KSB-621-2025
Team Normalizers: Sahaj Kaur, Dheeraj Shetty, Fahima Alizada, Marco Capoccia
12/13/2025

Next-Day ROAS Forecasting for Data-Driven Budgeting and Bidding Decisions

Executive Summary:

This project develops a predictive framework to guide daily advertising decisions by forecasting next-day Return on Advertising Spend (ROAS) in the Amazon Sponsored Products ecosystem. Operating in a fast-paced, auction-based environment, advertisers face continual shifts in consumer behavior, platform competition, and campaign configuration. These dynamics make reactive and rule-based decision-making insufficient for effective budget pacing, bidding, and campaign activation. The analysis addresses this challenge by identifying which campaign-level and product-level factors most strongly influence next-day ROAS and determining how these patterns can be used to inform strategic adjustments.

The modeling dataset was constructed at the ASIN × Day level, combining product performance data (e.g., impressions, clicks, conversions, attributed sales) with campaign-level attributes (e.g., daily budget, bidding strategy, campaign status). A time-based partition was used to simulate real-world forecasting, ensuring that only past data informed future predictions. Exploratory analysis revealed that ROAS is volatile, nonlinear, and only weakly correlated with raw spend, highlighting the need for models that capture complex interactions among engagement, cost, and control variables.

A range of modeling approaches was evaluated. Linear models, including OLS, Ridge, and Lasso, failed to generalize and produced low explanatory power. In contrast, tree-based ensemble methods—particularly XGBoost and Random Forest—achieved substantially better accuracy and robustness over time. Partial Least Squares (PLS) regression offered the strongest linear benchmark after log-transforming the ROAS target. Ultimately, XGBoost was selected as the final model based on its superior performance across all evaluation metrics and stability diagnostics.

The results show that next-day ROAS is primarily driven by engagement quality (CTR, CVR), spend efficiency (CPC, total cost), and campaign-level settings. These factors interact in nonlinear ways, with diminishing returns evident when cost rises outpace improvements in engagement. By embedding next-day ROAS forecasts into daily workflows, advertisers can shift from reactive strategies to forward-looking, data-driven optimization—improving return on spend, reducing wasted budget, and enabling more adaptive campaign management.

Project Motivation and Client Questions:

Digital advertisers operate in highly dynamic environments where daily campaign performance fluctuates in response to changes in consumer behavior, auction competitiveness, and platform algorithms. Managing budgets, adjusting bidding strategies, and determining which campaigns to activate or pause require reliable, data-driven insights rather than reactive decision-making. To address these operational challenges, the client seeks a predictive framework that identifies the key drivers of next-day Return on Advertising Spend (ROAS) to inform strategic budget pacing, bidding optimization, and campaign management.

The central question guiding this analysis is: **Which campaign-level and product-level variables best predict next-day ROAS, and how can these insights inform budget pacing, bidding strategy, and campaign management decisions?**

Supporting analytical questions include: **How stable are ROAS prediction patterns over time when using a time-based training and testing split? Which campaign- and product-level variables most strongly influence next-day ROAS according to model feature importance and component loadings?**

This project aims to develop actionable, model-driven recommendations that enable the client to allocate spend more efficiently, optimize bidding strategies, and proactively manage campaigns based on expected performance rather than historical heuristics.

Industry Context and Background:

Retail media networks (RMNs) have transformed digital advertising by allowing brands to reach consumers directly on e-commerce platforms through first-party shopper data. Platforms such as Amazon Ads enable "closed-loop" attribution, linking ad exposure to verified sales outcomes and positioning retail media as one of the most measurable and performance-oriented advertising channels (AI Digital, 2024).

Industry spending on retail media continues to expand rapidly, with U.S. revenues projected to reach tens of billions by 2025 and Amazon maintaining a dominant share (eMarketer, 2025). As budgets shift toward these high-frequency, auction-based ecosystems, advertisers face increasing complexity. Daily campaign performance fluctuates based on bidding strategies, campaign budgets, and consumer demand volatility—conditions that make reactive or rule-based decision-making insufficient (Skai, 2025). Recent research in marketing analytics highlights the value of predictive modeling for managing this complexity. Machine learning and regression-based approaches can forecast key performance indicators, optimize resource allocation, and improve real-time decision-making in competitive digital marketplaces (ResearchGate, 2024; SCITEPRESS, 2024).

Within this context, the present project builds a predictive framework for next-day Return on Advertising Spend (ROAS) at the ASIN × Day level. By quantifying how campaign-level controls and product-level engagement metrics jointly affect performance, the analysis supports the client's need to convert high-volume, volatile advertising data into actionable insights that guide budget pacing, bidding optimization, and campaign management.

Data and Key Variables:

Data Sources & Integration:

This analysis integrates two internal Amazon Advertising datasets: (1) the Sponsored Products Product Ads Daily file, which contains product-level metrics such as impressions, clicks, cost, attributed conversions (14-day attribution window), and attributed sales (14-day), and (2) the Sponsored Products Campaigns Daily file, which provides campaign-level attributes such as campaign budget, bidding strategy, budget type, and campaign status. After standardizing column names to lowercase snake_case and parsing the report date into a proper date format, the two datasets were left-joined on **campaign_id** and **report date** (rather than coded names such as *campaignid* or *reportdate*). This ensured that all ASIN×Day (Amazon Standard Identification Number by date) observations were retained, while campaign-level settings were inherited for each matched day. From the merged dataset, several key performance indicators were derived: CTR (clicks ÷ impressions), CVR (conversions ÷ clicks), CPC (cost ÷ clicks), and ROAS (sales ÷ cost). Each metric required explicit checks for zero denominators, since impressions or clicks may legitimately be zero on low-traffic days. Infinite values resulting from division by zero were flagged and addressed during cleaning. Each row in the resulting table represents a single ASIN×Day observation containing both product-level outcomes and contextual campaign attributes, forming the analytical foundation for all subsequent modeling.

Cleaning, Validation, and Transformations:

Preprocessing emphasized data integrity and numerical stability. All column names were standardized using janitor, and date fields were parsed with lubridate. To prevent invalid computations, divisions by zero were replaced with safe defaults (0 or NA), and infinite values were coerced to NA. Duplicate ASIN–date pairs were inspected to verify entity integrity, and referential integrity was confirmed across the join keys. A total of 682 rows with no ad activity—defined as having zero impressions, clicks, cost, conversions, or attributed sales—were removed, as these non-contributing records often distort rate-based metrics and hinder model learning. No formal imputation or rescaling was applied at this stage, since feature normalization and encoding were performed within each modeling pipeline to maintain flexibility across algorithms.

Exploratory Data Analysis (EDA):

Exploratory analysis focused on understanding the distributional properties, relationships, and temporal behavior of the key performance metrics. Impressions, clicks, and cost were highly right-skewed, indicating a long-tail distribution typical of digital advertising environments where a small number of ASINs capture the majority of traffic and spend. CTR and CVR values were tightly concentrated, suggesting consistent engagement behavior across campaigns, whereas ROAS exhibited high variance and extreme outliers, underscoring its nonlinear and volatile nature. Correlation analysis revealed strong associations among clicks, cost, CTR, and CVR, confirming that engagement quality is a primary driver of attributed conversions and sales. In contrast, ROAS showed weak linear correlation with most variables, implying that linear models may be insufficient to capture its underlying dynamics. Temporal inspection indicated modest variation in CPC and ROAS by day of week, with limited evidence of seasonality. Overall, the EDA characterized the dataset as skewed, interaction-heavy, and nonlinear—conditions favoring tree-based ensemble models such as XGBoost and Random Forest.

Sampling and Partitioning:

To emulate real-world forecasting conditions, the unified ASIN×Day dataset was partitioned using a time-based blocked split. Seventy percent of the earliest observations were assigned to the training set, while the remaining thirty percent were reserved for testing. This approach prevents look-ahead bias by ensuring that only past data informs future predictions and supports the intended next-day ROAS forecasting structure. Split diagnostics confirmed similar feature distributions and comparable ASIN representation across subsets, ensuring representativeness and validity for model evaluation.

Modeling Results:

Modeling Approach and Evaluation Criteria

To identify the most accurate and generalizable predictor of next-day Return on Advertising Spend (ROAS), a range of supervised learning models were evaluated using a common feature set. Predictors included engagement metrics (e.g., impressions, clicks, cost, CTR, CVR), campaign-level controls (e.g., budget, bidding strategy, status), and temporal indicators (e.g., day of week, month). A 70/30 time-based split was used to preserve temporal order and emulate forward-looking deployment scenarios.

Model performance was assessed using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2). RMSE emphasized large deviations and captured sensitivity to outliers; MAE offered a more intuitive interpretation of average prediction error; R^2 quantified variance explained. Together, these metrics provided a comprehensive view of accuracy and generalization performance.

Modeling Results

Initial experiments with linear models—Ordinary Least Squares (OLS), Ridge, and Lasso—showed consistently poor performance, with R^2 values below 0.25 and high RMSEs, even when including interaction terms. This suggests that next-day ROAS is governed by nonlinear dynamics that these models cannot adequately capture. The limited improvement from interaction terms further reinforces the idea that engagement and cost variables interact in more complex, context-sensitive ways.

Non-parametric methods such as k-Nearest Neighbors and individual decision trees also failed to deliver strong results. These models proved unstable when applied to temporally split data, often overfitting to local fluctuations without capturing generalizable patterns. Their lack of robustness and high sensitivity to feature scaling and noise made them unsuitable for reliable next-day ROAS forecasting.

Tree-based ensemble models delivered considerably better predictive power. Random Forest achieved an RMSE of 24.98, MAE of 1.55, and R^2 of 0.84. The model benefitted from bootstrap aggregation, which reduced variance while preserving high feature interpretability. Cost, CTR, and campaign budget consistently emerged as top predictors, aligning with known drivers of spend efficiency.

XGBoost outperformed all other models, achieving an RMSE of 18.21, MAE of 1.91, and R^2 of 0.85. Its gradient boosting structure—with regularization and sequential error correction—allowed it to capture nonlinearities, diminishing returns, and subtle interaction effects more effectively than Random Forest. Among linear approaches, Partial Least Squares (PLS) with a $\log_{1+}$(ROAS) transformation emerged as the best-performing alternative, achieving an RMSE of 0.55, MAE of 0.39, and R^2 of 0.73 in log-space. Its strength lay in handling multicollinearity and extracting stable latent factors that aligned with ROAS variance.

Model Assumption and Stability Checks

Each of the top-performing models was evaluated for residual behavior, feature importance consistency, and temporal generalization. XGBoost underwent robust hyperparameter tuning via cross-validation and early stopping, with regularization terms to guard against overfitting. Residual analysis showed no systematic bias or heteroskedasticity, and the prediction errors remained stable across the forecast window. Feature rankings were consistent across tuning configurations, with cost, clicks, and CTR repeatedly emerging as dominant signals.

Random Forest offered similar conclusions. Although slightly less accurate than XGBoost, it demonstrated minimal train–test drift and strong residual stability. Its performance benefitted from ensemble averaging, which suppressed overfitting and made its outputs more resilient to daily campaign volatility. Feature importances aligned closely with those identified by XGBoost, providing convergent validation of model structure and inference.

PLS, while linear, performed reliably after the log transformation. The use of 29 cross-validated components captured relevant variance while controlling for overfitting. Residual plots indicated approximate homoscedasticity and a lack of structural bias. Although the model lacked the flexibility to uncover thresholds or nonlinear breakpoints, it served as a useful and interpretable benchmark. Taken together, the diagnostic results confirm that XGBoost offers the strongest combination of accuracy, generalization, and practical stability for modeling ROAS in this setting.

Key Takeaways

Modeling confirms that next-day ROAS is shaped by three core factors: engagement quality (CTR, CVR), spend efficiency (cost, CPC), and campaign controls (budget, bidding strategy, status). Both XGBoost and Random Forest consistently identified CTR, cost, and campaign budget as top predictors, with nonlinear effects—returns taper when cost outpaces engagement. This underscores the value of predictive, adaptive decision-making over static rules.

Budget Allocation
ASINs with strong engagement metrics (CTR, CVR) consistently yielded higher ROAS. Models ranked CTR and cost among the most important drivers. Budgets should be focused on these high-efficiency ASINs, while those with poor performance—high CPC or low CTR—should see reduced spend. Forecasted ROAS can support smarter pacing and budget distribution.

Bidding Strategy
XGBoost showed that CPC and ROAS follow a diminishing return curve: higher bids don't always lead

to better outcomes. Predictive ROAS allows for more strategic bidding—raising bids when returns are likely and pulling back when efficiency drops—helping to control spend volatility and improve ROI.

Campaign Activation
Although not a top predictor globally, campaign status interacted with other variables. Campaigns running during low-engagement periods tended to underperform. Tying activation to predicted ROAS can reduce waste and ensure campaigns run during high-return windows.

Strategic Recommendation

Integrating predictive modeling into the daily optimization process offers clear value. XGBoost provides the strongest foundation for deployment, with Random Forest as a reliable secondary model. Used together, they enable proactive budget, bid, and activation decisions—shifting from reactive management to a forward-looking, data-driven approach that enhances return on advertising spend.

Limitations and Risks

Model results should be interpreted with several caveats. ROAS outcomes are based on a 14-day attribution window, which can blur day-to-day causality and overstate short-term effects. Because campaign controls (budgets, bids, and status) may change multiple times within a day while data are aggregated daily, some residual confounding likely remains. A subset of ASIN×Day records also showed sparse activity, creating noisy rate metrics (CTR, CVR, ROAS) despite filtering safeguards; applying minimum volume thresholds or hierarchical pooling could improve stability. Finally, seasonal and promotional fluctuations may shift performance regimes that a single time split cannot capture. Future iterations should adopt rolling or time-series cross-validation to assess robustness under changing conditions.

Conclusion & Next Steps

The analysis found that next-day ROAS is most strongly influenced by a combination of engagement quality (CTR, CVR), cost efficiency (CPC, total spend), and campaign-level controls (budget, bidding strategy). These predictors interacted in nonlinear ways that were best captured by tree-based models. XGBoost, in particular, delivered the highest predictive accuracy and stability, making it well-suited for deployment in real-world forecasting and budget planning. Feature importance patterns remained consistent across validation runs, reinforcing the reliability of key variables.

Time-based testing confirmed that model patterns generalized well across unseen days, with minimal overfitting or drift. Among linear approaches, Partial Least Squares (PLS) regression provided the most stable alternative, effectively reducing skew and multicollinearity, though its predictive power lagged behind ensemble methods. Together, the models confirmed that ROAS behavior is not easily captured by static heuristics—adaptive, data-driven methods are required to forecast campaign outcomes accurately.

Moving forward, integrating these models into campaign workflows can support dynamic budget allocation, bid optimization, and smarter activation strategies. To enhance long-term performance, future iterations should incorporate longer timeframes, lagged variables, and uplift modeling to better isolate causality and optimize for incremental return. These steps would allow the modeling system to evolve into a robust, forward-looking decision tool for marketing operations.

References

AI Digital. (2025, March 24). 2025 media trends report: The rise of retail media and walled gardens. AI Digital. https://www.aidigital.com/blog/media-trends-2025

Cramer-Flood, E. (2025, May 29). Worldwide retail media ad spending 2025. eMarketer. https://www.emarketer.com/content/retail-media-ad-spending-forecast-trends-h2-2025

Skai & Path to Purchase Institute. (2025). 2025 state of retail media report. Skai. https://skai.io/wp-content/uploads/2025/02/SORM-Report_2025-1-1.pdf

ADTelligent. (2025, June 19). Retail media market outlook 2025: Key data and growth patterns. ADTelligent. https://adtelligent.com/blog/retail-media-market-outlook

Affiliate Summit. (2025, October 24). How AI is transforming retail media in 2025. Affiliate Summit Blog. https://www.affiliatesummit.com/blogs/how-ai-is-transforming-retail-media-in-2025

Jha, A., et al. (2024). Machine learning–based optimization of e-commerce advertising campaigns. In Proceedings of the 11th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2024). SCITEPRESS. https://www.scitepress.org/Papers/2024/124567/124567.pdf

Apaza-Cáceres, J. A., Vilca Paja, L. A., Pillihuaman Huachaca, A., & Pedraza Choque, D. A. (2026). Machine learning applications in digital advertising performance optimization: A systematic literature review. International Journal of Data and Network Science, 10(1), 273–290. https://growingscience.com/beta/ijds/7842-machine-learning-applications-in-digital-advertising-performance-optimization-a-systematic-literature-review.html

Herhausen, D., Klein, J. F., Dos Santos, M. A. O., & Stummer, C. (2024). Machine learning in marketing: Recent progress and future directions. Journal of Business Research, 171, 114117. https://doi.org/10.1016/j.jbusres.2023.114117

Oldenburg, F., Han, Q., & Kaiser, M. (2023). Interpretable deep learning for forecasting online advertising costs: Insights from the competitive bidding landscape. arXiv preprint arXiv:2302.05762. https://arxiv.org/abs/2302.05762

Swetha, K. (2025). AI-based advertisement optimization and performance analytics. Asian Journal of Applied Science and Technology (AJAST), 9(1), 112–118. https://ajast.net/data/uploads/52192.pdf

Appendix

EDA Visualizations

Figure 1: Correlation Matrix of Key Amazon Ads Metrics

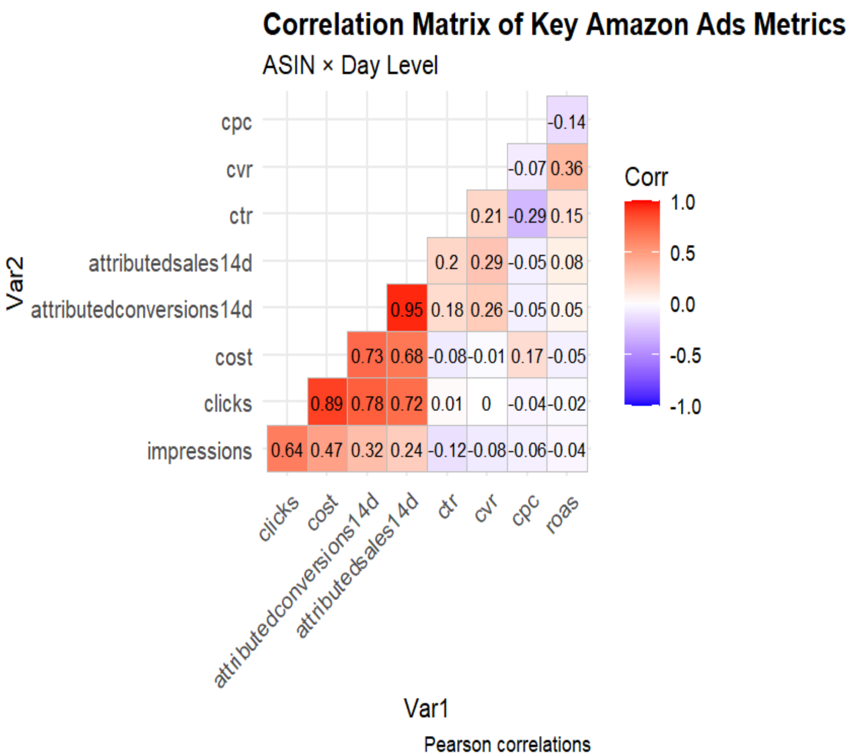

Figure 2: Daily Trends in Clicks, Cost, Impressions, Sales

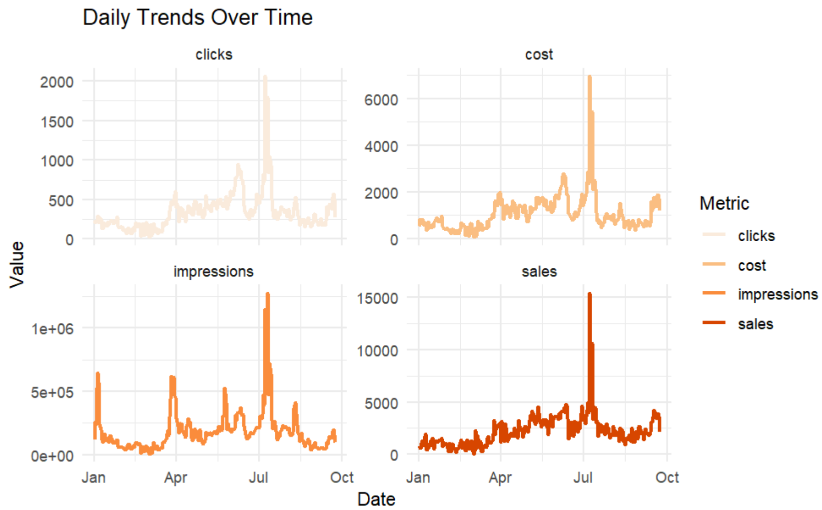

Figure 3: Weekday Effects on Key Performance Indicators(KPIs)

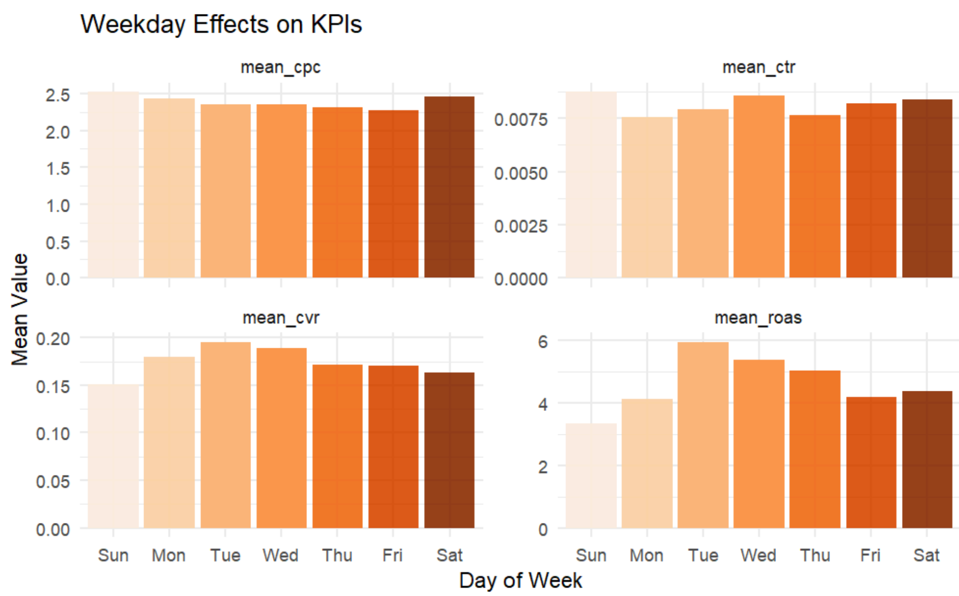

Table 1: Summary Stats Train vs Test Sets

Metric	Train – Mean	Train – SD	Test – Mean	Test – SD
Impressions	**5,130**	**14,257**	**4,626**	**20,894**
Cost ($)	**30.2**	**65.5**	**25.9**	**71.4**
CPC ($)	**2.49**	**2.37**	**2.21**	**2.18**
Attributed Conversions	**2.1**	**4.9**	**2.3**	**6.8**
Attributed Sales ($)	**58.9**	**140.0**	**61.5**	**157.0**
CVR	**0.17**	**0.35**	**0.18**	**0.36**
ROAS	**4.5**	**23.5**	**4.7**	**34.9**

Random Forest Tuning Parameters:

- **mtry** → number of variables tried at each split
- **nodesize** → minimum observations per terminal node

Table 2: Random Forest Hyperparameter Tuning Results

mtry	nodesize	RMSE ↓	R^2 ↑
1	5	30.24	0.74
3	5	24.24	0.85
7	5	23.33	0.85
1	10	30.43	0.70

3	10	24.49	0.85
7	**10**	**23.02**	**0.86 ← Best**

Key takeaways

- Increasing **mtry** substantially improves performance
- **nodesize = 10** slightly improves generalization (lower RMSE)
- Performance stabilizes → **model is robust**, not overfit
- Best RF is **not arbitrary** — it is empirically selected

Random Forest Tuning and Selection

To address model tuning and robustness, we evaluated multiple Random Forest configurations by varying the number of predictors sampled at each split (mtry $\in \{1, 3, 7\}$) and terminal node size (nodesize $\in \{5, 10\}$), while holding the number of trees fixed at 300.

Model performance improved materially as mtry increased, indicating that ROAS prediction benefits from allowing more feature interactions at each split. Larger node sizes marginally improved test‑set RMSE, suggesting better generalization.

The best performing configuration (mtry = 7, nodesize = 10) achieved an RMSE of 23.02 and an R^2 of 0.86 on the hold-out test set. Performance differences across nearby configurations were modest, indicating that Random Forest predictions are stable and not overly sensitive to tuning choices.

Gradient Boosted Model

Table 3: Comparison of XGBoost Runs Across Hyperparameter Settings

Run	eta	max_depth	child_weight	RMSE	R^2
1	0.01	3	1	26.3014	0.6960
2	0.05	3	1	28.1500	0.5708
3	0.10	3	1	28.0330	0.5719
4	0.01	4	1	27.4331	0.6209
5	0.05	4	1	28.8038	0.5225

6	0.10	4	1	28.6332	0.5575
7	0.01	6	1	28.6107	0.5373
8	0.05	6	1	27.4202	0.6309
9	0.10	6	1	30.3303	0.3603
10	0.01	3	5	20.6698	0.8413
11	0.05	3	5	18.4478	0.8467
12	**0.10**	**3**	**5**	**18.0303**	**0.8803**
13	0.01	4	5	21.8267	0.8365
14	0.05	4	5	18.4115	0.8490
15	0.10	4	5	18.5021	0.8539
16	0.01	6	5	22.4273	0.8307
17	0.05	6	5	18.2030	0.8722
18	0.10	6	5	18.5177	0.8399

Tuning & Parameters

max_depth:
Controls how deep each tree in the model can grow. Deeper trees can capture more complex patterns but may lead to overfitting.
→ *Tuned to test model performance with increasing complexity (values tried: 3, 4, 6).*

child_weight:
Minimum sum of instance weights (hessian) needed in a child node. Higher values make the model more conservative.
→ *Used to control overfitting by testing values of 1 (default) and 5 (more regularization).*

eta (learning rate):
Shrinks the contribution of each tree to slow down the learning process. Smaller values improve generalization but require more trees.
→ *Explored at 0.01, 0.05, and 0.10 to balance speed and accuracy.*

Tuning results indicate that model performance improved significantly with higher min_child_weight values, particularly when combined with moderate tree depth and learning rate. The best-performing configuration was Run 12 (eta=0.10, max_depth=3, min_child_weight=5), achieving the lowest RMSE (18.03) and highest R^2 (0.88). This suggests that regularization via child weight played a key role in controlling overfitting, while a shallower tree structure and faster learning rate enabled the model to generalize more effectively.

Figure 5 & 6: Feature Importance Rankings from Final XGBoost Model