

Data Science for Funsies!

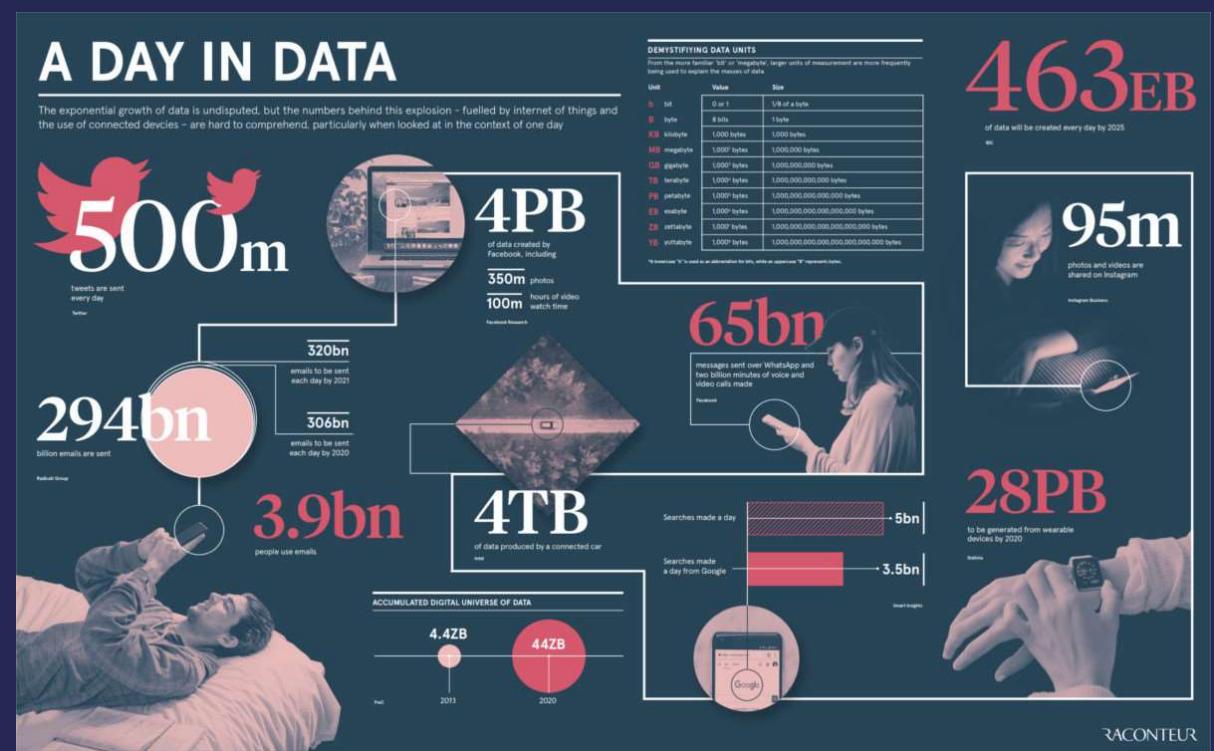
Brian Cornet

March 24th, 2021

Here's what's up!

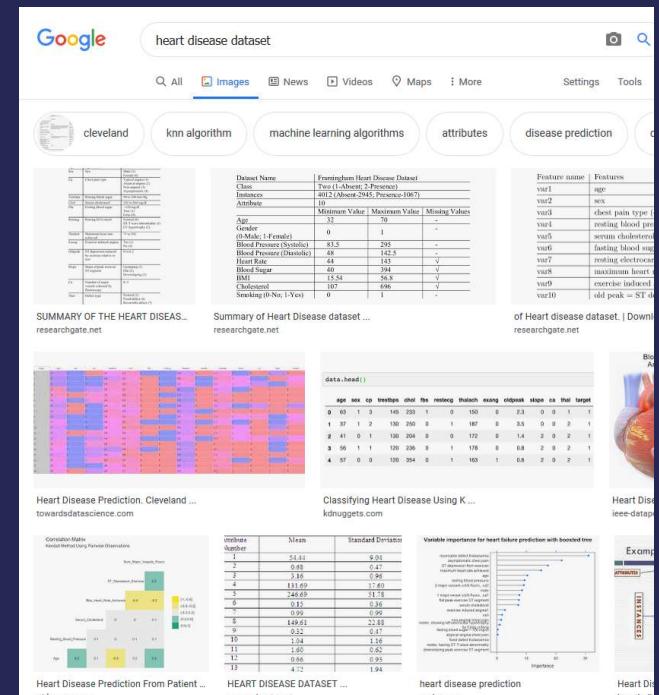
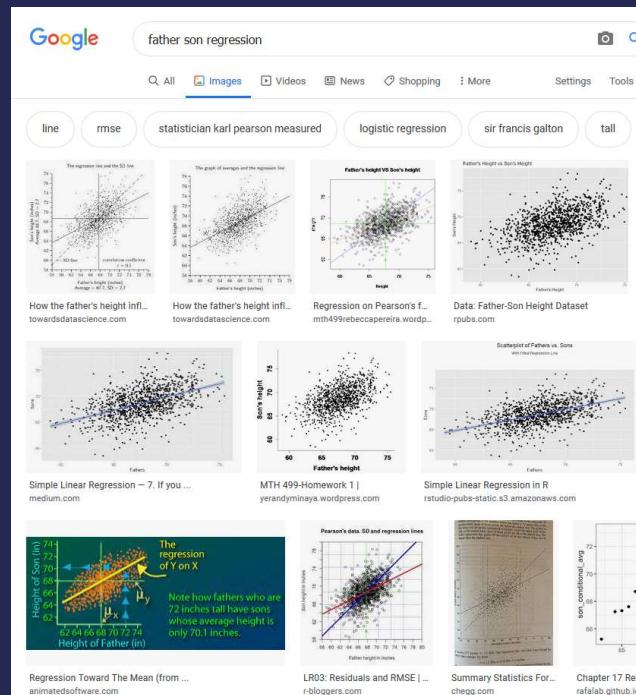
Data Science in the Real World

- **Data Science:** a huge, growing field for analyzing and predicting trends. Used in research, business, etc. with countless opportunities and high pay.
- Real data sets are **HUGE!**
 - Millions, billions, and yes, kajillions of rows or entries
 - Hundreds of columns or attributes to consider
 - Often spans multiple sets or introduces new data in real-time



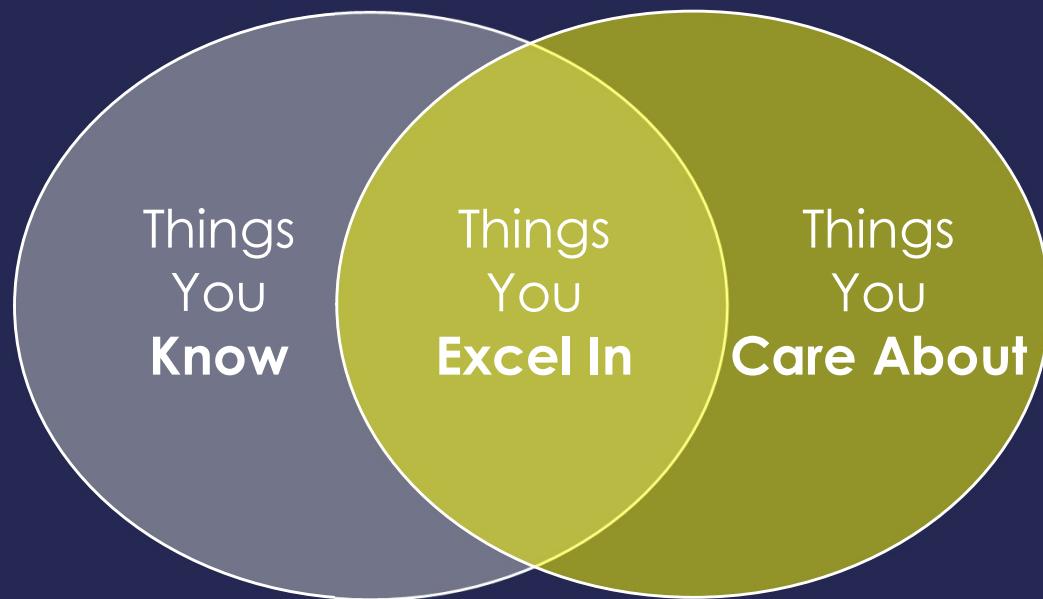
Data Science in Education

- Typically uses **small, curated, uninteresting** data sets for easy reproducibility and evaluation
- As introductory material, it's okay, but we can do better!



But **how** can we do better?

Make it Personal!



Let's Talk Advantages: Domain Knowledge

- **Domain Knowledge:** topical info about the data itself unrelated to Data Science
 - When you're already familiar with your data, it's easier to work with it!

1	Kinsect Extract Color extract associated with the part.	9	Base HP / Solo HP Solo HP is the typical HP found in singleplayer, single monster optional quests. The actual HP is calculated from quest specific modifiers and player scaling.																								
2	Part Names and States Colored parts can be tenderized. Multiple zones may be apart of the same group.	10	Base Status Values Initial Values, Tolerance Increase; Max Tolerance scaled by quest and player scaling.																								
3	Base Flinch Values Flinch threshold of the part. These values are modified by quest specific flinch modifiers, and player scaling.	11	Stamina, Exhaust, and Eldersel 1 st cell: monsters stamina. 2 nd cell: fatigue duration. Last cell: stamina damage. Exceeding the exhaust threshold will reduce the monsters stamina. The monster will become fatigued once stamina hits zero. Most elders do not have stamina; none can be directly exhausted. Eldersel values will be displayed instead.																								
4	Trip / Knock Down Number of flinches required to trip / knock down a monster; the duration in seconds associated with it. Trip Key: A [Fl/Swim, B]Break, * [After Break, Chrg] Charging	12	Traps & Item Data CC duration for various items and environmental traps.																								
5	Break & Sever Values Base values to destroy or sever the part. These are multiplied by the same factor as the flinch values.	13	Enrage Data Parameter changes during enraged state.																								
6	Notes 1. L/R denotes left and right are independent. 2. %SXH is the HP threshold to break the part. 3. Time denotes a particular states timer. 4. Some zones share flinch/break thresholds.	14	Tenderize Formula Raw hitzones are increased following this formula. Ex. Nergigant Head $\text{Hitzone} = \text{int}(0.75 * 45) + 25 = 33 + 25 \Rightarrow 58$																								
8	Miscellaneous Stats 1. RWT: Skill level required to negate roar, wind and tremor. Not all attacks will require this level. 2. HP Capture threshold. 3. HP Limp threshold 4. Clagger threshold is 1/10 base HP multiplied by the quest specific HP modifier. Changes based on rage, fatigue and HP.	15	Special Notes Relevant notes related to special mechanics.																								
		16	Raw and Elemental Hitzone Data These values correspond to the percent damage taken from each attack type.																								
<table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th></th> <th>Base</th> <th>Enraged</th> <th>Fatigued</th> </tr> </thead> <tbody> <tr> <td>100-90%</td> <td>90-80%</td> <td>80-70%</td> <td>70-60%</td> </tr> <tr> <td>60-50%</td> <td>50-40%</td> <td>40-30%</td> <td>30-20%</td> </tr> <tr> <td>20-10%</td> <td>10-0%</td> <td>-10%</td> <td>-20%</td> </tr> <tr> <td>+25%</td> <td>+25%</td> <td>+10%</td> <td>+10%</td> </tr> <tr> <td>-25%</td> <td>-25%</td> <td>-10%</td> <td>-25%</td> </tr> </tbody> </table>					Base	Enraged	Fatigued	100-90%	90-80%	80-70%	70-60%	60-50%	50-40%	40-30%	30-20%	20-10%	10-0%	-10%	-20%	+25%	+25%	+10%	+10%	-25%	-25%	-10%	-25%
	Base	Enraged	Fatigued																								
100-90%	90-80%	80-70%	70-60%																								
60-50%	50-40%	40-30%	30-20%																								
20-10%	10-0%	-10%	-20%																								
+25%	+25%	+10%	+10%																								
-25%	-25%	-10%	-25%																								



Oh, it's more
Monster Hunter.
Again. Thanks Bri.

Let's Talk Advantages: Stay Focused

When you don't care, it's a slog. It's probably even less interesting when you show it off.

Tid	Refund	Marital Status	Taxable Income	Defaulted Borrower
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

(a) Record data.

TID	ITEMS
1	Bread, Soda, Milk
2	Beer, Bread
3	Beer, Soda, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Soda, Diaper, Milk

(b) Transaction data.

Projection of x Load	Projection of y Load	Distance	Load	Thickness
10.23	5.27	15.22	27	1.2
12.65	6.25	16.22	22	1.1
13.54	7.23	17.34	23	1.2
14.27	8.43	18.45	25	0.9

(c) Data matrix.

item	Document 1	Document 2	Document 3							
team	3	0	5	0	2	6	0	2	0	2
win										
loss										
home										
away										
neutral										
series										

(d) Document-term matrix.

RMSE:

Lasso: 4.68423

Ridge: 4.78482

Multilinear: 5.05724

- Coefficients for Multilinear and Lasso are very similar despite Lasso being much more accurate.
- Ridge coefficients are completely different but has accuracy between Multilinear and Lasso.

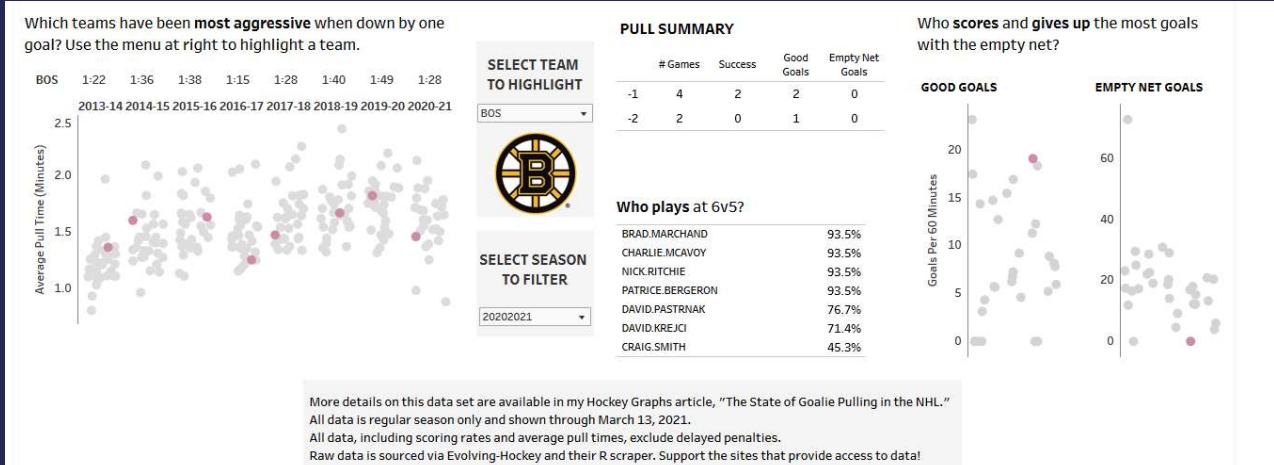
Comparison of Models

MULTI	Coefficients	RIDGE	Coefficients	LASSO	Coefficients
(Intercept)	36.341145004	(Intercept)	22.532806300	(Intercept)	34.261899143
CRIM	-0.108413345	CRIM	-0.443331300	CRIM	-0.097275622
ZN	0.045844929	ZN	0.593433800	ZN	0.041110363
INDUS	0.000000000	INDUS	-0.764716700	INDUS	0.004880904
CHAS	2.718716303	CHAS	0.749432900	CHAS	2.674445338
NOX	-17.376023429	NOX	-1.413996800	NOX	-16.277238692
RM	3.801578840	RM	3.212365700	RM	3.876056335
AGE	0.000000000	AGE	-0.313094100	AGE	0.000000000
DIS	-1.492711460	DIS	-2.553963600	DIS	-1.382960187
RAD	0.299608454	RAD	0.842445200	RAD	0.249649243
TAX	-0.011777973	TAX	-0.751022600	TAX	-0.009769353
PTRATIO	-0.946524570	PTRATIO	-1.809904000	PTRATIO	-0.929752122
B	0.009290845	B	0.897452200	B	0.009007352
LSTAT	-0.522553457	LSTAT	-2.892787400	LSTAT	-0.522861808
lambda	NA	lambda	0.67776537	lambda	0.03073584
nonzero	11	nonzero	13	nonzero	11
mse	25.57571	mse	22.89447	mse	21.94201
rmse	5.05724	rmse	4.78482	rmse	4.68423

Who cares?

Let's Talk Advantages: Stay Focused

But when you DO care, it's like you're not even working!



<https://medium.com/nightingale/so-you-want-to-make-a-hockey-data-viz-dda7b347f117>

How to train an artificial neural network to play Diablo 2 using visual input?

[Ask Question](#)

Asked 9 years, 8 months ago

Active 12 months ago

Viewed 37k times

I'm currently trying to get an ANN to play a video game and I was hoping to get some help from the wonderful community here.

140

I've settled on Diablo 2. Game play is thus in real-time and from an isometric viewpoint, with the player controlling a single avatar whom the camera is centered on.

80

To make things concrete, the task is to get your character x experience points without having its health drop to 0, where experience point are gained through killing monsters. Here is an example of the gameplay:



Let's Talk Advantages: Improve Your Well-Being

Real data from my credit card history!

Many services let you download your personal data records as .csv or Excel files.

You can use these to not only improve your DS skills, but also improve your quality of life!

...or not!

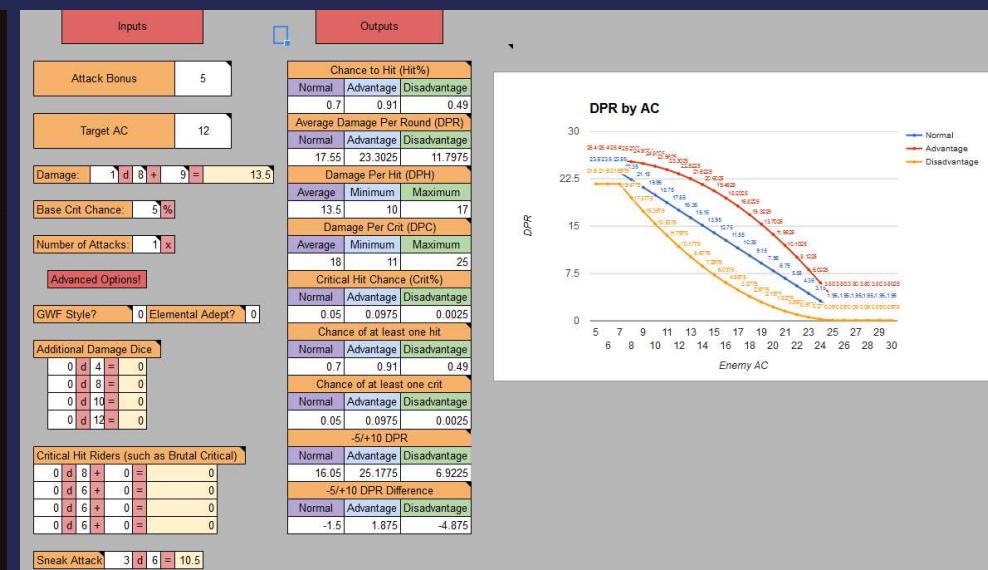


Remember: Nothing is a Waste of Time

- Simply trying is a good experience – cultivate your skills by doing **something**
- This is a good look for interviews!

```
377 xlsx = pd.ExcelFile('Dks_data.xlsx')
378 DS = {}
379 for i in xlsx.sheet_names:
380     DS[i] = pd.read_excel(xlsx, i).set_index("Id")
381
382 DS['Goods']['English Name'] = DS['Goods']['English Name'].str.replace("Pyromancy: |Miracle: |Sorcery: ","")
383 dlcspells = {3710:"Dark Orb", 3720:"Dark Bead", 3730:"Dark Fog", 3740:"Pursuers", 4530:"Black Flame"}
384 DS['Goods'].loc[dlcspells,['English Name']] = DS['Goods'].index.to_series().map(dlcspells)
385 magic = DS['Goods'][DS['Goods']['Magic Id'] > 0]['English Name'].copy() # index = magicId for all spells
386 magic = pd.DataFrame(magic).join(DS['Magic'][[['RefCategory','RefId']]])
387 bullets = {}
388 for i in magic.index:
389     k = []
390     v = []
391     if magic.loc[i,'RefCategory'] == 1:
392         ref = magic.loc[i,'RefId']
393         while ref > 0:
394             k.append(ref)
395             b = DS['Bullets'].loc[ref]
396             att = b.AtkId_Bullet
397             if att > 5 and att not in v:
398                 v.append(b.AtkId_Bullet)
399             ref = b.HitBulletID
400     bullets[i] = [i,k,v]
401 bullets = pd.DataFrame.from_dict(bullets,orient='index')
402 bullets.columns = ['Id','Bullets','Attacks']
403 magic = magic.join(bullets.set_index('Id'))
```

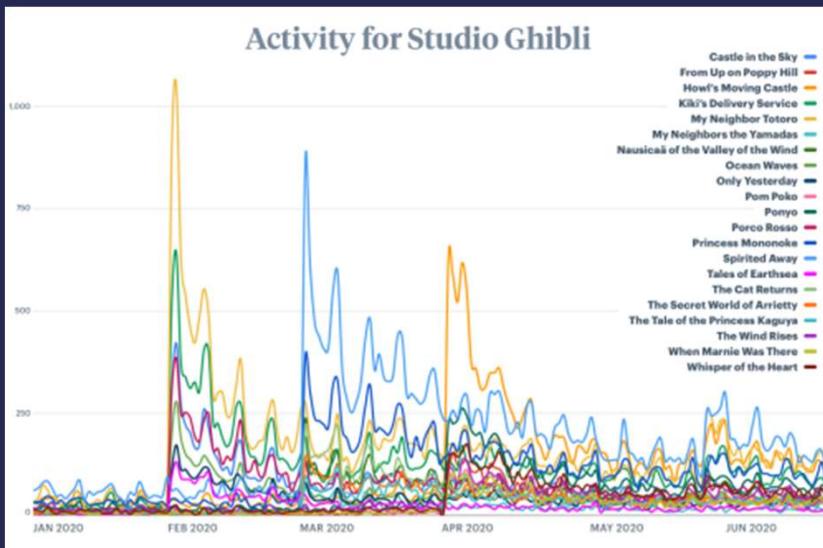
Coding!



Visualizing!

Remember: Nothing is a Waste of Time

- Anything you have an interest in can be relevant in the world of business
- There's no such thing as a dumb topic – YOU decide if it's worth doing



Your favorite movies!



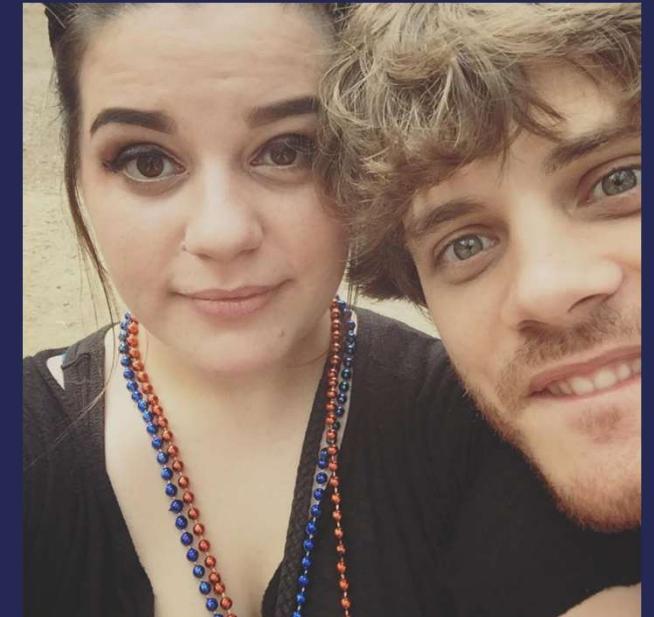
Your favorite breakfast!

Remember: Nothing is a Waste of Time

- You tend to meet the people you'd **want** to meet when you do what you love!



Make friends!



Find true love!