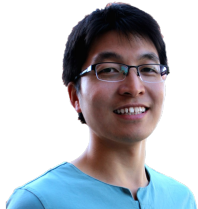


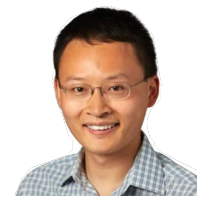
AAAI 2023 Tutorial: Economics of Data and ML



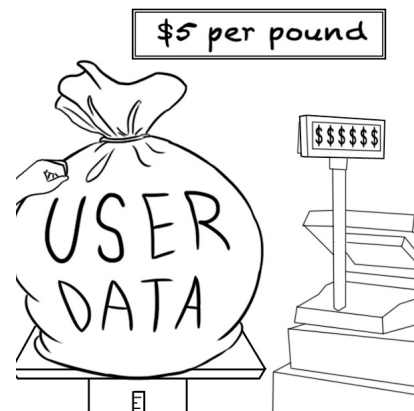
Haifeng Xu (Chicago)



Shuran Zheng (CMU)

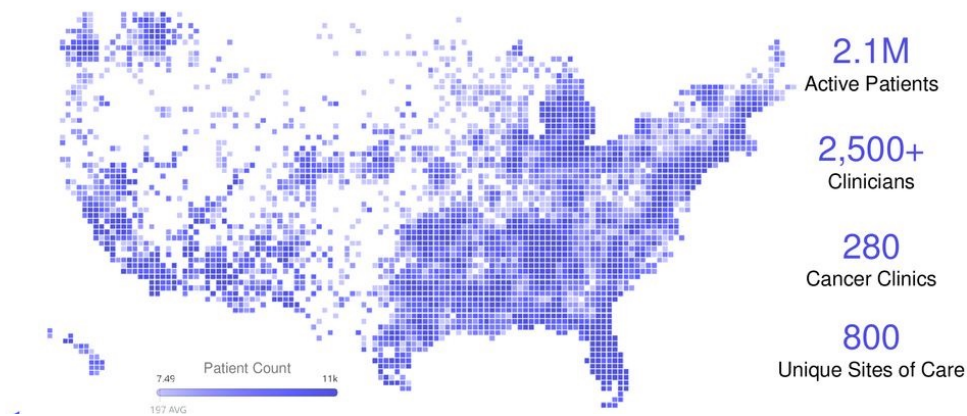


James Zou (Stanford)



2/8/2023

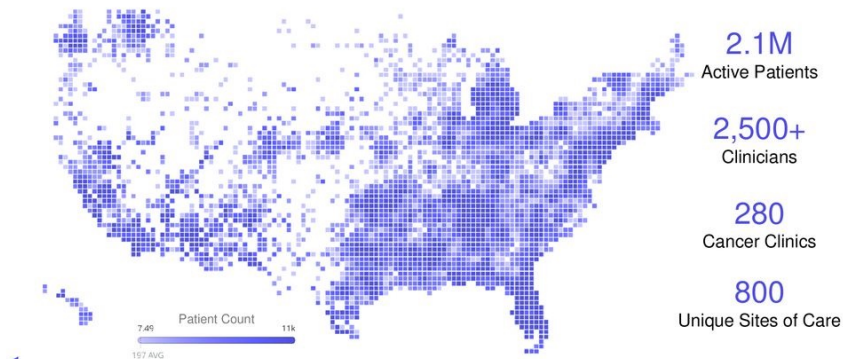
How much would you pay for this data and information?



Proprietary access to medical data of 2M cancer patients

If you are a health AI company, how much would you pay for this data?

How much would you pay for this data and information?



FORTUNE

RANKINGS MAGAZINE NEWSLETTERS VIDEO PODCASTS CONFERENCES COVID-19

COMMENTARY

Why police violence against Black people persists—
and what can be done about it

FINANCE

'Second stimulus check calculator' tells you what
you'd receive if the HEROES Act passes

FINANCE

It's official: This was the best quarter for stocks since
1998

HEALTH • FLATIRON HEALTH

Why Drug Giant Roche's \$1.9 Billion Deal to Buy Data Startup Flatiron Health Matters

BY SY MUKHERJEE

February 16, 2018 10:22 AM PST

Factors impact the value of data

1. How hard is it to create this data
2. Data quality
3. How rich is the data (diversity, longitudinal)
4. How much does this data improve my model/analytics

Data sellers and markets

The screenshot displays the AWS Marketplace interface. At the top, there is a navigation bar with the AWS Marketplace logo, a search bar, and a user greeting "Hello". Below the navigation bar, there are several menu items: "Categories", "Delivery Methods", "Solutions", "Migration Mapping Assistant", "Your Saved List", "Partners", "Sell in AWS Marketplace", and "Amazon Web Services".

The main content area is divided into a left sidebar and a main product display area. The sidebar contains the following sections:

- All Categories**: Data Products (12)
- Filters**
- Vendors**:
 - TruFactor (4)
 - NinthDecimal (3)
 - X-Mode (2)
 - Epsilon (2)
 - Gravy Analytics (1)
- Pricing Plan**:
 - Monthly (10)
 - Free (7)
 - Annual (3)
- Delivery Method**:
 - AWS Data Exchange (12)

The main product display area shows two data products:

- US Mobile Location Visitation Data - The Home Depot**: Sold by Gravy Analytics. Price \$10,000 | 1 month subscription available. This product includes anonymous visitation data in the US to The Home Depot, starting with Feb 2020. More data can be included on a historical and go-forward basis. Explore mobile location visitation data. Refine your understanding and use cases based on the available data and insights you can...
- Home Improvement (US Only)**: Sold by X-Mode. Price \$240,000 | 12 month subscription available. home improvement, furniture, hardware, garden, couch, bed, home depot, home goods, ikea

>3000 data products from free to \$millions/year

Data sellers and markets



Market Data

Regulated – products filed with SEC

i.e. ITCH, FPGA, SoupBinTCP

Priced by user type

\$380M



Index Data

Less Regulated

i.e. Nasdaq 100 (and QQQ)

Priced by Assets Under Management

\$230M



Investment & Analytics Data

Unregulated

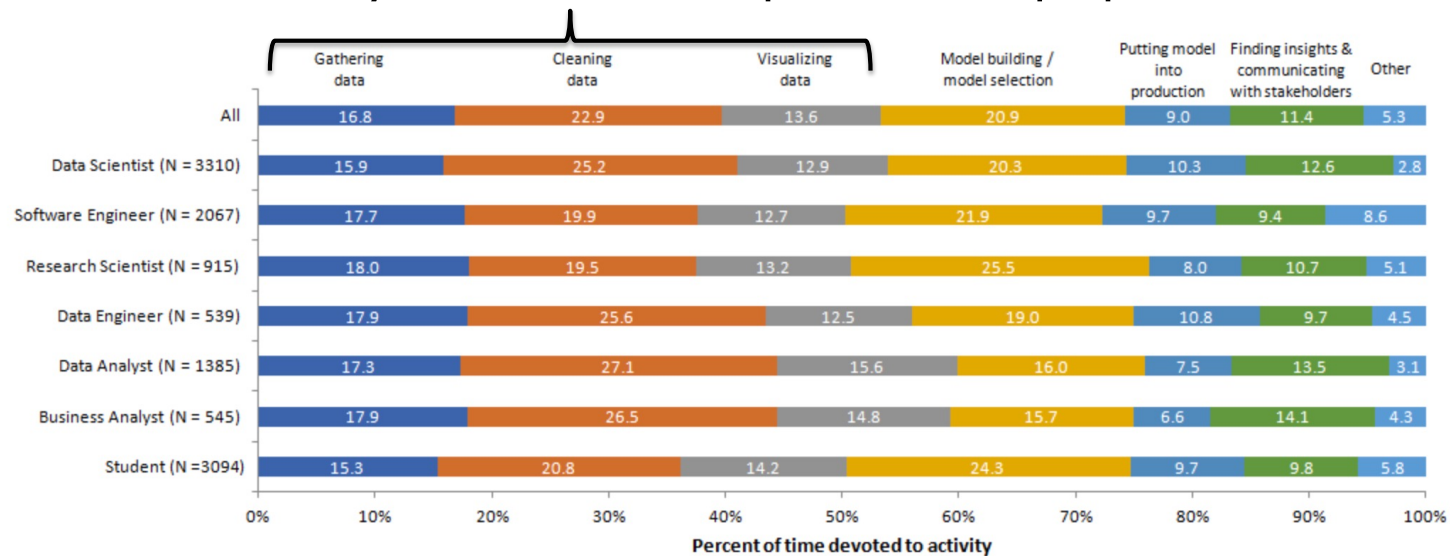
i.e. Quandl Alternative Data

Priced by value/experiment

\$150M

Getting the right data is the biggest challenge for ML

ML Survey: >50% of time is spent on data prep



What data to collect?

Which data sources are high quality?

How to audit, clean and filter data?



Tutorial outline: economics of data and ML

Part I: Data buyer's perspective.

- What data is the most useful? Statistical data valuation
- How to quantify the value of information.

Short break

Part II: Data seller's perspective.

- How to price information.
- How to collect truthful data.

Short break

Part III: economics of ML

- Market for ML-as-a-service

Learning objectives

After this tutorial, you will:

- Understand how the **value of data and information** is modeled in CS, statistics, and economics.
- Gain practical **tools for assessing value of different types of data** for your application.
- Appreciate where research is heading and **open problems**.