STATISTICAL ANALYSIS OF FACTORS IMPACTING HOTEL REVIEW SCORES IN THE LAS VEGAS STRIP

Ian Johnston and Peter Emanuelli

Sponsor: Dr. Victoria Woodard

University of Notre Dame

Introduction

Since Las Vegas is a large hub for gambling and travel that welcomes more than 41 million visitors a year, we conducted a study to look at the factors which would best help to predict TripAdvisor ratings for hotels on the Las Vegas Strip. A high hotel rating would likely translate to higher foot traffic in the corresponding hotel and thus a higher income.

Why Is This Important?

- Resource allocation
- Advertising
- Building Plans

Project Questions:

- What makes some of these hotels more popular than others?
- Is there a way to predict how much a guest will enjoy a hotel?

Previous Research

Previous research on this topic by Coehlo, Moro, and Rita focused more on the visitors' demographics themselves rather than the actual hotels, and other factors such as the day of the week of the review. This led to a discussion about how hotel management can best respond to customers on sites like TripAdvisor to improve their likelihood of receiving a positive review score without much discussion on improvements to the hotels themselves.

Moro, S., Rita, P., & Coelho, J. (2017). Stripping customers' feedback on hotels through data mining: The case of Las Vegas Strip. Tourism Management Perspectives, 23, 41-52.

Methodology

The dataset used for this logistic regression model is a collection of hotel review scores from TripAdvisor's website, and information about each of the hotels on the strip. All reviews were regarding stays in the year 2015. 24 reviews were randomly selected for each of the 21 hotels, two for each month of the year, giving a sample size of 504.

The response variables were review scores, which were split into "unfavorable" (1-3 stars) and "favorable" (4-5 stars). The independent variables included but were not limited to: many of binary amenity indicator variables, number of helpful votes given to a review by other TripAdvisor users, type of travelers (friends, business, couples, etc.).

Project Steps

Initial model with all hotel characteristics

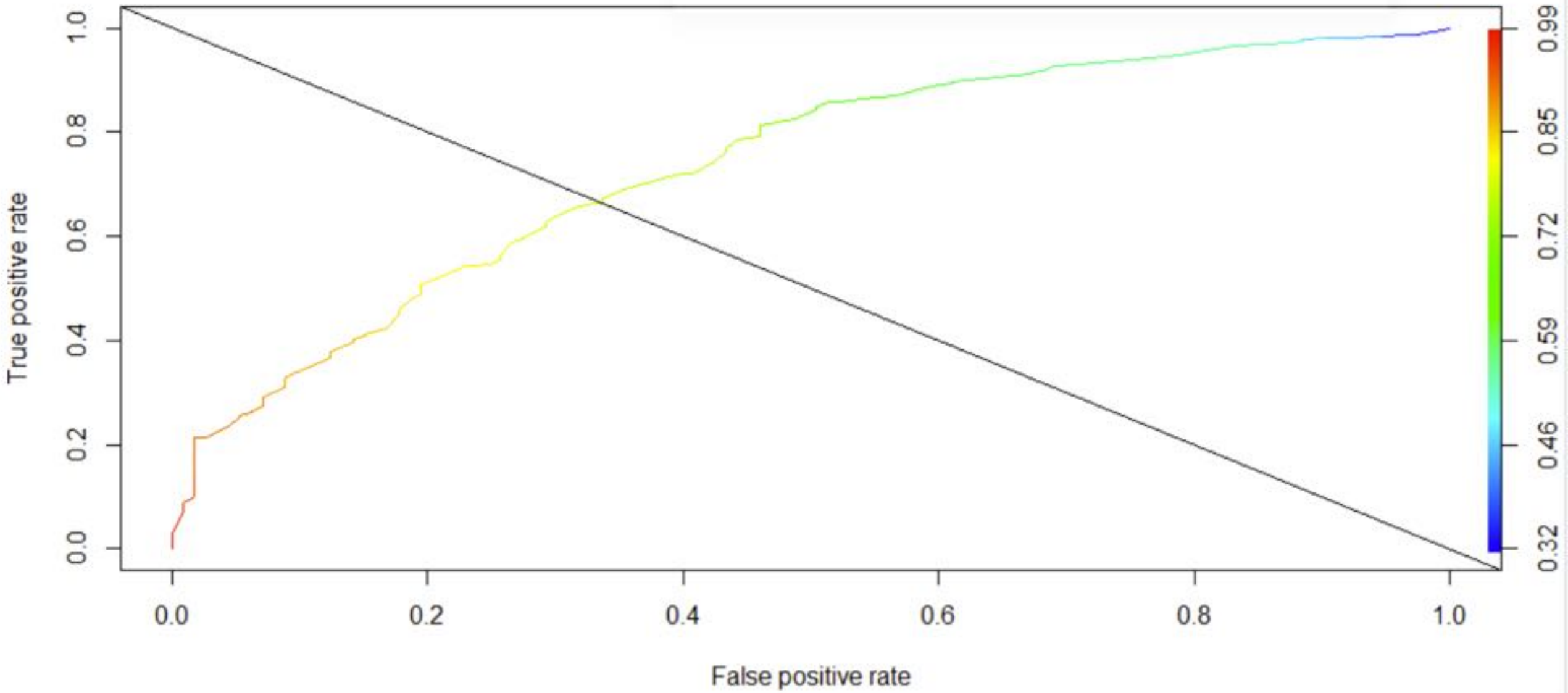
AIC selection to remove variables with high p values

Hosmer-Lemeshow Test shows no evidence reduced model was an inadequate fit

ROC curve to find optimal cutpoint of roughly 0.75

$$ln\left(\frac{\pi(favorable)}{1-\pi(favorable)}\right) = -2.000 + 1.109(Pool) + 1.302(Free\ Internet) + 1.162\ (3.5\ stars) + 0.349\ (4\ stars) + 0.186(4.5\ stars)$$
$$+ 0.869\ (5\ stars) + 0.763(Couple) - 0.033(Families) + 0.978(Friends) + 0.577(Solo) - 0.382(Casear's\ Palace)$$
$$+ 0.703(Encore) - 0.656(Excalibur) - 0.051(Flamingo) - 0.275(Boulevard) + 1.27(Marriott\ Grand\ Chateau)$$
$$- 0.343(Paris) - 0.179(The\ Cosmopolitan) - 0.087(The\ Palazzo) + 1.347(The\ Venetian) - 0.341(The\ Westin)$$
$$- 0.188(Treasure\ Island) + 0.352(Trump\ International) + 0.559(Wynn)$$

Figure 1 (ROC Curve):



Results, Discussion, and Next Steps

- *Biggest Positive Review variables:*
  - Pool
  - Free Internet
  - 3rd Party 5 Star Ratings
  - Hotel Names (Hypothesized as Reputation)
  - Guests Trip Classified as Friends
- *Model Limitations*
  - Binary Response Variable
  - Potential Correlation Between 3rd Party Star Ratings and Amenities
  - Unquantifiable Hotel Reputation
- *Next Steps/Future Research*
  - Use of Categorical Response Variables (Potentially in a Proportional Odds Model)
  - Looking at Other Locations
  - Inclusion of Financial Data

Model Summary

On average, TripAdvisor reviews for hotels on the Vegas Strip were positively influenced by facts about the hotel like the star rating, the existence of pools and free internet, and hotel name (and likely the name recognition that comes with it). The reviews were also influenced by the type of traveler that stayed at the hotel with solo, couple, and friends being the types that, on average, resulted in higher review scores.

Overall, the data suggests that hotel owners on the Vegas Strip should look into focusing on the areas listed above in order to give themselves the best foot forward to achieve more favorable ratings for their hotel, which will likely correspond to more success for the hotel.

References

Downtown Vegas Alliance. “Fun Facts.” https://downtown.vegas/visitors-guide/fun-facts/

Moro, S., Rita, P., & Coelho, J. (2017). Stripping customers' feedback on hotels through data mining: The case of Las Vegas Strip. Tourism Management Perspectives, 23, 41-52. https://novaresearch.unl.pt/en/publications/stripping-customers-feedback-on-hotels-through-data-mining-the-ca

Moro, S., Rita, P. “Las Vegas Strip Dataset.” (2017). University of California Irvine Machine Learning Repository. http://archive.ics.uci.edu/ml/datasets/Las+Vegas+Strip#