DATA ANALYTICS REPORT

# Facebook Algorithm A/B Test Analysis

*Hypothesis Testing | Statistical Analysis | Actionable Insight*

**Jordan Foltz**
Jordy3338@gmail.com
linkedin.com/in/jordan-foltz

March 2026

# Introduction

**Facebook ran a controlled experiment to evaluate a new video-serving algorithm.**

Historically, 14% of users qualify as Engaged Watchers — users who watch a meaningful amount of video on their feed. The business goal was to determine whether Algorithm B could raise that rate to 17%. A higher engagement rate expands video inventory and increases advertising revenue.

Users were randomly assigned to one of two groups. The control group received Algorithm A. The test group received Algorithm B. Each user was recorded as either an Engaged Watcher (1) or not (0).

# Hypotheses

Before reviewing any data, I defined both hypotheses. This step locks in the decision rule and prevents conclusions from being shaped by the data after the fact.

**NULL HYPOTHESIS ($H_0$)**

Algorithm B does not increase the proportion of Engaged Watcher users compared to Algorithm A.

Any observed difference is the result of random variation, not a genuine improvement.

$$H_0 : p_B \leq p_A$$

**ALTERNATIVE HYPOTHESIS ($H_a$)**

Algorithm B increases the proportion of Engaged Watcher users compared to Algorithm A.

The improvement is real and not the product of random variation.

$$H_a : p_B > p_A$$

**Why one-tailed? The test is one-tailed because the business question is directional. Facebook wants to know if Algorithm B is better, not simply different. A one-tailed test is more powerful when the direction of the expected effect is known in advance.**

# Minimum Detectable Effect

**Before examining any results, I ran a power analysis to confirm the experiment was large enough to be trusted.**

The Minimum Detectable Effect (MDE) is the smallest relative improvement the experiment must be able to detect. It determines the required sample size per group. Skipping this check risks missing a real improvement and incorrectly concluding that nothing changed.

MDE CALCULATION

```
MDE = (Target Rate - Baseline Rate) / Baseline Rate
MDE = (0.17 - 0.14) / 0.14 = 0.2143 = 21.43%
```

At a 5% significance level and 80% statistical power, this requires approximately 1,800 users per group.

# Pivot Table Results

I built a pivot table in Excel to count users by group and tally Engaged Watchers. This produced the raw figures used in the significance test.

| Algorithm | Total Users | Engaged Watchers | Engagement Rate |
|---|---|---|---|
| A — Control | 1,819 | 260 | 14.29% |
| B — Test | 1,792 | 293 | **16.35%** |
| **Grand Total** | **3,611** | **553** | **15.31%** |

SAMPLE SIZE CHECK

Group A: 1,819 users. Group B: 1,792 users. Both groups meet the required minimum of 1,800 users. The experiment carries sufficient statistical power to detect the targeted improvement.

# Two-Proportion Z-Test

### With the data in hand and sample sizes confirmed, I ran a two-proportion Z-test.

This test is appropriate because the outcome is binary and both groups are independent. With over 1,800 users per group, the Central Limit Theorem applies. The Z-test measures how many standard errors separate the two observed proportions. If the p-value falls below 5%, the null hypothesis is rejected.

### Calculation Steps

| Step | Formula | Result |
|---|---|---|
| 1. Observed rates | `p_A = 260 / 1,819  p_B = 293 / 1,792` | 14.29% \| 16.35% |
| 2. Pooled proportion | `p = (260 + 293) / (1,819 + 1,792)` | 0.1531 |
| 3. Standard error | `SE = sqrt( p(1-p) x (1/nA + 1/nB) )` | 0.01199 |
| 4. Z-statistic | `Z = (p_B - p_A) / SE` | 1.716 |
| 5. P-value (one-tailed) | `P(Z > 1.716)  from standard normal` | 0.0431 |
| 6. Decision (alpha = 0.05) | `Is p-value < 0.05?` | **Reject H0** |

# Statistical Conclusion

### The p-value of 0.043 falls below the 5% significance threshold. The null hypothesis is rejected.

There is only a 4.3% probability of observing a difference this large by random chance alone, assuming Algorithm B had no real effect. That probability is below our threshold. Algorithm B produces a statistically significant increase in Engaged Watcher users.

> **INTERPRETING THE P-VALUE**
> If this experiment were repeated 100 times on users where Algorithm B made no real difference, a result this extreme would appear roughly 4 times. That is rare enough to conclude the improvement is genuine.

# Actionable Insights

### The data is clear. Algorithm B outperforms Algorithm A and the result is statistically significant.

The test was properly powered, the sample sizes were sufficient, and the p-value falls below the predetermined threshold. There is a strong statistical basis for acting on these findings.

| +2.1pp | +14.4% | 0.043 |
|---|---|---|
| Absolute Lift<br>**Percentage point increase** | Relative Lift<br>**vs. Algorithm A** | P-Value<br>**Threshold: 0.05** |

# Recommendation

### Algorithm B should be deployed to the full user base.

The experiment confirms a real and meaningful improvement in video engagement. Algorithm B converts 2.1 additional percentage points of users into Engaged Watchers. At Facebook's scale, that translates to a significant expansion of ad inventory across pre-roll and mid-roll placements.

Post-launch monitoring is recommended to confirm that results hold at full scale and to detect any unintended effects on content diversity or session quality.

**FINAL VERDICT**

Reject Ho. Algorithm B produces a statistically significant increase in Engaged Watcher users at the 5% significance level. The business case for a full rollout is supported by the data.

# Tools and Methods

| | |
|---|---|
| Two-proportion Z-test | Power analysis (80%, alpha = 5%) |
| Pivot table (Microsoft Excel) | A/B Sample Size Calculator |
| Minimum Detectable Effect | A/B Significance Calculator |
| Central Limit Theorem | One-tailed hypothesis testing |