

ANALYTICS 101

THOMSON NGUYEN (@ITSTHOMSON)
FOUNDER/CEO, FRAMED DATA

TEAM



Thomson Nguyen
Co-Founder and CEO



Previously: Data Scientist at Lookout Mobile Security, Data Scientist at Causes.
Educated at UC Berkeley (Mathematics) & University of Cambridge (Mathematics).



 **Lookout**
causes



Elliot Block
Co-Founder and CTO



Previously: Program Manager at Microsoft, Software Engineer at Causes.
Educated at UC Berkeley (Sociology, Computer Science) & University of Washington.



 **Microsoft**
causes

ADVISORS



Kevin Mahaffey
CTO, Lookout



Sumon Sadhu
Dir. Intelligence, Quid Inc.

 **Combinator**

**Initialized
CAPITAL**

 **Google
ventures**

 创新工场
INNOVATION WORKS

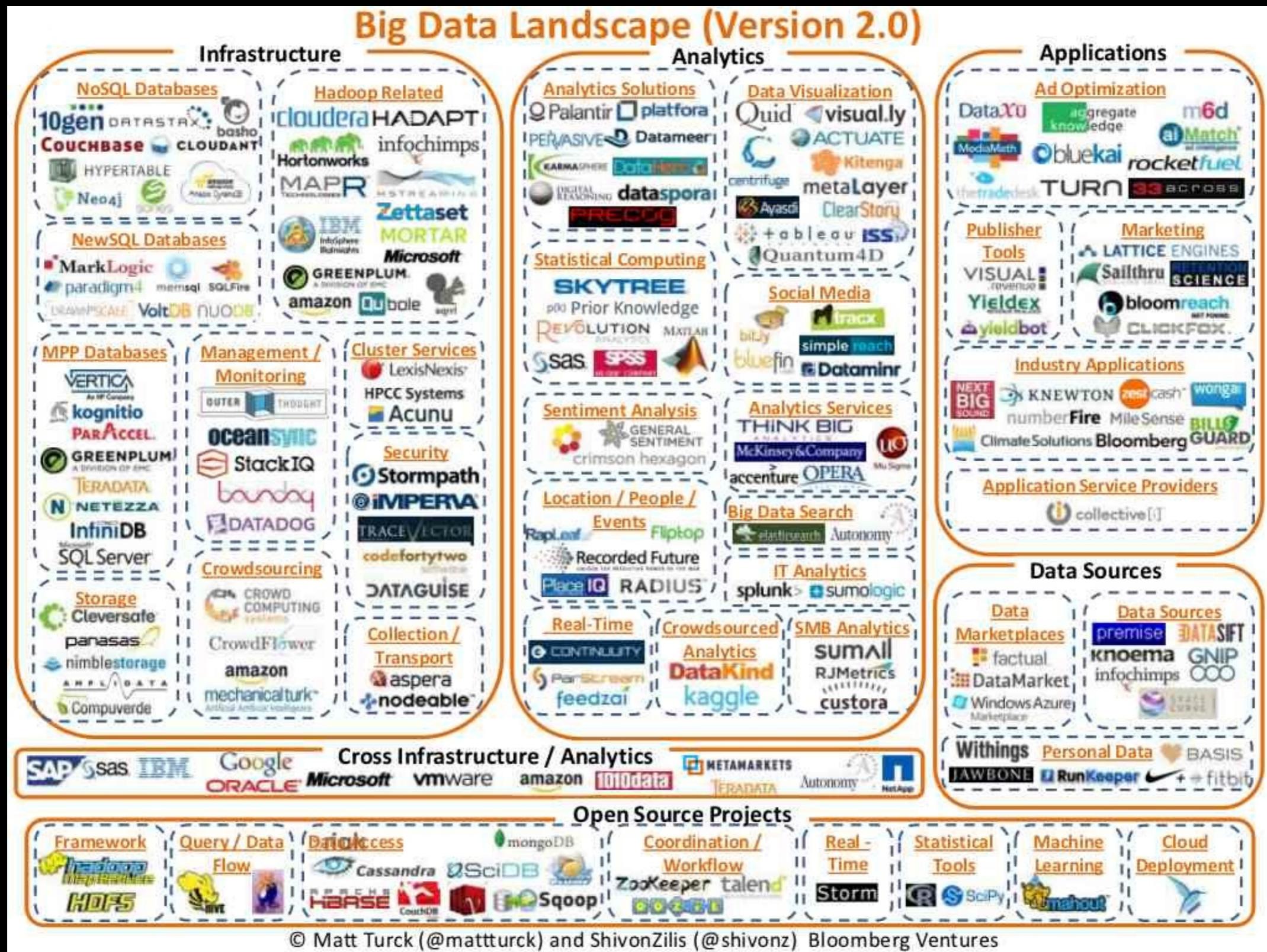
 **jotter**

Paul Buchheit, Kevin
Mahaffey, Sumon Sadhu,
Sean Byrnes, Klaus von Sayn-
Wittgenstein, Zac Auger

INVESTORS

**We tell businesses why their
customers leave, and when
they're about to leave.**

The Big Data landscape is commoditized:



Few companies focus on adding direct business value.

HOW DO I BECOME A DATA SCIENTIST?

OR, SOME SKILLS I FOUND USEFUL TO BE ONE

**Michael E. Driscoll**

@medriscoll

**Following**

Data scientists: better statisticians than
most programmers & better programmers
than most statisticians bit.ly/NHmRqu

[@peteskomoroch](#)



Reply



Retweet



Favorite



More



Pocket



[~undefined behavior~](#)

@silentbicycle



[Follow](#)

"What is a 'Data Scientist'? An analyst who
lives in California." -[@edmundjackson](#)
[#clojure_conj](#)

[Reply](#) [Retweet](#) [Favorite](#) [More](#)

RETWEETS

19

FAVORITES

5



6:00 AM - 16 Nov 2012



James Wickett

@wickett



[Follow](#)

RT @cote: “A data scientist is just an analyst who lives in California.” In Texas they are called data charlatans

[Reply](#) [Retweet](#) [Favorite](#) [More](#)

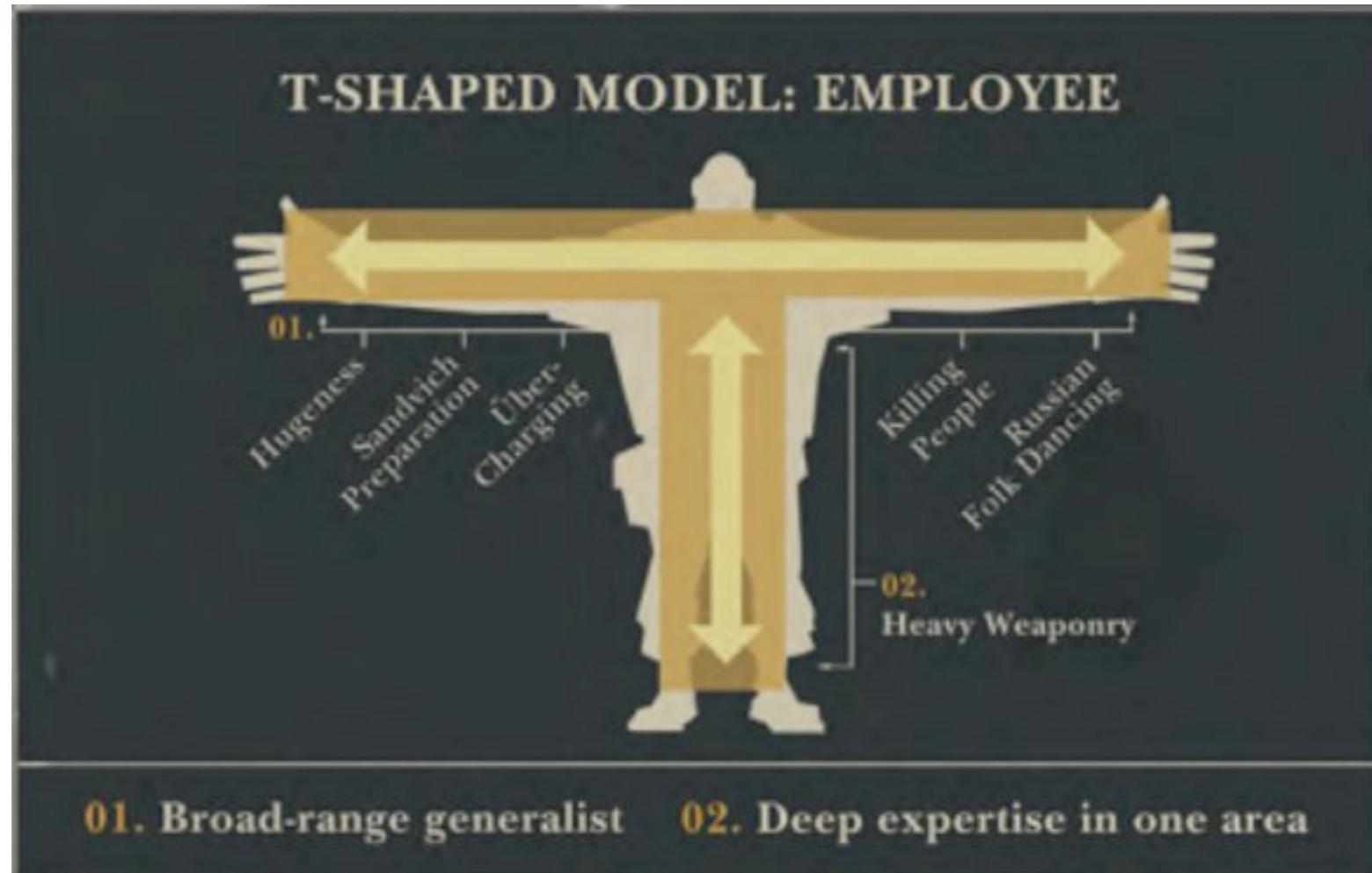
RETWEETS

8



6:39 AM - 4 Nov 2012

- › Statistical and machine learning knowledge
- › Engineering experience
- › Academic curiosity
- › Product sense
- › Storytelling
- › Cleverness



(Valve Software)





Academic curiosity: a desire to go beneath the surface and discover and distill a problem down into a very clear set of hypotheses that can be tested.

Golden Empty Nesters



Financially secure couples, many close to retirement,
living in sought after suburbs



Golden Empty Nesters contains wealthy older people living in large detached houses, often in choice residential locations in semi rural settings.

Who We Are	
Age	55-64 (16.42%)
Marital Status	Married (68.75%)
Household Composition	Couples, no children (23.71%)
Length of Residency	9+ years (62.22%)
Health	Good diet and health Drink alcohol daily
Our Education	
Adults	Degree level
Children	University admissions
Our Work Lives	
Occupations	Director, small company Female manager/professional
Our Finances	
Household Income	£25,000 - £49,999
Benefits	State pension
Indebtedness	Low
Where We Live	
Type of Property	Detached house
Council Tax Bands	Bands E-G
Home Ownership	Own home outright
House Value	£205k
Location	Semi-rural
Our Home Lives	
Car ownership	3 or more cars
Travel & holidays	Cruise
Leisure Interests	Grandchildren Classical music
Media	Connected to the Internet Daily Telegraph
Weltanschauung "How We View the World"	
Our Neighbourhood	Good place to live Neighbours help each other
Charities	Animal welfare Religious
Crime	Likely to occur away from home
Environment	Pay more for eco-friendly products
Fear of Redundancy	Not at all concerned

Storytelling: the ability to use data to tell a story and to be able to communicate it effectively to nontechnical stakeholders.

HOW TO BE A DATA SCIENTIST

16

The screenshot shows a product page for the game "Angry Birds Star Wars Premium HD (Kindle Tablet Edition)" on the Amazon Appstore for Android. The main content includes the game's cover art, title, price (\$2.99), and customer reviews. To the right, there's a sidebar for the Cyber Monday Deals Week, compatibility information, and a 'Customers Who Bought This Item Also Bought' section featuring other popular games.

Cyber Monday Deals Week [Shop now](#)

Presented by Amazon.com Rewards Visa Card

Hello, Thomson Your Account [Cart](#) [Wish List](#)

Go

Angry Birds Star Wars Premium HD (Kindle Tablet Edition)

by [Rovio Entertainment Ltd.](#)

Platform: Android Rated: [Guidance Suggested](#)

[4.5★ \(402 customer reviews\)](#) | [Like \(318\)](#)

Price: \$2.99

Sold by: Amazon Digital Services, Inc.

Available instantly for your Android device

Buy now with 1 Click

Compatibility with your devices

Thomson's Kindle

motorola DROID BIONI...

See all devices

How buying works

Redeem a gift card or promotion code & view balance

Add to Wish List

Share [Email](#) [Facebook](#) [Twitter](#) [Pinterest](#)

[Send us feedback about the Amazon Appstore for Android]

Click for larger image and other views

Share your own related images

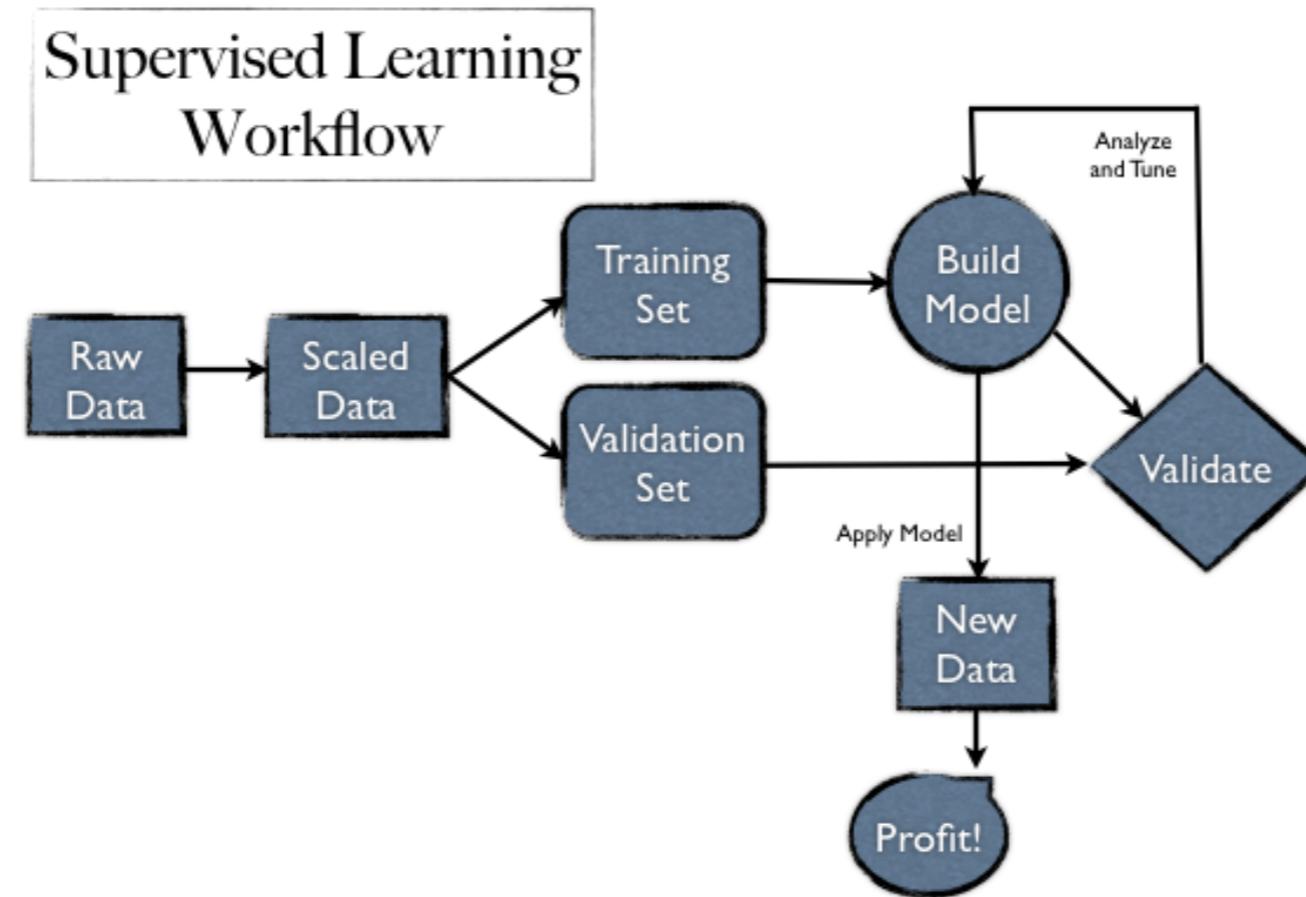
By placing your order, you agree to our [Terms of Use](#)

Customers Who Bought This Item Also Bought

Page 1 of 10

Product	Developer	Rating	Price
Angry Birds Star Wars (Ad-Free)	Rovio Entertainment Ltd.	4.5★ (116)	\$0.99
Angry Birds Space HD (Kindle Tablet Edition)	Rovio Entertainment Ltd.	4.5★ (1,352)	\$2.99
The Amazing Spider-Man (Kindle Tablet Edition)	Gameloft	4.5★ (329)	\$6.99
Skylanders Cloud Patrol (Kindle Tablet Edition)	Activision Publishing	4.5★ (166)	\$0.99
Bad Piggies Premium HD (Kindle Tablet Edition)	Rovio Entertainment Ltd.	4.5★ (302)	\$2.99
Plants vs. Zombies (Kindle Tablet Edition)	PopCap Games	4.5★ (1,075)	\$0.99
Angry Birds Rio (Ad-Free)	Rovio Entertainment Ltd.	4.5★ (2,935)	\$0.99
Minecraft - Pocket Edition	Mojang	4.5★ (2,278)	\$6.99

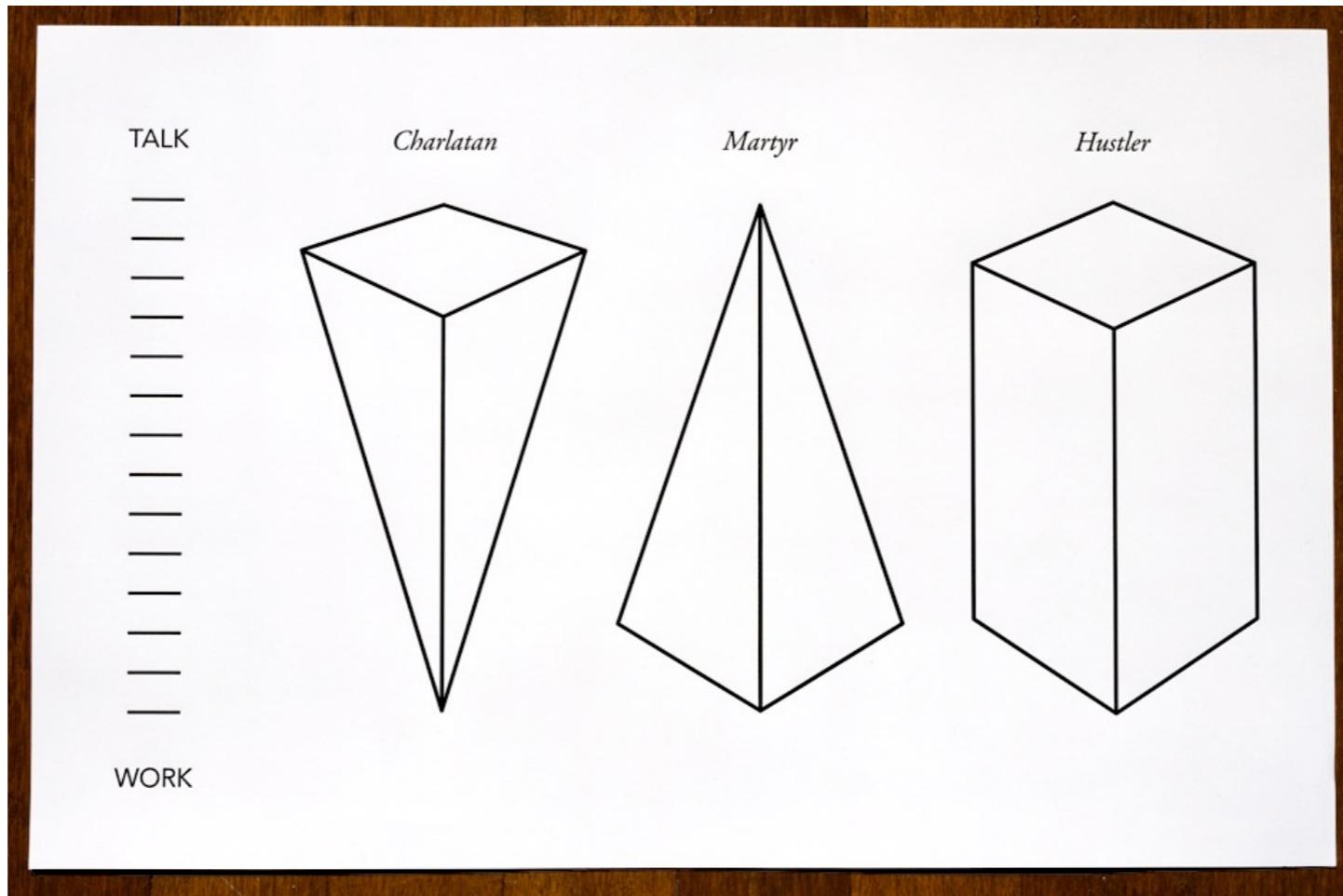
Product sense: the ability to use this story to create a new product, or change an existing product in a way that improves company goals and metrics.



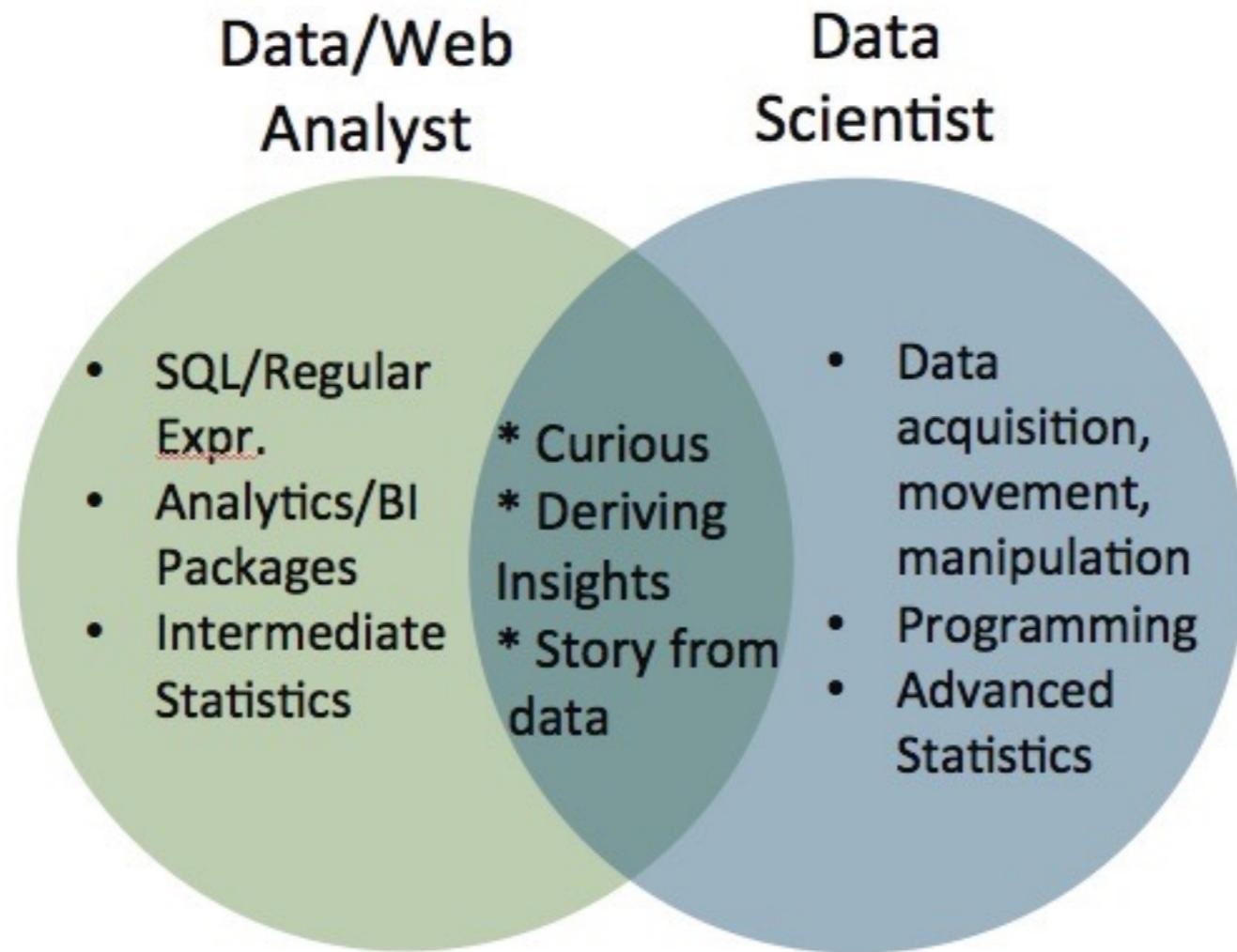
Statistical and machine learning knowledge: the domain expertise required to acquire data from different sources, create a model, optimize its accuracy, validate its purpose, and confirm its significance.

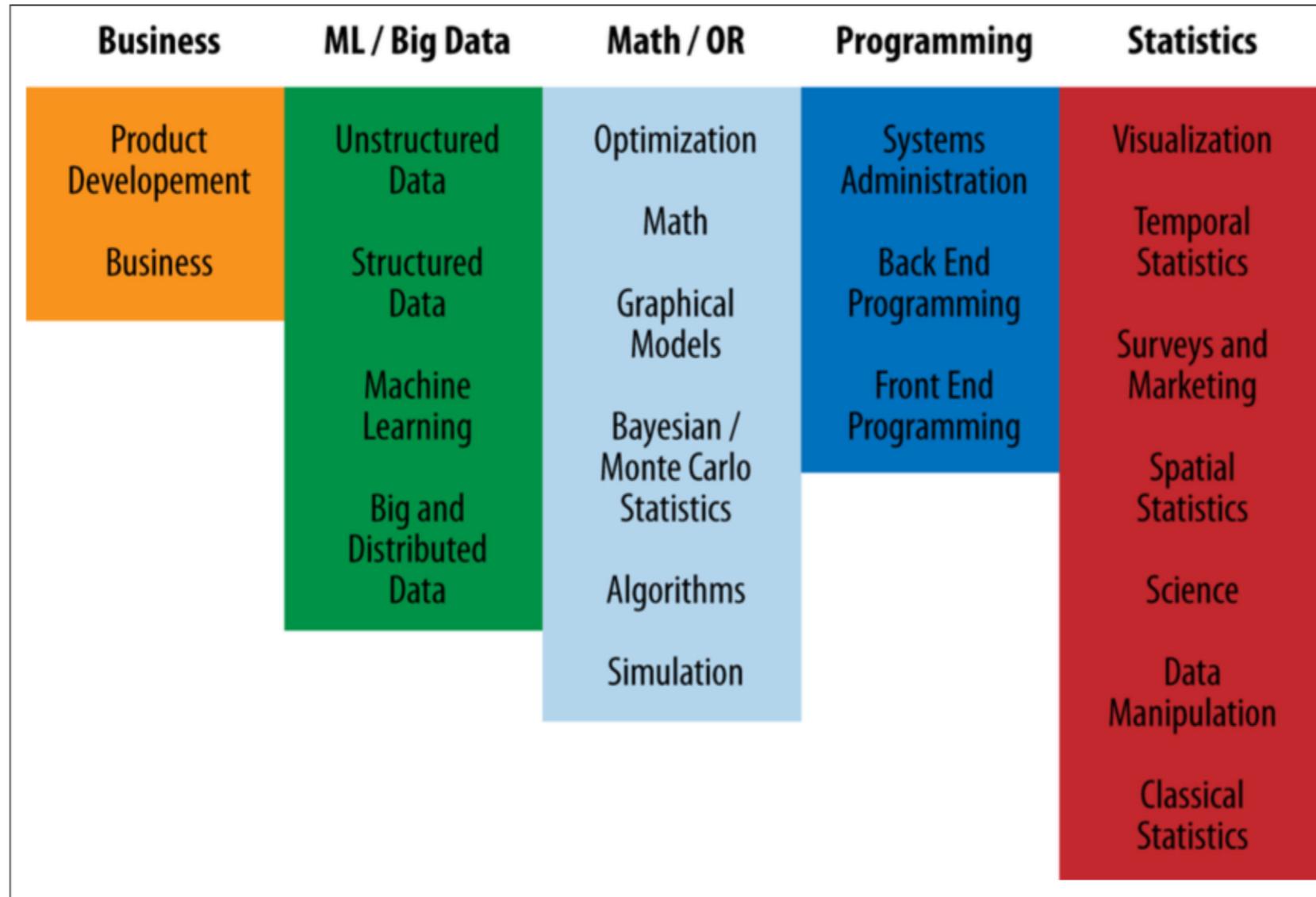
```
> control.rfe.rf <- rfeControl(functions = hhp.functions,
  rerank = TRUE,
  workers = 256,
  method = "repeatedcv",
  number = 25,
  returnResamp = "final",
  computeFunction = mclapply,
  computeArgs=list(mc.preschedule = FALSE,
                  mc.set.seed = FALSE)
)
> sizes <- seq(200, 24442 ,by=100)
> model.1 <- rfe(model.training, model.training.response, sizes,
  metric = "HHP", maximize = FALSE,
  rfeControl = control.rfe.lm)
```

Engineering experience: the coding chops necessary to implement and execute a statistical model and its validation.



Cleverness: the creativity to do all of these things on a deadline and on constrained resources.





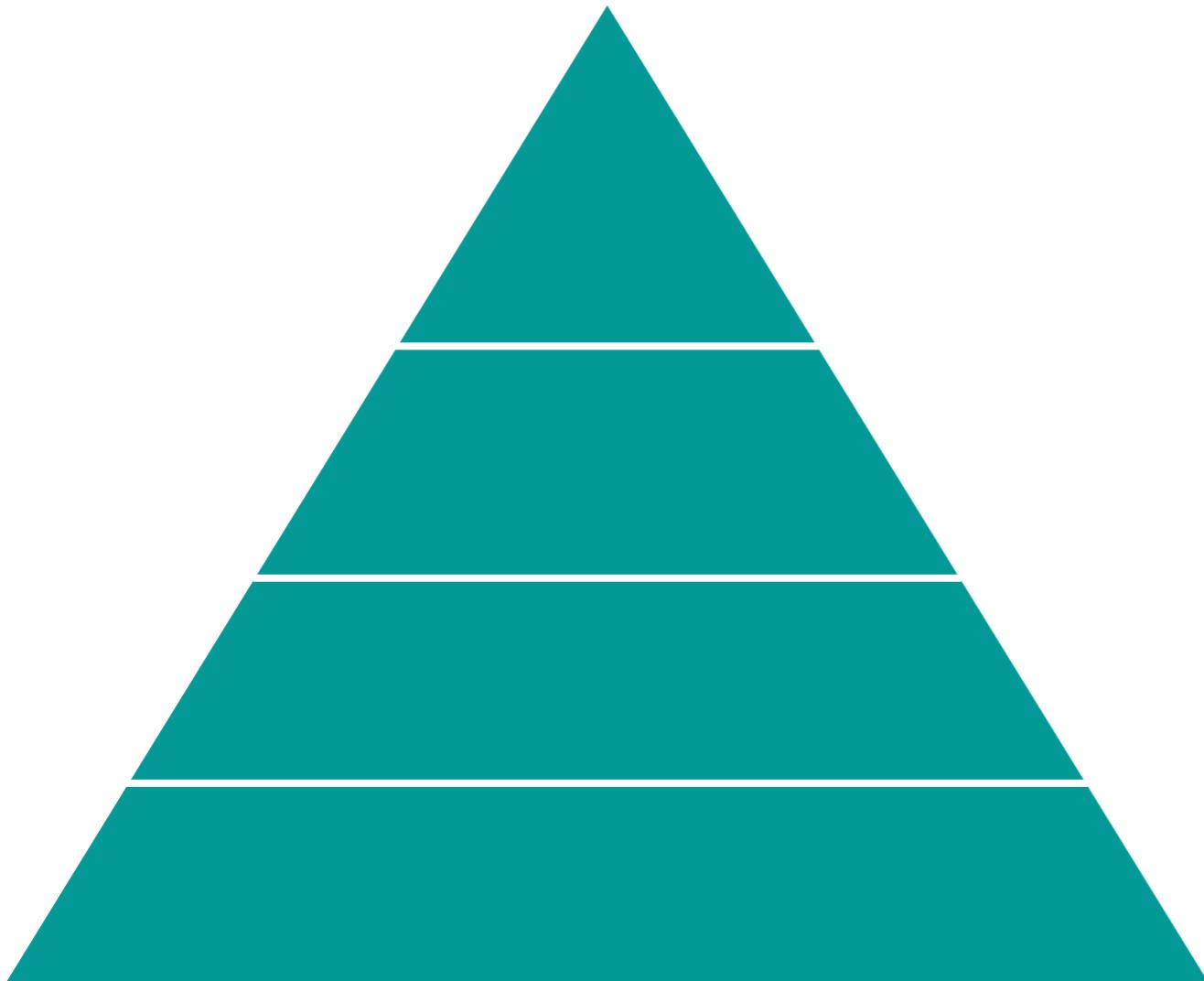
ANALYTICS 101

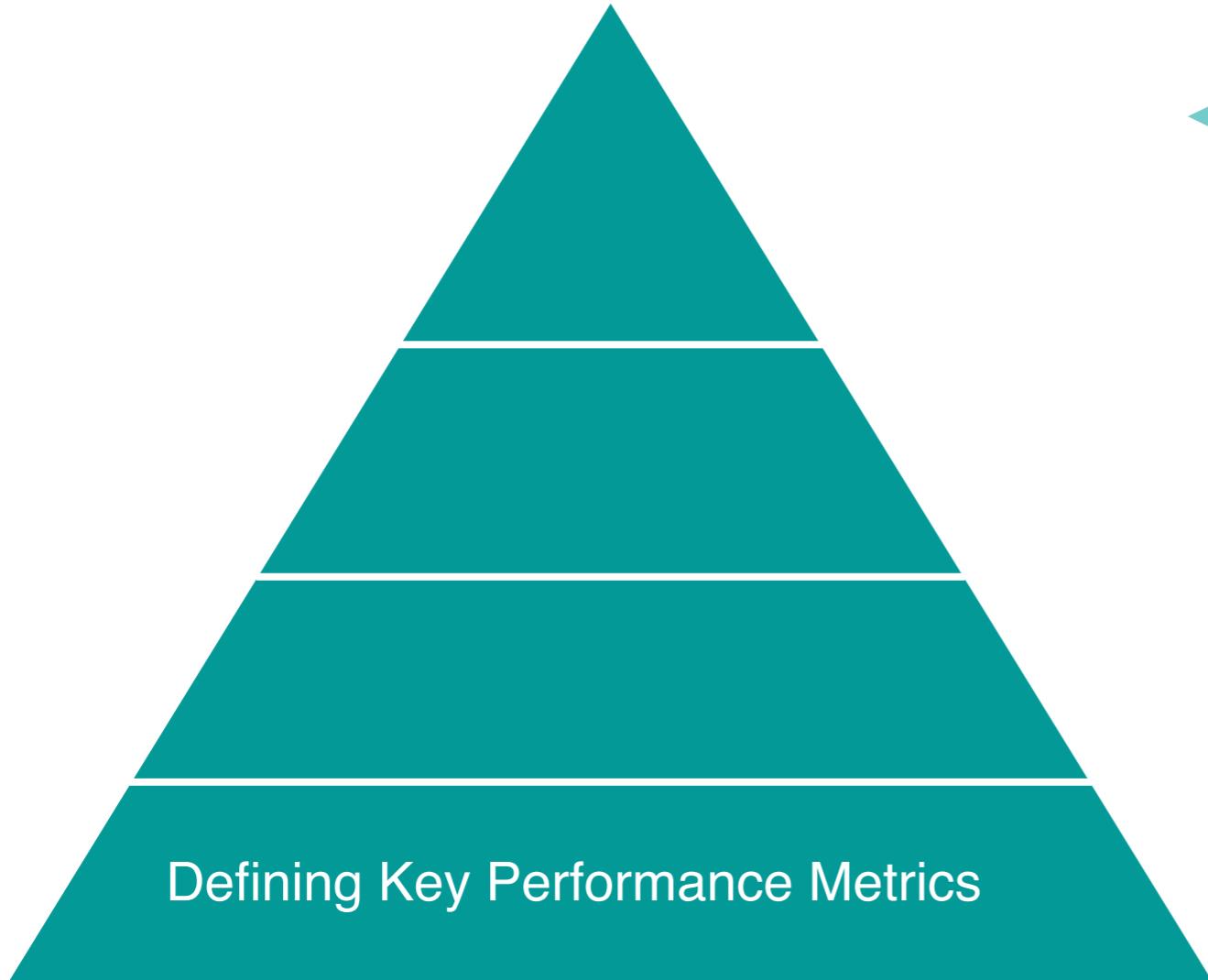
THE DATA PYRAMID

OR, THE ROAD TO BEING A DATA-DRIVEN COMPANY

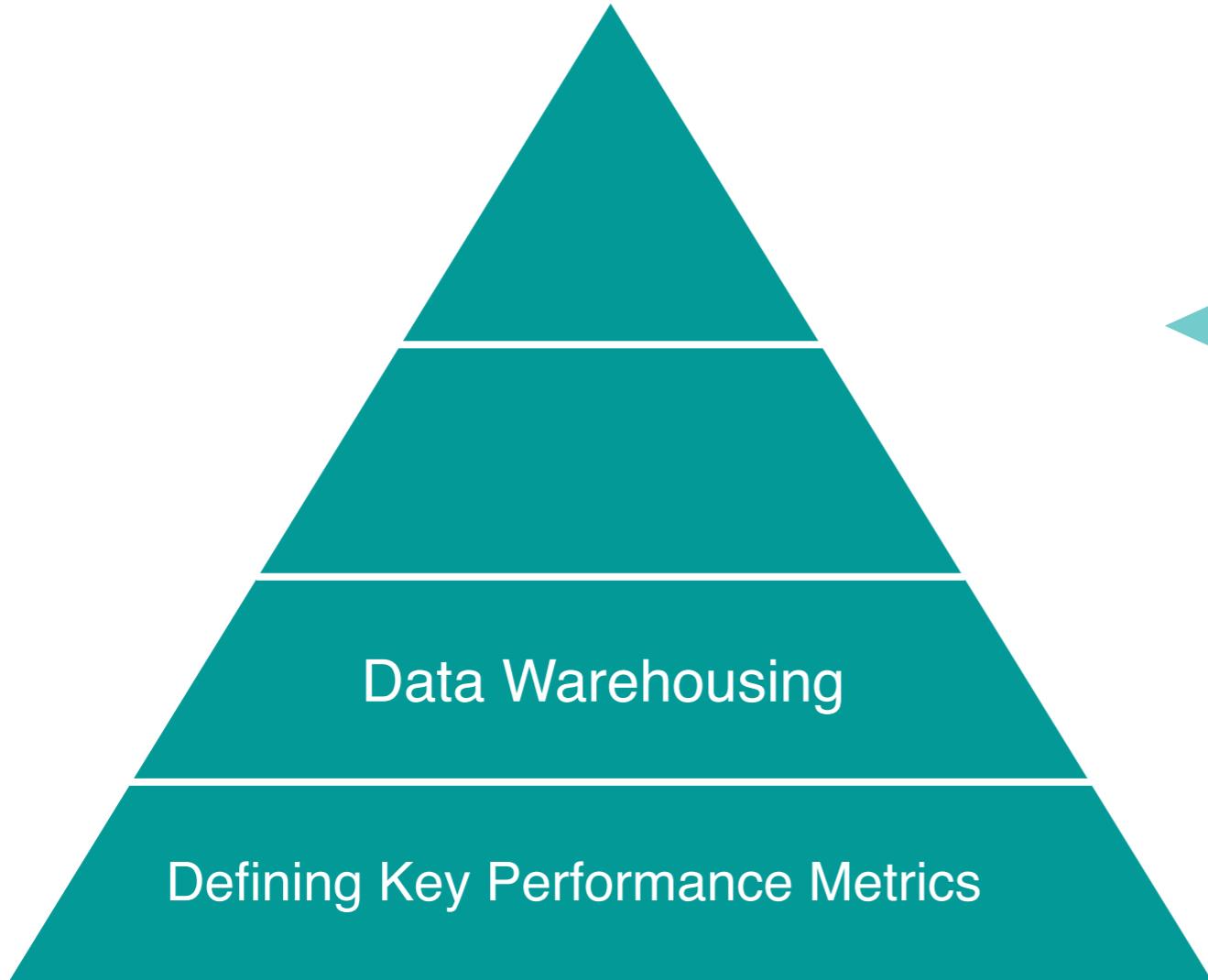
THE DATA PYRAMID

4





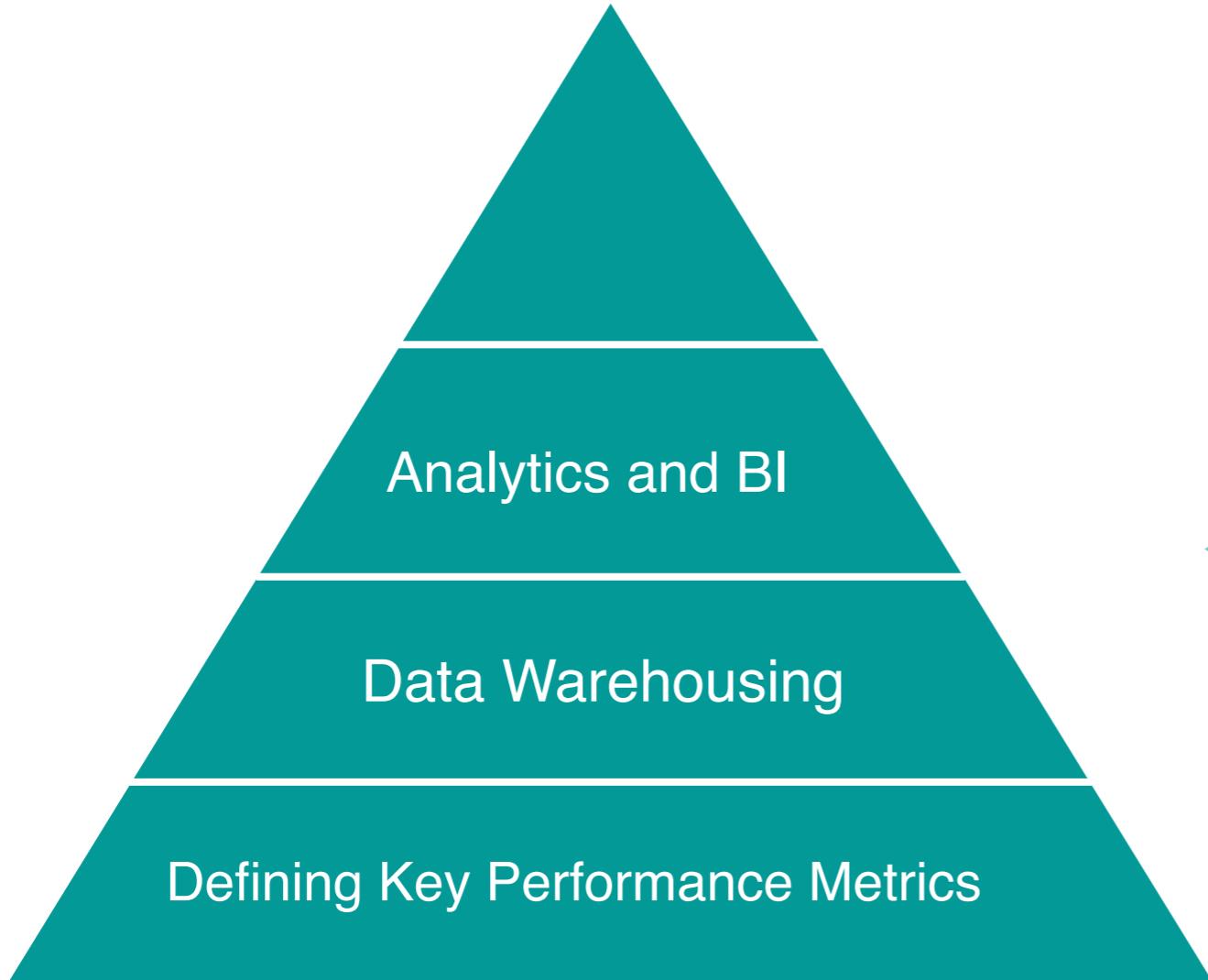
- Find out what makes a company successful
- Align entire company goals and roadmap around these KPIs
- Examples: Daily revenue, DAU/MAUs, Average Revenue Per User, Time spent on site, Site Uptime, etc.



- Taking production data from disparate sources and replicating it into a central database
- Schema is planned and arranged according to KPIs defined earlier
- Single source of truth for all analysis and queries (i.e., no one is using production)

THE DATA PYRAMID

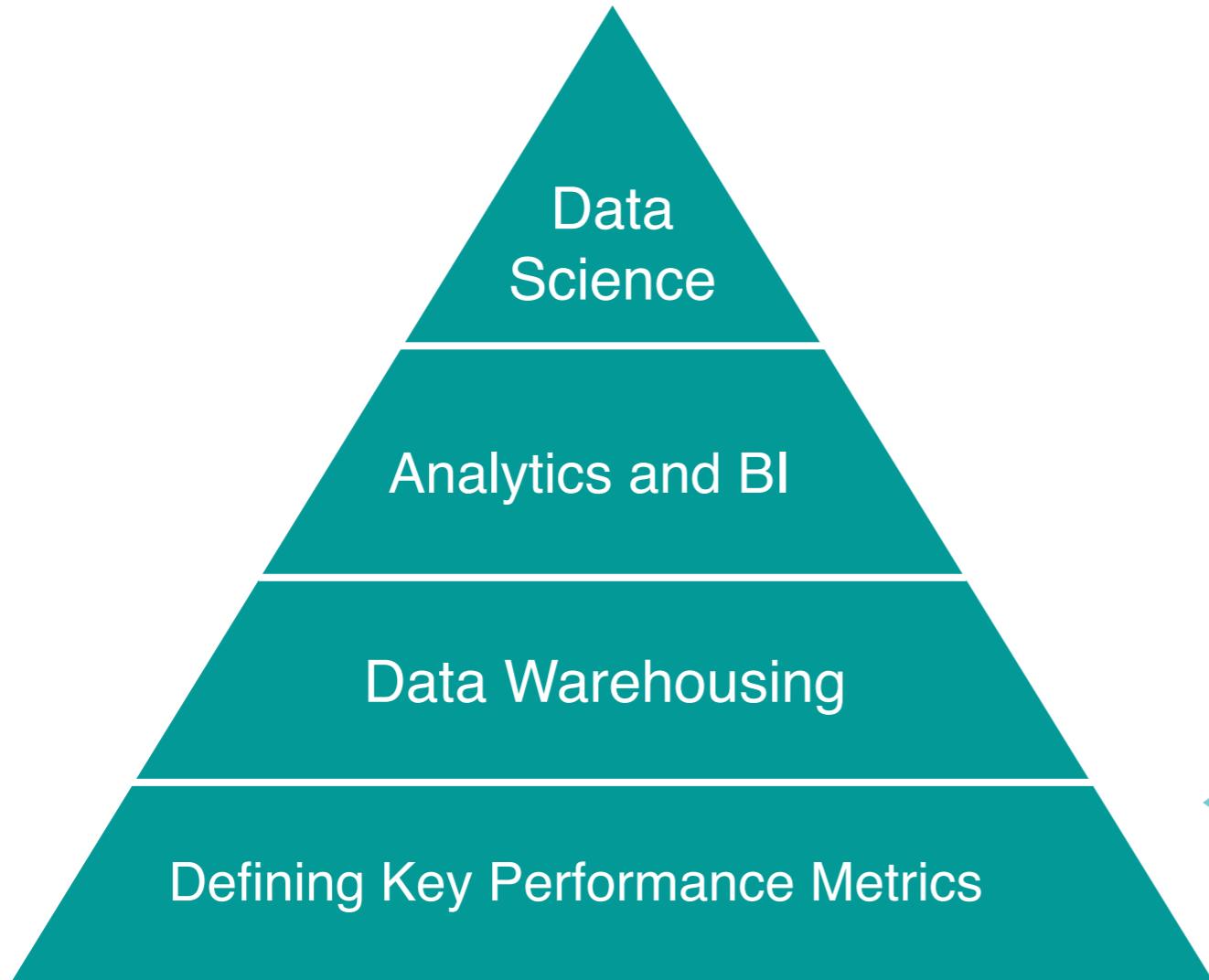
7



- From a data warehouse, we can create KPI reports for nontechnical stakeholders (decision makers, policy wonks, electorates)
- We can also analyze certain metrics and tables and make hypotheses about specific trends in our data.

THE DATA PYRAMID

8



- We can create features and variables from our hypotheses and using machine learning, we can write algorithms and models to:
 - Explain the past, or
 - Predict the future.
- These models can be user-facing (product features), or internal.

WHAT EVENTS SHOULD I INSTRUMENT?

Everything!

```
analytics.track("Loaded page",
    ("user": 5723,
     "page_name": "preferences");

analytics.track("Bought an item",
    ("user": 5723,
     "item_name": "Fire Phone",
     "cost": 399);
```



Dave McClure
500 Startups

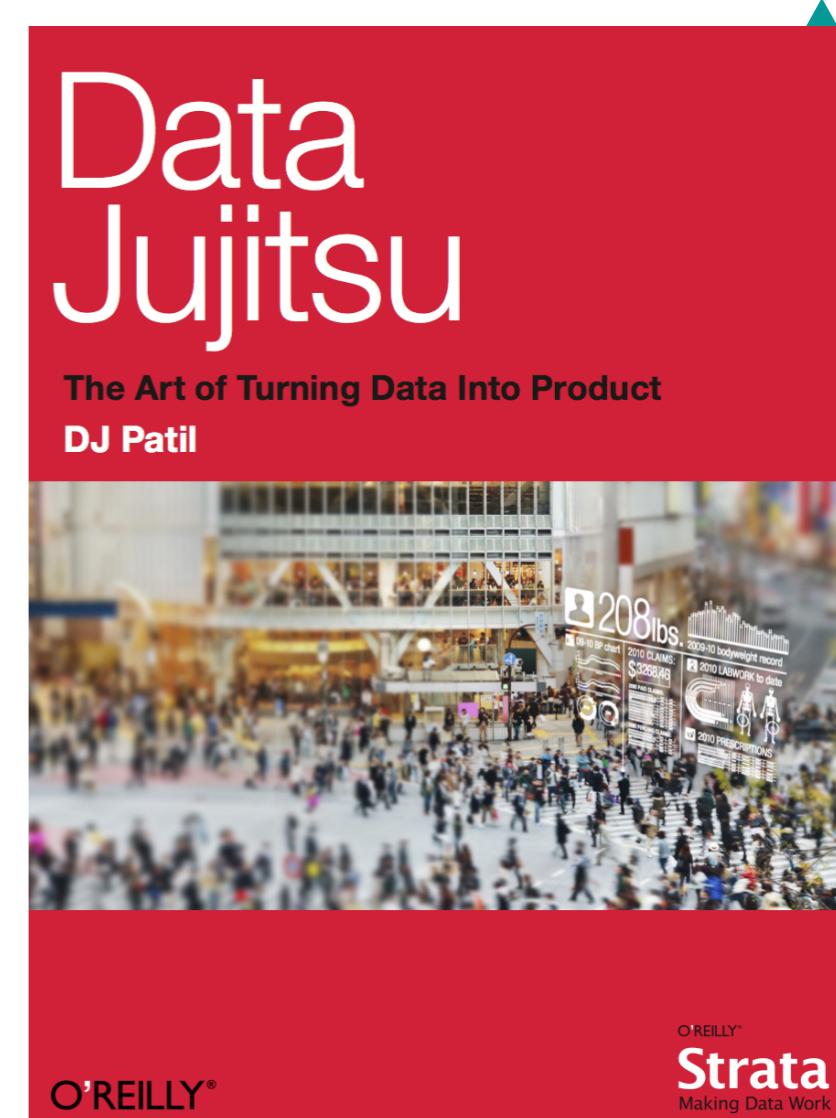
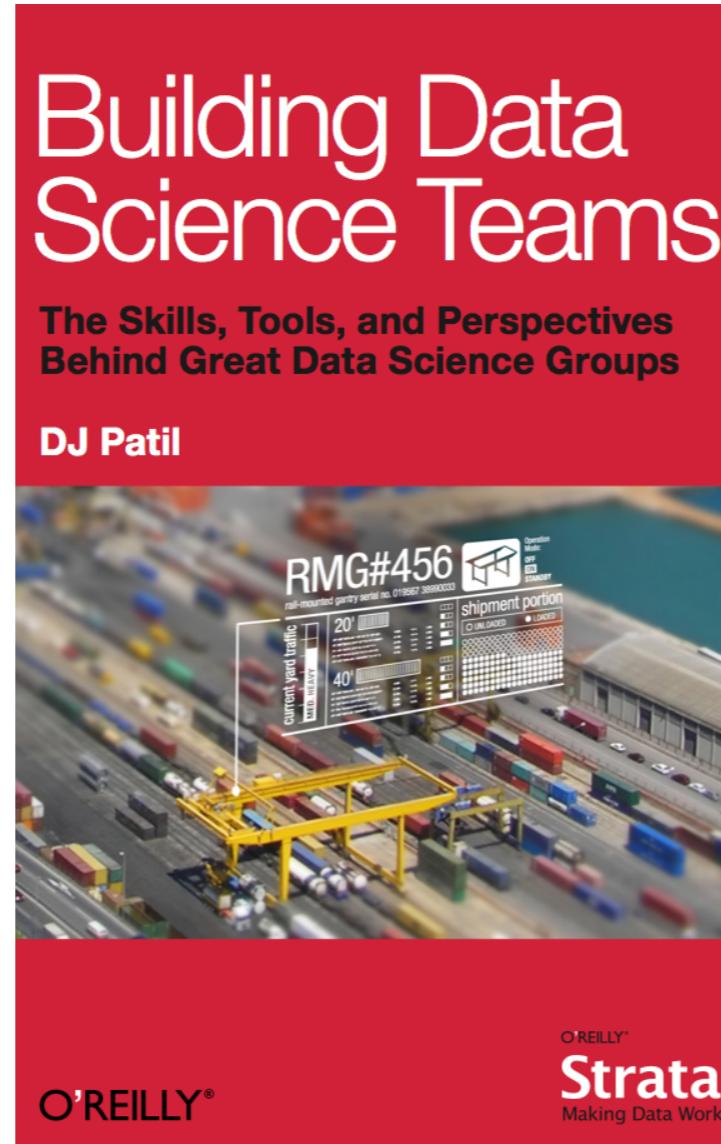
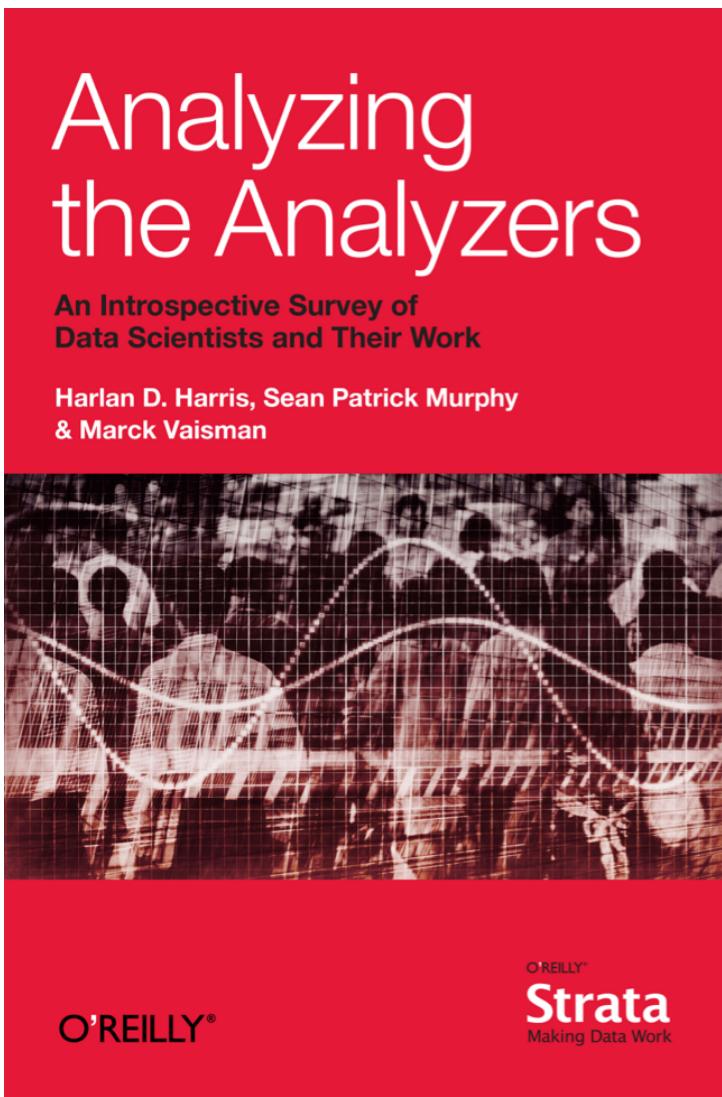
Example Conversion Metrics

Category	User Status	Conv %	Est. Value
Acquisition	Visit Site (or landing page, or external widget)	100%	\$.01
Acquisition	Doesn't Abandon (views 2+ pages, stays 10+ sec, 2+ clicks)	70%	\$.05
Activation	Happy 1 st Visit (views X pages, stays Y sec, Z clicks)	30%	\$.25
Activation	Email/Blog/RSS/Widget Signup (anything that could lead to repeat visit)	5%	\$1
Activation	Acct Signup (includes profile data)	2%	\$3
Retention	Email Open / RSS view → Clickthru	3%	\$2
Retention	Repeat Visitor (3+ visits in first 30 days)	2%	\$5
Referral	Refer 1+ users who visit site	2%	\$3
Referral	Refer 1+ users who activate	1%	\$10
Revenue	User generates minimum revenue	2%	\$5
Revenue	User generates break-even revenue	1%	\$25

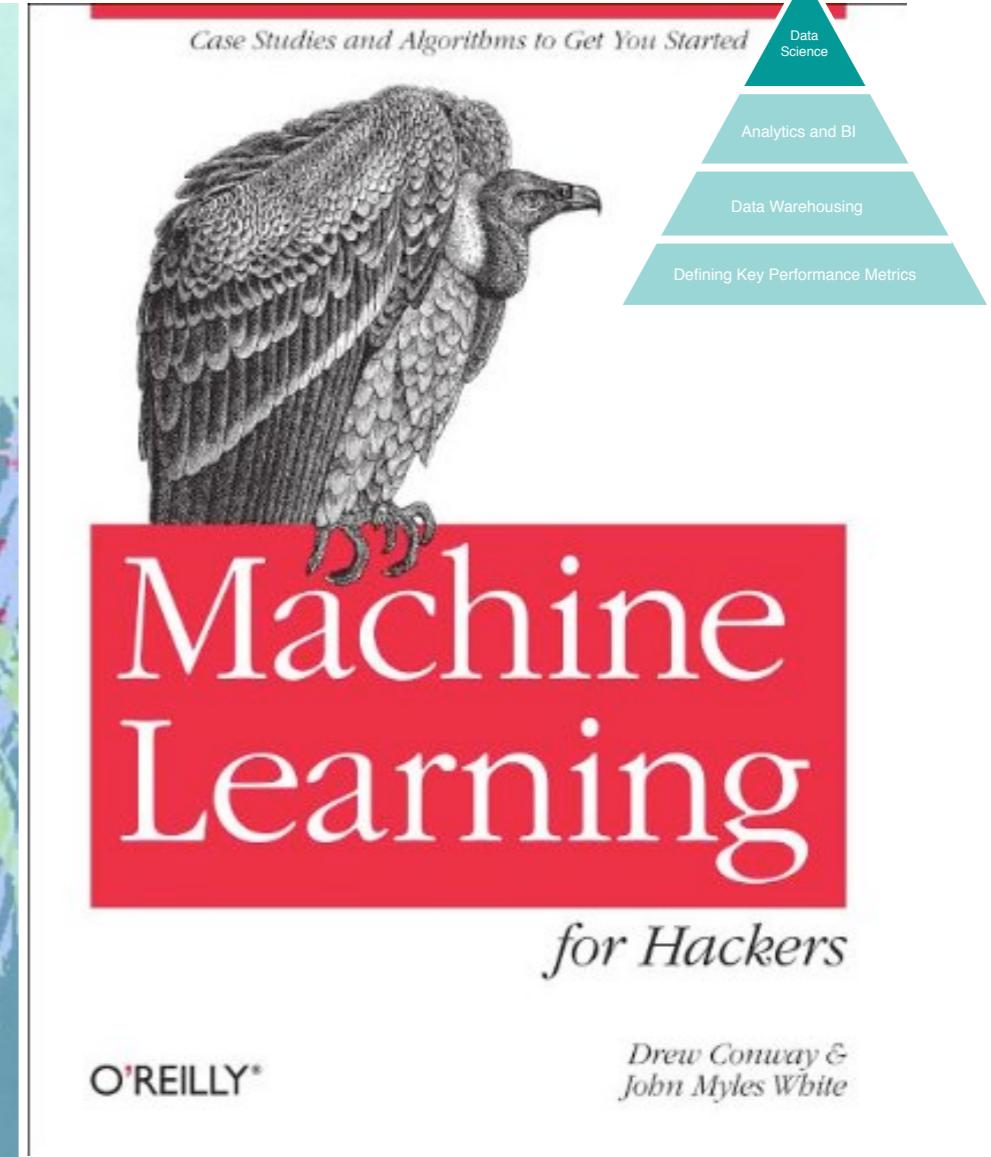
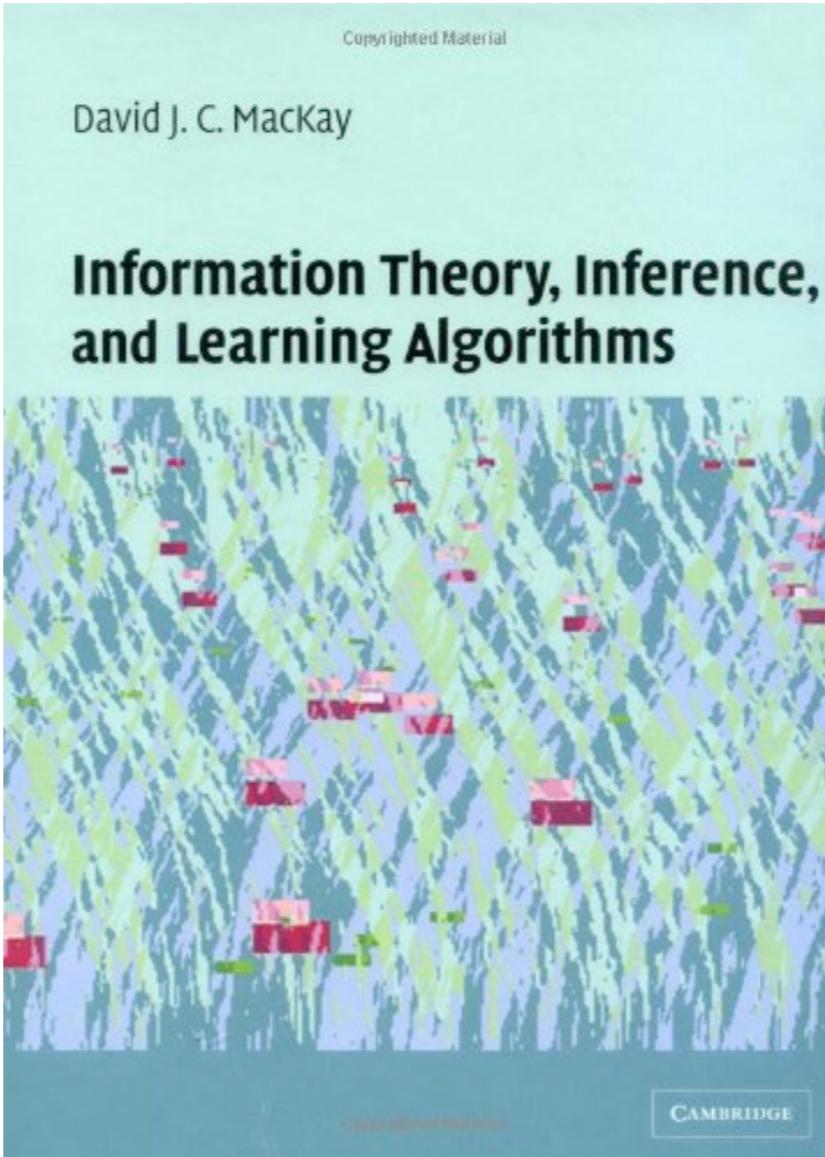
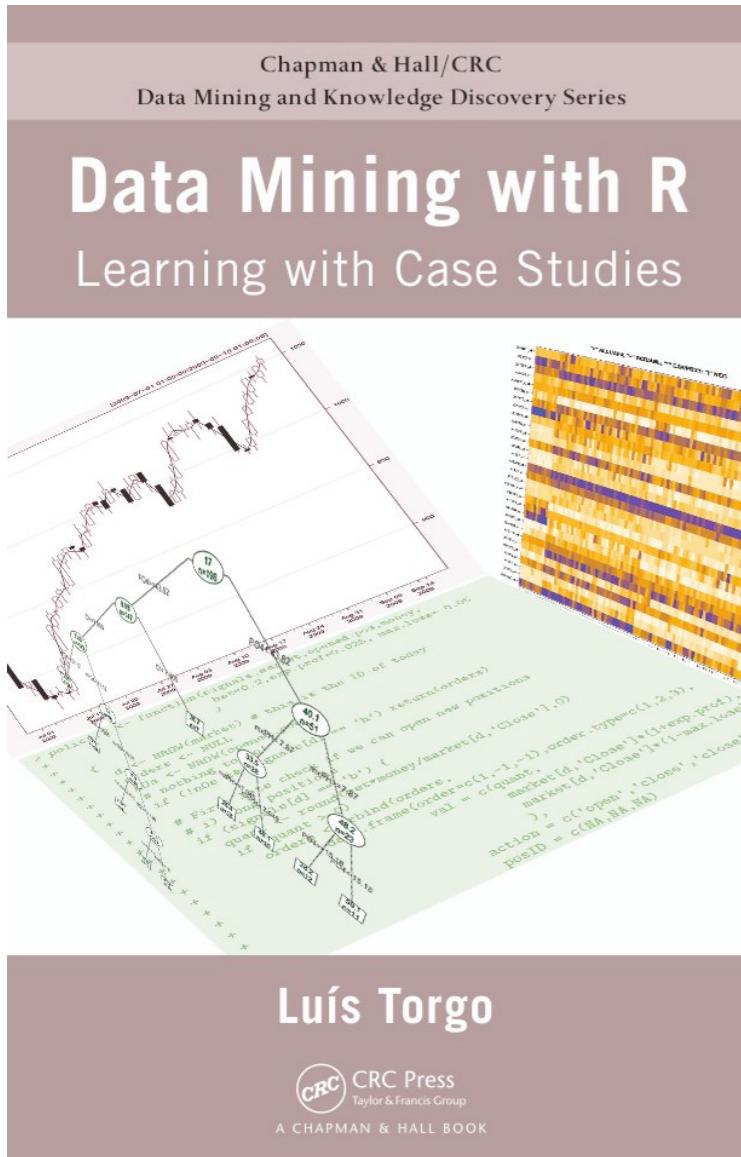
FURTHER RESOURCES

WHERE TO GO FROM HERE

FURTHER RESOURCES - BOOKS



FURTHER RESOURCES - BOOKS



THANKS!

E-MAIL ME QUESTIONS: THOMSON@CANTAB.NET

OR SEND ME A TWEET: @ITSTHOMSON

WE'RE HIRING! [HTTP://WWW.FRAMED.IO](http://WWW.FRAMED.IO)