

A photograph of a modern living room with exposed red brick walls. In the foreground, a woman sits on the floor, looking at a laptop. In the center, two men sit on a dark green sofa; one is using a laptop while the other looks on. In the background, a woman stands near a fireplace mantel, talking on a phone. The room is decorated with several potted plants and a large mirror above the fireplace.

Data Science Bootcamp

Intro to Data Science

Contents

- Big data history
- What is data science?
 - Drew Conway diagram
- Data analytics concepts
- Data analyst workflow
- Case studies

Big Data



Big data: The next frontier for innovation, competition, and productivity

May 1, 2011 | Report

<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

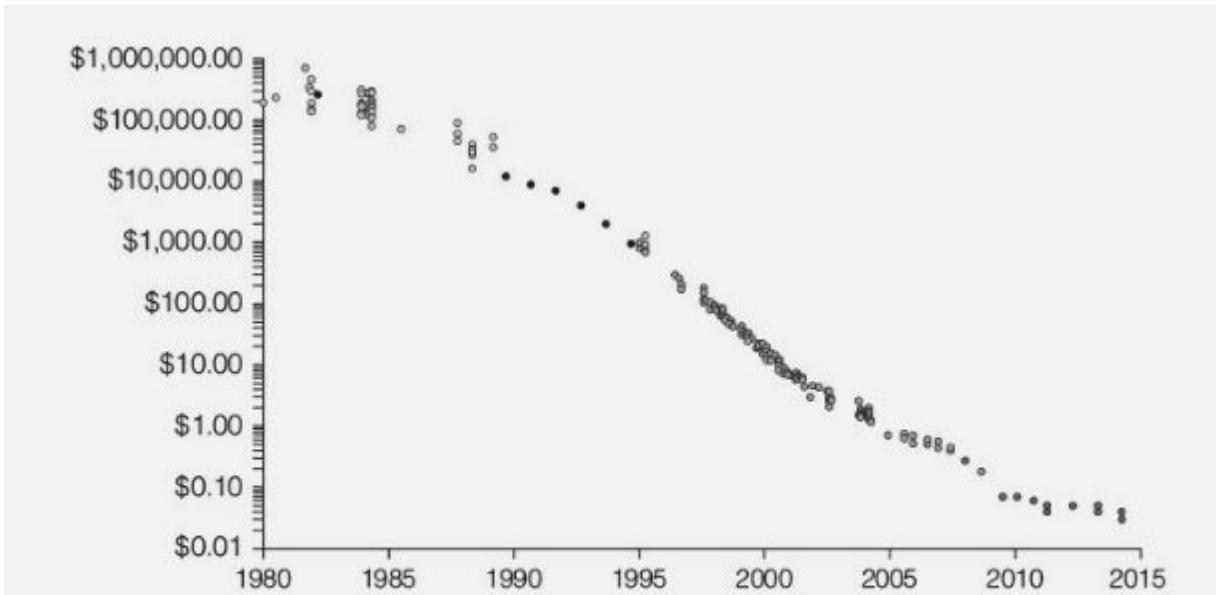
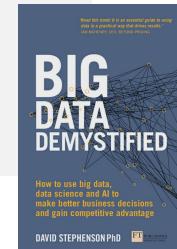


Figure 1.3 Historic cost of disk storage per GB (log scale).⁹





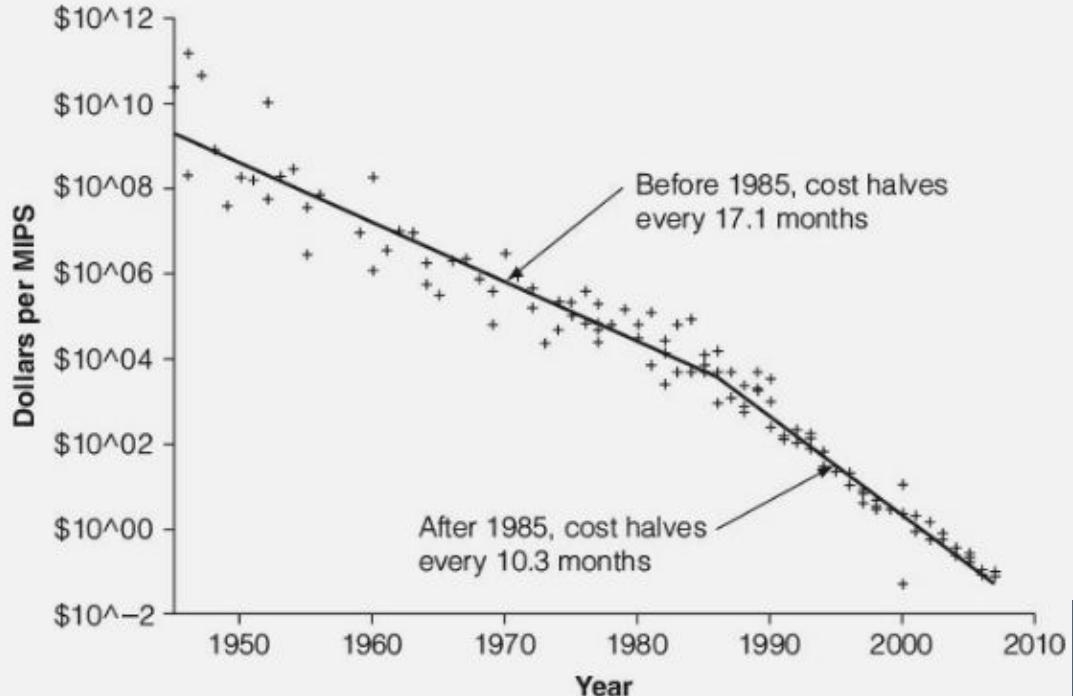
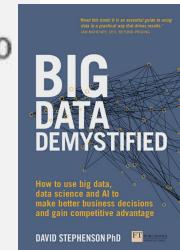
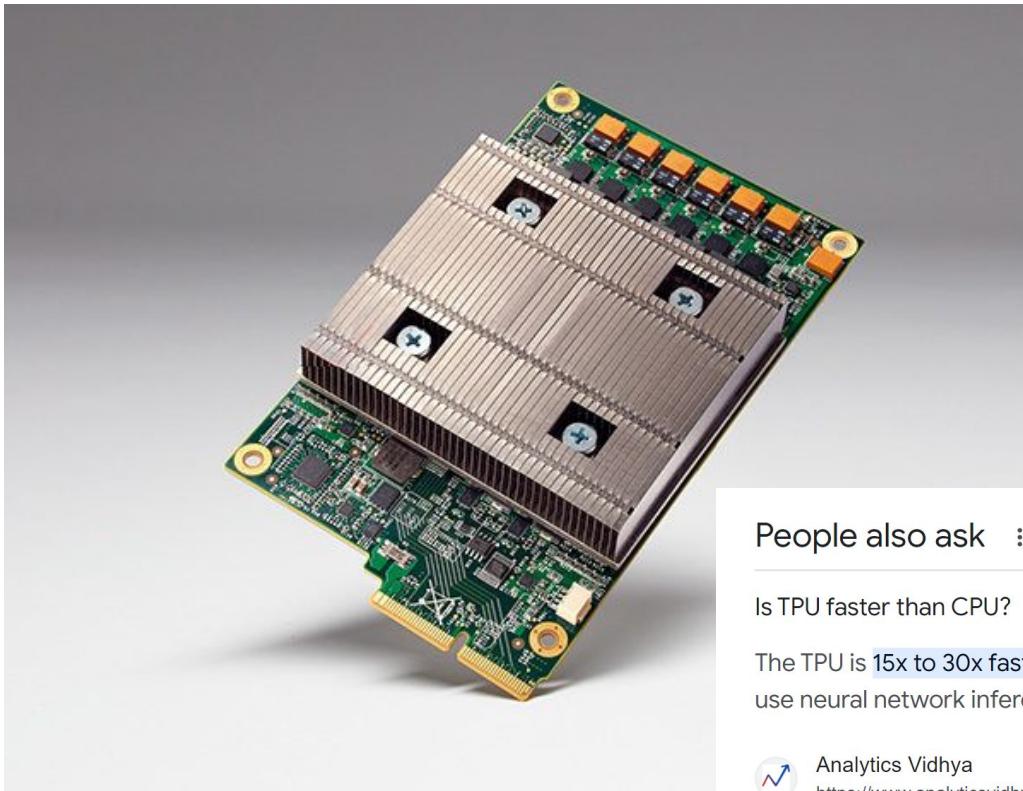


Figure 1.5 Historic cost of processing power (log scale).¹⁰





People also ask :

Google

Is TPU faster than CPU?

The TPU is [15x to 30x faster than current GPUs and CPUs](#) on production AI applications that use neural network inference. Aug 20, 2565 BE

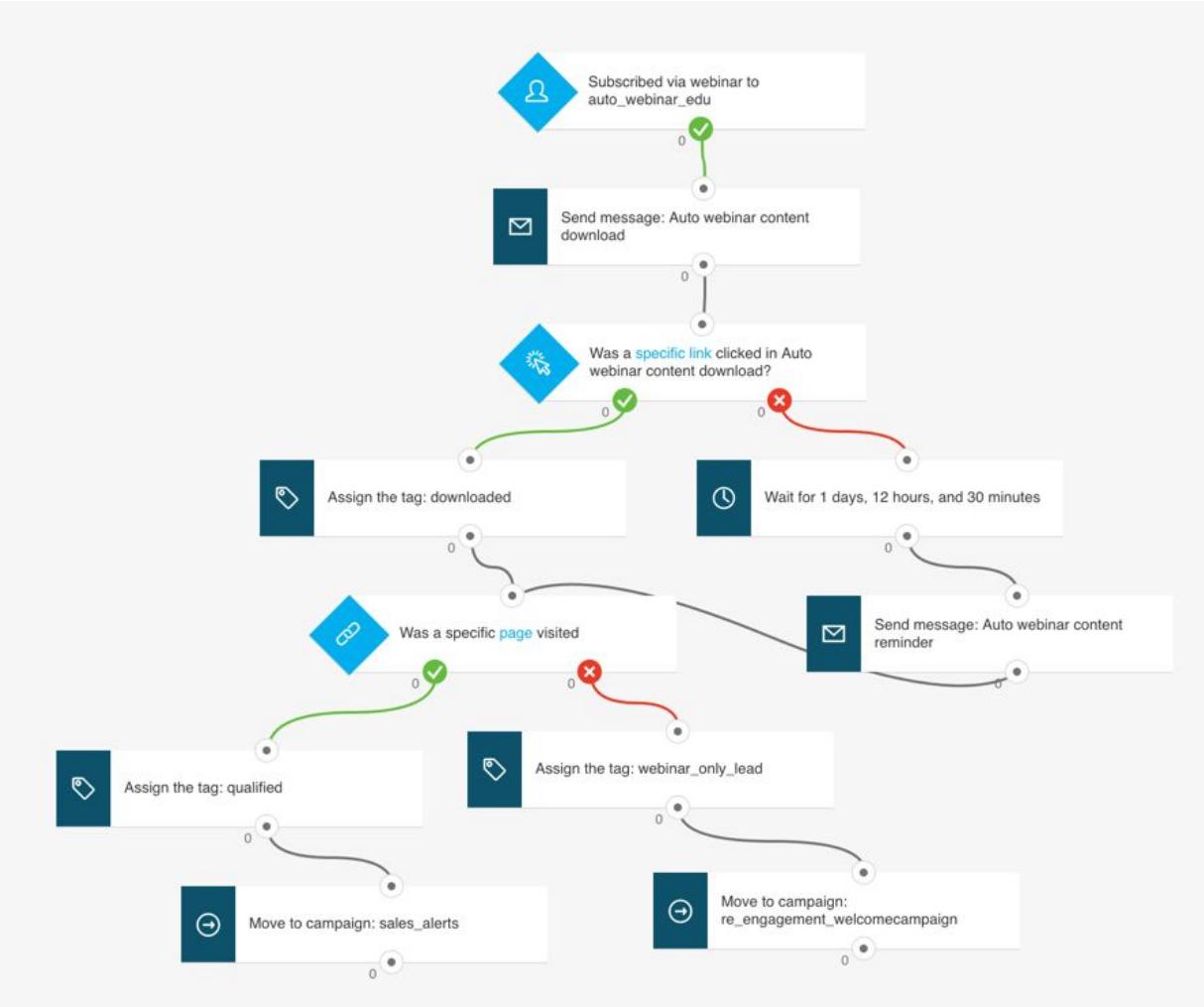


Analytics Vidhya

<https://www.analyticsvidhya.com> › blog › 2022/08 › e... ▾

Evolution of TPUs and GPUs in Deep Learning Applications -





At MailChimp, we retrain our models offline once a quarter, test them, and then promote them into production. In R, it takes me a few hours to train the model. And even though we as a company have terabytes of data, the model's training set, once prepped, is only 10 gigabytes, so I can even train the model on my laptop. Crazy.

When BIG become **small**



how big is google search query each day



All

Images

Shopping

Videos

News

More

Tools

About 247,000,000 results (0.48 seconds)

The latest data shows that Google processes over 99,000 searches every single second (Internet Live Stats, 2022). This makes more than 8.5 billion searches a day (Internet Live Stats, 2022). Let's also take a look at how Google's searches per year have progressed.

Jan 13, 2566 BE



Oberlo

<https://www.oberlo.com> › Blog



[10 Google Search Statistics You Need to Know in 2023 - Oberlo](#)

Explore what
Thailand
is searching for
right now

พิธีเปิดซีเกมส์ 2023

Explore

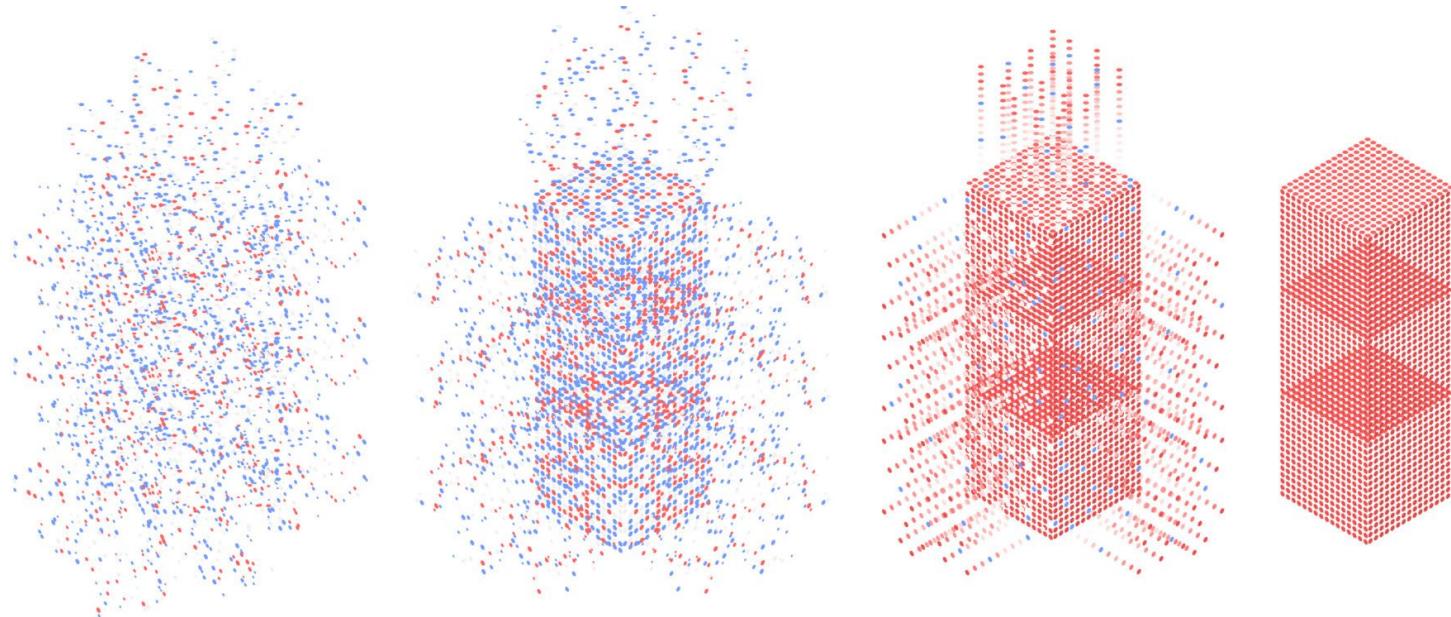
Search interest, past 24 hours

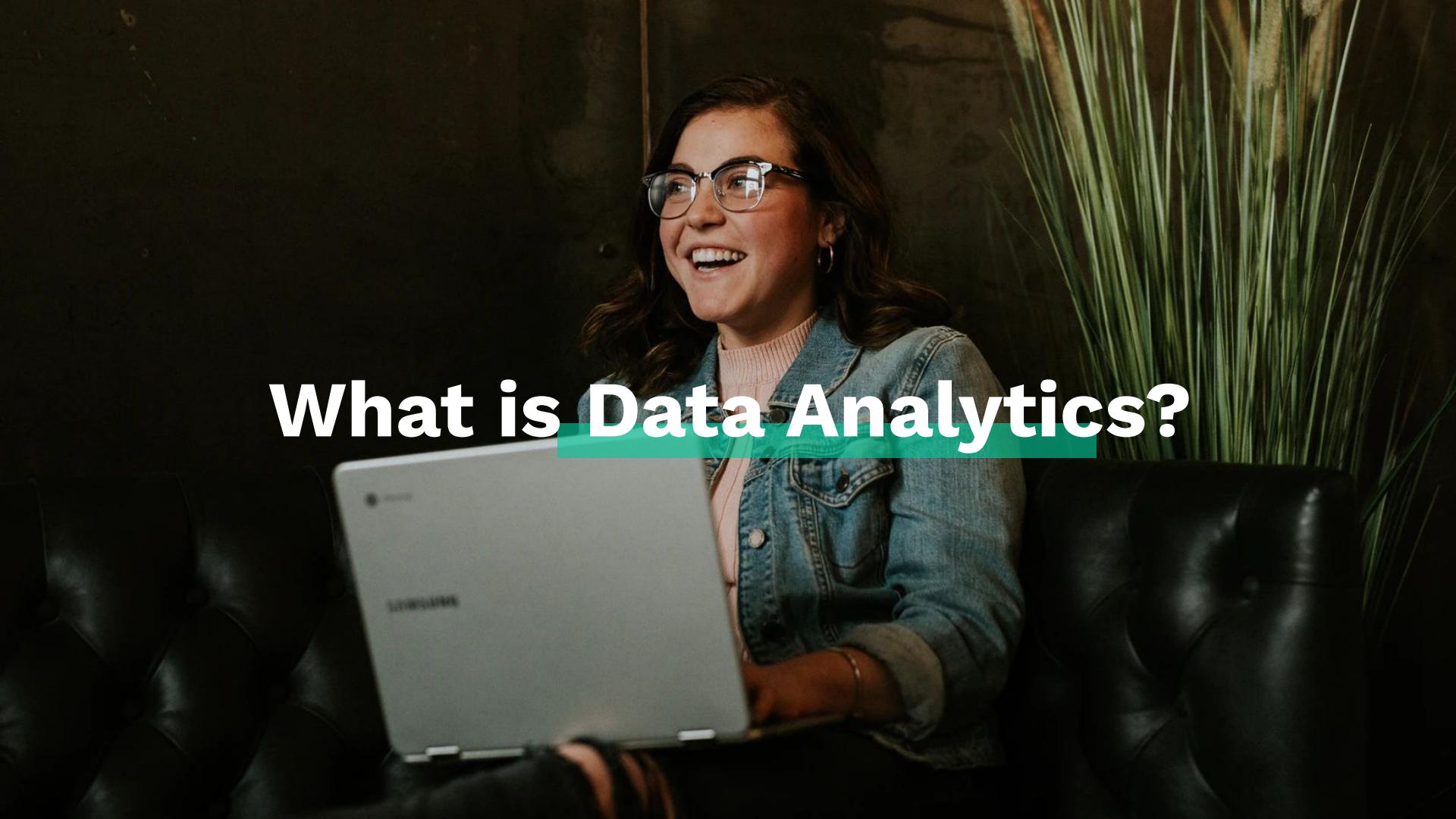


<https://trends.google.com/home>



When big data becomes small



A photograph of a young woman with long brown hair and glasses, wearing a denim jacket over a pink turtleneck. She is smiling and looking towards the right. She is sitting on a black leather sofa, holding a silver laptop. The background is dark and out of focus, showing some plants.

What is Data Analytics?



CareerFoundry

<https://careerfoundry.com> › blog › what-is-data-analyt...



What is Data Analytics? A Complete Guide for Beginners

Apr 6, 2566 BE — Data analytics is **the process of analyzing raw data in order to draw out meaningful, actionable insights**. Learn more in this complete ...



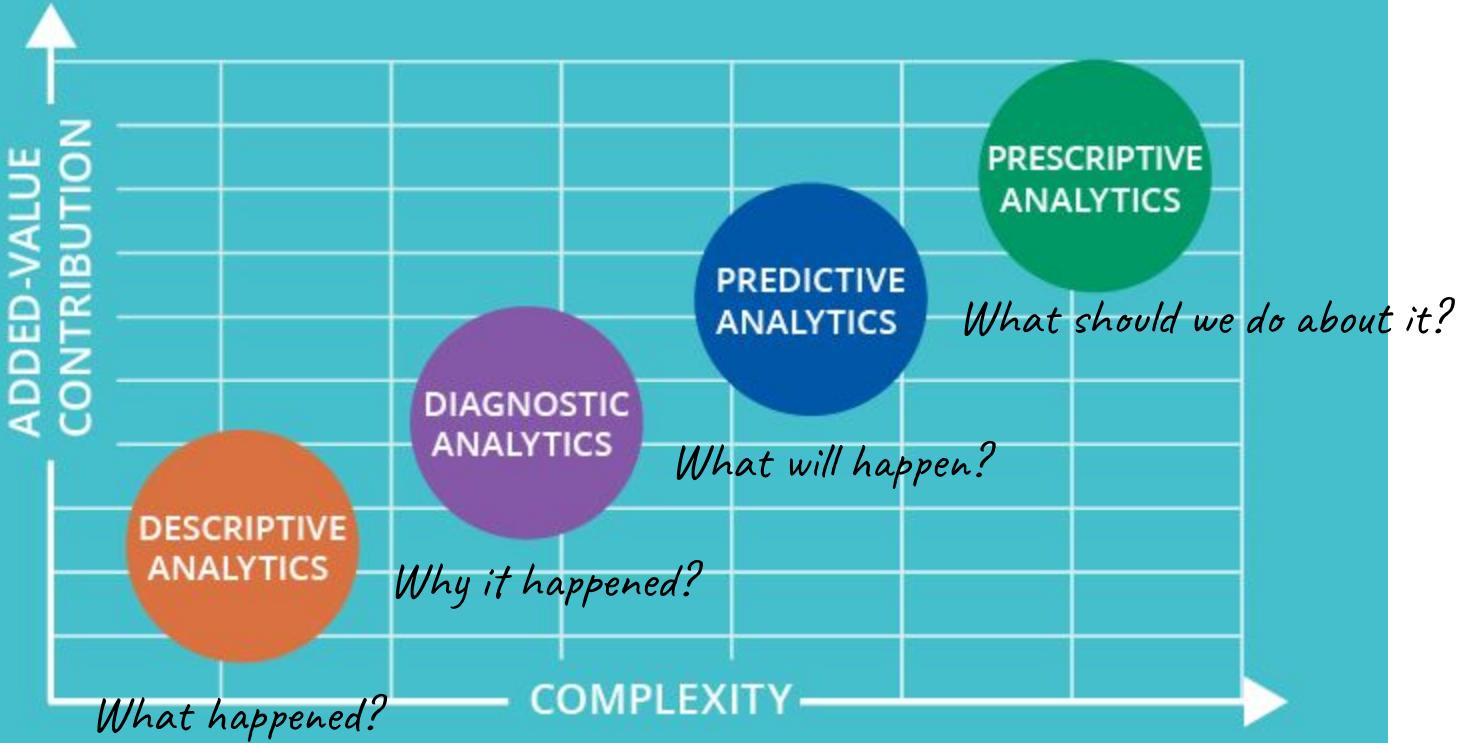
CompTIA

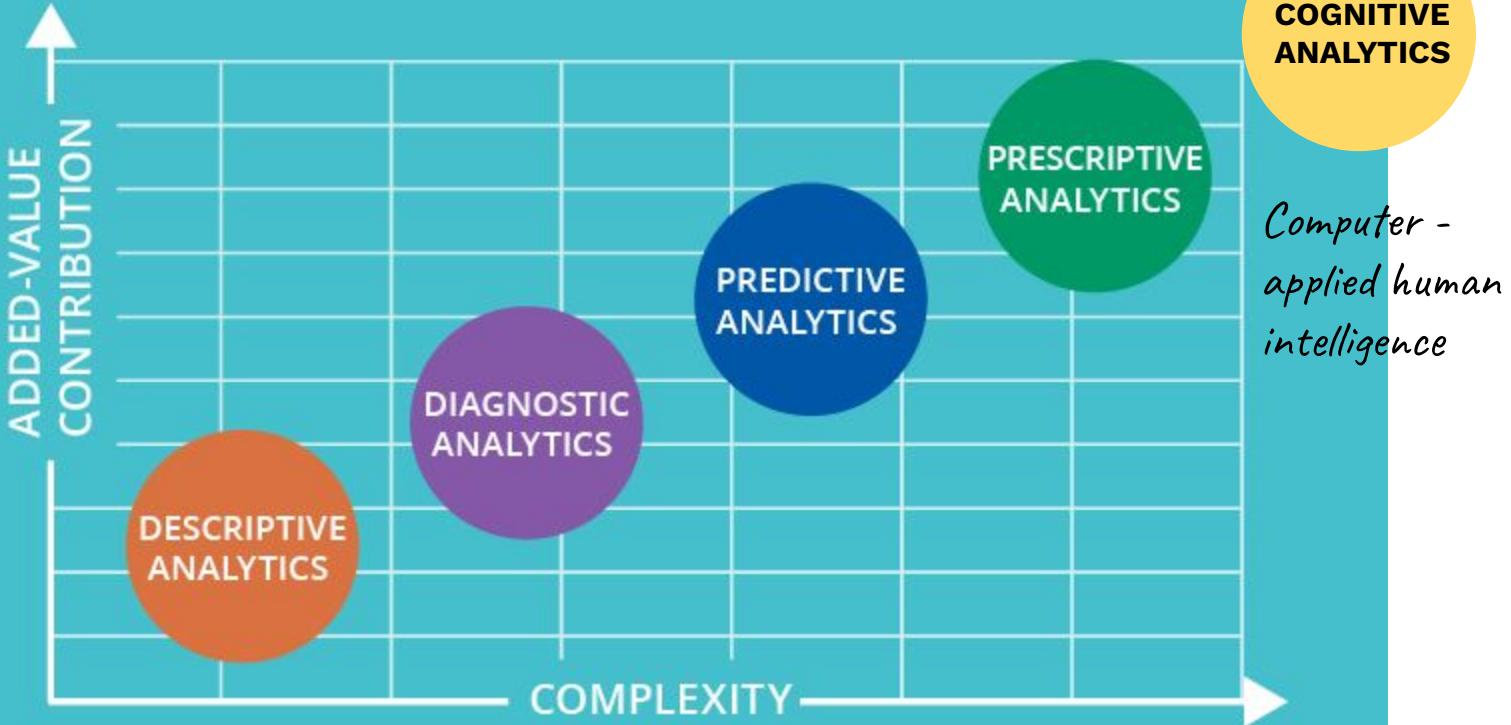
<https://www.comptia.org> › content › guides › what-is-...



What Is Data Analytics

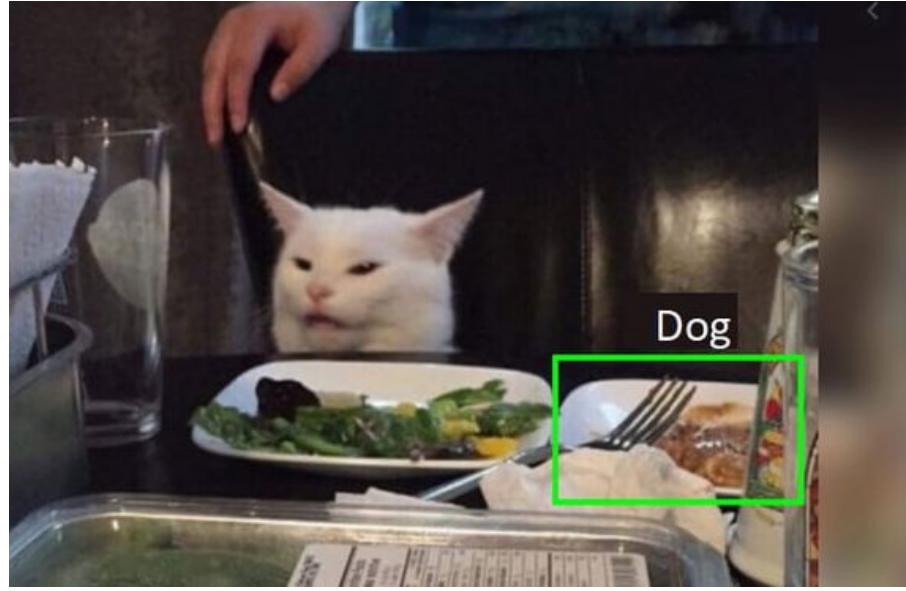
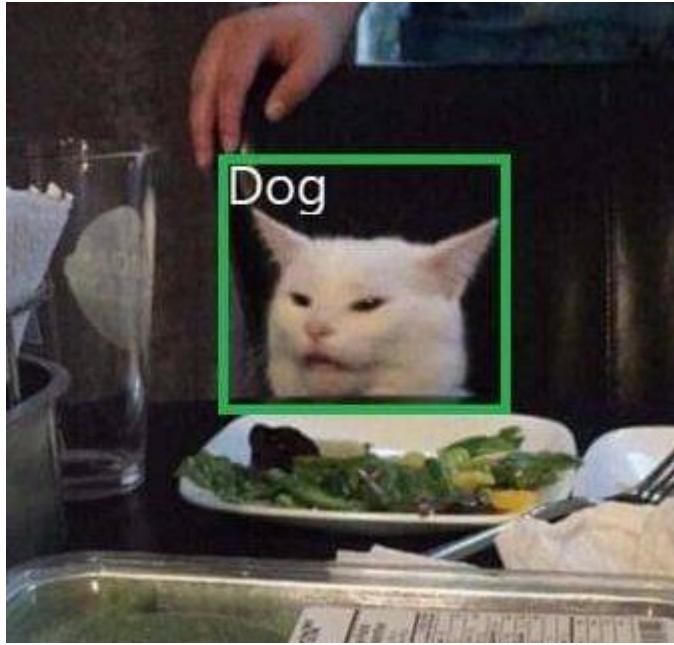
In short, data analytics is **the broad field of using data and tools to provide insights to make informed decisions**. Terms like data analysis, data science and ...





AGI vs. ANI





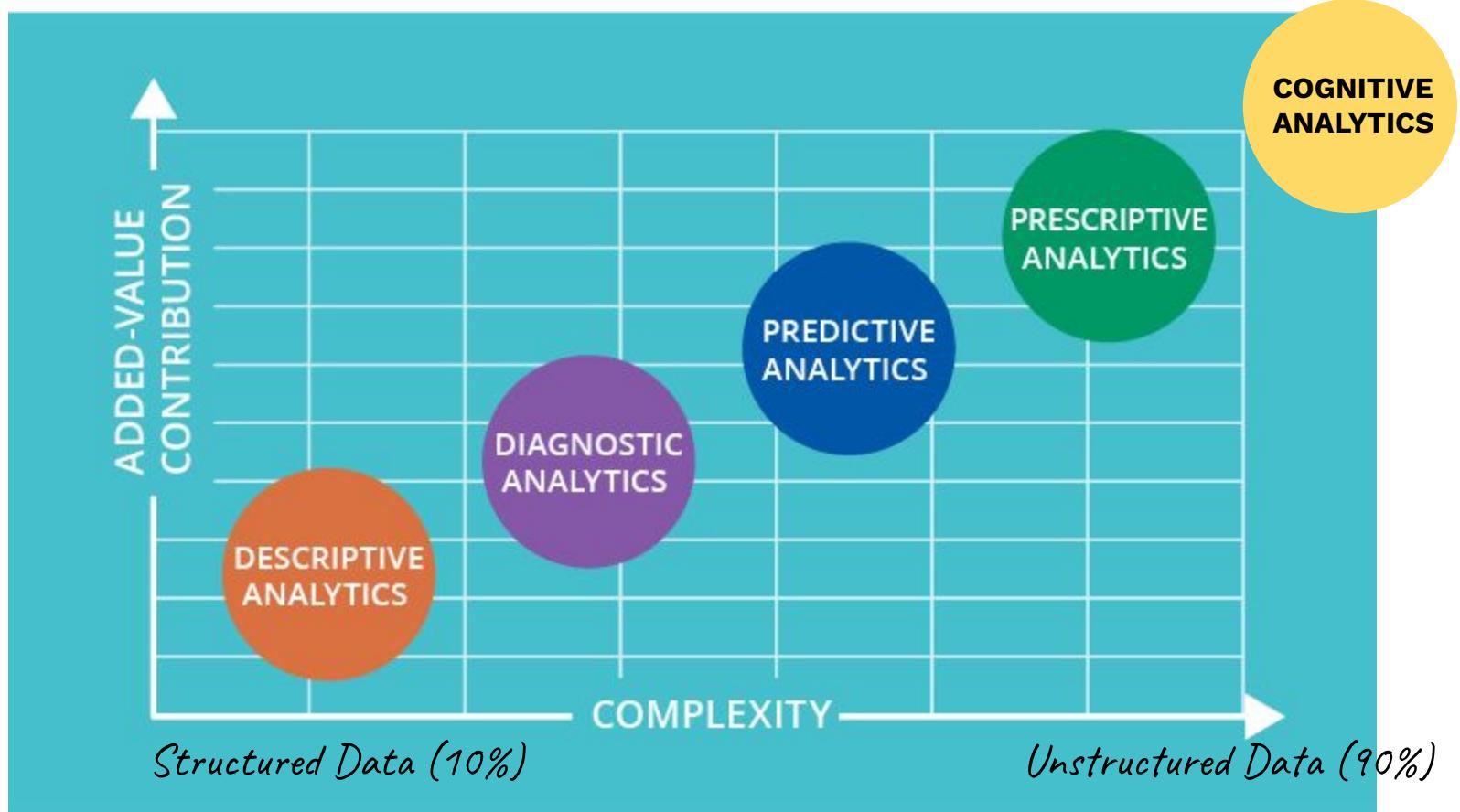
WIRED STAFF

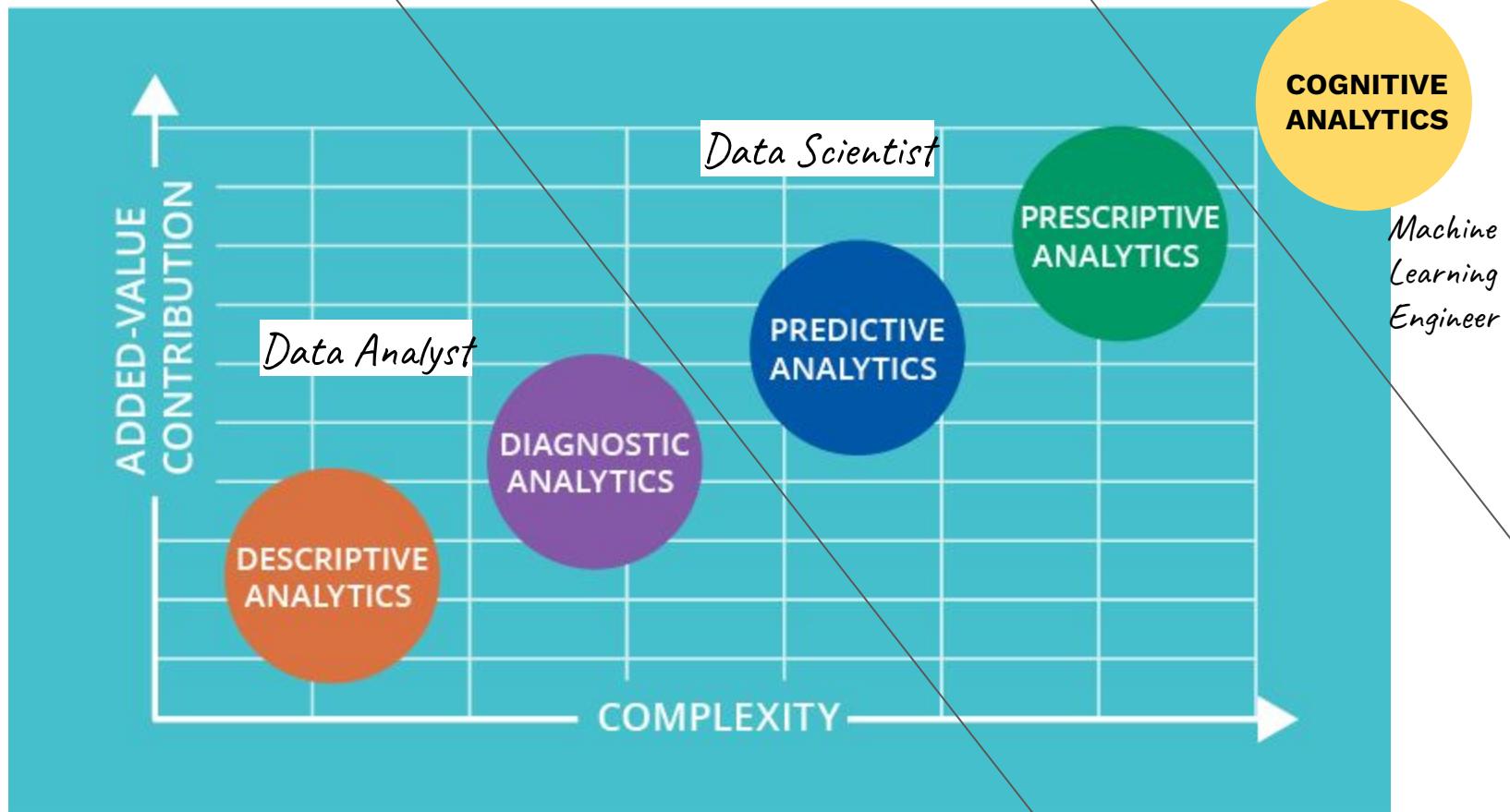
06.26.12 11:15 AM

Google's Artificial Brain Learns to Find Cat Videos

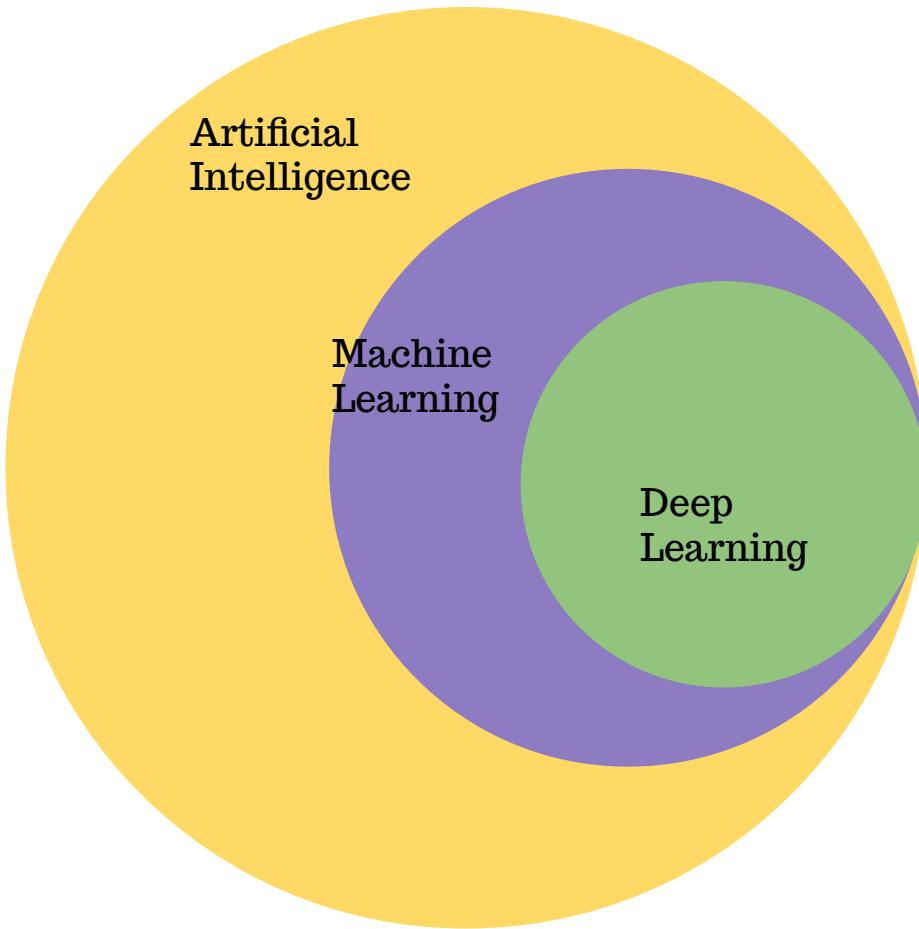


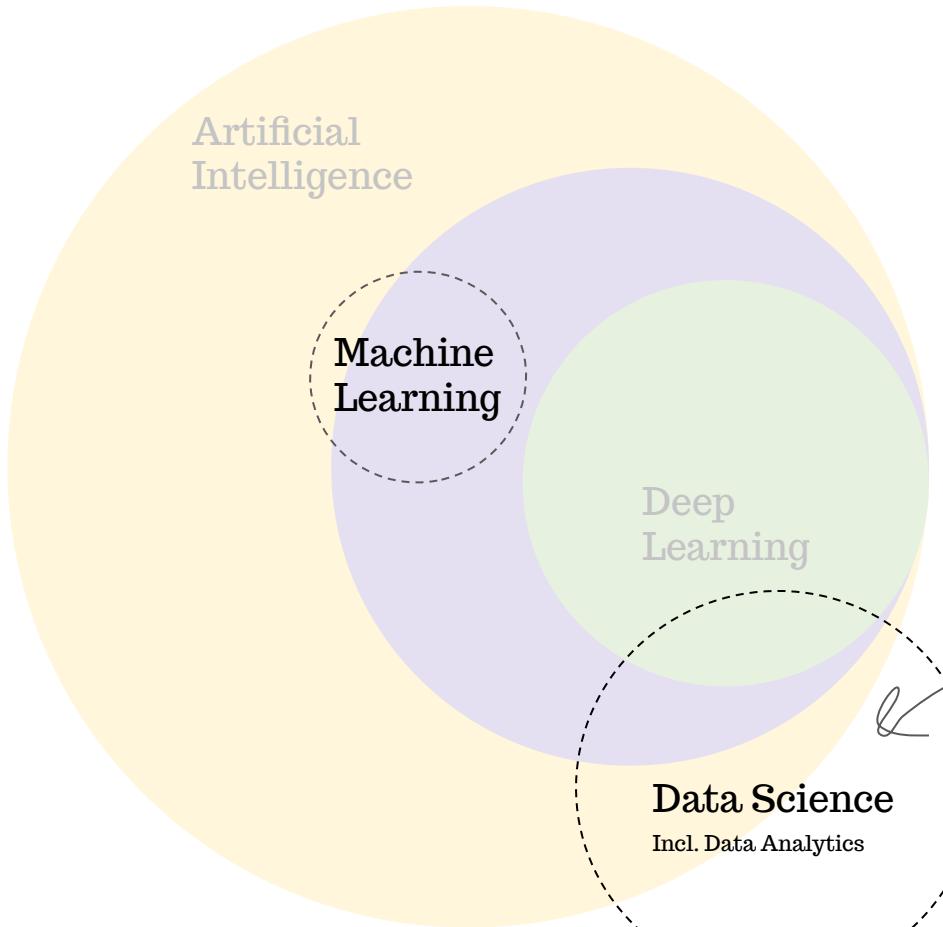
<https://www.wired.com/2012/06/google-x-neural-network/>





Data Engineer





Scaled automation
(boring tasks)

Our Focus



Extract meaningful and
actionable insights from data





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 ก็สามารถระบุสินค้าที่มีการซื้อยะยาในกลุ่มผู้หญิงมีครรภ์ได้แล้ว 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	ชิงค์	สำลี	...	ตั้งครรภ์
2	0	1	1	1	0	...	ไข่
3	0	0	1	0	0	...	ไม่ไข่
4	1	1	1	0	1	...	ไข่
5	1	1	0	1	1	...	ไข่
6	1	0	0	1	0	...	ไม่ไข่

ด้วยข้อมูลที่ Target ใช้สร้าง pregnancy score

$Pregnancy Score = f(\text{โลชั่นแบบไม่มีกลิ่น}, \text{วิตามิน}, \text{แคลเซียม}, \text{ชิงค์}, \text{สำลี}, \dots)$

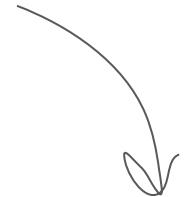
<https://datarockie.com/2019/08/10/case-study-target-supermarket/>

ผลการกำนันย

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

ข้อมูลลูกค้าใหม่ (new data)

ลูกค้ากี่เราจะส่งคูปองไปให้



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

ข้อมูลลูกค้าใหม่ (new data)



Prediction

		Actual	
		No	Yes
Prediction	No	850	20
	Yes	10	120

n=1000

Churn Prediction (The same problem)

churn	accountl	internatio	voicemail	numbervr	totaldaym	totaldayc	totaldaycl	totaleven	totalevec	totalevecd	totalnight	totalnightc	totalnightcl	totalintlm	totalintlc	totalintcl	numbercu
No	128	no	yes	25	265.1	110	45.07	197.4	99	16.78	244.7	91	11.01	10	3	2.7	1
No	107	no	yes	26	161.6	123	27.47	195.5	103	16.62	254.4	103	11.45	13.7	3	3.7	1
No	137	no	no	0	243.4	114	41.38	121.2	110	10.3	162.6	104	7.32	12.2	5	3.29	0
No	84	yes	no	0	299.4	71	50.9	61.9	88	5.26	196.9	89	8.86	6.6	7	1.78	2
No	75	yes	no	0	166.7	113	28.34	148.3	122	12.61	186.9	121	8.41	10.1	3	2.73	3
No	118	yes	no	0	223.4	98	37.98	220.6	101	18.75	203.9	118	9.18	6.3	6	1.7	0
No	121	no	yes	24	218.2	88	37.09	348.5	108	29.62	212.6	118	9.57	7.5	7	2.03	3
No	147	yes	no	0	157	79	26.69	103.1	94	8.76	211.8	96	9.53	7.1	6	1.92	0
No	117	no	no	0	184.5	97	31.37	351.6	80	29.89	215.8	90	9.71	8.7	4	2.35	1
No	141	yes	yes	37	258.6	84	43.96	222	111	18.87	326.4	97	14.69	11.2	5	3.02	0
Yes	65	no	no	0	129.1	137	21.95	228.5	83	19.42	208.8	111	9.4	12.7	6	3.43	4

P(Pregnant)

$$\text{sig}(t) = \frac{1}{1+e^{-t}}$$

sig(t)

1.0

0.8

0.6

0.4

0.2

Threshold

t

-8

-6

-4

-2

2

4

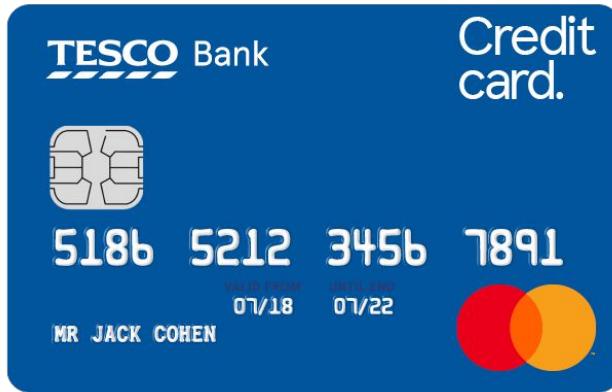
6

8



Case Study

02 - Tesco



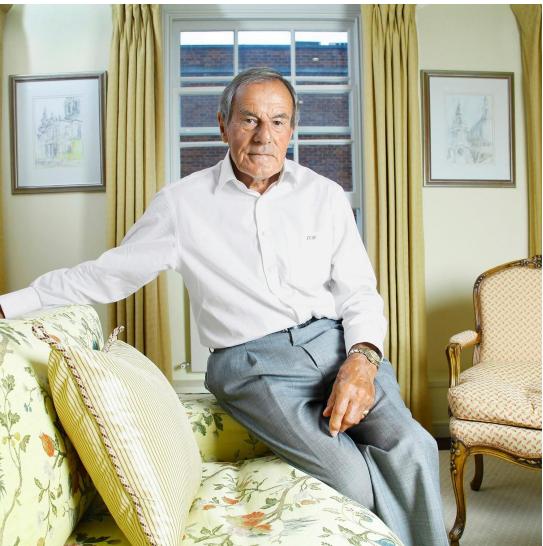
Tesco/ Dunnhumby - The very first company who uses data analytics and science with commercially success

ID

Amount

Store

Customer



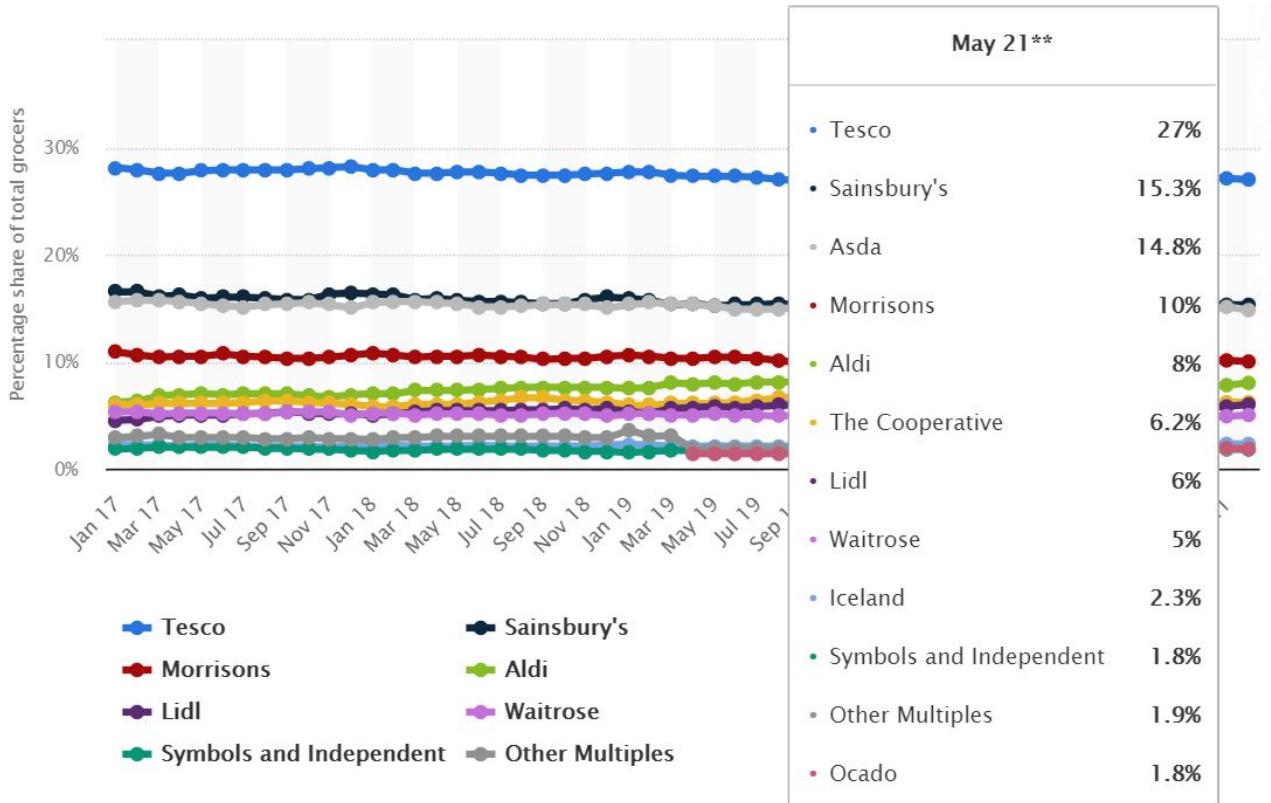
History [edit]

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]

Tesco Market Share 27% in 2021



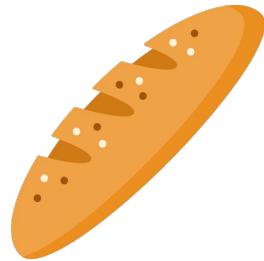
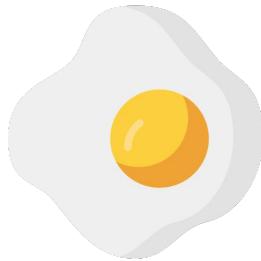
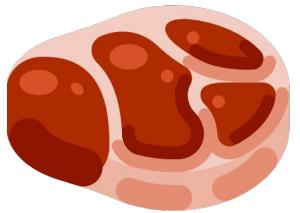
Strategic Data Acquisition

สมัคร ฟรี
สมัครง่าย
สมัครวันนี้ รับทันที
3,000 แต้ม*

The banner features a large orange background with white and yellow text. On the right side, a young man in a white t-shirt and grey hoodie holds up a white ALL member card in his left hand and a smartphone displaying the ALL member app in his right hand. The 7-Eleven logo is visible at the bottom right of the banner.

ALL member
ออลล์ เมมเบอร์

Item Set



Item ไหนขายดีที่สุด?

ถ้าลูกค้าซื้อ bread เดี๋ยวจะซื้อ item อะไรต่อ?

Basket 1



Basket 2



Basket 3



Basket 4



Basket 5



Basket 1



Basket 2



Basket 3



Basket 4



Basket 5



Item ไหนขายดีที่สุด?

Basket 1



Basket 2



Basket 3



Basket 4



Basket 5



Item ไหนขายดีที่สุด?



$$= 4 / 5 = 80\%$$



$$= 4 / 5 = 80\%$$

Basket 1



Basket 2



Basket 3



Basket 4



Basket 5



Item Set ไหนขายดีที่สุด?

Basket 1



Basket 2



Basket 3



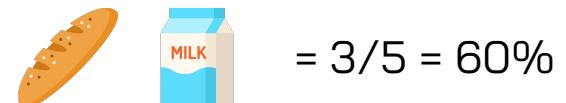
Basket 4



Basket 5



Item Set ไหนขายดีที่สุด?



$$= 3/5 = 60\%$$



$$= 3/5 = 60\%$$

Basket 1



Basket 2



Basket 3



Basket 4



Basket 5



ถ้าลูกค้าซื้อขนมปัง เดี๋ยวจะซื้อ item อะไรต่อ?

IF ... THEN ...

IF { **bread** } THEN { **milk** }

Basket 1



Basket 2



IF {**bread**} THEN {**milk**} = $3/4 = 75\%$

Basket 3



IF {**bread**} THEN {**egg**} = $3/4 = 75\%$

Basket 4



IF {**bread**} THEN {**beer**} = $2/4 = 50\%$

Basket 5



Basket 1



Basket 2



Basket 3



IF {bread, milk} THEN {egg} =

Basket 4



Basket 5



	lhs	rhs	support	confidence	lift	count
[1]	{WOBBLY CHICKEN}	=> {METAL}	0.001261773	1	443.82000	28
[2]	{WOBBLY CHICKEN}	=> {DECORATION}	0.001261773	1	443.82000	28
[3]	{DECOUPAGE}	=> {GREETING CARD}	0.001036456	1	389.31579	23
[4]	{BILLBOARD FONTS DESIGN}	=> {WRAP}	0.001306836	1	715.83871	29
[5]	{WOBBLY RABBIT}	=> {METAL}	0.001532153	1	443.82000	34
[6]	{WOBBLY RABBIT}	=> {DECORATION}	0.001532153	1	443.82000	34
[7]	{FUNK MONKEY}	=> {ART LIGHTS}	0.001712406	1	583.97368	38
[8]	{ART LIGHTS}	=> {FUNK MONKEY}	0.001712406	1	583.97368	38
[9]	{BLACK TEA}	=> {SUGAR JARS}	0.002072912	1	238.61290	46
[10]	{BLACK TEA}	=> {COFFEE}	0.002072912	1	69.34687	46

Big Data = **What**

Small Data = **Why**

A photograph of a night camping scene. In the foreground, there are two tents: a white one on the left and a grey one next to it. A campfire is burning between them, with two people sitting around it. The background is filled with dark, silhouetted trees against a starry sky.

Case Study

03 - Netflix





Netflix Prize

Home Rules Leaderboard Register Update Submit Download

Leaderboard

Display top leaders.

Rank	Team Name	Best Score	% Improvement	Last Submit Time
------	-----------	------------	---------------	------------------

No Grand Prize candidates yet

Grand Prize - RMSE <= 0.8563

1	PragmaticTheory	0.8584	9.78	2009-06-16 01:04:47
2	BellKor in BigChaos	0.8590	9.71	2009-05-13 08:14:09
3	Grand Prize Team	0.8593	9.68	2009-06-12 08:20:24
4	Dace	0.8604	9.56	2009-04-22 05:57:03
5	BigChaos	0.8613	9.47	2009-06-15 18:03:55

Progress Prize 2008 - RMSE = 0.8616 - Winning Team: BellKor in BigChaos

6	BellKor	0.8620	9.40	2009-06-17 13:41:48
7	Gravity	0.8634	9.25	2009-04-22 18:31:32
8	Opera Solutions	0.8640	9.19	2009-06-09 22:24:53
9	xvector	0.8640	9.19	2009-06-17 12:47:27
10	BruceDengDaoCiYIYou	0.8641	9.18	2009-06-02 17:08:31
11	Ces	0.8642	9.17	2009-06-12 23:04:25
12	majia2	0.8642	9.17	2009-06-15 03:35:00
13	xiangliang	0.8642	9.17	2009-06-13 16:35:35
14	Feeds2	0.8647	9.11	2009-06-16 22:21:19
15	Just a guy in a garage	0.8650	9.08	2009-05-24 10:02:54
16	Team ESP	0.8653	9.05	2009-06-16 05:25:11
17	pengpengzhou	0.8654	9.04	2009-05-05 18:18:03
18	NewNetflixTeam	0.8657	9.01	2009-05-31 07:30:22
19	J.Dennis Su	0.8658	9.00	2009-03-11 09:41:54
20	Vandelay Industries !	0.8658	9.00	2009-05-11 00:43:14

Apr 16, 2012, 05:53pm EDT

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

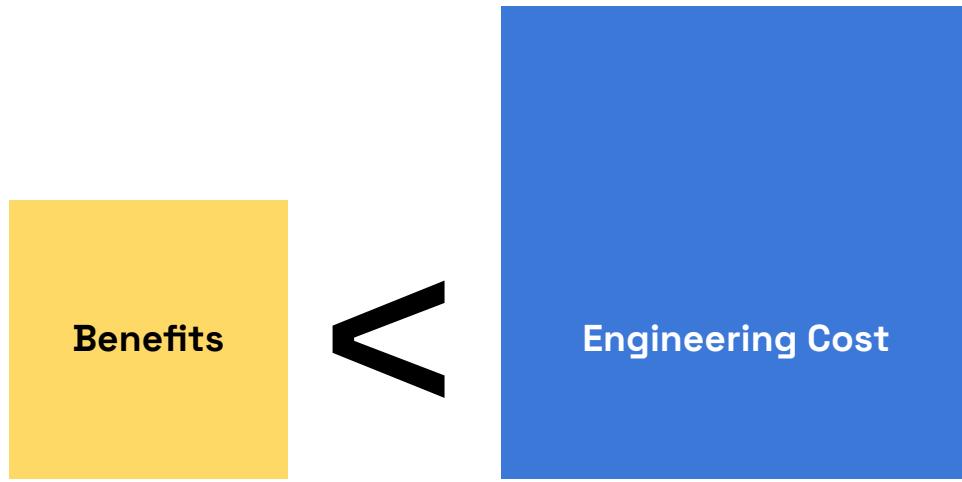


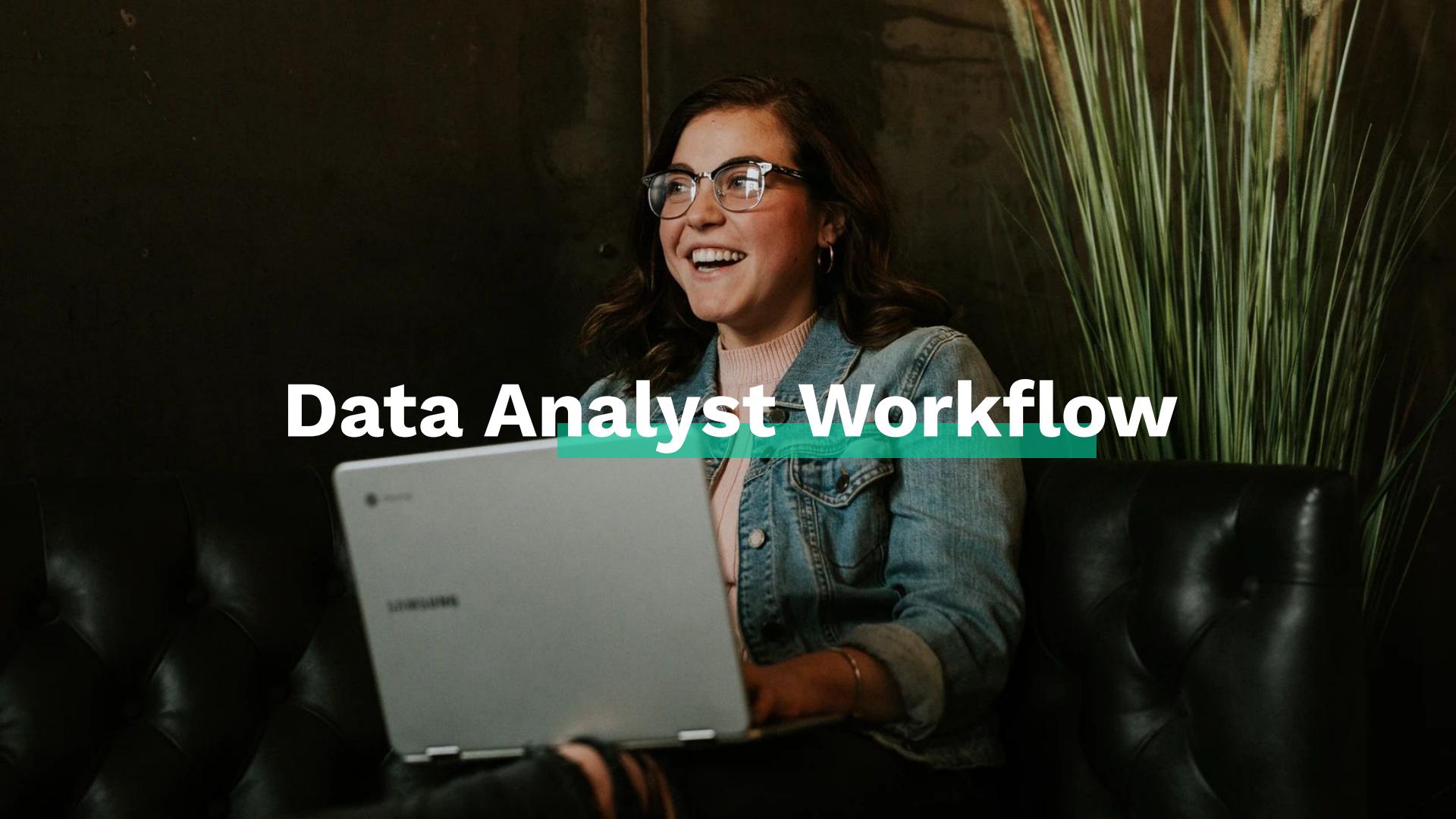
Ryan Holiday Contributor ⓘ

CMO Network - OLD

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

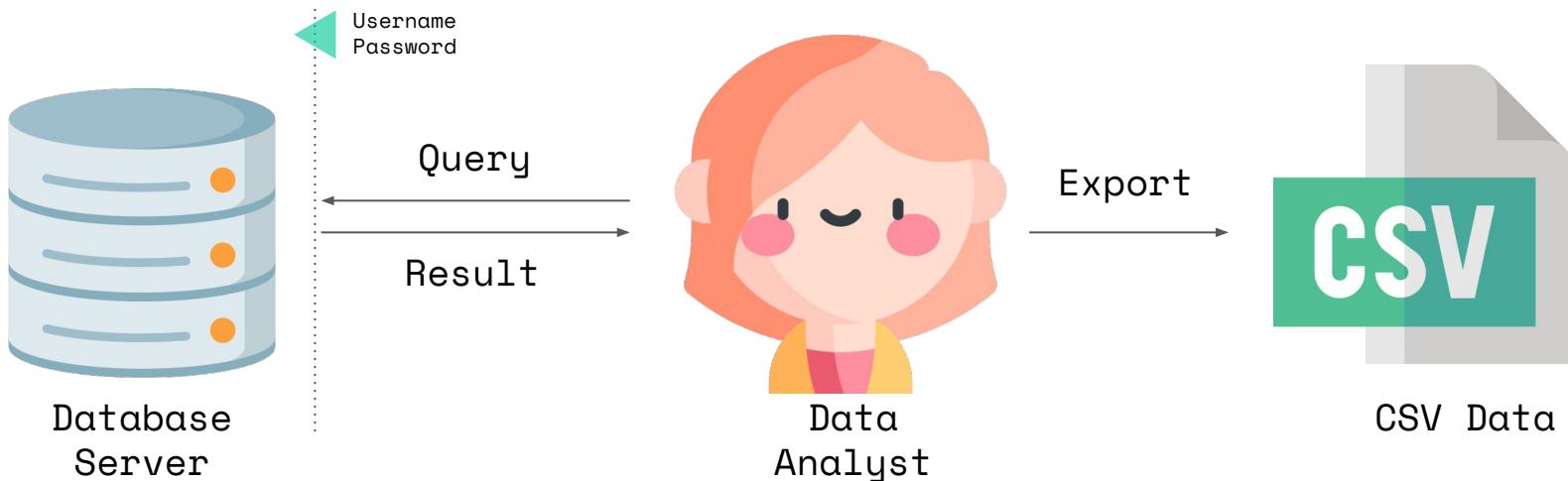
<https://www.forbes.com/sites/ryanholiday/2012/04/16/what-the-failed-1m-netflix-prize-tells-us-about-business-advice/#e24108a73c9a>



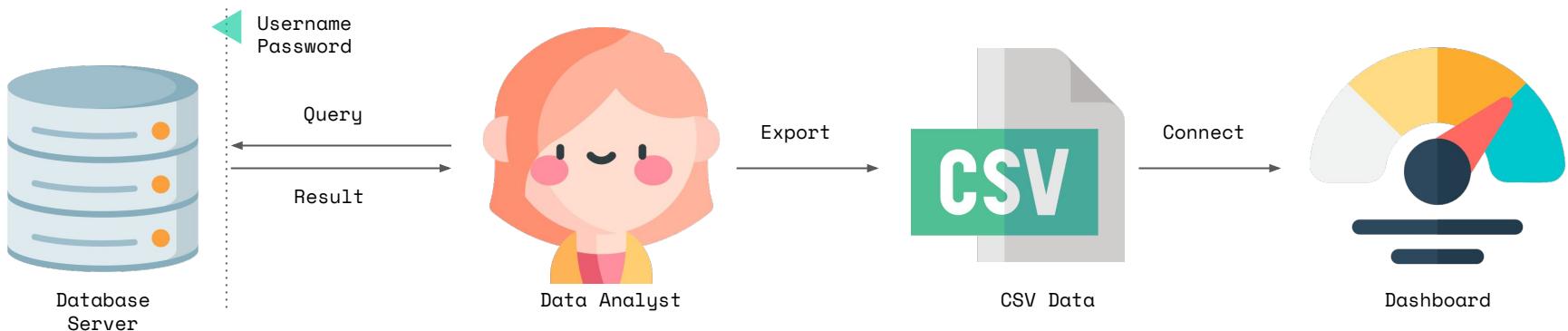
A photograph of a young woman with long brown hair and glasses, wearing a denim jacket over a pink turtleneck. She is smiling and looking towards the right. She is sitting on a black leather sofa, working on a silver laptop. The background is dark and out of focus, showing some vertical plants.

Data Analyst Workflow

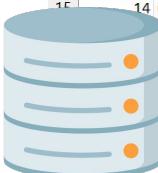
Data Analyst Workflow



Data Analyst Workflow V2



Structured data

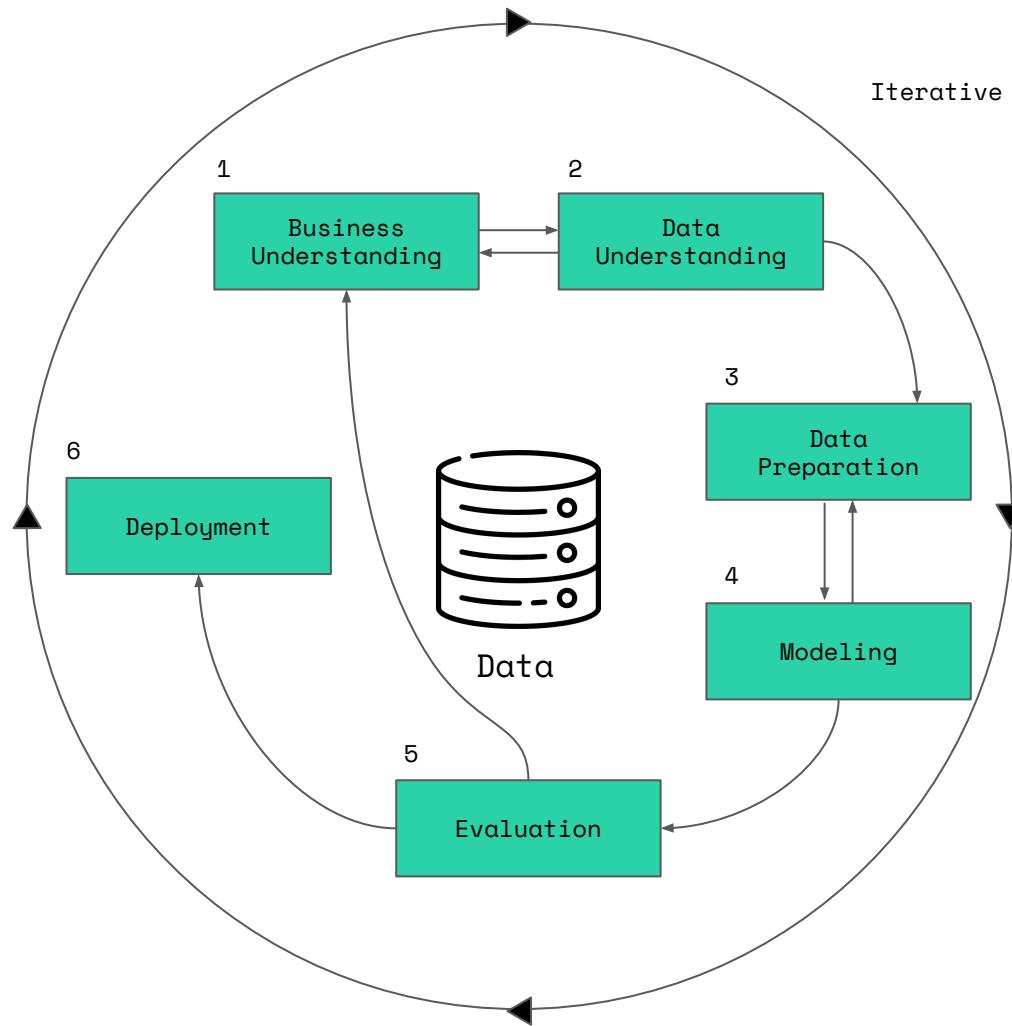


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer	Customer Segment	Country	City	State	Postal Code	Region	Product ID	Category	Sub-Catag	Product Name	Sales	Quantity	Discount	Profit	
2	1	CA-2016-1	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-BO-1	Furniture	Bookcases	Bush Some	261.96	2	0	41.9136
3	2	CA-2016-1	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-CH-1	Furniture	Chairs	Hon Delux	731.94	3	0	219.582
4	3	CA-2016-1	6/12/2016	6/16/2016	Second Class	DV-13045	Darrin Van	Corporate	United States	Los Angeles	California	90036	West	OFF-LA-10	Office Supply	Labels	Self-Adhes	14.62	2	0	6.8714
5	4	US-2015-1	10/11/2015	10/18/2015	Standard	CSO-20335	Sean O'Do	Consumer	United States	Fort Lauderdale	Florida	33311	South	FUR-TA-1C	Furniture	Tables	Bretford C	957.5775	5	0.45	-383.031
6	5	US-2015-1	10/11/2015	10/18/2015	Standard	CSO-20335	Sean O'Do	Consumer	United States	Fort Lauderdale	Florida	33311	South	OFF-ST-10	Office Supply	Storage	Eldon Fold	22.368	2	0.2	2.5164
7	6	CA-2014-1	6/9/2014	6/14/2014	Standard	BH-11710	Brosina Hc	Consumer	United States	Los Angeles	California	90032	West	FUR-FU-1C	Furniture	Furnishing	Eldon Express	48.86	7	0	14.1694
8	7	CA-2014-1	6/9/2014	6/14/2014	Standard	BH-11710	Brosina Hc	Consumer	United States	Los Angeles	California	90032	West	OFF-AR-1C	Office Supply	Art	Newell 32"	7.28	4	0	1.9656
9	8	CA-2014-1	6/9/2014	6/14/2014	Standard	BH-11710	Brosina Hc	Consumer	United States	Los Angeles	California	90032	West	TEC-PH-1C	Technology	Phones	Mitel 5320	907.152	6	0.2	90.7152
10	9	CA-2014-1	6/9/2014	6/14/2014	Standard	BH-11710	Brosina Hc	Consumer	United States	Los Angeles	California	90032	West	OFF-BI-10	Office Supply	Binders	DXL Angle-	18.504	3	0.2	5.7825
11	10	CA-2014-1	6/9/2014	6/14/2014	Standard	BH-11710	Brosina Hc	Consumer	United States	Los Angeles	California	90032	West	OFF-AP-10	Office Supply	Appliances	Belkin F5C	114.9	5	0	34.47
12	11	CA-2014-1	6/9/2014	6/14/2014	Standard	BH-11710	Brosina Hc	Consumer	United States	Los Angeles	California	90032	West	FUR-TA-1C	Furniture	Tables	Chromcraft	1706.184	9	0.2	85.3092
13	12	CA-2014-1	6/9/2014	6/14/2014	Standard	BH-11710	Brosina Hc	Consumer	United States	Los Angeles	California	90032	West	TEC-PH-1C	Technology	Phones	Konftel 25	911.424	4	0.2	68.3568
14	13	CA-2017-1	4/15/2017	4/20/2017	Standard	CAA-10480	Andrew All	Consumer	United States	Concord	North Carolina	28027	South	OFF-PA-10	Office Supply	Paper	Xerox 196*	15.552	3	0.2	5.4432
15	14	CA-2016-1	12/5/2016	12/10/2016	Standard	CM-15070	Irene Mad	Consumer	United States	Seattle	Washington	98103	West	OFF-BI-10	Office Supply	Binders	Fellowes P	407.976	3	0.2	132.5922
	JS-2015-1	11/22/2015	11/26/2015	Standard	CHP-14815	Harold Pav	Home Office	United States	Fort Worth	Texas	76106	Central	OFF-AP-10	Office Supply	Appliances	Holmes Re	68.81	5	0.8	-123.858	
	JS-2015-1	11/22/2015	11/26/2015	Standard	CHP-14815	Harold Pav	Home Office	United States	Fort Worth	Texas	76106	Central	OFF-BI-10	Office Supply	Binders	Storex Dur	2.544	3	0.8	-3.816	
	CA-2014-1	11/11/2014	11/18/2014	Standard	CPK-19075	Pete Kriz	Consumer	United States	Madison	Wisconsin	53711	Central	OFF-ST-10	Office Supply	Storage	Stur-D-Sto	665.88	6	0	13.3176	
	CA-2014-1	5/13/2014	5/15/2014	Second Class	AG-10270	Alejandro I	Consumer	United States	West Jordan	Utah	84084	West	OFF-ST-10	Office Supply	Storage	Fellowes S	55.5	2	0	9.99	
	CA-2014-1	8/27/2014	9/1/2014	Second Class	ZD-21925	Zuschuss D	Consumer	United States	San Francisco	California	94109	West	OFF-AR-1C	Office Supply	Art	Newell 34"	8.56	2	0	2.4824	

90% of the time

CRISP - DM

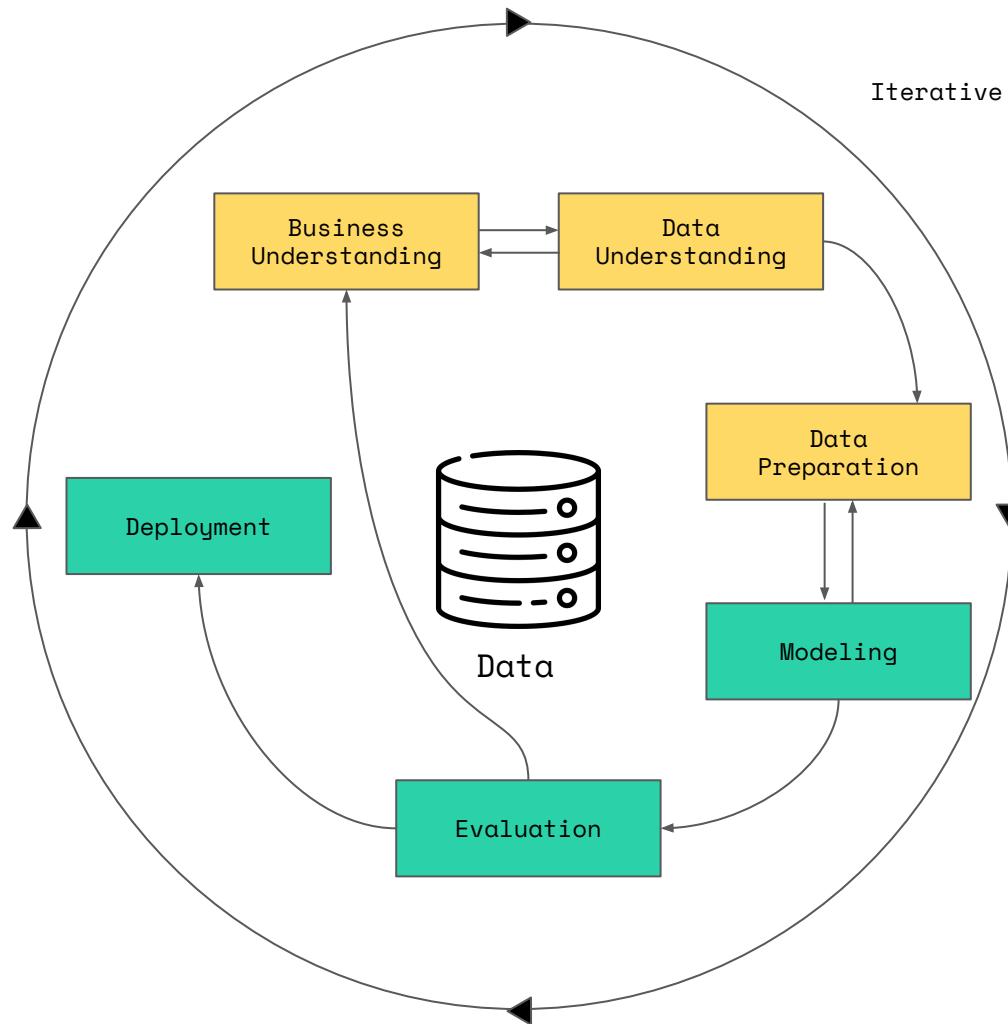
Cross industry standard process for
data mining



CRISP - DM

Cross industry standard process for data mining

80% spent on data collection, cleaning and organizing data



A photograph of a campsite at night. Two tents are set up in a clearing surrounded by trees. A campfire is burning in the center, with two people sitting around it. The scene is dimly lit by the fire and some ambient light, with stars visible in the dark sky.

Case Study

04 - Cambridge Analytica

Project Alamo

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



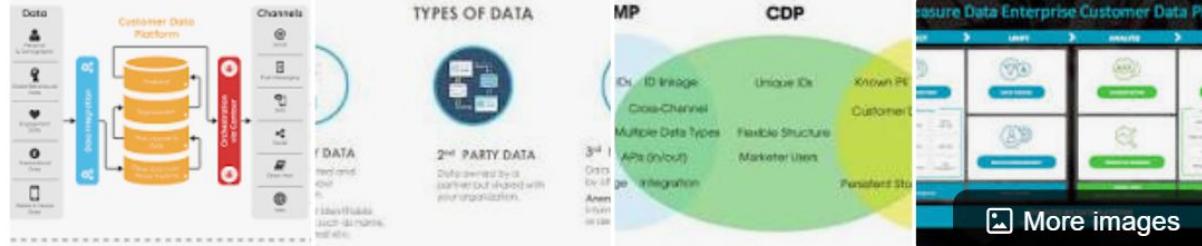
Project Alamo

From Wikipedia, the free encyclopedia

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

claimed to have
5,000 data points



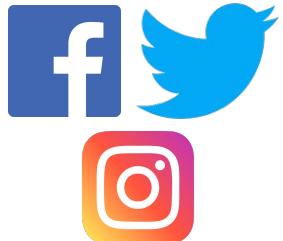


CDP data

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](https://en.wikipedia.org/wiki/Customer_data_platform)

[Customer data platform - Wikipedia](https://en.wikipedia.org/wiki/Customer_data_platform)



Digital Footprint

TED



[Carole Cadwalladr: Facebook's role in Brexit -- and the threat to democracy | TED Talk](#)



14:19



A portion of the American flag is visible in the top left corner.

blognone



Cambridge
Analytica



EVERYBODY LIES

BIG DATA, NEW DATA,
AND WHAT THE INTERNET CAN TELL US ABOUT WHO
WE REALLY ARE

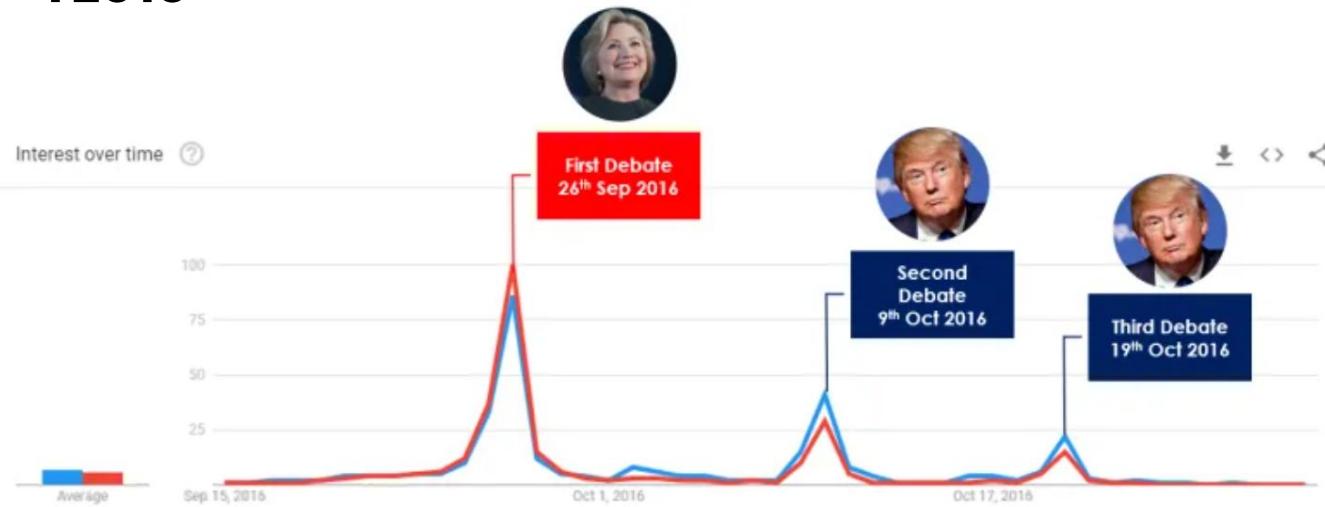


SETH STEPHENS-DAVIDOWITZ

FOREWORD BY STEVEN PINKER

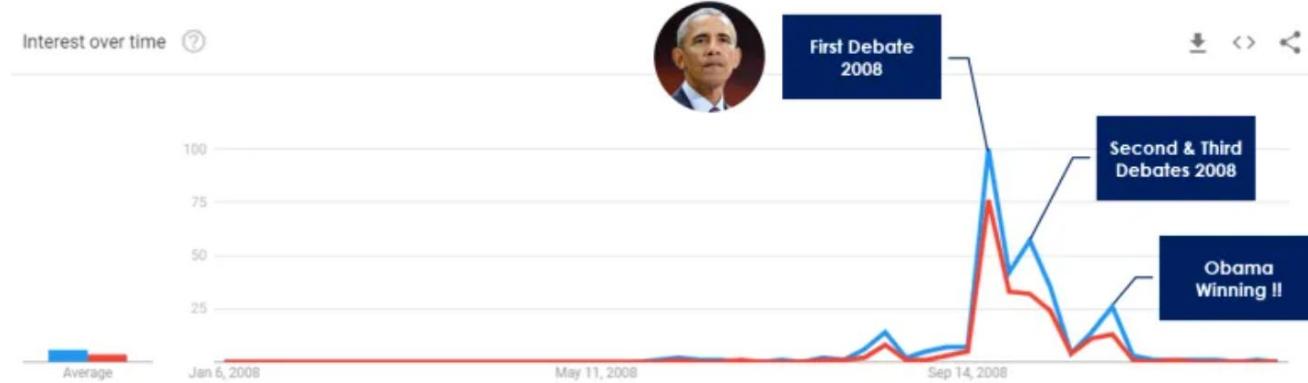
How? The order in which the candidates appear.
Our research suggests that a person is significantly
more likely to put the candidate they support first in a
search that includes both candidates' names.

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>



Will AI Replace My Job?





CaseyNeistat ✓

@casey • American YouTuber

🔔 Subscribed ▾



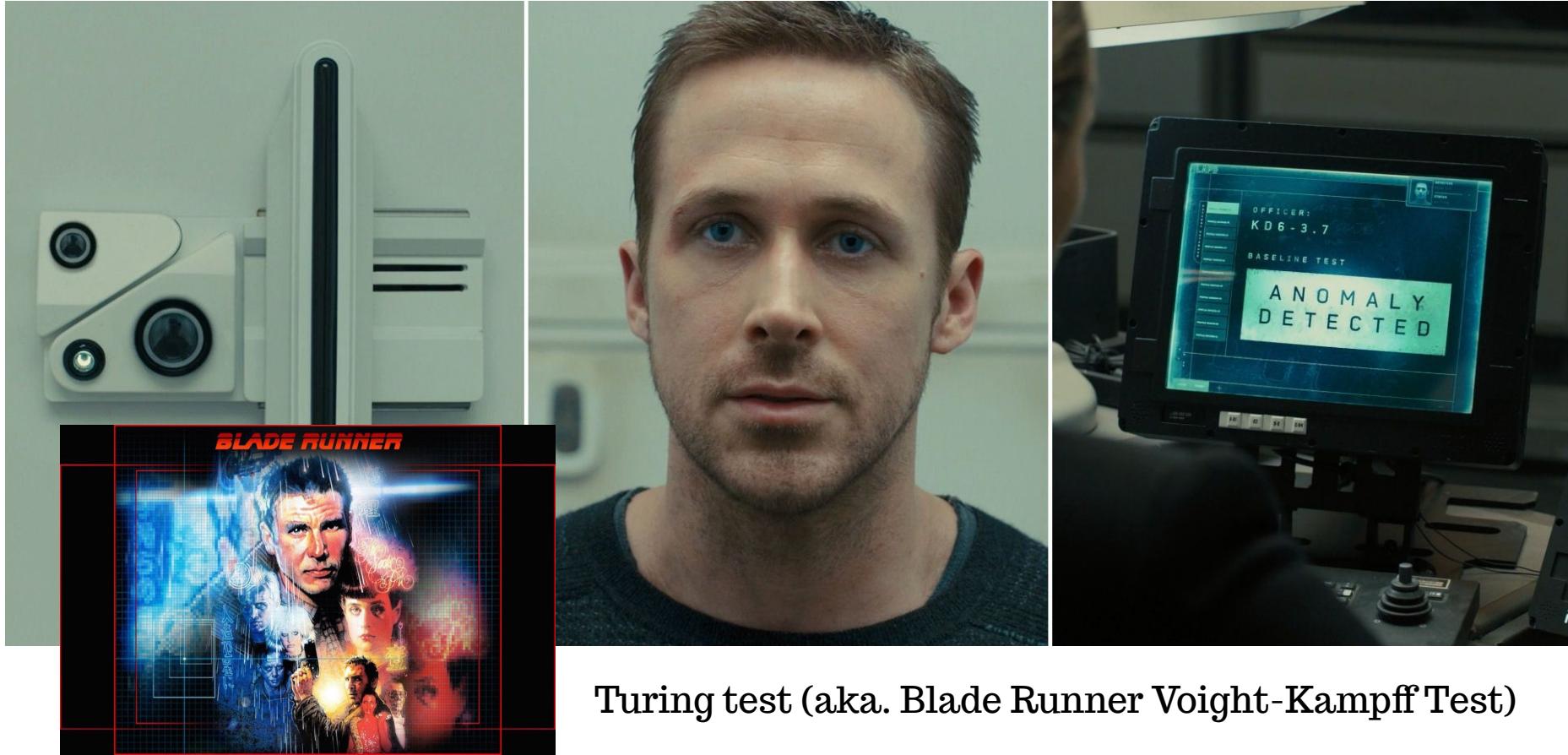


It has no humanity, no
depth, no soul

เป็นวีดีโอที่ห่วยที่สุดกีบ้ม(กู)เคย์ทำเลย 555+

that was the worst that was the worst
video I've ever made that video sucked

CASEY
NEISTAT



Turing test (aka. Blade Runner Voight-Kampff Test)



Geoffrey Hinton
@geoffreyhinton

Extrapolating the spectacular performance of GPT3 into the future suggests that the answer to life, the universe and everything is just 4.398 trillion parameters.

3:26 AM · Jun 11, 2020

666 Retweets

80 Quotes

3,761 Likes



82 Bookmarks

<https://twitter.com/geoffreyhinton/status/1270814602931187715?lang=en>

A young woman with long brown hair and glasses is smiling broadly while sitting on a dark leather couch. She is wearing a pink turtleneck and a blue denim jacket. She is holding a silver laptop on her lap. The background is dark, with some green plants visible on the right side.

How to Learn





PRODUCTIVITY

สรุป 8 เทคนิคการเรียนรู้อย่างมีประสิทธิภาพโดย Barbara Oakley



BY KASIDIS SATANGMONGKOL
April 5, 2019



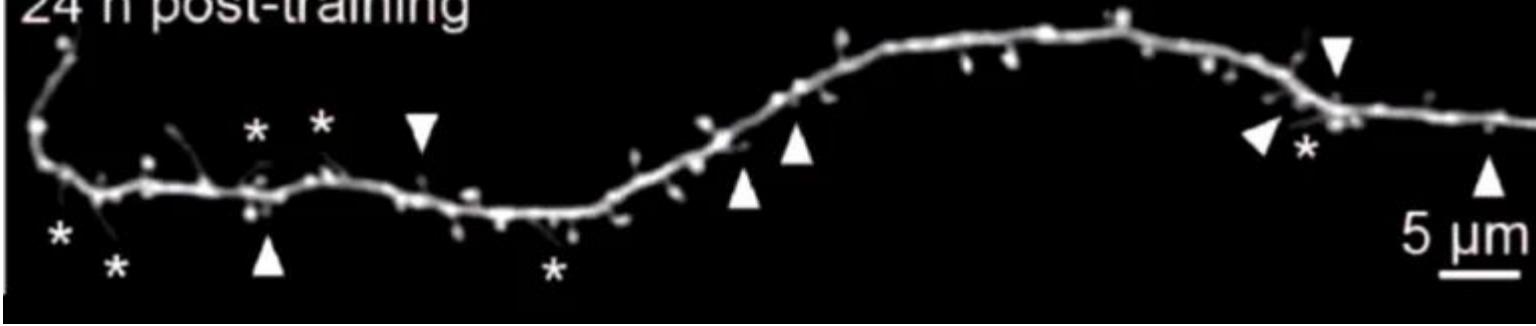
20 Comments

[สรุป 8 เทคนิคการเรียนรู้อย่างมีประสิทธิภาพโดย Barbara Oakley | DataRockie](#)

Pre-training



24 h post-training



- Good **Sleep**
- Good **Food**
- Good **Exercise**
- Drink a lot of **Water**





Learn **One Step** at a Time



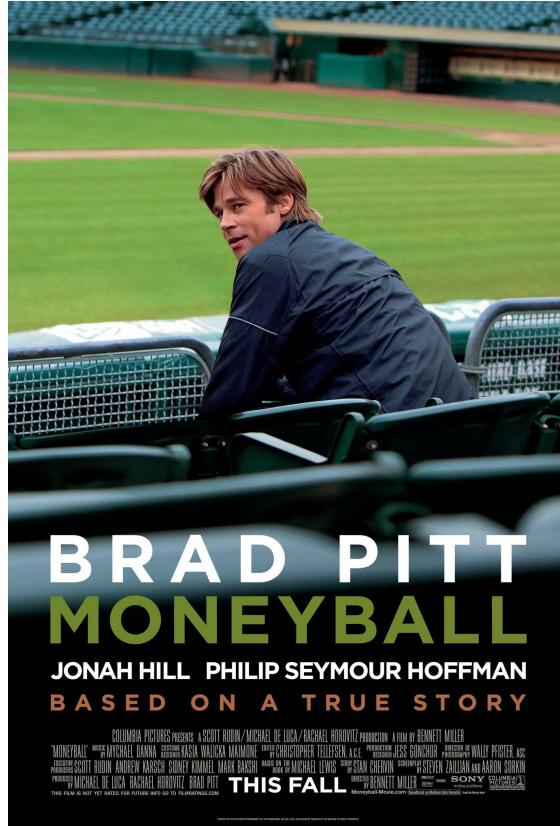
A photograph of a young woman with long brown hair and glasses, wearing a denim jacket over a pink turtleneck. She is smiling and looking towards the right. She is sitting on a dark leather couch, holding a silver laptop. The background is dark with some green plants visible.

Final Case Study



Moneyball

The key to success is getting the right data and finding the right attributes.



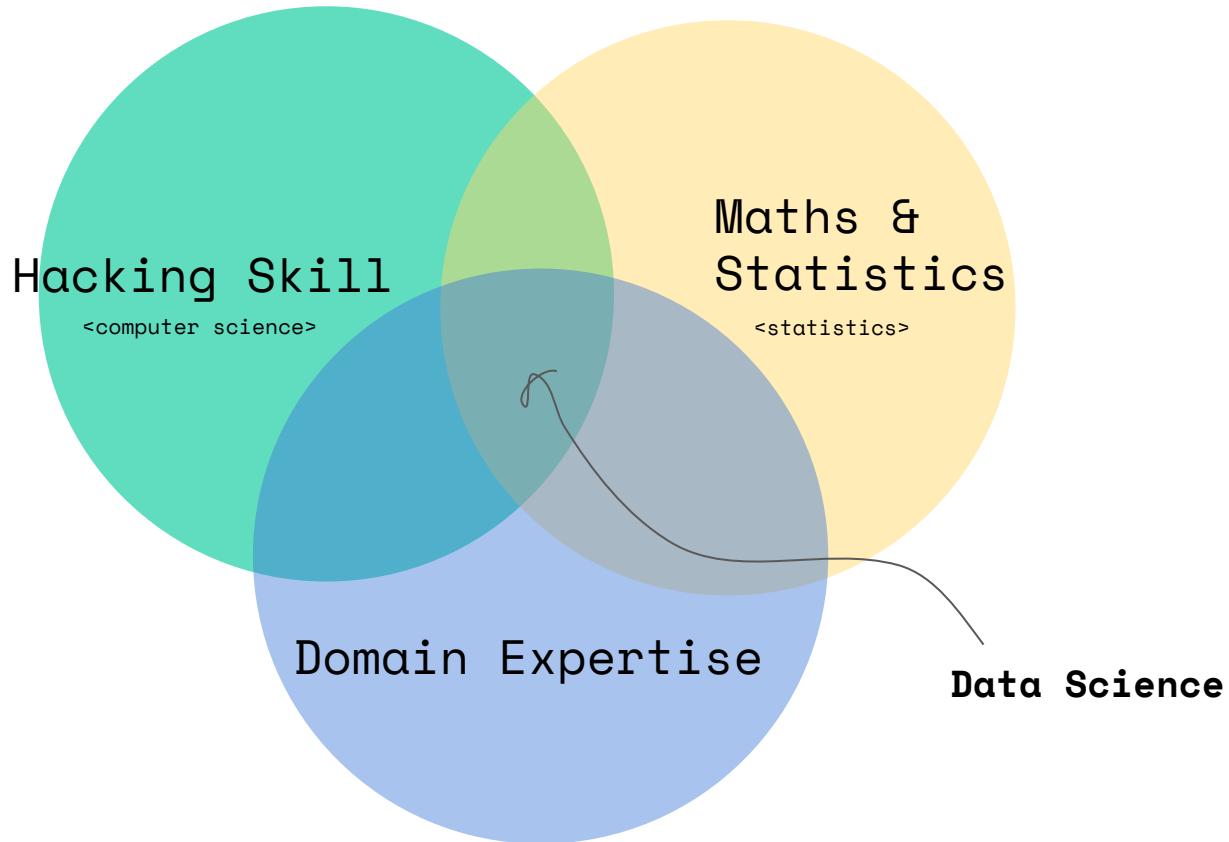


Coach

Finding great players,
undervalued prices

Economist

Using a [linear] regression
model to find right attributes



Drew Conway

[The Data Science Venn Diagram — Drew Conway](#)

THE DATA SCIENCE VENN DIAGRAM

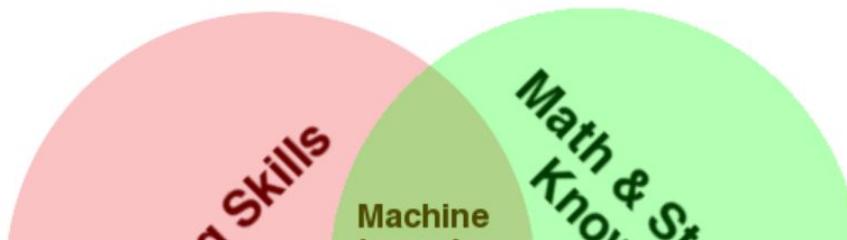
On Monday I—humbly—joined a group of NYC's most sophisticated thinkers on all things data for a half-day **unconference** to help O'Reilly organize their upcoming **Strata** conference. The break out sessions were fantastic, and the number of people in each allowed for outstanding, expert driven, discussions. One of the best sessions I attended focused on issues related to teaching data science, which inevitably led to a discussion on the skills needed to be a fully competent data scientist.

As I have said before, I think the term "data science" is a bit of a misnomer, but I was very hopeful after this discussion; mostly because of the utter lack of agreement on what a curriculum on this subject would look like. The difficulty in defining these skills is that the split between substance and methodology is ambiguous, and as such it is unclear how to distinguish among hackers, statisticians, subject matter experts, their overlaps and where data science fits.

What is clear, however, is that one needs to learn a lot as they aspire to become a fully competent data scientist. Unfortunately, simply enumerating texts and tutorials does not untangle the knots. Therefore, in an effort to simplify the discussion, and add my own thoughts to what is already a crowded market of ideas, I present the Data Science Venn Diagram.



Drew Conway







Data Science Skills

Data



Insights

Insights

Insights

Insights

Insights

ยินดีต้อนรับทุกคนเข้าสู่ **Bootcamp**
ที่เข้มที่สุด ในประเทศไทย

A photograph of a modern living room with exposed red brick walls. In the foreground, a woman sits on the floor, looking at a laptop. In the center, two men sit on a dark green sofa; one is using a laptop while the other looks on. To the left, a woman stands by a window, talking on her phone. The room is decorated with several potted plants and a small white side table.

Data Science Bootcamp

Intro to Data Science