fields of the future
speaker series

data science
with Kevin Hussain Hassnain

Industry Expert

Senior Data Scientist,
LivaNova, PLC

High School / Undergrad Students & Parents:
Join Kevin to hear more about:

- the field of data science
- what a data scientist does
- Kevin's educational journey
- characteristics of a successful data scientist professional
- applications and challenges of data science in the healthcare industry

Saturday, May 23, 2020
2 p.m. PT / 4 p.m. CT / 5 p.m. ET

Aga Khan Education Board for the United States of America

IPN

Register to attend:
akebsw.org/fof2020

Zoom meeting details will be shared after registration

What is Data Science?

By: Kevin Hassnain

Agenda

Introduction

- Personal & Professional Journey

What is Data Science?

- Definitions
- Data Jiujitsu

Essential skills and traits of an elite Data Scientist

- Education
- Continued Learning
- Passion Project
- Traits

Real world Data Science Applications and Challenges

- Amazing Tale
- Dr. Google
- Challenges

Future of Data Science

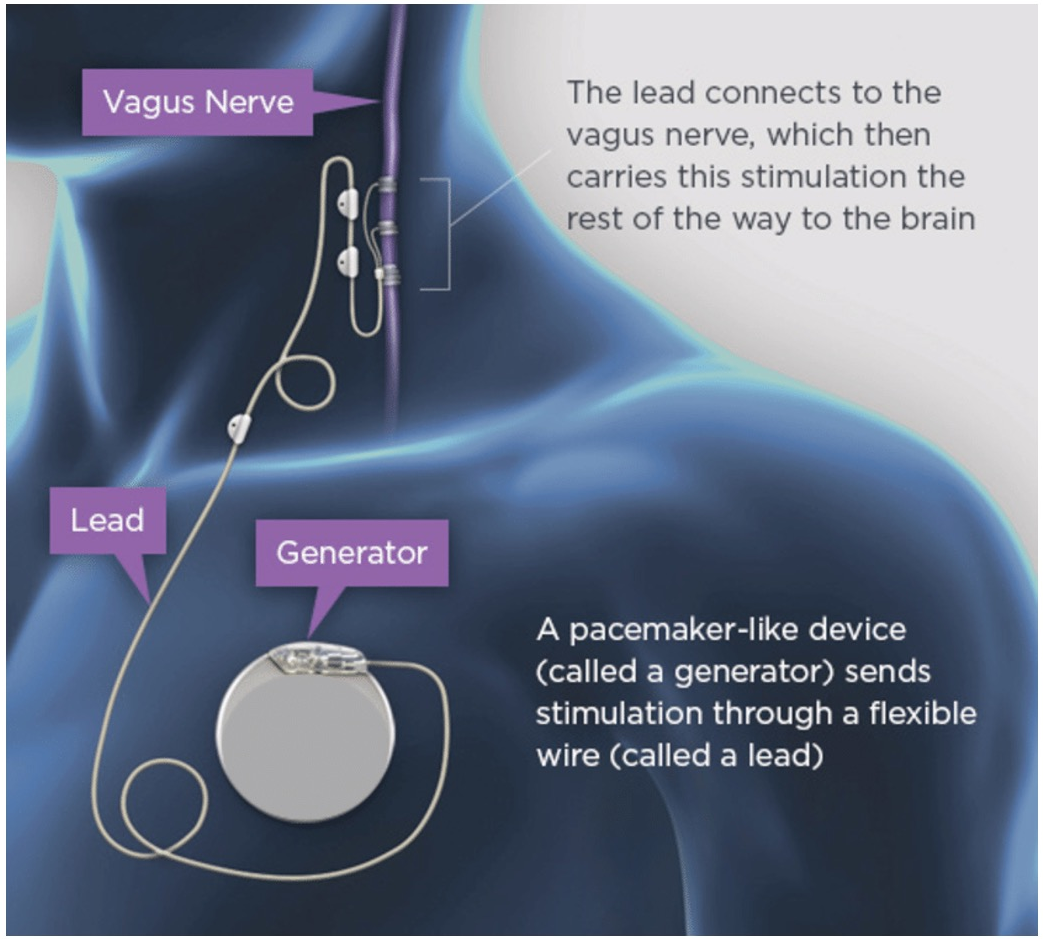
- Technological advancements
- Precision Medicine

Aga Khan Education Board
for the Southwestern United States

TEXAS A&M
UNIVERSITY

LivaNova
Health innovation that matters

LivaNova - Data Science Responsibilities



LivaNova

Health innovation that matters

What is Data?

Structured data

OlympicsData.csv - Notepad

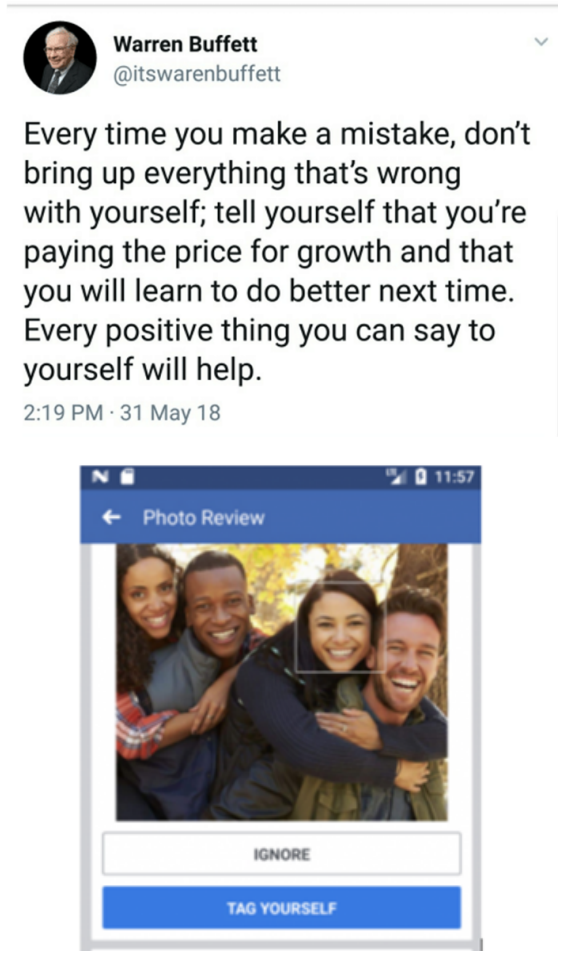
File Edit Format View Help

```
MedalistID;Edition;Athlete;NOC;Gender;Event;Medal;Season;MedalValue;SportI
1;1900;HALMAY, Zoltan;HUN;Men;1500m freestyle;Bronze;Summer;3;1;56;
2;1900;JARVIS, John Arthur;GBR;Men;1500m freestyle;Gold;Summer;1;1;56;
3;1900;WAHLE, Otto;AUT;Men;1500m freestyle;Silver;Summer;2;1;56;
4;1900;DROST, Johannes;NED;Men;200m backstroke;Bronze;Summer;3;1;56;
5;1900;HOPPENBERG, Ernst;GER;Men;200m backstroke;Gold;Summer;1;1;56;
6;1900;RUBERL, Karl;AUT;Men;200m backstroke;Silver;Summer;2;1;56;
7;1900;HALMAY, Zoltan;HUN;Men;200m freestyle;Silver;Summer;2;1;56;
8;1900;RUBERL, Karl;AUT;Men;200m freestyle;Bronze;Summer;3;1;56;
9;1900;LANE, Frederick C.V.;AUS;Men;200m freestyle;Gold;Summer;1;1;56;
10;1900;KEMP, Peter;GBR;Men;200m obstacle event;Bronze;Summer;3;1;56;
11;1900;WAHLE, Otto;AUT;Men;200m obstacle event;Silver;Summer;2;1;56;
12;1900;LANE, Frederick C.V.;AUS;Men;200m obstacle event;Gold;Summer;1;1;5
13;1900;FREY, Julius;GER;Men;200m team swimming;Gold;Summer;1;1;56;
14;1900;HAINLE, Max;GER;Men;200m team swimming;Gold;Summer;1;1;56;
15;1900;HOPPENBERG, Ernst;GER;Men;200m team swimming;Gold;Summer;1;1;56;
16;1900;SCHÖNE, Max;GER;Men;200m team swimming;Gold;Summer;1;1;56;
17;1900;VON PETERSDORFF, Herbert;GER;Men;200m team swimming;Gold;Summer;1;
18;1900;BERTRAND;FRA;Men;200m team swimming;Silver;Summer;2;1;56;
19;1900;CADET;FRA;Men;200m team swimming;Silver;Summer;2;1;56;
20;1900;HOCHEPIED, Maurice;FRA;Men;200m team swimming;Silver;Summer;2;1;56
21;1900;HOCHEPIED, Victor;FRA;Men;200m team swimming;Silver;Summer;2;1;56;
22;1900;VERBECKE;FRA;Men;200m team swimming;Silver;Summer;2;1;56;
23;1900;HOUBEN;FRA;Men;200m team swimming;Bronze;Summer;3;1;56;
24;1900;LEUILLIEUX, Jean;FRA;Men;200m team swimming;Bronze;Summer;3;1;56;
25;1900;MARTIN, Louis;FRA;Men;200m team swimming;Bronze;Summer;3;1;56;
26;1900;MERCHEZ, Désiré;FRA;Men;200m team swimming;Bronze;Summer;3;1;56;
27;1900;TARTARA;FRA;Men;200m team swimming;Bronze;Summer;3;1;56;
```

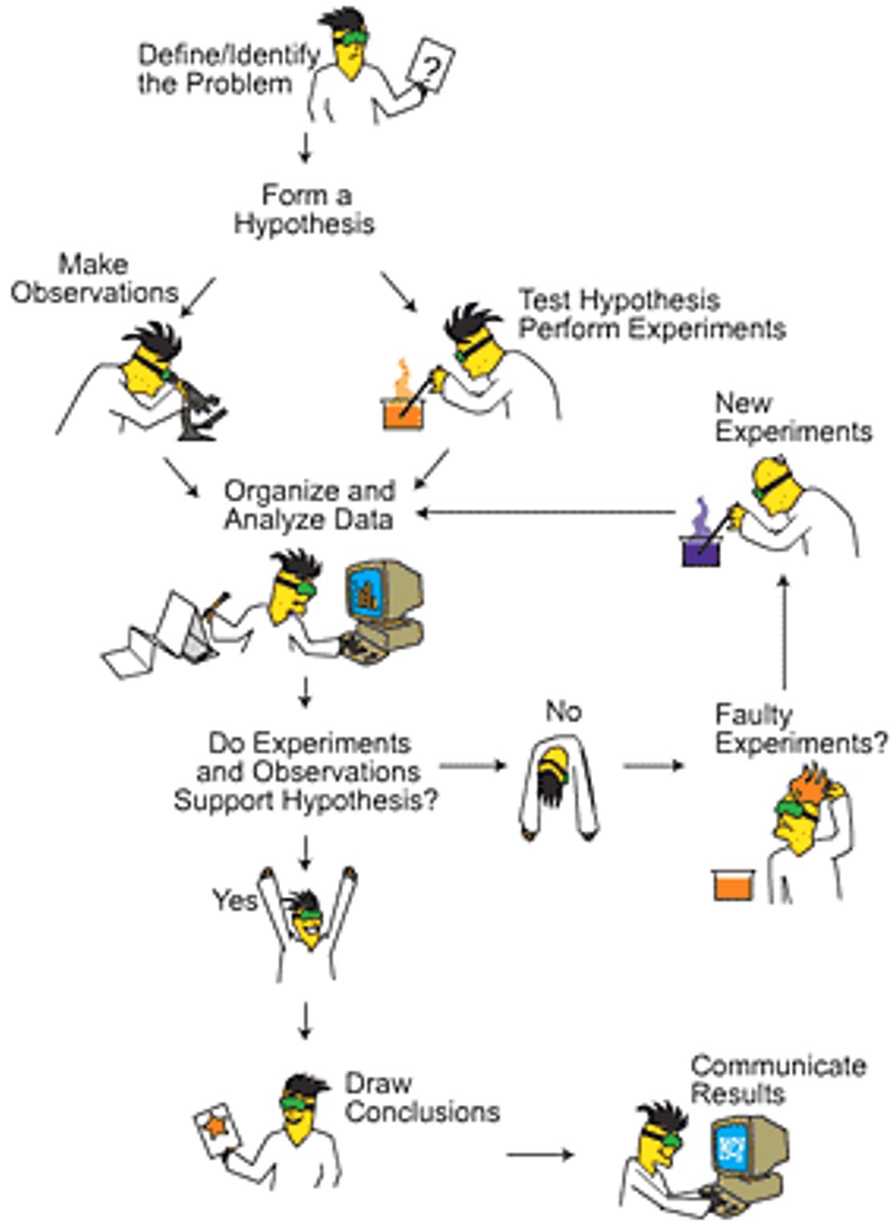
B2 | Calculus 101

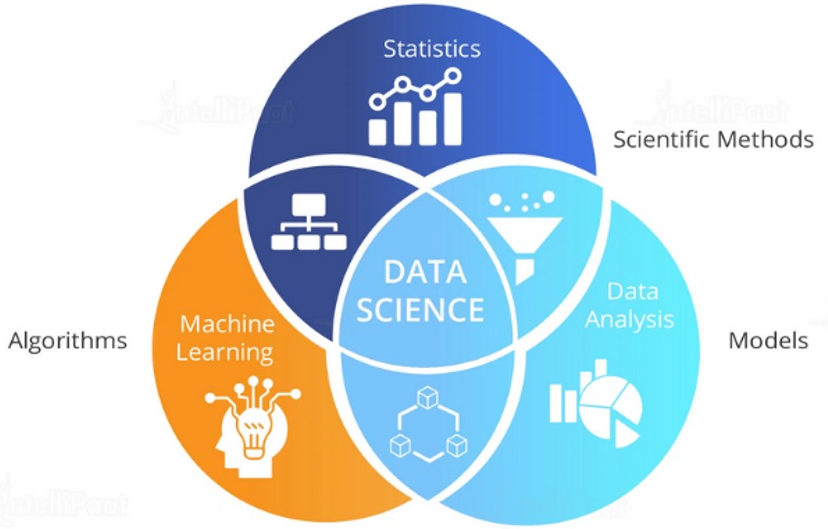
	A	B	C	D	E
1	Student ID	Class	Teacher	Final Grade	
2	VM4128	Calculus 101	Mr. Crasdale	77	
3	CH8965	Ancient Greece	Mr. Young	81	
4	SG9555	Accounting 101	Ms. Morley	76	
5	AB5235	Calculus 101	Mr. Crasdale	86	
6	KD0656	Accounting 101	Mr. Harris	98	
7	JG1183	Accounting 101	Mr. Harris	98	
8	HE7976	Masters of Philosophy	Mr. Crasdale	78	
9	NR5090	Calculus 101	Mr. Porter	79	
10	AJ3549	Accounting 101	Mr. Harris	85	
11	CW8495	English Literature	Mr. Johnson	66	
12	NR5090	Calculus 101	Mr. Porter	88	
13	QT6233	Accounting 101	Ms. Morley	73	
14	MP9632	English Literature	Ms. Rendson	84	
15	DK7492	Accounting 101	Mr. Richards	71	
16	RE3968	Ancient Greece	Mr. Young	75	
17	AP8356	Accounting 101	Mr. Richards	88	
18	FL1423	Masters of Philosophy	Ms. Untermeyer	93	
19	BW2559	Calculus 101	Mr. Porter	86	
20	QL9026	Calculus 101	Mr. Crasdale	77	
21	FE4903	English Literature	Ms. Appleson	74	
22	TY3987	Calculus 101	Mr. Porter	74	
23	EE4688	English Literature	Mr. Johnson	73	
24	WD0448	Ancient Greece	Mr. Young	76	
25	GY7754	English Literature	Mr. Johnson	66	

Unstructured data



What is Science?





Data Jiujitsu

It's not the size of the data, It's the size of the outcome

Education

Pick a relevant major

Preferably STEM major: Math, Statistics, Engineering or Physical Sciences

Major doesn't have to be specifically math-focused and data-focused, but you must be able to apply data to the field you study.

Pick the right classes

Focus on two main subjects during undergrad: Statistics and computer science

Data Science courses include: Time Series & Categorical Data Analysis, Survival analysis, Supervised & Unsupervised ML methods and Multivariate Data Analysis

Programming Language & Tools: Python, R, SQL, SAS, EXCEL macro language (VBA) & JMP

Real world Application

Show genuine interest in the focus of research

Display curiosity by talking with professors after class or office hours

Participate in Professor led research or create your own passion project

Learn from doing and early failing (build confidence)

Continued Learning

Graduate School is almost mandatory for Data Science Placement

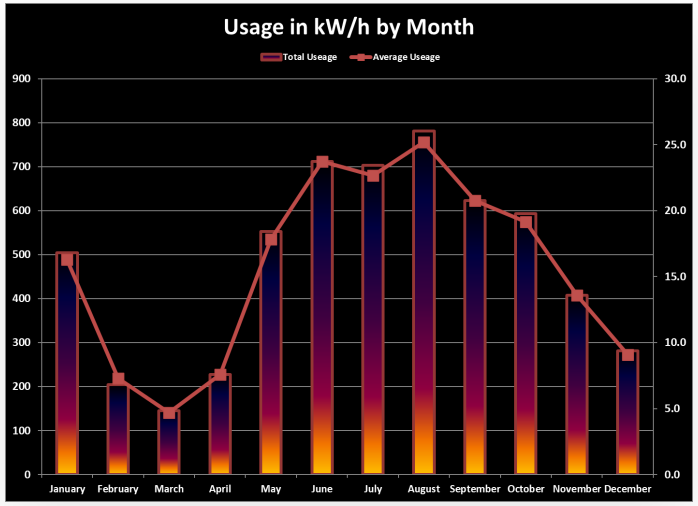
~91% of Data Scientist placements are from graduate program

Continued learning even beyond traditional schooling

Continued Learning

- **Kaggle Competitions**
  - Crowd source platform to attract, nurture, train and challenge data scientists from all around the world to solve data science, machine learning and predictive analytics problems
- **Online Learning Portals: Data Camp**
  - Learn data science at your own pace from real world examples
- **Online Certificate Programs**
  - MIT Certificate Program: Data Science & Big Data Analytics
    - https://learn-xpro.mit.edu/data-science
  - DOJO: Data Science & Data Engineering
    - https://datasciencedojo.com/
- **Annual Conferences**
  - Example: Predictive Analytics Workshop
    - https://www.predictiveanalyticsworld.com/machinelearningweek/workshops/predictive-analytics-applied-an-online-introduction/
  - Industry led experts provide pertinent trainings and discuss new trends and discoveries in Data Science
  - Students can get full scholarships or super discounted rates!!

Example of a Passion Project



ENERGY USAGE CALCULATOR

Location	My House		
Cost/kWh	0.08	$	
Total Rating/Hour	4,305.00	Watt	

Maximum consumption per day	19.78	kWh
Average consumption per day	19.28	kWh
Total Consumption per month	578.25	kWh
Total Monthly Cost	43.37	$

No	Appliances	Rating (W)	Hourly Usage per Day	# of Units	Consumption per Day	Day Frequency Usage per Week	Consumption per Week	Day Frequency Usage per Month	Consumption per Month	Monthly Cost
1	Television - Samsung	150.00	5.0	1	0.75	7.0	5.25	30.0	22.50	$ 1.69
2	Television - Sony	125.00	6.0	1	0.75	7.0	5.25	30.0	22.50	$ 1.69
3	Television - Samsung 2	480.00	1.0		0.48		3.36		14.40	$ 1.08
4	WiFi Modem	10.00	24.0		0.24		1.68		7.20	$ 0.54
5	Cable TV Setup Box	25.00	24.0		0.60		4.20		18.00	$ 1.35
6	Internet Modem	10.00	24.0	1	0.24		1.68		7.20	$ 0.54
7	Refrigerator	105.00	3.0	1	0.32	7.0	2.21		9.45	$ 0.71
8	Water Heater	1,000.00	0.5	1	0.50	5.0	2.50		15.00	$ 1.13
9	Microwave	600.00	24.0	1	14.40	7.0	100.80		432.00	$ 32.40
10	Dishwasher	1,200.00	1.0	1	1.20	5.0	6.00	20.0	24.00	$ 1.80
11	Dispenser	600.00	0.5	1	0.30	5.0	1.50	20.0	6.00	$ 0.45
12			4.0	1		7.0				
13										
14										
15										
16										
17										
18										
19										
20										

Instructions <-- *Using smart meter texas website, dump the usage data below to estimate bill for next month*

Date	Start Read	End Read	Usage(kWh)
1/1/2018	3554.408	3588.294	33.886
1/2/2018	3588.294	3632.133	43.839
1/3/2018	3632.133	3662.295	30.162
1/4/2018	3662.295	3684.926	22.631
1/5/2018	3684.926	3708.497	23.571
1/6/2018	3708.497	3728.459	19.962
1/7/2018	3728.459	3743.893	15.434
1/8/2018	3743.893	3759.368	15.475
1/9/2018	3759.368	3772.394	13.026
1/10/2018	3772.394	3782.828	10.434
1/11/2018	3782.828	3792.288	9.46
1/12/2018	3792.288	3807.968	15.68
1/13/2018	3807.968	3826.309	18.341
1/14/2018	3826.309	3846.745	20.436
1/15/2018	3846.745	3858.927	12.182
1/16/2018	3858.927	3889.781	30.854
1/17/2018	3889.781	3922.395	32.614
1/18/2018	3922.395	3961.397	39.002
1/19/2018	3961.397	3988.915	27.518
1/20/2018	3988.915	3998.294	9.379
1/21/2018	3998.294	4002.577	4.283
1/22/2018	4002.577	4006.68	4.103
1/23/2018	4006.68	4010.825	4.145
1/24/2018	4010.825	4014.882	4.057
1/25/2018	4014.882	4022.258	7.376
1/26/2018	4022.258	4029.969	7.711
1/27/2018	4029.969	4035.255	5.286
1/28/2018	4035.255	4039.473	4.218
1/29/2018	4039.473	4045.619	6.146
1/30/2018	4045.619	4050.29	4.671

USAGE INSIGHTS

Volt EP Plan E

	Usage (kWh)	Fixed	Multiplier	Total (without taxes)
1	1-500	$ 40.00	$ -	$ 40.00
501	501-1000	$ 80.00	$ -	$ 80.00
1001	1001-1500	$ 120.00	$ -	$ 120.00
1501	1501-2000	$ 160.00	$ -	$ 160.00
	2000+	$ 160.00	$0.125	$160 + x*0.125

Switch	ON
Select Month of Interest	February
Select Year of Interest	2018

Website	https://www.smartmetertexas.com
Username	kevinhassnain
Password	

Dashboard

Current Usage (kWh)	218.938
Days Elapsed	31
Days in Billing Cycle	31
Average usage per Day (kWh)	7.1
Total Projected usage in kW	218.9
Total Project Bill	$ 40.00

Rank by Season

Season	Coverage	Season Rank
Autumn	Nov - Jan	1
Spring	Feb - Apr	4
Summer	May - Jul	2
Winter	Aug - Oct	3

high rank correspond to high usage

Rank by Month

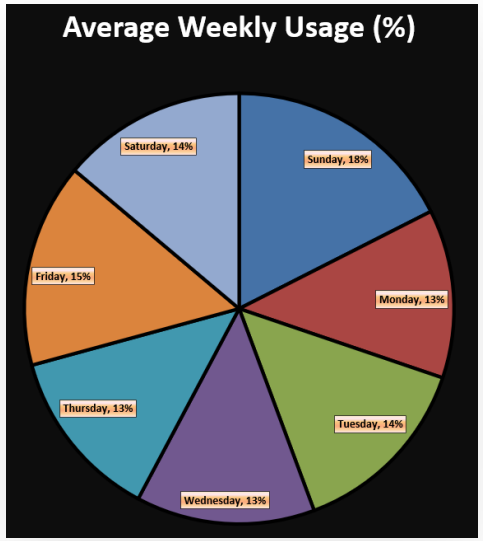
Season	Season Rank
January	7
February	11
March	12
April	10
May	6
June	2
July	3
August	1
September	4
October	5
November	8
December	9

high rank correspond to high usage

Rank by Day of the Week

Season	Season Rank
Sunday	1
Monday	7
Tuesday	3
Wednesday	5
Thursday	6
Friday	2
Saturday	4

high rank correspond to high usage



Traits of an Elite Data Scientist

- **Domain Base Understanding**
- **Analytics & Hypotheses driven**
- **Curiosity & Creativity**
- **Storytelling ability**

What do I enjoy most about this career

Knowns & Unknowns

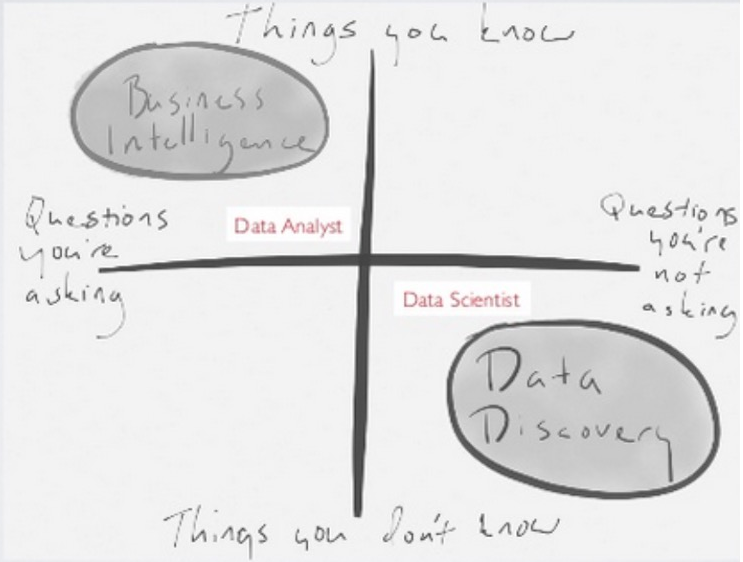
There are known knowns. These are things we know that we know.

There are known unknowns. That is to say, there are things that we know we don't know.

But there are also unknown unknowns. There are things we don't know we don't know

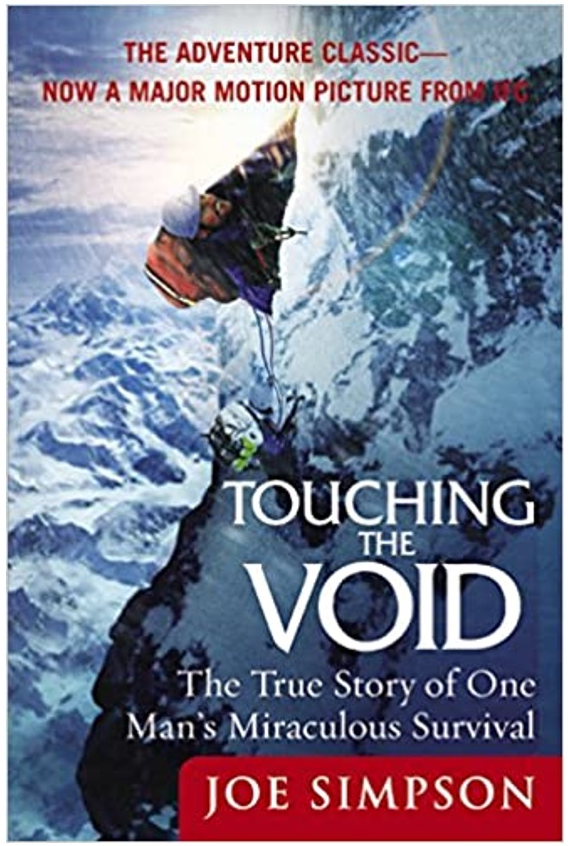
Donald Rumsfeld

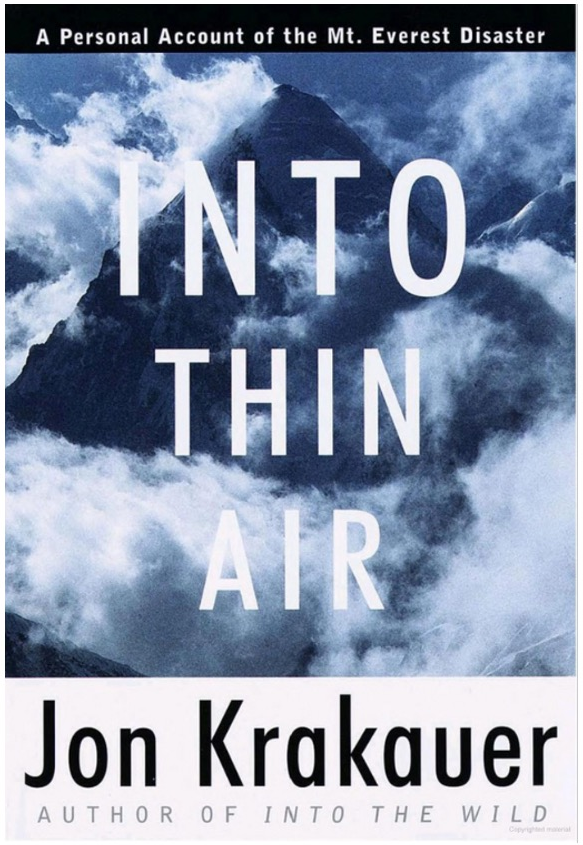
Data Discovery



new reality for Business Intelligence and Big Data, Applied Data Labs

The Tale of an Amazing Recommendation

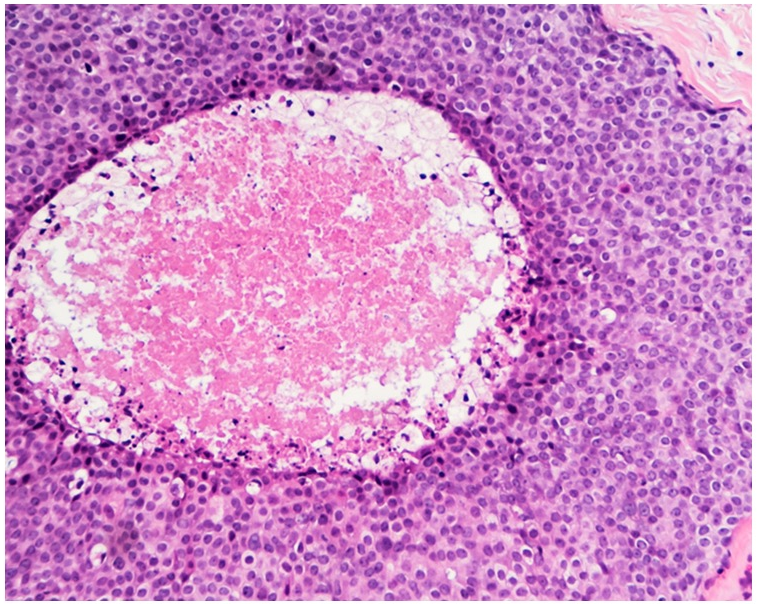




https://www.wired.com/2004/10/tail/

New sheriff in town – Dr. Google

- Google's Deep Learning AI predict abnormal biopsies of a breast cancer patient with much better Accuracy than a Human Pathologist
  - 89% vs. 73%
  - Significantly quicker than the human review



Google Health

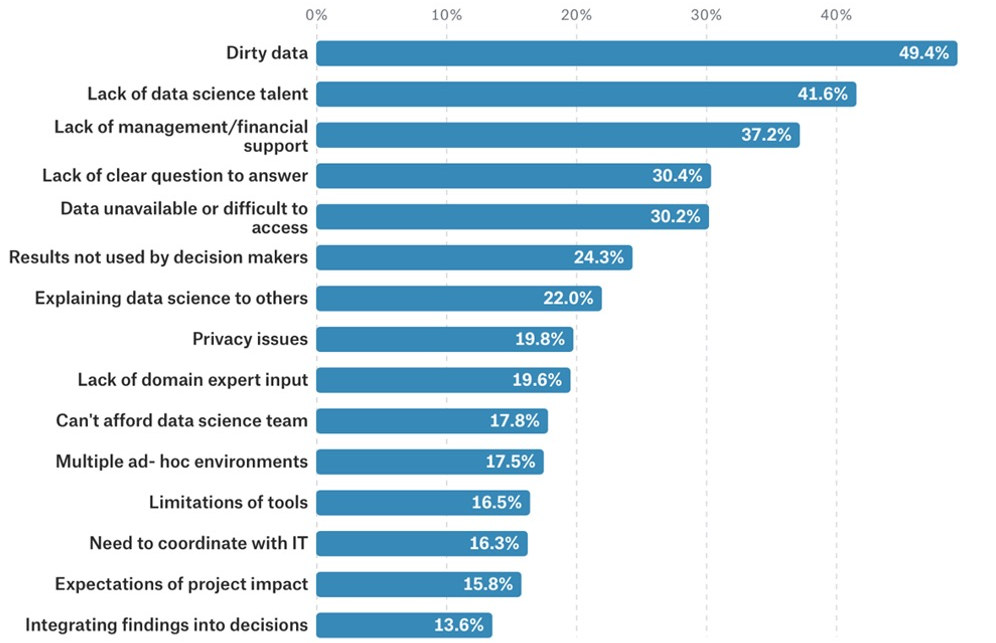
Google Deep

Neowin

https://www.mercurynews.com/2017/03/03/google-computers-trained-to-detect-cancer/

Applications of Data Science in Other Industries

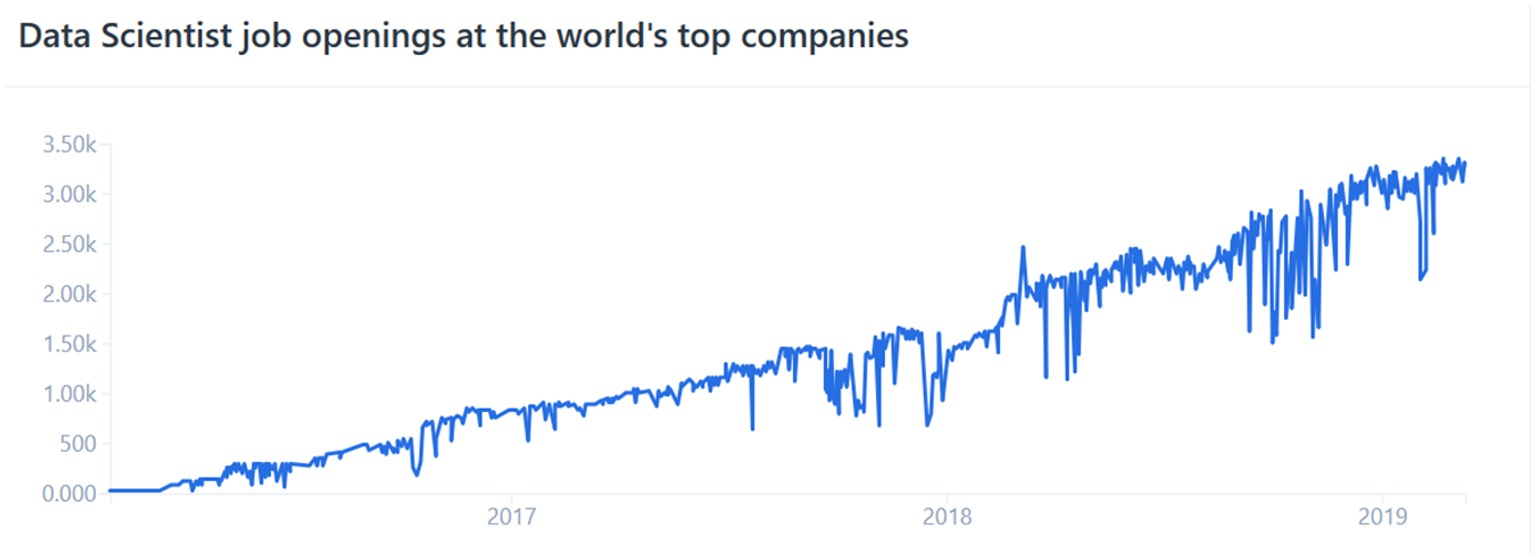
Data Science Challenges



7,376 responses

Only displaying the top 15 answers. There are 7 answers not shown.

Demand for Data Scientists



- Data scientist roles have grown more than 650 percent since 2012, according to an analysis on LinkedIn

https://www.forbes.com/sites/karstenstrauss/2017/12/07/the-fastest-growing-jobs-and-the-skills-needed-to-do-them/#6d9694f44e79

- IBM Predicts Demand For Data Scientists Will Soar 28% By end of 2020

https://www.forbes.com/sites/louiscolumbus/2017/05/13/ibm-predicts-demand-for-data-scientists-will-soar-28-by-2020/#749925ee7e3b

Technological advancements over time

M Model 350 disk file with a storage space of 5MB from 1956 and a Micro SD Car

1GB = 1024 MB

Moore's Law

the theory suggested that the speed of computer processors would double every two years. The transistors inside of computer chips would continue to decline in cost and size but increase in power.

Advancements in Healthcare – Precision Medicine

- Medicine continues to undergo a digital transformation
  - Electronic Health Records (EHR)
  - Data collected by wearable devices
- Collaboration between Data Scientist & Practitioners
- Use Data Science methods to enable more precise targeting of subgroups for diagnosis and treatment options
- Provide the right treatments to the right patients at the right time

Additional Resources & Contact

Structured vs. Unstructured Data
https://www.geeksforgeeks.org/difference-between-structured-semi-structured-and-unstructured-data/
Data Jujitsu
https://www.oreilly.com/data/free/data-jujitsu.csp
How to become a Data Scientist?
https://towardsdatascience.com/how-to-become-a-data-scientist-3f8d6e75482f
DataCamp: Learn Data Science with Python Certificate Program
https://www.datacamp.com/tracks/data-scientist-with-python
KAGGLE: Intro to Data Science – Step by Step guide to explore Titanic Exercise
https://www.kaggle.com/ydalat/titanic-a-step-by-step-intro-to-machine-learning
Harvard Business Review: Outlook of Data Science in the Future
https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century

CONTACT

- Email: kevinhassnain@gmail.com

data science
with Kevin Hussain Hassnain

fields of the future
speaker series

Q & A

Please submit your questions in the chat box below

Aga Khan Education Board
for the United States of America

IPN

Fields of the Future is brought to you by Aga Khan Education Board, Ismaili Professionals Network, and Ismaili Student Network

data science

with Kevin Hussain Hassnain

fields of the future
speaker series

Thank you for attending!

Please complete a brief evaluation

akebsw.org/feedback

Aga Khan Education Board
for the United States of America

IPN

Fields of the Future is brought to you by Aga Khan Education Board, Ismaili Professionals Network, and Ismaili Student Network

the.ismaili
USA

SOCIAL DISTANCING

Only leave home for work or for essential needs.

Wash your hands and disinfect surfaces regularly.

Maintain social distancing when in public.

Cover your face with a cloth face covering when in public.

Call a doctor if you are feeling unwell.

Call the ACCESS hotline at 1-844-55-ACCESS with any questions or concerns.

STAY HOME. SAVE LIVES.

1-844-55-ACCESS