Power Users, Long Tail Users, and Everything In Between

Choosing Meaningful Metrics and KPIs for Product Strategy

Dror A. Guldin | Alon Nir
PyData Amsterdam
14 Sep 2023

Alon Nir
alonn@spotify.com

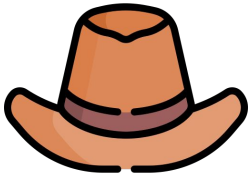
Spotify

Dror A. Guldin
dguldin@meta.com

Meta

Agenda



Introduction

The Hat Trick

Takeaways and Q&A

Introduction

The Insights Team

BI Analyst

User Researcher

Data Scientist

~~ML Engineer~~ Gen AI Tamer 💰

Economists

The Insights Team

BI Analyst

User Researcher

Data Scientist

~~ML Engineer~~ Gen AI Tamer 💰

Economists

The Different Flavours of Data Scientists

BI Analyst

User Researcher

Data Scientist

~~ML Engineer~~ Gen AI Tamer

Economists

- Business Intelligence and/or Strategy
- Marketing Science
- Product (aka “Product Analytics”)
- Infra and/or Security
- Payments & Fraud
- People Analytics

The Different Flavours of Data Scientists

BI Analyst

User Researcher

Data Scientist

~~ML Engineer~~ Gen AI Tamer

Economists

Business Intelligence and/or Strategy

Marketing Science

Product (aka “Product Analytics”)

Infra and/or Security

Payments & Fraud

People Analytics

Understand

Identify

Execute

Understand

Identify

Execute

Hindsight

What happened?

Why did it happen?

What are we going to do about it?

What happens next?

Understand

Hindsight

What happened?

Why did it happen?

What are we going to do about it?

What happens next?

Identify

Execute

Foresight

What do we want to happen?

How can we get there?

What do we know? What do we need to know?

What should we do next?

Things a Data Scientist (Analytics) Does

1

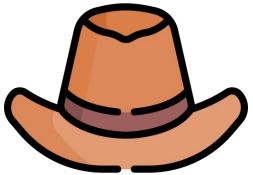
The MSc hat

Making sure the stats are tight
The “science” in “data science”
The numbers wiz
The cool kids in school

2

The MBA hat

Decisions, impact and strategy
Data in context,
transformed into actions



3

The Practitioner’s hat

Get stuff done
Beat analysis paralysis
Execute and iterate
Be the trusted advisor

The MSc Hat:
Choosing Significant Metrics

What Makes a Good Metric?

1. **Clear & Simple**
2. **Measurable, Reliable & Accurate:** when it happens, the metric says so; when the metric says so, it happened
3. **Timely & Sensitive**
4. **Informative & Actionable:** we understand what to do / what happens when the metric moves
5. **MECE** (Mutually Exclusive, Collectively Exhaustive) whenever breakdowns are involved
6. **Reproducible & Standardised** - with one “canonical” source of truth across the org
7. **Lorem:** dolor sit amet, consectetur adipiscing elit sed doloremque laudantium
8. **Ipsum:** Sed ut perspiciatis unde omnis iste natus error sit voluptatem accusantium


GOOD IDEA
BAD IDEA

Clear & Simple?

Let's use NPS (Net Promoter Score) as an example:

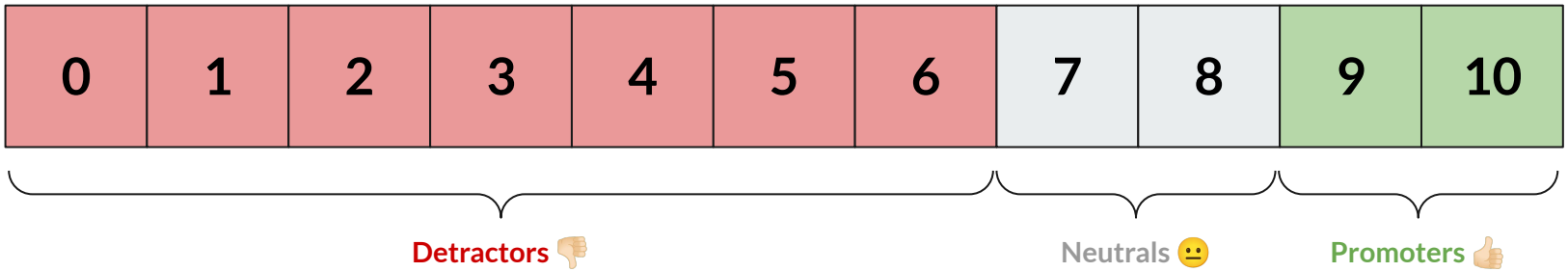
"On a scale of 0 to 10, how likely are you to recommend this PyData talk to a friend?"

0	1	2	3	4	5	6	7	8	9	10

Clear & Simple?

Let's use NPS (Net Promoter Score) as an example:

"On a scale of 0 to 10, how likely are you to recommend this PyData talk to a friend?"



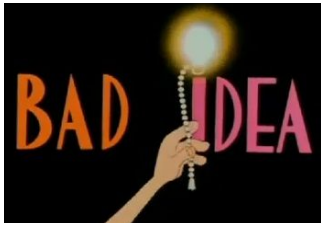
Clear & Simple?

Let's use NPS (Net Promoter Score) as an example:

"On a scale of 0 to 10, how likely are you to recommend this PyData talk to a friend?"

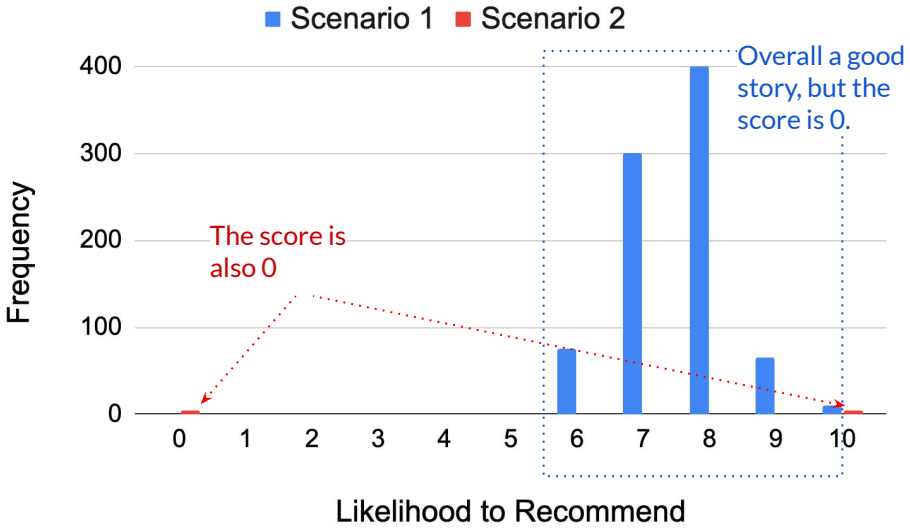
NPS = 100 * [(% of Promoters) - (% of Detractors)]

Clear & Simple?

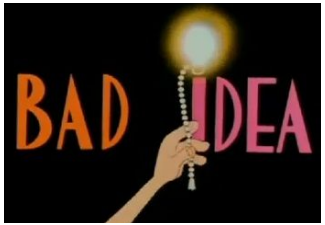


NPS (Net Promoter Score) surveys:

A very convoluted (to do, and to convery) way to gauge user satisfaction. Throws away valuable information. Introduces a third party to the company-user relationship.



Clear & Simple?



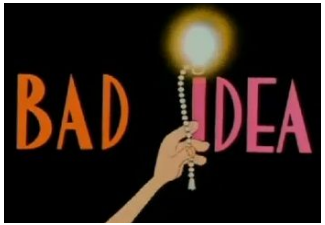
NPS (Net Promoter Score) surveys:

A very convoluted (to do, and to convery) way to gauge user satisfaction. Throws away valuable information. Introduces a third party to the company-user relationship.

Customer satisfaction surveys,

Retention and engagement metrics

Measurable, Reliable & Accurate

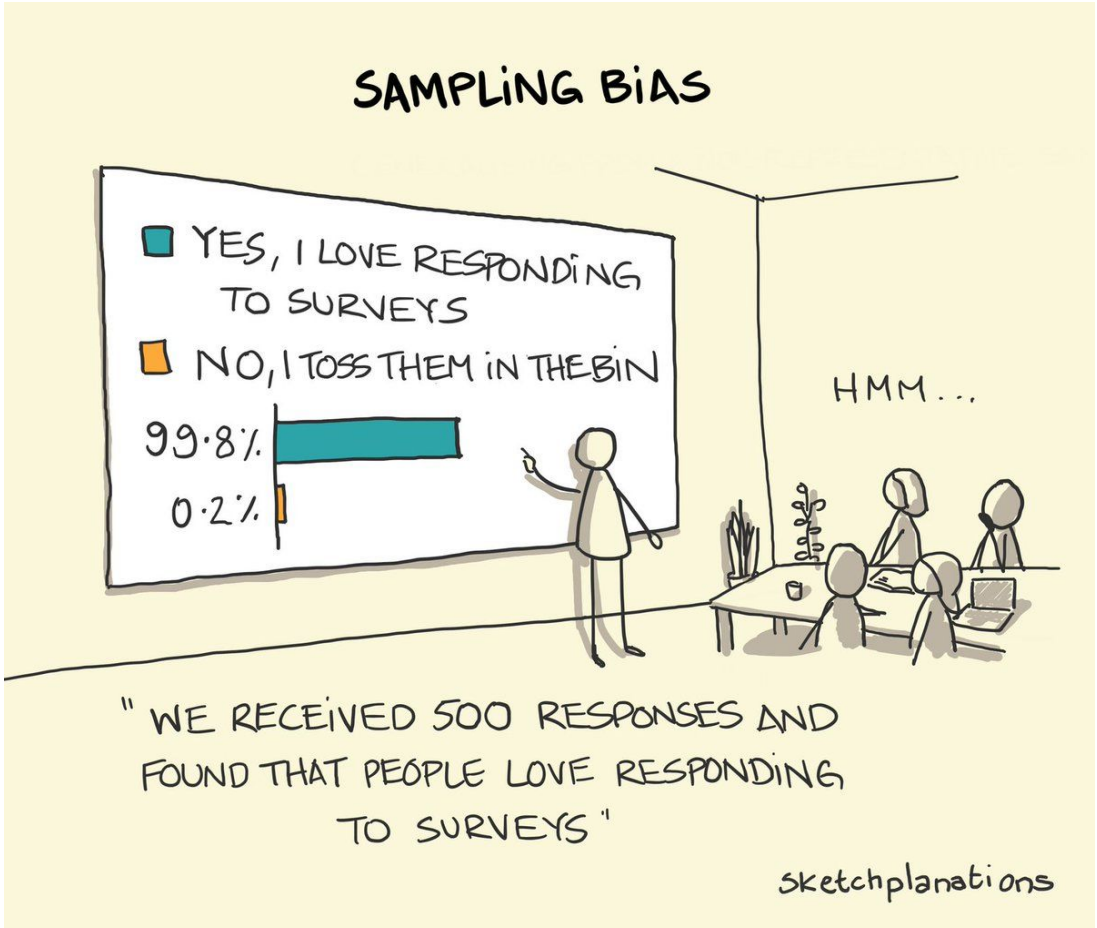


Customer satisfaction surveys (at face value)

A metric that ignores many edge cases, that is based on unreliable data sources/pipelines, that is subjective by nature, prone to selection bias, survivorship bias, or one which is very noisy (e.g. you get different values when running it at different hours of the day).

Customer satisfaction surveys (de-biased)

When the metric value is 5%, it means that it's really "5%" in the population.

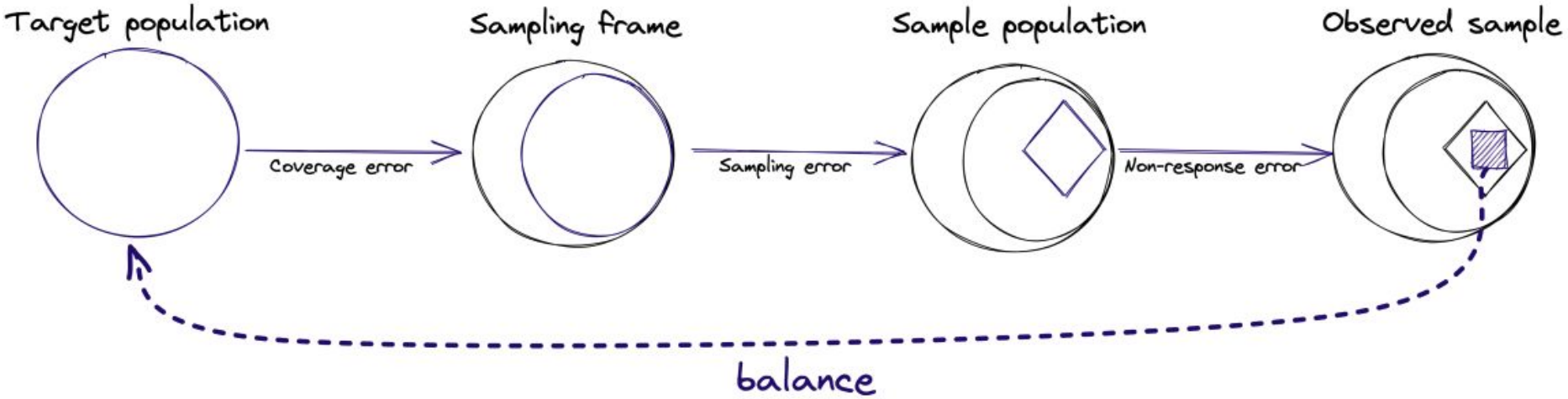


https://sketchplanations.com/sampling-bias

Shameless plug:

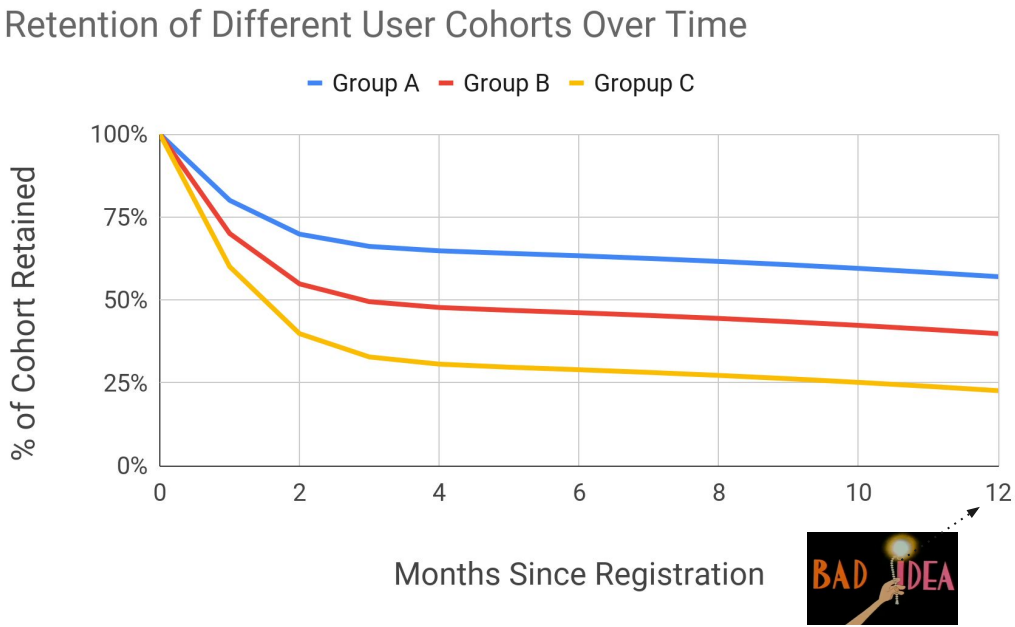
balance

`balance` - a Python package for balancing biased data samples



Timely? e.g. Long-term retention metrics

Wait 12 months to obtain 12 months retention data.



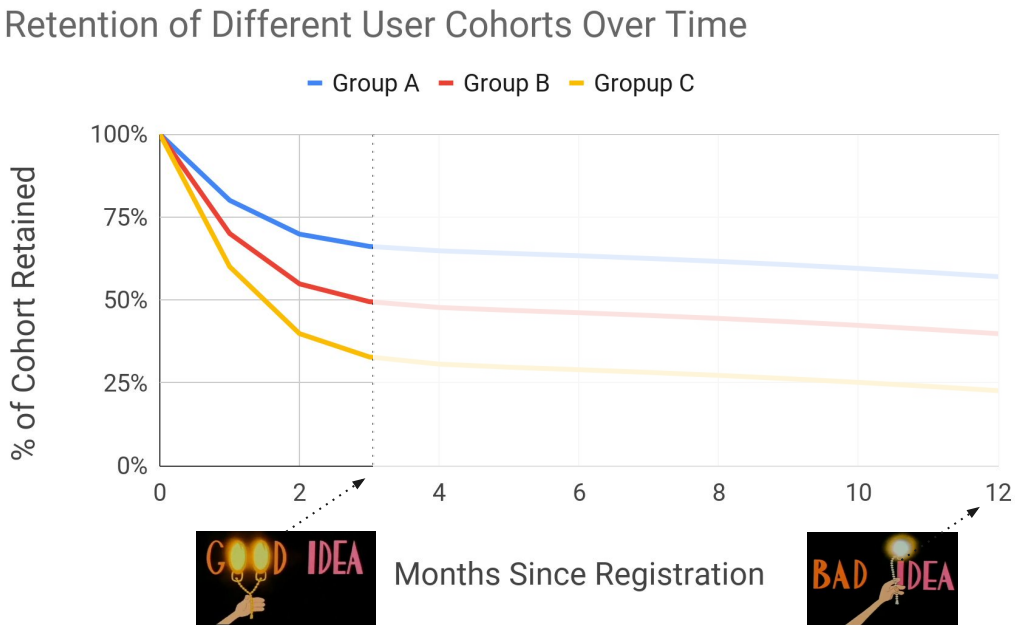
*See this post for an analysis of >5000 mobile apps.

Timely? e.g. Long-term retention metrics

Wait 12 months to obtain 12 months retention data.

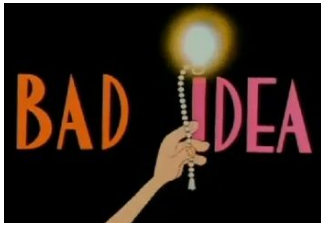
Better metrics will be proxy metrics which move faster, and serve as “good enough” predictors for long-term retention.

Good enough now > Perfect but late



*See this post for an analysis of >5000 mobile apps.

Informative & Actionable? e.g. Bing search



In an experiment*, Bing search recorded an uptick of 10% in queries per user, and 30% (!!) in revenue per user.

The only problem? A bug degraded search results.

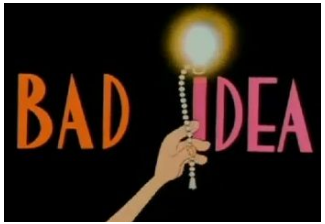
Bad idea: degrade search results on purpose.

In addition to engagement, measure task (e.g. search) success and retention.

Track the long run, as *short term gains* can lead to *long term harm*.

*https://www.researchgate.net/publication/237838307_Trustworthy_Online_Controlled_Experiments_Five_Puzzling_Outcomes_Explained

MECE (Mutually **E**xclusive**, C**ollectively **E**xhaustive**)**



Breakdown your audience into overlapping groups (e.g. Teens, Young Adults & Adults), and/or cover some of the audience

This won't allow for complete accounting whenever things move.

Every user is part of one category (*Collectively Exhaustive*) and only one category (*Mutually Exclusive*) per metric

E.g. Age <9, 9-12, 13-19, 20-30, 30+

Standardisation

Teams across the company should speak the same language, and their words need to mean the same.

Classic examples: retention, active users.

Standardisation through playbooks and tools solve for that.

https://engineering.atspotify.com/2023/08/coming-soon-confidence-an-experimentation-platform-from-spotify/
https://engineering.atspotify.com/2020/10/spotifys-new-experimentation-platform-part-1/

Does Your Metric Meet the “Scientific” Bar?

- ❑ Measurable, Reliable & Accurate
- ❑ Clear & Simple
- ❑ Timely & Sensitive
- ❑ Informative & Actionable
- ❑ MECE
- ❑ Reproducible & Standardised

The MBA Hat:
Setting The Right KPIs / OKRs

Businessy Considerations Around KPIs & OKRs

1. The KPIs / OKRs philosophies
2. North Star Metrics Drive Strategy
3. Goodhart's Law
4. Different Users Drive Different Metrics

Differences Between KPIs & OKRs

01

KPIs (Key Performance Indicators): How do we measure the ongoing success of the operation? KPIs tend to be more static (and so comparable over longer periods of time), and focus on maintaining and improving existing performance levels

OKRs (Objectives & Key Results): Focused more on setting bold goals "beyond the ongoing operation"

Both KPIs & OKRs leverage **measurable and quantifiable "north star metrics" for the organization**

KPI: y/y revenue growth

OKR:
Objective - become the #1 service for customer experience in the Netherlands
Key Result - increase customer satisfaction score by 10%

North Star Metrics Drive Strategy

02

“Management is doing things right; leadership is doing the right things” Peter Drucker

What metrics do you see in company’s quarterly reports? What do you/would you put in yours?

- **Facebook /IG: Daily & Monthly Active Users**
- **Spotify: Premium Subscribers & Ad-Supported MAU**
- **YouTube: Minutes watched**

Goodhart's Law: All Metrics Are Imperfect

03

"When a measure becomes a target, it ceases to be a good measure" (Charles Goodhart, 1981)

"Not everything that can be counted counts, and not everything that counts can be counted" (William Bruce Cameron, 1963)

Practical Implications:
All metrics are going to have limitations. Stay focused on what's good for users not what's good for metrics.

Real-World Example: North Star Metric for Threads

“On Threads … We saw unprecedented growth out of the gate and more importantly we’re seeing more people coming back daily than I’d expected. And now, we’re focused on retention and improving the basics. And then after that, we’ll focus on growing the community to the scale we think is possible. Only after that will we work on monetization. We’ve run this playbook many times before - with Facebook, Instagram, WhatsApp, Stories, Reels, and more – and this is as good of a start as we could have hoped for, so I’m really happy with the path we’re on here.

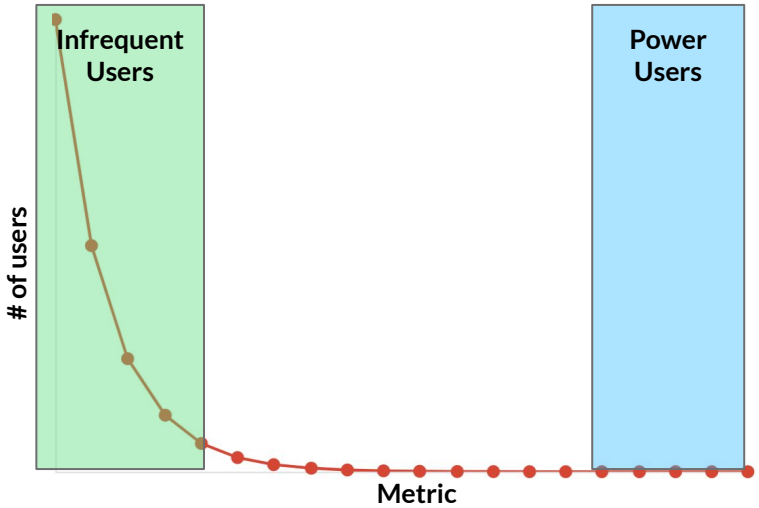
Mark Zuckerberg, in Meta’s Second Quarter 2023 Results Conference Call (pdf)

https://s21.q4cdn.com/399680738/files/doc_financials/2023/q2/META-Q2-2023-Earnings-Call-Transcript.pdf

Different Users Drive Different Metrics

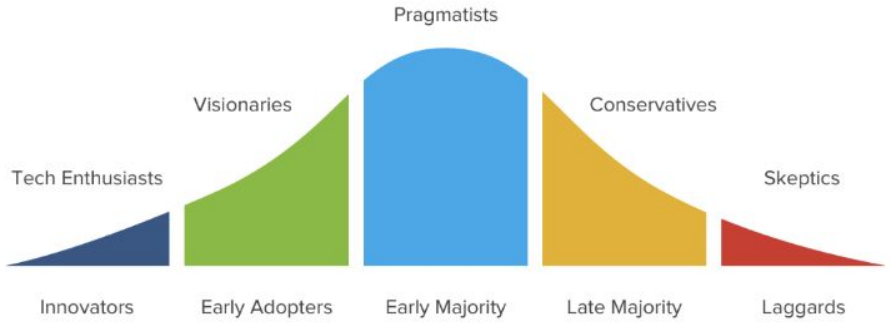
04

- **Retention metrics are driven by marginal users** (and retention metrics are #1 for driving growth)
- **Engagement metrics are driven by power users** (and engagement metrics are #1 for building awesome products)
- **Ratio metrics easily create the wrong dynamics**
- **Fancier metrics can be more "accurate", but also harder to reason about** (remember Goodhart!)



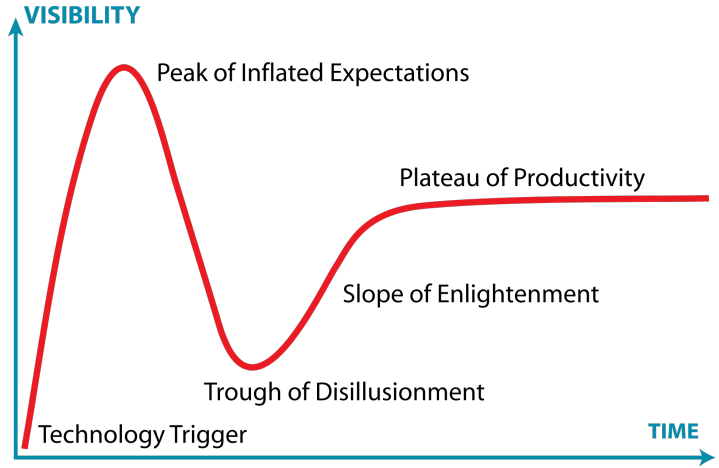
Practical Implications:
Figure out the segment of users you want your team to focus on.
Always look at the underlying distribution & mechanism of changes (Are we doing things better/worse? Is there a user mix-shift?)

… and Users and Usage Patterns Change Over Time



Product Diffusion Curve

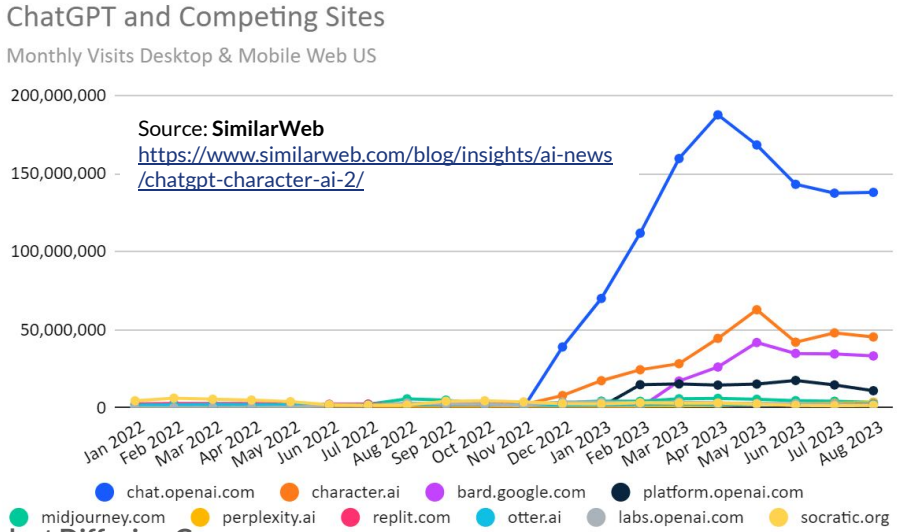
https://www.free-power-point-templates.com/articles/new-product-diffusion-curve-slide-for-powerpoint/



Gartner Research's Hype Cycle diagram

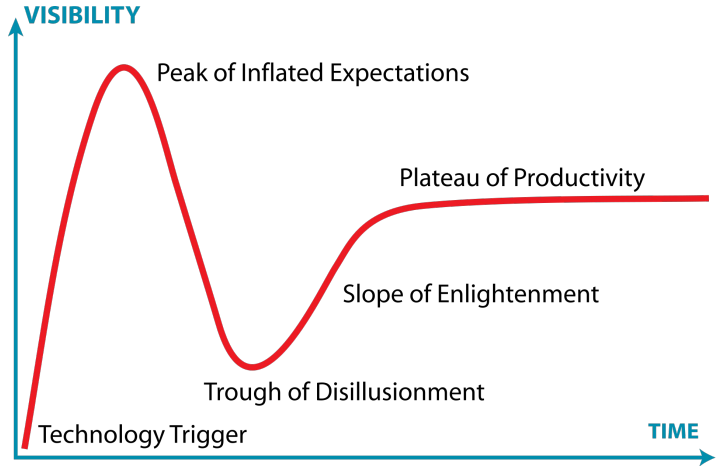
By Jeremykemp, https://en.wikipedia.org/wiki/Gartner_hype_cycle, CC BY-SA 3.0

… and Users and Usage Patterns Change Over Time



Product Diffusion Curve

https://www.free-power-point-templates.com/articles/new-product-diffusion-curve-slide-for-powerpoint/

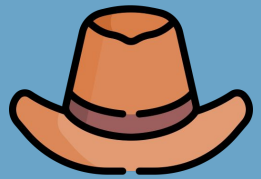


Gartner Research's Hype Cycle diagram

By Jeremykemp, https://en.wikipedia.org/wiki/Gartner_hype_cycle, CC BY-SA 3.0

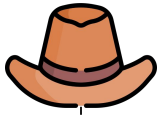
Does Your Metric Make Enough Business Sense?

- ❏ Should you go with KPIs or OKRs?
- ❏ Is it clear in strategic direction this “north star” is leading?
- ❏ Are you continuously aware of Goodhart’s Law?
- ❏ Do you understand which user segments move your metric (and how)?



Tying things together: Data-Driven Decision Making In Practice

Leveraging Data to Inform Decision-Making



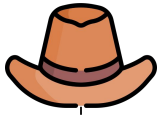
Without data, the DS is just a person with an opinion

1. Prioritise measuring & evaluating impact
2. Build the datasets you need
3. Sniff your data
4. Conduct back-of-the-envelope opportunity sizing

Without an opinion, the DS is just a person with data

5. Acknowledge when you don't really need data (just because you can, doesn't mean you should)
6. Leverage more than "just" data (UXR, product vision, ...)
7. Have a "strong opinion, weakly held" mindset
8. It doesn't have to be "science" but it can't be "alchemy"

Leveraging Data to Inform Decision-Making



Without data, the DS is just a person with an opinion

1. Prioritise measuring & evaluating impact
2. Build the datasets you need
3. Sniff your data
4. Conduct back-of-the-envelope opportunity sizing

Without an opinion, the DS is just a person with data

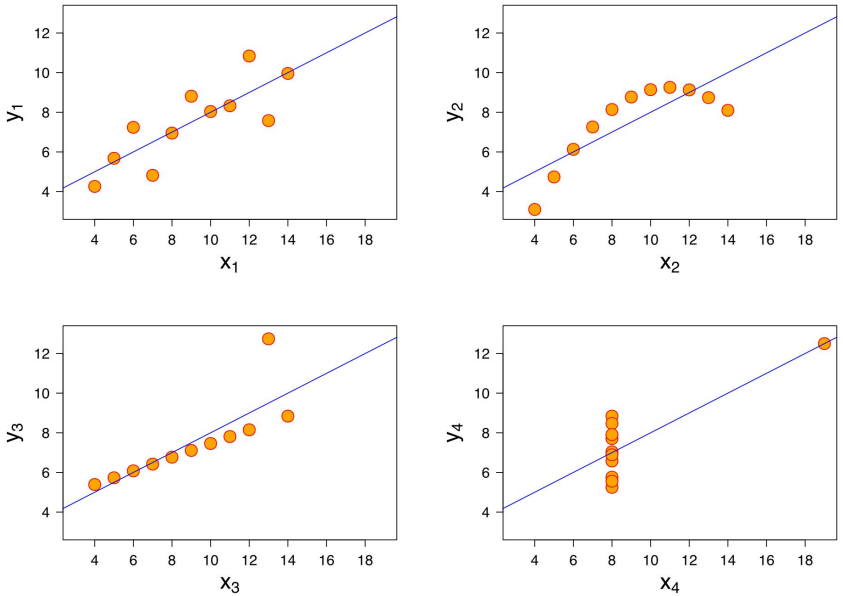
5. Acknowledge when you don't really need data (just because you can, doesn't mean you should)
6. Leverage more than "just" data (UXR, product vision, ...)
7. Have a "strong opinion, weakly held" mindset
8. It doesn't have to be "science" but it can't be "alchemy"

Sniff Your Data

Twyman's law: "Any figure that looks interesting or different is usually wrong" … "the more unusual or interesting the data, the more likely they are to have been the result of an error of one kind or another"

Sniff your data by

- Looking at raw data samples
- Looking at a variety of summary statistics
- Visualizing your data



Anscombe's quartet
By Schultz, Anscombe's quartet - Wikipedia, CC BY-SA 3.0

Back-of-the-Envelope Opp Sizing

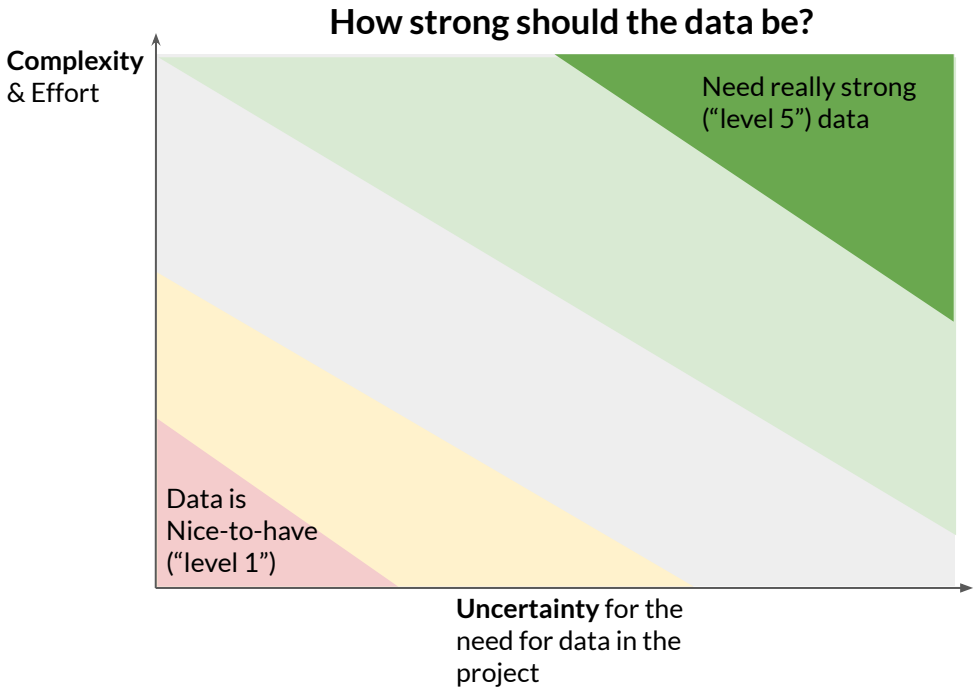
- Approximate the impact on your north star metric
- Do this before your team commits to a project
- Incorporate this in your periodic planning
- Use simplified assumptions
- Don’t try to be accurate, try to be quick (and close enough)

Acknowledge When You Don't Really Need Data

Help your org beat analysis paralysis

It doesn't have to be "science" but it can't be "alchemy"

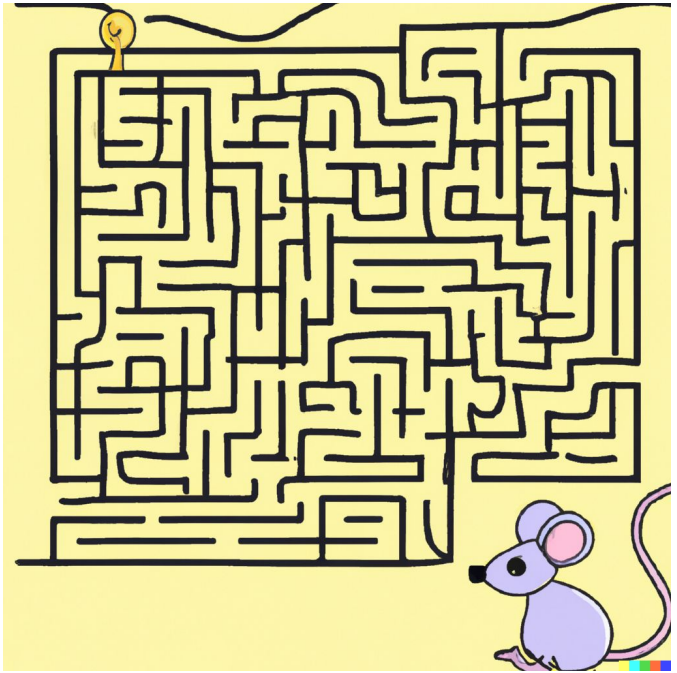
Keep correcting! Drive for continuous monitoring, optimization & experimentation. Help the team "fail fast" and correct course



Strong Opinion, Weakly Held

"**Allow your intuition to guide you to a conclusion, no matter how imperfect — this is the "strong opinion" part. Then –and this is the "weakly held" part– prove yourself wrong.**" Paul Saffo

- What is your working assumption based on information you currently have?
- What new data will be "cheapest" to get, and that might convince you otherwise?



Walk the Talk

- ❏ Prioritise measuring & evaluating impact
  - ❏ Sustain a culture of experimentation
- ❏ Conduct opportunity sizing
- ❏ Build the datasets you need
- ❏ Sniff your data
- ❏ Acknowledge when you don’t really need data
- ❏ Leverage more than “just” data
- ❏ Keep a “Strong opinion, loosely held” mindset
- ❏ It doesn’t have to be “science” but it can’t be “alchemy”

Takeaways

Take-Home Messages

1. **Without data you're just a person with an opinion.** Make sure the stats are tight (that's the "science" part of "data science"). Sniff your data!

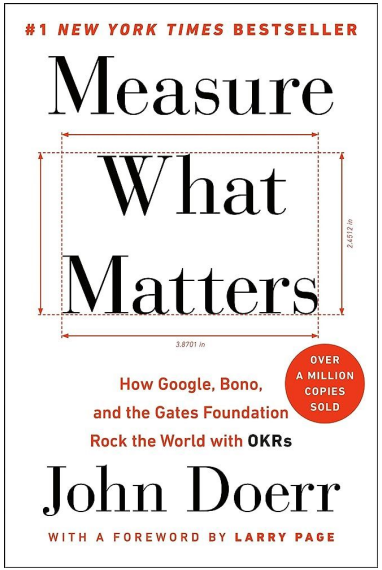
2. **Without an opinion you're just a person with data**. Help your org avoid analysis paralysis

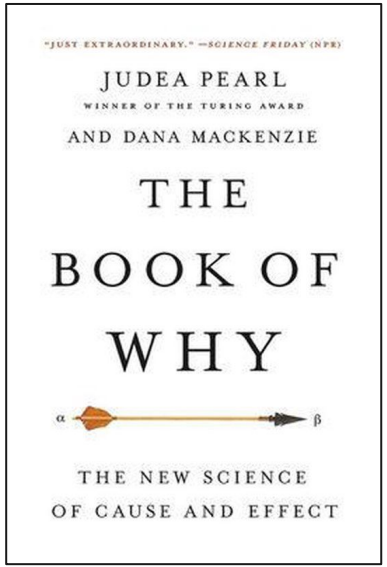
3. **Defining north star metrics == steering an org.** Make sure your org is optimising for users, not for metrics. Don't focus on the numbers, focus on what they represent.

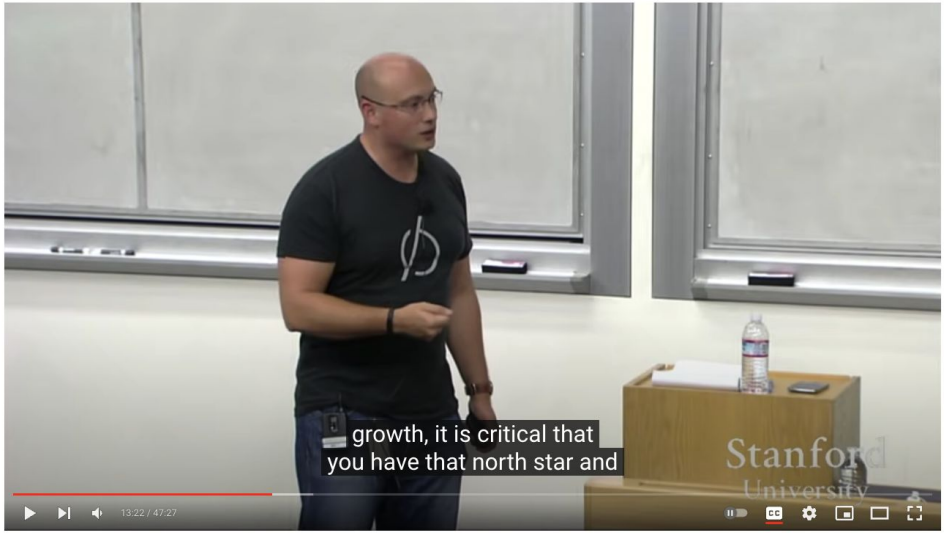
4. **All metrics are going to have limitations.** Don't look for the perfect KPI, look for the one that will best help your org focus.

5. **Product analytics doesn't have to be "science" but it can't be "alchemy"**

Interested in Diving Deeper?







Lecture 6 - Growth (Alex Schultz)

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

Questions?

Thank You!

Slides, check-lists, GIFs of puppies, goodies →
https://github.com/alonnir/PyData-Amsterdam-2023/

This slide intentionally left blank

Image credits

Takeaway icons created by Freepik - Flaticon

Wizard icons created by Freepik - Flaticon

Gentleman icons created by Freepik - Flaticon

Lock icons created by Freepik - Flaticon

Qa icons created by Freepik - Flaticon

Image credits

Hat icons created by Freepik - Flaticon

Hello icons created by Kalashnyk - Flaticon