Advanced Feature Flagging: It’s All About The Data

DAVE KAROW

Continuous Delivery Evangelist

split

FEATURE FLAG REVIEW

A quick review of feature flags and rollout strategies

01

HOW YOU MEASURE MATTERS

Don’t believe everything you see

02

FLAGS + DATA = EXPERIMENTATION

Attribution, calculation and analysis, automatic 24x7

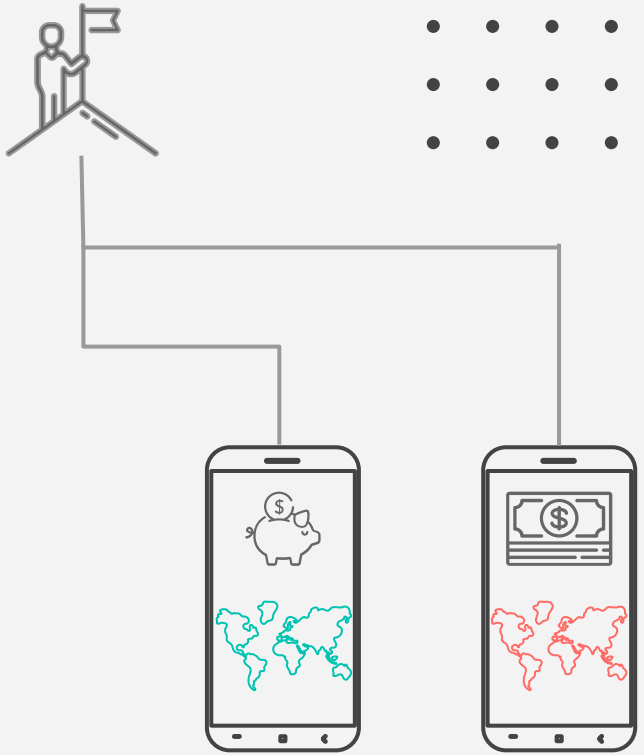
03

FEATURE FLAGS

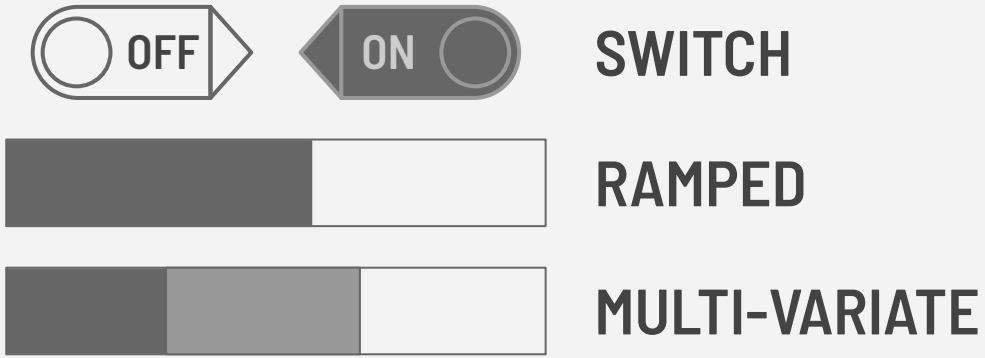
DEPLOY	RELEASE
When you want to	When you are ready

`if{} else{}` STATEMENT

Separates code deploy from feature release



TYPES OF FLAGS



ROLLOUT FRAMEWORK

SPEED
How quickly can we reach a decision?

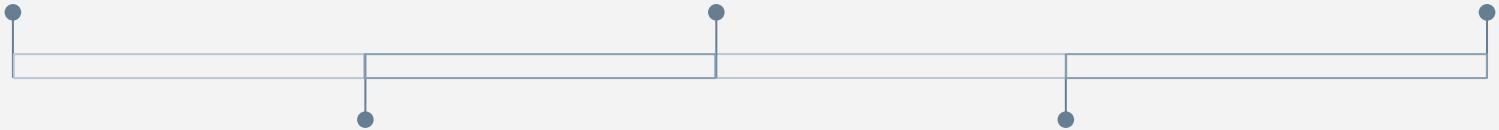
QUALITY
How confident can we be of that decision?

RISK
How can we minimize bad outcomes?

CREDIT
SQR: Balancing Speed, Quality and Risk in Online Experiments
Ya Xu, et.al

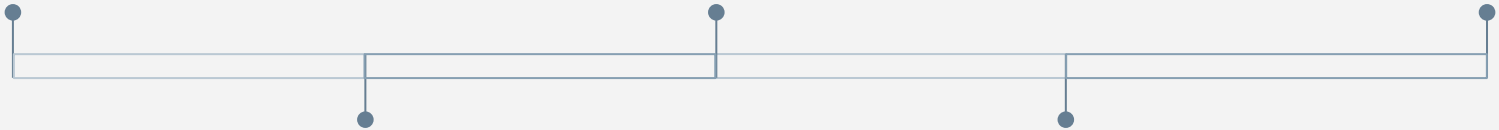
PHASES OF ROLLOUT

DEPLOY
Code deployed,
no exposure



PHASES OF ROLLOUT

DEPLOY
Code deployed,
no exposure



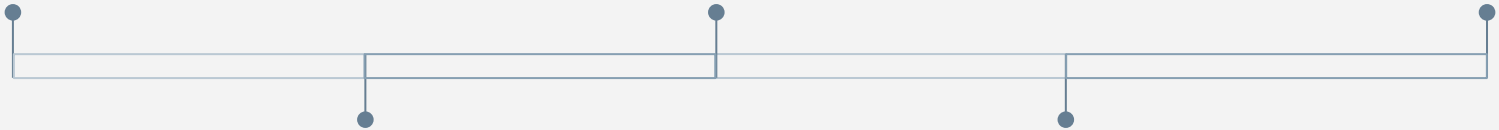
ERROR MITIGATION
1-50% Ramp
Identify bugs/crashes

PHASES OF ROLLOUT

DEPLOY
Code deployed,
no exposure

ERROR MITIGATION
1-50% Ramp
Identify bugs/crashes

MEASURE
Maximum Power Ramp
Understand impact



PHASES OF ROLLOUT

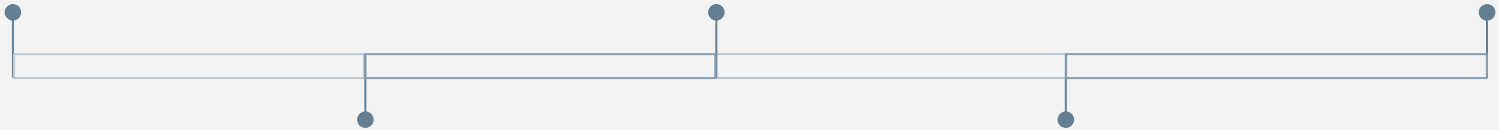
DEPLOY
Code deployed,
no exposure

ERROR MITIGATION
1-50% Ramp
Identify bugs/crashes

MEASURE
Maximum Power Ramp
Understand impact

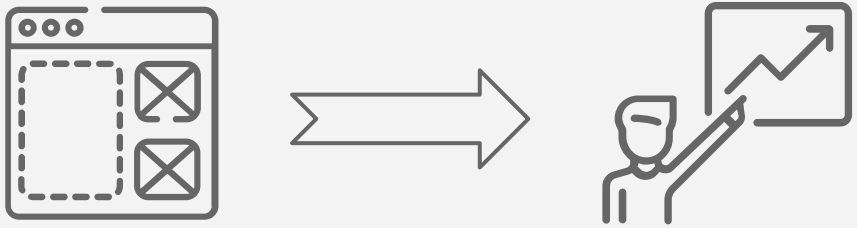
SCALE MITIGATION
50-100% Ramp
Identify scaling issues

RELEASE
Complete rollout



HOW YOU MEASURE MATTERS

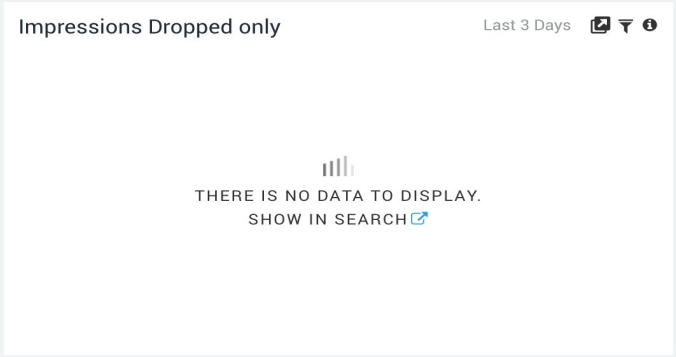
DON’T BELIEVE EVERYTHING YOU SEE...

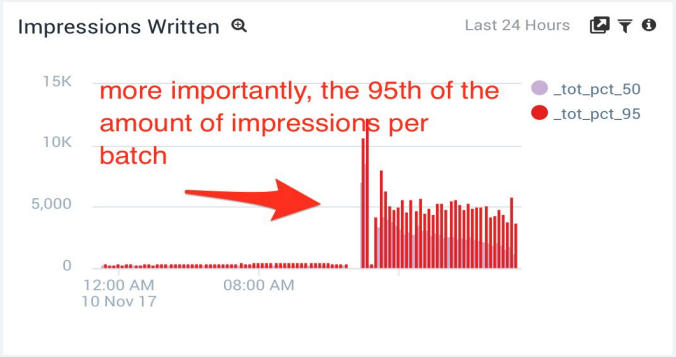


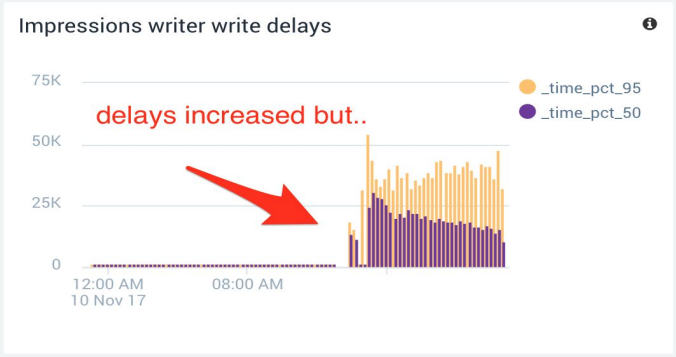
New Release

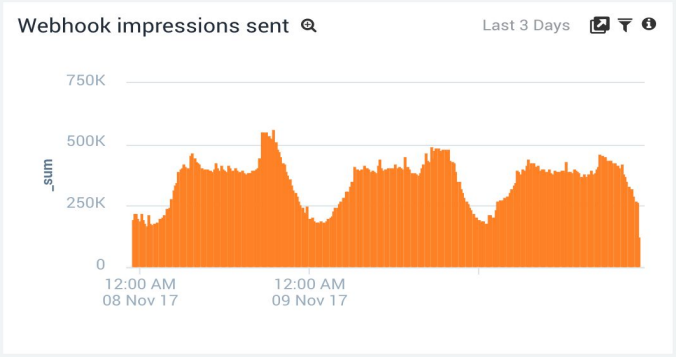
Metrics Change

“Can’t we just change things and monitor what happens?”









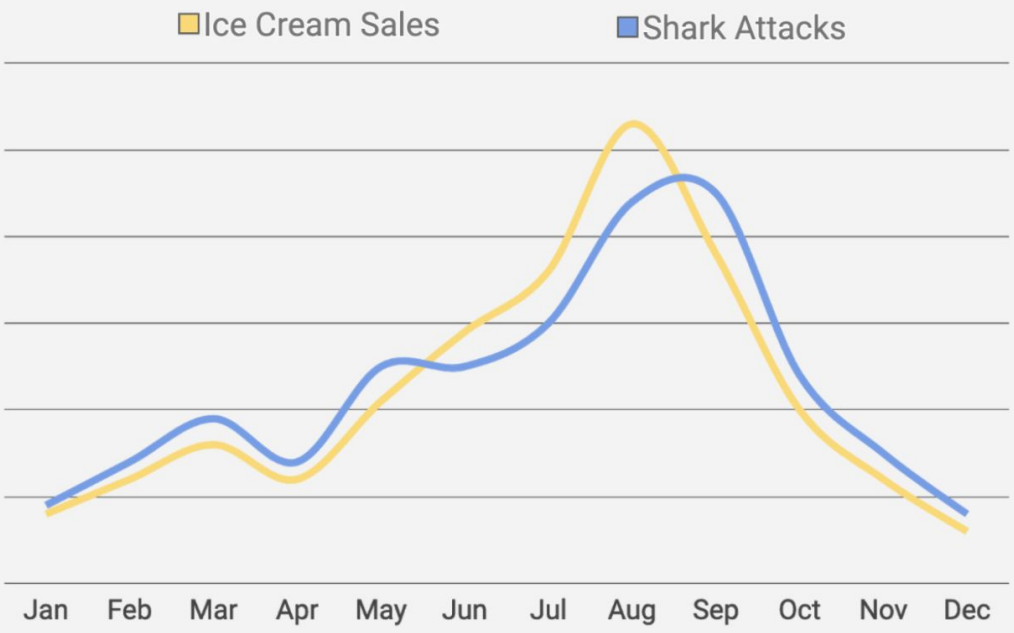
SYSTEMS DASHBOARD

FEATURE ENABLED TO 100%

FEATURE ENABLED TO 100%

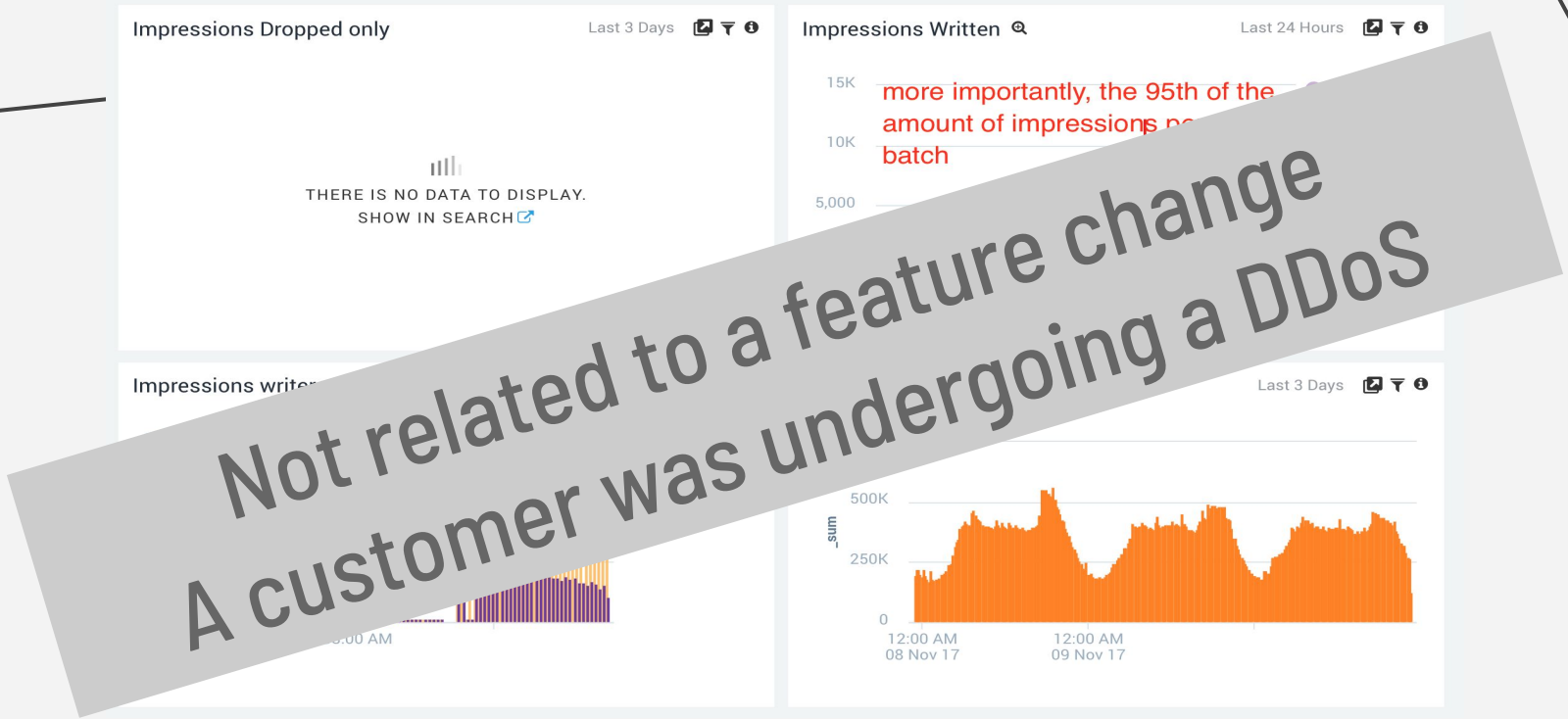
FEATURE ENABLED TO 5%

CORRELATION IS NOT CAUSATION



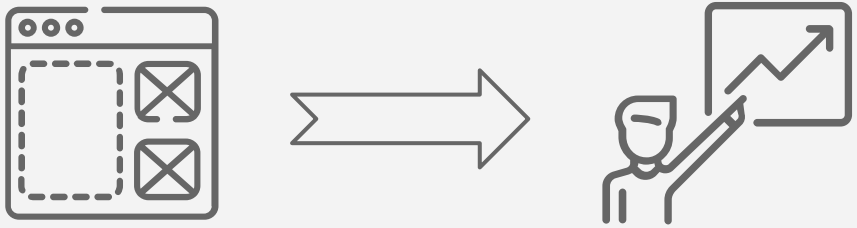
Both ice cream sales and shark attacks increase when the weather is hot and sunny, but they are not caused by each other

They are both caused by good weather, with lots of people at the beach, both eating ice cream and swimming in the sea



SYSTEMS DASHBOARD

DON’T BELIEVE EVERYTHING YOU SEE...



New Release

Metrics Change

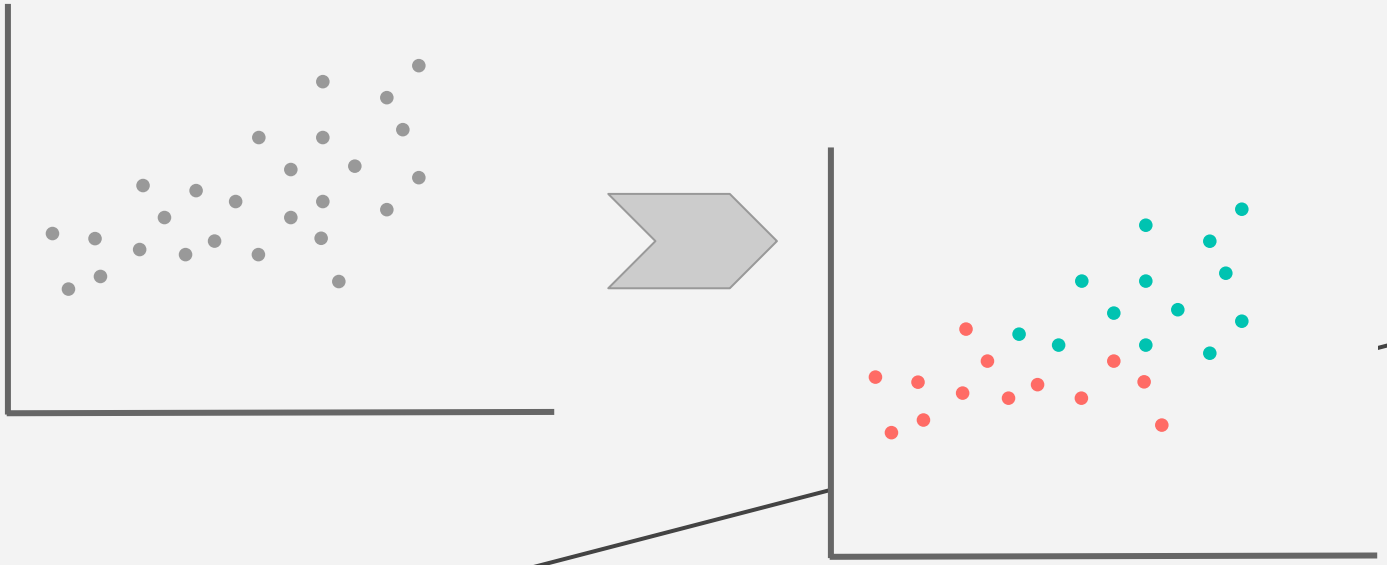
“Can’t we just change things and monitor what happens?”

DON’T BELIEVE EVERYTHING YOU SEE...

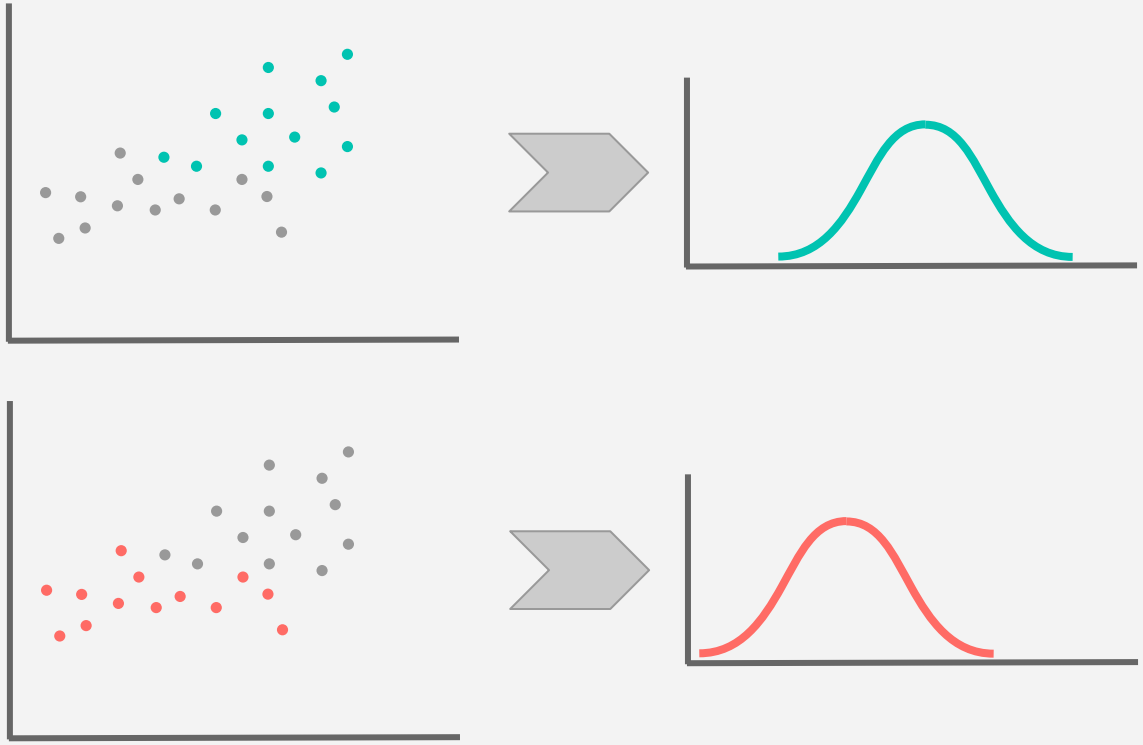
- Product changes
- Marketing campaigns
- Global Pandemics
- Nice Weather

MEASURING CAUSALITY

ATTRIBUTION



CALCULATION



ANALYSIS

TAGGING METRICS

Most dashboarding tools allow tagging of data for segmentation

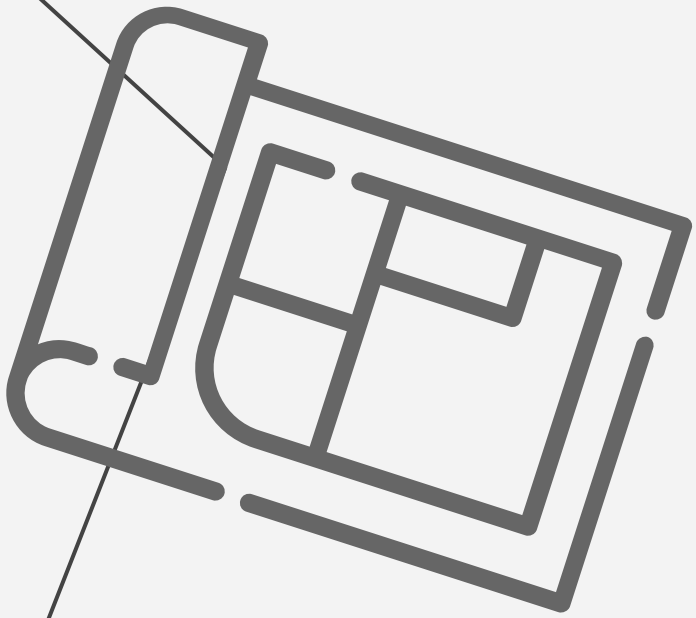
AD-HOC ANALYSIS

Storing feature attribution in BI database for querying

EXPERIMENTATION PLATFORM

Homegrown or SaaS offerings provide both data collection and statistical analysis

FLAGS + DATA = EXPERIMENTATION

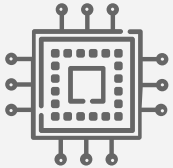


ANATOMY OF AN EXPERIMENTATION PLATFORM

TARGETING SYSTEM

TELEMETRY SYSTEM

STATISTICAL ENGINE

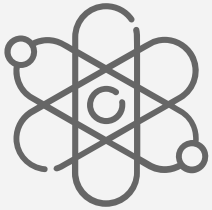


MANAGEMENT CONSOLE

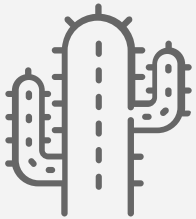
CREDIT
Understanding Experimentation Platforms [Aijaz, Stuart, Jewkes]

TARGETING SYSTEM

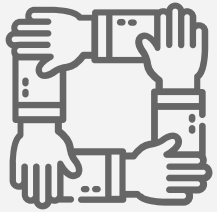
FAST



RANDOM



STICKY

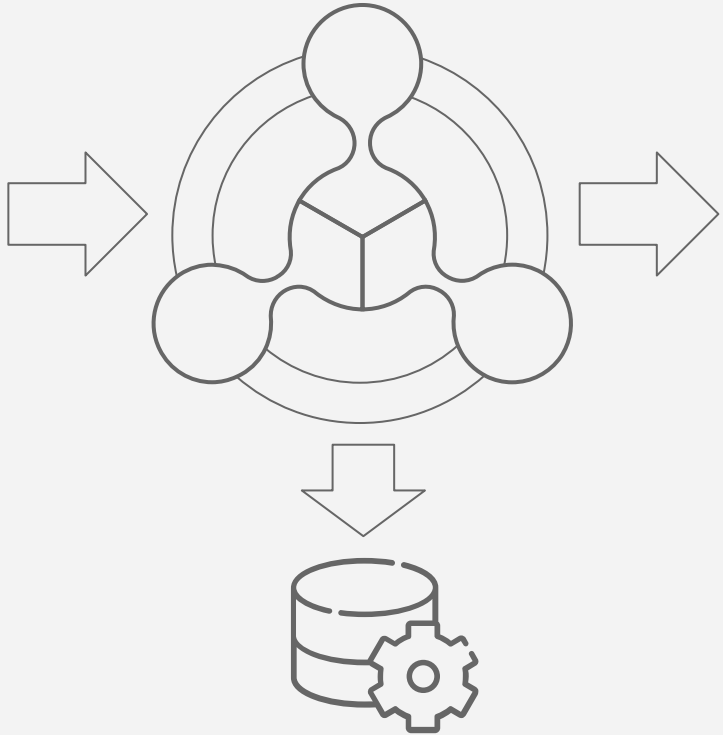


RELIABLE

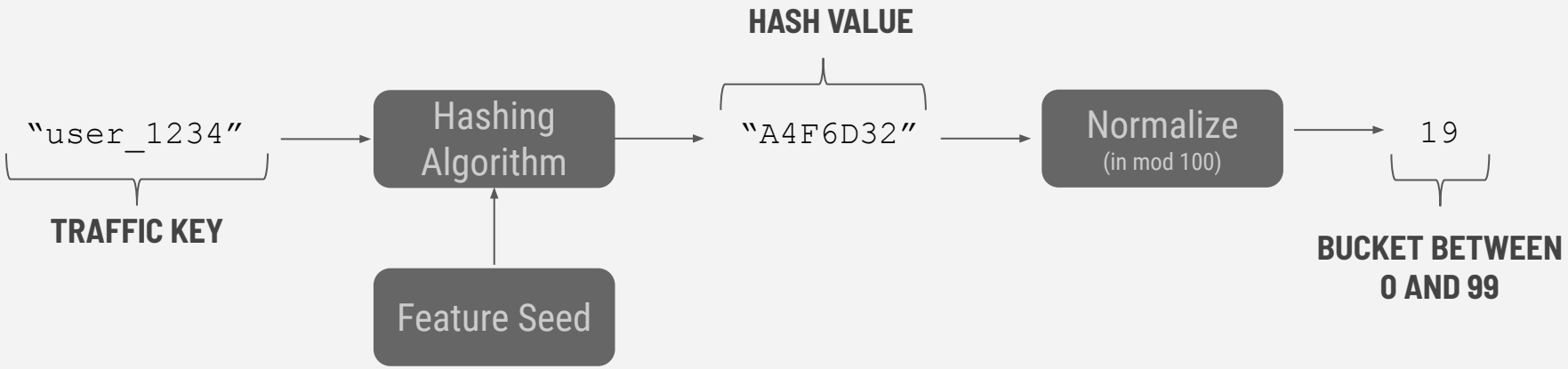
TARGETING SYSTEM

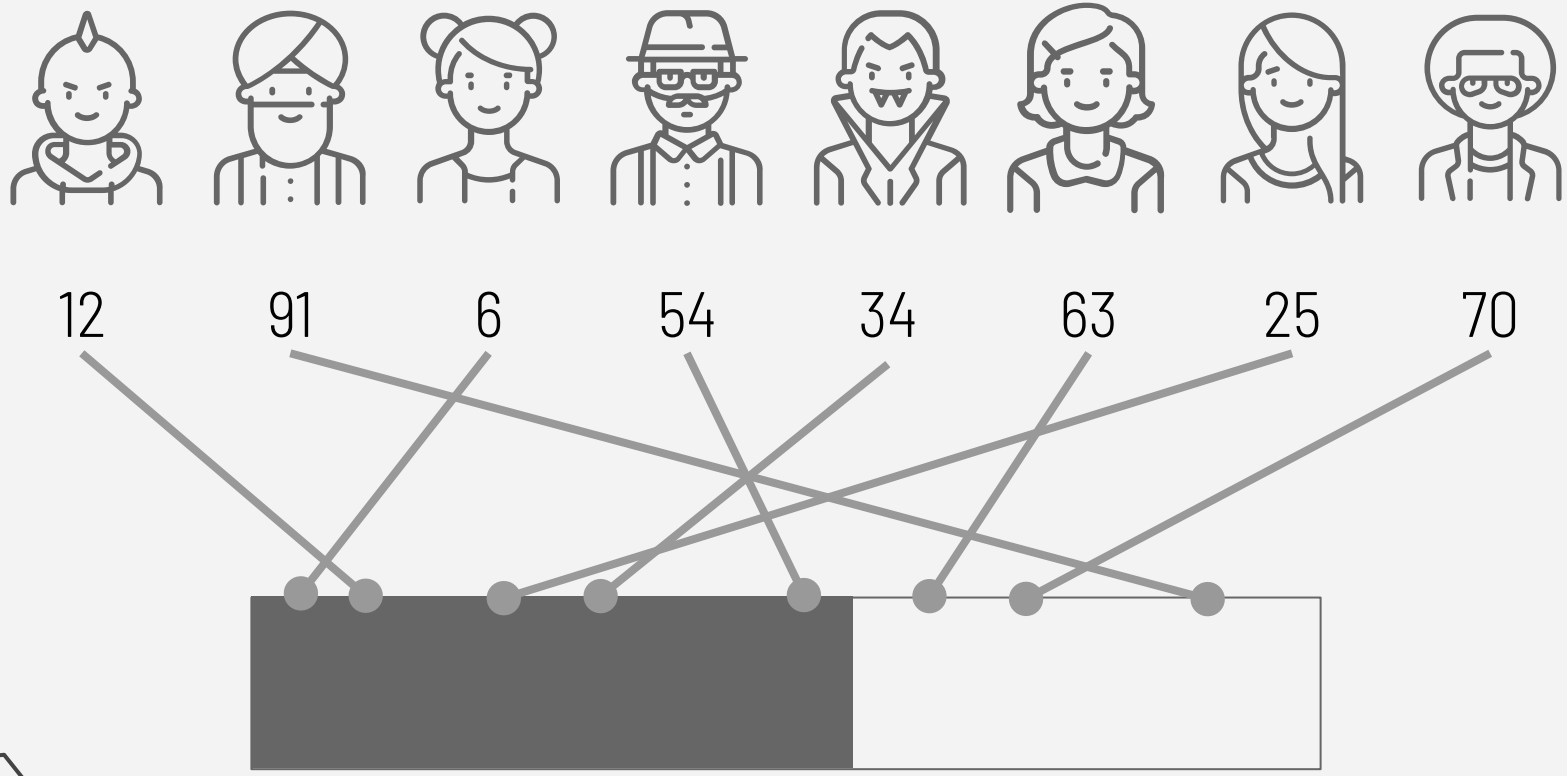
FEATURE

TRAFFIC KEY



ASSIGNMENT


HASH VALUE
"user_1234"
Hashing Algorithm
"A4F6D32"
Normalize
(in mod 100)
19
TRAFFIC KEY
Feature Seed
BUCKET BETWEEN 0 AND 99


12
91
6
54
34
63
25
70

TELEMETRY SYSTEM

track (String key, String eventType, double value)

WRAPPER

Centralize tracking across tools

Ensure reliable delivery

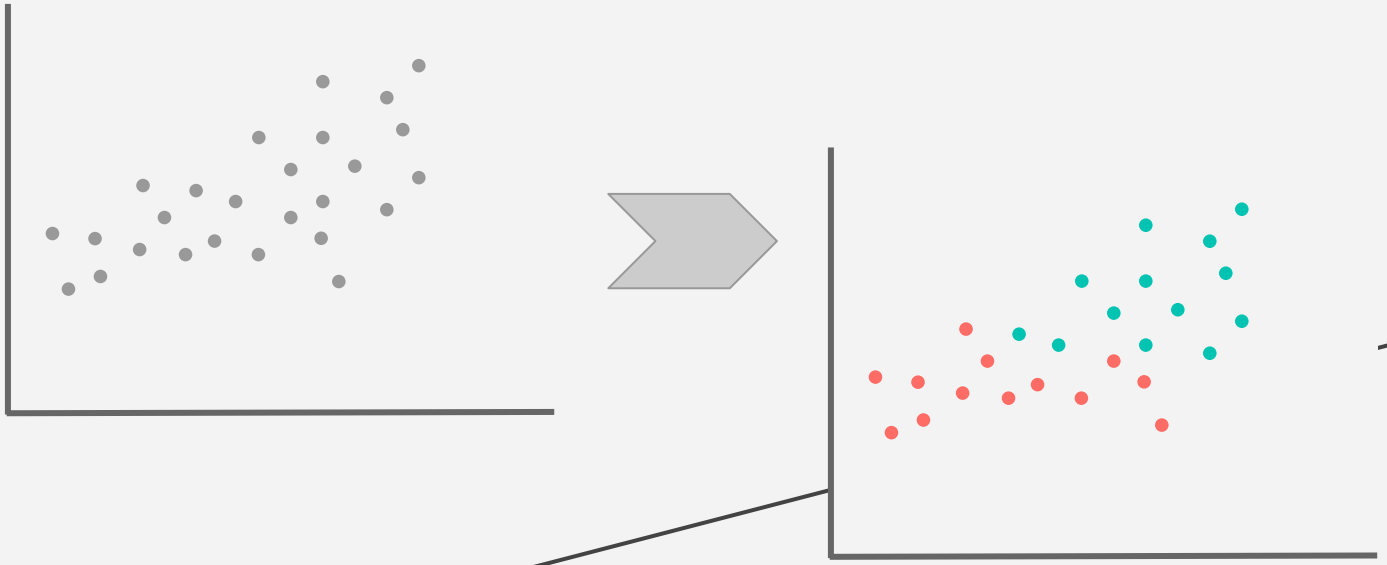
Identify and exclude malicious traffic

SERVICE

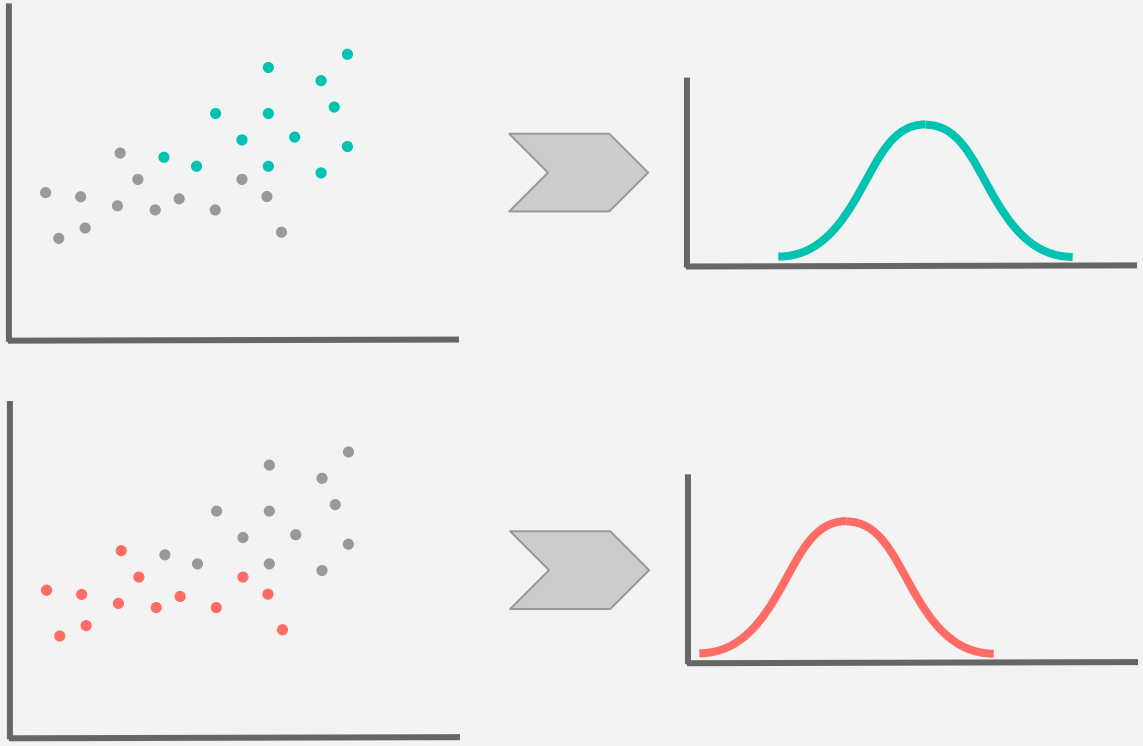
Receive batches of events

Store in warehouse for processing

STATISTICAL ENGINE

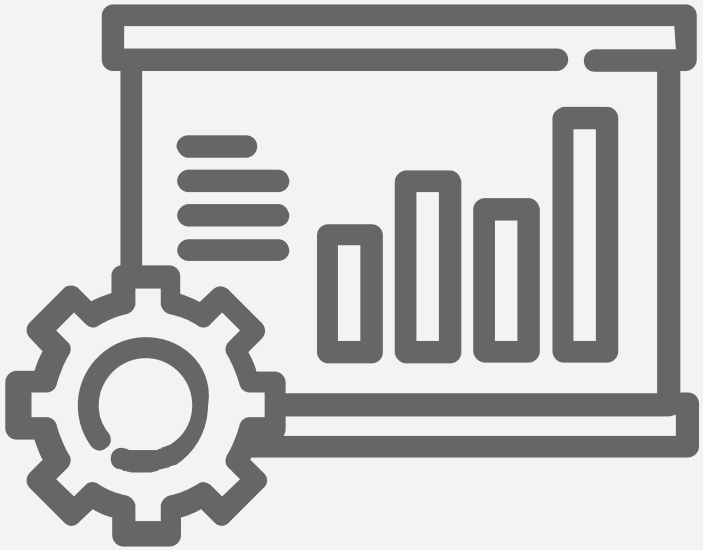


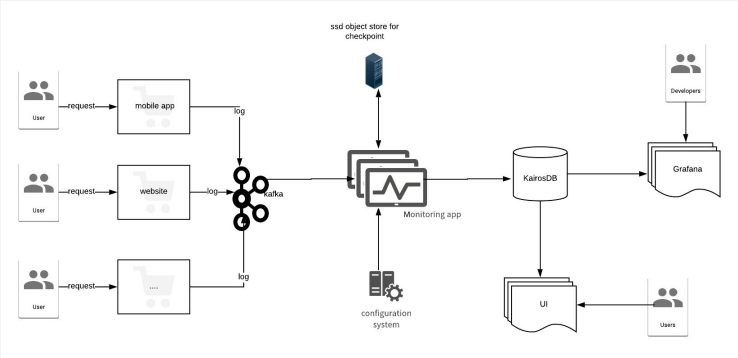
STATISTICAL ENGINE



STATISTICAL ENGINE

MANAGEMENT CONSOLE





Test to learn

- **Purpose:** Understand customer behavior and validate or invalidate a hypothesis
- Feature likely only launches if it is better than the production experience
- Ex. Guest Cart

Test to launch

- **Purpose:** Mitigating risk by phasing out the rollout to customers, and ensuring no bugs are introduced
- Feature is necessary to launch for the business. Will launch if it is doing no harm
- Ex. ADA

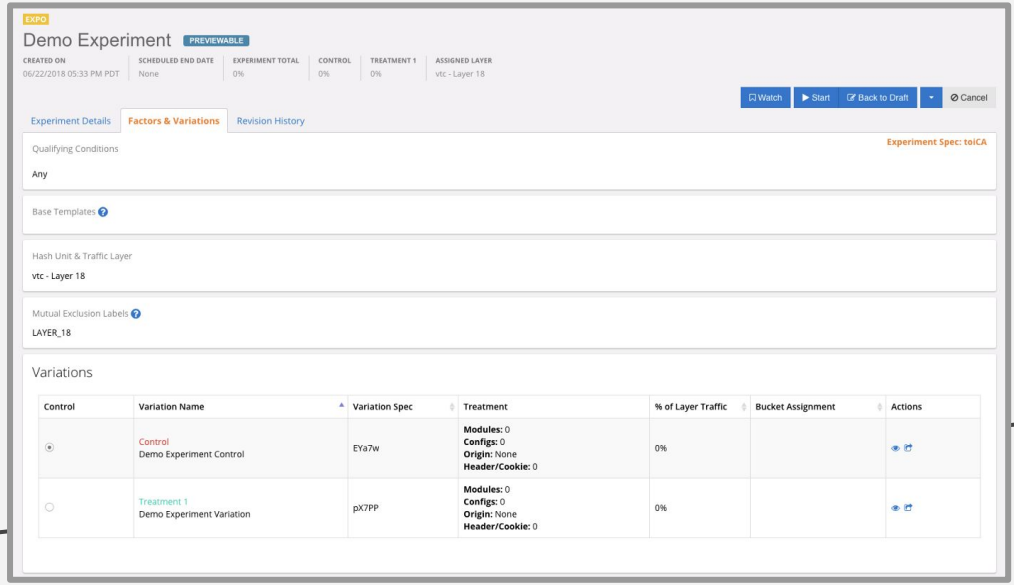
It's All About the Data

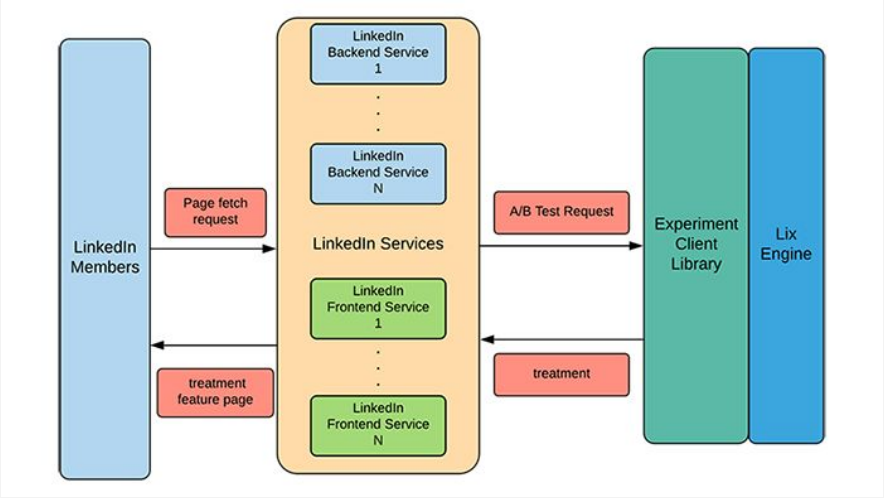
When someone runs an A/B test, they expect results, and whether they are positive or negative, they need to be correct. One challenge we faced in building our own platform was building trust in the results. With the support of a strong product analytics team, and a revamped data pipeline, we were able to achieve that.

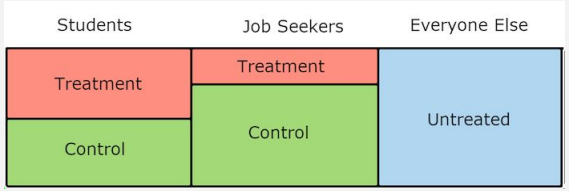
Experiment Results in Expo UI

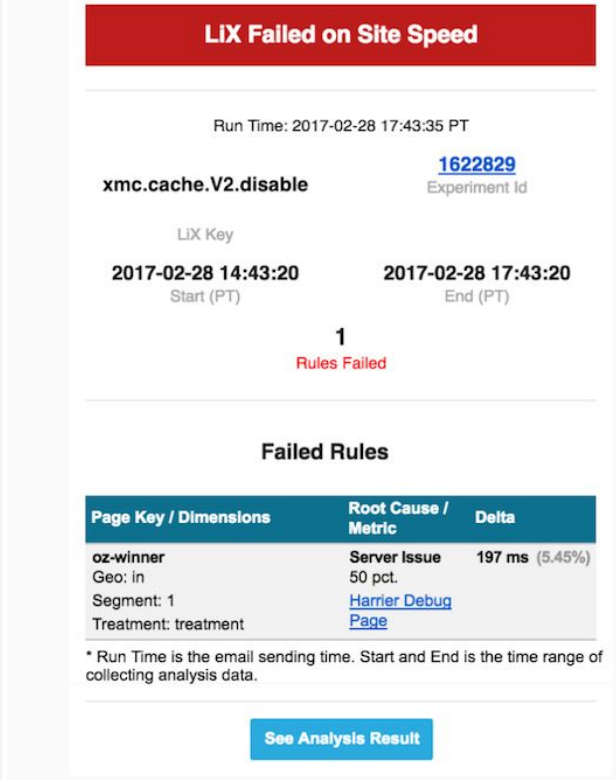
EXAMPLE IMPLEMENTATION

Walmart EXPO



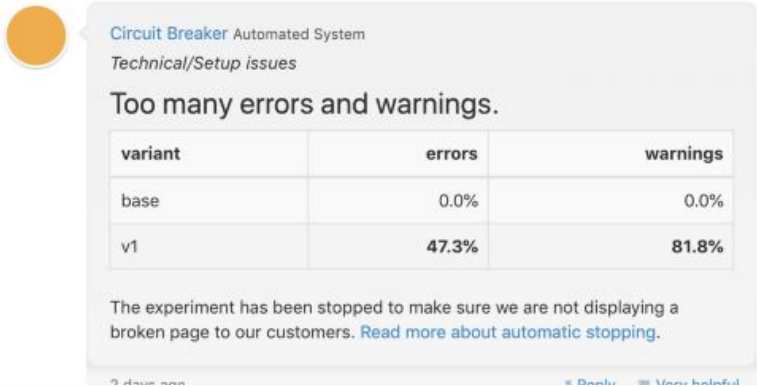


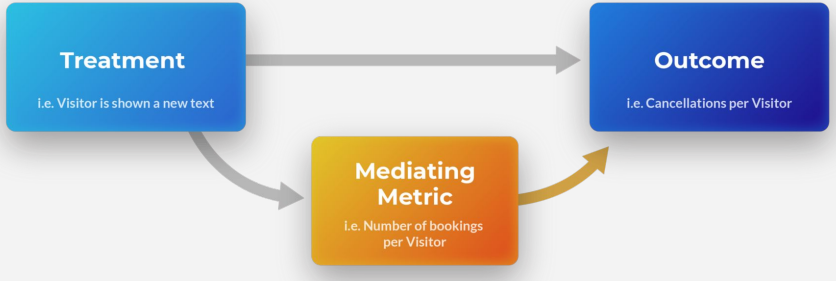




EXAMPLE IMPLEMENTATION

Linked In LiX

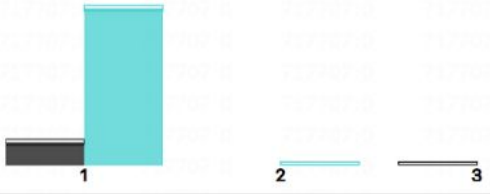




request_warning_count - Number of warnings triggered

code link

717707 (1h9m)	Exp. visitors	Goal server requests	Data validity	Sum	Exp. average	Goal average	Bar Graph
base (50%)	**31 167** (49.66%)	**22** (13.41%)	[BASE]	24 warnings	770.0452×10^{-6} warnings	1.0909 warnings	
variant 1 (50%)	**31 594** (50.34%)	**142** (86.59%)	**More data points** in this variant. This is **bad**, and the values are **no longer comparable.**	143 warnings	4.5262×10^{-3} warnings	1.007 warnings	



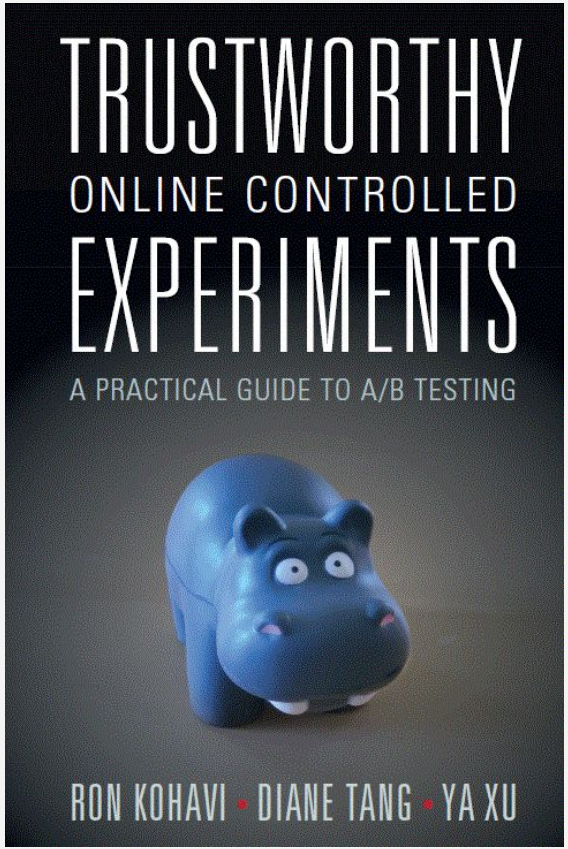
EXAMPLE IMPLEMENTATION

Booking.com

LESSONS LEARNED

@ Microsoft + Google + LinkedIn

https://experimentguide.com/



split®

KILL THE RELEASE NIGHT

Decoupling deployment from release make release rituals go away.

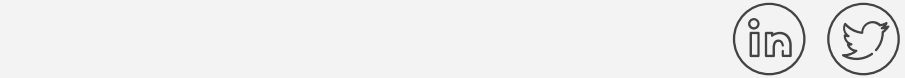
AUTOMATE DELIVERY WITH DATA

Independent feature rollouts are orchestrated by data, not people.

EVERY FEATURE IS AN EXPERIMENT

Targeted feature rollout allows for rapid A/B tests and customer insights.

THANKS!



Let’s move to discussion!

linkedin.com/in/davekarow
dave.karow@split.io
@davekarow

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**