Experiment Doc Template

Experiment Proposal Section

Experiment Summary

Fill out the sections below at the start of experiment planning.

We plan to change [this part] of the product by altering [this property of it] for [this subset of users]. Our goal is to improve [a metric we care about], without [hurting this other metric that we also care about].

Experiment Owners

PM:
DS:
Eng:
Design:

Why?

State the Problem we are solving
What is the problem we are trying to address with this experiment?

Example: We have seen pledge conversion hold stagnant for the last 3 months, we are looking to drive up conversion in order to increase creators' earnings.

State the Hypothesis for this experiment
What underlying hypotheses about user behavior support these product changes?

We believe that [this thing] will have an impact on [the metric] because of [a belief which was informed by [this briefly presented or linked evidence]]. We will know whether our hypothesis was right when we observe [this change in a metric], and our hypothesis will be refuted if we observe [this other change].

State what we are Learning from this experiment

Why should these product changes be run as an experiment?

☐ We need to know the impact with precision

☐ This change has a potential downside

☐ What we learn from this change will determine/impact future product decisions

Are there past results that informed this proposal?

- Experiment result link
- Analysis link
- External research link

🚨**Do not continue to fill out this doc until you have completed the above section and gotten tripod alignment (see checklist below)** 🚨

Experiment Plan

Make sure the sections below are filled out before the experiment is launched.

Experiment Design

Randomization Unit	Fill in how the experiment will be triggered - i.e. what is the identifier that we will use to randomize the participants: This is an experiment on creators / patrons /site visitors / email recipients; identified by user_id / device_id / campaign_id
Platform	Mobile Android: yes/no Mobile iOS: yes/no Mobile Web: yes/no Web: yes/no
Eligibility Criteria	What does someone have to do to be included in the experiment? Do they need to satisfy a certain segment like language, locale, currency, new/existing, etc.?
Assignment Criteria	At what point does a user/device get bucketed into

	either the control or the treatment group?
	Please have data science and engineering sign off on the QA'd assignment criteria prior to launching the experiment to avoid any errors.
Response Metric	What will we measure for each bucket of the experiment – for example: pledge amount, number of patrons, etc. If the response rate is conversion then what is the length of the conversion window?
	Provide a link to the Amplitude chart that you have pre-created for monitoring the response.

Description of Variants

The most minimal experiment that can validate our hypothesis (i.e. sufficient test) involves changing [this part] of the product in [these ways]. We are also interested in changing [this other thing], but will save it for a follow-up based on learnings from this experiment.

[Here is the link to our sketches/wireframes/screenshots/PR]

Runtime

Lift
The minimum lift we want to detect is <X%>, which corresponds to an increase in the response metric from <Y%> to <Z%>. (There is a tradeoff between precision and runtime - to detect a smaller lift will require a longer runtime.)

How long will the test take to detect the lift at the planned variant split? (Link to Runtime Calculator. On high-stake, complicated experiments, you can also

simulate the experiment.)

This test should need to run for Z days. We think that Y users/devices/creators/dollars will be exposed.

[Screenshot of the runtime calculator here]

When do we want to run the experiment? Are there other experiments that may interact with it?

This experiment will run from YYYYMMDD to YYYYMMDD. It needs to run after this other related experiment.

Risks

Could this experiment have adverse effects that we're not measuring directly? Does this require input from creator-facing teams?

We should remember to segment by X to properly understand results. We should check in with team Y while implementing this.

Post-Experiment Decisions

What are the conditions under which we will ship or not ship this change?

We will ship the Puppy variant if it has a statistically significant lift over the control group. Otherwise it is not worth the additional product complexity brought about by the puppies.

Are there followups that depend on the results of this experiment?

If the puppy variant is successful, we will explore more experiments on bones. If the puppy variant is flat, we will shift our focus to kittens.

Results

What were the results of this experiment? Which version won, and by how much?
What is the final recommendation and conclusion? Any followups needed?

Metrics assessment - were any additional metrics impacted by this change?

(Include link to metrics template)

Links

- Link to analysis, (when completed)
- Link to follow up experiments
- Links to PRs

Experiment Checklist

Before you start implementing an experiment (PM, DS)

☐ Fill out experiment proposal section (PM). You may want to check in with the research team to see if there's past research that could inform this experiment. (Link to scoping the purpose of an experiment.)

☐ Align with tripod on why this should be run as an experiment and that it is feasible to build as an experiment(PM)

☐ Fill out experiment plan section (PM)

 ☐ Consult with DS & Research on the eligibility and response metric definition

 ☐ Create an amplitude dashboard with baseline conversion for the response metric. (Link to the template.)

 ☐ Add a screenshot of the runtime calculator based on metrics baseline

☐ Share concept and experiment doc in #experiments . Add experiment to go/experiment-tracking, where you can also check for interactions with other

teams' experiments (PM).

☐ Get DS sign-off on concept and check with product marketing, CH, and/or creator-facing-teams on experiment risks (PM). See this guide to cross-functional experiments for a list of potential stakeholders.

☐ If your experiment is high user impact, you should let the content support team know about the experiment and the potential implications, risks, or users questions by submitting a content request at go/contentsupport

Turning on experiment (Eng)

☐ Implement experiment

☐ Implement logging

☐ QA logging

    ☐ QA logging for the response metric

    ☐ QA logging needed for analysis

    ☐ Make a version of the Amplitude experiment tracking template for the specific experiment (Eng or PM)

☐ Update experiment doc with final product screenshots/gifs

☐ Update experiment doc with start and end dates; set slack reminder for end date

☐ Activate the experiment (go live!) by creating a version

☐ Spot-check amplitude dashboard

☐ Notify #community-happiness and tag @datascience with a screenshot and description of experiment once the experiment is live

Decision and communication (DS, PM)

☐ Communicate timeline for when the analysis will be complete to the PM (DS)

☐ Analyze results and share with team (DS)

☐ Complete experiment doc with results (PM)

☐ Share what was learned alongside the rollout decisions to #experiments and possibly with #analytics (PM)

    ☐ If the experiment was particularly fascinating, buggy, disastrous, amazing - schedule time to bring insights to #product-brain and to Experiment Traffic Control (ping #experiments).

Cleanup after the experiment ends (Eng)

- [ ] Archive the experiment in the dashboard
- [ ] Clean up the code in the frontend and backend