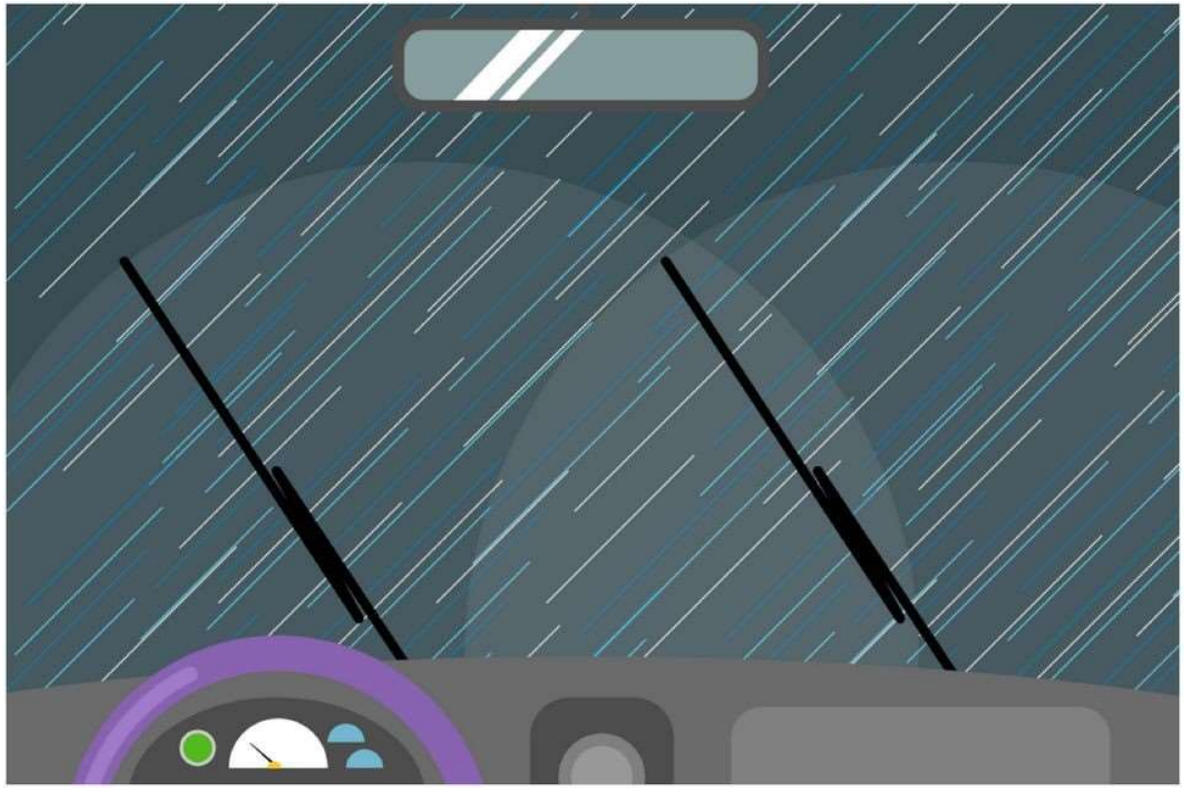
DATADOG

Automatically Detecting Anomalies and Outliers in Real-Time

Outline

- Monitoring
- Alerting
- Outlier vs. Anomaly Detection
- Outlier Detection Algorithms
- Anomaly Detection Algorithms

Monitor Everything



Monitor Everything

Datadog gathers performance data from all your application components.


AWS


Docker


CoreOS


Chef


Puppet


Github


Pagerduty


Nagios


Go


Postgres


Java


VMware


Redis


MySQL

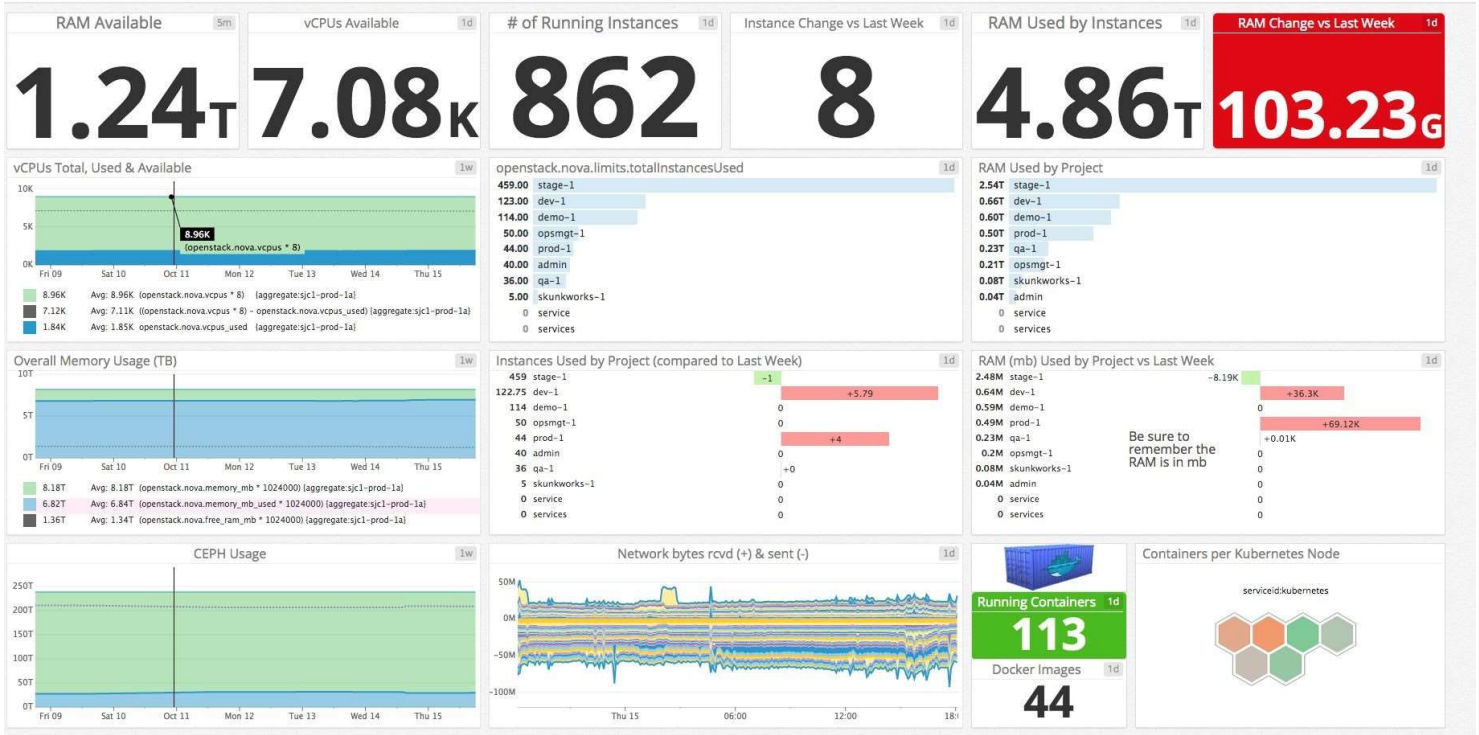

Apache


Tomcat

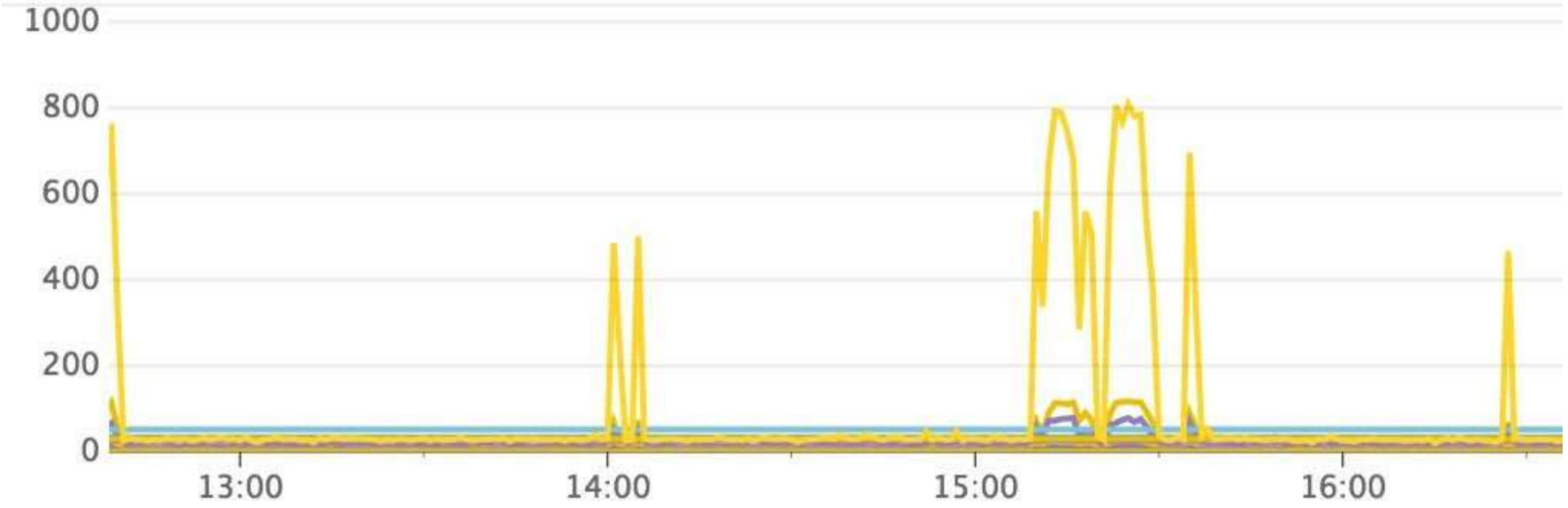

MongoDB

New Relic

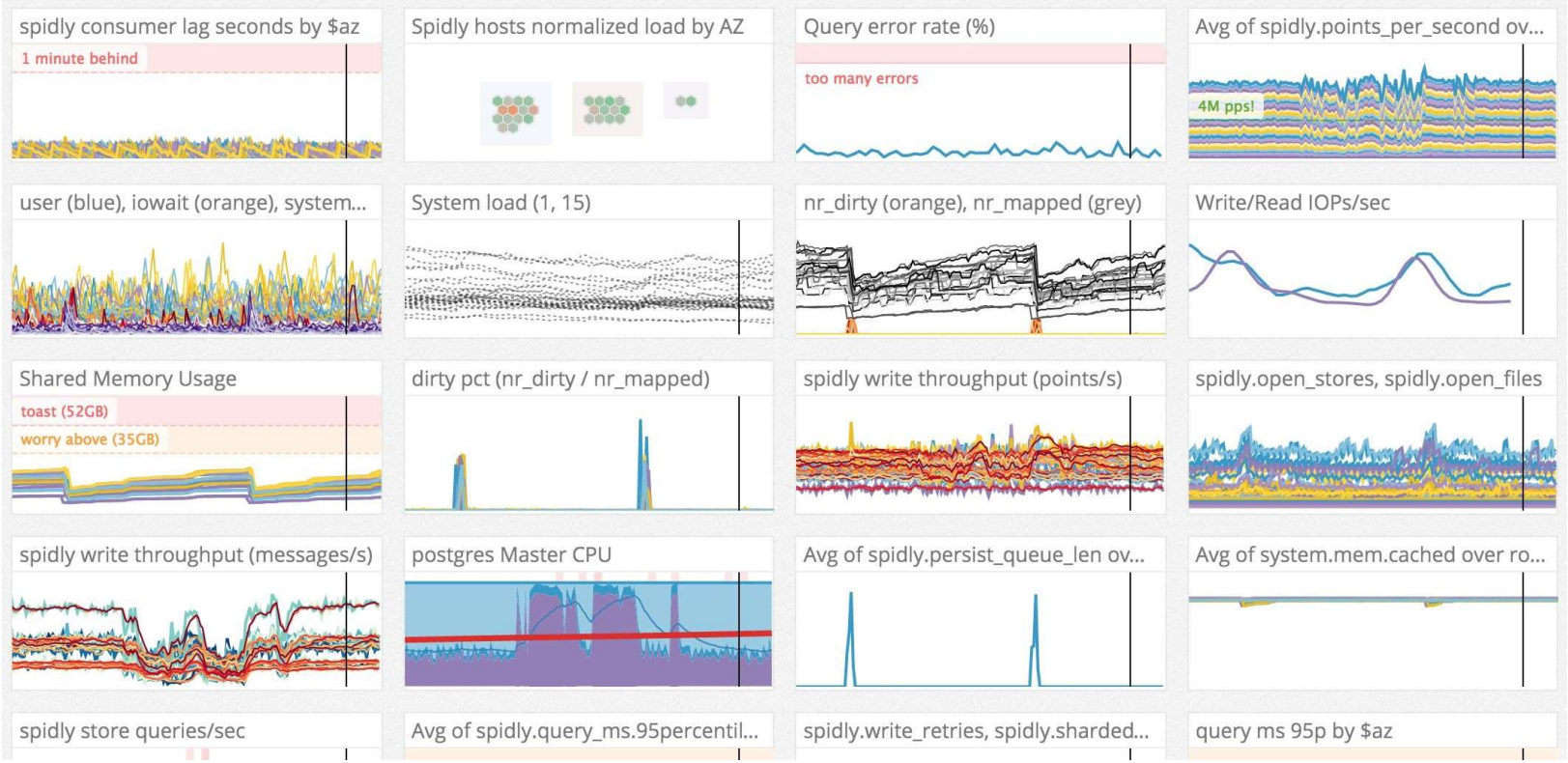
Monitor Everything



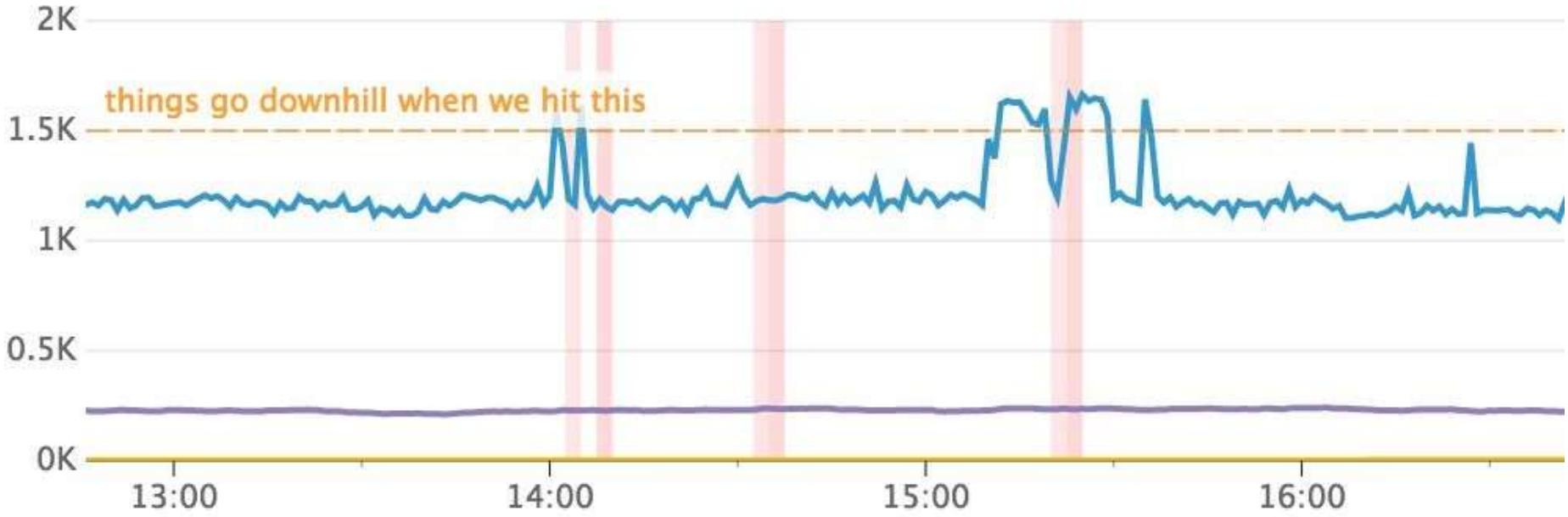
Monitor Everything



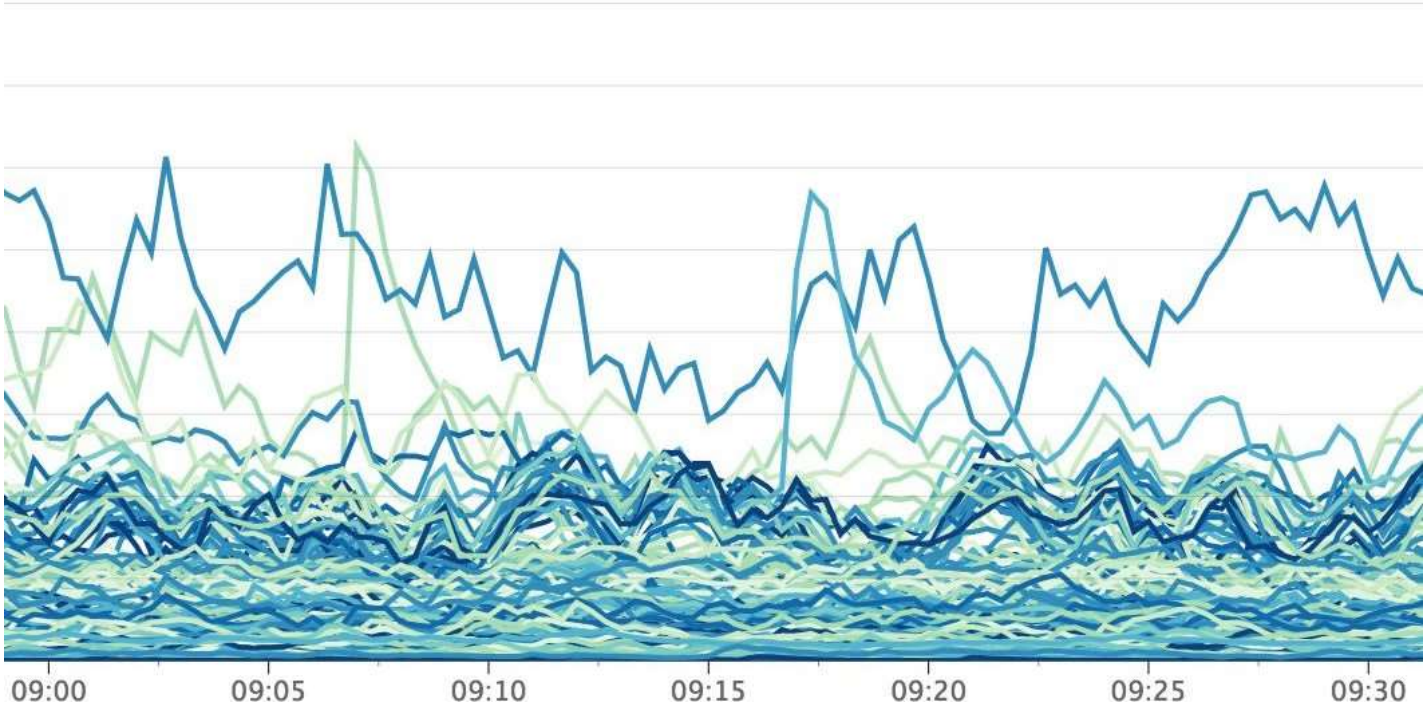
Monitor Everything?



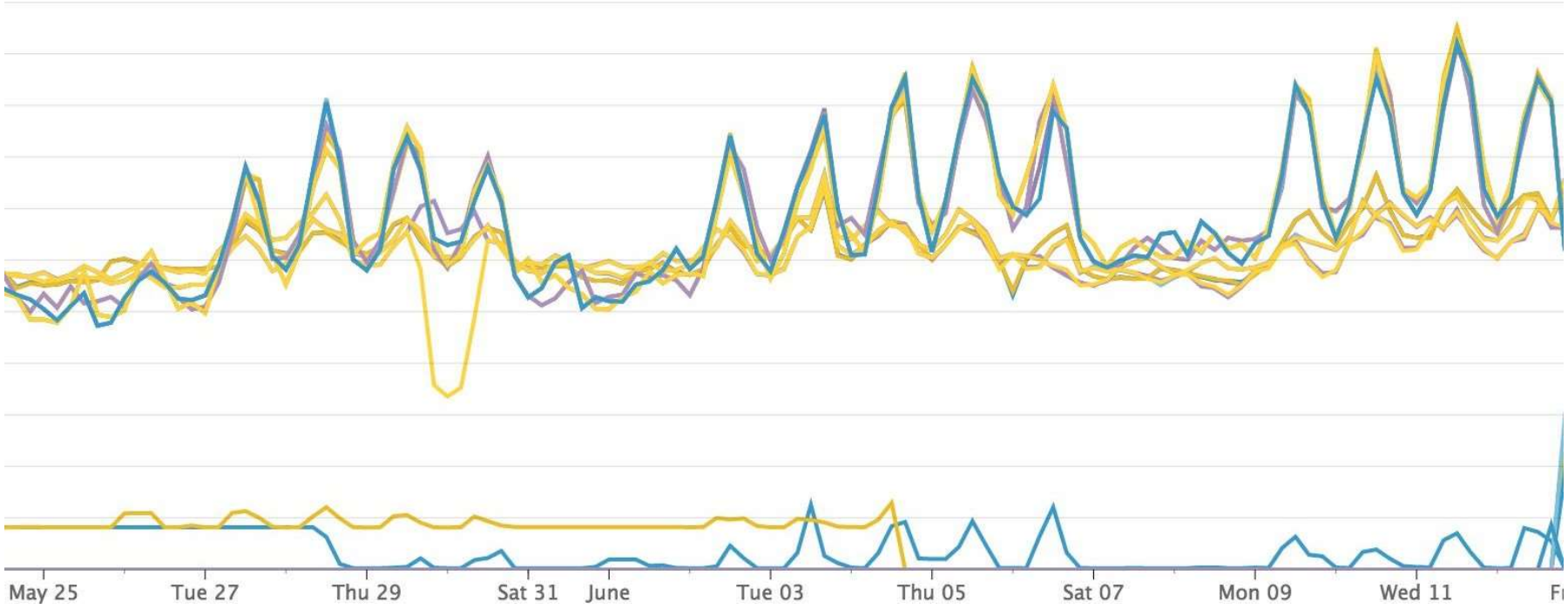
Alerting


2K
things go downhill when we hit this
1.5K
1K
0.5K
0K
13:00
14:00
15:00
16:00

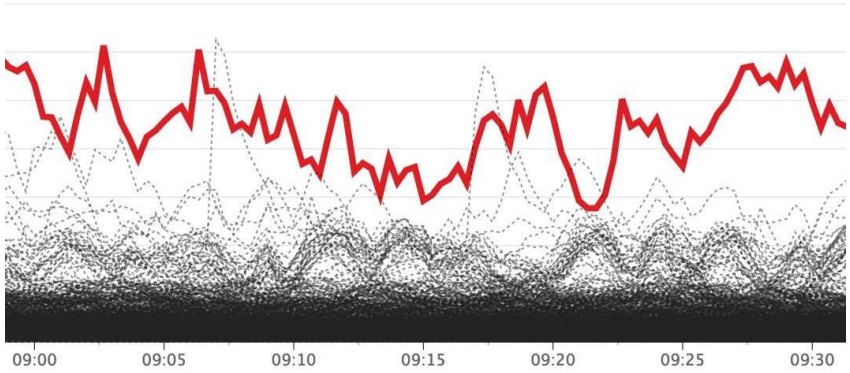
Alerting?

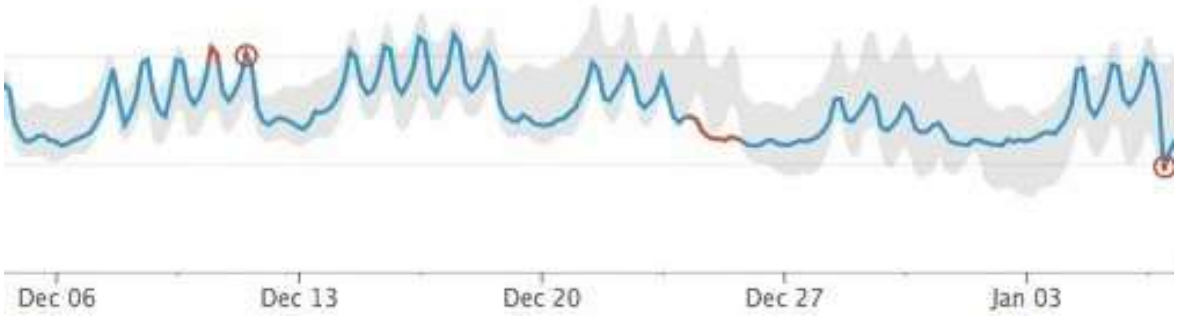


Alerting?



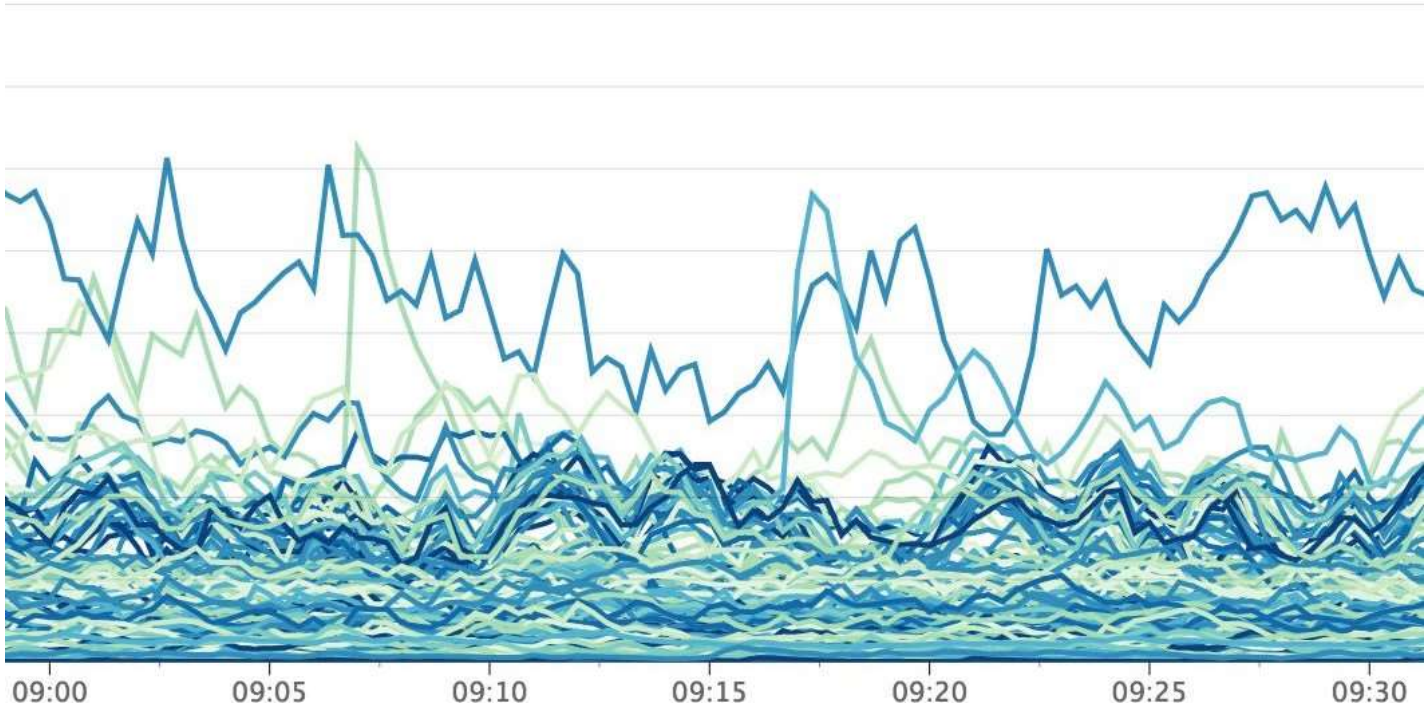
Outlier and Anomaly Detection



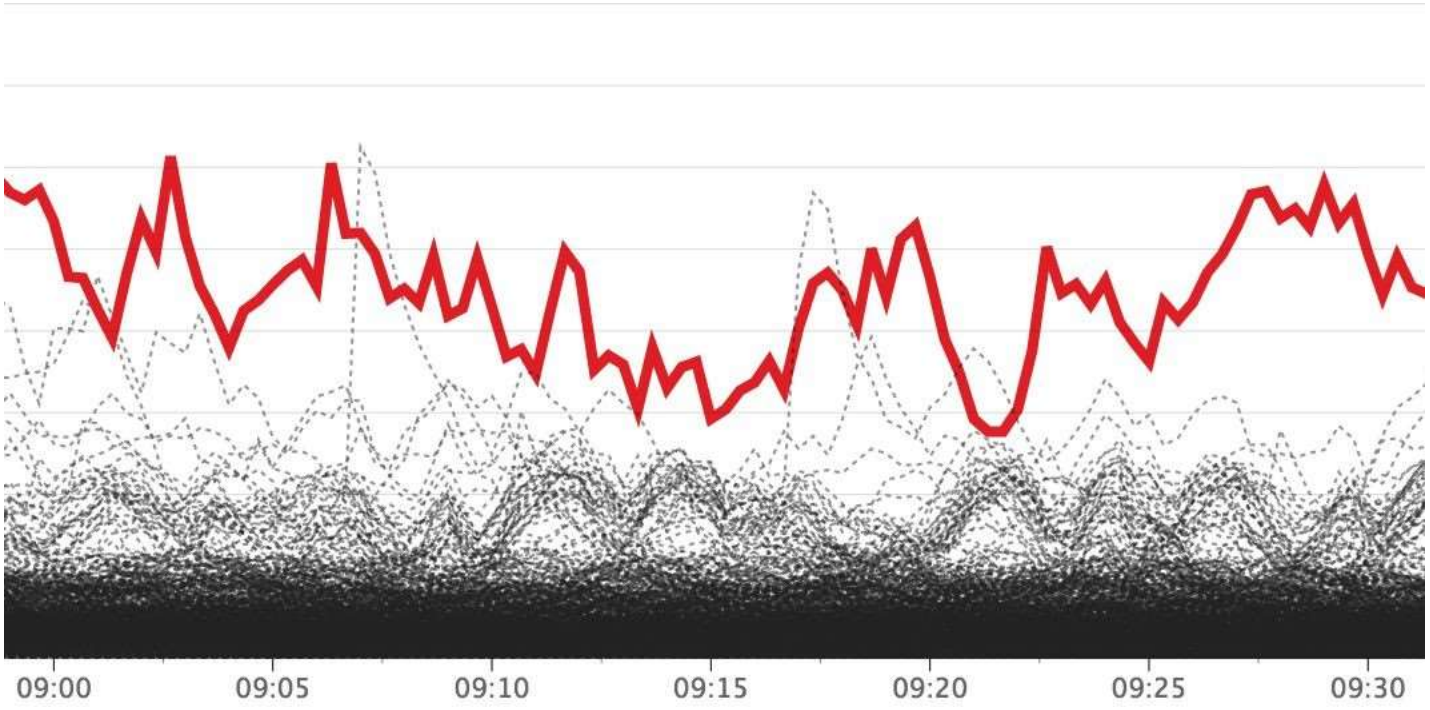


Outlier Detection

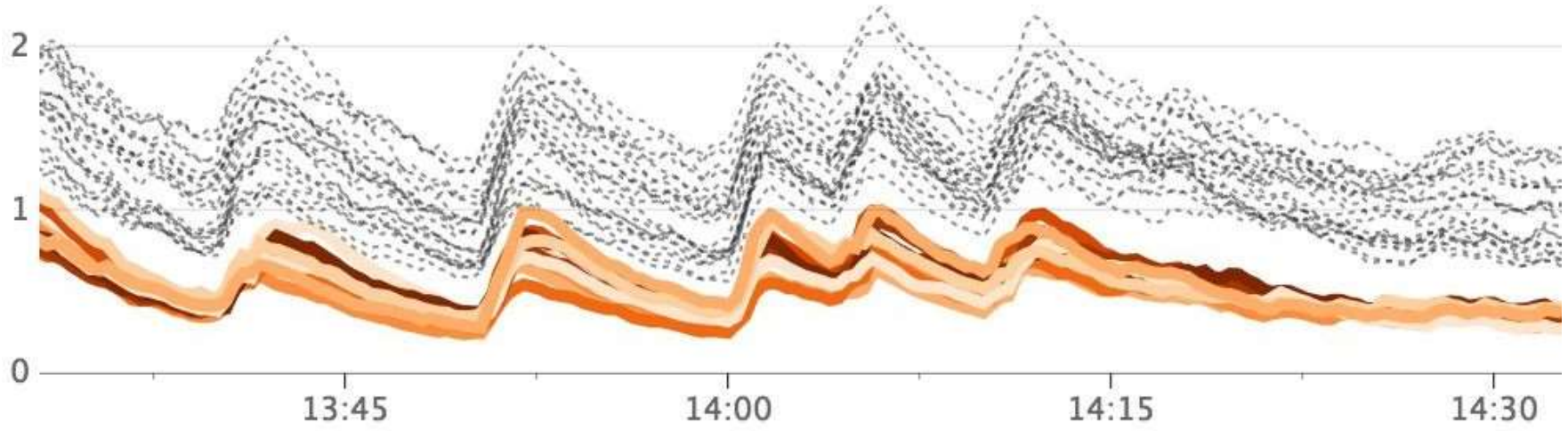
Outlier Detection



Outlier Detection



Outlier Detection



Outlier Detection Algorithms

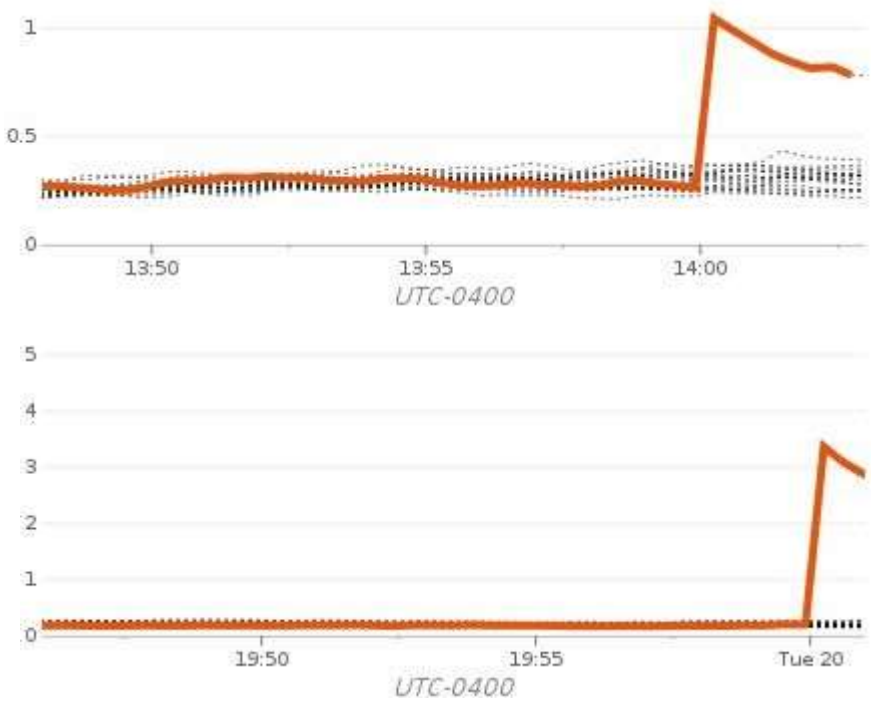
MAD

median absolute deviation

DBSCAN

density-based spatial clustering of applications with noise

Robust Outlier Detection Algorithms



Median Absolute Deviation

$$MAD(D) = median(\{|d_i - median(D)|\})$$

Median Absolute Deviation

$$MAD(D) = median(\{|d_i - median(D)|\})$$

$$D = \{1, 2, 3, 4, 5, 6, 100\}$$

Median Absolute Deviation

$$MAD(D) = median(\{|d_i - median(D)|\})$$

$$D = \{1, 2, 3, 4, 5, 6, 100\}$$

$$median = 4$$

Median Absolute Deviation

$$MAD(D) = median(\{|d_i - median(D)|\})$$

$$D = \{1, 2, 3, 4, 5, 6, 100\}$$

$$median = 4$$

$$deviations = \{-3, -2, -1, 0, 1, 2, 96\}$$

Median Absolute Deviation

$$MAD(D) = median(\{|d_i - median(D)|\})$$

$$D = \{1, 2, 3, 4, 5, 6, 100\}$$

$$median = 4$$

$$deviations = \{-3, -2, -1, 0, 1, 2, 96\}$$

$$abs\ deviations = \{0, 1, 1, 2, 2, 3, 96\}$$

Median Absolute Deviation

$$MAD(D) = median(\{|d_i - median(D)|\})$$

$$D = \{1, 2, 3, 4, 5, 6, 100\}$$

$$median = 4$$

$$deviations = \{-3, -2, -1, 0, 1, 2, 96\}$$

$$abs\ deviations = \{0, 1, 1, 2, 2, 3, 96\}$$

$$MAD = 2$$

Median Absolute Deviation

$$MAD(D) = median(\{|d_i - median(D)|\})$$

$$D = \{1, 2, 3, 4, 5, 6, 100\}$$

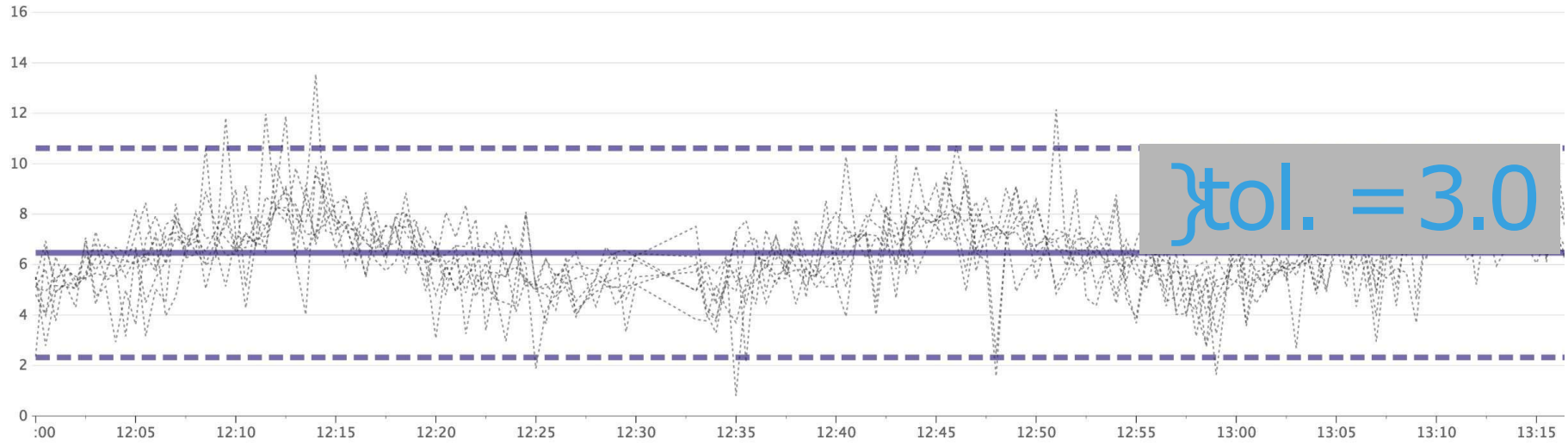
$$median = 4$$

$$deviations = \{-3, -2, -1, 0, 1, 2, 96\}$$

$$abs\ deviations = \{0, 1, 1, 2, 2, 3, 96\}$$

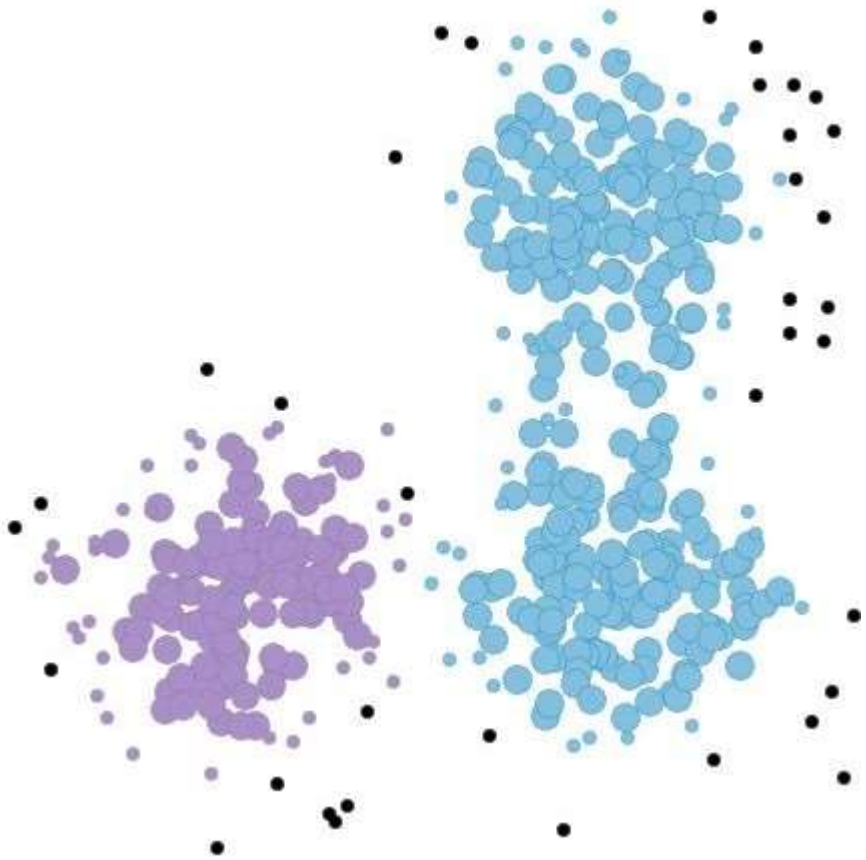
$$MAD = 2 \qquad (std\ dev = 33.8)$$

Median Absolute Deviation



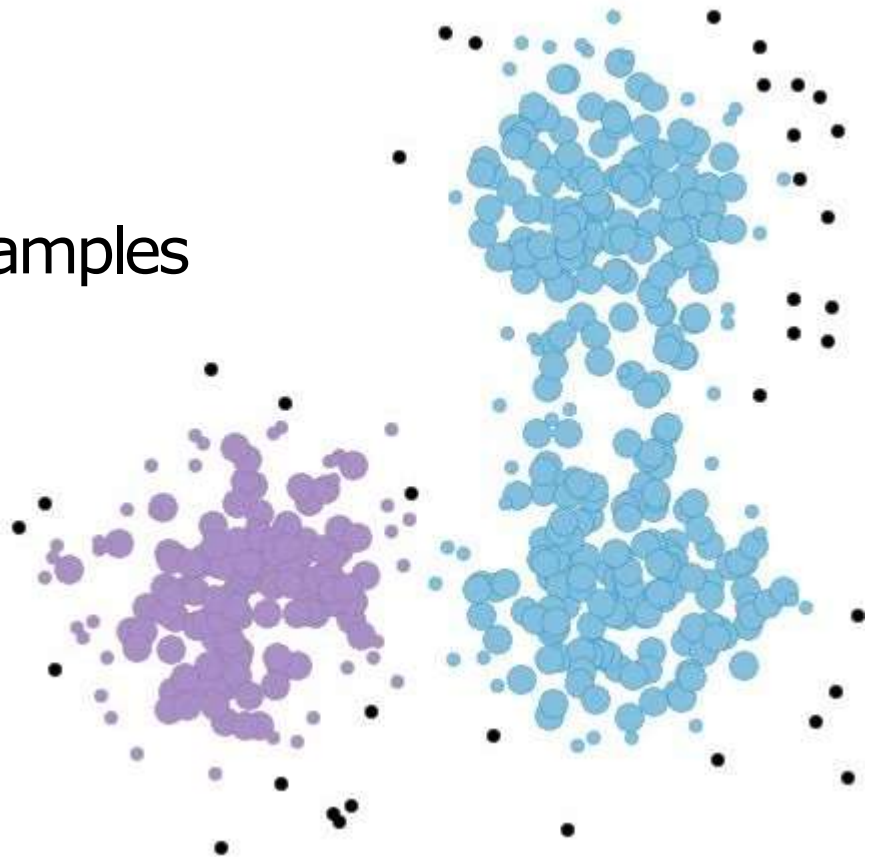
Parameters: Tolerance, Pct

DBSCAN

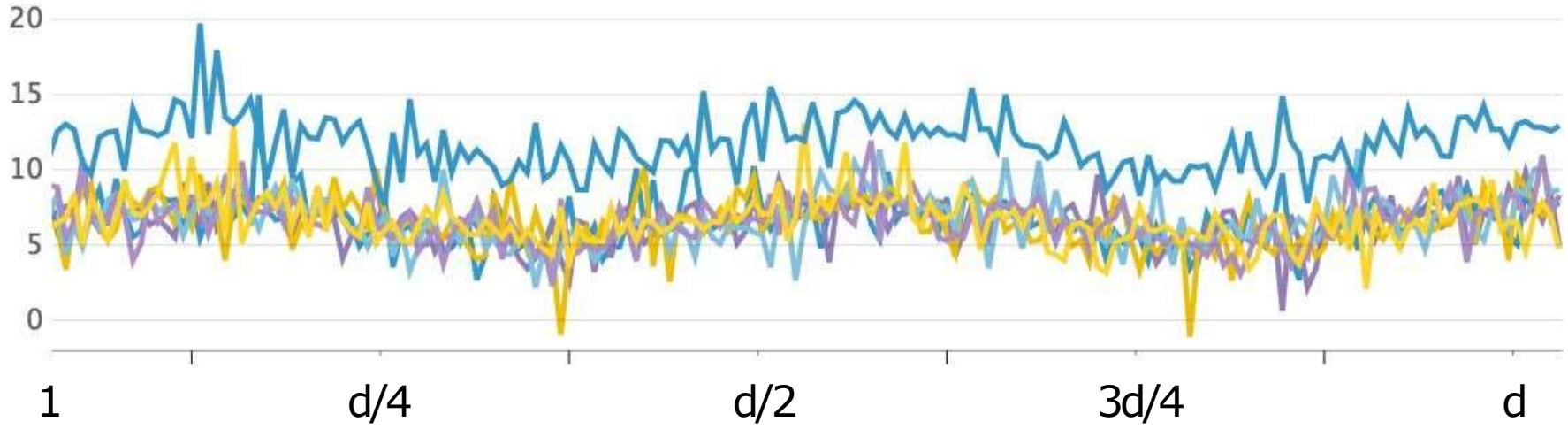


DBSCAN

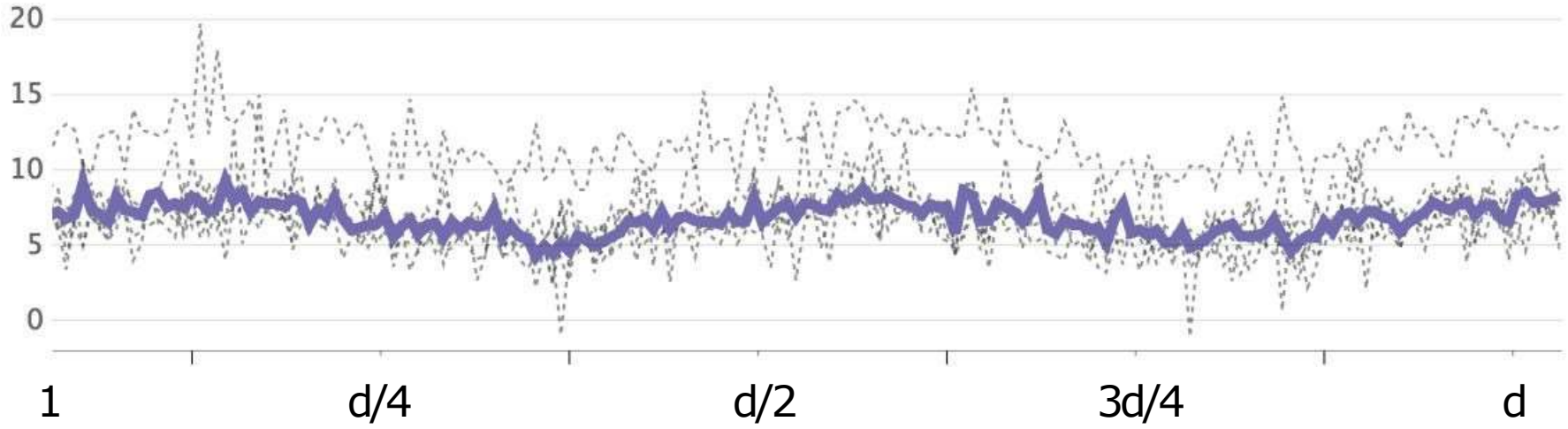
Parameters:
epsilon, min_samples



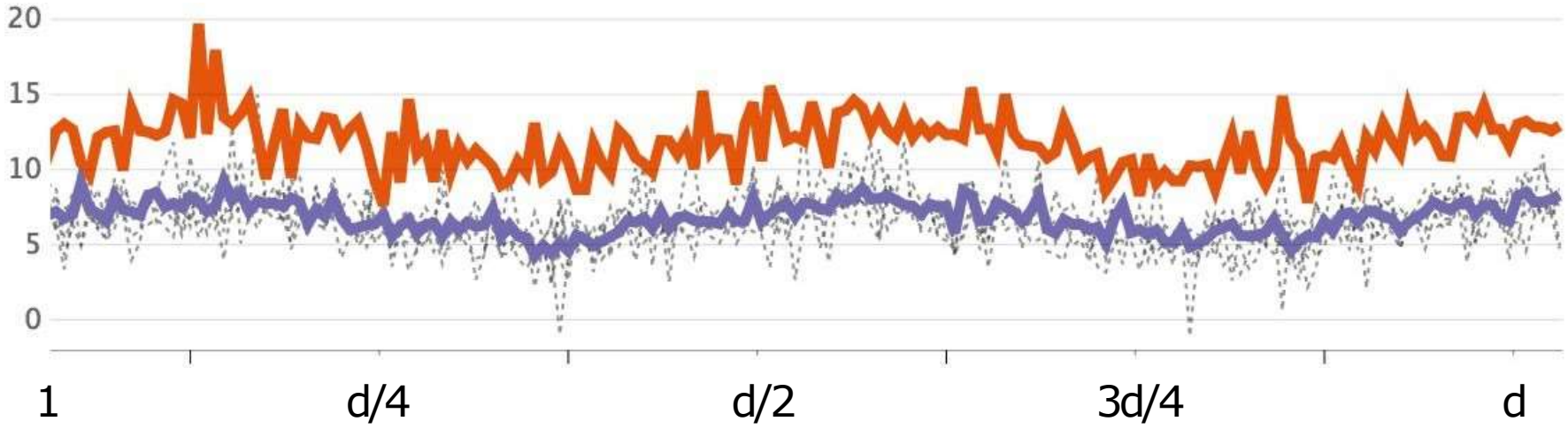
DBSCAN



DBSCAN

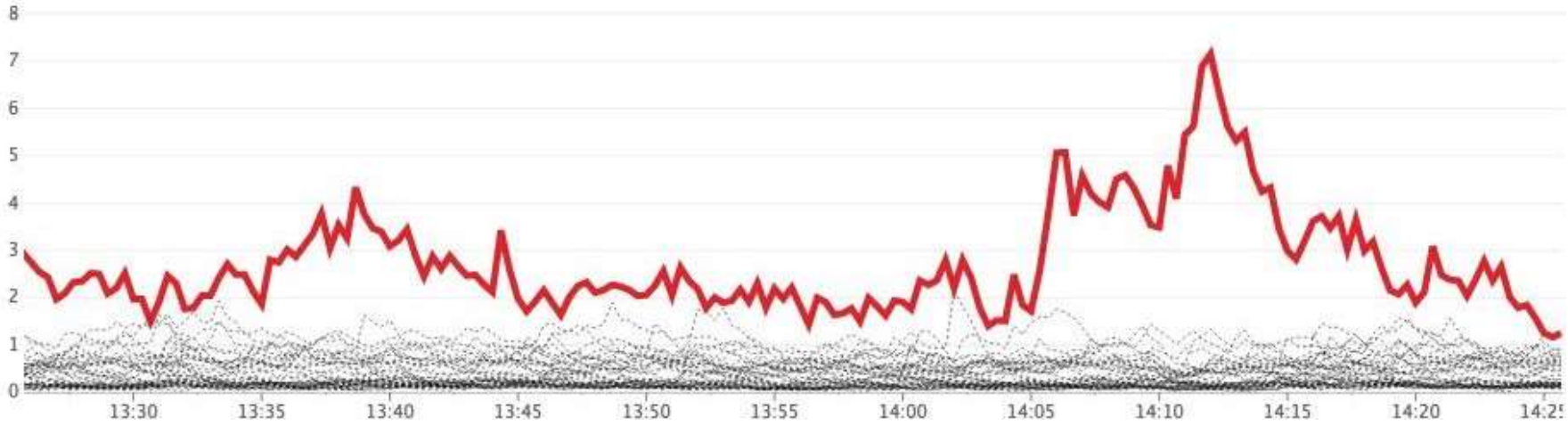


DBSCAN

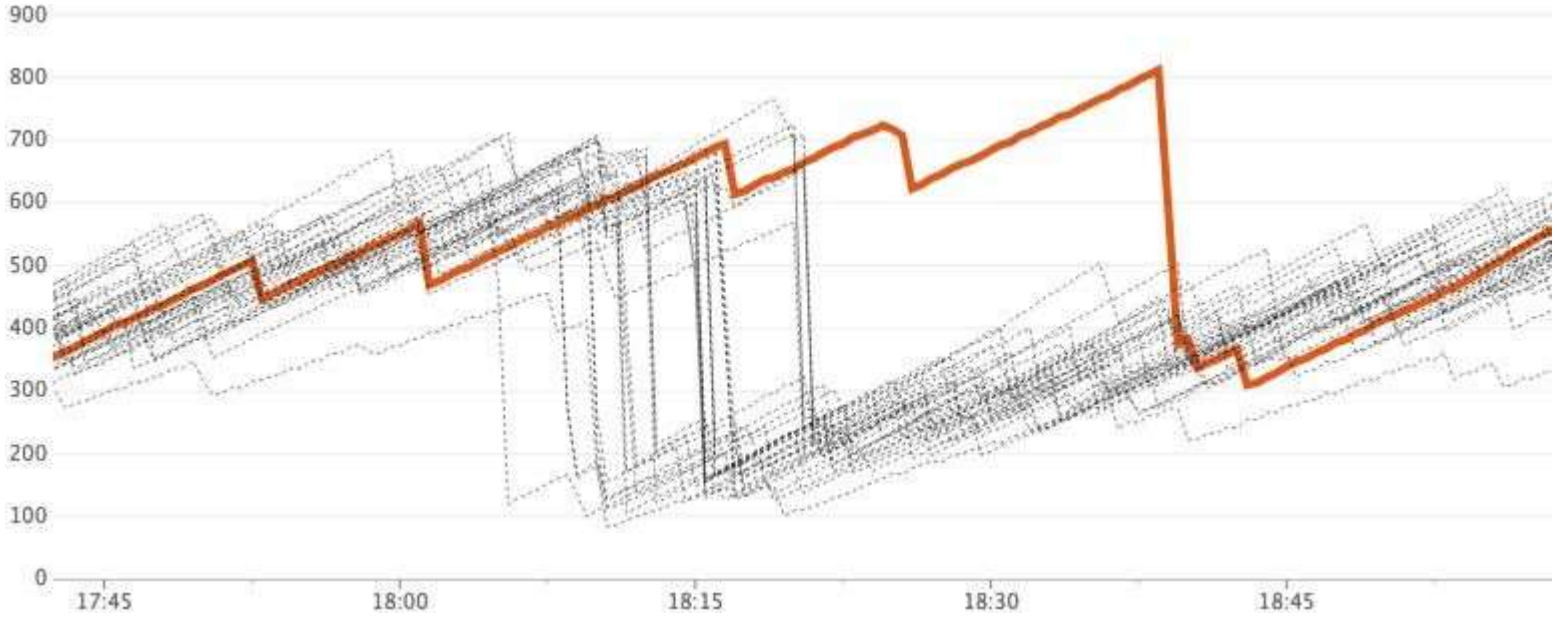


‹ ~median(dist from median series) ×tolerance

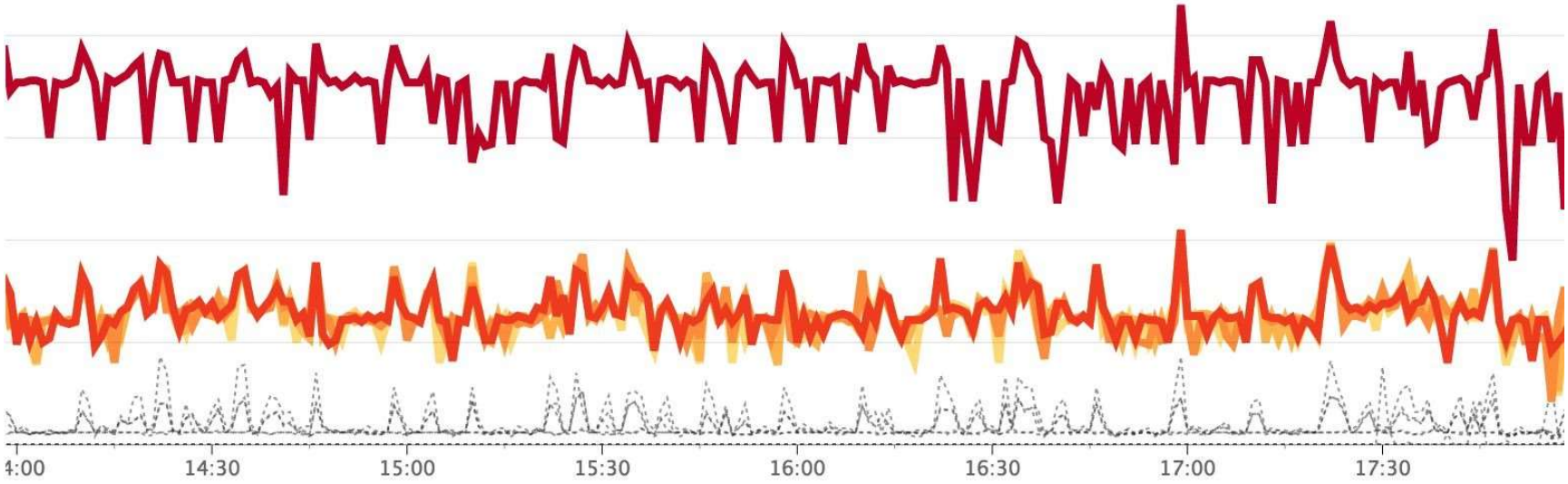
MAD or DBSCAN?



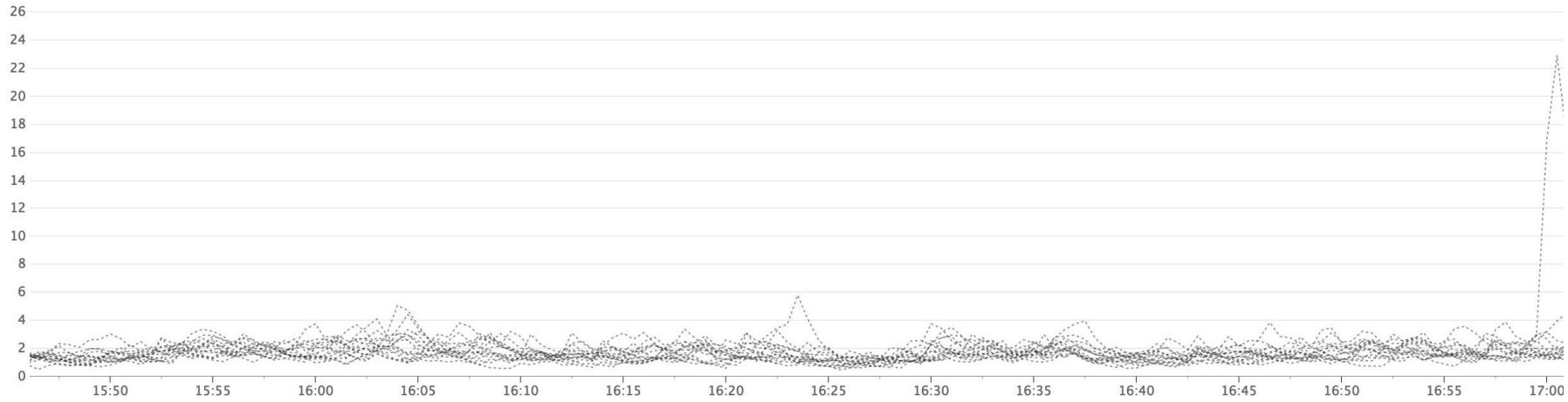
MAD or DBSCAN?



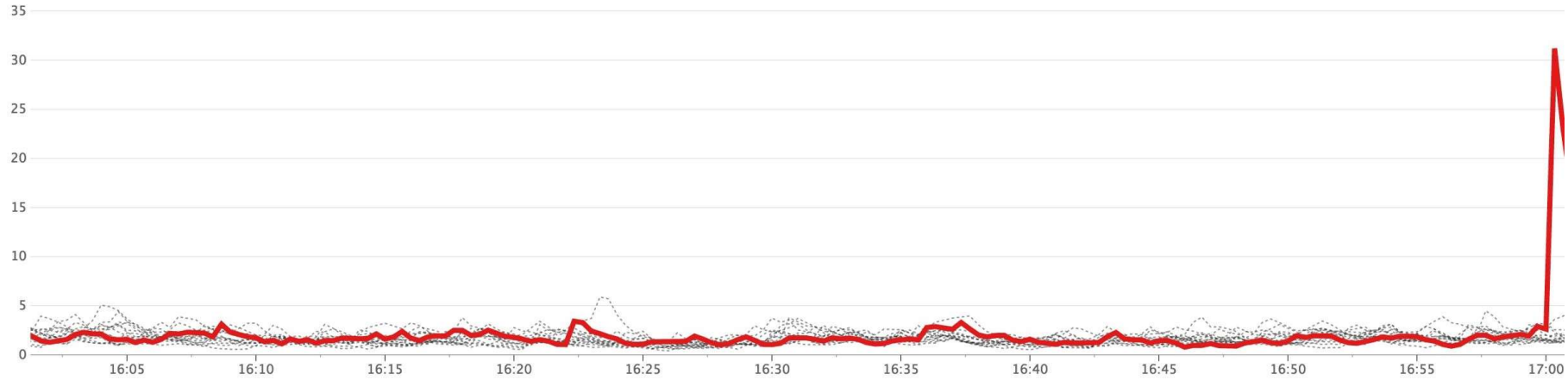
Some subtleties



Some subtleties

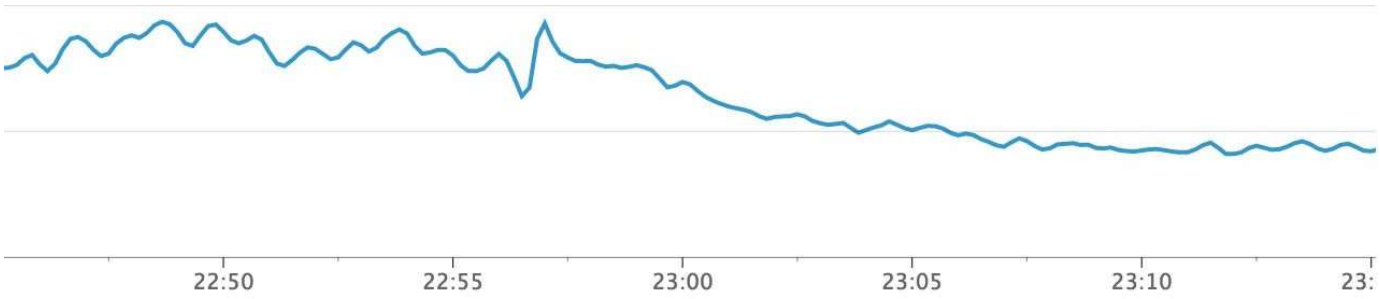


Some subtleties



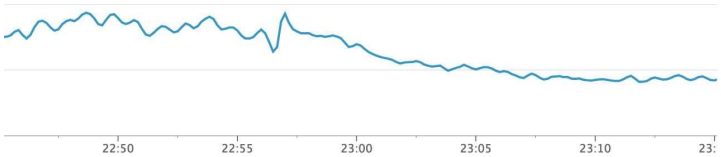
Anomaly Detection

An Investigation

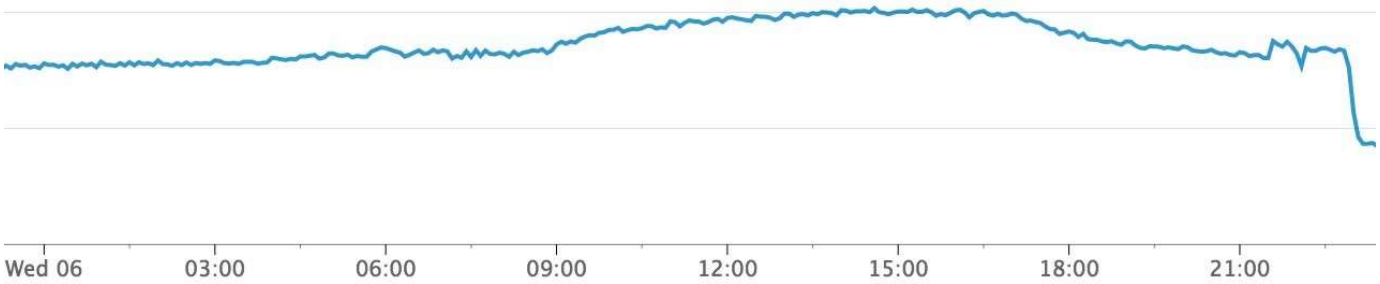


past 30 minutes

An Investigation

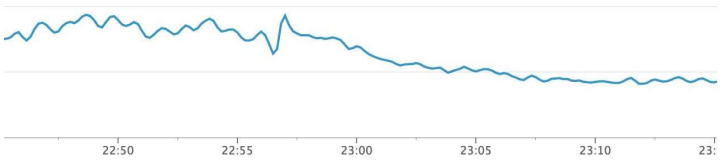


past 30 minutes

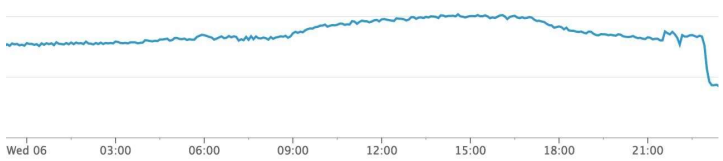


past day

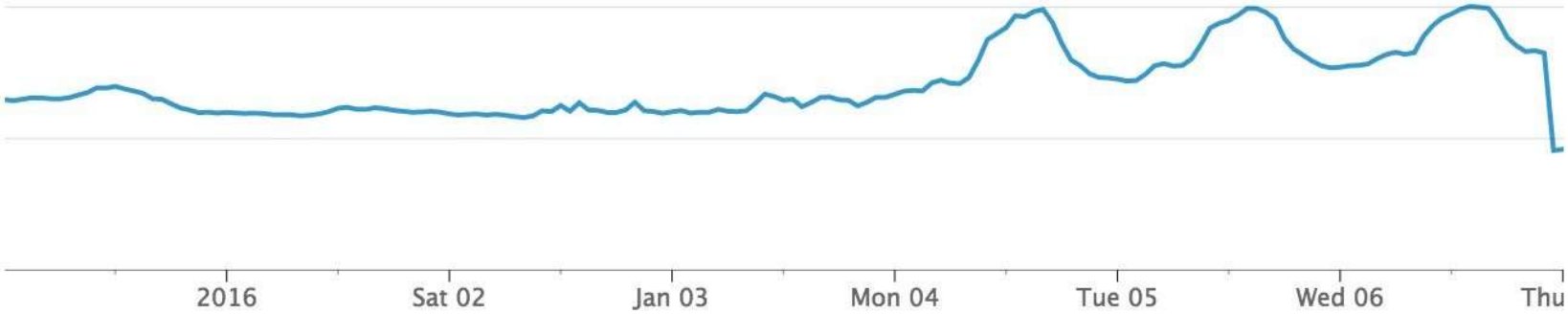
An Investigation



past 30 minutes

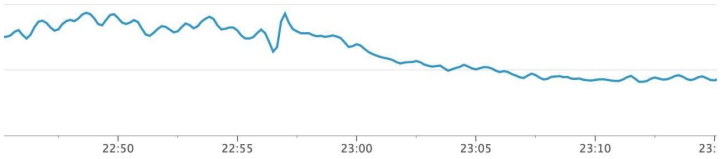


past day

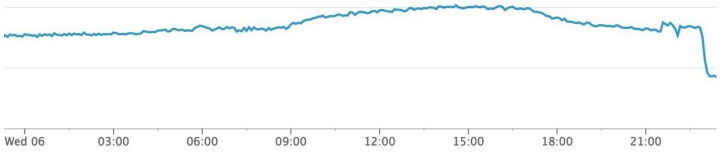


past week

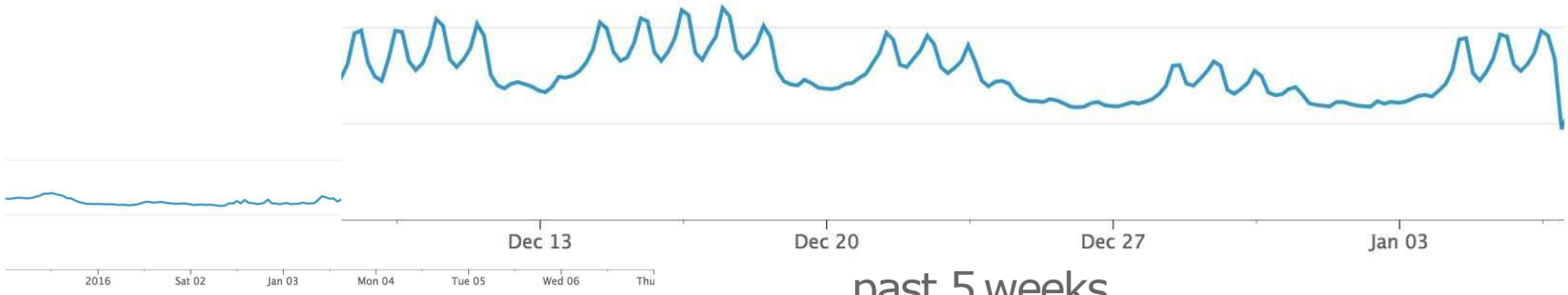
An Investigation



past 30 minutes



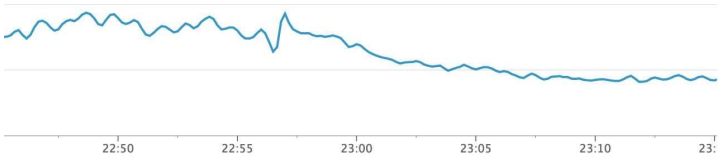
past day



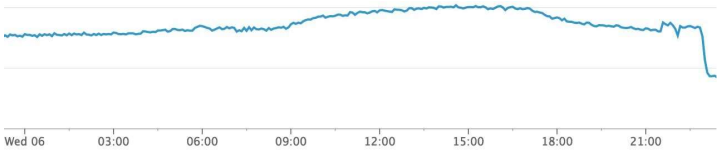
past week

past 5 weeks

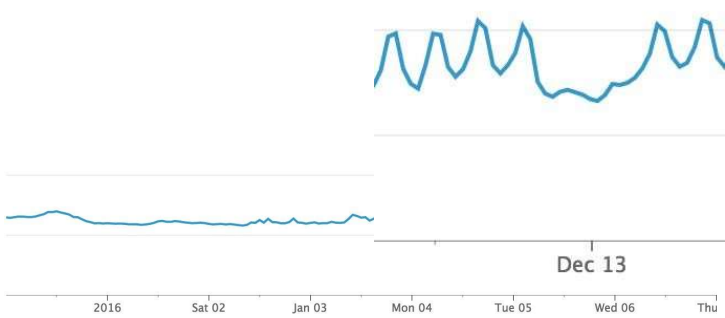
An Investigation



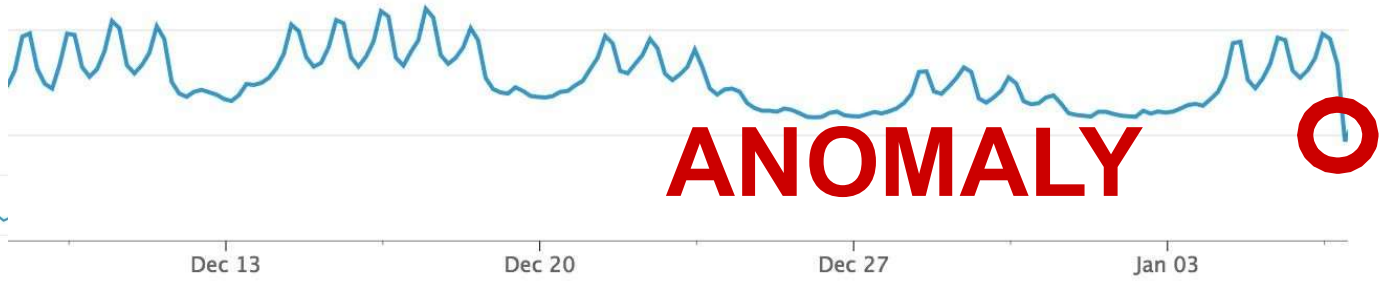
past 30 minutes



past day



past week



ANOMALY

past 5 weeks

Anomalies

A time series point is an anomaly if:

- Given the past points in the series (®®®®®), the point in question (®) is unlikely given your model of the past;

Anomalies

A time series point is an anomaly if:

- Given the past points in the series (®®®®®), the point in question (®) is unlikely given your model of the past;

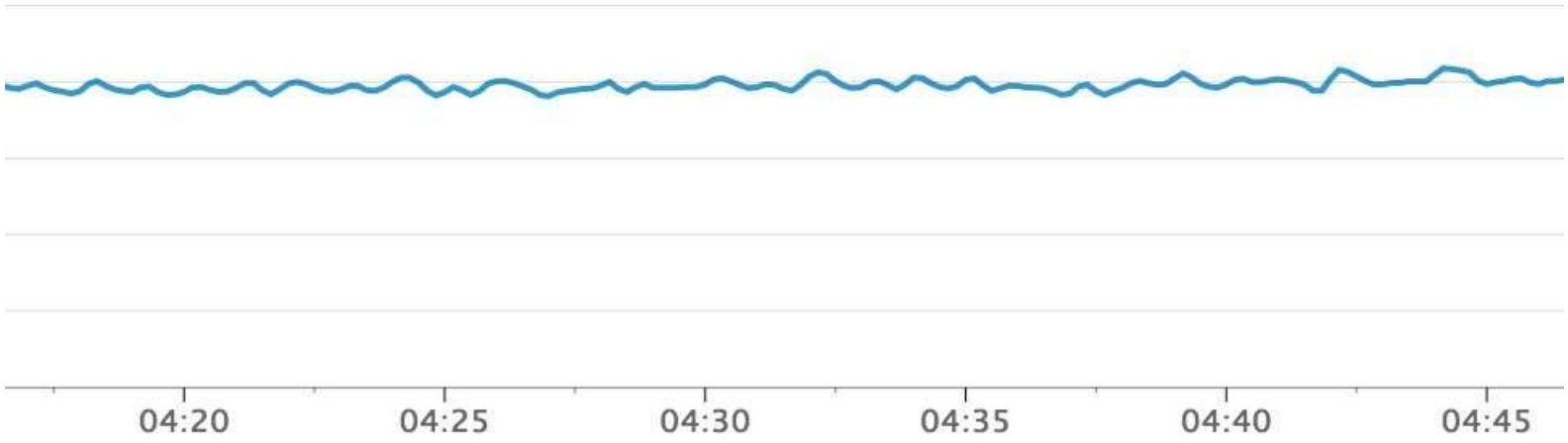
and you should alert on a set of anomalies if:

- they are a symptom of an issue you care about (Î).

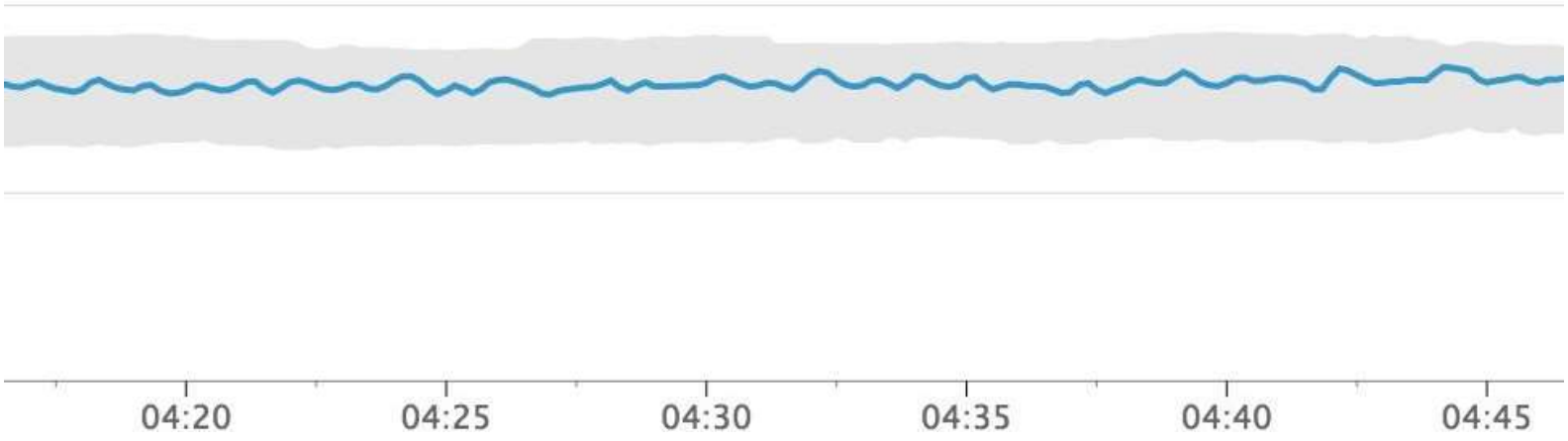
Our Approach

1. Extract as much signal as we can from the time series.
2. Use robust statistical measures when creating the model.
3. Give the user control over when they get alerted.

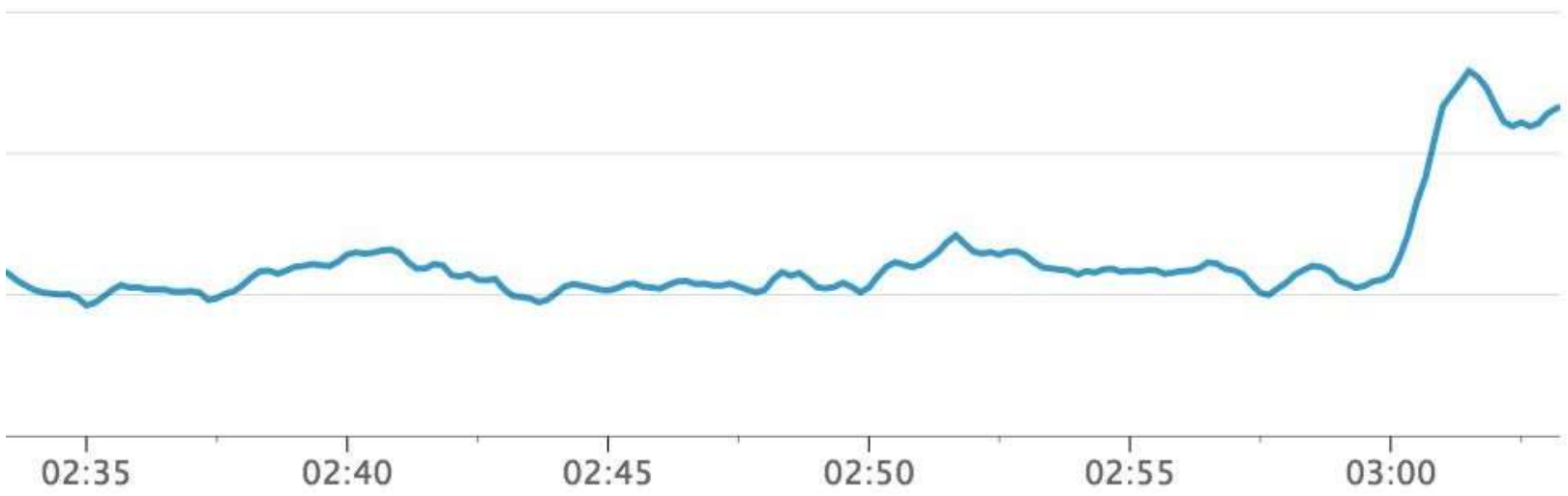
What's Normal?



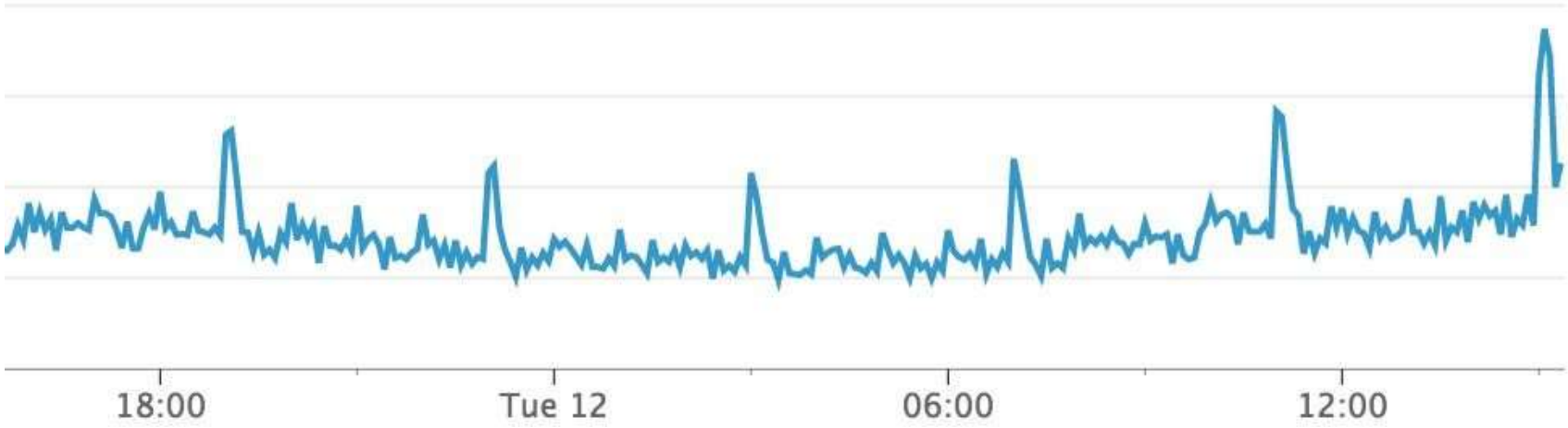
What's Normal?



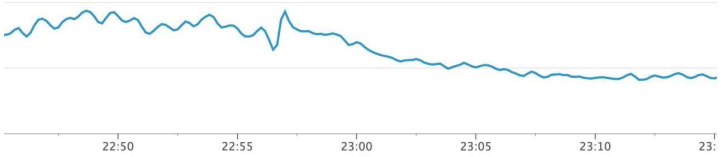
What's Normal?



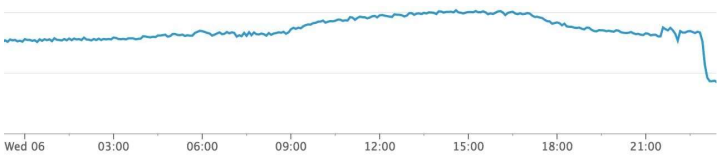
What's Normal?



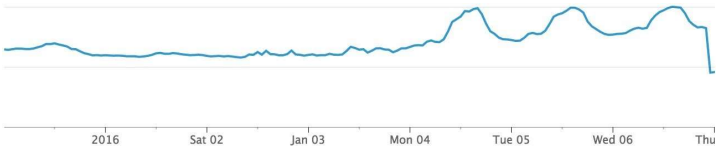
Past Performance...



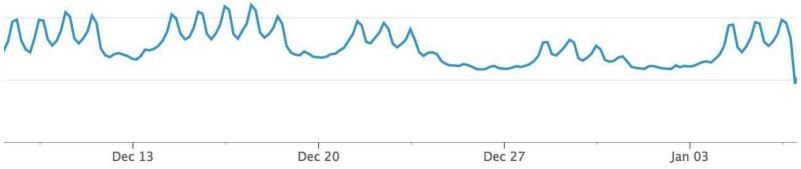
past 30 minutes



past day

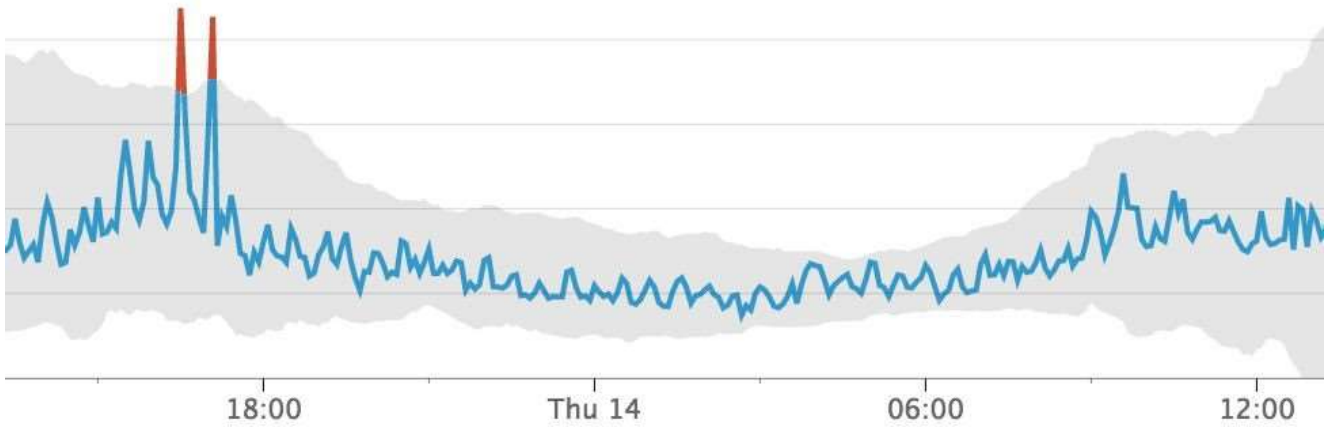


past week

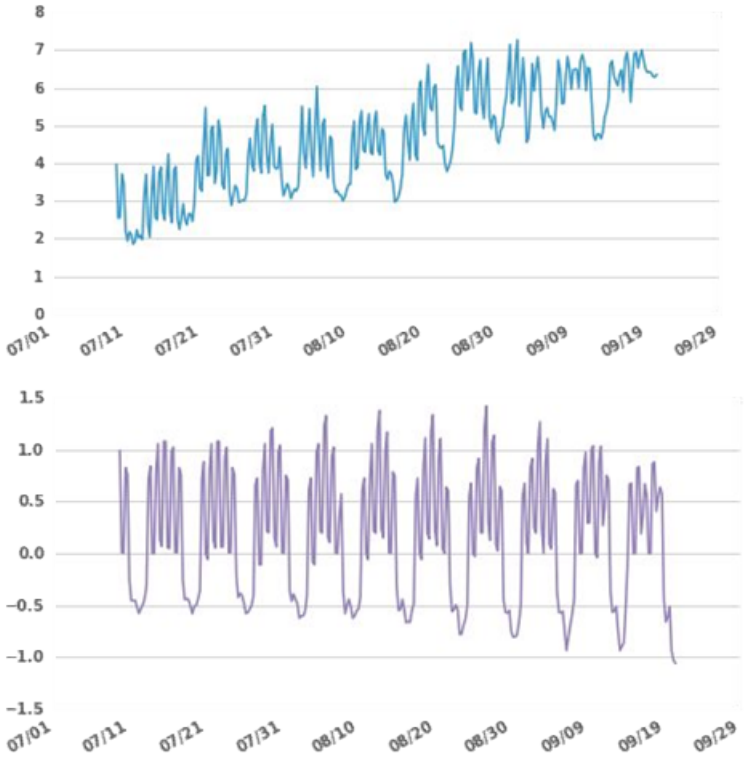


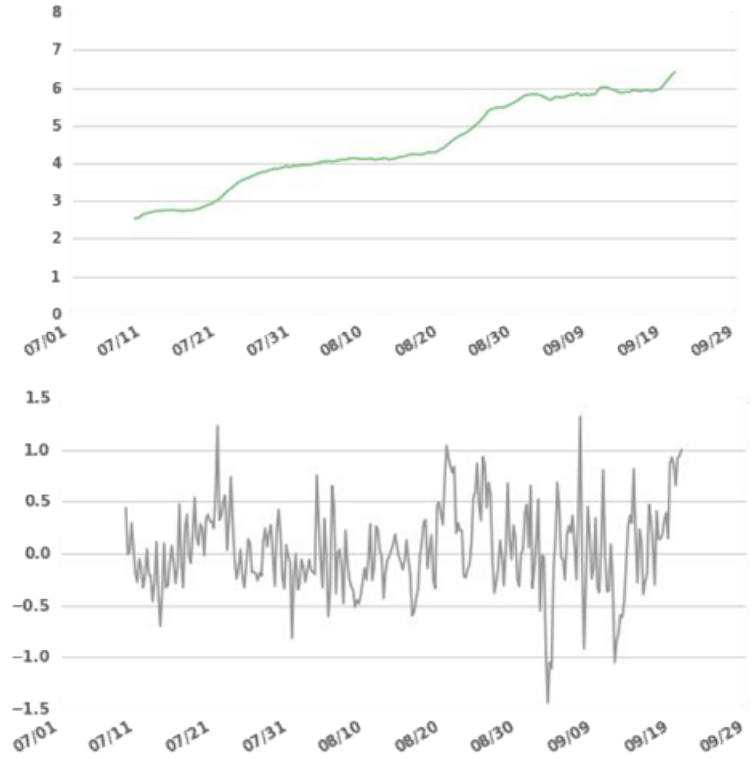
past 5 weeks

Past Performance...

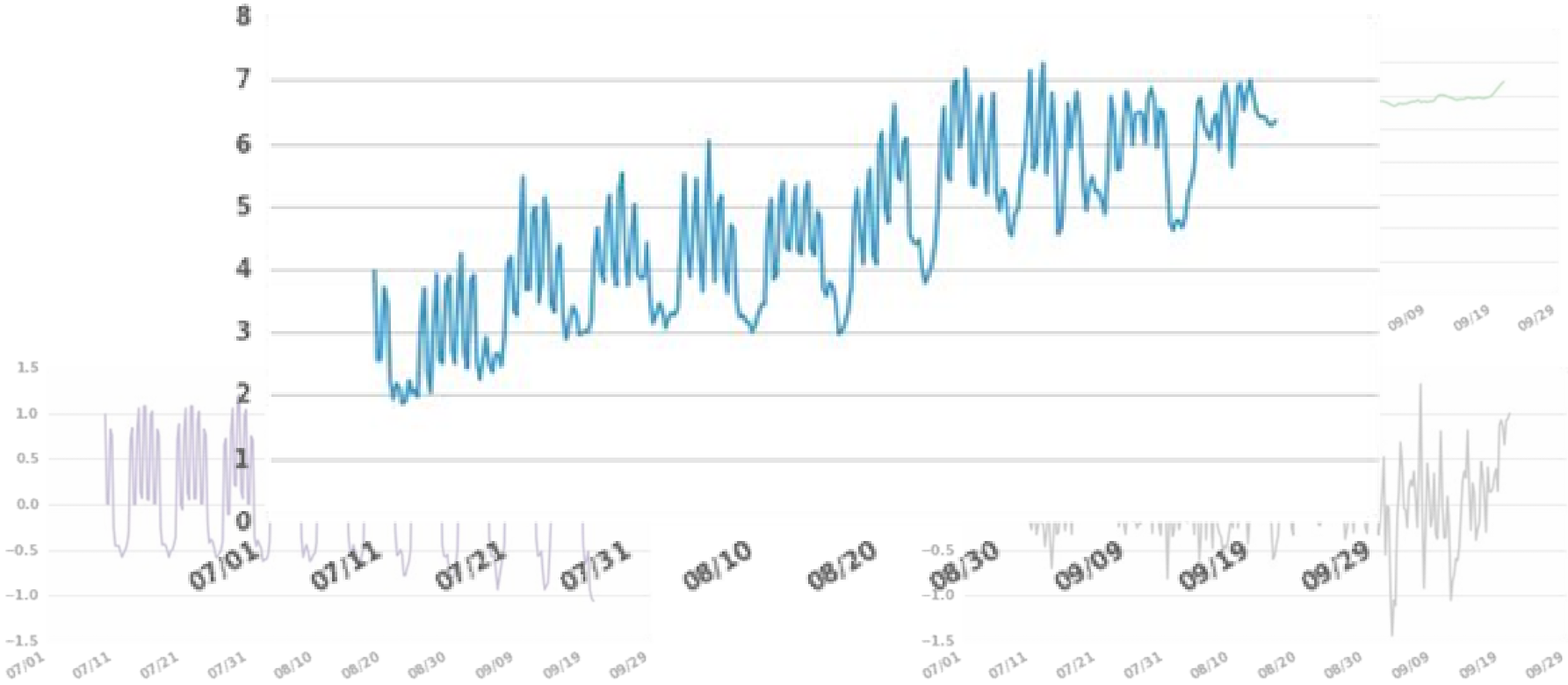


Decomposition



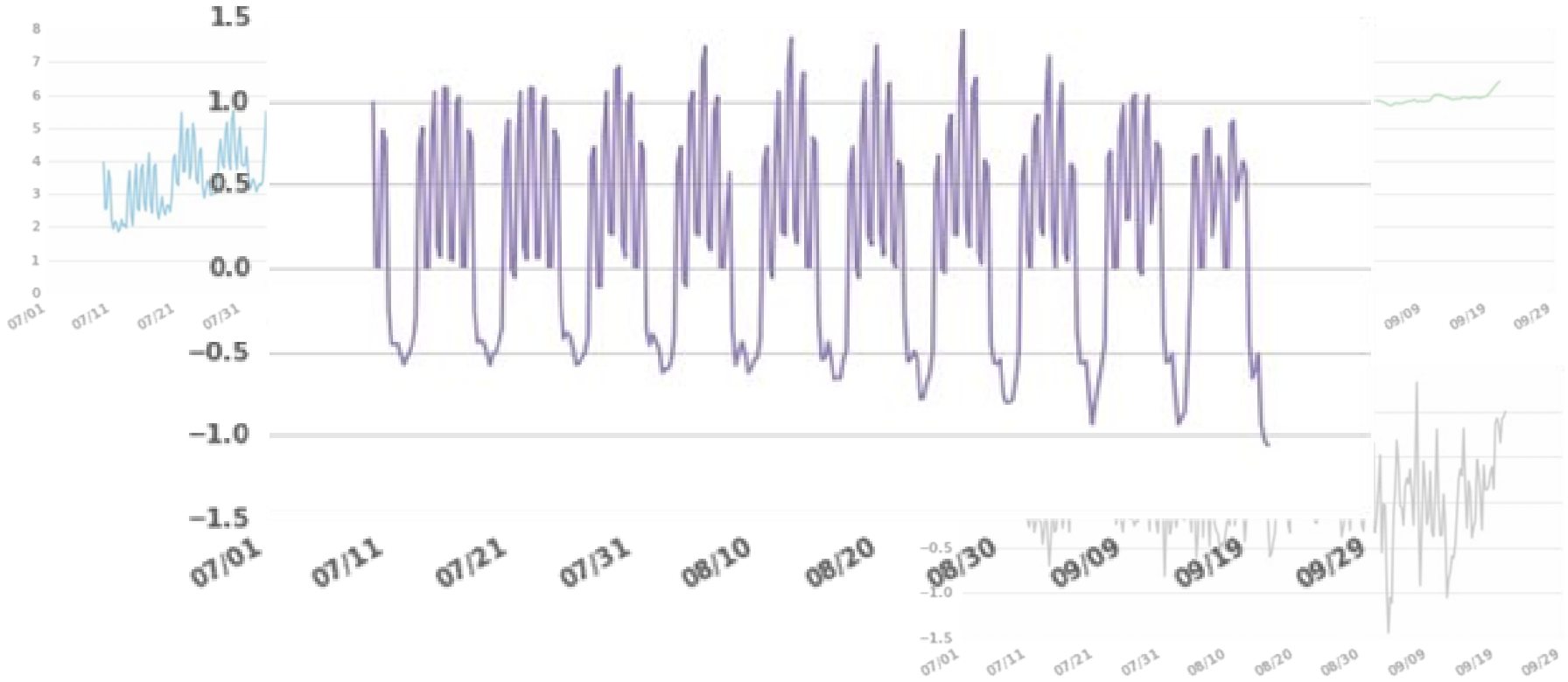


Decomposition

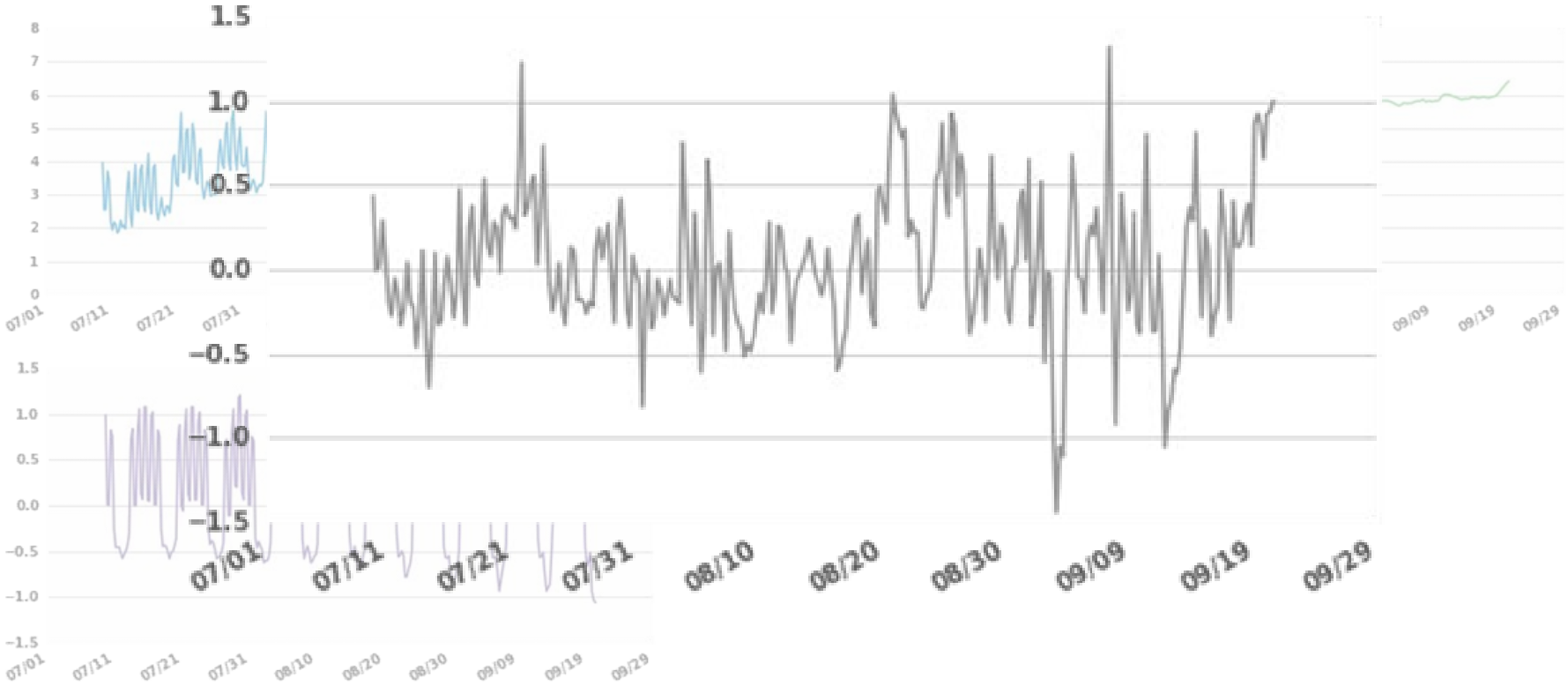


Decomposition

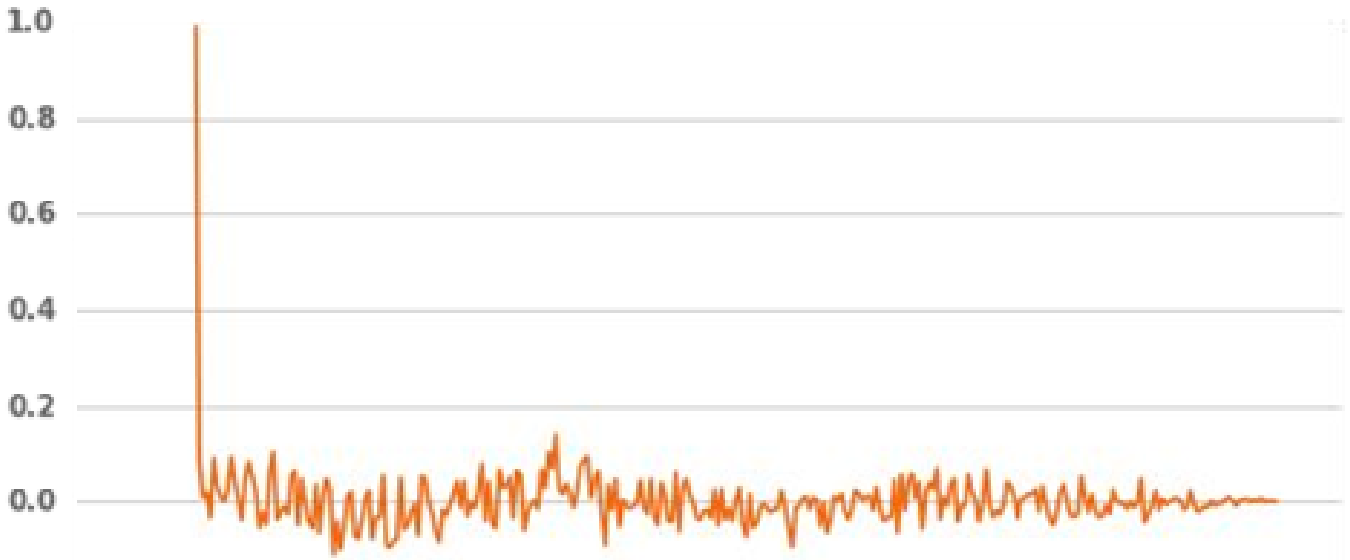
Decomposition



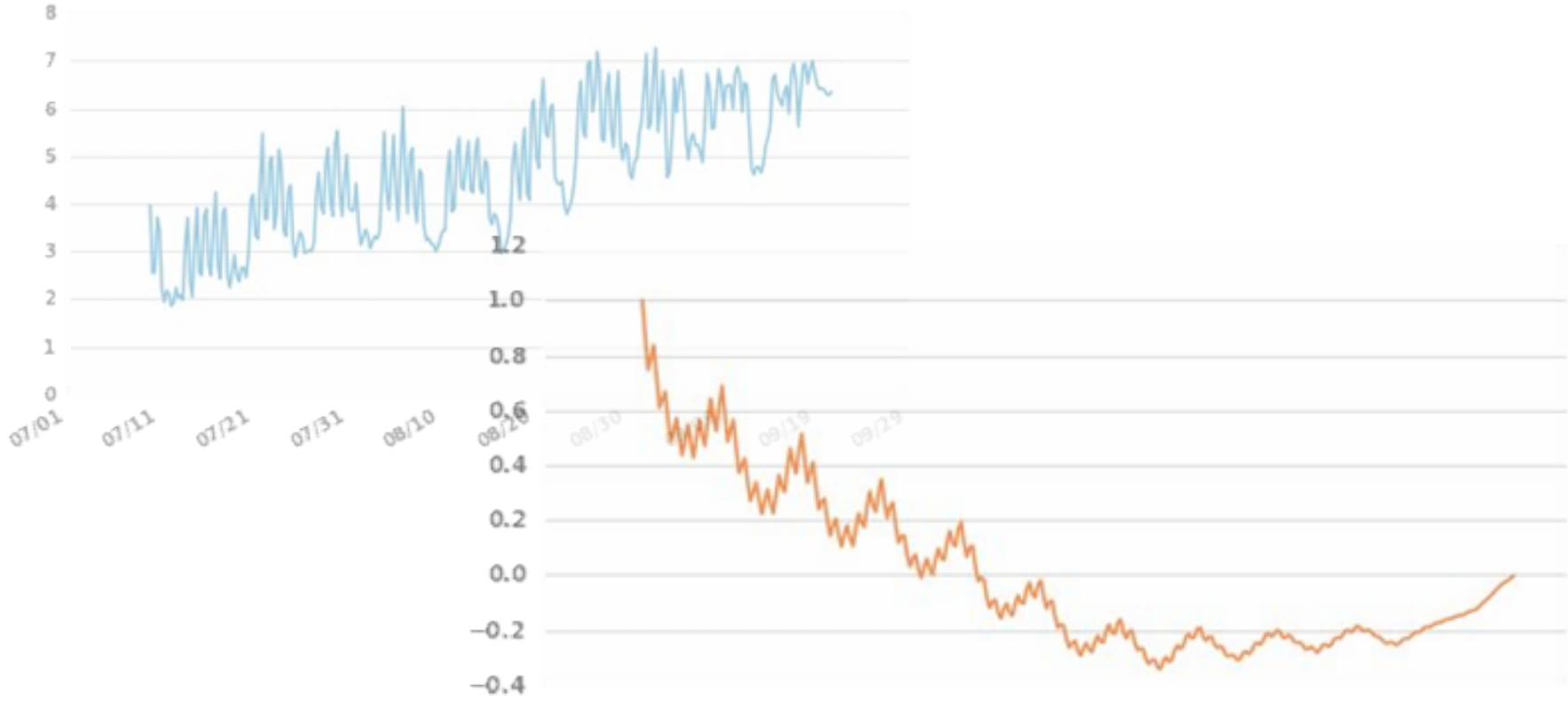
Decomposition



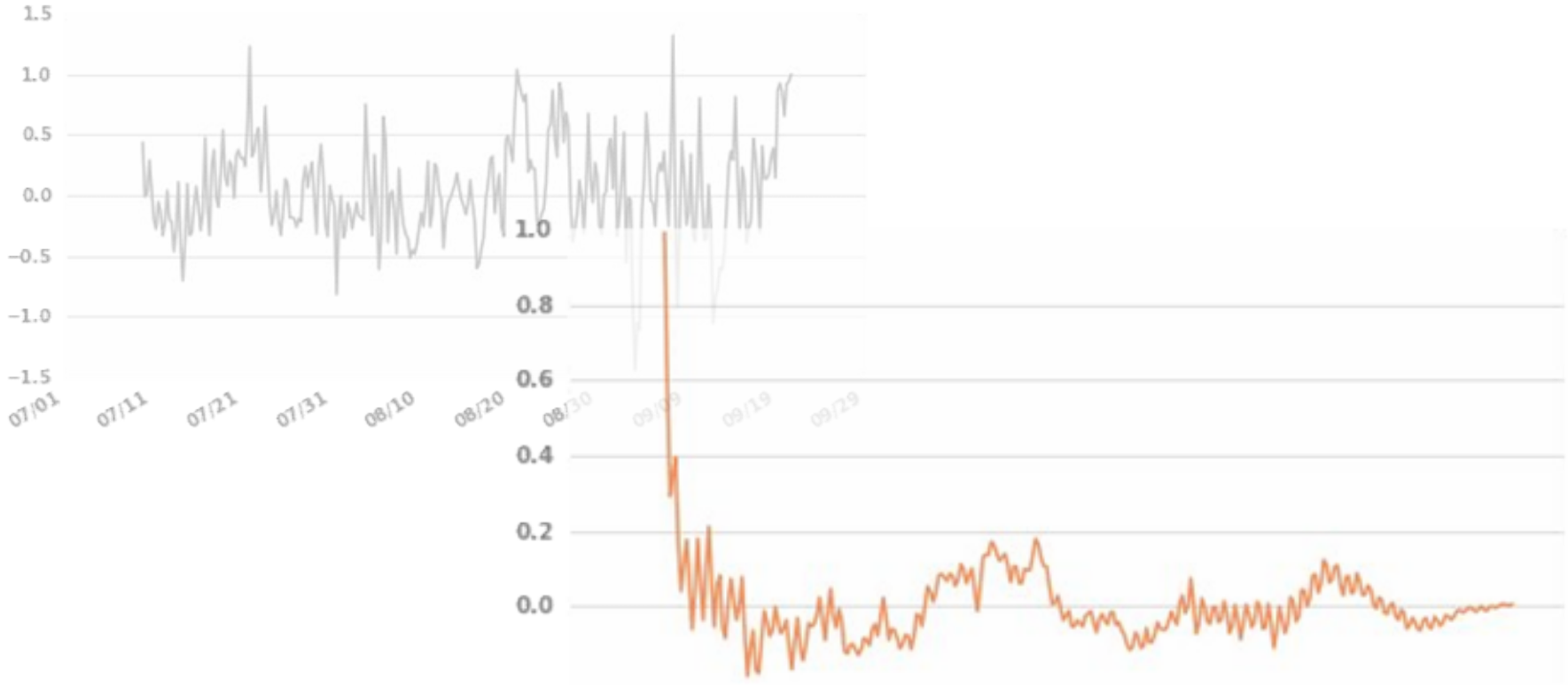
Autocorrelation



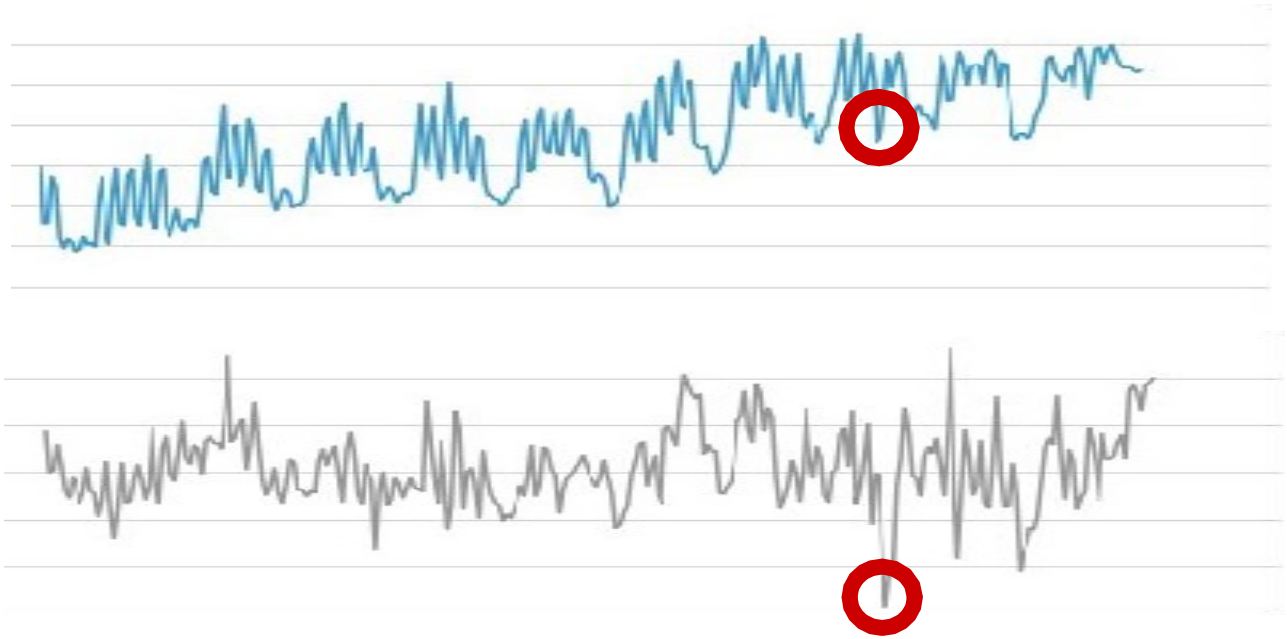
Signal vs. Noise



Signal vs. Noise

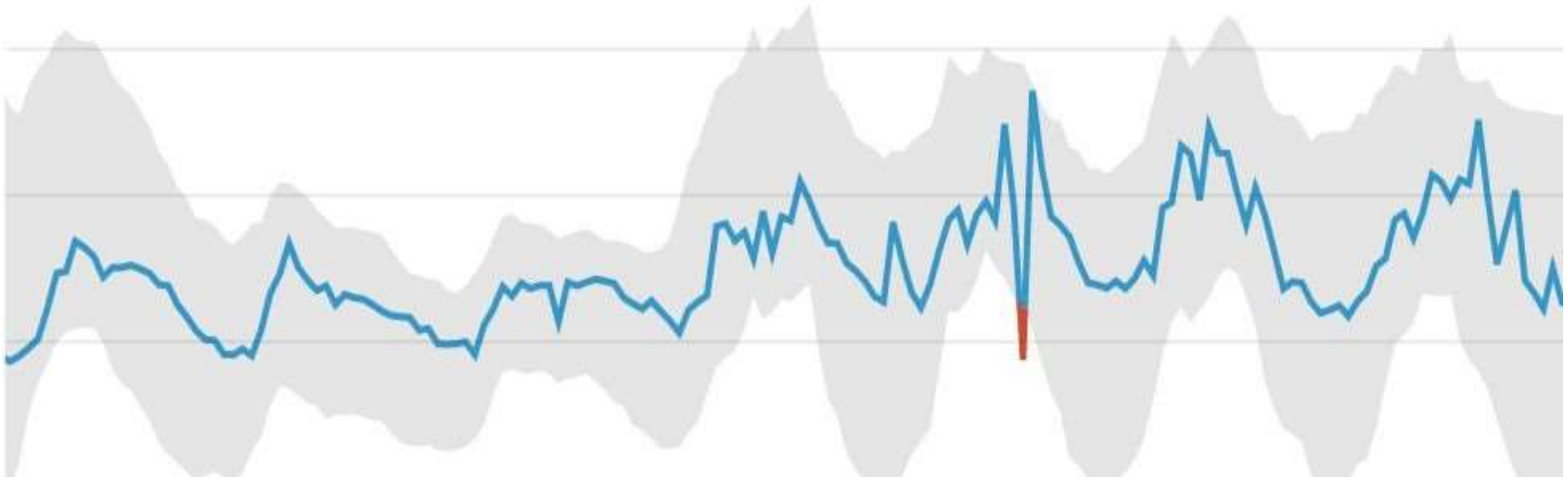


Signal vs. Noise vs. Signal

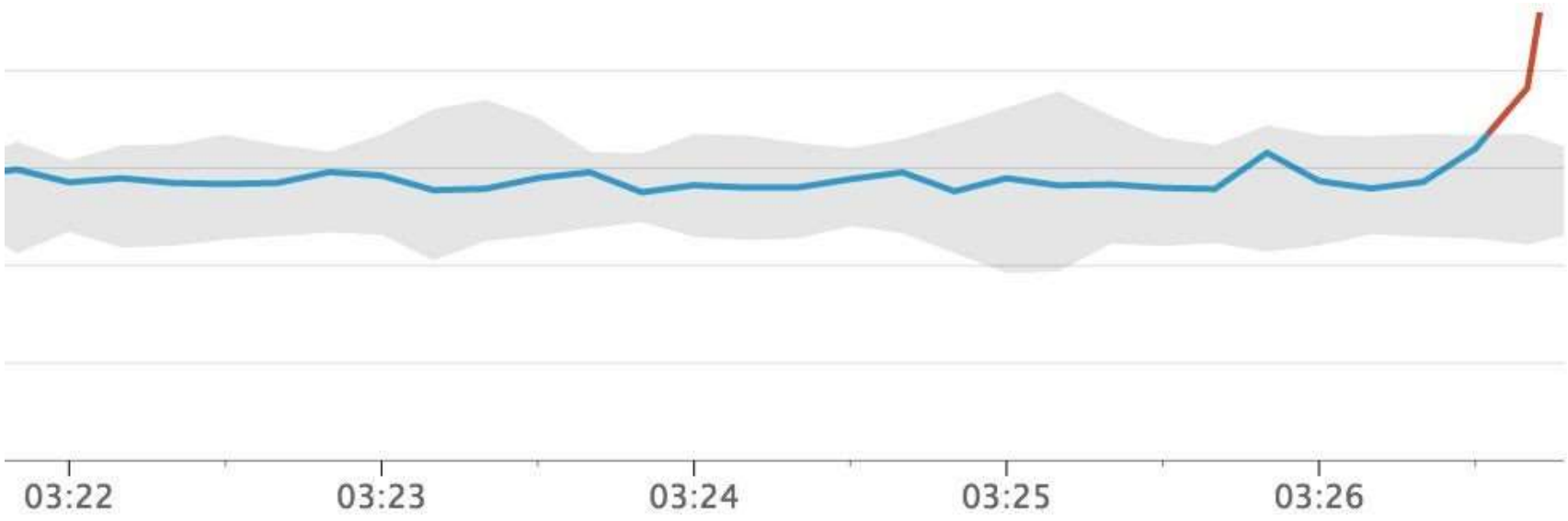


Real-time Anomaly Detection

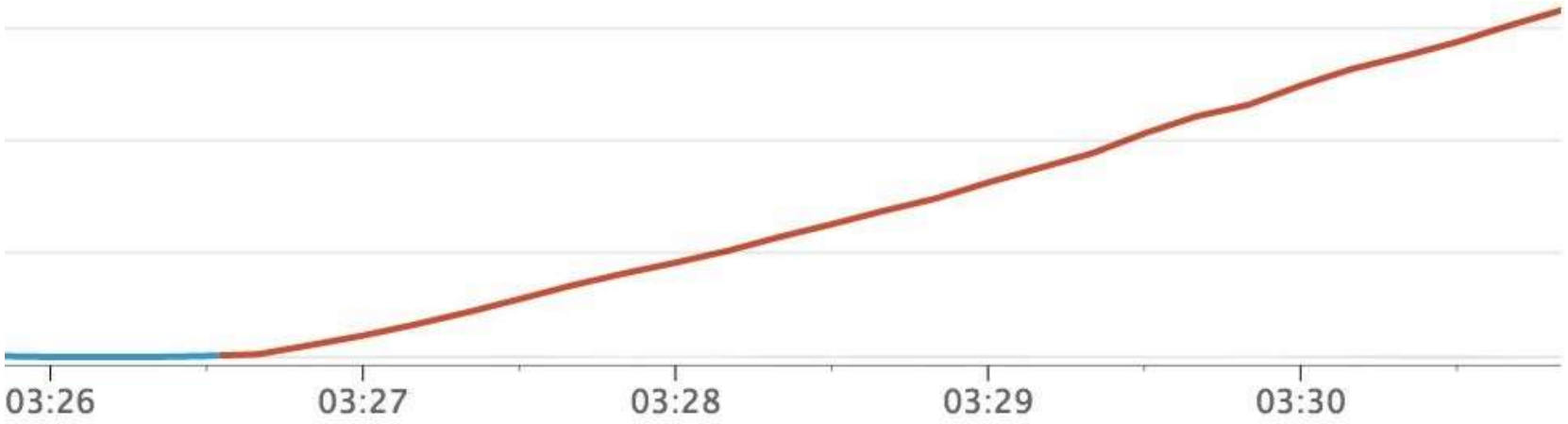
Anomaly Detection



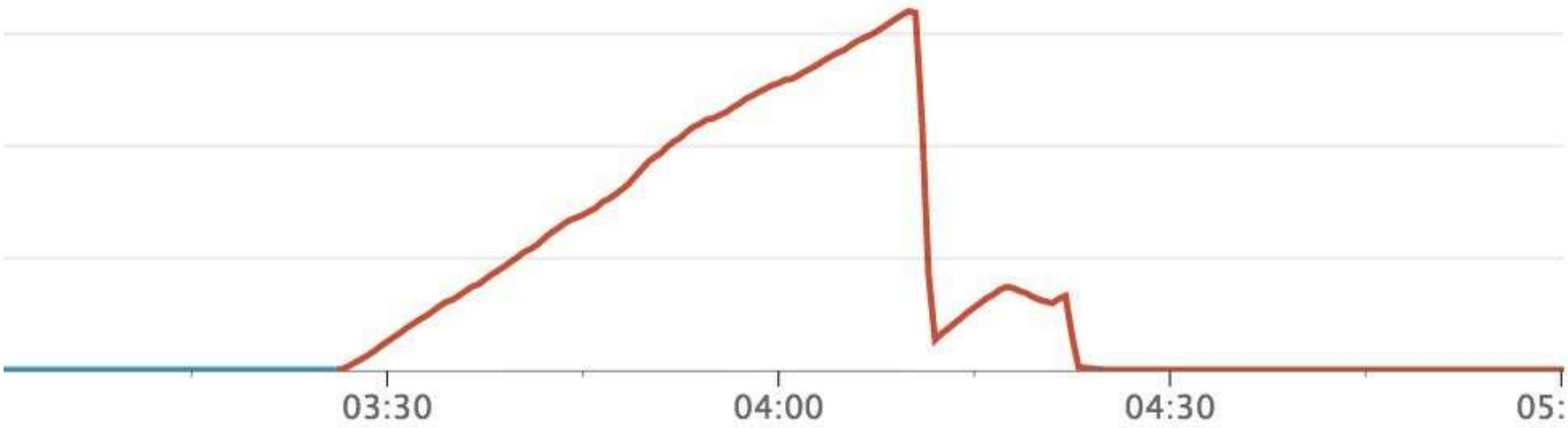
Robust Anomaly Detection



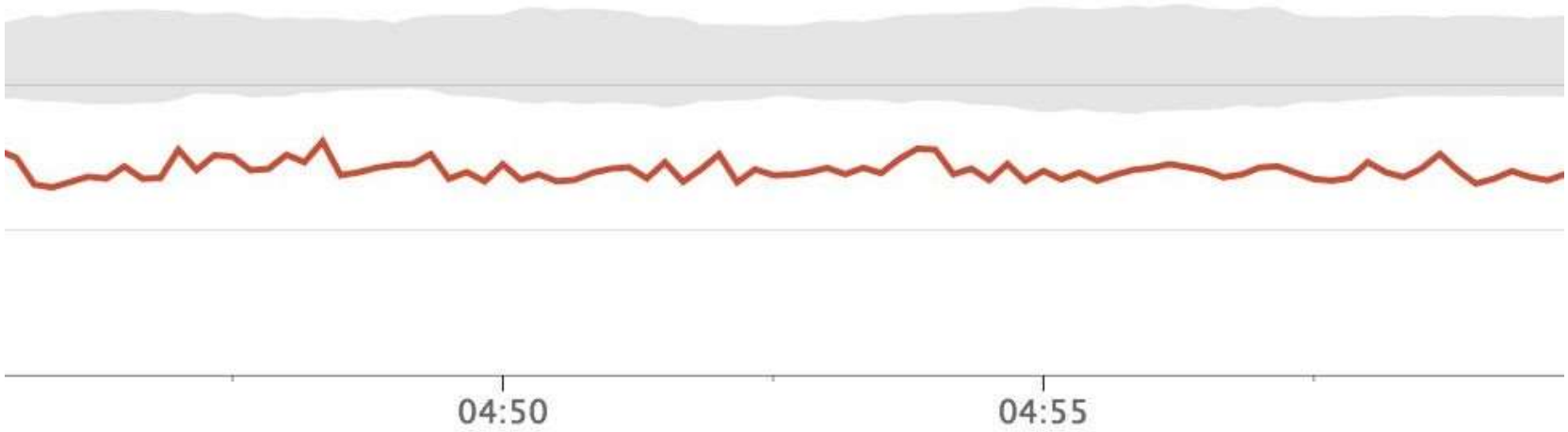
Robust Anomaly Detection



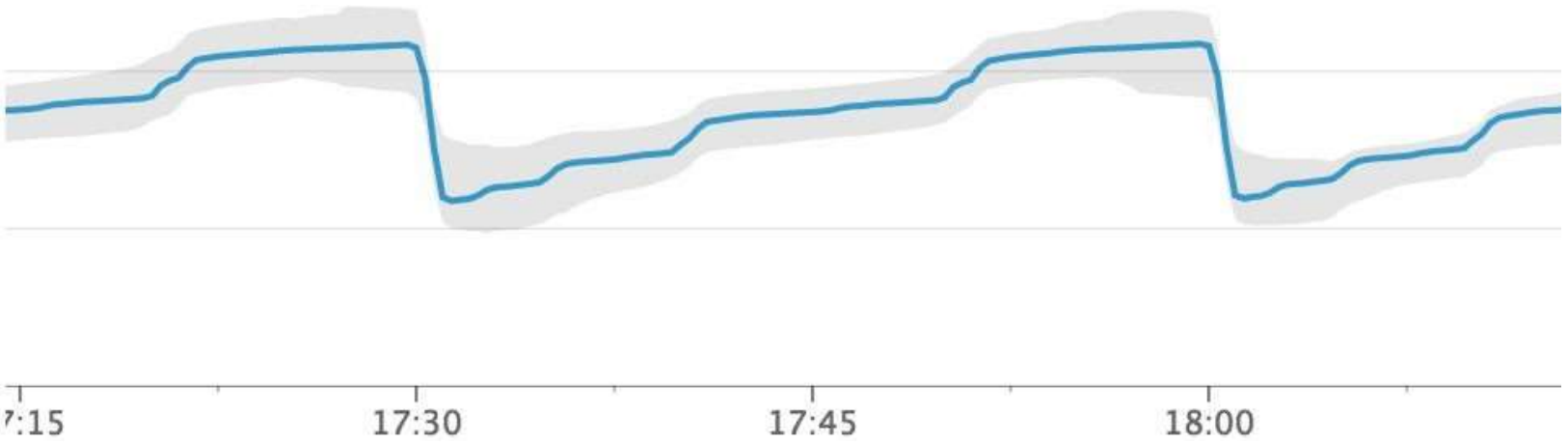
Robust Anomaly Detection



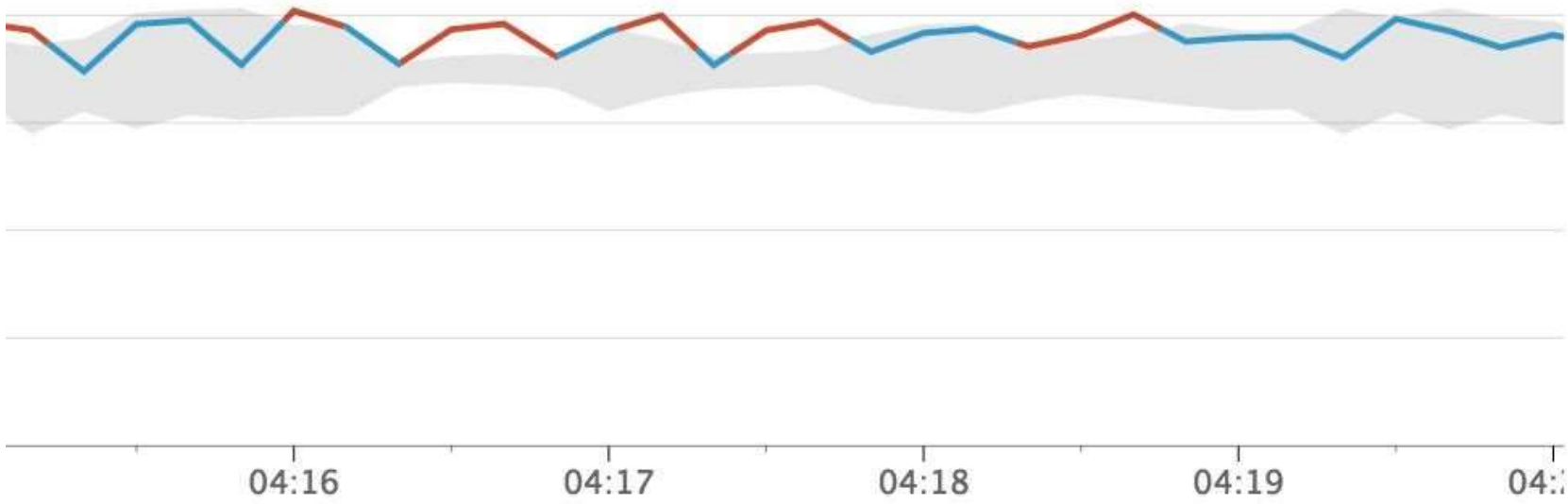
Robust Anomaly Detection



Alerting



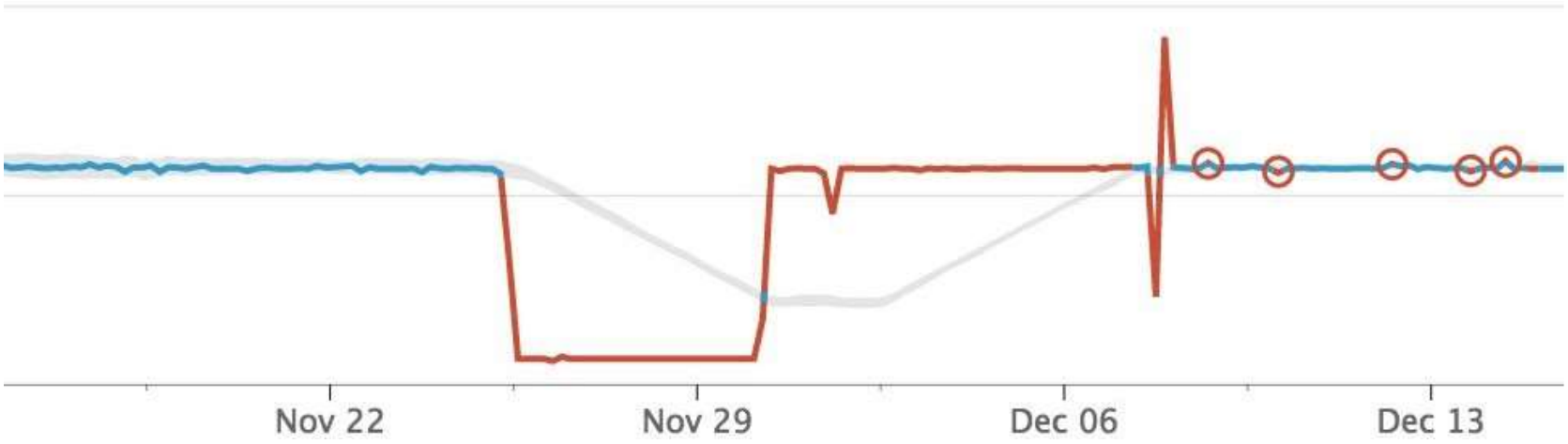
Alerting



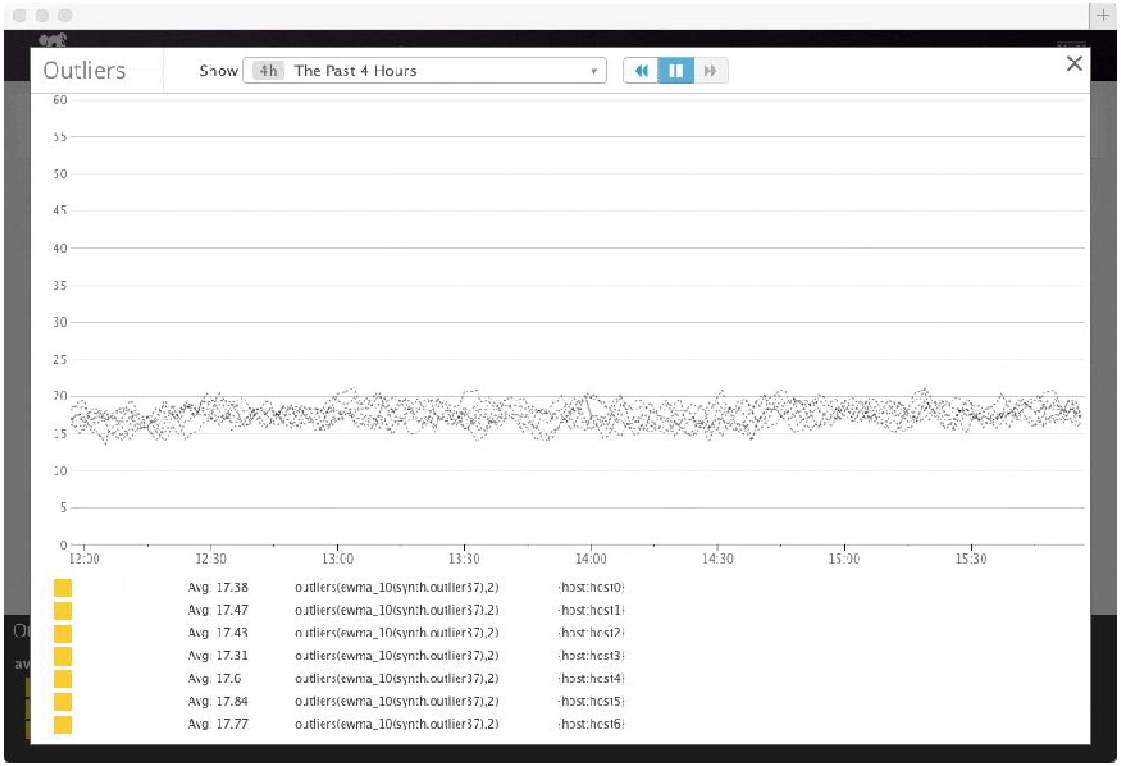
Recap

- Extract as much signal as you can.
- Use robust statistical measures.
- Alert judiciously.
- Don’t over-optimize.

Anomalies or Noise?



Thanks!


Outliers
Show
4h
The Past 4 Hours
60
55
50
45
40
35
30
25
20
15
10
5
0
12:00
12:30
13:00
13:30
14:00
14:30
15:00
15:30
Avg: 17.38 outliers(ewma_10(synth.outlier37),2) host:host10
Avg: 17.47 outliers(ewma_10(synth.outlier37),2) host:host11
Avg: 17.43 outliers(ewma_10(synth.outlier37),2) host:host12
Avg: 17.31 outliers(ewma_10(synth.outlier37),2) host:host13
Avg: 17.6 outliers(ewma_10(synth.outlier37),2) host:host14
Avg: 17.84 outliers(ewma_10(synth.outlier37),2) host:host15
Avg: 17.77 outliers(ewma_10(synth.outlier37),2) host:host16

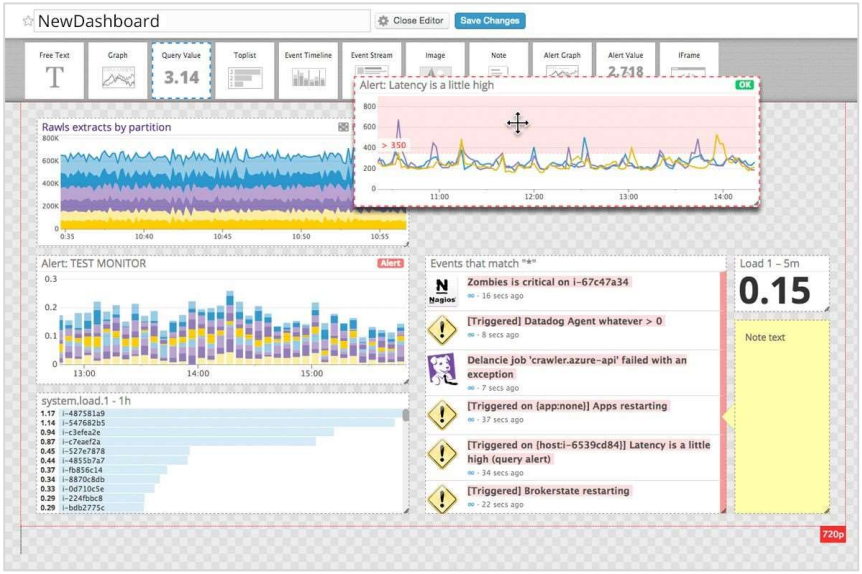
Appendix

See It All In One Place

Your Servers, Your Clouds, Your Metrics, Your Apps, Your team. Together.

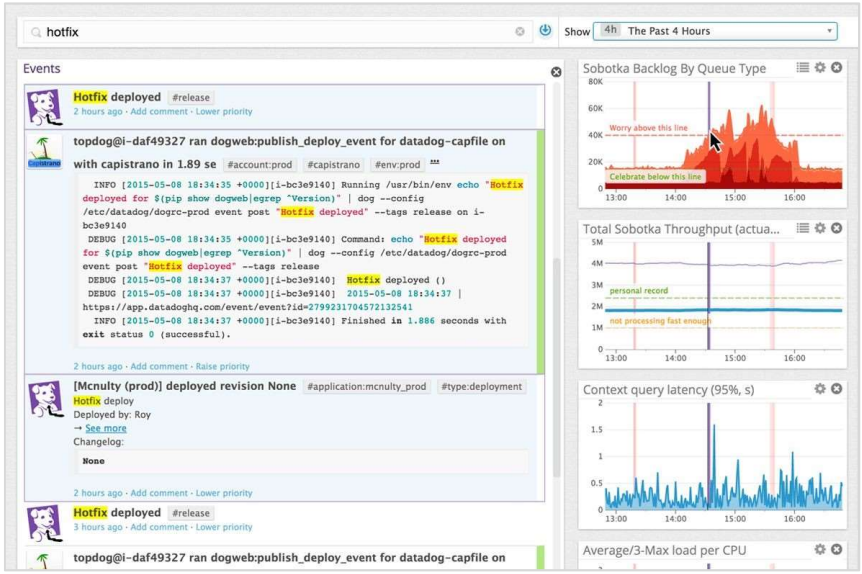
DASHBOARDS

Build Real-Time Interactive Dashboards



CORRELATION

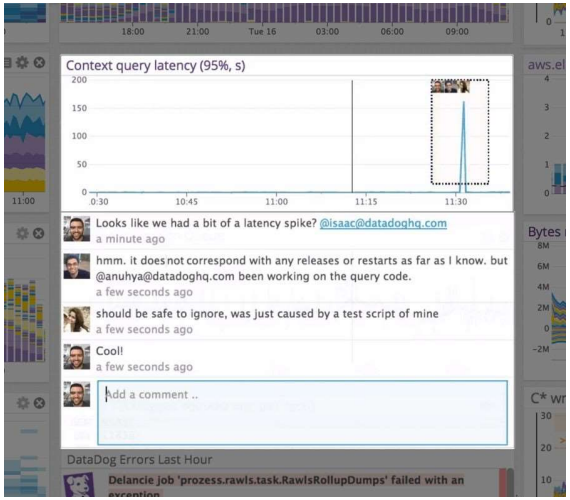
Search And Correlate Metrics And Events



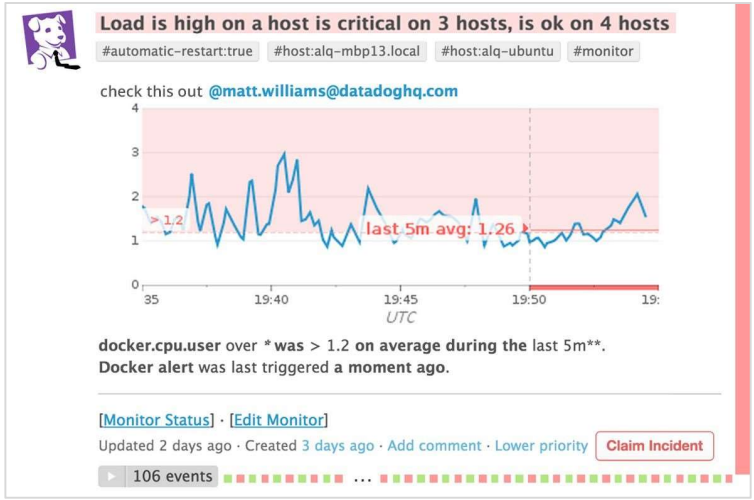
See It All In One Place

Your Servers, Your Clouds, Your Metrics, Your Apps, Your team. Together.

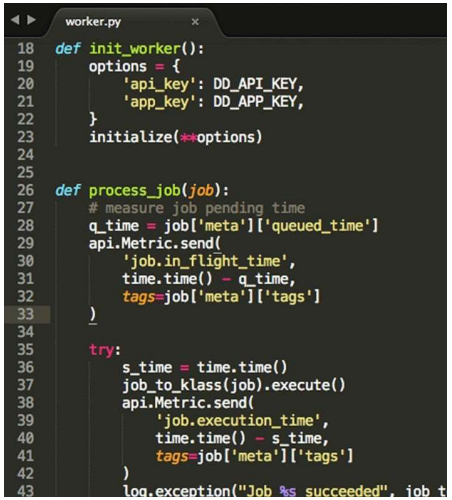
COLLABORATION
Share What You Saw, Write What You Did



METRIC ALERTS
Get Alerted On Critical Issues



DEVELOPER API
Instrument Your Apps,
Write New Integrations



Flexible Pricing

To Match Your Dynamic Infrastructure.

Free	Pro	Enterprise
Up to 5 Hosts	**Up to 500 Hosts**	**500+ Hosts**
	$15 Per Host / Month	Contact us for pricing: +1 866 329 4466 sales@datadoghq.com
1 Day retention	13 Month retention	Customized retention
Custom metrics and events	Custom metrics and events	Custom metrics and events
	Metric alerts*	Metric alerts*
Discussion group supported	Email supported	Email and phone supported