

Unveiling the Links: TripAdvisor Ratings and Extrinsic Factors on Vacation Rentals in British Columbia, Canada - Analysis Report

Audrey Nkrumah, Anna Brabender, Saiyara Islam, and Fernando Lopez Contla

The purpose of this project was to explore and analyze data from TripAdvisor and Geoapify APIs to examine the impact that nearby businesses and attractions have on a rental's rating. More specifically, we aimed to uncover any discernible relationship between a property's rating and its proximity to external locations, including tourist landmarks, activity centers, entertainment spots, water spots, and spa locations. Our hypotheses for the expected relationships were: If a vacation rental receives a high number of ratings, then the average rating of the rental is expected to be higher. Similarly, we expected similar relationships to be observed as the number of entertainment spots, spa spots, activity centres, water bodies, and tourist attractions as they increase around rental property.

During our research stage, we read various tourism journal articles exploring factors affecting the satisfaction of tourists. According to Chu et al., they found that reviews were a helpful tool for tourists to judge experiential goods since these allow for two-way information exchange (2022). They also found that reviews often mentioned parking accessibility, and it inspired us to explore whether public transport might influence a location's rating. Similarly, Van der Zee & Bertocchi's findings demonstrated that the closure of a tourist landmark had a detrimental effect on attendance to nearby businesses. Taking this into consideration we decided to take a step further and explore if the proximity of vacation rentals to tourist landmarks had any impact on their rating.

The project was conducted using Python in a Jupyter Notebook environment. The primary libraries utilized included Pandas for data manipulation, Matplotlib and Seaborn for data visualization, and requests for API interactions. The TripAdvisor and Geoapify APIs were used to gather relevant data. Data from TripAdvisor was used to extract information about rental properties in British Columbia. We ultimately cleaned this data and standardized the data to include the property name, location, rating, and other relevant information such as Longitude and Latitude coordinates. We removed any rental locations with zero reviews or if there were duplicate locations with the same property name. We then used these coordinates and the Geoapify API to search for external businesses near each rental. The businesses we searched for included activity centres (sports clubs and community centers), entertainment spots (cinemas, zoos, museums, etc.), natural water bodies (rivers, lakes, springs, etc.), spa locations, and tourist landmarks (religious temples, city hall, bridges, etc.).

. Preliminary data about the mean values of the external locations surrounding our top 20 rentals demonstrated that tourism landmarks had the highest mean, followed by water bodies and activity centres. Lastly, public transit and spa spots had the lowest mean values. In order to confirm the proximity of these locations impacted rental ratings. We then created individual scatter plots to explore the relationship between the total number of businesses belonging to any of the external location genres and their impact on the property's rating. We can observe the summarised findings in Table 1.

Table 1: Summarised analysis for each of the examined relationships.

Relationship	R-values	P-values	Hypothesis Accepted
Average rating versus review count	0.0176	0.1932	Null
Average rating versus number of entertainment spots	0.0003	0.8455	Null
Average rating versus number of spa spots	0.0695	0.0087	Null
Average rating versus number of activity centres	0.0082	0.3743	Null
Average rating versus number of tourism spots	0.0021	0.6502	Null
Average rating versus number of water spots	0.0003	0.8765	Null

Based on these findings there was a strong implication that the proximity to tourist sites, activity centres, entertainment spots, water spots, and spa locations does not influence a rental's rating. However, the rentals were spread across different geographical locations within British Columbia, including urban, suburban, and rural areas. Most of the rentals were concentrated around Vancouver Island and Vancouver. Given that the further locations in rural areas might have been too far apart, we needed to account for sampling bias in our data. Based on Van der Zee & Bertocchi's research it's possible that the lack of tourist landmarks could have had a detrimental effect on the number of reviews for our rural locations (2018). This would explain

why our data appears skewed in favour of rentals located near tourist landmarks. Hernández et al. found that in relation to tourist landmarks, geographic proximity between property rentals and tourist landmarks aren't always close together (2018). This provides an alternate explanation to the distribution of our data and aligns with our interpretation of the findings suggesting that landmarks might not be very influential to a tourist's decision to stay there.

Sampling bias is possibly the most harmful factor for our correlation analysis. We should have taken it into consideration since only a select group of tourists or locals such as influencers and bloggers might be the ones contributing to the reviews of a location. As a result, these locations form a review network, they are likely to attract consecutive visits and reviews leading to skewed perceived popularity (Van der Zee & Bertocchi, 2018). There might be more factors that limited the significance of our findings. Such as the audience of each reviewing database. Some scholars found most reviews posted on TripAdvisor are positive; hence, the rating distribution is skewed towards the positive end of the scale (Filieri et al., 2020). Taking these flaws into consideration, for any future studies we plan to include more databases such as Google Reviews and Yelp. We also wanted to change the external locations to include more modern services such as glamping, sports events, and medical tourism (Sangkaew & Zhu, 2020).

In conclusion, our project successfully mined data from TripAdvisor and Geoapify APIs to analyze the relationship between rental ratings in British Columbia and their proximity to external locations. While some weak correlations were observed, further analysis is needed to establish causation and identify other factors influencing ratings. This led us to reject our hypotheses and accept the null hypotheses for all the questions we aimed to explore in this analysis. This project provides a foundation for future investigations into the factors contributing to the success of rental properties in this region.

References

- Çakırtaş, M., & Ozdemir, M. K. (Eds.). (2021). Big Data and Social Media Analytics. *Lecture Notes in Social Networks*. doi:10.1007/978-3-030-67044-3
- Chu, M., Chen, Y., Yang, L., & Wang, J. (2022). Language interpretation in travel guidance platform: Text mining and sentiment analysis of TripAdvisor reviews. *Frontiers in psychology*, 13, 1029945. <https://doi.org/10.3389/fpsyg.2022.1029945>
- Filieri, R., Acikgoz, F., Ndou, V., & Dwivedi, Y. (2020). Is TripAdvisor still relevant? the influence of review credibility, review usefulness, and ease of use on consumers' continuance intention. *International Journal of Contemporary Hospitality Management*. doi:10.1108/IJCHM-05-2020-0402
- Hernández, J. M., Kirilenko, A. P., & Stepchenkova, S. (2018). Network approach to tourist segmentation via user generated content. *Annals of Tourism Research*, 73, 35–47. doi:10.1016/j.annals.2018.09.002
- Sangkaew, N., & Zhu, H. (2020). Understanding Tourists' Experiences at Local Markets in Phuket: An Analysis of TripAdvisor Reviews. *Journal of Quality Assurance in Hospitality & Tourism*, 1–26. doi:10.1080/1528008x.2020.1848747
- Van der Zee, E., & Bertocchi, D. (2018). Finding patterns in urban tourist behaviour: a social network analysis approach based on TripAdvisor reviews. *Information Technology & Tourism*. doi:10.1007/s40558-018-0128-5
- .