Decision model, meet the real world

Testing optimization models for use in production environments

September 20, 2023

👋 Hello and welcome

Ryan O'Neil
CTO at Nextmv

Nicole Misek
Engineering VP at Nextmv

Stories to get started
We'll set the stage with stories from the field

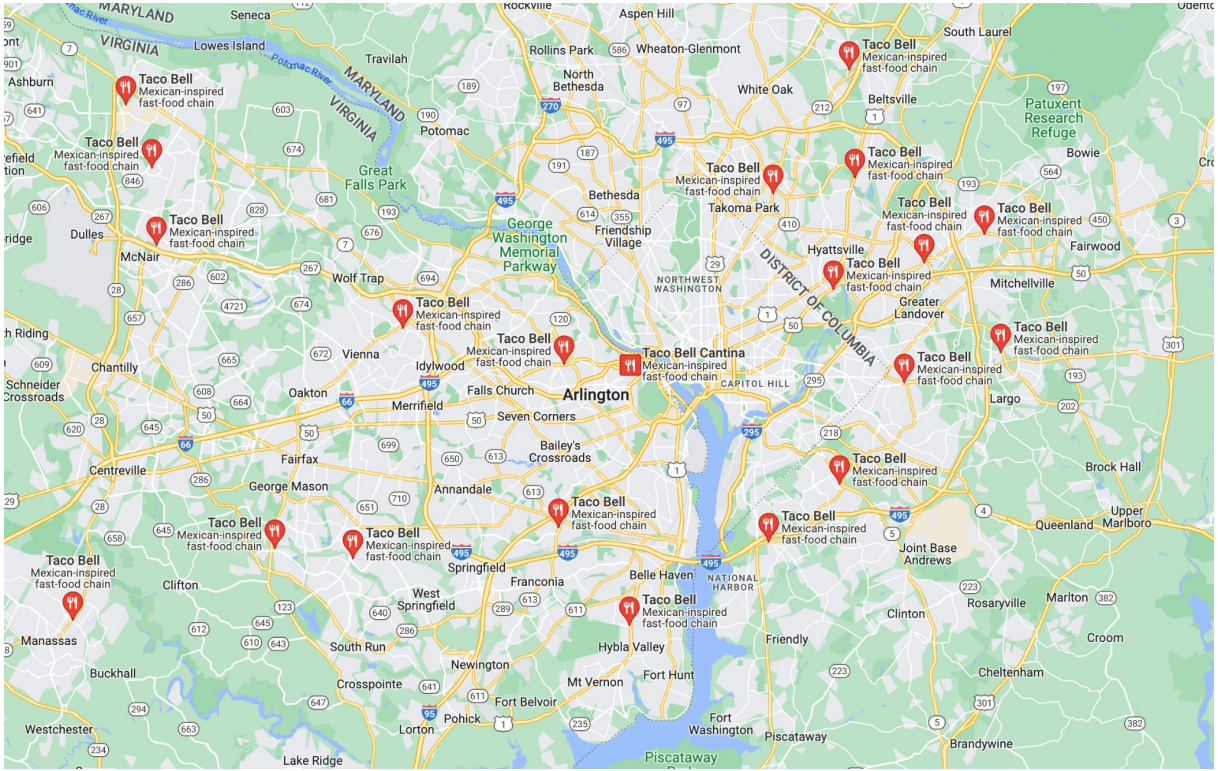
Why testing is hard
Reasons to test and the challenges that arise

Testing and CI/CD for decision models
What an opinionated testing experience looks like

Q&A time
Your time to shine: ask questions, give feedback!

First, story time

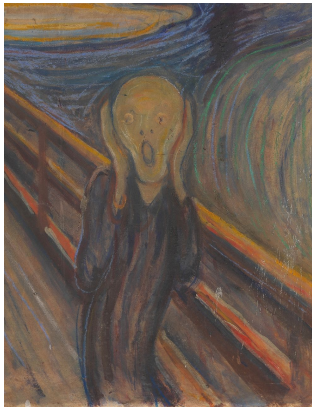
A tale of tacos and fried chicken



One Friday night in Metropolis...

FRIDAY

Dinner! 6PM



Testing is hard

What is decision model testing?

A set of techniques for **reducing risk and building confidence** to achieve stakeholder buy-in with decision models and drive improving solutions.

Why does decision model testing matter?

- Teams need a clear, repeatable path to production
- Stakeholders require confidence in outcomes
- Mistakes are expensive

Why is decision model testing hard?

- Technical setup and maintenance (SRE stakeholder)
- Complexity of analysis (OR stakeholder)
- Buy-in and alignment (Product stakeholder)
- Problem drift (Operational stakeholder)

Systems behave differently depending on environment and traffic patterns. Since the behavior of utilization can change at any time, sampling real traffic is the only way to reliably capture the request path. To guarantee both authenticity of the way in which the system is exercised and relevance to the current deployed system, **Chaos strongly prefers to experiment directly on production traffic**.

Source: [Principles of Chaos Engineering]

How we think about testing

Common types of decision model testing

Shadow testing

Run candidate model alongside production model

Switchback testing

Switch between treatment and control models over time

Batch experiments

Run exploratory experiments on one or more models

Scenario testing

Identify outcomes for different inputs, models, or decisions

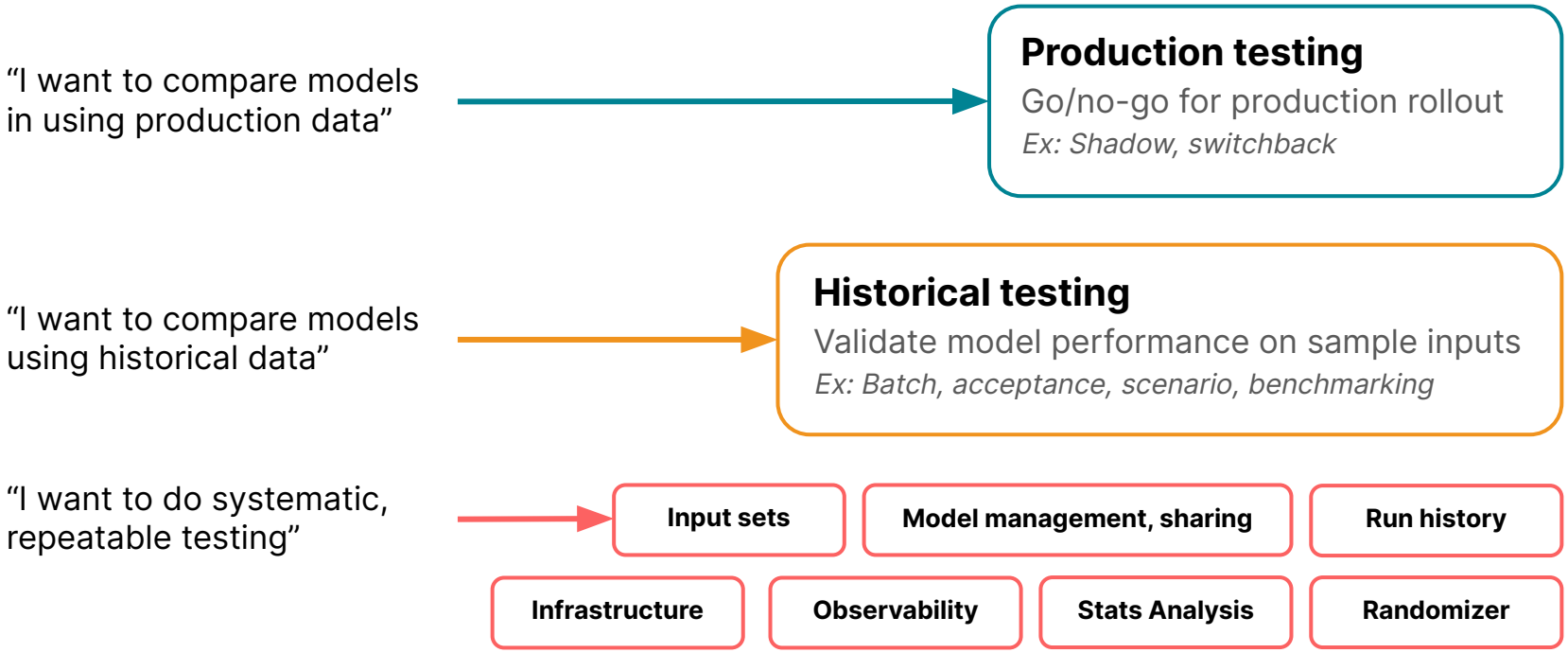
Acceptance testing

Determine if business KPIs are met by a new model

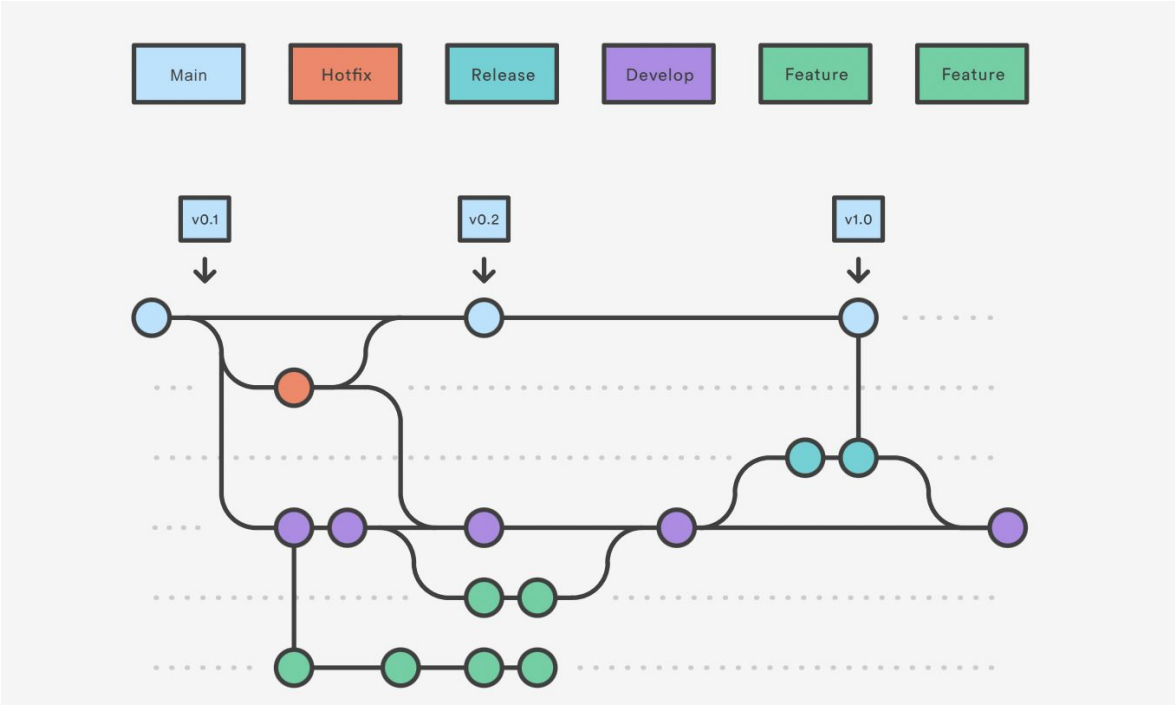
Benchmarking

Compare scale, speed, performance of models and solvers

Decision model testing framework

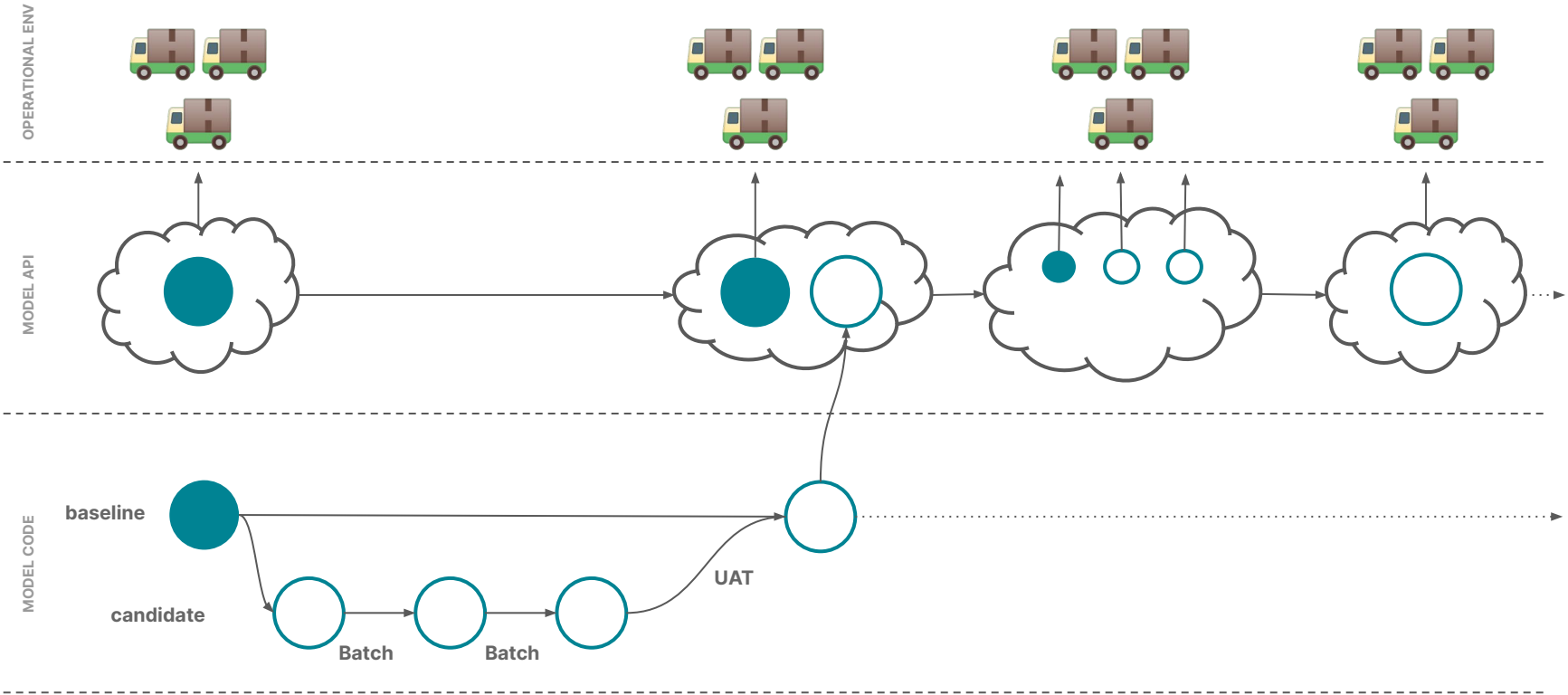


Inspiration: Git Flow

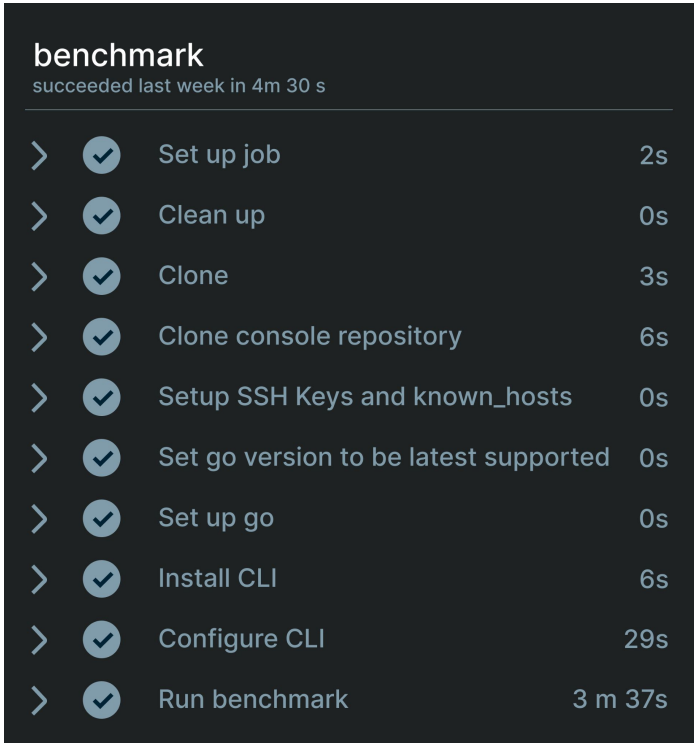


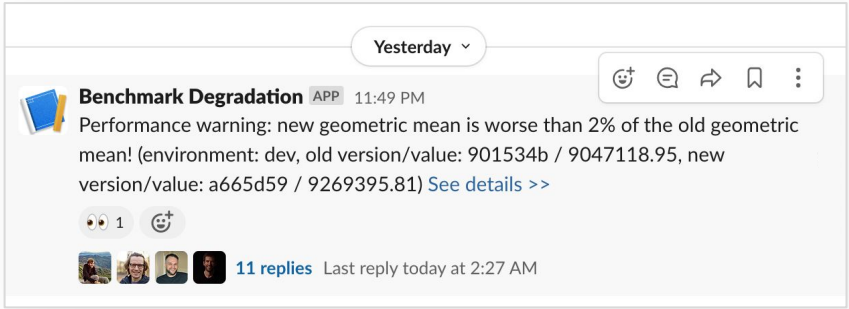
Source: Atlassian

DecisionOps model development workflow



DecisionOps test workflow in practice





Batch experiments

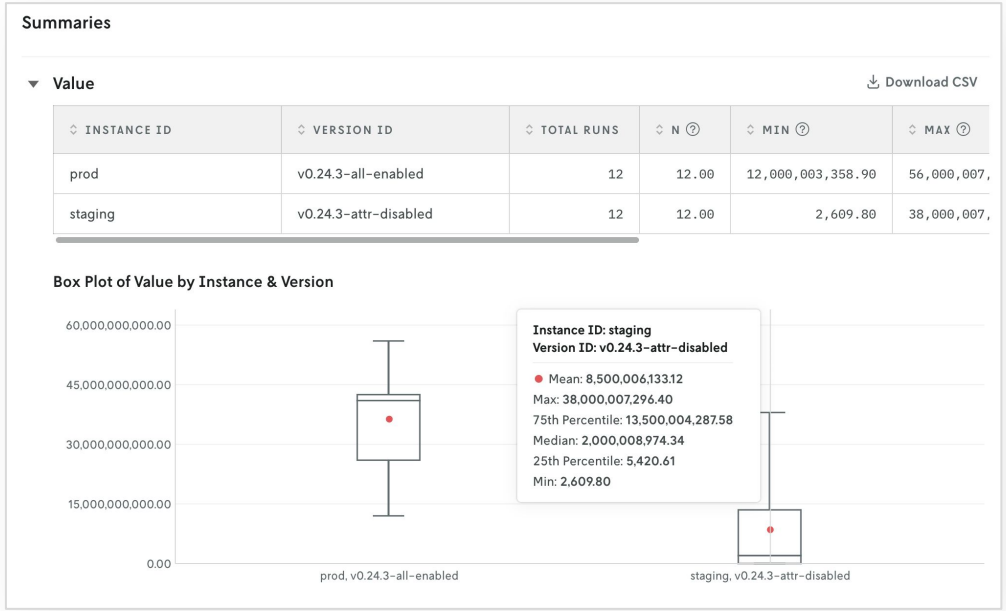
Early, ad hoc exploration of model change impact across output metrics

CHARACTERISTICS

- ✘ Operational decisions
- ✘ Production conditions
- ✘ Online data inputs
- ✘ Acceptance criteria
- ✔ Historical data inputs

CONFIDENCE

★★☆☆☆



Acceptance testing

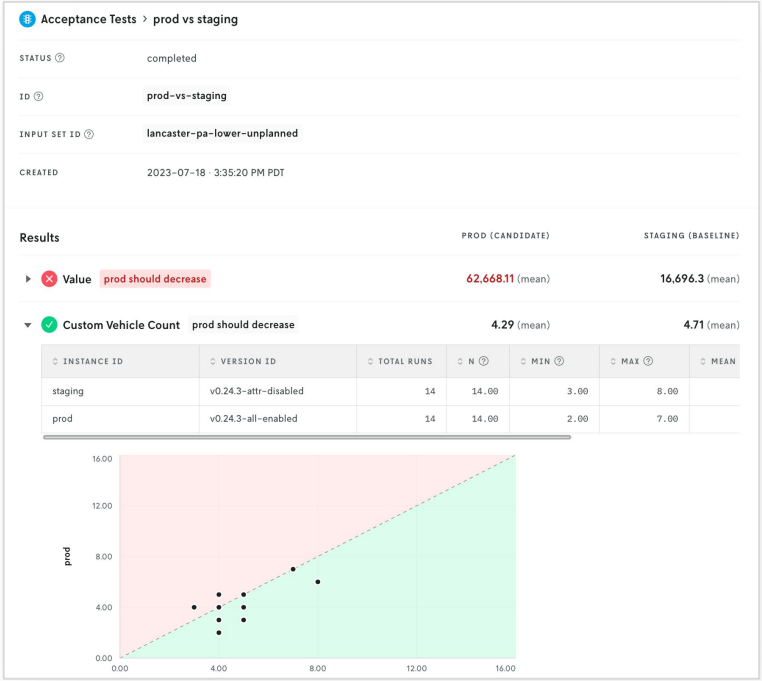
Determines if business requirements/KPIs are met by a new model

CHARACTERISTICS

- ✘ Operational decisions
- ✘ Production conditions
- ✘ Online data inputs
- ✔ Acceptance criteria
- ✔ Historical data inputs

CONFIDENCE

★★★☆☆



Scenario testing

Identify outcomes for a range of model inputs or configuration

CHARACTERISTICS

- ✘ Operational decisions
- ✘ Production conditions
- ✘ Online data inputs
- ✔ Acceptance criteria
- ✔ Historical data inputs

CONFIDENCE

```
...

"vehicles":
   {
      "capacity": "20"
...
```

```
...

"vehicles":
   {
      "capacity": "300"
...
```

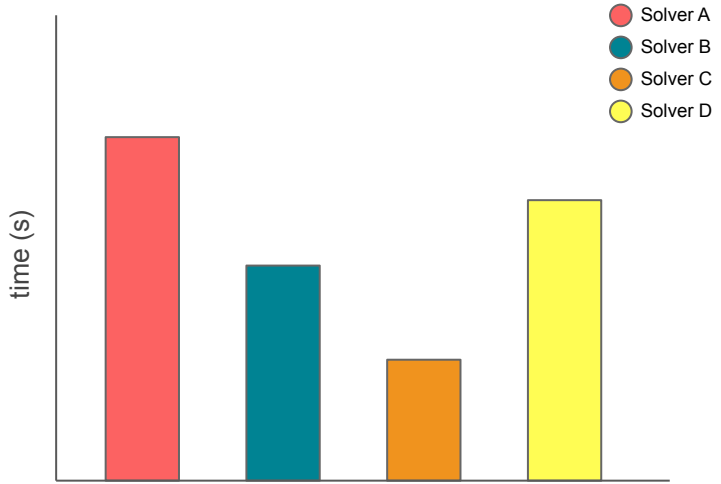
Benchmarking

Compare scale, speed, performance of models and solvers

CHARACTERISTICS

- ✘ Operational decisions
- ✘ Production conditions
- ✘ Online data inputs
- ✔ Acceptance criteria
- ✔ Historical data inputs

CONFIDENCE



Shadow testing

Run candidate model alongside production model, without production impact

CHARACTERISTICS

✘ Operational decisions

✔ Production conditions

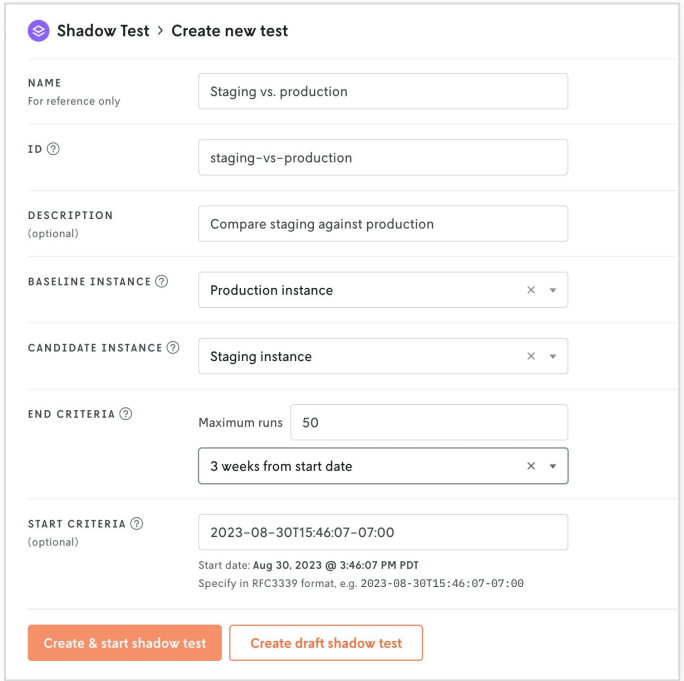
✔ Online data inputs

✔ Acceptance criteria

✔ Historical data inputs

CONFIDENCE

★★★★☆



Switchback testing

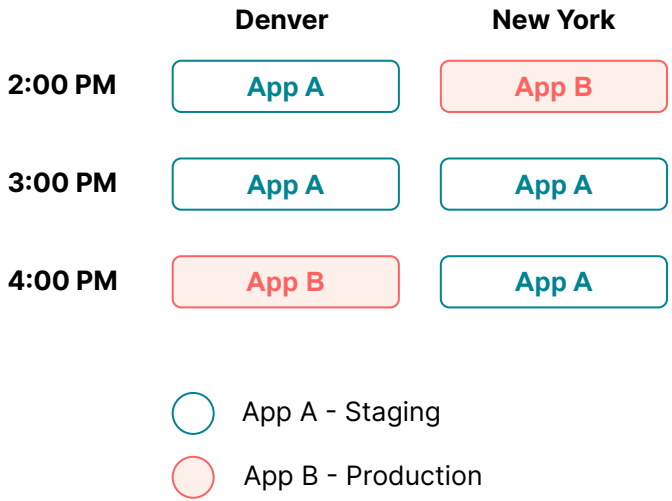
Test treatment and control models over time/regions

CHARACTERISTICS

- ✔ Operational decisions
- ✔ Production conditions
- ✔ Online data inputs
- ✔ Acceptance criteria
- ✔ Historical data inputs

CONFIDENCE

★★★★★



🌮 Summary

- We shared some stories from the field
- We talked about why testing decision models is hard
- We presented a framework for how testing can fit into your decision model development workflow

QUESTIONS?


nextmv

⚙️ Components to test on

APPS

~ REPOS

VERSIONS

~ RELEASE SNAPSHOTS

INSTANCES

~ BRANCHES