



galvanize

# Building Data Science Teams: Lessons from Unicorn Hunting in the Wild

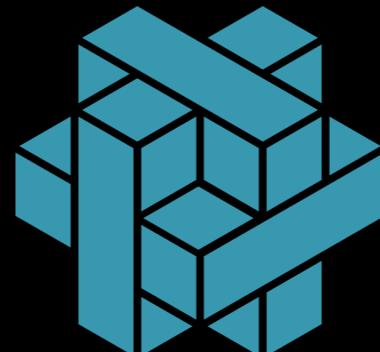
@galvanize

Sept 29, 2015

Hi, I'm Katie Kent  
@k80kent



Director, Educational Outcomes  
@galvanize



Director of Partnerships  
Zipfian Academy

# LEARNING OBJECTIVES

- Why it's hard to find data scientists
- Know your Data Scientist subtypes.
- Know how to bait them
- Where to set your traps

# THE STORY OF THE DATA SCIENCE UNICORN

## (ABRIDGED VERSION)



**what:**

**data scientist**

job title, keywords or company

**where:**

**San Francisco, CA**

city, state, or zip

## **data scientist jobs in San Francisco, CA**

### My recent searches

[data scientist - san francisco, ca - 113 new](#)

[data scientist - Houston, TX - 15 new](#)

[neo4j data - San Francisco, CA - 5 new](#)

[Magento Web Developer - 36 new](#)

[developer javascript - boulder, CO - 30 new](#)

[developer javascript rails - boulder, CO - 5 new](#)

[developer javascript rails - Longmont, CO - 3 new](#)

[software developer - Longmont, CO - 37 new](#)

[data science mathematics phd - San Francisco, CA - 7 new](#)

**New! Join Indeed Prime** - Get offers from great tech companies

Jobs 1 to 10 of 2,138

### **Data Engineer - Large-Scale 1TB w/ Hive/Pig/HDFS**

7,800 jobs?

**Stepping Up Solutions - San Francisco, CA**

Work closely with Product Management, Data Analysts and Data Scientists to ensure that implemented solutions are built as desired....

**Easily apply**

30+ days ago - [email](#)

**Sponsored by Stepping Up Solutions**

### **Head of Data Sciences**

**Credit Karma - San Francisco, CA**

Hire, coach and motivate a team of data scientists to grow their skills and careers. Strong leadership skills with 5+ years of experience hiring and managing...

30+ days ago - [email](#)

**Sponsored by Credit Karma**

### **Staff Data Scientist - Personalization**

**Chegg** 17 reviews - Santa Clara, CA

The Staff Data Scientist focusing on Personalization will use data mining and machine learning expertise to define, design and develop personalization pipeline...

# A Perfect Storm



# Storage is cheap.

Year	Capacity (GB)	Cost per GB (USD)
1992	0.08	\$3,827.20
1997	2.1	\$157.00
2002	80	\$3.74
2007	750	\$0.35
2012	3,000	\$0.05

# Data is valuable.

## SURGE PRICING

Demand is off the charts! Rates have increased to get more Ubers on the road.

7.75X

THE NORMAL RATE

\$93 MINIMUM FARE

+\$5.81 / MIN + \$23.25 / MILE

NETFLIX Watch Instantly • Just for Kids • Taste Profile • DVDs

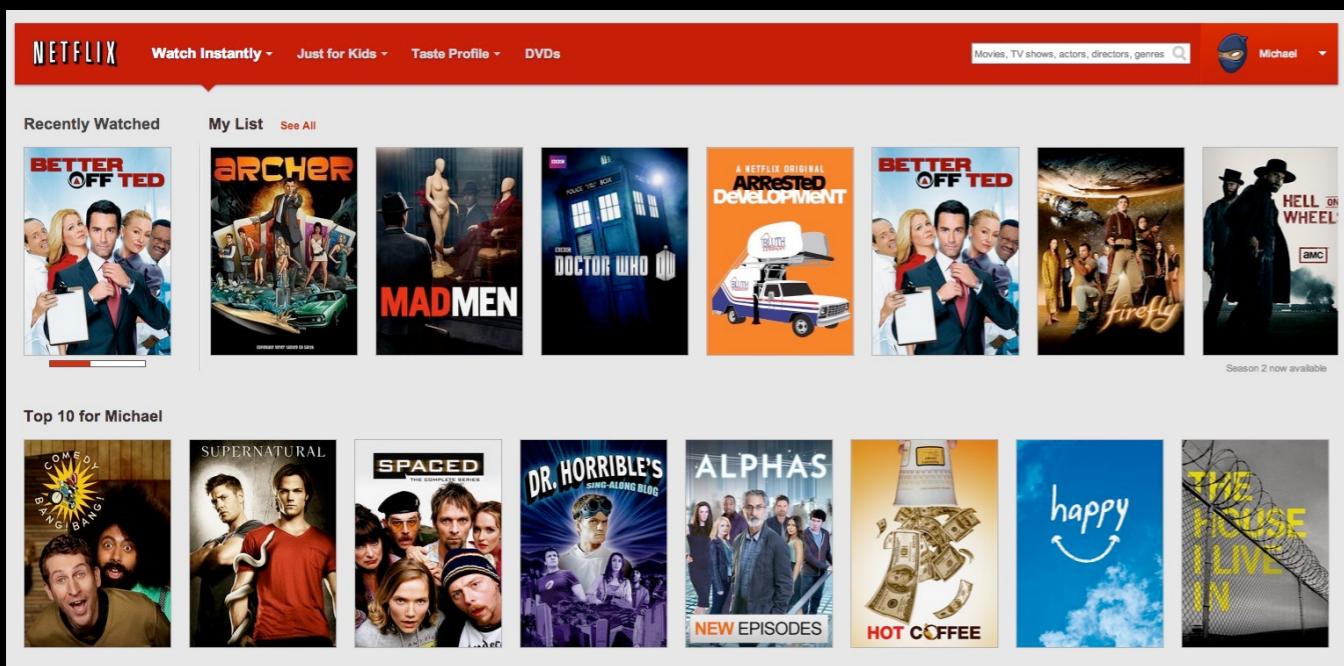
Recently Watched My List See All

BETTER OFF TED ARCHER MADMEN ARRESTED DEVELOPMENT DOCTOR WHO BETTER OFF TED FIREFLY HELL ON WHEEL

Top 10 for Michael

COMEDY BANG! BANG! SUPERNATURAL SPACED DR. HORRIBLE'S SING-ALONG BLOG ALPHAS HOT COFFEE THE HOUSE I LIVE IN

Season 2 now available



## People You May Know beta

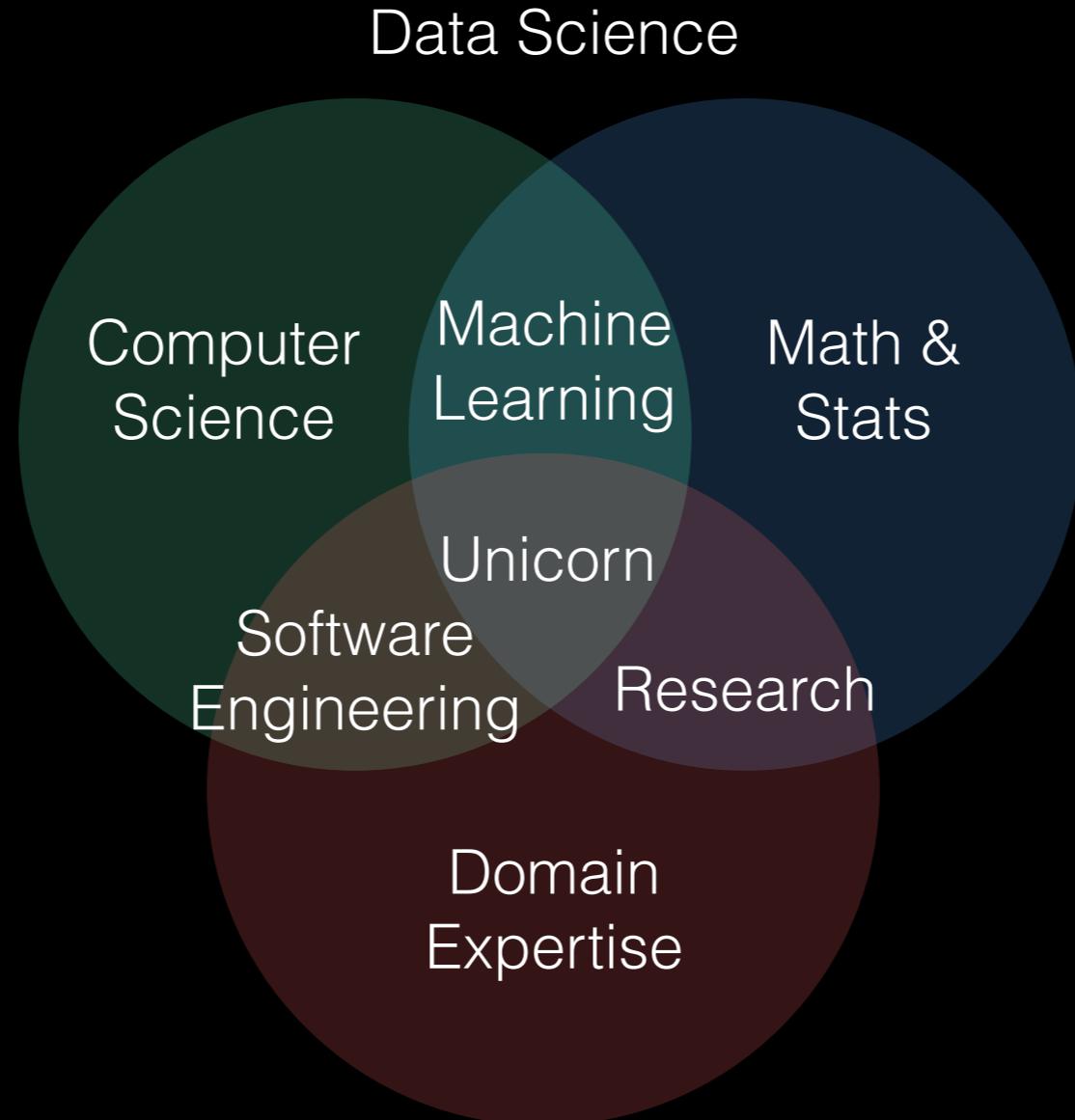
See people from different parts of your professional life



All Suggestions / LinkedIn (12)

 <a href="#">Connect</a>	Brad Mauney <small>2nd</small> Senior Product Manager, Search & Social Graph at LinkedIn Mountain View, California
 <a href="#">Connect</a>	Sam Shah <small>2nd</small> Principal Engineer at LinkedIn Mountain View, California
 <a href="#">Connect</a>	Vinodh Jayaram <small>2nd</small> Software Engineering Manager at LinkedIn Mountain View, California
 <a href="#">Connect</a>	Andy Chen <small>2nd</small> Software Engineer at LinkedIn Mountain View, California
 <a href="#">Connect</a>	Albert Wang <small>2nd</small> Senior User Experience Designer at LinkedIn Mountain View, California
 <a href="#">Connect</a>	Tan Nhu <small>2nd</small> Senior Web Developer at LinkedIn Mountain View, California
 <a href="#">Connect</a>	10 shared connections
 <a href="#">Connect</a>	78 shared connections

# Data Scientists are a rare breed.



# Data Scientists are a rare breed.



SUPERB\*  
wallpapers

Traditional educational institutions  
struggle to teach technology, esp. the  
interdisciplinary practice of data science.

# LESSONS FROM UNICORN HUNTING IN THE WILD

# The Professional Unicorn Hunters Club

**facebook.**

**Linkedin**

**NETFLIX**

 **yelp**

 **galvanize**

 **airbnb**

**Google**



**U B E R**

**STITCH FIX**



**Etsy**

**IBM**

 **trulia**

1

Learn the Data Scientist  
sub-types.

\*and how they work together.

“Data scientist” means different things to different people.

# The Data Science Skills Map

Translation — Product — Complex Analysis — Software Engineering

# The Data Science Skills Map

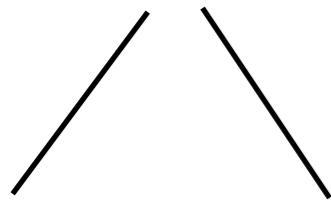
Translation — Product — Complex Analysis — Software Engineering



Visualization,  
Communication,  
& Evangelism

# The Data Science Skills Map

Translation — Product — Complex Analysis — Software Engineering



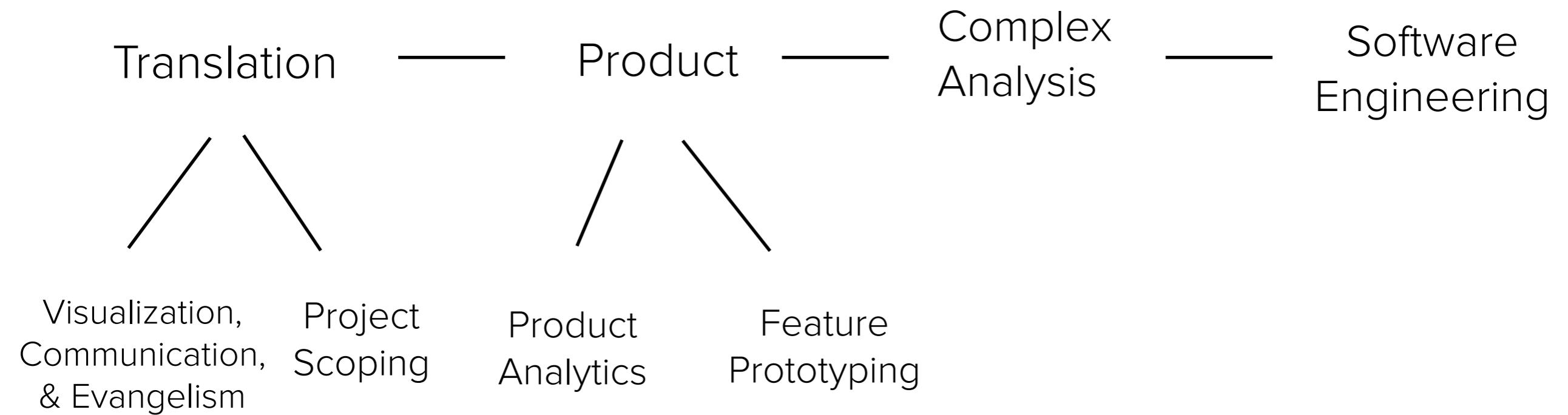
Visualization,  
Communication,  
& Evangelism

Project  
Scoping

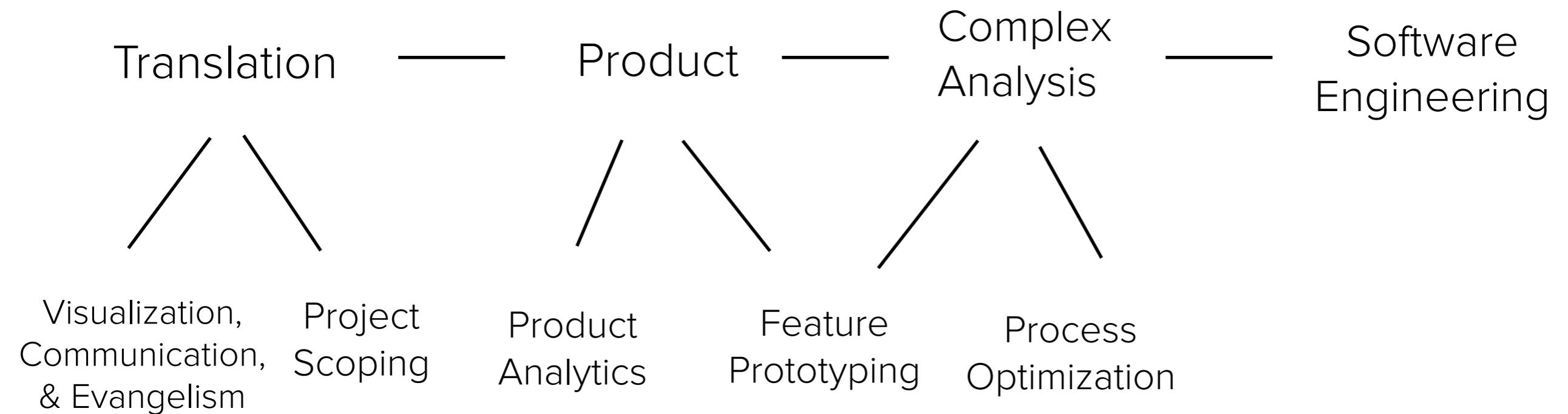
# The Data Science Skills Map



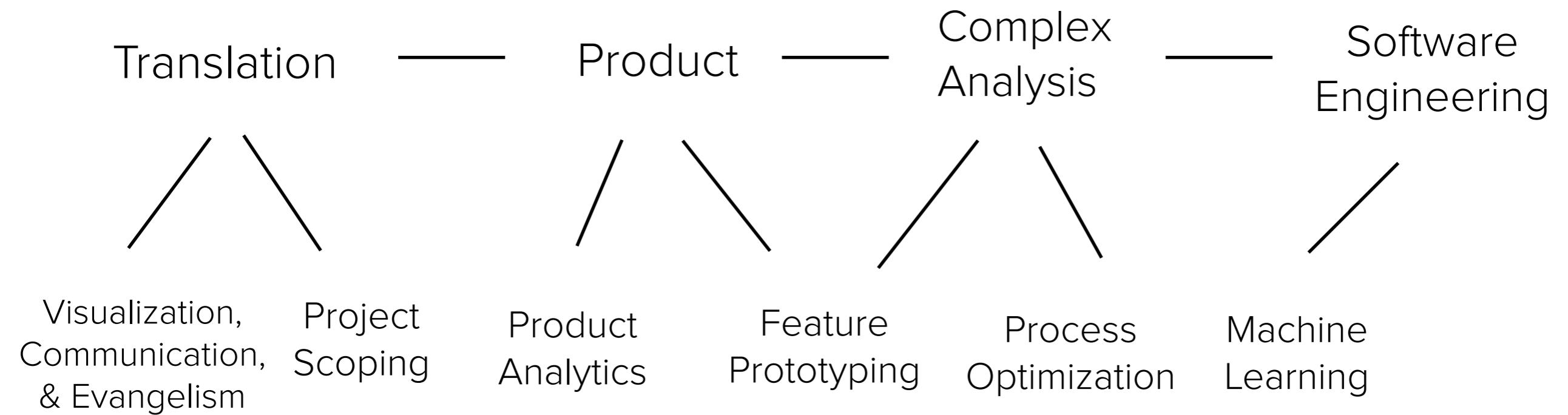
# The Data Science Skills Map



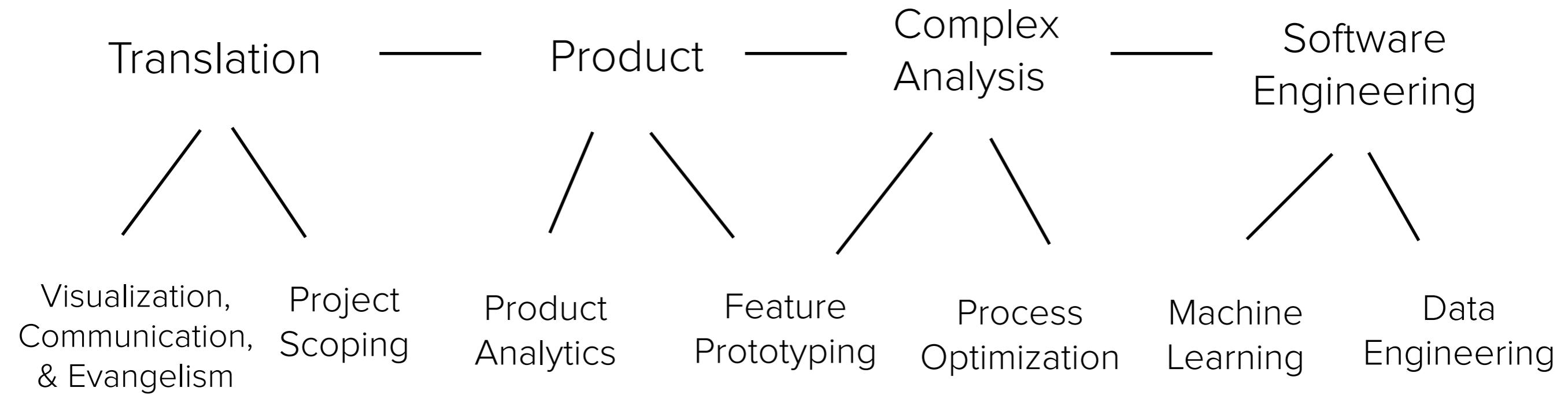
# The Data Science Skills Map



# The Data Science Skills Map

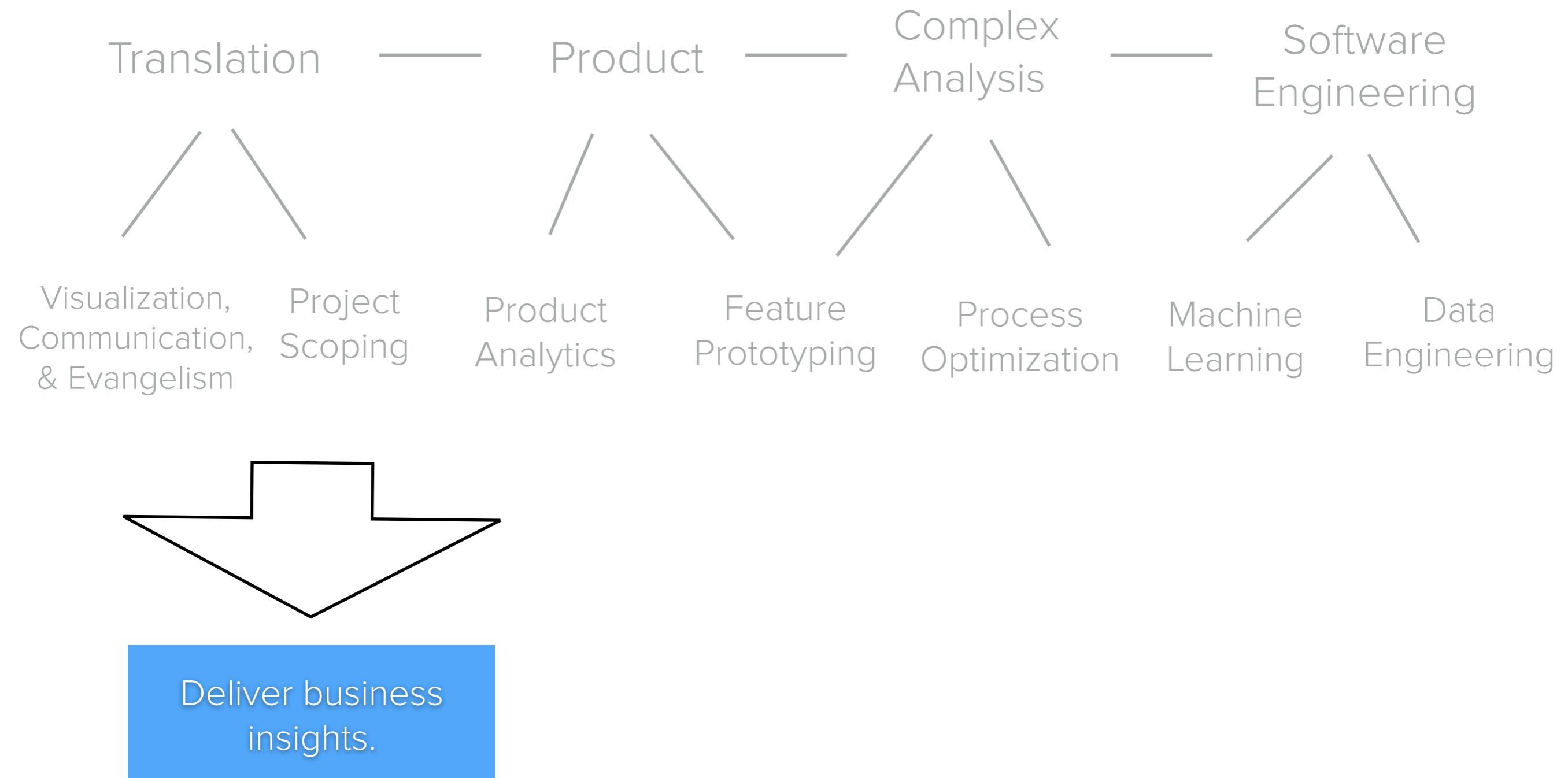


# The Data Science Skills Map

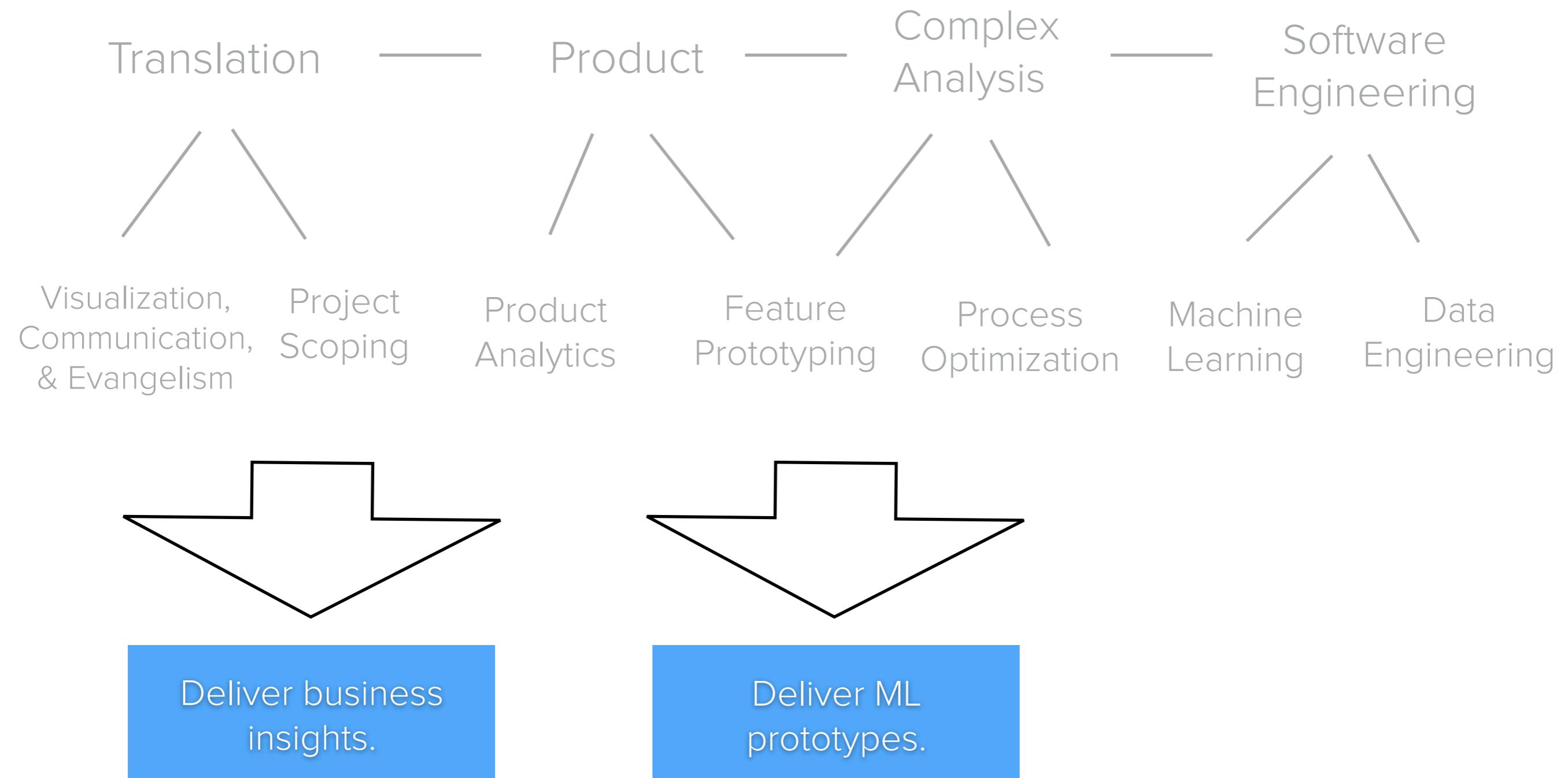


The skills required for a specific role depend on the ultimate deliverables.

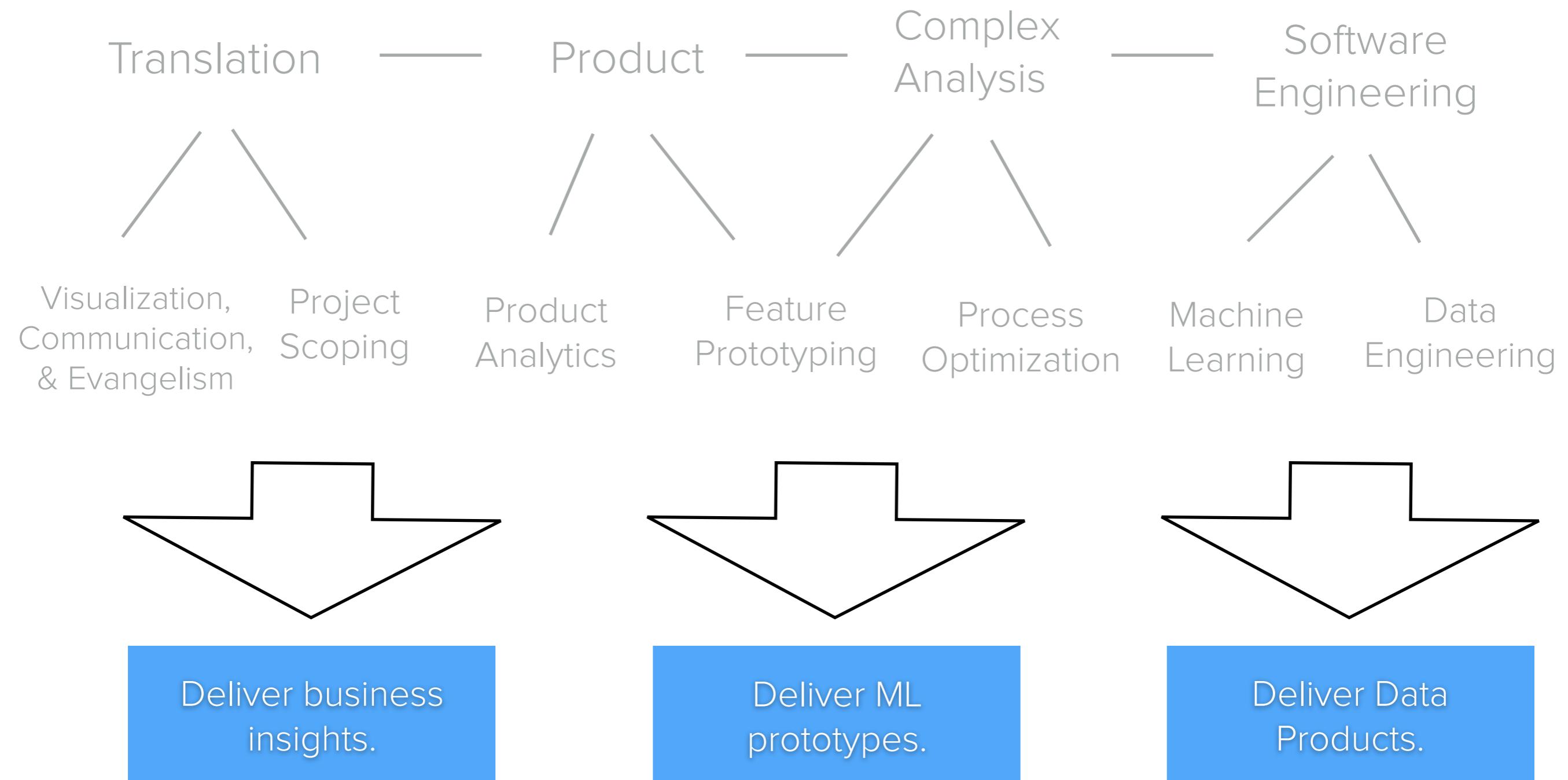
# Different skills = different deliverables.



# Different skills = different deliverables.



# Different skills = different deliverables.



At scale, each of these are their own  
data science team.

Translation

Product

Complex  
Analysis

Software  
Engineering

Visualization,  
Communication,  
& Evangelism

Project  
Scoping

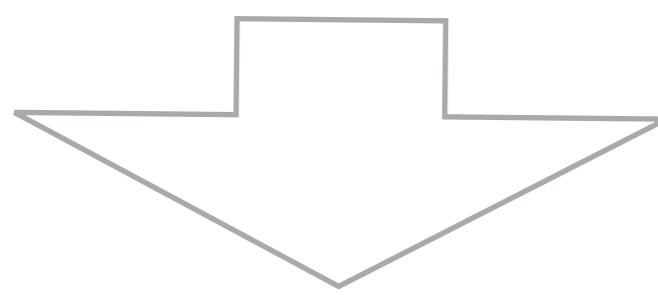
Product  
Analytics

Feature  
Prototyping

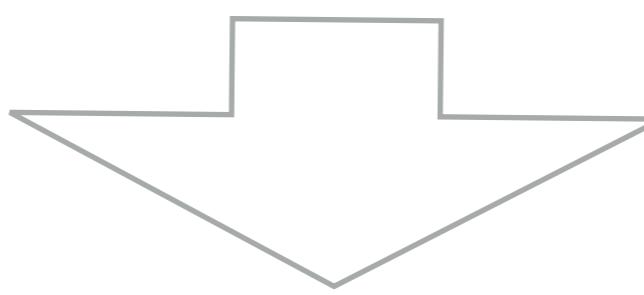
Process  
Optimization

Machine  
Learning

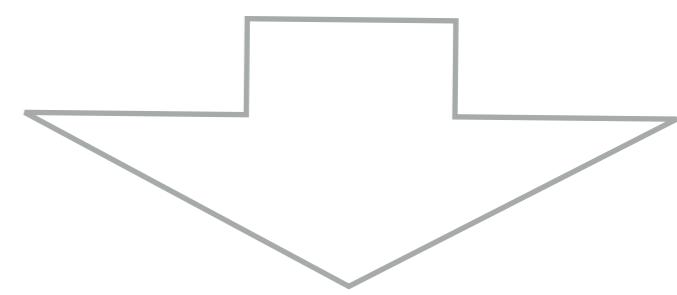
Data  
Engineering



Deliver business  
insights.



Deliver ML  
prototypes.



Deliver Data  
Products.

Insights Team

Translation

Product

Complex  
Analysis

Software  
Engineering

Visualization,  
Communication,  
& Evangelism

Project  
Scoping

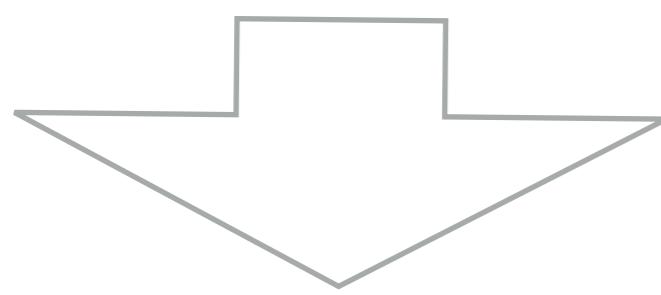
Product  
Analytics

Feature  
Prototyping

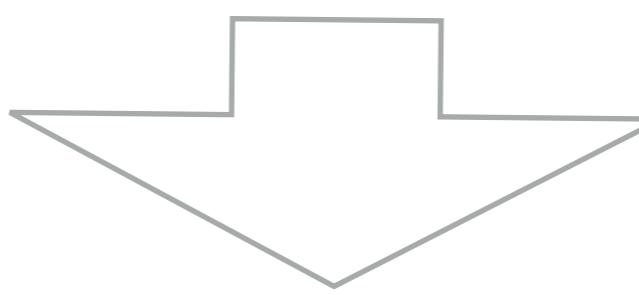
Process  
Optimization

Machine  
Learning

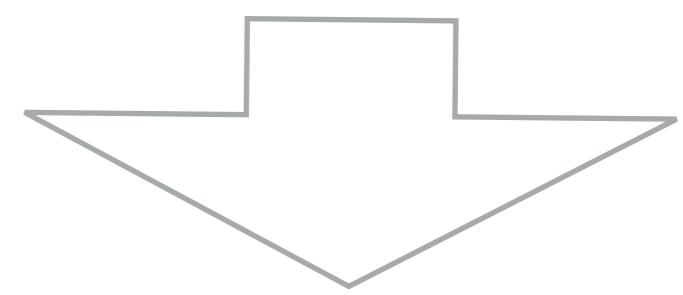
Data  
Engineering



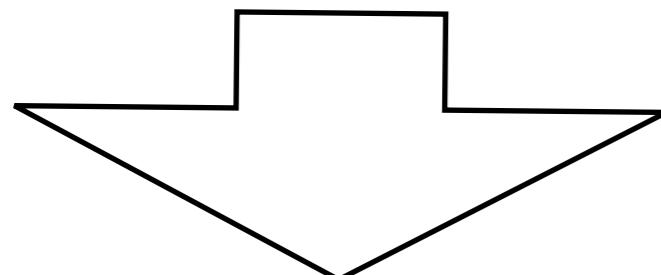
Deliver business  
insights.



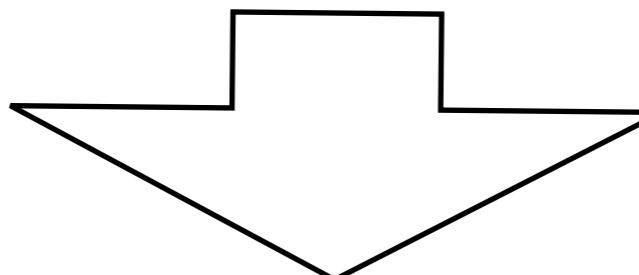
Deliver ML  
prototypes.



Deliver Data  
Products.



Insights Team



Product Data Science  
Team

Translation

Product

Complex  
Analysis

Software  
Engineering

Visualization,  
Communication,  
& Evangelism

Project  
Scoping

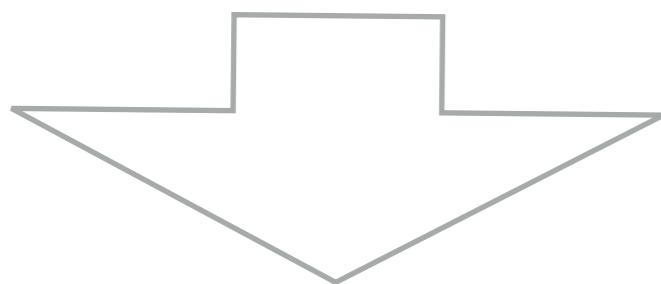
Product  
Analytics

Feature  
Prototyping

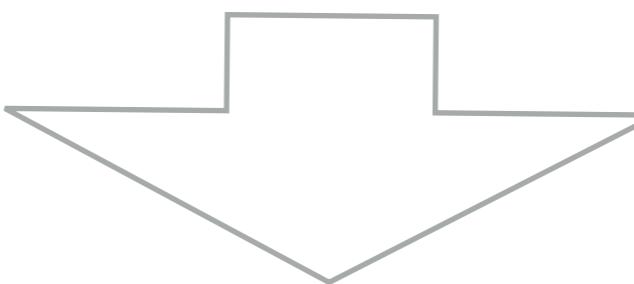
Process  
Optimization

Machine  
Learning

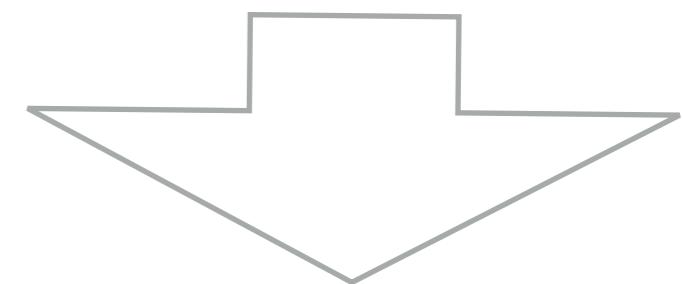
Data  
Engineering



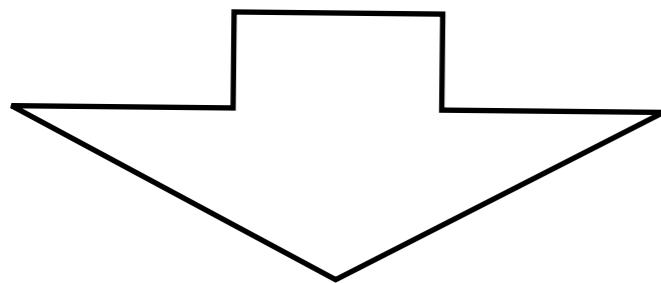
Deliver business  
insights.



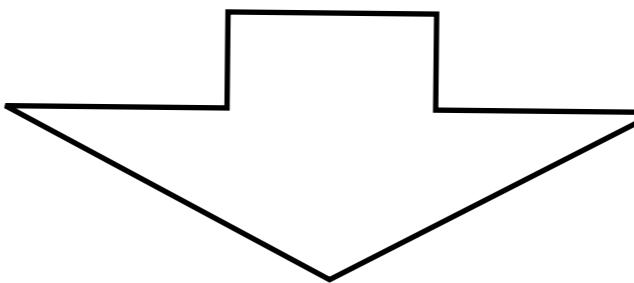
Deliver ML  
prototypes.



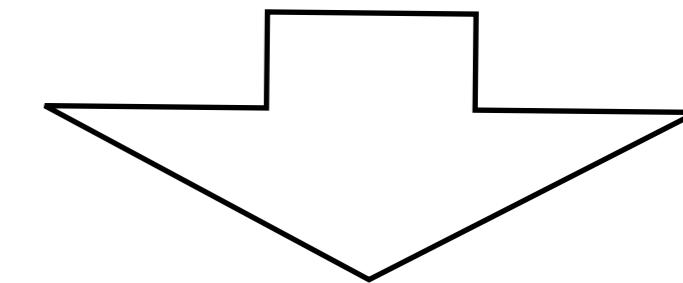
Deliver Data  
Products.



Insights Team



Product Data Science  
Team



Data Engineering  
Team

Target your species before you go  
hunting.

2

How to Bait & Snare Them

Great data scientists often have 3 simultaneous offers, so salesmanship is important.

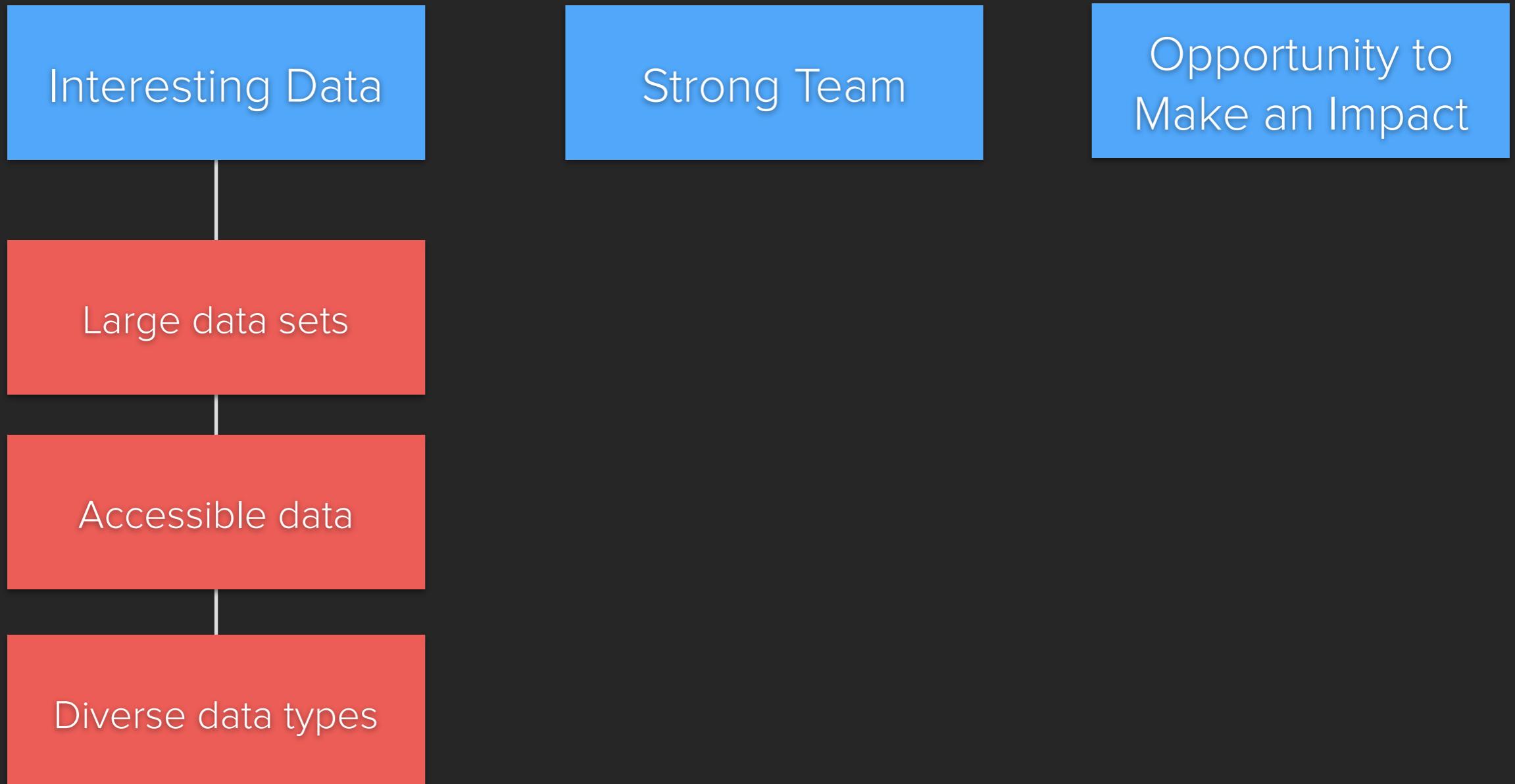
# What (Great) Data Scientists Want

Interesting Data

Strong Team

Opportunity to  
Make an Impact

# What (Great) Data Scientists Want



Sell the data you have. Acquire what  
you don't.

# How to talk about your data set

SITUATION

Interesting Data

ACTION

Sell that!

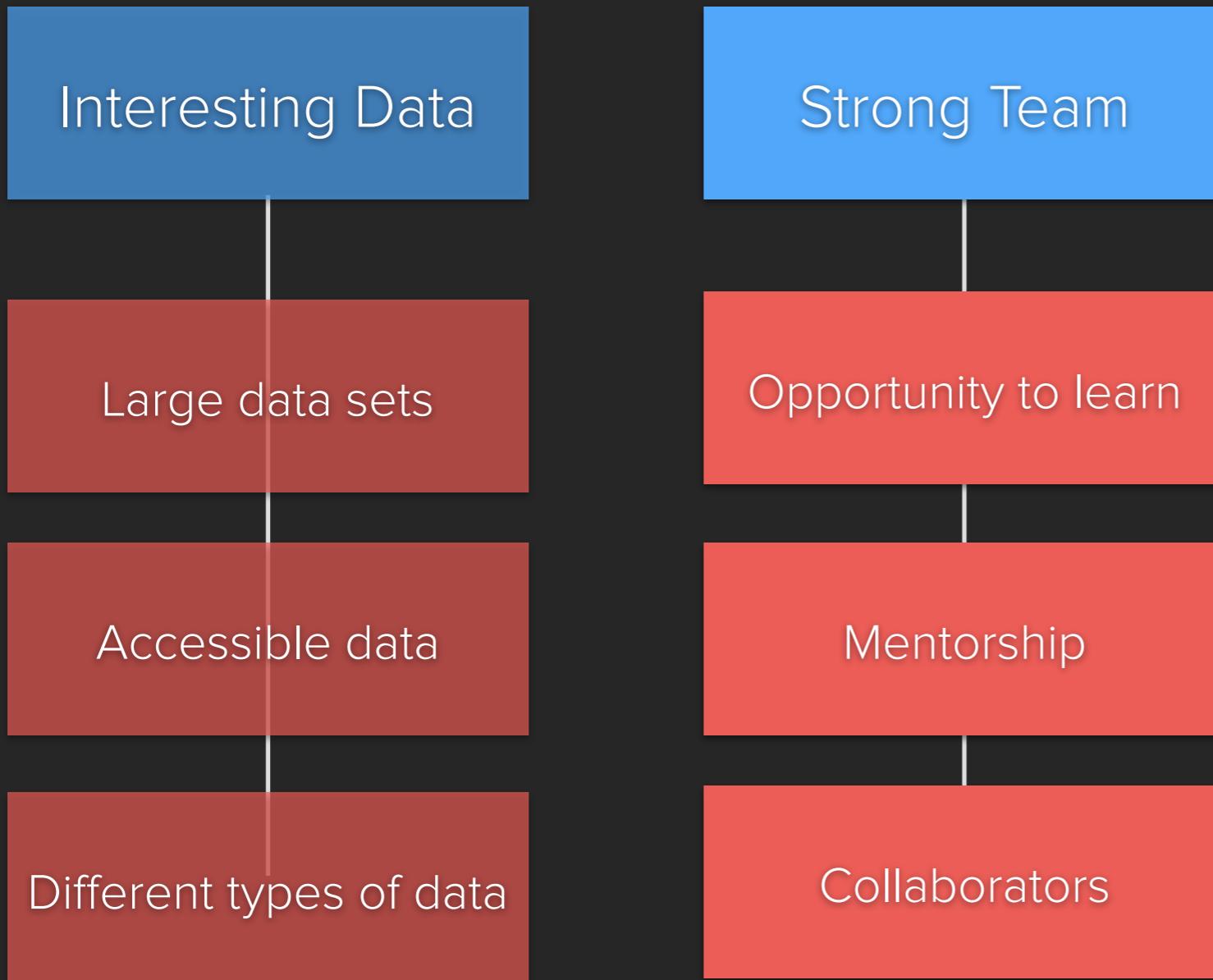
Small Data

Auxiliary sources

Inaccessible data

Fix infrastructure  
first.

# What (Great) Data Scientists Want



Sell what you have on the team. Build  
what you don't.

# How to talk about the team

## SITUATION

Awesome team.

## ACTION

Sell that!

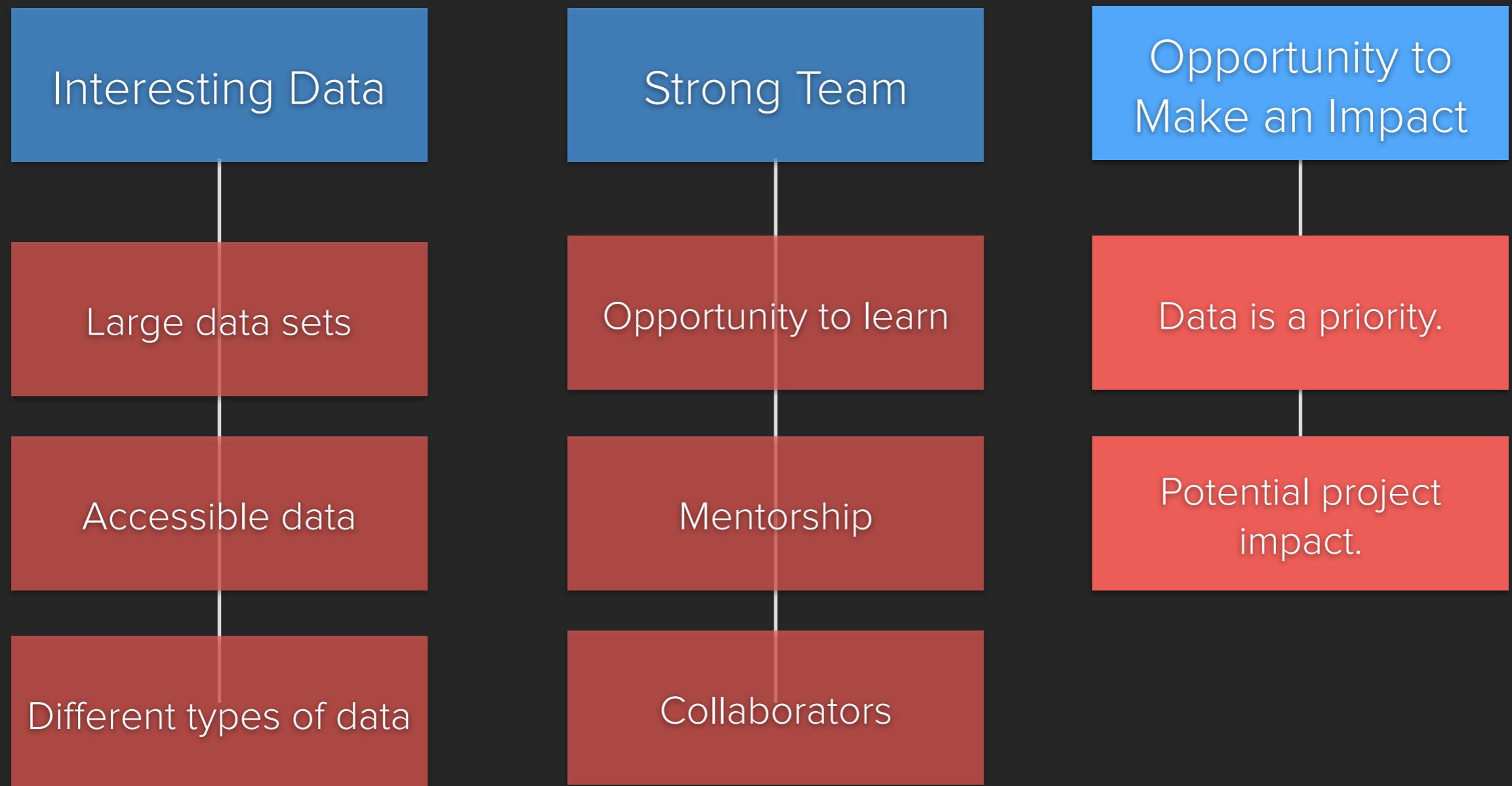
No mentorship.

Training budget

No daily  
collaborators.

Designate partners.

# What (Great) Data Scientists Want



# How to demonstrate impact opportunity

## SITUATION

Data not a priority.

## ACTION

Hire an evangelist.

Low-hanging fruit  
are picked.

>

Seek a researcher  
to push boundaries.

Chip at the hiring stakeholder  
knowledge gap.

3

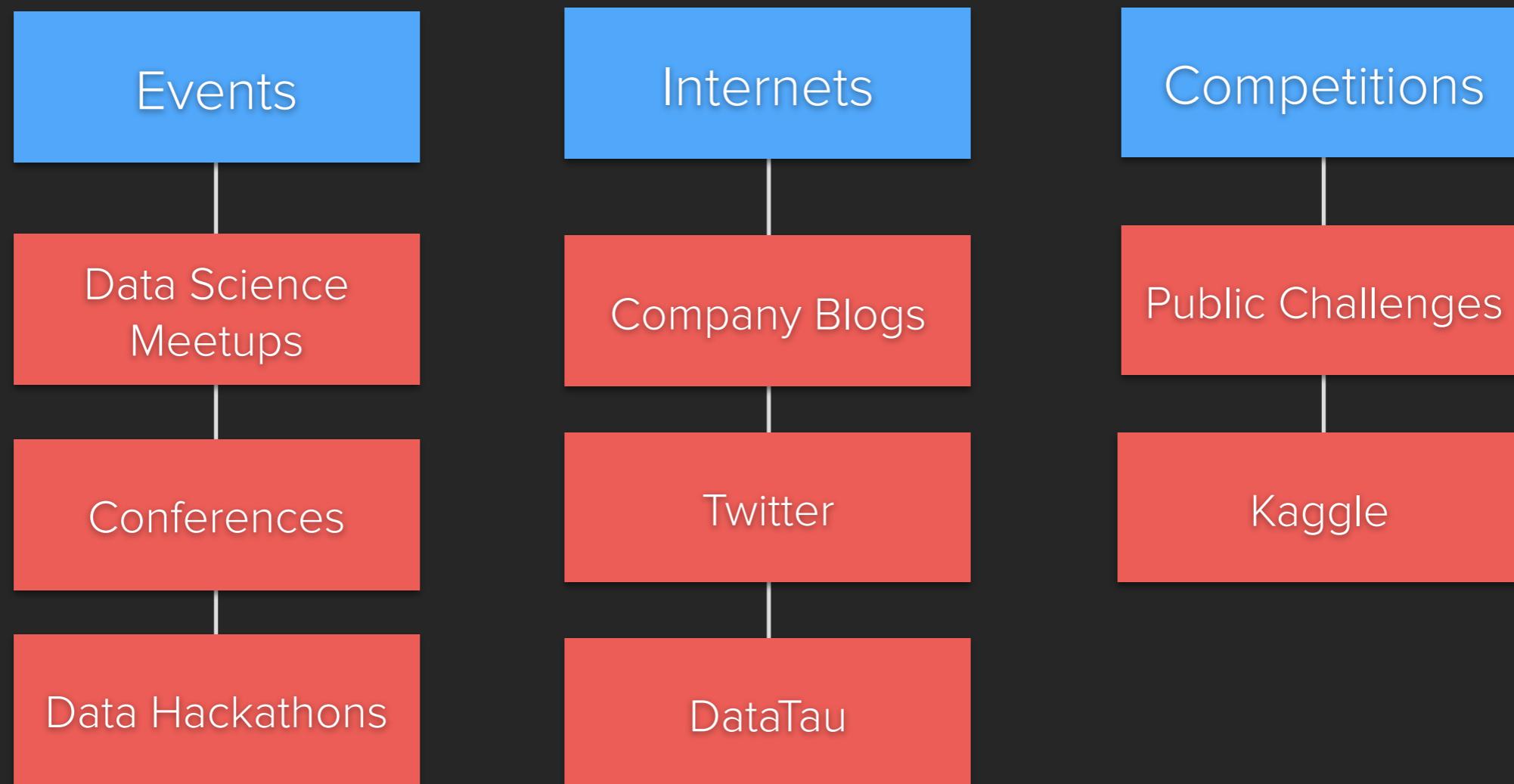
# Build a Talent Community

Savvy data science teams share content about their work in places where data scientists hang out.

‘I get about 300 applications for every posting.’

- Todd Holloway, Data Science Lead at Trulia for 4 years  
who runs the SF Data Mining Meetup

# Where (Great) Data Scientists Hang Out



# Where (Great) Data Scientists Hang Out



# Speak at a Data Science Meetup

NYC Data Science

NYC Open Data

Economics & Big  
Data

NYC Women in ML  
& DS

NYC Big Data  
Science

Data Skeptics

<http://www.quora.com/What-are-the-best-data-science-big-data-meetups-in-the-San-Francisco-Bay-Area>



Links and  
summaries!

# Speak or Sponsor Data Science Conferences

O'Reilly Strata

MLConf

PyData

OSDC

Predictive  
Analytics World

DataWeek

<http://www.kdnuggets.com/meetings/>

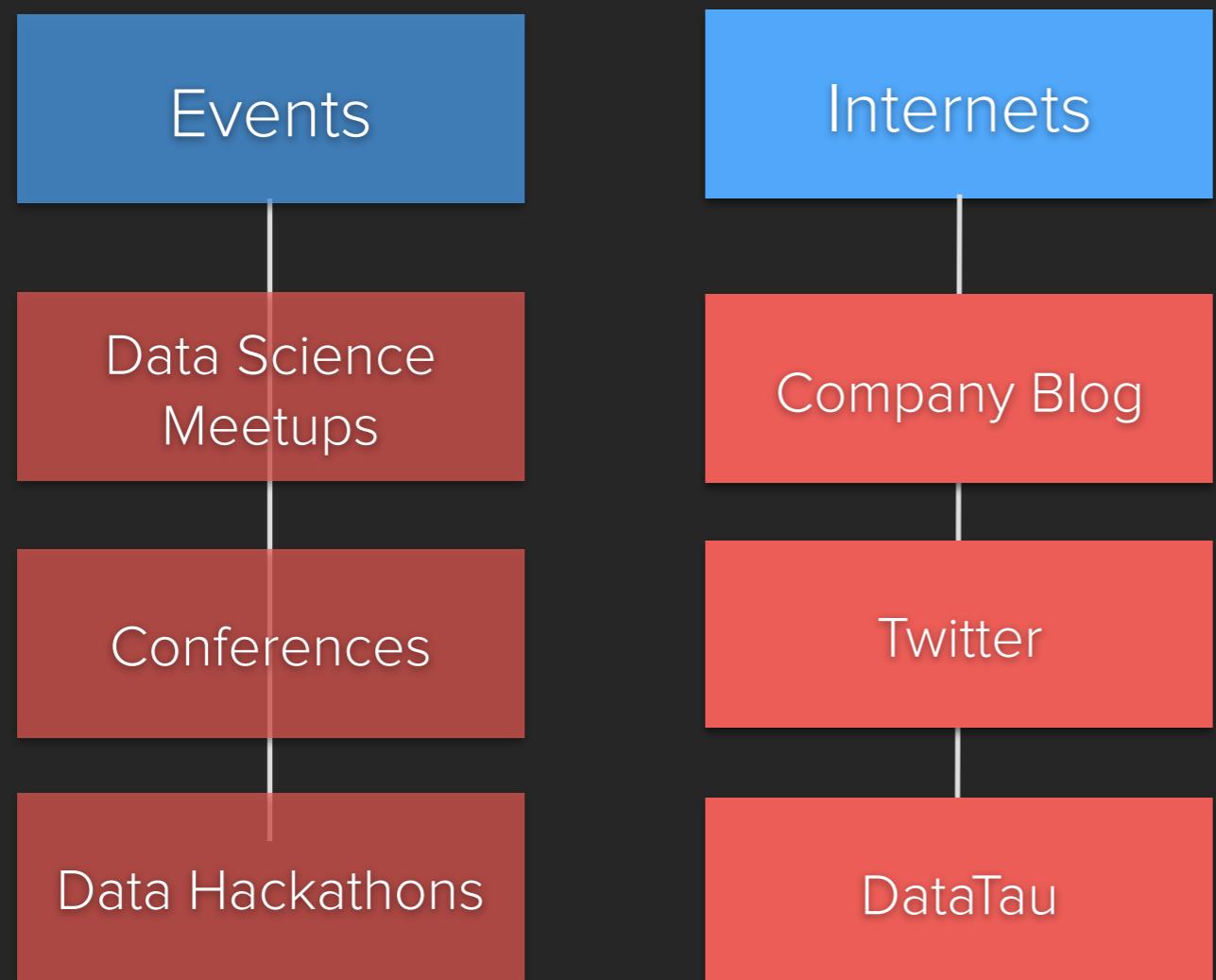
# Sponsor a Data Hackathon

DataKind

Bayes Impact

PyData

# Where (Great) Data Scientists Hang Out



# Make a Tech Blog Like These

Airbnb Nerds

Stitchfix  
Technology

Netflix Technology

- [nerds.airbnb.com/data/](http://nerds.airbnb.com/data/)
- [http://technology.stitchfix.com/blog/  
2015/03/31/advice-for-data-scientists/](http://technology.stitchfix.com/blog/2015/03/31/advice-for-data-scientists/)
- <http://techblog.netflix.com/>

# Follow Data Science Twitter Celebrities

- @hmason - Formerly Accel
- @mrogati - VP of Data, Jawbone
- @djpatil - Chief Data Scientist, The White House
- @peteskromoroch, Formerly Linkedin
- @josh\_wills, Dir. Data Science, Cloudera
- @johnmyleswhite, Co-author of Machine Learning for Hackers
- @hackingdata, Formerly Chief Scientist at Cloudera
- @bigdata, Chief Data Scientist at O'Reilly Media

[giladlotan.com/2012/11/mapping-twitters-python-data-science-communities/](http://giladlotan.com/2012/11/mapping-twitters-python-data-science-communities/)

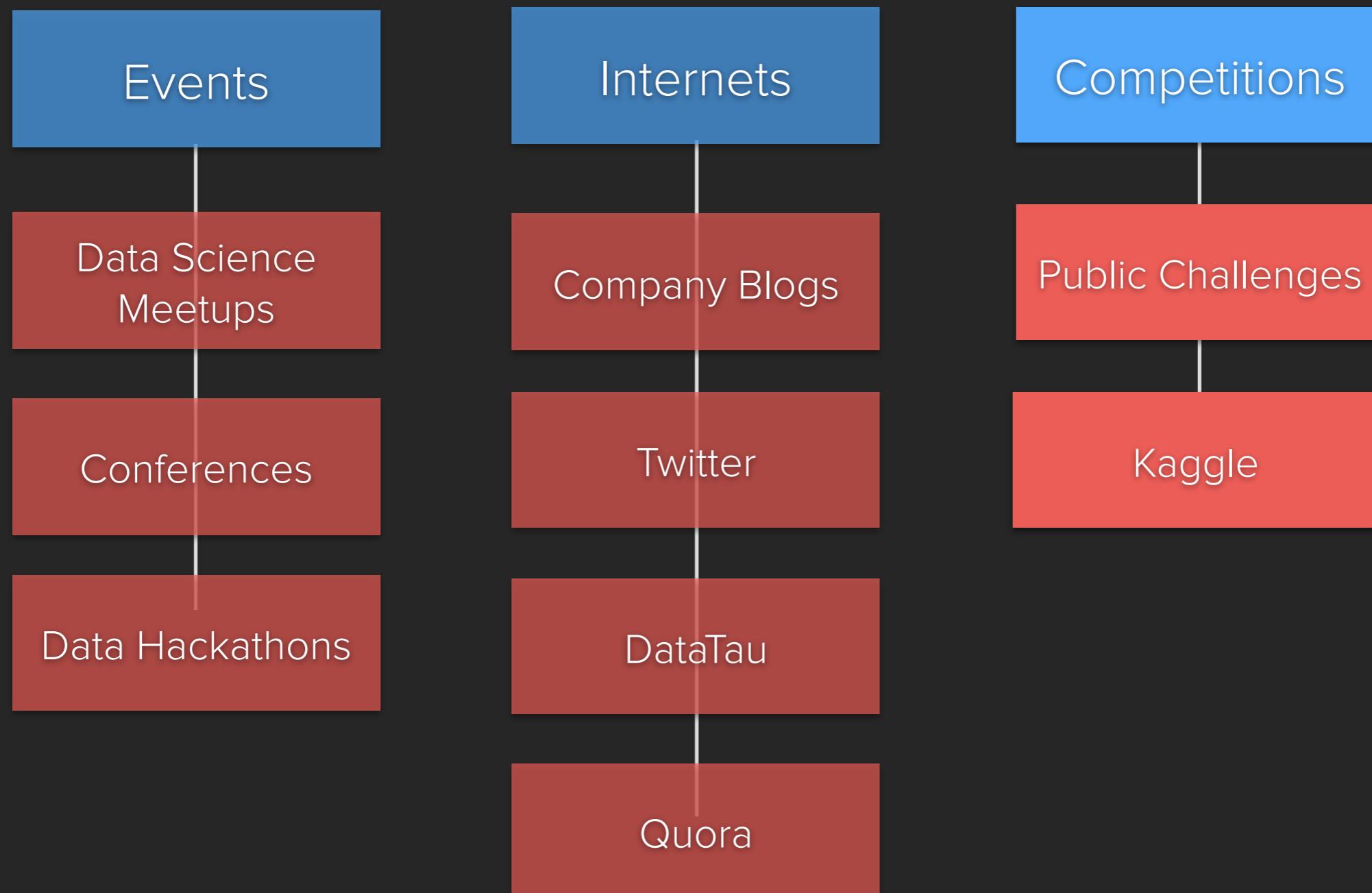
# Post to datatau.com

 DataTau new | comments | leaders | submit

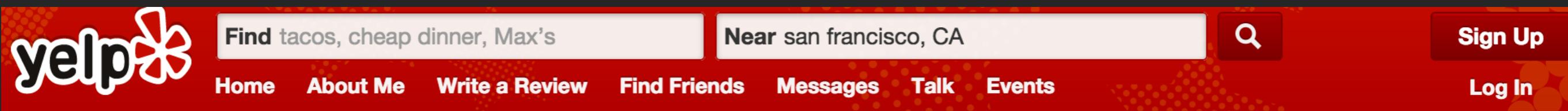
login

1. [▲ Cross-validation != magic](#) (andrewgelman.com)  
3 points by don 3 hours ago | 1 comment
2. [▲ Interview with Chris Wiggins, chief data scientist at the New York Times](#) (simplystatistics.org)  
5 points by jtleek 1 day ago | discuss
3. [▲ Ask DT: Who is hiring? \(June 2015\)](#)  
8 points by sbairishal 1 day ago | 9 comments
4. [▲ Introduction to statistics, in python.](#) (work.thaslwander.at)  
8 points by Lofkin 1 day ago | discuss
5. [▲ Clustering 733 Million Web Pages on a single machine](#) (de-vries.id.au)  
4 points by helium 1 day ago | discuss
6. [▲ High Frequency Trading Strategies](#) (jonathankinlay.com)  
2 points by 1\_over\_n 20 hours ago | discuss
7. [▲ An illustrated introduction to the t-SNE algorithm](#) (oreilly.com)  
5 points by cast42 1 day ago | discuss
8. [▲ Python library for interactive topic model visualization. Port of the R LDavis](#) (github.com)  
5 points by cast42 1 day ago | discuss
9. [▲ The ultimate reading list for finance](#) (filll.com)  
10 points by fillldotcom 3 days ago | 1 comment
10. [▲ Tutorial on hyperparameter search](#) (dato.com)  
10 points by alicez 4 days ago | 4 comments
11. [▲ Ask DT: How to excel in interviews?](#)  
3 points by thelittlec 1 day ago | 2 comments
12. [▲ Out-of-core Learning and Model Persistence using scikit-learn](#) (github.com)  
2 points by cast42 1 day ago | 1 comment
13. [▲ So, you need a statistically significant sample?](#) (stitchfix.com)  
15 points by astrobiased 6 days ago | discuss
14. [▲ 6 Tricks I Learned From The OTTO Kaggle Challenge](#) (medium.com)  
17 points by Thony 8 days ago | 2 comments
15. [▲ Extrapolating Data with Day-of-Week Effects](#) (periscope.io)  
8 points by jonathanbishop 5 days ago | discuss
16. [▲ Meta bag algorithm](#) (kaggle.com)  
4 points by mikeskim 4 days ago | 2 comments
17. [▲ Why I Left My \[Data Science\] Master's Program](#) (linkedin.com)  
15 points by tfturing 8 days ago | 1 comment
18. [▲ Automatically generate album covers with deep learning](#) (apassant.net)  
3 points by apassant 3 days ago | 1 comment

# Where (Great) Data Scientists Hang Out



# Public Open Data Challenges



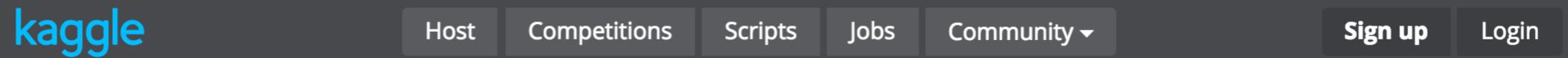
The image shows the top navigation bar of the Yelp website. It features the Yelp logo on the left, followed by two search input fields: one for "Find tacos, cheap dinner, Max's" and another for "Near san francisco, CA". A magnifying glass icon is positioned between the search fields. To the right are four buttons: "Sign Up", "Log In", and two other buttons that are partially visible.

## Yelp Dataset Challenge

**Yelp Dataset Challenge is doubling up: Now 10 cities across 4 countries!** Two years, four highly competitive rounds, over \$35,000 in cash prizes awarded and [several hundred peer-reviewed papers](#) later: the Yelp Dataset Challenge is doubling up. We are proud to announce our latest dataset that includes information about local businesses, reviews and users in 10 cities across 4 countries. The Yelp Challenge dataset is much larger and richer than the [Academic Dataset](#). This treasure trove of local business data is waiting to be mined and we can't wait to see you push the frontiers of data science research with our data.



# Kaggle



Welcome to Kaggle's data science competitions.

New to Data Science?

[Tutorials on the Titanic competition »](#)

Want to learn from other's code?

[Kaggle's top rated scripts »](#)



## Download

Choose a competition & download the training data.



## Build

Build a model using whatever methods and tools you prefer.



## Submit

Upload your predictions. Kaggle scores your solution and shows your score on the leaderboard.

### Active Competitions

#### All Competitions

### Active Competitions



 Springleaf  
Lending made personal

 Deloitte.

#### Springleaf Marketing Response

Determine whether to send a direct mail piece to a customer

20 days  
1785 teams  
994 scripts  
\$100,000

#### Western Australia Rental Prices

Predict rental prices for properties across Western Australia

2 months  
38 teams  
\$100,000

\*Bonus Lesson!\*

4

# Know How to Pick Them

If you interview data scientists, go  
read this article:

[firstround.com/review/how-to-consistently-hire-  
remarkable-data-scientists/](https://firstround.com/review/how-to-consistently-hire-remarkable-data-scientists/)

# The New Data Science Interview Process

Less bias, less time, more accurate.

Check for a Pulse

Take Home Assignment

Onsite Interviews

Offer

\*Double Bonus  
Lesson!\*

5

# Galvanize Can Help

Galvanize trains data scientists and data engineers through immersive educational programs.

12-week Data Science Immersive (fka Zipfian Academy)

12-week Data Engineering Immersive)

GalvanizeU Masters' of Science in Data Science

# HOW WE CAN CONTRIBUTE

- Pipeline of trained talent
- Data Science community hub
- Custom training

[galvanize.com](http://galvanize.com)



# THANK YOU

KATIE KENT | DIRECTOR, EDUCATIONAL OUTCOMES  
@K8OKENT

[katie@galvanize.com](mailto:katie@galvanize.com)