From analytics to action:

How I built a major gift prospect model, and how my client organization used it

Matthew Dubins,
- Founder of **Donor Science Consulting**

Greg Wagland,
- Manager of Annual Giving of **The Mustard Seed**

APRA CANADA CONFERENCE
OCTOBER 17-19, 2018

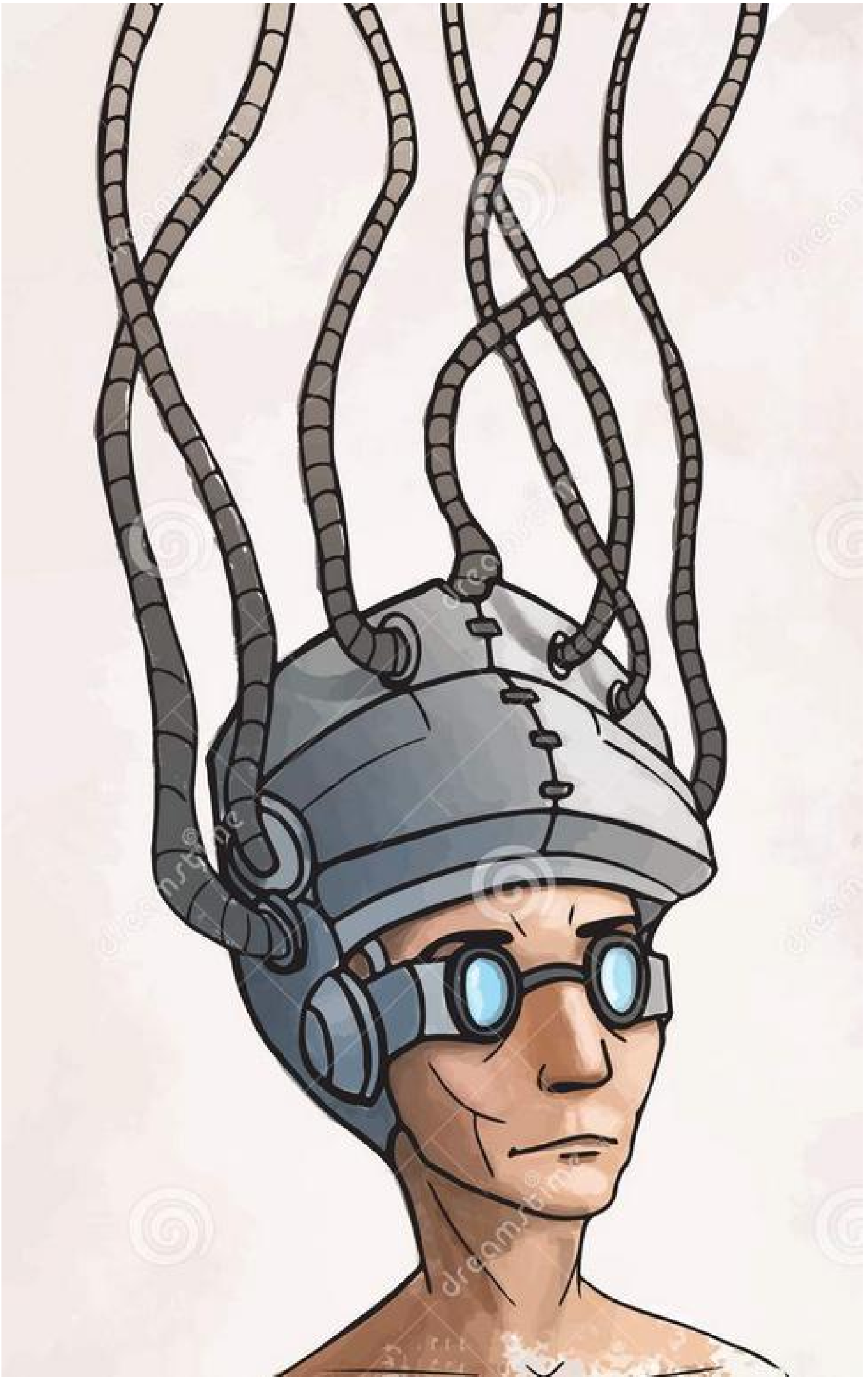
SESSION DATE: 18-Oct-2018
SESSION TIME: 1:45 pm - 2:45 pm

First, an introduction!

Some things to know about me:

- I'm crazy about data analytics
- 1st job: "Run Data and Research Coordinator" for Canadian Breast Cancer Foundation
- After, I did some very in depth analytics and modelling for all sorts of non-profits as a consultant at KCI and Blakely Inc.
- In May 2017, I really went off the deep end and started my own analytics consulting business I called **Donor Science Consulting**

Just look at me now!



CANADA CONFERENCE
OCTOBER 17-19, 2018

One client, two prospect models

Enter my client, The Mustard Seed (represented by Greg Wagland)

The need: identify new major gift prospects for upcoming campaign

Two prospect model solution --> 2 major gift scores:

1) **Behaviour Fitness Model (BFit):** Use database characteristics to predict $10k+ giving over 5yr period
2) **Financial Fitness Model (FinFit):** Use Environics Wealthscapes characteristics to predict the same giving

Behaviourial/Biographical Data + Finances Data = Extra confidence in prioritizing prospects for approach.

My process for creating major gift prospect models

3 Stage Process:

Stage Number	Stage Description
1	Commencement, Conversations & Confirmation
2	Data Prep & Decision Making
3	Model Training, Validation, Discussion of Results, & Scoring

Simple…. right?

Not quite…

<u>Stage 1</u>

1. Confirm Objective
2. Discuss available data sources
3. Enough data for the model?
4. Gather required database exports
5. **Look** at the data!
6. Check in with the client with any concerns or your thumbs up

<u>Stage 2</u>

1. Summarize all data sources by donor ID and year if possible
2. Decide which variables you want to test, transform, or drop for modelling
3. Compute new predictor variables if wanted
4. Merge all data sources together
5. **ACK!** So much data

<u>Stage 3</u>

1. Create training and testing samples
   a. Oversampling needed?
2. Feature selection
3. Initial stab at training a rather large predictive model
4. Train a much smaller but more effective model
5. Model validation
6. Summary of interesting insights coming out of model
7. Discuss results
8. Score all eligible records

Don't forget….

Document your work!!!!

Stage 1: The Initial Conversation

- **Questions** to be answered here:
  1. What do we want?
  2. Do we have the data to carry forward with the project?

- For this project, the **answers** were:
  1. Create prospect ranks that sort people by how much they look like a $5k or $10k+ donor
  2. It looked like we had enough to model $5k+ donors

- Data sources used: biographical details, gift history of active donors, solicitation history, contact preferences, event participation, constituency codes

- Environics Wealthscapes wasn't discussed at this point

Stage 1: Looking at the Data

Summary of gift history by year

Year	Num Donors	Num Donations	Donation Total	Median Donation	Top 3 Donations
2000	1,973	3,918	$1,072,280	$50	$100,000 ; $50,000 ; $27,000
2001	3,599	5,859	$1,121,127	$50	$50,000.00 ; $27,000.00 ; $24,285.86
2002	5,736	10,741	$1,486,021	$50	$20,244.18 ; $16,000.00 ; $14,000.00
2003	8,037	14,526	$2,040,178	$50	$75,840.00 ; $32,913.00 ; $32,895.34
2004	9,631	19,885	$2,563,490	$40	$61,494.00 ; $50,000.00 ; $43,957.17
2005	11,270	26,264	$3,725,339	$40	$181,205 ; $93,084 ; $50,000
2006	14,255	37,235	$5,003,624	$39	$200,000 ; $103,396 ; $59,253
2007	16,364	45,871	$6,149,415	$39	$300,000 ; $123,942 ; $100,000
2008	19,032	58,344	$7,105,956	$39	$400,000 ; $131,742 ; $81,051
2009	21,648	65,890	$8,231,085	$37	$300,000 ; $198,271 ; $108,332
2010	23,659	70,285	$8,568,353	$37	$300,000 ; $75,700 ; $73,035
2011	22,483	66,922	$10,904,261	$37	$953,770 ; $561,659 ; $250,000
2012	27,717	80,058	$10,946,064	$40	$400,000 ; $250,000 ; $241,527
2013	35,178	91,996	$13,653,799	$40	$500,765 ; $498,870 ; $400,000
2014	34,278	84,252	$12,645,990	$40	$498,030 ; $118,970 ; $100,383
2015	36,052	84,542	$13,307,260	$40	$813,989 ; $200,538 ; $128,752
2016	35,283	85,173	$14,030,216	$40	$775,450 ; $400,000 ; $252,153
2017	34,787	85,610	$13,467,009	$40	$300,000 ; $200,000 ; $134,627
2018	9,558	16,441	$2,061,275	$32	$100,000 ; $80,645 ; $60,000
2024	1	1	$0	$0	$0 ; ;

Stage 1: Looking at the Data

Summary of key biographical characteristics

Tenure	# Donors
NA	1593
0	513
1	7152
2	7856
3	8817
4	7613
5	6126
6	4463
7	2242
8	3864
9	4302
10	3492
11	3518
12	3391
13	2380
14	2760
15	3027
16	2767
17	2178
18	1710

Recency	N
NA	1593
0	9495
1	26469
2	12869
3	10811
4	9251
5	9029
6	53
7	27
8	31
9	32
10	14
11	28
12	14
13	15
14	9
15	11
16	7
17	1
18	5

Deceased	N
No	78267
Yes	1497

Monthly Donor Status	N
	74635
Active	3576
Held	67
Terminated	1486

TMS City	N
	90
Calgary	49468
Edmonton	27048
Red Deer	3158

Quantile	First Gift Amount
0%	0
10%	19
20%	21
30%	25
40%	39
50%	50
60%	53
70%	100
80%	100
90%	249
100%	813989

Key Indicator	N
I	73169
O	6595

ConstituencyCodeShort	N
Business	4842
Charity/Foundation	237
Church	532
Group/Club/Society/Union/Association	650
Group/Club/Society/Union/Association\n	8
Individual	73215
School	280

Province	City	N
AB	Calgary	39355
AB	Edmonton	18157
AB	Red Deer	2157
AB	Sherwood P	2016
AB	St. Albert	1273
		1010
AB	Airdrie	904
AB	Cochrane	868
AB	Okotoks	717
AB	Spruce Grov	551

Province	N
AB	76547
	1246
BC	762
ON	457
ab	188
SK	180
MB	52
NS	38
QC	38

Province
AB
BC
ON
SK
MB
NS
QC
TX

Stage 1: Looking at the Data

Simple summaries for each and every field in a table

	A	B	C	D	E	F	G	H	I
1	Column Names	Column Types	Number Non-null Records	Number Unique Values	Sum Total	Average Value	Median Value	Max Value	Sample Values
2	Postal Code	character	78657	34856					T5N0B6, T5X1H9, T2E6E7, T1V1J2, T6W0L2
3	Constituent Import ID	character	79764	79764					00001-593-0000451490, EDM-593-0000044918, 78238-079-0000185387, 00001-593-0000366314, 00001-593-0000485719
4	Constituent ID	integer	79764	79764	12202997757	152988.79	184451	232899	194481, 150678, 15177, 130577, 223279
5	Organization Name	character	6592	6418					, , , ,
6	Key Indicator	character	79764	2					I, I, I, I, I
7	Deceased	character	79764	2					No, No, No, No, No
8	Mailing Address On File	character	78662	2					Yes, Yes, Yes, Yes, Yes
9	City	character	78704	1127					Edmonton, Edmonton, Calgary, High River, Edmonton
10	Province	character	78518	50					AB, AB, AB, AB, AB
11	Home Phone	character	50724	2					Yes, Yes, Yes, ,
30	TMS City	character	79674	4					Edmonton, Edmonton, Calgary, Calgary, Edmonton
43	# Gala Donations	integer	79764	10	958	0.01	0	41	0, 0, 0, 0, 0
44	$ Gala Donations	numeric	79764	146	3062643.27	38.4	0	106500	0, 0, 0, 0, 0
45	# Golf Donations	integer	79764	18	994	0.01	0	78	0, 0, 0, 0, 0
46	$ Golf Donations	numeric	79764	176	1090080.26	13.67	0	92250	0, 0, 0, 0, 0
47	# Gifts Lifetime	integer	79764	205	826354	10.36	3	9126	2, 32, 26, 5, 1
48	$ Given Lifetime	numeric	79764	23509	129070668.7	1618.16	246.145	2814522.99	40, 5941, 463.44, 1347, 14
49	Last Gala Gift Date	POSIXct	634	209					2017-11-16, 2009-09-11, 2017-11-07, 2010-05-26, 2016-11-04
50	Last Gala Gift Amount	numeric	634	92	1748821.11	2758.39	900	43500	100, 1500, 1000, 275, 1500
51	Last Gala Gift Appeal ID	character	315	5					, , , ,
52	Last Golf Gift Date	POSIXct	354	110					2017-06-08, 2016-06-21, 2012-07-13, 2017-06-08, 2016-07-21
53	Last Golf Gift Amount	numeric	354	90	319614	902.86	250	25000	300, 50, 5500, 2000, 50
54	Last Golf Gift Appeal ID	character	354	10					, , , ,
55	Last Gift Date Lifetime	POSIXct	78171	2012					2015-12-18, 2017-12-19, 2018-03-30, 2014-12-29, 2016-12-26
56	Last Gift Amount Lifetime	numeric	78171	1831	18329817.45	234.48	50	498030.3	20, 400, 20, 270, 14
57	Last Gift Appeal ID Lifetime	character	78171	520					15EWCOCT1, E1711N1H, C1802M1H, 14CHCDEC1, EAGMEAL
58	First Gift Date	POSIXct	78171	4762					2014-01-14, 2011-10-20, 2004-10-12, 2010-11-16, 2016-12-26
59	First Gift Amount	numeric	78171	1769	14921497.74	190.88	50	813989.14	20, 120, 13.1, 311, 14
60	First Gift Appeal ID	character	78171	897					13EGADEC1, EOINT, rRSC04, CAGHONOUR, EAGMEAL
61	Largest Gift Date Lifetime	POSIXct	78171	4519					2014-01-14, 2017-12-19, 2013-10-03, 2010-11-16, 2016-12-26
62	Largest Gift Amount Lifetime	numeric	78171	2288	31691769.47	405.42	100	953769.6	20, 400, 37.32, 311, 14
63	Largest Gift Appeal ID Lifetime	character	78171	928					13EGADEC1, E1711N1H, 13CHCAUG2B, CAGHONOUR, EAGMEAL
64	First_Gift_Year	numeric	78171	20	157185206	2010.79	2012	2018	2014, 2011, 2004, 2010, 2016

Stage 1: # of Target vs All Donors

1,777 donors who gave $5k+ bw 2013 - 2017

778 donors who gave $10k+ bw 2013 - 2017

79,763 donors in the entire dataset

Stage 2: Data Prep

- To create the model, summary characteristics of each donor need to be compiled and then merged into 1 dataset

- Take each source dataset, transform so that each row is a donor, each column is a kind of summary overall or by year for that donor

- The result --> A super wide dataset!

Dataset - A

ID	Name	Height
1	A	1
3	B	2
5	C	2
7	D	2
9	E	2

Dataset - B

ID	Name	Weight
2	A	2
4	B	3
5	C	4
7	D	5

Left Join : Merged Dataset

ID	Name	Height	Weight
1	A	1	.
3	B	2	.
5	C	2	4
7	D	2	5
9	E	2	.

Stage 2: Decision Making

Hang on to your seats

This is going to be big!

Stage 3: Training & Testing Samples

What's a Training Sample?

A random portion of your compiled dataset that you use to teach the predictive modelling algorithm the factors that best distinguish Target vs. Non-Target donors.

What's a Testing Sample?

An untouched random portion of your compiled dataset that you did not use to train your predictive modelling algorithm. Instead, you use it assess how well your model does at distiguishing Target vs. Non-Target donors in a new dataset.

Stage 3: Training & Testing Samples

What is oversampling?

Rare events (like $10k+ giving) can make it hard to build a useful model.   Oversampling is the practice of randomly and repeatedly selecting records from the Target group so that you can amp up the power of the model to detect those rare events.

Stage 3: Training & Testing Samples

Source Data

60%

Oversampled Modeling Data
- 481 $10k+ donors
- 2,405 extra records, randomly selected from orig 481
- 47,044 non-$10k+ donors

60%

Oversampled training data
- 1,723 $10k+ donors
- 28,210 non-$10k+ donors

40%

Oversampled testing data
- 1,163 $10k+ donors
- 18,834 non-$10k+ donors

40%

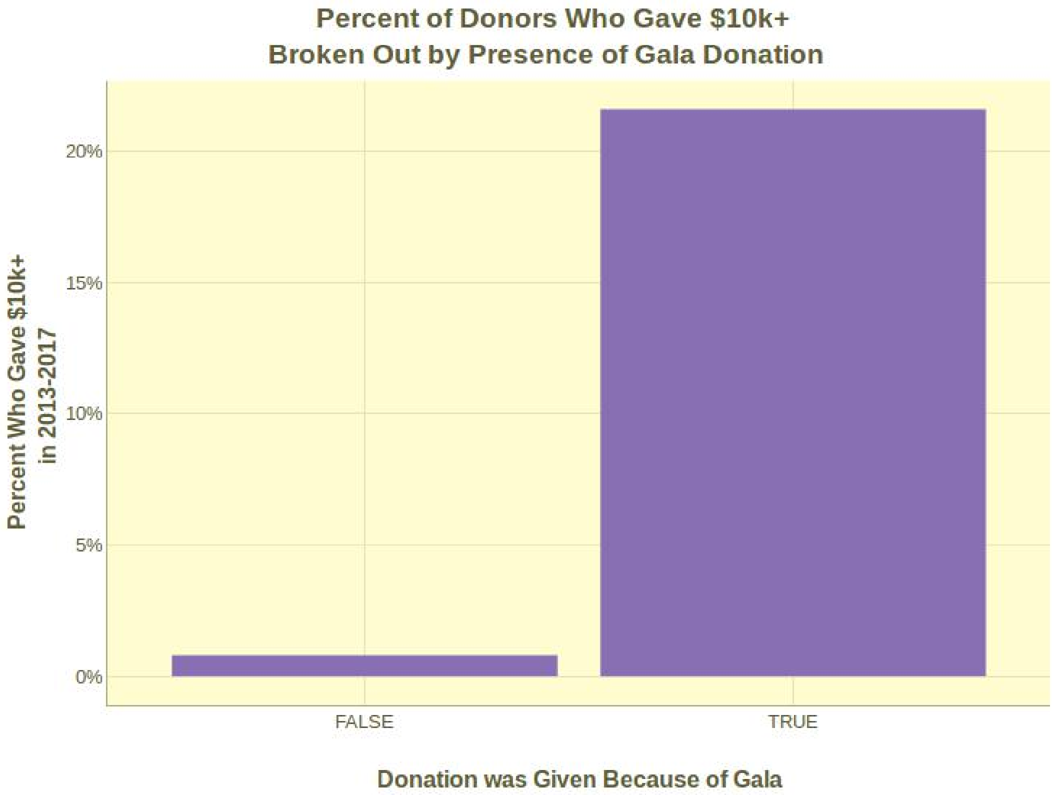
Raw, non-oversampled testing data
- 285 $10k+ donors
- 31,267 non-$10k+ donors

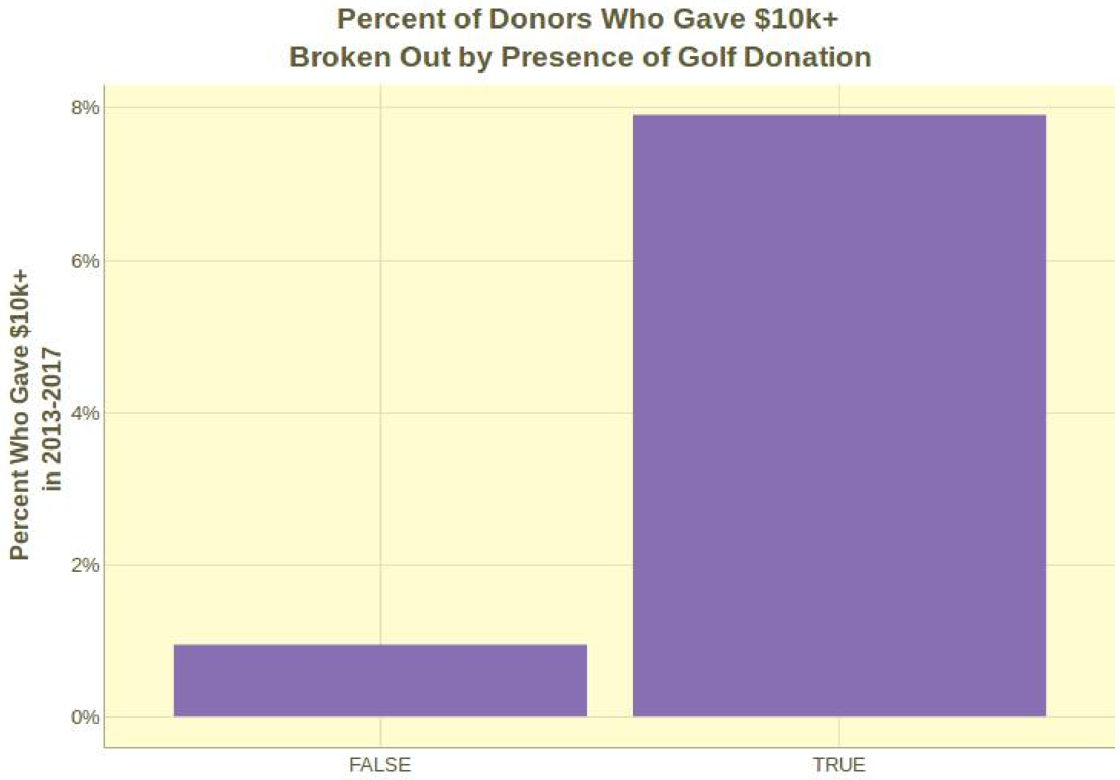
Stage 3: Feature Selection, Model Summaries & Validation

Time to look at

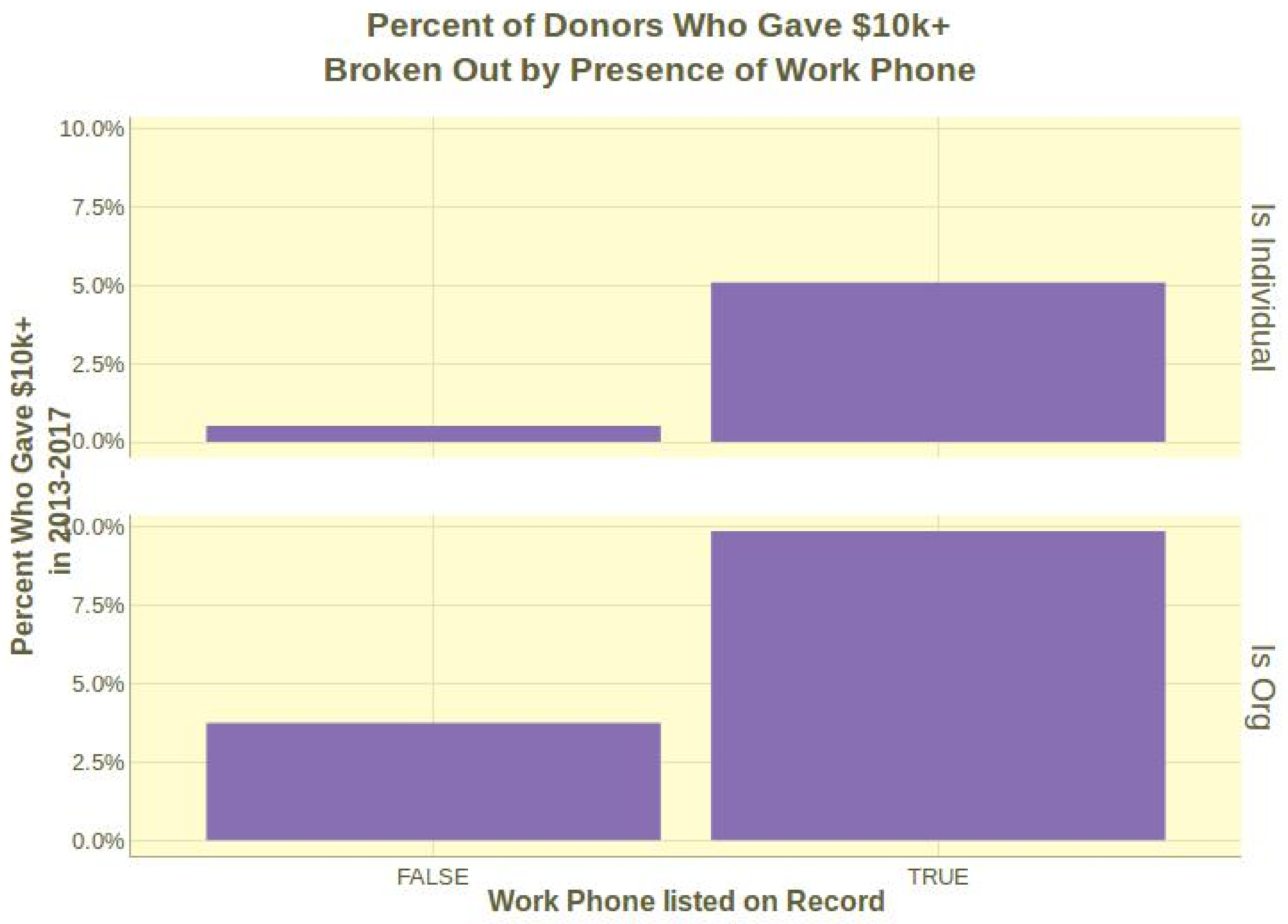
a spreadsheet again :)

Stage 3: Model Insights

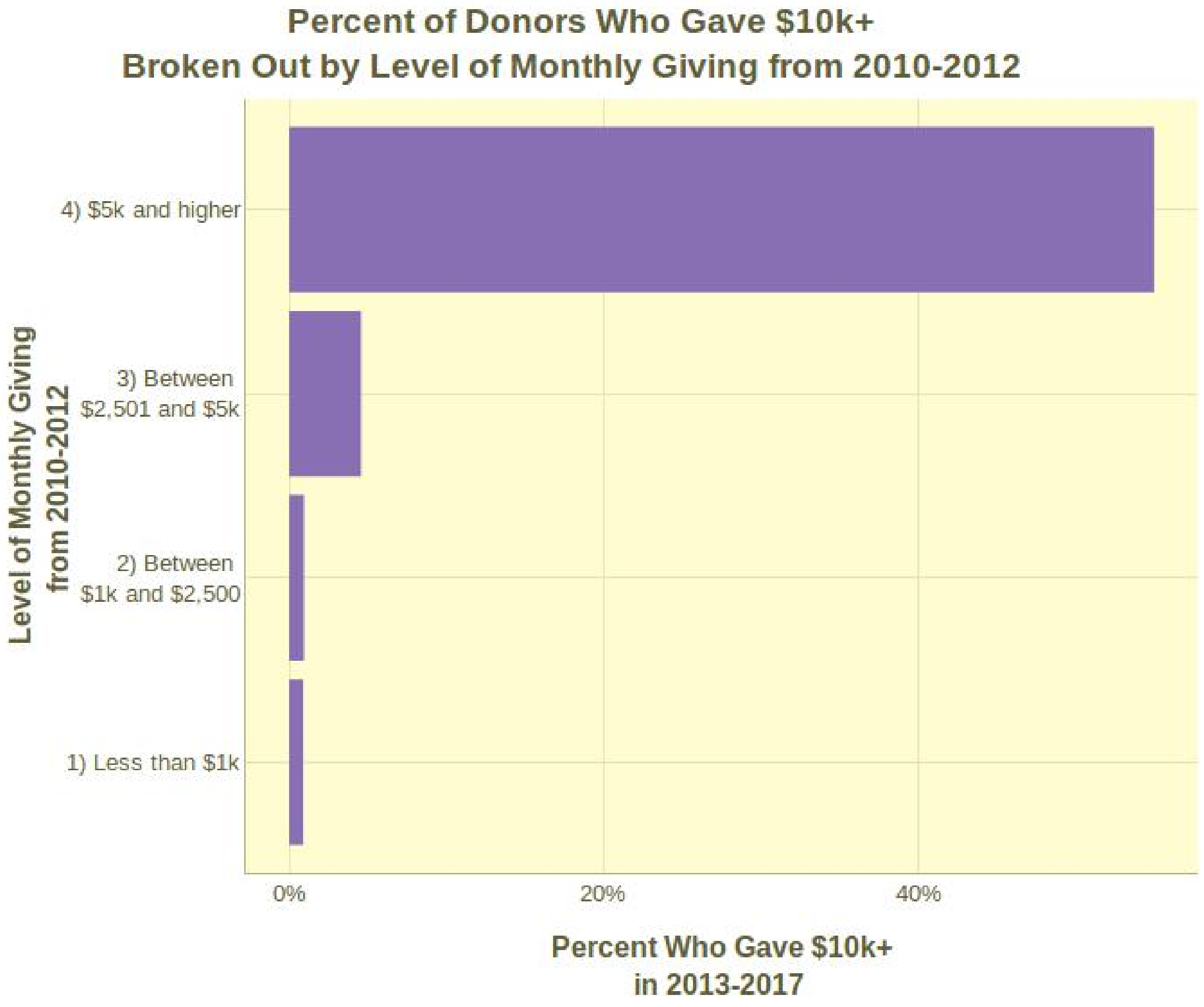




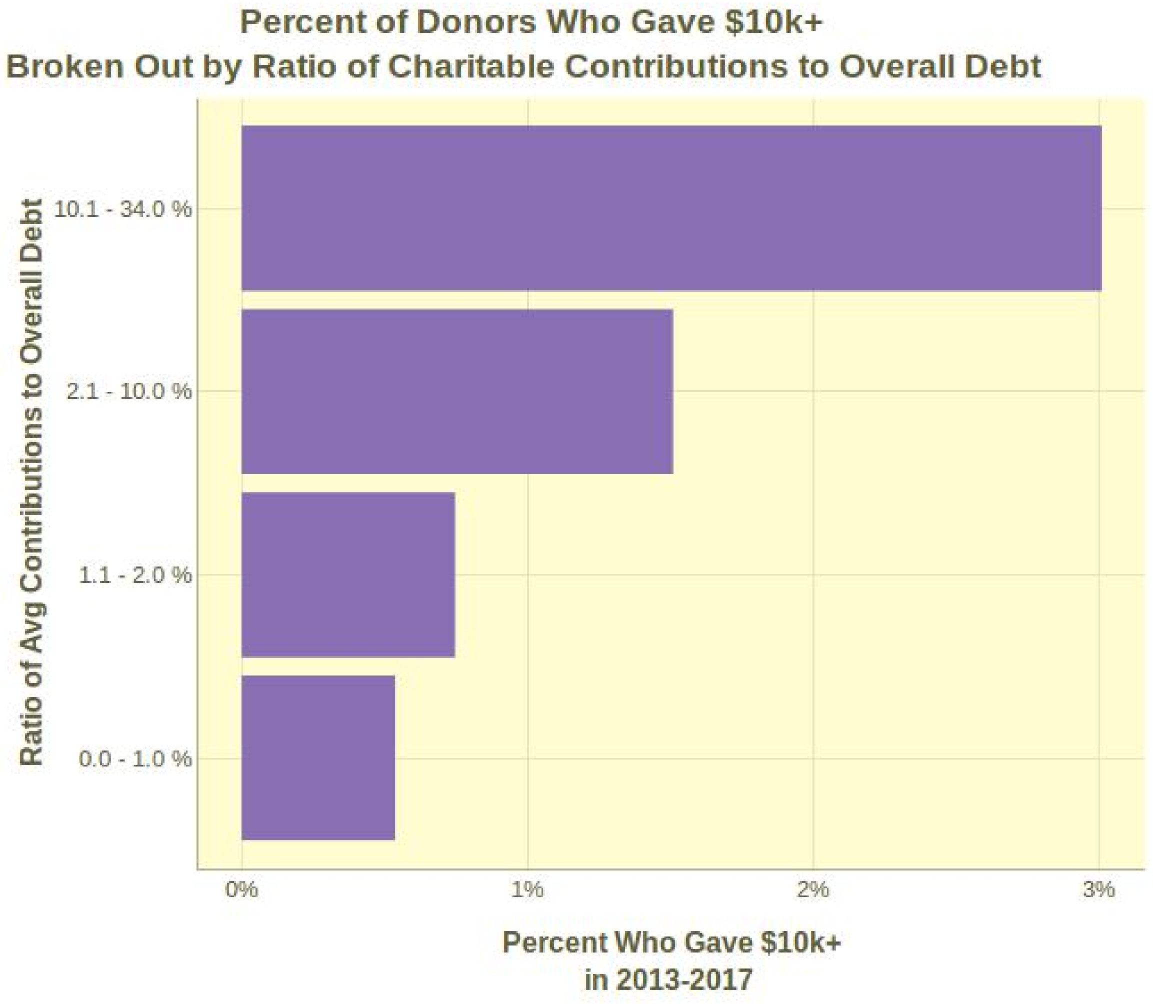
Stage 3: Model Insights



Stage 3: Model Insights



Stage 3: Model Insights



Stage 3: Model Scoring

- Final step to use the models to score the donors bw 0 and 1 (or 0 and 100 if you like).
- After scoring, use the score to rank them into groups of 10 (deciles).

- **Who did I score?**
  - Living donors with no recent history of $10k+ giving to enable discovery of new prospects.

- **What is this colourful table?**
  - Using BFit and FinFit decile ranks, we prioritized who to approach first (those with highest lvls of engagement and most suitable finances for $10k+ giving)

  - Each cell = # of donors scored w a particular BFit and FinFit decile

Is Business Addr?	BFit Decile	FinFit - 1	FinFit - 2	FinFit - 3	FinFit - 4	FinFit - 5	FinFit - 6	FinFit - 7	FinFit - 8	FinFit - 9	FinFit - 10
No	1	5	12	2	3	1	1	1	2	3	2
No	2	2	6	6	5	4	2	6	3	5	1
No	3	3	17	10	13	3	7	8	6	3	4
No	4	7	12	15	8	9	5	8	11	7	2
No	5	6	24	20	8	11	14	17	10	15	6
No	6	12	30	22	19	24	12	16	21	15	13
No	7	15	35	27	30	39	26	17	18	23	27
No	8	24	34	46	57	53	56	46	44	32	44
No	9	122	322	244	270	259	220	221	230	212	201
No	10	2597	8126	7383	7325	7369	7414	7445	7134	7811	7512

…. Then the Torch was Passed!

Donor Science Consulting

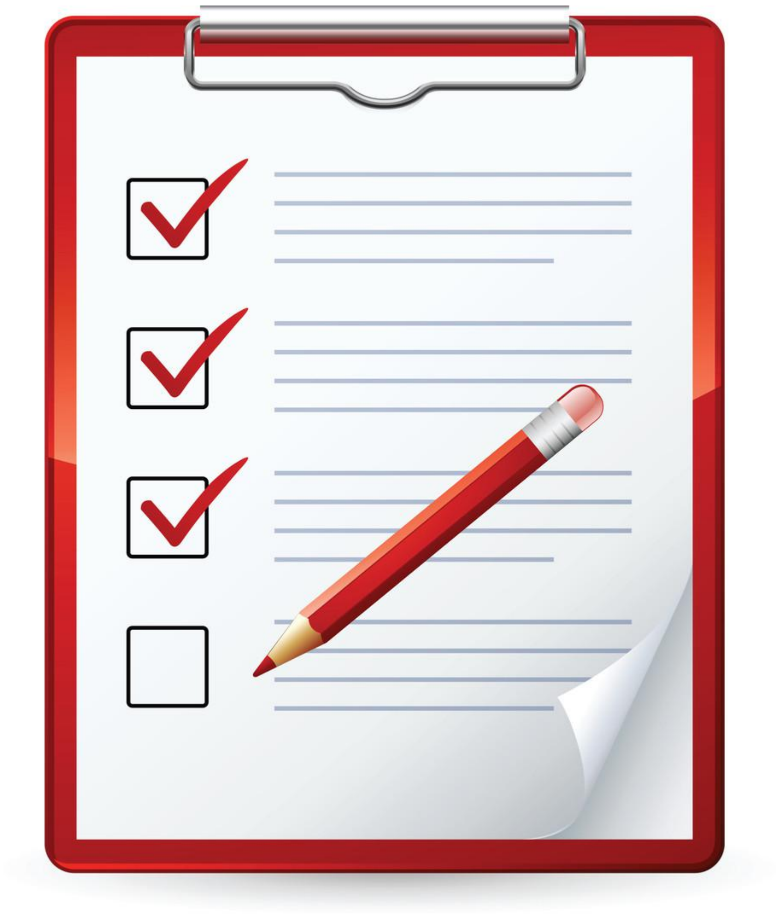
THE MUSTARD SEED

Major Themes for Greg After Delivery of Analytics

- Timing
  - Acceptance at all levels takes a lot of time
- Trust
  - Trusting relationship with MG Team crucial to make sure results are used
- Strategy for communicating results to MG Team
  - Communicating key model insights
  - Piecemeal introduction of prospect names

Greg's Internal Vetting of Model

- Reviewing the model with the data scientist, asking as many questions as possible
- Ensure that you know how to explain the model to the MG team
- Perform additional checks on the model to ensure the result is trustworthy



Thinking About Your MG Team

- A preliminary step is to bring the top 5 prospects to the MG Team

- Find someone in the team to champion the model

- Do they accept? Start feeding up to 10 prospects every 2 weeks

- Coordinate with Prospect Researcher

And then the hard work of cultivating your donors begins!

Questions?