



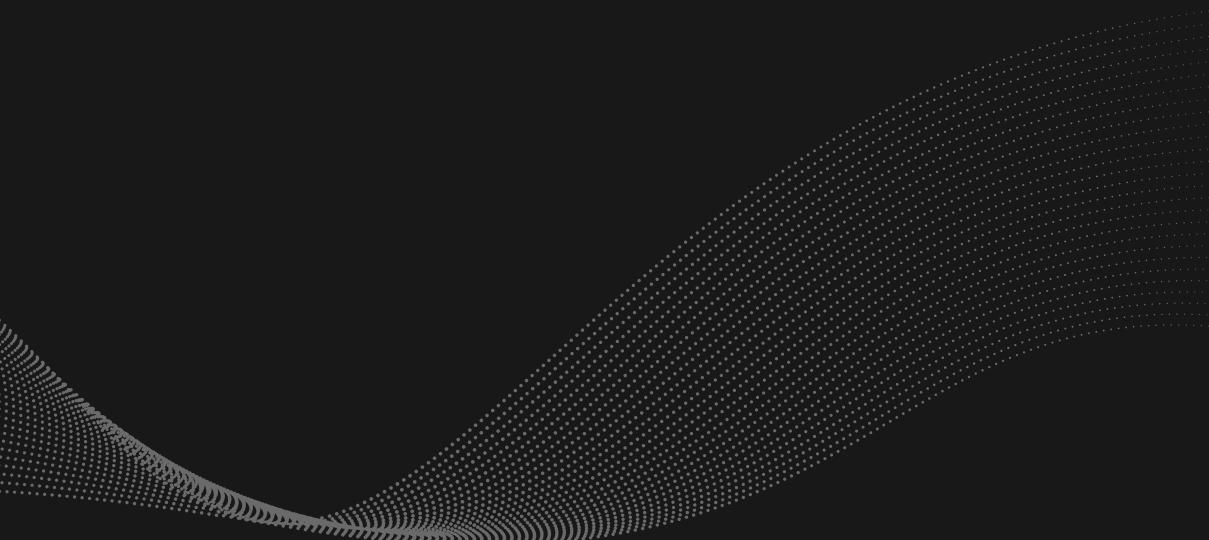
Lab 07: Monitoring

The MLOps Course

02 October 2025

Table of Contents

- 01 Introduction
- 02 Data Drift
- 03 Tools for monitoring
- 04 Conclusion



Why Monitor Your ML Models?

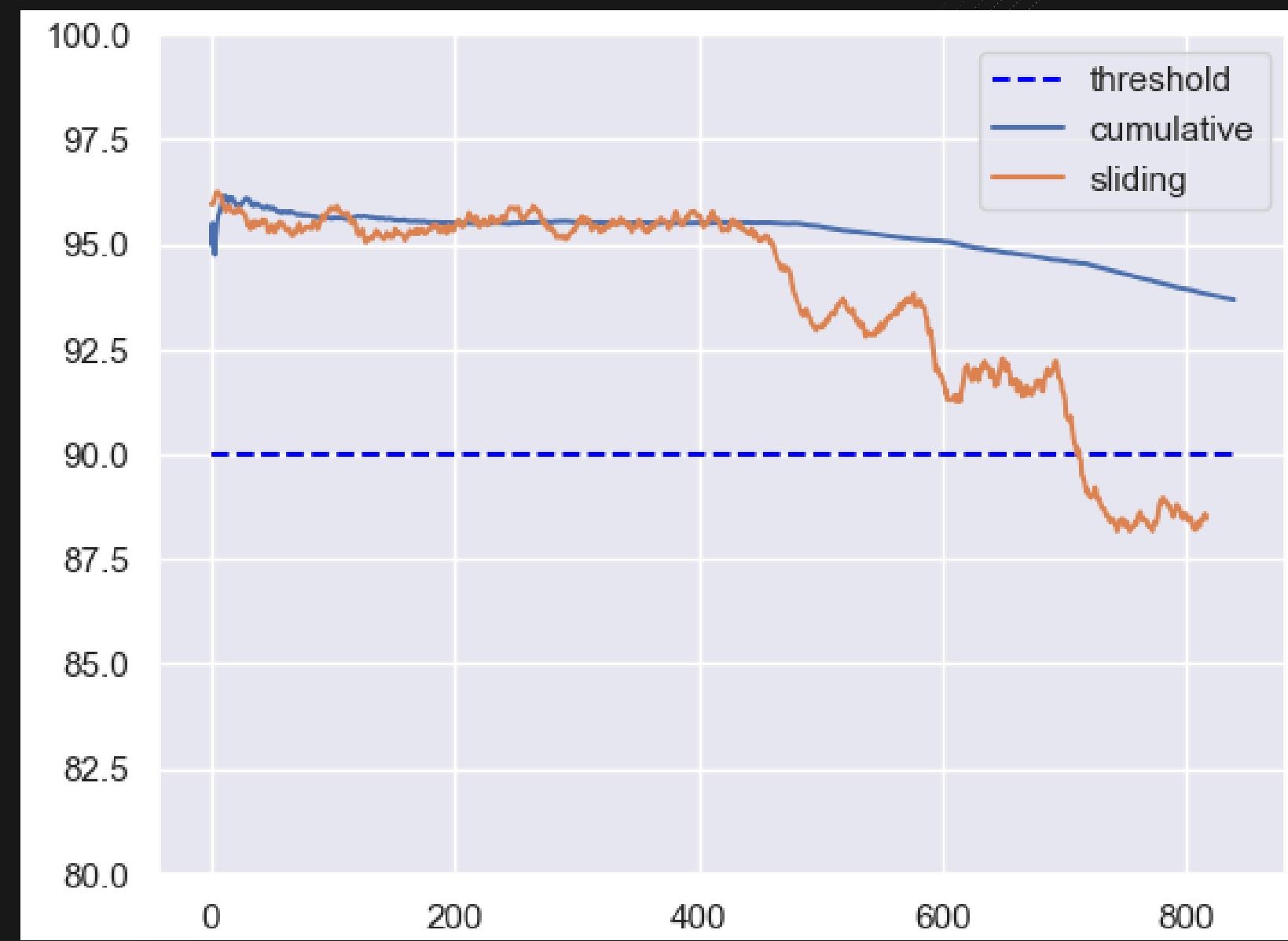
- system health
- performance
 - Models are trained on historical data, but the world changes.

==> real work begins once we deploy to production

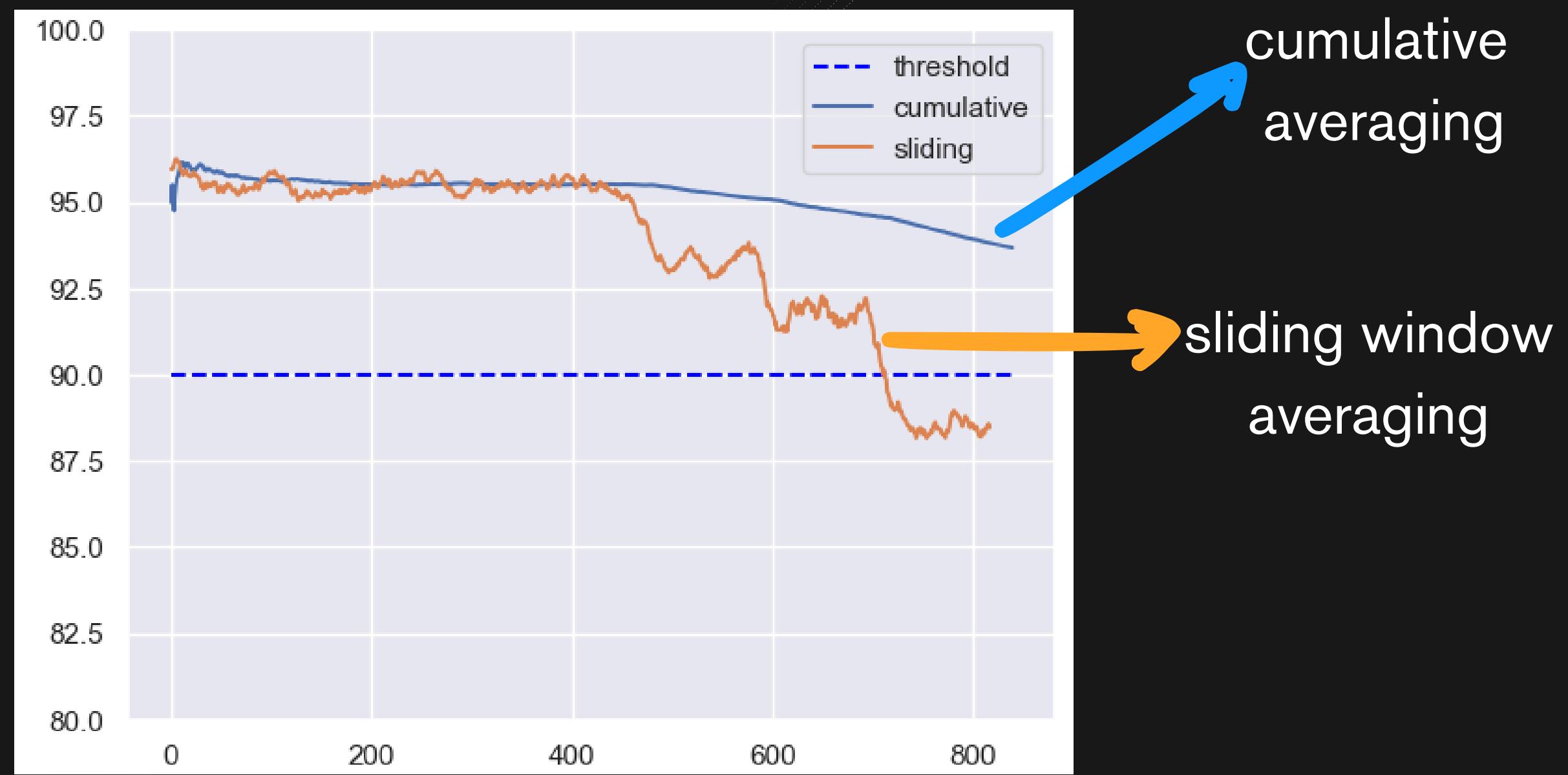
What is Data Drift

- production data \neq original/reference data

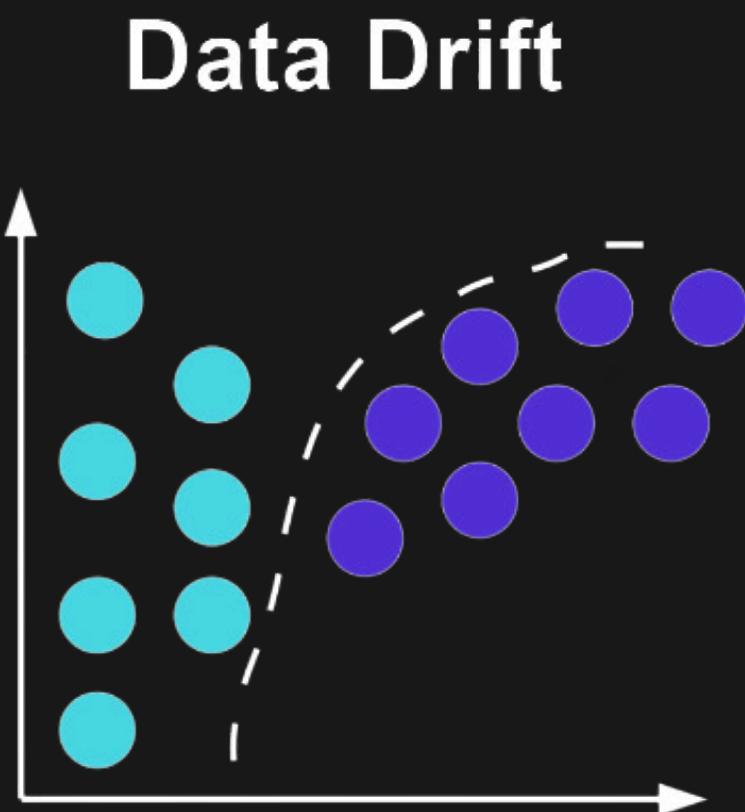
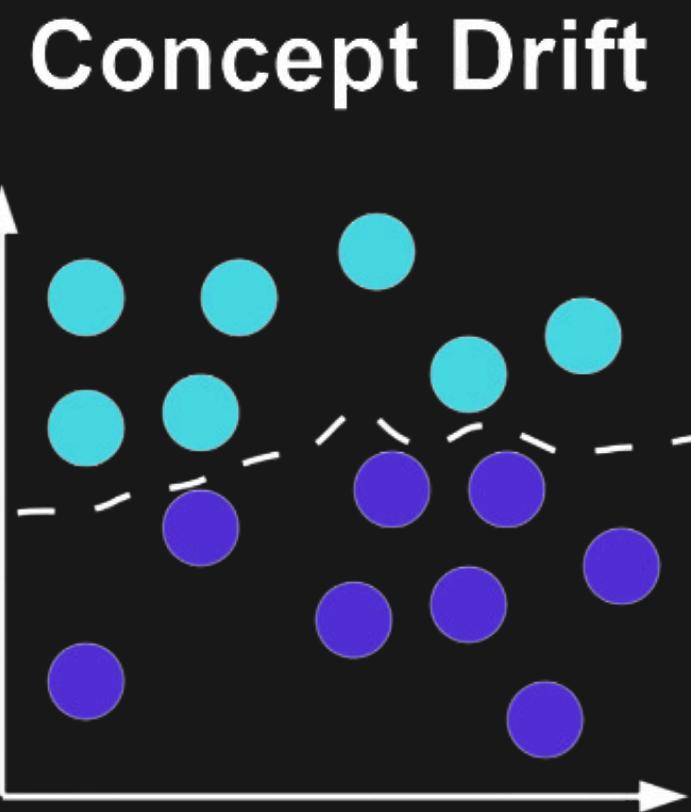
What is Data Drift



What is Data Drift



Types of drift



Alibi-detect

- drift detection:
 - initialize a drift detector with reference data
 - apply drift detector to current/prod data

Alibi-detect

► Univariate:

- **Chi-Squared** (for categorical data)
- **Kolmogorov-Smirnov**
- **Cramér-von Mises**
- **Fisher's Exact Test** (for binary data)

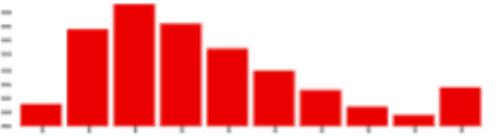
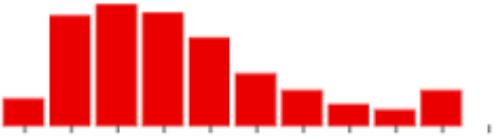
► Multivariate:

- **Least-Squares Density Difference (LSDD)**
- **Maximum Mean Discrepancy (MMD)**

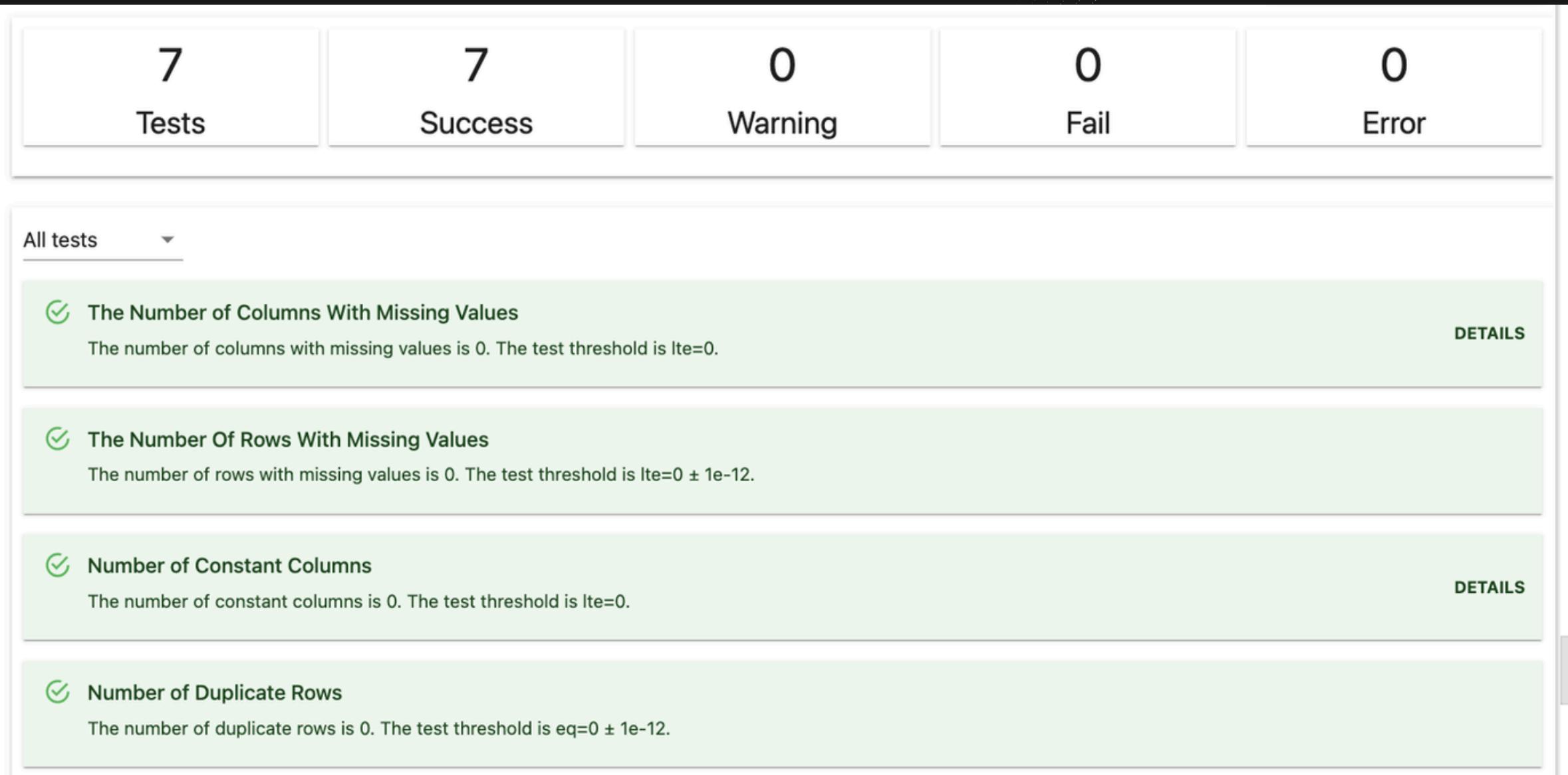
Evidently

- track data drift and quality
- dashboards
- reports
- tests
- very interactive

Evidently Reports

Feature	Type	Reference Distribution	Current Distribution	Data Drift	Stat Test	Drift Score
> target	num			Not Detected	Wasserstein distance (normed)	0.031996
> prediction	num			Not Detected	Wasserstein distance (normed)	0.013115
> AveBedrms	num			Not Detected	Wasserstein distance (normed)	0.055536

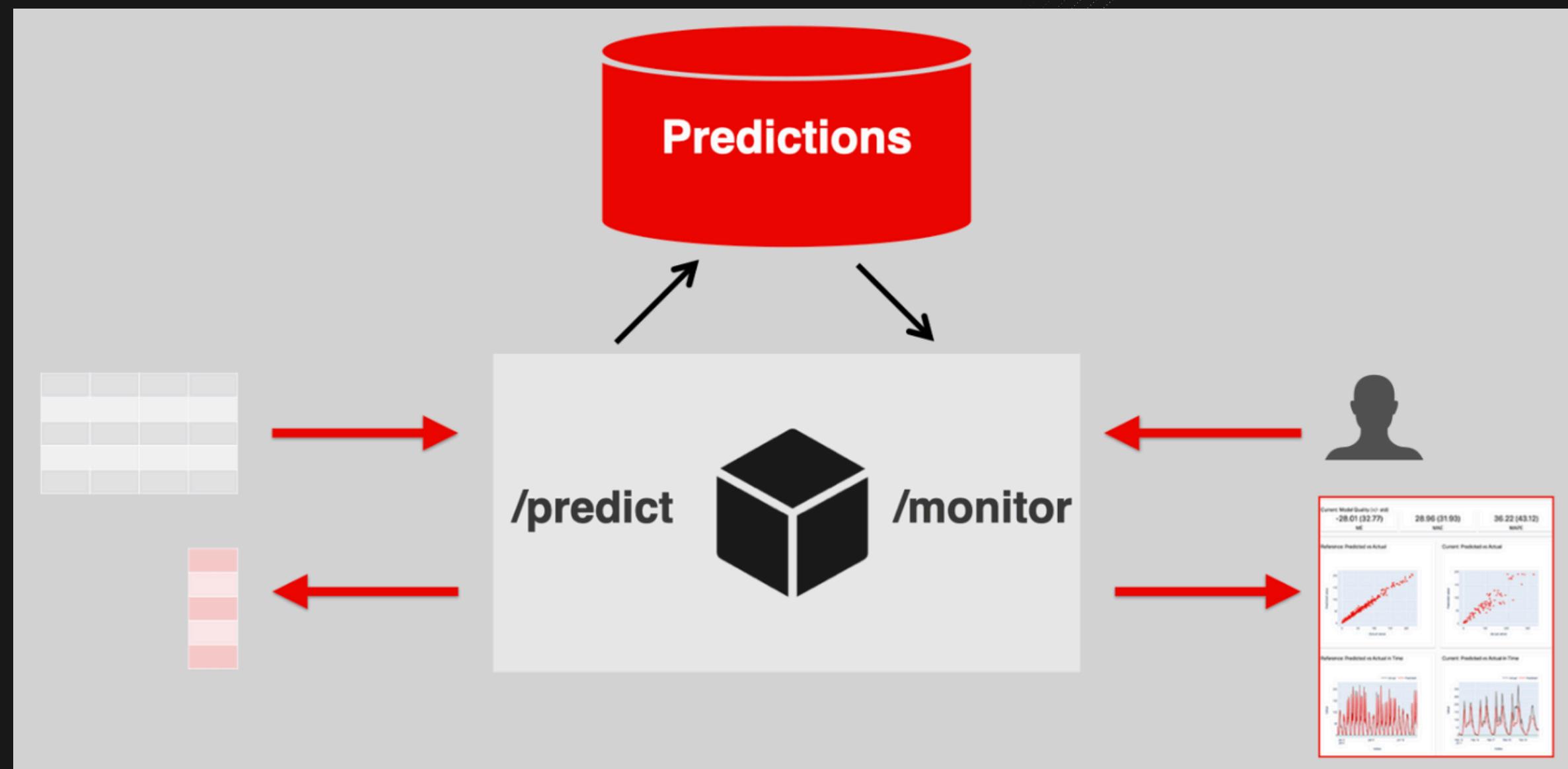
Evidently Tests



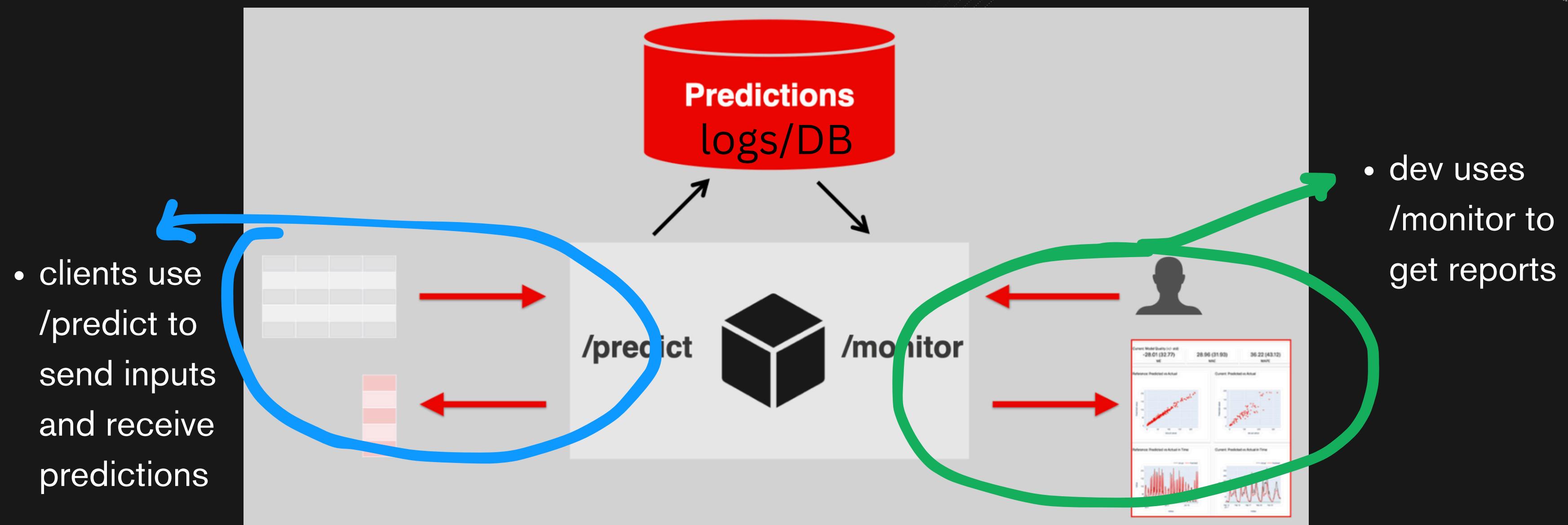
The screenshot shows the Evidently Tests interface. At the top, there are five summary boxes: 'Tests' (7), 'Success' (7), 'Warning' (0), 'Fail' (0), and 'Error' (0). Below this, a dropdown menu is set to 'All tests'. The main area lists four test results, each with a green checkmark icon:

- The Number of Columns With Missing Values**: The number of columns with missing values is 0. The test threshold is $\text{lte}=0$. [DETAILS](#)
- The Number Of Rows With Missing Values**: The number of rows with missing values is 0. The test threshold is $\text{lte}=0 \pm 1e-12$.
- Number of Constant Columns**: The number of constant columns is 0. The test threshold is $\text{lte}=0$. [DETAILS](#)
- Number of Duplicate Rows**: The number of duplicate rows is 0. The test threshold is $\text{eq}=0 \pm 1e-12$.

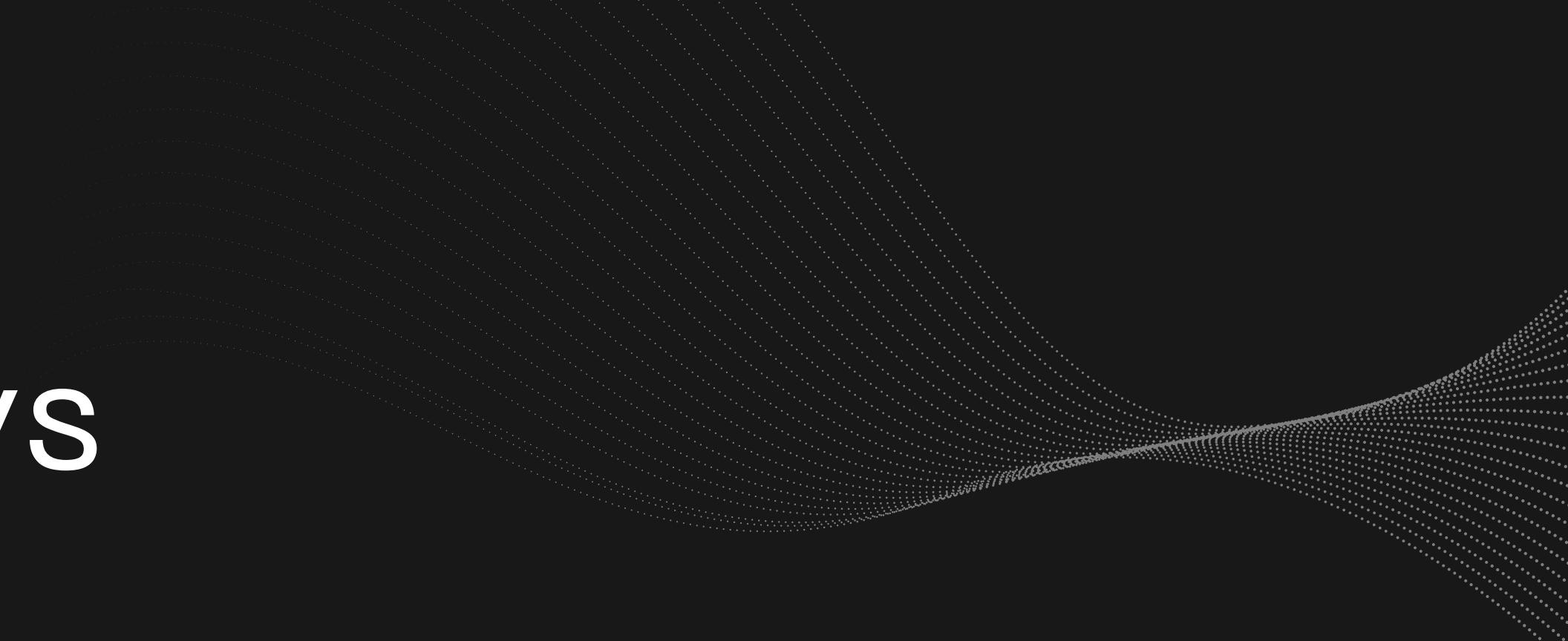
How its all connected?



How its all connected?



Key Takeaways



Data drift is inevitable,
monitoring is essential.