Making Data Science Useful

Cassie Kozyrkov, Chief Decision Scientist, Google

Tip #1 of 10

Don’t forget the point of data science.

Take a practical view of data science.

A map of data science? No

SQL

Descriptive Analytics

Python

Machine Learning

R

Statistical Inference

A map of data science? No

Histogram

Descriptive Analytics

Neural network

Machine Learning

Student's t-test

Statistical Inference

A map of data science!

None	Many	Few
		
Descriptive Analytics	Machine Learning	Statistical Inference

A map of data science!

Get inspired

Descriptive Analytics

Make a recipe

Machine Learning

Decide wisely

Statistical Inference

Tip #2 of 10

Inspiration is cheap, rigor is expensive.

Conclusions about your data: cheap
Conclusions beyond your data: expensive

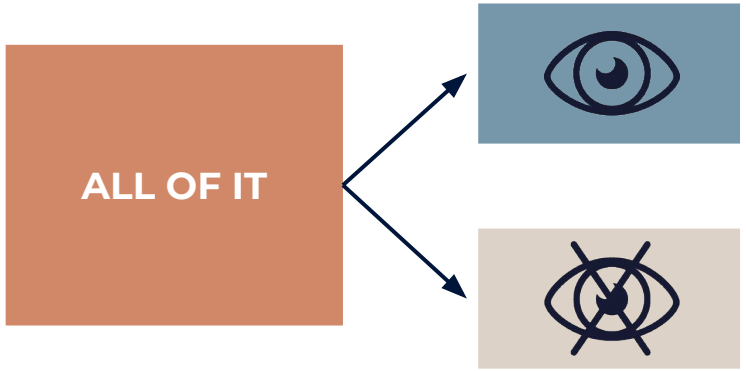
Have your cake and eat it too.

Tip #3 of 10

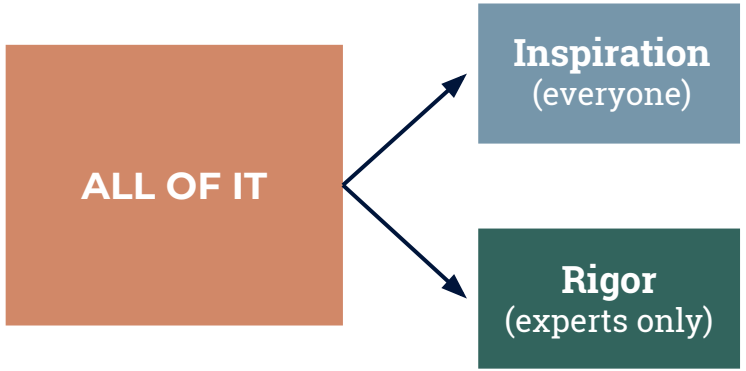
Split your data!

Don't trust "insights" without it. Restrict access.

Split your data.



Split your data.



Tip #4 of 10

Incentivize your entire workforce to look at ~~data~~ information.

Share where possible and make access easy.

Tip #5 of 10

Rigor begins with the decision-maker.

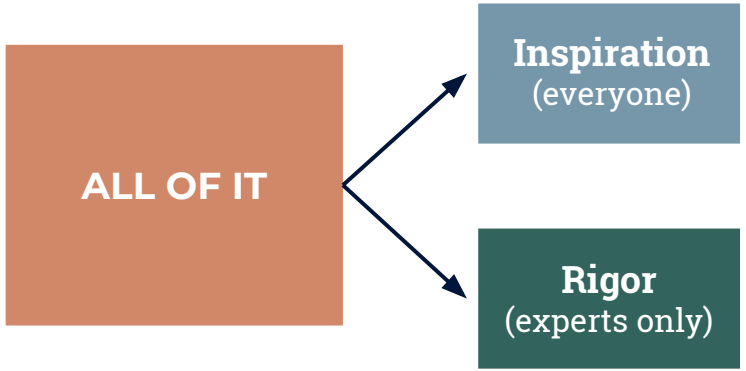
Avoid rigor for rigor's sake.

Tip #6 of 10

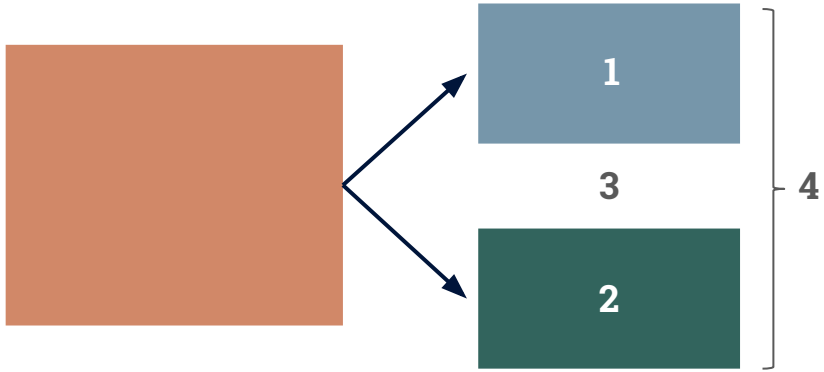
Understand how decision-making is delegated.

Trust is broken by misalignment.

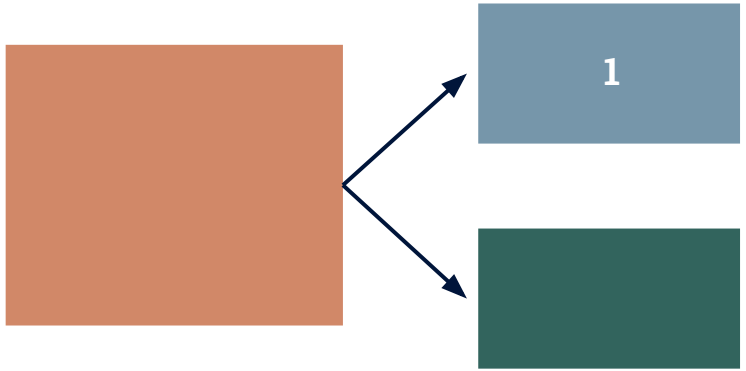
Know your role



Know your role

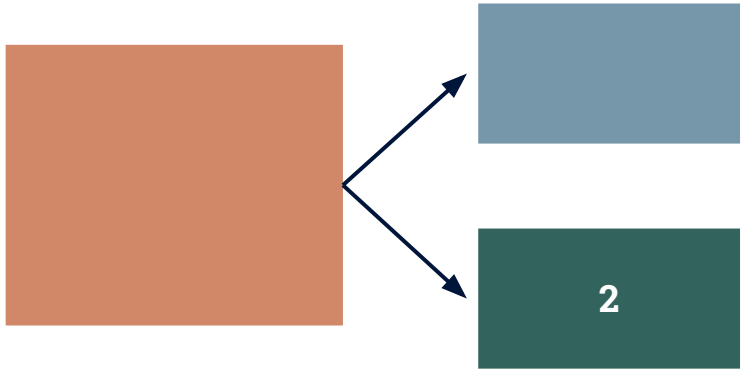


Time investor



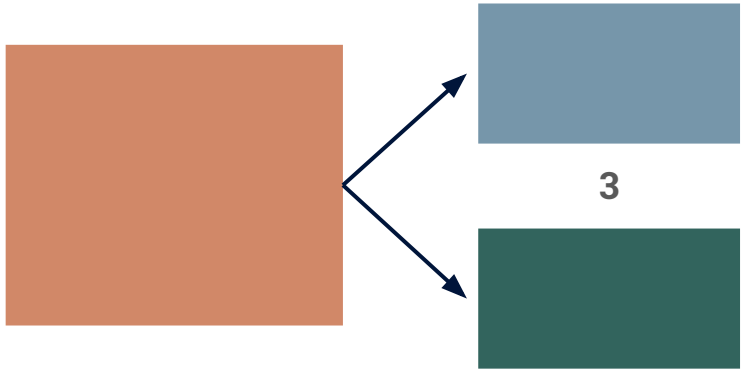
Find something actionable quickly.

Decision supporter



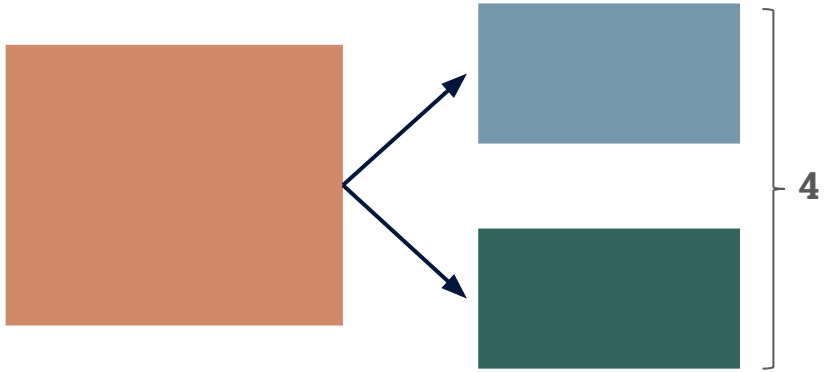
Provide the decision-maker's rigor.

Decision-maker



Exercise judgment about what's useful.

Full data science leader

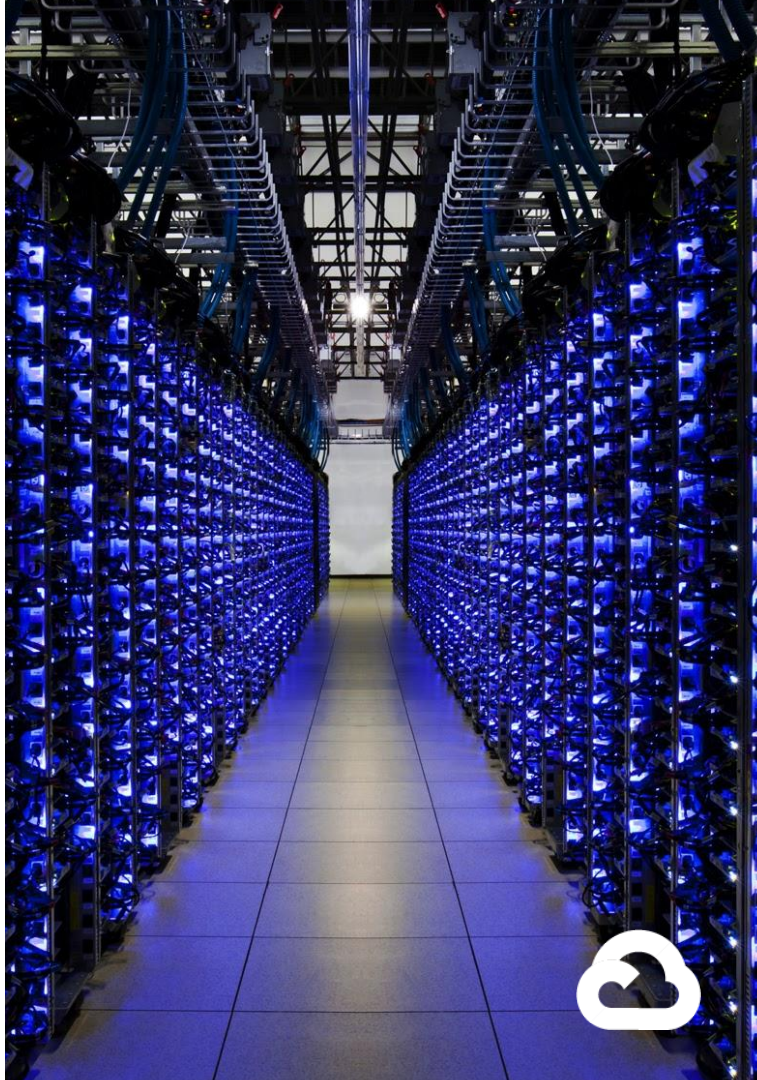


Master all three focus areas.

Tip #7 of 10

Harness the power of large datasets.

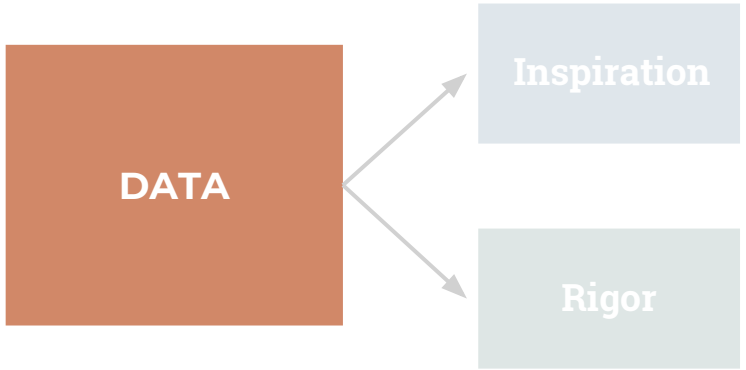
The history of data science is a history of data-splitting.



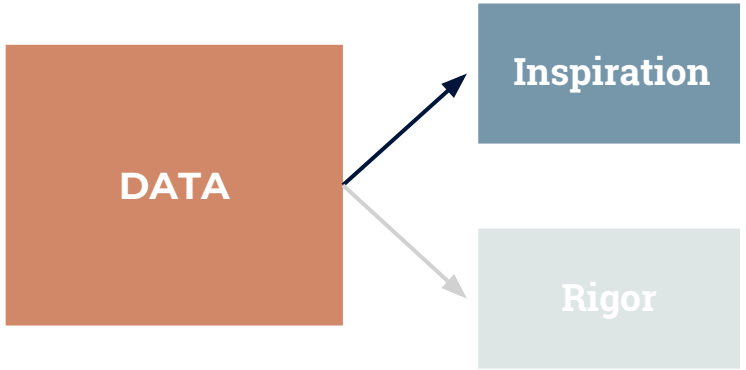
No datasets.

NO DATA

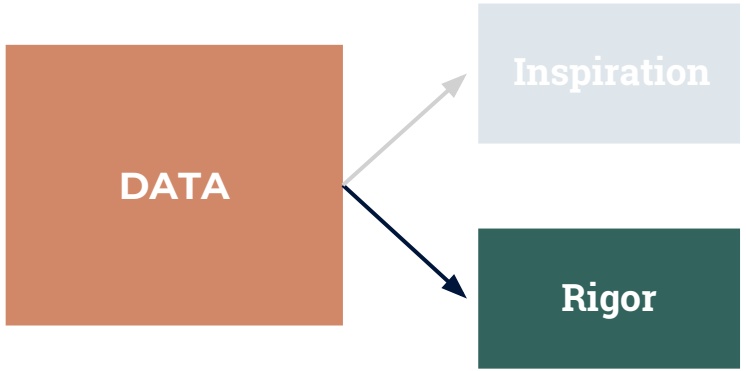
One dataset.



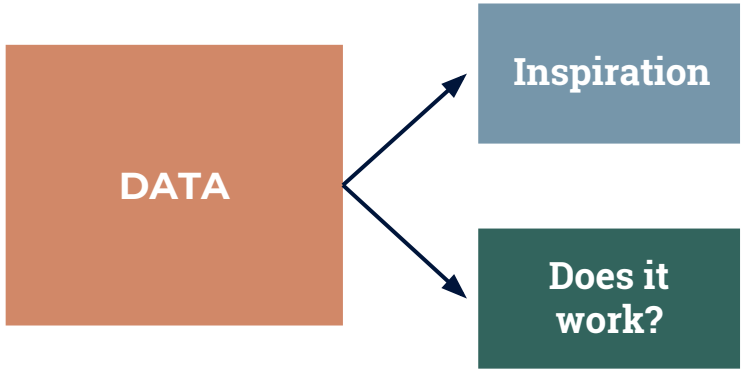
One dataset.



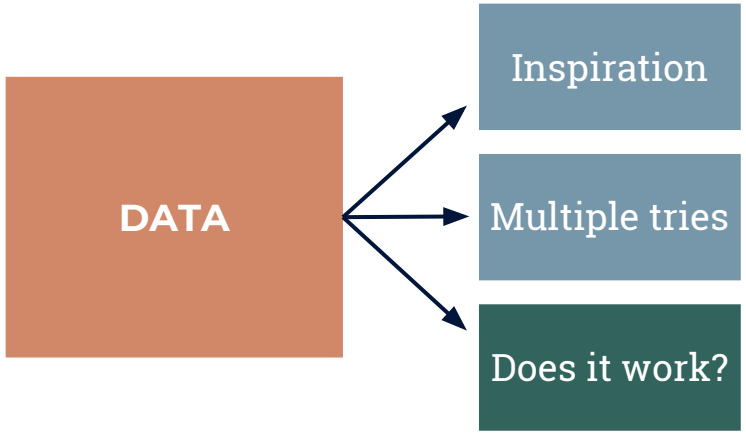
One dataset.



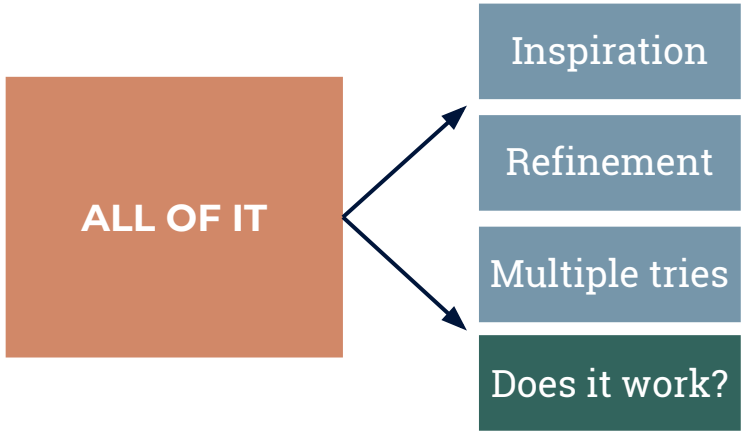
Two datasets.



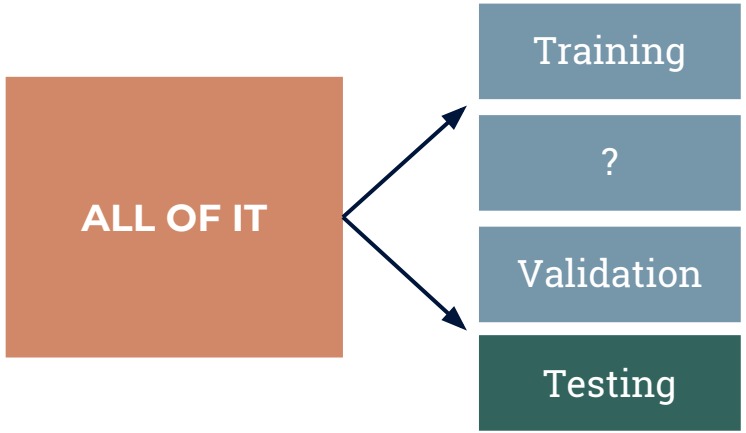
Three datasets.



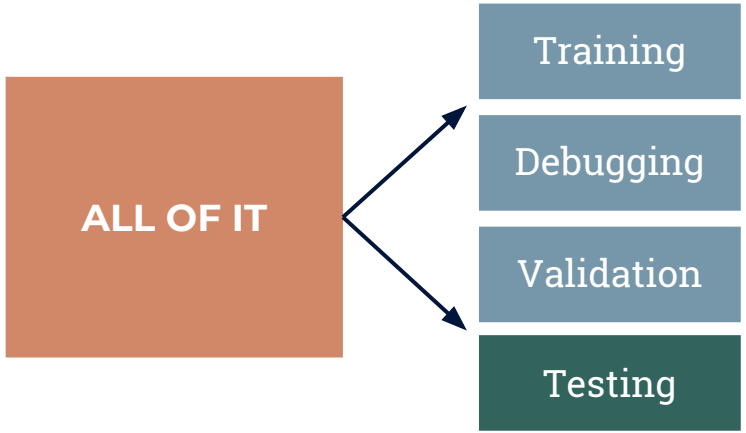
Four datasets.



Four datasets.



Four datasets.



A map of data science

Get inspired

Descriptive Analytics

Make a recipe

Machine Learning

Decide wisely

Statistical Inference

Applied AI and machine learning

Select

Descriptive Analytics

Train

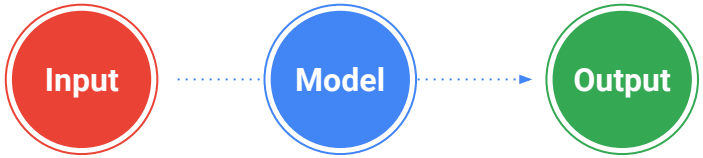
Machine Learning

Test

Statistical Inference

Tip #8 of 10

Do things in the right order.



1. ***Outputs*** → *Decision-Maker*
2. ***Performance*** → *Statistics Leader*
3. ***Inputs*** → *Analytics Leader*
4. ***Models*** → *Machine Learning Leader*

Tip #9 of 10

Take data quality seriously.

You are at the mercy of data engineering.

Tip #10 of 10

Testing is the best basis for trust.

Make sure it works on new data.
Apply rigorous statistical principles.


Thank you !
@quaesita