Executive Summary

Binary Recommendation: Do NOT release Wonder Crafters as-is.

- Primary outcome (Day-14 Average Revenue per User, ARPU) shows no lift from either treatment; Control_1 remains highest (Plot D14-ARPU).
- Day-14 Minutes per User are ~8.0–8.2 minutes for all arms (Plot D14-Minutes) — differences are trivial.
- Daily ARPDAU (Average Revenue per Daily Active User) and full-window ARPU confirm the same ranking: Control_1 ≥ MoreExpensive ≈ Control_2 > Control_3 ≈ LessExpensive (Plot A1, A3; Payer KPIs table).
- Treatments massively increase exposure and shift gem-spend mix toward crafting, but these changes do not translate to more minutes or more revenue per user (Funnel & Gem-Mix outputs).

Data & Methods

We deduplicated assignment to earliest test_entry_date, sanitized durations (remove negatives; recompute total duration from parts camp/level/event) in the query "Temporary Tables for Analysis" and then computed user-level post-entry outcomes.

- Assignment deduplication & duration cleaning: Temporary Views abtest_state_one and duration_clean.
- Test interval & sample size checks: calendar coverage and daily assignment balance verified.
- Eligibility (≥700 Dragon Power) and exposure (any feature event) computed at user-level.
- Engagement: Minutes per Daily Active User (Minutes/DAU) calculated as total minutes of play (camp, level, event) divided by DAU; **minutes-mix (Plot E2) computed as the share of playtime spent in camp, level, and event per variant per day**.
- Monetization: ARPDAU time series (Plot A1), payer rate (Payer KPIs table), ARPU per user; user-level Day-14 ARPU via bootstrap confidence intervals (CIs) to account for skew in revenue distribution (Plot D14-ARPU).
- Controls are reported separately (Control_1, Control_2, Control_3) because monetization levels differ across them. **Control_1 serves as the benchmark control in the narrative.**
- "Daily ARPDAU.sql" and "Engagement.sql" queries use explicit NULLIF guards to avoid divide-by-zero.

Randomization & Power Checks

- Assignment window spans 2023-10-18 → 2023-12-06; daily assignment steady and balanced by arm (Randomization Health & Test Interval outputs).
- Sample size per arm ≈ 386k users — sufficient power for percent-level effects (Test Interval output).
- No obvious Sample Ratio Mismatch (SRM) or time skew (e.g., one variant disproportionately assigned early in the test, another assigned more toward the end) issues observed in the daily assignment plots/tables.

Eligibility & Exposure

- Eligibility (≥700 Dragon Power) uniform across arms at ~39.7–40.0%.
- Exposure (any feature event, post-entry): Control ≈2.6% of all users (≈6.3% of eligibles); LessExpensive ≈25.2% all (≈61.4% eligibles); MoreExpensive ≈40.2% all (≈94.2% eligibles).
- Interpretation: Treatments (especially MoreExpensive) **massively increase exposure**.

Feature Usage Funnel

- LessExpensive: ~41.9k recipe selections → 36.2k crafting starts → 35.7k item collects.
- MoreExpensive: ~77.4k recipe selections → 69.5k crafting starts → 68.9k item collects.

- Controls: only a few hundred events at each stage.
- Median delay Selection → Start ≈0.05–0.10 hours (players typically begin crafting **within minutes**).
- Median delay Start → Collect ≈0 hours (most recipes completed/collected the same day → either **short timers or gem skips**).
- Interpretation: Low friction — once players engage, they progress quickly through the funnel with minimal barriers.

Engagement

- Minutes per DAU (daily mean ± bootstrap 95% CI): Treatments show an insignificant dip vs best control (Plot E1).
- Minutes-mix: ~72% of time spent in events across arms; MoreExpensive shows slightly more camp time (Plot E2).
- Minutes per DAU time series are nearly identical across variants (Plot E3).

Bootstrap Confidence Intervals (CIs)[1] are estimated by repeatedly resampling users with replacement to capture the variability of skewed metrics (e.g., revenue, minutes). This method **avoids assuming** normality and provides robust error bars in the plots. Overlapping CIs across variants indicate no statistically meaningful differences.

Monetization

- Payer KPIs (full window): payer rate ~9.85–9.93% across arms (near-identical). ARPU ranking: Control_1 highest; LessExpensive lowest; MoreExpensive ≈ Control_2 (Payer KPIs table).
- Daily ARPDAU: Control_1 leads overall; MoreExpensive between Control_1 and Control_2/3; LessExpensive trails (Plots A1, A3).
- User-level D14 ARPU with bootstrap 95% CIs: Control_1 highest; MoreExpensive near Control_2 but still below Control_1; LessExpensive lower. Error bars overlap (Plot D14-ARPU).

Economy Mix (Gem Spend)

- Share of crafting-related gem spend in total gem spend: Controls ≈0.035–0.19%; LessExpensive ≈0.45%; MoreExpensive ≈5.92%.
- Interpretation: **strong reallocation toward crafting economy in treatments (especially MoreExpensive); this mix shift did not produce higher revenue per user.**

Recommendation

Do NOT release Wonder Crafters as-is.

- High exposure and large gem-mix shifts do not translate into higher revenue per user (D14 ARPU) or meaningfully higher engagement (Minutes/DAU).
- MoreExpensive is the least-bad option (near Control_2) but fails to beat the best control (Control_1).

Appendix Reference

All figures cited (A1, A3, D14-ARPU, D14-Minutes, E1–E3) and CSV summaries (see "Outputs" -> "Summary Tables" subfolder) can be found in the folder "Outputs". These materials include all plots and summary CSV files referenced in this report.

[1] Ho, J. (2019, July 4). Bootstrap confidence intervals. https://acclab.github.io/bootstrap-confidence-intervals.html#:~:text=Introducing%20the%20bootstrap%20confidence%20interval&text=That%20is%20to%20say%2C%20we,difference%20by%20performing%20bootstrap%20resampling.