Modeling Board Game Success: A Regression Study

A Machine Learning Project for University Course

Group Members: Abdullah Emre Esen, Osman Sefa Özdemir, Ali Köçer

Motivation: Understanding Board Game Success

1

Predictive Power

Why do some board games become wildly popular while others fade? Predicting success offers insights.

2

Data-Driven Design

Move beyond intuition. Utilize data to inform game design and publishing decisions.

3

Market Insights

Rapid growth in board game industry: market valued at $17.3 billion by 2028

Problem Statement: Beyond Intuition

Predicting Average Ratings

We aim to predict the average rating of board games using easily accessible game characteristics.

Limitations of Intuition

- Subjective and open to bias
- Doesn't scale with dataset size
- Misses complex, non-linear relationships

Value of Explainable Models

Understanding the contribution of features, not only the predictions, is important for useful insights.

Research Questions & Hypotheses

→ **RQ1: Predictive Accuracy**

Can regression models accurately predict average board game ratings?

→ **H1: Improved Performance**

Advanced models (Random Forest, XGBoost) will outperform Linear Regression.

→ **RQ2: Feature Impact**

Which game mechanics and themes significantly influence ratings?

→ **H2: Non-Linear Effects**

Mechanics and themes will show non-linear relationships with ratings.

→ **RQ3: Explainability**

How stable are predictions and feature importances across different data splits?

→ **H3: Temporal Drift**

Models trained on older data may generalize poorly to newer games without adaptation.

Dataset Overview: BoardGameGeek

Source

Kaggle - BoardGameGeek dataset (GAMES, MECHANICS, THEMES).

Scope

• **GAMES.csv** (~22k games): General features in games (e.g Play Time, Year, Min-Max Players).

• **MECHANICS.csv**: many-to-many mapping between games and mechanics (e.g., Dice Rolling, Worker Placement).

• **THEMES.csv**: many-to-many mapping between games and themes (e.g., Fantasy, Economic).

Target Variable

Average Rating: Continuous variable, reflecting community sentiment.

Ethical Considerations

This data is public, but we need to know that users might not always be objective.

```
GAMES
    BGGId           BoardGameGeek game ID
    Name        Name of game
    Description     Description, stripped of punctuation and lemmatized
    YearPublished       First year game published
    GameWeight      Game difficulty/complexity
    AvgRating       Average user rating for game
    BayesAvgRating      Bayes weighted average for game (x # of average reviews applied)
    StdDev          Standard deviation of Bayes Avg
    MinPlayers      Minimum number of players
    MaxPlayers      Maximun number of players
    ComAgeRec       Community's recommended age minimum
    LanguageEase    Language requirement
    BestPlayers     Community voted best player count
    GoodPlayers     List of community voted good plater counts
    NumOwned    Number of users who own this game
    NumWant         Number of users who want this game
    NumWish         Number of users who wishlisted this game
    NumWeightVotes      ? Unknown
    MfgPlayTime     Manufacturer Stated Play Time
    ComMinPlaytime      Community minimum play time
    ComMaxPlaytime      Community maximum play time
    MfgAgeRec       Manufacturer Age Recommendation
    NumUserRatings      Number of user ratings
    NumComments     Number of user comments
    NumAlternates       Number of alternate versions
    NumExpansions       Number of expansions
    NumImplementations  Number of implementations
    IsReimplementation  Binary - Is this listing a reimplementation?
    Family          Game family
    Kickstarted     Binary - Is this a kickstarter?
    ImagePath       Image http:// path
    Rank:boardgame      Rank for boardgames overall
    Rank:strategygames  Rank in strategy games
    Rank:abstracts      Rank in abstracts
    Rank:familygames Rank in family games
    Rank:thematic       Rank in thematic
    Rank:cgs    Rank in card games
    Rank:wargames       Rank in war games
    Rank:partygames     Rank in party games
    Rank:childrensgames  Rank in children's games
```

Feature Description: Building Blocks

Numeric Features

- year_published
- min_players, max_players
- min_play_time, max_play_time
- log_play_time
- players_range

Multi-label Categories

- mechanics (e.g., "Dice Rolling", "Area Control")
- themes (e.g., "Fantasy", "Sci-Fi")

Popularity Control

- users_rated: Captures game visibility and engagement.
- Crucial for understanding rating legitimacy.

Data Processing & Feature Engineering

01

Data prepare

Merged games.csv, mechanics.csv, and themes.csv using the unique BGGId.

Rows with missing target values (AvgRating) or essential numeric attributes were removed to ensure data integrity.

02

Data Cleaning

Filtered entries with logic errors (e.g., MaxPlayers < MinPlayers or MinPlayers < 1).

Missing values in categorical features (mechanics and themes) were filled with zero to indicate absence.

03

Dimensionality Reduction

To manage high dimensionality, only the **top 80 most frequent mechanics** and **top 80 themes** were retained.

Less frequent attributes were discarded to reduce noise and improve model focus.

04

Derived Features

players_range: Difference between maximum and minimum players.

is_solo_supported: Binary indicator for single-player support.

log_play_time: Log-transformed playtime to handle skewness.

play_time_per_player: Ratio of playtime to max players.

age_norm: Ordinal encoding for age recommendations.

05

Outlier Handling

Applied clipping to prevent extreme values from dominating the learning process:

- **Playtime**: Clipped to the range **[5, 600] minutes**.
- **Max Players**: Capped at **50 players**.

06

Data split

The processed dataset was split into **80% training** and **20% test** sets.

Final features and labels were stored in a compressed NumPy archive (.npz) to ensure consistency across experiments.

Modeling Approach: From Baseline to Advanced

1

Linear Regression (OLS)

Baseline model for interpretability and comparison.

Provides initial understanding of linear relationships.

2

Random Forest

Ensemble method, robust to outliers and captures non-linearities.

Handles feature interactions effectively.

3

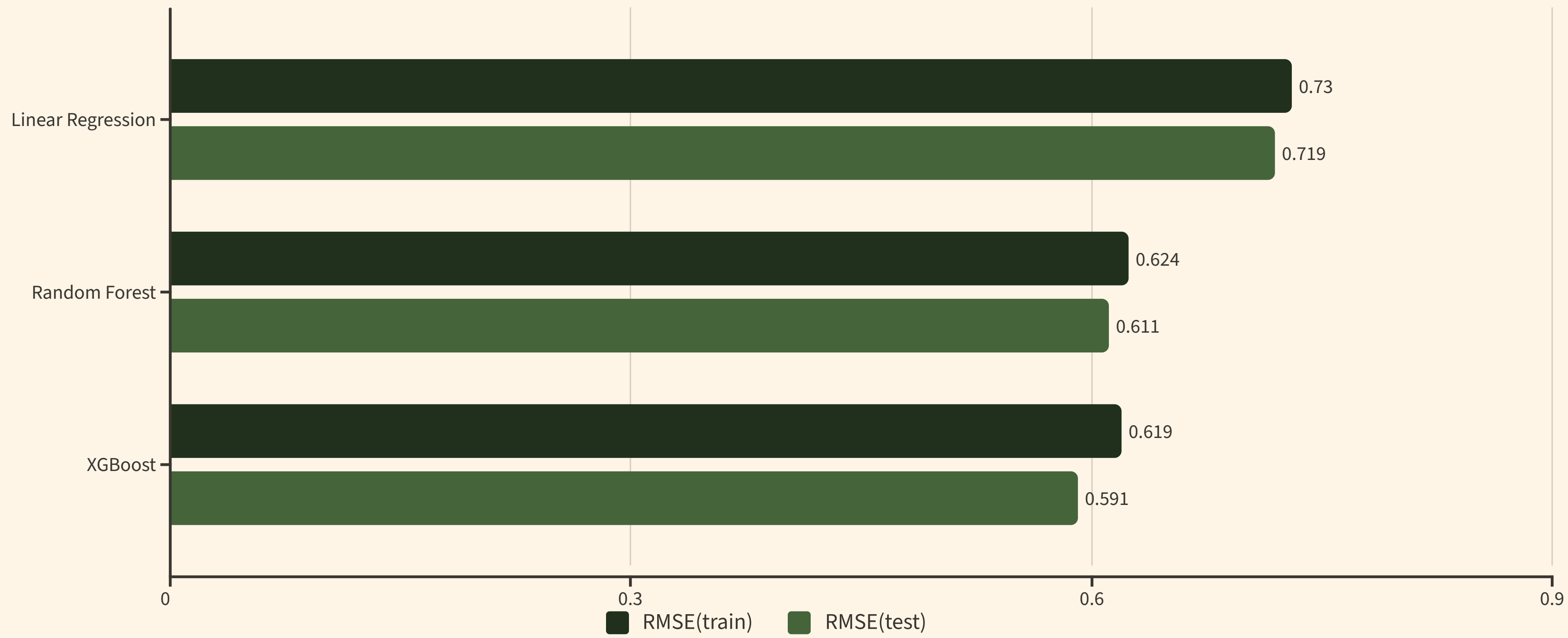
XGBoost

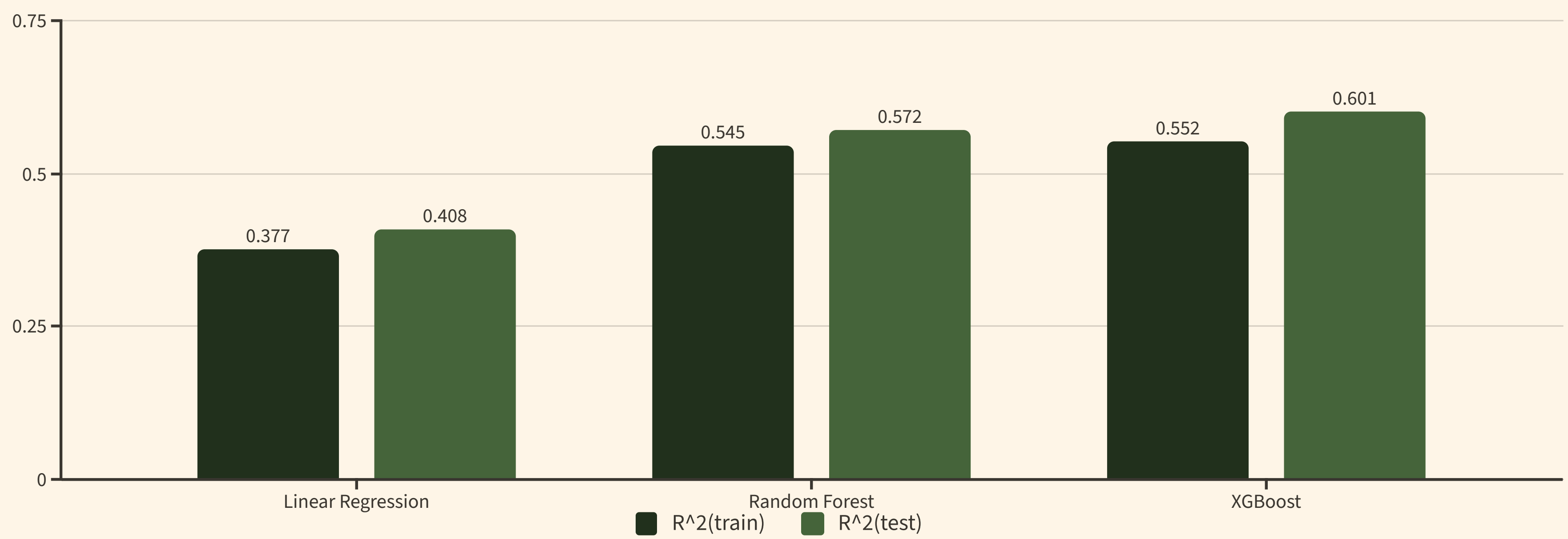
Gradient boosting: state-of-the-art performance, handles complex data patterns.

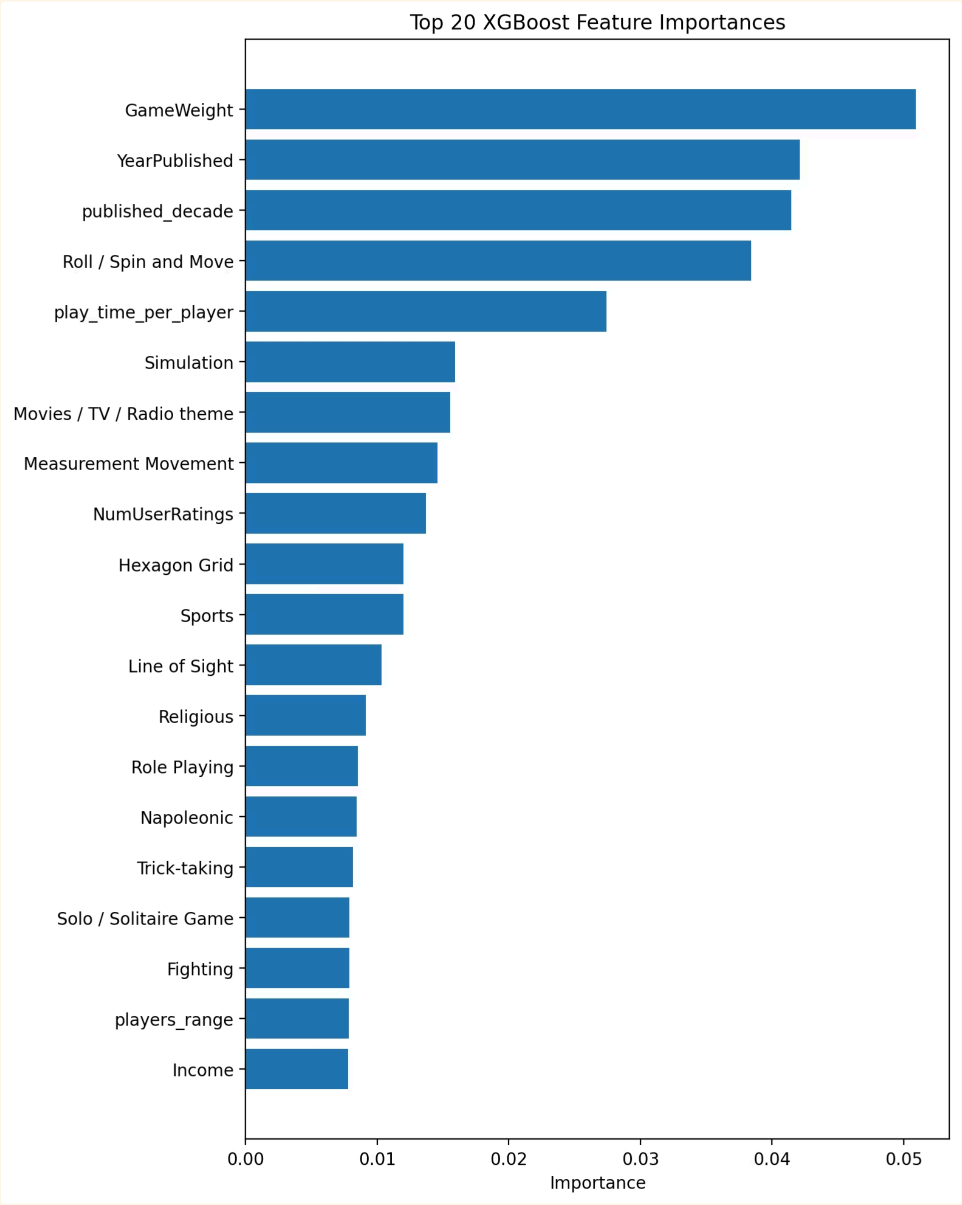
Often achieves highest predictive accuracy.

Model Performance Overview

Train Shape :(17540 , 176) Test Shape:(4385, 176)

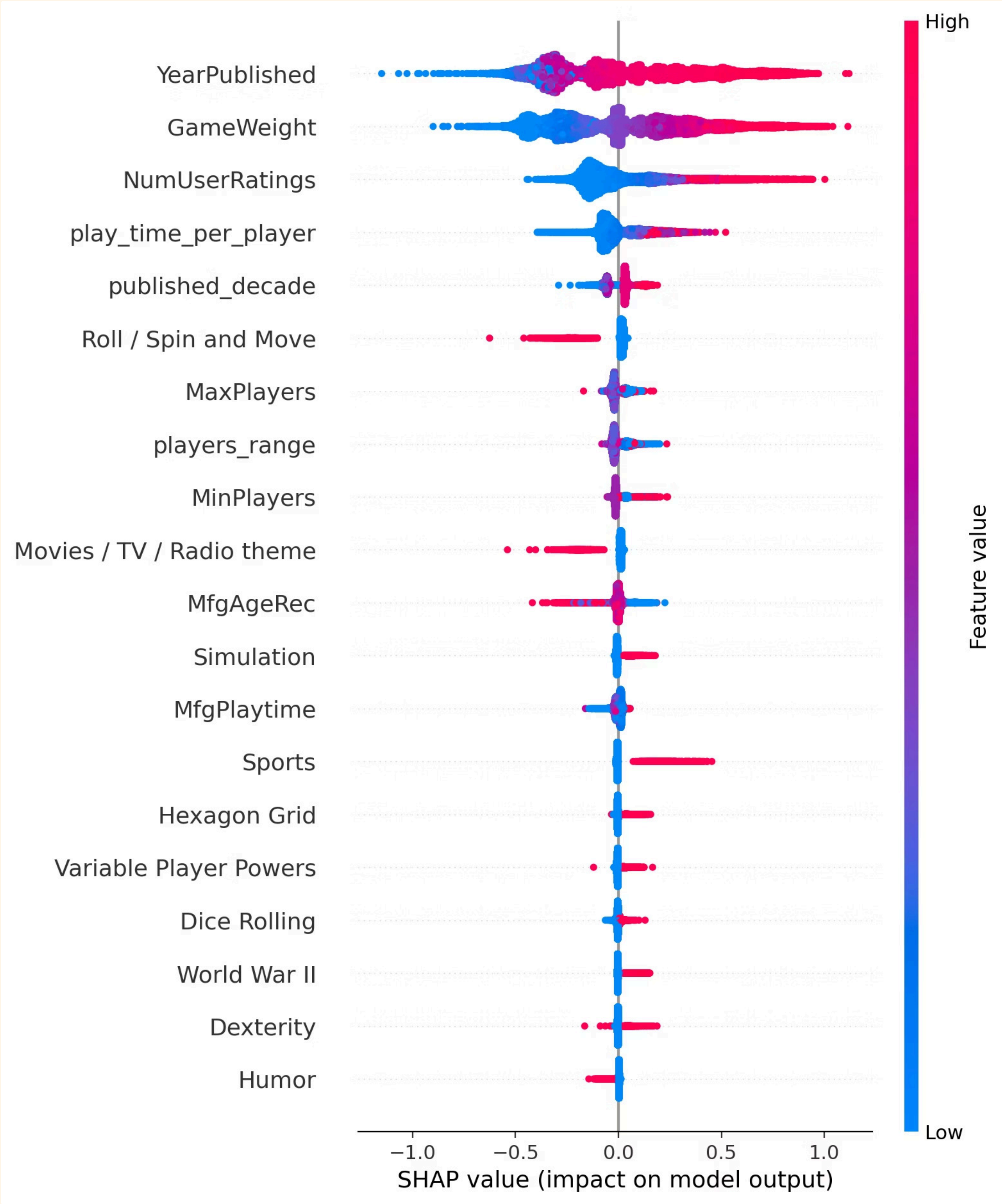






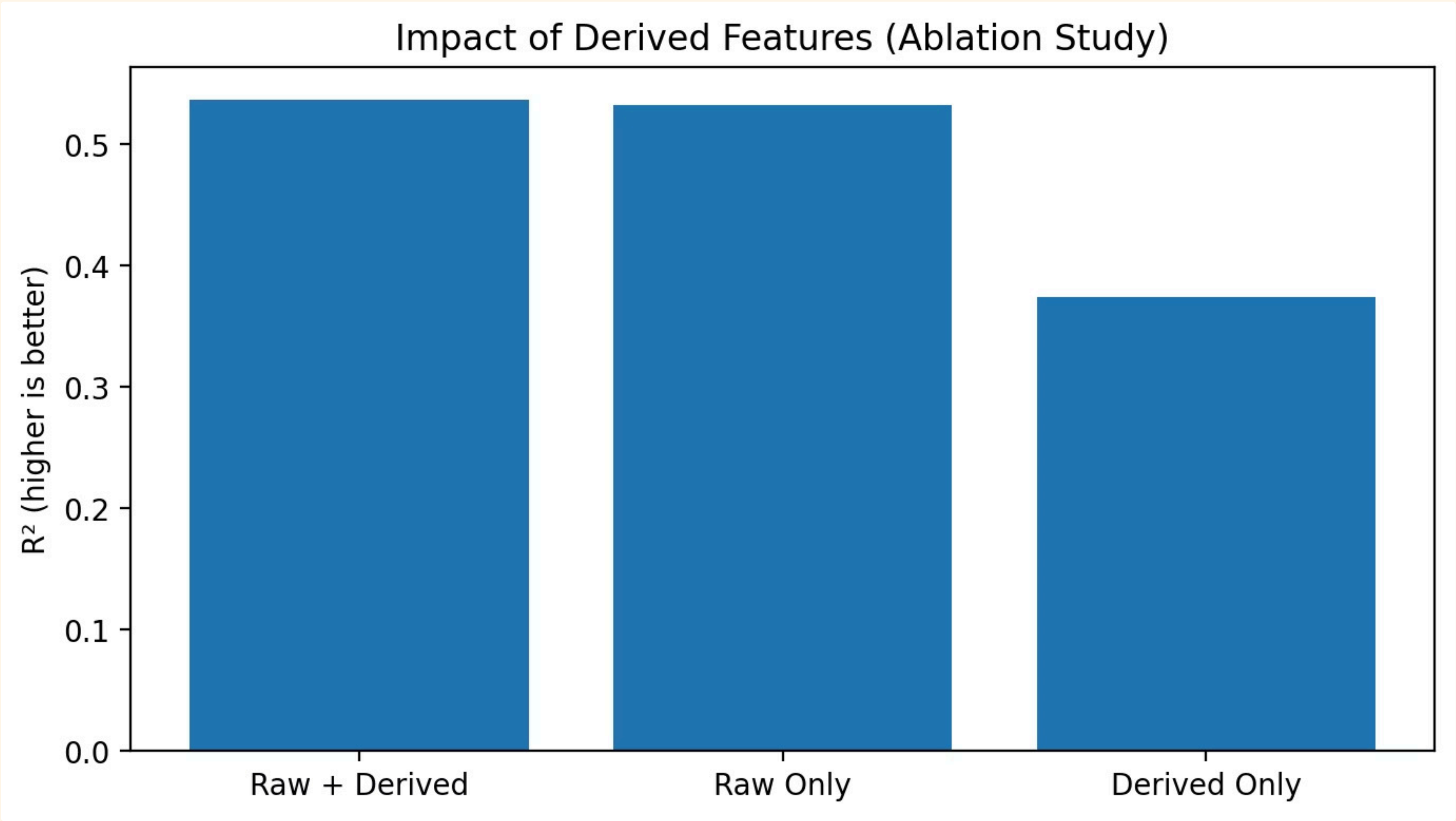
Feature Importances

GameWeight, YearPublished, and Roll/Spind and Move are the most influential predictors of board game ratings. Several mechanics and themes contribute moderately, showing that gameplay characteristics also shape model decisions.



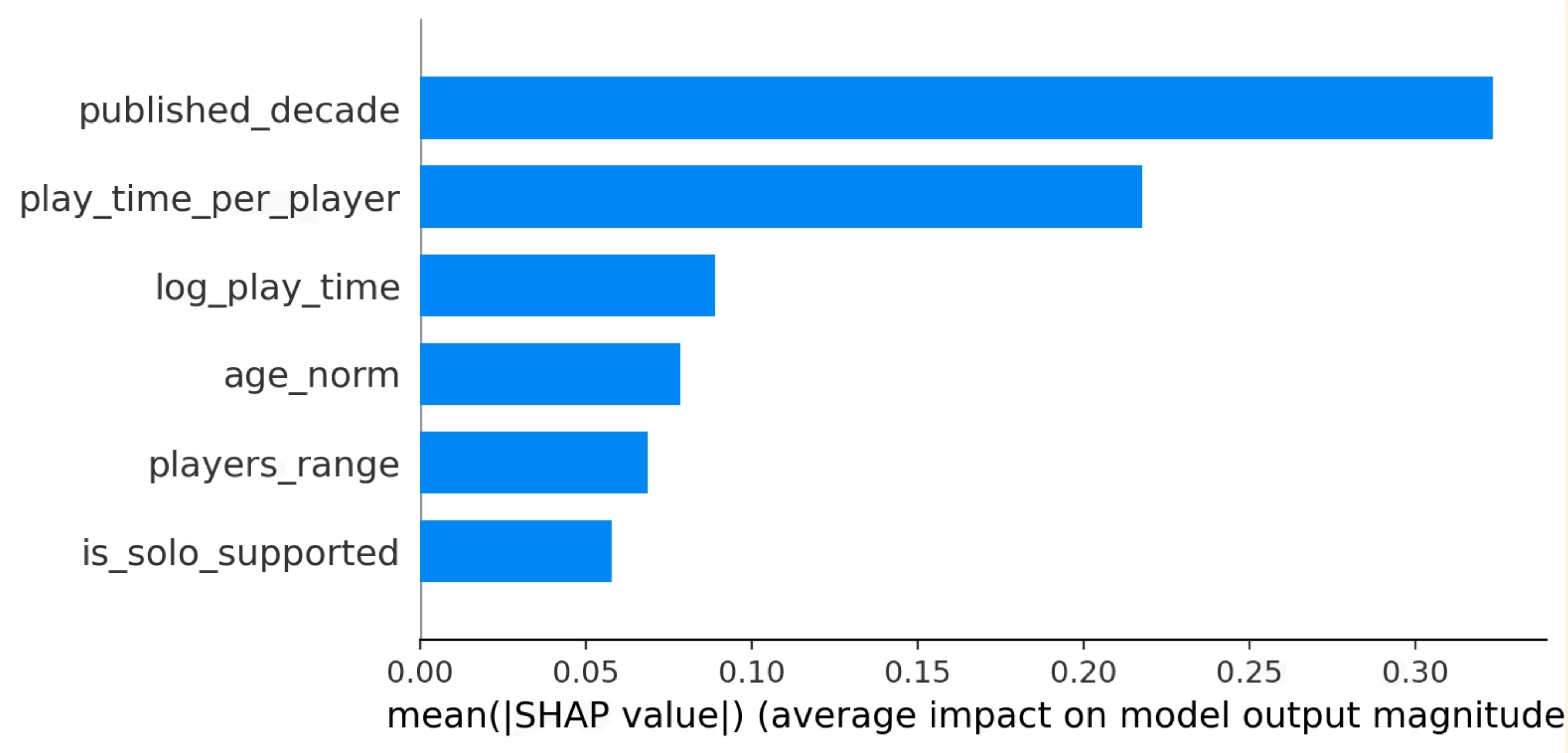
SHAP Summary Plot - Global Feature Impact

GameWeight, YearPublished, and NumUserRatings demonstrate the strongest global impact on model predictions. High feature values generally push predictions upward, while lower values tend to reduce the expected rating.



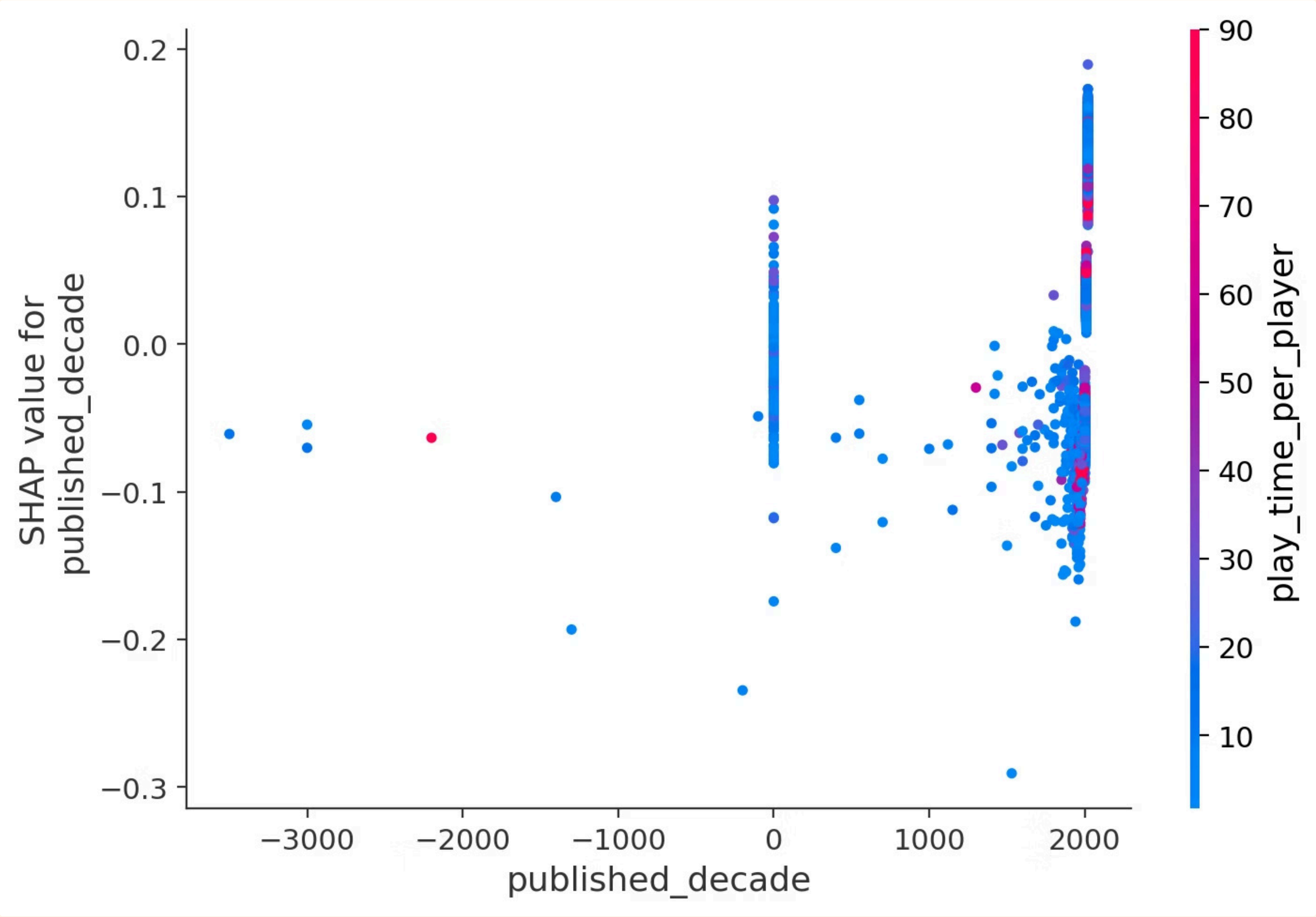
Ablation Study

The ablation study confirms that the **'Raw + Derived'** configuration yields the optimal performance (**R^2 = 0.537, RMSE = 0.632**), proving that while derived features alone are insufficient (R^2 = 0.374), they effectively complement the raw data to maximize predictive accuracy.



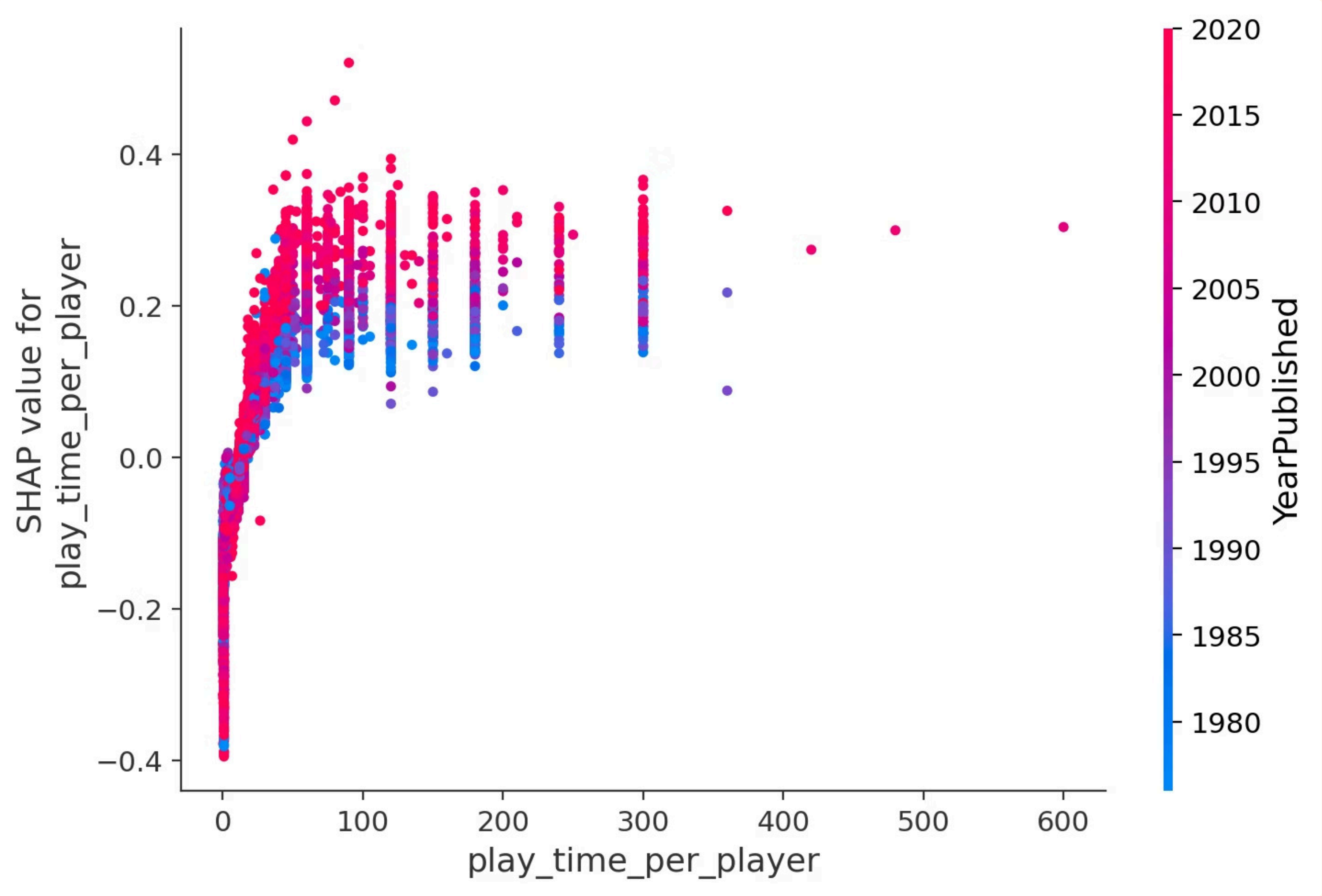
Importance of Derived Features

Among the derived features, `published_decade` (capturing modernization trends) and `play_time_per_player` (representing game flow) exhibit the strongest influence on the model.



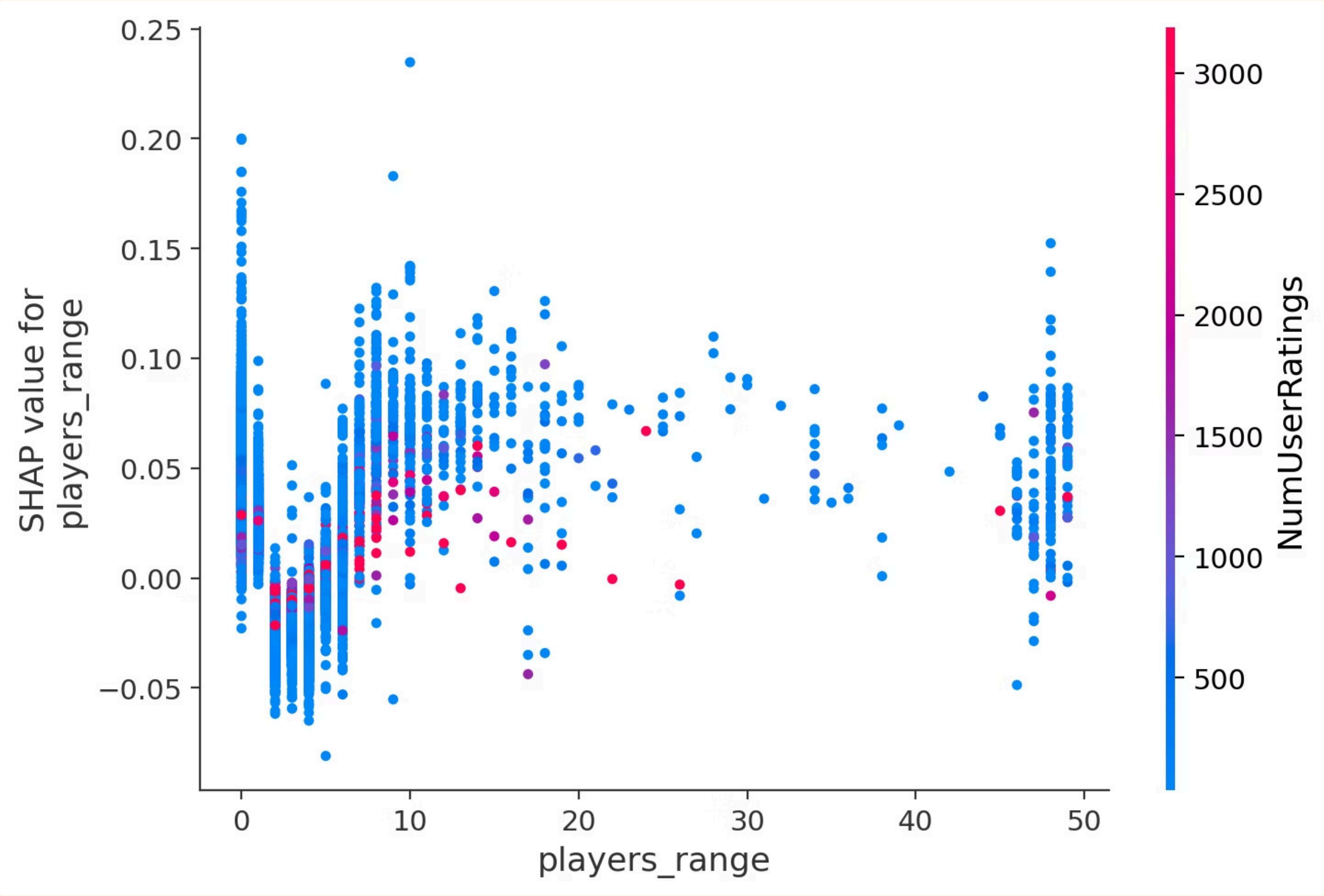
Impact of Decade

SHAP analysis indicates a systematic increase in predicted ratings as the published_decade approaches the present, suggesting that modern games tend to receive higher user evaluations.



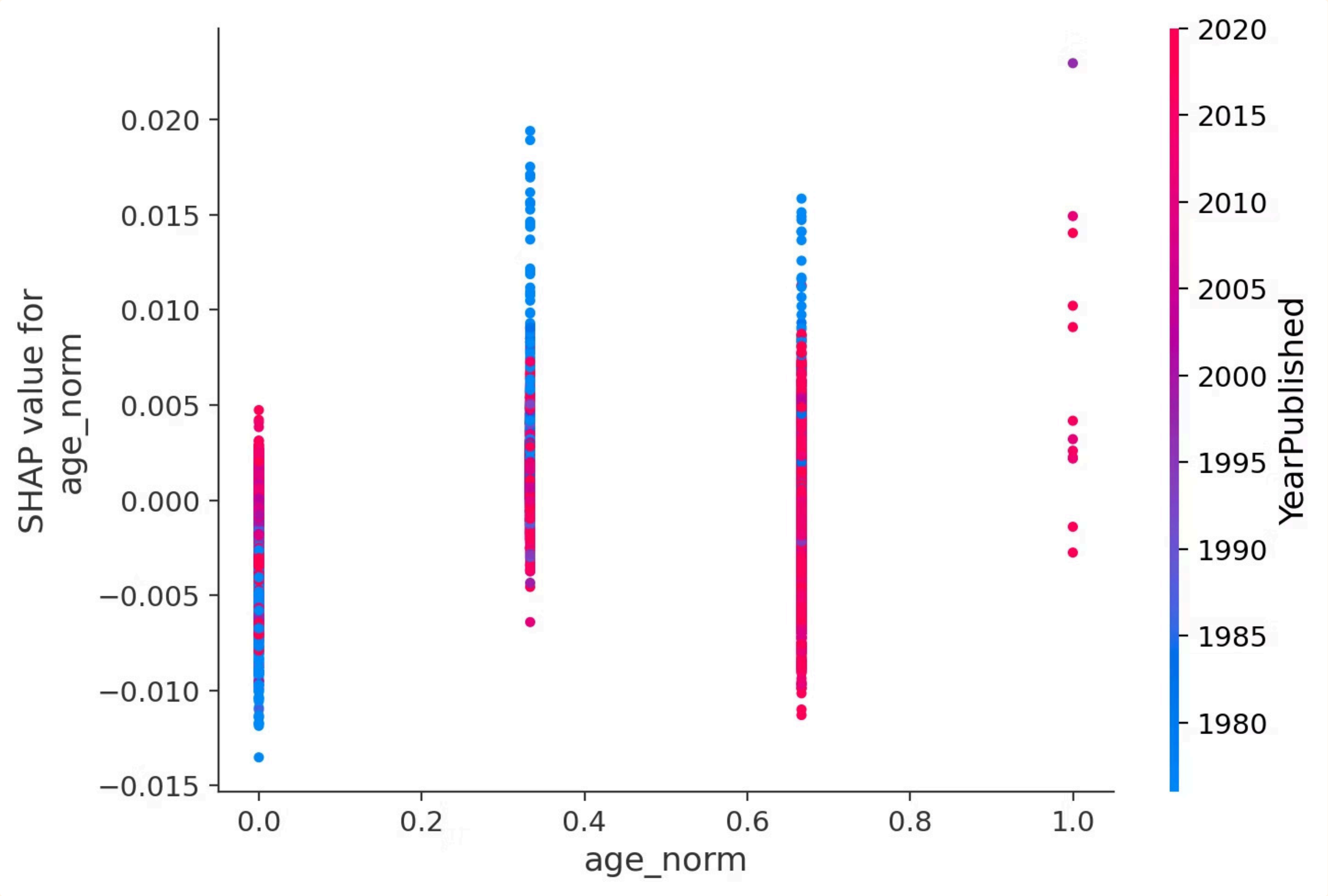
Impact of Time Per Player

While very short play_time_per_player values negatively impact predictions, games that allow sufficient time for strategic depth (approx. 30–60 mins) show a positive contribution to the model output.



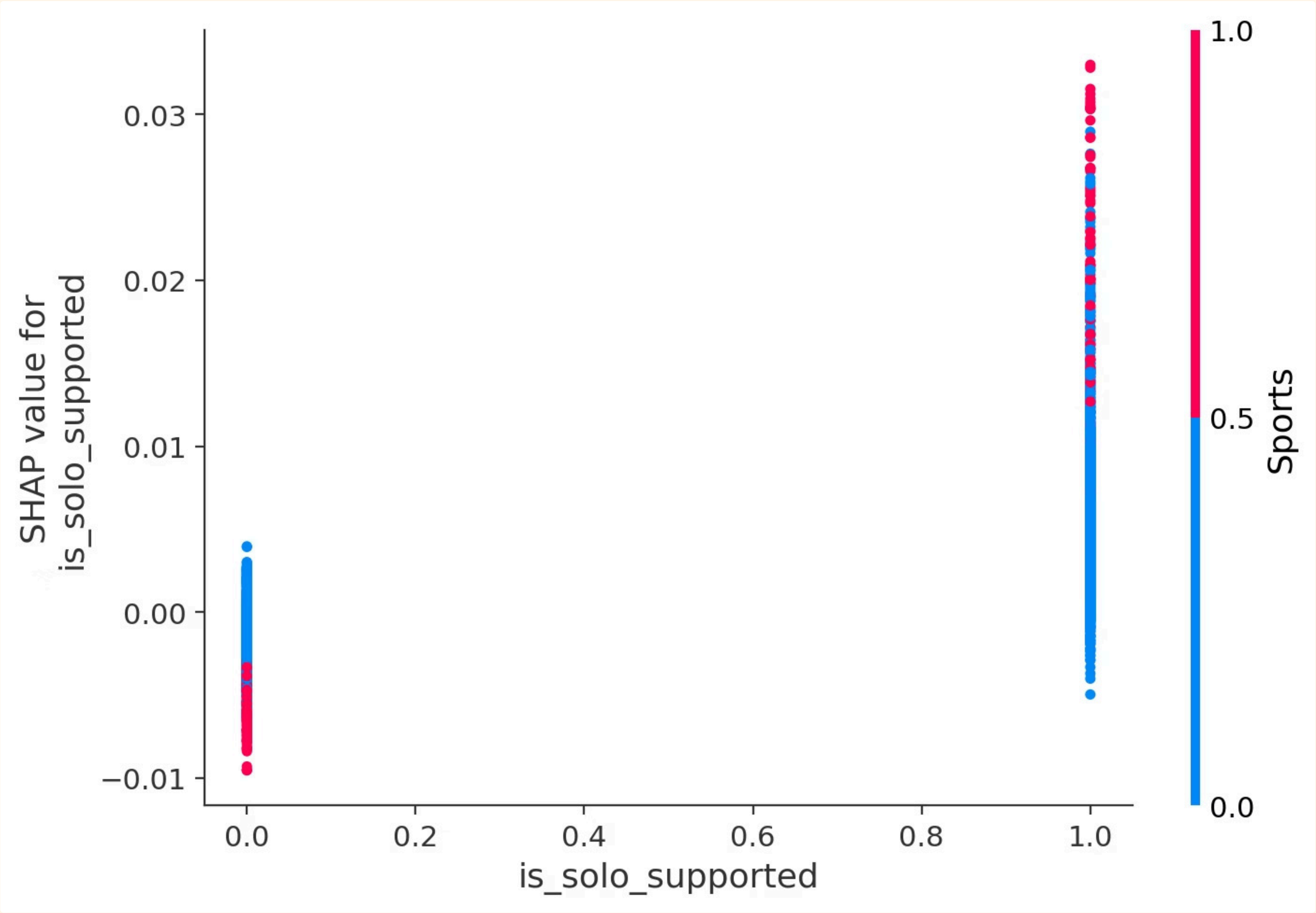
Impact of Player Range

A wider players_range initially increases the predicted rating by indicating flexibility, though this positive effect reaches saturation at higher values.



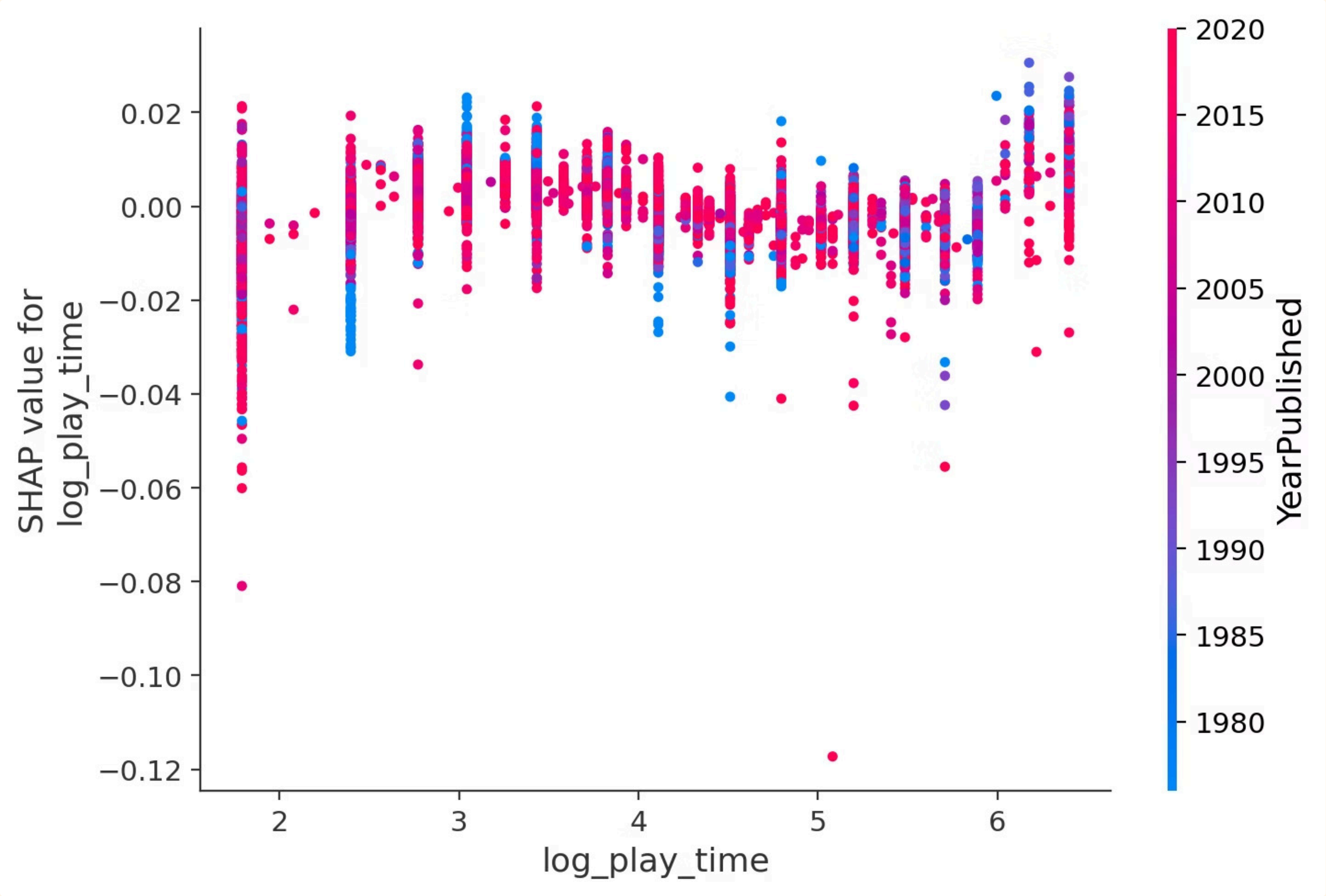
Impact of Age Recommendation

As the target audience age (age_norm) increases, a slight upward trend in predicted ratings is observed, likely correlating with greater game complexity and audience maturity.



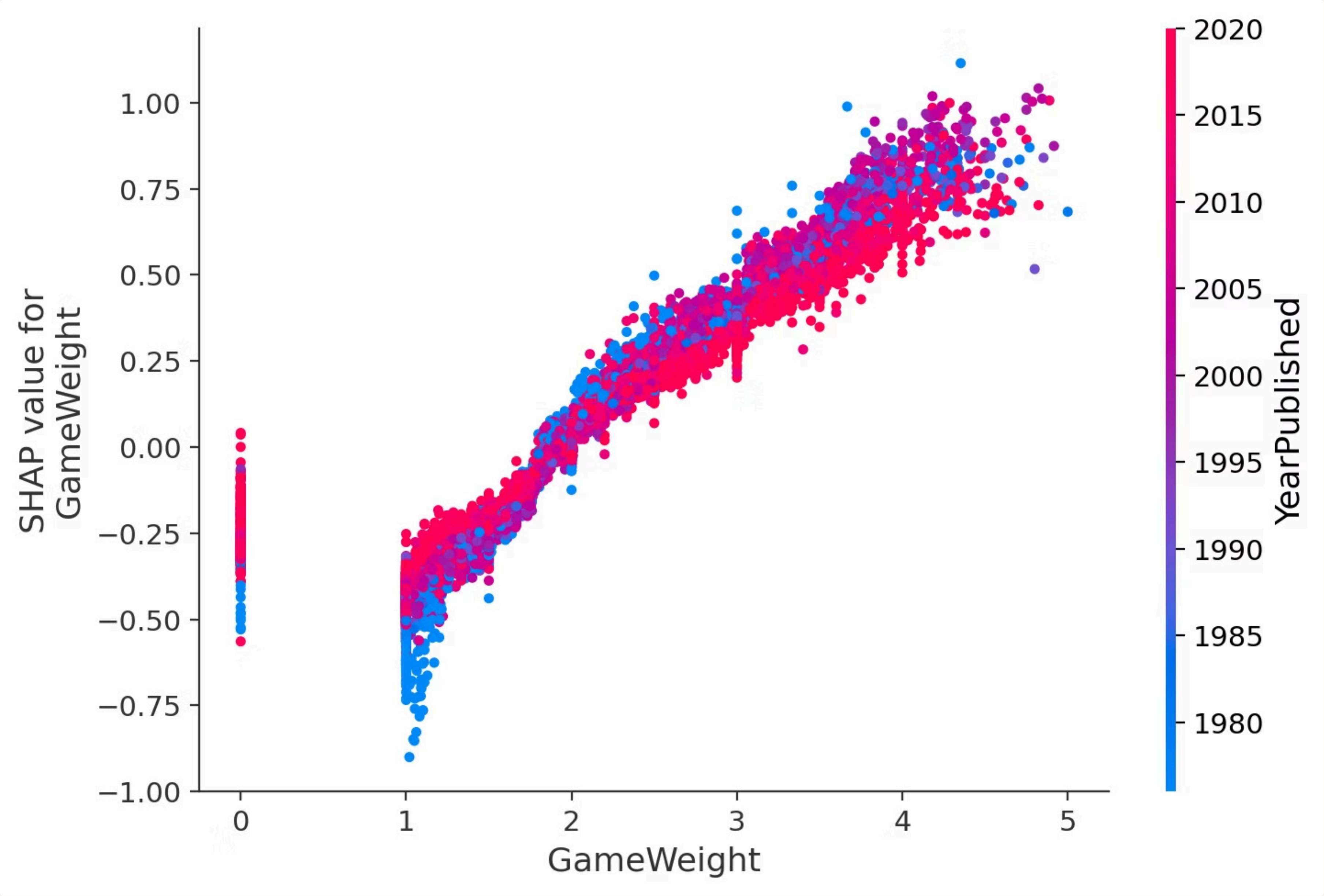
Impact of Solo Support

Games that support solo play (`is_solo_supported=1`) exhibit higher SHAP values compared to those that do not, indicating a positive contribution to the predicted rating.



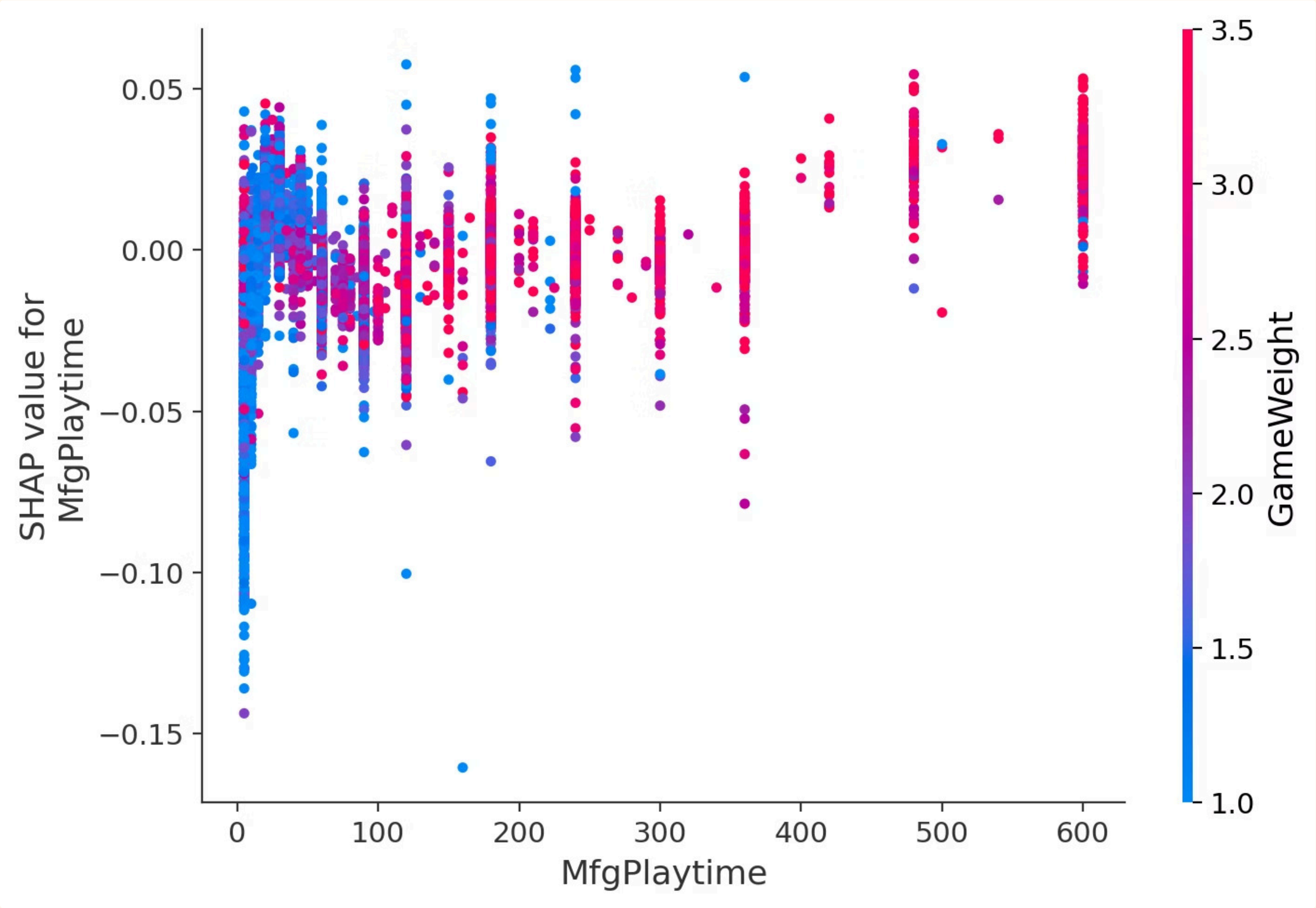
Impact of Total Playtime

Analysis of `log_play_time` suggests that extremely short games are penalized by the model, whereas games with moderate duration show a neutral to slightly positive impact.



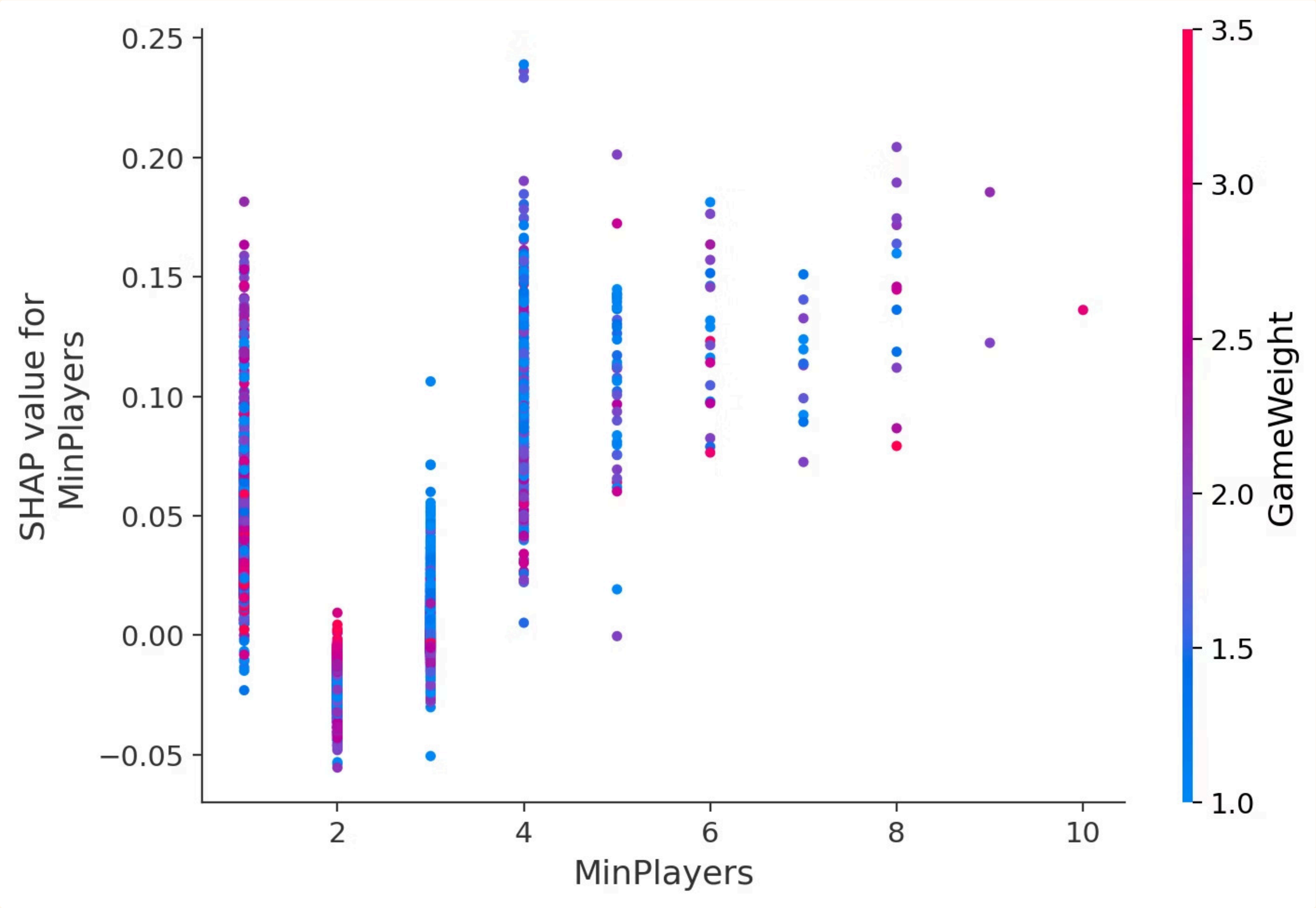
SHAP Dependence Plot – GameWeight

GameWeight shows a clear positive relationship with SHAP values, meaning complex games are predicted to receive higher ratings. The pattern is consistent across decades, with modern games generally reinforcing this trend.



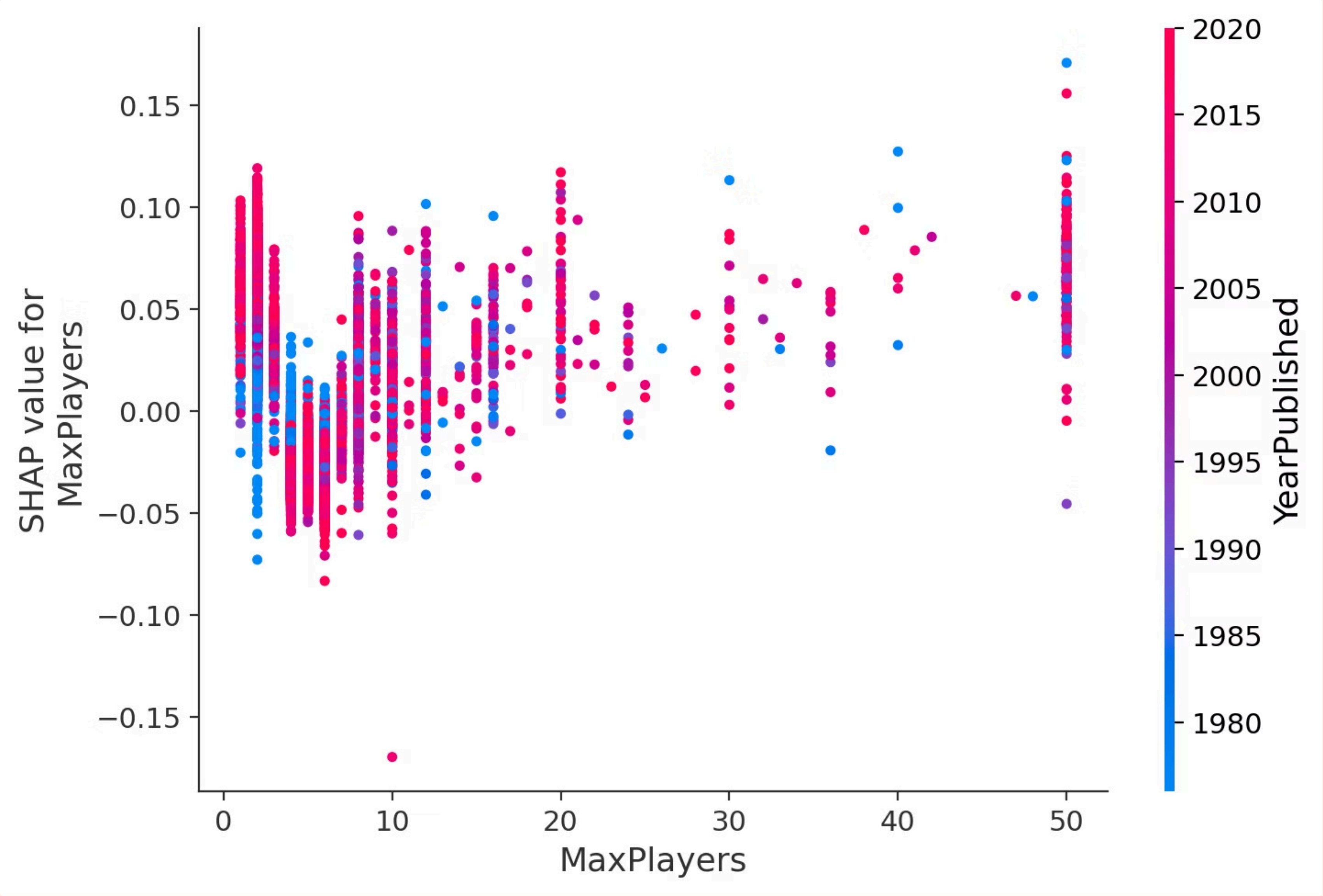
SHAP Dependence Plot – MfgPlaytime

Longer playtime tends to slightly increase SHAP values, meaning longer games may receive higher predicted ratings. The effect is stronger for complex games, indicated by higher GameWeight values appearing with more positive SHAP impacts.



SHAP Dependence Plot – MinPlayers

Games that support one player (solo play) produce positive SHAP effects, suggesting higher predicted ratings. As the minimum player requirement increases, the SHAP influence becomes slightly negative.



SHAP Dependence Plot – MaxPlayers

Most values cluster between 2–6 players, and the SHAP impact remains close to zero, indicating that maximum player count has limited influence on predicted ratings. Extremely high player counts are rare and do not significantly change the model output.

Conclusion: Decoding the Formula of Success

Our comprehensive analysis of board game data has uncovered significant factors influencing game ratings, moving beyond subjective intuition to data-driven insights.

1

Validating Predictive Power (RQ1 & H3)

Our analysis confirmed that machine learning can effectively predict board game success. **XGBoost outperformed Linear Regression significantly** ($R^2 \approx 0.60$), validating our hypothesis that the relationship between game features and ratings is non-linear and complex. This proves that success is not random; it is partially engineered through specific design choices.

2

The "Complexity" Factor (RQ2)

The most dominant predictor of a high rating is **Game Weight (Complexity)**. Contrary to the belief that "simpler is better" for mass appeal, the BGG community heavily favors games that offer strategic depth and decision-making challenges.

3

The Evolution of Quality

Year Published is a top-tier predictor. There is a clear temporal trend: Modern games receive systematically higher ratings than older classics, suggesting a rise in industry standards and design quality over the last decade.

Limitations & Potential Improvements

Limitations of the Study

- **Unexplained Variance:** While our XGBoost model achieved an R^2 **score of 0.60**, approximately 40% of the variance remains unexplained. This suggests that board game success relies heavily on subjective factors not present in tabular data.
- **Data Bias:** The dataset relies on BoardGameGeek user ratings. Since this community consists mostly of dedicated hobbyists, the ratings may not fully reflect the preferences of the general "casual" audience.

Potential Improvements

- **Text Analysis (NLP):** To capture the subjective quality of games, we could incorporate Natural Language Processing (NLP) to analyze user reviews instead of relying solely on numerical metadata.
- **Visual Data Integration:** Using Deep Learning (CNNs) to process images of game boxes and components could help quantify the impact of visual aesthetics on the average rating.

Final Conclusion

> The project successfully validated that board game ratings are predictable using regression models. We confirmed that game mechanics and complexity ("Weight") are far more significant predictors than simple intuition suggests.