Distil Data

Data Science in Fundraising and Data Ethics

Data for social good meet up

15th October 2019

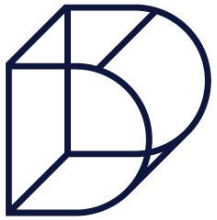
Welcome discussion

Why should we care about ethics for data science (in fundraising)?

Who am I?

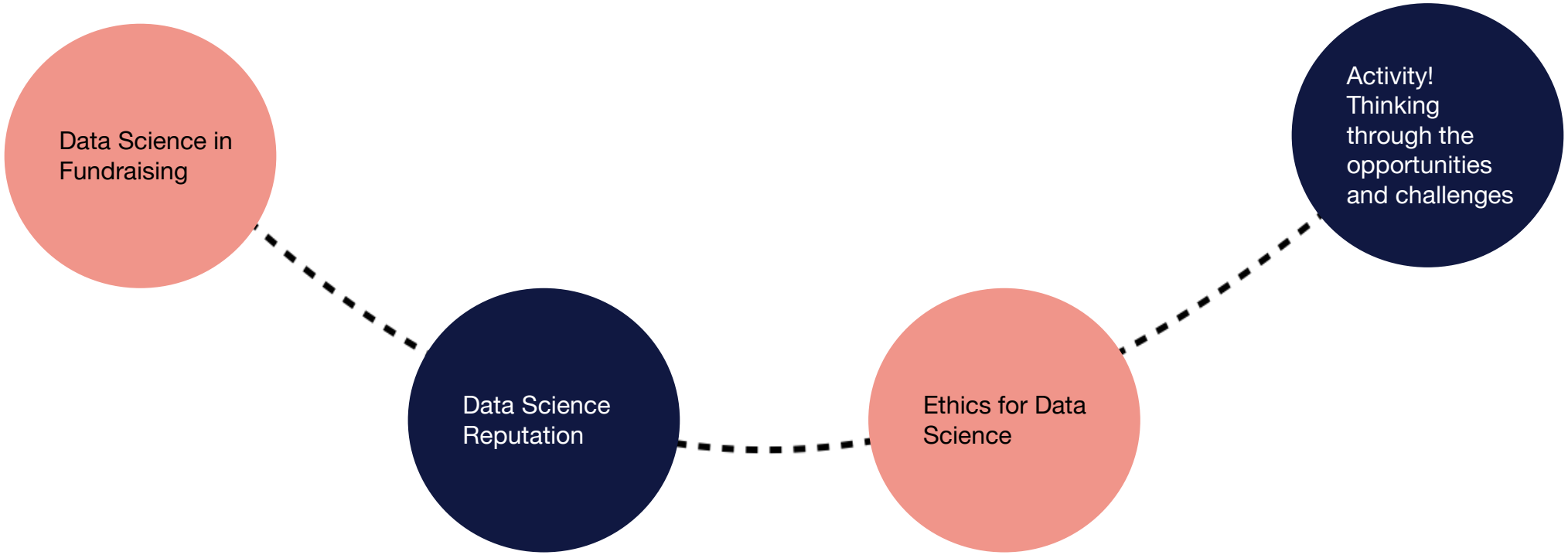
James Orton

- Teaching data science with General Assembly.
- Mentoring junior data scientists. Reach out if you're new to data science.
- Applying data science to fundraising and marketing, through my consultancy Distil Data.
- Looking for ways to change the world for better through data science for social good!



Distil Data

Agenda


Data Science in Fundraising
Data Science Reputation
Ethics for Data Science
Activity! Thinking through the opportunities and challenges

Why is data science important to fundraisers now?

Potential return on investment is now significant for Fundraisers.

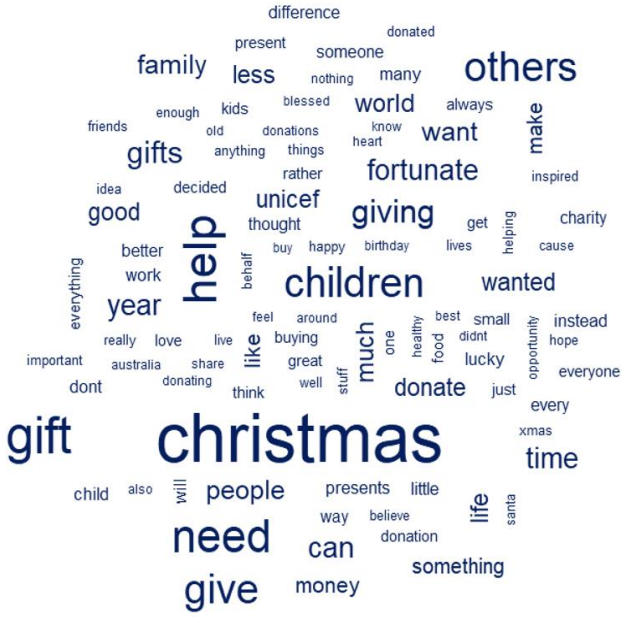
- Breakthroughs in methods and improved accuracy in an uncertain world.
- Increasing power of open-source tools like R and Python reduces barriers to entry.
- Availability of data

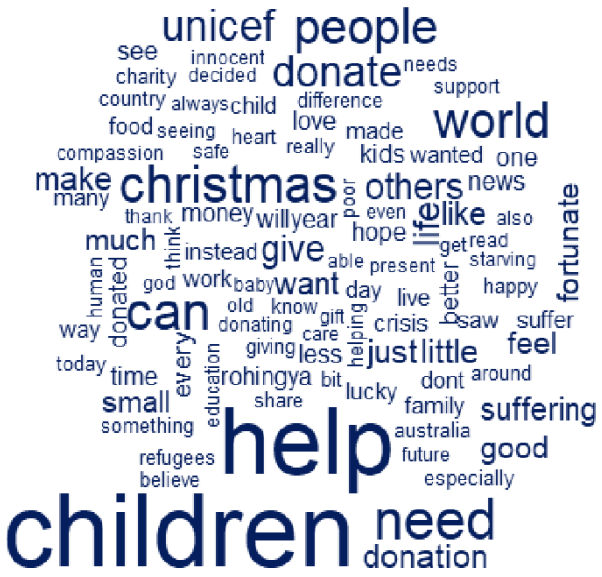
Bringing data science to life…

Examples

Using natural language processing

Reviewing the free text comments on supporter motivations across different products.





A game!

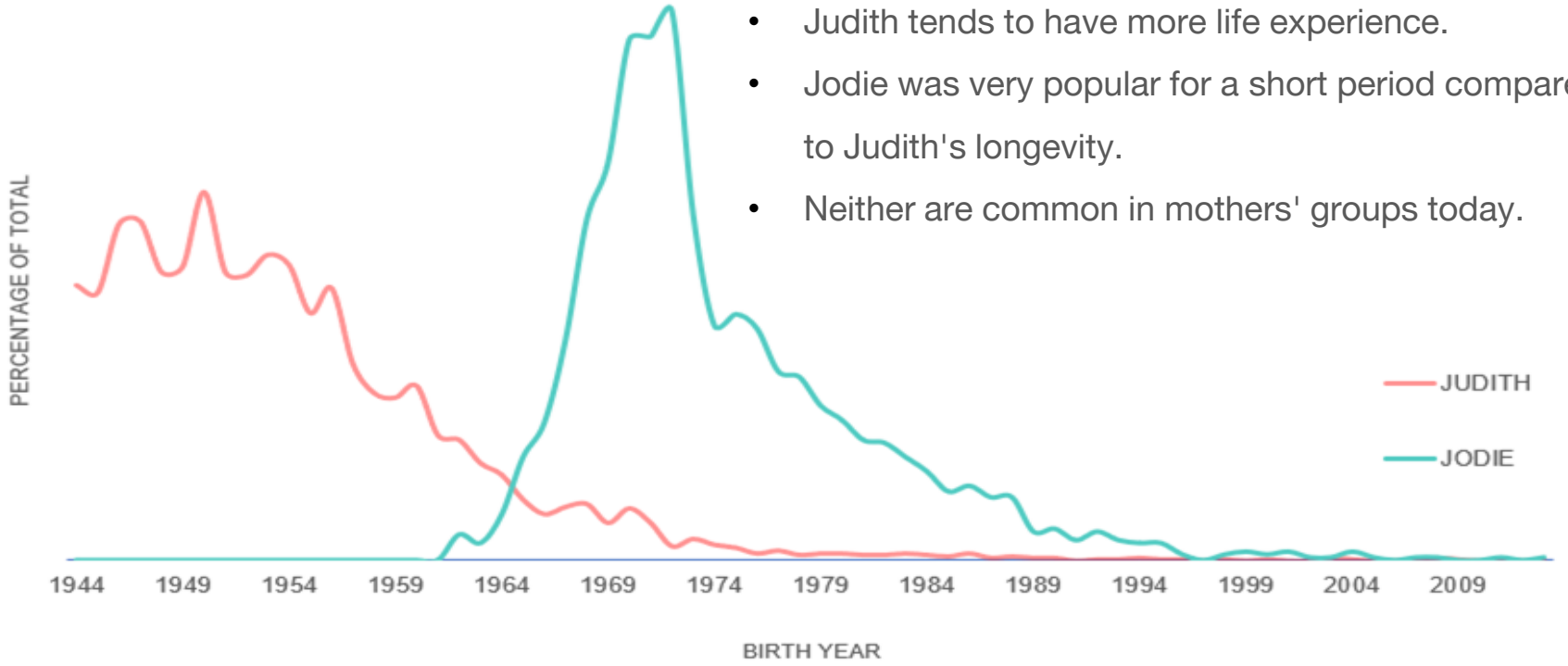
Who doesn’t mind telling us their name (and their age)? Don’t be shy

https://distildata.io/how-we-can-predict-age-from-a-first-name/

Age Prediction: How it works...

- Inspired by similar models in the US and UK.
- Collates publicly available births and deaths data for Australia.
- Calculates likely age distribution of people alive by a given name.

Jodie vs Judith



- Judith tends to have more life experience.
- Jodie was very popular for a short period compared to Judith's longevity.
- Neither are common in mothers' groups today.

Predictive methods and optimisation

- Enabling a long and valuable supporter journey:

  - Who to target? Which supporters are likely to respond to which fundraising asks?
  - When to target them, is Christmas or Tax time better for this supporter?
  - What to target them with? Are they interested components of our work? Are they more receptive to direct mail or email communications?
  - How much to ask for?

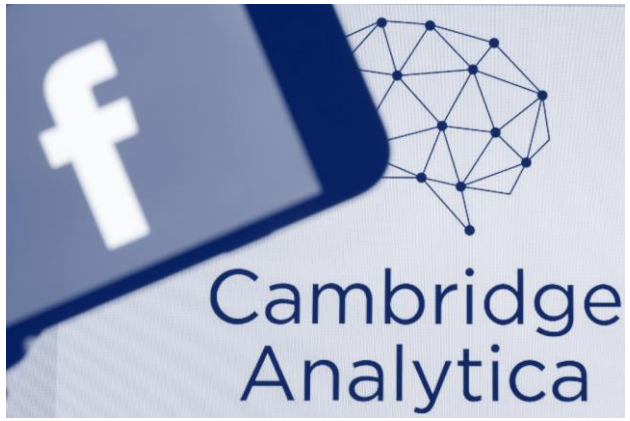
- We can measure their accuracy and be confident in our expected results.

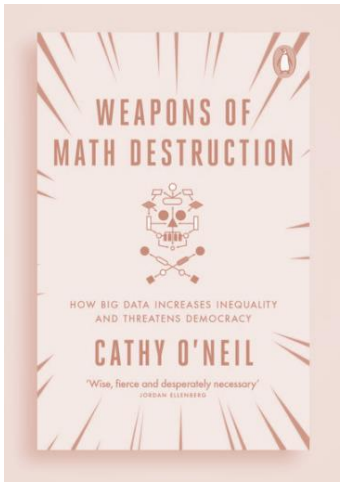
Data science is awesome!

What is the problem?

The dark side of data science

At this critical time, data scientists must demonstrate that the work we do is for social good.





- Data science has become de-humanised
- Trust in data science and the use of personal data is at an all-time low.
- As data scientists and people working with data for social good – how do we navigate this complex world?

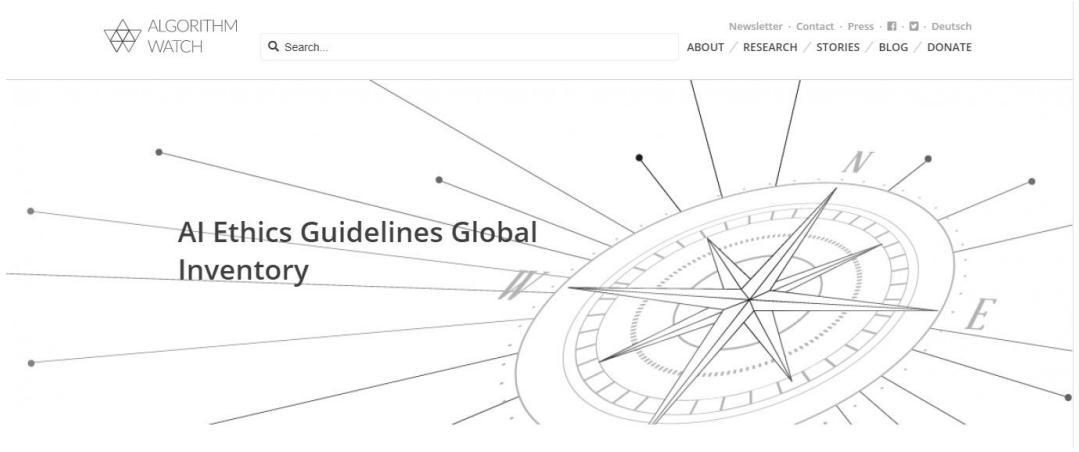
Ethics for data science

Ethics has become a hot topic in data science.

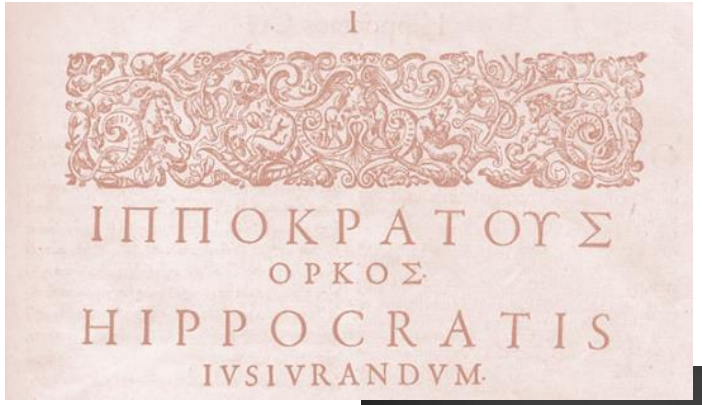
The next few slides will summarise the existing data ethics resources.

There are over 80 different ethical frameworks!

How do we apply data ethics to fundraising?



Ethics Framework



Three frequently discussed approaches for ensuring ethical data science are:

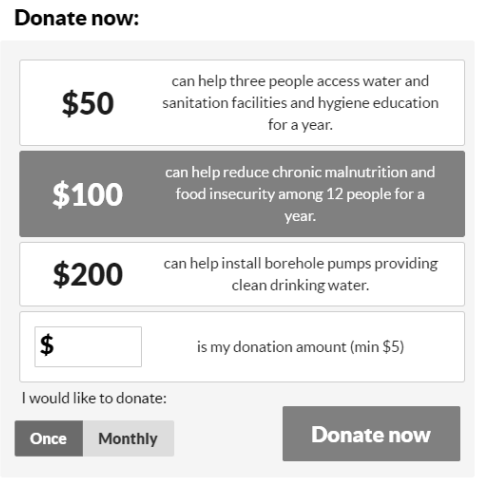
1. Regulation.
2. Oaths and principles.
3. Checklists.

As data scientists and people working with data we should ask **“How would I like my own data to be used?”**

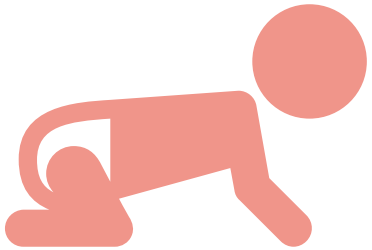
Ethics Considerations

What are (some of) the common themes?

- Consent.
- Data Security.
- Treatment of protected groups in terms of fairness and error rates.
- Transparency.
- Explainability.



Group exercise: Scenario



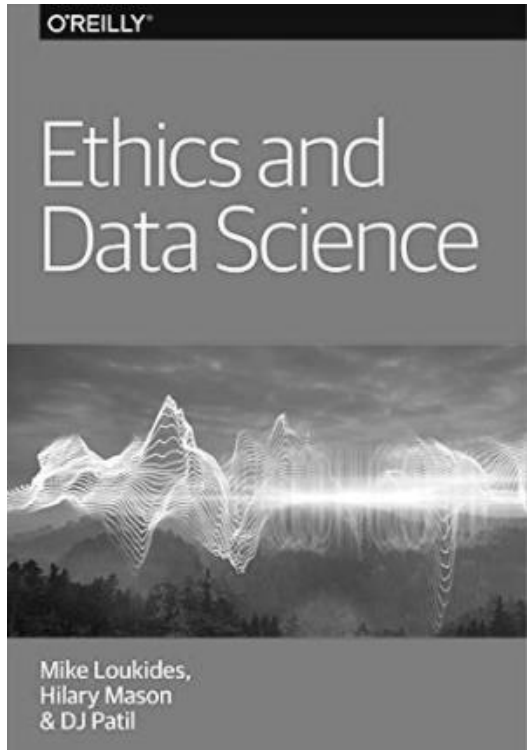
We are working for a large international development charity whose primary fundraising product is child sponsorship via a monthly donation.

We have been asked to use data science to: “Maximise fundraising net income per supporter”.

- What projects and activities should we consider?
- What are the potential ethical considerations?

Ethics and Data Science checklist

- Have we listed how this technology can be attacked or abused?
- Have we tested our training data to ensure it is fair and representative?
- Have we studied and understood possible sources of bias in our data?
- Does our team reflect diversity of opinions, backgrounds and kinds of thought?
- What kind of user consent do we need to collect to use the data?
- Do we have a mechanism for gathering consent from users?
- Have we explained clearly what users are consenting to?
- Do we have a mechanism for redress if people are harmed by the results?
- Can we shut down this software in production if it is behaving badly?
- Have we tested for fairness with respect to different user groups?
- Have we tested for disparate error rates among different user groups?
- Do we test and monitor for model drift to ensure our software remains fair over time?
- Do we have a plan to protect and secure user data?



Links

Ethics and Data Science Book:

https://www.oreilly.com/library/view/ethics-and-data/9781492043898/ch01.html

https://github.com/oreillymedia/ethics-datascience/ < free version ☺

Great research from Turing Institute and others:

https://www.turing.ac.uk/research/data-ethics

https://www.rss.org.uk/Images/PDF/get-involved/2019/Data%20Ethics%20Landscape.pdf

Ethics guidelines inventory:

https://algorithmwatch.org/en/project/ai-ethics-guidelines-global-inventory/

Top news and blogs:

http://aiweekly.co/ < has a whole ethics section in the weekly email

https://www.theguardian.com/technology/2018/jul/07/artificial-intelligence-can-tell-your-sexuality-politics-surveillance-paul-lewis < thought provoker

Cathy O'Neil:

https://www.youtube.com/watch?v=gdCJYsKlX_Y

Thank you

Distil Data

James Orton
Founder and Data Scientist

M. 0432 795 658

E. james@distildata.io

W. distildata.io

Li. linkedin.com/in/jamesortonthedataman