

# Letters from the Data Science trenches

AI William Tammsaar  
Data Scientist at eAgronom

# Who am I?

- Data Scientist at eAgronom
- Occasionally consult on the topics of behavioral analytics and recommender systems
- Previous experience founding a startup, doing UX and Front-end development

# Topics for this talk

- What does doing data science as a job look like
- What do words really mean?
- What is the value of statistics?
- Why transparency is great for everyone.

# THE DATA SCIENCE HIERARCHY OF NEEDS

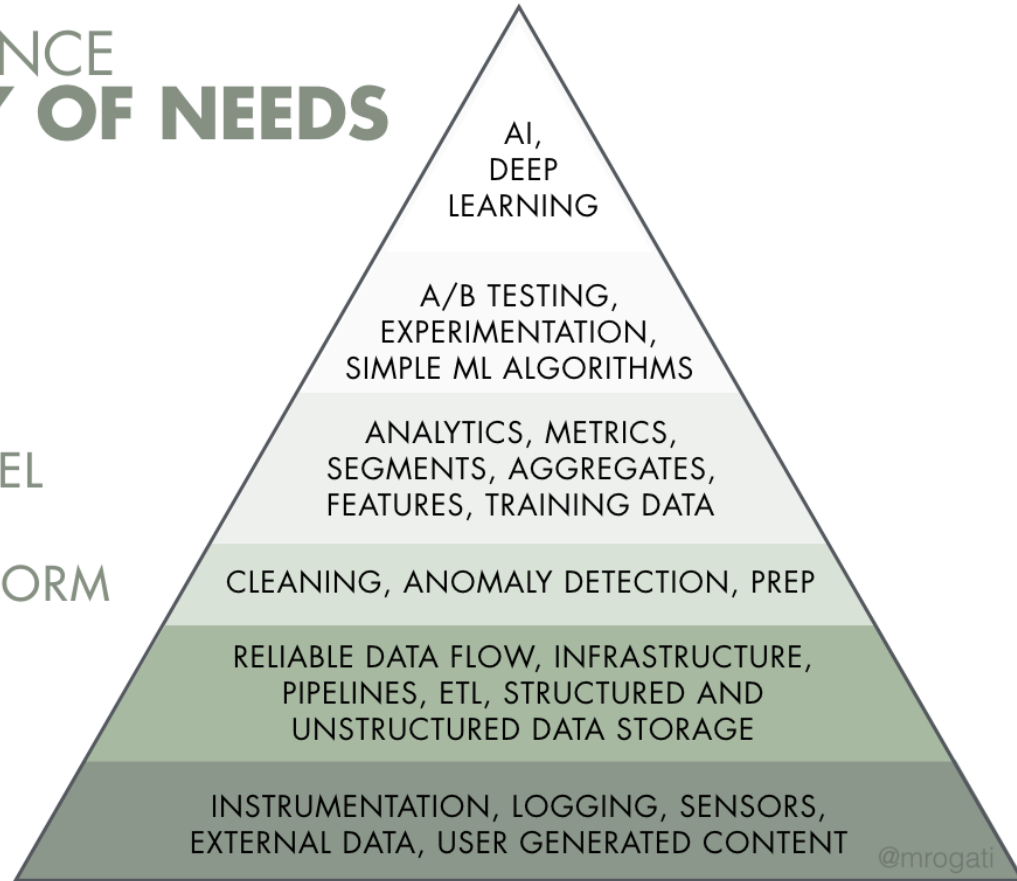
LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT



Source: Monica Rogati, [aipyramid.com](http://aipyramid.com)

**Predict  $X$  based on  $Y$**

**‘Wait what do you mean by X?’**



Source: <https://giphy.com/gifs/ryan-gosling-eating-girlfriends-zdIGTldD1mi4>

# World of Wordcraft

- The fact we don't really know what things mean is painfully obvious when we try to measure them
- You really have to make the right set of assumptions



**Exercise: What is an  
'active user'?**

# 'Active user'

- Might be for instance Monthly Active Users if we want a really high number for investors
- Might be a term for a user who is **actually using the product** as intended
- Can be used as an aspirational term. We want users to be active.

**Exercise: What is a  
'Web session'?**

# Practicalities of AB testing

**What is the  
value of AB testing?**

# AB testing values

- Measures impact (adds accountability)
- Makes sure there are no unexpected consequences (reduces risk)
- Lets you avoid talking to users (allows you to move fast)

**Bad AB testing**

# **The Psychology Replication Crisis**



# Lessons learned for us

- Set a reasonable sample ahead of time, if you start adding samples you risk getting a false positive
- Be really mindful of when you want things to be true, this clouds your judgement

**AB testing in a bad situation**



Source: <https://giphy.com/gifs/reaction-imgfave-RddAJiGxTPQFa>

# Stop over-quantifying

- Sometimes you can actually just move faster by talking to your users
- AB testing is cool, but its just a tool in a very large toolbox

**Don't let your coworkers  
explain small variances**

# Thinking Fast and Slow

- Kahneman's story of Israeli fighter pilots
- The Regression Fallacy
- The need for a story
- Practicality of a t-test

**Rounding it out:  
The value of transparency**

**Growing eAgronom's paid  
hectares by 20% overnight**



# Transparency

- We make almost all information available inside the company
- The company OKR sheet
- Noticing problems

**Thank you!**  
**Questions? Remarks?**

**We are hiring an experienced  
mobile dev btw ;)**