



The Path to 10,000 Experiments

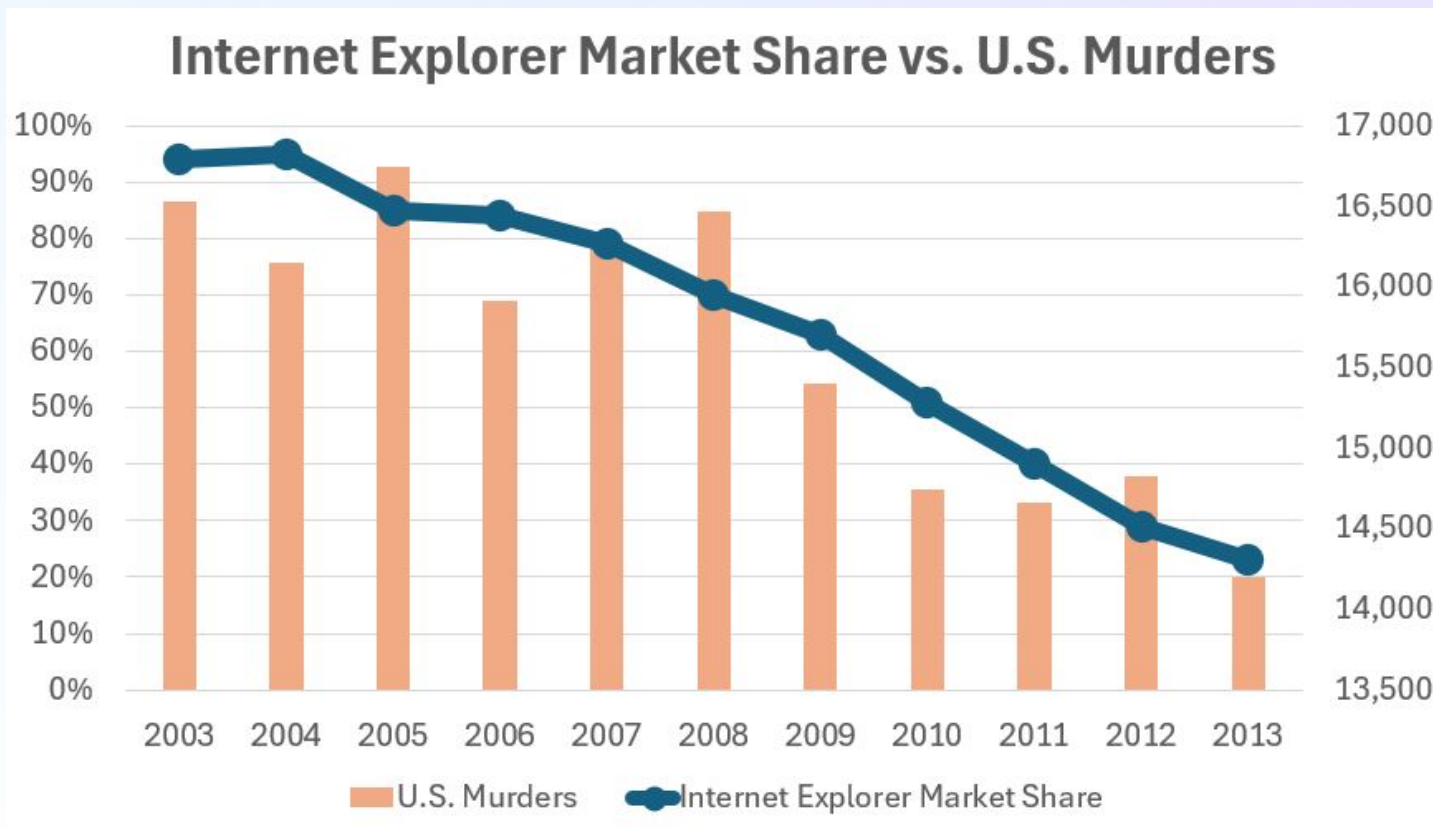
LDX3 2025

Aaron Silverman

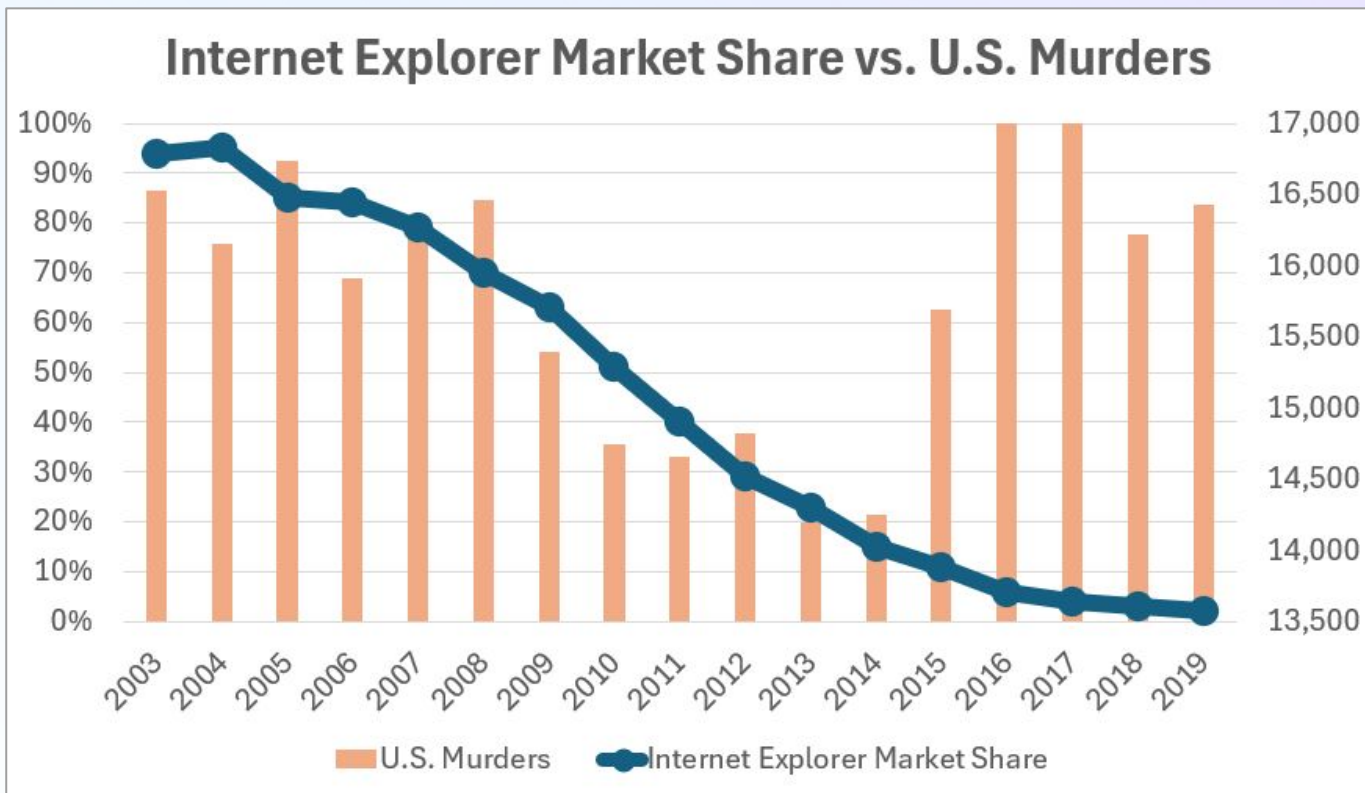
eppo

 DATADOG

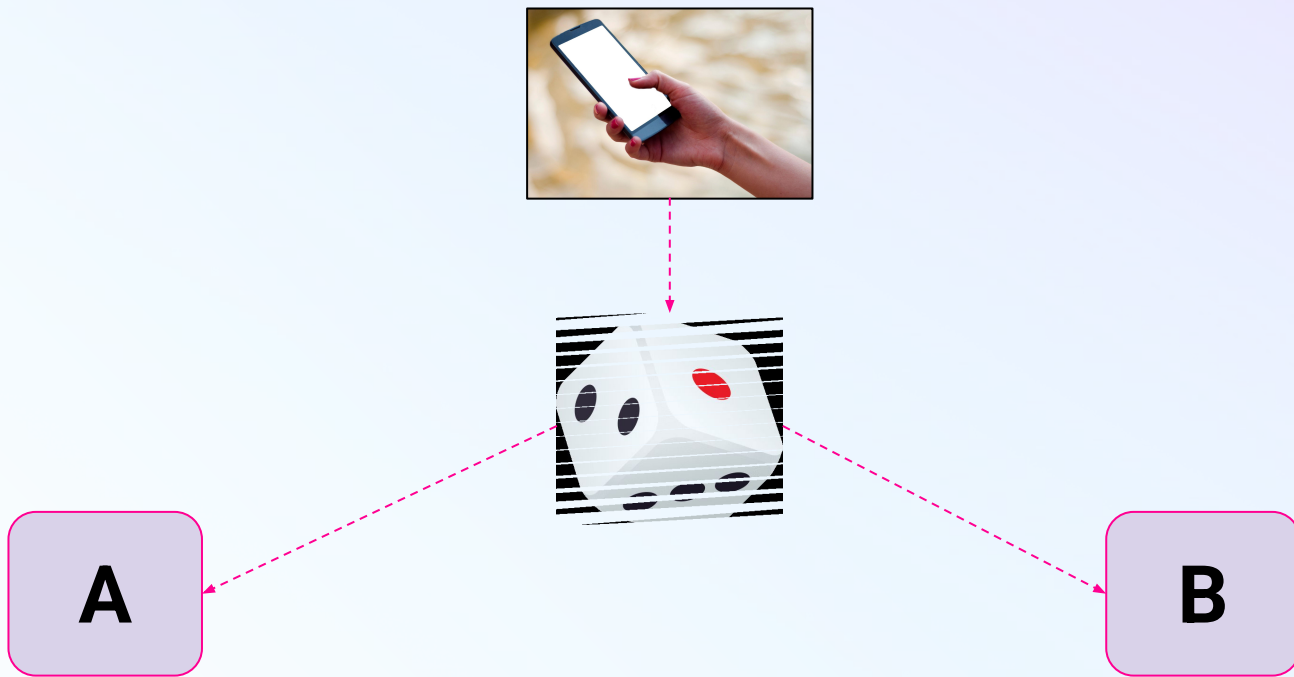
It's natural to look for relationships



Correlation does not imply causation



Establish causation with randomized controlled trials



**If you are not testing
you are not learning!**

Launching today: Skito Burrito



Time to start testing!

We'll start with simple A/B tests

Bye Bye Hunger,
Hello Burritos!

Download
App

Bye Bye Hunger,
Hello Burritos!

Download
App

A/B test our social posts



Skito Burrito

Don't let hunger stop
you from shredding
the gnar!



Skito Burrito

Don't let hungry kids
ruin your family ski
day!

A/B test email campaigns



hi@skitoburrito.com

You're about to get
tasty burritos fast!



hi@skitoburrito.com

You're about to get
nutritious burritos fast!

A/B test promotions



50% Off!


Start
Order




Buy One
Get One
Free!

Start
Order

A/B test new features




Most
Ordered




Meat Lovers

Order
Now



Recommended
For You



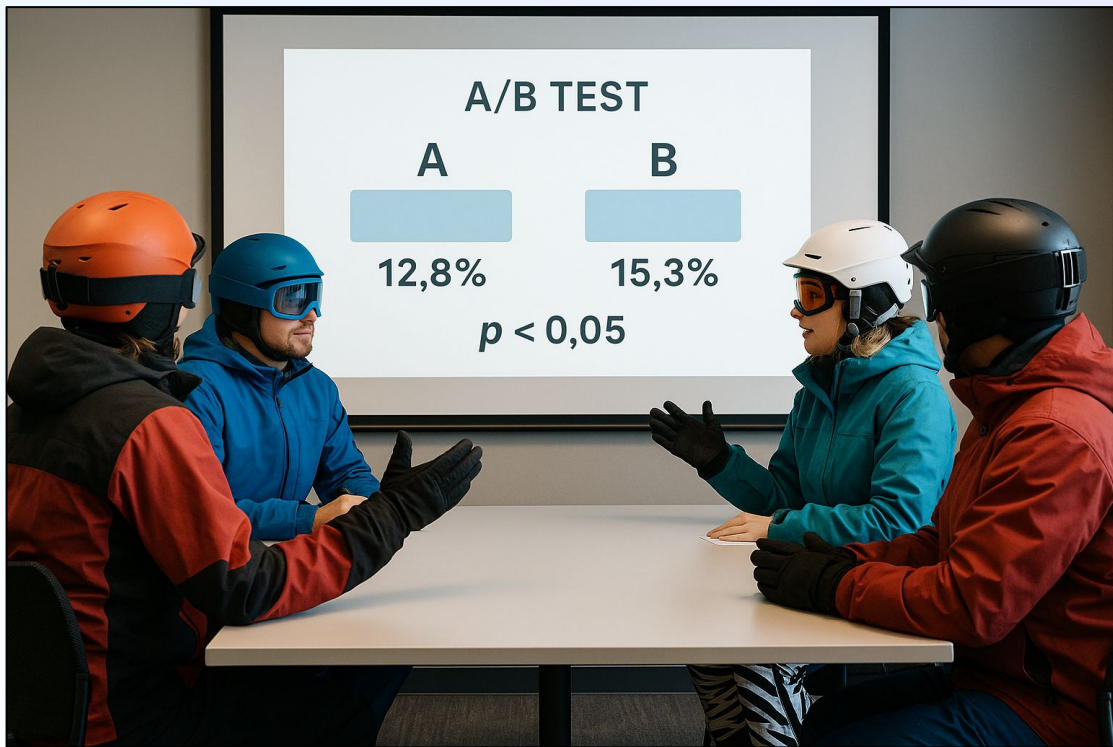
Vegetarian

Order
Now

***“The best way to have
a good idea is to have
lots of ideas”***

**–Linus Pauling
Two-time Nobel Prize winner**

Make experimentation part of company culture



**Tests are only as good
as their data and
evaluation**

Automated diagnostics for clean data

Experiment Check-o-matic:

- ✓ Traffic is balanced across variants
- ✓ Subjects are only being exposed to one variant
- ✓ Pre-experiment data is similar across variants

(plus 20 additional checks...)

Experiment platforms for consistency



Default settings and decision flows

Run Time

☒ Set a default run time

14 days

☒ Set a minimum threshold before the experiment status can be "re"

Min. duration 7 days

Min. sample size 500 subjects per variant

The min. sample size cannot be less than 100 subj

Stats Analysis Plan

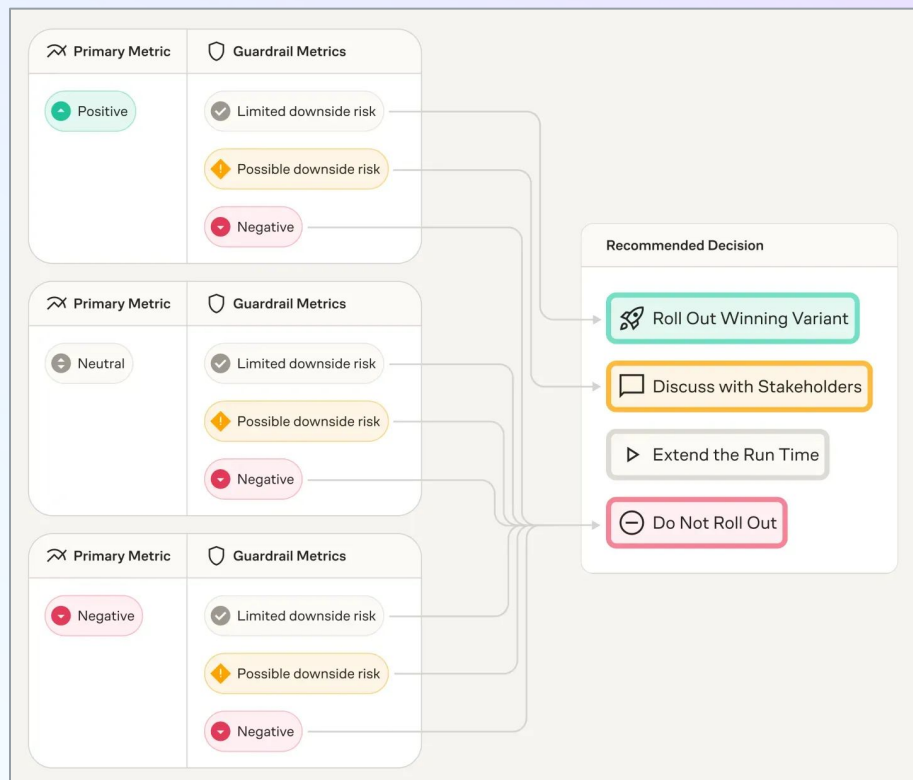
* Confidence Interval Method ⓘ

Sequential

* Confidence Level ⓘ

95%

☒ CUPED ⓘ



Things are going well!

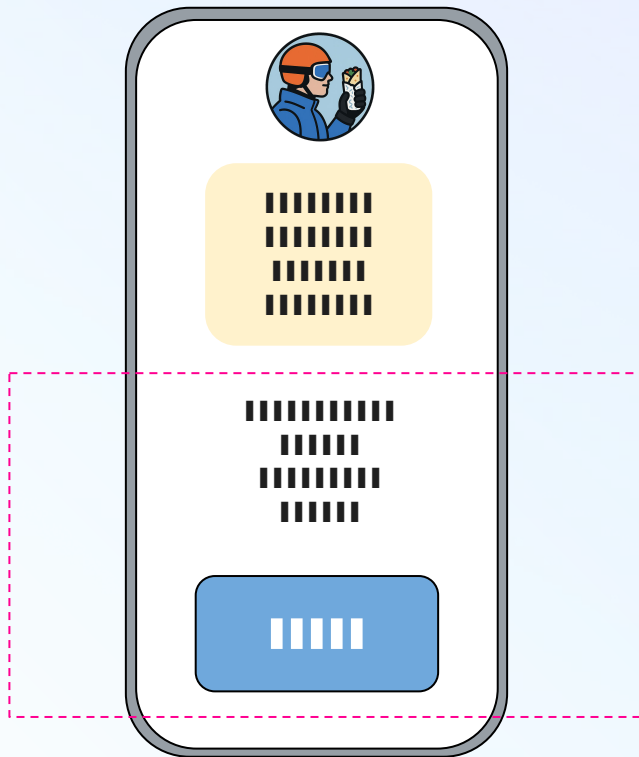


Time to go beyond simple A/B tests

Holdouts to measure cumulative impact of our experimentation



Mutual exclusion layers to avoid unwanted experimentation conflicts



Checkout Promotion Layer

Contextual bandits to personalize the best choice based on the user

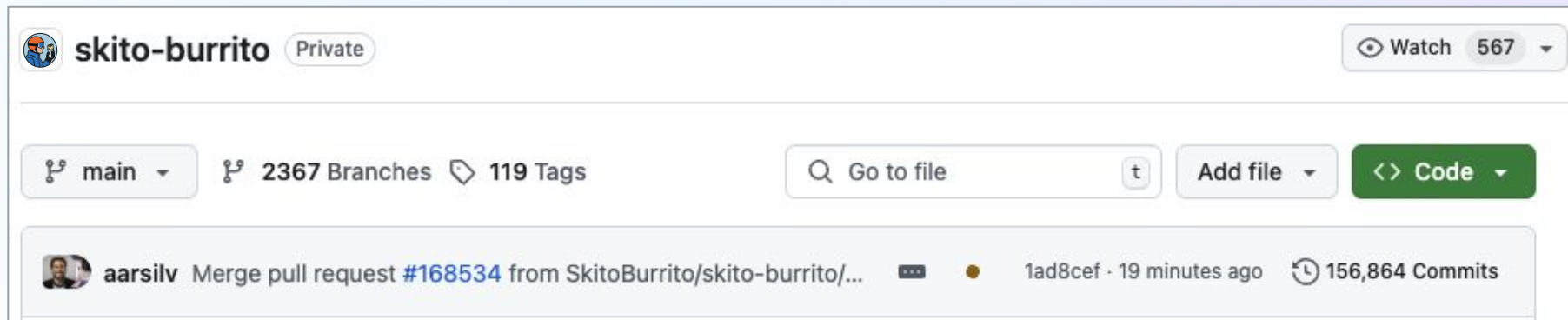


**This is not as hard as it
sounds**

We've hit it big!

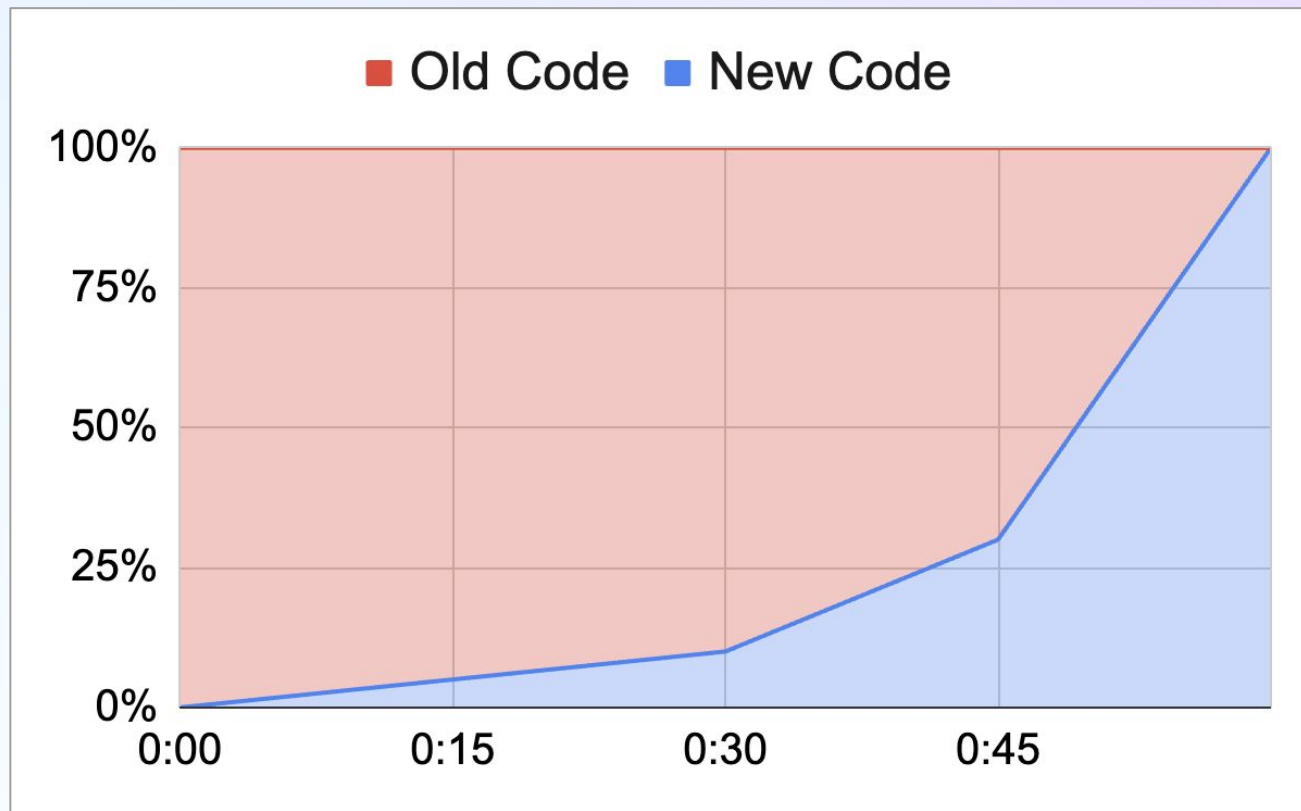


Our global product development team is shipping constantly!



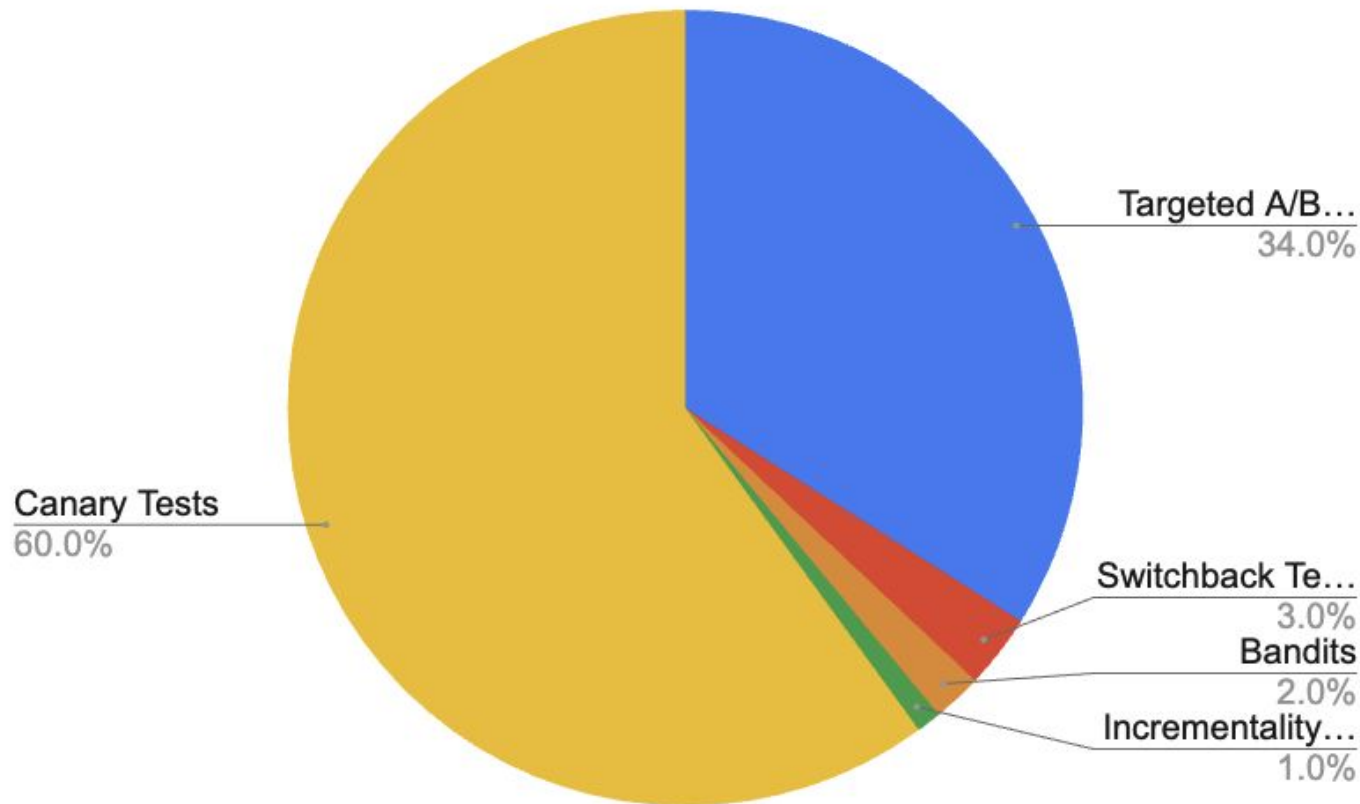
**Need to move fast,
safely, at scale**

Canary testing to quickly catch problems



**Every change should
be an experiment**

Skito Burrito's 10,000 experiments



Experimentation culture is a critical part of sustained success



**Build experimentation
muscles and let
tooling accelerate you**

Thanks!

eppo

