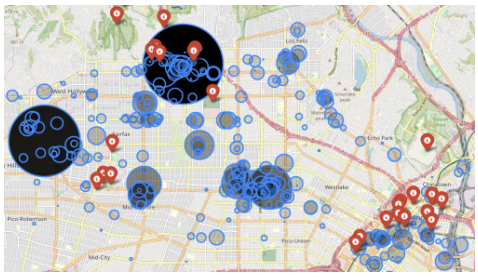
(a) Anchorage



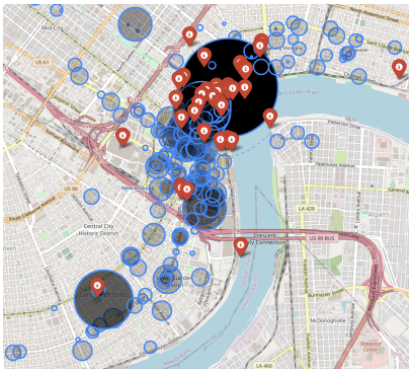
(b) Chicago



(c) Corpus Christi

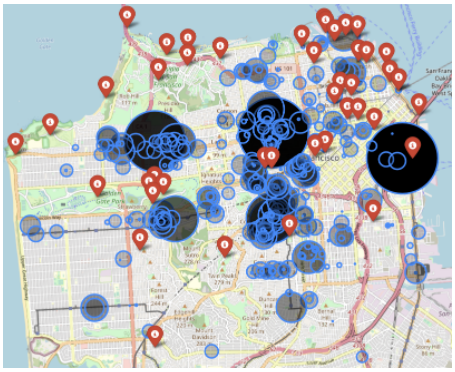


(d) Los Angeles

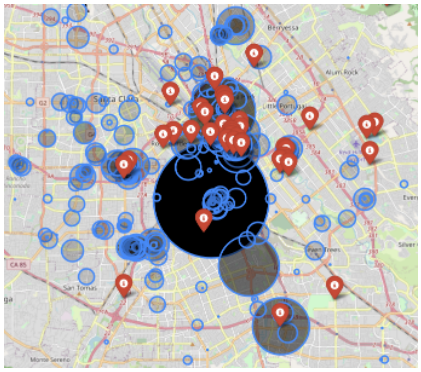


(e) New Orleans

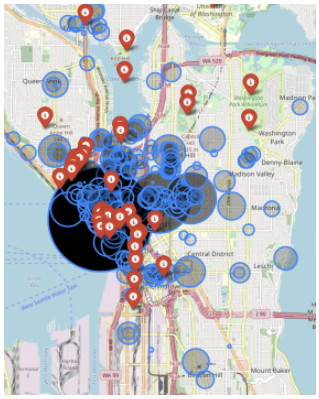
(f) New York City



(g) San Francisco

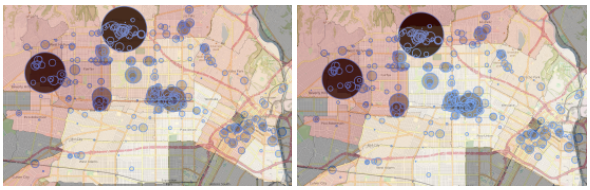


(h) San Jose



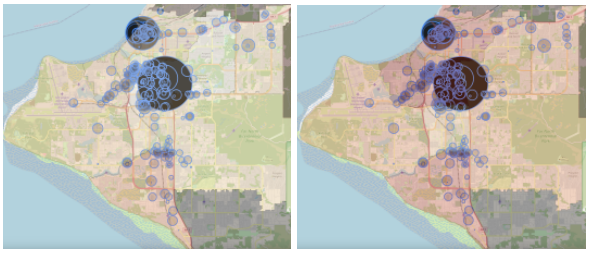
(i) Seattle

Figure 6: Exposure of restaurants in each city. Note that the red marker in the figure shows the popular attraction in each city



(a) Education (b) White Population

Figure 9: Heatmap for Los Angeles



(a) White Population (b) Education

Figure 7: Heatmap for Anchronage

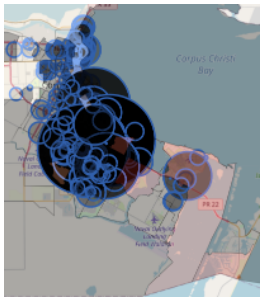


Figure 8: Education heatmap for Corpus Christi

Table 14: Results of the regression analysis of average exposure with hotspot for Los Angeles

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

Table 15: Results of the regression analysis of hotspot with other sensitive attributes for Los Angeles

	Dependent variable:
	Hotspot
	Logistic
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

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 educated neighborhoods. Interestingly none of the demographic features were significant with hotspot. Figure 9 shows the heatmap for Los Angeles. We can see visually that major hotspots are in highly educated (shown in Figure 9a) and restaurants with higher exposure are more in less white neighborhoods, hence confirming the validity of our regression analysis (shown in Figure 9b). 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 [102]. 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 *H3*, and businesses that are in the zip codes that are in a hotspot, are in highly educated neighborhoods, hence we can reject *H4*.

New Orleans: Figure 6e 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 16 shows the results of our linear regression model. We can see similar results, that were obtained in the previous cities,

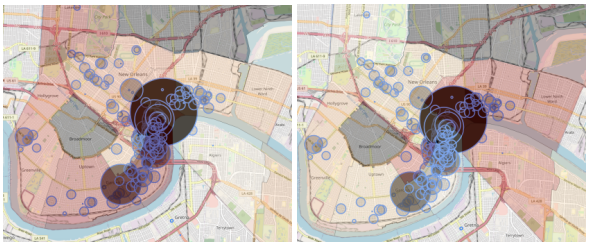
Table 16: Results of the regression analysis of average exposure with hotspot for New Orleans

	Dependent variable:
	average exposure
	Linear
Hotspot	0.039 (0.000)***
Observations	281
Adjusted R-squared	0.04479
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 Orleans

	Dependent variable:
	Hotspot
	Logistic
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

where average exposure is positively correlated to the 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 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 [86]. Figure 10



(a) Education (b) Unemployment

Figure 10: Heatmap for New Orleans

shows the heatmap for New Orleans. We can see visually that major hotspots are in highly educated (shown in Figure 10a) and highly unemployed neighborhoods, (shown in Figure 10b). 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 suppory for *H3*, and businesses in zip codes where there is a hotspot, are in highly educated and highly unemployed neighborhoods and diverse neighbourhoods, hence we find a partial support for *H4*.

New York City: Figure 6f 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 18 shows the results of our linear regression model.

Table 18: Results of the regression analysis of average exposure with hotspot for New York City

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

Table 19: Results of the regression analysis of hotspot with other sensitive attributes for New York

	Dependent variable:
	Hotspot
	Logistic
WN	-2.077 (0.018)*
BN	-2.046 (0.027)*
AIN	0.597 (0.359)
AN	1.218 (0.031)*
HED	N/A
HUne	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 19 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. Upon further examination we found that the percentage of people with bachelor's degrees for all the zip codes was higher than that of the whole city average i.e., 39.6% based on Census [20]. Figure 11 shows the heatmap for New York City.

(a) Education **(b) Wealth**

Figure 11: Heatmap for New York City

We can visually see that major hotspots are all scattered in highly educated neighborhoods(shown in Figure. 11a), hence confirming why we did not obtain any result from our regression model and highly wealthy neighborhoods, (shown in Figure 11b). 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 *H3*, 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 *H4*.

San Francisco: Figure 6g 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 20 shows the results of our linear regression model. We can see similar results, that were

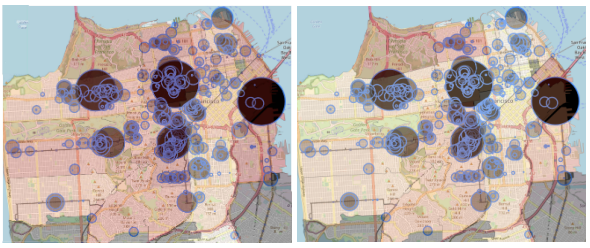
Table 20: Results of the regression analysis of average exposure with hotspot for San Francisco

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

Table 21: Results of the regression analysis of hotspot with other sensitive attributes for San Francisco

	Dependent variable:
	Hotspot
	Logistic
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)***
HUne	0.826 (0.368)
HWe	1.491 (0.049)*
Note:	*p 0.05; **p 0.01; ***p 0.001

obtained in the previous cities, where average exposure is positively correlated to the hotspots of the city. Table 21 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 12: Heatmap for San Francisco

Figure 12 shows the heatmap for San Francisco. We can visually see that major hotspots are all scattered in highly educated (shown in Figure 12a) and highly wealthy neighborhoods, (shown in Figure 12b). 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 *H3*, we were only able to partially find support for *H4*.

San Jose: Figure 6h shows the exposure of restaurants in the city of San Jose. Using the DBSCAN algorithm, we were able to

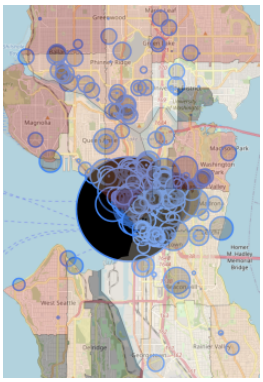


Figure 13: Wealth heatmap for Seattle

Table 22: Results of the regression analysis of average exposure with hotspot for San Jose

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

identify 151 clusters. Using our threshold, we obtained 10 hotspots in the city of San Jose. Table 22 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, hence we cannot reject our *H3*.

Seattle: Figure 6i 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 23 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 24 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 23: Results of the regression analysis of average exposure with hotspot for Seattle

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

Table 24: Results of the regression analysis of hotspot with other sensitive attributes for Seattle

	Dependent variable:
	Hotspot
	Logistic
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 13 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 *H3*, we reject our *H4*.