

Problem: Lately, we've noticed a troubling trend: our site's average session duration has been declining steadily over the past two months, dropping from an average of 4 minutes to just under 3 minutes. Since session duration is a key indicator of user engagement on our platform, this decline is a big concern for our team—and for revenue, because fewer minutes spent often translates to fewer ad impressions and less likelihood of a purchase. Our suspicion is that recent changes to our homepage layout might be confusing users or failing to highlight our most engaging content.

1. Context & Background

1. Business Context

- **Problem:** Session duration fell from ~4 minutes to ~3 minutes over two months.
- **Platform:** Online site featuring content and e-commerce opportunities.
- **Suspected Cause:** Recent homepage layout changes might be confusing users or failing to highlight engaging content.

2. Why This Test Now?

- **Performance Issue:** Lower session duration often means fewer ad views and lower purchase likelihood.
- **Competitive Risk:** If users aren't engaged, they may switch to alternatives.
- **Business Impact:** Revenue could decline because ad impressions and conversions typically correlate with session duration.

Key Takeaway: We need to determine if redesigning or reverting the homepage layout can reverse the declining engagement trend.

2. Objective

High-Level Goal:

- Increase the **average session duration** back up to (or beyond) 4 minutes.

Secondary Goals (optional but useful to track):

- **Bounce Rate:** Decrease bounce rate (the percentage of users leaving quickly).
- **Click-through Rate (CTR)** on recommended or featured content.
- **Time on Key Pages:** E.g., product pages or articles we especially want users to see.

Why This Matters:

- A direct link exists between session duration and revenue (ads, purchases).
- Ensuring users spend more time on the site can lead to higher satisfaction and retention.

3. Hypothesis

If we modify the homepage layout to highlight engaging content and simplify navigation, **then** we expect the average session duration to **increase** from about 3 minutes to at least 4 minutes.

Rationale

1. **User Feedback:** Customers and support staff have noted confusion about the new homepage design.
 2. **Analytics:** The decline started after the last homepage update, suggesting the design may be a root cause.
 3. **Industry Best Practices:** Showcasing popular or personalized content on the homepage typically improves engagement.
-

4. Metrics Selection

1. **Primary Metric**
 - **Average Session Duration:** The main metric we want to improve. Measured as total time spent on the site per session.
 2. **Secondary Metrics**
 - **Bounce Rate:** If bounce rate decreases, it suggests the homepage design is encouraging users to explore.
 - **CTR on Featured Content:** If users are drawn to highlighted articles or product links, we'll see an uptick in clicks.
 3. **Guardrail Metrics**
 - **Page Load Time:** Ensure the redesign doesn't slow down performance.
 - **Error Rates:** Check for any increase in site errors or broken links.
-

5. Randomization Strategy

- **Segmentation:** Decide whether this test applies to all users or specific segments (e.g., new vs. returning users).
- **Assignment:**
 - A simple **50/50** split is often best for a standard A/B test.
 - If you suspect mobile vs. desktop usage drastically differs, consider **block randomization** so each device type is balanced in Control vs. Treatment.

6. Sample Size Determination

1. **Minimum Detectable Effect (MDE)**
 - Suppose we want to detect an increase in session duration from 3 minutes to **3.5 or 4 minutes**. Decide on your smallest “worthwhile” improvement.
 2. **Baseline & Variability**
 - Past data shows a baseline of 3 min (down from 4).
 - Check how much session duration fluctuates typically—this variability informs the required sample size.
 3. **Statistical Significance & Power**
 - Choose common values: $\alpha = 0.05$, **power = 0.80**.
 - Use an online sample size calculator (e.g., Optimizely, Google’s calculator) to find how many sessions you need in each group to detect the chosen MDE with 80% power.
 4. **Traffic Projections & Duration**
 - Estimate how many daily sessions you get.
 - Determine how many days/weeks you’ll need to reach the required sample in both Control and Treatment.
-

7. Pre-Test Validation (A/A Testing & Instrumentation Check)

1. **A/A Test:**
 - Optionally run a short test splitting users **50/50** with the **same** homepage experience in both groups.
 - Confirm that you’re not seeing weird differences in session duration that could suggest instrumentation or randomization issues.
 2. **Instrumentation Checks:**
 - Verify session duration is tracked correctly (e.g., ensure the tracking script fires at session start and updates on user activity).
 - Check that bounce rate and other events are correctly logged.
 3. **Baseline Confirmation:**
 - Compare A/A test data to the known 3-minute average to ensure consistency.
-

8. Experiment Setup

1. **Technical Implementation**
 - Possibly use a **feature flag** or A/B testing platform.
 - Control Group: Current homepage layout.
 - Treatment Group: Redesigned homepage that highlights top content and simplifies navigation.
 2. **Version Control / Environments**
 - Test the redesign on a staging environment first to ensure no major bugs.
 - Then enable the feature flag in production for your defined user percentage.
 3. **Hypothesis Registration**
 - Document the null hypothesis (no difference in session duration) and the alternative hypothesis (session duration improves by X minutes).
-

9. Running the Experiment

1. **Real-Time Monitoring**
 - Track session duration, bounce rate, and guardrail metrics in a dashboard.
 - Set alerts if bounce rate or error rates spike suddenly.
 2. **Avoid Excessive Peeking**
 - Don't base decisions on the first day or two of data.
 - Define checkpoints (e.g., 1 week, 2 weeks) or wait until you've reached the calculated sample size.
 3. **Handling Novelty Effects**
 - Users might be curious initially or confused. Let the test run for at least 1-2 weeks to get stable data.
 4. **Logging & Documentation**
 - Note any external events, marketing campaigns, or site outages that might affect session duration.
-

10. Data Analysis & Interpretation

1. **Statistical Significance (p-value)**
 - Perform a test (z-test or t-test) comparing average session duration in Control vs. Treatment.
 - If $p < 0.05$, we conclude a statistically significant difference.
2. **Effect Size**
 - How much did session duration increase? (e.g., from 3 min to 3.8 min).
 - Is this difference practically meaningful (does it meet or exceed your MDE)?
3. **Segment Analysis**

- Check if improvements differ by device type, user region, or new vs. returning users.
 - You may discover the layout redesign especially helps mobile users, for instance.
4. **Guardrail Check**
- Confirm page load time hasn't spiked.
 - Ensure there isn't a big jump in error rates or user complaints.
-

11. Decision & Implementation

1. **Go/No-Go Decision**
 - **Go:** If session duration improved significantly and no major guardrails were broken, proceed with rollout.
 - **No-Go:** If there's no improvement or negative impacts, revert the homepage changes.
 2. **Rollout Strategy**
 - **Phased Rollout:** If you're cautious, roll out to 10% → 25% → 50% → 100% of users, monitoring metrics each step.
 - **Full Rollout:** If results are very strong and you're confident, you can launch to all users at once.
 3. **Communication**
 - Present your findings to stakeholders (product, marketing, executives).
 - Show the difference in session duration, p-value, secondary metrics, and next steps.
-

12. Post-Implementation Monitoring

1. **Long-Term Observation**
 - Keep an eye on session duration after the feature is at 100%.
 - Watch for any gradual decline or external events that might affect user behavior.
 2. **Ongoing Optimization**
 - If you see smaller subgroups that didn't benefit, run follow-up tests (e.g., a dedicated layout for mobile).
 - If performance drifts downward again, investigate additional improvements.
 3. **Feedback Loop**
 - Gather qualitative feedback from surveys, customer support.
 - Merge that with quantitative data to refine the homepage design further.
-

13. Additional Considerations

1. Edge Cases & Low Traffic

- If certain user segments (e.g., users on older devices) are too small, data might be inconclusive.
- Consider lengthening test duration or grouping similar segments together.

2. Data Quality

- Ensure no spam or bot traffic is skewing session duration.
- Double-check analytics scripts for all relevant browsers and device types.

3. Multiple Variants

- You might want to test more than one homepage redesign concept. In that case, use a multi-armed bandit approach or just run separate A/B tests for each variant.

4. Privacy & Compliance

- If tracking user behavior in detail, ensure compliance with GDPR/CCPA where applicable.