

Building a Data Science Idea Factory

How to prioritize the portfolio of a large, diverse, and opinionated data science team

Strata Data Conference
San Jose, California
March 7, 2018

Katie Malone
Director of Data Science R&D
Civis Analytics
[@multiarmbandit](#) &
[lineardigressions.com](#)

Skipper Seabold
Director of Data Science R&D
Civis Analytics
[@jseabold](#)



We help organizations integrate data science into their business so they can find truth and take action

We revolutionized the presidential campaign process. Now we're revolutionizing data science for businesses.

- Founded in 2013, Civis Analytics is a data science technology and advisory company with offices in Chicago and Washington, D.C.
- We provide technology to operationalize data science for our own and our clients' data science teams.
- Today, Civis provides applied data science services and a platform for outcomes-oriented data science.



How our team created a balanced, organizationally-aligned data science roadmap

Many companies struggle in their efforts to become more data-driven because leaders fail to see the value that data science teams can provide and data science teams fail to see kind of value the business needs.

Preliminary	Early	Initial	Ongoing
Broad business alignment The data science team works to get a broad understanding of the overall business goals and the provide an honest assessment of the value that data science can deliver.	Deep technical assessment The data science team translates business goals into data science projects and develops a deep understanding of the risks involved.	Team & Organizational Consensus The data science team aligns around the goals and projects. They prioritize the projects into a roadmap, which is broadly communicated.	Execution & Continued Communication Armed with the business context and an understanding of the moving technical pieces and where things can go wrong, the projects get underway.

Our team and their challenges

This talk is a bit of a deep dive into our team, but the challenges generalize to other organizations. We know because we've worked with them and we've seen it in practice.

- Data science research and development team
 - Center of excellence model
 - Balance bottom-up science-driven ideas with business goals
- Civis Analytics
 - Data science technology and consulting
 - We work across industries and challenges



Leaders in organizations and stakeholders in analytics projects are thinking about the business objectives

- **Getting started with analytics**
 - How can I make better data-driven decisions?
 - I don't really know what data science is. I hear it's great though!



Leaders in organizations and stakeholders in analytics projects are thinking about the **business objectives**

- **Getting started with analytics**
 - How can I make better data-driven decisions?
 - I don't really know what data science is. I hear it's great though!
- **Realizing returns from investments in analytics**
 - How can I help my data scientists or data science teams understand our business objectives?
 - How can I have an active dialogue with my data scientists so we work together toward shared goals?



Data scientists need to balance methodological and technical excellence with practicality and usability

- **Doing great data science**
 - How, and when, should I get scientific feedback on my work from my peers?
 - What other great ideas are floating around the organization that I might be able to help with?



Data scientists need to balance methodological and technical excellence with practicality and usability

- Doing great data science
 - How, and when, should I get scientific feedback on my work from my peers?
 - What other great ideas are floating around the organization that I might want to help with?
- Making it relevant to the organization
 - If I have more autonomy than top-down direction, how do I ensure that my work has a big impact?
 - How do I advocate for projects that I think will have a big impact?



Managers of data scientists bridge the communication gap between stakeholders and data scientists

- Proving value up the org chart
 - How can I translate the business needs into a project roadmap for my team?
 - Everyone is happier when I proactively manage expectations with my boss, and can communicate the tradeoffs when we're making decisions.



Managers of data scientists bridge the **communication gap** between stakeholders and data scientists

- Proving value up the org chart
 - How can I translate the business needs into a project roadmap for my team?
 - Everyone is happier when I proactively manage expectations with my boss, and can communicate the tradeoffs when we're making decisions.
- **Keeping the team happy and productive**
 - What is the right balance on my team of skill development, R&D, and needing to get important things done?
 - Technology and data science moves really fast, my team knows more than me!



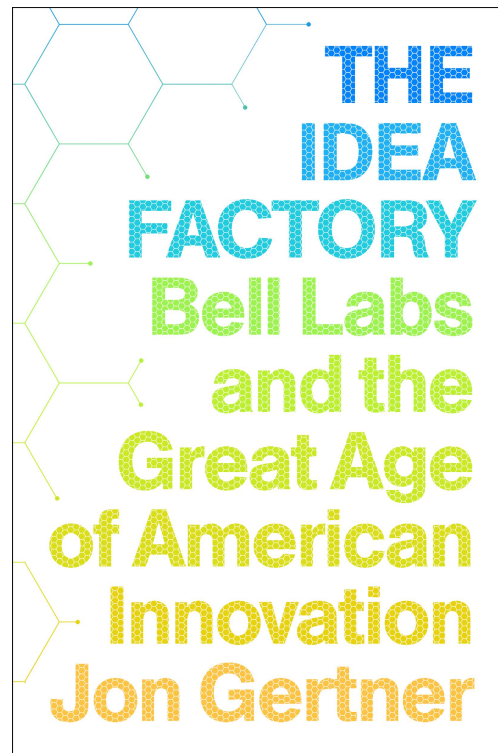
The Idea Factory is a process we created to better align around data science project selection

Where do our ideas come from?

How do we decide which projects to work on?

How do we manage our projects for success?

“The Idea Factory is the worst form of project prioritization, except for all the others”



Building an Idea Factory

Plans are useless, but planning is indispensable



Effective communication is key to success

Ideas come from many places. Make sure your team is talking to the rest of the organization.



Write a **value calculus** to define the benefits of success

Your team will benefit from a deep understanding of how they provide value to the organization and how you measure that value.

External users

What are the technical and scientific improvements that will make our products better?

Are we seeing user adoption and engagement?

Internal users

Can you build software that allows your colleagues be more efficient in delivering for clients?

Are we observing those efficiency gains?

Civis

Could this project raise our company profile?

Did a blog post drive site views? Do we see adoption of our open source package?

Our Team

Is this a sufficiently difficult and interesting problem?

Is the team happy with their work and career progression?



Write a **risk calculus** to define the costs of success

Your team can help you think beyond time and materials. Understanding the risks will help you balance projects and continue to monitor for success along the way.

Technical risk

Does it rely on a library, language, or framework no one else uses?

Do you understand the quality of the data?

Market risk

Do we understand the problem space, the solution space, and the user needs enough to provide the value that will drive adoption?

Legal and compliance risk

Do we have access to the data that we need? Does this meet our security requirements?

Process risk

Does this project require coordination and alignment with another department's roadmap?



Evaluate projects together to keep a balanced portfolio

Resource constraints are real, and you are going to have to make trade-offs. Keep in mind the different types of initiatives you need to deliver value today and in the future.

Methodological research

What is the next advancement in machine learning or statistics?

Methodological development

How do I make a new statistical or machine learning breakthrough usable?

Technical research

What are the tools that will enable my data science efforts in the future?

Technical development

How can I get those tools up and running in my existing stack?



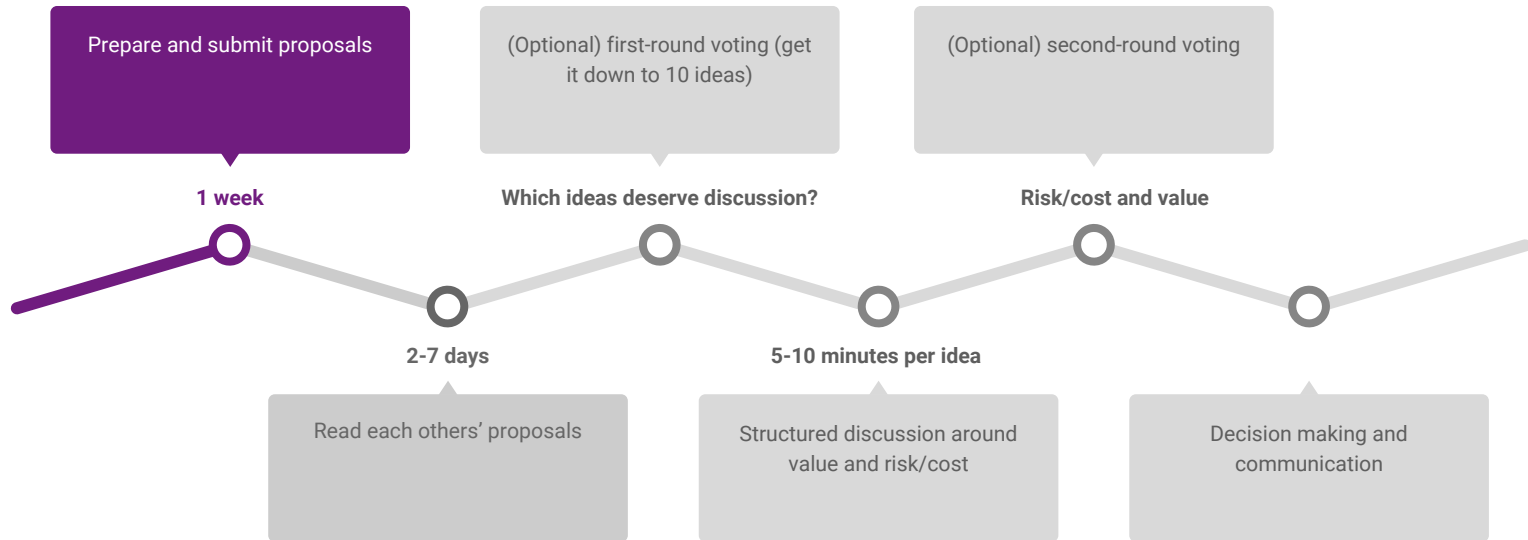
Running an Idea Factory

Create alignment and then let your team run wild



What your team does: the submission process

Force people to make hard choices early. Brevity can be a liberating constraint and allow your team to be really creative.



What makes a good project proposal, part 1

This is our template, feel free to steal shamelessly or make your own. But we do suggest having a template. It lowers the cognitive overhead associated with writing, so the focus is on the idea.

Field	What it is	Commentary
Proposer(s)	Team member(s) and any collaborators from outside the data science team	We favor proposals with that are co-proposed by data scientists and other stakeholders
Project name	Need to know what to call it	Don't get too cute here
One-sentence description	A brief, non-salesy description of the project	If you can't describe in a sentence what you want to build, it's probably not very well-defined. This field was hugely valuable when we got 40+ ideas.
Who would use it?	Who are your users, and why do they care	User-centered ideation encourages projects that will get used. If they don't get used, what's the point?
Deliverables	Very concretely, what will the data scientists create?	Writing this down at the outset helps prevent the "lagging last 10%" problem.
Why we should do this	The affirmative case for doing this project--what's the value proposition?	The best answers here lay out the case for doing the project in terms of agreed-upon business priorities.

What makes a good project proposal, part 2

One non-obvious field: proposer's desire to be project lead, if approved. This field helps people register strong ownership of ideas, so their teammates know if an idea is free for whoever wants it vs. already has a presumed owner.

Field	What it is	Commentary
Estimated timeline	Rough estimate of how long it would take to build	For especially long or high-risk projects, break it into pieces
Resources needed	Resources beyond the scope of the team itself--e.g. new computing resources, or time from other teams	Everything will go more smoothly if people are considering these dependencies upfront
Technical scope + Risks	Where could this fail? Both big technical requirements and other sources of risk	Honesty is critical here, reviewers should be expected to dig into the risks
Project leadership	Does the project proposer want to be in charge of leading the project if it's approved?	Help distinguish projects where the proposer feels real ownership from "cool idea for anyone who's interested" (both are good!)



What makes a bad project proposal



What makes a bad project proposal

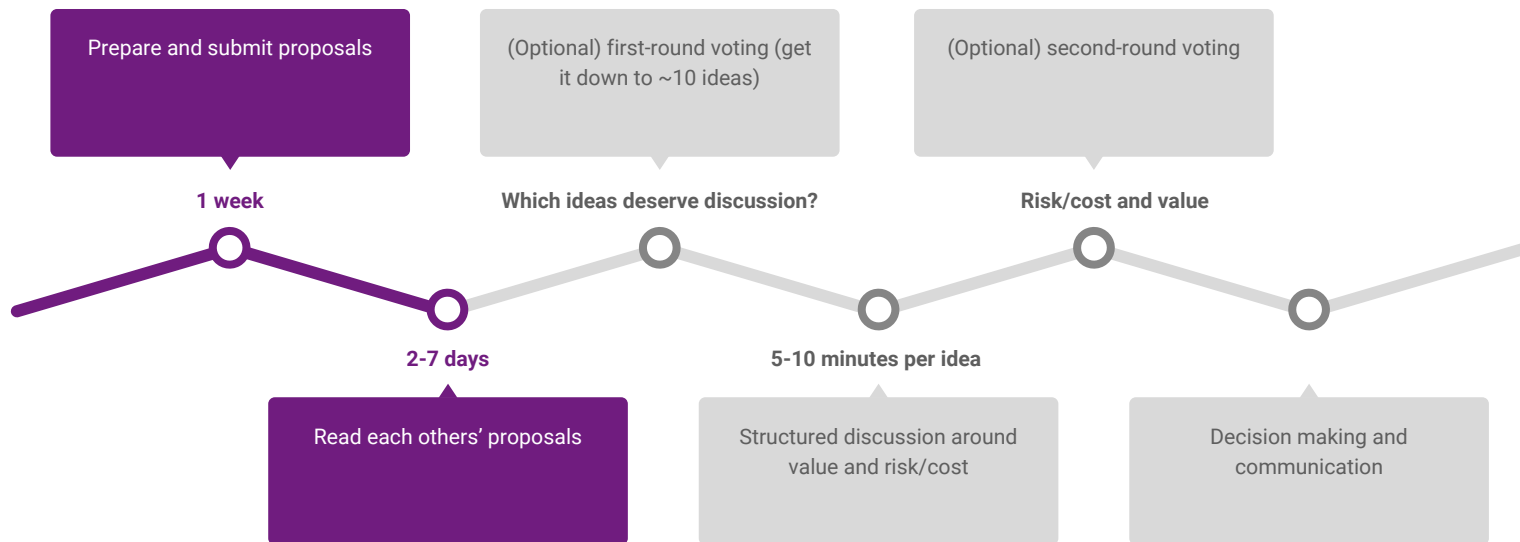
Nothing. There are no bad proposals. Only proposals that may not yet be great.

Next slide.



What your team does: the review process

Crafting good proposals takes time, space, and collaboration. We added in a week long review process after all the proposals were in.



Gathering constructive feedback will make better proposals

In every step of the process, your team is always better together. Get them thinking about ideas that they haven't yet.

Why

The review phase ensures that every idea gets some consideration by and feedback from someone who didn't author it. Having familiarity with many proposals will facilitate discussion later.

How

Create proposals using a tool that allows public commenting, ask people to read the proposals and offer up any questions or comments.

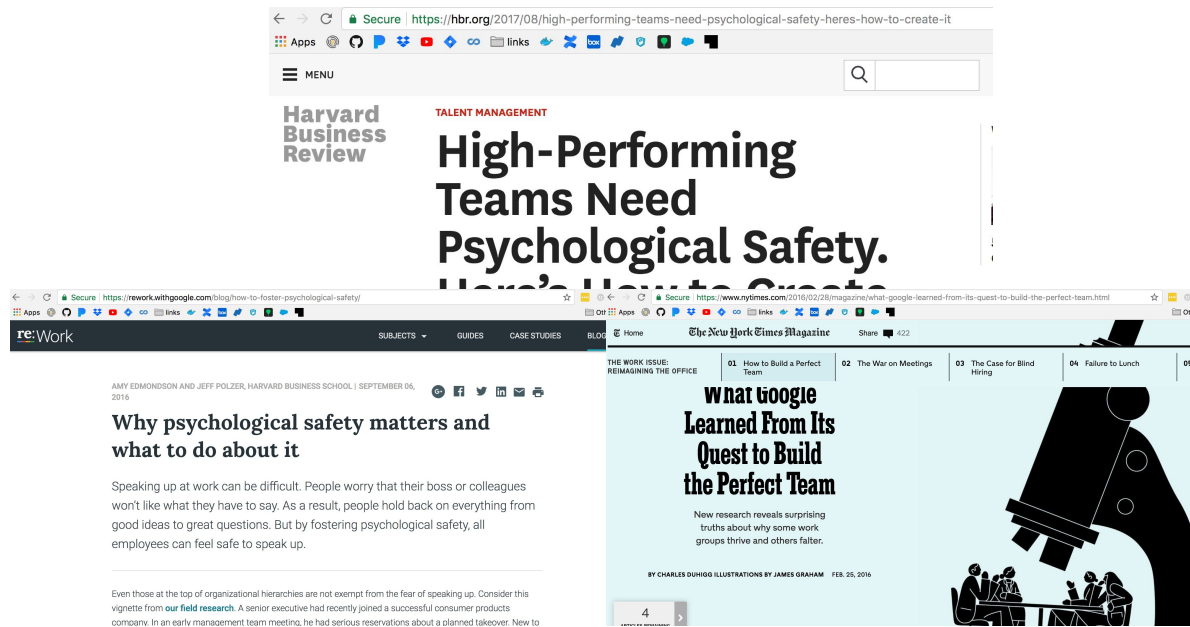
Pro Tip

Assign at least one reviewer to each proposal – it ensures that every proposal gets someone thinking about it and asking questions.



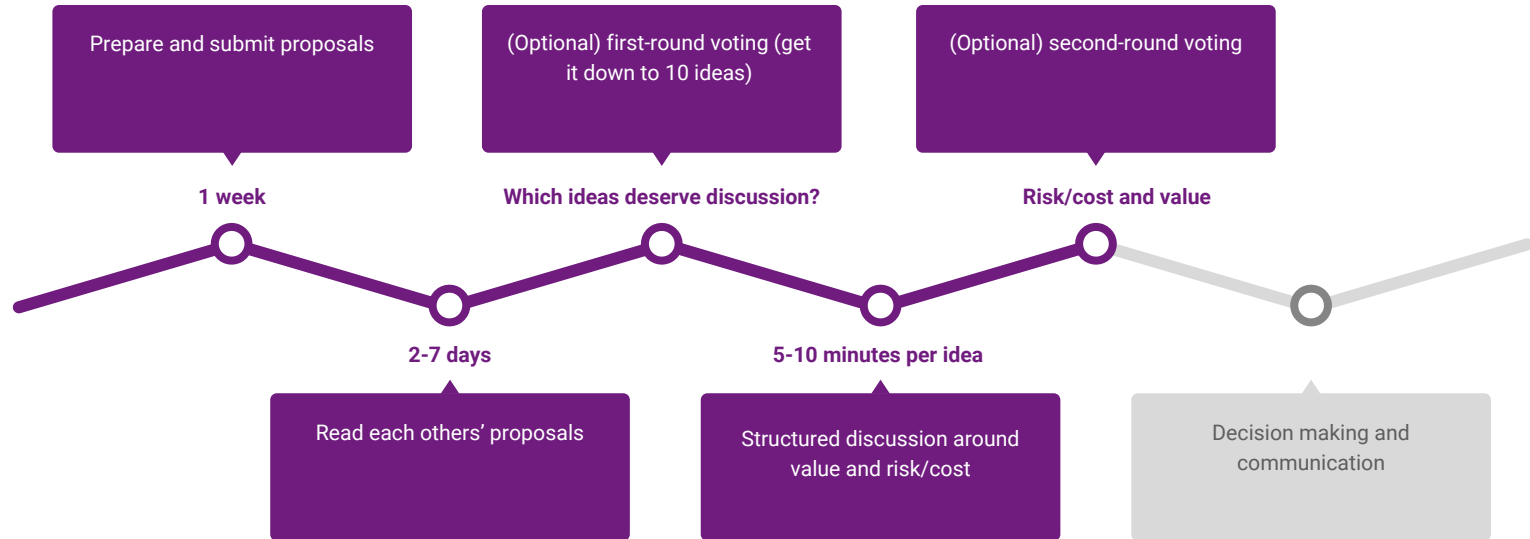
An aside on **psychological safety**: make a code of conduct

Encouraging the exchange of constructive feedback in a psychologically safe environment is one of the best things you can do for the dynamism and creativity of your team.



What your team does: the voting process

Voting narrows the field so you can discuss the best proposals in more detail and helps surface the views of the team to the leads and managers.



Discussion and voting capture the collective wisdom

	VALUE		RISK	
	H	L	H	L
6	1	5	5	1
15	1	1	1	5
9	1	1	1	5
12	1	1	1	5
23	1	1	1	1
7	1	1	1	1
8	1	1	1	1
24	1	1	1	1
18	1	1	1	1
11	1	1	1	1

The whole point of the exercise is to involve the team in decision-making.

Voting is a quick and quantifiable way of gauging enthusiasm and prioritizing.

Discussion helps inform voting.

Round 1 voting: Which ideas should advance to discussion? Approval voting

Discussion: 5-10 minutes per idea (time it)

Round 2 voting: High/low, risk/value

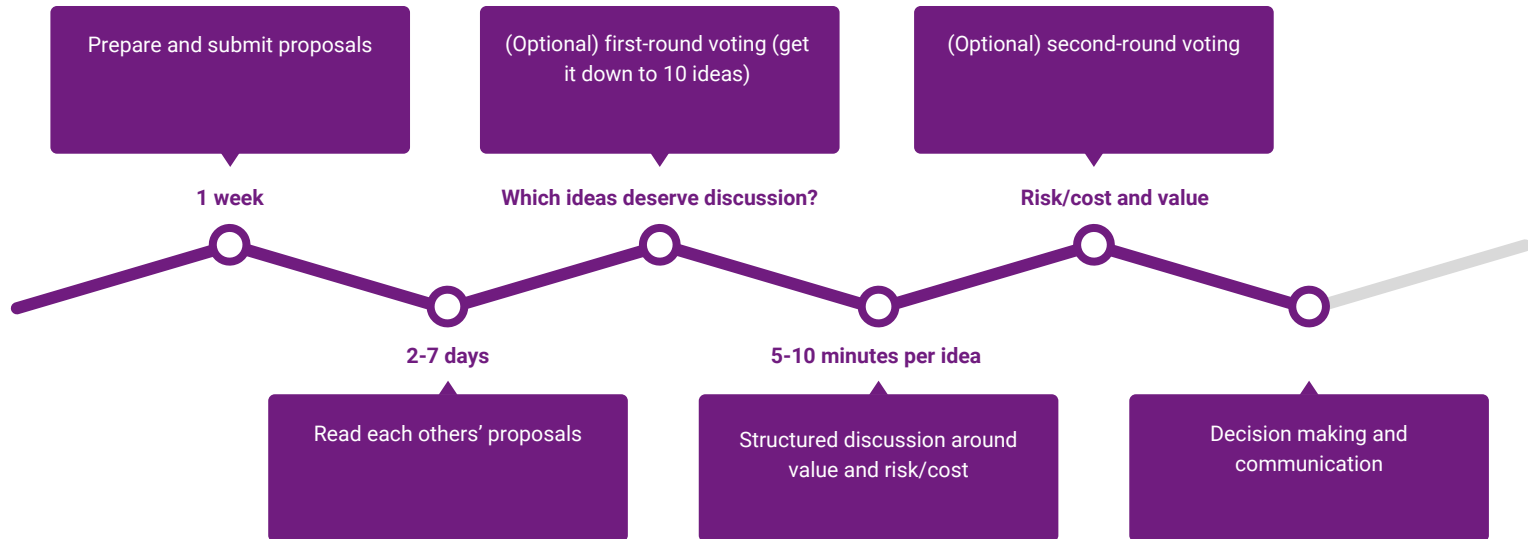
our actual vote
counting sheet

you don't have to get
fancy



What your team does: understanding the decisions

This is a long process – it's easy to lose momentum at the end. Don't. Communicate the final decisions that were made and let everyone know why.



There is no such thing as overcommunication

To be an good leader, you have to repeat things. If you find yourself saying the same things over and over, people may just be starting to get it.

1

discussion with
our team leads

5

example proposals

2

talks to the team to
motivate and explain the
process

2

long-form (5+ pages)
documents about what,
why, how

1

webpage for collecting
proposal submissions

3

hours of discussion,
public voting

6+

private conversations

1

talk to the department,
explaining project
choices



Reflecting on two iterations of the Idea Factory

What all this work actually produced



R&D Selected Projects and Results

What's the payoff of all this work? A really cool portfolio of projects. Already appearing at a data science conference near you (cough cough)...

1. Predicting a particular type of transactions using a new **factorization machines** implementation
2. Making causal, not just correlative, **attribution** with digital ad data
3. Modeling customer **churn** with recurrent neural networks
4. Pay down **tech debt** on a high-traffic package

Factorization Machine Model

$$E(Y) = \beta_0 + \sum_{j=1}^p X_j \beta_j + \sum_{j=1}^p \sum_{l=j+1}^p X_j X_l v_j^T v_l$$

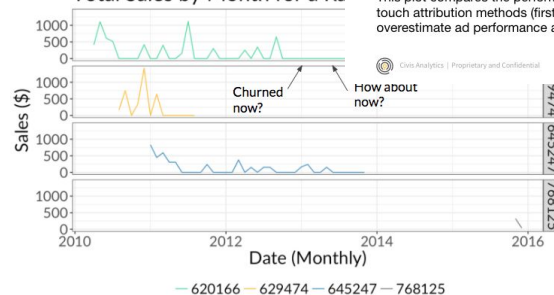
- Interaction vector v_j for each feature
- Interaction terms estimated by computing dot product of low-dimensional representation
 - Compute unseen interactions**
 - Inform
 - Recall
- Scoring is even!
- Link function
 - L2 pe
 - logist

Fancy Trick

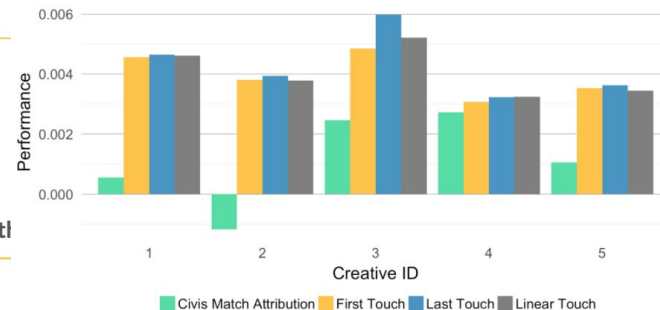
What we are typically working with

How do we define churn?

Total Sales by Month for a Random Sample of Advertisements



Ad Effectiveness by Creative ID and Attribution Method



This plot compares the performance of ad campaigns using Civis Match Attribution and three common touch attribution methods (first, last, and linear touch). Commonly used touch attribution methods often overestimate ad performance and are unable to capture backlash.



Technology and Product Roadmapping

We used a very similar model for our recent product roadmapping to great effect.

Similarities

Diversity of opinion.
Many different needs.
Balancing quick wins vs.
riskier long-term
investments.



Differences

Input by business unit.
More go-to market focus.
Product design sprints.
Top-down vision more
important.



Outcome

Identified a strong opportunity to invest in
some of our best point solutions in
addition to areas of our platform as an
enabling technology.



Closing thoughts

Ensuring data science has a lasting impact on the way that an organization operates takes a lot of work.

- Get everyone speaking the same language by establishing a **shared context**.
- Make sure your data science team understand the **business goals**.
- Be realistic about what you can achieve by **including the data scientists early** in your business planning processes.
- Create an environment of **psychological safety** by establishing a **code of conduct** for discussions and trying out some exercises in **participatory decision-making**.
- **Balance** your projects by taking into account multiple objectives.



THANK YOU

[@multiarmbandit](#)
cmalone@civisanalytics.com
lineardigressions.com



[@jseabold](#)
sseabold@civisanalytics.com

