Final_Paper

Contents

This project investigates how the volume of photos attached to Yelp reviews relates to a restaurant's average ratings, and if the type of picture (menu, food & drink, or environment) affects this relationship.

1. Research Motivation

The influence of online reviews and electronic word of mouth (eWOM) on consumer perceptions and decision-making is evident in today's digital society (Wang et al., 2021). As such, this phenomenon has been widely researched in marketing. The influence of reviews is especially pronounced in the case of "experience goods". For example, restaurant services are perceived as riskier to evaluate before purchase (Weisskopf, 2018).

Previous findings suggest that by reading restaurant reviews, customers are able to reduce this perceived risk (Parikh et al., 2014). Further supporting this claim, Luca (2016, p.3) demonstrates that "a one-star increase in Yelp rating leads to a 5–9 percent increase in revenue".

Photos have become a key component of online reviews as they add a visual element that allows users to better communicate their opinions (Li et al., 2021). However, while text-based reviews have been extensively studied, the association between the number of photos included in reviews and a restaurant's average rating remains underexplored. Additionally, current literature is limited with regard to how different types of photos may impact this relationship. To address this gap in the literature, the following research question has been formulated:

2. Research Question

How does the total number of photos included in Yelp reviews influence a restaurant's average rating, and to what extent does the type of photo (food, environment, menu) moderate this relationship?

3. Managerial Relevance

From a managerial and business perspective, the stakeholders, namely restaurant owners and managers, may gain a deeper understanding from the findings of this project on how to potentially raise ratings per restaurant, reduce average rating volatility and increase visibility and engagement. Luca (2016) concluded that the average rating of a restaurant guides customers to make informed decisions. Ultimately, this may lead to an improved reputation and subsequently increased customer demand.

Furthermore, managers can enhance their marketing strategy by strategically encouraging customers to publish reviews with photos on platforms such as "Yelp" or "Tripadvisor" as it reduces customer uncertainty. To encourage reviews that include photos, managers may add QR codes on the menu or receipts that could incentivize customers with rewards for the uploaded photos. Additional innovative approaches are table toppers and Wi-Fi login prompts inviting the customers to share a photo of the food or environment. By prompting customers to share additional photos, managers may be more inclined to detect recurring issues, creating a feedback loop.

4. Data

4.1 Data Sourcing

This project uses two data sets obtained from the Yelp Open Dataset. One data set contains business-related information, while the other provides detailed photo data, including both the images and their associated classifications, such as "food," "drink," "menu," "inside," and "outside."

4.2 Data Preparation and Variables

To answer the research question, these two data sets were merged, cleaned, and transformed. This process resulted in "final_dataset.csv," which serves as the data set used for analysis. This CSV file includes 29,374 observations across 10 variables. An overview of these variables is provided below.

Table 1: Overview of Variables in Final Dataset

Variable	Description	Data_Class
business_id	The unique Yelp ID of the business	Character
name	The business name as shown on Yelp	Character
attributes	The map on Yelp of a restaurant's amenities, services, and policies	List
categories	The list of Yelp categories of cuisines for the business	Character
stars	The average Yelp scale star rating (1–5)	Numeric
review_count	The total number of Yelp reviews	Numeric
environment	The number of environment photos	Numeric
food & drink	The number of food & drink photos	Numeric
menu	The number of menu photos	Numeric

5. Data Exploration

Before conducting the analysis, it is essential to gain a comprehensive understanding of the data set to identify its characteristics, patterns, and potential limitations.

The data contains 29,374 restaurant observations, each with measures of average Yelp rating (1–5 scale) and photo counts. These counts are broken down by category (food & drink, environment, and menu), and also include a total photo count. The mean restaurant rating is approximately 3.74 stars, indicating a moderate overall quality perception. The average total number of photos per restaurant is about 59. However, it is important to note that the distribution is heavily right-skewed. This means that many restaurants have only a few photos, while a small number have hundreds. This uneven distribution is important context as it indicates that although the large sample size guarantees statistical precision the size of the effects should be considered realistically.

Additional data exploration and visualizations conducted using ggplot2 are documented in the data_exploration file.

6. Method

This project employs linear regression analysis as the primary research method, chosen for its suitability in examining the relationship between restaurant ratings and photo-related variables while allowing for straightforward interpretation of effects.

To address the research question, a series of linear regression models were estimated. First, a simple linear regression model was used to obtain a baseline understanding of the relationship between the average star rating of a restaurant and its total number of photos.

The baseline model is specified as follows:

$Stars = \beta 0\ + \beta 1\ (Total\ Photos) + \varepsilon$

Second, a multiple linear regression model was estimated. This second model expands upon the baseline by including additional predictors, such as photo category (food and drink, environment, and menu), and an interaction term. The inclusion of an interaction term allows for an assessment of whether the effect of photo quantity varies depending on the dominant photo category. This modeling approach allows for a better understanding of both the combined and conditional effects of photo content on restaurant ratings.

As such, the main model is formally specified as follows:

$Y = \beta 0 + \beta 1 (Photos) + \beta 2 (PhotoCategory) + \beta 3 (Photos \times PhotoCategory) + \varepsilon$

where Y represents the average restaurant rating, $Photos$ denotes the total number of photos, and $PhotoCategory$ indicates the dominant photo type category. The interaction term $\beta 3(Photos \times PhotoCategory)$ captures whether the influence of photo quantity on ratings differs across categories.

Lastly, an interaction model was estimated to more explicitly assess whether the effect of photo volume on ratings varies by the predominant type of photo. This model introduces interaction terms between the total number of photos and the dominant photo category, allowing for testing whether the slope of the relationship between total photos and star ratings differs across food-, menu-, and environment-dominant restaurants, rather than assuming a uniform effect.

The model is specified as follows: $Stars = \beta 0 + \beta 1 \, (Photos \ Centered) + \beta 2 \, (Category \ Dominant) + \beta 3 \, (Interaction) + \varepsilon$

7. Analysis

7.1 Baseline Model

To establish a preliminary understanding of the relationship between the total number of photos in reviews and the average star rating of restaurants, a baseline linear regression model was estimated. The output of the linear regression is presented below:

```
Call:
lm(formula = stars ~ total_photos, data = final_dataset)

Residuals:
    Min      1Q Median      3Q     Max
-4.683  -0.577   0.113   0.423   1.423

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.5673406  0.0052475  679.82   <2e-16 ***
total_photos  0.0096877  0.0004532   21.38   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7782 on 29372 degrees of freedom
Multiple R-squared:  0.01532,	Adjusted R-squared:  0.01529
F-statistic:   457 on 1 and 29372 DF,  p-value: < 2.2e-16
```

Model Interpretation

The coefficient for Total Photos is positive and highly significant ($\beta = 0.0097, p < .001$), indicating that restaurants with more photos tend to receive slightly higher average ratings. More specifically, each additional photo uploaded to a restaurant's Yelp page is associated with a 0.0097 point increase in its average star rating. In practical terms, an increase of 50 photos corresponds to approximately +0.48 stars on the 1–5 rating scale.

4

Explanatory Power: The model's explanatory power is very low ($R^2 = 0.015$), indicating that photo quantity alone explains only a small fraction of the variation in ratings. Most of the variation in ratings is driven by other factors (food quality, service, price, location, etc.), not simply by the number of photos. Thus, while the relationship is statistically overwhelming due to the sample size, the actual effect in practice is quite small.

Despite this limited explanatory power, the positive relationship aligns with theoretical expectations that photos in online reviews help consumers better evaluate restaurants, reducing perceived risk and increasing confidence in their decision-making, which in turn influences average ratings.

7.2 Main Model

To examine whether different types of photos contribute differently to restaurant ratings, a multiple linear regression model was estimated. The dependent variable remains the average star rating, while the independent variables represent photo counts across distinct categories: Food & Drink, Menu, and Environment. As such, this model builds upon the baseline results by decomposing the overall photo effect into more specific photo types.

The output of the regression is presented below:

```
Call:
lm(formula = stars ~ food_and_drink + menu + environment, data = final_dataset)

Residuals:
    Min     1Q  Median     3Q    Max
-5.4679 -0.5722  0.0717  0.4278  1.4281

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.5715547  0.0052523 680.000   <2e-16 ***
food_and_drink 0.0003278  0.0008089   0.405    0.685
menu           0.1327520  0.0154143   8.612   <2e-16 ***
environment    0.0221495  0.0011212  19.755   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7752 on 29370 degrees of freedom
Multiple R-squared:  0.02308,Adjusted R-squared:  0.02298
F-statistic: 231.3 on 3 and 29370 DF,  p-value: < 2.2e-16
```

Model Interpretation:

When controlling for all photo categories, menu photos exhibit the strongest positive association with average ratings. Specifically, each additional menu photo is associated with a 0.13-star increase, holding the number

of environment and food & drink photos constant. Environment photos also show a positive effect, though smaller in magnitude, with each additional photo contributing approximately +0.022 stars. In contrast, food & drink photos do not have a statistically significant effect, suggesting that once menu and environment photos are accounted for, additional images of food provide little incremental predictive value.

These results suggest different functional roles for each photo type: - Menu photos likely reduce informational uncertainty, allowing consumers to evaluate prices, portion sizes, and offerings before visiting, which increases confidence and trust. - Environment photos convey restaurant atmosphere and cleanliness, shaping emotional expectations. - Food photos, though abundant in the data set, may suffer from saturation and variable quality, explaining their negligible incremental impact.

Explanatory Power: The inclusion of photo types as moderators increases the model's explanatory power to $R^2 = 0.023$, representing almost a 50% improvement over the baseline model. While the overall explained variance remains small, the findings from this model indicate that photo type carries meaningful information beyond quantity.

7.3 Interaction Model

To assess whether the effect of photo volume on ratings varies by the predominant type of photo, an interaction model was estimated, introducing interaction terms between the total number of photos and the dominant photo category. This specification allows testing whether the slope of the relationship between total photos and star ratings differs across food-, menu-, and environment-dominant restaurants, rather than assuming a uniform effect.

```
Call:
lm(formula = stars ~ photos_centered * photo_category_dominant,
    data = final_dataset)

Residuals:
    Min      1Q  Median      3Q     Max
-5.2442 -0.5275  0.0921  0.4725  1.4834

Coefficients:
                                                          Estimate Std. Error
(Intercept)                                               3.744442   0.008312
photos_centered                                           0.008216   0.000878
photo_category_dominantFood_and_Drink                    -0.175589   0.009923
photo_category_dominantMenu                               0.164804   0.108742
photos_centered:photo_category_dominantFood_and_Drink     0.002652   0.001024
photos_centered:photo_category_dominantMenu               0.027144   0.024425
                                                         t value Pr(>|t|)
(Intercept)                                              450.504  <2e-16 ***
photos_centered                                            9.358  <2e-16 ***
photo_category_dominantFood_and_Drink                   -17.696  <2e-16 ***
photo_category_dominantMenu                               1.516  0.1296
photos_centered:photo_category_dominantFood_and_Drink     2.590  0.0096 **
photos_centered:photo_category_dominantMenu               1.111  0.2664
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7738 on 29368 degrees of freedom
Multiple R-squared:  0.02657,Adjusted R-squared:  0.02641
F-statistic: 160.3 on 5 and 29368 DF,  p-value: < 2.2e-16
```

Model Interpretation:

For environment-dominant restaurants (the reference group) with an average number of photos (5.8), the expected star rating is 3.74. Each additional photo above the mean, for restaurants where environmental photos is the dominant group, is associated with a small but statistically significant increase in rating of 0.0082 (p < 0.001), consistent with the baseline finding that more photos slightly raise ratings. The model further indicates that restaurants dominated by food and drink photos have lower average ratings than environment-dominant restaurants ($-0.176 stars, p < 0.001$), while menu-dominant restaurants show a small positive difference (+0.165 stars), though this is not statistically significant ($p = 0.130$).

The interaction terms indicate that the effect of additional photos differs by photo type. For food & drink-dominant restaurants, each extra photo contributes an additional 0.0027 to the star rating beyond the effect for environment-dominant restaurants ($p = 0.0096$), suggesting a slightly stronger positive relationship. For menu-dominant restaurants, each extra photo adds 0.027 points compared to environment-dominant restaurants, but this effect is not significant ($p = 0.266$), indicating that photo quantity does not meaningfully alter ratings in this category.

Explanatory Power: The model explains approximately 2.6% of the variance in star ratings ($R^2 = 0.026$) and is statistically significant. While modest, this shows that both the number and type of photos contribute meaningfully to predicting ratings. The low R² also highlights that many other factors, such as food quality, service, and pricing, likely influence ratings beyond what is captured by visual content.

7.4 Overview of Findings and Managerial Recommendations

Across all three models, the findings consistently demonstrate that the total number of photos on Yelp has a significant effect on average restaurant ratings. The baseline model establishes that restaurants with more photos tend to receive higher average ratings, supporting the idea that visual information enhances perceived credibility and reduces consumer uncertainty.

When investigating photo type, the results reveal that not all photos contribute equally. Menu photos have the strongest positive relationship with ratings, suggesting that they help customers form realistic expectations about offerings and prices. Environment photos also contribute positively, likely because they signal cleanliness, atmosphere, and other attributes that shape expectations before a visit. By contrast, food photos appear less influential once other types are controlled for, possibly due to oversaturation or inconsistent quality.

The interaction model further refines these insights by showing that while the overall photo–rating relationship is robust across categories, the strength of the effect varies slightly depending on which type of photo dominates a restaurant's gallery. Food-dominant restaurants start with lower ratings but benefit more from increasing photo volume, whereas menu-dominant restaurants start higher but do not gain additional advantage from extra photos.

Taken together, the findings suggest that managers should can implement strategies to improve perceived reputation and attract potential customers. Practical steps include: - Encouraging customers to upload diverse photos, especially of menus and interiors. - Prompting visual engagement through in-store QR codes, receipts, or social media campaigns. - Monitoring photo composition over time to maintain a balanced gallery that aligns with brand positioning.

8. Conclusion

This project investigated the relationship between the number and type of photos in Yelp reviews and restaurants' average star ratings. The analysis demonstrates that photos play a meaningful role in influencing ratings, but their impact varies by content category. Menu and environment photos are most influential, while food and drink photos alone have limited effect.

From a managerial perspective, these findings provide actionable guidance for restaurants seeking to enhance their online reputation. Encouraging customers to upload specific types of photos through incentives or prompts can increase perceived quality, foster trust, and ultimately improve ratings. For review platforms such as Yelp, incorporating photo type alongside quantity into recommendation algorithms could improve the accuracy and relevance of suggestions for users.

Overall, this study highlights the value of user-generated visual content in online reviews and its role in shaping consumer decision-making, providing both practical insights for businesses and a foundation for further research on visual eWOM in the restaurant industry.

9. Limitations and Future Research

While the models identify significant relationships between photo characteristics and Yelp ratings, several limitations remain.The analysis is correlational, not causal, which means restaurants with higher ratings may simply attract more photos. Moreover, unobserved factors such as location, price level, cuisine, and review text sentiment were not controlled for, potentially biasing estimates. Lastly, the linear specification also assumes constant effects across photo volumes, which may overlook diminishing returns. Future research could incorporate elements such as photo quality,or sentiment to better capture how photos truly shape consumer perceptions over time.

References:

Li, C., Kwok, L., Xie, K. L., Liu, J., & Ye, Q. (2021). Let Photos Speak: The Effect of User-Generated Visual Content on Hotel Review Helpfulness. Journal of Hospitality & Tourism Research, 47(4), 109634802110191. https://doi.org/10.1177/10963480211019113

Luca, M. (2016). Reviews, Reputation, and Revenue: The Case of Yelp.com. Harvard Business School NOM Unit Working Paper, 12(016). https://doi.org/10.2139/ssrn.1928601

Parikh, A., Behnke, C., Vorvoreanu, M., Almanza, B., & Nelson, D. (2014). Motives for reading and articulating user-generated restaurant reviews on Yelp.com. Journal of Hospitality and Tourism Technology, 5(2), 160–176. https://doi.org/10.1108/jhtt-04-2013-0011

Wang, Y., Kim, J., & Kim, J. (2021). The financial impact of online customer reviews in the restaurant industry: A moderating effect of brand equity. International Journal of Hospitality Management, 95, 102895. https://doi.org/10.1016/j.ijhm.2021.102895

Weisskopf, D. J.-P. (2018, September 30). Online Customer Reviews: Their Impact on Restaurants. Hospitalityinsights.ehl.edu. https://hospitalityinsights.ehl.edu/online-customer-reviews-restaurants