# Hotel Analytics Report

## 1. Introduction and Project Goal

The goal of this project is to analyze hotel review for hotels in Las Vegas and to extract insights into hotel performance, customer behavior, and review patterns. The analysis focuses on identifying high- and low-performing hotels, understanding how traveler characteristics influence ratings, exploring relationships between review variables, and predicting hotel review scores using a simple linear regression model.

The dataset contains user reviews with information about reviewer location, traveler type, hotel characteristics, amenities, and ratings. The project combines descriptive analytics, visual data storytelling, relationship exploration, and predictive modeling.

## 2. Dataset Overview

The dataset includes reviews from multiple continents, with most reviewers coming from North America (294 reviews) and Europe (118 reviews). Smaller contributions come from Oceania, Asia, Africa, and South America.

Average review scores are relatively high across all continents, ranging from approximately 3.4 to 4.4, indicating generally positive hotel experiences. North America and Europe dominate the dataset, making their averages more reliable, while smaller continents (e.g., South America, Africa) should be interpreted with caution due to low sample size.

Most reviews were written by:

- Couples (dominant traveler type),
- during March–May,
- mostly on Tuesdays, Wednesdays, and Sundays.

## 3. Hotel Performance Insights

**Top Performing Hotels**

The best-performing hotels based on average score and review volume include:

- Wynn Las Vegas
- The Venetian Las Vegas Hotel
- Marriott’s Grand Chateau
- Encore at Wynn Las Vegas
- Wyndham Grand Desert

These hotels consistently achieve average scores above 4.4, indicating high guest satisfaction.

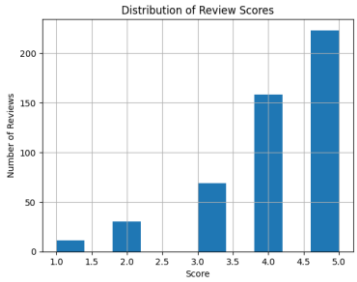
**European Reviewer Perspective**

When focusing only on European reviewers, similar luxury hotels appear at the top, confirming consistent quality perception across regions.

**Low Performance Despite Full Amenities**

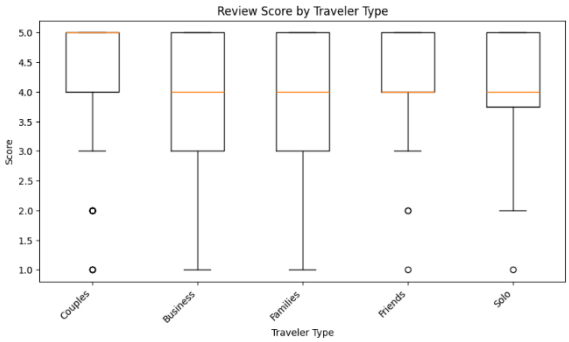
Some hotels offering all major amenities (spa, gym, casino, tennis court) still show comparatively lower average scores. This suggests that amenities alone do not guarantee high ratings, and service quality or guest expectations play an important role.

# 4. Visual Data Storytelling




### Distribution of Review Scores

The histogram shows a strong concentration of review Scores at 4 and 5, indicating that most guests rate their hotel experience positively. Low ratings (1–2) occur rarely, suggesting that strongly negative experiences are uncommon in the dataset.




### Scores by Traveler Type

The boxplot reveals that all traveler types rate hotels similarly, with median Scores close to 4 across all groups. Couples and friends show slightly more consistent ratings, while business and family travelers display greater variability, indicating more mixed experiences.

# 5. Relationship Exploration

### Member Years vs. Helpful Votes

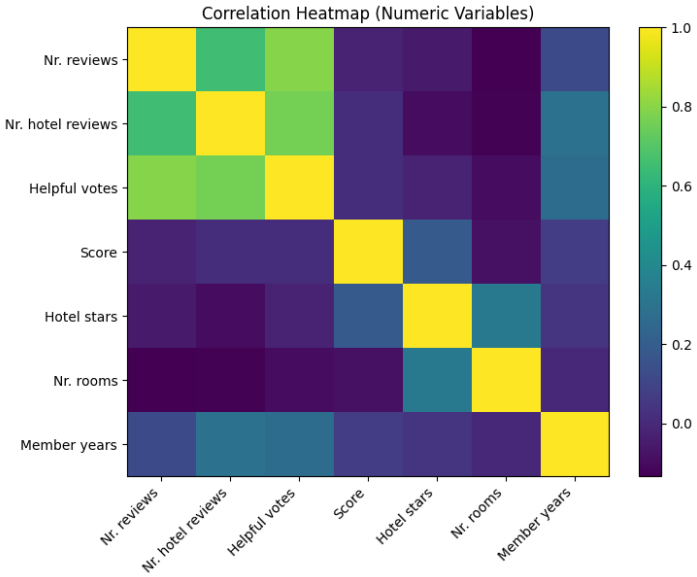
The analysis shows no strong relationship between years of membership and helpful votes. Experienced members do not automatically receive more helpful votes, suggesting that content quality matters more than seniority.

### Casino vs. Score

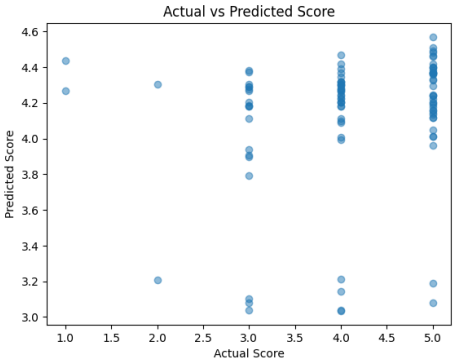
Hotels with a casino have slightly higher average scores, but the difference is small. Casinos may contribute to perceived value, but they are not the main driver of satisfaction.

### Correlation Heatmap (Numeric Variables)

The correlation heatmap shows strong relationships between number of reviews, hotel reviews, and helpful votes, indicating that more active reviewers tend to receive more helpful votes. In contrast, the Score variable shows very weak correlations with most numeric features, including hotel size and member years.



# 6. Hotel Score Prediction




A linear regression model was trained using hotel stars, number of rooms, member years, helpful votes and amenity flags (pool, gym, spa, casino, free internet, tennis court).

The model was evaluated on a test set.

Model Performance

- $R^2 \approx 0$ (slightly negative)
- MSE relatively high

This means the model cannot meaningfully explain the variation in scores. The prediction plot confirms this: predicted values are clustered and do not follow the diagonal well.

**Variable Importance**

Amenities such as free internet and pool show the strongest positive coefficients, while spa and room count show weak or negative effects. However, because the model performance is poor, these coefficients should be interpreted cautiously.

**Interpretation**

Hotel review scores are likely influenced by factors not present in the dataset, such as:

- text sentiment of reviews,
- service quality,
- price,
- location,
- expectations of travelers.

| | feature | coefficient |
|---|---|---|
| 0 | Free internet | 0.993850 |
| 1 | Pool | 0.981986 |
| 2 | Gym | 0.297263 |
| 3 | Hotel stars | 0.184756 |
| 4 | Tennis court | 0.147143 |
| 5 | Member years | 0.064103 |
| 6 | Casino | 0.009401 |
| 7 | Helpful votes | -0.002153 |
| 8 | Nr. rooms | -0.009963 |
| 9 | Spa | -0.203366 |

# 7. Conclusion

This project successfully combined descriptive analysis, visualization, relationship exploration, and prediction into one structured analytics pipeline. Key findings include:

- Review scores are generally high and skewed toward positive ratings.
- Top hotels are consistent across continents.
- Amenities alone do not guarantee high scores.
- Reviewer experience does not strongly influence helpful votes.
- Linear regression cannot effectively predict review scores with the available features.

The results highlight the limitations of structured data for modeling customer satisfaction.