



Experiment Design and Results Template

| What This Is | Step by Step Guide |
|--|---|
| <p>This template is brought to you by Split, a unified feature flagging and experimentation platform. By pairing feature flags with data measurement capabilities, Split is helping enterprise-level businesses release features quickly, more intelligently, and with less risk.</p> <p>In this template, we'll walk you through the steps that will help you create solid, well-thought out experiments. The template will provide you a framework to ensure all considerations are met, helping teams to maximize learnings, consider the bigger picture, plan for all outcomes, and easily scale your experimentation program.</p> | <ol style="list-style-type: none">1. Create a Hypothesis2. Perform a Power Analysis3. Provide Experiment Information4. Provide Design References5. Capture Metrics Impact and Results6. Final Conclusion, Learnings and Next Steps |

Hypothesis

What hypothesis are you trying to validate with this experiment? Need help with writing a hypothesis? Visit Split's Help Center:

<https://help.split.io/hc/en-us/articles/360055681831-Constructing-a-Hypothesis>

| | |
|------------------------|--|
| Problem - What? | |
| Strategy - How? | |
| Tactics - If? | |
| Result - Then? | |
| Hypothesis: | |



Baseline Data: Power Analysis

Visit Split's Help Center for more details regarding performing a power analysis prior to running experiments:

<https://help.split.io/hc/en-us/articles/360034040851-Sample-size-and-sensitivity-calculators>

| | |
|---|--|
| Sample size <i>How many <traffic type> do you expect to be included in the test given the targeting criteria and the intended run time?</i> | |
| Minimum detectable effect of key metric <i>What is the result of the power analysis for the key metric? How large does the impact need to be in order to reach statistical significance?</i> | |
| Experiment Settings <i>What will you use for the experiment settings? What is the minimum for calculating statistical significance? What will you use for the significance threshold (p-value)? Will you apply multiple comparison corrections (recommended)?</i> | |
| Expected Run Dates & Duration <i>How long will you run the test? (e.g. 7 days, 2 weeks, etc.) When do you expect to run the experiment?</i> | |

Experiment Information

| | |
|--|--|
| Experiment name <i>What is your experiment called?</i> | |
|--|--|



| | |
|--|--|
| Type of Experiment <i>Optimization, A/B/n, Validation/Do No Harm, Painted Door Research, Backend changes, Progressive rollout, UX</i> | |
| Experiment URL <i>What is the URL of the experiment configuration?</i> | |
| Status <i>What is the status of this project?</i> PRE-PRODUCTION/ QA & INTERNAL TESTING / RAMPING / EXPERIMENTING /100% RELEASED PERMANENT /KILLED / REMOVED FROM CODE | |
| Experiment Owner <i>Who is a good point of contact for more information?</i> | |
| Description <i>What is being changed, or what new feature are you testing?</i> | |
| Location of Experiment/Point of Exposure <i>Does it exist on a certain page / area of the product?</i> | |
| Treatment(s) <i>What are the variations of software treatments?</i> | |
| Primary Key Metric(s) <i>Which metric(s) will be impacted?</i> <i>What is your Overall Evaluation Criteria?</i> | |
| Secondary Metric(s) <i>Which metric(s) are most likely to be impacted?</i> | |
| Guardrail metric(s) <i>What are you monitoring to ensure has not been impacted?</i> | |



| | |
|---|--|
| Traffic Type and Allocation <i>What traffic type is being used for the experiment? e.g., account, if all users in an organization should see the same treatment</i> | |
| Targeting Criteria: Rules, Segments <i>Who will be included in the test? Are there considerations of cohorts or audiences? (e.g. account types, user designations)</i> | |
| Approvals <i>Who should be approving and launching your experiment?</i> | |
| Jira Ticket <i>Link to Jira url</i> | |
| Other notes and considerations <i>List any particular points of worry, especially in terms of circumstances that would entail shutting down the experiment, or considerations of other stakeholders or business owners.</i> | |




Design References

| | |
|--|--|
| Treatment A: <i>Add a screenshot for reference</i> | Treatment B: <i>Add a screenshot for reference</i> |
|--|--|



| | |
|--|--|
| | |
|--|--|

Metrics Impact and Results

| Metrics Scorecard | | | | |
|--|---|---|--|---|
| Metric Name <i>Overall Evaluation Criteria</i> | <Metric Value A> <i>Baseline or Control</i> | <Metric Value B> <i>Variation</i> | Δ% <i>Increase/ Decrease</i> | Significance    <i>p-value</i> <i>Desired, Undesired, or Inconclusive?</i> |
| Key Metric: <Name> | | | | |
| Key Metric: <Name> | | | | |



| | | | | |
|------------------------------------|--|--|--|--|
| Guardrail Metric: <Name> | | | | |
| Guardrail Metric: <Name> | | | | |

| Metric Export PDF | |
|--|--|
| <i>Attach a copy of the results exported from the Split app through "Share results" button</i> | |
| Export Date | |
| Split Version | |
| Targeting Rule | |
| Comparison treatment | |
| Compared against baseline treatment | |
| <i>Copy of exported results</i> | |

Conclusion and Learnings

| |
|--|
| Highlights and Learnings |
| <i>What did you learn? What are the key take aways? Other notes to share? Was there any significant degradation seen in any of the metrics (primary or guardrail)?</i> |
| <ul style="list-style-type: none">● Notes● Notes |
| Next Steps |
| <i>Did the results pass the decision criteria? Will this change/feature be rolled out to all users? What knowledge was gained that suggests future experiments in this area?</i> |
| <ul style="list-style-type: none">● Notes● Notes |