Temporal Popularity-Based Recommender Systems for e-Commerce: A Comprehensive Evaluation

Mustafa Keskin, Enis Teper and Sinan Keçeci
Hepsiburada, Turkey

Keywords: Recommender Systems, Popularity-Based Recommendation, Temporal Dynamics, Trend Detection, e-Commerce.

Abstract: We explore popularity-based recommendation strategies for e-commerce, using a year of sales logs to evaluate three baselines: most popular, recently popular, and decay popular products. We also propose trend popular products, a novel method that captures emerging preferences by analyzing weekly sales changes. Our evaluation on a subsequent month of orders shows that approaches considering recency or time decay are more effective than simple popularity. The trend-aware method further enhances performance, demonstrating that lightweight, popularity-driven models can offer effective and clear recommendation strategies for e-commerce

1 INTRODUCTION

The rapid growth of e-commerce platforms has significantly increased the importance of effective recommendation and ranking systems. With millions of users interacting with vast product catalogs, understanding customer behavior and providing relevant suggestions has become a critical factor for improving user satisfaction, engagement, and overall sales. Traditional approaches to recommendation often rely on collaborative filtering or content-based techniques. However, these methods may struggle to capture the temporal dynamics of user interactions or the evolving popularity trends of products.

To address these challenges, recent research emphasizes the integration of popularity-based signals into recommendation pipelines. Metrics such as most popular, recent popular, and decayed popularity provide valuable insights into item attractiveness by accounting for both historical demand and temporal recency. The most popular metric highlights globally trending products, recent popularity captures short-term surges in demand, while decayed popularity balances long-term and short-term interest by applying a time-based decay function.

In the context of e-commerce, these popularity-driven signals are particularly effective because user purchasing decisions are often influenced by collective behavior and temporal patterns. For instance, seasonal demand spikes, newly launched products, or fast-fading trends can be captured more effectively by incorporating recency and decay-based measures. Leveraging such features not only enhances recommendation quality but also supports business goals such as increasing conversion rates and promoting new or relevant items.

This paper explores the application of popularity-based methods on large-scale sales data from an e-commerce platform. We first compute user-level and global metrics, including most popular, recent popular, and decayed popularity. Then, we evaluate their effectiveness in capturing user preferences and improving ranking performance. Our findings demonstrate the practical value of these approaches in building efficient, interpretable, and scalable recommendation systems for real-world e-commerce environments.

2 RELATED WORKS

Recent literature on recommender systems emphasizes the limitations of standard popularity baselines and the advantages of incorporating temporal dynamics such as recency and decay. Ji et al. (Ji et al., 2020) challenge the conventional “MostPop” baseline by showing how its effectiveness significantly improves when modified to consider item popularity relative to the user’s interaction time. They introduce RecentPop and DecayPop, both of which yield superior performance on MovieLens datasets and especially benefit

Keskin, M., Teper, E. and Keçeci, S.
Temporal Popularity-Based Recommender Systems for e-Commerce: A Comprehensive Evaluation.
DOI: 10.5220/0014286900004848

In *Proceedings of the 2nd International Conference on Advances in Electrical, Electronics, Energy, and Computer Sciences (ICEEECS 2025)*, pages 20-24
ISBN: 978-989-758-783-2

users with sparse interaction histories.

Earlier works also explore personalization of popularity signals. (Anelli et al., 2018) propose a time-aware personalized popularity approach that incorporates both item popularity among similar users and its temporal dynamics. This method performs comparably to advanced collaborative filtering approaches in top-N recommendation settings. (Balloccu et al., 2022) focus on explanation quality in recommender systems by factoring in recency, item popularity, and diversity when re-ranking explanation outputs, ultimately enhancing explanation relevance without sacrificing recommendation utility.

Other lines of research extend beyond popularity baselines to mitigate popularity bias in recommendations. (Han et al., 2024) introduce PopSI, a popularity-aware, multi-behavior framework that uses an orthogonality constraint to separate item popularity features from latent representations. This reduces bias while maintaining high recommendation accuracy on e-commerce datasets.

More broadly, time-aware decay functions are widely applied in context-aware recommender systems. (Hassan et al., 2022) integrate bias and decay strategies into collaborative filtering (e.g., MF, KNN, SLIM) to emphasize recent user actions in e-commerce contexts, reporting improved precision, recall, and MAP especially for decay-based models.

3 METHODOLOGY

3.1 Problem Setting and Data Splits

We use the last one year of transaction logs from an e-commerce platform to build time-aware popularity signals and evaluate their ability to forecast next-month orders. For each calendar month t, we construct features from a training window ending at the last day of $t-1$ and produce a ranked list of items to forecast orders in month t. Unless otherwise stated, the primary analysis uses a sliding window where the training span is the previous 12 months $[t-12, t-1]$, and the test span is the subsequent 1 month $[t, t]$. We report averages across all available monthly folds.

Let $i \in I$ denote an item and $o_{i,\tau}$ the number of orders for item i on day. τ.

3.2 Popularity Signals

3.2.1 Most Popular

Most popular captures long-horizon demand by aggregating orders over the last 12 months:

$$\mathrm{MostPopular}(i) = \sum_{\tau \in [t-12m, t-1d]} o_{i,\tau}. \tag{1}$$

This score is robust but insensitive to fast changes in demand (Ji et al., 2020).

3.2.2 Recent Popularity (RecentPop)

RecentPop emphasizes short-term surges using only the last 3 months:

$$\mathrm{RecentPop}(i) = \sum_{\tau \in [t-3m, t-1d]} o_{i,\tau}. \tag{2}$$

Compared to MostPop, this favors newly trending or seasonal items (Ji et al., 2020).

3.2.3 Decayed Popularity (DecayPop)

DecayPop balances recency and volume by exponentially down-weighting older interactions in the last 3 months:

$$\mathrm{DecayPop}(i;\lambda) = \sum_{\tau \in [t-3m, t-1d]} o_{i,\tau}\, e^{-\lambda \Delta(\tau,t)}, \tag{3}$$

where $\Delta(\tau, t)$ is the age (in days) between τ and the end of month $t-1$, and $\lambda > 0$ is a decay rate. We parameterize λ via a half-life h days:

$$\lambda = \ln(2)/h. \tag{4}$$

3.3 Trend-based Popularity Estimation (TrendPop)

We incorporated a trend-aware approach that captures short-term shifts in product demand. The method compares each product's order volume over two consecutive weekly windows. Specifically, the number of orders in the current week and the previous week are aggregated using a rolling window over daily sales. Products are then ranked by their weekly counts, and a *trend score* is computed as:

$$\mathrm{trend\_score} = \frac{\mathrm{previous\_week\_rank} - \mathrm{this\_week\_rank}}{\mathrm{this\_week\_rank}}. \tag{5}$$

This metric highlights products whose rank has improved compared to the previous week, indicating increasing popularity. To ensure robustness, we filter out products with fewer than five daily sales and retain only those with positive trend scores. Finally, the top-N trending products are selected based on the highest trend scores.

Table 1: Comparison of Popularity Methods.

Method	HitRate@10	HitRate@100	HitRate@1000	Recall@10	Recall@100	Recall@1000
MostPop	0.0657	0.1260	0.2953	0.0252	0.1413	0.1547
RecentPop	**0.0715**	0.1358	**0.3172**	**0.0294**	**0.1656**	**0.1547**
DecayPop	0.0713	**0.1374**	0.3123	0.0290	0.1568	0.1524
TrendPop	0.0036	0.0172	0.0795	0.0006	0.0046	0.0268

3.4 Ranking and Forecasting Task

Given a scoring function we rank all items in descending order and treat the task as next-month order forecasting for top-K planning:

$$\pi_s = \mathrm{argsort}_{i\in I}\big(s(i)\big). \tag{6}$$

Ground-truth for month t is the set of items ordered in t. Because these signals are global (non-personalized), evaluation is item-level rather than user-level.

3.5 Evaluation

3.5.1 Metrics

We adopt standard top-K ranking metrics:

- **Hit Rate @ K (HR@K):**

$$\mathrm{HR@}K = \frac{1}{|\mathcal{G}_t|}\sum_{i\in\mathcal{G}_t} \mathbb{1}\{i \in \pi_s^{(K)}\}, \tag{7}$$

  where $\mathcal{G}_t$ is the set of items ordered in month t.
- **Recall @ K:**

$$\mathrm{Recall@}K = \frac{\sum_{i\in\pi_s^{(K)}} o_{i,t}}{\sum_{i\in I} o_{i,t}}. \tag{8}$$

We report results for $K \in \{10, 100, 1000\}$. HR@K and Recall@K is our primary metrics.

Table 1 presents the comparison of four baseline methods—MostPop, RecentPop, DecayPop, and TrendPop—using hit rate (HR) and recall at different cutoff thresholds ($K = 10, 100, 1000$).

From the results, it is evident that all three main baselines (MostPop, RecentPop, and DecayPop) achieve modest scores at smaller cutoff values (e.g., $K = 10$), which is expected given the sparsity of user interactions and the wide diversity of product choices. Among these methods, RecentPop consistently outperforms MostPop across all metrics, confirming that recency is an important factor in modeling user demand and capturing evolving product preferences. DecayPop, which applies a temporal decay weighting to interactions, achieves the best overall performance. Its advantage becomes more pronounced at larger cutoff values (e.g., $K = 1000$), indicating that decay weighting provides a more nuanced balance between long-term popularity and short-term recency effects.

In contrast, the TrendPop method performs poorly across all metrics, with HR and recall values significantly lower than the other approaches. This suggests that relying on week-over-week changes in ranking introduces excessive volatility and fails to provide stable recommendations. Overall, the results highlight the importance of incorporating temporal dynamics into popularity-based methods, with DecayPop demonstrating the most robust performance across evaluation metrics.

3.6 User Segmentation Methodology

To better understand user behavior and their engagement with products, we segmented users into four activity-based categories according to their historical purchasing frequency:

Table 2: User Segmentation Categories.

Segment	Description
Segment 1	Low Activity Users
Segment 2	Medium Activity Users
Segment 3	High Activity Users
Segment 4	Very High Activity Users

This segmentation helps in analyzing patterns of product popularity across different user types. For example, low-activity users might be targeted with incentives to increase engagement, whereas very high-activity users could indicate loyal customers who frequently purchase popular products.

Table 3 reports the same evaluation metrics, but broken down by user activity segments. This analysis reveals notable differences in model effectiveness depending on user type.

For Segment 1 (users with few historical orders), all methods achieve relatively high HR and recall compared to other segments. This indicates that popularity-based methods are particularly effective for cold-start or low-activity users, as their preferences are less clearly defined and general popularity signals provide strong recommendations. Among the baselines, RecentPop and DecayPop outperform MostPop, showing the added value of incorporating temporal information even for less active users.

Table 3: Segment-Level Evaluation Results.

Segment	Method	HitRate@10	HitRate@100	HitRate@1000	Recall@10	Recall@100	Recall@1000
1	MostPop	0.1795	0.2074	0.2981	0.0768	0.0981	0.1743
1	RecentPop	**0.1821**	0.2097	**0.3116**	**0.0793**	**0.1003**	**0.1858**
1	DecayPop	0.1814	**0.2124**	0.3109	0.0785	0.1022	0.1838
1	TrendPop	0.0012	0.0065	0.0340	0.0003	0.0033	0.0194
2	MostPop	0.0567	0.0945	0.2177	0.0224	0.0464	0.1306
2	RecentPop	**0.0616**	0.0995	**0.2350**	**0.0263**	**0.0495**	**0.1433**
2	DecayPop	0.0599	**0.1024**	0.2312	0.0252	0.0512	0.1410
2	TrendPop	0.0097	0.0334	0.1422	0.0009	0.0064	0.0351
3	MostPop	0.0323	0.0838	0.2404	0.0114	0.0387	0.1292
3	RecentPop	**0.0387**	0.0911	**0.2611**	**0.0160**	**0.0426**	**0.1424**
3	DecayPop	0.0376	**0.0936**	0.2569	0.0158	0.0439	0.1404
3	TrendPop	0.0027	0.0123	0.0600	0.0006	0.0043	0.0256
4	MostPop	0.0295	0.1290	0.3835	0.0071	0.0380	0.1387
4	RecentPop	0.0376	0.1428	**0.4145**	0.0120	0.0435	**0.1538**
4	DecayPop	**0.0392**	**0.1479**	0.4069	**0.0125**	**0.0438**	0.1512
4	TrendPop	0.0018	0.0083	0.0431	0.0004	0.0034	0.0220

For Segments 2 and 3 (moderately active users), the advantage of RecentPop and DecayPop over MostPop becomes more evident. These users benefit from models that can better capture evolving product trends while still leveraging accumulated popularity information. Nevertheless, performance is overall lower than in Segment 1, suggesting that medium-activity users represent a more challenging group, as their preferences are neither fully new nor as consistent as those of highly active users.

Finally, for Segment 4 (highly active users), all methods perform poorly, with very low HR and recall values. This highlights a major limitation of simple popularity-based models: they fail to serve power users who likely expect more personalized recommendations. In this group, the differences between MostPop, RecentPop, and DecayPop are less pronounced, indicating that none of the baselines are sufficient to address the complexity of heavy-user behavior. TrendPop, again, performs poorly across all segments, confirming its instability and lack of practical utility.

In summary the segment-level analysis underscores that popularity-based methods especially those incorporating temporal dynamics are effective for less active users but inadequate for heavy users. This finding suggests that hybrid or personalized approaches would be necessary to achieve strong performance across the full spectrum of user types.

4 CONCLUSIONS

This paper evaluated the effectiveness of four popularity-based recommendation methods, MostPop, RecentPop, DecayPop, and TrendPop, on a large-scale e-commerce dataset. Our findings demonstrate that models incorporating temporal dynamics significantly outperform a simple long-term popularity baseline. Specifically, RecentPop and DecayPop consistently achieved higher Hit Rate and Recall scores, confirming that recent transaction data is a more powerful predictor of future product demand. DecayPop, by applying an exponential decay function, offered a slightly more robust balance between long-term popularity and short-term trends, particularly at larger recommendation list sizes.

In contrast, our proposed TrendPop model, designed to capture emerging trends by analyzing weekly rank changes, performed poorly across all metrics. This suggests that the high volatility of weekly sales data introduces significant noise, making simple rank-based trend detection an unreliable strategy for stable recommendations.

Perhaps the most critical insight comes from our user segmentation analysis. We found that popularity-based methods are highly effective for low-activity and new users, where personalized signals are sparse. However, their performance drastically diminishes for highly active "power users," who likely expect more tailored and diverse suggestions. This underscores a fundamental limitation of non-personalized approaches: while they serve as excellent and computationally efficient baselines for the cold-start problem, they are insufficient for retaining engaged customers. In summary, our work highlights that recency-aware popularity models are valuable components of a recommendation system, but they must be complemented by personalized strategies to cater to the full spectrum of user behavior.

5 FUTURE WORKS

Based on the findings of this study, several promising avenues for future research emerge. The primary focus should be on developing more sophisticated and personalized models that address the limitations of global popularity signals. A natural next step is to create hybrid systems that dynamically serve efficient DecayPop recommendations to new users while deploying personalized algorithms like collaborative filtering for established users. In a similar vein, popularity itself can be personalized by calculating it within specific user segments based on demographics or past behavior. Furthermore, the goal of identifying emerging products remains crucial; instead of simple rank changes, this could be revisited using robust time-series analysis methods like STL decomposition to reliably detect upward trends while filtering out statistical noise.

Beyond algorithmic enhancements, a second critical research avenue involves rigorous validation and optimization. This includes a comprehensive analysis of hyperparameters, such as lookback windows and decay rates, to fine-tune model performance for different product categories or market dynamics. Ultimately, the true effectiveness of any proposed model must be validated beyond offline metrics. It is essential to conduct online A/B testing to measure the real-world impact of these strategies on key business metrics like conversion rates, user retention, and overall engagement, providing definitive evidence of their value in a production environment.

ACKNOWLEDGEMENTS

This project was made possible by the individual contributions of each member of the recommendation team within Hepsiburada technology group. Also, this project would not have been possible if the technology group management of Hepsiburada had not supported and encouraged the recommendation team in innovation.

REFERENCES

Anelli, V. W., Noia, T. D., Sciascio, E. D., Ragone, A., and Trotta, J. (2018). Local popularity and time in top-n recommendation. arXiv preprint.

Balloccu, G., Boratto, L., Fenu, G., and Marras, M. (2022). Post processing recommender systems with knowledge graphs for recency, popularity, and diversity of explanations. *arXiv preprint*.

Han, Y., Xu, B., Wang, Y., and Gao, S. (2024). Towards popularity-aware recommendation: A multi-behavior enhanced framework with orthogonality constraint. *arXiv preprint*.

Hassan, A. Y., Fadel, E., and Akkari, N. (2022). Exponential decay function-based time-aware recommender system for e-commerce applications. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 13(10):602–603.

Ji, Y., Sun, A., Zhang, J., and Li, C. (2020). A re-visit of the popularity baseline in recommender systems. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '20)*, page 4 pages.