Are Yelp Ratings Worth Your Trust?

Using Ordered Logit Regression to Model Las Vegas Restaurant Yelp Ratings and Inspection Data in 2017

This paper will investigate in Yelp ratings and several factors that might affect the rating of Food and Beverage (F&B) businesses in Las Vegas. The research question is whether one can trust ratings on Yelp when choosing restaurants, coffee places, or catering services. To answer this question, I first introduce relevant data of 1368 local F&B businesses from Yelp data set and Southern Nevada Health District. Secondly, data visualization will be provided to discover the relationship among variables of interest. I will also present a linear regression and an ordered logit regression to assess the relationship among Yelp ratings and Inspection data. The idea is to verify if Yelp reviews and ratings really reflect the conditions and violations of restaurants. Eventually, model setting and improvement will be discussed.

I Introduction

Yelp - the crowd-sourced reviews platform has been around for more than 10 years. People go on Yelp to share, read and exchange opinions about local businesses especially F&B businesses. As Yelp is becoming one of the biggest resources for casual decision making about picking restaurants and services, the amount of concerns about its reliability is also rising. There are a large number of factors that can go into a rating on Yelp. One rating may be solely based on tangible aspects such as food presentation, interior design, and restaurant decoration. Whereas, others may rate a coffee shop using their own personal emotions and judgements on intangible attributes such as the coffee’s taste, customer services and general atmosphere.

In this analysis, I will focus on a single aspect, which is the cleanliness of an establishment. Cleanliness can be both visible and invisible to consumers. For example, it is easy for a customer to see if the floor and table are clean, but there is no way one can tell what is going on in the kitchen, or are there germs in their food. Therefore, if the Yelp ratings is reliable, they should somehow reveal a part of the cleanliness conditions of a restaurant. In other words, the rating would be more credible if it reflects the cleanliness of a restaurant.

II Data Source

I retrieved data from two different sources in January, 2018. The first one is Yelp data set, which is published by the company themselves. The data is offered for students and researchers for educational purposes. It contains information about local businesses in 11 metropolitan areas in the United States. There are 6 subsets in the original dataset, which is business, reviews, user, tip, photos, and checkin. This research will only use the business data set. The business data set contains 156,639 observations with 101 variables such as Name, Stars (ratings) Business ID, Address, City, Category, etc. The number of variables in this data set are enormous due to 96 variables of hours and attributes. Because I am interested in observations that are in Las Vegas , so I subset the data set, keep only those in Las Vegas. I also drop irrelevant variables. Subsequently, Yelp data set used for this research includes 26,066 observations with 13 variables.

The second data set is from Southern Nevada Health District. The health district conducts unannounced inspections of food establishments such as restaurants, bars, taverns, snack bars, food warehouses, health food stores, markets and permanent outdoor barbeques at least once a year. Inspections are posted online with both cumulative and most recent results. The data set contains 153,100 observations with 15 variables such as Inspection Grade, Inspection Demerits, Violations, Restaurant Name, and Neighborhood.

Finally, I merge two datasets together using Restaurant Name as a key. The final data set that will be presented in this paper includes 1,368 observations with 14 variables.

III Data Summary

The table below describes the final data set after cleaning and merging. There are a couple of notes that is worth making here. Firstly, even though I merged both datasets to one, I still keep geographical coordinates from both datasets since they are different. The difference may come from the difference in the way two institutions gather their resources. Secondly, the way grading system works is that inspection findings may result in an “A” grade, “B” grade or “C” grade status. “A” grade indicates that The establishment has earned between 0-10 demerits on their last inspection, “B” grade - the business has earned between 11 and 20 demerits or identical consecutive critical or major violations.

“C” grade - the establishment has earned between 21 and 40 demerits, has identical consecutive critical or major violations or more than 10 demerits on a “B” grade re-inspection. As inspection is a continuous process, it is helpful to keep both cumulative and most recent Inspection Results and Grades for analysis. In the dataset, there is an “X” grade, which the Health District themselves do not explain clearly its meaning. Thirdly, Stars or Ratings in the dataset is treated as numeric but it is in fact, factor with 9 levels (1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5). However, I keep it as numeric at first, so the regression results would be easier to interpret. Later on, when running Logistic Regression we will consider stars as factor. Lastly, demerits refer to the number of food safety and cleanliness violations, I change the names of variables from “demerits” to “violations” to make it easier for readers to keep track.

Table 1: Data Description

Variable	Description	Example
Restaurant Name	Name of the restaurant (Character)	Hawaiian Barbecue
Neighborhood	Geographical area (Factor with 6 levels)	Centennial, Strip, etc.
Stars	Cumulative ratings of the business as of January, 2017 (Numeric/Factor)	1.5, 2.0, 5.0, etc.
Review Count	Number of reviews the business received (Integer)	15, 1994, 195, etc.
Category Name	Name of business’s category (Factor with 30 levels)	Bakery sales, Restaurants, Buffet, etc.

Current Violations	Cumulative number of cleanliness and food safety violations (Integer)	3, 0, 10, etc.
Current Grade	Cumulative cleanliness and food safety grade (Factors with 4 levels)	a, b ,c ,x
Inspection Violations	Most recent number of cleanliness and food safety violations	3, 0, 10, etc.
Inspection Grade	Most recent cleanliness and food safety grade (Factors with 4 levels)	a, b ,c ,x
Address	Address of the business (Character)	5115 Spring Mountain Rd 103
Zip	Zip code (Integer)	80109,
Longitude	Numeric	-115
Latitude	Numeric	36.1
Location.1	Character	(34.470918, 117.204886)

Table 2: Descriptive Analysis - Numeric variables

```
Descriptive statistics
==========================================================
Statistic                N    Mean  St. Dev.  Min    Max
----------------------------------------------------------
stars                  1,368  3.8     0.7     1.0    5.0
review_count           1,368 133.6   249.0     3    3,838
Current.Violations     1,368  5.7     4.8      0      51
Inspection.Violations  1,368  8.2     7.5      0      58
longitude              1,368 -115.2   0.1    -115.6 -115.0
latitude               1,368  36.1    0.1     36.0   36.4
----------------------------------------------------------
```

Table 2 above shows a simple descriptive analysis of the data. One thing that is worth noticing here is the difference between Current Violations (Cumulative violations) and Inspection Violations. The mean of the recent Inspection Violations is 8.2, which is higher than the mean of Curren Violations (5.7). The recent violation also seems to fluctuate a lot more than the cumulative violations.

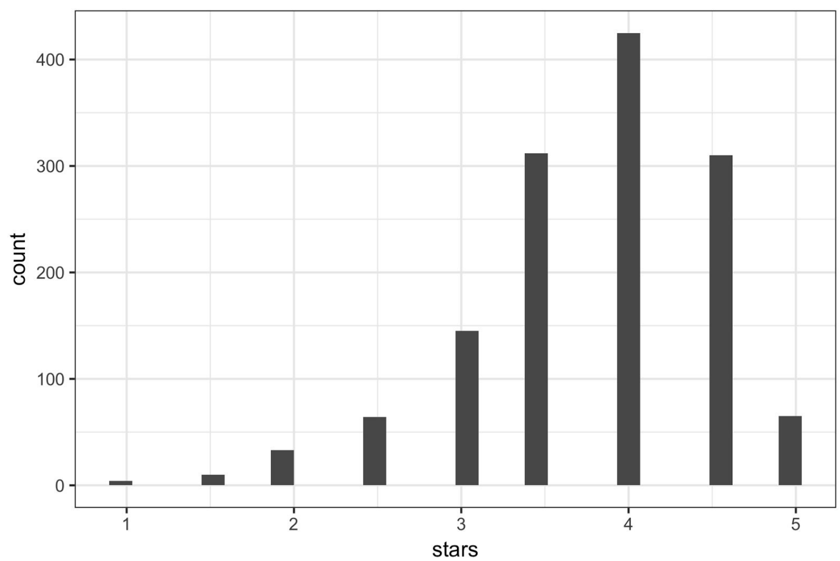
Table 3: Descriptive Analysis - Ratings and Categorical variables

Variable	Value	Count	Percentage
Inspection.Grade	a	1,149	83.99
Inspection.Grade	b	146	10.67
Inspection.Grade	c	59	4.31
Inspection.Grade	x	14	1.02
Current.Grade	a	1,330	97.22
Current.Grade	b	18	1.32
Current.Grade	c	15	1.10
Current.Grade	x	5	0.37

Based on Table 3, Grade A accounts for most of the Inspection Grade and Current Grade. Interestingly, it seems like the cumulative inspection grade is relatively higher than the recent grade. In other words, there are more B, C and X within recent inspection grade.

IV Data Visualization

Figure 1: Stars (Ratings) Distribution



We usually think that Yelp is a space for complainers. However, looking at the simple distribution graph above, it is obviously that the data highly skew to the left. There are actually more 3.5 - 5 stars than lower stars such as 1 or 1.5 stars.

Figure 2: Average Yelp Review Star Ratings by Category

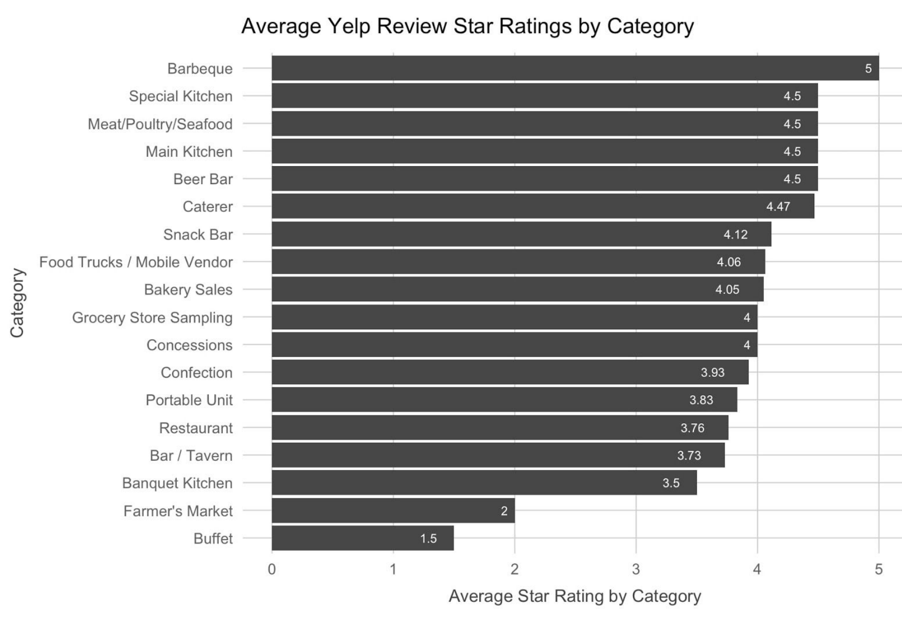
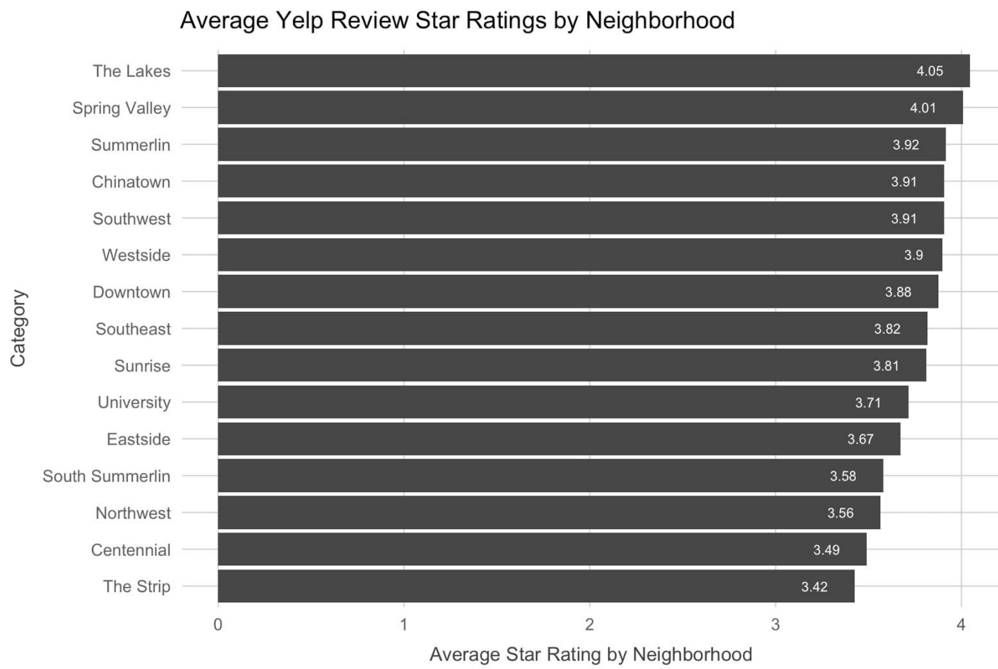


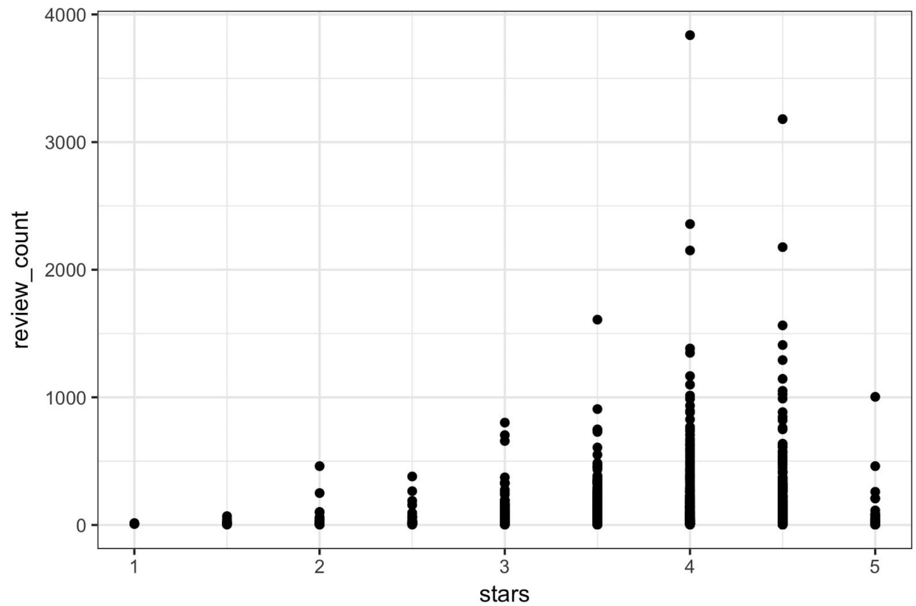
Figure 2 indicates that the average stars are different among different category groups. For example, Buffet restaurant tends to get lower ratings while Special Kitchen, Seafood, Barbeque are likely to get higher stars. We might expect there are some significant coefficients if we enter the category variable into regression as dummy variables.

Figure 3: Average Yelp Review Star Ratings by Neighborhood



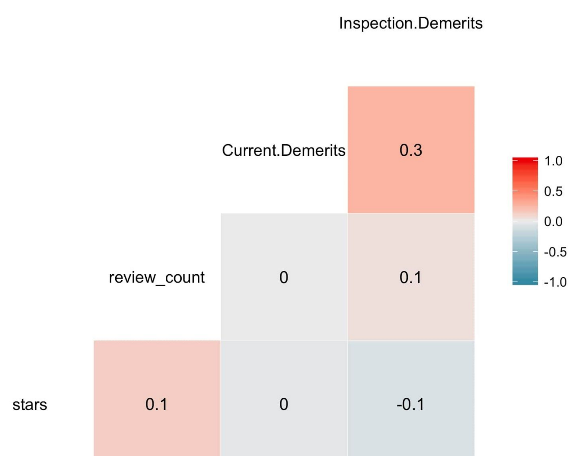
Restaurants at The Strip - the famous area for tourists turn out to get the worst rating among all neighborhoods. However, there are 237 neighborhood missing values (out of 1,368 sample), we cannot tell if area where the restaurant operates has an effect on its rating.

Figure 4: Stars and review count plot



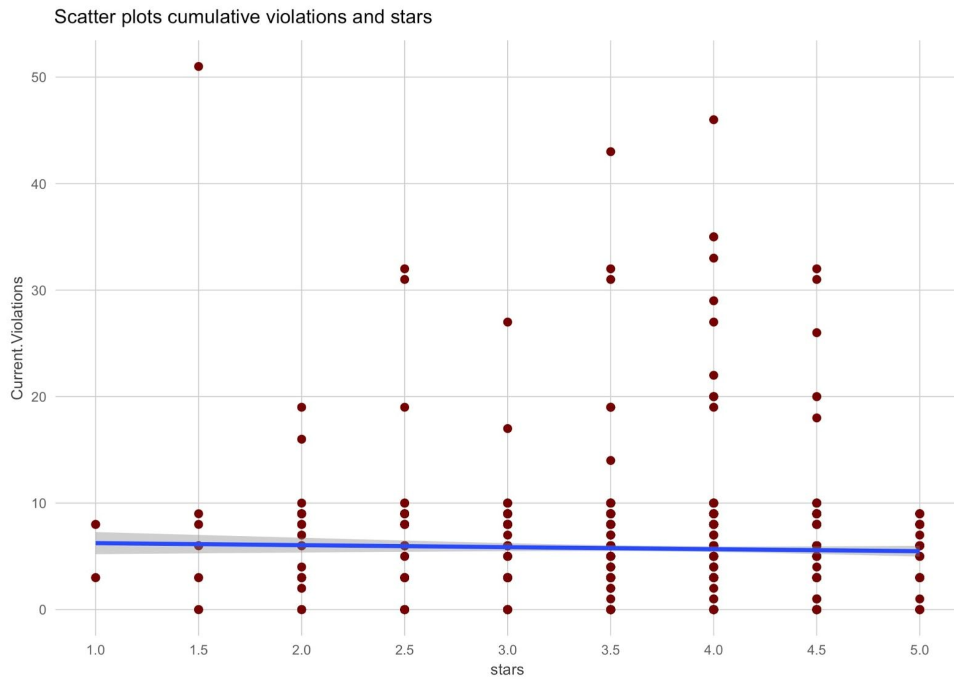
It is hard to affirm that the more reviews the business gets, the more stars it receives. Based on Figure 4, there seems like an upward trend but we definitely need more statistically proof to know if number of reviews has an impact on ratings.

Figure 5: Correlation plot



According to Figure 5, there is negative correlation between most recent or Inspection Demerits and Stars. Whereas, there is a positive relationship between stars and review count.

Figure 6: Scatter plots cumulative violations and stars



Both correlation plot and scatter plot imply that there is no relationship between cumulative violations and inspection violations. On the other hand, Figure 5, the scatter plot between most recent violations and stars show a downward trend. The trend line is clearly more significant than the line in Figure 4.

Figure 7: Scatter plots most recent violations and stars

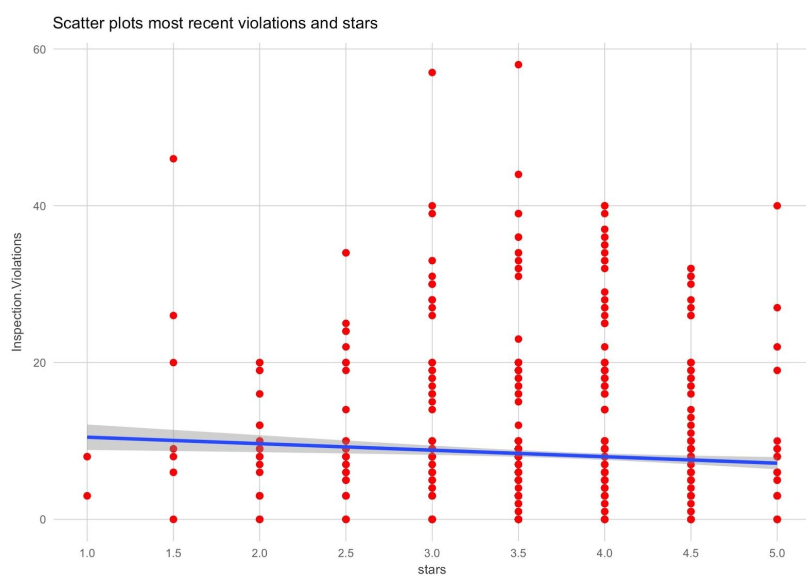
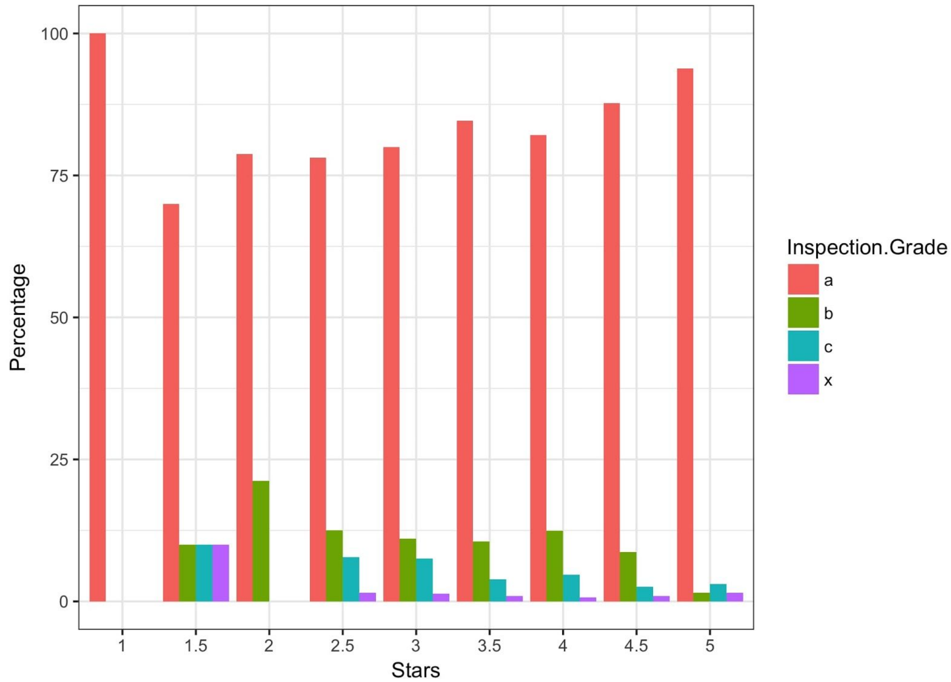
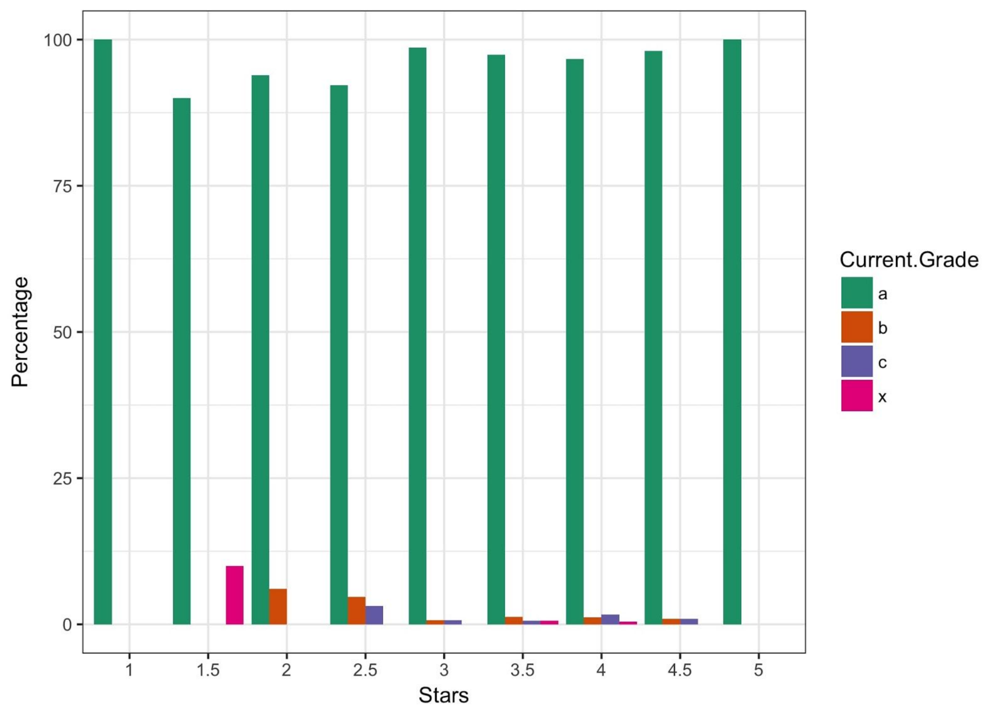


Figure 8: Review Percentage and Stars within Inspection Grade



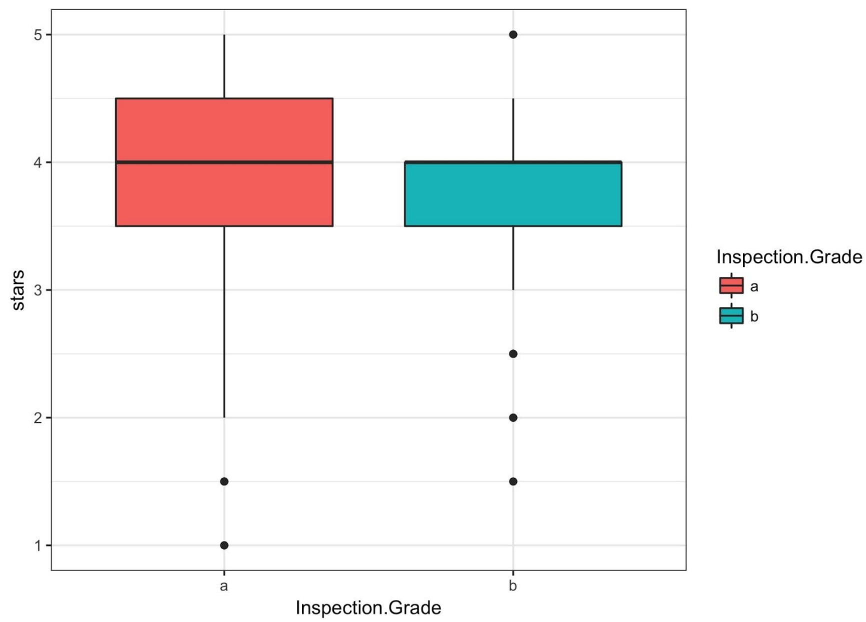
Here it is clear that Grade A accounts for the majority of the data. We also see start from 4 stars to 5 stars, there are less B and C grade than other ratings, which gives us hope that Yelp reviews really reflect the Inspection results. After 4 stars, the number of grade A restaurant also increases. Interestingly, 100% restaurant with 1 stars get grade A. It is because there are a lot more factors going into a Yelp review than just food safety and cleanliness.

Figure 9: Review Percentage and Stars within Cumulative Grade



Once again, grade A accounts for the majority of data among all stars. However, the numbers of cumulative grade B and C are significantly lower than most recent grade B and C. One reason might be if restaurants do not improve their cleanliness and keep getting downgrade, they eventually go out of business. The Southern Nevada Health District states on their official website that for an extremely excessive number of violations or an uncontrolled imminent health hazard, a closure of the facility may result. Here we can also see that there is no 5-star-rating restaurant get Grade B or C. The number of restaurants with Grade B and C gradually decrease towards higher stars.

Figure 10: Review Percentage and Stars within Cumulative Grade



Interestingly, when grouping all the “b”, “c”, and “c” Inspection Grade together as “b”, I find out that restaurant with Inspection grade “a” get higher rating stars comparing to the rest of the Inspection Grade. Overall, most restaurants with grade lower than “a” get lower rating even though the median of both groups is 4. There’s only several outliers within Group “b” get higher than 4 stars - rating.

V Model Building and Implementation

1. OLS

Stars = B0 + B1 x review_count + B2 x Current.Violations + B3 x Inspection.Violations + B4 x Inspection.Grade (dummy) + B5 X Current.Grade + B6 x Neighborhood (dummy) + B7 x Category.Name (dummy)

Figure 11: Regression result (all variables included model)

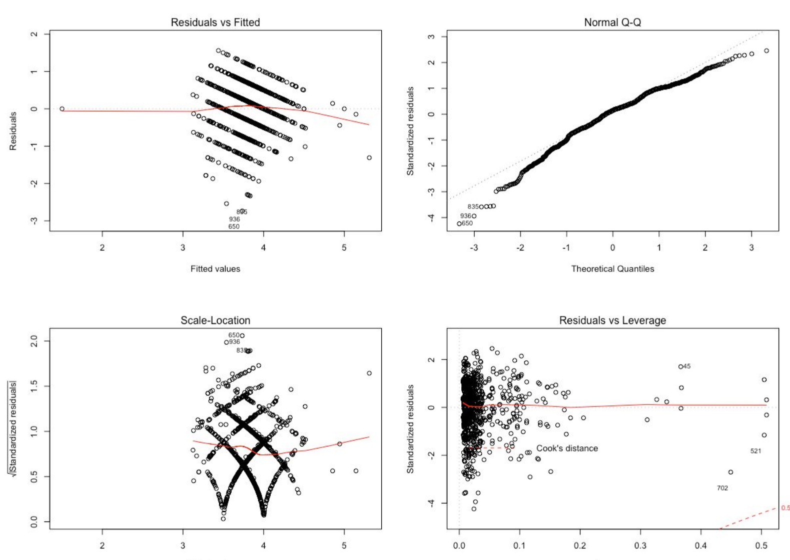
	Dependent variable:
	stars
Inspection.Violations	-0.014**
	(0.006)
Current.Violations	0.007
	(0.007)
Inspection.Gradea	-0.152
	(0.252)
Inspection.Gradeb	-0.106
	(0.221)
Inspection.Gradec	0.002
	(0.217)
Current.Gradea	0.189
	(0.378)
Current.Gradeb	-0.186
	(0.392)
Current.Gradec	0.033
	(0.397)
Category.NameBanquet Kitchen	-0.549
	(0.494)
Category.NameBar / Tavern	-0.203
	(0.188)
Category.NameBarbeque	1.031
	(0.681)
Category.NameBuffet	-2.445***
	(0.779)
Category.NameCaterer	0.328
	(0.251)
Category.NameConcessions	1.117**
	(0.495)
Category.NameConfection	0.118
	(0.317)
Category.NameFood Trucks / Mobile Vendor	0.184
	(0.255)
Category.NameGrocery Store Sampling	-0.014
	(0.679)
Category.NameMain Kitchen	0.598
	(0.678)
Category.NameMeat/Poultry/Seafood	0.626
	(0.681)
Category.NamePortable Unit	-0.106
	(0.288)
Category.NameRestaurant	-0.202
	(0.172)
Category.NameSnack Bar	0.214
	(0.188)
Category.NameSpecial Kitchen	0.014
	(0.428)
neighborhoodCentennial	0.024
	(0.179)
neighborhoodChinatown	0.383**
	(0.166)
neighborhoodDowntown	0.377**
	(0.164)
neighborhoodEastside	0.227
	(0.164)
neighborhoodNorthwest	0.086
	(0.185)
neighborhoodSoutheast	0.304*
	(0.159)
neighborhoodSouthwest	0.364**
	(0.169)
neighborhoodSpring Valley	0.504***
	(0.161)
neighborhoodSummerlin	0.458**
	(0.193)
neighborhoodSunrise	0.316*
	(0.181)
neighborhoodThe Lakes	0.513**
	(0.252)
neighborhoodThe Strip	-0.107
	(0.169)
neighborhoodUniversity	0.283
	(0.209)
neighborhoodWestside	0.396**
	(0.161)
review_count	0.0004***
	(0.0001)
Constant	3.635***
	(0.430)
Observations	1,131
R^2	0.139
Adjusted R^2	0.109
Residual Std. Error	0.653 (df = 1092)
F Statistic	4.650*** (df = 38; 1092)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 11 shows result of the final model. I ran 4 different models. Firstly, I start with simple model with 2 explanatory variables (Violations variables) then keep adding up until getting this final model. The other model results will be attached in the Appendix.

Based on the regression result, there are some significant coefficients such as Inspection Violations, Review Counts, some of the Category Name, and some of the neighborhood. Specifically, keeping everything else the same, when the recent violation goes up by 1 unit, the rating star decrease by 0.014 unit. Similarly, when the review count increases by 1 unit, the rating star goes down by 0.014 unit. Regarding category, Buffet restaurant tends to get 2.445 stars lower than the base, and it is statistically proven. Food and Beverages business locating in Chinatown, Downtown, SouthEast, SouthWest, Spring Valley, Summerlin, Sunrise, and The Lakes have higher stars rating than the base. Surprisingly, according to this model, Inspection Grade and Current Grade have no influence on rating stars.

R-squared of the model is 0.139, which means that the model explain only 1.39% the variability of the response data around its mean. Nonetheless, I totally expect that the R-squared value will be low as rating reviews, eventually, are human opinions. Any field that attempts to predict human behavior, such as psychology, typically has R-squared values lower than 50%. Humans are simply harder to predict than other subjects.

Figure 12: Residuals Plot and QQ plot



Here are Residual Plots and QQ Plot. The residual plot looks problematic as residuals do not distribute randomly around the line. QQ plot generally look like a straight line. To make sure that the model is BLUE (Best linear unbiased estimate), I also check some assumptions such as Homoscedasticity, Link Function, etc.

2. Logistics Regression (Ordered Logit)

There is intrinsic ordering in stars variable, so a more appropriate model would be ordered logit model. As the outcomes (stars) are not continuous, the outcomes will only fall in to the category of 9 different rating from 1 - 5 stars. I run Ordered Logit Regression with an assumption that the error term has logistic distribution.

Figure 13: Significant Coefficients (Ordered Logit model)

	Value	Std..Error	t.value	pval
neighborhoodCentennial	-1.045863875	2.984415e-01	-3.504419e+00	0.00
neighborhoodNorthwest	-0.836115411	3.513113e-01	-2.379984e+00	0.02
neighborhoodSouth Summerlin	-1.187895356	4.514111e-01	-2.631516e+00	0.01
neighborhoodThe Strip	-1.335608714	2.563766e-01	-5.209558e+00	0.00
review_count	0.001090398	2.446992e-04	4.456074e+00	0.00
Category.NameBarbeque	16.778250678	2.648550e-07	6.334883e+07	0.00
Category.NameBuffet	-5.945963526	2.969485e-01	-2.002355e+01	0.00
Category.NameCaterer	1.422950942	6.482650e-01	2.195014e+00	0.03
Category.NameConcessions	16.850535560	7.845771e-07	2.147722e+07	0.00
Category.NameMain Kitchen	1.859511846	7.866131e-01	2.363947e+00	0.02
Category.NameMeat/Poultry/Seafood	1.962084087	7.582810e-01	2.587542e+00	0.01
Inspection.Violations	-0.042184090	1.702151e-02	-2.478281e+00	0.01
1|1.5	-7.521731021	7.650661e-01	-9.831478e+00	0.00
1.5|2	-5.773532743	4.599428e-01	-1.255272e+01	0.00
2|2.5	-4.507776561	3.824871e-01	-1.178543e+01	0.00
2.5|3	-3.356643800	3.575807e-01	-9.387094e+00	0.00
3|3.5	-2.385638186	3.497233e-01	-6.821503e+00	0.00
3.5|4	-1.130924887	3.453873e-01	-3.274368e+00	0.00
4.5|5	2.540070195	3.593819e-01	7.067886e+00	0.00

The figure above sums up the significant result of Ordered Logit Regression. The values here are basically marginal effects. It shows the change in probability when explanatory variables increases by one unit. I will also use relative risk ratios instead of coefficients because they allow an easier interpretation of the logit coefficients. They are the exponentiated value of the logit coefficients. Keeping all other variables constant, when the restaurant is located in Centennial, Northwest, South Summerlin, and The Strip, it is more likely to be in a lower category (getting lower rating stars). Additionally, restaurants with large number of reviews tend to get higher rating stars. Regarding category, restaurants within Barbeque, Caterer, Main Kitchen, Concessions, and Meat and Seafood category increase the chance of getting higher stars. Interestingly, keeping all other variables constant, when Inspection violations increases one unit, it is 0.96 times more likely to be in a higher category. The coefficient is significant. The Residual Deviance is 3640.941 and AIC is 3732.941. These numbers will help us decide which model would be appropriate. However, in this case, I only run one ordered logistic model.

VI Discussion

The better model is the ordered logit model as it treats “stars” the way it is supposed to be (as a categorical variable with orders). Even though, the final model supports the hypothesis that rating stars reflect part of the food safety and cleanliness conditions of Food and Beverages business, there are a lot of shortcomings in the model itself. Firstly, the sample size (1,368) is relatively small comparing to the whole Las Vegas restaurant population (about 26,000). Secondly, there are obscurities in in terms of the way that Health District calculate Current Grade and Inspection Grade (we have no idea what Grade X is). Thirdly, there might be multicollinearity among Inspection Violations, Inspection Grade, Current Grade and Inspection Grade as the number of violations is a part of final Inspection and Current Grades. Fourthly, by including a lot of dummy variables in the model, I run up against the degree of freedom. Fifthly, while the distribution among neighborhood is quite spread out, the distribution among category is highly skewed. Some categories such as Restaurants and Bar have a decent number of observations whereas Food truck and Farmer Markets have one and two observation. Moreover, there are a lot of NAs in Category Name (about 240), which makes it hard to claim anything about this variable. Finally, the credibility of data from Yelp challenge data set has yet been discovered. They claim that data collected at a certain point of the year but we do not know the exact date, which can be a huge red flag when merging with real time data such as Inspection data.

Beside all the downsides, there are some areas to explore in the future research. For example, a sentiment analysis on the reviews may reveal if customers talk about food safety and cleanliness concerns in their feedback. As mentioned before, there are a lot of other factors go into a Yelp review, thus; more variables could be included in the regression. To mitigate the troubles arising from Yelp challenge data set, an extra web scraping for real time data step might help.

VII Conclusion

All in all, there is a reverse relationship among the number of recent Demerits(violations) and rating stars, indicating partly of the consumers' reviews can be explained by food safety and cleanliness. However, the result driven from such a small sample size is not sufficient to say that whether Yelp reviews and stars are trustworthy or not. There are several things that the visualization process revealed such as the differences stars among categories and among neighborhoods.

APPENDIX

Table 1a: OLS Regression

	Dependent variable:			
	stars			
	(1)	(2)	(3)	(4)
Inspection.Violations	-0.007***	-0.007	-0.007	-0.014**
	(0.003)	(0.006)	(0.006)	(0.006)
Current.Violations	-0.001	0.003	0.008	0.007
	(0.004)	(0.006)	(0.006)	(0.007)
Inspection.Gradea		-0.056	-0.041	-0.152
		(0.251)	(0.245)	(0.252)
Inspection.Gradeb		-0.086	-0.041	-0.106
		(0.217)	(0.212)	(0.221)
Inspection.Gradec		-0.032	0.021	0.002
		(0.214)	(0.208)	(0.217)
Current.Gradea		0.523	0.134	0.189
		(0.373)	(0.370)	(0.378)
Current.Gradeb		0.180	-0.277	-0.186
		(0.381)	(0.380)	(0.392)
Current.Gradec		0.477	-0.077	0.033
		(0.374)	(0.375)	(0.397)
Category.NameBakery Sales			-0.458	
			(0.704)	
Category.NameBanquet Kitchen			-0.999	-0.549
			(0.842)	(0.494)
Category.NameBar / Tavern			-0.760	-0.203
			(0.691)	(0.188)
Category.NameBarbeque			0.501	1.031
			(0.972)	(0.681)
Category.NameBuffet			-3.042***	-2.445***
			(0.803)	(0.779)
Category.NameCaterer			-0.034	0.328
			(0.709)	(0.251)
Category.NameConcessions			-0.500	1.117**
			(0.794)	(0.495)
Category.NameConfection			-0.576	0.118
			(0.735)	(0.317)
Category.NameFarmer's Market			-2.498**	
			(0.973)	
Category.NameFood Trucks / Mobile Vendor			-0.408	0.184
			(0.709)	(0.255)
Category.NameGrocery Store Sampling			-0.500	-0.014
			(0.972)	(0.679)

Category.NameMain Kitchen	0.002	0.598
	(0.972)	(0.678)
Category.NameMeat/Poultry/Seafood	0.144	0.626
	(0.974)	(0.681)
Category.NamePortable Unit	-0.655	-0.106
	(0.710)	(0.288)
Category.NameRestaurant	-0.712	-0.202
	(0.688)	(0.172)
Category.NameSnack Bar	-0.363	0.214
	(0.692)	(0.188)
Category.NameSpecial Kitchen	0.021	0.014
	(0.794)	(0.428)
neighborhoodCentennial		0.024
		(0.179)
neighborhoodChinatown		0.383^{**}
		(0.166)
neighborhoodDowntown		0.377^{**}
		(0.164)
neighborhoodEastside		0.227
		(0.164)
neighborhoodNorthwest		0.086
		(0.185)
neighborhoodSoutheast		0.304^{*}
		(0.159)
neighborhoodSouthwest		0.364^{**}
		(0.169)
neighborhoodSpring Valley		0.504^{***}
		(0.161)
neighborhoodSummerlin		0.458^{**}
		(0.193)
neighborhoodSunrise		0.316^{*}
		(0.181)
neighborhoodThe Lakes		0.513^{**}
		(0.252)
neighborhoodThe Strip		-0.107
		(0.169)
neighborhoodUniversity		0.283
		(0.209)
neighborhoodWestside		0.396^{**}
		(0.161)
review_count		0.0004^{***}
		(0.0001)

Constant	3.861***	3.378***	4.405***	3.635***
	(0.033)	(0.370)	(0.780)	(0.430)
Observations	1,368	1,368	1,368	1,131
R^2	0.006	0.011	0.071	0.139
Adjusted R^2	0.005	0.005	0.054	0.109
Residual Std. Error	0.705 (df = 1365)	0.705 (df = 1359)	0.687 (df = 1342)	0.653 (df = 1092)
F Statistic	4.225** (df = 2; 1365)	1.826* (df = 8; 1359)	4.132*** (df = 25; 1342)	4.650*** (df = 38; 1092)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 1b: OLS Assumptions

	Value	p-value	Decision
Global Stat	137.662	0.000e+00	Assumptions NOT satisfied!
Skewness	91.913	0.000e+00	Assumptions NOT satisfied!
Kurtosis	30.493	3.351e-08	Assumptions NOT satisfied!
Link Function	13.261	2.710e-04	Assumptions NOT satisfied!
Heteroscedasticity	1.995	1.578e-01	Assumptions acceptable.

Table 2a: Ordered Logit Model

	Dependent variable:
	stars
neighborhoodCentennial	-1.046***
	(0.298)
neighborhoodChinatown	-0.085
	(0.244)
neighborhoodDowntown	0.052
	(0.236)
neighborhoodEastside	-0.422*
	(0.234)
neighborhoodNorthwest	-0.836**
	(0.351)
neighborhoodSouth Summerlin	-1.188***
	(0.451)
neighborhoodSoutheast	-0.219
	(0.206)
neighborhoodSouthwest	-0.122
	(0.258)
neighborhoodSpring Valley	0.337
	(0.214)
neighborhoodSummerlin	0.027
	(0.349)
neighborhoodSunrise	-0.039
	(0.330)
neighborhoodThe Lakes	0.563
	(0.561)
neighborhoodThe Strip	-1.336***
	(0.256)
neighborhoodUniversity	-0.405
	(0.417)
review_count	0.001***
	(0.0002)
Category.NameBanquet Kitchen	-1.661
	(1.193)
Category.NameBar / Tavern	-0.679*
	(0.371)
Category.NameBarbeque	16.778***
	(0.00000)
Category.NameBuffet	-5.946***
	(0.297)
Category.NameCaterer	1.423**

	(0.648)
Category.NameConcessions	16.851***
	(0.00000)
Category.NameConfection	0.273
	(0.736)
Category.NameFood Trucks / Mobile Vendor	0.513
	(0.648)
Category.NameGrocery Store Sampling	-0.187
	(1.413)
Category.NameMain Kitchen	1.860**
	(0.787)
Category.NameMeat/Poultry/Seafood	1.962***
	(0.758)
Category.NamePortable Unit	-0.227
	(0.701)
Category.NameRestaurant	-0.570*
	(0.315)
Category.NameSnack Bar	0.662*
	(0.383)
Category.NameSpecial Kitchen	0.941
	(1.302)
Current.Violations	0.019
	(0.017)
Current.Gradeb	-0.938*
	(0.536)
Current.Gradec	-0.433
	(0.701)
Current.Gradex	-0.869
	(0.848)
Inspection.Violations	-0.042**
	(0.017)
Inspection.Gradeb	0.231
	(0.280)
Inspection.Gradec	0.540
	(0.526)
Inspection.Gradex	0.781
	(0.714)
Observations	1,131
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 2b: Logit Coefficients

	Dependent variable:
	stars
neighborhoodCentennial	0.351***
	(0.298)
neighborhoodChinatown	0.919
	(0.244)
neighborhoodDowntown	1.054
	(0.236)
neighborhoodEastside	0.656*
	(0.234)
neighborhoodNorthwest	0.433**
	(0.351)
neighborhoodSouth Summerlin	0.305***
	(0.451)
neighborhoodSoutheast	0.803
	(0.206)
neighborhoodSouthwest	0.885
	(0.258)
neighborhoodSpring Valley	1.401
	(0.214)
neighborhoodSummerlin	1.028
	(0.349)
neighborhoodSunrise	0.961
	(0.330)
neighborhoodThe Lakes	1.756
	(0.561)
neighborhoodThe Strip	0.263***
	(0.256)
neighborhoodUniversity	0.667
	(0.417)
review_count	1.001***
	(0.0002)
Category.NameBanquet Kitchen	0.190
	(1.193)
Category.NameBar / Tavern	0.507*
	(0.371)
Category.NameBarbeque	19,350,923.000***
	(0.00000)
Category.NameBuffet	0.003***
	(0.297)
Category.NameCaterer	4.149**

	(0.648)
Category.NameConcessions	20,801,498.000***
	(0.00000)
Category.NameConfection	1.314
	(0.736)
Category.NameFood Trucks / Mobile Vendor	1.670
	(0.648)
Category.NameGrocery Store Sampling	0.829
	(1.413)
Category.NameMain Kitchen	6.421**
	(0.787)
Category.NameMeat/Poultry/Seafood	7.114***
	(0.758)
Category.NamePortable Unit	0.797
	(0.701)
Category.NameRestaurant	0.566*
	(0.315)
Category.NameSnack Bar	1.939*
	(0.383)
Category.NameSpecial Kitchen	2.563
	(1.302)
Current.Violations	1.019
	(0.017)
Current.Gradeb	0.391*
	(0.536)
Current.Gradec	0.649
	(0.701)
Current.Gradex	0.419
	(0.848)
Inspection.Violations	0.959**
	(0.017)
Inspection.Gradeb	1.260
	(0.280)
Inspection.Gradec	1.717
	(0.526)
Inspection.Gradex	2.183
	(0.714)
Observations	1,131
Note:	*p<0.1; **p<0.05; ***p<0.01