Drivers of Perceived Rental Property Quality: Analysing shitrentals.org Reviews

540178003

***Abstract*—This study investigates factors influencing tenants' perceived quality of rental properties listed on shitrentals.org, a platform where users can leave anonymous reviews and assign score ratings to the rental properties. Data from 1,000 properties located in 3 suburbs (Redfern, Newtown, Camperdown) of Sydney, Australia with diverse pricing structures were analyzed. All properties were unit/flat type and managed by either agency or private landlord. Ordinal regression models were conducted. The results revealed that the best predictors are weekly price, and specific issues (lease violations, poor condition, pest problems) mentioned in customer reviews. Weekly prices were found to have positive impacts on score ratings, suggesting tenants might view more expensive properties as higher quality. Conversely, negative complaints mentioning specific issues like lease violations, poor condition, pest problems were found to lower the score ratings. Among those specific issues, lease violations were found to be the most significant, followed by pest problems and poor conditions. By understanding the factors influencing those ratings, tenants can make more informed decisions by utilizing observed score ratings. Also, landlords can gain valuable insights to identify areas to improve and potentially increase rental income.** [1]

1. Introduction

Finding a suitable rental property is a top priority for many people. Tenants make decisions based on various factors, including perceived quality, which this study defines as the score ratings (on a 5-point scale) of rental properties. These scores provide a snapshot of how past tenants perceived the property. Understanding factors influencing perceived quality is important for both tenants and landlords.

For tenants:

While score ratings can offer valuable insights, limited research on their underlying factors can be detrimental for tenants to make decisions. By conducting this study, we aim to empower tenants to be able to make better-informed decisions when choosing properties.

For landlords:

A study by Chris Anderson and Saram Han [1] found a significant relationship between customer perceptions and hotel performance. The study also indicated that if hotels get 1 additional score rating on a 5-point scale, they can increase their prices by 11.2 percent while still maintain the same occupancy or market share. Additionally, a study by Christopher D. Ittner and David F. Larcker [2] also suggested that customer satisfaction is a leading indicator of companies' accounting performance (revenues, profit margins, return on sales). Furthermore, a study by Indawati Lestari and Maharani Maharani [3] indicated that online customer ratings have a significant positive association with purchase intention.

Therefore, by understanding factors influencing tenant perceptions, landlords can gain valuable insights that lead to effective targeted property improvements. These improvements can potentially boost score rating, allowing landlords to charge higher prices and gain larger profits while still attracting the same number of tenants.

This paper delves into key factors that drive score ratings for rental properties. We examine various features including weekly rental price, suburb, number of bedrooms, and property management type (private landlord vs agent)

Additionally, we analyzed contents of tenants' reviews, and identified specific themes that potentially influence ratings.

We aim to enable significant benefits for both tenants and landlords. Tenants will be better equipped to understand and utilize rating scores to select properties that suit their needs. Landlords can leverage insights to improve property attractiveness, which ultimately leads to higher profit and success in the rental market.

2. Methods

This study investigates the factors influencing tenants' rating scores for rental properties. Rating scores are discrete ordinal variables ranging from 1 (bad) to 5 (good). Due to the inherent ordering of scores, the ordinal logistic regression model was chosen as the primary statistical method.

Unlike standard logistic regression, ordinal logistic regression takes into account the order of each category. Therefore, it enhances a deeper analysis of how changes in each factor affect the probability of each property getting a higher or lower score.

After we conducted Exploratory Data Analysis (EDA), multiple ordinal logistic regression models were built using combinations of various features, focusing on factors that potentially influence scores.

Price per Week: A strong positive correlation (0.96) between weekly price and rating score was observed. Additionally,

1. I acknowledge the use of ChatGPT [https://chat.openai.com/], Gemini [https://gemini.google.com/] to curate the code, create the themes from words, help suggesting when errors occured, and help suggesting grammatical, words choice improvements in the report.

Figure 1 shows that properties receiving higher score ratings (4-5) are aggregated within the higher price range (around 300-450 AUD per week), while those with lower score ratings (1-2) fall within the lower price range (100-200 AUD per week). Therefore, we infer that weekly price may influence the score of each property, and it was included as a feature in the models.

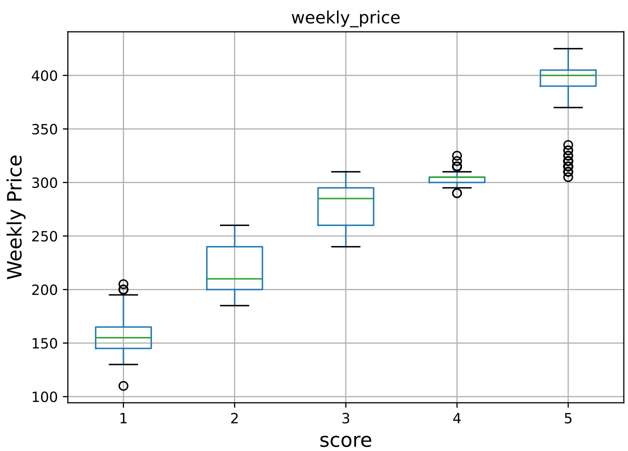

Figure 1. Distribution of Weekly Price by score

Average Price Per week Per bedroom: Considering our dataset includes both one-bedroom and two-bedroom properties, relying on the overall weekly price alone might be misleading. Thus, we calculated the average weekly price per bedroom. However, incorporating this feature into our model did not significantly improve its performance, as being shown in Table 1. Instead, the overall accuracy was higher when using the original weekly price.

TABLE 1. ACCURACY AND F1-SCORE OF VALIDATION DATA FOR VARIOUS FEATURE SETS

Features	**Accuracy**	**F1-Score**
weekly_price	0.868	0.88
Average_weekly_price_per_bedroom	0.308	0.19
weekly_price, suburb	0.868	0.88
Average_weekly_price_per_bedroom , suburb	0.604	0.6

[a]see Appendix Table 7 for more models.

Suburbs: During initial EDA, we found a relationship between suburb and rating score, with Newtown showing the lowest range of score, while Redfern is the highest, as shown in Figure 2. However, further analysis revealed that the score differences between suburbs were linked with their different price variations, as shown in Figure 3. This argument is also supported by correlations between weekly price and suburb variables, showing a correlation of -0.59 with Newtown and 0.73 with Redfern. In addition, as shown in Table 1, trials of including suburbs in the model which already have weekly prices did not lead to better prediction.

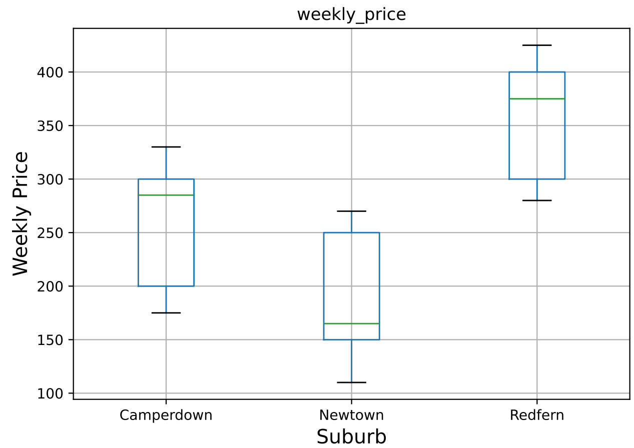

Figure 2. Distribution of Weekly Price by suburb

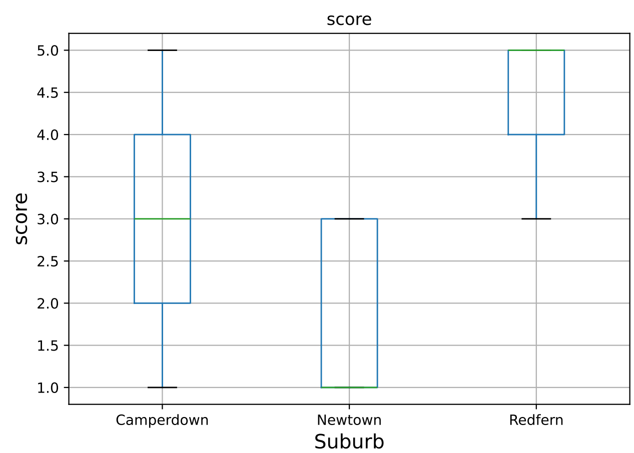

Figure 3. Distribution of Score by suburb

Number of bedrooms, lessor type, agency names: No significant relationships could be found between rating scores and these features.

Review Text Sentiment: A study by Sameh Al-Natour and Ozgur Turetken [6] has shown that sentiment analysis can be utilized in detecting the overall tone of customer perceptions. Therefore, we hypothesized that the sentiment expressed in tenants' reviews (positive or negative) could influence rating scores. To investigate this, we created new features by identifying 250 words with the highest TF-IDF scores in the reviews. (250 accounts for 25 percent of all words) These words were then categorized into positive and negative sentiment categories.

Figure 4 illustrates the distribution of negative review text's rating scores. The frequency of negative sentiment detected is more on lower scores, suggesting a possible connection between negative sentiment and lower ratings.

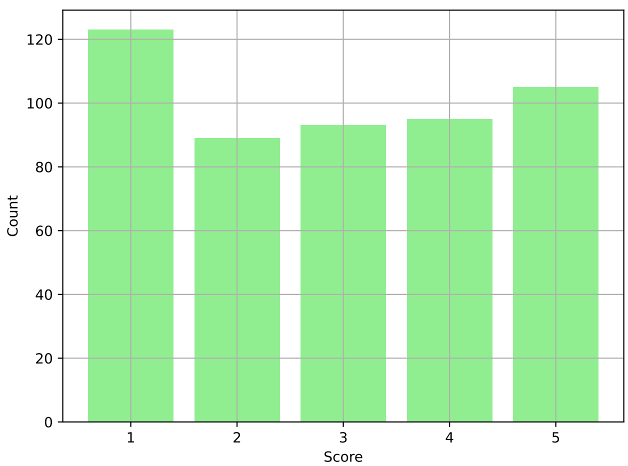

Figure 4. Distribution of Scores for Negative Sentiment

While we did not observe a strong relationship between positive sentiment review and rating scores during EDA (See Appendix, Figure 6), we still have a hypothesis that it may influence scores. Thus, we tried exploring impacts of review's sentiment by conducting several models which incorporated different combinations of features. The models have satisfactory accuracy and F1-score as shown in Table 2.

TABLE 2. ACCURACY AND F1-SCORE OF VALIDATION DATA FOR VARIOUS FEATURE SETS

Features	**Accuracy**	**F1-Score**
positive_sentiment, negative_sentiment, weekly_price	0.864	0.87000
weekly_price, negative_sentiment	0.872	0.88000
weekly_price, positive_sentiment	0.860	0.87000

[a]see Appendix Table 7 for more models.

Compared to the model that includes only weekly price, as shown in Table 1, incorporating negative sentiment improved both the accuracy and F1-score of the model. The results confirm our initial hypothesis and what we found in EDA – negative sentiment tends to have a significant impact on rating scores. For positive sentiment, it did not improve the accuracy and F1-score, indicating that this feature might not significantly influence rating scores.

Review Text Themes: To further improve our analysis, we delved deeper by identifying specific themes from keywords found within review text. We classified keywords found into 5 following negative themes :

1) Pest Problems - reviews mentioning issues related to pests such as cockroaches or rodents.
2) Poor Conditions - reviews mentioning problems related to property's physical state.
3) Lease Violation - review having keywords indicating breaches of the rental agreement.
4) Maintenance Issues - review mentioning issues related repairing responsibilities that are not met.
5) Lack of Amenities - review having keywords of absences or inadequacy of expected features.

EDA revealed that certain themes show a pattern with rating scores. As shown in Figure 5 , customers who mentioned pest problems, poor conditions, maintenance issues are more likely to give lower scores.

For lease violations, although we cannot observe significant trends with score, this theme captures the largest amount of text reviews, signifying frequent concerns among customers. For lack of amenities, the captured data is minimal and no clear patterns can be observed. It is important to note that we are capturing review text based on specified keywords; thus, bar charts can only tell frequency of words mentioned, not impacts or importance of each word on scores.

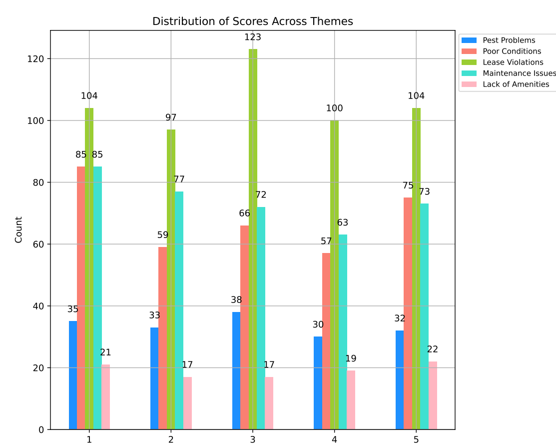

Figure 5. Distribution of keyword themes mentioned in review text

From seeing some patterns in specific themes of negative keywords, we have a hypothesis that specific themes of positive keywords might have patterns as well. Thus, we conducted positive themes as below:

1) Good location - reviews mentioning words related to convenience and positive descriptors of locations
2) Good security - reviews mentioning words related to safety, peacefulness, and general positive feedback.

As shown in Table 3, Models incorporating some specific theme features, along with weekly price, led to better results compared to those using only negative sentiment, as seen in Table 2.

The model with the highest accuracy and F1-score for the validation data (0.888 and 0.89572, respectively) includes the features of lease violations, poor condition, pest problems, and weekly price.

Adding positive theme features of good security and good location to this model did not change the accuracy or F1-score. Furthermore, the P-value for good location is 0.448 while for good security is 0.126, both of which are relatively high. Additionally, the P-values for other features in the model remained unchanged, suggesting those positive theme features might not be as influential as expected.

Therefore, we choose the final model, comprising the features of weekly price, lease violations, poor conditions, and pest problems.

TABLE 3. ACCURACY AND F1-SCORE OF VALIDATION DATA FOR VARIOUS FEATURE SETS

Features	Accuracy	F1-Score
weekly_price, pest_problems, poor_condition, lease_violation	0.876	0.88403
weekly_price, pest_problems, lease_violation, lack_of_amenities	0.872	0.8076
weekly_price, poor_condition, lease_violation, lack_of_amenities	0.876	0.88403
weekly_price, maintenance_issues, poor_condition, pest_problems, lease_violation	0.888	0.89572
weekly_price, maintenance_issues, pest_problems, lease_violation	0.876	0.88463
weekly_price, maintenance_issues, poor_condition, pest_problems, lease_violation, lack_of_amenities	0.88	0.88802
weekly_price, maintenance_issues, lease_violation, lack_of_amenities	0.868	0.87857
weekly_price, maintenance_issues, poor_condition, lack_of_amenities	0.872	0.88142
weekly_price, good_security, good_location, pest_problems, poor_condition, maintenance_issues	0.888	0.89572
weekly_price, poor_condition, lease_violation, pest_problems, lack_of_amenities	0.876	0.88403
weekly_price, poor_condition, maintenance_issues , lease_violation, pest_problems, lack_of_amenities	0.88	0.88804

[a]see Appendix Table 7 for more models.

3. Results and Discussion

After retraining the chosen model, the accuracy and F1-score of test datasets went up to 0.9 and 0.90299, respectively.

High accuracy in test datasets suggests that a model can handle unseen data quite well, with estimated errors of 10 percent.

A high F1-score indicates a high performance of the model, being able to balance a good value of precision and recall metrics.

A classification matrix, as shown in Table 4, was created to visualize where the model performs poorly. By analyzing it, we can better pinpoint which score predictions need improvement.

Overall, the confusion matrix shows that our model performs well, with most misclassifications occurring in adjacent scores. The model performs the best on predicting 5-score rating, while for other scores, there are still rooms for improvement. Our final model includes weekly price, lease violations, poor condition, and pest problems as features. As shown in Table 5. Weekly price is the only feature with a positive coefficient, indicating higher rental prices are associated with better scores, while the remaining negative feature themes have negative coefficients, suggesting negative associations with scores.

TABLE 4. CONFUSION MATRIX OF FINAL MODEL'S TEST SET

	Predicted Score				
True Score	1	2	3	4	5
1	55	6	0	0	0
2	2	41	5	0	0
3	0	1	47	7	0
4	0	0	4	41	0
5	0	0	0	0	41

The most significant features are weekly price and lease violations, evidenced by their lowest p-values (approximately 0), followed by pest problems (0.034) and poor condition (0.216). We can interpret that pricing and lease violation issues might have a stronger influence on scores, while poor condition issues have the weakest effect.

Most of our findings align with the results from the EDA: weekly price is highly correlated with scores, and certain themes' keywords in reviews are associated with lower scores. Interestingly, while the impact of lease violations was not clear during the EDA, the final model revealed it to have a strong influencing factor on scores.

TABLE 5. SUMMARY OF FINAL MODEL

Features	coefficient	std err	z	P> |z|
weekly_price	0.3244	0.027	11.870	0.000
lease_violations	-1.0237	0.279	-3.673	0.000
poor_condition	-0.3483	0.282	-1.237	0.216
pest_problems	-0.7394	0.350	-2.114	0.034

To gain better understanding of each feature's impact on scores, we computed odd ratios of each feature as shown in Table 6.

TABLE 6. ODDS RATIOS OF FINAL MODEL

Features	Odds Ratios
Weekly Price	1.383169
Lease Violations	0.359259
Poor Condition	0.705891
Pest Problems	0.477399

We can interpret odd ratios as follows:

- When weekly price increases by one unit, odds of getting a higher score increase by approximately 38.32 percent.
- When lease violations are present, odds of getting a higher score decrease by approximately 64.07 percent.
- When poor conditions are present, odds of getting a higher score decrease by approximately 29.41 percent.
- Pest Problems: When pest problems are present, odds of getting a higher score decrease by approximately 52.26 percent.

4. Conclusion and Limitations

From our final model, we can conclude that weekly prices positively influence a property's rating score, inferring that customers may perceive higher quality in properties with higher prices. Conversely, the presence of issues like lease violations, pest problems, and poor conditions negatively impact the scores, indicating that to increase customer satisfaction, improvements in these areas might be needed.

Lease violation mentioning was found to have a significant negative association with scores. Its presence leads to a decrease in the odds of higher scores by 64.07 percent, highlighting that customers might view lease issues as highly important factors. This finding underscores that landlords may need to prioritize improving this issue when aiming to enhance customer rating scores.

To sum up, the study highlights factors that tenants view as important and can influence their perceived property quality. By addressing these factors, landlords can efficiently improve the property's quality, leading to better customer satisfaction and rating scores. By understanding this, tenants can make better decisions in choosing property as well.

The study still has limitations. First, only a small portion of data is captured by some review text themes, such as lack of amenities. This leads to restricted abilities of models to comprehensively capture the impact of those keywords. Improvements could be made by expanding keyword lists.

Additionally, there were duplicated review texts, which occurred as the data were generated through simulation. The instructor noted that this redundancy resulted from limitations in data generation. However, if these were real datasets, immediate investigations about the data accuracy would be necessary, as review texts from different customers should not be completely identical.

Another limitation is ordinal regression model that we used, has a proportional odds assumption, meaning it assumes the effect of each independent variable on the log-odds of moving between any two score categories is constant. This might not always be true. For instance, the impact of a specific issue existence on a tenant's decision between a score of 2 vs 1 might differ significantly from its influence on the decision between 5 vs 4. For future improvements, the appropriate tests for this assumption [5] should be incorporated.

Finally, potential data bias exists, as datasets were restricted to only 3 suburbs in Australia and only one type of property (unit/flat). It is uncertain whether the customers' perceptions are similar for other suburbs or other property types. It's important to note that shitrentals.org possesses wider rages of property types and suburbs. Future research could improve this limitation by utilizing larger and more diverse datasets. This could involve expanding data collection geographically and across different property types.

References

[1] Chris Anderson & Saram Han. (2016). Hotel Performance Impact of Socially Engaging with Consumers. Cornell Hospitality Report, 16. https://sha.cornell.edu/faculty-research/centers-institutes/chr/research-publications/hotel-performance-impact-socially-engaging-with-consumers

[2] Christopher D. Ittner & David F. Larcker. (1998). Are Nonfinancial Measures Leading Indicators of Financial Performance? An Analysis of Customer Satisfaction. Journal of Accounting Research, 36. https://doi.org/10.2307/2491304

[3] Indawati Lestari & Maharani Maharani. (2023). Investigating the Effect of Customer Reviews and Online Customer Ratings on Purchase Intention: Mediating Role of Word of Mouth. International Journal of Finance Economics and Business, 2, 313–321. https://doi.org/10.56225/ijfeb.v2i4.285

[4] Learn LaTeX in 30 minutes. (n.d.). Overleaf. https://www.overleaf.com/learn/latex/Learn_LaTeX_in _30_minutes

[5] Liu, A., He, H., Tu, X. M., & Tang, W. (2023). On testing proportional odds assumptions for proportional odds models. General psychiatry, 36(3), e101048. https://doi.org/10.1136/gpsych-2023-101048

[6] Sameh Al-Natour & Ozgur Turetken. (2020). A comparative assessment of sentiment analysis and star ratings for consumer reviews. International Journal of Information Management, 54. https://doi.org/10.1016/j.ijinfomgt.2020.102132

[7] Learn LaTeX in 30 minutes. (n.d.). Overleaf. https://www.overleaf.com/learn/latex/Learn_LaTeX_in _30_minutes

[8] Shitrentals. (n.d.). https://www.shitrentals.org/

Appendix

1. Figures and Tables

TABLE 7. ACCURACY AND F1-SCORE OF VALIDATION DATA FOR VARIOUS FEATURE SETS

Features	Accuracy	F1-Score
weekly_price	0.868	0.88000
suburb	0.396	0.31000
weekly_price, suburb	0.868	0.88000
Average_weekly_price_per_bedroom	0.308	0.19
Average_weekly_price_per_bedroom , suburb	0.604	0.6
weekly_price, 1, 2	0.864	0.87000
weekly_price, 2	0.872	0.88000
weekly_price, 1	0.86	0.87000
1, 2, Average_weekly_price_per_bedroom, suburb	0.608	0.60000
weekly_price, 1, 2, n_bedrooms	0.864	0.87000
weekly_price, 1, 2, suburb	0.864	0.87000
weekly_price, 1, 4, 6, 5, 3	0.86	0.86841
weekly_price, 1, 4, 6, 5, 3, 7	0.868	0.87637
weekly_price, 3, 4, 5	0.876	0.88403
weekly_price, 3, 4, 7	0.86	0.86841
weekly_price, 3, 5, 7	0.872	0.8076
weekly_price, 4, 5, 7	0.876	0.88403
weekly_price, 6, 4, 3	0.864	0.87340
weekly_price, 6, 4, 3, 5	0.888	0.89572
weekly_price, 6, 3, 5	0.876	0.88463
weekly_price, 6, 4, 3, 5, 7	0.88	0.88802
weekly_price, 6, 5, 7	0.868	0.87857
weekly_price, 6, 4, 7	0.872	0.88142
weekly_price, 8, 9, 2	0.864	0.87374
weekly_price, 8, 9, 3, 4, 6	0.888	0.89572
weekly_price, 4, 5, 3, 7	0.876	0.88403
weekly_price, 4, 6, 5, 3, 7	0.88	0.88804

[a] 1-9 are theme words features as explained in section 2.

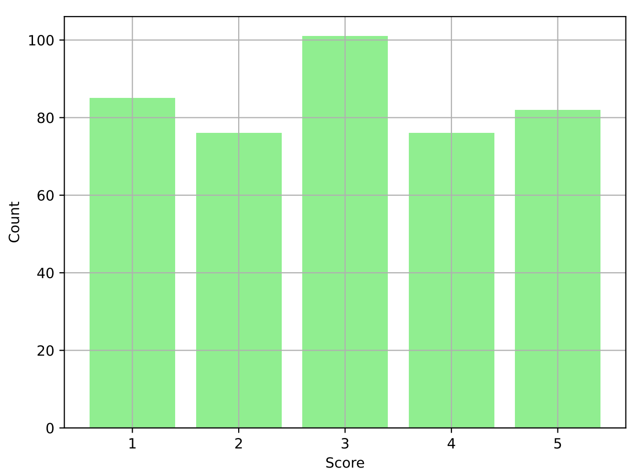

Figure 6. Distribution of Scores for Positive Sentiment

2. Acronyms

In Table 7, we use numbers as follows to represent theme words features

1) Positive sentiment
2) Negative sentiment
3) Pest Problems
4) Poor Conditions
5) Lease Violation
6) Maintenance Issues
7) Lack of Amenities
8) Good Location
9) Good security