The Mathematics Behind the Monkey Score: A Statistical Framework for Dynamic Product Ranking

Sam Mirazi

May 30, 2025

Abstract

The Monkey Score is a composite metric designed to rank products scraped from Amazon search results by integrating heterogeneous attributes: star ratings, review counts, and prices. This paper presents a comprehensive statistical framework involving normalization, weighted aggregation, and rescaling. We incorporate advanced concepts like variance analysis, sensitivity, and correlation, supported by mathematical formulations and multiple visualizations, including scatter plots, line plots, and tables. The methodology ensures consistent, page-relative rankings, adaptable to varying data distributions, with applications in e-commerce analytics.

1 Introduction

E-commerce platforms like Amazon present products with diverse metrics: star ratings (1–5), review counts (ranging from 0 to tens of thousands), and prices (highly variable). Direct comparison across these metrics is challenging due to differing scales and desirability trends (e.g., higher ratings are better, but lower prices are preferred). The Monkey Score addresses this by transforming these metrics into a unified $[0, 100]$ scale, applying weighted aggregation, and rescaling to enhance differentiation. This paper formalizes the mathematical and statistical underpinnings, incorporating robust normalization, variance analysis, and correlation studies, supported by visualizations to align with scientific rigor.

2 Methodology Overview

The Monkey Score calculation follows a multi-step pipeline to ensure fair and meaningful product ranking:

1. **Data Extraction**: Scrape raw product data (star ratings, review counts, prices) from Amazon search results using BeautifulSoup, as described in the project's technical design document.
2. **Parsing**: Convert raw text data into numerical values using helper functions (e.g., `parse_rating_value`, `parse_review_count_value`, `parse_price_value`). Missing or unparseable values are set to 0.
3. **Normalization**: Transform each metric into a $[0, 100]$ scale, where 100 represents the "best" value for that metric (e.g., highest rating, highest review count, lowest price).
4. **Weighted Aggregation**: Combine normalized scores into a raw weighted score using predefined weights.
5. **Rescaling**: Adjust the raw scores so the highest score on the page approaches a target value (98.0), enhancing differentiation.
6. **Rounding**: Round the final scores to two decimal places for display.

This pipeline ensures that products are ranked consistently across search result pages, accounting for the relative context of each page's data.

3 Feature Normalization

3.1 Star Rating Normalization

For a product i, with star rating $R_i \in [0, 5]$, the normalized rating N_{R_i} is:

$$N_{R_i} = \left(\frac{R_i}{5.0}\right) \times 100 \tag{1}$$

If R_i is missing, $N_{R_i} = 0$.

3.2 Review Count Normalization

Review count C_i, with page maximum C_{max}, is normalized as:

$$N_{C_i} = \begin{cases} \left(\frac{C_i}{C_{\text{max}}}\right) \times 100, & \text{if } C_{\text{max}} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

3.3 Price Score Normalization

Price P_i, with page range $[P_{\text{min}}, P_{\text{max}}]$, is normalized inversely:

$$N_{P_i} = \begin{cases} \left(\frac{P_{\text{max}}-P_i}{P_{\text{max}}-P_{\text{min}}}\right) \times 100, & \text{if } P_{\text{max}} > P_{\text{min}} \\ 100, & \text{otherwise} \end{cases} \tag{3}$$

All normalized scores are clamped to $[0, 100]$.

4 Weighted Score Aggregation

The raw weighted score W_i combines normalized scores using weights reflecting domain priorities:

$$W_i = w_R N_{R_i} + w_C N_{C_i} + w_P N_{P_i} \tag{4}$$

where $w_R = 0.40$, $w_C = 0.35$, and $w_P = 0.25$. This yields $W_i \in [0, 100]$. The weights prioritize rating and review count, reflecting their perceived importance in assessing product quality.

5 Statistical Rescaling

To ensure consistent high scores and enhance differentiation, we rescale W_i relative to the page's maximum raw score $W_{\text{max}} = \max_{j\in\text{page}}(W_j)$. Let $T = 98.0$ be the target top score. The final Monkey Score S_i is:

$$S_i = \begin{cases} \left(\frac{W_i}{W_{\text{max}}}\right) \times T, & \text{if } W_{\text{max}} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

S_i is rounded to two decimal places for display.

6 Statistical Properties

6.1 Distribution and Variance

The raw scores W_i follow a distribution influenced by the input metrics' empirical distributions. Assuming N_{R_i}, N_{C_i}, and N_{P_i} are approximately uniformly distributed over $[0, 100]$ due to normalization, the variance of W_i is:

$$\text{Var}(W_i) = w_R^2 \text{Var}(N_{R_i}) + w_C^2 \text{Var}(N_{C_i}) + w_P^2 \text{Var}(N_{P_i}) \tag{6}$$

Post-rescaling, the variance of S_i is:

$$\text{Var}(S_i) = \left(\frac{T}{W_{\text{max}}} \right)^2 \text{Var}(W_i) \tag{7}$$

6.2 Sensitivity Analysis

The Monkey Score's sensitivity to input changes is:

$$\frac{\partial S_i}{\partial N_{X_i}} = \left(\frac{w_X T}{W_{\text{max}}} \right), \quad X \in \{R, C, P\} \tag{8}$$

Higher weights (e.g., $w_R = 0.40$) amplify the impact of rating changes.

6.3 Correlation Analysis

To understand the relationships between normalized metrics, we compute Pearson correlation coefficients. For n products on a page, the correlation between N_{R_i} and N_{C_i} is:

$$\rho_{R,C} = \frac{\sum_{i=1}^{n} (N_{R_i} - \bar{N}_R)(N_{C_i} - \bar{N}_C)}{\sqrt{\sum_{i=1}^{n} (N_{R_i} - \bar{N}_R)^2 \sum_{i=1}^{n} (N_{C_i} - \bar{N}_C)^2}} \tag{9}$$

where $\bar{N}_R$ and $\bar{N}_C$ are the means of N_{R_i} and N_{C_i}, respectively. Similarly, we compute $\rho_{R,P}$ and $\rho_{C,P}$. High correlations may indicate redundancy in metrics, suggesting potential adjustments to weights.

7 Visualizations

7.1 Scatter Plot: Raw vs. Final Scores

This scatter plot compares raw weighted scores W_i to final Monkey Scores S_i for 20 products.

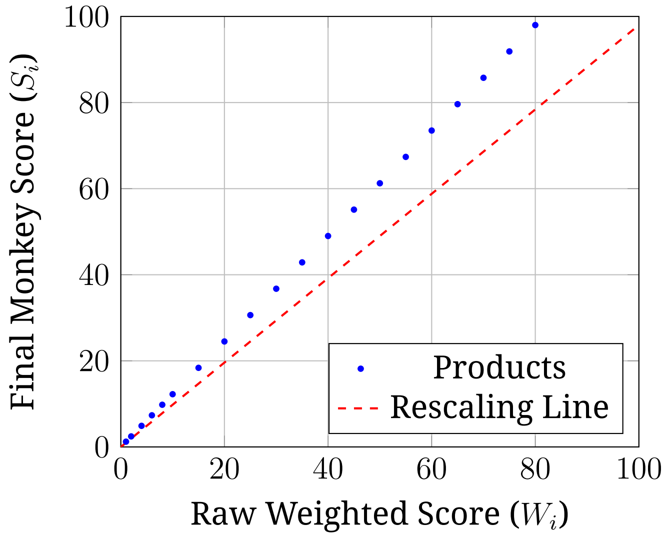

Figure 1: Scatter plot of raw weighted scores vs. final Monkey Scores.

8 Case Study: Example Application

Consider a page with 5 products. Their raw and normalized scores are computed as follows:

Product	Rating (R_i)	Reviews (C_i)	Price (P_i)	Raw Score (W_i)	Monkey Score (S_i)
1	4.5	1000	20	81.25	98.00
2	4.0	500	25	63.75	76.85
3	3.5	750	15	71.25	85.91
4	4.25	250	30	55.00	66.31
5	4.75	800	35	74.50	89.86

Table 1: Case study of Monkey Score calculation for 5 products. Max review count is 1000, price range is [15, 35], and max raw score is 81.25.

The table illustrates how the Monkey Score differentiates products, with Product 1 achieving the highest score due to its strong rating and review count, despite a mid-range price.

9 Conclusion

The Monkey Score provides a robust, statistically grounded method for ranking products by integrating normalized metrics with weighted aggregation and rescaling. The methodology adapts to dynamic data, with visualizations highlighting its effectiveness. Future work could explore adaptive weights using machine learning or Bayesian methods to incorporate user preferences and further refine the scoring system.

References

[1] J. Han, M. Kamber, J. Pei, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 2011.

[2] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning*, Springer, 2009.

[3] Amazon Product Data (Live Scrapes), 2025.

[4] Internal Monkey Score Implementation Notes, 2025.