Feature Store Observability

What It Is & Why It Matters

Aman Khan, Group Product Manager, Arize AI

FEATURE STORE SUMMIT 2022

Organized by HOPSWORKS

arize

Feature Store Monitoring with ML Observability

A little bit about me...

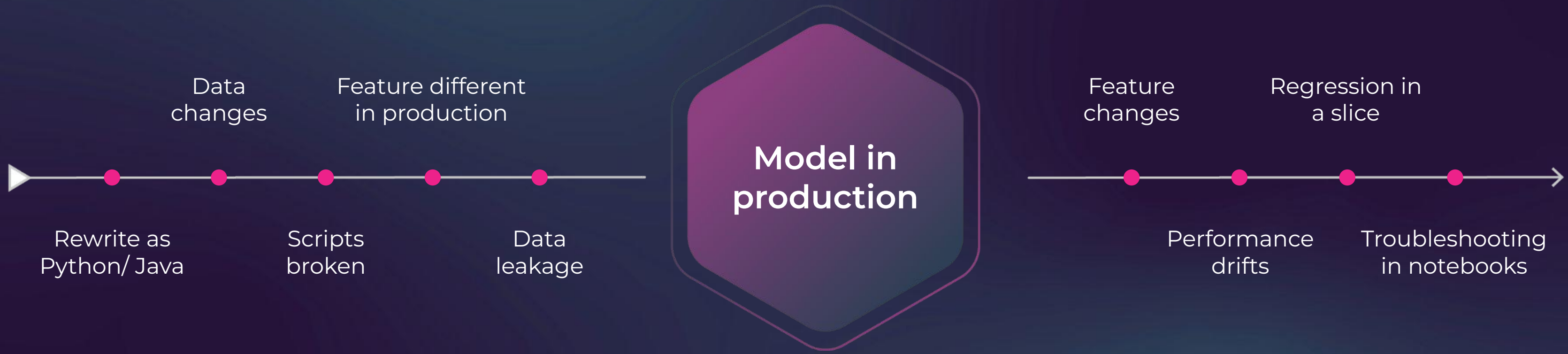
Aman Khan | Group Product Manager @ Arize AI | aman@arize.com

Agenda

1. Challenges with Productionalizing Machine Learning (ML) models

2. Approaching Feature Store Observability:
   a. Data Quality
   b. Drift Analysis
   c. Performance & Explainability

The road to production is challenging



But the path after production is equally difficult

ML Observability

Software that helps teams automatically monitor AI, understand how to fix it when it's broken, and improve data & models.

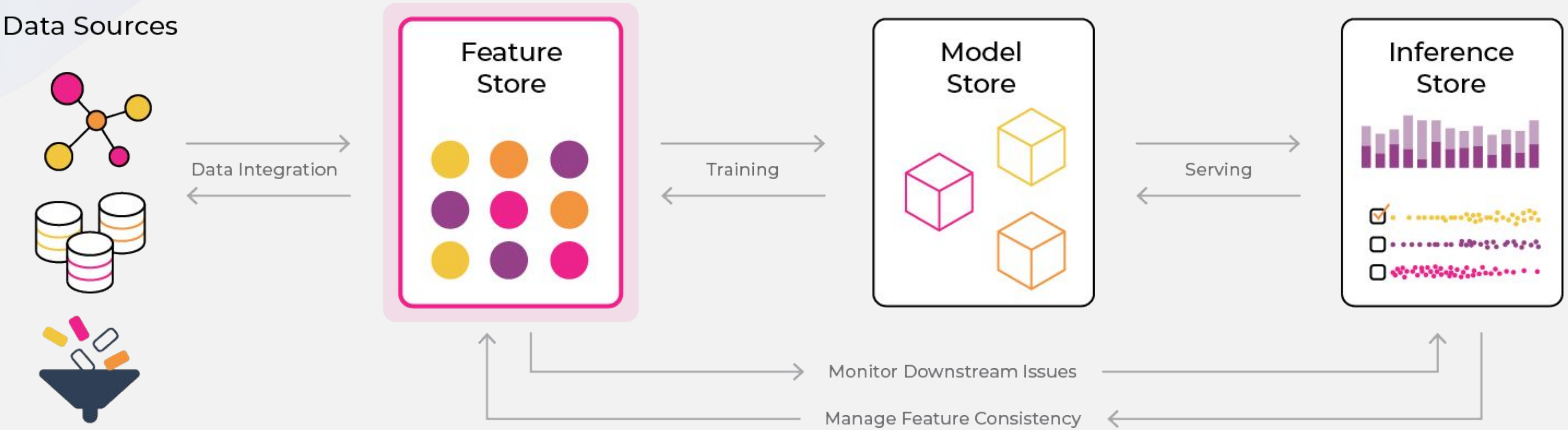
Observability vs Monitoring

Monitoring alerts you with issues. However, it doesn’t troubleshoot and root cause your problematic areas.

Observability goes deeper. It helps you get to the root cause and resolve the underlying issues.

Feature Store Observability

An ***inference store*** is a central place to monitor and identify issues within your feature store



Why does this matter

84%

Takes >1 week to detect & fix a model issue

51%

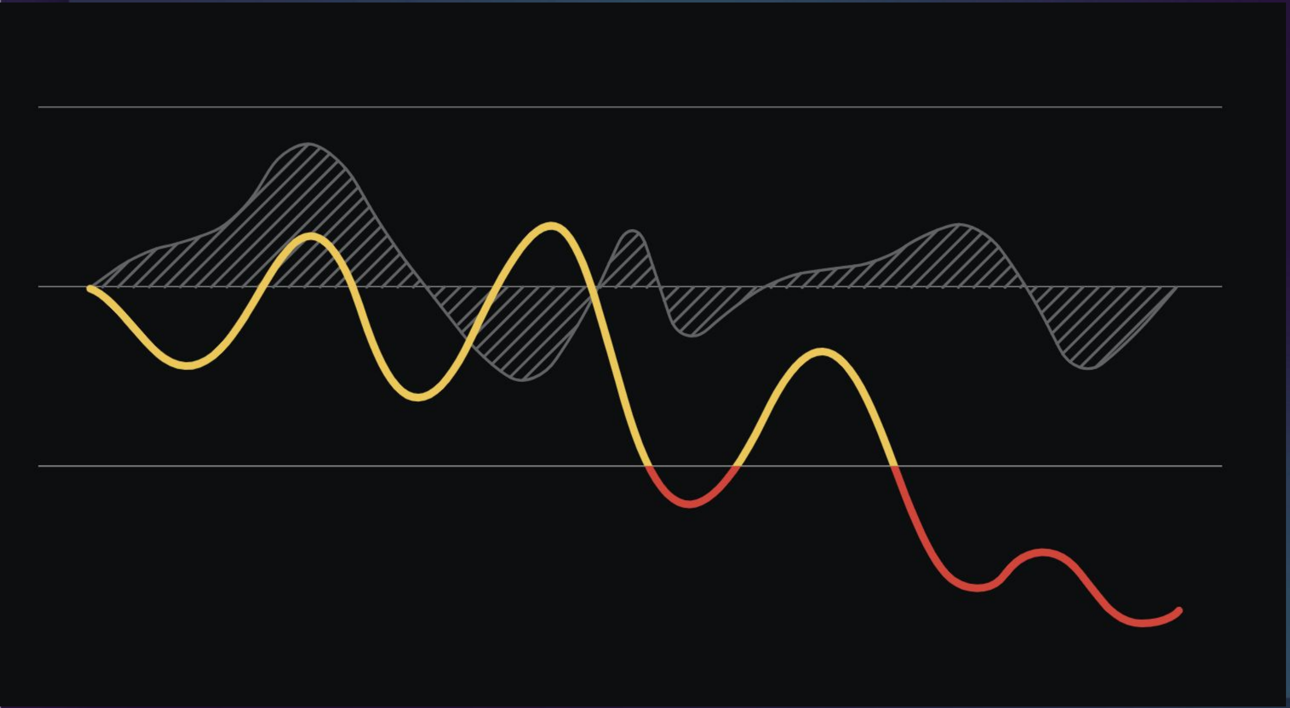
Want deeper capabilities to monitor & RCA drift

54%

Say business execs can't quantify AI ROI

- > 1000 companies surveyed

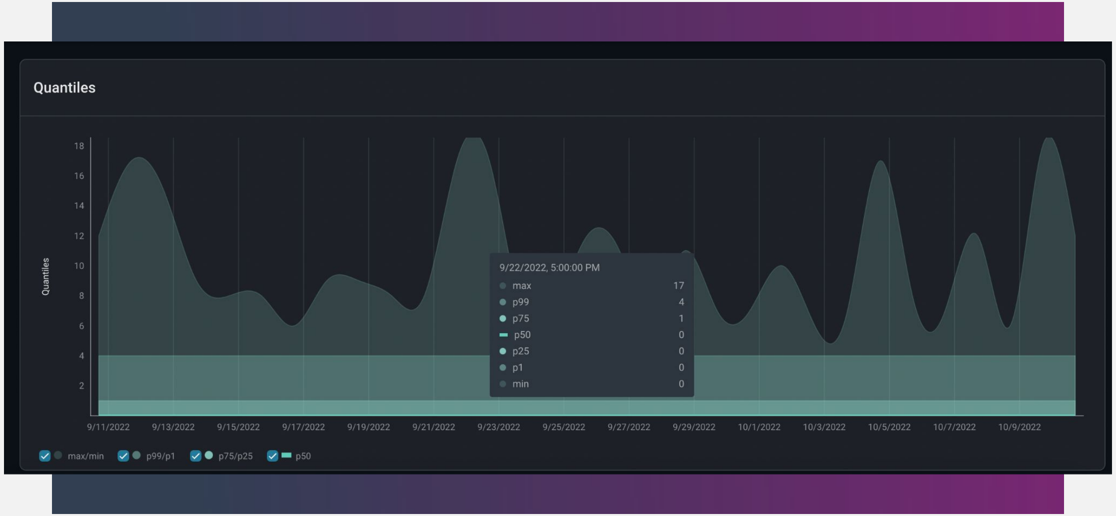
Approaching Feature Store Monitoring: Data Quality



Data Quality Checks in Production

Common issues:

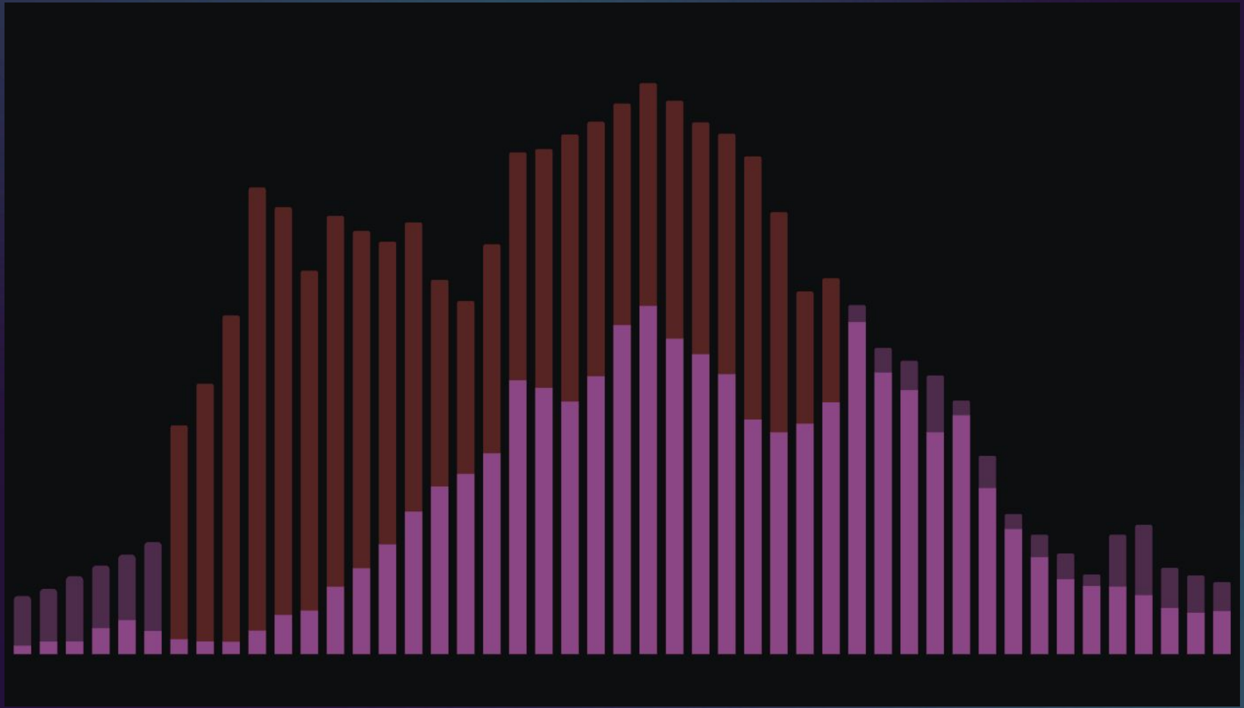
- Missing Data
- Invalid Data
- Noisy Data
- Duplicated Data
- Out of Range Violations
- Cardinality Changes
- Type Mismatch
- Unexpected Traffic



Ways to catch these in production:

- Profile your data. Profile the frequency and distribution of values in your dataset.
- Data Quality Monitoring

Approaching Feature Store Monitoring: Feature Drift



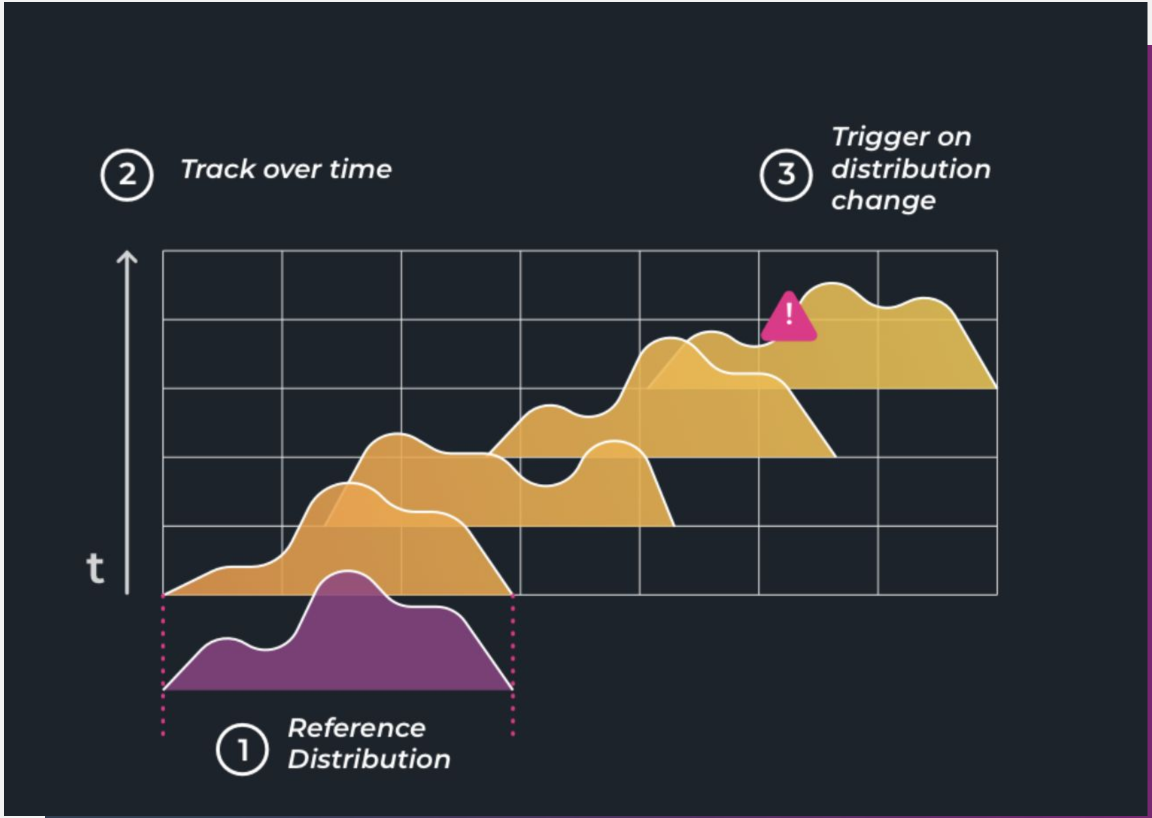
What is Drift and why is it important?

Drift is the change in a distribution over time.

Data is not static
Accounting for drift is important to make sure your models stay relevant.

Feature drift measures the changes in your feature distribution.

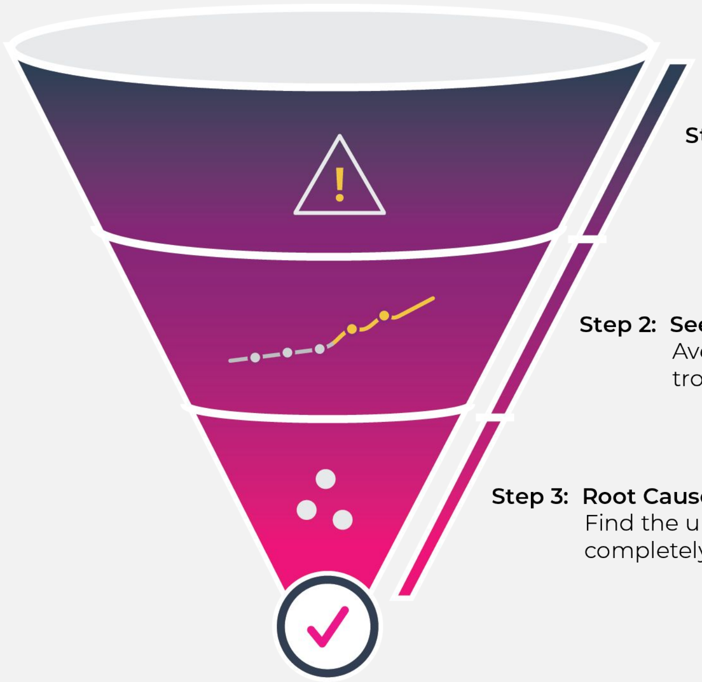
Feature drift is very common and can happen at anytime, yet remain undetected.



How do we measure Drift?

Population Stability Index (PSI)	**Kullback-Leibler (KL) divergence**	**Wasserstein / Earthmovers distance**	**Jensen-Shannon Distance**
Kolmogorov-Smirnov test (or KS test)	Chi-squared test	Manhattan Distance	Euclidean Distance

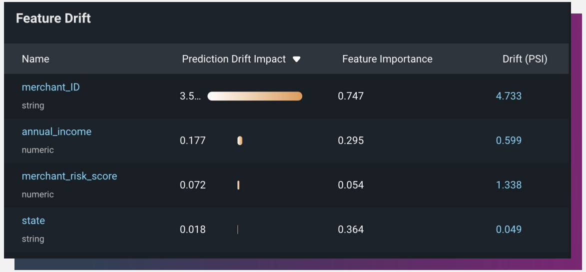
Approaching Drift Resolution



Step 1: Compare To Something You Know
Select a dataset from training, validation or production.

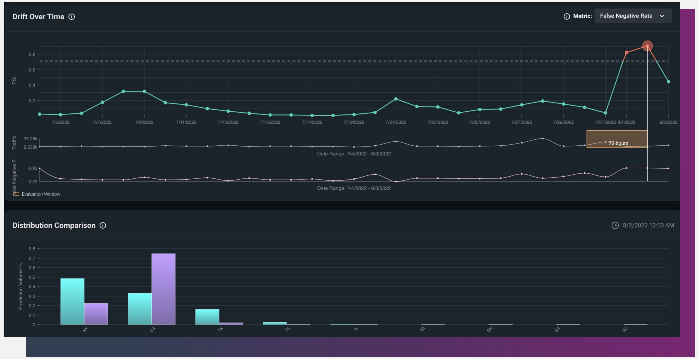
Step 2: See Performance Breakdown By Slices
Averages can lie; look to slices to troubleshoot.

Step 3: Root Cause & Resolve
Find the underlying cause and fix it completely and quickly.

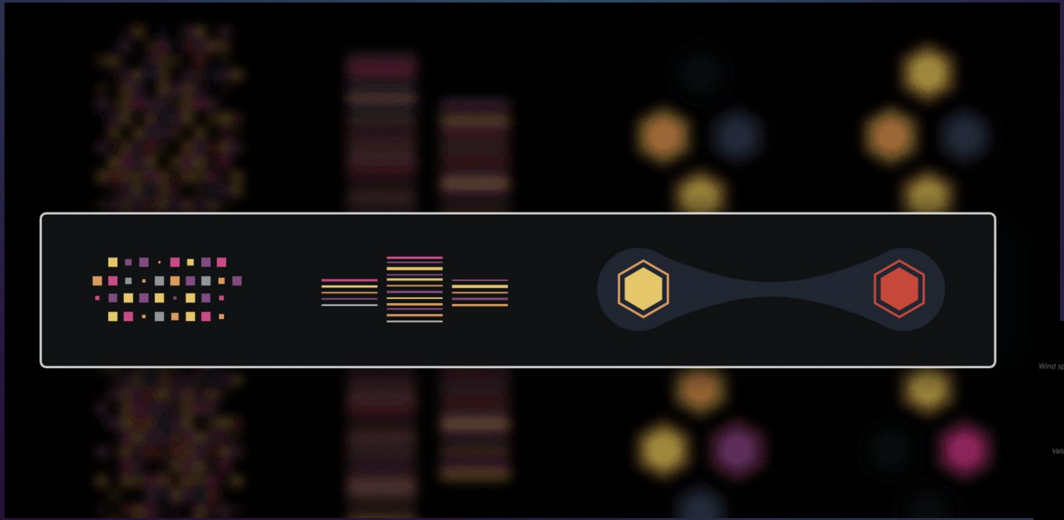


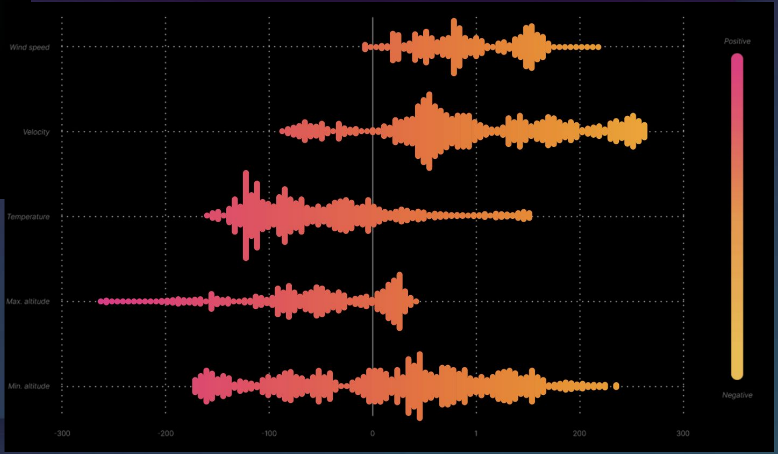
Feature Drift

Name	Prediction Drift Impact	Feature Importance	Drift (PSI)
merchant_ID string	3.5…	0.747	4.733
annual_income numeric	0.177	0.295	0.599
merchant_risk_score numeric	0.072	0.054	1.338
state string	0.018	0.364	0.049



Explainability, Performance Tracing, & Fairness



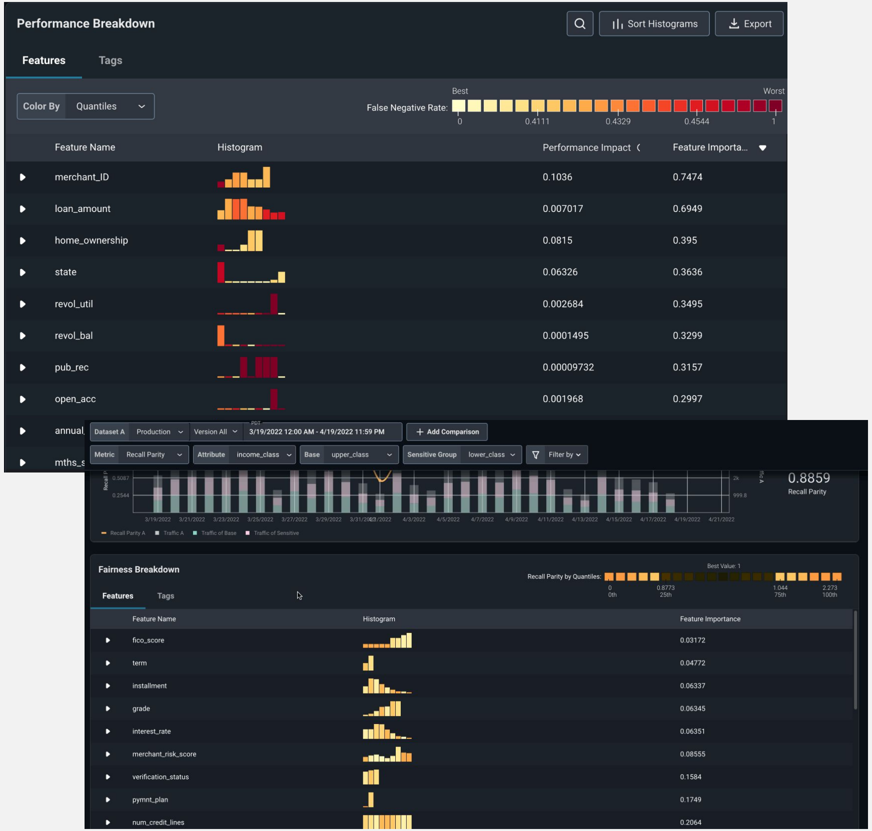


Using Explainability To Understand Feature Impact

- Explainability is a technique designed to determine which model **feature or combination of features led to a specific model decision.**

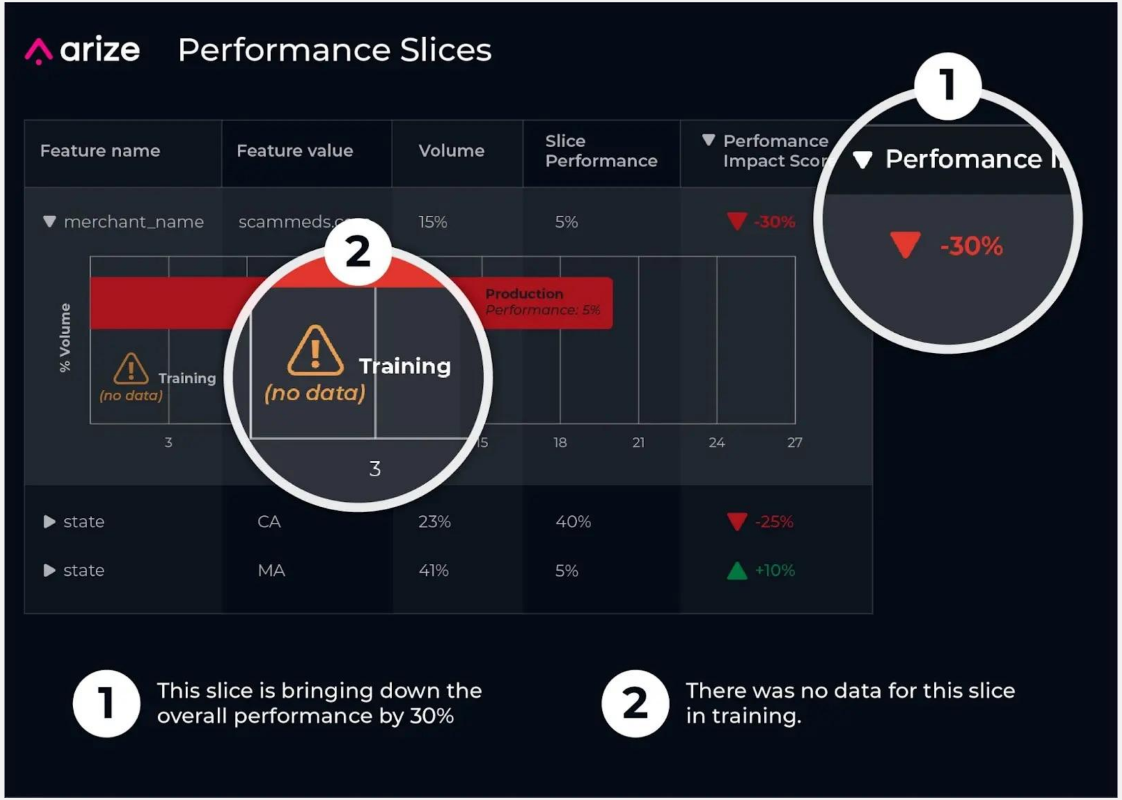
- Explainability helps us understand drift, performance degradation, and fairness in terms the **most important model features**.
    - **Not all features are equal.**

- Does the ranking of feature importance change for a specific cohort?



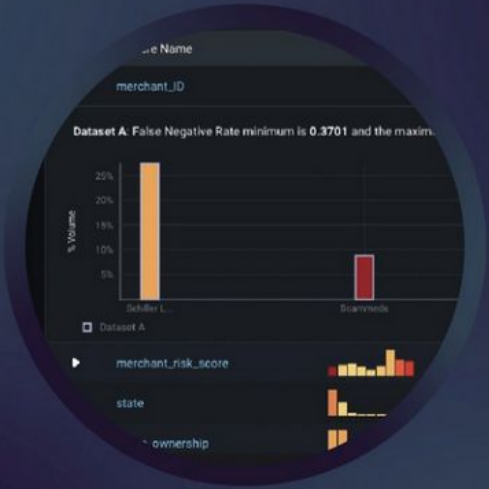
How Features Affect Model Performance

- A dataset slice identifies a subset of your data within each feature that may behave qualitatively differently than the rest
- A single poor performing slice can drastically impact model performance



arize

ML Performance Tracing



1. Compare to something you know
Select datasets from training, validation or production.

2. See performance breakdown by slice
Averages can lie; look to slices to troubleshoot.

3. Root cause & resolve
Find the underlying cause and fix it completely and quickly.

ML Observability for Unstructured data, alongside structured data

Monitor every embedding

Easily identify data has changed

Identify pattern of change

Proactively identify data quality issues

Export problem sample

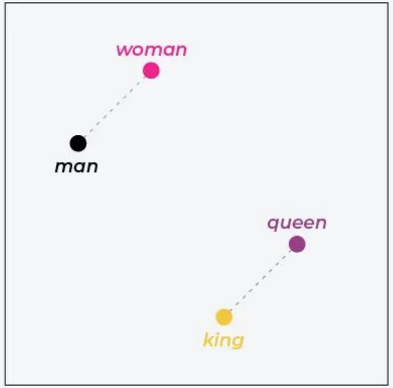
Quickly fix upstream data issue

What are **embeddings?**

- They provide a common mathematical representation of your data
- They compress your data
- They preserve relationships within your data

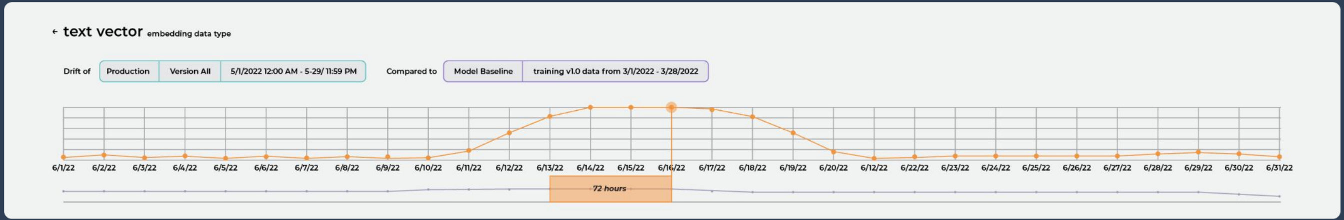
word	living being	feline	human	gender	royalty	verb	plural
man	0.6	-0.2	0.8	0.9	-0.1	-0.9	-0.7
woman	0.7	0.3	0.8	-0.7	0.1	-0.5	-0.4
king	0.5	-0.4	0.7	0.8	0.9	-0.7	-0.6
queen	0.8	-0.1	0.8	-0.9	0.8	-0.5	-0.9

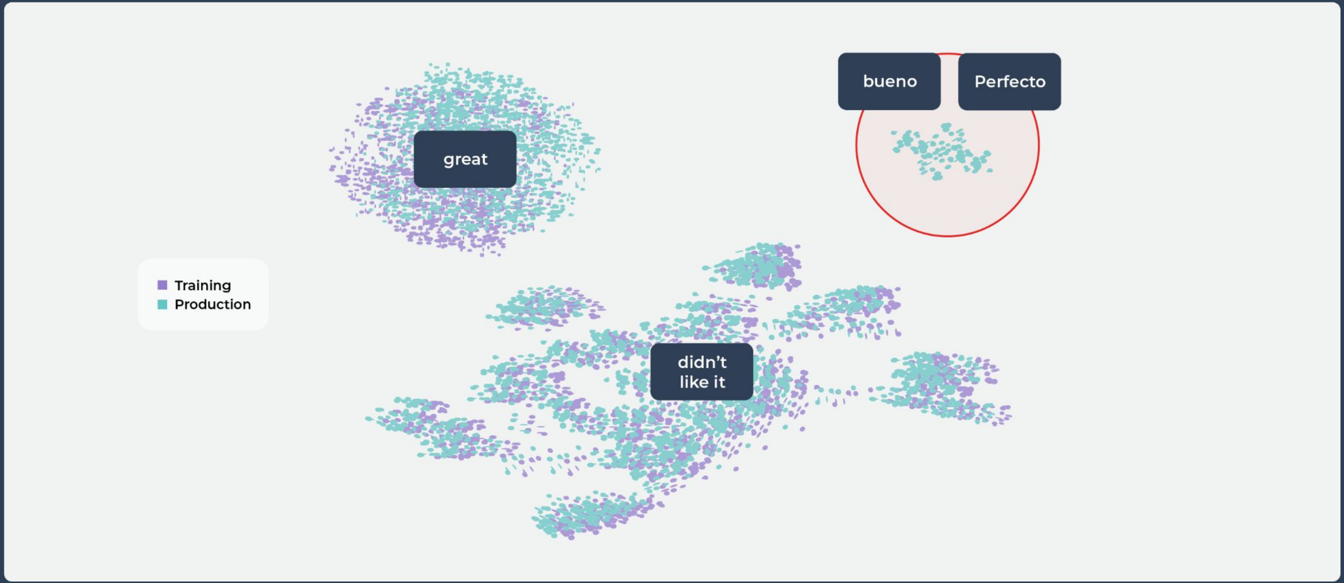
Word embedding



Visualization of word embedding

Interactive visualizations to identify pattern of change





In Summary

1. Focus on your customers and data science use cases

2. Feature store monitoring is critical to downstream model performance and ensuring upstream quality

3. Looking across data quality, drift and performance metrics enables root cause analysis

Thank you!

Arize AI Community
arize-ai.slack.com

Join our community Slack for Arize updates, events and platform support!

Sign up for a free account

https://bit.ly/3zx1ZCQ