




SILICON VALLEY
DATA SCIENCE

FOR THE INTERNET OF THINGS, THE MORE THINGS THE MERRIER

Data Science Pop-up in Austin

John Akred, CTO

13 April 2016



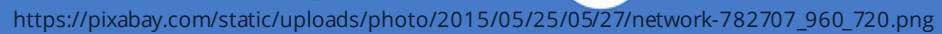
Silicon Valley Data Science transforms
business through data science and
engineering.



OUR TEAM



4



The Internet of Things (IoT) is the **network of physical objects**—devices, vehicles, buildings and other items—**embedded with**

- **electronics,**
- **software,**
- **sensors, and**
- **network connectivity**

that enables these objects to collect and exchange data.

Source:

https://en.wikipedia.org/wiki/Internet_of_Things
"[Internet of Things Global Standards Initiative](#)". ITU. Retrieved 26 June 2015.

WHAT MAKES A THING AN IOT THING?



ER docs get heart rate info from Fitbit, save patient's life

Doctors provided strategic treatment after the man suffered a seizure.

by Valentina Palladino - Apr 8, 2016 7:16am PDT

Share

Tweet

Email

62

A connected thing
can do useful stuff
for us, sometimes
really useful.

<http://arstechnica.com/science/2016/04/er-docs-get-heart-rate-info-from-fitbit-save-patients-life/>



ALFRED SACCHETTI, MD



An ecosystem of connected things can do useful stuff for us too.

<https://jawbone.com/blog/napa-earthquake-effect-on-sleep/>

8.25.14

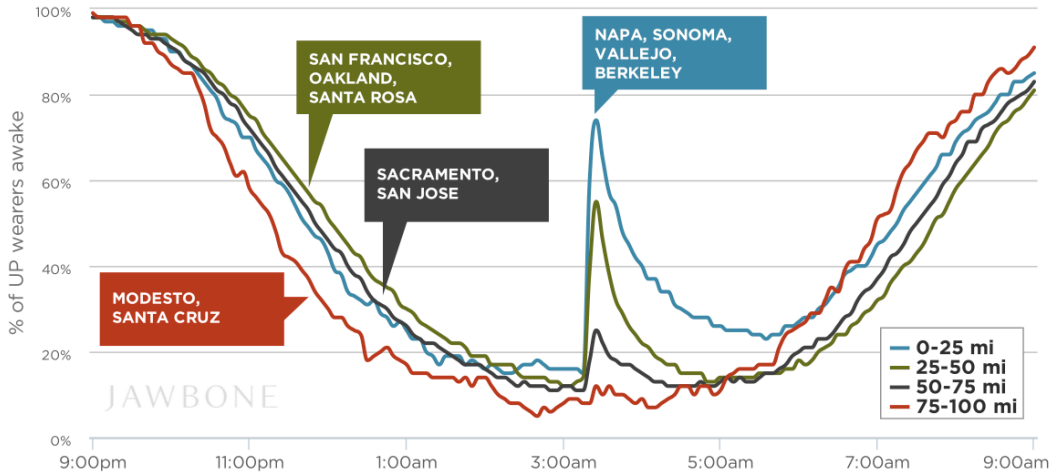
EUGENE MANDEL

DATA, NEWS

+6737



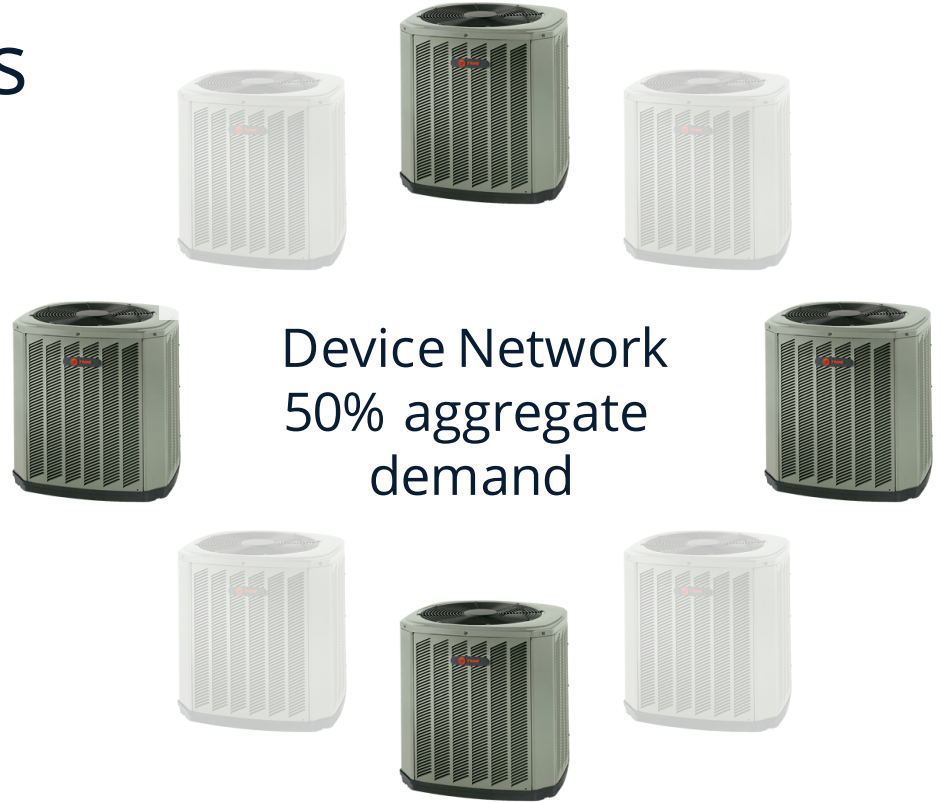
The [South Napa Earthquake](#) was the strongest to hit Northern California in 25 years. Our data science team wanted to quantify its effect on sleep by looking at the data recorded by [Jawbone UP](#) wearers in the Bay Area who track their sleep patterns.



Value of Connected Devices



Single Device
\$100s/year



Why more = better? DATA RESILIENCY

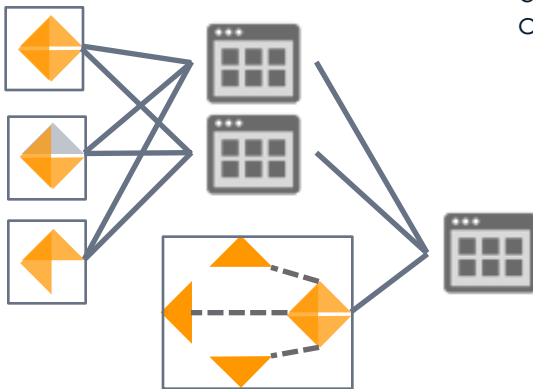
Stovepipe: One-to-one relationship from data source to product.



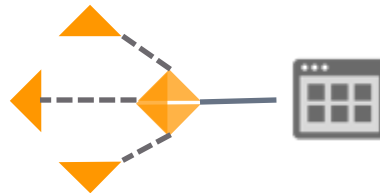
Hard Failure: If the data source is broken, so is the app.



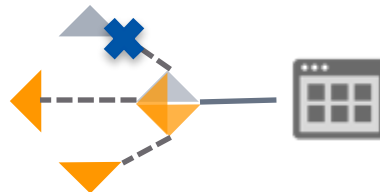
Production data services abstract the probabilistic integration of overlapping data sources. We call this model a **Data Mesh**.



Multi-sourced: Redundancy of overlapping data sources makes your products more resilient.



Graceful Degradation: If a data source breaks, there is a backup and your app continues to function.



Caltrain Rider

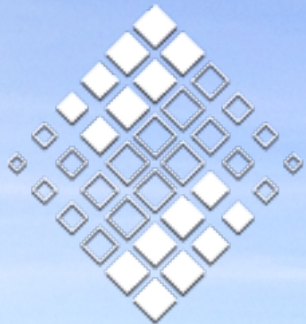


START		END	
Burlingame		San Francisco	
⇌	Local	Limited	Bullet
	\$5.25 one way (\$.50 less with Clipper, \$10.50 day pass)		My Commute
Time to train	Scheduled times		Riding time
Riding this train	3:38PM - 4:04PM #257 - Limited		26
Riding this train	4:07PM - 4:40PM #159 - Local		33
40	4:53PM - 5:31PM #263 - Limited		38
1:04	5:17PM - 5:43PM		26



- Commuter rail between San Francisco and San Mateo and Santa Clara counties ~30 stations
- 118 passenger cars
- 60% ≥ 30 years old
- 2014 weekday ridership is 52,019 people daily
- On-time performance is about 92%
- **No reliable real-time status information**
- **API outage between April 5th and June 2nd**





HOW DO WE KNOW IF THE TRAIN IS ON TIME?



home | site map | bids/contracts | jobs | contact Search



Schedules

Stations

Fares

Rider Information

Projects & Plans

About Caltrain

Friday, Apr. 16: Caltrain will run a Bus Bridge between Hayward Park, San Mateo and Burlingame Stations. Go to caltrain.com/SanMateoBusBridge for more info.

News

Construction, Installation and Maintenance:
April 9 to April 15, 2016

April 8, 2016

Open Letter from CEO Jim Hartnett

Caltrain Executive Director Testifies On Draft High-Speed Rail Business Plan

April 6, 2016

San Mateo's Poplar Avenue Rail Bridge to be Replaced; Bus Bridge Provided

April 6, 2016

Caltrain Ready for Return of Regular Season Baseball at AT&T Park

April 5, 2016

IT CAN TAKE
UP TO ONE MILE
TO STOP A TRAIN. See Tracks? Think Train!



See Tracks? Think Train!
caltrain.com/safety
#CaltrainSafe



Real-time Departures

as of 5:33 AM

Choose a station and click Select

California Ave

Select

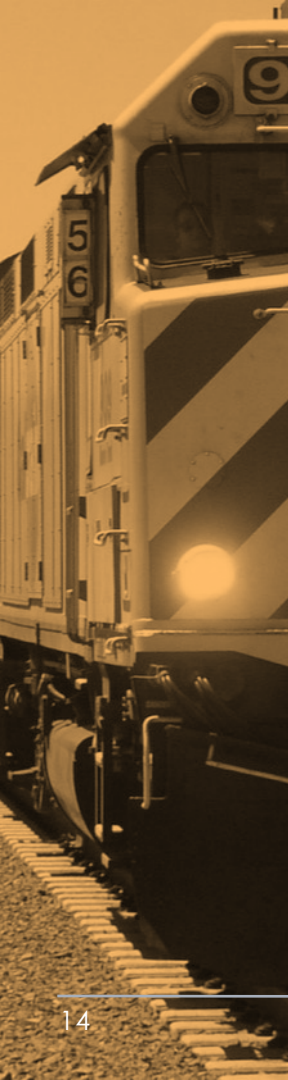
Highlights

Weekday Timetable
Weekend Timetable
Board of Directors
Agendas & Minutes

Mobile Schedules
Title VI
Advisory Committees
CAC, CEMOF, BAC

SUBSCRIBE to E-mail notifications!

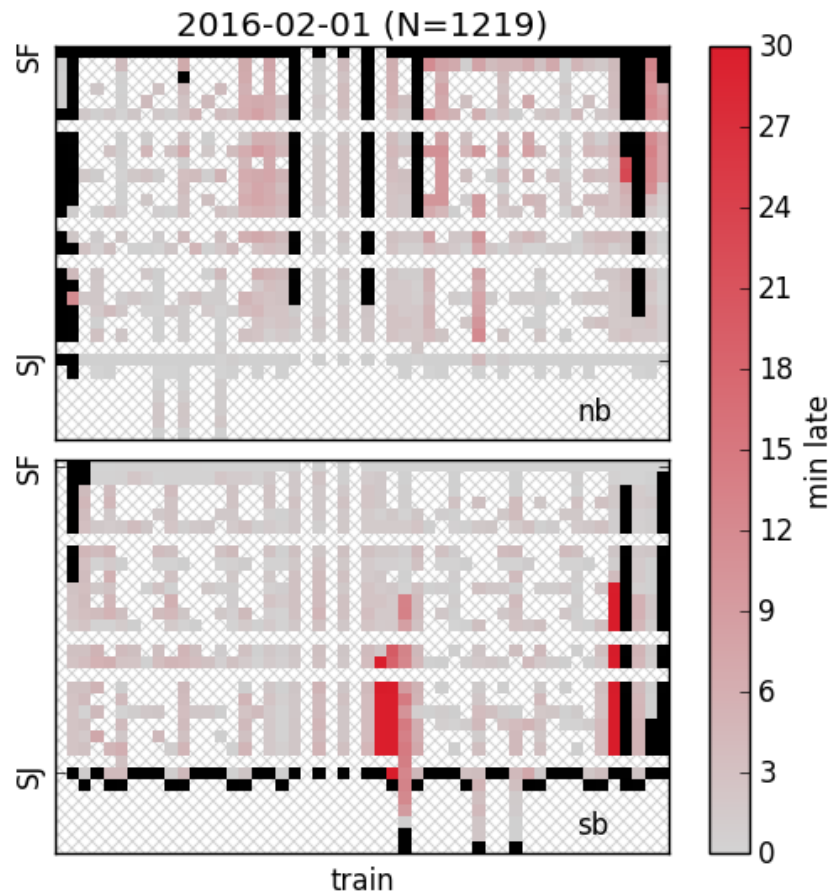




MISSING DATA?

The departure data:

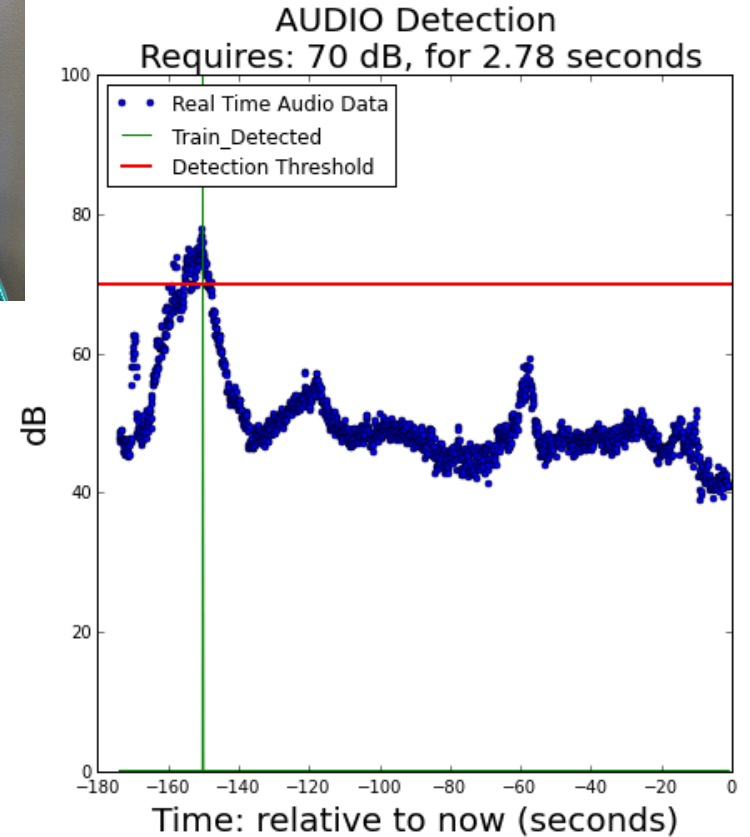
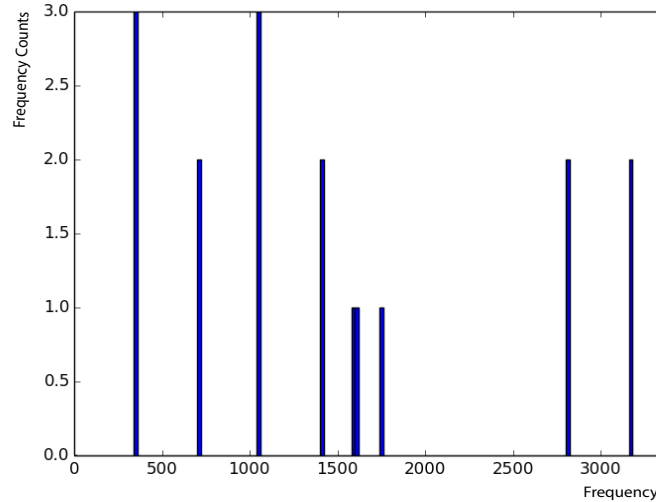
- mostly “on time”
- when significant delays occur, they often appear suddenly
- often missing >10% per day
- ~1 minute resolution



AUDIO



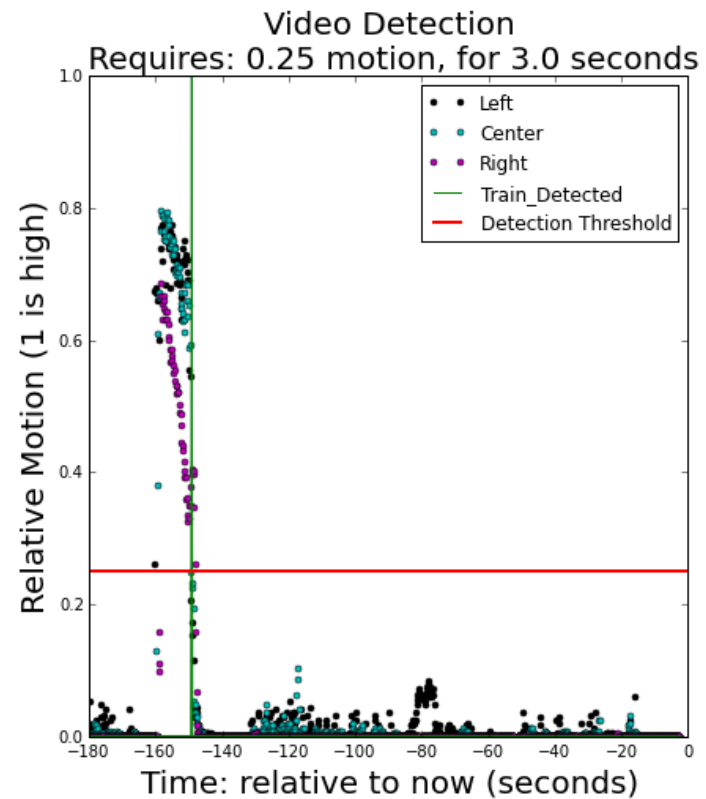
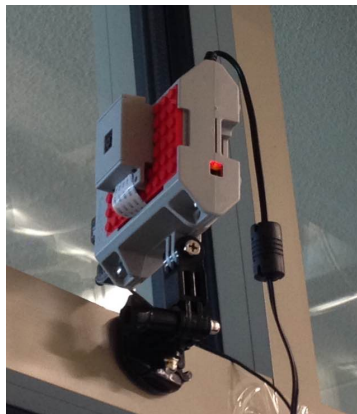
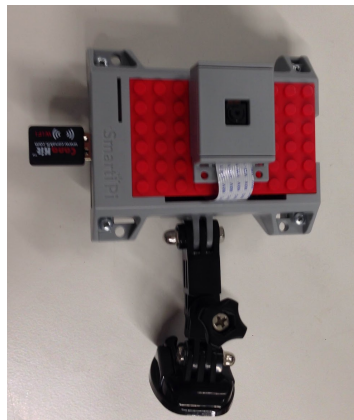
Histogram of Whistle Frequencies Over a Period of Time



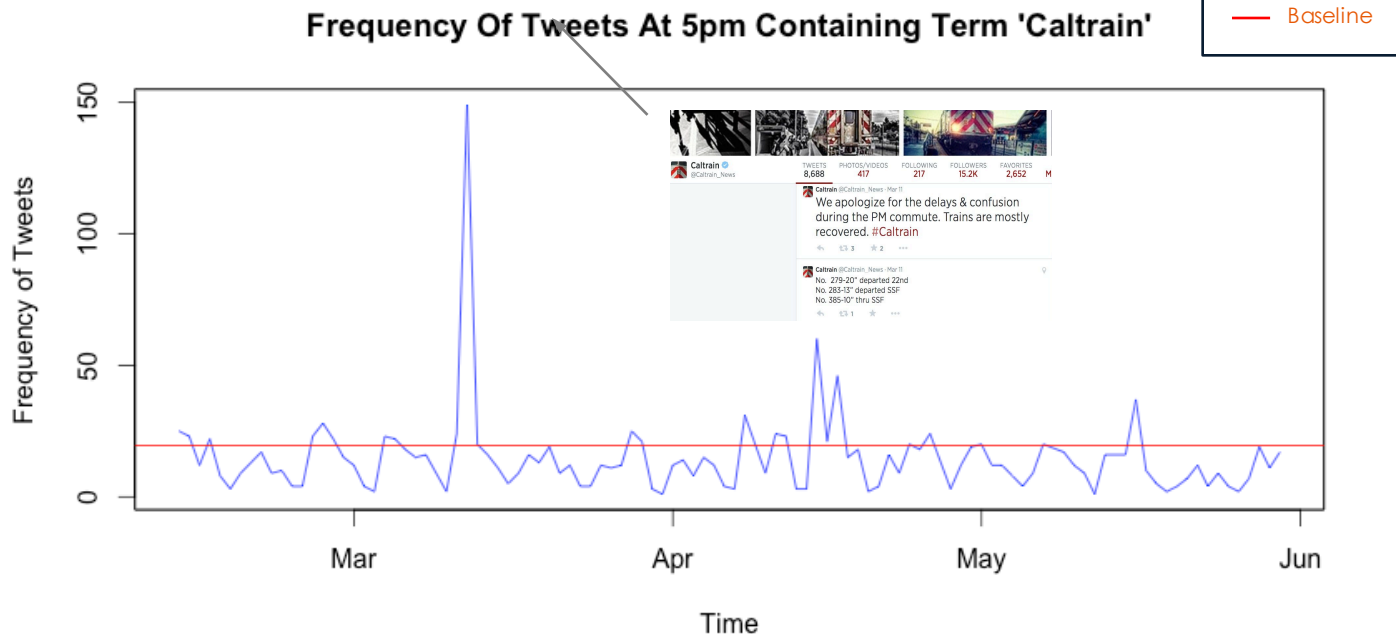
VIDEO DETECTION - FRAME ANALYSIS



VIDEO

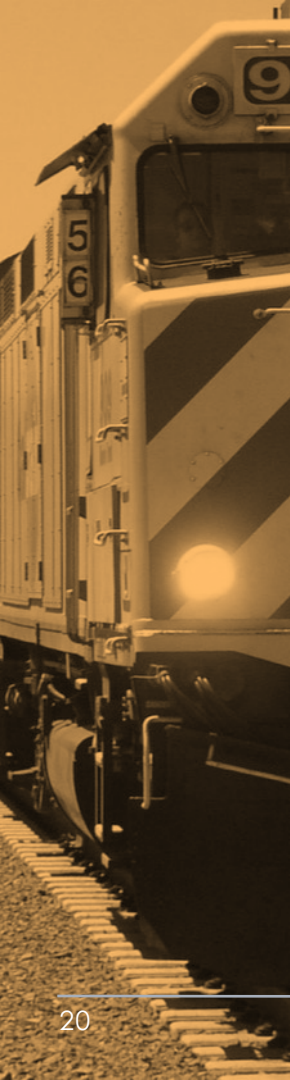


SOCIAL MEDIA AS SENSOR

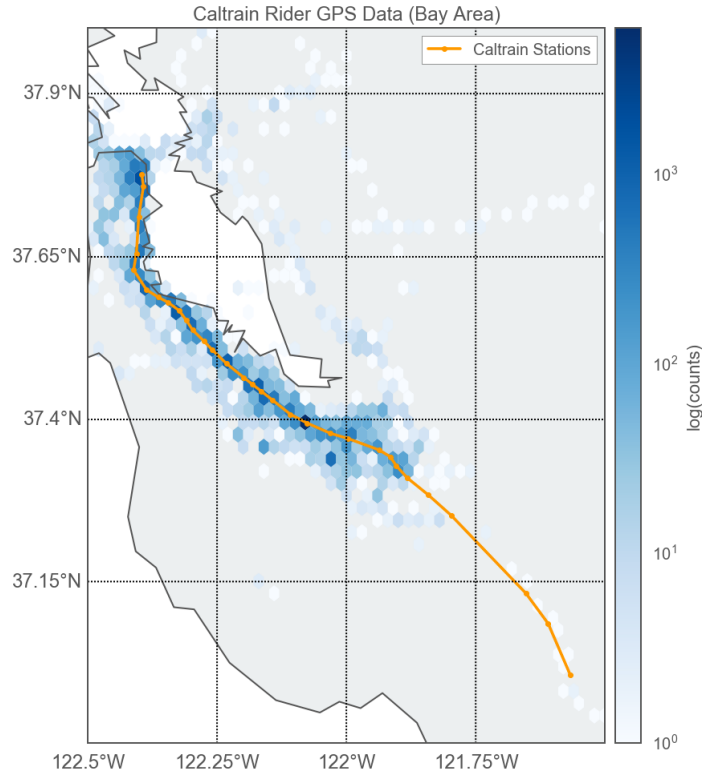


A close-up, low-angle shot of a train engine on tracks. The engine is white with black and grey diagonal stripes. A large, glowing headlight is visible. The number '56' is printed on the side of the engine. The tracks are made of gravel and metal rails. The background is a warm, orange-yellow sky.





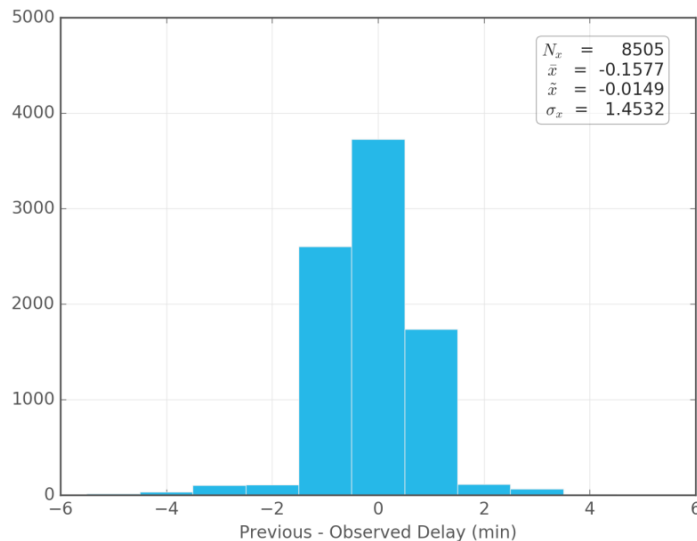
THE MOST IMPORTANT SIGNAL



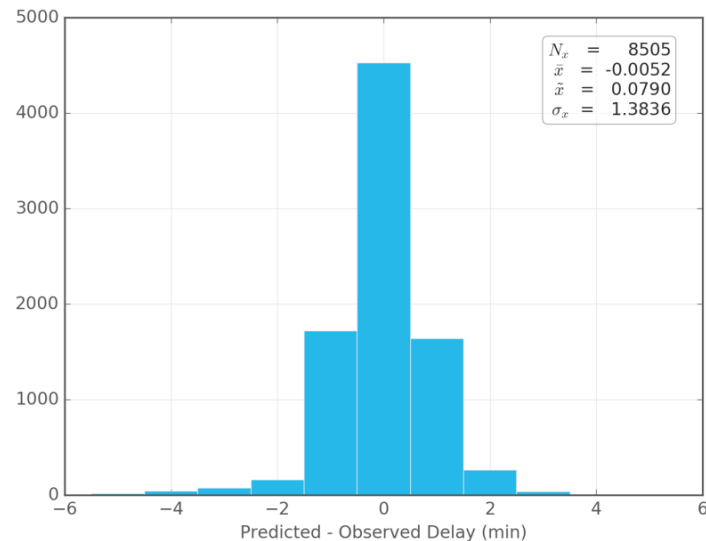
Where is
the train
right
now?

PREDICTION MODEL: NEURAL NETWORK

previous delay
residuals



neural network
prediction residuals





SILICON VALLEY
DATA SCIENCE

Yes, we're hiring!
info@svds.com

THANK YOU

John Akred
[@BigDataAnalysis](#)