

User Engagement Analysis Report

Executive Summary

This analysis evaluates how learner engagement changed between Q2 2021 and Q2 2022 on an online education platform, focusing on **minutes watched** as the primary engagement metric. The study compares **free** and **paid** learners, analyzes engagement distributions, performs hypothesis testing, and identifies actionable business implications.

Key Engagement Findings

Segment	2021 Mean	2022 Mean	Change	Test Result
Free Learners	A	B	X%	Significant / Not Significant
Paid Learners	C	D	Y%	Significant / Not Significant

Insights

- Paid learners consistently show much higher engagement than free learners.
- Distribution analysis reveals strong right-skew driven by highly active users.
- Confidence intervals provide reliability and variance understanding.

Statistical Testing

Two-sample t-tests were performed to evaluate whether 2022 engagement increased compared to 2021.

Free Learners

- Hypothesis: $H_1 : \mu_{2022} > \mu_{2021}$
- p-value: `p_free`
- Interpretation: [Reject / Fail to Reject] $H_0 \rightarrow$ [Increase / No confirmed increase]

Paid Learners

- Hypothesis: $H_1 : \mu_{2022} > \mu_{2021}$
- p-value: `p_paid`
- Interpretation: [Reject / Fail to Reject] $H_0 \rightarrow$ [Increase / No confirmed increase]

Engagement & Certificates

- Pearson correlation: `r = ...`
- Relationship strength: Weak / Moderate / Strong
- Interpretation: More engaged learners tend to earn more certificates, though correlation does not imply causation.

Key Issues Identified

1. Large performance gap between free and paid learners suggests different motivation levels.
2. Growth measurement is sensitive to outliers; trimming affects interpretation.
3. Confidence intervals may overlap, reducing certainty of measured improvements.
4. Engagement-certificate correlation may reflect learner motivation rather than product features.

Recommendations

Immediate

- Strengthen engagement nudges for free users.
- Analyze feature changes between 2021–2022 for contribution insights.
- Investigate segmentation by learner type (new vs returning).

Strategic

- Improve subscription conversion pathways for medium-engagement users.
- Introduce gamification and certificates to motivate free users.
- Conduct A/B experiments to avoid Type II errors and validate improvements.

Methodology

- Data Source: CSV exports from SQL database
- Tools: Python (Pandas, NumPy, SciPy, Matplotlib, Seaborn)
- Cleaning: Outlier trimming at the 99th percentile
- Statistical Tests: One-sided t-tests, 95% confidence intervals
- Metrics: Minutes watched, certificates issued

Conclusion

The results of the hypothesis tests provide statistical evidence regarding year-over-year engagement trends. Where p-values fall below the 0.05 significance threshold, we can confidently conclude that engagement in 2022 is significantly higher than in 2021. When p-values exceed this threshold, we cannot reliably confirm improvement, suggesting that any observed differences may be due to random variation rather than real behavioral change.

The accuracy of hypothesis conclusions is strengthened by the large sample sizes and confidence interval estimation, which reduce uncertainty. However, overlapping confidence intervals and skewed distributions highlight that effect sizes should be interpreted cautiously. Overall, the combination of hypothesis testing, confidence intervals, and robust outlier filtering delivers a dependable assessment of engagement growth and supports evidence-based business decisions.

Future analyses through controlled A/B testing can further improve confidence, reduce noise from external factors, and minimize the risk of Type II error when evaluating feature impact or product strategy.