



* Part of this slide is modified from a slide of Prof.Natawut



Introduction to Data Science

2110446 Data Science and Data Engineering

Assoc. Prof. Peerapon Vateekul, Ph.D.
Department of Computer Engineering,
Faculty of Engineering, Chulalongkorn University

Peerapon.v@chula.ac.th

www.cp.eng.chula.ac.th/~peerapon/



Outline

- Introduction
 - Data is important
 - Data Science Definition by Dr.Virote
 - Data Science Definition by Aj.Natawut
- Big Data
- Data Science Process & Data Science Trend





Introduction



Data is important (in 2017)!

The Economist Topics ▾ Current edition More ▾

Regulating the internet giants

The world's most valuable resource is no longer oil, but data

The data economy demands a new approach to antitrust rules



David Parkins

Print edition | Leaders >
May 6th 2017

[Twitter](#) [Facebook](#) [LinkedIn](#) [Email](#) [Print](#)

- Alphabet (Google's parent company), Amazon, Apple, Facebook and Microsoft
- \$25bn in net profit in the first quarter of 2017
- Amazon captures half of all dollars spent online in America.
- Google and Facebook accounted for almost all the revenue growth in digital advertising in America last year



Data is important (in 2018)! (cont.)

The New Oil

Jennifer Presley Executive Editor, E&P Magazine Hart Energy Thursday, November 1, 2018 - 6:40am



With a number of successful projects under its collective belt, the oil and gas industry is proving Big Data is more than just a buzzword. (Source: Makhnach_S/Shutterstock.com; Design by Felicia Hammons)

<https://www.epmag.com/new-oil-1720651>

Data Science
(AI,ML,DM)
+
Big Data



Who analyzes these data!

Harvard
Business
Review



ARTWORK: TAMAR COHEN, ANDREW J BUBOLTZ, 2011, SILK SCREEN
ON A PAGE FROM A HIGH SCHOOL YEARBOOK, 8.5" X 12"

DATA

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

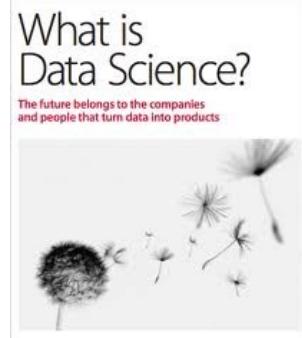
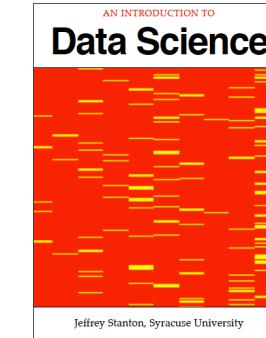
WHAT TO READ NEXT



Competing on Analytics



What is Data Science?



■ Data

- Facts and statistics collected for reference or analysis



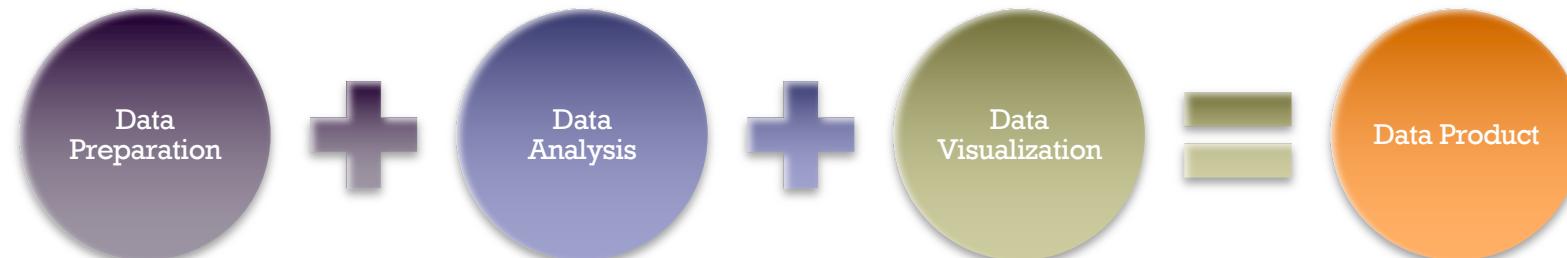
■ Science

- A systematic study through observation and experiment



■ Data Science

- The scientific exploration of data to extract meaning or insight,
- and the construction of software to utilize such insight in a business context.





What is Data Science? (cont.)

1. Transform data into **valuable insights**
2. Transform data into **data products**
3. Transform data into **interesting stories**



Ta Virot Chiraphadhanakul
Data Scientist, Facebook

Code Mania 2 (01), Jan-2015



1) Transform data into valuable insights

How Obama Raised \$60 Million by Running a Simple Experiment



By Dan Siroker
November 29, 2010

A/B Testing Case Studies, Multivariate Testing, Political Campaign

The image displays two versions of the Obama '08 campaign website's landing page. Both versions feature the same top navigation bar with the "OBAMA '08" logo and a "GET INVOLVED" section featuring a photo of Barack Obama at a campaign rally.

The left version shows a "SIGN UP" button positioned below the sign-up form fields (Email Address and Zip Code). An arrow points from the word "Button" to this red button.

The right version shows a "LEARN MORE" button instead of a "SIGN UP" button. An arrow points from the word "Button" to this red button.

Both versions include a "JOIN THE MOVEMENT" call-to-action, sign-up fields for Email Address and Zip Code, and a "CONTINUE TO WEBSITE" link at the bottom. The overall design is blue-themed with white text.



1) Transform data into valuable insights (cont.)



BUSINESS

Amazon introduces next major job killer to face Americans

By James Covert, Linda Massarella and Bruce Golding

December 5, 2016 | 9:59pm | Updated



The Amazon Go storefront
Amazon

Ta Virot Chiraphadhanakul
Data Scientist, Facebook
Code Mania 2, Jan-2015



Amazon's new supermarket will blow your mind — and cost America jobs
Why employment may be optional in the near future
Amazon's futuristic grocery store could spell employment doom
Amazon introduces next major job killer to face Americans
Amazon's latest idea could kill off jobs forever



2) Transform data into data products

CHASE

Action required: Please confirm activity.

FRAUD PROTECTION SERVICES

Chase Sapphire
Account Ending: XXXX

We want to help keep your account secure so we continuously monitor it for possible fraudulent activity. We're writing to verify whether the transaction below was authorized by you or another Cardmember. Click **YES** below if you

YES **NO**

This email is from Chase. Chase.com

Outlook Search

New message Empty folder Mark all as read Undo

Favorites Junk Email Filter

Items in Junk Email will be deleted after 30 days

Today

Inbox 45

Junk Email 128

Drafts

Sent Items

Deleted Items

Archive

Conversation Hist...

Notes

Work At Home Opportunities
New work from home pro... 1:47 PM
ATTENTION: HOME WORKERS ARE NEEDED ...

Client service
NETFLIX SURVEY 1:40 PM
Congratulations pv_shopper! Stand a chance ...

Thank You Costco
Re: Costco Has a Surprise Fo... 12:01 PM
Congratulations! You can get a \$100 Costco ...

Client service
- Are you a friend of Amazo... 8:43 AM
amazon 2019 Shopping Survey Congratulati...

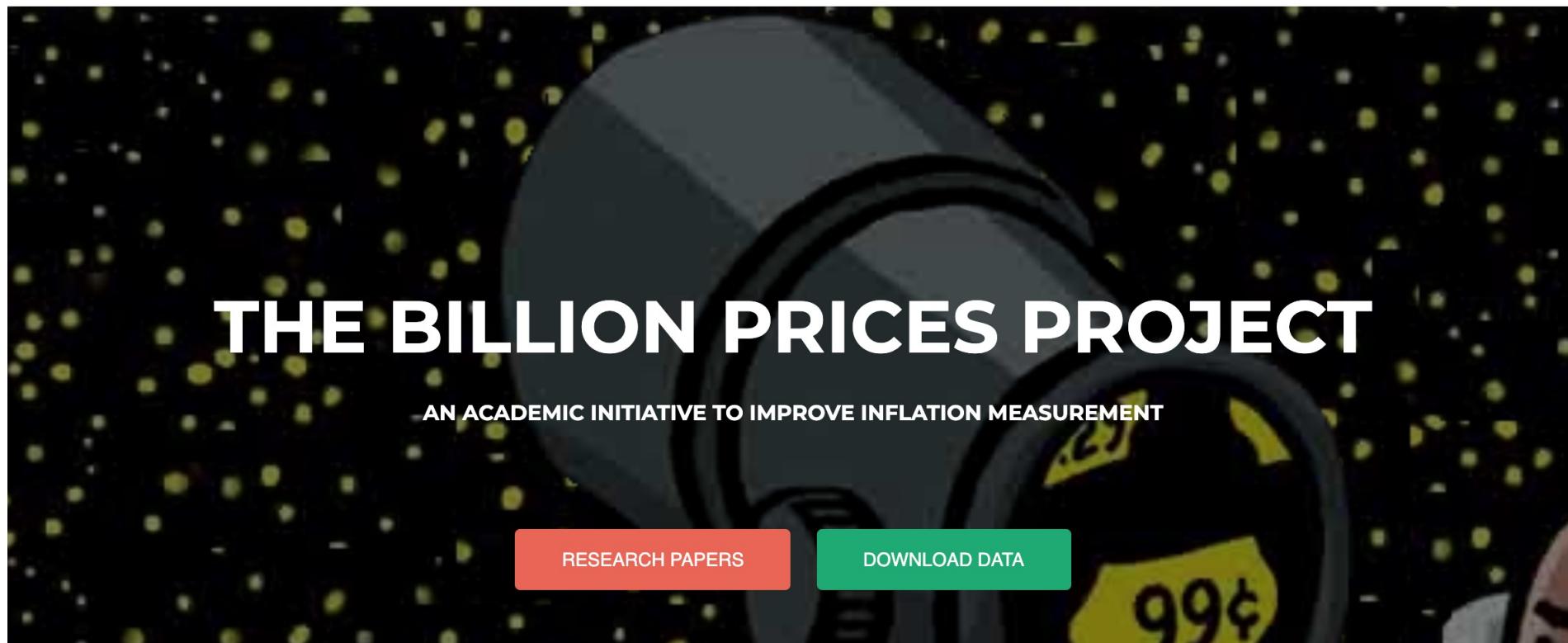


3) Transform data into interesting stories

Consumer Price Index (CPI) - Inflation

The Billion Prices Project

Home Our Public Data Our Research News



<http://www.thebillionpricesproject.com/>

mini Big C 22-28 Feb 15

แลกซื้อ สุดคุ้ม

เมื่อซื้อสินค้าที่บิ๊กซี ครบ 79 บาทขึ้นไป แลกซื้อสินค้าต่อไปได้ตามเงื่อนไขที่ระบุไว้ด้านใน

Product Examples:
 - Milo: 69 Baht
 - USA: 35 Baht
 - Pocky: 10 Baht
 - ชาบูบาร์บี: 19 Baht
 - ชาบูบาร์บี: 7 Baht

ลดสุดไปเลย

ลดสุดๆ ไม่ต้องรอ!

Product Examples:
 - ชาบูบาร์บี: 17 Baht
 - ชาบูบาร์บี: 22 Baht
 - ชาบูบาร์บี: 26 Baht
 - ชาบูบาร์บี: 15 Baht
 - ชาบูบาร์บี: 285 Baht
 - ชาบูบาร์บี: 117 Baht
 - ชาบูบาร์บี: 39 Baht
 - ชาบูบาร์บี: 49 Baht
 - ชาบูบาร์บี: 51 Baht
 - ชาบูบาร์บี: 86 Baht
 - ชาบูบาร์บี: 10 Baht
 - ชาบูบาร์บี: 39 Baht
 - ชาบูบาร์บี: 109 Baht
 - ชาบูบาร์บี: 99 Baht
 - ชาบูบาร์บี: 49 Baht
 - ชาบูบาร์บี: 179 Baht
 - ชาบูบาร์บี: 39 Baht
 - ชาบูบาร์บี: 10 Baht
 - ชาบูบาร์บี: 83 Baht
 - ชาบูบาร์บี: 39 Baht
 - ชาบูบาร์บี: 16 Baht
 - ชาบูบาร์บี: 35 Baht
 - ชาบูบาร์บี: 57 Baht
 - ชาบูบาร์บี: 52 Baht
 - ชาบูบาร์บี: 47 Baht

เดือนพฤษภาคม ลดกันทั่ว ใช้คูปองลดกันทั่ว 150 บาท ฟรี!

สิทธิ์เพิ่ม ลดเพิ่ม 10 บาท*

สิทธิ์เพิ่ม ลดเพิ่ม 500 บาท*

เงื่อนไขของสิทธิ์เพิ่มลดเพิ่ม จำกัด 100 ชุด / 1 สาขา ต้องซื้อสินค้าที่บิ๊กซี ครบ 150 บาท จึงจะสามารถใช้สิทธิ์เพิ่มลดเพิ่มได้ ต้องซื้อสินค้าที่บิ๊กซี ครบ 150 บาท จึงจะสามารถใช้สิทธิ์เพิ่มลดเพิ่มได้

สังกัด โลหะบ้านคุณ

TESCO Lotus express

ลด กัน เดือน

วันที่ 28 ก.ย. - 29 ต.ค. 60

Product Examples:
 - น้ำอัดลม ตราเสือ ขนาด 420 มล. ลดเหลือ 24 Baht
 - ชาอุตสาห์ ตราเสือ ขนาด 380 กรัม ลดเหลือ 19 Baht
 - ชาอุตสาห์ ตราเสือ ขนาด 170, 180 มล. ลดเหลือ 41 Baht
 - โนร์มีสก์ ขนาด 225 มล. ลดเหลือ 64 Baht
 - ชาอุตสาห์ ตราเสือ ขนาด 420 มล. ลดเหลือ 11 Baht
 - ชาอุตสาห์ ตราเสือ ขนาด 420 มล. ลดเหลือ 85 Baht
 - ชาอุตสาห์ ตราเสือ ขนาด 420 มล. ลดเหลือ 115 Baht
 - ชาอุตสาห์ ตราเสือ ขนาด 420 มล. ลดเหลือ 198 Baht
 - ชาอุตสาห์ ตราเสือ ขนาด 420 มล. ลดเหลือ 48 Baht
 - น้ำยาล้างหน้า ตราเสือ ขนาด 100 มล. ลดเหลือ 145 Baht
 - น้ำยาล้างหน้า ตราเสือ ขนาด 100 มล. ลดเหลือ 89 Baht
 - น้ำยาล้างหน้า ตราเสือ ขนาด 100 มล. ลดเหลือ 96 Baht
 - ยาสีฟัน Colgate ขนาด 149 Baht
 - ยาสีฟัน Colgate ขนาด 149 Baht
 - ยาสีฟัน Colgate ขนาด 105 Baht
 - น้ำยาล้างผัก ตราเสือ ขนาด 500 มล. ลดเหลือ 15 Baht
 - น้ำยาล้างผัก ตราเสือ ขนาด 500 มล. ลดเหลือ 33 Baht
 - น้ำยาล้างผัก ตราเสือ ขนาด 500 มล. ลดเหลือ 29 Baht
 - น้ำยาล้างผัก ตราเสือ ขนาด 500 มล. ลดเหลือ 139 Baht
 - น้ำยาล้างผัก ตราเสือ ขนาด 500 มล. ลดเหลือ 115 Baht
 - น้ำยาล้างผัก ตราเสือ ขนาด 500 มล. ลดเหลือ 13.50 Baht

กิ๊บปี๊ส ได้รับการยกย่องเป็น Trusted Brand
 ซึ่งเป็นอุตสาหกรรมที่พึ่งรักษาก้าวให้คงอยู่ 5 ปี สาขาญี่ปุ่นที่แรก ใจภัยยังคงรับรองต่อไป

กิ๊บปี๊ส บริษัทไทยก้าวไกล จำกัด สำหรับ 30 ปี - 19 ก.ย. 2560

Product Examples:
 - กิ๊บปี๊ส ชาบูบาร์บี ขนาด 400 กรัม ลดเหลือ 12 Baht
 - กิ๊บปี๊ส ชาบูบาร์บี ขนาด 400 กรัม ลดเหลือ 22 Baht
 - กิ๊บปี๊ส ชาบูบาร์บี ขนาด 400 กรัม ลดเหลือ 39 Baht
 - กิ๊บปี๊ส ชาบูบาร์บี ขนาด 400 กรัม ลดเหลือ 99 Baht
 - กิ๊บปี๊ส ชาบูบาร์บี ขนาด 400 กรัม ลดเหลือ 79 Baht
 - กิ๊บปี๊ส ชาบูบาร์บี ขนาด 400 กรัม ลดเหลือ 119 Baht
 - กิ๊บปี๊ส ชาบูบาร์บี ขนาด 400 กรัม ลดเหลือ 129 Baht
 - กิ๊บปี๊ส ชาบูบาร์บี ขนาด 400 กรัม ลดเหลือ 59 Baht
 - กิ๊บปี๊ส ชาบูบาร์บี ขนาด 400 กรัม ลดเหลือ 135 Baht
 - กิ๊บปี๊ส ชาบูบาร์บี ขนาด 400 กรัม ลดเหลือ 49 Baht
 - กิ๊บปี๊ส ชาบูบาร์บี ขนาด 400 กรัม ลดเหลือ 219 Baht
 - กิ๊บปี๊ส ชาบูบาร์บี ขนาด 400 กรัม ลดเหลือ 42 Baht
 - กิ๊บปี๊ส ชาบูบาร์บี ขนาด 400 กรัม ลดเหลือ 59 Baht
 - กิ๊บปี๊ส ชาบูบาร์บี ขนาด 400 กรัม ลดเหลือ 21 Baht
 - กิ๊บปี๊ส ชาบูบาร์บี ขนาด 400 กรัม ลดเหลือ 95 Baht
 - กิ๊บปี๊ส ชาบูบาร์บี ขนาด 400 กรัม ลดเหลือ 99 Baht
 - กิ๊บปี๊ส ชาบูบาร์บี ขนาด 400 กรัม ลดเหลือ 42 Baht
 - กิ๊บปี๊ส ชาบูบาร์บี ขนาด 400 กรัม ลดเหลือ 59 Baht
 - กิ๊บปี๊ส ชาบูบาร์บี ขนาด 400 กรัม ลดเหลือ 63 Baht
 - กิ๊บปี๊ส ชาบูบาร์บี ขนาด 400 กรัม ลดเหลือ 79 Baht
 - กิ๊บปี๊ส ชาบูบาร์บี ขนาด 400 กรัม ลดเหลือ 15 Baht
 - กิ๊บปี๊ส ชาบูบาร์บี ขนาด 400 กรัม ลดเหลือ 123 Baht
 - กิ๊บปี๊ส ชาบูบาร์บี ขนาด 400 กรัม ลดเหลือ 129 Baht
 - กิ๊บปี๊ส ชาบูบาร์บี ขนาด 400 กรัม ลดเหลือ 65 Baht
 - กิ๊บปี๊ส ชาบูบาร์บี ขนาด 400 กรัม ลดเหลือ 30 Baht



The Billion Prices Project: Using Online Prices for Measurement and Research *

Alberto Cavallo

MIT and NBER

Roberto Rigobon

MIT and NBER

This Version: April 8, 2016

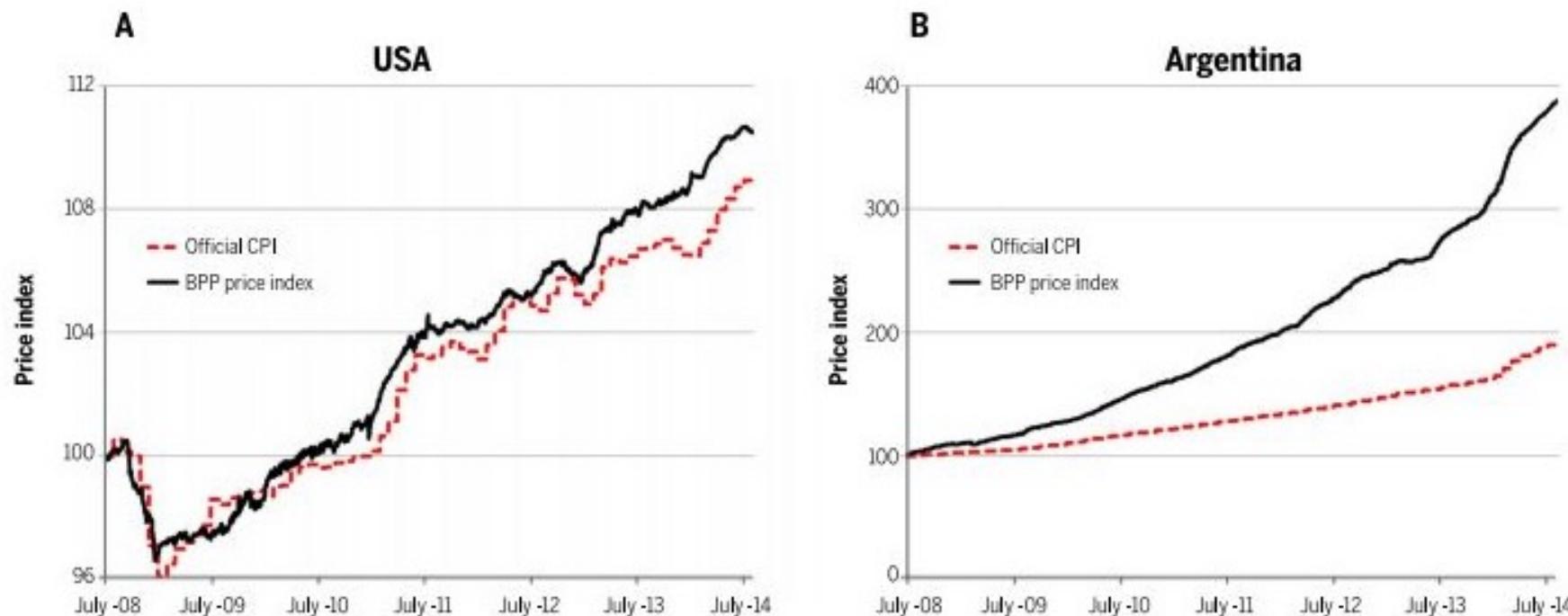
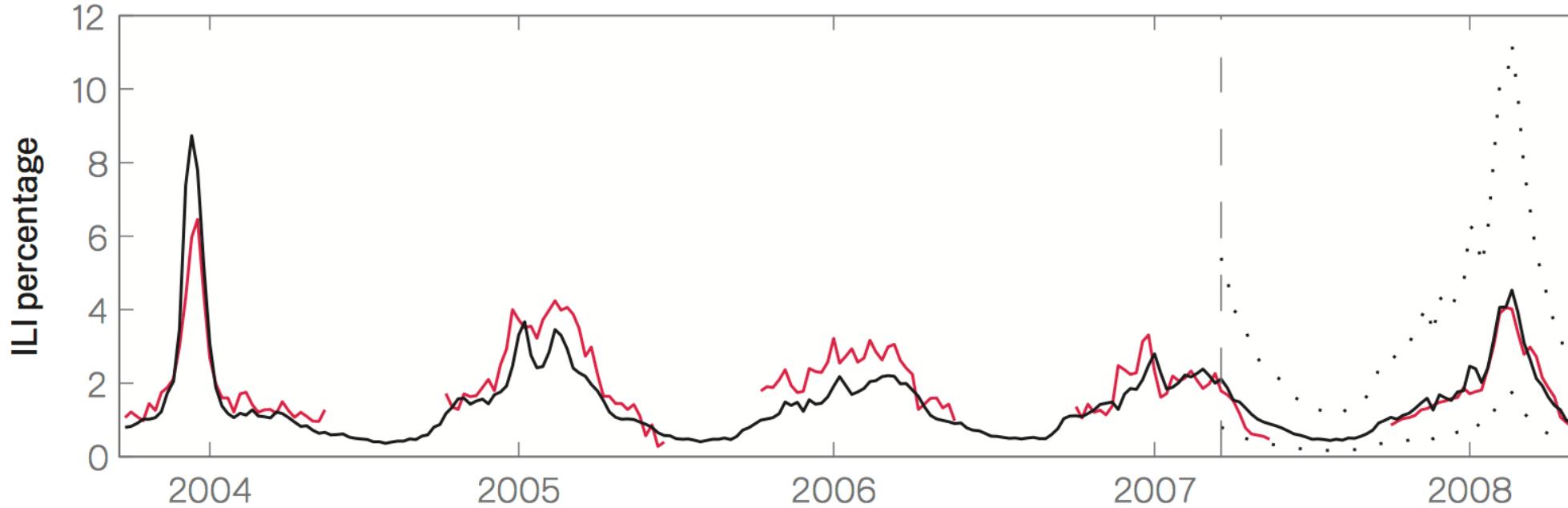


Fig. 2. BPP price index. Dashed red lines show the monthly series for the CPI in the United States (A) and Argentina (B), as published by the formal government statistics agencies. Solid black lines show the daily price index series, the "State Street's PriceStats Series" produced by the BPP, which uses scraped Internet data on thousands of retail items. All indices are normalized to 100 as of 1 July 2009. In the US chart, the two series track

each other quite closely, although the BPP index is available in real time and at a more granular level (daily instead of monthly). In the plot for Argentina, the indices diverge considerably, with the BPP index growing at about twice the rate of the official CPI. [Updated version of figure 5 in (18), provided courtesy of Alberto Cavallo and Roberto Rigobon, principal investigators of the BPP]



Google Flu Trend



Ginsberg, Jeremy; Mohebbi, Matthew H.; Patel, Rajan S.; Brammer, Lynnette;
Smolinski, Mark S.; Brilliant, Larry (19 February 2009). "Detecting influenza
epidemics using search engine query data". *Nature*. **457** (7232): 1012–1014.



What are they using data science for?

1. Measurement
2. Insights
3. Data Products





1) Measurement

- To make a decision based on data
- Aka. benchmarking
- Turning qualitative information into quantitative values
 - Usually called metrics or indicators
- Direct and indirect measurement

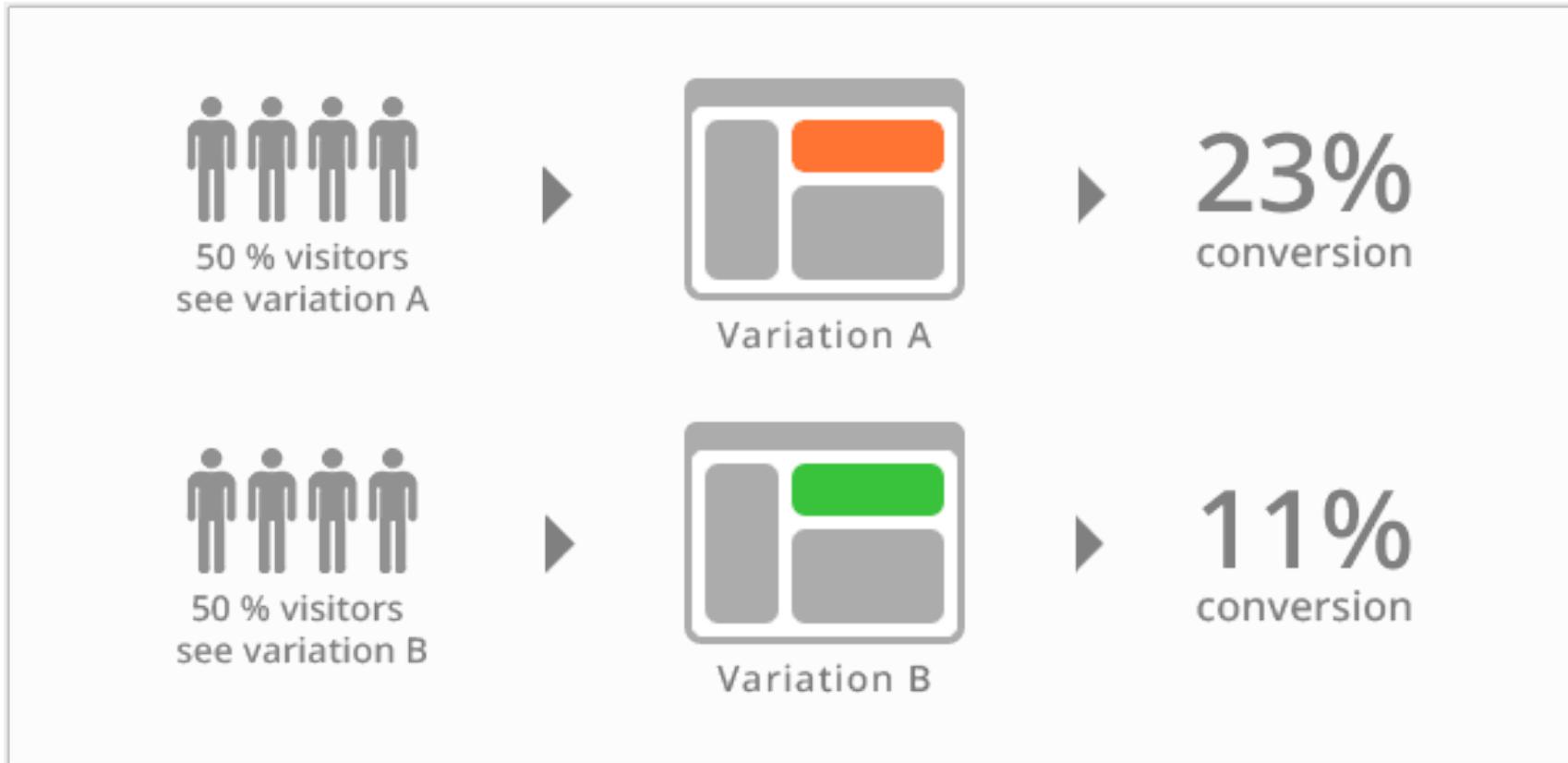


Why do we need to measure?

- Comparison between alternatives (**make a selection**)
 - Choosing which notebook to buy
- Comparison after **improvement** or tuning
 - Should I add memory to my notebook?
- **A/B Testing** (split testing)
 - Let the actual users decide their preferences
 - Very popular for UI design



A/B Testing



Source: <https://vwo.com/ab-testing/>



Example: SimCity

1. Remove product banner: SimCity sees 43% more conversions without hero banner image

Control

The screenshot shows the official EA SimCity website. A large red box highlights a prominent banner at the top left that reads "PRE-ORDER AND GET \$20 OFF YOUR NEXT PURCHASE". Below this, there are two main purchase options: "SIMCITY™" for \$59.99 and "SIMCITY™ DIGITAL DELUXE EDITION" for \$79.99. Each option includes a "BUY NOW" button and a choice between PC Download and PC Physical. To the right, there's a section titled "DIGITAL DELUXE EDITION INCLUDES" listing four items: HEROES AND VILLAINS SET, FRENCH CITY SET, GERMAN CITY SET, and BRITISH CITY SET. At the bottom, there are sections for "Key Features", "WHAT IS SIMCITY?", and "DEPTH OF SIMULATION".

Variation

The screenshot shows the same SimCity website, but the pre-order banner from the control version has been removed. In its place, a large blue circular overlay on the right side displays the text "43% increase in checkouts". The rest of the page layout remains identical to the control version, featuring the two main purchase options, the digital deluxe edition inclusion list, and the footer sections.



Example: SmartWool

3. Use a well-defined grid layout for your online shopping experience: Uniform product page images increase ARPV 17% for SmartWool

Control

The 'Control' version of the SmartWool product page features a grid layout for socks. The main image at the top shows two pairs of socks being worn while running. Below it, a grid displays various sock models categorized by activity (Run, Hike, On-Snow, etc.) and cushioning levels (Micro, Mini, Crew, Over-The-Calf). Each item includes a small image, the product name, and price.

Variation

The 'Variation' version of the SmartWool product page includes a red border around the grid area. It adds several filtering options: 'Sort By' (Price, Favorites), 'Refine your Search' dropdown, and category filters for 'CATEGORY' (Lifestyle Socks, Performance Socks), 'ACTIVITY' (Hike, On-Snow, Cycle, Run, Fish/Hunt, Around Town), 'COLOR' (color swatches), 'CUSHION' (Non-Cushion, Ultra Light Cushion, Light Cushion, Medium Cushion, Heavy Cushion, Extra Heavy Cushion), 'HEIGHT' (Micro, Mini, 3/4 Crew, Crew, Over The Calf), and 'SIZE' (XS). The grid layout remains similar to the control version but is presented within a more structured and detailed interface.

17%
increase in
average
revenue per
visitor



2) Insights

<https://blogs.scientificamerican.com/guest-blog/9-bizarre-and-surprising-insights-from-data-science/>

- **Good understanding of user behavior** can lead to new product development or improvements of the existing products

- Walmart -- Pop-Tarts before a hurricane
 - Prehurricane, Strawberry Pop- Tart sales increased about sevenfold

- Financial startup -- Typing with proper capitalization indicates creditworthiness
 - Online loan applicants who complete the application form with the correct case are more dependable debtors

- Starbucks use customer purchase information from My Starbucks Mobile Apps to figure out new products

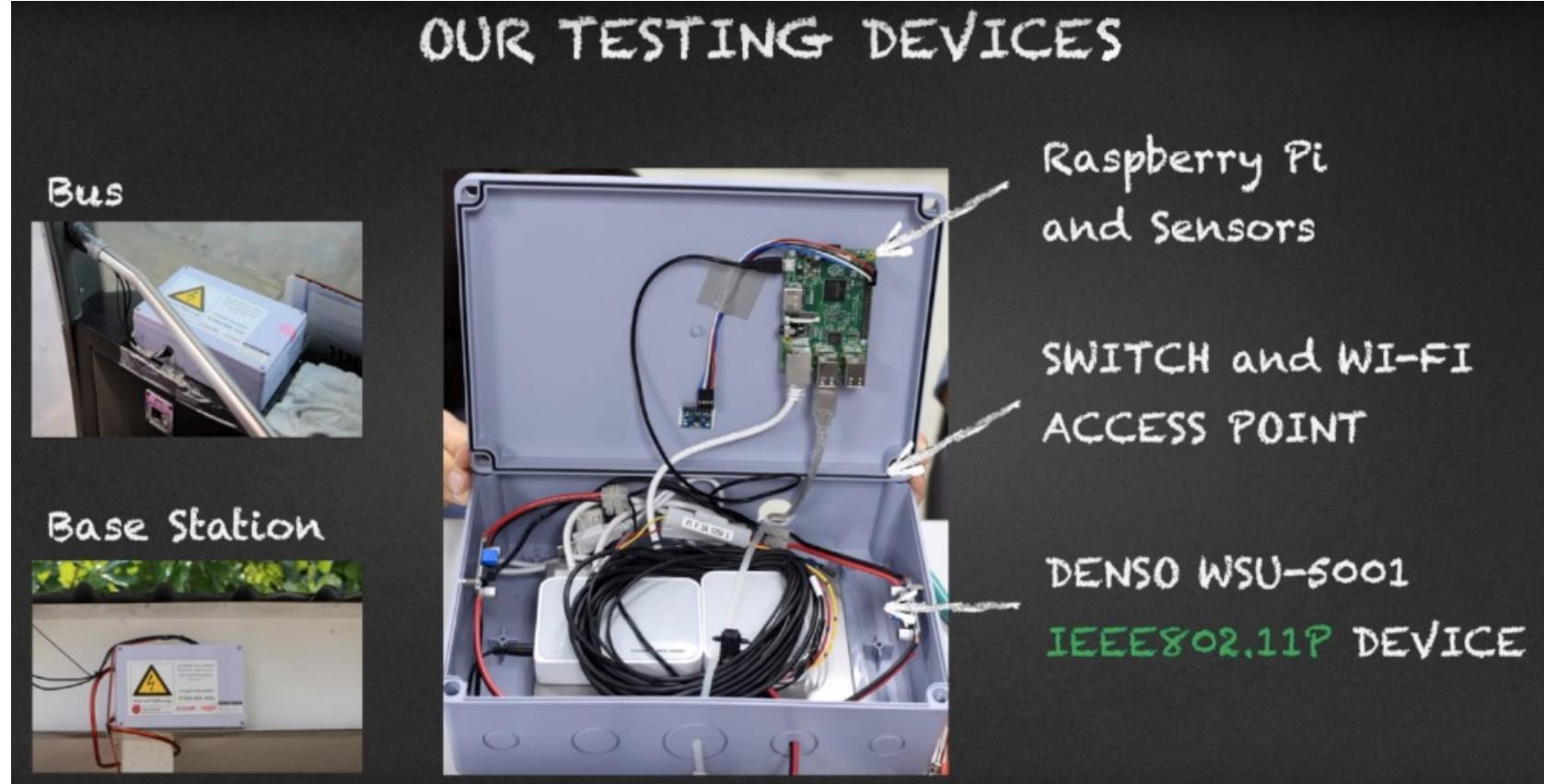


Example: Tracing Traffic





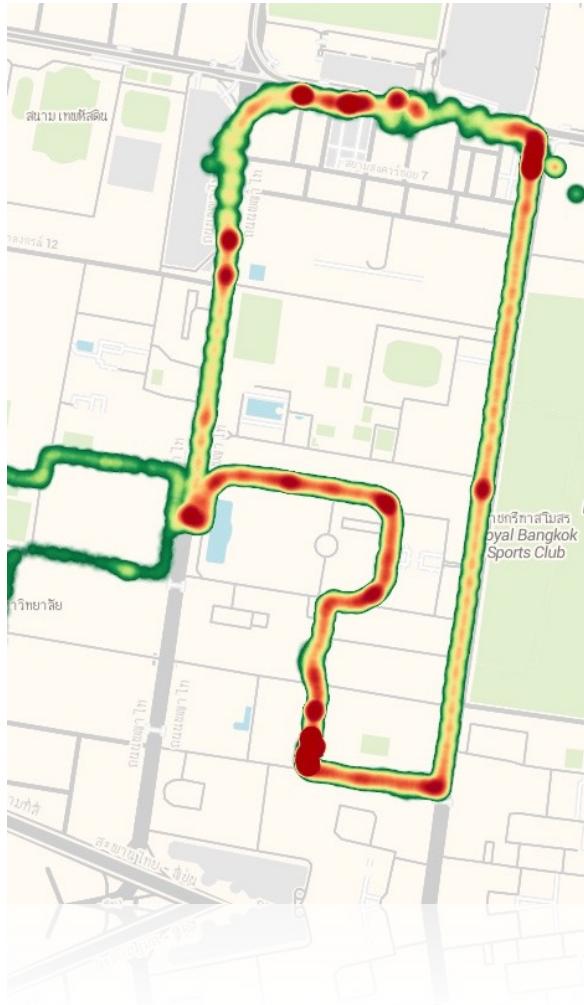
Example: Tracing Traffic



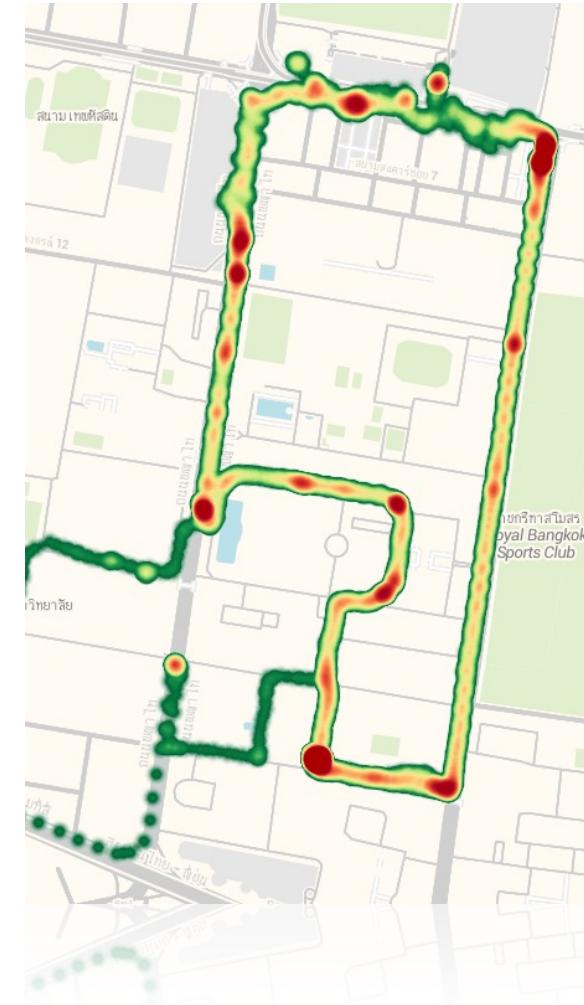


GPS Average Speed

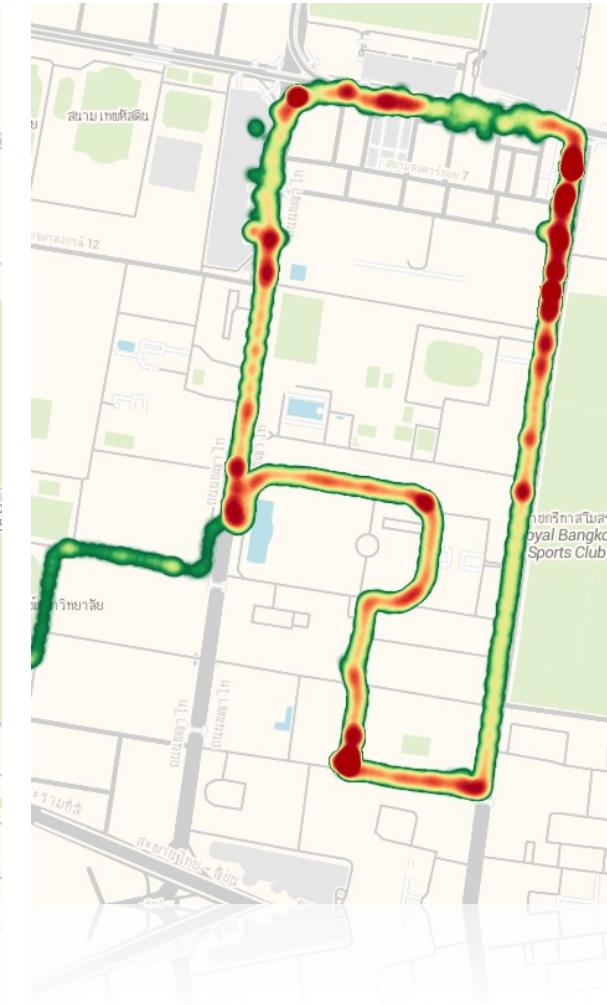
6:00-10:00



10:00-15:00



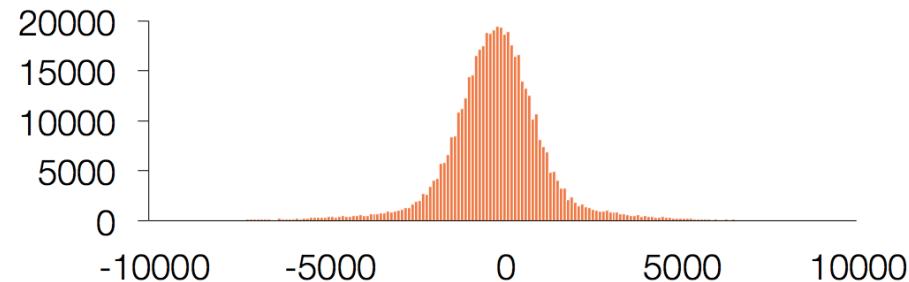
15:00-18:00



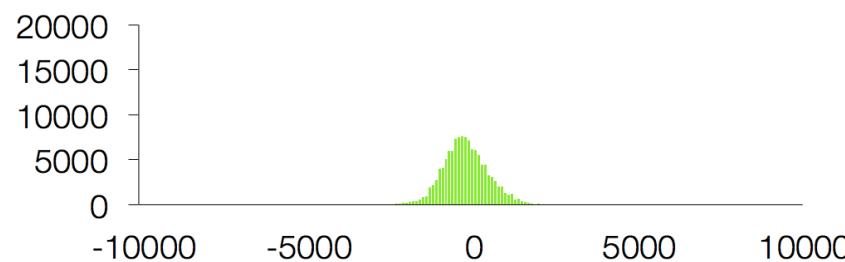
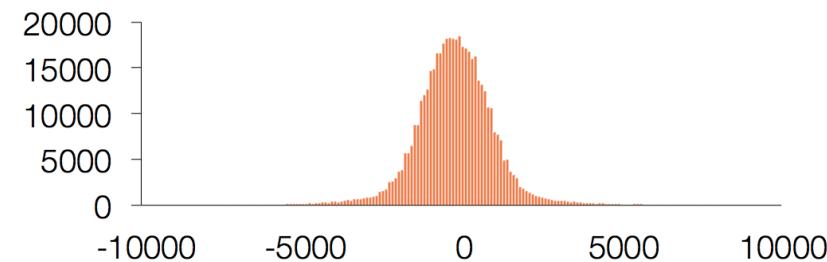
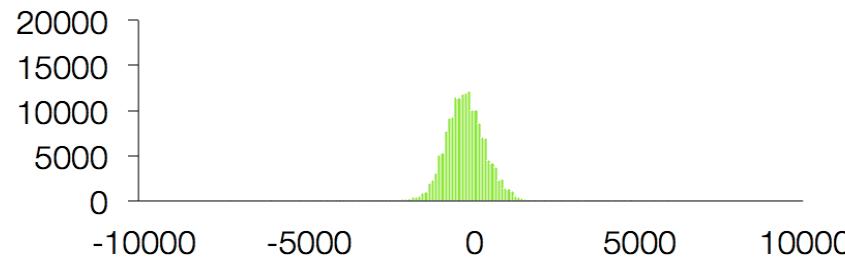
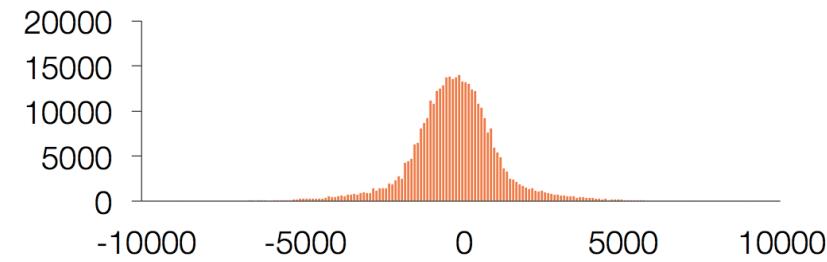
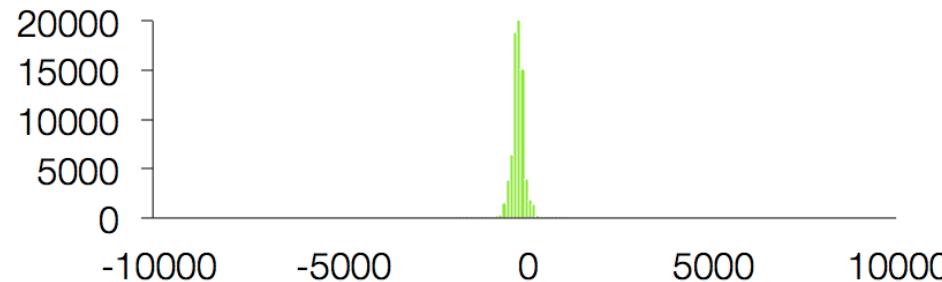


Bus Drivers' Behaviors

Bus A



Bus B





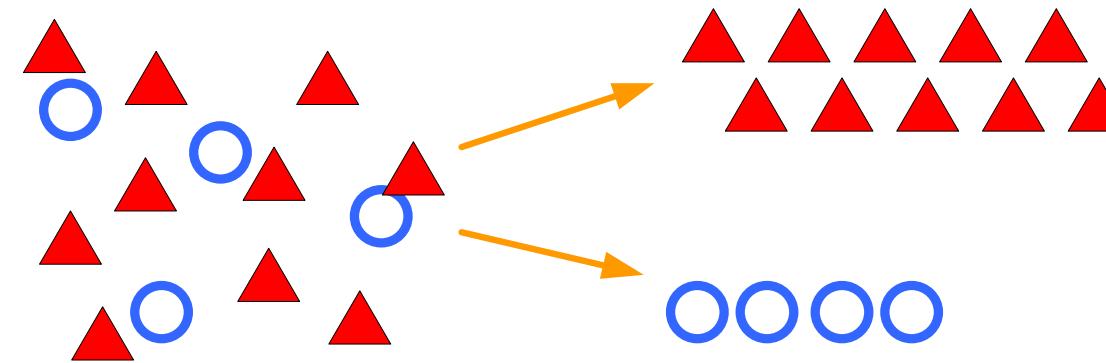
3) Data Products

- An application or system that uses data to provide “intelligent” products or services, which create more data that can be further used
- **Machine learning** plays an important role in building great data products



Machine Learning Classification

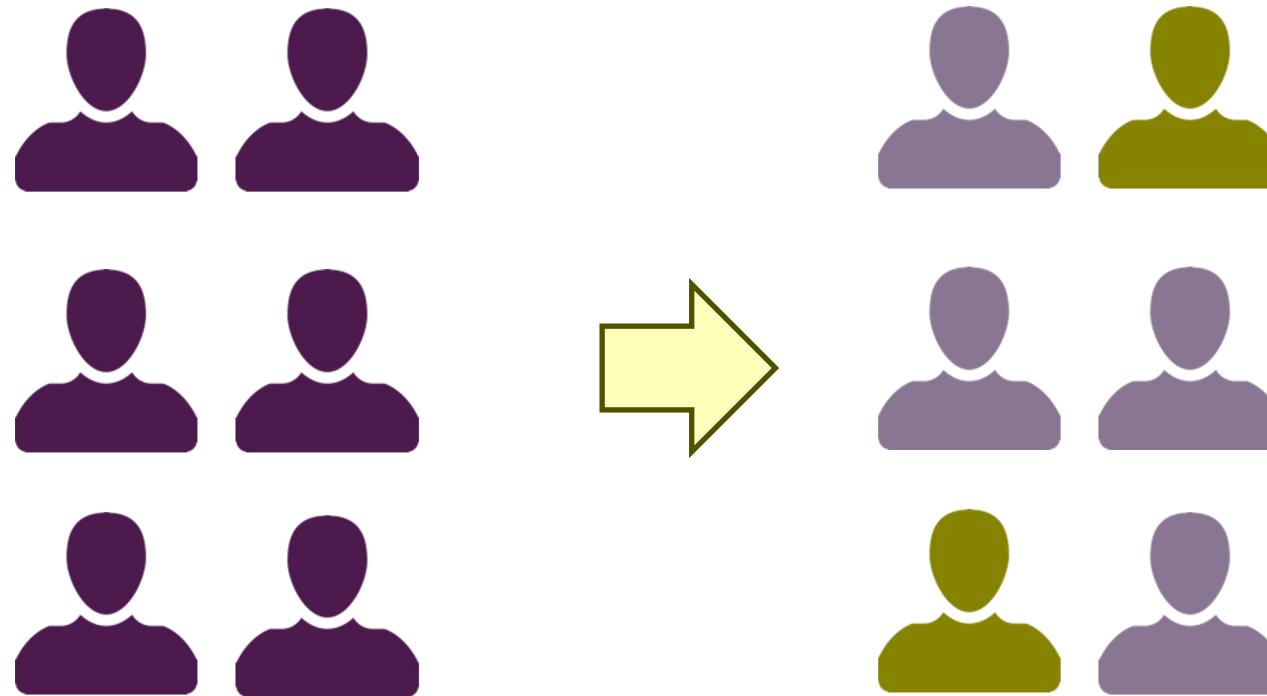
- Identify to which set of categories a new observation belong

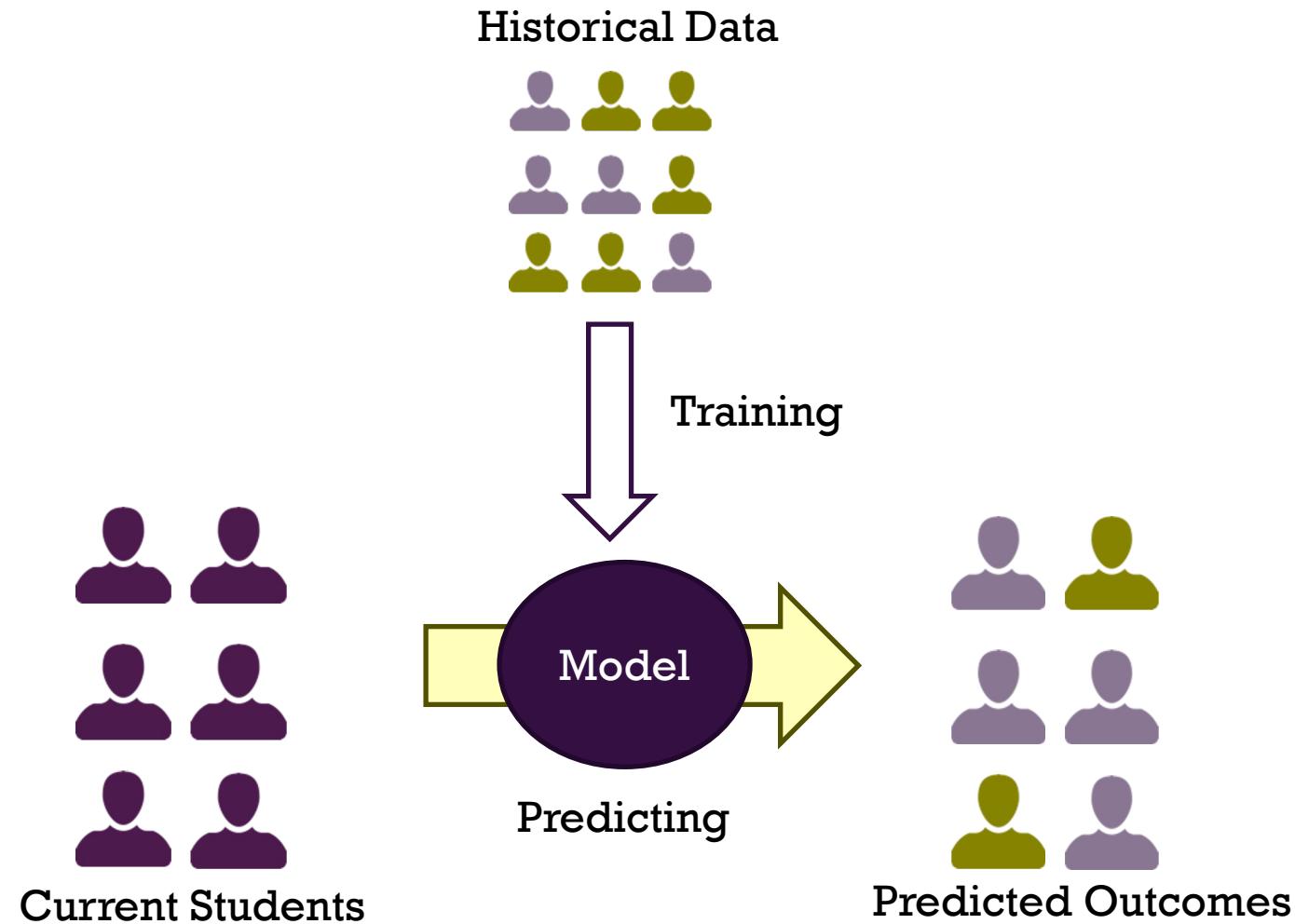


- Example: spam filtering, customer churn prediction, complaint classification



Example: Students Grade Prediction





$$\frac{OS \times Data\ Struct \times Prog}{9} > 7$$

Example: Amazon Recommendation

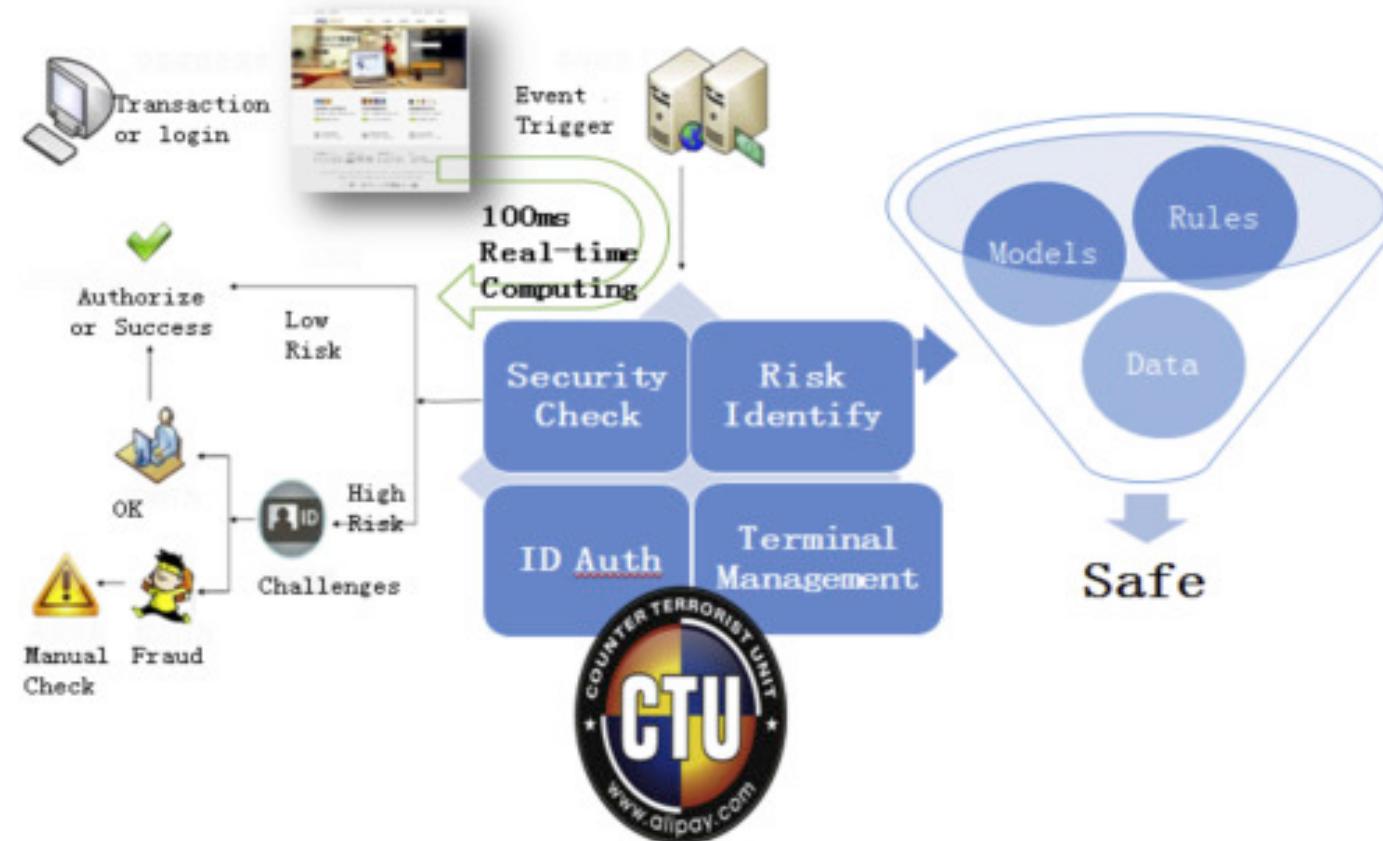
- Amazon sells 480M products (485k new products per day)
- Use recommendation systems to bring products to customers
- Analyze data from 300M customers
 - Purchase history
 - Reviews / Ratings
 - Search history
 - Views

The screenshot shows the top navigation bar of the Amazon website, including links for 'Browsing History', 'Natawut's Amazon.com', 'Today's Deals', and 'Gift Cards'. Below the header, a banner for 'Natawut's Amazon' is displayed, along with a message encouraging users to sign in for order status and balances. The main content area features several recommendation cards:

- Computer & Technology Books**: 92 ITEMS. Includes a book cover for 'Hadoop Application Architectures' by Mark Grover, Ted Malaska, Jonathan Seidman & Gwen Shapira.
- Science & Math Books**: 51 ITEMS. Includes a book cover for 'Storytelling with Data' by Cole Nussbaumer Knaflic.
- Storytelling**: A Guide on How to Tell a Story, With Storytelling Techniques and Storytelling Secrets. Includes a book cover for 'Own the Room' by Robert J. Knell.
- Introductory Machine Learning**: Fully revised and expanded version. Includes a book cover for 'Machine Learning for Absolute Beginners' by Robert J. Knell.



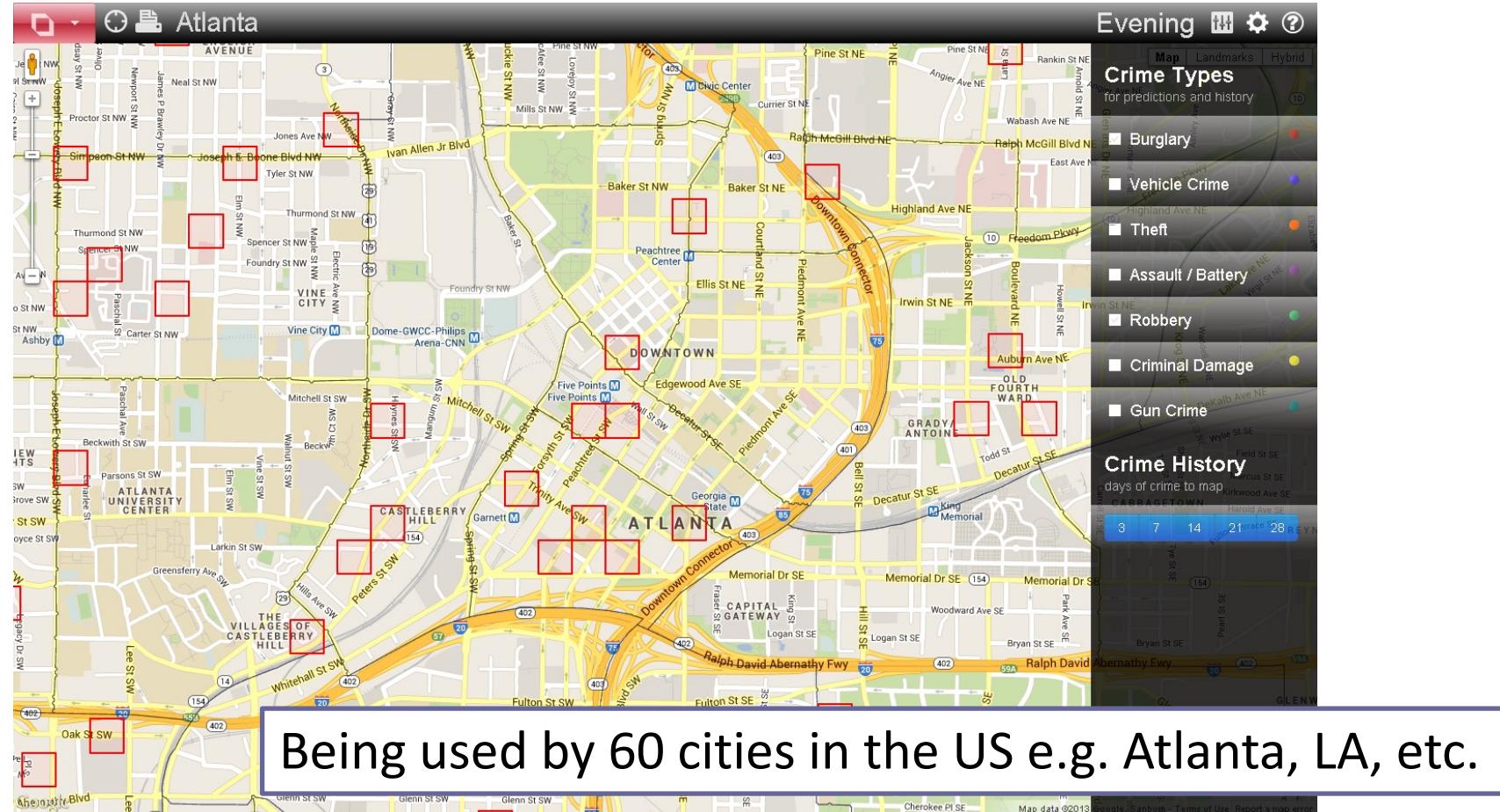
Case study: Alibaba Fraud Detection



Source: <http://www.sciencedirect.com/science/article/pii/S2405918815000021>



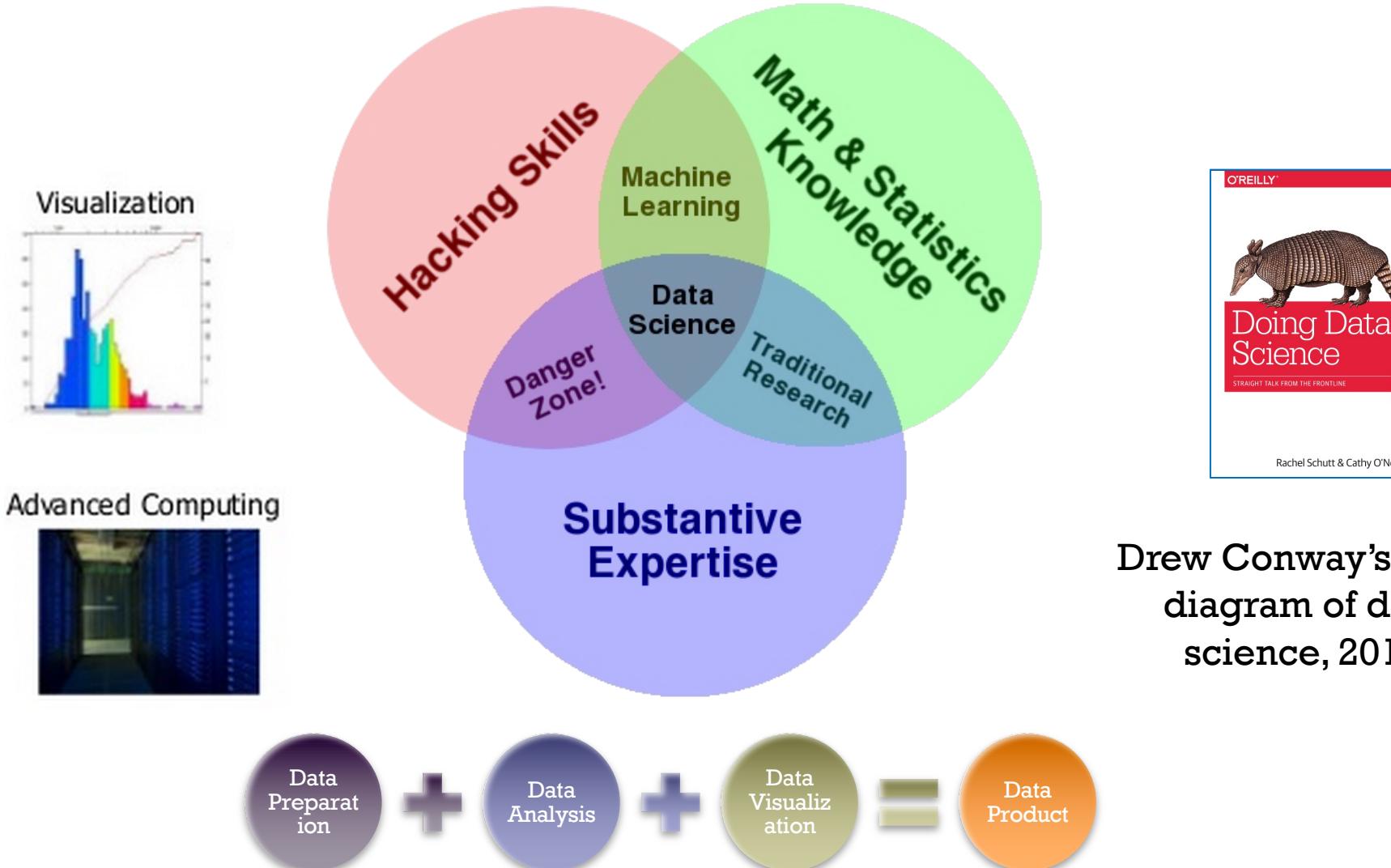
Case study: Predictive Policing



Source: <http://www.forbes.com/sites/ellenhuet/2015/02/11/predpol-predictive-policing>



Drew Conway's Data Science Venn diagram (Skills)





The Most In-Demand Skills for Data Scientists in 2021

Results from webscraping over 15,000 Data Scientist job postings

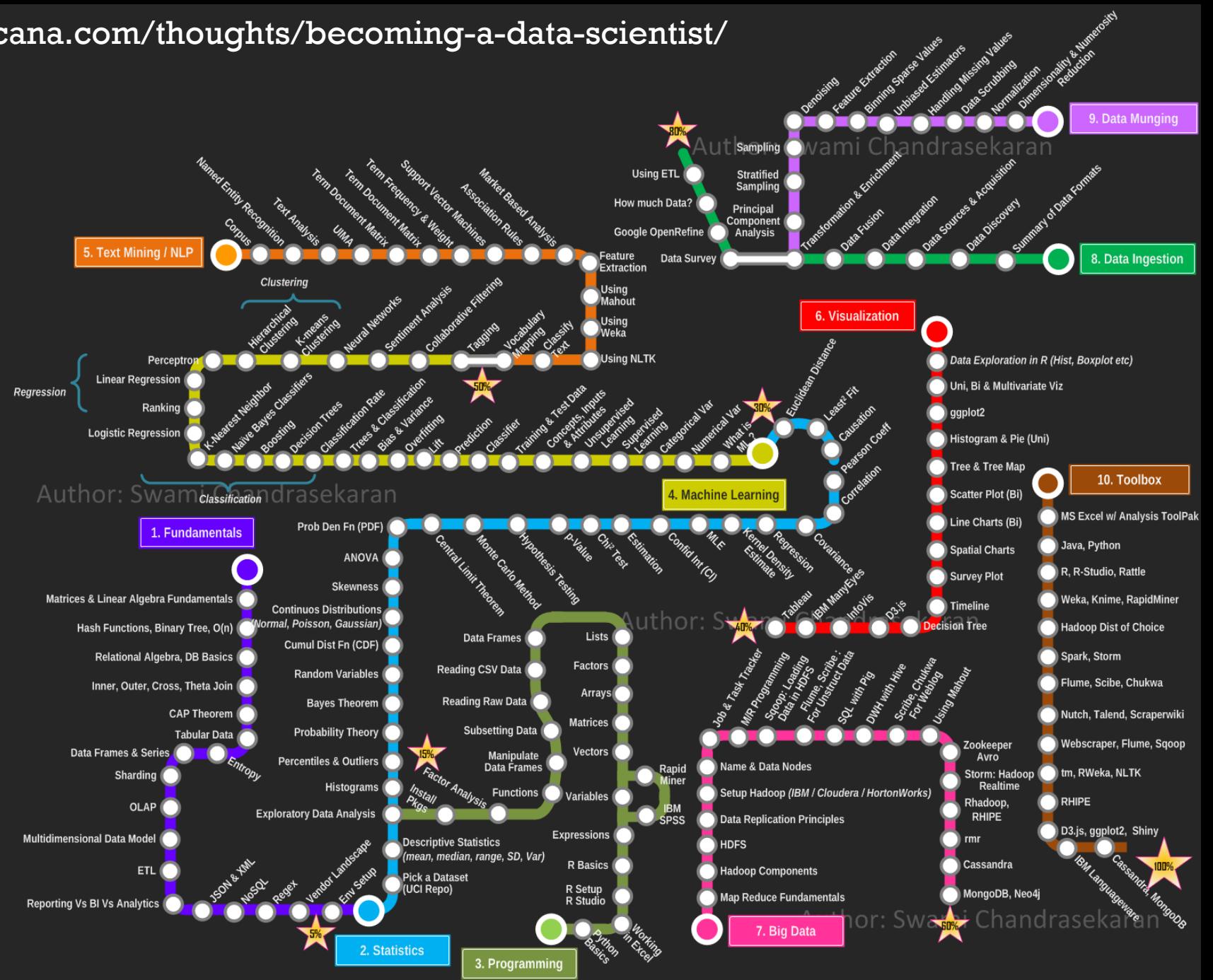


Terence Shin Mar 22, 2021 · 5 min read *



10 Most In-Demand Data Science Skills in 2021







Needed Skills (cont.)



Harvard
Business
Review

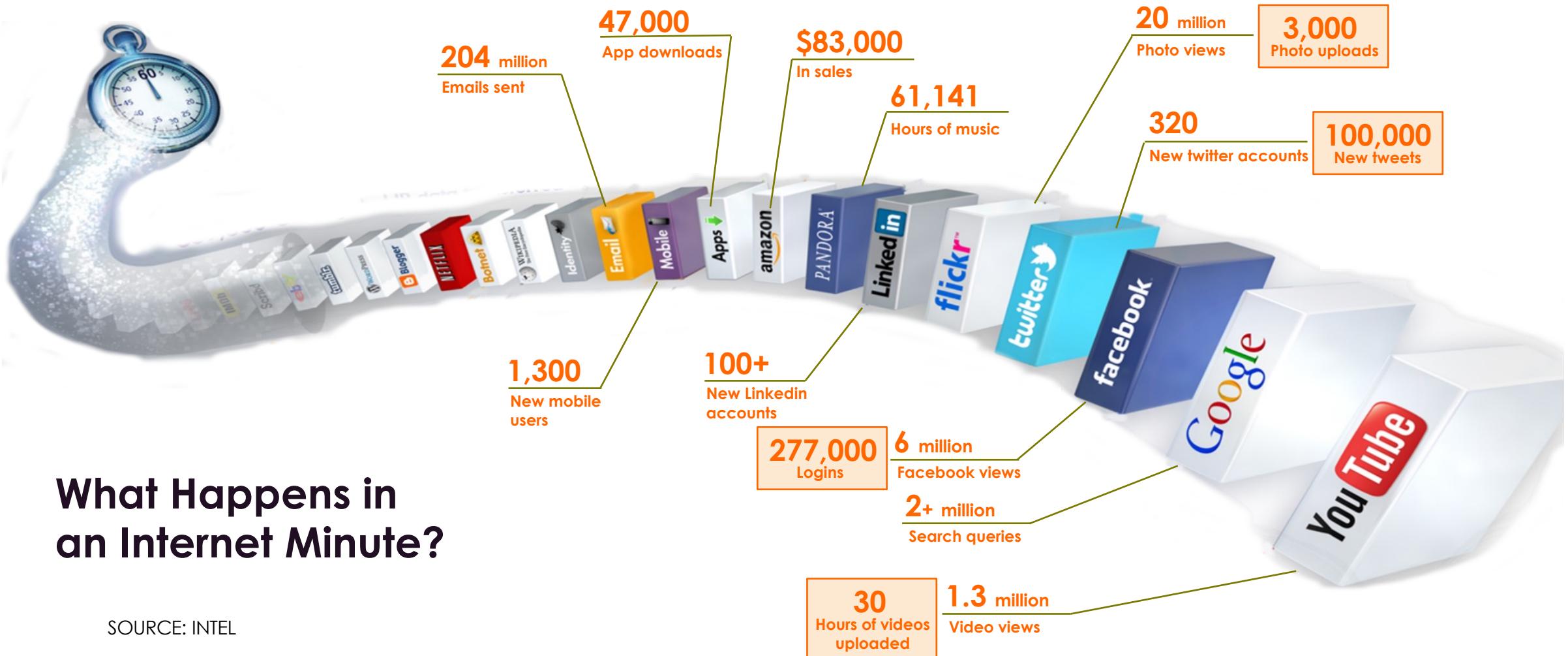
**Data Scientist:
The Sexiest Job of
21st Century is
Tedious, and the
Needs to Change**

By Sean Kandel
From the April 2014 issue

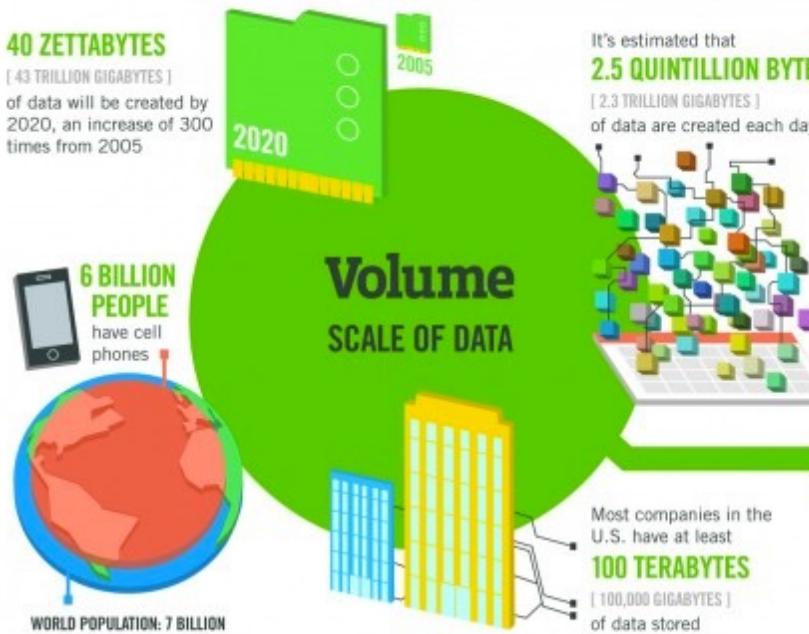


Big Data

Big Data Explosion



40 ZETTABYTES
[43 TRILLION GIGABYTES]
of data will be created by
2020, an increase of 300
times from 2005



The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume**, **Velocity**, **Variety** and **Veracity**.

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015
4.4 MILLION IT JOBS
will be created globally to support big data, with 1.9 million in the United States



The New York Stock Exchange captures
1 TB OF TRADE INFORMATION during each trading session

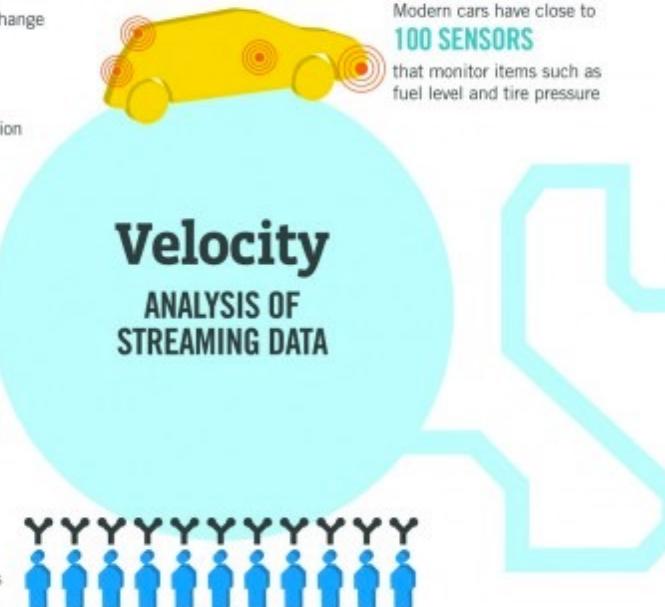


Velocity ANALYSIS OF STREAMING DATA

By 2016, it is projected there will be

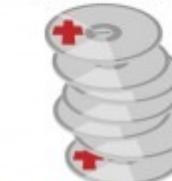
18.9 BILLION NETWORK CONNECTIONS

– almost 2.5 connections per person on earth



As of 2011, the global size of data in healthcare was estimated to be

150 EXABYTES
[161 BILLION GIGABYTES]



30 BILLION PIECES OF CONTENT

are shared on Facebook every month



Variety DIFFERENT FORMS OF DATA



By 2014, it's anticipated there will be
420 MILLION WEARABLE, WIRELESS HEALTH MONITORS

4 BILLION+ HOURS OF VIDEO

are watched on YouTube each month



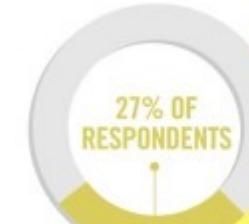
400 MILLION TWEETS

are sent per day by about 200 million monthly active users

Poor data quality costs the US economy around
\$3.1 TRILLION A YEAR



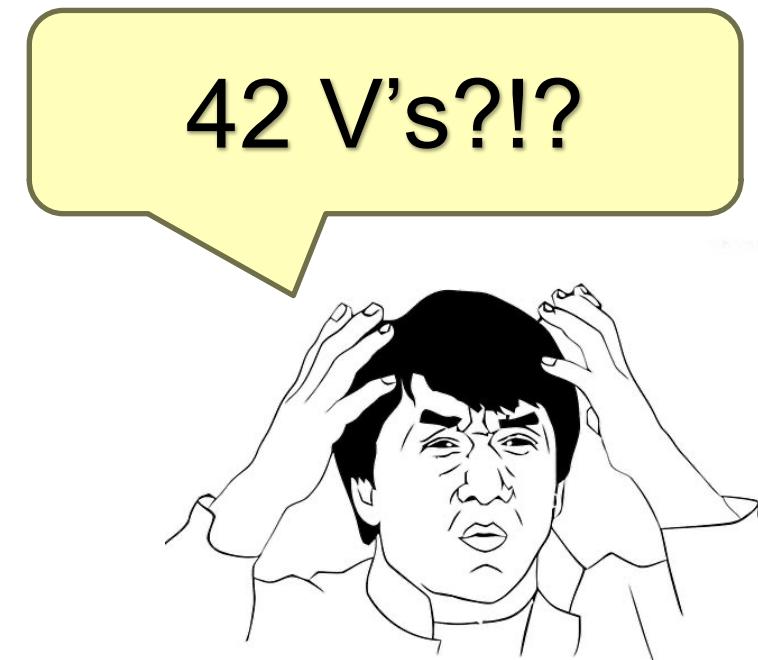
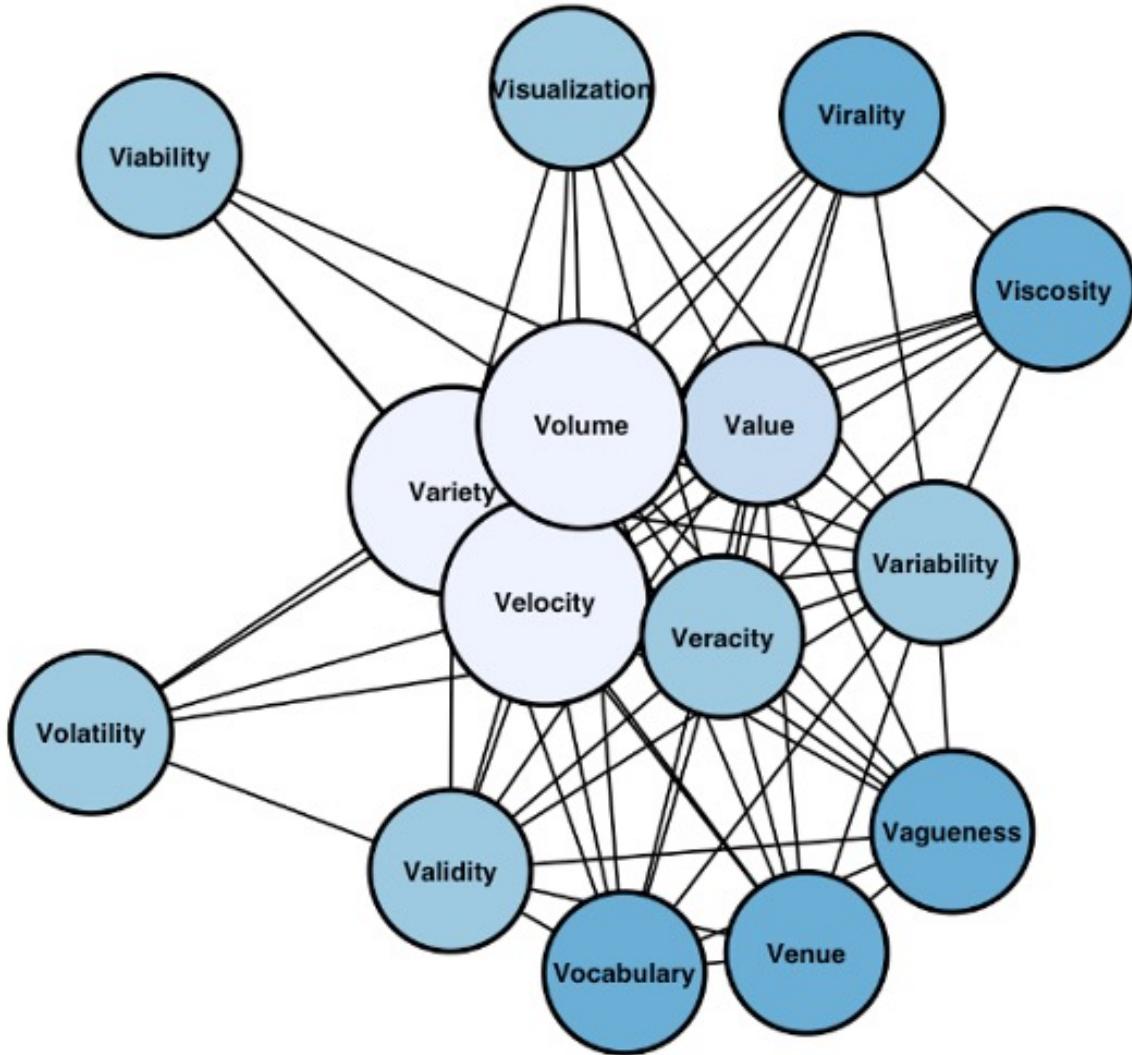
1 IN 3 BUSINESS LEADERS
don't trust the information they use to make decisions



Veracity UNCERTAINTY OF DATA

in one survey were unsure of how much of their data was inaccurate

Now 42 V's of Big Data



Big Data Driver: Internal + External Data





A Better Way to Know

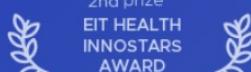
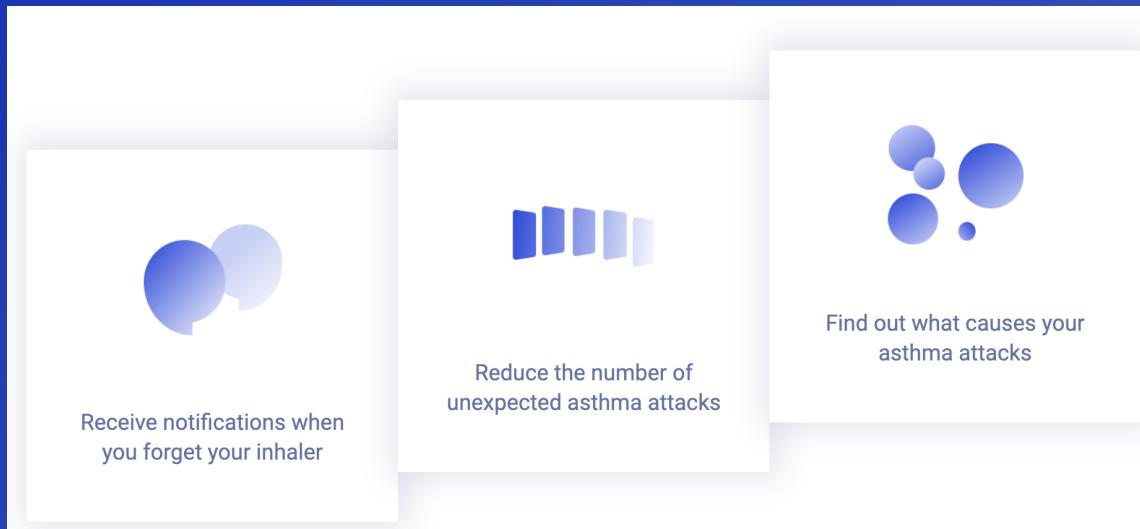
We understand what it's like to hover over the crib at night. That's why we invented the Owlet Smart Sock. It's a better way to check on your baby and smarter way to know they're okay.

Every Beat. Every Breath.

KNOWING IS BEST

Control your asthma with FindAir smart inhaler

Full Control. Less asthma attacks. Better life.



2nd prize
EIT HEALTH
INNOSTARS
AWARD
2018



1st prize
UPC DIGITAL
IMAGINATION
CHALLENGE
2017



Best Pitch
SMART
HEALTH
BUDAPEST
2017



1st prize
INNOLABS
DIGITAL HEALTH
HACKATHON
2017



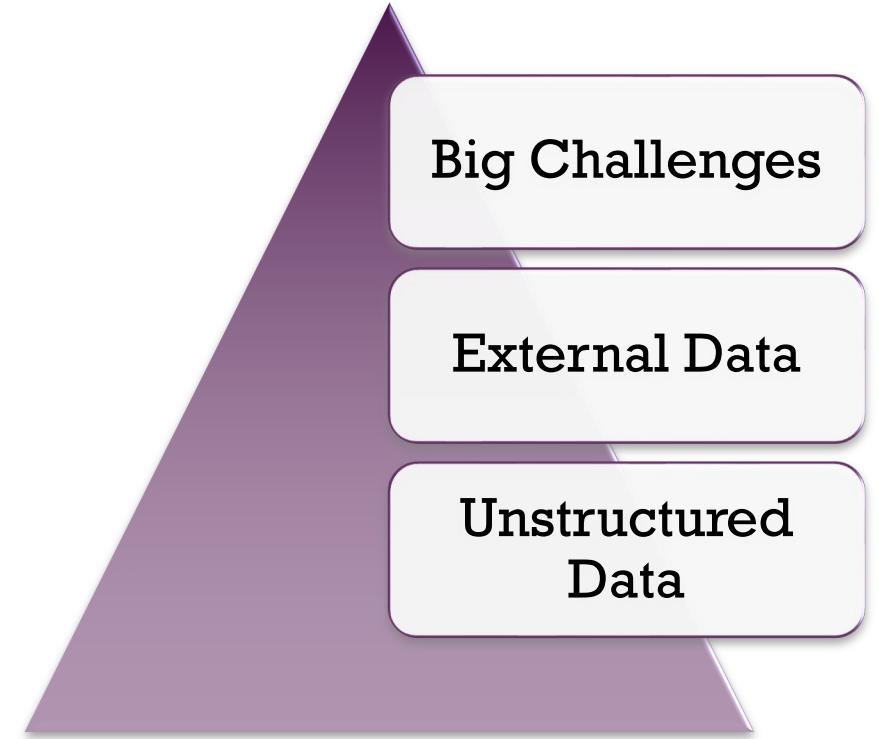
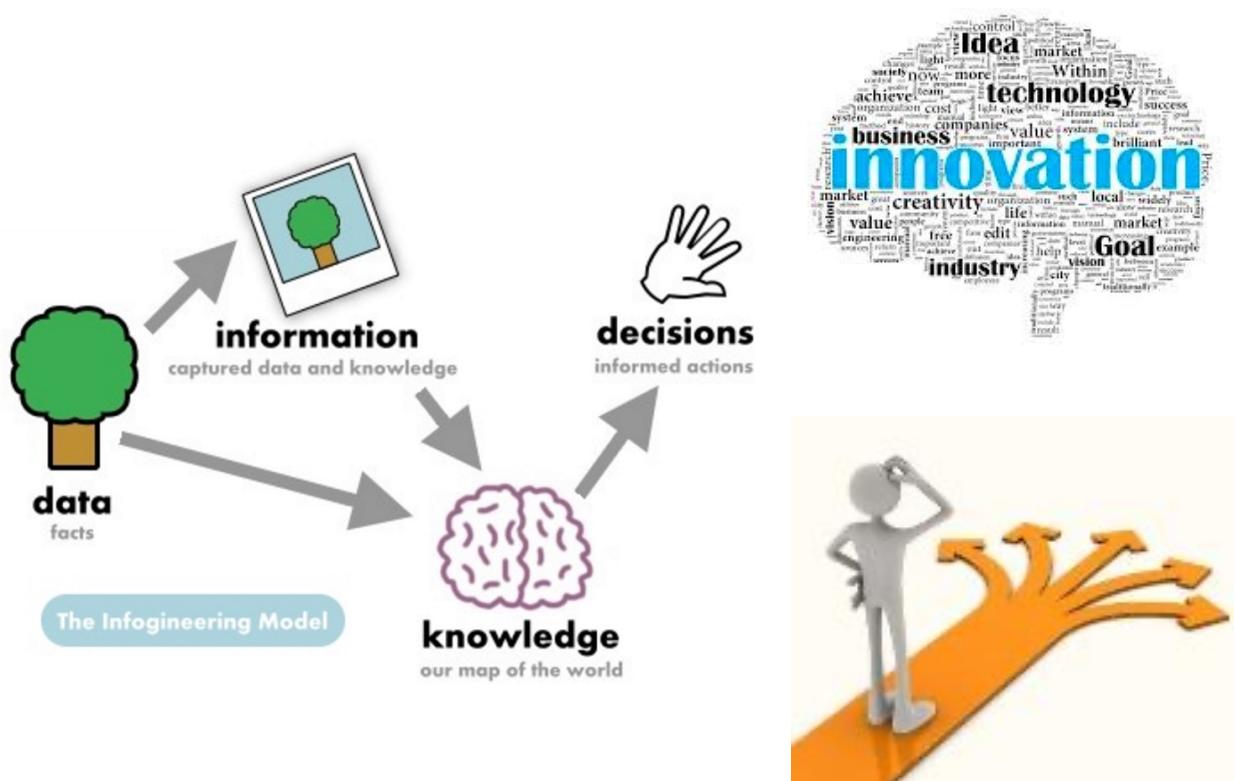
1st prize
INNOLABS
DIGITAL HEALTH
HACKATHON
2017



1st prize
MEDTRENDS
TOP TRENDS
2018

Big Data Analytics

- It is a process of examining **Big Data** to uncover useful information and knowledge.
- More data means better decision!



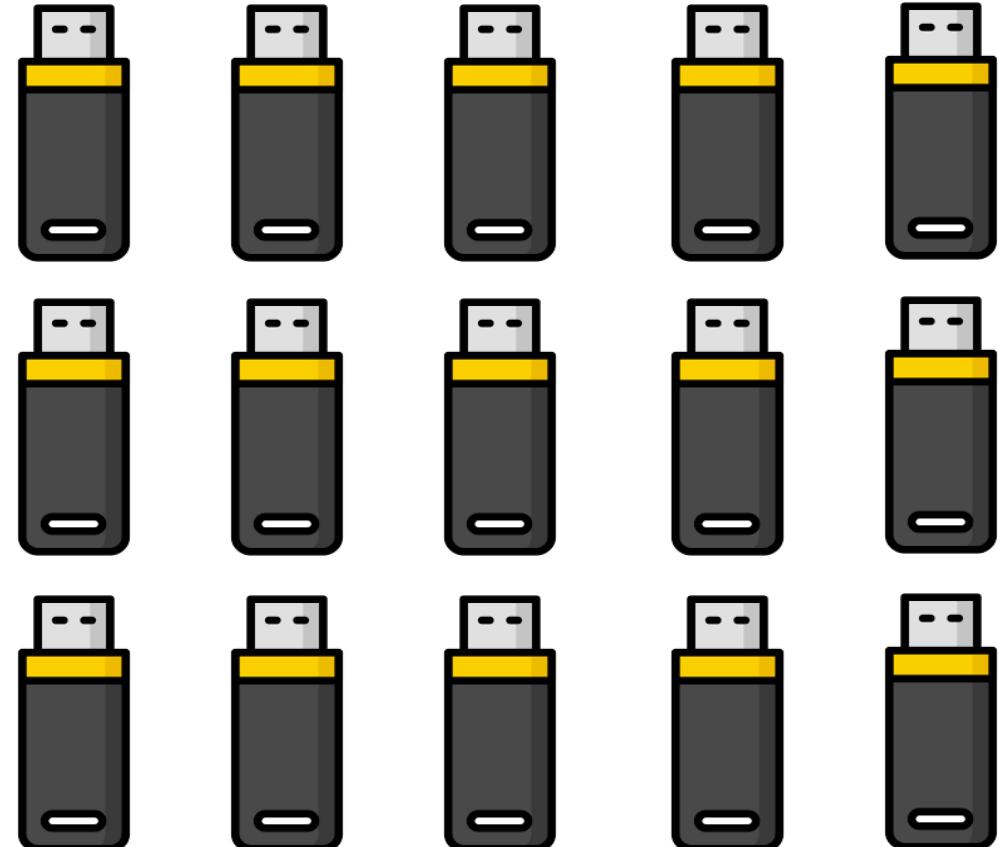
Big Data Challenges

Same tasks, but much more difficult!

2MB



200TB



Big Data Solution



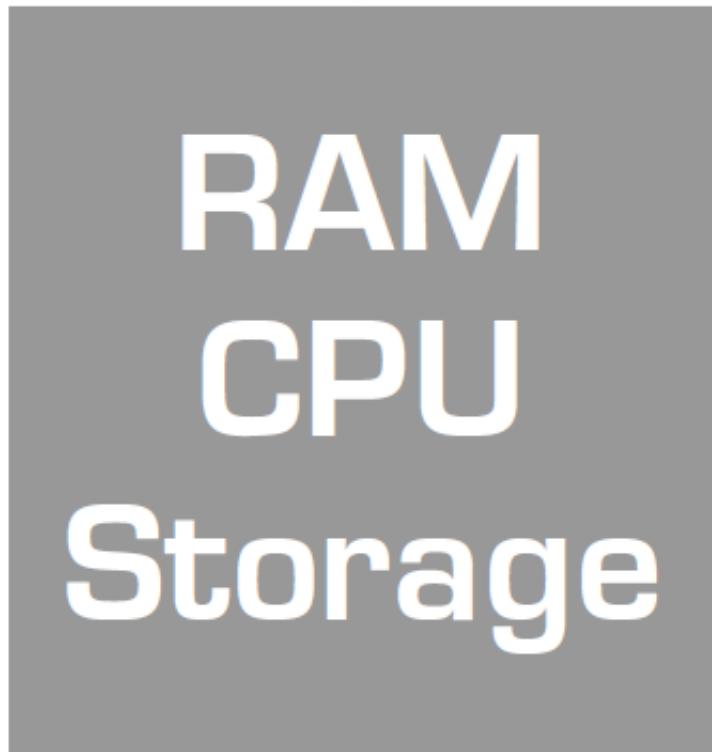
INFRASTRUCTURE



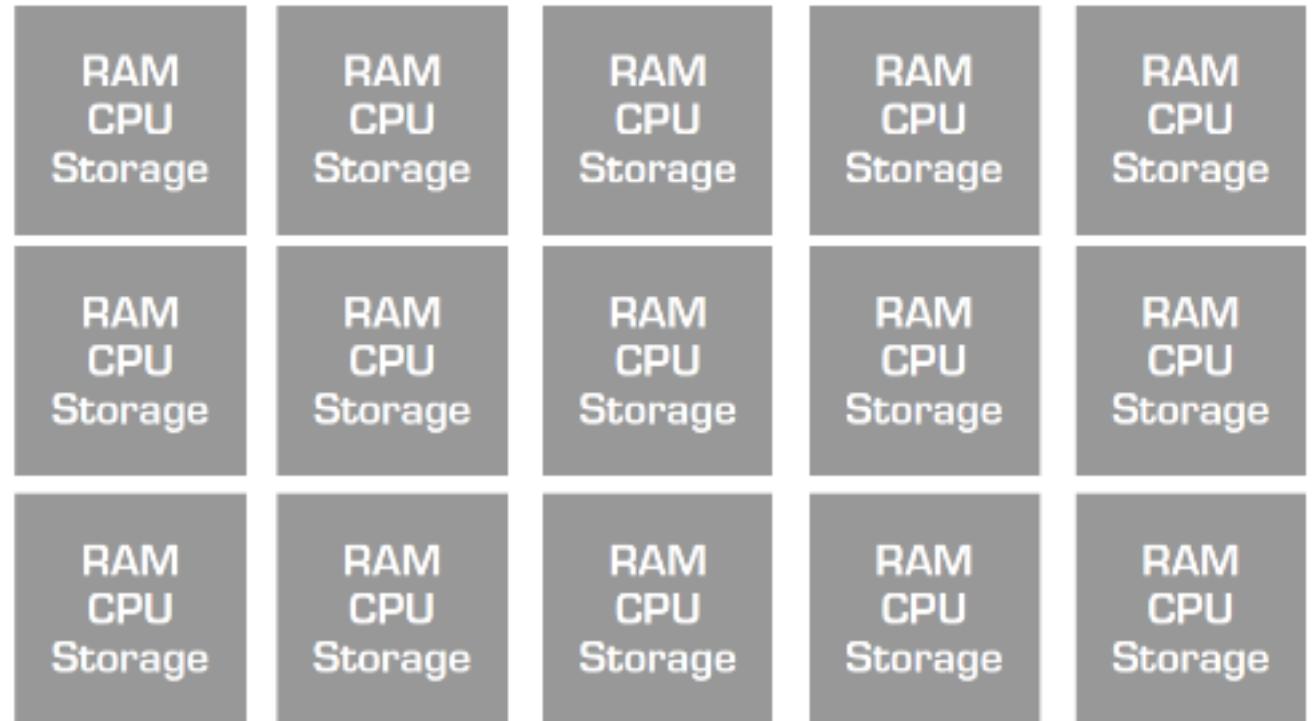
ALGORITHM

Big Data Solution (cont.)

Scale-out Infrastructure



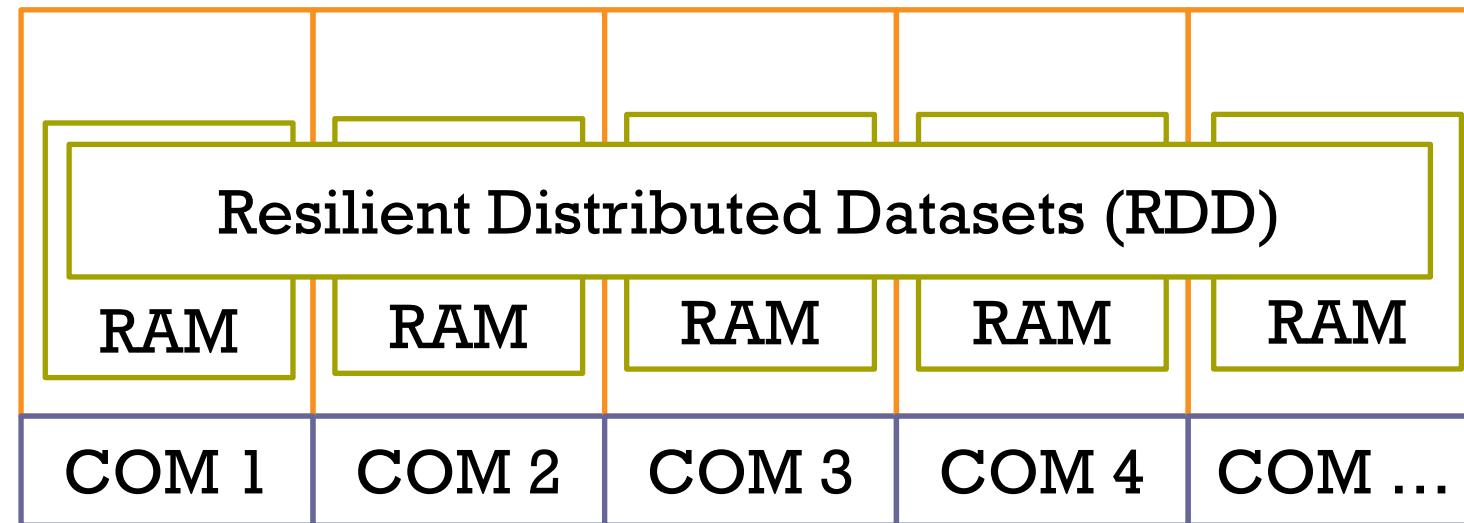
Vertical Scaling
(Scale-up)



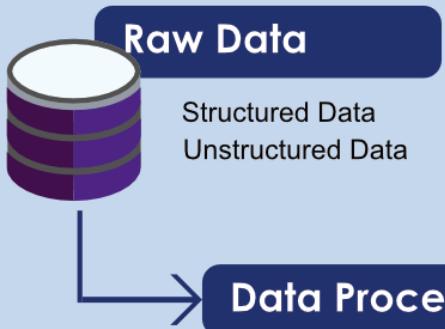
Horizontal Scaling
(Scale-out)

Big Data Solution (cont.)

In-memory & Distributed Computing



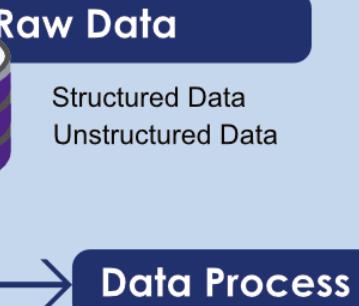
DATA ENGINEER



SQL
NoSQL
Python

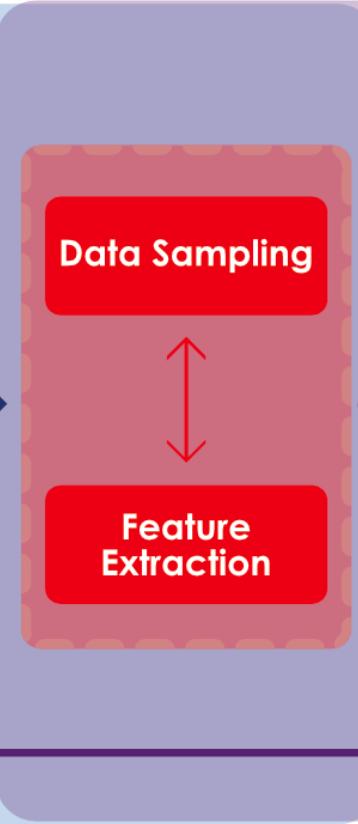
Hadoop
Spark

- Clustering
- Cleaning
- Structuring
- Streaming
- Extract ,Transform and Load (ETL)



Data Storage

DWH / Cloud

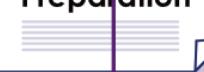


DATA SCIENTIST

Training Data



Data Preparation



Testing Data



Model Creation/ Machine Learning



Model Validation

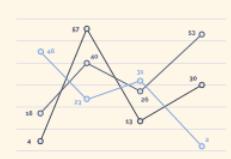
Final Model

DATA ANALYST

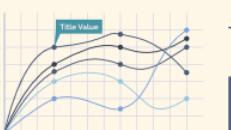
Data Visualization
BI
Application
Report
Action



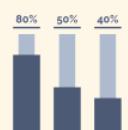
Nunc vitae eleifend massa. Quisque nisi ex, viverra ut velit a. efficitur cursus nisl. Integer pretium nisl. turpis. vitae segnits nisl. mattis luctus.



Your Title
Vivamus ultrices hendrerit neque erat tristique. Duis justo libero, hendrerit a turpis nec, dictum pellentesque massa. Nunc vitae eleifend massa.



Your Title
Proin efficitur turpis, vel pulvinar turpis, et tempus maximus nec. Lorem ipsum dolor sit amet.



Your Title
Proin efficitur turpis, vel pulvinar turpis, et tempus maximus nec. Lorem ipsum dolor sit amet.

BigData Ecosystem





<https://blog.datath.com/data-engineer-guide/>



Top Chef Thailand ตอนสุดท้าย ที่ผู้เข้าแข่งขันต้องช่วยกันทำงานเป็นทีม – ขอบคุณรูปจาก one31

Data Engineer ก็เหมือนกับผู้ช่วยเชฟ มีหน้าที่จัดเตรียมข้อมูลจากแหล่งต่าง ๆ มารวมกันไว้ในจุดเดียว โดยต้องทำให้ข้อมูลมีความถูกต้อง และดูและระบบทว่าทำงานได้ไม่เกิดปัญหาอะไร (ในชีวิตจริงนี่ต่อให้เราวางแผนมาดีแค่ไหน เจอข้อมูลเยอะ ๆ วันดีคืนดีก็ล้มได้ครับ T_T)



Data Science Process

+

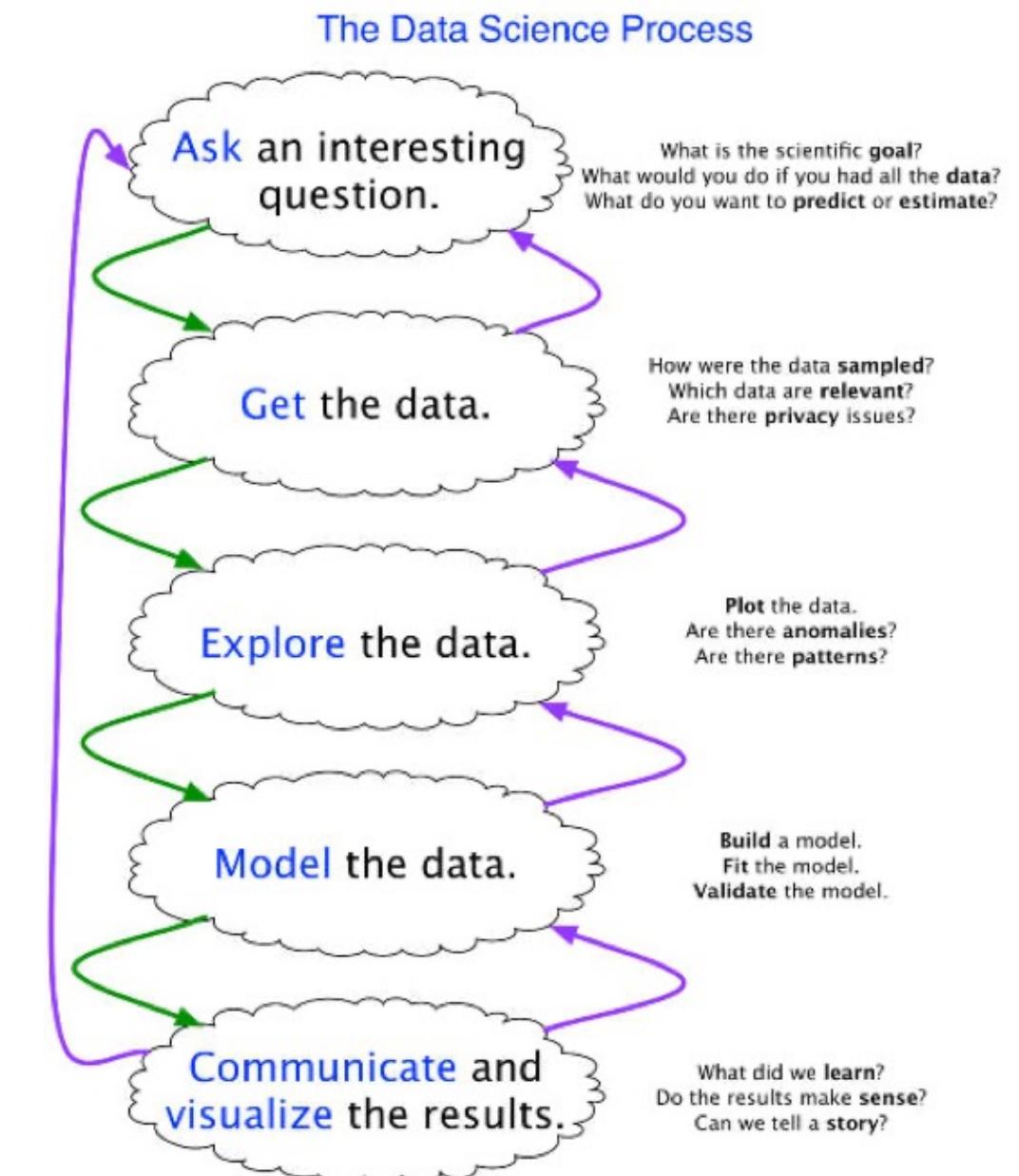
Data Science Process

Dr.Virote

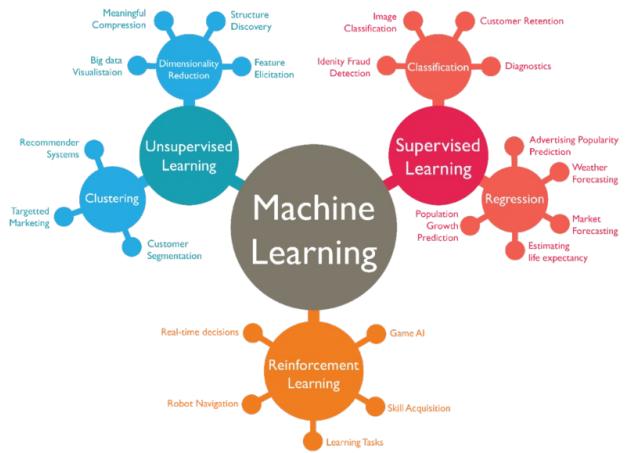
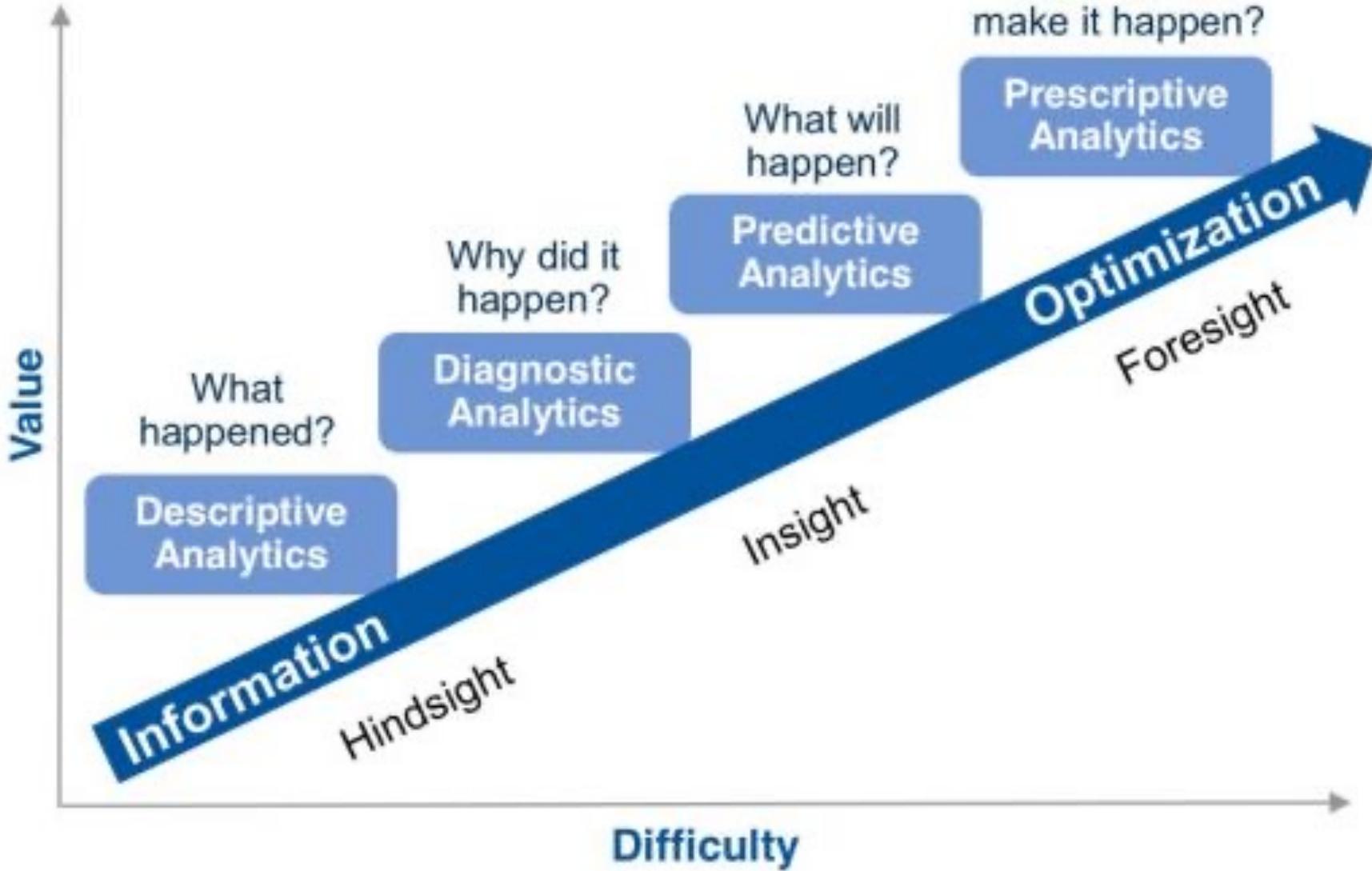
1. Transform data into **valuable insights**
2. Transform data into **data products**
3. Transform data into **interesting stories**

Aj.Natawut

1. Measurement (**decision**)
2. Insights (**knowledge**)
3. Data Products (**Innovation, Intelligent**)



Data Analytics



BIG DATA





Types of Data Science Projects

Valuable insights

- Data visualization
- Analytical skills & storytelling
- Infographic



Advanced analytics

- AI/Machine Learning/Deep Learning
- Prediction, Forecasting, Clustering, etc.



Ask a question about your data

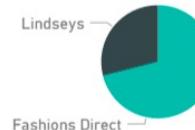
Total Stores
NEW & EXISTING STORES

104

This Year's Sales
NEW & EXISTING STORES

\$22M

This Year's Sales
BY CHAIN



New Stores, New Stores Target
YEAR TO DATE

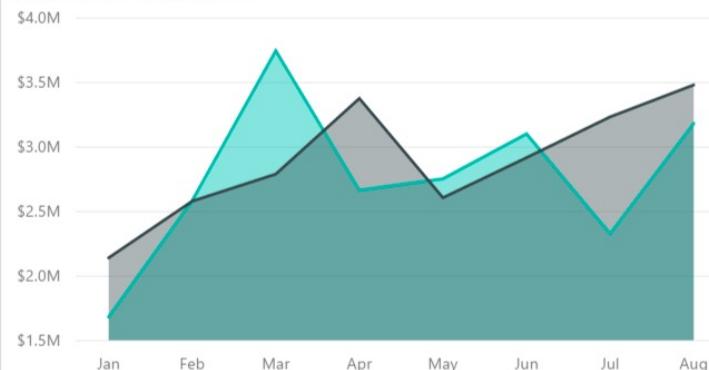


This Year's Sales
NEW STORES ONLY

\$2M

This Year's Sales, Last Year's Sales
BY FISCAL MONTH

This Year Sales Last Year Sales



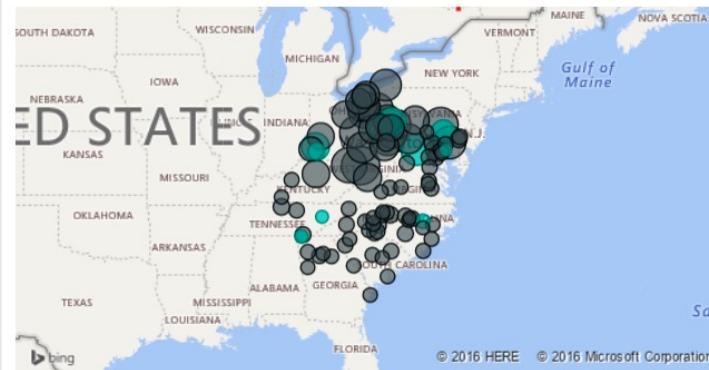
Total Sales Variance %, Sales Per Sq Ft, This Year's Sales
BY DISTRICT



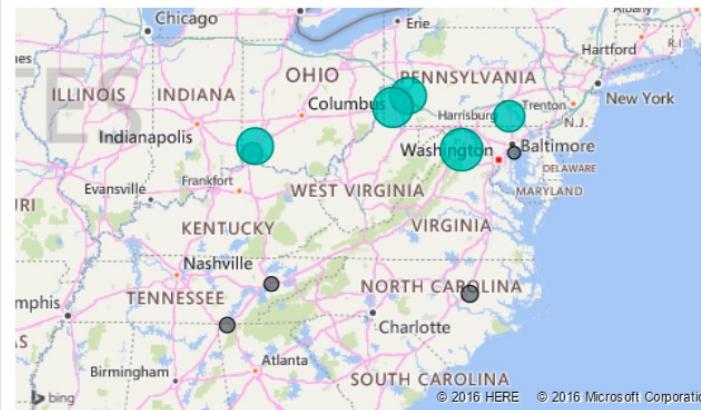
New Stores
NEW STORES ONLY

10

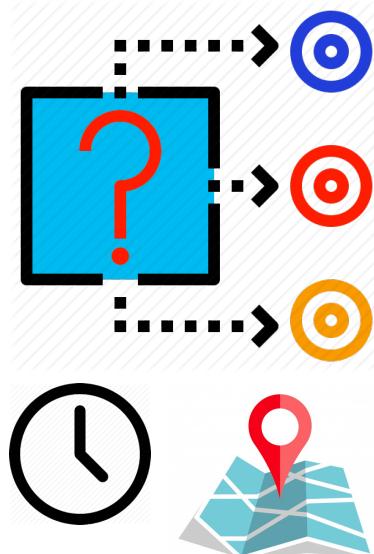
This Year's Sales
BY POSTAL CODE, STORE TYPE



This Year's Sales
BY CITY, CHAIN



Sales Per Sq Ft
BY NAME



Magic Quadrant for Analytics and Business Intelligence Platforms.



Source: Gartner (Feb 2019 and 2020)



Figure 1: Magic Quadrant for Analytics and Business Intelligence Platforms

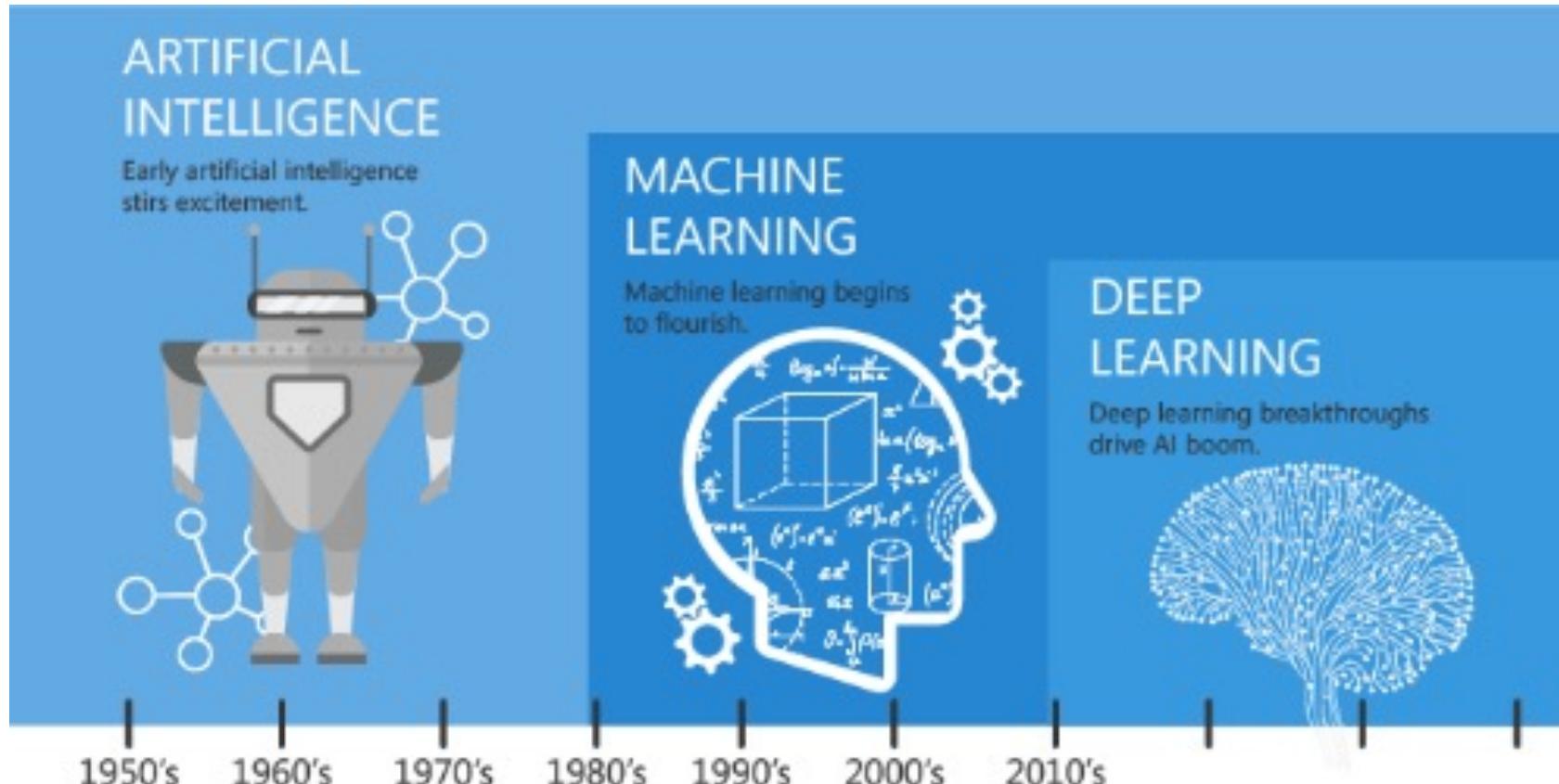


Source: Gartner (February 2021)



AI, Machine Learning, and Deep Learning

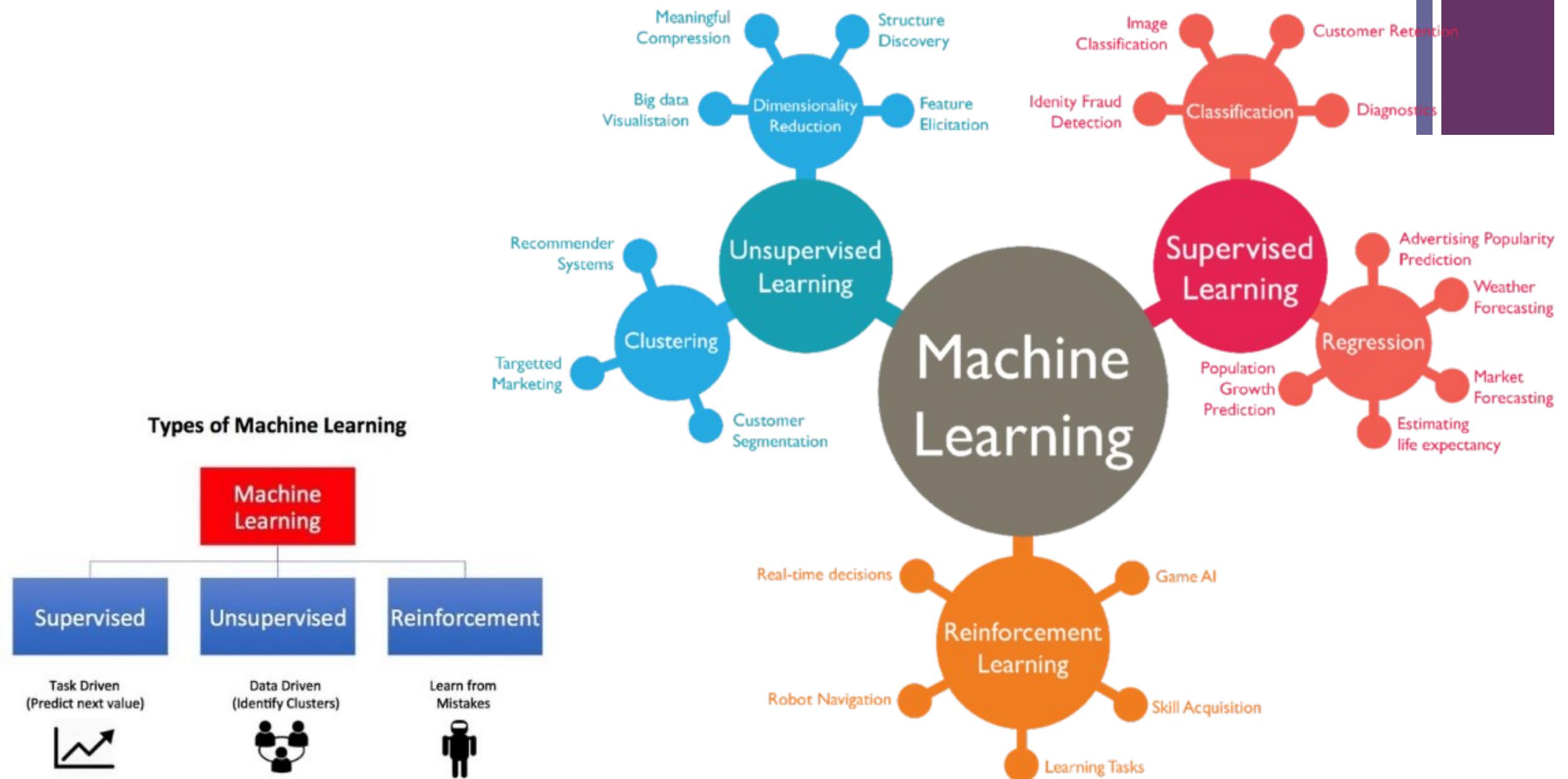
“Machine learning is the science of getting computers to act without being explicitly programmed.” — Stanford University
<https://towardsdatascience.com/cousins-of-artificial-intelligence-dda4edc27b55>



Since an early flush of optimism in the 1950's, smaller subsets of artificial intelligence - first machine learning, then deep learning, a subset of machine learning - have created ever larger disruptions.

+ Machine Learning (cont.)

60



Gartner's Top Strategic Technology Trends for 2023

Optimize

 **Digital Immune System**

 **Applied Observability**

 **AI TRiSM**

Scale

 **Industry Cloud Platforms**

 **Platform Engineering**

 **Wireless-Value Realization**

 **Sustainable Technology**

Pioneer

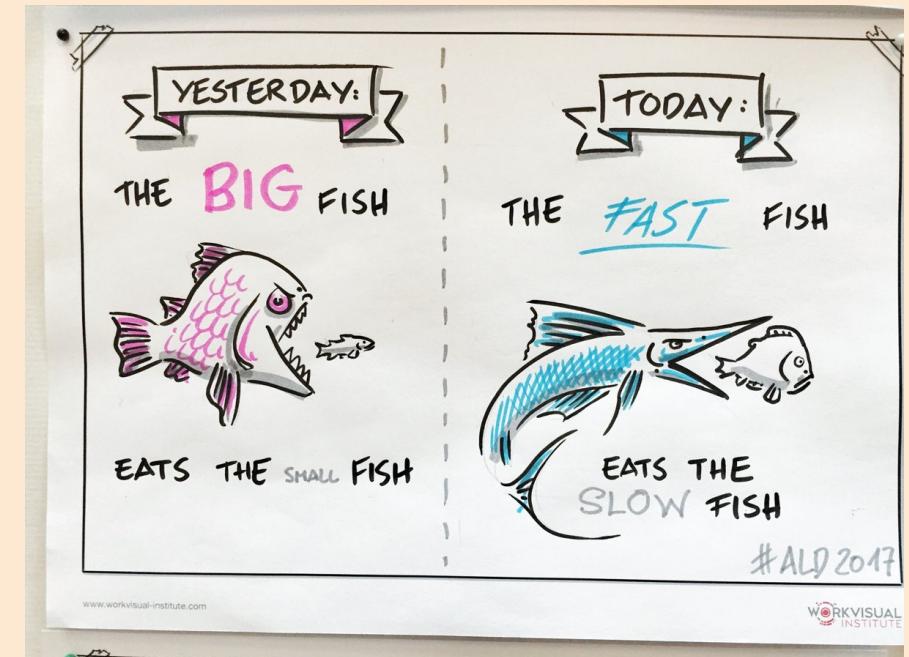
 **Superapps**

 **Adaptive AI**

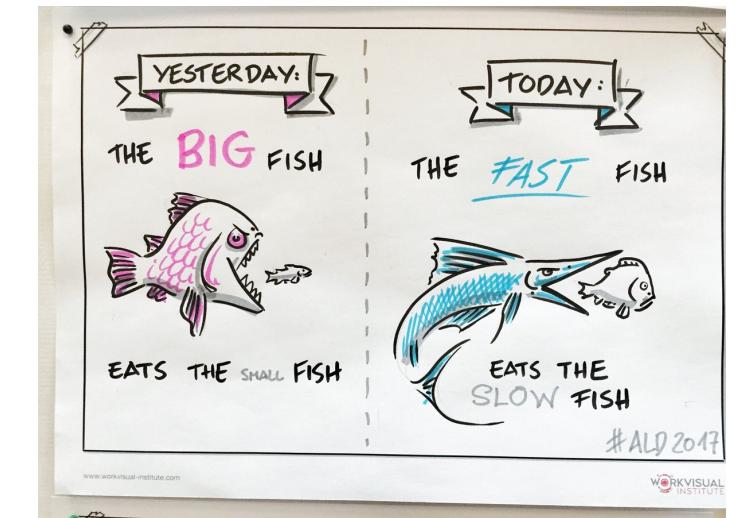
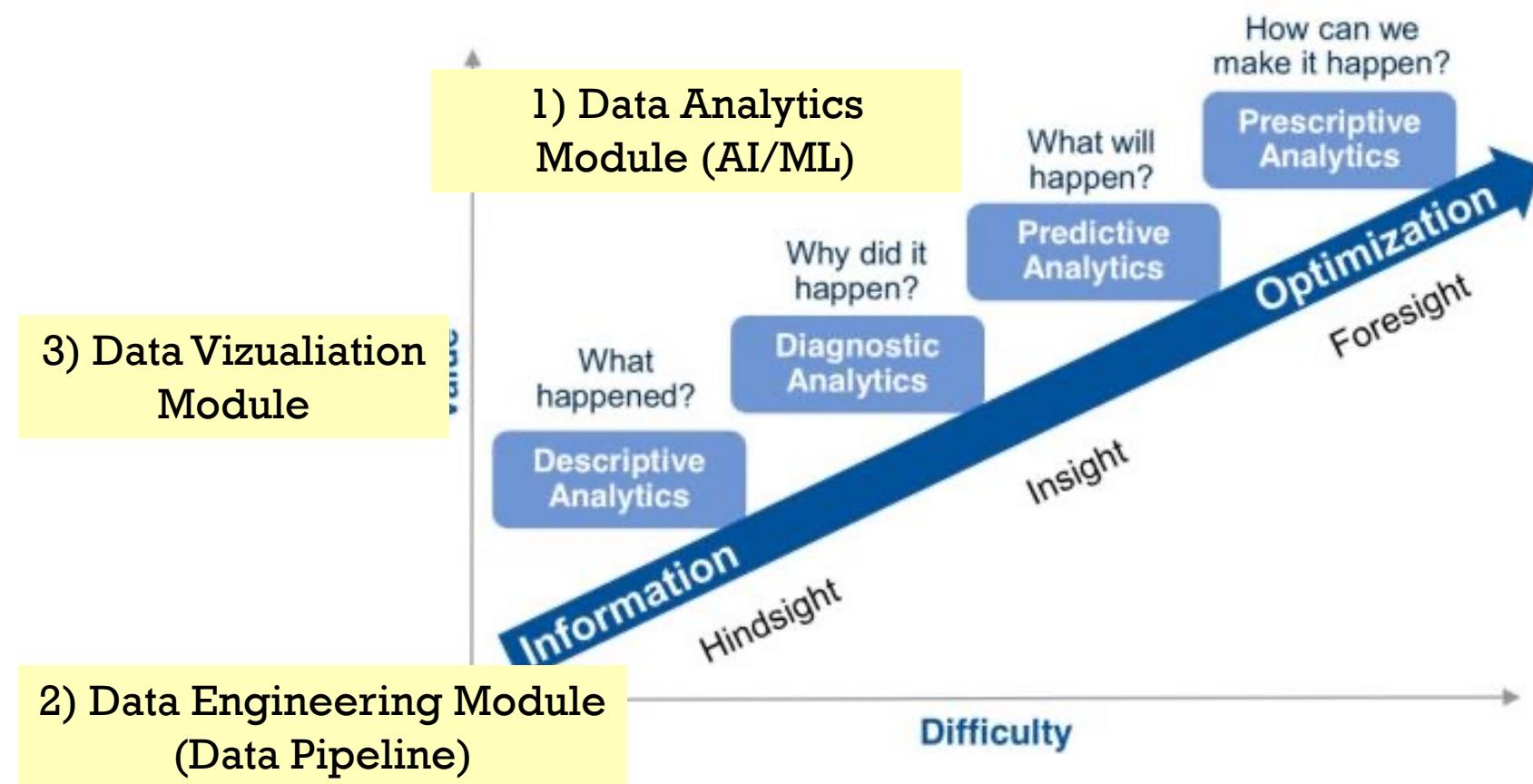
 **Metaverse**

Data Trend in 2023 (cont.)

- Big data is even bigger.
- AI is the key component.
- Knowledge without action is meaningless.
- Fast, Flexible, Scalable!



Conclusion





Any questions? 😊