IN-GAME PURCHASE ANALYSIS

Identifying Behavioral and Demographic Drivers of In-Game Purchases in RPG Environments

RESEARCH

PURPOSE

Identify behavioral and demographic drivers of in-game purchases. Our goal is to balance player immersion (Engagement) with structural barriers (Progression Friction) to drive conversion without damaging retention.

QUESTIONS

RQ1: Does faster leveling reduce the need for paid boosts?

RQ2: Is a holistic Engagement Score more predictive than raw session metrics?

RQ3: Which ML model best reveals the "Why" behind spending?

FEATURES

ENGAGEMENT SCORE

“How deeply does a player engage on average"

LEVEL EFFICIENCY

"How rapidly does this player level up?"

SESSION INTENSITY

“How deeply does a player engage during a single session relative to their login frequency?"

WEEKLY COMMITMENT

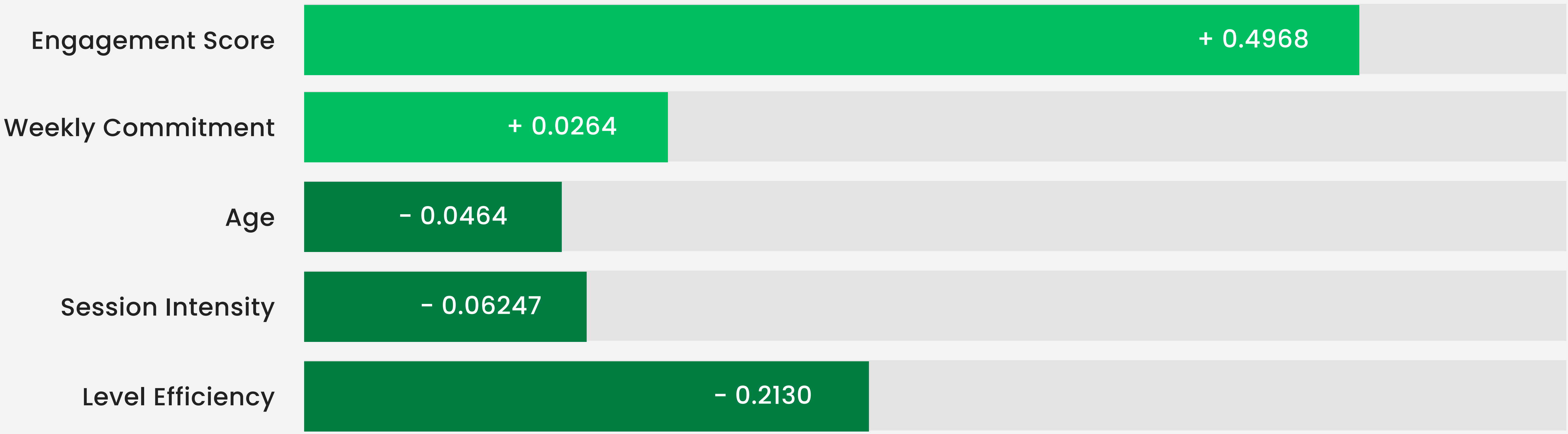
"How many total minutes per week does a player invest in the game?"

AGE

“How old is the player"

SVM Coefficients

The Linear Support Vector Machine (SVM) reveals the general direction of influence for each metric on purchase likelihood.



Key Finding: **High engagement** is the primary macro driver, while high efficiency discourages spending.

Accuracy: 48.55%

Random Forest

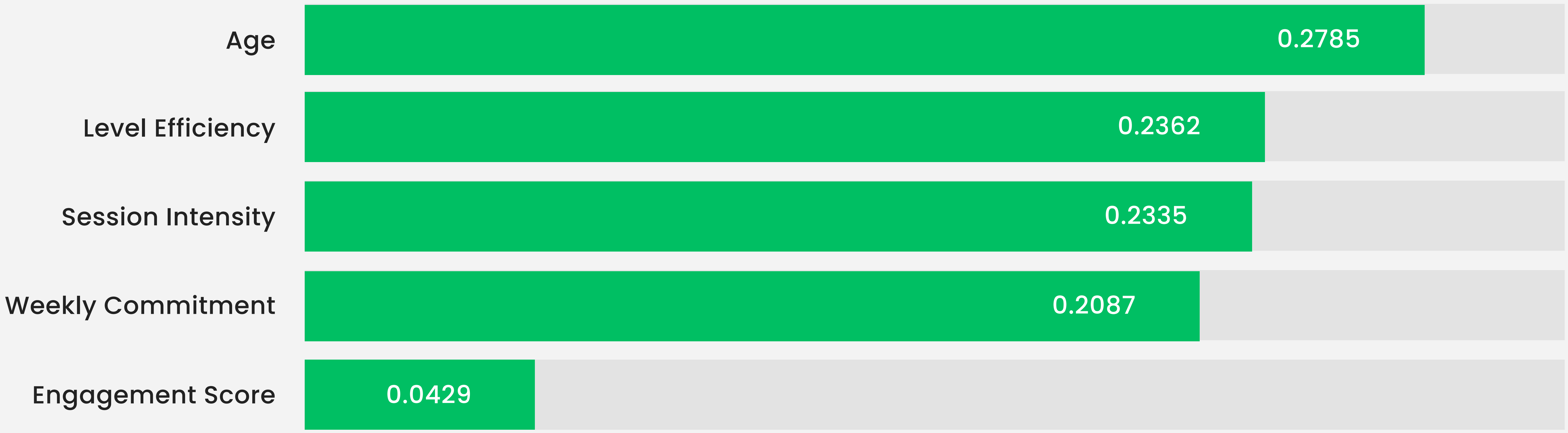
Identifies feature importance

Key Finding: This confirms that specific **Progression Hurdles** are better predictors of spending than holistic activity averages.

Accuracy: 48.72%

GBT

Identifies feature importance and non-linear Patterns



Key Finding: Predictive precision improves when **Demographic** and **Progression metrics** interact.

Accuracy: 51.79%

MODEL SUMMARY

Model Benchmarking Comparison

Model	Accuracy	Top Predictor	Analytical Value
Linear SVM	48.55%	Engagement Score	Reveals Macro Direction (Pos/Neg)
Random Forest	48.72%	Level Efficiency	Highlights Progression Impact
GBT	51.79%	Age	Captures Complex Micro-interactions

The Accuracy Paradox (~50%)

- Human Complexity: Spending is driven by **unobserved factors** (mood, social influence, events) that logs cannot capture.
- Synthetic Limits: **Synthetic data** lacks the subtle "behavioral nuances" of real-world telemetry.

RQ1 ANSWER

Does faster leveling reduce the need for paid boosts?

Across all models, Level Efficiency showed a negative correlation. Specifically, higher negative coefficients in SVM prove that effortless progression eliminates the psychological "Necessity" for boosts.

Conclusion: "Efficient players might be poor payers."

RQ2 ANSWER

Is a holistic Engagement Score more predictive than raw session metrics?

While SVM identifies the Engagement Score as a macro-driver for segmentation, our most accurate model (GBT) captures non-linear intent more effectively by analyzing the interaction between Age and progression metrics.

Conclusion: "Composite scores provide direction and granular data provides precision."

RQ3 ANSWER

Which ML model best reveals the "Why" behind spending?

- SVM (Linear): Best for establishing intuitive policy directions.
- GBT (Ensemble): Captures complex micro-interactions (e.g., age-specific friction thresholds) with highest accuracy (51.79%).

Conclusion: "GBT identifies intricate feature interactions for high-precision targeting, while SVM establishes intuitive policy frameworks based on directional coefficients."

SUMMARY

Accuracy & Strategic Validity

The Accuracy Paradox (~50%)

Precision is capped by unobserved human factors (mood, social influence, finance) and synthetic data constraints.

Linear Insight (SVM)

Establishes the Macro-Direction.
Identifies Engagement (+) and Efficiency (-) as the primary directional drivers.

Non-Linear Insight (RF/GBT)

Provides Micro-Precision.
Identifies Age and LevelEfficiency as the most critical predictors, capturing complex interactions that simple scores miss.

Strategic Synthesis

Maximize engagement while managing friction and Focus on specific age groups hitting progression walls."