

E-Commerce Platform Ranking Based on Target-Based Performance Indicators

Ufuk Bölükbaş^{1,*}, Yasin Çetinkaya¹, Kaan Umut Adaş¹,

¹ Department of Industrial Engineering, Yildiz Technical University, Istanbul, Türkiye

ARTICLE INFO

Article history:

Received 24 July 2026

Received in revised form 11 September 2026

Accepted 16 October 2026

Available online 21 October 2026

Keywords:

E-commerce; Multi-criteria decision making; Analytic Hierarchy Process; CRITIC; TOPSIS; Exponentially weighted moving average

ABSTRACT

Managing a broad product assortment on an e-commerce platform requires category performance to be judged jointly across several, frequently conflicting, indicators rather than through a single sales figure, which turns category ranking into a multi-criteria decision-making problem. In such problems the outcome is driven above all by how the criteria weights are obtained, yet applications in the e-commerce domain almost always adopt either a subjective or an objective weighting scheme, rarely both. This study evaluates and ranks thirty-two product categories within an active e-commerce firm using seven target-based key performance indicators (KPIs), where each criterion value is operationalized as the ratio of realized performance to the predetermined target. Furthermore, the paper provides a comparative analysis of the outcomes yielded by subjective versus objective weighting methods applied to the same decision-making problem. Criteria weights are derived from the pairwise judgements of two domain experts through the Analytic Hierarchy Process (AHP) and, independently, from the statistical structure of the decision matrix through the Criteria Importance Through Intercriteria Correlation (CRITIC) method; both weight vectors are then supplied to a common TOPSIS ranking. Three months data are condensed into a single decision matrix using an exponentially weighted moving average ($\lambda = 0.7$) so that the most recent period dominates while earlier trends are retained. The two methods produce almost opposite weight structures — AHP concentrates roughly two-thirds of the weight in gross order count and net sales, whereas CRITIC assigns the largest weights to conversion rate and session count — yet the resulting rankings agree strongly (Spearman's $\rho = 0.878$, mean absolute rank shift of 3.7 positions, four of the top five categories shared). Categories that perform consistently across all criteria are insensitive to the weighting scheme, whereas categories with sharply unbalanced criteria profiles move by up to ten positions. The study delivers a systematic and reproducible decision-support model for category prioritisation and shows that the choice between subjective and objective weighting matters selectively rather than uniformly.

1. Introduction

The rapid growth of electronic commerce over the past decade has turned the management of an online platform into an increasingly complex decision problem. For a company that operates a

* Corresponding author.

E-mail address: bolukbas@yildiz.edu.tr

wide assortment, monitoring aggregate sales volume is no longer sufficient: net sales, order and customer counts, session volume and conversion rate all must be read together at the category level.

These indicators frequently point in different directions—a category that is strong on one indicator may lag badly on another—so category success cannot be judged against a single yardstick and instead calls for a holistic assessment in which several criteria are weighed simultaneously.

Evaluating and ranking alternatives against multiple conflicting criteria is, by definition, a multi-criteria decision-making (MCDM) problem. Within any MCDM application, the stage that most directly determines the final ranking is the derivation of the criteria weights. Weighting approaches are conventionally grouped into subjective methods (which elicit weights from the preferences of decision makers, such as the Analytic Hierarchy Process - AHP) and objective methods (which derive weights solely from the statistical structure of the data, such as CRITIC). TOPSIS is frequently adopted as a common ranking backbone due to its conceptual simplicity and flexibility. However, while e-commerce applications typically adopt either a subjective or an objective weighting scheme, combining both to compare their outcomes under identical conditions remains scarce. Furthermore, most evaluations rely on a static cross-section of data and focus on independent platforms rather than product categories within a single firm.

The present study addresses these gaps in a real setting. The motivating company receives category-level performance reports on a regular basis, but the reports are read selectively: experts examine only the categories they consider a priority and only the indicators they personally regard as important. There is no systematic structure in which all criteria enter with defined weights and all categories are ranked end to end in a reproducible manner. This creates two problems. Because the assessment depends on the attention of a single analyst and on a self-selected set of indicators, the results are subjective and difficult to replicate; and because category performance fluctuates over time, a static view of one period fails to reflect current trends and periodic mobility. The problem is therefore defined as the development of a systematic decision-support model that combines several performance criteria with defined weights, incorporates temporal information, and generates the category success ranking automatically.

The contribution of the paper is fourfold. (i) The same e-commerce category ranking problem is solved separately with subjective AHP–TOPSIS and objective CRITIC–TOPSIS on an identical criterion set, category set, data set and temporal treatment, so that any difference in the rankings is attributable to the weighting approach alone. (ii) Daily and monthly target-based performance data are processed with an exponentially weighted moving average (EWMA), which brings temporal information into the decision matrix while keeping the ranking a single-shot procedure. (iii) The agreement between the two rankings is quantified rather than described, using Spearman's rank correlation, the mean absolute rank shift and top-k overlap. (iv) The pattern of agreement is explained: it is shown that the sensitivity of a category to the weighting scheme is governed by the dispersion of its criterion-level ranks combined with its overall performance level, which yields a simple diagnostic for practitioners. The remainder of the paper is organised as follows. Section 2 reviews the relevant literature. Section 3 sets out the problem definition, the data, the EWMA aggregation, the AHP and CRITIC weighting procedures and the TOPSIS ranking. Section 4 reports and discusses the results. Section 5 concludes and outlines directions for further work.

2. Literature Review

MCDM techniques are built from three basic components—alternatives, criteria and their weights—and are widely applied in business and engineering to performance assessment and prioritization. In a review of the field, Dalbudak and Rençber [1] report that TOPSIS is one of the most frequently applied techniques. Its conceptual simplicity and the ease with which it accepts any weight

vector make it a natural ranking backbone whenever different weighting schemes are to be compared under identical conditions. The decisive influence of the weighting choice was demonstrated concretely by Ersoy [2], who found that ranking differences between methods become pronounced whenever the alternative set changes, highlighting the need to carefully consider subjective, objective, and integrated models, as proposed by Zoraghi *et al.*, [3] and Marković *et al.*, [4].

Subjective methods, particularly AHP and its network-structured extension ANP, are frequently encountered in e-commerce and retail decision problems. Burnaz and Topcu [5] addressed store location selection in retailing, while Tyagi *et al.*, [6] applied an AHP–TOPSIS hybrid to identify the alternative that would most improve e-supply-chain performance. Where judgemental uncertainty is high, fuzzy extensions have been preferred. For instance, Li and Sun [7] prioritised design factors of a B2C e-commerce website with fuzzy AHP and TOPSIS-Grey, identifying service quality as the most critical factor. Similarly, Ishak *et al.*, [8] evaluated the service quality of Indonesian platforms, and Ocampo *et al.*, [9] evaluated online marketing strategies. The common feature of these studies is that criteria weights are established from expert assessments, making the resulting rankings dependent on the judgement of the decision maker.

Among objective weighting methods, CRITIC, developed by Diakoulaki *et al.*, [10], is the most prominent. The method quantifies the information content of a criterion through contrast intensity (standard deviation) and conflict (correlation). Krishnan *et al.*, [11] proposed D-CRITIC based on distance correlation to produce more valid weights. In most applications, CRITIC is paired with a ranking method. Görçün [12] combined CRITIC with EDAS to evaluate urban transport vehicles, while Yin *et al.*, [13] directly integrated CRITIC with TOPSIS for community governance evaluation. Closer to the e-commerce context, Arisoy and İç [14] proposed a modified TOPSIS model to rank e-commerce platforms from categorical data, and Chodak [15] examined the application of multi-criteria evaluation to e-commerce logistics, drawing attention to the importance of the weight choice.

As summarised in Table 1, the multi-criteria decision-making (MCDM) literature related to e-commerce employs both subjective and objective weighting methods, frequently pairing them with TOPSIS for the final ranking of alternatives. TOPSIS was the most frequently applied technique in the review of forty-seven studies by Dalbudak and Rençber [1], which supports its use as a common ranking method. Subjective AHP-based approaches combined with TOPSIS have been used to identify the highest-priority alternative for e-supply-chain performance [6] and to evaluate online marketing strategies in the hospitality sector [9]. Under uncertainty, fuzzy extensions have been preferred, as in the prioritisation of design factors for B2C e-commerce websites by Li and Sun [7] and the assessment of e-commerce service quality by Ishak *et al.*, [8].

On the objective side, the CRITIC method of Diakoulaki *et al.*, [10] was later refined by Krishnan *et al.*, [11] through distance correlation, and has been paired with EDAS [12] and TOPSIS [13]. Closer to the present context, Arisoy and İç [14] ranked e-commerce platforms with a modified TOPSIS model, Özs Salman *et al.*, [16] evaluated 25 e-commerce websites in Turkey on eleven technical and user-centric indicators, and Chodak [15] validated multi-criteria models on data from an operating online store while emphasising the importance of the weight choice. The influence of the weighting method on the outcome was shown directly by Ersoy [2], and integrated subjective–objective models were proposed by Zoraghi *et al.*, [3] and Marković *et al.*, [4]. Crucially, however, the e-commerce studies in Table 1 evaluate platforms or websites rather than the product categories of a single firm, and studies that compare subjective AHP and objective CRITIC weights within a shared TOPSIS framework on the same decision problem remain scarce. The present study builds on these foundations by making this comparison at the category level, thereby isolating the specific effect of the weighting scheme on e-commerce category evaluation.

Table 1
Summary of the literature

Author(s)	Year	Purpose of the study	Method	Main findings
Özsalman <i>et al.</i> , [16]	2026	Performance evaluation of e-commerce websites	MAXC, SITDE, CORASO (objective)	Evaluated 25 e-commerce websites in Turkey using 11 technical and user-centric indicators.
Arisoy & İç [14]	2025	Ranking e-commerce platforms from categorical data	Modified TOPSIS, VIKOR, MOORA (objective)	The proposed TOPSIS model reflects categorical data more holistically and produces consistent rankings.
Chodak [15]	2024	Supplier and assortment evaluation in e-commerce	Multi-criteria evaluation on real data	Models validated with data from an operating online store; importance of the weight choice emphasised.
Yin <i>et al.</i> , [13]	2023	Evaluating smart-community governance performance	CRITIC–TOPSIS (objective)	Indicator weights from CRITIC and a performance ranking of seven communities from TOPSIS.
Dalbudak & Rençber [1]	2022	Literature review of MCDM methods	Review of 47 studies	TOPSIS is the most frequently used method (35%), followed by AHP and VIKOR.
Ersoy [2]	2022	Effect of the weighting method on MCDM outcomes	Entropy, CRITIC, MEREC, SD, IDDSW + SAW	When the alternative set changes, the weighting method can materially alter the ranking.
Krishnan <i>et al.</i> , [11]	2021	Improving CRITIC to produce more valid weights	D-CRITIC (objective)	Distance correlation yields more valid weights and greater stability on large decision matrices.
Ocampo <i>et al.</i> , [9]	2021	Evaluating online marketing strategies in hospitality	AHP–TOPSIS (subjective)	Customer and business perspectives integrated to prioritise strategies.
Ishak <i>et al.</i> , [8]	2021	Assessing the service quality of e-commerce platforms	Fuzzy AHP–TOPSIS (subjective)	The platform scoring highest on the product-return criterion offers the best service quality.
Görçün [12]	2021	Selecting metro and tram vehicles for urban transport	CRITIC–EDAS (objective)	Objective weighting simplifies problems containing a large number of indicators.
Marković <i>et al.</i> , [4]	2020	Ranking alternatives for business excellence and sustainability	Fuzzy PIPRECIA + CRITIC + I-distance (integrated)	Integrating subjective and objective methods yields more balanced weights.
Li & Sun [7]	2020	Prioritising design factors of a B2C e-commerce website	Fuzzy AHP + TOPSIS-Grey (subjective)	Service quality identified as the most critical factor; the integrated model performs well under uncertainty.
Tyagi <i>et al.</i> , [6]	2014	Selecting the alternative that improves e-supply-chain performance	AHP–TOPSIS (subjective)	Investment in web-based technologies ranked as the highest-priority alternative.
Zoraghi <i>et al.</i> , [3]	2013	Evaluation of service quality in hotels	Fuzzy MCDM (subjective + objective)	Combining subjective and objective weights compensates for the shortcomings of either approach used alone.
Burnaz & Topcu [5]	2006	Retail store location selection in apparel retailing	ANP (subjective)	Shopping centres preferred over central business districts; accessibility and competition level identified as decisive criteria.
Diakoulaki <i>et al.</i> , [10]	1995	Deriving criteria weights independently of subjective judgement	CRITIC	Objective weights obtained by combining standard deviation and inter-criteria correlation; results close to those of principal component analysis.

3. Methodology

This section describes the integrated MCDM model built to evaluate and rank the product categories of an e-commerce company based on performance. The study is quantitative, applied and

case based. The proposed model is constructed as two parallel strands: a subjective strand, in which criteria weights are established from expert judgement with AHP and categories are ranked with TOPSIS, and an objective strand, in which the weights are derived from the statistical structure of the data with the CRITIC method introduced by Diakoulaki *et al.*, [10] and the categories are ranked with the same TOPSIS procedure developed by Hwang and Yoon [17]. In both strands the performance data of different periods are merged into a single decision matrix by EWMA, so that the temporal performance trend enters the ranking process. Because both strands share the problem definition, the data set, the category set and the temporal treatment, the effect of subjective versus objective weighting on the ranking can be isolated. Figure 1 summarises the flow.

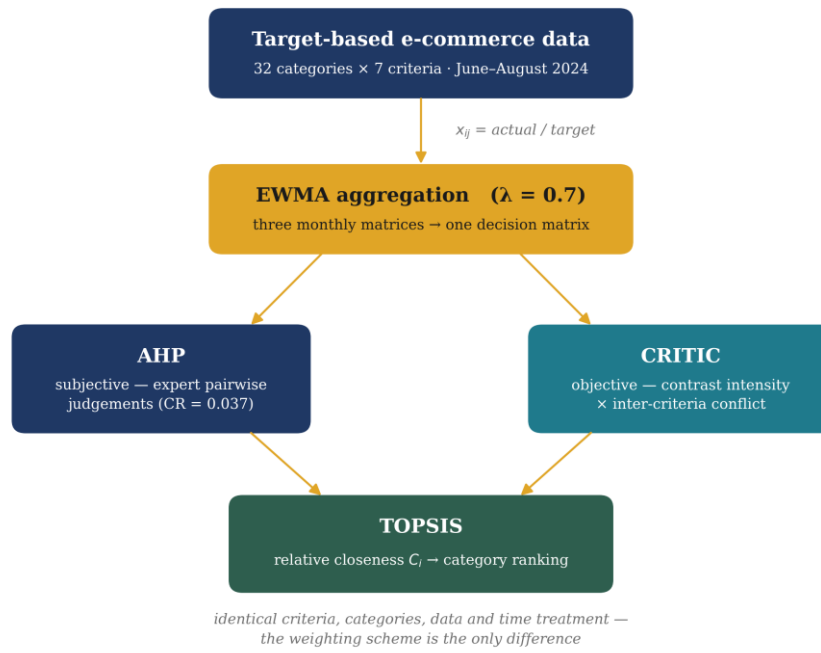


Fig. 1. Flowchart of the proposed integrated decision model

3.1 Problem Definition and Decision Variables

The decision problem is defined as an MCDM problem in which m alternatives are evaluated against n criteria. The alternatives are the 32 product categories of the e-commerce platform ($m = 32$) and the criteria are the 7 indicators that measure category performance ($n = 7$). The initial decision matrix X is written as

$$X = [x_{ij}]_{m \times n}, \quad i = 1, \dots, m; \quad j = 1, \dots, n \quad (1)$$

where x_{ij} denotes the value of the i -th alternative on the j -th criterion. The starting point is the realised (actual) value of each of the 32 categories on each of the seven criteria. Because the categories differ greatly in scale, comparing raw realised values directly would be misleading. For every category–criterion pair the company sets a target (budget) value in advance, and the realised value is divided by that target:

$$x_{ij} = \text{Realised}_{ij} / \text{Target}_{ij} \quad (2)$$

The ratio is dimensionless and is interpreted as a percentage: a value of 1.00 means the target was met exactly, 1.20 that was exceeded by 20%, and 0.80 that realisation fell 20% short. Categories of very different sizes are thereby transferred onto a common, comparable scale, and the evaluation

is carried out in terms of each category's achievement against its own target rather than in absolute magnitudes.

The seven criteria and their definitions are given in Table 2. All criteria are benefit-oriented, so an increase in the ratio means that the category has met its target on that criterion to a greater degree and has performed better.

Table 2
Evaluation criteria and their definitions

Criteria	Abbr.	Definition.	Direction
Net Sales	NS	Actual sales revenue generated by the products sold in the category after returns and cancellations are deducted; indicates the contribution of the category to total turnover.	Benefit
First-Party Net Sales	1P	Net revenue from transactions in which products are sold directly from the company's own inventory, that is, goods the platform itself procures and sells.	Benefit
Third-Party Net Sales	3P	Net revenue from products sold by external sellers on the marketplace; indicates the contribution of the marketplace business model to the category.	Benefit
Gross Order Count	GOC	Total number of orders placed by customers in the category before cancellations and returns; reflects the size of demand and purchase frequency.	Benefit
Customer Count	CUS	Number of distinct customers purchasing from the category; a customer buying several times is counted once. Indicates the actual customer base reached.	Benefit
Session Count	SES	Number of visits to the category pages (site traffic). Every visit in which a user enters and browses the site counts as one session; measures interest in the category.	Benefit
Conversion Rate	CVR	Share of users visiting the category pages whose visit ends in a purchase; the core efficiency measure of how effectively incoming traffic is converted into sales.	Benefit

The following paragraphs summarizes how and for what purpose the evaluation criteria presented in Table 2 are used across different studies;

Dolega *et al.*, [18] investigate the relationship between social media marketing activities and business performance using 12 months of data from a major online retailer. The study separately examines website traffic, order volume, and sales revenue, demonstrating the importance of distinguishing between customer visits and actual commercial outcomes. The findings provide an empirical basis for considering traffic, order volume, and sales revenue as separate indicators of e-commerce category performance.

Etro [19] examines the product selection decisions of online marketplaces operating under both first-party (1P) and third-party (3P) business models. The study analyzes the economic incentives underlying a platform's decision to sell products directly or facilitate sales by independent sellers. It provides a conceptual foundation for distinguishing between first-party and third-party sales in the performance evaluation of e-commerce categories. Etro [20] reviews the economic literature on hybrid e-commerce marketplaces, focusing on platforms that simultaneously sell their own products and host third-party sellers. The study discusses the revenue models associated with direct product sales and commissions from third-party transactions, as well as the competitive implications of platform self-preferencing. It provides additional conceptual support for distinguishing first-party and third-party business activities when evaluating e-commerce platform performance.

Di Fatta *et al.*, [21] examine the determinants of conversion rates in small and medium-sized e-commerce websites using 1,184 observations from six websites. The study investigates the factors

associated with the proportion of website visitors who complete a purchase. It supports the use of conversion rate as an indicator of how effectively website traffic is translated into purchasing activity. McDowell *et al.*, [22] investigate the relationship between retail website design features and online conversion rates using regression analysis. The findings indicate that website features associated with an immersive browsing experience are positively related to the conversion of visitors into customers. The study supports the use of conversion rate as an indicator of the effectiveness of an e-commerce website in generating purchases from incoming traffic.

Perdikaki *et al.*, [23] examine the relationship between store traffic, labor, and sales performance using data from an apparel retailer. The study decomposes sales volume into conversion rate and basket value, demonstrating that traffic and conversion efficiency represent distinct dimensions of retail performance. The findings provide a conceptual and empirical basis for evaluating website traffic, sales revenue, and conversion rate as separate performance criteria, although the study focuses on physical retail stores rather than online marketplaces.

Lohse and Spiller [24] investigate how online store interface features influence website traffic and sales using data from 28 online retail stores. The study identifies relationships between interface design characteristics, store traffic, and monthly sales revenue. Its findings support the distinction between website traffic as a measure of customer interest and sales revenue as a financial performance outcome in online retailing.

Unal and Park [25] investigate the effects of one-click purchasing on customer behavior using quasi-experimental data from an online retailer over a 35-month period. The findings indicate that one-click purchasing increases purchase activity by encouraging customers to buy more frequently and purchase more items. The study also links these changes to increased website visits and page views, supporting the joint consideration of purchase activity and website engagement in e-commerce performance evaluation.

In evaluating these performance dimensions, the choice of MCDM methodology is critical. Existing empirical literature frequently emphasizes that relying strictly on subjective or objective weighting schemes in isolation can bias ranking outcomes. Subjective methods effectively capture domain expertise and strategic priorities, whereas objective methods extract the inherent information content and contrast intensity from empirical data. Guided by precedents in the literature—such as Tyagi *et al.*, [6], who examined e-commerce performance by juxtaposing AHP and CRITIC, and the extensive body of work validating TOPSIS as a robust ranking engine (e.g., Hwang & Yoon, [17]; Dalbudak & Rençber, [1]; Yin *et al.*, [13])—this study establishes an integrated evaluation framework. Specifically, the Analytic Hierarchy Process (AHP) is employed to elicit strategic, expert-driven weights; the CRITIC method is implemented to capture objective, data-driven weights; and TOPSIS is utilized as the common ranking backbone under identical conditions. This methodological integration ensures a rigorous, reproducible, and balanced assessment that aligns with both managerial judgement and statistical data structure.

3.2 Data Preparation for Case

The data are drawn from the category-level operational performance records of an operating e-commerce company. For every category–criterion pair the raw records contain both the realised value and the internally predetermined target (budget) value; the criteria values entering the decision matrix are derived as the ratio of these two quantities in accordance with Eq. (2). To protect the commercial confidentiality of the company, the raw realised and target values are not disclosed; the analysis is presented entirely in terms of the dimensionless ratios computed from them, and the company is not named.

In the AHP–TOPSIS strand the ranking is carried out on monthly category-level performance data for June, July and August. Each month is organised as an independent decision matrix, so that the movement of category rankings between periods can be observed. The pairwise comparison judgements used to determine the criteria weights were obtained from two senior business professionals with extensive domain expertise working within the company. Both have direct experience of category performance indicators and of company priorities, which supports the reliability of the AHP judgements.

In the CRITIC–TOPSIS strand the data are treated in two stages. In the first stage, one week of daily data is used as a pilot to test sensitivity to parameter choices, and four scenarios are compared. In the second stage, once the scenario has been selected, the final evaluation is carried out on the three-month (June–August) monthly data. In this strand the criteria weights are computed objectively from the statistical structure of the decision matrix, without recourse to expert opinion.

3.3 Constructing the Decision Matrix with EWMA

The exponentially weighted moving average (EWMA) is used to merge observations from different periods into a single decision matrix. EWMA summarises the current trend of the categories in a single value by assigning higher weight to recent periods than to earlier ones. The method is defined recursively as

$$EWMA_t = \alpha \cdot x_t + (1 - \alpha) \cdot EWMA_{t-1} \quad (3)$$

where x_t is the value realised in period t , $EWMA_{t-1}$ is the EWMA value of the previous period, and α is the smoothing factor taking values in $[0, 1]$. The computation starts from the earliest period and proceeds towards the most recent one; the value written into the decision matrix is the EWMA value of the most recent period at the end of this chain. A large α makes the system highly responsive to recent data, whereas a small α preserves the historical trend for longer. In this study $\alpha = \lambda = 0.7$ is used, which distributes the contribution to the final value as approximately 70% for the most recent month (August), 21% for the middle month (July) and 9% for the earliest month (June), as shown in Figure 2.

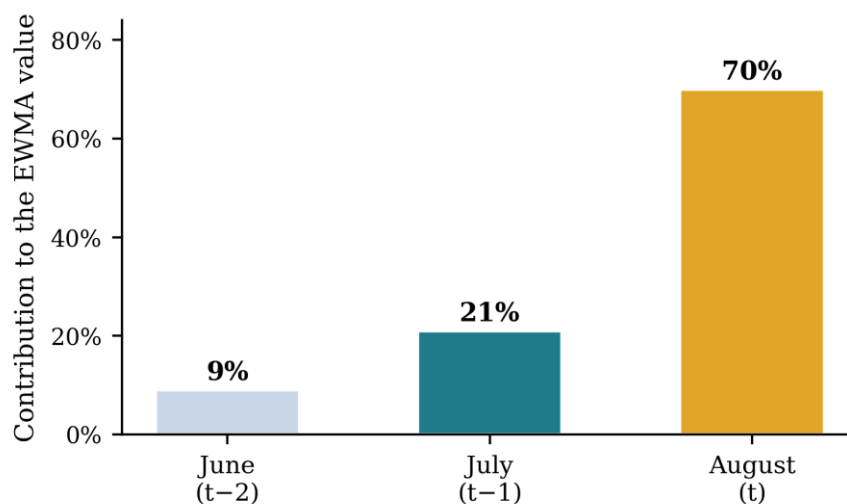


Fig. 2. Contribution of each month to the EWMA value ($\lambda = 0.7$)

3.4 Subjective Weighting: The Analytic Hierarchy Process

AHP determines the relative importance of decision criteria from the pairwise comparisons of decision makers and renders the consistency of those judgements' measurable [26]. Every pair of criteria is compared in terms of how much more important one is than the other, and the

comparisons are quantified on Saaty's fundamental 1–9 scale. A value of 1 indicates that two criteria are equally important and a value of 9 that one is more important than the other; intermediate values (3, 5, 7) and their reciprocals (1/3, 1/5, ...) express the direction and intensity of the preference. The scale is defined in Table 3.

Table 3

Saaty's fundamental 1–9 pairwise comparison scale

Intensity of importance	Definition.
1	Equally important
3	Moderately more important
5	Strongly more important
7	Very strongly more important
9	Absolutely more important
2, 4, 6, 8	Intermediate values between adjacent judgements
Reciprocals (1/3, 1/5, ...)	Corresponding value for the reverse direction of the comparison

Each of the two experts compared the seven criteria pairwise, giving one judgement for each of the 21 comparisons. The judgements of the two experts were converted into a single consensus judgement using the geometric mean, the standard aggregation procedure for group decisions in AHP Aczél, J., & Saaty, T. L., [26]. The geometric mean is preferred over the arithmetic mean because it preserves the reciprocity property of ratio-scale comparison judgements. The aggregated comparison value is computed as

$$a_{ij}^{(b)} = \sqrt{(a_{ij}^{(1)} \cdot a_{ij}^{(2)})} \quad (4)$$

where $a_{ij}^{(1)}$ and $a_{ij}^{(2)}$ denote the judgements of the first and second expert on the importance of criterion i relative to criterion j , and $a_{ij}^{(b)}$ the aggregated judgement. Applying this operation to all 21 comparisons yields the aggregated pairwise comparison matrix. From the aggregated judgements an $n \times n$ comparison matrix A with unit diagonal satisfying the reciprocity condition is constructed:

$$A = [a_{ij}]_{n \times n}, \quad a_{ij} = 1, \quad a_{ji} = 1 / a_{ij} \quad (5)$$

Each column of the matrix is normalised by its column total,

$$\bar{a}_{ij} = a_{ij} / \sum_{k=1}^n a_{kj} \quad (6)$$

and the criteria weights are obtained as the arithmetic mean of each row of the normalised matrix:

$$w_i = \left(\frac{1}{n}\right) \cdot \sum_{j=1}^n \bar{a}_{ij}, \quad \sum_{i=1}^n w_i = 1 \quad (7)$$

Here w_i is the weight of the i -th criterion and the weights sum to unity. This weight vector forms the input to the TOPSIS ranking in the AHP–TOPSIS strand.

Because the pairwise comparisons rest on expert judgement, it is necessary to check whether the judgements contradict one another logically. AHP provides this check through the consistency ratio (CR). The principal eigenvalue of the comparison matrix is computed as

$$\lambda_{max} = \left(\frac{1}{n}\right) \cdot \sum_{i=1}^n (A \cdot w)_i / w_i \quad (8)$$

the consistency index (CI) as

$$CI = \frac{\lambda_{max} - n}{(n-1)} \quad (9)$$

and the consistency ratio as the ratio of CI to the random index RI, the mean consistency index of randomly generated matrices of the same order:

$$CR = \frac{CI}{RI} \quad (10)$$

For a matrix with seven criteria ($n = 7$) the value $RI = 1.32$ is used [26]. Satisfying the condition $CR \leq 0.10$ indicates that the expert judgements are consistent to an acceptable degree.

3.5 Objective Weighting: The CRITIC Method

The essential difference of the objective strand from AHP is that the weights are computed directly from the statistical properties of the decision matrix — contrast intensity and inter-criteria correlation — without any expert judgement. Criteria values expressed as ratios may, for some category–criterion pairs, contain extreme values far above target. Such outliers can disproportionately affect both the EWMA decision matrix and the standard deviation and correlation structure on which the CRITIC method rests Krishnan *et al.*, [11], so that a criterion weight comes to be determined by a single observation. To limit this effect, an optional pre-processing step applied before the EWMA computation sets values exceeding a specified upper bound (cap) equal to that bound, while values below the bound are preserved unchanged:

$$x_{ij}^{(c)} = \min(x_{ij}, c) \quad (11)$$

The cap c is set in this study at twice the target ($c = 2.00$). Whether or not the cap is applied constitutes one of the two dimensions of the scenario design, the other being the level of the EWMA smoothing factor ($\lambda = \alpha$). Crossing two levels of each dimension gives a 2×2 factorial design with four scenarios (Table 4). The scenarios were compared on the pilot data to expose the effect of the parameter choices on the criteria weights and to identify the most responsive configuration.

Table 4

Scenario design (2×2 factorial) for constructing the CRITIC decision matrix

Scenario	Outlier capping	Smoothing factor ($\lambda = \alpha$)
Scenario 1	Applied (cap = 2.00)	0.3 (conservative)
Scenario 2	Not applied (original data)	0.3 (conservative)
Scenario 3	Applied (cap = 2.00)	0.7 (recency-sensitive)
Scenario 4	Not applied (original data)	0.7 (recency-sensitive)

CRITIC determines the weights objectively by analysing the internal structure of the data in the decision matrix [10]. The method begins with min–max normalisation of the decision matrix. Since all criteria are benefit-oriented, the following transformation is applied:

$$x_{ij}^* = \frac{(x_{ij} - \min_i x_{ij})}{(\max_i x_{ij} - \min_i x_{ij})} \quad (12)$$

On the normalised matrix, the contrast intensity of each criterion is measured by the sample standard deviation σ_j :

$$\sigma_j = \sqrt{\left[\left(\frac{1}{m-1}\right) \cdot \sum_{i=1}^m (x_{ij}^* - \bar{x}_j^*)^2\right]} \quad (13)$$

The relationship between criteria is measured by the Pearson correlation coefficient r_{jk} :

$$r_{jk} = \frac{\sum_i (x_{ij}^* - \bar{x}_j^*)(x_{ik}^* - \bar{x}_k^*)}{\sqrt{[\sum_i (x_{ij}^* - \bar{x}_j^*)^2][\sum_i (x_{ik}^* - \bar{x}_k^*)^2]}} \quad (14)$$

The total quantity of information carried by each criterion, C_j , is computed as the product of contrast intensity and conflict:

$$C_j = \sigma_j \cdot \sum_{k=1}^n (1 - r_{jk}) \quad (15)$$

and the criterion weights are obtained by normalising the information quantities:

$$w_j = \frac{C_j}{\sum_{k=1}^n C_k}, \quad \sum_{j=1}^n w_j = 1 \quad (16)$$

Criteria that exhibit high variability, or that show low or negative correlation with the other criteria, receive higher weights. This multiplicative structure means that the weight is determined not by variability alone nor by independence alone, but by the two conditions being satisfied together.

3.6 Ranking the Alternatives with TOPSIS

TOPSIS is a ranking method resting on the principle that each alternative should be as close as possible to the positive-ideal solution and as far as possible from the negative-ideal solution [17]. The criteria weights obtained from AHP in the first strand and from CRITIC in the second are used at this step in the same way; the steps of the method are identical for both approaches. In the first step, the decision matrix is normalised column-wise by the vector norm:

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^m (x_{ij})^2}} \quad (17)$$

The normalised values are multiplied by the criteria weights to form the weighted normalised matrix:

$$v_{ij} = w_i \cdot r_{ij} \quad (18)$$

Since all criteria are benefit-oriented, for each criterion the largest value forms the positive-ideal solution (A^+) and the smallest value the negative-ideal solution (A^-):

$$A^+ = \{\max_i v_{ij}\}, \quad A^- = \{\min_i v_{ij}\} \quad (19)$$

The Euclidean distances (separation measures) of each alternative from the positive-ideal and negative-ideal solutions are computed as

$$S_i^+ = \sqrt{(\sum_{j=1}^n (v_{ij} - v_j^+)^2)} \quad (20)$$

$$S_i^- = \sqrt{(\sum_{j=1}^n (v_{ij} - v_j^-)^2)} \quad (21)$$

and the relative closeness coefficient of each alternative to the ideal solution as

$$C_i = S_i^- / (S_i^+ + S_i^-) \quad (22)$$

The value C_i lies in $[0, 1]$; the closer it is to 1 the nearer the alternative is to the positive ideal, and the closer to 0 the nearer to the negative ideal. Alternatives are ordered by C_i from largest to smallest and the category with the highest coefficient is deemed the most successful. The influence of the weights is visible in Eq. (18): because each criterion column is multiplied by its own weight, high-weight criteria dominate the distances to the ideal and negative-ideal solutions. In AHP–TOPSIS the ranking is therefore shaped by expert priorities, and in CRITIC–TOPSIS by the statistical structure of the data.

3.7 Comparing the Two Rankings

Because the two strands differ only in their weight vectors, the divergence between their outputs can be quantified directly. Three measures are used. Agreement over the whole ordering is measured by Spearman's rank correlation coefficient. As a non-parametric measure of statistical dependence, this coefficient evaluates the strength and direction of the monotonic relationship between two sets of rankings [27]. It yields a value between -1 and +1, where +1 indicates perfectly identical rankings, -1 indicates completely opposite rankings, and 0 implies no association. The coefficient is calculated as

$$\rho = \frac{1 - 6 \sum_{i=1}^m d_i^2}{(m \cdot (m^2 - 1))} \quad (23)$$

where d_i is the difference between the ranks assigned to category i by the two methods. The magnitude of individual displacement is measured by the mean absolute rank shift, and agreement at the decision-relevant end of the ordering by the overlap of the top- k sets. Finally, to explain rather than merely describe the pattern of agreement, each category is characterised by the range of its criterion-level ranks, defined as the worst rank minus the best rank across the seven criteria. A small range denotes a balanced profile and a large range an unbalanced one.

3.8 Implementation

The AHP–TOPSIS computations were carried out in Microsoft Excel, with each stage of the method placed on a separate worksheet so that every link in the calculation chain remains auditable at formula level: geometric-mean aggregation of the pairwise comparisons, column normalisation and row averaging for the AHP weights, and vector normalisation, weighted multiplication, identification of the ideal solutions and computation of the Euclidean distances for TOPSIS. The CRITIC–TOPSIS computations were implemented in Python using the pandas, NumPy and openpyxl libraries, with each stage of the method written as a separate function so that the methodological flow and the code structure correspond one to one: EWMA construction of the decision matrix Eq. (3), min–max normalisation Eq. (12), CRITIC weighting Equations (13–16) and TOPSIS ranking Equations (17–22). The final outputs of both strands are reported as formatted spreadsheets.

3.9 Assumptions

The assumptions common to both strands are as follows: (i) all criteria are benefit-oriented and are treated as having equal measurement reliability; (ii) category performances are accepted as comparable in the form of the ratio of the realised value to the target; and (iii) the evaluation is confined to the data of June, July and August, so the findings are specific to the period and the company examined. The AHP–TOPSIS strand carries the additional assumption that expert judgements are period-independent, and the AHP weights are therefore held common across the three months; only the decision matrix changes from month to month. Since different expert combinations could produce different weight structures, the robustness of the results is audited through the internal consistency of the two experts' judgements (CR). In the CRITIC–TOPSIS strand the criteria weights are computed from the matrix under evaluation, so the chosen period and data structure shape the weights directly; this is an inherent property of data-dependent objective weighting.

4. Results

This section applies the model defined above step by step with the two weighting approaches. The first part establishes the AHP criteria weights from the aggregated pairwise comparison matrix, tests the consistency of the judgements and applies TOPSIS separately to each of the three months

and then to the EWMA-aggregated matrix. The second part implements CRITIC, compares the four scenarios on the pilot data, computes the objective weights on the three-month EWMA matrix and ranks the categories with TOPSIS. The third part compares the two rankings under equal conditions.

4.1 AHP Criteria Weights and Consistency

The pairwise comparison judgements provided by the two experts on the seven criteria were aggregated by the geometric mean of Equation (4) into a single comparison matrix, presented in Table 5. The diagonal elements of the matrix are unity, and reciprocal cells are inverses of one another; the value in a cell expresses the relative importance of the row criterion over the column criterion.

Table 5

Aggregated pairwise comparison matrix

Criteria	GOC	NS	CUS	1P	3P	SES	CVR
Gross Order Count (GOC)	1.000	3.464	4.472	5.657	4.899	6.000	5.477
Net Sales (NS)	0.289	1.000	3.464	5.657	4.000	6.000	5.477
Customer Count (CUS)	0.224	0.289	1.000	2.828	0.816	1.414	1.414
First-Party Net Sales (1P)	0.177	0.177	0.354	1.000	0.258	0.707	0.707
Third-Party Net Sales (3P)	0.204	0.250	1.225	3.873	1.000	2.449	2.449
Session Count (SES)	0.167	0.167	0.707	1.414	0.408	1.000	1.414
Conversion Rate (CVR)	0.183	0.183	0.707	1.414	0.408	0.707	1.000

Each column of the aggregated matrix was divided by its column total Eq. (6) and the arithmetic mean of each row of the normalised matrix was taken to obtain the criteria weights, or priority vector Eq. (7). The resulting weights are reported in descending order in Table 6.

Table 6

AHP criteria weights

Criteria	Weight	Percentage
Gross Order Count	0.3936	39.36%
Net Sales	0.2617	26.17%
Third-Party Net Sales	0.1102	11.02%
Customer Count	0.0844	8.44%
Session Count	0.0567	5.67%
Conversion Rate	0.0525	5.25%
First-Party Net Sales	0.0409	4.09%
Total	1.0000	100.00%

Examining the weight structure, gross order count carries by far the highest importance at 39.36%, followed by net sales at 26.17%; together the two criteria account for roughly two-thirds of the total weight. This reflects the fact that the experts, in judging category performance, give priority to the size of demand (order volume) and to the contribution to revenue. Third-party net sales (11.02%) and customer count (8.44%) receive intermediate weights, while session count, conversion rate and first-party net sales remain comparatively low; the lowest weight (4.09%) belongs to first-party net sales.

The internal consistency of the expert judgements was tested with Equations (8)–(10). The principal eigenvalue of the aggregated comparison matrix was computed as $\lambda_{\max} = 7.2929$, giving a consistency index of $CI = 0.0488$; with $RI = 1.32$ for a seven-criteria matrix, the consistency ratio is $CR = 0.0370$. Since this value lies below the acceptability threshold of 0.10, the expert judgements are consistent, and the weight vector can be used with confidence (Table 7).

Table 7

AHP consistency indicators

Indicator	Value
Principal eigenvalue ($\lambda_{\max}$)	7.2929
Consistency Index (CI)	0.0488
Random Index (RI, $n = 7$)	1.32
Consistency Ratio (CR)	0.0370
Assessment	Consistent ($CR < 0.10$)

4.2 AHP–TOPSIS Rankings Based on Monthly Data

With the AHP weights held fixed, TOPSIS was applied separately to June, July and August. The decision matrix of each month consists solely of that month's realised-to-target ratios, so the weights remain constant while the matrix differs from month to month. No cap was applied to outliers in this strand: values realised well above target were preserved as they stood, since they reflect performance that occurred in the period concerned. The reason for examining the three months separately rather than reducing them to a single composite period is that category rankings can change from month to month, and that change is itself a finding worth examining. Table 8 reports the ideal solution vectors obtained for all months equally weighted, for the EWMA-aggregated AHP run of Section 3.3 and for the CRITIC run of Section 3.5.

Table 8

Positive-ideal (A^+) and negative-ideal (A^-) solutions for the three TOPSIS runs

Run	Solution	NS	GOC	CUS	SES	CVR	1P	3P
AHP – Equally Weighted	A^+	0.0635	0.0979	0.0219	0.0178	0.0103	0.0202	0.0274
AHP – Equally Weighted	A^-	0.0302	0.0444	0.0097	0.0065	0.0071	0.0034	0.0093
AHP – EWMA	A^+	0.0626	0.0942	0.0190	0.0137	0.0117	0.0163	0.0264
AHP – EWMA	A^-	0.0348	0.0547	0.0118	0.0073	0.0064	0.0008	0.0138
CRITIC – EWMA	A^+	0.0224	0.0235	0.0252	0.0465	0.0524	0.0587	0.0329
CRITIC – EWMA	A^-	0.0124	0.0136	0.0156	0.0248	0.0286	0.0030	0.0171

For every category the Euclidean distances to the positive-ideal and negative-ideal solutions (S^+ and S^-) were computed with Equations (20)–(21) and the relative closeness coefficient C_i derived from them with Eq. (22). The closeness coefficients and ranks of all 32 categories for the three months are presented together in Table 9, ordered by the June ranking so that the movement of each category between months can be read directly.

Table 9

Monthly AHP–TOPSIS closeness coefficients and ranks

Category	June*		July		August	
	C_i	Rank	C_i	Rank	C_i	Rank
MDA	0.8302	1	0.5733	6	0.1154	31
Watches, Eyewear & Accessories	0.7632	2	0.2251	23	0.3498	19
Outdoor & Marine	0.7256	3	0.3658	18	0.1520	29
Footwear & Bags	0.7057	4	0.0513	32	0.4022	16
Gold Store	0.6248	5	0.6749	4	0.9223	1
Furniture	0.5386	6	0.1951	24	0.4514	12
Automotive Accessories Store	0.5255	7	0.6858	3	0.5823	3
Mobile Phones	0.5101	8	0.3857	17	0.3215	21
Apparel	0.5046	9	0.1482	27	0.1028	32
Game Consoles	0.4969	10	0.8121	1	0.4276	15
Sporting Goods	0.4802	11	0.1371	28	0.2686	22
TV	0.4800	12	0.7281	2	0.6957	2
Hardware	0.4790	13	0.4638	13	0.4462	13

Table 9
Continued

Category	June*		July		August	
	C_i	Rank	C_i	Rank	C_i	Rank
Mother & Baby	0.4732	14	0.2716	21	0.2166	24
Pet Shop	0.4316	15	0.2398	22	0.5046	6
SDA	0.4244	16	0.5097	11	0.4551	9
Home Textiles & Décor	0.4165	17	0.1245	30	0.2299	23
Jewellery	0.3807	18	0.1613	26	0.5390	5
Health & Beauty	0.3798	19	0.3879	16	0.4541	10
Cosmetics	0.3587	20	0.1663	25	0.1493	30
Toys	0.3312	21	0.4168	15	0.3433	20
Tablets	0.3198	22	0.3489	19	0.4306	14
Computers	0.3182	23	0.6319	5	0.3793	18
Staple Goods	0.3140	24	0.5109	10	0.1625	27
Photo & Camera	0.3137	25	0.5202	9	0.4698	8
Food & Beverage	0.2935	26	0.5510	7	0.3803	17
Books	0.2728	27	0.4367	14	0.1610	28
Wearable Technology	0.2541	28	0.1146	31	0.2016	26
Music Products	0.2109	29	0.2926	20	0.4925	7
Glassware	0.1124	30	0.1338	29	0.2077	25
Office & Stationery	0.1110	31	0.4965	12	0.4541	11
Mobile Phone Accessories	0.0525	32	0.5279	8	0.5448	4

Categories in Table 9 are sorted in descending order based on their June performance (rank). Rankings in subsequent months reflect this baseline ordering to facilitate direct tracking of monthly rank mobility.

In June the highest relative closeness coefficient belongs to MDA ($C_i = 0.8302$), which is therefore closest to the ideal solution and the most successful category of the period, followed by watches, eyewear and accessories ($C_i = 0.7632$); mobile phone accessories occupies the bottom of the ranking ($C_i = 0.0525$). The spread of the coefficients from 0.0525 to 0.8302 shows that the categories separated markedly in terms of performance during the month. In July the top position is taken by game consoles ($C_i = 0.8121$) and in August by the gold store ($C_i = 0.9223$): the leading category changes every month, which points to substantial periodic fluctuation in category performance.

Reading the three months together, several movement patterns stand out. MDA falls from 1st in June to 6th in July and 31st in August, swinging from the very top of the ranking to close to the bottom within three months. Mobile phone accessories follow the opposite trajectory, rising from last place (32nd) in June to 8th in July and 4th in August. The gold store remains stable near the top throughout (5th, 4th and 1st), reaching the summit in August, while hardware is the most stable category of all, holding 13th place in all three months. Figure 3 traces these trajectories.

This mobility vindicates the decision to treat the three months separately: an assessment based on a single month, or on the plain average of the months, would have concealed the periodic fluctuation and created a misleading impression of stability. The effect of the weight structure on the ranking is also visible — categories exceeding their targets on the two highest-weighted criteria, gross order count and net sales, tend to rise in the overall ordering.

4.3 AHP–TOPSIS Ranking Based on EWMA-Aggregated Data

Alongside the separate monthly evaluation, the three months of data were merged into a single decision matrix by EWMA with $\lambda = 0.7$ and re-ranked with the AHP weights. This approach assigns the highest weight to the most recent month (August, 70%) while damping July (21%) and June (9%) progressively and so reduces the periodic fluctuation to a single recency-weighted ordering. The

resulting 32×7 EWMA decision matrix is reported in Appendix A. It was first vector-normalised and then multiplied by the AHP criteria weights to form the weighted normalised matrix; the ideal solution vectors are given in the second block of Table 8 and the full ranking in the left-hand block of Table 14.

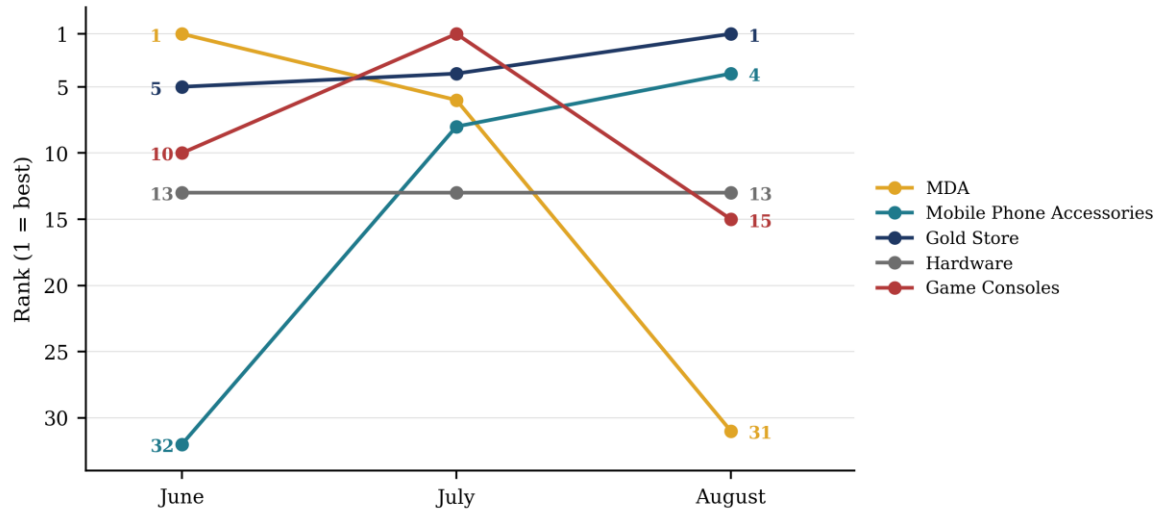


Fig. 3. Monthly rank mobility of selected categories under AHP-TOPSIS

In the EWMA-based evaluation the gold store shows the strongest performance ($C_i = 0.8539$), followed by TV ($C_i = 0.7110$) and the automotive accessories store ($C_i = 0.6264$); apparel occupies the bottom of the ordering ($C_i = 0.0982$). Merging the three months into a single EWMA matrix reduces the fluctuation observed in the monthly rankings to one recency-weighted ordering, so that categories that were consistently strong over the period gather at the top. The leading category, which changed every month in the separate monthly evaluation, becomes stably identified as the gold store, reflecting the strong performance it sustained in the most recent period.

4.4 CRITIC: Scenario Sensitivity and Criteria Weights

To observe the effect of the scenario design on the criteria weights, CRITIC weights were computed for each of the four scenarios on one week of daily data. The daily resolution sharpens sensitivity to parameter changes. The results are compared in Table 10.

Table 10

CRITIC criteria weights under the four pilot scenarios (seven-day data)

Criteria	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Net Sales	0.0718	0.0715	0.0723	0.0714
First-Party Net Sales (1P)	0.4504	0.4310	0.3915	0.3941
Third-Party Net Sales (3P)	0.0813	0.0756	0.0814	0.0788
Gross Order Count	0.0703	0.0697	0.0734	0.0722
Customer Count	0.0682	0.0689	0.0705	0.0691
Session Count	0.0876	0.0854	0.0879	0.0860
Conversion Rate	0.1704	0.1980	0.2230	0.2284
Total	1.0000	1.0000	1.0000	1.0000

Moving from Scenario 1 to Scenario 4 — that is, removing the cap and raising the smoothing factor — changes the weight structure. The weight of first-party net sales declines ($0.4504 \rightarrow 0.3941$) while the weight of conversion rate rises markedly ($0.1704 \rightarrow 0.2284$). Scenario 4, in which raw data

are combined with the high smoothing factor, therefore distributes weight more evenly across the criteria instead of concentrating it in a single criterion, and shows the greatest sensitivity to the current performance trend. On these grounds the conditions of Scenario 4 (raw data, $\lambda = 0.7$) were adopted for the final three-month evaluation.

Under the selected conditions the three months of monthly data were arranged chronologically and the EWMA formula applied with $\lambda = 0.7$, producing the single 32×7 decision matrix of Appendix A. CRITIC was then applied to that matrix: the contrast intensity σ , the total conflict $\Sigma(1 - r)$, the information quantity C and the final weight w were computed for each criteria (Table 11).

Table 11

CRITIC intermediate quantities and criteria weights (three-month EWMA matrix)

Criteria	σ_j	$\Sigma(1 - r_{jk})$	C_j	Weight w_j	Weight (%)
Conversion Rate	0.1928	5.7867	1.1158	0.2319	23.19%
Session Count	0.2294	3.9637	0.9091	0.1890	18.90%
First-Party Net Sales (1P)	0.1624	4.2986	0.6981	0.1451	14.51%
Third-Party Net Sales (3P)	0.2139	3.0376	0.6497	0.1351	13.51%
Customer Count	0.2413	2.1927	0.5291	0.1100	11.00%
Gross Order Count	0.2165	2.1491	0.4653	0.0967	9.67%
Net Sales	0.1988	2.2323	0.4437	0.0922	9.22%

The criterion carrying the highest weight over the three-month period is conversion rate ($w = 0.2319$), followed by session count ($w = 0.1890$). Both owe their prominence to high conflict values: conversion rate ($\Sigma(1 - r) = 5.7867$) and session count (3.9637) correlate weakly with the remaining criteria and therefore carry more distinctive information, which CRITIC rewards with a higher weight. Net sales ($w = 0.0922$) and gross order count ($w = 0.0967$), by contrast, receive the lowest weights because of their strong correlation with the other sales and demand indicators.

That the weight is set by the quantity of distinctive information rather than by variability alone can be read directly from Table 11. The contrast intensity of net sales ($\sigma = 0.1988$) is slightly higher than that of conversion rate ($\sigma = 0.1928$), yet its weight is roughly a quarter as large (0.0922 against 0.2319). The source of the difference is the conflict component: net sales correlates strongly with the other sales indicators and so has low conflict ($\Sigma(1 - r) = 2.2323$), whereas conversion rate is largely independent of the other criteria and has the highest conflict of all (5.7867). Likewise, customer count, which has the highest contrast intensity of any criterion ($\sigma = 0.2413$), receives only an intermediate weight (0.1100) because it moves almost in lockstep with gross order count ($r = 0.9865$).

The inter-criteria relationships underlying the weights are set out in the correlation matrix of Table 12 and visualised in Figure 4.

Table 12

Pearson correlation matrix among the criteria

	NS	1P	3P	GOC	CUS	SES	CVR
Net Sales (NS)	1.0000	0.3722	0.8915	0.8901	0.8657	0.6432	0.1051
First-Party Net Sales (1P)	0.3722	1.0000	0.1186	0.4011	0.4022	0.1577	0.2496
Third-Party Net Sales (3P)	0.8915	0.1186	1.0000	0.7067	0.6941	0.6048	-0.0534
Gross Order Count (GOC)	0.8901	0.4011	0.7067	1.0000	0.9865	0.6234	0.2431
Customer Count (CUS)	0.8657	0.4022	0.6941	0.9865	1.0000	0.5985	0.2604
Session Count (SES)	0.6432	0.1577	0.6048	0.6234	0.5985	1.0000	-0.5913
Conversion Rate (CVR)	0.1051	0.2496	-0.0534	0.2431	0.2604	-0.5913	1.0000

Table 12 shows that the criteria fall into two groups with markedly different dependence structures. The sales and demand indicators are strongly and positively related: gross order count and customer count are almost perfectly correlated ($r = 0.9865$), and net sales moves closely with

gross order count ($r = 0.8901$), third-party net sales ($r = 0.8915$) and customer count ($r = 0.8657$). First-party net sales, by contrast, is only weakly related to the other criteria (r between 0.12 and 0.40), and conversion rate is also largely independent of them, with correlations close to zero for net sales ($r = 0.1051$) and third-party net sales ($r = -0.0534$). The only notable negative relationship in the matrix is that between conversion rate and session count ($r = -0.5913$), indicating that categories attracting more sessions tend to convert a smaller share of them into orders. Figure 4 presents the same matrix as a heat map.

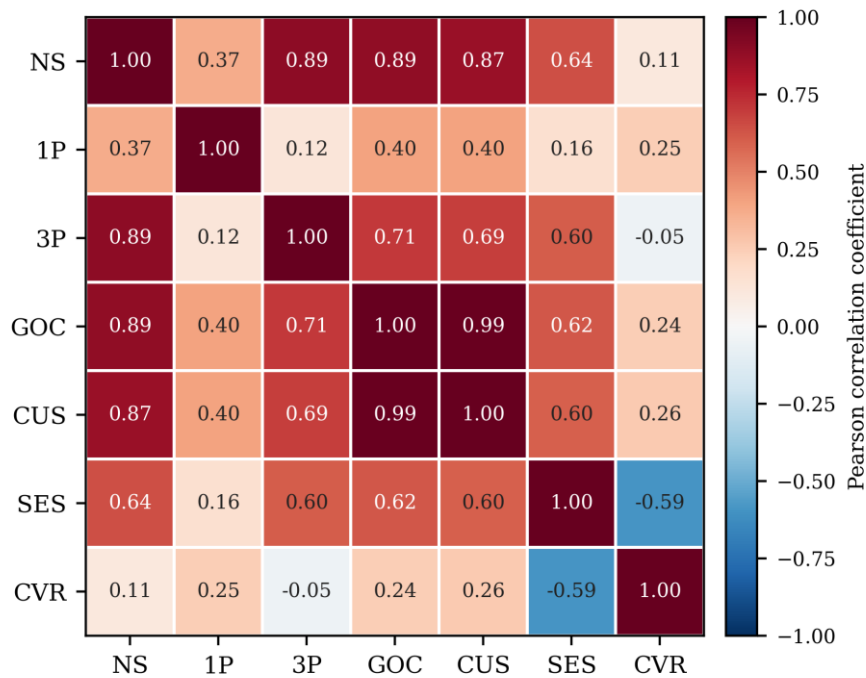


Fig. 4. Heat map of the inter-criteria Pearson correlation matrix

The almost perfect positive relationship between gross order count and customer count ($\rho = 0.99$) and the high positive values in the net sales–3P block reveal that these criteria largely duplicate one another's information. The strong negative relationship that conversion rate carries with session count ($\rho = -0.59$) confirms visually that this criterion carries information independent of and distinctive from the others. This structure explains why CRITIC assigns the highest weights to conversion rate and session count.

4.5 CRITIC–TOPSIS Ranking

Using the CRITIC weights, the decision matrix was vector-normalised and then converted into the weighted normalised matrix; the ideal solutions are reported in the third block of Table 8 and the full ranking in the right-hand block of Table 14. Over the three-month period the strongest performance belongs to game consoles ($C_i = 0.7385$), which separates itself by a clear margin from the gold store in second place ($C_i = 0.5670$) and stands out as the dominant category of the period. It is followed by the automotive accessories store (0.5108), the pet shop (0.4840) and mobile phone accessories (0.4800). At the bottom of the ordering are home textiles and décor ($C_i = 0.1798$), glassware (0.2209) and apparel (0.2360). The spread of the coefficients from 0.1798 to 0.7385 again indicates clear differentiation between categories, although a substantial number of categories cluster around 0.40, showing that performance runs close together in the middle segment.

How the closeness coefficient is formed can be made concrete at the two ends of the ranking. For game consoles in first place the distance to the positive ideal is $S^+ = 0.0210$ and to the negative

ideal $S^- = 0.0592$, giving $C_i = 0.0592 / (0.0210 + 0.0592) = 0.7385$; the category is markedly close to the ideal. For home textiles and décor in last place, $S^+ = 0.0621$ and $S^- = 0.0136$ give $C_i = 0.0136 / (0.0621 + 0.0136) = 0.1798$, placing the category far closer to the negative ideal. The heavily weighted conversion rate and session count criteria play the dominant role in determining these distances through the weighted normalised matrix.

4.6 Comparing the Subjective and Objective Rankings

For a sound comparison the two methods must be placed on an equal footing. The comparison therefore uses the EWMA-based AHP–TOPSIS results, in which the AHP strand employs the same EWMA decision matrix ($\lambda = 0.7$) as CRITIC. Both methods thus share the same criteria, the same category set, the same performance data and the same temporal treatment; the only difference between them is how the criteria weights are determined, so that any difference in the rankings reflects the weighting approach alone.

The weight vectors produced by the two methods are markedly different and largely opposed. AHP concentrates weight in a small number of criteria: gross order count (39.36%) and net sales (26.17%) alone account for roughly two-thirds of the total. CRITIC distributes weight more evenly and assigns the largest shares to conversion rate (23.19%) and session count (18.90%). Table 13 places the two weight vectors side by side.

Table 13
Comparison of the AHP and CRITIC criteria weights

Criteria	AHP	CRITIC	Difference (AHP – CRITIC)
Gross Order Count	39.36%	9.67%	+29.69
Net Sales	26.17%	9.22%	+16.95
Third-Party Net Sales	11.02%	13.51%	–2.49
Customer Count	8.44%	11.00%	–2.56
Session Count	5.67%	18.90%	–13.23
Conversion Rate	5.25%	23.19%	–17.94
First-Party Net Sales	4.09%	14.51%	–10.42

The differences show that the two approaches prioritise the criteria in almost inverted fashion. Gross order count, to which AHP gives the highest weight, is one of the lowest-weighted criteria under CRITIC (9.67%); similarly, net sales, second under AHP, receives the lowest CRITIC weight of all (9.22%). Conversely, conversion rate and session count, the two criteria CRITIC weights most heavily, are among the least important under AHP. The contrast arises because the two methods determine weight by different logics. AHP reflects the experts' judgement of company priorities and treats demand volume and revenue contribution as direct indicators of success. CRITIC sets the weight of a criterion by the quantity of distinctive information it carries — its variability and whether it correlates weakly with the other criteria. Indicators such as net sales and gross order count are highly correlated with the other sales indicators and therefore receive low CRITIC weights, while conversion rate, behaving largely independently of the rest, earns the highest weight.

Despite this pronounced contrast in the weight structures, the two rankings obtained on the identical EWMA decision matrix agree more closely than might be expected. Spearman's rank correlation between them is $\rho = 0.878$, the mean absolute rank shift is 3.7 positions, and the largest single displacement is 10 positions. Table 14 sets the two orderings side by side with the rank shift for every category, and Figure 5 displays the same information graphically.

Table 14

EWMA-based AHP–TOPSIS versus CRITIC–TOPSIS: closeness coefficients, ranks and rank shifts

Category	AHP–TOPSIS (EWMA)		CRITIC–TOPSIS		Δ rank
	C _i	Rank	C _i	Rank	
Gold Store	0.8539	1	0.5670	2	1
TV	0.7110	2	0.4556	10	8
Automotive Accessories Store	0.6264	3	0.5108	3	0
Game Consoles	0.5151	4	0.7385	1	3
Mobile Phone Accessories	0.4861	5	0.4800	5	0
SDA	0.4592	6	0.4236	15	9
Photo & Camera	0.4574	7	0.4780	6	1
Hardware	0.4460	8	0.4773	7	1
Pet Shop	0.4387	9	0.4840	4	5
Jewellery	0.4352	10	0.4691	9	1
Health & Beauty	0.4239	11	0.4287	13	2
Computers	0.4184	12	0.4145	16	4
Office & Stationery	0.4132	13	0.4075	17	4
Music Products	0.4115	14	0.4701	8	6
Furniture	0.3987	15	0.4452	12	3
Food & Beverage	0.3964	16	0.4019	19	3
Tablets	0.3924	17	0.4254	14	3
Watches, Eyewear & Accessories	0.3526	18	0.4539	11	7
Footwear & Bags	0.3442	19	0.2686	28	9
Mobile Phones	0.3376	20	0.4041	18	2
Toys	0.3324	21	0.3584	22	1
MDA	0.2440	22	0.3322	25	3
Sporting Goods	0.2300	23	0.2834	27	4
Mother & Baby	0.2196	24	0.3136	26	2
Outdoor & Marine	0.2164	25	0.2392	29	4
Staple Goods	0.2064	26	0.3678	21	5
Home Textiles & Décor	0.1877	27	0.1798	32	5
Books	0.1844	28	0.3492	23	5
Wearable Technology	0.1589	29	0.3484	24	5
Cosmetics	0.1330	30	0.4005	20	10
Glassware	0.1284	31	0.2209	31	0
Apparel	0.0982	32	0.2360	30	2

The agreement is not uniform along the ordering. In the top ten the mean absolute shift is 2.9 positions, in the middle band (ranks 11–22) 3.9 positions and in the bottom ten 4.2 positions. Four of the five leading categories are shared — the gold store, the automotive accessories store, game consoles and mobile phone accessories appear in the top five of both rankings — and nine of the top ten coincide. Agreement is therefore strongest exactly where it matters most for prioritisation decisions.

4.7 Explaining Ranking Agreement Across Methods

The pattern of agreement and divergence is not random: it follows directly from the shape of each category's criteria profile. The categories that occupy the top of both rankings perform well not on one criterion but on most of them. The gold store ranks first among all 32 categories on net sales, gross order count, customer count and session count, with a mean rank across the seven criteria of about 4. The automotive accessories store never falls below ninth place on any criterion. Because such categories are strong on every criteria at once, they remain at the top whichever criteria is weighted heavily; they are, in other words, robust to a change in the weighting approach. This is why

the gold store and the automotive accessories store stand out in both rankings even though AHP weights the sales indicators and CRITIC the behavioural ones.

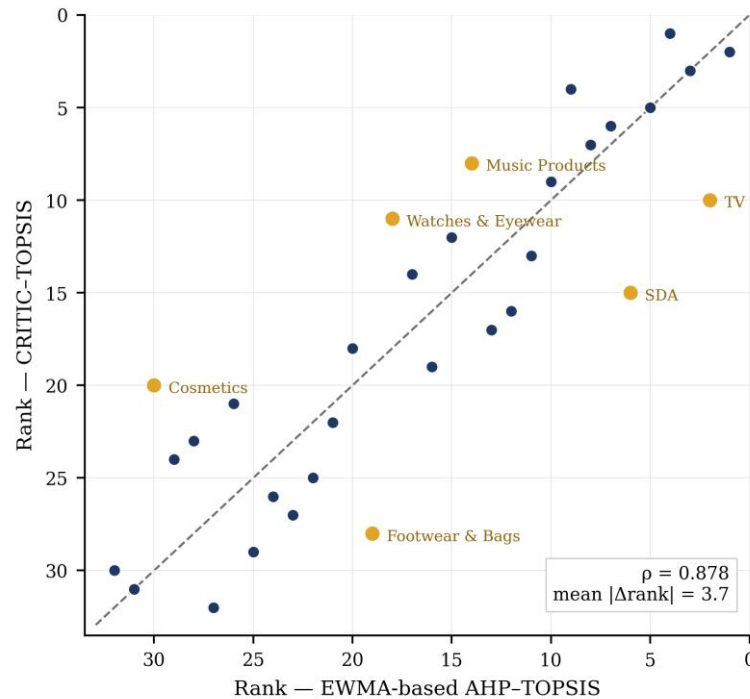


Fig. 5. Rank agreement between the EWMA-based AHP–TOPSIS and CRITIC–TOPSIS rankings. Categories shifting by six or more positions are labelled

The divergences in the middle of the ordering originate in the opposite profile: these categories perform very well on some criteria and very poorly on others. Where a category's position depends directly on which criterion receives the higher weight, its rank becomes sensitive. The clearest example is TV. On the sales indicators — net sales, gross order count and third-party net sales — TV is at or near the very top (2nd, 2nd and 1st respectively), but it falls to 24th on first-party net sales and 12th on conversion rate. AHP, which places 76.55% of the total weight on the three sales indicators, ranks the category second; CRITIC, which places 37.70% of the weight on conversion rate and first-party net sales alone, pushes it down to tenth. Cosmetics illustrates the same mechanism more sharply still: it is last (32nd) on every sales indicator yet third on first-party net sales and eleventh on conversion rate, and this stark imbalance makes it the category that moves most between the two methods (10 positions, from 30th under AHP to 20th under CRITIC). SDA and footwear and bags diverge by the same mechanism, the latter losing nine places under CRITIC because it ranks 31st on first-party net sales, a criterion CRITIC weights at 14.51% and AHP at only 4.09%. In the opposite direction, music products and watches, eyewear and accessories gain six and seven places respectively under CRITIC, both being led by strong conversion-rate performance (music products ranks first categories on that criterion). Figure 6 contrasts a balanced and an unbalanced profile directly.

Table 15 shows, however, that a wide range is a necessary but not a sufficient condition for rank sensitivity. The gold store has a range of 20, comparable to that of TV, yet moves only one position between the two methods. The reason is that it occupies the ceiling on five of the seven criteria: its single weak criterion — conversion rate, on which it ranks 21st — cannot dislodge it, because no re-weighting can move a category above first place or, given its dominance elsewhere, far below it. Rank sensitivity therefore requires two conditions to hold simultaneously: a wide dispersion of criteria-

level ranks and a mid-range overall performance level. Categories pinned at extreme of the distribution are largely immune to the weighting choice, while categories in the middle band with unbalanced profiles are the ones that move. This is a useful practical diagnostic: before acting on any single ranking, an analyst can compute the range statistic and identify at once which categories are weighting-artefacts and which are genuinely settled.

Table 15

Criteria-level rank profiles of representative categories (1 = best, 32 = worst)

Category	NS	GOC	CUS	SES	CVR	1P	3P	Range	$ \Delta \text{rank} $
Gold Store (balanced)	1	1	1	1	21	4	2	20	1
Automotive Accessories Store (balanced)	3	3	3	6	9	7	5	6	0
TV (unbalanced)	2	2	2	3	12	24	1	23	8
Cosmetics (unbalanced)	32	32	32	31	11	3	32	29	10
Footwear & Bags (unbalanced)	18	16	16	15	22	31	13	18	9

Range = worst criterion rank – best criterion rank. $|\Delta \text{rank}|$ is the displacement of the category between the EWMA-based AHP–TOPSIS and CRITIC–TOPSIS rankings.

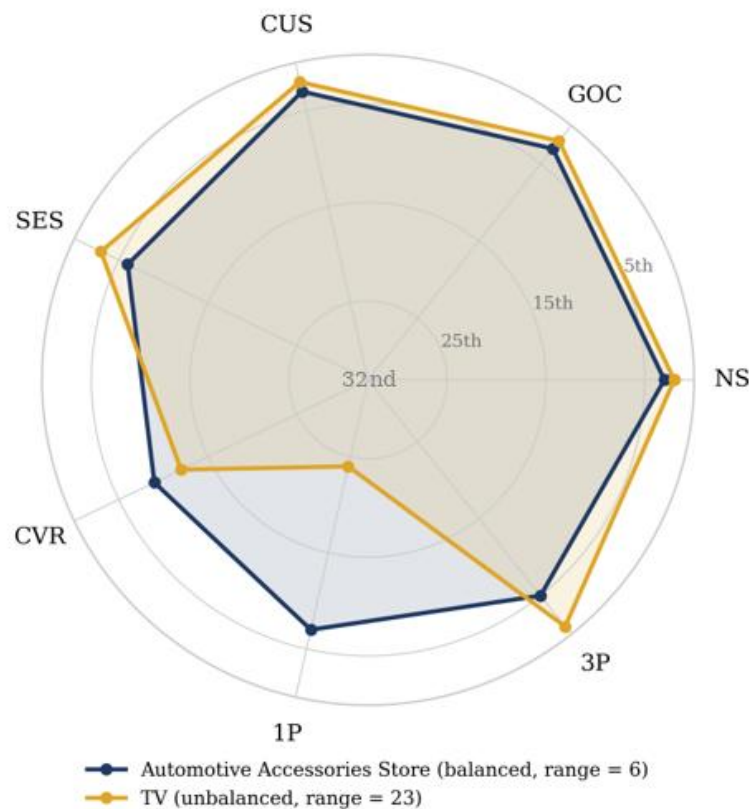


Fig. 6. Radar comparison of a balanced and an unbalanced category profile. The radial axis shows criterion-level rank, with the outer edge denoting the best rank (1st) and the centre denoting the worst rank (32nd) out of all 32 categories

4.8 Managerial Implications

Three implications follow for the company. First, the objective weight structure indicates that the most discriminating factor between categories is not absolute sales volume but the efficiency with which traffic is converted into sales, together with first-party sales performance. This suggests that

focusing resource and campaign planning on turnover magnitude alone may be inadequate, and that prioritising categories with high conversion efficiency may be the sounder course. Second, and in the opposite direction, the AHP results express what the business itself considers important and are directly interpretable for a decision maker who wishes to read category success through order volume and revenue. Presenting the two rankings together therefore offers a more complete decision support that evaluates category success from the standpoint of both strategic priority and data structure. Third, and most operationally, the categories on which the two methods agree — those with balanced profiles at the top of both orderings — can be prioritised with confidence, whereas the categories that move substantially between the methods should be examined criterion by criterion before any resource decision is taken, since their apparent position is an artefact of the weighting choice rather than a stable property of their performance.

Beyond the specific rankings, the model converts an evaluation practice that was previously dependent on a single analyst's attention into a systematic and repeatable mechanism. Once a performance report containing the criteria values arrives, the category success ranking is produced automatically according to the defined weights, and the assessment becomes an auditable decision-support process in which inter-criteria priorities are explicitly stated. In this respect the study does not eliminate the company's existing expert-driven practice but strengthens it with a quantified and automated instrument, providing a firmer basis for category prioritisation, resource allocation and performance monitoring decisions.

4.9 Limitations of the Study

Two features of the data deserve explicit mention, since they qualify the CRITIC–TOPSIS result. First, the first-party net sales column of the EWMA decision matrix contains a single extreme observation: game consoles realise 2.0200, twice its target and far above the next-highest value of 1.0251. Because Scenario 4 deliberately applies no cap, this observation alone defines the positive-ideal solution on a criterion weighted at 14.51%, which contributes materially to the large margin by which game consoles lead the CRITIC ranking ($C_i = 0.7385$ against 0.5670 for the runner-up). The scenario choice was justified on the grounds that the uncapped, recency-sensitive configuration spreads weight more evenly across the criteria, but the trade-off is that a single observation anchors one of the ideal-solution coordinates. Re-running the analysis under Scenario 3, in which the cap is applied, is the natural robustness check and should accompany any operational use of the ranking. Second, six categories record a first-party net sales ratio of exactly 1.0000, which is unlikely to arise from independent realisation and target values and probably reflects a reporting convention for categories with negligible first-party activity. Since first-party net sales carries a high CRITIC weight, verifying the provenance of these values is a prerequisite for relying on the objective ranking.

5. Conclusion

Beyond its methodological contribution, this study offers a brief observation on the Turkish e-commerce sector. The sector is characterised by continuous development and rapid change, while intense competition requires firms seeking to maintain a strong market position to monitor current trends, understand evolving consumer needs, and systematically analyse competitors. In this context, category-level analysis is particularly important because a platform's overall performance reflects the combined performance of categories with potentially different dynamics. Evaluating categories on a common, target-normalised basis enables firms to distinguish consistently strong categories from those whose performance depends disproportionately on a single criterion. This approach can complement trend monitoring and competitor analysis, thereby supporting more informed strategic decisions.

This study aimed to evaluate and rank the 32 product categories of an operating e-commerce company against seven performance criteria. The criterion values were calculated as the ratios of each category's realised performance to its predetermined target, thereby placing categories of different scales on a common basis for comparison. Criteria weights were determined independently using two approaches: the Analytic Hierarchy Process (AHP), which captures the subjective pairwise judgements of two domain experts, and the CRITIC method, which derives objective weights from the statistical characteristics of the data. In both approaches, the categories were ranked using TOPSIS on the same EWMA-aggregated decision matrix, ensuring that the weighting scheme was the only methodological difference between the two ranking procedures.

The AHP weights assign approximately two-thirds of the total weight to gross order count (39.36%) and net sales (26.17%), while the expert judgements demonstrate an acceptable level of internal consistency ($CR = 0.037$). In contrast, the CRITIC weights assign the largest shares to conversion rate (23.19%) and session count (18.90%). This difference reflects the information structure of the criteria: CRITIC assigns greater weight to criteria that exhibit substantial contrast intensity and provide information that is less redundant with other criteria. Under EWMA aggregation, the AHP–TOPSIS ranking places the gold store first ($C_i = 0.8539$), whereas the CRITIC–TOPSIS ranking places game consoles first ($C_i = 0.7385$).

The central finding is that, despite their contrasting weight structures, the two approaches produce closely aligned rankings (Spearman's $\rho = 0.878$; mean absolute rank difference = 3.7 positions), with four of the top five and nine of the top ten categories common to both rankings. This agreement is accompanied by systematic variation in rank sensitivity across categories. Categories that perform consistently well across all criteria are relatively insensitive to the weighting scheme, whereas categories with unbalanced performance profiles and intermediate overall performance can shift by as many as ten positions. In this dataset, greater rank sensitivity was associated with both a wider dispersion of criterion-level ranks and an intermediate overall performance level. This pattern provides practitioners with a useful diagnostic for identifying categories whose positions are more dependent on the choice of weighting method. In this respect, AHP addresses which criteria decision-makers consider important, whereas CRITIC identifies criteria that provide greater discriminating information within the observed data. Considering both approaches therefore offers a more comprehensive basis for decision-making than relying on either approach alone.

Several limitations should be acknowledged. First, the analysis is based on data from a single company over a three-month period, which limits the extent to which the findings can be generalised across firms and time periods. Second, the AHP weights were derived from the judgements of two experts, and a different expert panel could produce a different weighting structure. Third, the CRITIC weights were calculated from the same decision matrix used for ranking, making them dependent on the observed data: adding or removing categories could change the weights and, consequently, the ranking. Fourth, the smoothing factor was selected using a seven-day pilot analysis for the CRITIC approach and subsequently applied to the AHP approach without an equivalent sensitivity analysis. Finally, as noted in Section 3.9, an extreme observation in first-party net sales and a set of exactly unit ratios for the same criterion require verification before the objective ranking is used operationally.

These limitations point to several directions for future research. Applying the framework over longer periods and across multiple firms would help assess the generalisability of the findings and capture campaign-related effects and seasonal fluctuations. The comparison could also be extended to additional weighting methods, including objective approaches such as entropy, SWARA, and MEREC and subjective approaches such as BWM and fuzzy AHP. Similarly, incorporating ranking methods such as VIKOR, EDAS, and PROMETHEE alongside TOPSIS would help distinguish the effects

of weighting schemes from those of ranking procedures. Formal sensitivity analysis of TOPSIS results to perturbations in the weight vector, together with an examination of rank reversal when the set of categories changes, would provide further evidence of ranking robustness. Finally, hybrid weighting schemes that combine subjective and objective weights in predefined proportions offer a promising avenue for integrating expert priorities with data-driven information and developing a more comprehensive decision-support framework.

Acknowledgement

This research received no external funding. The authors would like to thank the experts at the participating e-commerce company for providing the operational data and pairwise comparison judgements used in the study.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Dalbudak, E., & Rençber, Ö. F. (2022). Çok kriterli karar verme yöntemleri üzerine literatür incelemesi. *Gaziantep Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 4(1), 1–16. <https://doi.org/10.55769/gauniibf.1068692>
- [2] Ersoy, N. (2022). Kriter ağırlıklandırma yöntemlerinin ÇKKV sonuçları üzerindeki etkisine yönelik gerçek bir hayat uygulaması. *Manas Sosyal Araştırmalar Dergisi*, 11(4), 1449–1463. <https://doi.org/10.33206/mjss.1026666>
- [3] Zoraghi, N., Amiri, M., Talebi, G., & Zowghi, M. (2013). A fuzzy MCDM model with objective and subjective weights for evaluating service quality in hotel industries. *Journal of Industrial Engineering International*, 9(1), 38. <https://doi.org/10.1186/2251-712X-9-38>
- [4] Marković, V., Stajić, L., Stević, Ž., Mitrović, G., Novarlić, B., & Radojičić, Z. (2020). A novel integrated subjective-objective MCDM model for alternative ranking in order to achieve business excellence and sustainability. *Symmetry*, 12(1), 164. <https://doi.org/10.3390/sym12010164>
- [5] Burnaz, S., & Topcu, Y. İ. (2006). A multiple-criteria decision-making approach for the evaluation of retail location. *Journal of Multi-Criteria Decision Analysis*, 14(1–3), 67–76. <https://doi.org/10.1002/mcda.401>
- [6] Tyagi, M., Kumar, P., & Kumar, D. (2014). A hybrid approach using AHP-TOPSIS for analyzing e-SCM performance. *Procedia Engineering*, 97, 2195–2203. <https://doi.org/10.1016/j.proeng.2014.12.463>
- [7] Li, R., & Sun, T. (2020). Assessing factors for designing a successful B2C e-commerce website using fuzzy AHP and TOPSIS-Grey methodology. *Symmetry*, 12(3), 363. <https://doi.org/10.3390/sym12030363>
- [8] Ishak, A., Ginting, R., & Wanli, W. (2021). Evaluation of e-commerce services quality using fuzzy AHP and TOPSIS. *IOP Conference Series: Materials Science and Engineering*, 1041, 012042. <https://doi.org/10.1088/1757-899X/1041/1/012042>
- [9] Ocampo, L., Besabella, O., Fallore, M., Guinandal, A. R., Merabueno, A., Himang, C., & Yamagishi, K. (2021). An integrated AHP-TOPSIS for evaluating online marketing strategies for the hospitality industry. *International Journal of Asian Business and Information Management*, 12(3), 489–510. <https://doi.org/10.4018/IJABIM.20210701.0a11>
- [10] Diakoulaki, D., Mavrotas, G., & Papayannakis, L. (1995). Determining objective weights in multiple criteria problems: The CRITIC method. *Computers & Operations Research*, 22(7), 763–770. [https://doi.org/10.1016/0305-0548\(94\)00059-H](https://doi.org/10.1016/0305-0548(94)00059-H)
- [11] Krishnan, A. R., Kasim, M. M., Hamid, R., & Ghazali, M. F. (2021). A modified CRITIC method to estimate the objective weights of decision criteria. *Symmetry*, 13(6), 973. <https://doi.org/10.3390/sym13060973>
- [12] Görçün, Ö. F. (2021). Evaluation of the selection of proper metro and tram vehicle for urban transportation by using a novel integrated MCDM approach. *Science Progress*, 104(1), 1–18. <https://doi.org/10.1177/0036850420950120>
- [13] Yin, J., Wang, J., Wang, C., Wang, L., & Chang, Z. (2023). CRITIC-TOPSIS based evaluation of smart community governance: A case study in China. *Sustainability*, 15(3), 1923. <https://doi.org/10.3390/su15031923>
- [14] Arısoy, İ., & İç, Y. T. (2025). Development of a new modified TOPSIS model to rank the e-commerce companies using the categoric data sets. *Electronic Commerce Research*. Advance online publication. <https://doi.org/10.1007/s10660-025-10046-8>
- [15] Chodak, G. (2024). Multi-criteria evaluation in e-commerce logistics. Springer.

- [16] Özsalman, E., Ulutaş, A., & Ünver, H. (2026). Performance evaluation of e-commerce websites based on multi-criteria decision-making methods. *Journal of Theoretical and Applied Electronic Commerce Research*, 21(6), 190. <https://doi.org/10.3390/jtaer21060190>
- [17] Hwang, C. L., & Yoon, K. (1981). *Multiple attribute decision making: Methods and applications*. Springer-Verlag. <https://doi.org/10.1007/978-3-642-48318-9>
- [18] Dolega, L., Rowe, F., & Branagan, E. (2021). Going digital? The impact of social media marketing on retail website traffic, orders and sales. *Journal of Retailing and Consumer Services*, 60, Article 102501. <https://doi.org/10.1016/j.jretconser.2021.102501>
- [19] Etro, F. (2021). Product selection in online marketplaces. *Journal of Economics & Management Strategy*, 30(3), 614–637. <https://doi.org/10.1111/jems.12428>
- [20] Etro, F. (2024). e-Commerce platforms and self-preferencing. *Journal of Economic Surveys*, 38(4), 1516–1543. <https://doi.org/10.1111/joes.12594>
- [21] Di Fatta, D., Patton, D., & Viglia, G. (2018). The determinants of conversion rates in SME e-commerce websites. *Journal of Retailing and Consumer Services*, 41, 161–168. <https://doi.org/10.1016/j.jretconser.2017.12.008>
- [22] McDowell, W. C., Wilson, R. C., & Kile, C. O., Jr. (2016). An examination of retail website design and conversion rate. *Journal of Business Research*, 69(11), 4837–4842. <https://doi.org/10.1016/j.jbusres.2016.04.040>
- [23] Perdikaki, O., Kesavan, S., & Swaminathan, J. M. (2012). Effect of traffic on sales and conversion rates of retail stores. *Manufacturing & Service Operations Management*, 14(1), 145–162. <https://doi.org/10.1287/msom.1110.0356>
- [24] Lohse, G. L., & Spiller, P. (1999). Internet retail store design: How the user interface influences traffic and sales. *Journal of Computer-Mediated Communication*, 5(2). <https://doi.org/10.1111/j.1083-6101.1999.tb00339.x>
- [25] Unal, M., & Park, Y.-H. (2023). Fewer clicks, more purchases. *Management Science*, 69(12), 7317–7334. <https://doi.org/10.1287/mnsc.2023.4716>
- [26] Saaty, T. L. (1980). *The analytic hierarchy process: Planning, priority setting, resource allocation*. McGraw-Hill. <https://doi.org/10.21236/ADA214804>
- [27] Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1), 72–101. <https://doi.org/10.2307/1412159>
- [28] Aczél, J., & Saaty, T. L. (1983). Procedures for synthesizing ratio judgements. *Journal of Mathematical Psychology*, 27(1), 93–102. [https://doi.org/10.1016/0022-2496\(83\)90028-7](https://doi.org/10.1016/0022-2496(83)90028-7)