



The Care and Feeding of Data Scientists: Concrete Tips for Retaining Your Data Science Team

Michelangelo D'Agostino
Senior Director, Data Science
@MichelangeloDA

September 13, 2018

Data Science Retention is a Real Problem

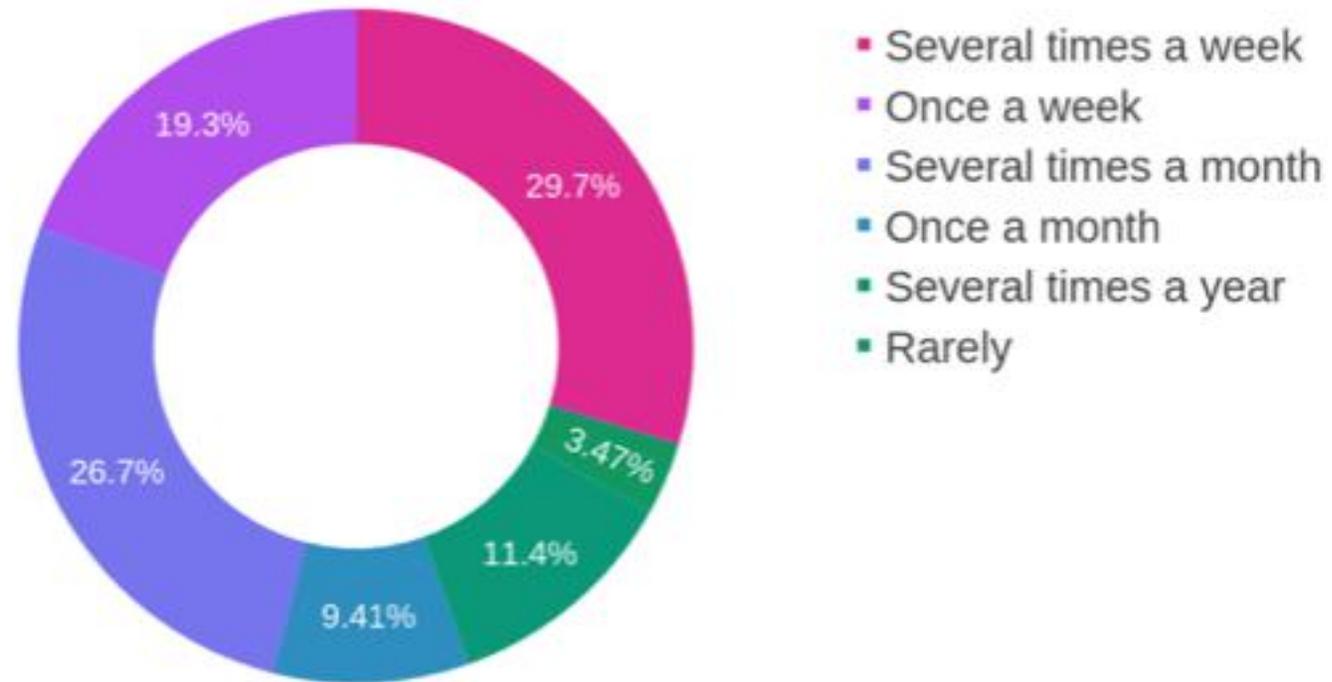
#StrataData



Strata
DATA CONFERENCE

Data Science Retention is a Real Problem

How often do you get contacted for new job opportunities?



Data from "Data Scientist Report 2018" by FigureEight

Data Science Retention is a Real Problem

How often do you get contacted for new job opportunities?



Data from "Data Scientist Report 2018" by FigureEight

Data Science Retention is a Real Problem



Data from <https://www.kdnuggets.com/2015/09/how-long-data-scientists-stay-jobs.html>,
<https://www.kdnuggets.com/polls/2015/how-long-stay-analytics-data-science-job.html>

Data Science Retention is a Real Problem

Once you lose someone, it takes a long time to find a replacement...

Job Category	Average Time to Fill Position (Days)
All	45
Data-Driven Decision Makers	48
Functional Analysts	40
Data Systems Developers	50
Data Analysts	38
Data Scientists & Advanced Analysts	46
Analytics Mangers	43

Data from <https://www.forbes.com/sites/louiscolombus/2017/05/13/ibm-predicts-demand-for-data-scientists-will-soar-28-by-2020/#1f8f99317e3b>

Data Science Retention is a Real Problem

Once you lose someone, it takes a long time to find a replacement...

Job Category	Average Time to Fill Position (Days)
All	45
Data-Driven Decision Makers	48
Functional Analysts	40
Data Systems Developers	50
Data Analysts	38
Data Scientists & Advanced Analysts	46
Analytics Mangers	43

Data from <https://www.forbes.com/sites/louiscolombus/2017/05/13/ibm-predicts-demand-for-data-scientists-will-soar-28-by-2020/#1f8f99317e3b>

“Nationally, we have a shortage of 151,717 people with data science skills.”

LinkedIn Workforce Report, August 2018 (<https://economicgraph.linkedin.com/resources/linkedin-workforce-report-august-2018>)



- Who Am I?
- Organizational Structure and Leadership for Data Science Teams
- Infrastructure and Tools
- To Agile or Not To Agile?
- Continuing Education for Data Scientists



Who Am I?



**Me circa
2007: more
science
than data...**

Me in November 2012...

Predictive Modeling and Data Mining Scientists/Analysts

 Tweet 173

analyze the campaign data to guide election strategy and develop quantitative, actionable insights that drive our decision-making. Looking for people at both the senior and junior level to join the campaign Analytics Dept through November 2012.

Company: Obama 2012 Presidential Campaign
Location: Chicago, IL
Web: www.barackobama.com



The Obama for America Analytics Department analyzes the campaign's data to guide election strategy and develop quantitative, actionable insights that drive our decision-making. Our team's products help direct work on the ground, online and on the air.

We are looking for Predictive Modeling/Data Mining Scientists and Analysts, at both the senior and junior level, to join our department through November 2012 at our Chicago Headquarters. We are a multi-disciplinary team of statisticians, predictive modelers, data mining experts, mathematicians, software developers, general analysts and organizers - all striving for a single goal: re-electing President Obama.

Using statistical predictive modeling, the Democratic Party's comprehensive political database, and publicly available data, modeling analysts are charged with predicting the behavior of the American electorate. These models will be instrumental in helping the campaign determine which voters to target for turnout and persuasion efforts, where to buy advertising and how to best approach digital media.

Our Modeling Analysts will dive head-first into our massive data to solve some of our most critical online and offline challenges. We will analyze millions of interactions a day, learning from terabytes of historical data, running

Strata
DATA CONFERENCE





We ran the first individualized presidential campaign.



- After 2012, I started the data science team at Braintree/Venmo, which was acquired by PayPal.



- At Civis, I ran the Data Science R&D team—20 top notch data scientists responsible for software, algorithms, and direct client consulting for political organizations, non-profits, and Fortune 500's.



- Building a growing team of 7 working on some of the hardest problems in e-commerce

Over my career, I've interviewed hundreds of data scientists, hired ~25, and have lost 2.

Amazon Prime for the other half of the internet: Our 6 million members get free two-day shipping, returns, and deals across a growing network of 140+ retailers.



#StrataData

 **SHOPRUNNER**

Strata
DATA CONFERENCE

MEN'S SHOES / SNEAKERS / MEN'S GRANDPRØ SPECTATOR SNEAKER

NEXT



ROLL-OVER TO ZOOM

#COLEHAAN



Men's GrandPrø Spectator Sneaker

\$170.00

Color: Woodbury Handstain



Size:

Width:

SELECT 

M 

[Size Guide](#)

ADD TO BAG

[Select a size/width](#)

[Add To Wishlist](#)

 GIVE A GIFT

FIND IN STORE

 SHOPRUNNER free 2-day shipping & free returns
[learn more](#) | [sign up](#)

[Write a Review](#) 

[Ask a Question](#) 

MEN'S SHOES / SNEAKERS / MEN'S GRANDPRØ SPECTATOR SNEAKER

NEXT



ROLL-OVER TO ZOOM

#COLEHAAN



Men's GrandPrø Spectator Sneaker

\$170.00

Color: Woodbury Handstain



Size:

Width:

SELECT

M

Size Guide

ADD TO BAG

Select a size/width

Add To Wishlist

GIVE A GIFT

FIND IN STORE

SHOPRUNNER free 2-day shipping & free returns
learn more | sign up

Write a Review

Ask a Question

COLE HAAN Women Men Sale GRAND Gift Guide Search

COLE HAAN
— has a limited time offer for you —
GET YOUR ORDER BY
Wed Nov. 15
with **FREE 2-DAY SHIPPING** by **SHOPRUNNER**

Sign up for ShopRunner and enjoy unlimited free 2-day shipping and free returns at 140+ stores.

1 Choose 1 of 3 ways to sign up FOR FREE

- COLE HAAN Free when you spend \$200 today
-  Free with your eligible card
-  Free with your PayPal account

2 Activate your Shoprunner account

First name Last name

email

password

[Activate this offer](#)

By clicking activate this offer, you are agreeing to the [Terms of Service](#) and [Privacy Policy](#)

ROLL-OVER TO ZOOM

Write a Review

#StrataData

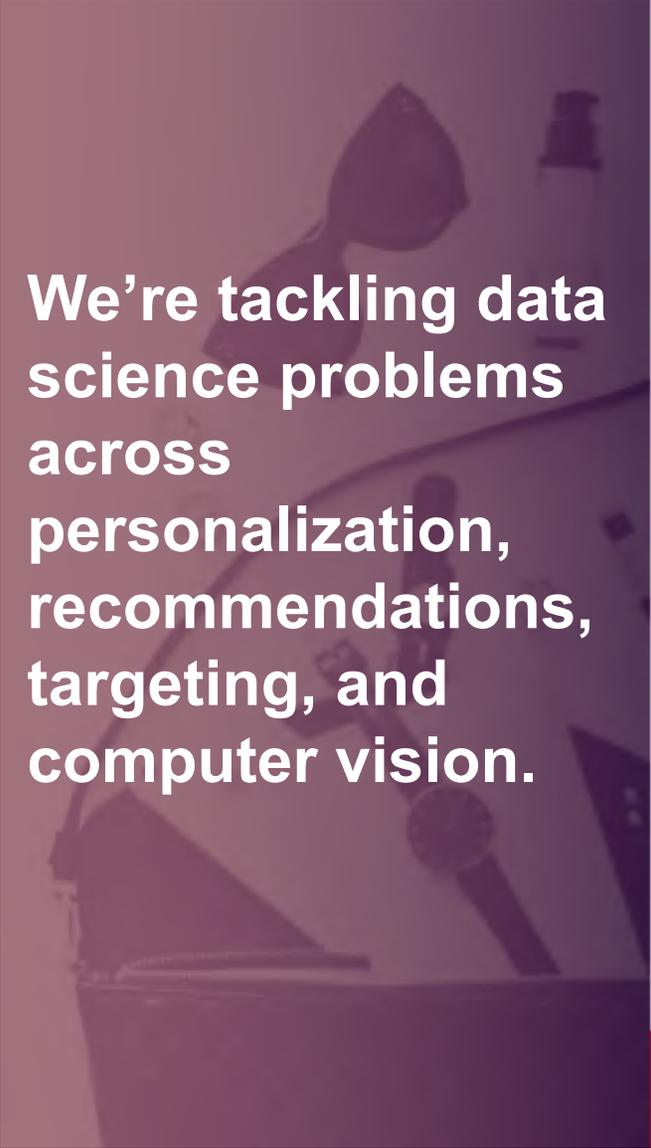


Strata
DATA CONFERENCE



Fundamentally,
we're a data and
technology
company.

- **We're collecting many terabytes of data a month**
 - granular, sku-level browsing and purchase data across our network of retailers
 - product catalog feeds of prices, inventories, product images, full text descriptions, etc. of ~10mm sku variants
- **We use this data to better personalize both our member and non-member experience.**



We're tackling data science problems across personalization, recommendations, targeting, and computer vision.

- **How do we intelligently surface retailers and brands to our members based on their past browsing and purchase behavior?**
- **How do we recommend the right product at the right time?**
- **What can we learn by applying computer vision to our corpus of ~10mm images, or NLP to the product page descriptions?**
- **Algorithms to support a new consumer mobile app and Chrome browser extension**



Our Data Science Tech Stack



docker



TensorFlow



WE KNOW A THING OR TWO

BECAUSE WE'VE SEEN A THING OR TWO



Organizational Structure and Leadership for Data Science Teams



Good Leaders
and the Right
Structure
Create Happy
Teams

- **Data science is fundamentally different from engineering**

- Our work is less well-defined, and often more experimental, iterative, and end-to-end in nature
- Data science teams do best with a data leader rather than an engineer or a product leader



Good Leaders
and the Right
Structure
Create Happy
Teams

- **Data science is fundamentally different from engineering**
 - Our work is less well-defined, and often more experimental, iterative, and end-to-end in nature
 - Data science teams do best with a data leader rather than an engineer or a product leader
- **Data science is inherently cross-functional, but resist fully dissolving your centralized data science team**
 - Data scientists get lonely without other data scientists to talk to
 - Hire your first data scientists in a pair, if possible



Good Leaders
and the Right
Structure
Create Happy
Teams

- **Data science is fundamentally different from engineering**
 - Our work is less well-defined, and often more experimental, iterative, and end-to-end in nature
 - Data science teams do best with a data leader rather than an engineer or a product leader
- **Data science is inherently cross-functional, but resist fully dissolving your centralized data science team**
 - Data scientists get lonely without other data scientists to talk to
 - Hire your first data scientists in a pair, if possible
- **Socialize data science with brown-bag talks about the basics of data science and what the team is doing**





Good Leaders
and the Right
Structure
Create Happy
Teams

- **Leave the reporting to the analytics or BI team**
 - This work is crucially important
 - Maslow's Hierarchy of Analytics
 - But it's not what data scientists sign up for 100% of the time



Good Leaders
and the Right
Structure
Create Happy
Teams

- **Leave the reporting to the analytics or BI team**
 - This work is crucially important
 - Maslow's Hierarchy of Analytics
 - But it's not what data scientists sign up for 100% of the time
- **Train your data scientist managers, and make sure to create a technical promotion track so you don't force them to become people managers**



Infrastructure and Tools



**Data Scientists
Will Leave If
They Don't
Have the Right
Tools To Do
Their Jobs**

- Data science work is **inherently experimental and elastic**, and it demands a certain set of tools



Data Scientists Will Leave If They Don't Have the Right Tools To Do Their Jobs

- Data science work is **inherently experimental and elastic**, and it demands a certain set of tools
- **Scalable infrastructure**
 - Set up **dedicated cloud provider accounts** to avoid IT bottlenecks when booting up bigger servers and clusters
 - If you're not cloud-based, forget it
 - If your data scientists have to work on their laptop exclusively or wait a few weeks to get a server from IT, they will leave



**Data Scientists
Will Leave If
They Don't
Have the Right
Tools To Do
Their Jobs**

▪ **Collaborative, interactive, and exploratory data science platforms**

- Domino Data Labs
- Databricks
- Mode Analytics
- RStudio Connect
- Civis Data Science Platform





**Data Scientists
Will Leave If
They Don't
Have the Right
Tools To Do
Their Jobs**

- **Collaborative, interactive, and exploratory data science platforms**

- Domino Data Labs
- Databricks
- Mode Analytics
- RStudio Connect
- Civis Data Science Platform



- **The cutting edge is happening in open source software—R, python, and Spark**



To Agile or Not to Agile?

Manifesto for Agile Software Development

“We are uncovering better ways of developing software by doing it and helping others do it. Through this work we have come to value:

- **Individuals and interactions** over processes and tools
- **Working software** over comprehensive documentation
- **Customer collaboration** over contract negotiation
- **Responding to change** over following a plan



That is, while there is value in the items on the right, we value the items on the left more.”

<http://agilemanifesto.org>

#StrataData

 SHOPRUNNER

Strata
DATA CONFERENCE

Manifesto for Agile Software Development

“We are uncovering better ways of developing software by doing it and helping others do it. Through this work we have come to value:

- **Individuals and interactions** over processes and tools
- **Working software** over comprehensive documentation
- **Customer collaboration** over contract negotiation
- **Responding to change** over following a plan



That is, while there is value in the items on the right, we value the items on the left more.”

“Deliver working software frequently, from a couple of weeks to a couple of months, with a preference to the shorter timescale.”

“Business people and developers must work together daily throughout the project.”

“Simplicity—the art of maximizing the amount of work not done—is essential.”

<http://agilemanifesto.org>

#StrataData

 **SHOPRUNNER**

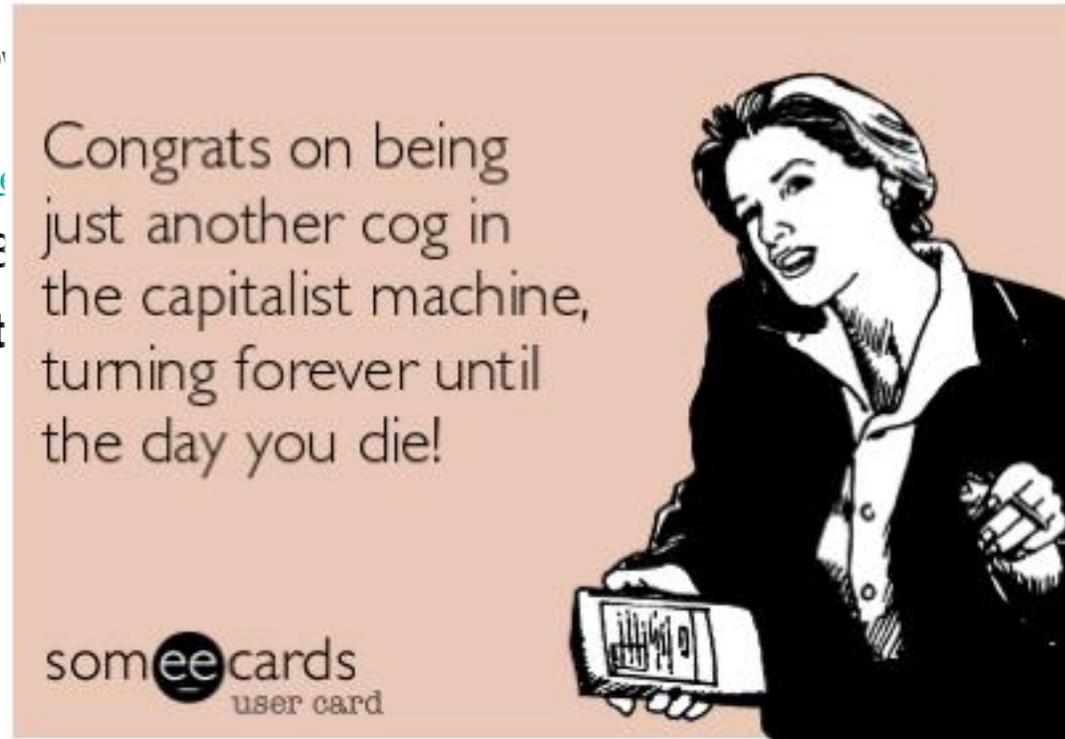
Strata
DATA CONFERENCE

Agile in Practice

- **Agile roles:** Scrum Master, Product Owner, etc.
 - “If the Product Owner is captain of the ship, then the Scrum Master is first mate. The Scrum Master is responsible for crew welfare and making sure team members follow protocol.” (<https://redbooth.com/blog/main-roles-agile-team>)
- **Agile meetings:** Backlog Grooming, Sprint Planning, Standups, Retros
- Tasks are **estimated**, **velocity** and “**sprint burndown**” are measured

Agile in Practice

- **Agile roles:** Scrum Master, Product Owner, etc.
 - “If the Product Owner and Scrum Master is not following the protocol.” (<https://r>)
- **Agile meetings:** Daily Scrum
- Tasks are **estimated**



first mate. The members follow



MichelangeloDAgostin

@MichelangeloDA



Question for my data scientist friends out there. How do you feel about "agile"?

27% It's great for DS teams

73% Agile mindset > process

0% Hate everything about it

11 votes • Final results

9:22 AM - 24 Aug 2018



Agile Mindset vs. Agile Ritual and Process

- **“Responding to change over following a plan”**
 - The data science lifecycle is iterative and constantly changing depending on what you find in the data or how early model results look.
 - Your data scientists need to be able to go where the data leads them, and they need to have the freedom to explore new ideas iteratively.

Agile Mindset vs. Agile Ritual and Process

- **“Responding to change over following a plan”**
 - The data science lifecycle is iterative and constantly changing depending on what you find in the data or how early model results look.
 - Your data scientists need to be able to go where the data leads them, and they need to have the freedom to explore new ideas iteratively.
- **“Deliver working software frequently, from a couple of weeks to a couple of months, with a preference to the shorter timescale.”**
 - Apps, internal visualizations, and incremental presentations are essential to getting buy-in for data science work.

Agile Mindset vs. Agile Ritual and Process

- **“Business people and developers must work together daily throughout the project.”**
 - Data scientists must work hand-in-hand with the business to understand how data is generated, how models will be used, and what will provide the most value.

Agile Mindset vs. Agile Ritual and Process

- **“Business people and developers must work together daily throughout the project.”**
 - Data scientists must work hand-in-hand with the business to understand how data is generated, how models will be used, and what will provide the most value.
- **“Simplicity—the art of maximizing the amount of work not done—is essential.”**
 - Deliver a simple logistic regression baseline before going off in a cave for six months to build your deep learning model.

Agile Mindset vs. Agile Ritual and Process

- **“Business people and developers must work together daily throughout the project.”**
 - Data scientists must work hand-in-hand with the business to understand how data is generated, how models will be used, and what will provide the most value.
- **“Simplicity—the art of maximizing the amount of work not done—is essential.”**
 - Deliver a simple logistic regression baseline before going off in a cave for six months to build your deep learning model. (Actually, don't ever do that.)





Continuing Education for Data Scientists

FOMO in Data Science is Real

It's cool, I'll just sit here
nursing my FOMO.



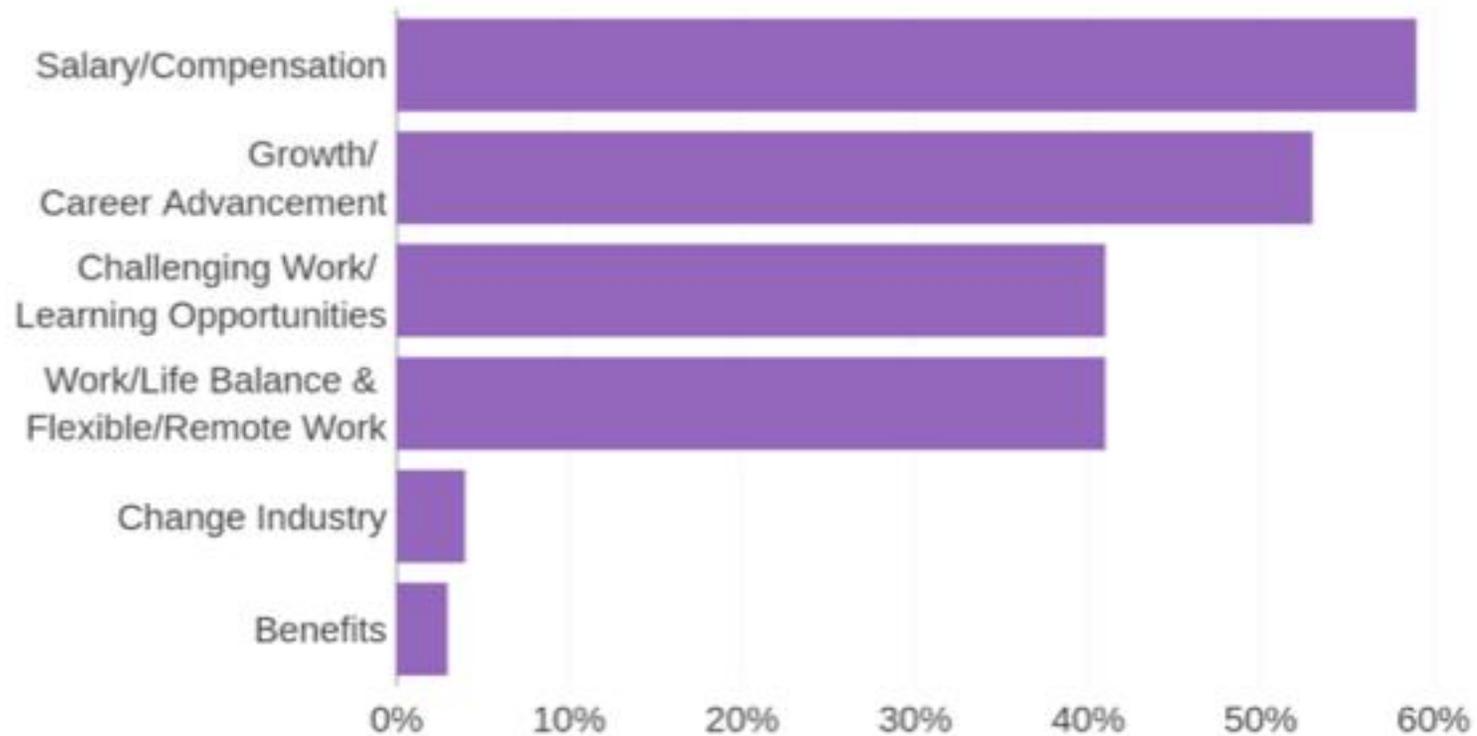
someecards
user card

FOMO [FOE-MOE]

NOUN

ANXIETY THAT AN EXCITING OR INTERESTING EVENT
MAY CURRENTLY BE HAPPENING ELSEWHERE.

What are the top two factors that would motivate you to change jobs?



Data from <https://www.burtchworks.com/2018/06/18/survey-results-what-motivates-analytics-pros-data-scientists-to-change-jobs/>



FOMO Is Real:
“My Company
Isn’t Doing Any
Cool Machine
Learning Stuff”

- Hire data scientists who care about having a **real, measurable impact on your business**, not just using the newest, shiniest tools



FOMO Is Real:
“My Company
Isn’t Doing Any
Cool Machine
Learning Stuff”

- Hire data scientists who care about having a **real, measurable impact on your business**, not just using the newest, shiniest tools
- Institute a **Journal Club**
 - very informal—can be done over lunch
 - people read a journal article or blog post and discuss it
 - open up to any interested party
 - great cross-pollination
 - “The Morning Paper”: <https://blog.acolyer.org/>



FOMO Is Real:
“My Company
Isn’t Doing Any
Cool Machine
Learning Stuff”

- Hire data scientists who care about having a **real, measurable impact on your business**, not just using the newest, shiniest tools
- Institute a **Journal Club**
 - very informal—can be done over lunch
 - people read a journal article or blog post and discuss it
 - open up to any interested party
 - great cross-pollination
 - “The Morning Paper”: <https://blog.acolyer.org/>
- **Set the example: read widely and share articles on Slack**
 - <http://bit.ly/DSNewsletters>



FOMO Is Real:
“My Company
Isn’t Doing Any
Cool Machine
Learning Stuff”

Useful Data Science Newsletters

- The Morning Paper
- Data Machina
- Data Science Weekly
- O'Reilly Data Newsletter
- O'Reilly AI Newsletter
- The Wild Week in AI
- Deep Learning Weekly
- The Data Science Roundup
- The Analytics Dispatch
- KDNuggets News
- PyData Newsletter
- FlowingData
- DataTau
- DataElixir

- <http://bit.ly/DSNewsletters>

FOMO Is Real:
“My Company
Isn’t Doing Any
Cool Machine
Learning Stuff”

- Hire data scientists who care about having a **real, measurable impact on your business**, not just using the newest, shiniest tools
- Institute a **Journal Club**
 - very informal—can be done over lunch
 - people read a journal article or blog post and discuss it
 - open up to any interested party
 - great cross-pollination
 - “The Morning Paper”: <https://blog.acolyer.org/>
- **Set the example: read widely and share articles on Slack**
 - <http://bit.ly/DSNewsletters>
- **DS Movie Night**





FOMO Is Real:
“My Company
Isn’t Doing Any
Cool Machine
Learning Stuff”

- **Allow each team member to take a quarterly hack week**
 - Hack days and two-day company hackathons don’t work well for data science
 - Explore a new software package or language, a new statistical technique or tool, or do something with the company’s data that they just haven’t had time to do
 - Must have a concrete outcome planned in advance—an application, a software prototype, a notebook documenting the research process for others to read, a blog post
 - Needs tracking and accountability—plan in JIRA and daily check-in on progress with a buddy
 - **Mandate an end of hack week presentation to the rest of the team**

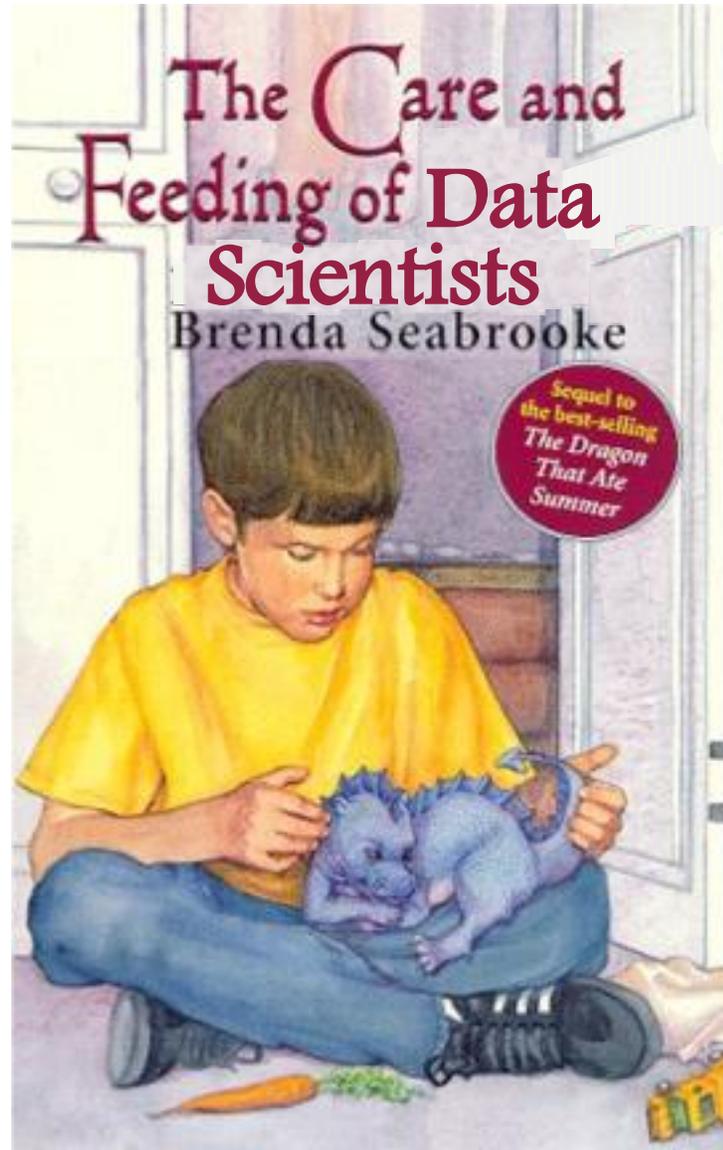
FOMO Is Real:
“My Company
Isn’t Doing Any
Cool Machine
Learning Stuff”

- **Have a conference policy**

- Encourage abstract submission and pay if it’s accepted
- Fund one conference a year if it’s in your budget
- Otherwise, use Journal Club as an opportunity to watch talk videos from conferences



You can keep your data scientists happy and productive.



The background of the slide is a sepia-toned photograph. It shows a close-up of a hand holding a pen, poised to write on a document. The document has some faint, illegible text on it. The overall tone is professional and focused.

Thanks! Questions?

[@MichelangeloDA](#)
mdagost@gmail.com

Also, we're hiring....