You get what you measure—why metrics are important

Kumar Srinivasamurthy
Bing & Cortana Engineering
@00kumars

Microsoft

Examples

Metric definition, education and reliability of monitors

Los Angeles Times CALIFORNIA & LOCAL L.A. NOW POLITICS BUSINESS NATION ENTERTAINMENT OPINION FOOD MORE

YOU ARE HERE: LAT Home → Collections

FROM THE ARCHIVES

Math Error Exaggerated TriZetto Loss Estimate
March 7, 2001

Math Error Inflated Ventura Blvd. Cost : Traffic: Mistake...
November 18, 1994

Mars Probe Lost Due to Simple Math Error

October 01, 1999 | ROBERT LEE HOTZ | TIMES SCIENCE WRITER

Email Share G+ Tweet Recommend 19

NASA lost its $125-million Mars Climate Orbiter because spacecraft engineers failed to convert from English to metric measurements when exchanging vital data before the craft was launched, space agency officials said Thursday.

A navigation team at the Jet Propulsion Laboratory used the metric system of millimeters and meters in its calculations, while Lockheed Martin Astronautics in Denver, which designed and built the spacecraft, provided crucial acceleration data in the English system of inches, feet and pounds.

https://mars.jpl.nasa.gov/msp98/news/mco990930.html

The New York Times SUBSCRIBE NOW LOG IN

ARCHIVES | 1983

JET'S FUEL RAN OUT AFTER METRIC CONVERSION ERRORS

By RICHARD WITKIN

https://www.nytimes.com/1983/07/30/us/jet-s-fuel-ran-out-after-metric-conversion-errors.html

Availability & Downtime

Availability %	Downtime per year	Downtime per month	Downtime per week	Downtime per day
90% ("one nine")	**36.5 days**	**72 hours**	**16.8 hours**	**2.4 hours**
95%	18.25 days	36 hours	8.4 hours	1.2 hours
97%	10.96 days	21.6 hours	5.04 hours	43.2 minutes
98%	7.30 days	14.4 hours	3.36 hours	28.8 minutes
99% ("two nines")	**3.65 days**	**7.20 hours**	**1.68 hours**	**14.4 minutes**
99.5%	1.83 days	3.60 hours	50.4 minutes	7.2 minutes
99.8%	17.52 hours	86.23 minutes	20.16 minutes	2.88 minutes
99.9% ("three nines")	**8.76 hours**	**43.8 minutes**	**10.1 minutes**	**1.44 minutes**
99.95%	4.38 hours	21.56 minutes	5.04 minutes	43.2 seconds
99.99% ("four nines")	**52.56 minutes**	**4.38 minutes**	**1.01 minutes**	**8.66 seconds**
99.995%	26.28 minutes	2.16 minutes	30.24 seconds	4.32 seconds
99.999% ("five nines")	**5.26 minutes**	**25.9 seconds**	**6.05 seconds**	**864.3 milliseconds**
99.9999% ("six nines")	**31.5 seconds**	**2.59 seconds**	**604.8 milliseconds**	**86.4 milliseconds**
99.99999% ("seven nines")	**3.15 seconds**	**262.97 milliseconds**	**60.48 milliseconds**	**8.64 milliseconds**
99.999999% ("eight nines")	**315.569 milliseconds**	**26.297 milliseconds**	**6.048 milliseconds**	**0.864 milliseconds**
99.9999999% ("nine nines")	**31.5569 milliseconds**	**2.6297 milliseconds**	**0.6048 milliseconds**	**0.0864 milliseconds**

https://en.wikipedia.org/wiki/High_availability

Simplified formula:

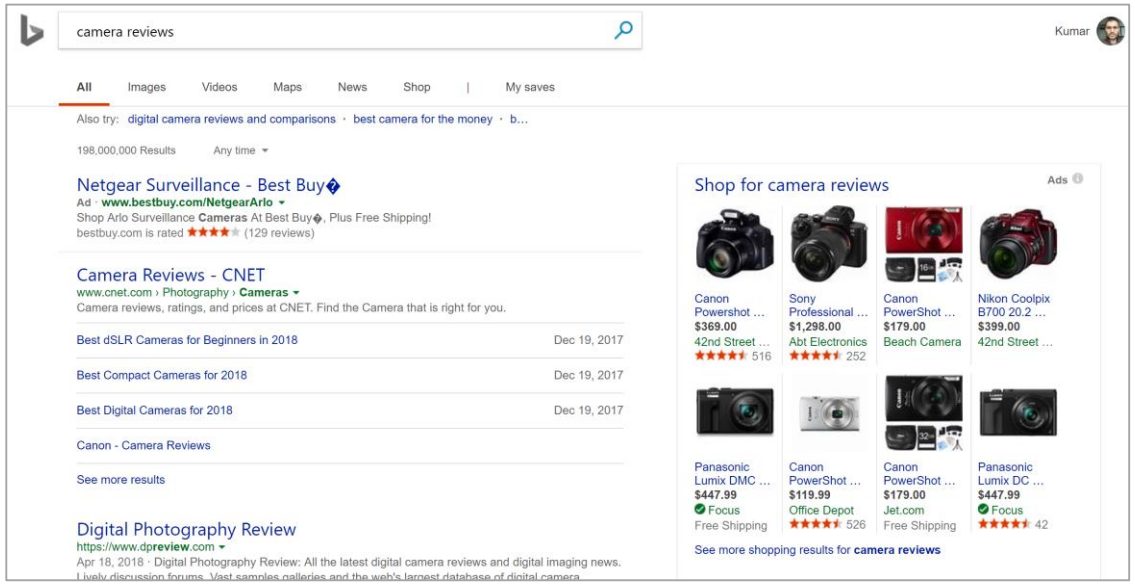
$$Availability = \frac{(Total\ Requests - Failed\ Requests)}{Total\ Requests}\ \%$$

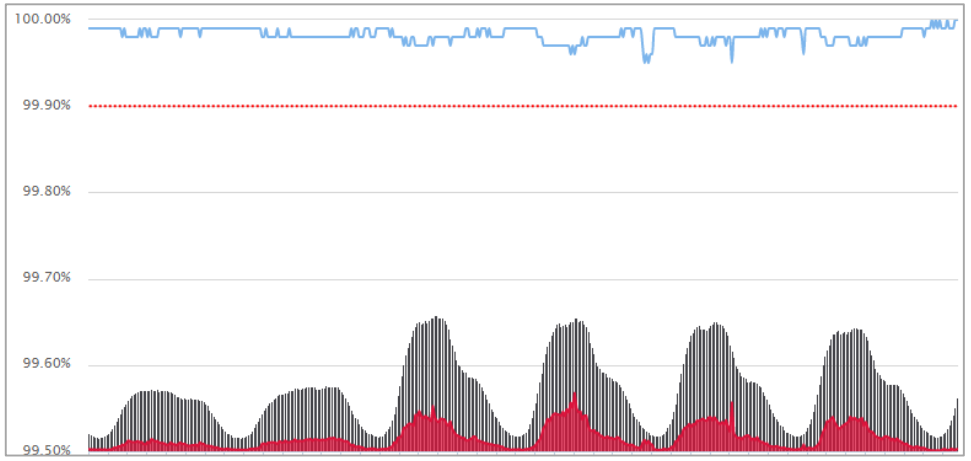
$$A = \frac{T-F}{T}\ \%$$

Microsoft

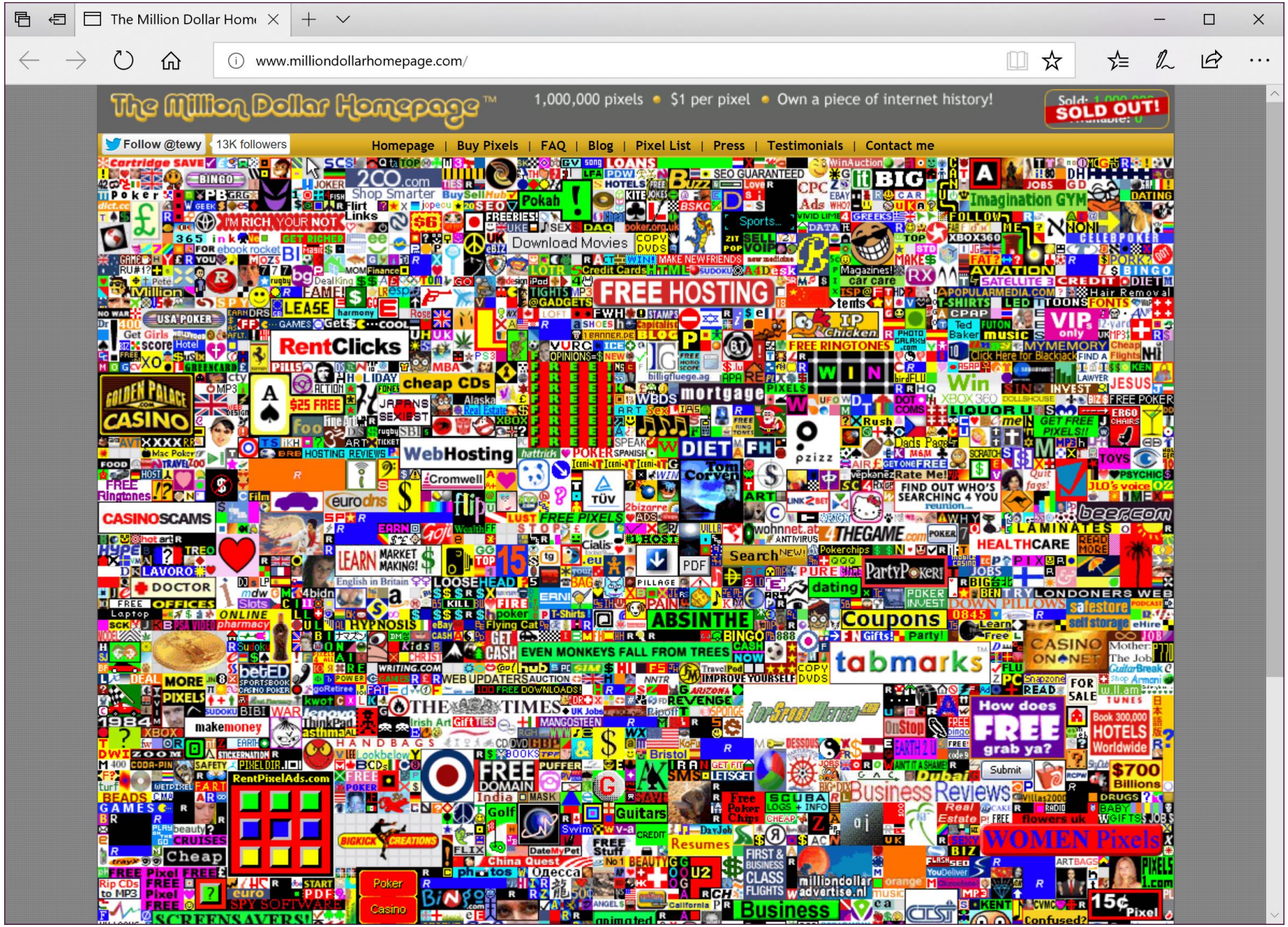
Hitting the target and missing the point

$$A = \frac{T - F}{T}\%$$





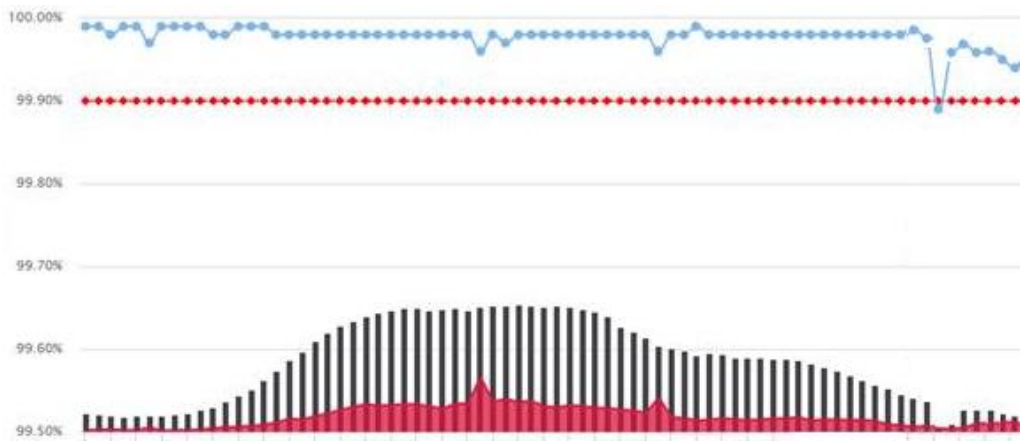
- Bad config led to no web results to the user, just ads
- Remember that one metric may not tell the whole story
- Opposing Metrics

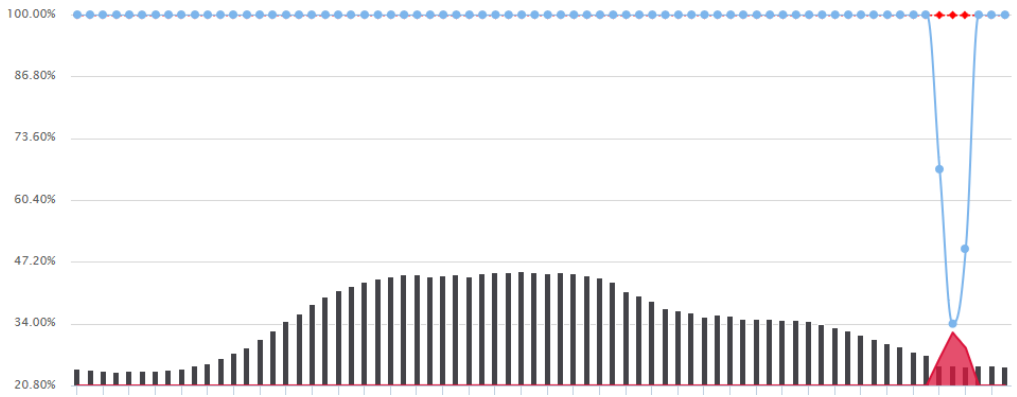


http://www.milliondollarhomepage.com/

How and where you measure matters

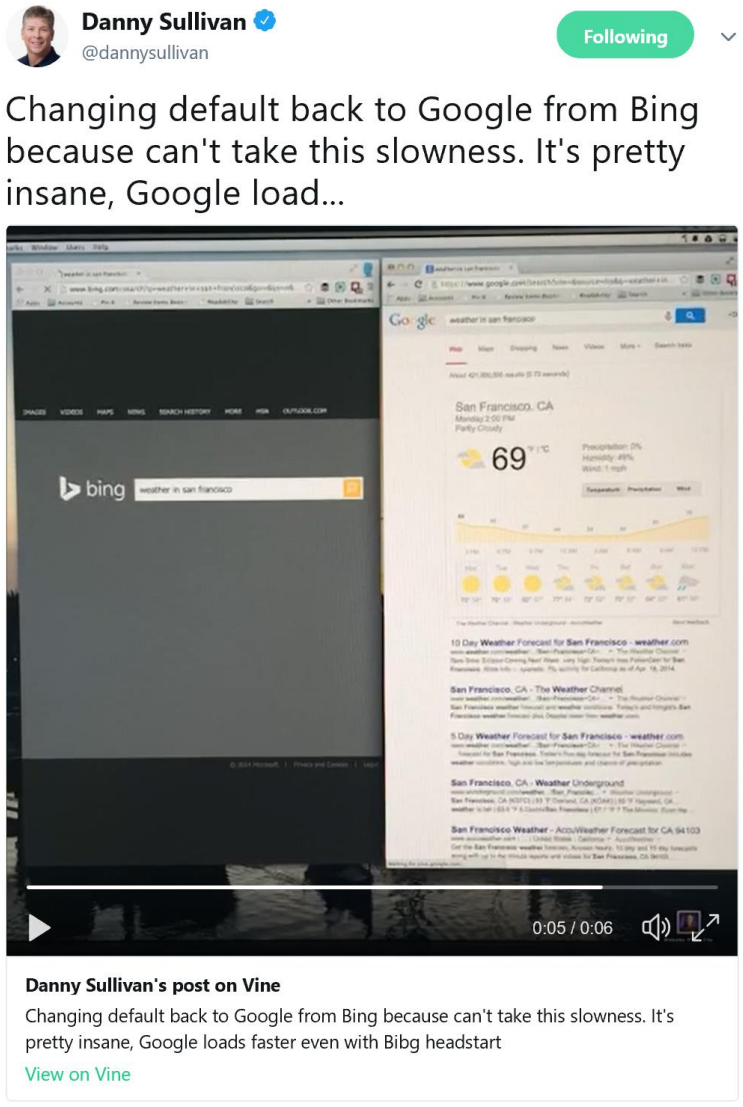
$$A = \frac{T - F}{T}\,\%$$





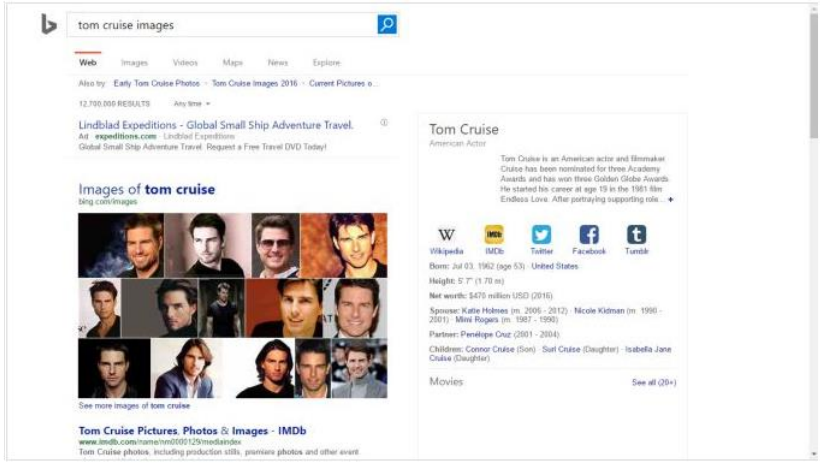
- Traffic black-holing - Drop at CDN/Edge
- High availability at service level, but drop in incoming volume
- *Anomaly detection* can identify changes in volume
- Outside-in monitoring should supplement internal metrics
- Incorrect sample sizes can also affect your metric

Gaps in measurement

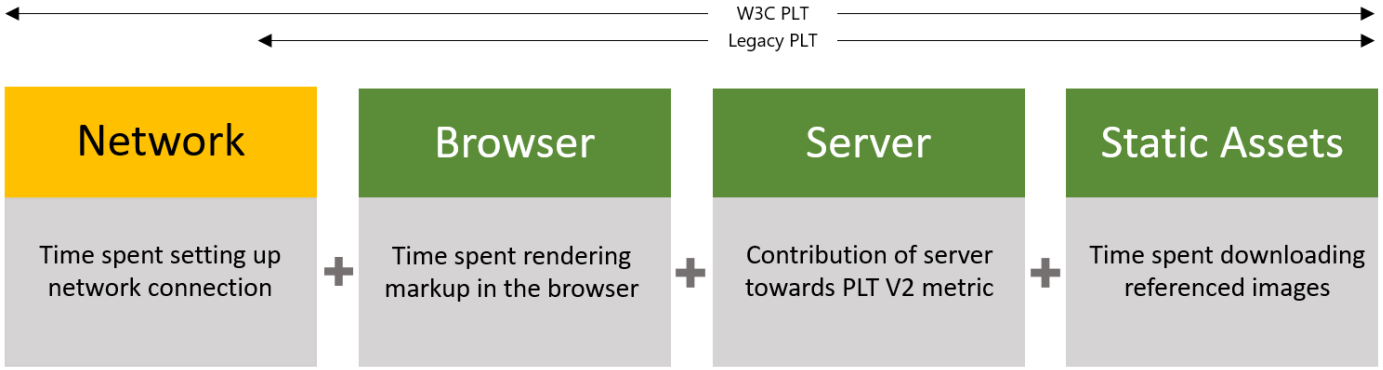


- Terrible slowness for a small segment of users
- PLT tracked at 75th or 95th didn't show issues
- Investigation and findings led to additional dimensions

Performance measurement journey

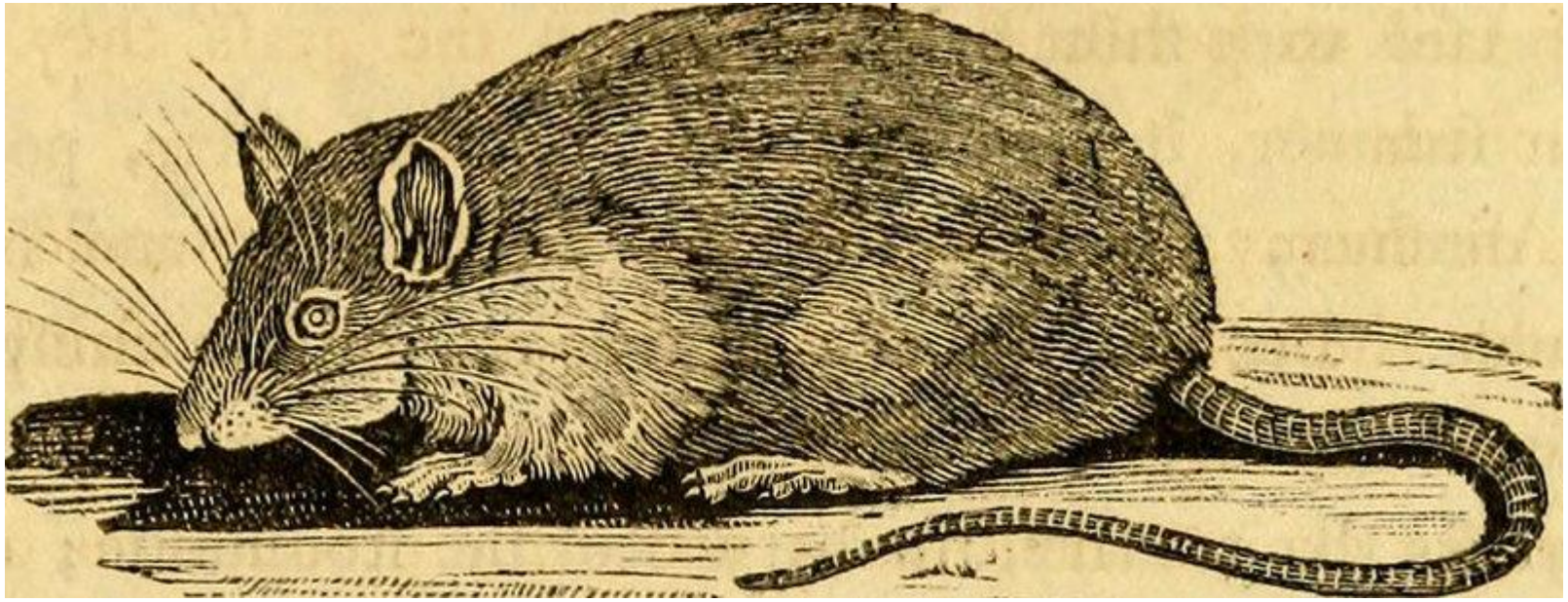


- Periodic update of metrics
- Above fold vs below fold



Microsoft

The great Hanoi rat massacre



https://flic.kr/p/ovusTD

More: http://freakonomics.com/media/vannrathunt.pdf
https://www.atlasobscura.com/articles/hanoi-rat-massacre-1902

Unintended consequences - Window tax

https://en.wikipedia.org/wiki/Window_tax#/media/File:Window_Tax.jpg

https://en.wikipedia.org/wiki/Window_tax#/media/File:Windows_in_Brighton_Street,_Edinburgh.jpg

Closing thoughts

- High data quality
  - Correctness, completeness and cleanliness
- Periodic metrics update
  - Metrics do get “gamed”
- Wrong incentives – unintended consequences
- Metrics overload
- Percentiles can hide problems
- Some metrics do flat-line. Important to let them not regress

Questions?

- kumars@microsoft.com
- @00kumars