Food, Service, and Weather too?
Influence of Weather on Yelp Restaurant Reviews

IMT 589D
December 4, 2013

Prepared by:
Tue Bui
Xian Ding
Ari Pollack
Zhen Wang

Abstract

Yelp, founded in 2004, was created to help people find "great local businesses." Potential customers could read the reviews and ratings of previous customers with the goal of deciding whether or not to patronage the business. Traditionally, it has been thought that the reviews a business received were based solely on the merits of the business itself. However there may be other external factors that come into play, that have the potential to bias an individual review. Here we explore the relationship between different weather parameters and the rating scores given to restaurants, in three different cities: Los Angeles, New York City, and Seattle. Variations in only two weather factors seemed to be correlated with different rating scores: daily average minimum temperature (TMIN) and average daily wind speed (AWND): as TMIN and AWND increased, the average rating decreased. However, when we examined Italian, Japanese, and Thai restaurants only, average daily maximum temperature and average daily precipitation changes also seemed to be correlated with varying review ratings. In addition, we explored how weather might influence reviewers differently in the three different cities. It appears that in Seattle, reviewers tend to write more reviews on cold days, while more reviews are written on warmer days in the other cities. However, individual weather factors did not seem to influence ratings differently in the different cities. Despite the long held belief that ratings are based on the inherent qualities of the restaurant, here we show that other factors might be important as well.

Introduction

The starting point for our research project centered on the question “how could we incorporate, analyze, and predict social data through data science”. Keeping this in mind, we thought about questions that penetrated into our social lives and how we could possibly answer these questions through the synthesis of different social datasets.

Yelp is a widely used social media service that provides a variety of online reviews for various types of businesses, restaurants being one of the most commonly reviewed categories. The company has made available a subset of their data for analysis by university researchers. There have been many student teams from all over the world competing against each other on how to take advantages of this data and make it more meaningful. Some of the teams have used advanced natural language processing techniques such as “Latent Dirichlet Allocation” to identify subtopics or hidden factors from user’s reviews and then used them as feature variables to predict the new user scores [1]. Others have utilized “Sentiment analysis” to extract subjective information from reviews to predict the impact on businesses attention based on the number of user reviews [2].

Our work will not focus on the content of user reviews for a particular businesses on Yelp, but rather concentrate on the relationships between weather and users’ ratings. Theoretically, the ratings on restaurants should focus on internal factors such as the quality of food, services and environment. However, it is also possible that ratings may vary due to external factors that are not directly related to the quality of the restaurant. Weather, one such external factor, may influence individual reviews. In order to investigate this we have proposed the following research question: Does weather influence Yelp restaurant reviews?

To answer the above research question, we designed four sub-questions:

1. What weather factors (temperature, wind, weather type, etc.) influence restaurant ratings?
2. Does weather impact the frequency of reviews?
3. Does weather impact rating frequency differently in Seattle, Washington, New York City, New York or Los Angeles, California?
4. Does weather impact ratings differently in Italian, Japanese, or Thai restaurants?

Methods

Data Sources

To answer our research questions, we collected data from the Yelp academic dataset[1] as well as the of National Oceanic and Atmospheric Administration’s (NOAA) National Climatic Data

[1] Yelp Academic Data Set (https://www.yelp.com/academic_dataset)

Center[2]. The Yelp dataset, contains information for businesses near 30 universities across the United States, was available as JSON formatted code. The weather data for Seattle, Los Angeles, and New York was obtained as CSV files. In order to proceed with our analysis we needed to extract and merge these two data sources into one single file, formatted such that each individual observation, a restaurant review, included all necessary attributes (for a description of the attributes and their units see appendix 2).

To accomplish this we utilized python scripts to parse and filter the JSON objects from Yelp and then exported it as two CSV files. All of the files (from Yelp and NOAA) were imported into Microsoft Access in order to join the individual reviews, with the corresponding restaurant and weather data. The integration process is shown in Appendix 1.

Statistical analysis

A linear regression model using all weather variables was developed to answer the first question. In addition we utilized the "step" function in R to automatically simplify the linear model using ANOVA and eventually chose the variables that had the highest correlation with the dependent variable.

To answer the second question related to the frequency of reviews, we utilized chi-square calculations to determine if the reviews fit into our expected frequency distribution. To investigate the impact of temperature we first defined hot and cold days as one standard deviation above or below the mean daily maximum temperature respectively. Assuming a normal distribution of temperatures we would expect the reviews to be equally distributed as well and this was the basis for the chi square analysis.

For the third question, we used Pearson's Chi-Square test to investigate the impact of weather on the frequency of ratings in the different cities. If temperature had a similar influence in the three cities we would expect the frequency of reviews on hot and cold days to be the same. The Chi-Square test compares the actual and expected frequencies to determine if there is a statistical difference between the two. To explore the impact of rain on the number of reviews we used univariate Chi-Square test for each city to examine this relationship since the number of rainy days in each city was different. To correct for this difference we standardized the number of reviews based on the observed number of rainy and dry days (i.e. # of rainy day observations/# of rainy days) and then compared this ratio between cities.

Finally, we utilized t-test and a univariate linear regression model to answer the last question which explores the impact of weather in different restaurant categories: Italian, Japanese, and Thai. These three categories were chosen as they had the highest frequency of reviews in the entire dataset.

Results

[2] National Oceanic and Atmospheric Administration Data Set (http://www.ncdc.noaa.gov/cdo-web/datasets)

Our data set that we used to analyze includes 7 years of observations between November 1, 2005 through October 15, 2012. Over this period of time there were 37,738 reviews of 652 restuarants completed in the three cities.

Research Question 1:

Descriptions: As shown in Table 1, at 5% significance level, average daily wind speed (AWND) and minimum temperature (TMIN) have statistically significantly linear relationship with Review Rating Scores. The coefficient for variable AWND is -0.00251, meaning that for one-unit increase in AWND, we expect a 0.00251 decrease in rating score for a review, holding all other independent variables constant. Similarly, for one-unit increase in TMIN, we expect a 0.000415 decrease in rating score for reviews.

Table 1: Summary of the entire dataset with regression coefficients and their corresponding P-values

Variables	Review Rating Score					Multivariate Regression Coefficient	P-Value
	1	2	3	4	5		
Avg TMAX	186.79	185.58	184.10	185.87	183.85	5.10E-05	0.7929
Avg TMIN	111.26	109.27	109.11	109.88	107.36	-4.15E-04	0.0486
Avg Snow	0.61	0.91	0.75	0.66	0.45	-1.06E-03	0.1466
Avg Precipitation	18.16	21.79	20.52	21.14	19.10	6.78E-05	0.4612
Avg Wind Speed	35.73	36.30	35.92	35.75	34.63	-2.51E-03	1.70E-08
No. of Observations in Rainy Days	952	1555	2788	4480	2687	1.58E-02	0.2868
No. of Observations in Snowy Days	80	135	243	364	202	-3.74E-02	0.388

Findings: Based on the P-values of variables in the table 1, we concluded that MIN temperature and wind speed have statistically significant correlation to the rating scores while for other variables such as MAX temperature, snow, precipitation, rainy, etc. we failed to reject the null hypothesis that their coefficients are non-zero.

Research Question 2:

	# Reviews	Expected Frequencies
Hot day	6212	0.5
Cold day	6726	0.5
Total	12938	1
	Chi-squared test for given probabilities	
	X-squared = 20.4202, df = 1, p-value = 6.217e-06	

Findings*: The univariate Chi-Square test shows that more users reviewed in the **cold** days than **hot** days.*

Research Question 3:

Table 2 describes the data from the perspectives of three cities selected. Restaurants in Los Angeles have the highest average score of 3.53 and highest number of reviews in our data set. As for weather factors, New York City has the most variable weather with

the largest standard deviation (SD) for maximum temperature (Avg TMAX), minimum temperature (Avg TMIN), precipitation and wind speed. Number of rainy days, number of non-rainy days and number of snowy days are also recorded for each of the three cities.

Table 2: summary of the dataset for each city

Cities	No. of Reviews	Avg Score (SD)	Avg TMAX (SD)	Avg TMIN (SD)	No. of Rainy Days	No. of Non-rainy Days	No. of Snowy Days	Precipitation (SD)	Avg Wind Speed (SD)
Overall	37,738	3.51 (1.18)	185.1 (72.51)	109.2 (64.33)	2,437.00	4,026	220	20.41 (72.82)	35.61 (14.86)
Seattle	10,558	3.38 (1.16)	152.3 (71.02)	71.38 (49.68)	1,180	1,000	87	26.98 (62.16)	33.75 (14.22)
Los Angeles	17,745	3.53 (1.18)	208.5 (48.85)	134.8 (36.90)	382	1,877	-	6.89 (36.61)	30.93 (9.79)
New York City	9,435	3.38 (1.16)	178.6 (100.53)	103.4 (91.86)	875	1,149	133	34.18 (102.33)	46.78 (17.67)

Findings:

- The Pearson Chi-Square test shows that the impact of temperature on the number of user reviews is significantly different in different cities.

	Reviews in Hot day	Reviews in Cold day
Seattle	1705	1638
New York	1819	1912
Los Angeles	2688	3176
Total	6212	6726
	Pearson's Chi-squared test X-squared = 23.8897, df = 2, p-value = 6.493e-06	

- 3 Chi Square tests show that there's no significant difference in the number of reviews in rainy days versus days without rain.

	Seattle		Los Angeles		New York	
	#Reviews	#Days	#Reviews	#Days	#Reviews	#Days
Rainy	5831	1180	2511	302	4120	875
Non-Rainy	4727	1000	15234	1877	5315	1149
Chi-Square.p	0.36		0.74		0.74	

Research Question 4:

Description:

To test the correlation on restaurant categories between review stars and the each of the weather factors including maximum daily temperature, minimum daily temperature, average wind speed and rain, below four linear regression models are applied to Italian, Japanese and Thai restaurants respectively.

$$Review\_stars_i = \alpha + \beta TMIN_i + \varepsilon_i \quad (1)$$
$$Review\_stars_i = \alpha + \beta TMAX_i + \varepsilon_i \quad (2)$$
$$Review\_stars_i = \alpha + \beta AWND_i + \varepsilon_i \quad (3)$$
$$Review\_stars_i = \alpha + \beta RAIN_i + \varepsilon_i \quad (4)$$

The results of the regression models is displayed in table 3, indicating that both maximum temperature and minimum temperature have significant negative relationship with the

average star count for Thai restaurants, and that Italian restaurants' star counts have a significant negative relationship with the wind speed. In addition, the results from our t-test (table 4) suggest that Thai restaurants tend to receive a higher number of stars on rainy days compared to non-rainy days. The significant relationships we find between those variables are also visualized in Figure 1.

Table 3 - correlation regression for the rating scores in different restaurant categories

Category	Max Temp (p-value)	Min Temp (p-value)	Wind Speed (p-value)	Rainy (Yes/No) (p-value)
Italian	0.0002 (0.399)	0.0001 (0.863)	-0.0029 (0.026)	-0.414 (0.347)
Japanese	0.0003 (0.296)	-0.0003 (0.378)	-0.0010 (0.54)	-0.0177 (0.698)
Thai	-0.0012 (<0.001)	-0.0021 (<0.001)	0.0013 (0.4)	0.2410 (<0.001)

Table 4 - T-test results to compare rating scores between rainy days and non-rainy days in each restaurant categories

P-value of T-test (rainy day review_stars vs non-rainy day review_stars)	
Italian	0.3443
Japanese	0.6976
Thai	<0.001

Findings: The impact of weather on the rating scores is different in different categories. Thai restaurants tended to have lower reviews as both the TMAX and TMIN increased. In addition reviews tended to be higher on rainy days compared to non rainy days. For italian restaurants there was a significant positive correlation between wind and review scores (i.e. as the wind increased so did the rating).

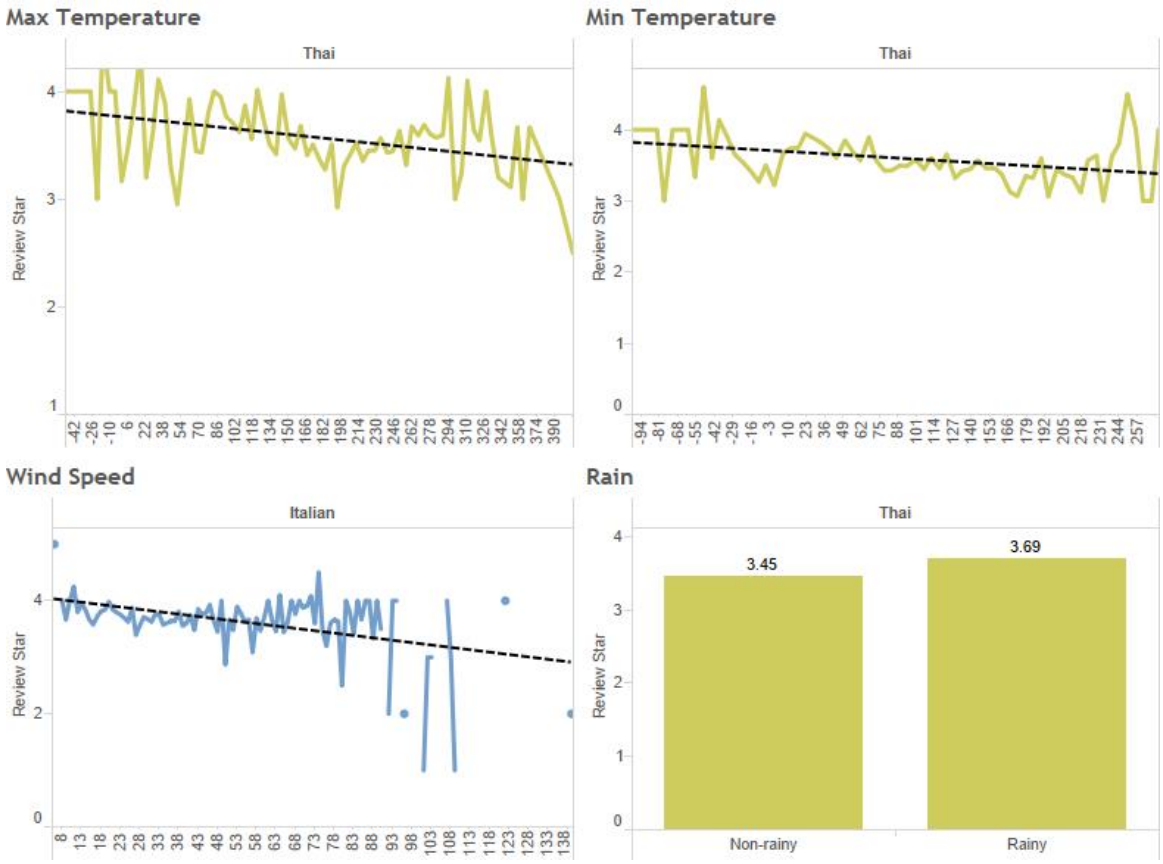


Figure 1: Correlation between rating scores and weather types

Discussion:

Our analysis demonstrates that weather, a non intrinsic factor, may play a role in influencing the restaraunt reviews on Yelp, answering our our primary research question. However, our analysis did not find a relationship between individual Yelp reviews and all of the weather factors we investigated. When considering all observations in all cities and all restaurant categories, we only found an association with two weather variables: daily minimum temperature and average daily wind speed. Despite the significant associations between these two variables and individual review scores, the impact on the individual scores is small and may play a minimal role in changing the individual scores over small temperature ranges. However, over larger ranges it seems to have a more pronounced impact. It is interesting that only one of the temperature variables was found to have a significant association with review scores, suggesting reviewers may be more influenced by the daily minimum temperature and not the maximum temperature.

The number of reviews for a particular restaurant on Yelp is just as important as the average score. The score for businesses with more reviews, is likely to be more accurate as it captures

a broader range of opinions. The opposite can be assumed for a restaurant with a smaller set of reviews. Therefore understanding how weather impacts Yelp reviews must also consider how it impacts not the rating itself but the frequency of reviews as well. Here our results are much more meaningful. Our results clearly show that more reviews occur on cold days compared to hot days. Providing more support for the theory colder temperatures plays a larger role in influencing Yelp reviews. However, this is only true in New York and Los Angeles, as in Seattle more reviews were written on warm days, our first hint that weather may not actually influence reviewers equally in different cities. Considering just Seattle and Los Angeles, one could argue that people prefer to write reviews in temperatures opposite the natural trend. For example Seattle tends to be colder on average than Los Angeles which tends to be warmer on average and therefore reviewers in Seattle are swayed to write reviews in warmer temps while the opposite is true for Los Angeles. However, this argument doesn't seem to apply to New York as its temperature has a wider range including both extremes over the course of a normal year. It would be interesting to investigate this finding in more detail and see if there is a difference review frequencies in different seasons. Perhaps in winter people review more in warmer days and in the summer on colder days. Given the significant relationship between temperature and review frequency, we would have expected to find similar results for other weather factors. However, when we compared the number of reviews recorded on rainy days versus those on dry days, we didn't find a significant difference. Given that we combined all of the data temporally and didn't separate it into simialr and smaller clusters we might have missed any potential relationships.

Our final question looked at how weather might impact different types of restaurants in different ways. We chose three categories to explore this hypothesis: Italian, Japanese, and Thai restaurants. These three categories were chosen since there was a large collection of these restaurants in the data set across all three cities. Similar to our other questions, our results were mixed. Thai restaurants seem to be influenced the most by a variety of weather factors, while no association with weather factors was found with Japanese restaurants. The inverse association with review scores and temperature (i.e as the temperature increases the review score decreases) could potentially be explained by the fact that Thai food is the spiciest of the three categories and that people tend to enjoy it more on colder and rainy days. Though one could argue that a similar finding would be expected for Italian cuisine as well. Many would consider it to be comfort food and comfort food tends to be preferred on cold rainy days, though this is likely highly variable based on one's background. Unfortunately, given the data used for this project it is not possible to get at the explanation behind these observed differences. A well designed and controlled experiment could potentially help investigate the relationships between different restaurant types and weather in more detail.

It is important to understand that our analysis can only identify associations between Yelp reviews and weather attributes. Just as we have hypothesized that weather may impact reviews, there may be other factors as well we are not accounting for that could skew the results. Since we utilized retrospective data and did not control for any of these potential confounding factors, we need to be careful about over interpreting our results. However, with

that being said, given the large number of observations in our data set, it likely limits both type I and type II errors. Though it is possible that both types of error have occurred in our analysis despite the large sample size. Type II errors may have resulted through misclassification and improper grouping. For example, as identified previously, we may have missed statistically significant relationships with specific weather factors, if we had explored smaller segments of our data, perhaps looking at seasons.

In addition, there were other limitations with our analysis. For example, our weather data was generalized for each city as it was obtained from a weather station located at the corresponding aiport. It is likely that the weather observations in our data set varied slightly from the weather observations at the local restaurant. In addition we matched weather data based on the date the review was written, which might have been different from the date that the reviewer actually visited the specific restaurant. Unfortunately we have no way of knowing how often this occurred, and therefore our results might be skewed. By taking a running three day average for each weather variable might help to address this shortcoming.

Our results provide a glimpse into the possible associations between weather and Yelp reviews. Given more time it would be interesting to explore this association in more detail. As mentioned previously, investigating temporal relationships between weather and Yelp reviews, might uncover additional relationships obscured by our current methods. In addition, exploring the impact of weather on additional restaurant categories might also prove interesting. Finally, one could further investigate the local impacts of weather on reviews both in different cities as well as different restaurant categories.

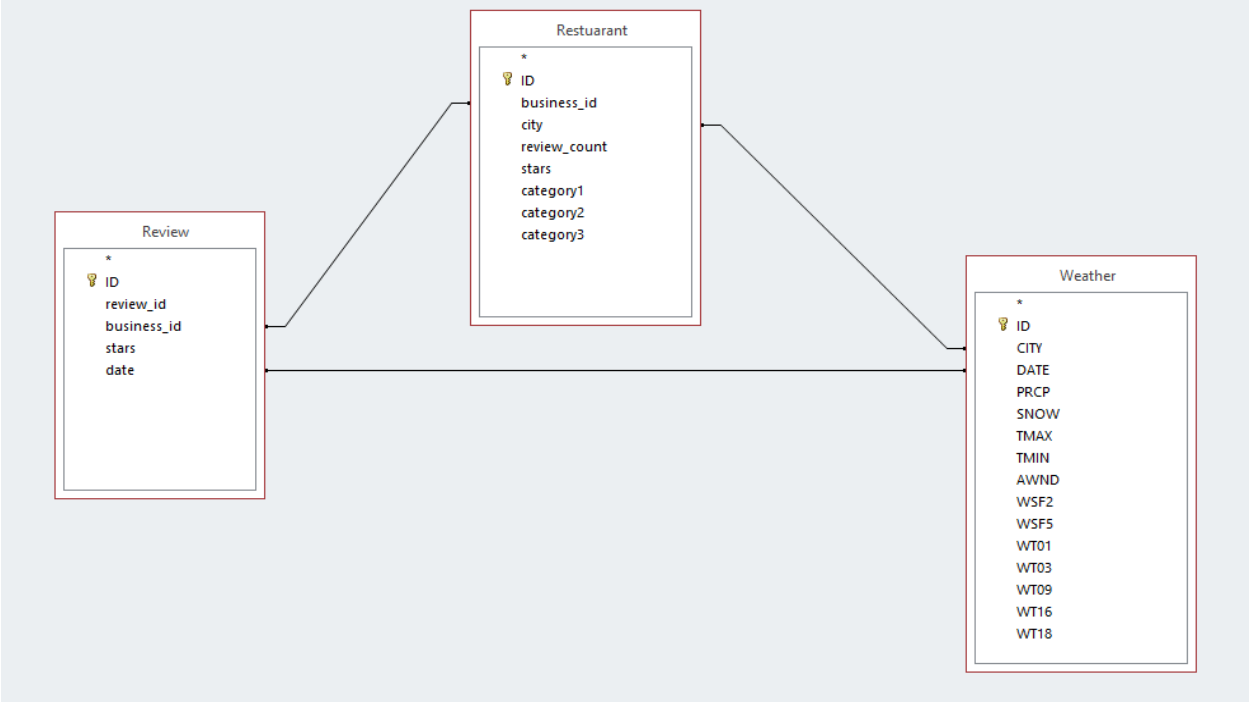
Conclusion

Our work clearly demonstrates an association between specific weather factors and reviews on the social media site Yelp. We were able to find significant relationships between weather and review scores, the frequency of reviews, reviews in different cities and in different restaurant categories. This suggests that ratings on Yelp and perhaps other similar sites might be influenced by extrinsic factors beyond the restaurant's control. Consumers of this information should be made aware of these associations so that they can make more informed decisions.

Reference:

[1] McAuley, Julian, and Jure Leskovec. "Hidden factors and hidden topics: understanding rating dimensions with review text." Extended version http://infolab.stanford.edu/~julian/pdfs/recsys13.pdf (2013).
[2]Hood, Bryan, Victor Hwang, and Jennifer King. "Inferring Future Business Attention."

Appendix 1 (Data set integration in MS Access)



Appendix 2 (Summary of the attributes of each observation)

attribute	description	min	max	mean	median	sd
review_id	unique id of each review	na	na	na	na	na
business_id	unique id of each restaurant	na	na	na	na	na
city	Los Angeles, New York, Seattle	na	na	na	na	na
Restaurant_review count	number of reviews of individual restaurants	2	841	200.5	150	176.67
Restaurant_stars	overall stars of individual restaurants	1	5	3.51	3.5	0.54
Category	category of the restaurants,	na	na	na	na	na

	sushi, BBQ, soup etc.					
Review stars	starts given by each review	1	5	3.51	4	1.18
date	the date of each review	11/1/2005	10/15/2012	na	12/10/2010	na
PRCP	precipitation (tenths of mm) on each review day	0	1699	20.41	0	72.82
SNOW	snowfall (mm)	0	264	0.66	0	9.17
TMAX	maximum temperature (tenths of degrees C)	-78	406	185.1	194	72.51
TMIN	minimum temperature (tenths of degrees C)	-133	300	109.2	117	64.33
AWND	average daily wind speed (tenths of meters per second)	0	139	35.61	33	14.86
WSF2	fastest 2-minute wind speed (tenths of meters per second)	27	416	80.38	76	25.22
WSF5	fastest 5-second wind speed (tenths of meters per second)	0	1520	103.7	98	58.86
WT01	fog, ice fog, or freezing fog	0	1	0.38	0	0.49
WT03	thunder	0	1	0.03	0	0.17
WT09	blowing or drifting snow	0	1	0.02	0	0.15
WT16	rain	0	1	0.33	0	0.47
WT18	snow, snow pellets, snow grains, or ice crystals	0	1	0.03	0	0.16