2018 Top 10 Business Intelligence Trends

Don’t Fear Artificial Intelligence

1.
Don’t Fear Artificial Intelligence

2.
Liberal Arts Impact

3.
Promise of NLP

4.
Multi-Cloud Debate

5.
Rise of the Chief Data Officer

6.
Crowdsourced Governance

7.
Data Insurance

8.
Data Engineer Role

9.
Location Internet of Things

10.
Academics Investment

Don’t Fear Artificial Intelligence

Moderator & Speakers

Andy Cotgreave

Technical Evangelism Director

Moderator

Chris Fraley

Senior Research Scientist

Eric Brochu

Senior Software Engineer

Who we are

Andy Cotgreave

Andy is a visual analytics expert who has been with Tableau since 2011 in various roles ranging from product consultant to social content manager. He is now Tableau's technical evangelism director. Prior to Tableau, he was a data analyst at the University of Oxford.

As a technical evangelist, Andy helps people see and understand their data using Tableau's innovative products. He shares his passion for visual analysis and technology with his writing, (e.g. Computerworld, on tableau.com, and his own blog), speaking at industry conferences like SXSW and Tableau’s own events.

Who we are

Eric Brochu

- PhD in CS from University of British Columbia
  - Thesis in Machine Learning, worked on models for nonlinear optimization problems using human input
- Joined Tableau in March 2016
  - Working on Recommendations team, helping design and implement recommender systems for tables & joins and data sources.

Who we are

Chris Fraley

- PhD in CS from Stanford University
  - 25+ years experience in statistical computing
- Joined Tableau in November 2015
  - Working on Analytics research and development unit
  - Contributed to Tableau's clustering algorithm

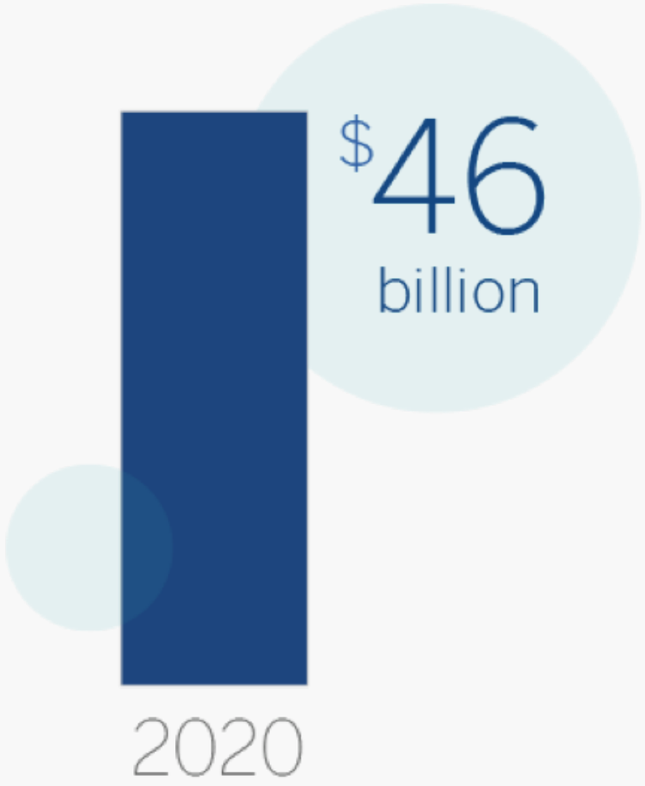
Agenda

1. Don’t Fear Artificial Intelligence
2. Machine Learning Definitions
3. Advantages & Limitations with Machine Learning
4. Machine Learning at Tableau
5. Open Discussion with Panel

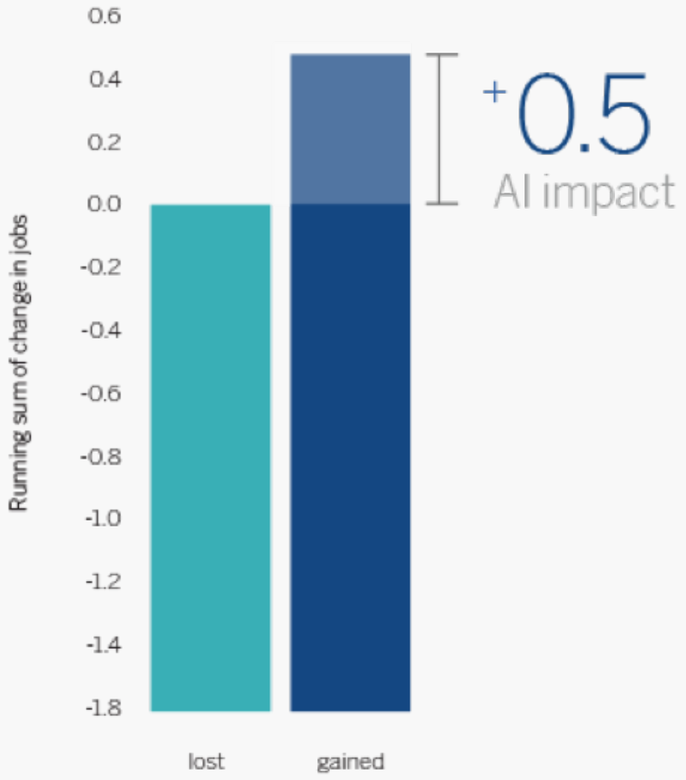
Machine Learning Introduction

AI & Machine Learning To Boom

IDC forecasts revenues from AI and machine learning systems to total $46 billion by 2020



In 2020, AI will become a positive net job motivator, creating 2.3 million jobs, Gartner reports



“The development of full artificial intelligence could spell the end of the human race. ”

- Stephen Hawking
(BBC Interview, 2014)

“The development of full artificial intelligence **could** spell the end of the human race. ”

- Stephen Hawking
(BBC Interview, 2014)

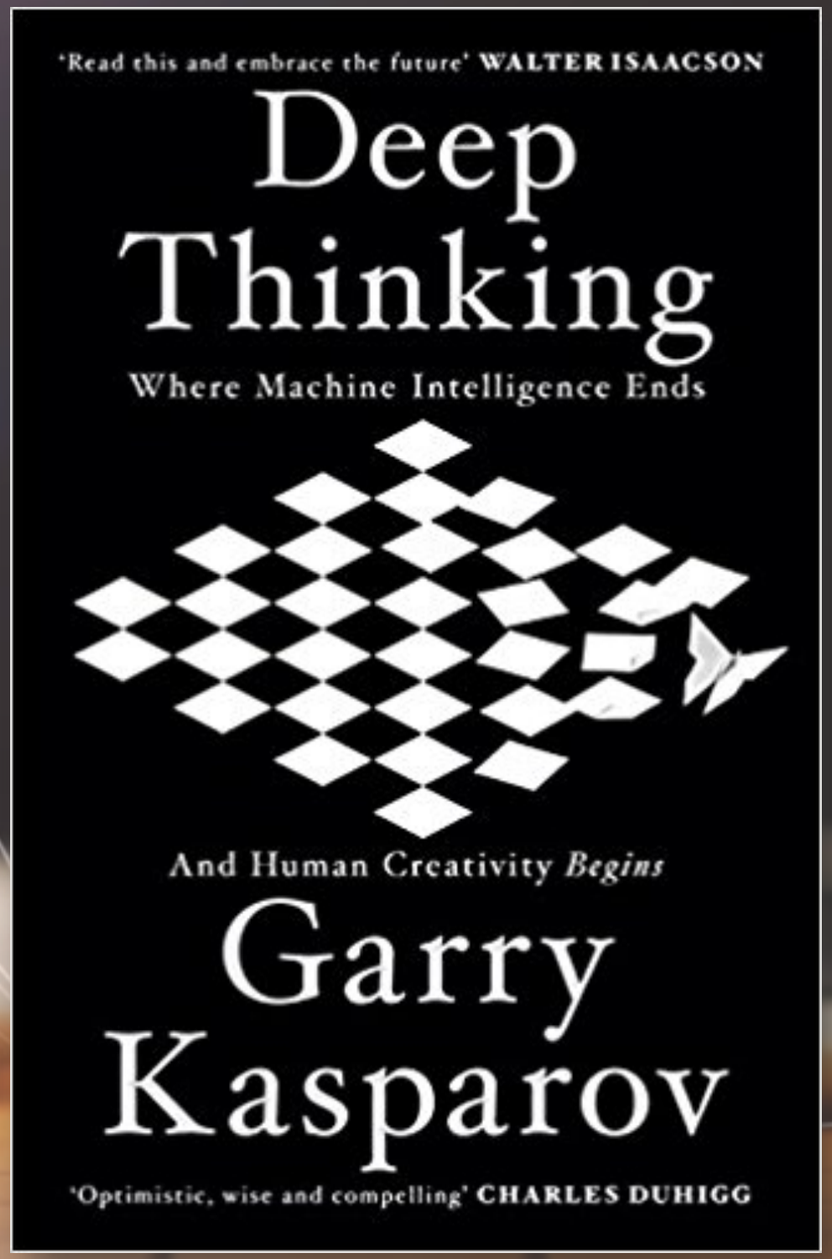
“I don’t work on not turning AI evil today for the same reason I don't worry about the problem of overpopulation on the planet Mars.”

Andrew Ng,
VP & Chief Scientist, Baidu


AlphaGo Zero
Starting from scratch

AlphaGo Zero

Starting from scratch



The Telegraph

ALL SECTIONS

Technology

More

Technology

Robots will take a third of British jobs by 2030, report says

The Starship delivery robot has already started transporting food orders

INDEPENDENT

News InFact Politics Voices Indy/Life Sport Business Video Culture

News › Science

The robots are coming – but will they really take all our jobs?

Artificial intelligence is going to transform the world, and our lives. But are we heading for a brave new world, or a science fiction horror-show?

David Barnett | Thursday 30 November 2017 06:15 GMT | 58 comments

168 shares

Like Click to follow The Independent Online

Menial, repetitive drudgery could be absorbed by robots, which would leave us humans more free time – but to do what? *Getty*

Last week, Chancellor Philip Hammond announced in the Autumn Budget a £500m package of investment into tech

PROMOTED STORIES

INDEPENDENT News InFact Politics Voices Indy/Life Sport Business Video Culture

News › Science

The robots are co
really take all our

Artificial intelligence is going to transform
new world, or a science fiction horror-show

David Barnett | Thursday 30 November 2017 06:15 GMT

168 shares

Menial, repetitive drudgery could be absorbed by robots, wh

Last week, Chancellor Philip Hammond announced in the Autumn Budget a £500m package of investment into tech

PROMOTED STORIES

The Telegraph

ALL SECTIONS

More

ritish jobs by

porting food orders

pwc | UK

Up to 30% of existing UK jobs could be impacted by automation by early 2030s, but this should be offset by job gains elsewhere in economy

Mar 24, 2017

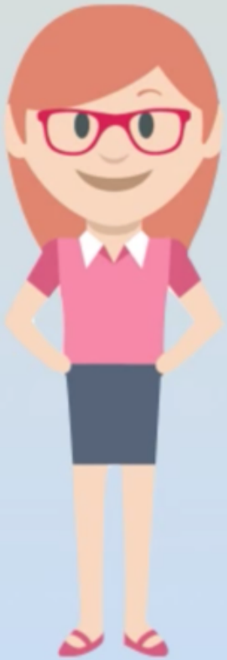
- Up to around 30% of existing UK jobs are susceptible to automation from robotics and Artificial Intelligence (AI) by the early 2030s, but in many cases the nature of jobs will change rather than disappear

Machine Learning Definitions

What is Machine Learning?

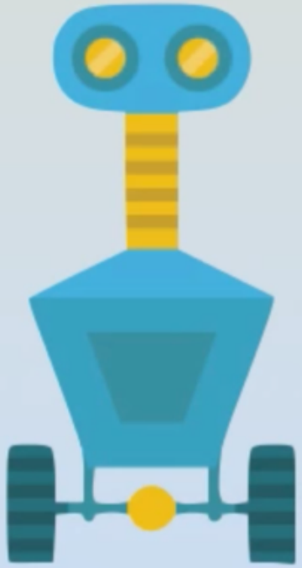
What is Machine Learning?

Learn from experience



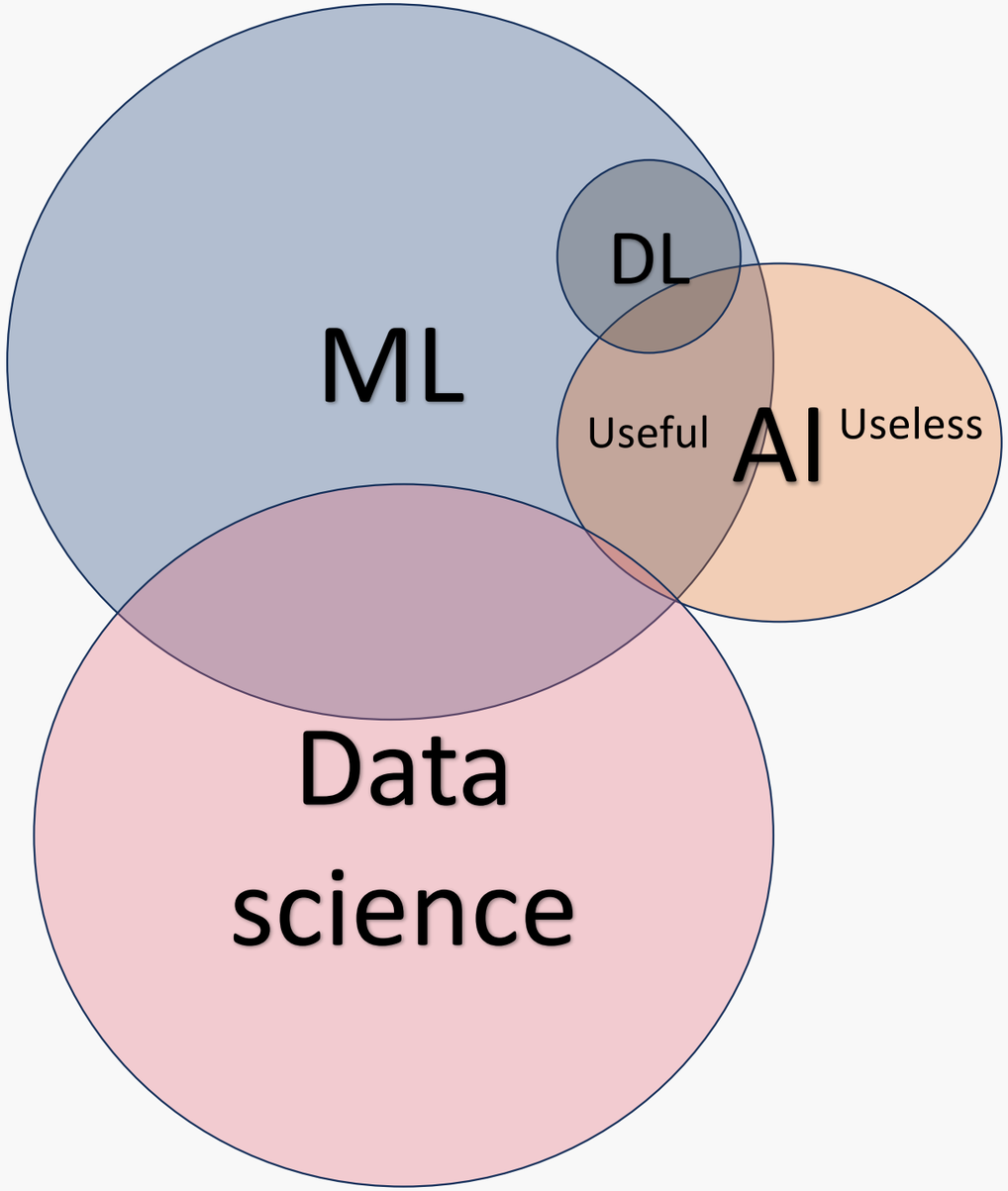
Learn from ~~experience~~ data

Follow instructions

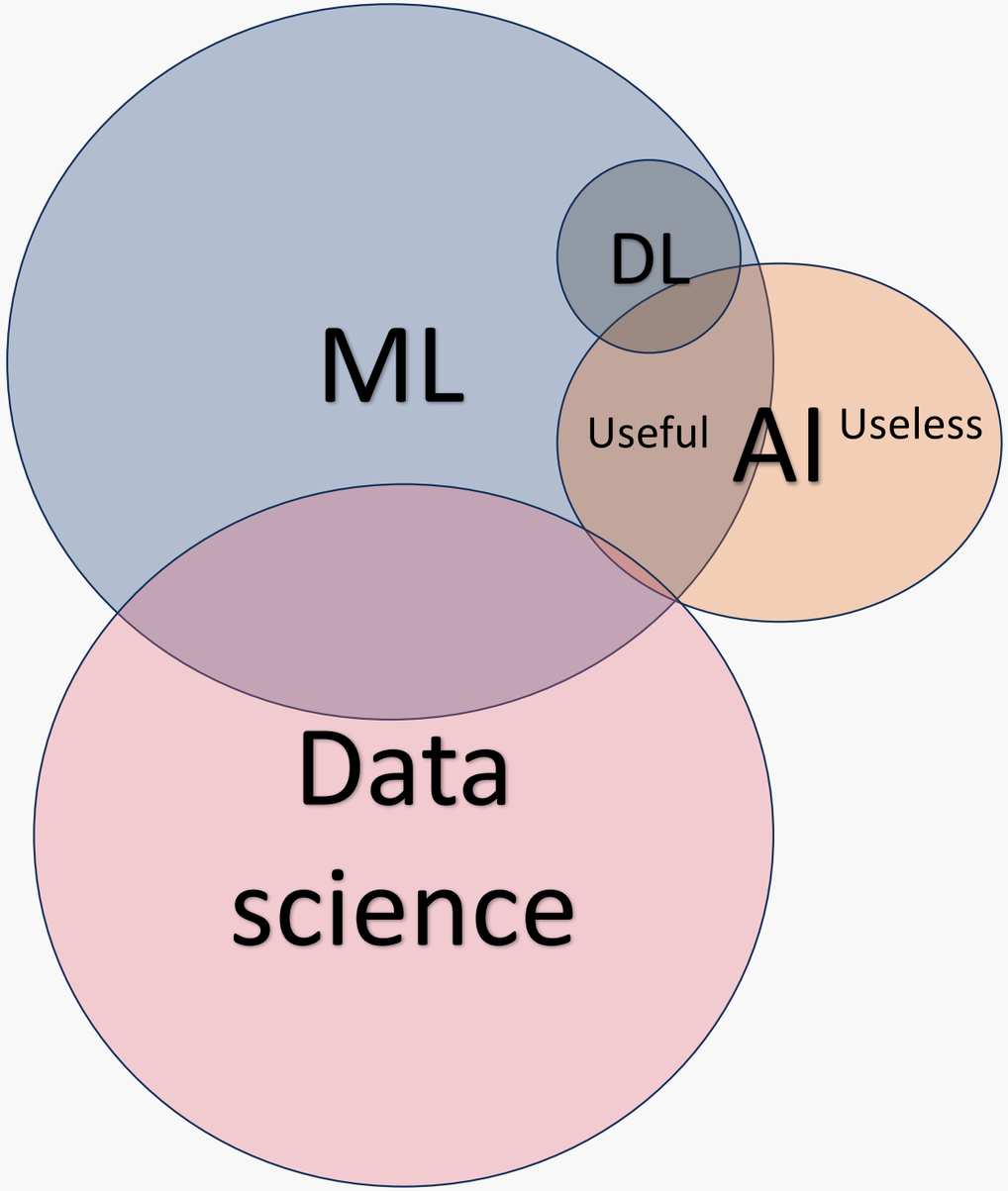


How does it differ from AI? Data science? Deep learning?

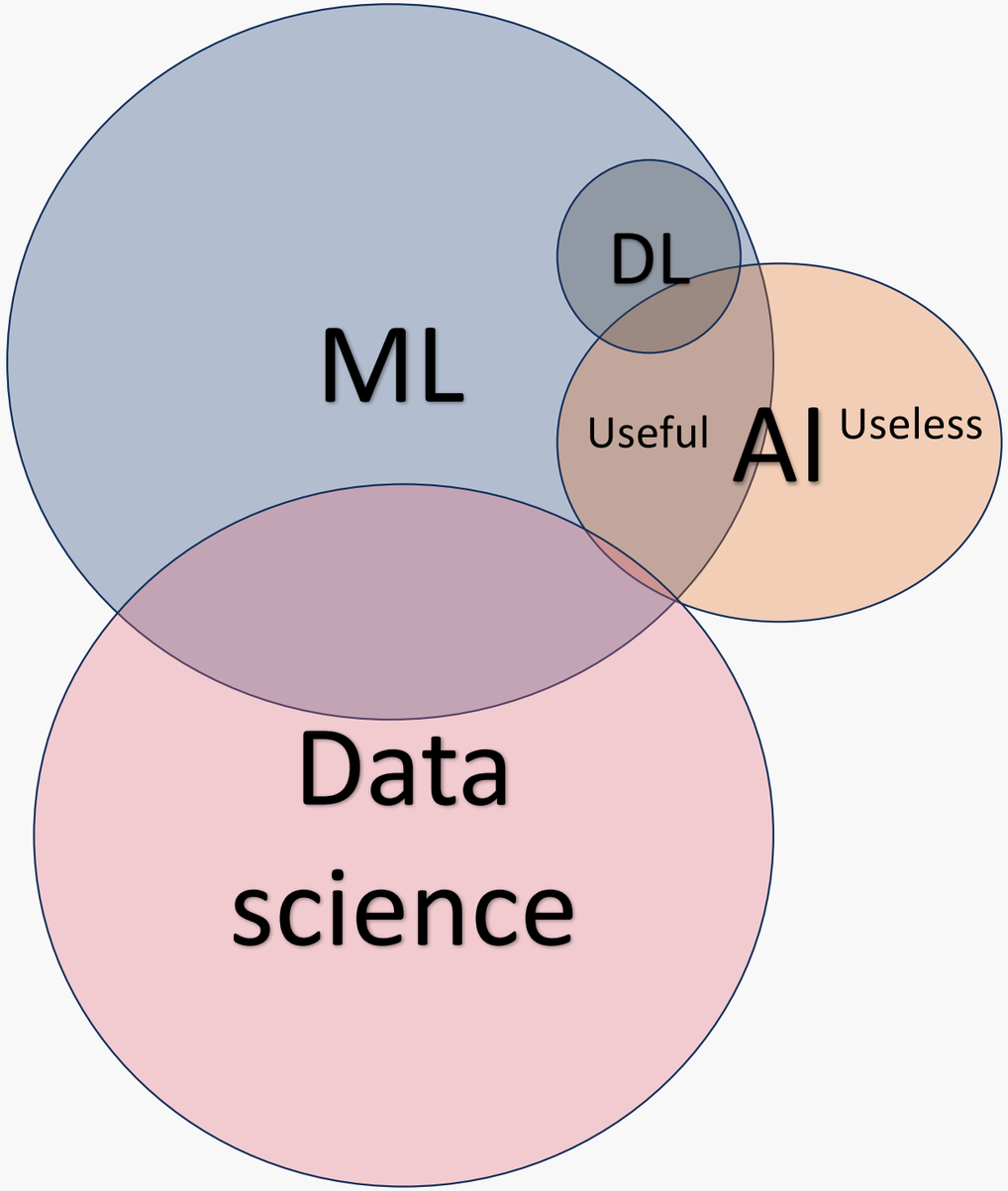
- Machine learning
  - Methods which are trained on past data to produce future output



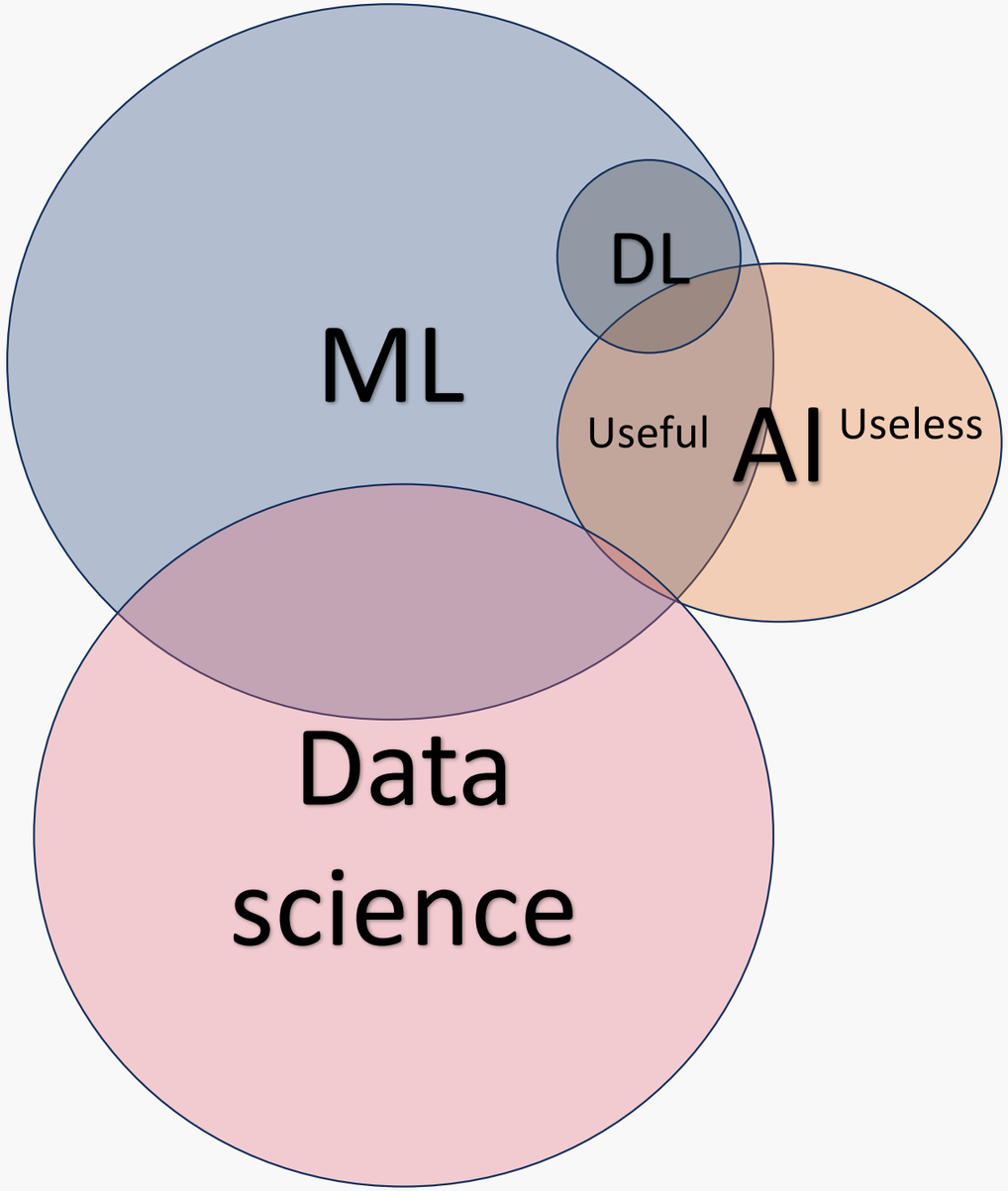
- Machine learning
  - Methods which are trained on past data to produce future output
- AI
  - Mimicking *human* understanding and behavior
  - No requirement to use ML
  - ML happens to work the best



- Machine learning
  - Methods which are trained on past data to produce future output
- Deep learning
  - Specific collection of Neural Net based ML methods
  - Pretty amazing results on certain problems



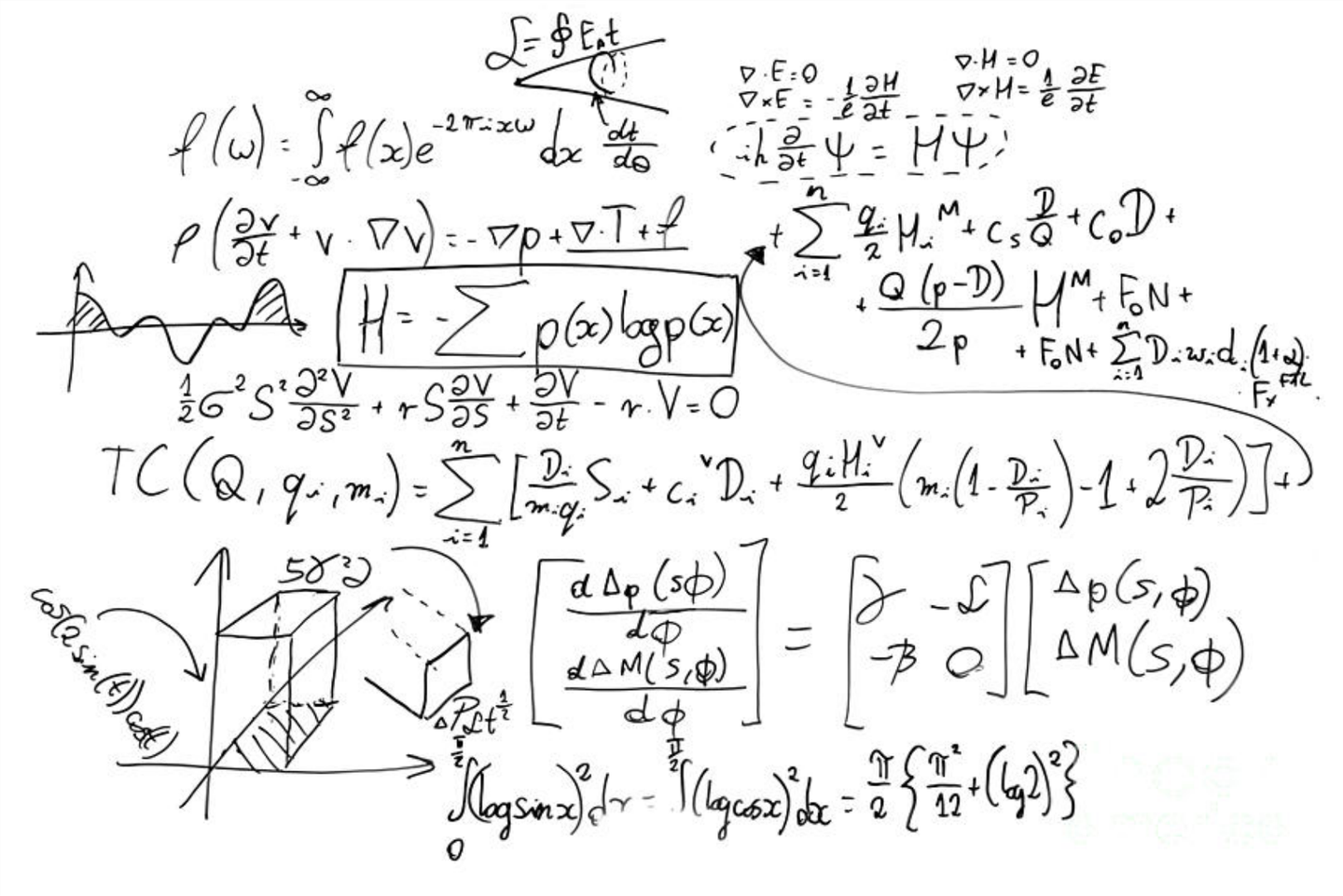
- Machine learning
  - Methods which are trained on past data to produce future output
- Data science
  - Full pipeline to solve problems with data
  - Often uses ML, focuses less on automation
- Statistics
  - Same / similar techniques
  - Usually different goals



Advantages / Limitations of ML

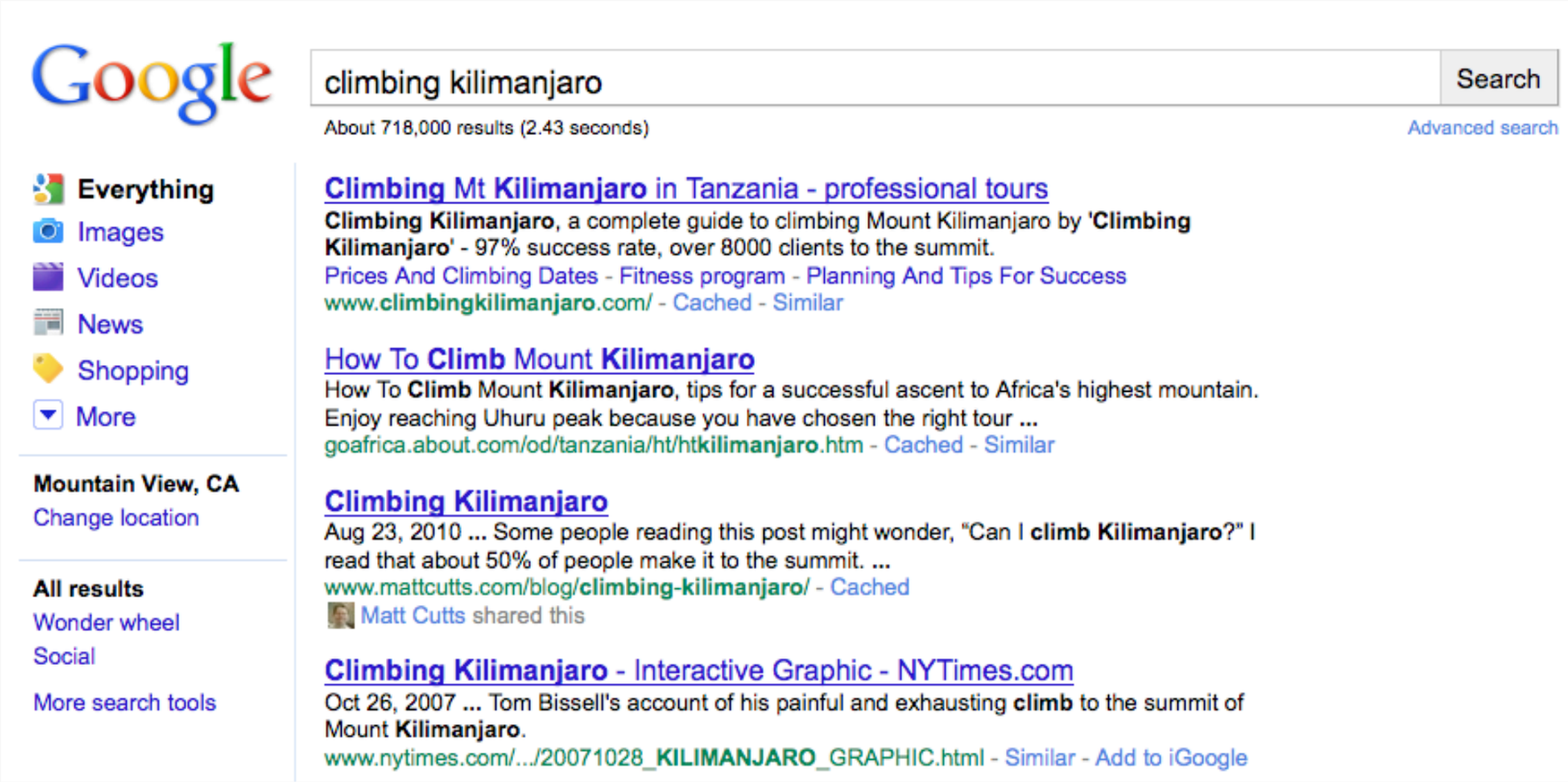
Advantages of Machine Learning

#1: Don’t have to develop math/stat algorithms from scratch



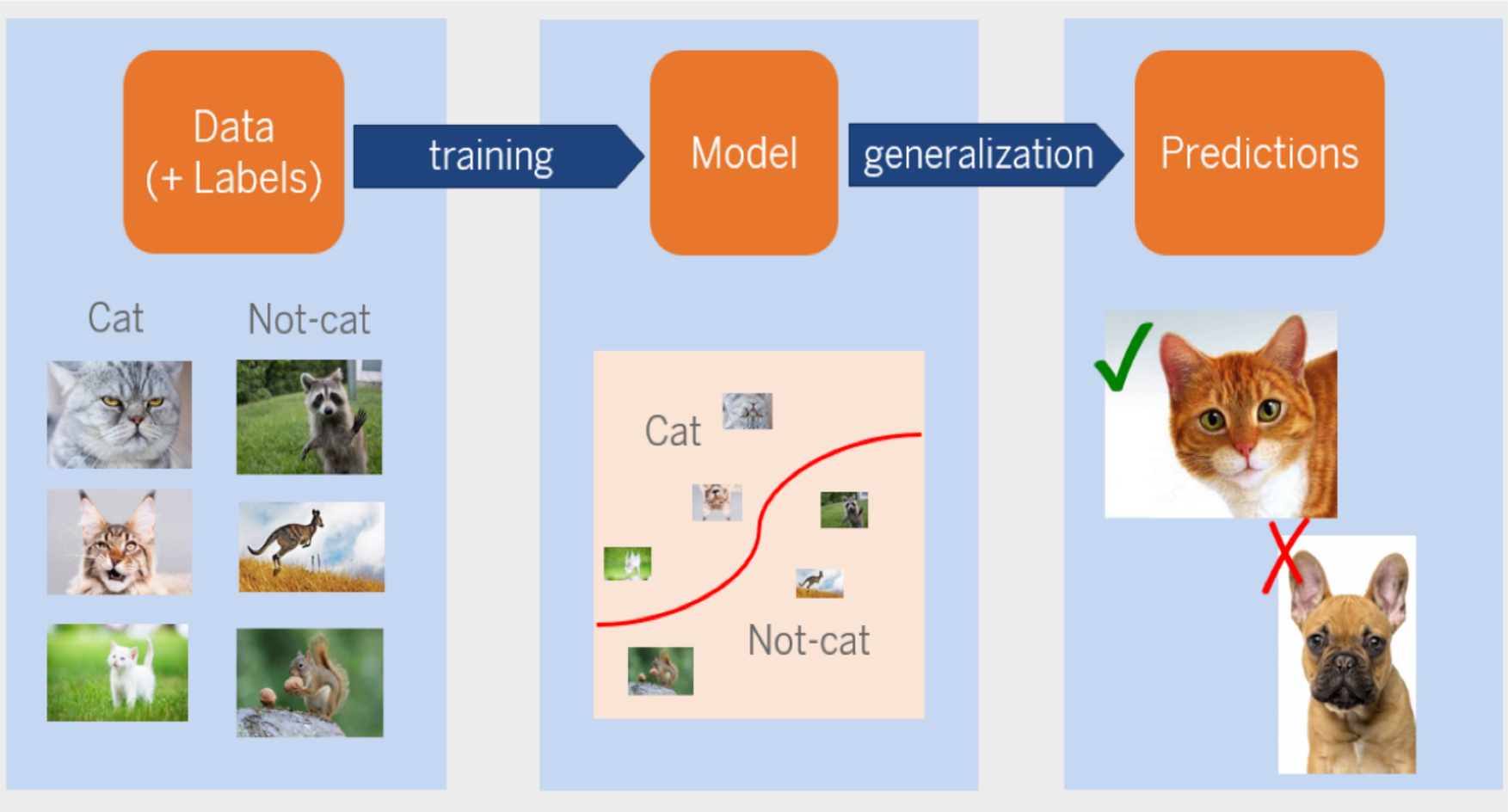
Advantages of Machine Learning

#2: Speed up decision making through automation



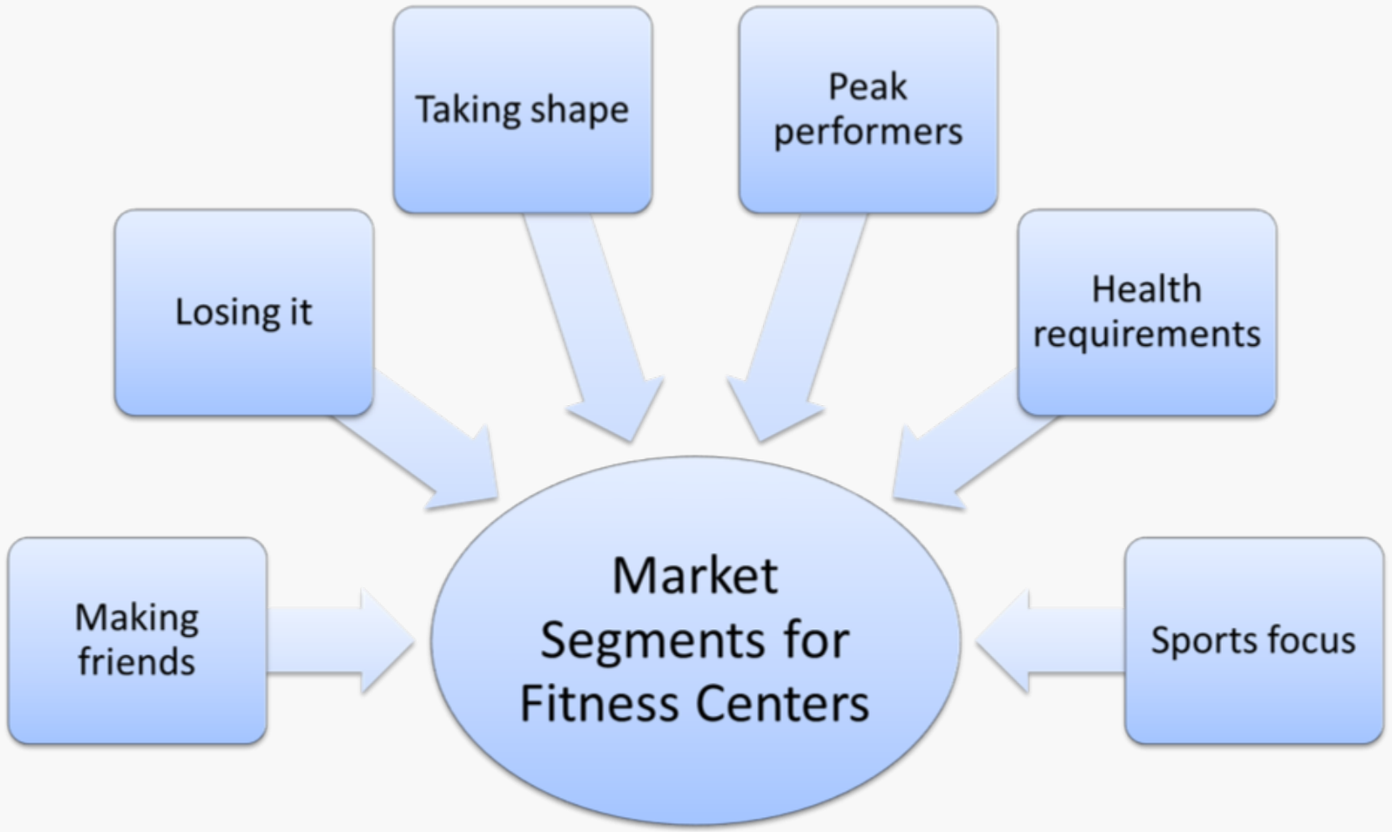
Advantages of Machine Learning

#3: Learn automatically through examples (e.g. classification)



Advantages of Machine Learning

#4: Discover patterns, like clustering/segmentation or outlier/anomaly detection

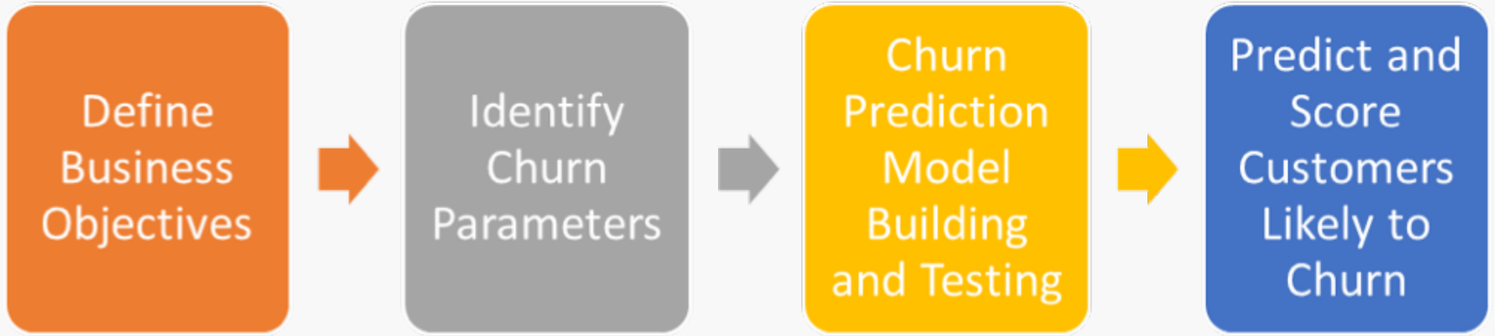


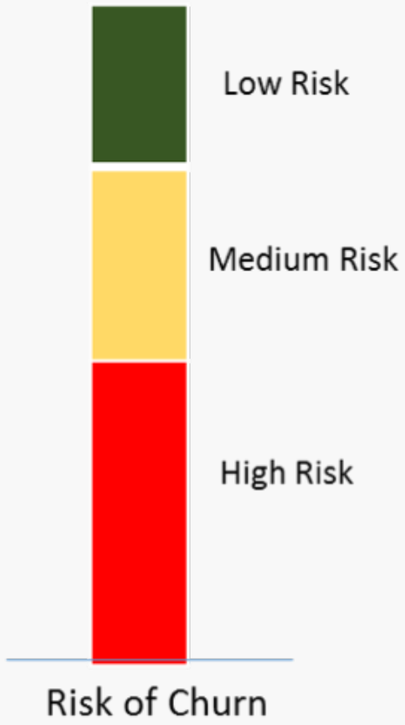
Advantages of Machine Learning

#4: Discover patterns, like clustering/segmentation or outlier/anomaly detection

Advantages of Machine Learning

#5: Predict outcomes





Advantages of Machine Learning

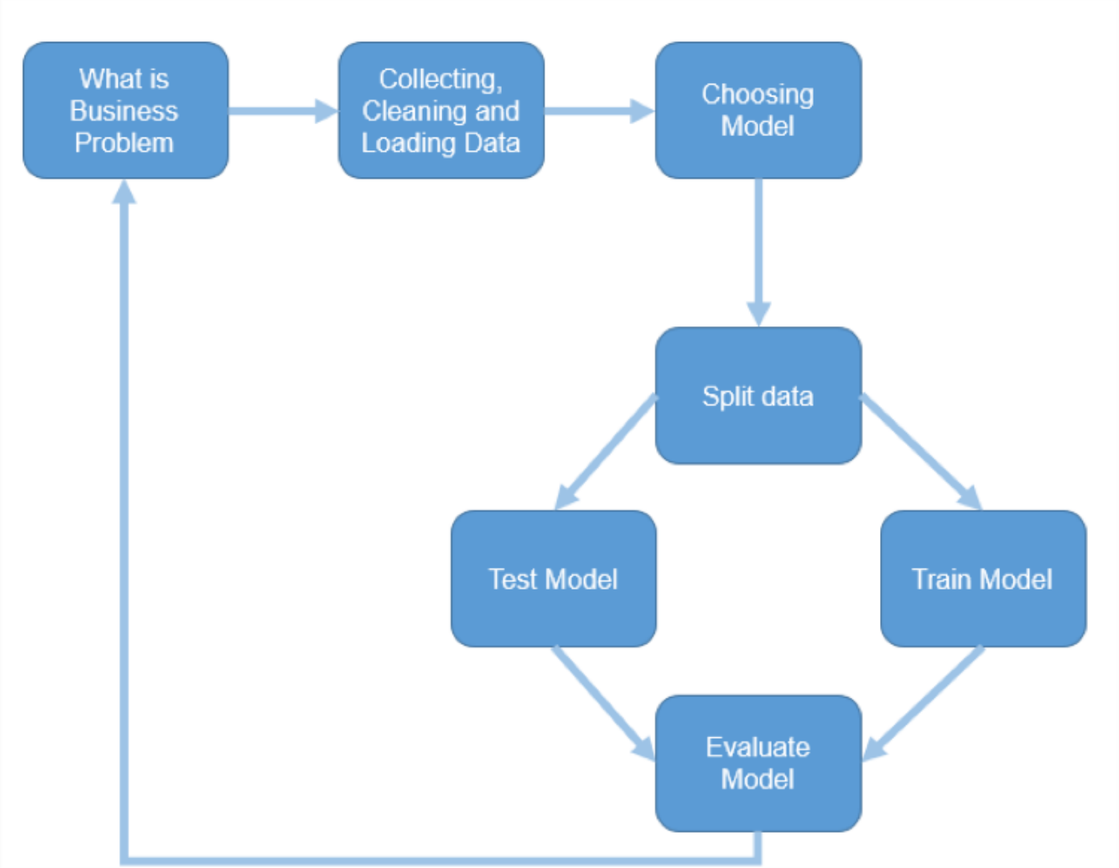
#6: Identify which inputs most affect outcomes

Limitations of Machine Learning

#1: Doesn't eliminate the importance of quantity and quality of data collected

Limitations of Machine Learning

#2: Need for validation and uncertainty quantification



Limitations of Machine Learning

#3: Decisions to be made when using machine learning methods:

Data transformation

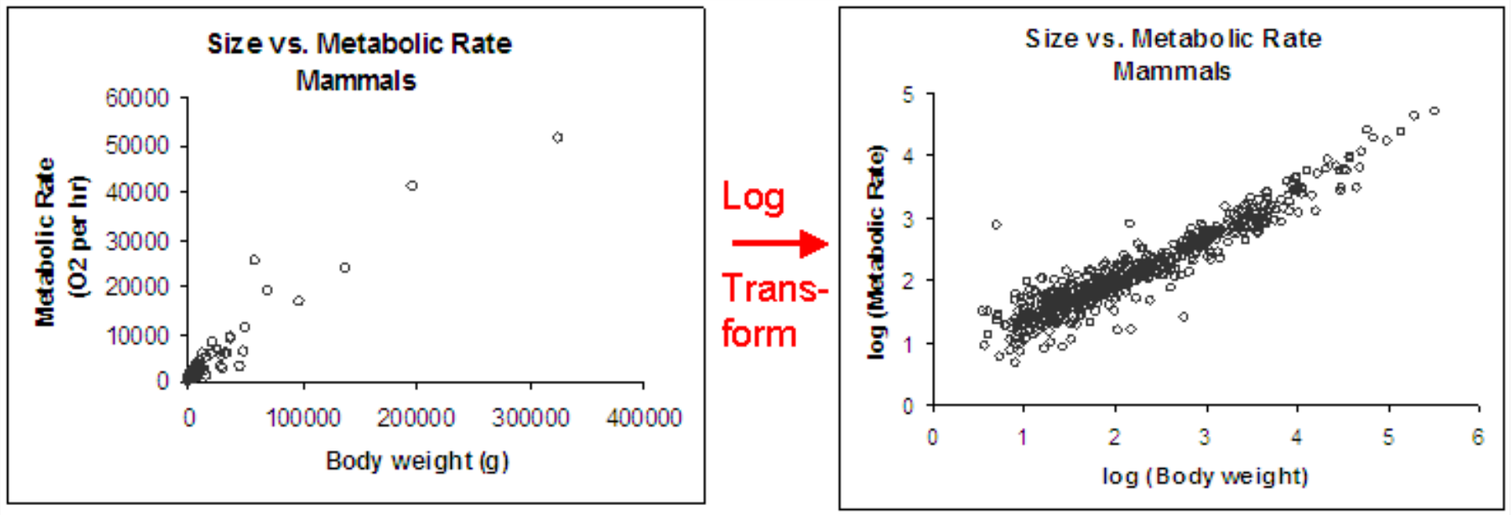
Model selection

Variable selection

Parameter settings

Handling missing data

Handling outliers/anomalies



Limitations of Machine Learning

#3: Decisions to be made when using machine learning methods:

Data transformation

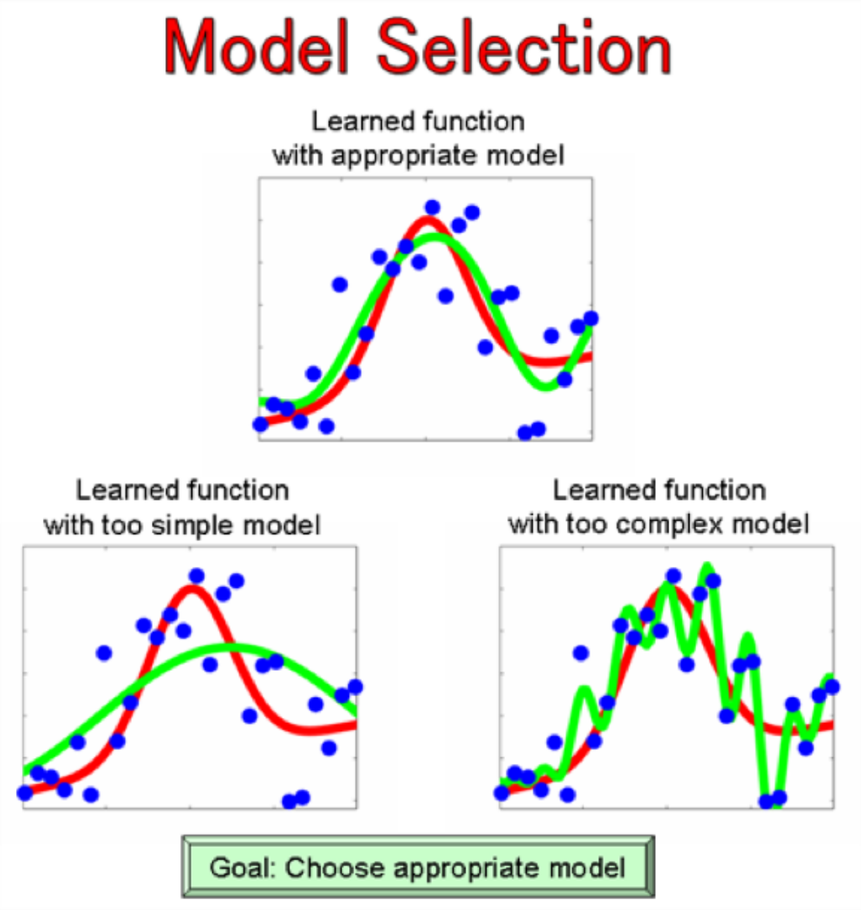
Model selection

Variable selection

Parameter settings

Handling missing data

Handling outliers/anomalies



Limitations of Machine Learning

#3: Decisions to be made when using machine learning methods:

Data transformation

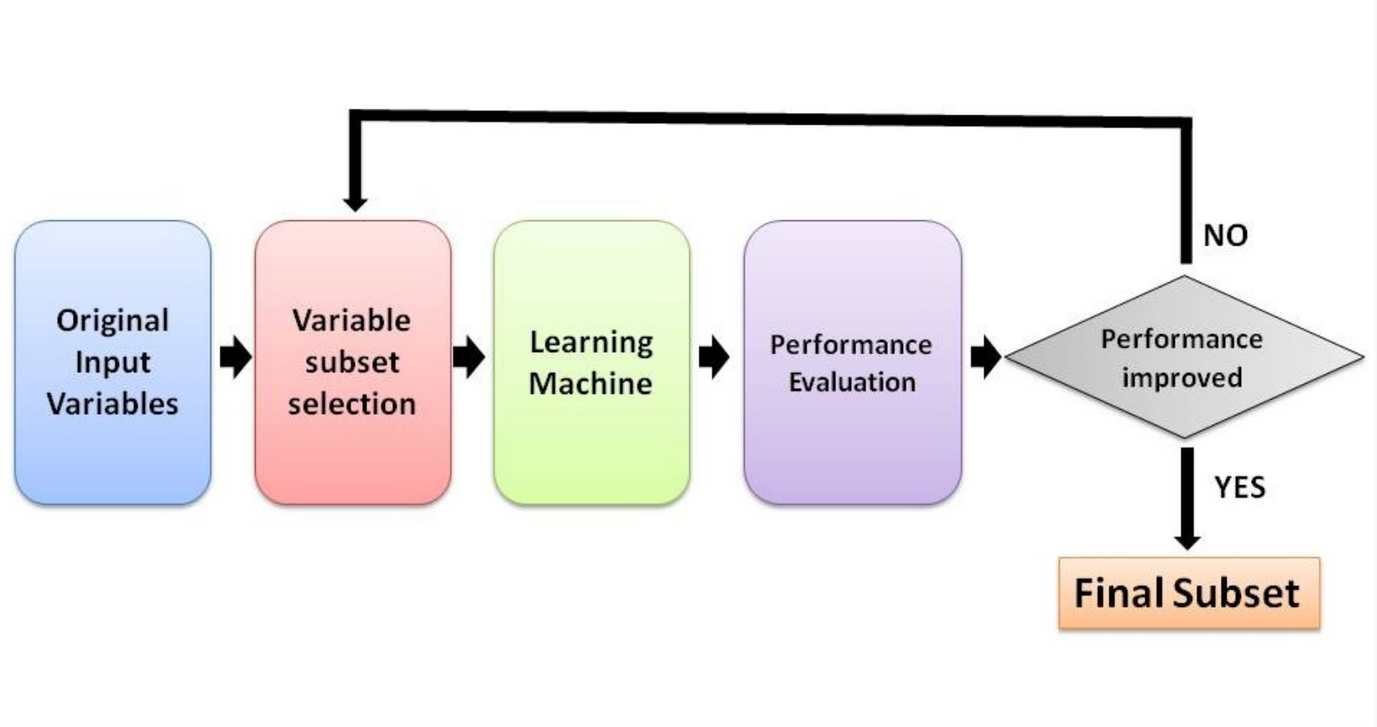
Model selection

Variable selection

Parameter settings

Handling missing data

Handling outliers/anomalies



Limitations of Machine Learning

#3: Decisions to be made when using machine learning methods:

Data transformation

Model selection

Variable selection

Parameter settings

Handling missing data

Handling outliers/anomalies

Hyper-parameter	Variants
Non-linearity	linear, tanh, sigmoid, ReLU, VLReLU, RReLU, PReLU, ELU, maxout, APL, combination
Batch Normalization (BN)	before non-linearity. after non-linearity
BN + non-linearity	linear, tanh, sigmoid, ReLU, VLReLU, RReLU, PReLU, ELU, maxout
Pooling	max, average, stochastic, max+average, strided convolution
Pooling window size	3x3, 2x2, 3x3 with zero-padding
Learning rate decay policy	step, square, square root, linear
Colorspace & Pre-processing	RGB, HSV, YCrCb, grayscale, learned, CLAHE, histogram equalized
Classifier design	pooling-FC-FC-clf, SPP-FC-FC-clf, pooling-conv-conv-clf-avepool, pooling-conv-conv-avepool-clf
Network width	$1/4$, $1/2\sqrt{2}$, $1/2$, $1/\sqrt{2}$, 1 ,$\sqrt{2}$, 2, $2\sqrt{2}$, 4, $4\sqrt{2}$
Input image size	64, 96, 128, 180, 224
Dataset size	200K, 400K, 600K, 800K, 1200K(full)
Batch size	1, 32, 64, 128, 256, 512, 1024
Percentage of noisy data	0, 5%, 10%, 15%, 32%
Using bias	yes/no

Limitations of Machine Learning

#3: Decisions to be made when using machine learning methods:

Data transformation

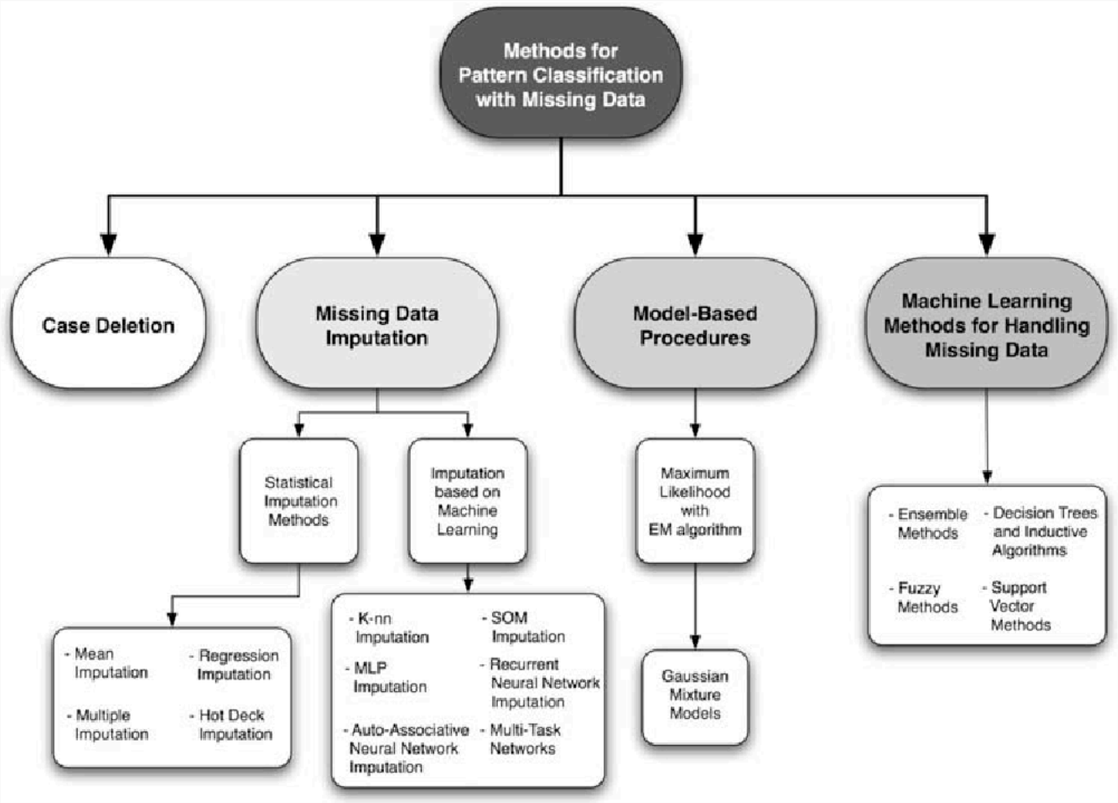
Model selection

Variable selection

Parameter settings

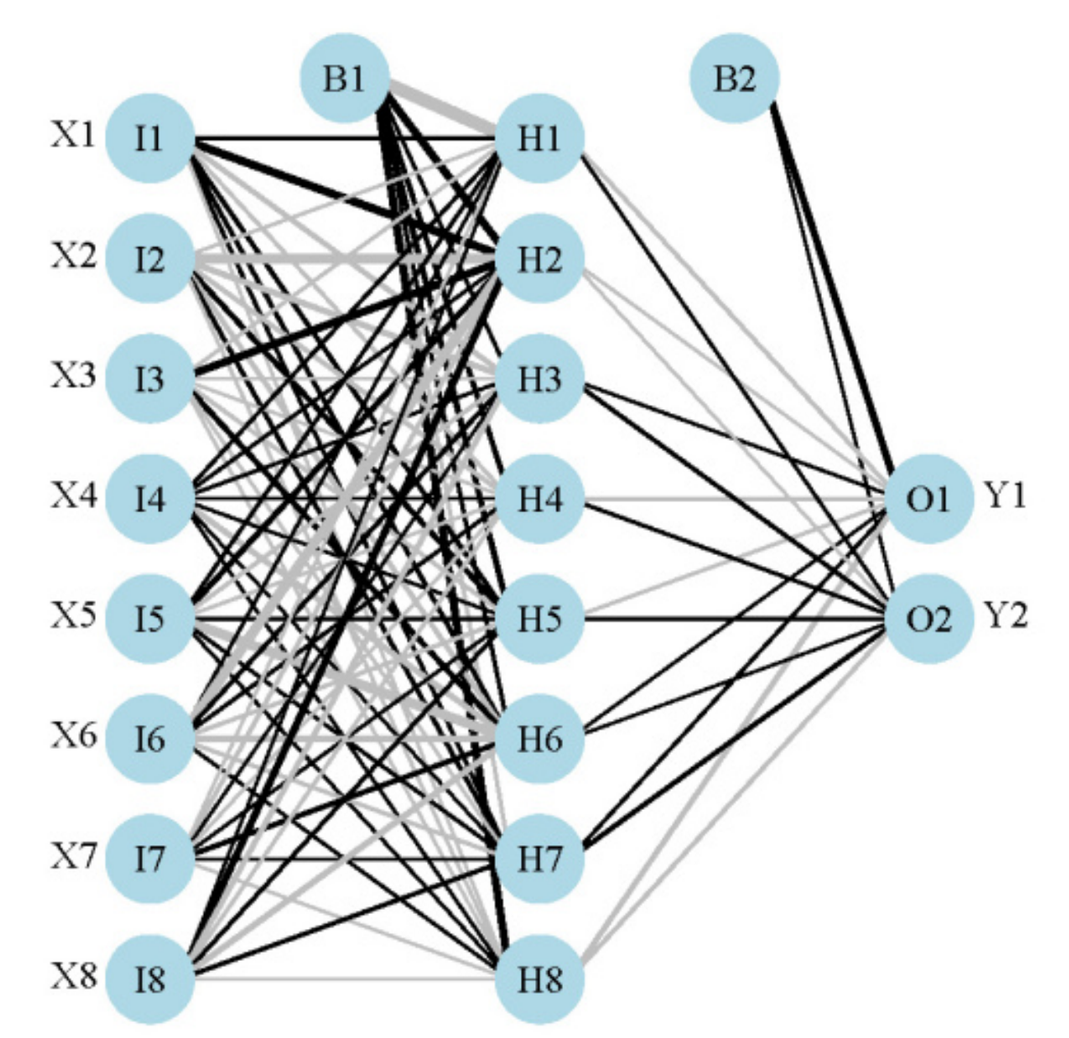
Handling missing data

Handling outliers/anomalies



Limitations of Machine Learning

#4: Machine Learning models can be hard to interpret



Machine Learning at Tableau

Smart @ Tableau

Human-Centered

- Augmentative.
- Empowering.
- Focus on things machines are good at and humans are not.
- Don’t interrupt me!

Contextual

- What is the task?
- Knowledge elicitation.
- What to infer and how?

Adaptive

- “Learning”
- More data = things can improve.
- The more we can learn, the better for everyone.

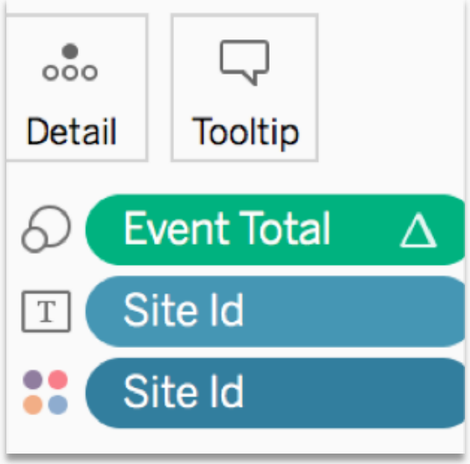
[T]he computer is the most remarkable tool that we've ever come up with. It's the equivalent of a bicycle for our minds.

~ Steve Jobs, *Memory & Imagination (1990)*

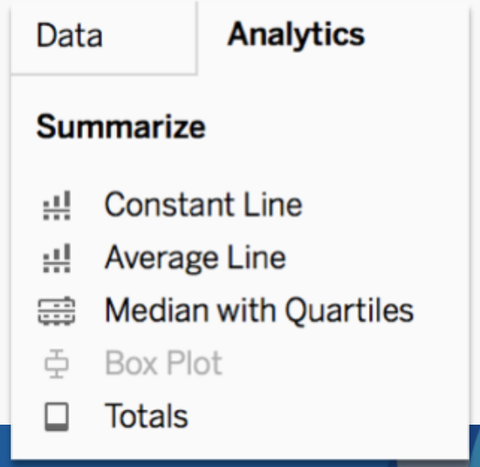
Ways to be Smart

Heuristics/rules

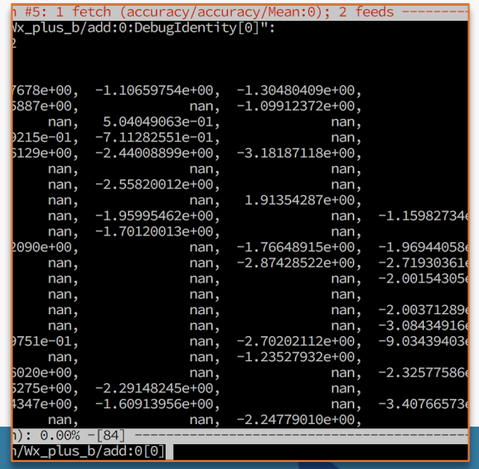
UX



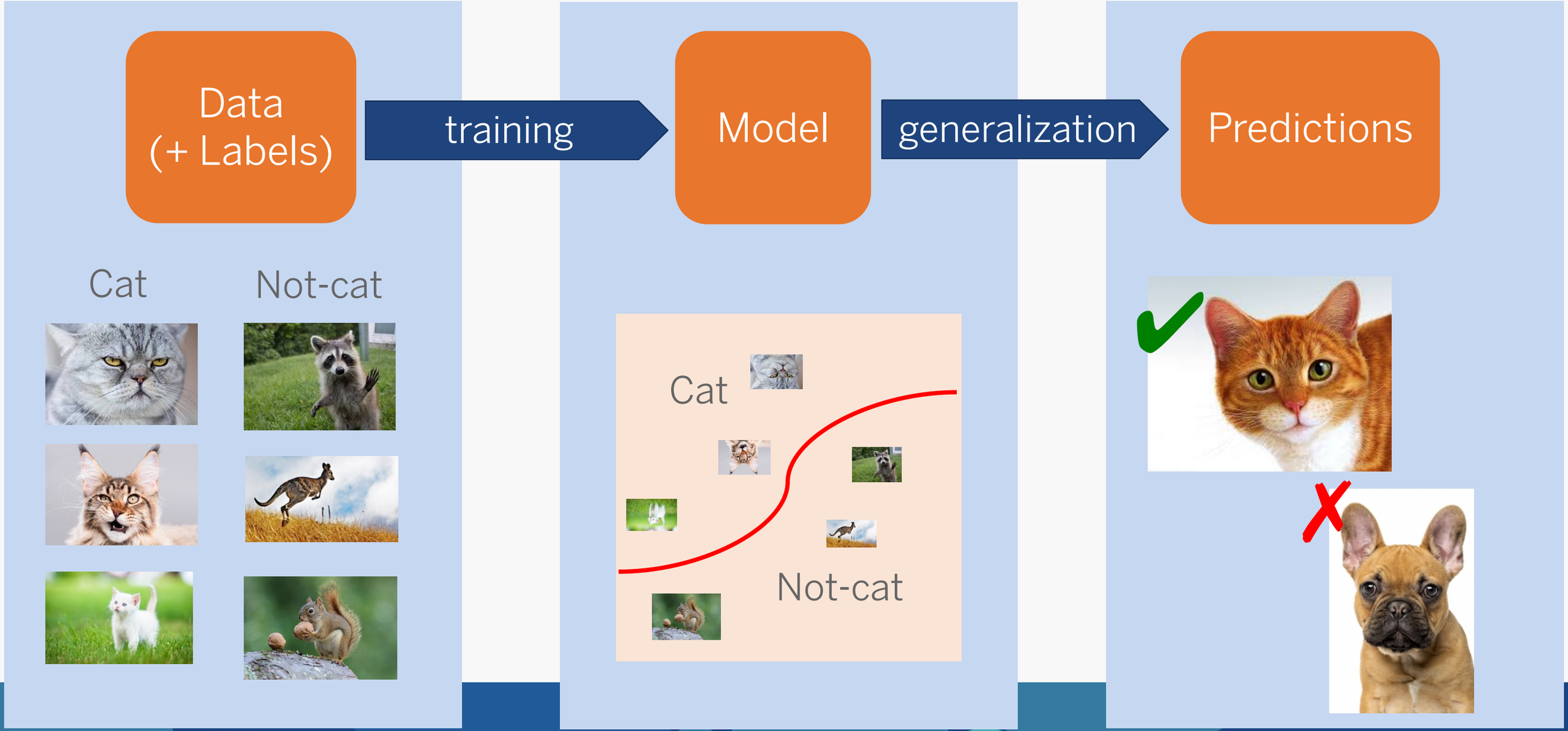
Data Science



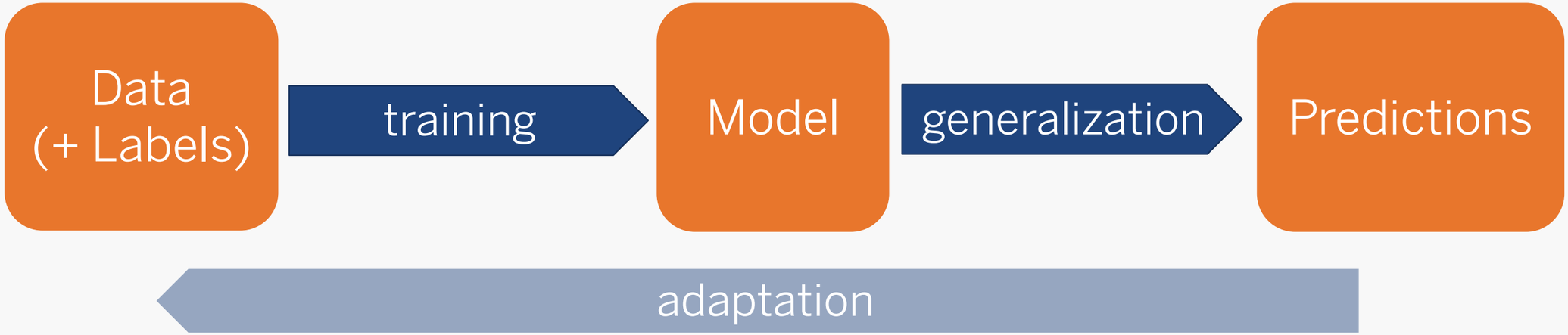
Machine Learning



Machine Learning: a very very short intro



Machine Learning Key Points

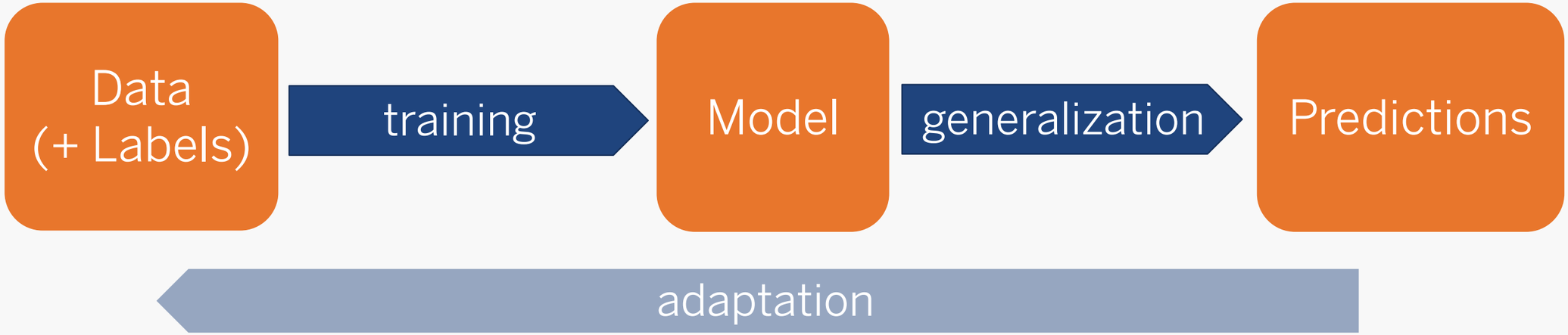


Data-Driven

Generalization

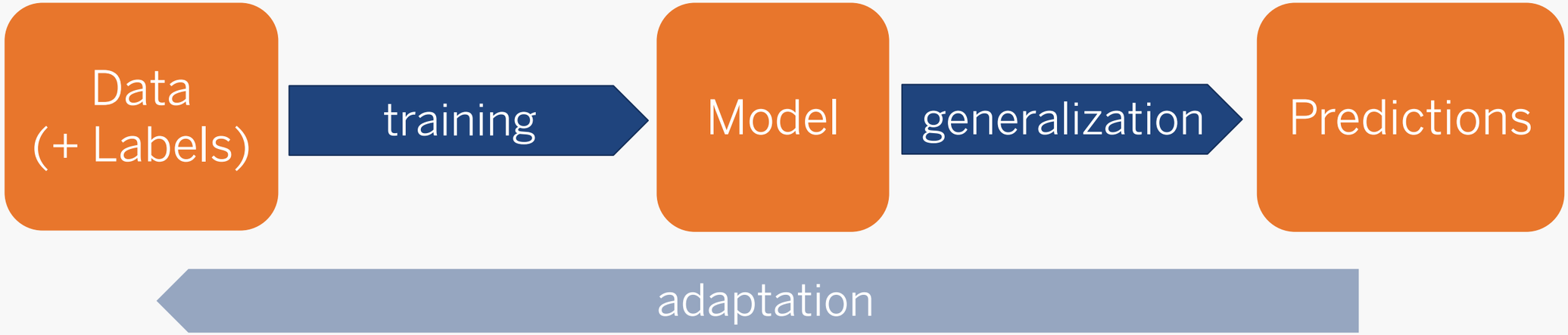
Adaptive

Machine Learning Key Points



Being data-driven, generalizable and adaptive doesn't just let us do things **better**, it enables us to do things that were previously **impossible** (or at least impractical).

Machine Learning Key Points



BUT, it's not going to solve all our problems

- Requires formally-specified problems
- Requires "good" data, and plenty of it

How to get from data + questions to features?

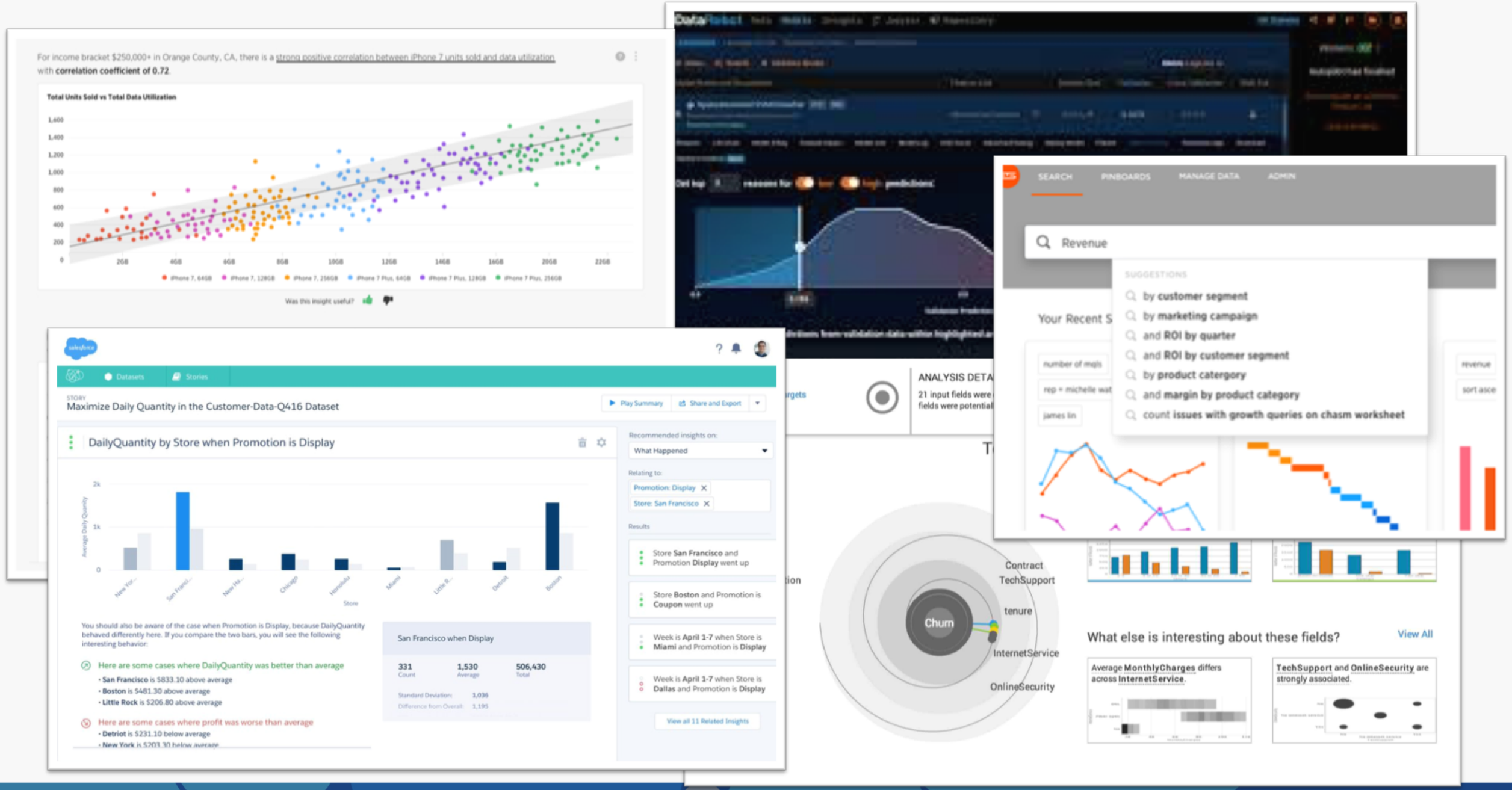
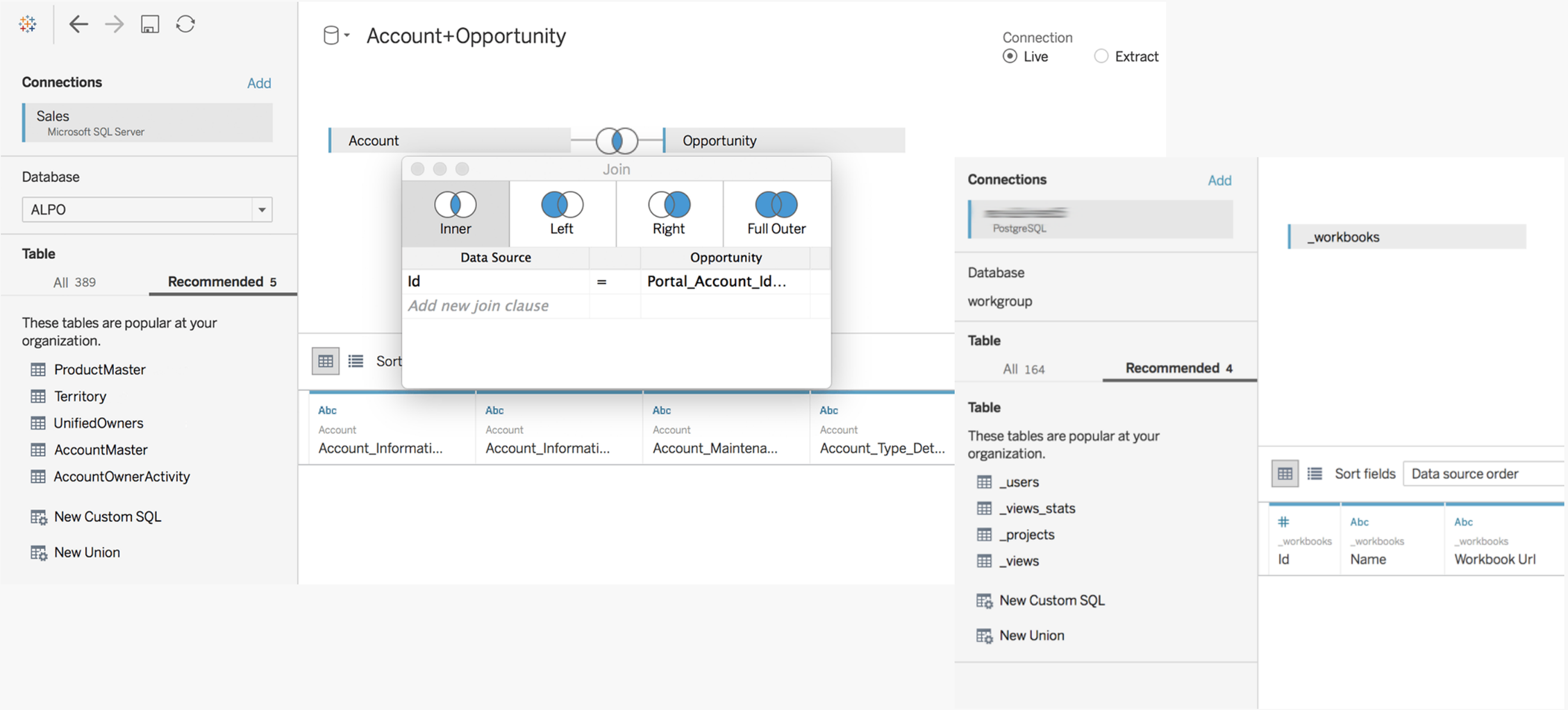

For income bracket $250,000+ in Orange County, CA, there is a strong positive correlation between iPhone 7 units sold and data utilization with correlation coefficient of 0.72.
Total Units Sold vs Total Data Utilization
Was this insight useful?
SEARCH
PINBOARDS
MANAGE DATA
ADMIN
Revenue
SUGGESTIONS
by customer segment
by marketing campaign
and ROI by quarter
and ROI by customer segment
by product catergory
and margin by product category
count issues with growth queries on chasm worksheet
Datasets
Stories
STORY
Maximize Daily Quantity in the Customer-Data-Q416 Dataset
Play Summary
Share and Export
DailyQuantity by Store when Promotion is Display
Recommended insights on:
What Happened
Relating to:
Promotion: Display
Store: San Francisco
Results
Store San Francisco and Promotion Display went up
Store Boston and Promotion is Coupon went up
Week is April 1-7 when Store is Miami and Promotion is Display
Week is April 1-7 when Store is Dallas and Promotion is Display
View all 11 Related Insights
You should also be aware of the case when Promotion is Display, because DailyQuantity behaved differently here. If you compare the two bars, you will see the following interesting behavior:
Here are some cases where DailyQuantity was better than average
San Francisco is $833.10 above average
Boston is $481.30 above average
Little Rock is $206.80 above average
Here are some cases where profit was worse than average
Detriot is $231.10 below average
New York is $203.30 below average
San Francisco when Display
331 Count
1,530 Average
506,430 Total
Standard Deviation: 1,036
Difference from Overall: 1,195
Churn
Contract
TechSupport
tenure
InternetService
OnlineSecurity
What else is interesting about these fields?
View All
Average MonthlyCharges differs across InternetService.
TechSupport and OnlineSecurity are strongly associated.

Tableau Recommendations



Recommendations

- What did we learn from working on recommendations?
- How did this help us think about ML? What decisions did we make?

Tableau File Data Server Help 100% Mon Jan 8 17:47

Tableau - Book1

Connect

To a File

- Microsoft Excel
- Text file
- JSON file
- PDF file
- Spatial file
- Statistical file
- More...

To a Server

- Tableau Server
- Oracle
- Amazon Redshift
- Google BigQuery
- Microsoft SQL Server
- More...

Saved Data Sources

- Sample - Superstore
- World Indicators

Open

Open a Workbook

Book3 [rec-e2e-...

Book3

Book2

Book1

Sample Workbooks

More Samples

Superstore

Regional

World Indicators

Discover

Training

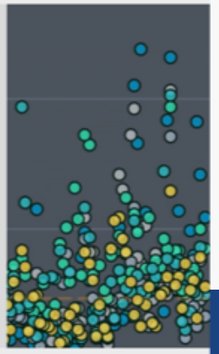
- Getting Started
- Connecting to Data
- Visual Analytics
- Understanding Tableau
- More training videos...

Resources

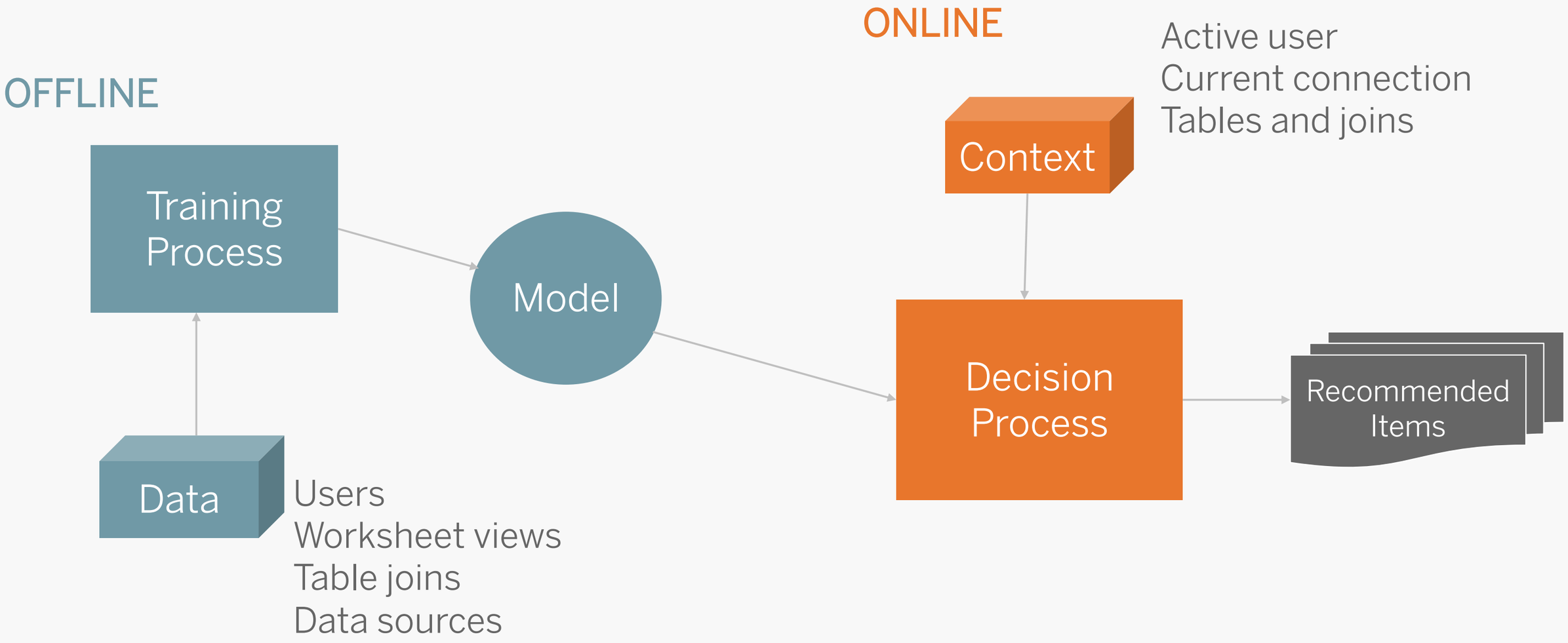
- Blog - Nominate the new class of Tableau Zen Masters!
- Tableau Conference - Register Now
- Forums

VIZ OF THE WEEK

2017: Year of the Alt-Coin →



Recommendations



Abstract Machine Learning system

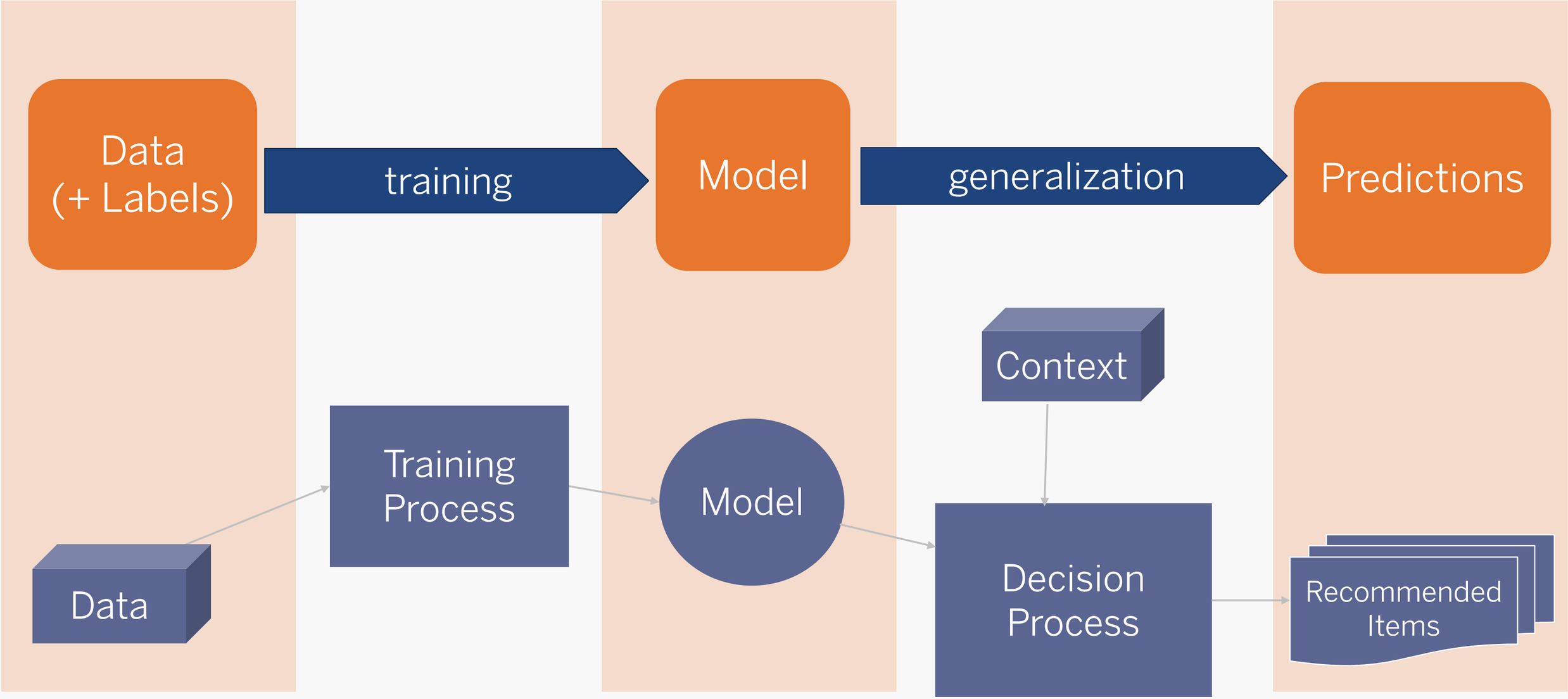


Tableau Recommendations system

Abstract Machine Learning system

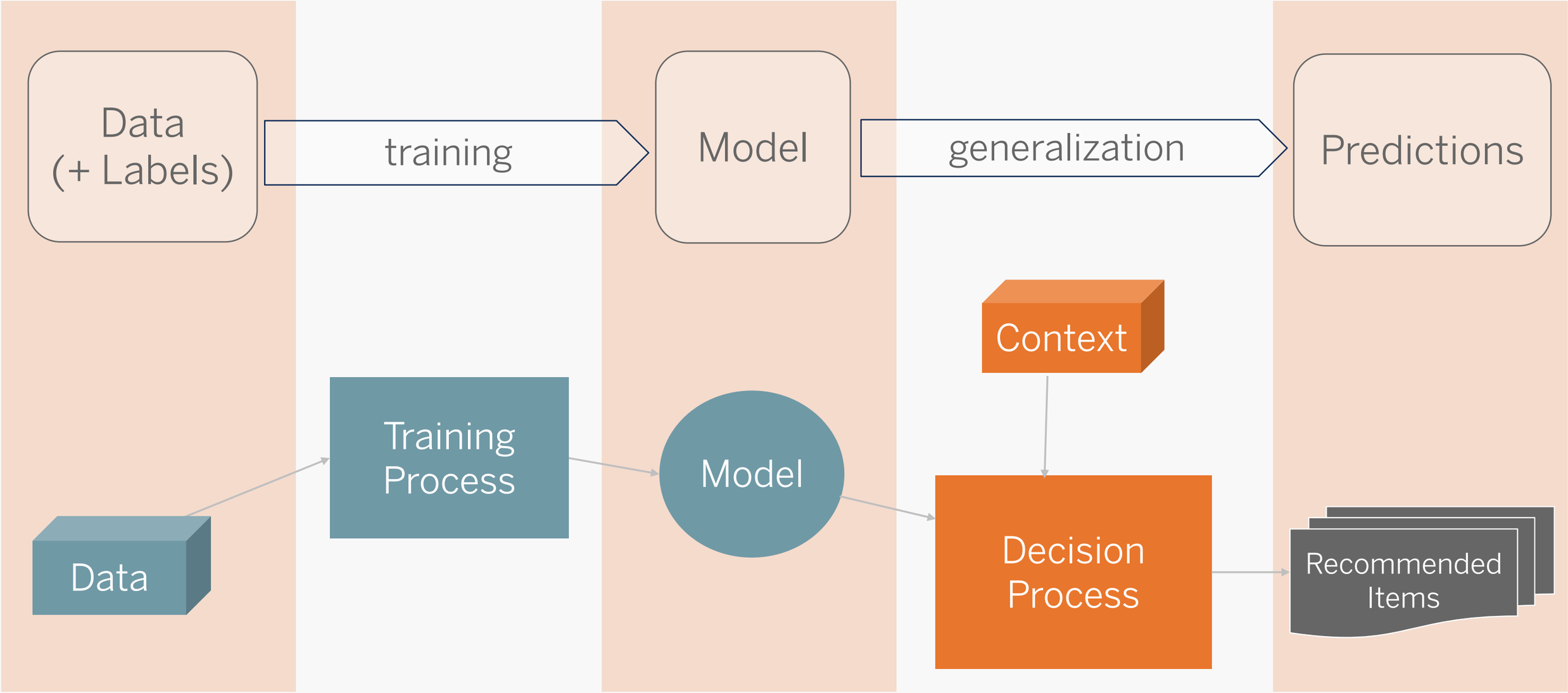
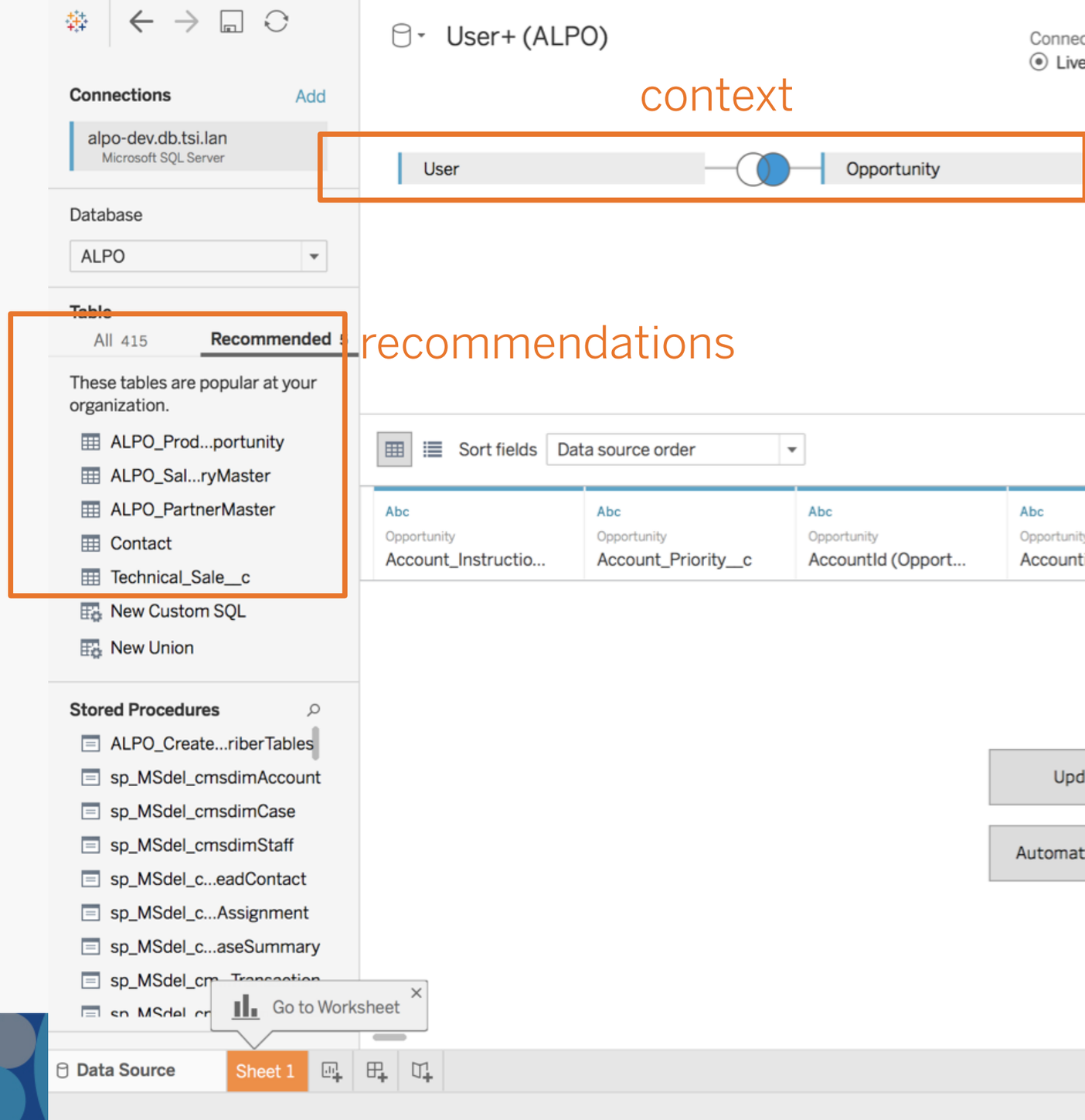


Tableau Recommendations system

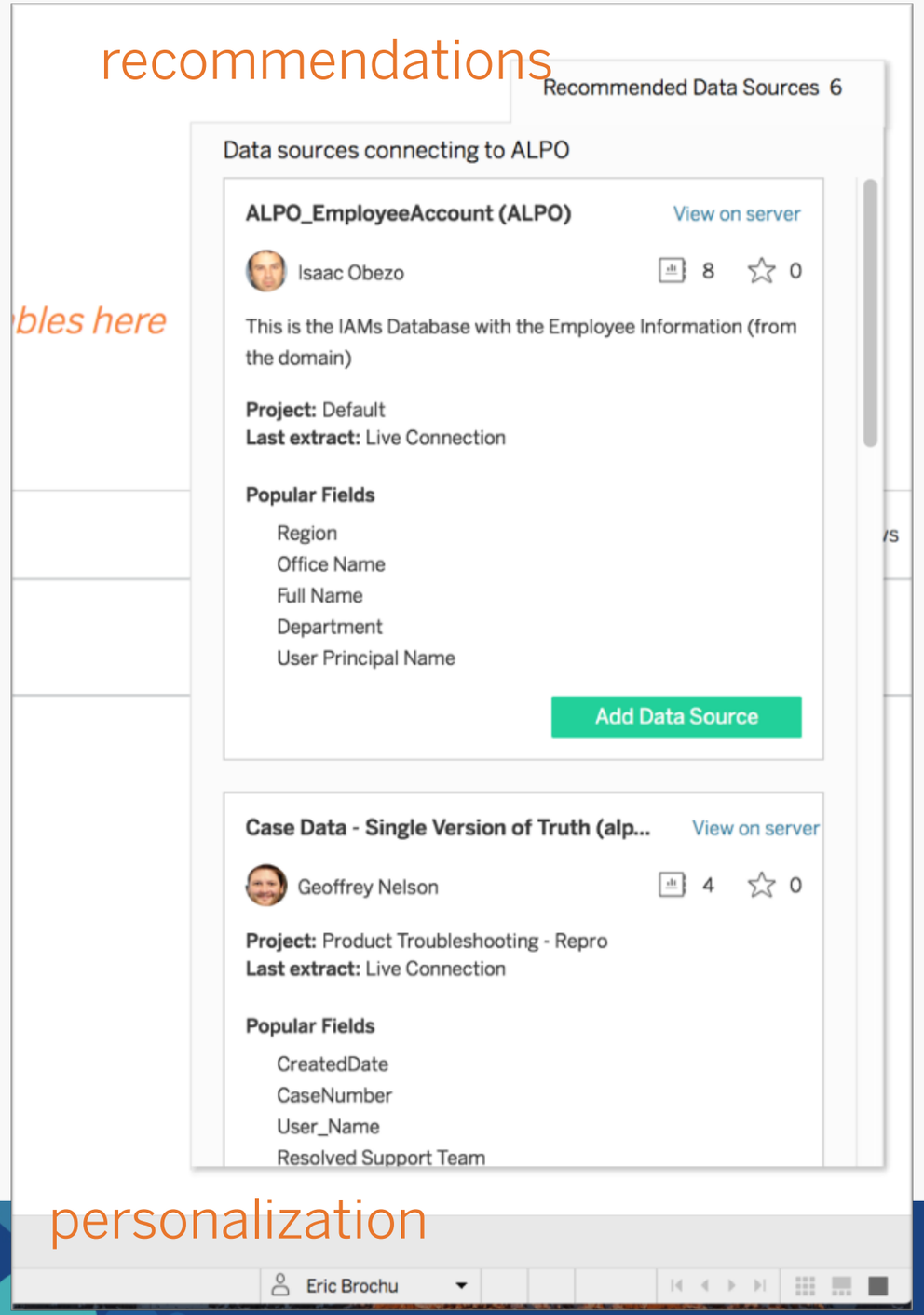
Table & Join Recommendations

- Surface popular tables and joins
- Training data = table join data
- Model = associations
- Prediction = popularity



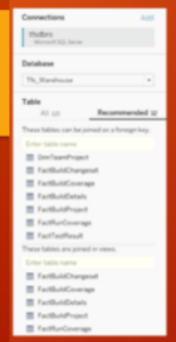
Data Source Recommendations

- Model = “People like you tend to like this.”
- Training data = user-workbook interactions
- Prediction = how likely user is to interact with item



6. Table join recommendations from foreign keys and database views

Xiaojian Wang

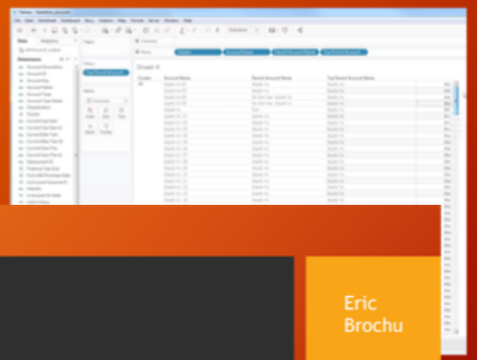


289-4749-be04-d51d71de6afc

1. Common data prep problem: **string deduplication**

Walther Maciel

- When data is typed in manually, especially by multiple individuals, it is bound to have inconsistencies in spelling and/or punctuation. This problem can waste many hours trying to coalesce all the different versions to a single one
- The approach that provided reasonable results in a timely manner was the token collision with double metaphone
- In a Nutshell: every string is tokenized by the double metaphone algorithm, which makes similar sounding strings



. KDA: Key drivers/influencers analysis (Live demo)

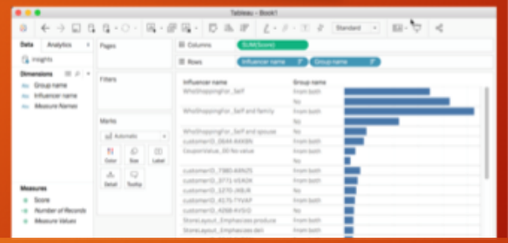
Rec Team

What is key driver analysis?

- KDA gives you the relative importance of an independent (driver) field as a predictor of a dependent (target) field. KDA looks for statistically significant linear relationships b/n drivers and targets. For continuous target variables, multivariate linear regression or variants (e.g. Kruskal) are common, for discrete target variables Chi-squared is common

What did we do?

- Suppose that the user is interested in comparing and contrasting drivers to sales in two different countries, or drivers for different marketing campaigns. In Tableau, the user may have a visualization of [Sa



2. Automatically Finding Joinable Datasources

David Mosimann

- An experiment to determine if, through data analysis, existing Tableau Server datasources can be automatically joined with the users current datasource
- Techniques using Bloom Filters were shown to be very promising
- Discussed with Anushka to determine applicability and the approach will be used in Maestro

3. Automatic Statistician

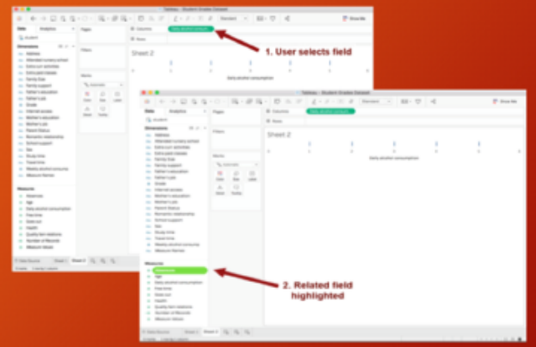
Eric Brochu

- Automatic Statistician is a tool that uses Machine Learning techniques to build an artificial intelligence system for data science
- It analyzes data points in a se combination of functions tha interesting attributes of the s periodicity, noise, change po communicates it to an analys
- There is some existing code f Machine Learning Group, wh
- This hackthon was to extend show how such a system cou help explain time series data Bayesian model selection, wh Eric's PhD thesis

5. Smart Pills

Vishaal Kapoor

- Smart Pills highlights fields related to those that you're working with so that you can get straight to the most insightful visualizations of your data. As you drag fields to the rows and columns shelves, relevant fields that you have not yet considered are highlighted so that you know there is an interesting relationship to explore.
- In real-time, an association rule mining algorithm that is discovering statistically significant associations between the fields you're considering and those you haven't yet looked at
- The workflow has similarities to Recommended Table & Joins, but for fields instead of Tables



leau.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=1bce9781-70bf-4f90-86bc-4581e161acaa

4. Content Model Graph Search

Mark Siegel

- Storing the full TWB/TDS content in a graph repository
  - Each node is an element (datasource, connection, field, worksheet)
  - Experimented with Titan and Neo4j
- Direct relationships between published or shared artifacts
  - Example: A direct reference from workbooks to their published datasource
  - Example: Connections to the same DB are shared
- Content searches in this model can specify exactly where they want to search
  - Example: find as a field vs find in title
- Content searches in this model can depict connections
  - Example: Ensure that all terms are descendants of a single worksheet

Video not recorded

Open Discussion with Panel

Discussion #1:

When it comes to machine learning for my organization, where do we start? How do I know it is the right investment? What should my team be focused on?

Discussion #2:

How do you see natural language processing (NLP) playing into machine learning?

Discussion #3:

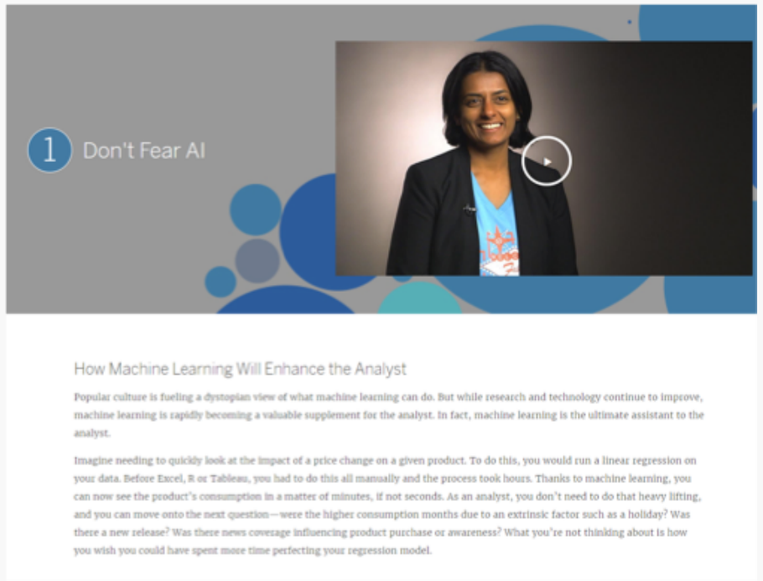
What gets you most excited when we can combine the power of machine learning and data analytics?

Additional Resources

BI Trends Additional Resources

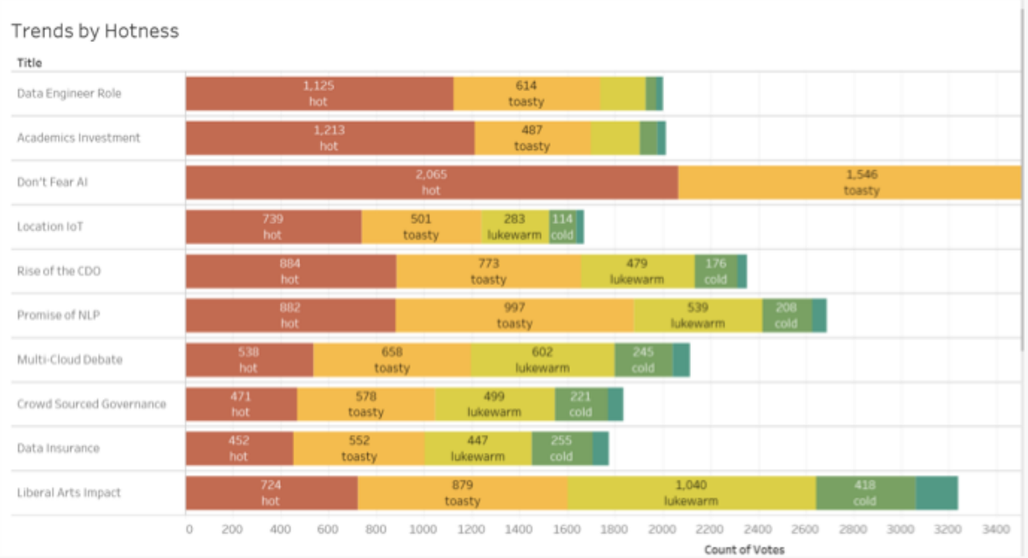
2018 BI Trends – Full Report

http://tabsoft.co/2A2Mwhm



BI Trends Voting Visualization

https://alpo/authoring/2018Top10BITrendsVotes/Votes?#1



BI Trends Additional Resources

2018 BI Trends – Facebook Live w/ Chief Product Officer Francois Ajenstat
http://tabsoft.co/2D653LF

We're chatting with Tableau Chief Product Officer Francois Ajenstat,...
16K views · November 16, 2017

BI Trends "11th Trend" – Facebook Live
http://tabsoft.co/2mlkBzV

The Tableau community has spoken! Find out what they chose...
12K views · December 13, 2017

Thank you