The Determinants of Yelp Ratings: An Analysis of Toronto Businesses in 2017

Thomas Keough

April 16 2023

1 Introduction

With broad access to review data on websites, online business ratings are becoming increasingly relevant to business owners. This is especially true of service-oriented businesses, such as restaurants, hair salons, and hotels. Customers prefer to visit well-reviewed businesses, which means businesses should care about their online ratings. By extension, this means that both businesses and customers are interested in identifying what causes a business to earn good reviews. This analysis seeks to understand the determinants of online business ratings by using metrics available on Yelp, one of the most popular websites for reviews.

This analysis using data acquired from Kaggle.com. The data was initially published by Yelp as part of a dataset challenge. It was last updated in

August 2017 and includes data from 11 cities in four countries, including the USA and Canada. Additional Yelp data on prices is gathered using the Yelp Fusion API. Population data is acquired from Statistics Canada. The dataset is composed of files pertaining to business attributes, hours of operation, reviews, check-ins, tips, and users. This analysis merges the files for business attributes, hours of operation, check-ins, prices, and population for businesses in Toronto, Ontario. The dependent variable is business ratings, measured in Yelp stars. The chosen covariates are business' total number of Yelp reviews, total weekly hours of operation, daily average number of Yelp check-ins, business prices as measured by the number of dollar signs on Yelp, and population by forward sortation area (FSA).

The results of this analysis demonstrate that the number of reviews, hours of operation, and number of check-ins are all positively correlated with business ratings. The effect of the number of reviews depends in part on the type of business. By splitting the observations into three categories (shops, restaurants, other), restaurants have the lowest average rating while shops have the highest. Neither business price levels nor the population of a business's FSA appears to be correlated with businesses ratings. In order to determine the strength and magnitude of these correlations, a multiple regression analysis should be conducted.

Many studies have been conducted using Yelp data thanks to it being easily accessible through Yelp's Fusion API. Luca (2016) found that a one-star increase in Yelp ratings generates between five and nine percent higher revenues for restaurants in Seattle. Luca utilized a regression discontinuity design to infer causality with this relationship. The pursuit of rating maximization is therefore worthwhile for restaurants (at least). However, Luca's study does not draw conclusions about what a business can do to increase its Yelp rating. Given that higher ratings create more revenue for restaurants, this study's objective stands to be very relevant for businesses in the internet age.

Other research has been able to identify what factors are negatively correlated with Yelp ratings. Byers, Mitzenmacher, and Zervas (2012) find that businesses who provide Groupon promotions suffer from a significant decline in Yelp ratings on average. They suggest that Groupon customers are often treated more poorly relative to others. Naturally, their result contributes to identifying the determinants of Yelp ratings since it illustrates that poor service is punished with negative Yelp reviews. Conversely, it is reasonable to assume that good service is to be rewarded with positive reviews. Though this is self-evident, it is an important foundation on which the intitution for rating predictors can be built. With this in mind, this paper seeks to measure "good service" through business accessible through Yelp. Vinson, Dale, and Jones (2019) discovered that reviewers tend to systematically bias their present reviews away from their previous reviews. Though their study is focused on an application to human cognition, it still reveals that holding the reviwer constant, Yelp reviews are in part determined by previous reviews. This demonstrates that Yelp ratings are not necessarily an objective measure of business quality. Therefore, there must factors that contribute to biases in reviews beyond the dichotomy of good versus bad service. These factors may be measurable through variables that account for customer access (i.e. weekly hours of operation) or social popularity (i.e. average number of check-ins). Thus, this paper will partially account for reviewer bias in analyzing predictors of Yelp ratings.

2 Data

This analysis seeks to identify the determinants of business ratings using quantitative data gathered from Yelp. The outcome variable is the business rating, which is referred to as ”stars” in the dataset. This variable takes a value between 1 and 5. It is discrete data; potential values must be multiples of 0.5. This analysis includes four covariates of interest.

The review system is an extremely prominent feature of Yelp and is used frequently, so the number of reviews should account for a significant proportion of the variation in business ratings. There are two valid hypotheses for the correlation between the number of reviews with business ratings. For instance, a business that has exceptionally good service can encourage Yelp users to leave positive reviews. This would suggest that the number of reviews is positively associated with business ratings. However, the opposite effect is also feasible; exceptionally poor service may encourage Yelp users to leave negative reviews, meaning the number of reviews would be negatively correlated with business ratings. The dominant hypothesis, should one exist, will manifest itself through an empirical analysis of the relationship between review count and ratings.

The average number of Yelp check-ins per day is a measure of business popularity. It is important to note that this calculation of the daily check-in average ignores days wherein businesses had zero check-ins. In other words, it is the average number of check-ins for days that had at least one check-in. "Checking in" is a Yelp feature that allows users to inform their Yelp following of the businesses they visit. When a user "checks in" to a business, their attendance is published on their profile. Users can also earn badges and special offers at businesses they check in at. These two components provide increased incentives for Yelp users to visit high-quality businesses; users can show off their attendance at a popular business and potentially earn discounts in the process. Comparatively, poor-quality businesses would have fewer check-ins since users would not want to publish their attendance at them. Therefore, this variable accounts for customer perception of businesses. Therefore, check-ins should be positively correlated with a business's ratings in theory. An OLS regression that utilizes this covariate and the number of reviews would provide insight on how the number of reviews affects a business's rating while holding constant the customers' perception of the business. This would provide insight on the causality of business ratings.

The weekly number of hours of operation represents customer access to businesses. Firms differ largely in hours of operation on a day-to-day basis, so using the weekly total of operating hours will account for those differences across the days of the week. This variable will provide insight on the effect of customer access to businesses, which may be a component in evaluating business quality from the perspective of customers. There are several hypotheses for the direction of the relationship between operating hours and business ratings. It could be that businesses that are open longer throughout the week earn more reviews because they can serve more customers. Therefore, the direction of the relationship would be dependent on the ambiguous effect of the number of reviews on business ratings. Alternatively, businesses that are open for fewer hours per week may instill a sense of exclusivity in its customers. For example, fine dining venues, nightclubs, and other businesses that have limited weekly operating hours may observe more positive reviews because attendees feel exclusive.

Often, the price of a product is a significant factor in a consumer's decision to buy. On Yelp, the price level of a business is denoted by an integer quantity of dollar signs. The minimum is one dollar sign, which indicates that the business's good or service is priced low. The maximum is four dollar signs, which suggests that the business is among the most expensive in the region. There are two hypotheses for the effect of prices on Yelp ratings. Firstly, the cheap-positive hypothesis: cheaper products may be positively correlated with business ratings. Customers who spend less money on a purchase may make them happier, which in turn improves their perspective on the quality of the quality of the business they were shopping at. Further-more, customers can shop more frequently at businesses that are affordable. Repeat customers may be more likely to leave positive reviews. On the other hand, the expensive-positive effect may dominate. Cheaper products may suggest low quality for some customers, and would thereby yield negative reviews. By contrast, businesses that sell high-priced goods often advertise themselves as being high-quality. This may affect customer perception and lead to better reviews, even if the objective quality of whatever good or service being transacted is no greater than that of a cheaper alternative. So regardless of a business's price level, it will appeal to some customers and repel others. Including business price level in this analysis would help identify whether the cheap-positive or expensive-positive effect dominates with regard to a business's Yelp rating. This would provide a clearer picture of the determinants of Yelp ratings. The price level of a business is also an important control variable for a multiple regression in this analysis. For example, cheaper businesses may see more customers on average compared to expensive businesses which suggests that cheaper businesses would have more reviews.

3 Summary Statistics

Table 1: Summary statistics, $n = 14850$

	Rating	Reviews	Hours per Week	Daily Check-ins	Price Level
mean	3.495	28.382	47.455	1.524	1.535
std	0.841	56.379	35.824	1.092	0.988
min	1.000	3.000	0.000	1.000	0.000
25%	3.000	5.000	0.000	1.000	1.000
50%	3.500	10.000	54.000	1.188	2.000
75%	4.000	28.000	74.500	1.556	2.000
max	5.000	1494.000	167.883	31.043	4.000

The summary statistics for ratings indicate that the data is discrete. Moreover, it demonstrates that the minimum and maximum ratings are 1 and 5 respectively. The average review is roughly 3.5 stars. Assuming that an average business should have a rating near the midpoint between the minimum and maxmimum, the mean may indicate that there is a systematic bias towards over-rating businesses by 0.5 stars at the aggregate level. The interquartile range is between 3 and 4 stars, which further suggests this bias; it demonstrates that half of all businesses in Toronto are above the rating midpoint. In other words, only 25% of business can be considered poor-quality (i.e. a rating below 3 stars).

There appears to be a great variation in the number of reviews across businesses but it is primarily due to an enormous right skew. The standard deviation (56) is double the mean review count (28). The maximum review count of 1494 is indicative of right skew since it is 53 times larger than the average. Moreover, 75% of all Toronto businesses have 28 reviews or less. This may yield low power in a regression analysis since the vast majority of businesses have very similar review counts, despite the extreme variation suggested by the standard deviation.

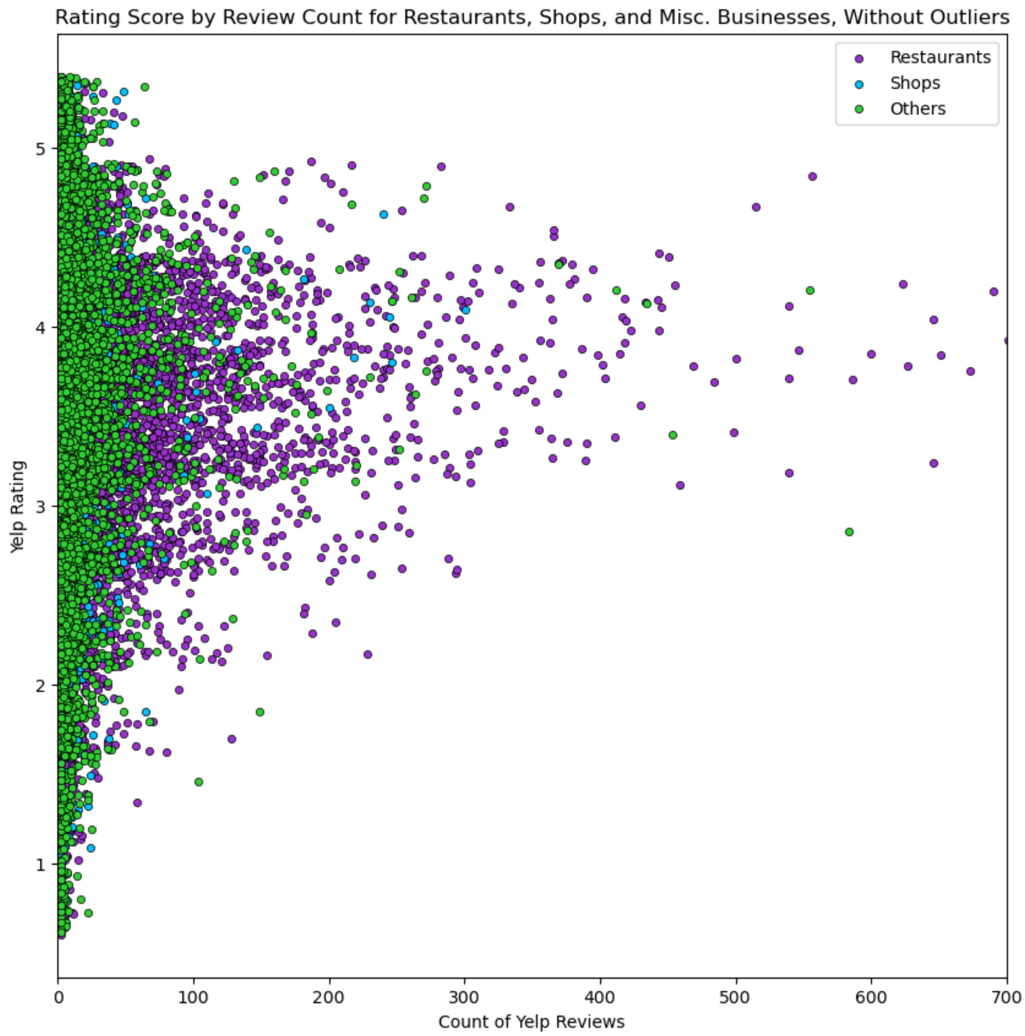
On average, businesses in Toronto are open for roughly 47 hours per week. The standard deviation is roughly 36 hours per week which indicates that there is a large variation in weekly operating hours across businesses. There are several signs of right skew in this variable, which may be the cause of the large standard deviation. The 50th percentile (54) is larger than the mean, which suggests that outliers on the right side of the distribution are positively biasing the mean. Moreover, this variable's lower bound of zero suggests that most of the variation would occur above the mean. The maximum value observed for this variable is 167.88, which is 0.12 hours less than the total number of hours in a week. This observation may need to be dropped from analysis if it biases the trends of interest, since the vast majority of businesses cannot operate for nearly every hour of the week.

There is very little variation across businesses in their daily check-in averages. Despite the mean and standard deviations being approximately 1.52 and 1.10 respectively, 75% of the observations lie between 1 and 1.55. There is a significant right skew in this data since the maximum value observed is roughly 31 check-ins on average. This variable may prove irrelevant in a regression analysis due to the uniformity of the data: there is not enough variation to properly calculate how an outcome variable changes given a change in the daily check-in average.

The price levels of most businesses in Toronto are two dollar signs or below on Yelp. Since there are many more observations at a price level of two or less compared to a price level of 3 or more, this analysis may be unable to determine any relationship between the price level and Yelp ratings. This is because there is a great lack of variation in the data. As the vast majority of businesses have an average price level of roughly 2, a regression including the price level would probably be unable to determine a statistically significant effect.

4 Visualizations

Figure 1: Rating by review count for restaurants, shops, and misc. businesses



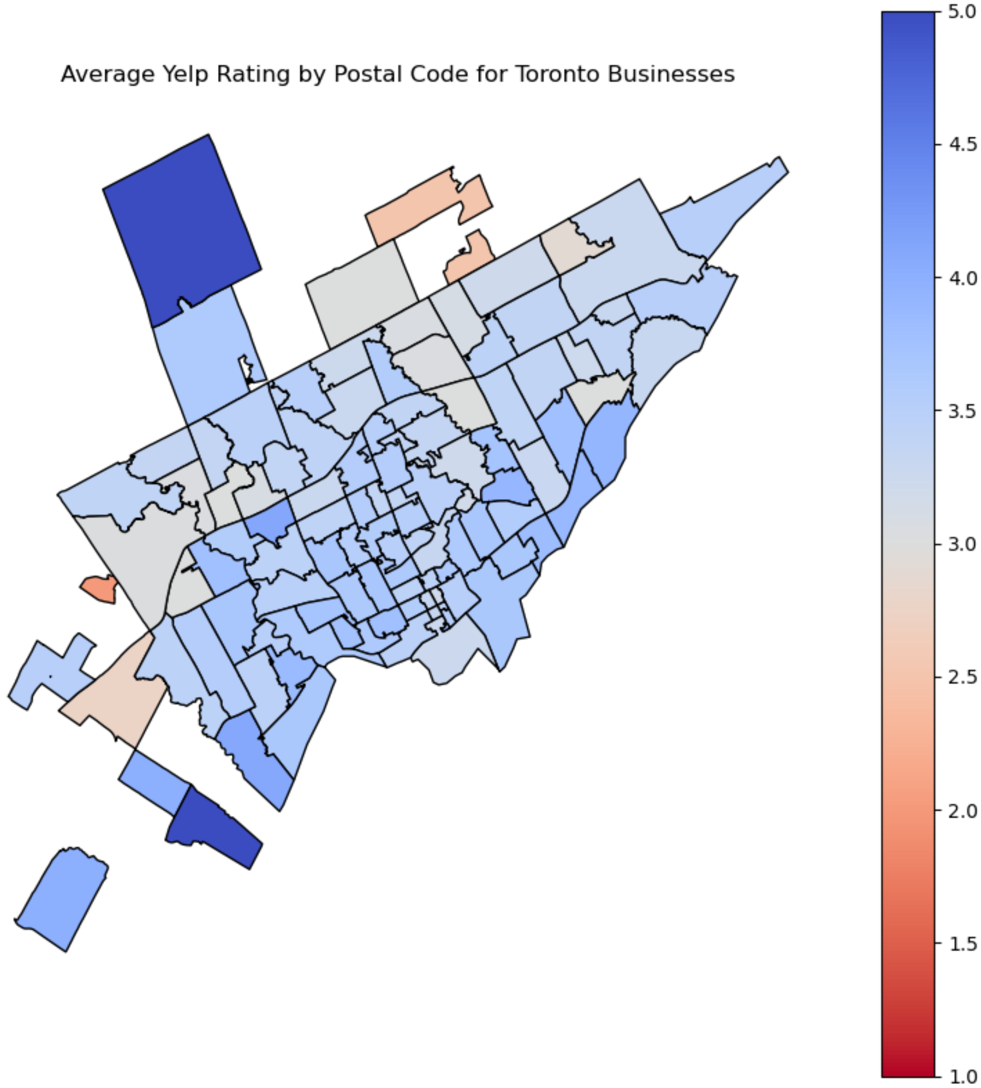
This scatter plot demonstrates the relationship between a business's count of Yelp reviews and its Yelp rating. Different colours are used to portray different types of businesses: restaurants are in purple, shops are in blue, and miscellaneous (non-shops, non-restaurants) are in green. Each type of business also has its own line of best fit, which is calculated with the OLS method. This plot ignores outliers that have more than 700 reviews.

For restaurants, there appears to be significant non-linearity in the distribution of Yelp ratings across reviews counts. Specifically, the distribution appears to be logarithmic. Each additional review provides much greater returns to ratings prior to the one-hundred review count threshold. Beyond this threshold, the returns of an additional review diminish dramatically. After a restaurant has earned roughly two-hundred reviews, the marginal benefit to ratings of an additional review is virtually zero on average.

This nonlinearity is not evident in the relationship between ratings and review count for shops or miscellaneous businesses. In fact, neither shops nor iscellaneous businesses appear to have any significant correlation between ratings and count of reviews. Shops appear to have significant variation across review counts but there are so few shops in the data that drawing conclusions about the strength of a relationship is difficult. Inversely, there are many miscellaneous business in the data but there is very little variation in their review counts. In reality, miscellaneous businesses may have a similar nonlinear relationship but are perhaps inherently less likely to receive greater numbers of reviews. In order to remedy this in a regression, additional controls should be included.

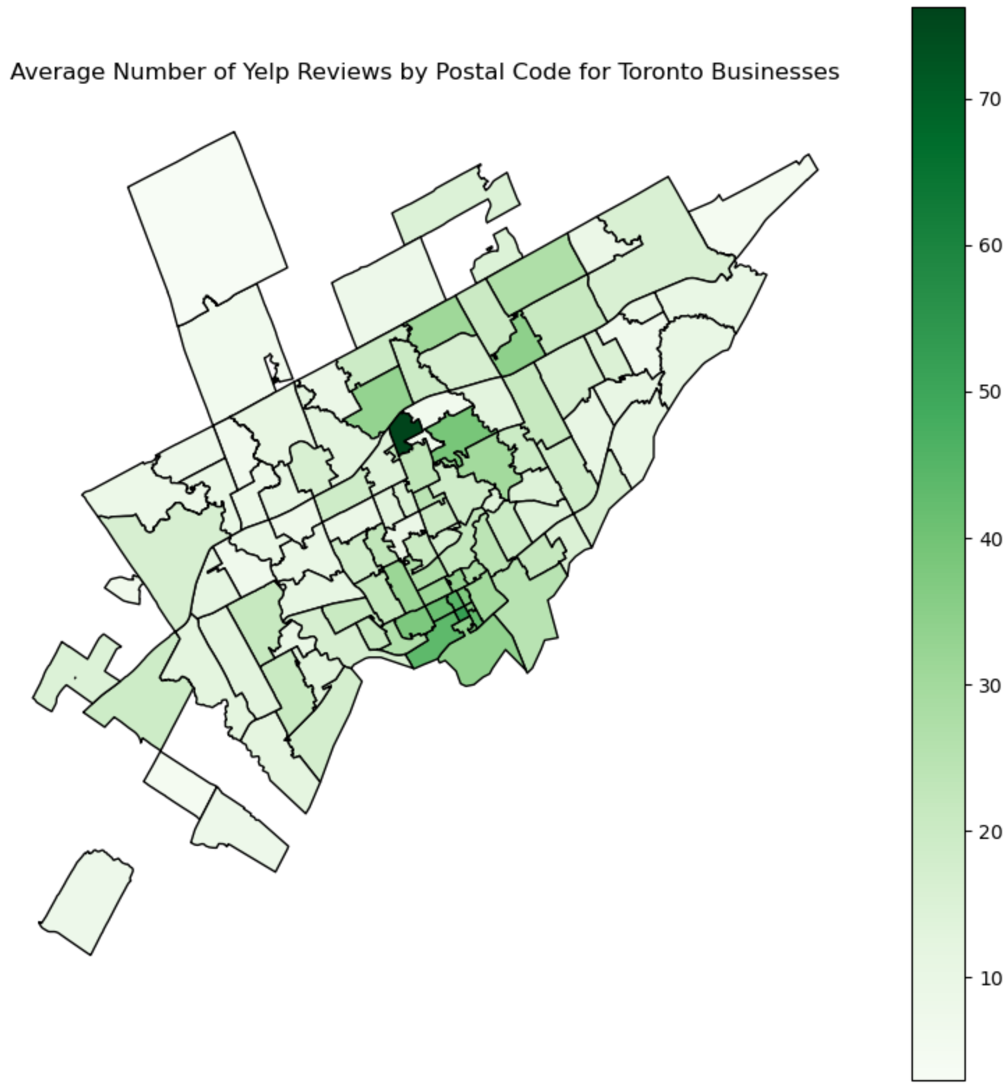
All in all, this plot provides significant insight with regard to this paper's main message. It demonstrates that the number of Yelp reviews that a restaurant has is correlated with its Yelp rating. However, there are significant diminishing returns to ratings for addtional reviews beyond two-hundred. Moreover, it shows that the effect of a determinant of ratings likely depends on the type of business, at least in the case of review counts. This is because shops and miscellaneous businesses differ from restaurants in regard to how an additional review affects their respective Yelp ratings.

Figure 2: Average Yelp rating by postal code



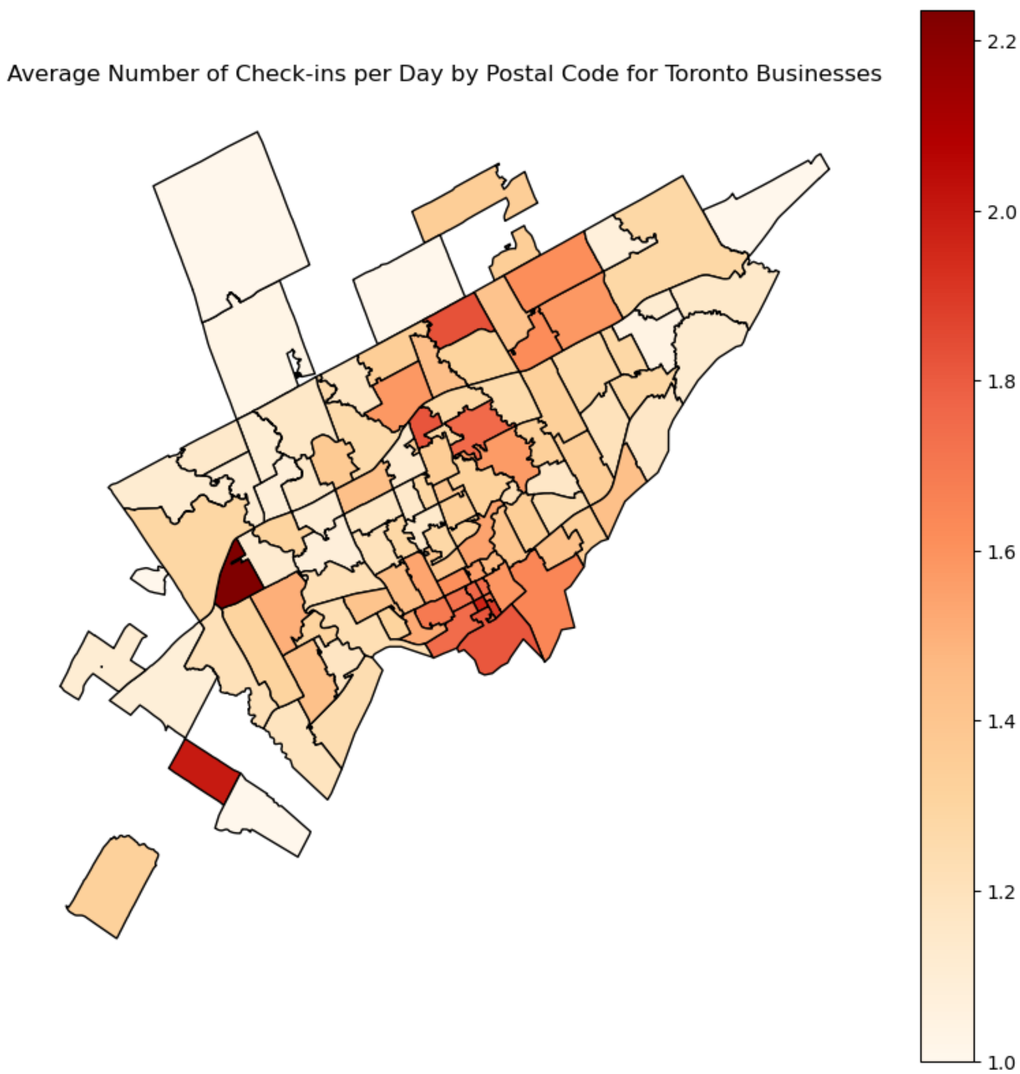
The map of average Yelp ratings by postal code demonstrates that physical geography does not play a significant role in determining Toronto businesses' ratings. Most postal codes have very typical ratings, especially in regions with greater population density. Outliers become more common as distance from downtown increases, but the direction of these outliers seems to be random. This suggests that the prominence of outliers may be the result of each business having relatively fewer reviews. With fewer reviews in these regions, the expected rating is more likely to deviate from the true mean rating.

Figure 3: Average number of reviews by postal code



Contrary to the map of average Yelp ratings, the average number of reviews seems to be influenced by geography. The postal codes with the highest average number of reviews are seen nearer to downtown Toronto. However, there are significant outliers to this trend in the city's northern regions. It is clear that there is significant variation in reviews across postal codes. This is likely because there are popular restaurants all over the city: regardless of where a business is located, it will likely earn more reviews if it is popular on Yelp.

Figure 4: Average number of reviews by Postal Code



Similar to the map for average number of reviews, the number of Yelp check-ins a business gets appears to be influenced by geography. The businesses that are located downtown have more check-ins per day on average. The number of check-ins seems to be much more varied than the number of reviews. This map has outliers of greater magnitude and with greater frequency since many postal codes outside of downtown seem to have high average daily check-in counts. This may be because check-ins are used infrequently on Yelp: given that the mean number of check-ins is 1.5 per day, only a few additional check-ins per day are required to be considered an outlier. One especially passionate customer could raise the average number of check-ins per day by 1 on their own. This theory is especially plausible for restaurants, since regular customers are commonplace in the food industry.

These maps demonstrate that there is great variation across postal codes in most of the covariates of interest. There is no clear reason why these metrics differ so significantly across regions. As a result, this analysis will include a suite of dummy variables to control for postal code variation.

5 Results

5.1 Ordinary Least Squares

In accordance with the theories of this paper so far, the count of reviews, the average number of daily check-ins, and the number of operating hours per week should be included in regressions for the Yelp rating. There is also evidence of a potential relationship between weekly operating hours and ratings. As a result, the inclusion of these primary variables of interest should explain some of the variation in Yelp ratings. Though the variables measuring postal code population and business price level are not obviously correlated with Yelp ratings, they are still important control variables. Including these in a regression model will enable the isolation of variation in the primary variables of interest. As a result, they may contribute to increasing statistical significance for regression results. Regression specification 1 includes the count of reviews, average number of check-ins per day, operating hours per week, and the price level:

$$stars_i = \beta_0 + \beta_1 reviews_i + \beta_2 checkins_i + \beta_3 hours_i + \beta_4 price_i$$

Regressions 2 builds on regression 1 by including a suite of dummies for postal codes:

$$stars_i = \beta_0 + \beta_1 reviews_i + \beta_2 checkins_i + \beta_3 hours_i + \beta_4 price_i + postal_{it}$$

Regression 3 uses a quadratic models for the count of reviews and check-ins:

$$stars_i = \beta_0 + \beta_1 reviews_i^2 + \beta_2 reviews_i^2 + \beta_3 checkins_i + \beta_4 checkins_i^2 + \beta_5 hours_i + \beta_6 price_i + postal_{it}$$

Regression 4 uses logarithmic models for the count of reviews and check-ins:

$$stars_i = \beta_0 + \beta_1 ln\_reviews_i + \beta_2 ln\_checkins_i + \beta_3 hours_i + \beta_4 price_i + postal_{it}$$

Table 2: Regression Results

	Dependent variable: stars			
	(1)	(2)	(3)	(4)
constant	3.380***	3.488***	3.473***	3.477***
	(0.020)	(0.011)	(0.014)	(0.015)
review_count	0.000***	0.001***	0.001***	
	(0.000)	(0.000)	(0.000)	
review2			-0.000*	
			(0.000)	
ln_review				0.034***
				(0.009)
daily_checkin_avg	0.028***	0.036***	0.053***	
	(0.008)	(0.007)	(0.012)	
checkin2			-0.001***	
			(0.000)	
ln_checkin				0.092***
				(0.021)
wk_op_hours	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
price_level	-0.020**	-0.026***	-0.028***	-0.033***
	(0.008)	(0.008)	(0.008)	(0.008)
Observations	14,774	14,774	14,774	14,774
R^2	0.012	0.058	0.058	0.058
Adjusted R^2	0.012	0.050	0.051	0.051
Residual Std. Error	0.836	0.820	0.820	0.819
F Statistic	74.659***	141.099***	72491.924***	104.767***

Notes: *p<0.1; **p<0.05; ***p<0.01
Standard errors in parenthesis.

The best specification among the ones constructed is regression 3. This is because it is tied for the greatest R^2 value, meaning it accounts for the most amount of variation among the specifications used. Regression 3's Fstat is also the largest, meaning the covariates included in its specification have the greatest joint significance. Nearly all of the covariates of interest are statistically significant at the 1 percent level, except for the squared term on review count which has a p-value greater than 0.05. By extension, it accounts for the nonlinear relationships between Yelp ratings, review count and average check-ins in the most statistically significant way.

The most obvious insight presented by this suite of regressions is that there is significant variation across postal codes. Regression 1, which does not control for postal codes, has an R^2 of 0.012. All specifications that control for the suite of dummies for postal codes have an R^2 of 0.058. This means that roughly five percentage points more variation in Yelp ratings are explained by differences across postal codes in Toronto, and therefore a business's postal code partially determines its Yelp rating. That said, 0.058 is still a very low R^2. This means that much more significant determinants of ratings exist which are unaccounted for in these regressions.

The number of reviews a business has is also a significant determinant of Yelp ratings in every specification, though to varying degrees of practical significance. Regressions 1 through 3 indicate that an increase of 100 in a business's count of reviews is correlated with a 0.1-point increase in the Yelp rating on average. Regression 4 suggests that a 1-percent increase in the number of reviews is correlated with a 0.034 point increase in ratings. As ratings are rounded to the nearest half-point, this correlation would only be visible on Yelp for businesses that have 500 more ratings. Considering that the average number of reviews is roughly 28.3 and the seventy-fifth percentile in the distribution of the review count variable is 28, the magnitude of the coefficient for the number of reviews is remarkably insignificant in the practical sense. Moreover, regression 3 suggests that the number of reviews may have diminishing returns to ratings since the coefficient on the squared term for review count is negative. However, this finding is not resolute because this coefficient has low statistical significance (P¡0.1). While review count is certainly a statistically significant predictor of ratings, it is not feasible for a business to increase its Yelp rating by increasing its number of reviews alone. This is because reviews are not inherently good or bad: without controlling for business quality, every additional review is equally likely to increase or decrease the business's rating.

The average daily number of check-ins is a significant predictor of Yelp ratings. Regression 3 indicates that there is significant non-linearity in the relationship between check-ins and ratings. This suggests that there is diminishing returns to ratings for the number of check-ins that a business receives. An increase of 10 in the average number of check-ins per day is correlated with a 0.43 increase in ratings, which is rather large considering that would almost be enough to increase the rating portrayed on a business's Yelp page. This means that the average number of daily check-ins is a significant predictor of ratings. This makes intuitive sense since a check-in is a clear indicator that a customer enjoys shopping at a business; if a customer didn't like a business, they wouldn't want to show off that they are shopping there, so they would not check in.

The number of weekly operating hours and price level are not relevant predictors of ratings. Each specification finds that one additional hour of operation per week is correlated with a 0.002 higher rating. Since there are only 168 hours in a week, operating hours can only be correlated with a 0.336 ($168 \cdot 0.002$) increase in ratings points at most. Similarly, the price level is an integer quantity between 0 and 4. Using the largest coefficient found for this variable (regression 4), the price level can be correlated with a 0.01 ($3 \cdot -0.033$) ratings point decrease at most. Curiously, the coefficient on the price level is negative and statistically significant in each regression. This suggests that higher prices are not conducive to higher ratings, though the negative effect they have on ratings is negligible.

All in all, the only practically significant determinants of ratings found are the average daily number of check-ins and postal code. Though statistically significant, the number of reviews, price level, and number of hours of operation per week are practically insignificant.

5.2 Regression Tree

Let j refer to the number of mutually-exclusive regions in the regression space. Let s refer to the optimal threshold on which to split the regression space in order to minimize the sum of squared residuals (RSS). The objective function is:

$$\min_{j,s}\left(\sum_{i:x_{i,j}\leq s, x_i\in R1}(stars_i - s\hat{tar}s_{R_1})^2 + \sum_{i:x_{i,j}>s, x_i\in R2}(stars_i - s\hat{tar}s_{R_2})^2\right)$$

where $\hat{y}_{R_1}$ and $\hat{y}_{R_2}$ refer to the average of Yelp ratings in each regression space:

$$s\hat{tar}s_{R_m} = Average(stars_i | x_i \in R_m)$$

R_1 and R_2 refer to partitions of the regression space given by:

$$R_1 = \{X | X_j < s\}$$

$$R_2 = \{X | X_j \geq s\}$$

X belongs to the chosen covariates:

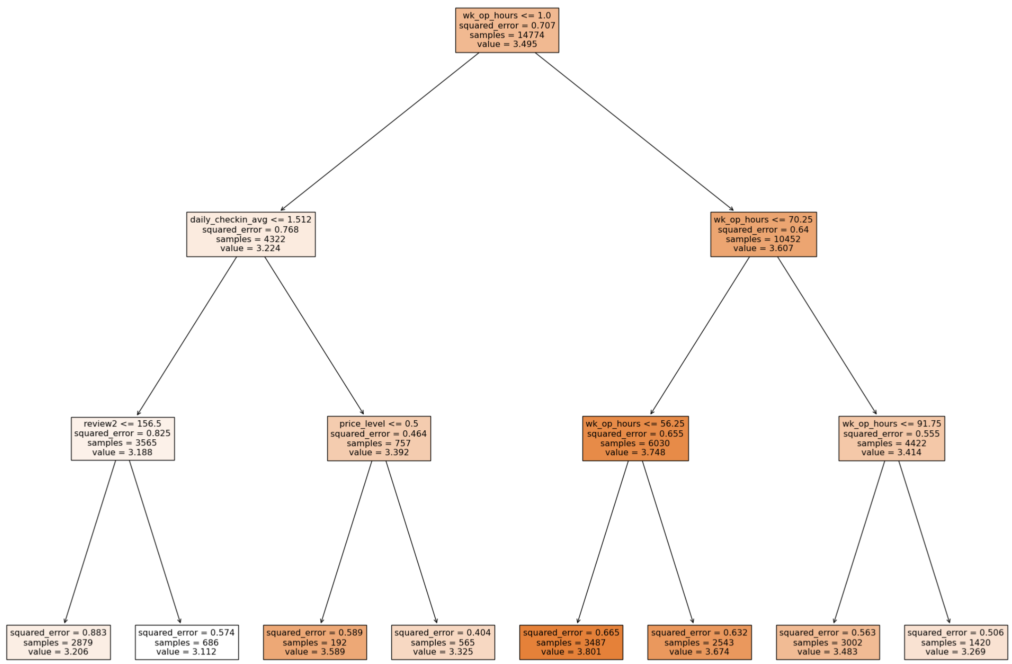
$$X \in \{review, checkin, hour, price\}$$

For each partition, the objective function is re-evaluated with sub-partitions of the regression space. The tree is then pruned according to the following function:

$$\min_{tree \subset T} \sum (\hat{f}(x) - stars)^2 + \alpha |T|$$

where α is chosen to minimize this equation and T is the number of terminal nodes in the tree.

There are three main regulization parameters in this model. They are the minimum leaf size, maximium tree depth, and the alpha parameter. The minimum leaf size contributes to preventing prediction from being too granular. Hyperspecific groupings have little relevance because they are incapable of producing significant predictions for observations that differ even slightly from them. Leafs must not be too small in order to ensure a reasonable prediction can be made about an observation that is not identical to the ones that were used to construct the regression tree. The maximum tree depth is important to regulate for a similar reason. Splitting on many variables is conducive to producing smaller leaves, which would make the predictions of a regression tree not generalizeable. The alpha parameter also contributes to balancing complexity and quality of the regression tree by increasing the MSE for highly-specified models. For each additional terminal node in a regression tree, the MSE is increased by alpha (which must be greater than 0).



The MSE for the regression tree is 0.651. The average error is therefore 0.81 stars ($\sqrt{0.651}$). As ratings must be between 1 and 5, this is extremely large for the error of prediction for each individual business's rating. This indicates that the chosen covariates are poor predictors of Yelp ratings.

Both the OLS and regression tree models predict very low magnitudes of correlation among most of the covariates. The range between the minimum and maximum value of ratings among leaf nodes is less than 0.7. Since the mean rating is 3.5, a variation of 0.7 is only enough to increase a business's displayed rating by 0.5. The regression tree indicates that this the range between the lowest and highest predicted values of business ratings, so most of the variation in ratings must not be determined by the chosen covariates. Similarly, the OLS models in section 4.1.3 have coefficients of very low magnitude. Though most are statistically significant, that is likely because the number of observations (¿14000) yields great statistical power. The OLS and regression tree models both indicate that the chosen covariates are not relevant determinants of business ratings.

The regression tree also demonstrates that much of the variation in ratings occurs in relation to the number of operating hours per week. This is evidenced by all subtrees of the right subtree, and the tree's root itself, being split on this variable. Moreover, the regression tree indicates that having fewer weekly operating hours is correlated with higher Yelp ratings. This runs contrary to the results of the OLS model, which indicated that weekly operating hours were slightly positively correlated with ratings. This is likely because the OLS model controls for the other covariates, whereas the regression tree does not. This suggests that operating hours are confounded by one of the other covariates in the analysis, since their influence on ratings is much larger when variables are not controlled for. As a result, the regression tree overestimates the returns of a business's operating hours to its rating.

6 Conclusion

Prior to this paper, the relevant literature did not identify how business metrics on Yelp may affect business ratings. This analysis demonstrates that the number of reviews and average number of check-ins per day appear to be slightly positively correlated with business ratings in Toronto on Yelp. According to the OLS results, there are precise null effects for weekly hours of operation on business ratings. In other words, high-quality businesses tend to have more reviews, are open for longer, and receive more check-ins on average relative to poor-quality businesses. Check-ins have a significant quadratic relationship with ratings, meaning there are diminishing returns to having high numbers of check-ins. Businesses that are located in high-population regions do not seem to gain any advantage or disadvantage with regard to ratings. That said, a relatively large portion of variation in ratings is captured by postal codes. Since there is no apparent correlation between geographical location and ratings in Toronto, it could be that variation across postal codes is essentially random. The price level of a business is negatively correlated with ratings but is neglible. While these relationships are all statistically significant, only the number of check-ins seems to have practical significance in determining ratings.

Since the number of reviews, weekly operating hours, postal code population, and price level are all relatively insignificant, this study has been largely unsucessful in identifying the determinants of Yelp ratings for Toronto businesses. Regression results indicated that this analysis accounted for no more than six percent of the variation in business ratings. It is clear that most of the data presented on Yelp is actually not relevant in predicting business's Yelp ratings. This makes intuitive sense, since businesses can very greatly in price, availability, and location; these descriptors, though relevant to consumers, are not inherent indicators of business quality. Businesses can be very popular and earn many reviews, but they may not be of high quality. Similarly, businesses can be open for long hours but are not necessarily better as a result. The price level of a business may be a prominent indicator of popularity, but is not a clear predictor of ratings. Cheaper products may be of lesser quality relative to expensive products, but expensive products are less accessible and not inherently better than cheaper products.

This study has also been unable to identify causal relationships. It is likely that Yelp ratings can themselves cause increase in the number of reviews or check-ins. Businesses with high ratings likely garner more attention, which may lead to an influx of customers willing to write reviews and check-in. Though they are significant, the prospect of reverse causation further suggests that these variables are not strong predictors of ratings.

The invalidity of these covariates demonstrates that business ratings are better predicted by variables that more precisely measure business quality. However, such variables may be more difficult to acquire. Since tangible measures of consumer experience (such as cost and access) are not reliable predictors of ratings, variables should be constructed based on more intagible metrics. For instance, text-based analysis of Yelp reviews may identify keywords and sentiments in reviews for well-performing businesses. By omtting words that obviously indicate a favourable rating like "good" or "fun", qualitative characteristics for businesses with high ratings may be identifiable. Yelp has an immense amount of review data for most businesses that are listed on it, so there would be no issues with data shortages. Due to Yelp Fusion's daily call limit of 5000, it may take weeks to acquire all the data needed. As a result, a text-based analysis of reviews may generate more insightful results regarding the determinants of ratings.

References

Byers, John W., Michael Mitzenmacher, and Georgios Zervas. 'The Groupon Effect on Yelp Ratings: A Root Cause Analysis'. In Proceedings of the 13th ACM Conference on Electronic Commerce, 248–65. EC '12. New York, NY, USA: Association for Computing Machinery, 2012. https://doi.org/10.1145/2229012.2229034.

Luca, Michael. 'Reviews, Reputation, and Revenue: The Case of Yelp.Com'. Harvard Business School Working Paper 12-016, 2016. https://www.hbs.edu/ris/Publication%20Fil 016_a7e4a5a2-03f9-490d-b093-8f951238dba2.pdf.

Vinson, David W., Rick Dale, and Michael N. Jones. 'Decision Contamination in the Wild: Sequential Dependencies in Online Review Ratings'. Behavior Research Methods 51, no. 4 (1 August 2019): 1477–84. https://doi.org/10.3758/s13428-018-1175-8.