

# A Novel Approach to Hotel Rating with Percentile Normalization and Customized Penalty Algorithm

Arun N. Nair  
Strata Digital  
stratadigital.com  
arunn@stratadigital.com

October 27, 2024

## Abstract

Selecting a hotel that meets specific quality expectations can be challenging due to variations in rating scales, reviewer biases, and regional differences. This paper presents a methodology for normalizing hotel ratings using percentile rankings and custom penalty functions. By integrating a no-penalty zone and adjusting the sharpness of penalties, the proposed approach enables travelers to prioritize top-performing hotels while considering a broader range of accommodations that meet their preferences.

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Challenges in Hotel Rating Systems</b>	<b>3</b>
<b>3</b>	<b>Methodology</b>	<b>4</b>
3.1	Percentile Rankings . . . . .	4
3.2	Weighted Normalized Ratings . . . . .	4
3.3	Custom Penalty Functions . . . . .	5
3.3.1	No Penalty Zone . . . . .	5
3.3.2	Adjusting Penalty Sharpness . . . . .	5
3.3.3	Piecewise Penalty Functions . . . . .	6
<b>4</b>	<b>Implementation</b>	<b>6</b>
4.1	Data Collection . . . . .	6
4.2	Parameter Selection . . . . .	6
4.3	Calculation Steps . . . . .	7

<b>5</b>	<b>Case Study</b>	<b>7</b>
5.1	Hotel Data . . . . .	7
5.2	Applying the Methodology . . . . .	8
5.3	Results and Analysis . . . . .	9
<b>6</b>	<b>Benefits and Applications</b>	<b>10</b>
<b>7</b>	<b>Conclusion</b>	<b>11</b>
<b>8</b>	<b>Future Work</b>	<b>11</b>
<b>A</b>	<b>Penalty Function Graph</b>	<b>11</b>
<b>B</b>	<b>Parameter Adjustment Guidelines</b>	<b>12</b>

# 1 Introduction

In the age of global travel, selecting suitable accommodations is an important aspect of trip planning. Travelers often rely on online ratings and reviews to make informed decisions. However, disparities in rating scales across platforms and regions can lead to confusion and misinterpretation of what constitutes a "good" hotel. Additionally, reviewer biases, cultural differences, and the recency of reviews can affect the accuracy of these ratings. This paper introduces a systematic approach to normalizing hotel ratings using percentile rankings and customizable penalty functions, enabling accurate, consistent and personalized hotel selections.

The primary objective of this methodology is to provide a comprehensive framework that travelers can use to identify suitable hotels without being overwhelmed by the inconsistencies present in the existing rating systems. By employing percentile rankings, the methodology aims to provide contextual insights, making hotel ratings comparable across various platforms and regions. Moreover, the use of customizable penalty functions ensures that travelers can adjust the importance of hotel rankings based on their personal preferences.

## 2 Challenges in Hotel Rating Systems

There are several challenges that make hotel rating systems complex and often unreliable for travelers. Some of the major challenges are highlighted below:

- **Variability in Rating Scales:** Different booking platforms and regions may use varying rating scales (e.g., 1-5 stars, 1-10 points), making direct comparisons difficult. A hotel rated 4 stars on one platform may not necessarily match a 4-star rating on another platform.
- **Reviewer Bias:** Cultural differences and personal expectations can influence reviewers' perceptions and ratings. For example, travelers from different parts of the world may have varying expectations regarding amenities, cleanliness, and service quality, which can lead to biases in their reviews.
- **Recency of Reviews:** Hotels may improve or decline over time, and older reviews may not reflect the current state of the hotel. Recent renovations, changes in management, or other improvements may not be captured if older reviews dominate the overall rating.
- **Lack of Context:** Absolute ratings do not indicate how a hotel performs relative to others in the same area. A rating of 4 stars may be excellent in a rural area but below average in a major metropolitan city. Without context, travelers may struggle to understand the relative quality of a hotel.
- **Inconsistent Standards:** Different platforms may have different standards for assigning ratings, which makes it challenging to compare ratings across multiple booking websites. A 5-star rating on one platform may not meet the same criteria as a 5-star rating on another.

These challenges necessitate a more standardized approach to hotel selection that takes into account the relative performance of hotels within a specific region and adjusts for biases and inconsistencies.

### 3 Methodology

The proposed methodology consists of several key components that work together to normalize hotel ratings and provide a consistent framework for hotel selection.

#### 3.1 Percentile Rankings

Percentile rankings position a hotel’s rating relative to its peers in the same region, providing context beyond absolute scores. This allows travelers to understand how well a hotel performs compared to other hotels in the same area.

- **Definition:** The percentile rank indicates the percentage of hotels that a particular hotel scores higher than. For example, a hotel with a percentile rank of 90% is better rated than 90% of the hotels in the same region.

- **Calculation:**

$$\text{Percentile Rank of Hotel } i = \frac{\text{Number of Hotels with Lower Ratings than } i}{\text{Total Number of Hotels}} \times 100\% \tag{1}$$

- **Advantages of Percentile Rankings:** Percentile rankings provide context by showing how a hotel compares to others in the same region. This helps travelers make more informed decisions by understanding the relative quality of a hotel rather than relying solely on an absolute rating.

#### 3.2 Weighted Normalized Ratings

Adjust ratings based on factors such as reviewer origin and recency to reflect the user’s preferences. By weighting reviews according to the user’s preferences, the methodology ensures that the final rating is tailored to individual needs.

- **Factors:**

- **International Reviews:** Assign higher weights to reviews from international travelers. For example, if a hotel is in India, and the traveler is from the United States, reviews from other international travelers may carry more weight than reviews from India. This is simply because international travelers often have similar expectations, making their feedback more aligned with what the traveler is seeking.
- **Recency:** Give more weight to recent reviews to capture the current quality of the hotel. Hotels can change significantly over time, and recent reviews are often more reflective of the current state of the property.

- **Weighted Rating Calculation:**

$$\text{Weighted Rating} = \frac{\sum(\text{Rating} \times w_{\text{international}} \times w_{\text{recent}})}{\sum(w_{\text{international}} \times w_{\text{recent}})} \quad (2)$$

- **Normalization:** The weighted ratings are then normalized to ensure that they fall within a consistent range, making comparisons easier.

### 3.3 Custom Penalty Functions

Introduce penalty functions to adjust the weighted normalized ratings based on percentile ranks, emphasizing top performers and penalizing lower-ranked hotels. Custom penalty functions allow travelers to prioritize hotels that meet certain criteria while avoiding those that fall below specific thresholds.

#### 3.3.1 No Penalty Zone

Define a percentile threshold ( $p_0$ ) above which hotels receive no penalty. This ensures that top-performing hotels maintain their high scores and are not penalized.

- **Purpose:** Preserve the ratings of top-performing hotels by not penalizing those that fall within the top percentile range.
- **Implementation:**

$$P(p) = \begin{cases} 0 & \text{if } p \geq p_0 \\ \text{Penalty Function} & \text{if } p < p_0 \end{cases} \quad (3)$$

- **Example:** If the no-penalty threshold is set at the 90th percentile, all hotels with a percentile rank of 90% or higher will not receive any penalty.

#### 3.3.2 Adjusting Penalty Sharpness

Control how rapidly penalties increase as percentile ranks decrease by adjusting the exponent ( $\delta$ ) in the penalty function. This allows travelers to customize how harshly lower-ranked hotels are penalized.

- **Higher  $\delta$ :** A higher value of  $\delta$  results in a sharper increase in penalties for lower-ranked hotels. This is useful for travelers who want to strictly avoid hotels that fall below a certain quality level.
- **Lower  $\delta$ :** A lower value of  $\delta$  results in a gentler increase in penalties. This is useful for travelers who are more flexible and willing to consider a broader range of hotels.

### 3.3.3 Piecewise Penalty Functions

Create a penalty function with different behaviors in specified percentile ranges. This allows for a more nuanced approach to penalizing hotels based on their rankings.

- **Gentle Penalty Zone:** Apply a mild penalty between  $p_0$  and an inflection point ( $p_1$ ). Hotels in this range are penalized slightly to reflect their lower ranking compared to top performers, but the penalty is not overly harsh.
- **Sharper Penalty Zone:** Apply a steeper penalty below  $p_1$ . Hotels in this range are penalized more significantly, reflecting their lower quality relative to other hotels in the region.
- **Continuity:** Ensure the penalty function is continuous at  $p_1$  to avoid sudden jumps in the penalty values. This ensures a smooth transition between the gentle and sharper penalty zones.

## 4 Implementation

The implementation of the proposed methodology involves several key steps, from data collection to calculating final scores for each hotel.

### 4.1 Data Collection

The first step in implementing the methodology is to collect data from various sources, including ratings, reviews, and other relevant information about hotels.

- **Ratings and Reviews:** Gather ratings from various platforms, such as Booking.com, TripAdvisor, and Expedia. Collect both numerical ratings and textual reviews to provide a comprehensive understanding of each hotel.
- **Percentile Calculation:** Compute the percentile rank for each hotel in the target region. This involves comparing the ratings of each hotel to those of other hotels in the same area.
- **Weights Determination:** Assign weights for international reviews and recency based on user preferences. For example, recent reviews may be assigned higher weights to reflect the current quality of a hotel.

### 4.2 Parameter Selection

Parameter selection is a critical aspect of the methodology, as it allows users to customize the hotel selection process based on their preferences.

- **No Penalty Threshold ( $p_0$ ):** The percentile threshold above which hotels receive no penalty. This threshold can be user-defined, e.g., 90%.

- **Inflection Point** ( $p_1$ ): The point at which the penalty sharpness changes. This can be adjusted to control how penalties are applied across different percentile ranges.
- **Penalty Coefficients** ( $\lambda_1, \lambda_2$ ): These coefficients determine the severity of penalties in different zones. Higher values result in more significant penalties.
- **Penalty Exponent** ( $\delta$ ): Controls the sharpness of the penalty below  $p_1$ . A higher value results in a steeper penalty.

### 4.3 Calculation Steps

The calculation process involves several steps to determine the final scores for each hotel.

1. **Compute Weighted Normalized Ratings:** Calculate the weighted ratings for each hotel based on reviewer origin and recency, and normalize the ratings.
2. **Apply Penalty Function:**
  - Determine which zone the hotel falls into based on its percentile rank.
  - Calculate the penalty using the appropriate formula based on the hotel's percentile rank.
3. **Calculate Final Score:**

$$\text{Final Score} = \text{Weighted Normalized Rating} - P(p) \quad (4)$$

4. **Rank Hotels:**
  - Sort hotels based on their final scores to generate a ranked list of accommodations.

## 5 Case Study

### 5.1 Hotel Data

To illustrate the proposed methodology, we present a case study involving a sample of hotels with percentile ranks ranging from 50% to 95% and initial weighted normalized ratings. The sample data is shown in Table 1.

Hotel	Percentile Rank ( $p$ )	Weighted Normalized Rating
A	95% (0.95)	8.8
B	90% (0.90)	8.5
C	85% (0.85)	8.2
D	80% (0.80)	8.0
E	75% (0.75)	7.8
F	70% (0.70)	7.5
G	65% (0.65)	7.3
H	60% (0.60)	7.0
I	55% (0.55)	6.8
J	50% (0.50)	6.5
K	45% (0.45)	6.3
L	40% (0.40)	6.0
M	35% (0.35)	5.8
N	30% (0.30)	5.5
O	25% (0.25)	5.3
P	20% (0.20)	5.0
Q	15% (0.15)	4.8
R	10% (0.10)	4.5

Table 1: Sample Hotel Data

## 5.2 Applying the Methodology

Parameters:

- **No Penalty Threshold** ( $p_0$ ): 90% (0.90)
- **Inflection Point** ( $p_1$ ): 70% (0.70)
- **Penalty Coefficients:**
  - $\lambda_1 = 2$  (Gentle penalty coefficient)
  - $\lambda_2 = 3$  (Sharper penalty coefficient)
- **Penalty Exponent** ( $\delta$ ): 2

Penalty Calculations:

- **No Penalty Zone** ( $p \geq 0.90$ ):

$$P(p) = 0 \tag{5}$$

- **Gentle Penalty Zone** ( $0.70 \leq p < 0.90$ ):

$$P(p) = \lambda_1 \times \left(1 - \frac{p}{p_0}\right) \tag{6}$$



- **Sharper Penalty Zone** ( $p < 0.70$ ):

$$P(p) = P(p_1) + \lambda_2 \times \left(1 - \left(\frac{p}{p_1}\right)^\delta\right) \quad (7)$$

- **Calculating  $P(p_1)$  for Continuity:**

$$P(p_1) = \lambda_1 \times \left(1 - \frac{p_1}{p_0}\right) = 2 \times \left(1 - \frac{0.70}{0.90}\right) = 0.4444 \quad (8)$$

### 5.3 Results and Analysis

The results of applying the penalty functions to the sample hotel data are shown in Table 2.

Hotel	Percentile Rank ( $p$ )	Weighted Normalized Rating	Penalty $P(p)$	Final Score
A	95% (0.95)	8.8	0	<b>8.8</b>
B	90% (0.90)	8.5	0	<b>8.5</b>
C	85% (0.85)	8.2	0.1	<b>8.1</b>
D	80% (0.80)	8.0	0.2	<b>7.8</b>
E	75% (0.75)	7.8	0.3	<b>7.5</b>
F	70% (0.70)	7.5	0.4	<b>7.1</b>
G	65% (0.65)	7.3	0.8	<b>6.5</b>
H	60% (0.60)	7.0	1.2	<b>5.8</b>
I	55% (0.55)	6.8	1.6	<b>5.2</b>
J	50% (0.50)	6.5	1.9	<b>4.6</b>
K	45% (0.45)	6.3	2.2	<b>4.1</b>
L	40% (0.40)	6.0	2.4	<b>3.6</b>
M	35% (0.35)	5.8	2.7	<b>3.1</b>
N	30% (0.30)	5.5	2.9	<b>2.6</b>
O	25% (0.25)	5.3	3.1	<b>2.2</b>
P	20% (0.20)	5.0	3.2	<b>1.8</b>
Q	15% (0.15)	4.8	3.3	<b>1.5</b>
R	10% (0.10)	4.5	3.4	<b>1.1</b>

Table 2: Final Scores After Applying Penalty Functions

#### Analysis:

- **Top Performers (Hotels A and B):**

– Remain unpenalized and retain their high scores. These hotels are clearly the best options for travelers who prioritize quality.

- **Middle Range (Hotels C to F):**

- Receive mild penalties, maintaining competitiveness. These hotels may still be viable options for travelers who are looking for a balance between quality and cost.
- **Lower Range (Hotels G to J):**
  - Penalties increase sharply, lowering final scores significantly. These hotels are less competitive and may only be suitable for budget-conscious travelers who prioritize cost over quality.

## 6 Benefits and Applications

The proposed methodology offers several benefits and can be applied in various scenarios to improve hotel selection for travelers.

- **Contextual Comparison:** Percentile rankings provide context by comparing hotels within the same region, allowing travelers to understand how a hotel performs relative to its peers. This helps in making more informed decisions based on the local market.
- **Customized Penalties:** Adjustable penalty functions allow users to tailor the importance of hotel rankings to their preferences. For example, travelers who prioritize high-quality accommodations can set a higher penalty for lower-ranked hotels, while those who are more flexible can reduce the penalty sharpness.
- **Balanced Decision-Making:** The methodology combines quantitative ratings with relative performance, aiding in more informed choices. This ensures that travelers are not solely reliant on absolute ratings but also consider how a hotel compares to others in the same area.
- **Flexibility:** Parameters such as the no-penalty threshold, inflection point, and penalty coefficients can be adjusted based on individual needs or changes in the hotel market. This flexibility allows travelers to customize the selection process to match their specific preferences and priorities.
- **Improved Transparency:** By using percentile rankings and penalty functions, the methodology provides a transparent framework for hotel selection. Travelers can understand how each hotel’s final score is calculated and adjust the parameters to see how different factors affect the rankings.
- **Applications in Travel Platforms:** This methodology can be integrated into travel booking platforms to provide more personalized hotel recommendations. By allowing users to set their preferences for weighting reviews and applying penalties, travel platforms can offer tailored recommendations that better match the needs of individual travelers.

## 7 Conclusion

The proposed methodology offers a systematic approach to normalizing hotel ratings by integrating percentile rankings and customizable penalty functions. By incorporating a no-penalty zone and adjusting the sharpness of penalties, travelers can prioritize top-performing hotels while considering a broader range of accommodations. This approach addresses the challenges posed by varying rating scales, reviewer biases, and regional differences, enabling more consistent and personalized hotel selection.

The flexibility of the methodology allows travelers to adjust the parameters based on their individual preferences, making it suitable for a wide range of use cases. Whether a traveler is looking for the highest quality accommodations or is more focused on finding a good value, the proposed approach provides a structured way to make informed decisions.

## 8 Future Work

There are several potential areas for future work to further improve and extend the proposed methodology:

- **Incorporating Textual Reviews:** Future work could involve incorporating sentiment analysis of textual reviews to complement the quantitative ratings. This would provide additional insights into the quality of a hotel, especially in areas that may not be captured by numerical ratings alone.
- **Integration with Price Sensitivity:** Incorporate price as a factor in the hotel selection process. By integrating price sensitivity, the methodology could help travelers find the best balance between quality and cost, providing recommendations that match both their quality expectations and budget.

## A Penalty Function Graph

*Note:* The graph below illustrates the penalty function with gentle and sharper penalty zones, showing how the penalty increases with decreasing percentile rank, becoming sharper below the 70th percentile.

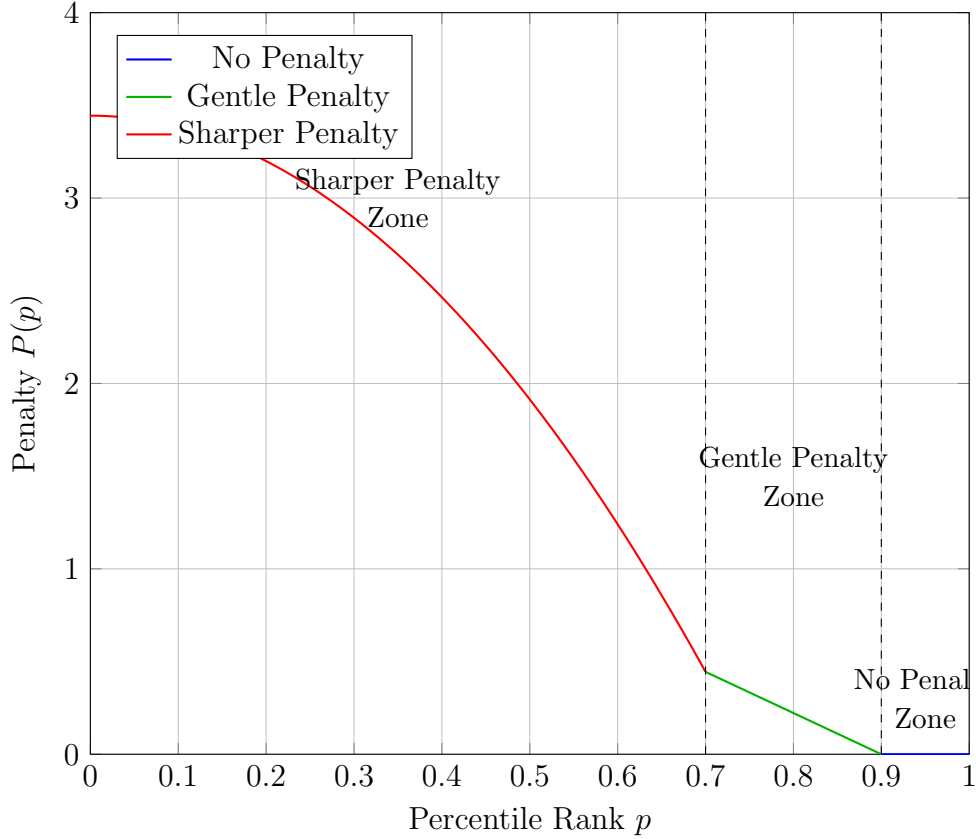


Figure 1: Penalty Function with Gentle and Sharper Zones

## B Parameter Adjustment Guidelines

- **No Penalty Threshold ( $p_0$ ):**
  - **Increase** to focus on top-tier hotels.
  - **Decrease** to widen the range of unpenalized hotels.
- **Inflection Point ( $p_1$ ):**
  - Adjust to control where the penalty sharpness changes.
- **Penalty Coefficients ( $\lambda_1, \lambda_2$ ):**
  - Modify to change the severity of penalties in respective zones.
- **Penalty Exponent ( $\delta$ ):**
  - **Increase** for a sharper penalty below  $p_1$ .
  - **Decrease** for a gentler slope.