Data Science Bootcamp

Kick Off - Intro to Data Science

Big Data

https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation#

- Volume
- Velocity
- Variety

5 MB hard disk drive - 1956

1 TB micro SD card - 2020

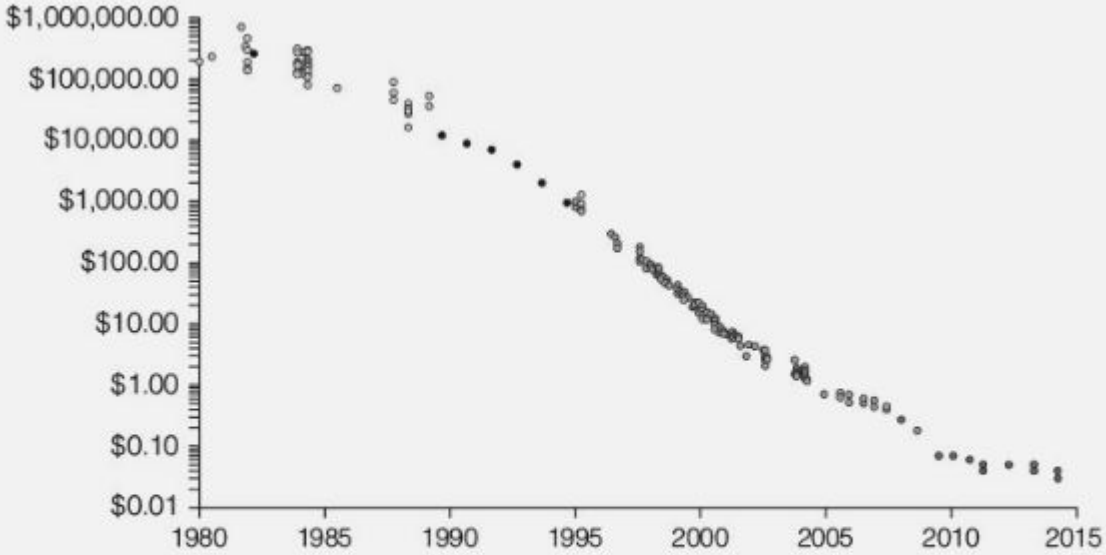


Figure 1.3 Historic cost of disk storage per GB (log scale).[9]

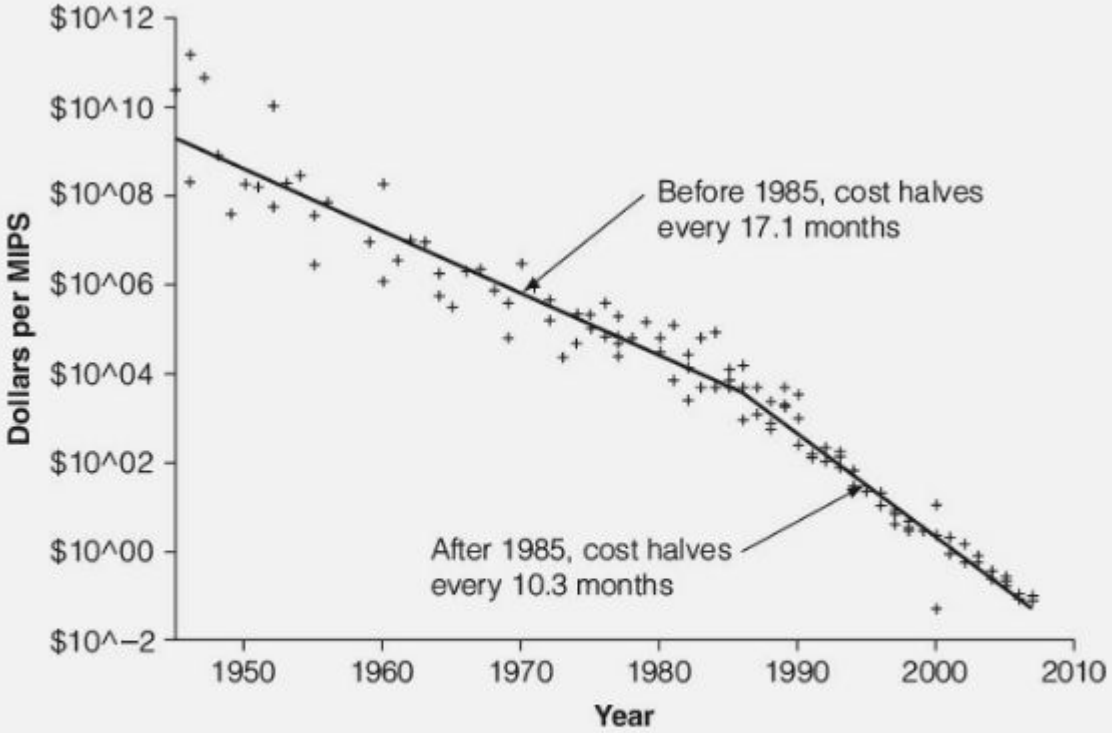


Figure 1.5 Historic cost of processing power (log scale).[10]

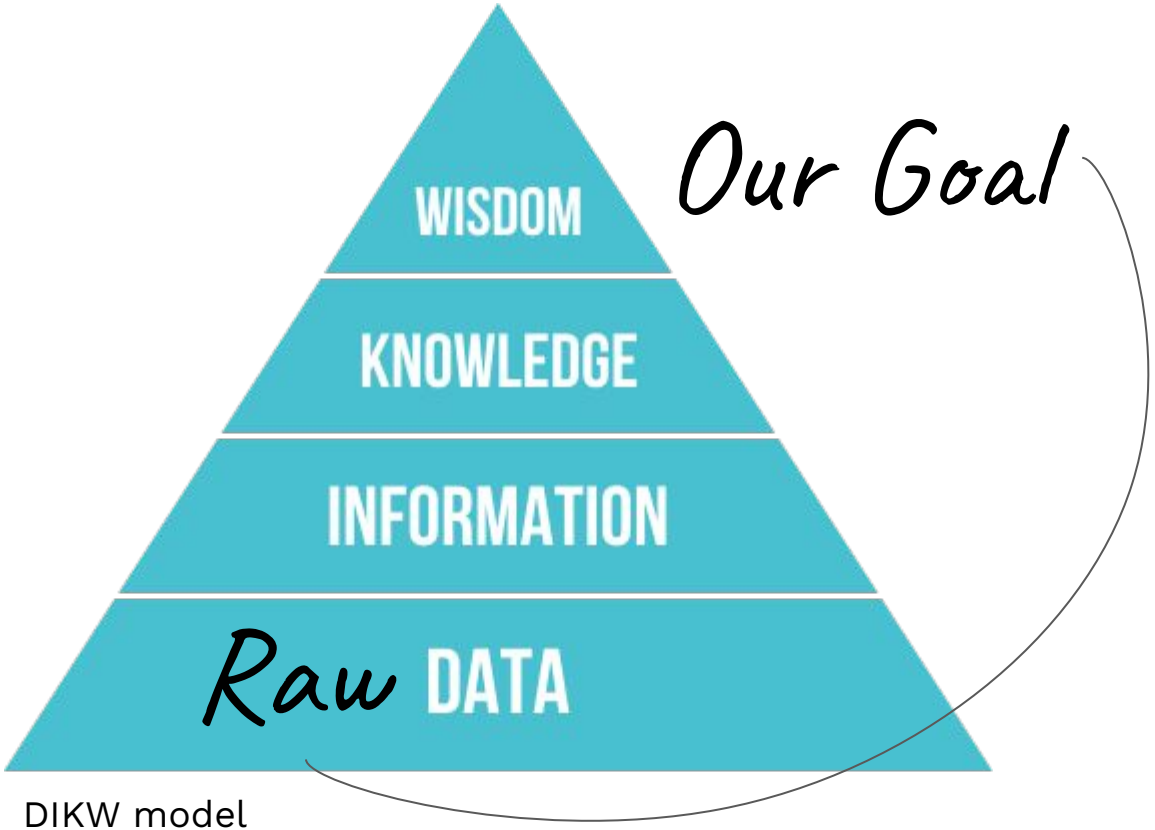
Big Data คือปัจจัยที่ทำให้ Data Skills กลายเป็นสิ่งสำคัญ

Data analytics is the **Science** of analyzing raw data in order to make conclusions about that information.

DA helps a business optimize its performance.

https://www.investopedia.com/terms/d/data-analytics.asp

Cassie Kozyrkov
Chief Decision Scientist @Google



DIKW model

Case Study

01 - Target

Case Study

01 - Target

Feb 16, 2012, 11:02am EST

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

Kashmir Hill Former Staff

Tech

Welcome to The Not-So Private Parts where technology & privacy collide

https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/#4e35ca8c6668

แค่นี้ Target ก็สามารถระบุสินค้าที่มีการซื้อเยอะๆในกลุ่มผู้หญิงมีครรภ์ได้แล้ว ในทาง data science เราเรียกสินค้า 25 ตัวนั้นว่า signal ที่เราใช้ทำ prediction (หรือ “The force is strong” ใน star wars ไม่เกี่ยว!)

	A	B	C	D	E	F	G
1	โลชั่น	วิตามิน	แคลเซียม	ซิงค์	สำลี	...	ตั้งครรภ์
2	0	1	1	1	0	...	ใช่
3	0	0	1	0	0	...	ไม่ใช่
4	1	1	1	0	1	...	ใช่
5	1	1	0	1	1	...	ใช่
6	1	0	0	1	0	...	ไม่ใช่

ตัวอย่างข้อมูลที่ Target ใช้สร้าง pregnancy score

Pregnancy Score = f(โลชั่นแบบไม่มีกลิ่น, วิตามิน, แคลเซียม, ซิงค์, สำลี, ...)

https://datarockie.com/2019/08/10/case-study-target-supermarket/

ผลการทำนาย

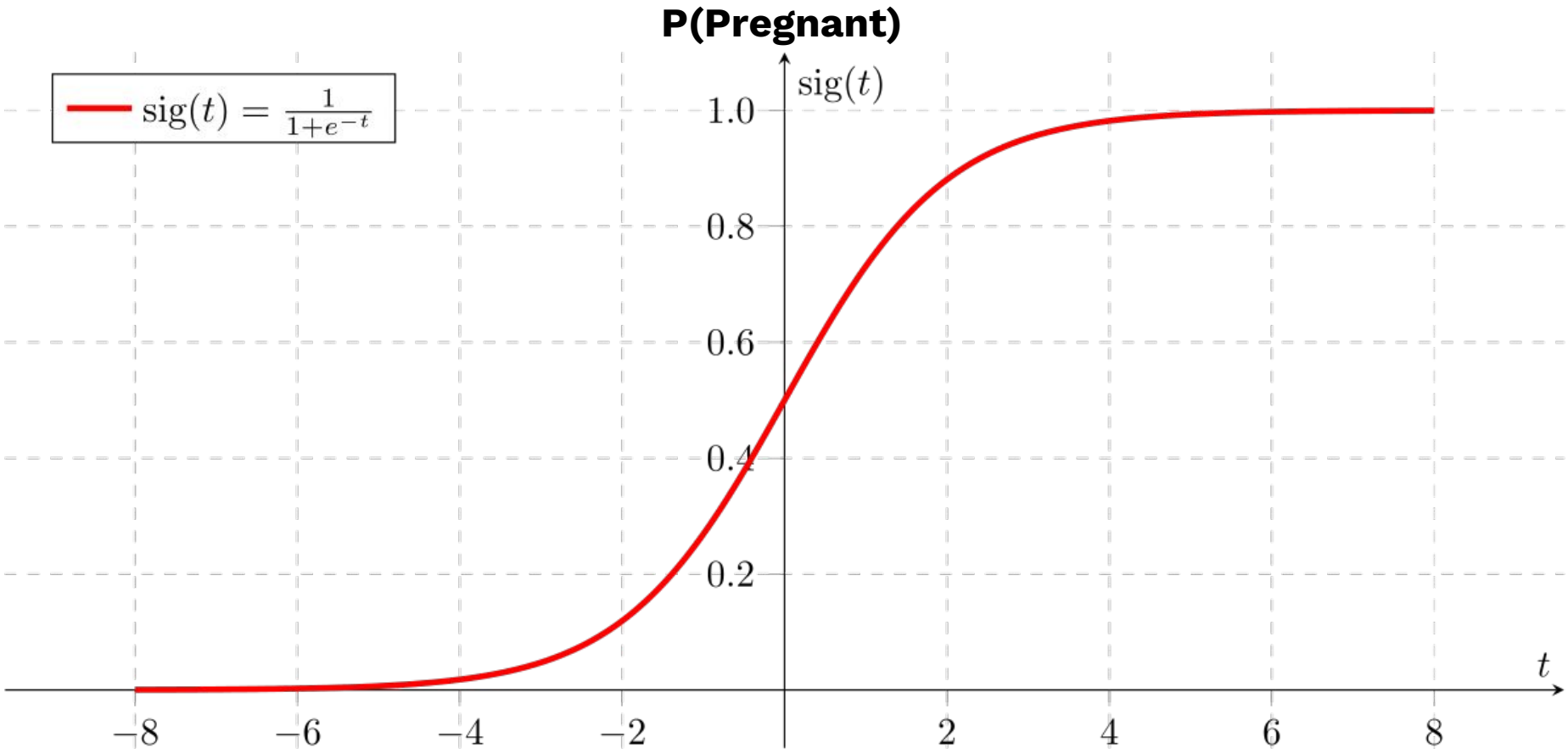
	A	B	C	D	E	F	Probability
1	โลชั่น	วิตามิน	แคลเซียม	ซิงค์	สำลี	...	
2	0	1	1	1	0	...	60%
3	0	0	1	0	0	...	88%
4	1	1	1	0	1	...	25%
5	1	1	0	1	1	...	39%
6	1	0	0	1	0	...	92%

ข้อมูลลูกค้าใหม่ (new data)

ลูกค้าที่เราจะ
ส่งคูปองไปให้

	A	B	C	D	E	F	Probability	Decision
1	โลชั่น	วิตามิน	แคลเซียม	ซิงค์	สำลี	...		
2	0	1	1	1	0	...	60%	**Yes**
3	0	0	1	0	0	...	88%	**Yes**
4	1	1	1	0	1	...	25%	No
5	1	1	0	1	1	...	39%	No
6	1	0	0	1	0	...	92%	**Yes**

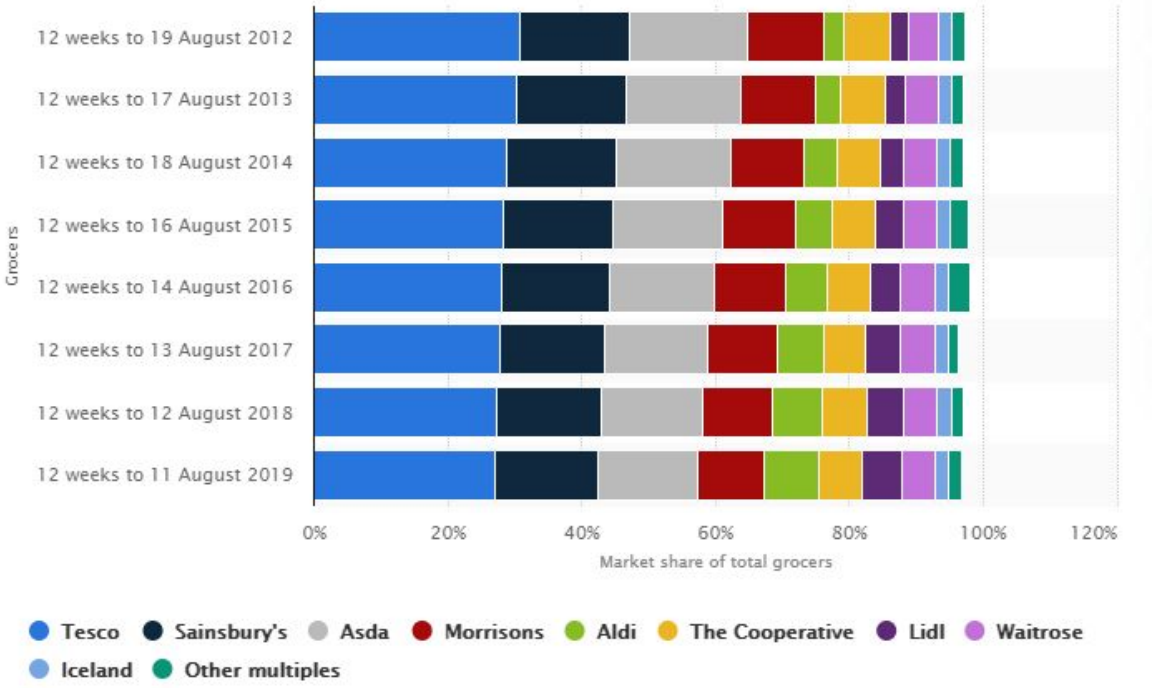
ข้อมูลลูกค้าใหม่ (new data)


P(Pregnant)
$\mathrm{sig}(t) = \frac{1}{1+e^{-t}}$
sig(t)
1.0
0.8
0.6
0.4
0.2
t
−8
−6
−4
−2
2
4
6
8

Case Study

02 - Tesco

Tesco Market Share 27% in 2019



https://www.statista.com/statistics/300656/grocery-market-share-in-great-britain-year-on-year-comparison/

Tesco/ Dunnhumby - The very first company who uses data analytics and science with commercially success

ID	Amount	Store	Customer

History [edit]

In 1993, Terry Leahy asked the Tesco marketing team to investigate the potential of loyalty cards. In the past Tesco had run Green Shield Stamps as a promotional tool which rewarded people for visits and spend but gained no customer information. The initial team, led by Grant Harrison, researched programmes across the world and developed a proposal which showed that a loyalty card could be very effective. The key change since the days of Green Shield Stamps was the ability to track individual customer behaviour cost-effectively using a magnetic stripe card.

In 1994, Harrison attended a conference where Clive Humby from marketing firm dunnhumby was speaking. Dunnhumby was already working with clients such as Cable & Wireless and BMW, and Harrison approached them to help with the loyalty card project.[2] Successful trials throughout 1994 led to the Tesco board asking Harrison and Humby to present to the annual Board strategy session.

The first response from the board came from Tesco's then chairman Lord MacLaurin, who said, "What scares me about this is that you know more about my customers after three months than I know after 30 years."[3]

Strategic
Data
Acquisition

Case Study

03 - Netflix

2009
NETFLIX
DATE: 09·21·09
PAY TO THE ORDER OF: BellKor's Pragmatic Chaos
$ 1,000,000 00
AMOUNT: ONE MILLION
00/100
FOR: The Netflix Prize
Reed Hastings

Apr 16, 2012, 05:53pm EDT

What the Failed $1M Netflix Prize Says About Business Advice

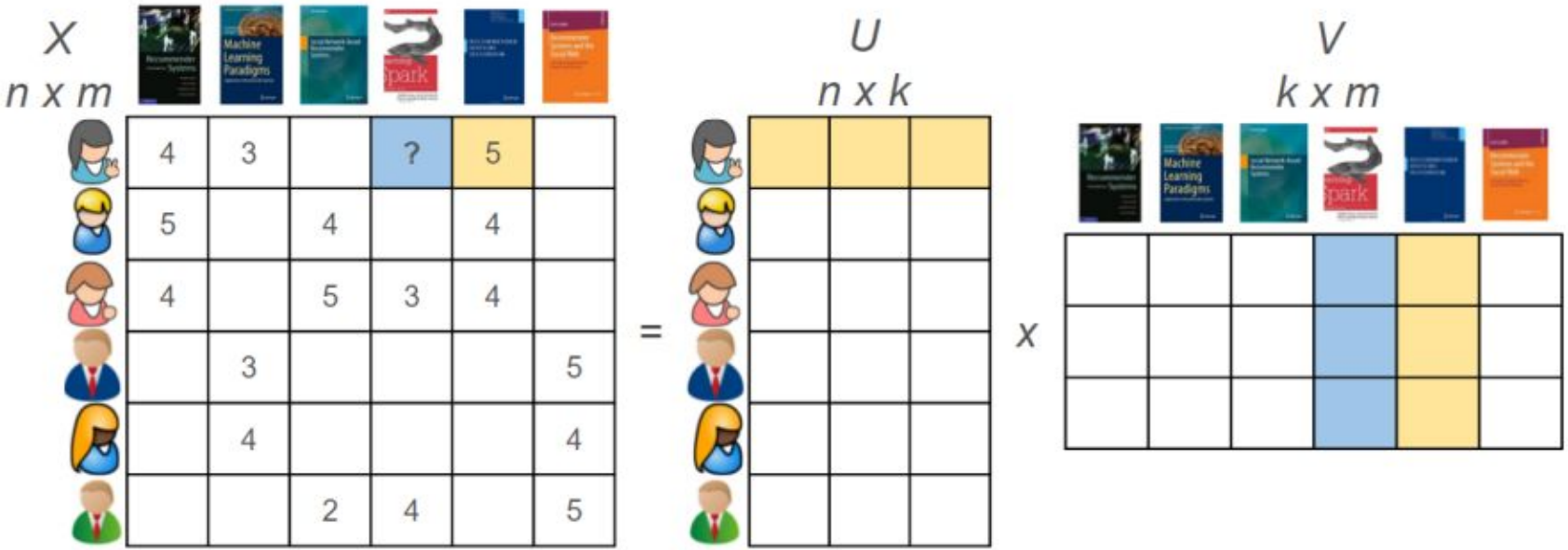
Ryan Holiday Contributor

CMO Network - OLD

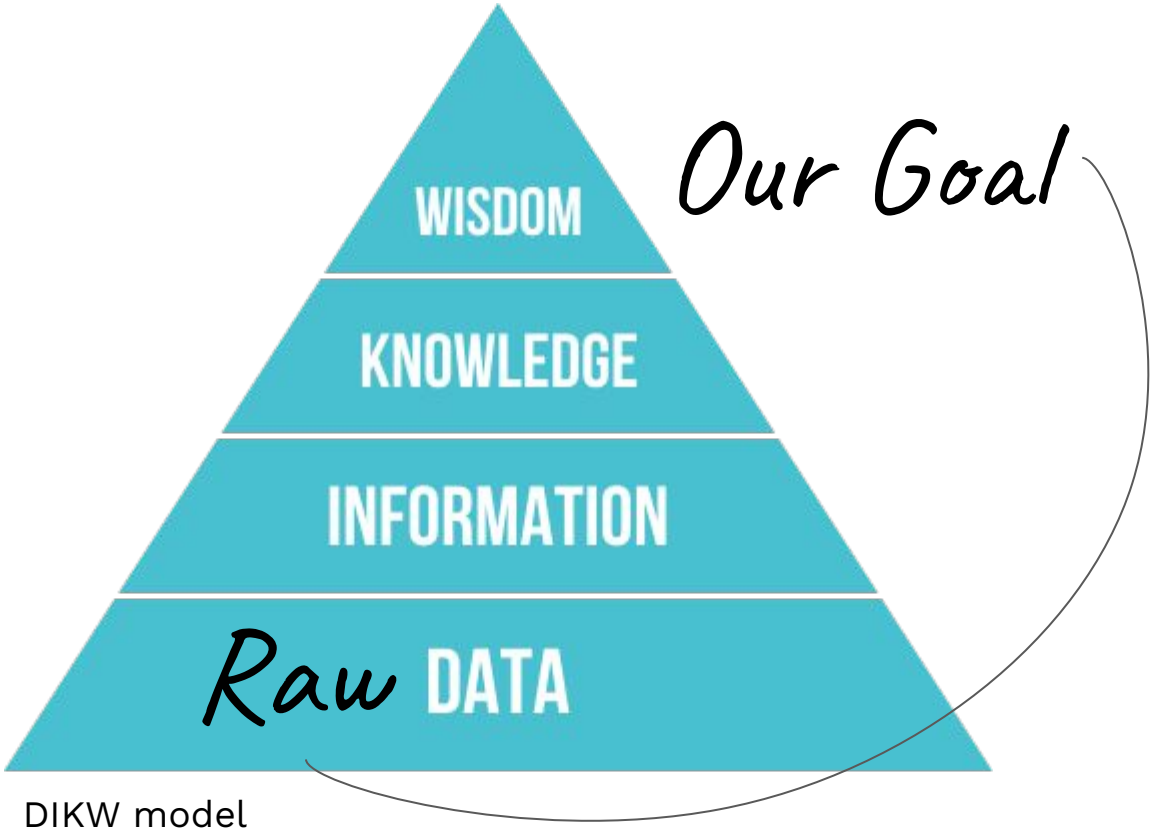
An inside look at the hidden side of marketing, PR and strategy.

https://www.forbes.com/sites/ryanholiday/2012/04/16/what-the-failed-1m-netflix-prize-tells-us-about-business-advice/#e24108a73c9a

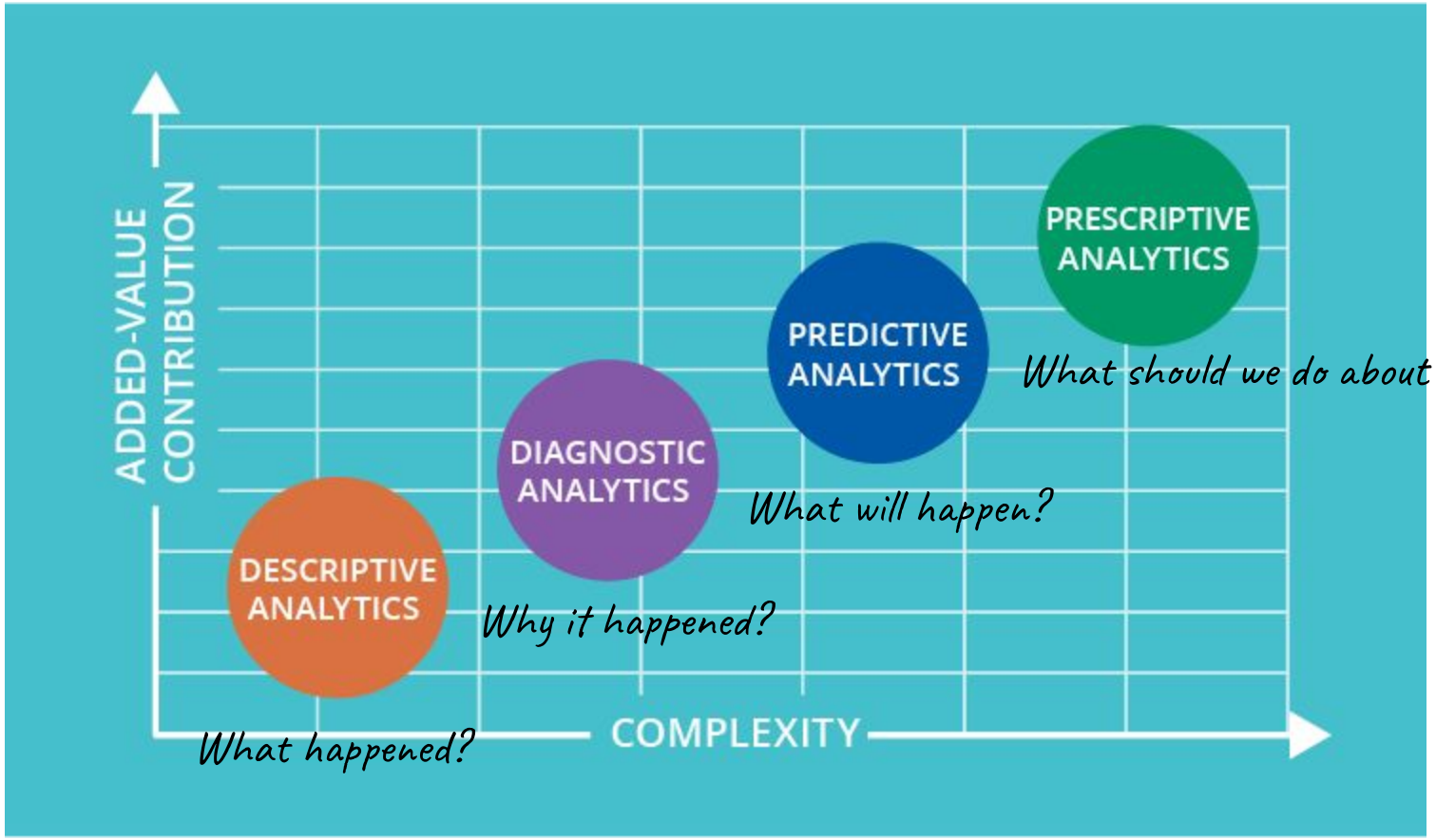
Collaborative Filtering = Recommender Systems

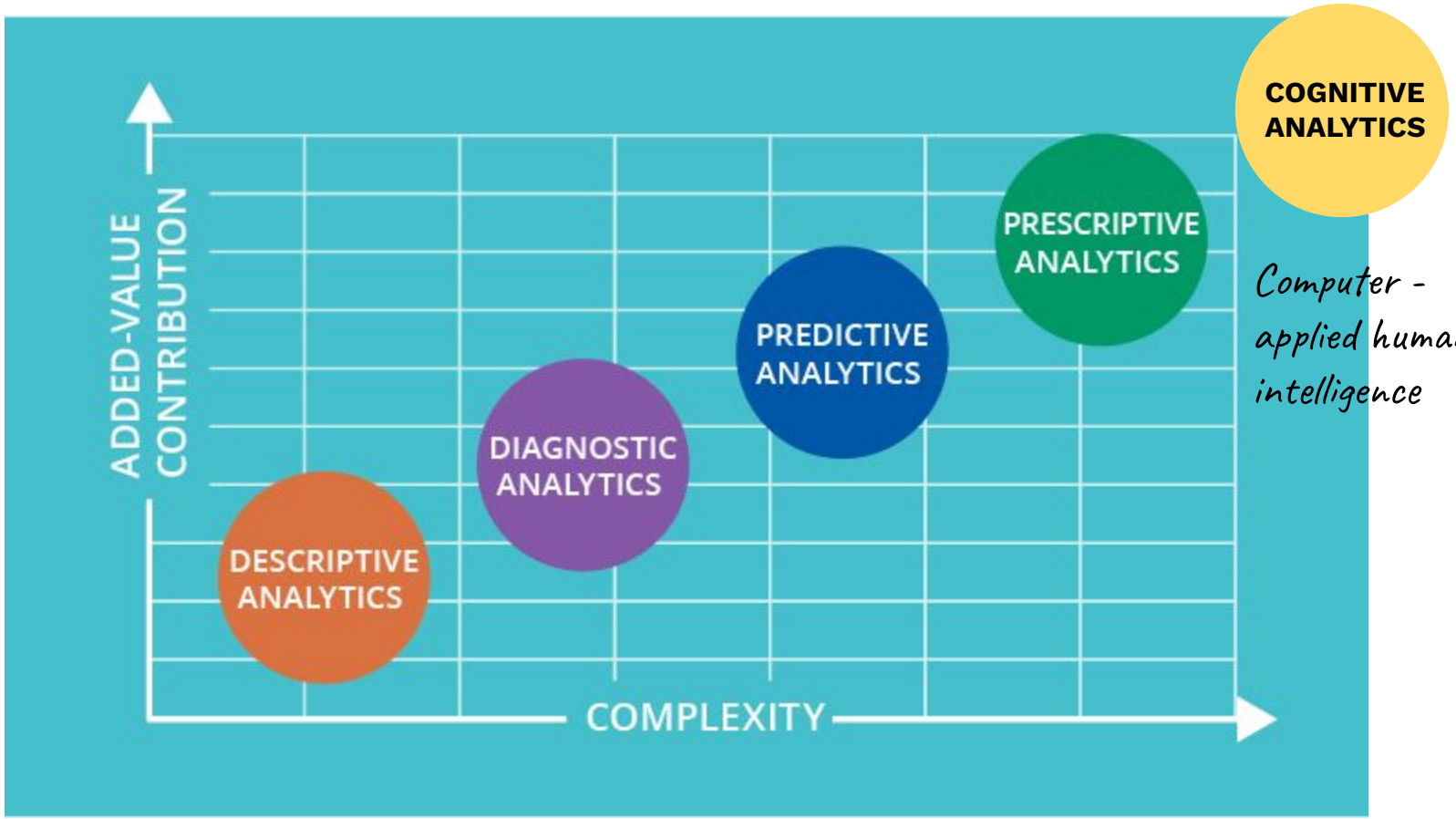


https://heartbeat.fritz.ai/recommender-systems-with-python-part-iii-collaborative-filtering-singular-value-decomposition-5b5dcb3f242b



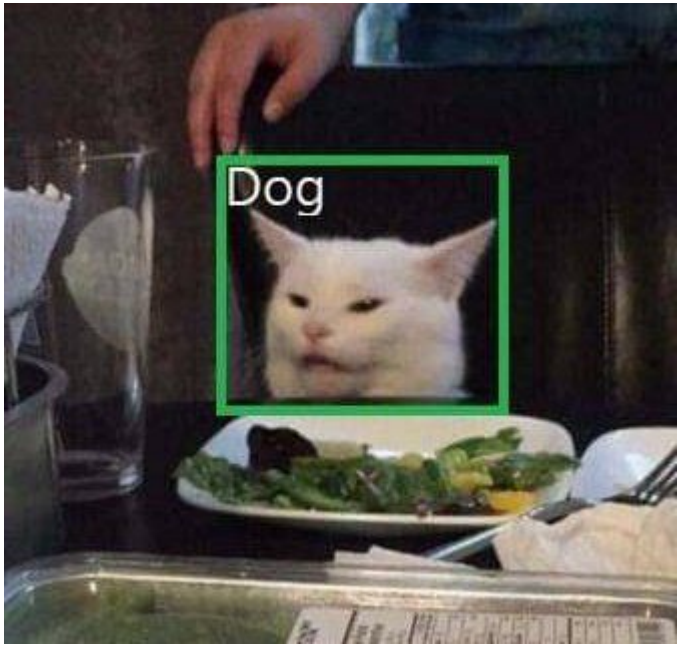
DIKW model

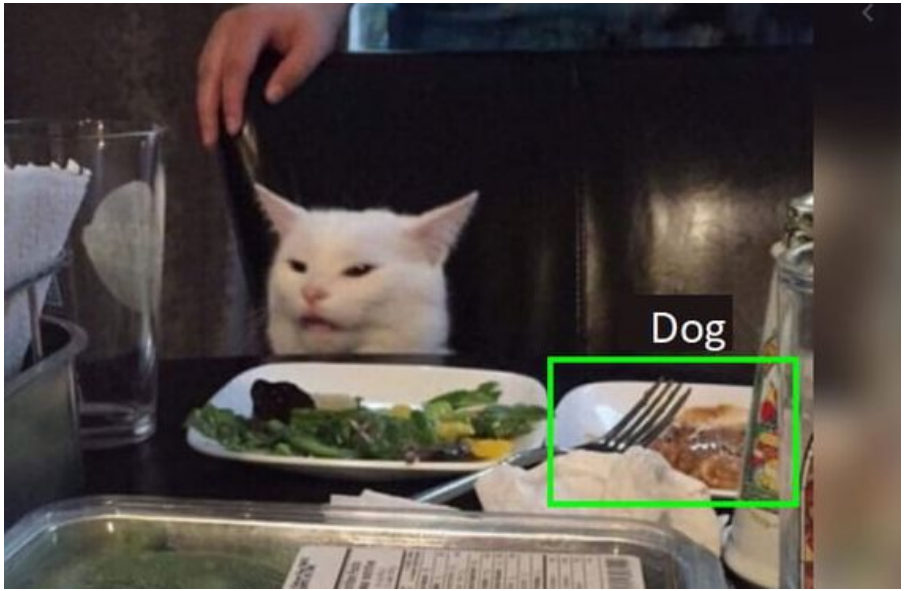

ADDED-VALUE CONTRIBUTION
COMPLEXITY
DESCRIPTIVE ANALYTICS
What happened?
DIAGNOSTIC ANALYTICS
Why it happened?
PREDICTIVE ANALYTICS
What will happen?
PRESCRIPTIVE ANALYTICS
What should we do about it?



Computer - applied human intelligence

AGI vs. ANI

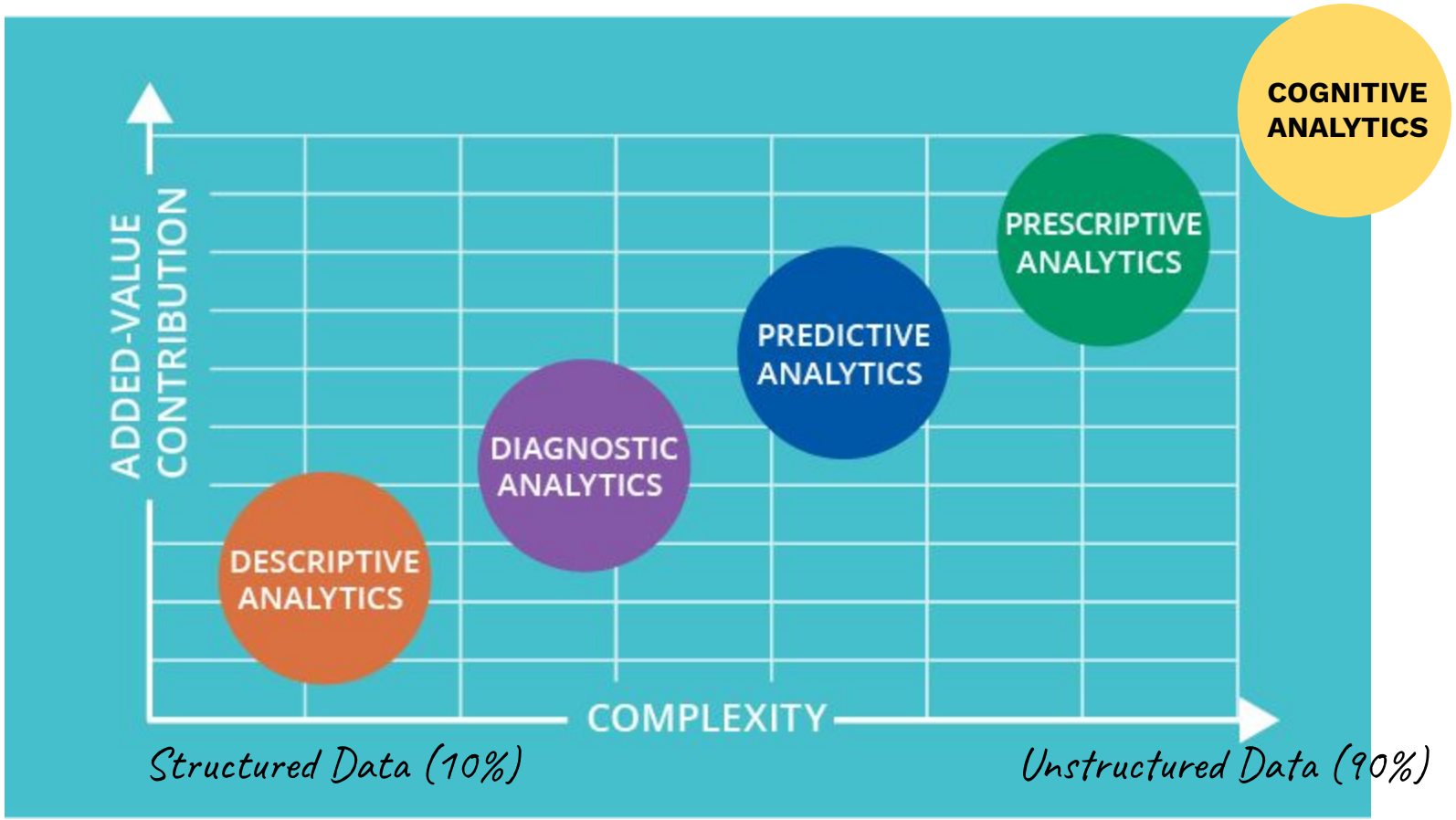

Dog

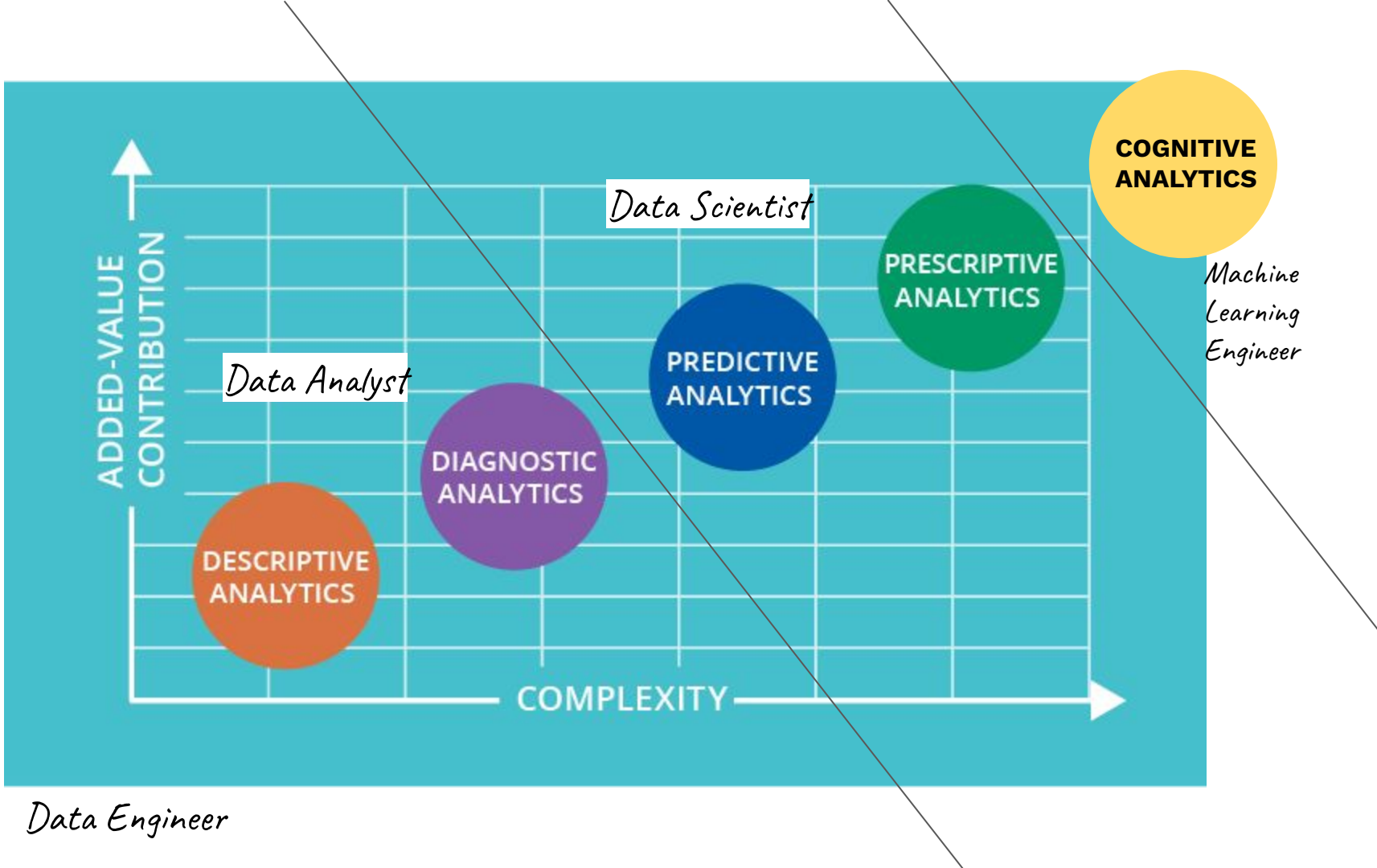

Dog

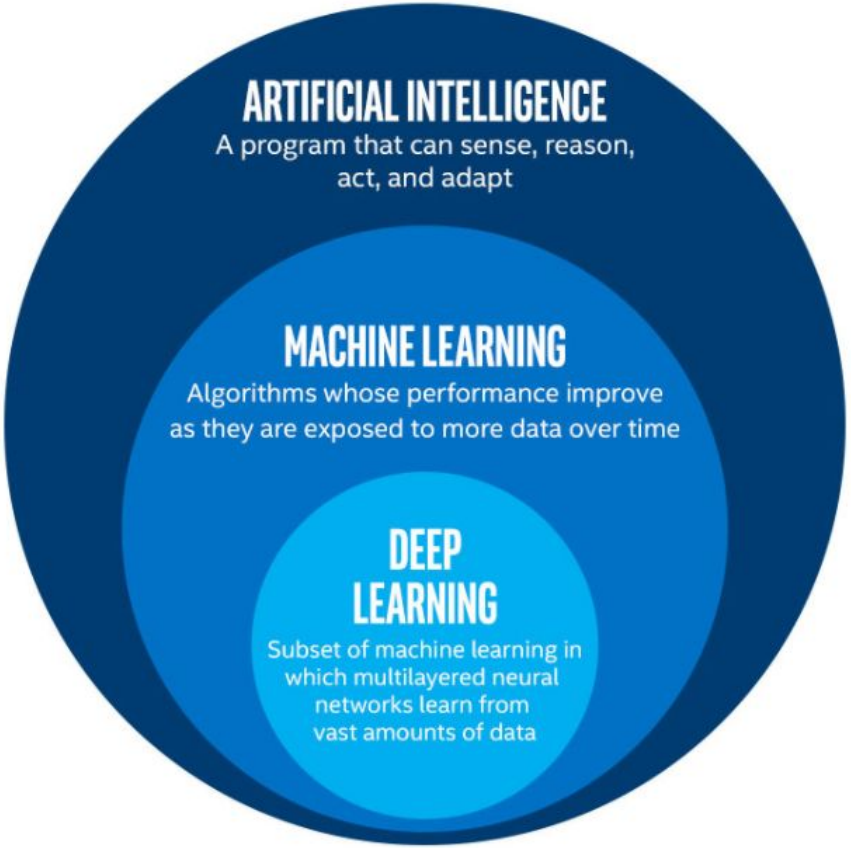
WIRED STAFF 06.26.12 11:15 AM

Google's Artificial Brain Learns to Find Cat Videos

https://www.wired.com/2012/06/google-x-neural-network/

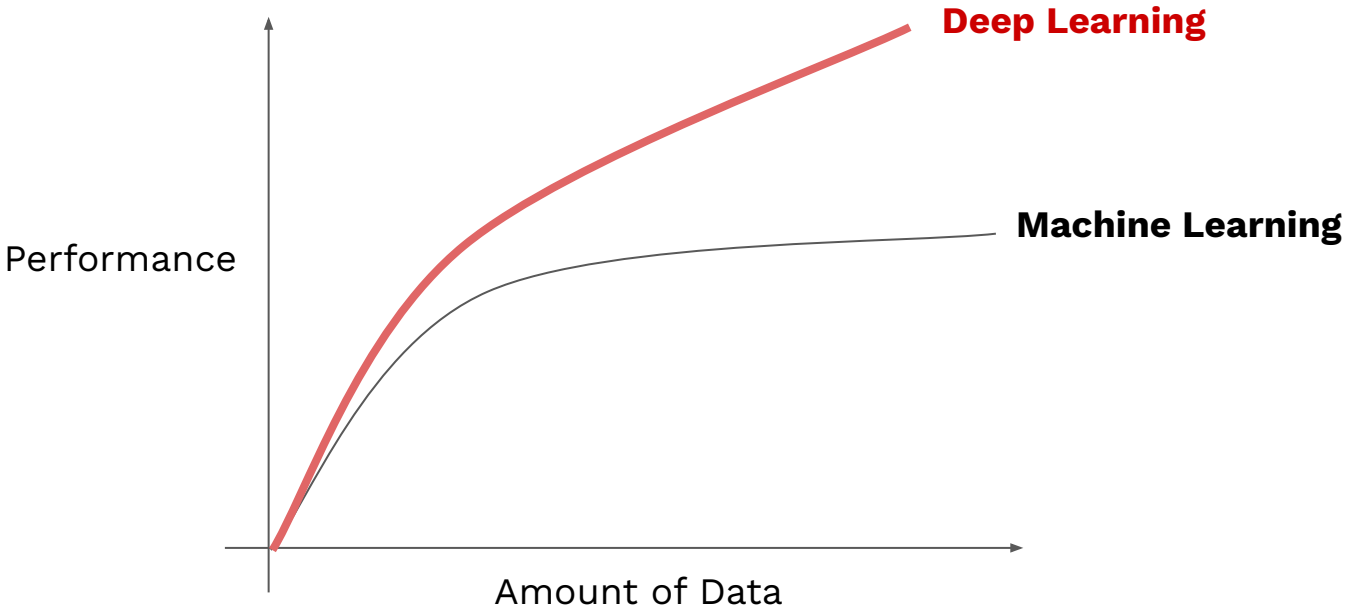

COGNITIVE ANALYTICS
PRESCRIPTIVE ANALYTICS
PREDICTIVE ANALYTICS
DIAGNOSTIC ANALYTICS
DESCRIPTIVE ANALYTICS
ADDED-VALUE CONTRIBUTION
COMPLEXITY
Structured Data (10%)
Unstructured Data (90%)

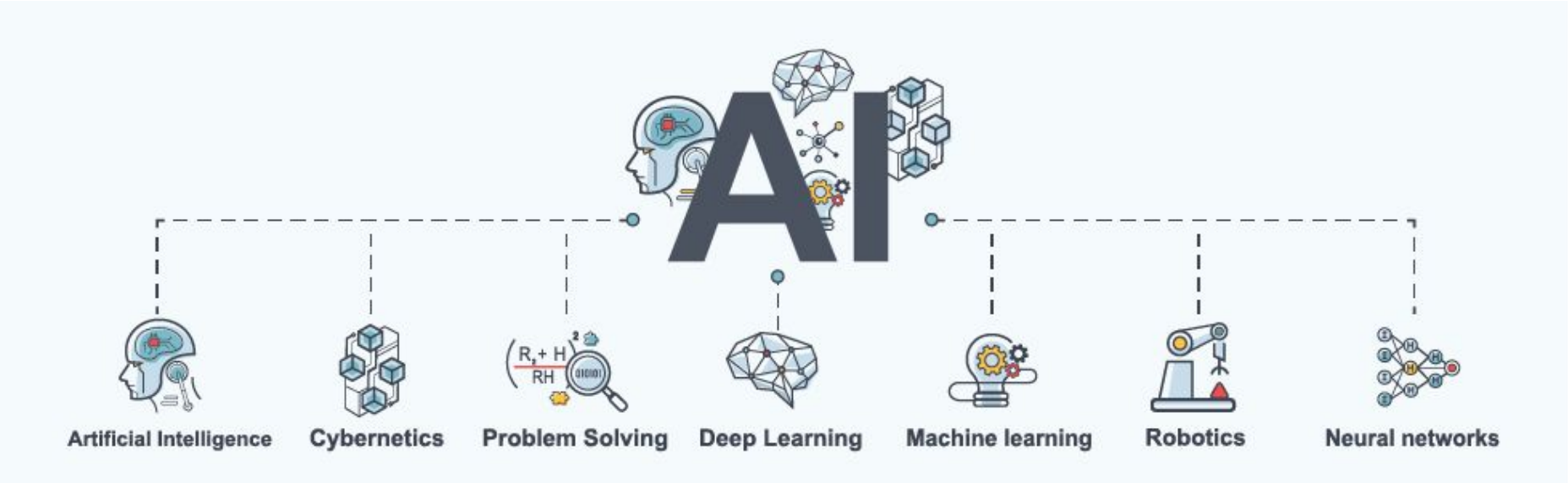

COGNITIVE ANALYTICS
Machine Learning Engineer
Data Scientist
PRESCRIPTIVE ANALYTICS
PREDICTIVE ANALYTICS
Data Analyst
DIAGNOSTIC ANALYTICS
DESCRIPTIVE ANALYTICS
ADDED-VALUE CONTRIBUTION
COMPLEXITY
Data Engineer


ARTIFICIAL INTELLIGENCE
A program that can sense, reason, act, and adapt
MACHINE LEARNING
Algorithms whose performance improve as they are exposed to more data over time
DEEP LEARNING
Subset of machine learning in which multilayered neural networks learn from vast amounts of data

Traditional Programming

Machine Learning

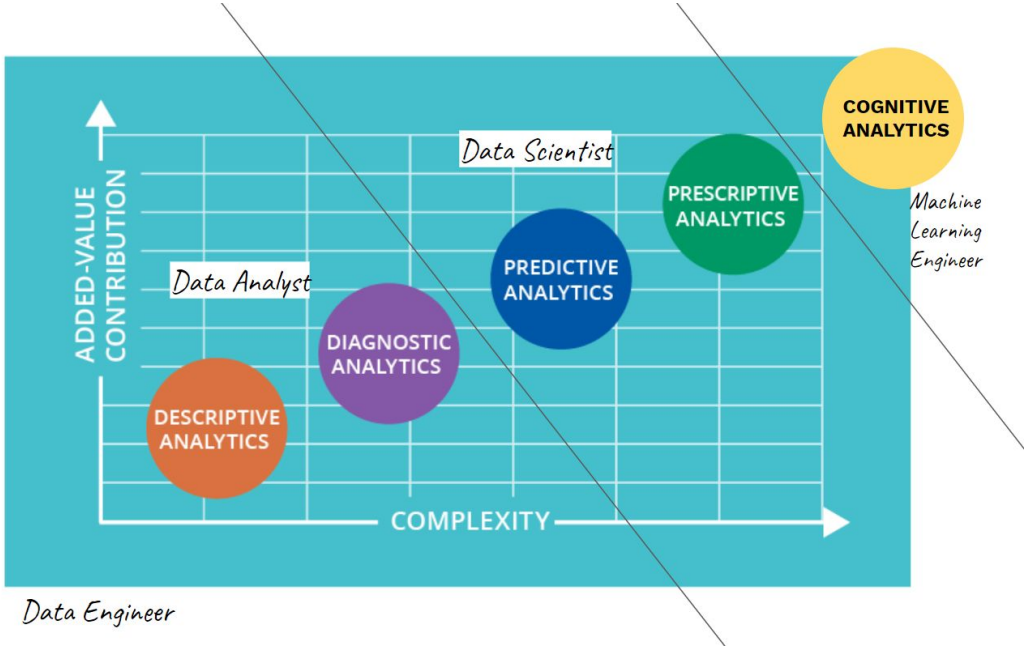

Deep Learning
Machine Learning
Performance
Amount of Data



https://blockruption.com/artificial-intelligence-and-machine-learning/?lang=en

Data Analytics is the whole process to turn raw data into products/ decisions

Using science and technology



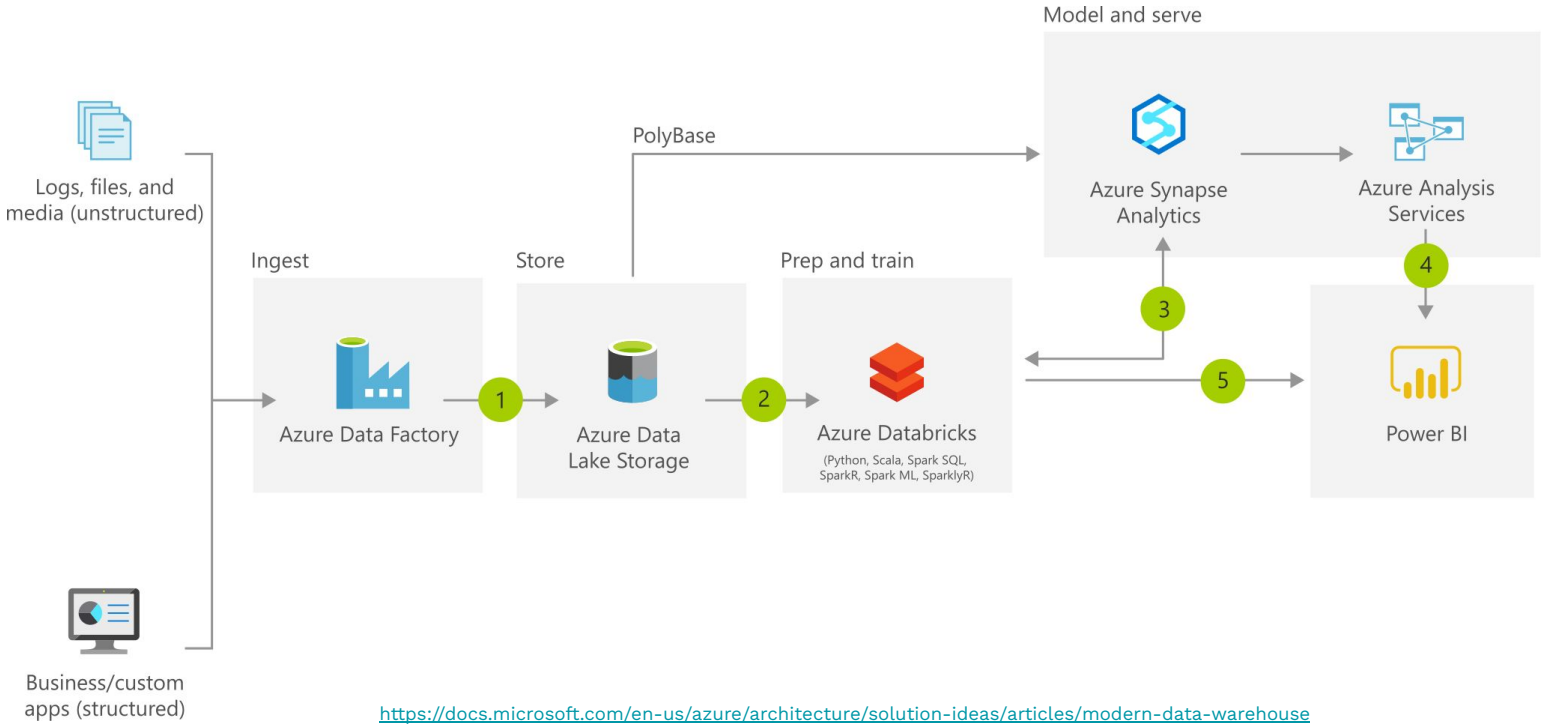
AWS Certified Data Analytics - Specialty

Validate your ability to design, build, secure, and maintain analytics solutions on AWS that are efficient, cost-effective, and secure

Schedule an exam

ออกแบบ สร้าง ดูแลระบบ Data Analytics (Process)

Modern Data Warehouse Architecture



https://docs.microsoft.com/en-us/azure/architecture/solution-ideas/articles/modern-data-warehouse

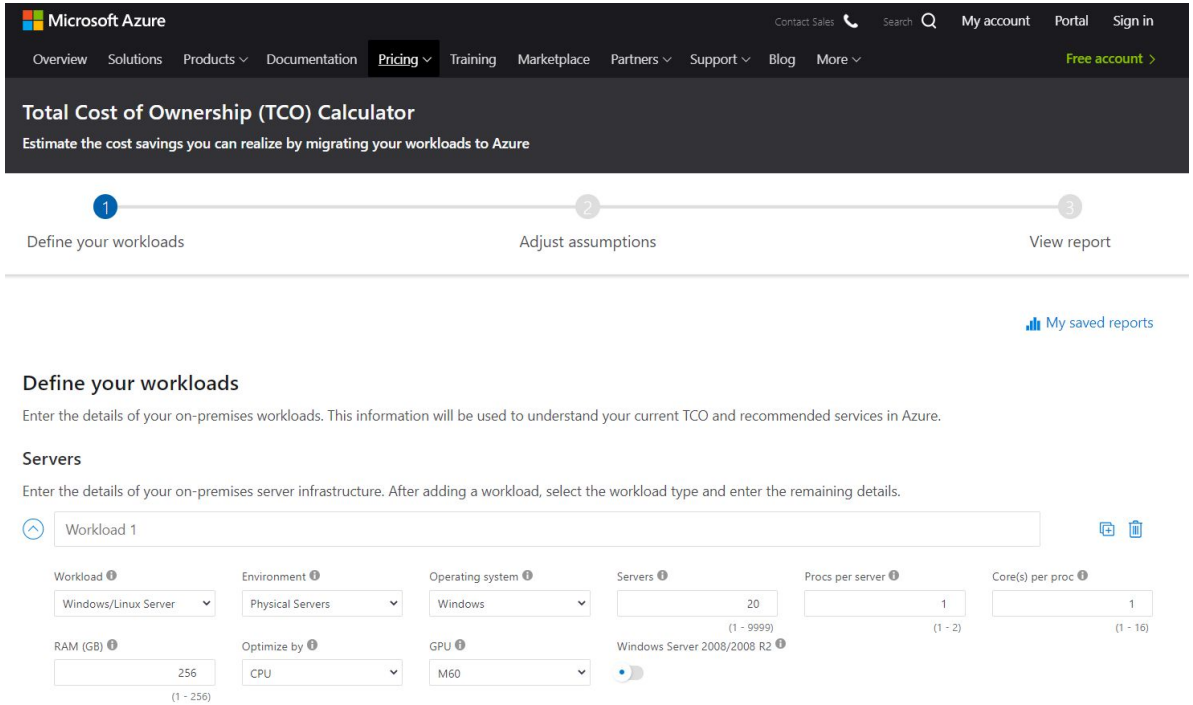
How to enable **Data Analytics** in an organization in 2020+?

1. Do it yourself - Private Cloud
2. Rental model - Public Cloud
3. **Mixed the two - Hybrid Cloud**

Leading cloud providers: AWS, Azure, GCP

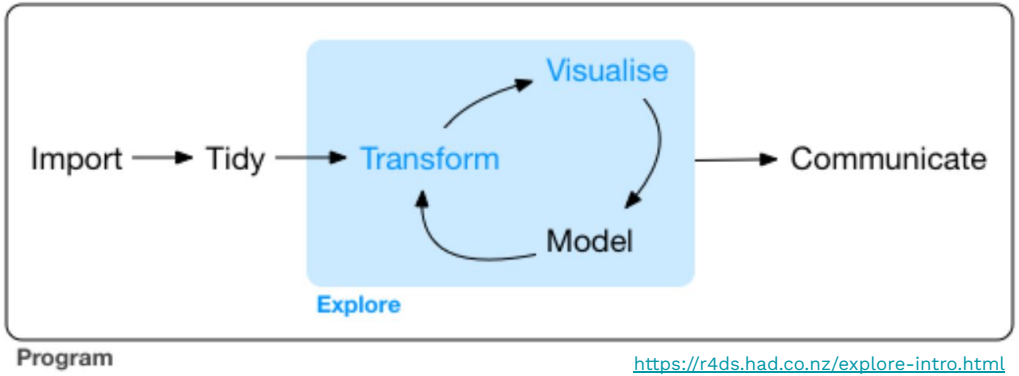
Total Cost of Ownership (TCO)

- **Cap**ital **Ex**pense
- **Op**erational **Ex**pense



https://azure.microsoft.com/en-us/pricing/tco/calculator/

Data Science is a set of **skills** that we use to analyze data and communicate our results.

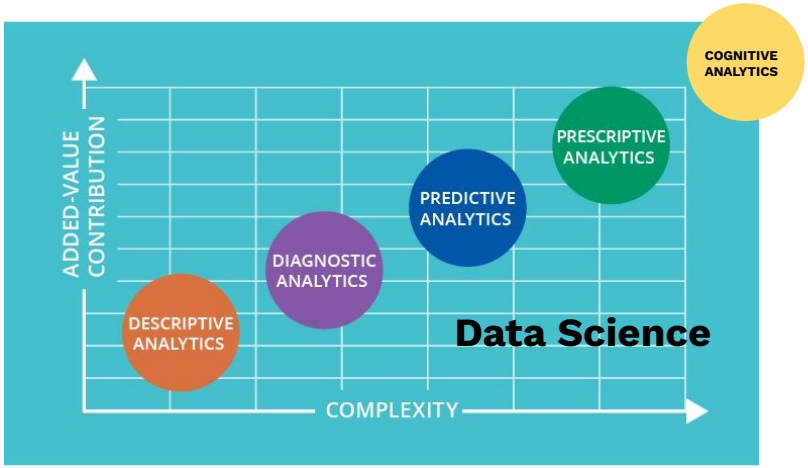


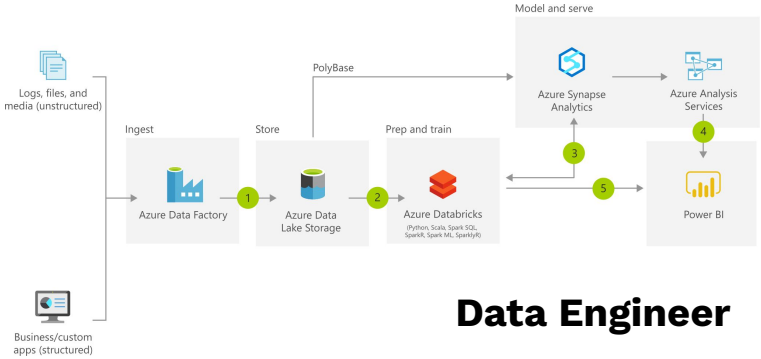
https://r4ds.had.co.nz/explore-intro.html

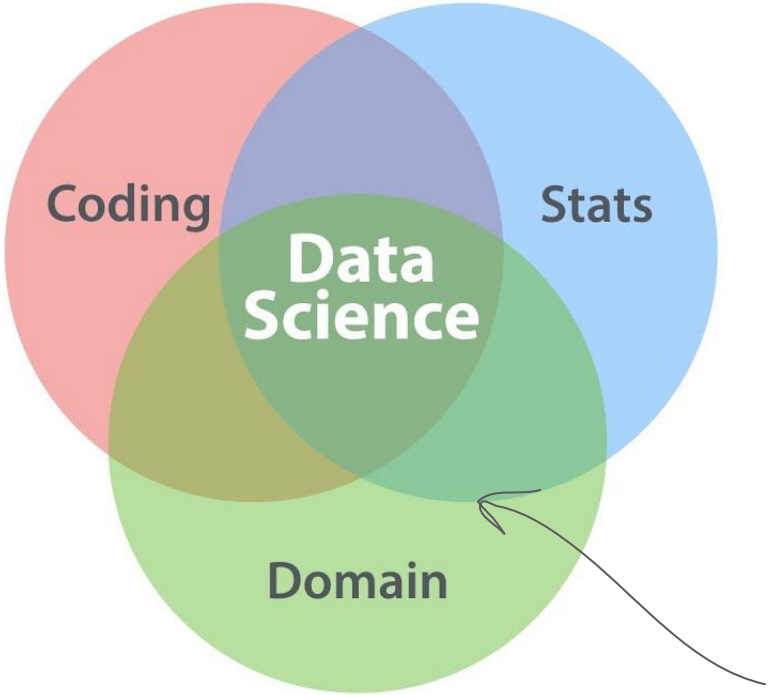
Data Analytics

Data analytics is the **<u>Science</u>** of analyzing raw data in order to make conclusions about that information.

DA helps a business optimize its performance.



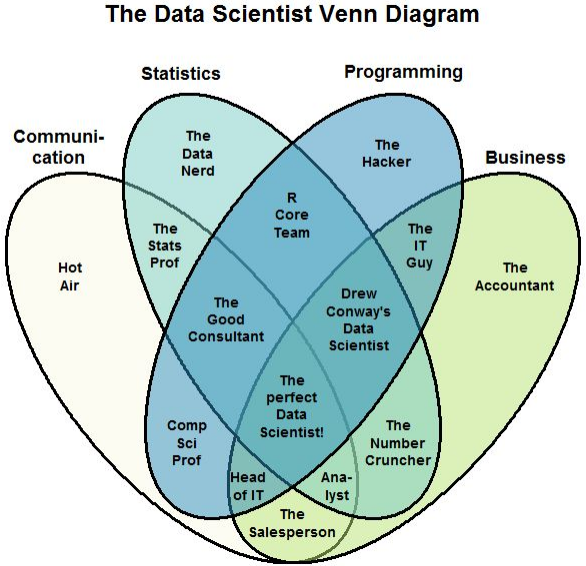


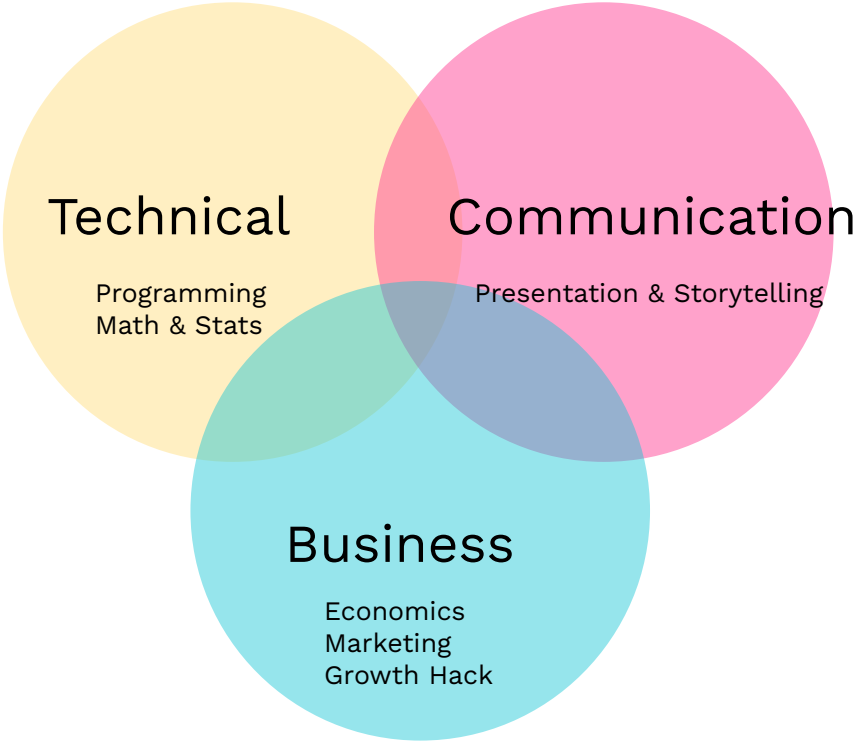


Made by Drew Conway

It Keeps Expanding

1. Business/ Domain
2. Statistics
3. Programming (to do data wrangling etc.)
4. Algorithms
5. Data Visualization & Storytelling
6. **Cloud Knowledge +**
7. **Growth Mindsets +**





Our School Version

Case Study

04 - Cambridge Analytica

Project Alamo

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

Project Alamo

From Wikipedia, the free encyclopedia

"Project Alamo" was a database of voter information created for Donald Trump's 2016 presidential campaign

CDP data

A customer **data** platform (**CDP**) is a type of packaged software which creates a persistent, unified customer database that is accessible to other systems. **Data** is pulled from multiple sources, cleaned and combined to create a single customer profile. This structured **data** is then made available to other marketing systems.

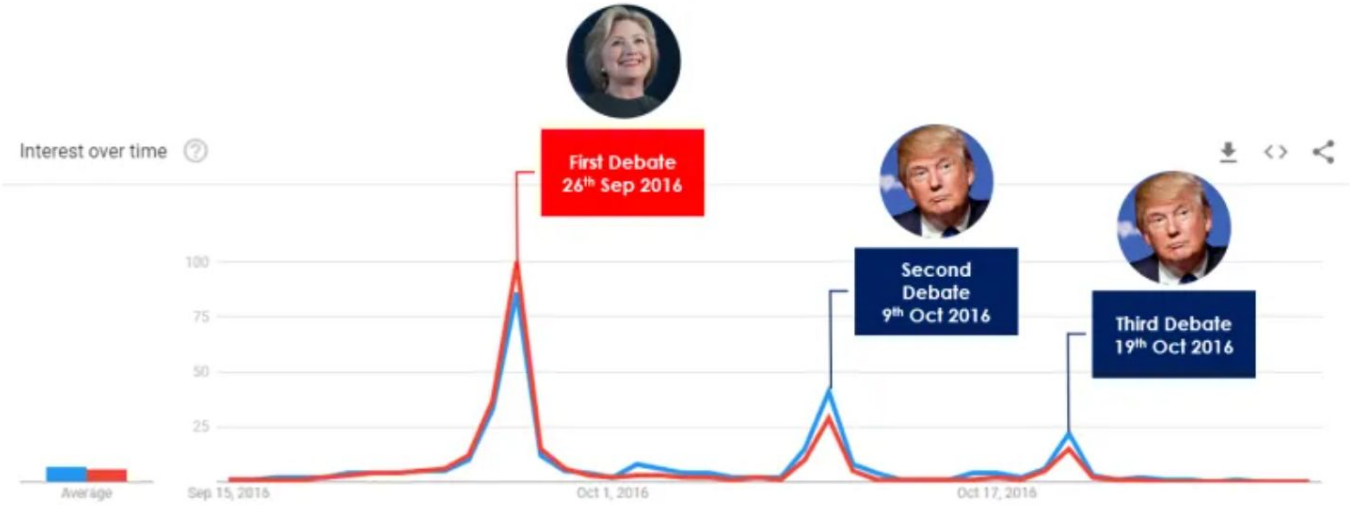
en.wikipedia.org › wiki › Customer_data_platform

Customer data platform - Wikipedia

Digital Footprint

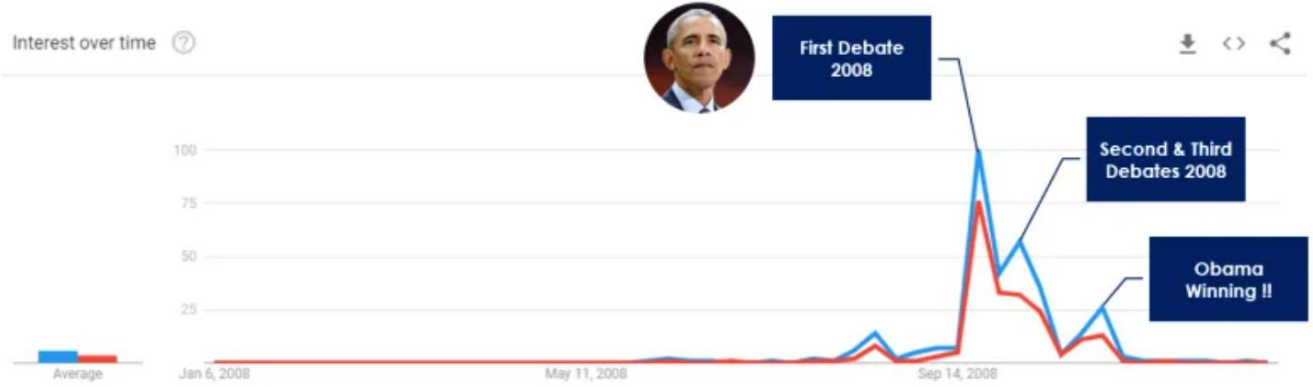
blognone
Cambridge
Analytica

Y2016



https://trends.google.com/trends/explore?date=2016-09-15%202016-10-30&geo=US&q=%22Trump%20Clinton%20Debate%22,%22Clinton%20Trump%20Debate%22

Y2008



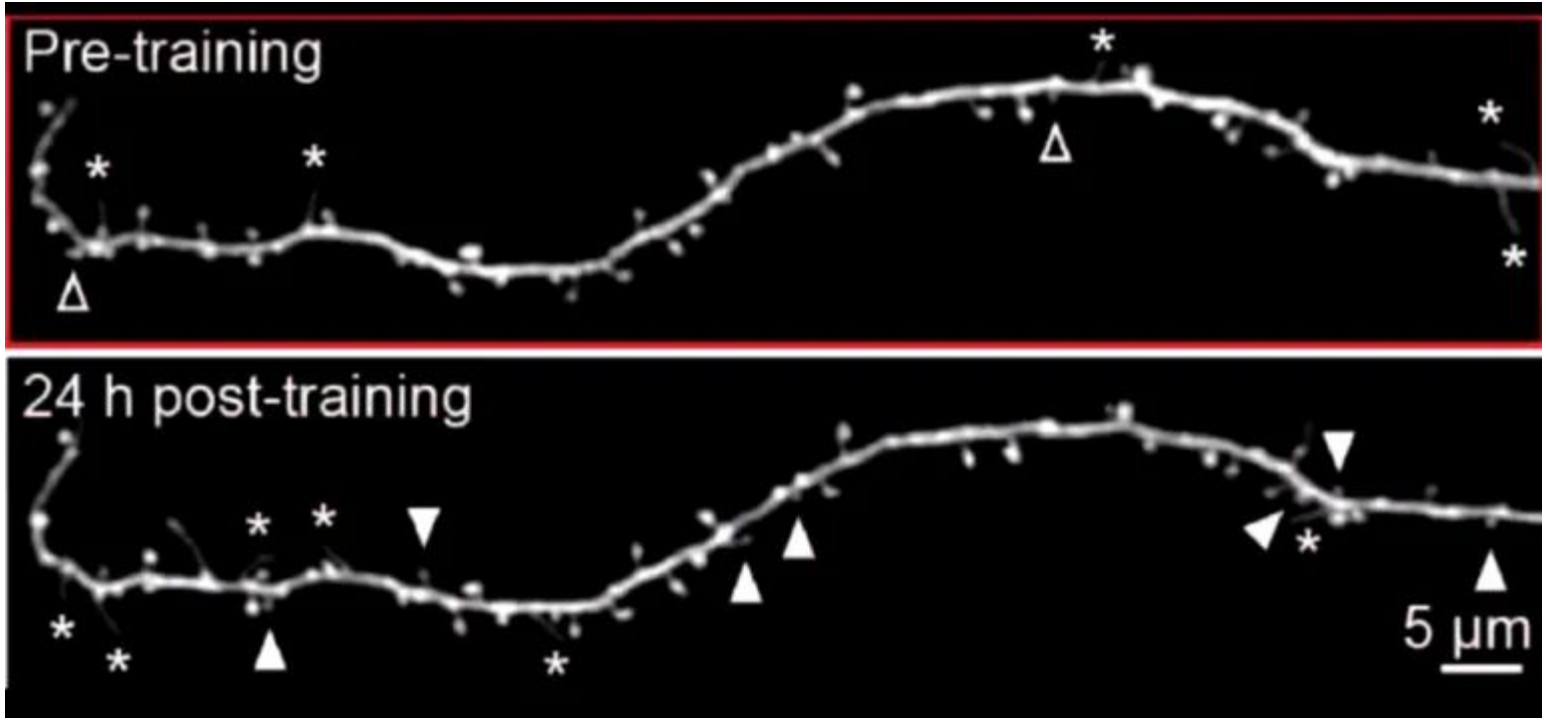
https://trends.google.com/trends/explore?date=2008-01-01%202008-12-31&geo=US&q=%22Obama%20Mccain%20Debate%22,%22Mccain%20Obama%20Debate%22

Final Tips

Learning how to learn

Barbara Oakley

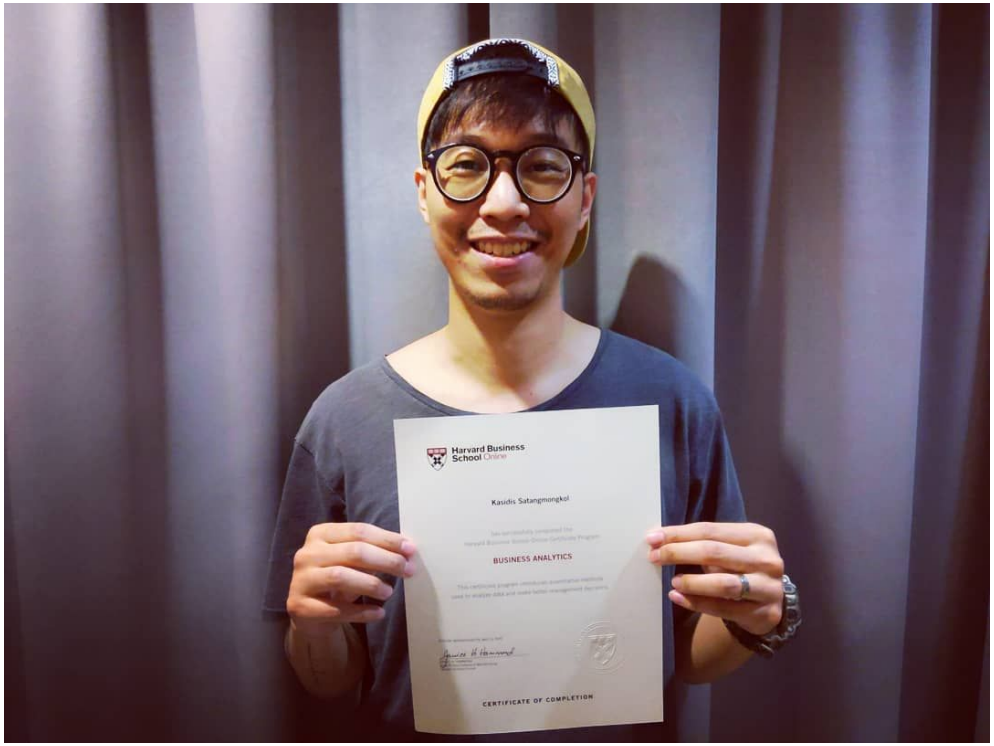
https://datarockie.com/2019/04/05/learning-how-to-learn-barbara-oakley/


Pre-training
24 h post-training
5 µm

Learn **One Step** at a Time

Contents

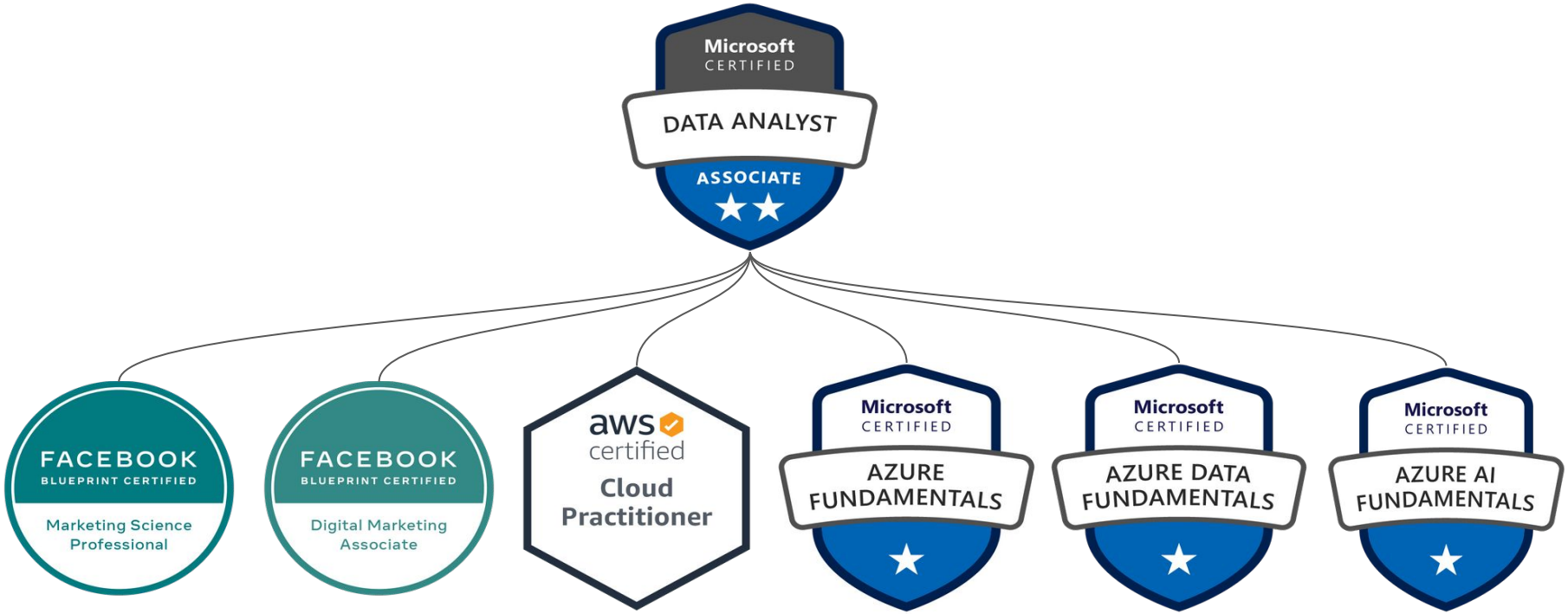
- Meet Toy! Your instructor
- Learning path 0-4 months
- What tools we have on our school
  - Podia https://course.datarockie.com (~~Notion~~) NEW
  - YouTube Live (~~Google Meet~~)
- Introduction to Data Science
  - Big Data
  - Cloud Computing
  - What is Data Science?
  - 5 Phases of Data Analytics
  - Skills & Tools e.g. demo Google Trends
  - Data Roles
  - Case Studies: Target, Clubcard, NetFlix, Cambridge Analytica
- Tips for learning online successfully



Toy

- ~~32~~ 33 years old
- Economist
- Data Analyst
- Reading
- Blogging

Data Analyst



Statistics/ Testing | Marketing | Cloud Concept | Cloud Concept | Data Concept | AI Concept

This badge was issued to KASIDIS SATANGMONGKOL on 4 October 2020.
Expires on 4 October 2022

Share

•••

Microsoft Certified: Azure Data Scientist Associate

Issued by Microsoft

Earning the Azure Data Scientist Associate certification validates the skills and knowledge to apply Azure's machine learning techniques to train, evaluate, and deploy models that solve business problems.

Skills

- Azure
- Azure Machine Learning
- Data Scientist
- Machine Learning
- ML Ops

Additional Details

https://www.youracclaim.com/badges/e8b733ba-5d27-40e8-83f5-4785bb7faa54/public_url

CMMU Mahidol
June 10, 2016 ·

6 CMMU students received the prestigious ERASMUS+ scholarship to study for 1 term at Ludwig Maximilians University (LMU) in Munich, Germany

LMU is ranked 75 among the top 100 universities in the world (QS ranking)

Mr. Borwornsuk Vinittumkul, Mr. Nondtawat Prapasawas, Ms. Rapeepan Luangthada and Mr. Tad Hatchaleelaha from Entrepreneurship Management (EN) Mr. Kasidis Satangmongkol and Ms. Witchaya Wongwai from Marketing and Management (MM).

88 2 Comments 23 Shares

NEW Podia

- YouTube Live

Data Science Bootcamp #2

- Sprint 1 - Spreadsheets (10 Hours)
- Sprint 2 - Databases (10 Hours)
- Sprint 3 - Programming for Data Science (12-15 Hours)
- Sprint 4 - Data Visualization (5 Hours)
- Sprint 5 - Business Statistics (15 Hours)
- Sprint 6 - Machine Learning (5-10 Hours)
- Sprint 7 - Python
- Sprint 8 - Dashboard
- Sprint 9 - Exam Preparation

Google Sheets 102

- Download Exercise File
- EP01 - Import Data
- EP02 - COUNTIFS
- EP03 - SUMIFS
- EP04 - FILTER
- EP05 - SORT
- EP06 - VLOOKUP Exact Match
- EP07 - VLOOKUP Approximate Match
- EP08 - INDEX MATCH
- EP09 - DATE
- EP10 - DATEDIF

Google Sheets 103

- Download Exercise File
- EP01 - Intro to QUERY function
- EP02 - Select Data
- EP03 - Filter Data
- EP04 - Dealing with NULL values
- EP05 - Aggregate Data
- EP06 - Aggregate Data by Group
- EP07 - Label
- EP08 - Pivot
- EP09 - Order By

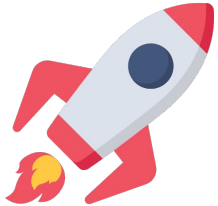
Data Science Bootcamp by Toy DataRockie

Login

Enrollment is closed

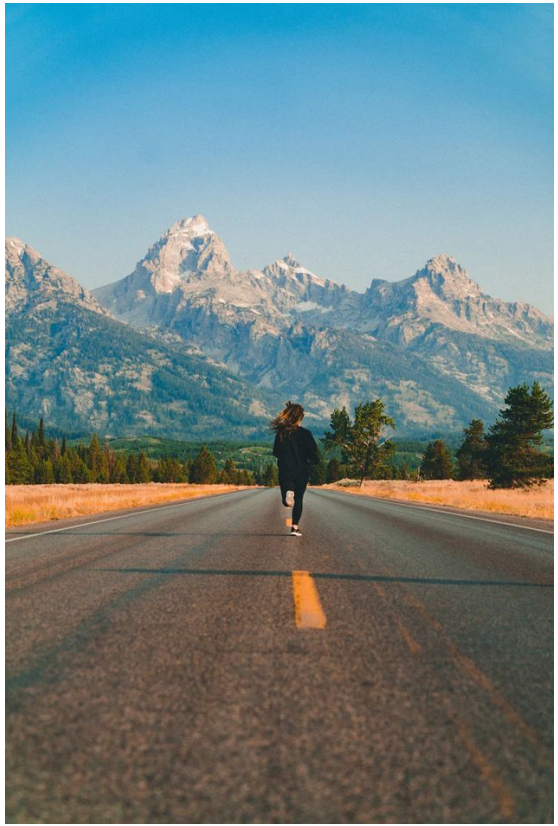
Data Science Bootcamp

Intensive data science bootcamp batch 4



Spreadsheets + SQL	Programming	Stats + ML	Dashboard
1st month	2nd month	3rd month	4th month

What is **Sprint?**



Google

what is sprint agile

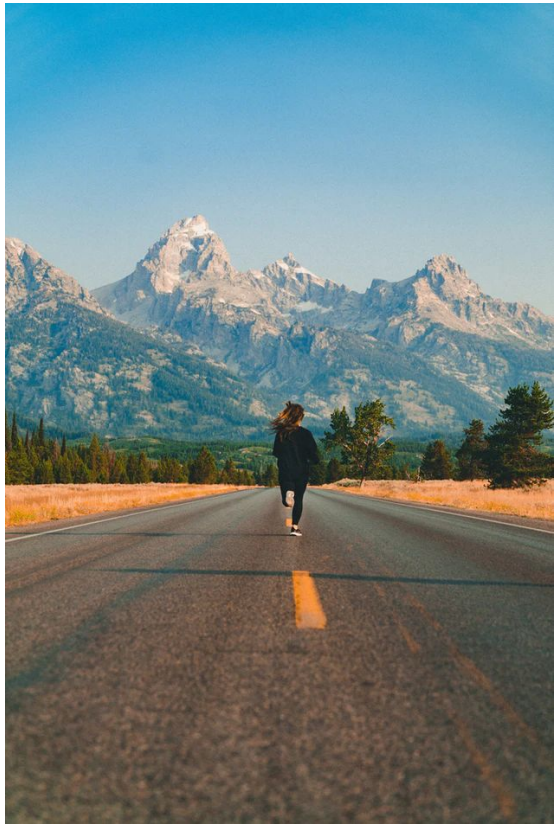
All | Images | Videos | Shopping | News | More | Settings | Tools

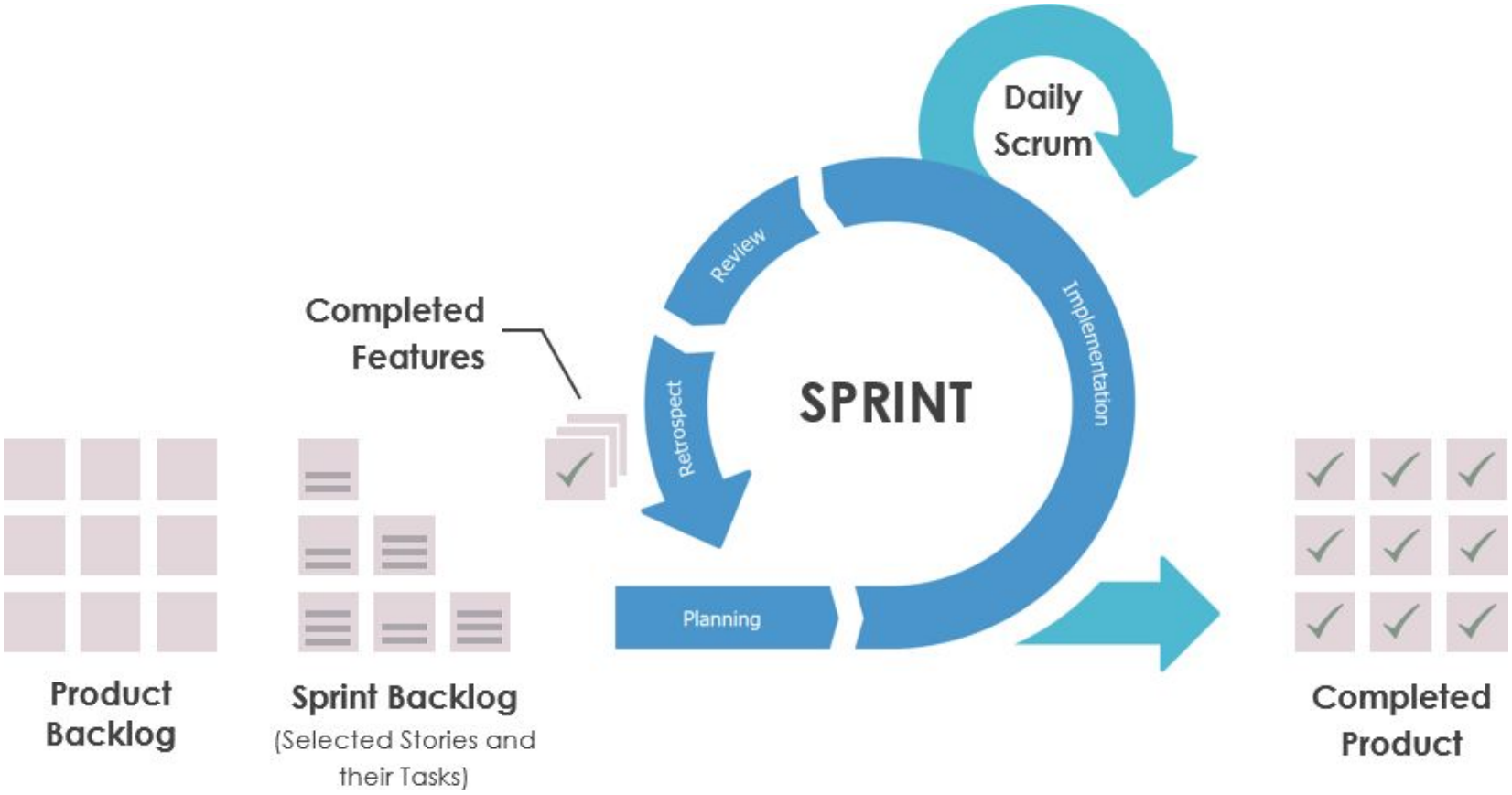
About 11,300,000 results (0.55 seconds)

In **Agile** product development, a **sprint** is a set period of time during which specific work has to be completed and made ready for review. Each **sprint** begins with a planning meeting. ... The duration of a **sprint** is determined by the **scrum** master, the team's facilitator and manager of the **Scrum** framework.

https://searchsoftwarequality.techtarget.com › Scrum-sprint

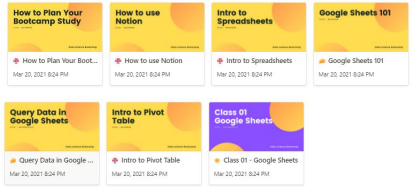
What is a Sprint (Agile Software Development)? Definition ...




Daily Scrum
Review
Implementation
Retrospect
SPRINT
Planning
Completed Features
Product Backlog
Sprint Backlog
(Selected Stories and their Tasks)
Completed Product

Sprint

Planning



Refinement

Can I meet the deadline on time?

Review

What have you done good/ bad?