

FROM THE PRESIDENTIAL CAMPAIGN TRAIL TO THE ENTERPRISE

Building Effective Data-Driven Teams

Katie Malone

Director of Data Science Research
and Development



2012: THE CAMPAIGN

Winning the White House (again) with data



THE OBAMA CAMPAIGN, 2012

The first presidential campaign to be built around data and analytics

Presidential campaigns are not unlike modern businesses

- ⦿ The pressure is high, the timelines are tight, and priorities shift often.

But a successful Presidential campaign doesn't end with the whole team moving to D.C.

- ⦿ Sometimes there's a bigger impact in "going corporate."



This talk is about how campaign-style analytics are being used today, for advocacy organizations and legacy enterprises alike.



THE BIGGEST CAMPAIGN ANALYTICS PROBLEMS GENERALIZE BEYOND POLITICS

WHO TO TARGET

- ⦿ Campaigns want to understand who their voters and donors are, but parties, PACs, secretaries of state, and advocacy groups all use their own data systems and conventions.

WHAT TO SAY

- ⦿ Naive tests of message effectiveness often overestimate effects and miss backlash, so outreach dollars go toward ineffective messages.

WHO TO TALK TO

- ⦿ Not all targets are created equal. Someone who likes your candidate but might not turn out to vote is different from a very engaged voter who might be persuaded to volunteer or donate.

WHERE TO SAY IT

- ⦿ Resources are limited. If your voters are watching Adult Swim, that's probably a better way to reach them than buying a Super Bowl ad.



2014:HEALTHCARE OUTREACH

Estimating the numbers of
uninsured Americans



After the passage of the Affordable Care Act, a national healthcare non-profit needed to run a campaign to inform the public about and, ultimately, influence them to sign up for healthcare. Here's how we brought some campaign-style analytics to that problem.



ENROLL AMERICA NEEDED TO KNOW HOW MANY PEOPLE WERE UNINSURED IN THE U.S. AND WHERE THEY LIVED



THE PROBLEM

The organization needed to know where a high number of uninsured individuals reside so as to target enrollment outreach efforts



THE GOAL

To provide estimates by demographic groups and geographic areas so the organization could design outreach and allocate resources effectively



OUR CHALLENGES

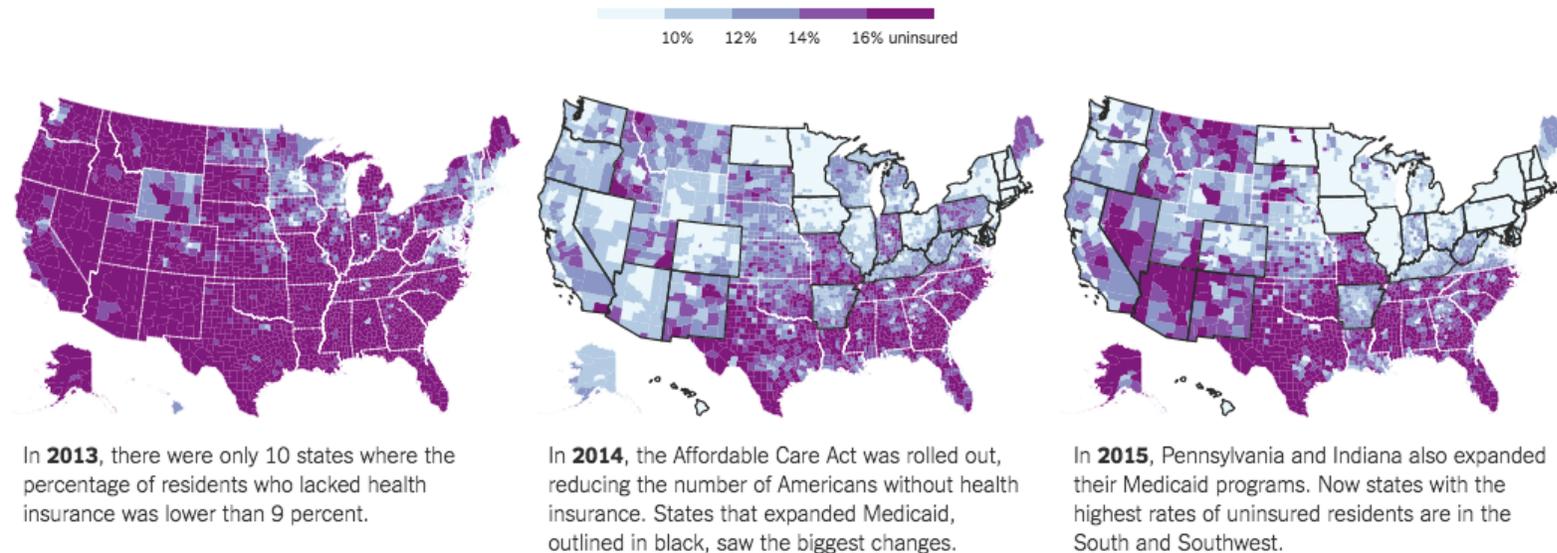
Being able to target outreach to uninsured Americans is difficult because data on the uninsured are limited and when it is available, it is often outdated for the purposes of designing programs



OUR SOLUTION...

We combined survey data and consumer data to model each individual's likelihood of being insured

Percentage Uninsured, by County, 2013 to 2015



- Our model of the uninsured has been repeated every year since 2013 and used to inform strategic resourcing decisions by the client
- It has also been used to identify trends in insured status
- Results have also been featured in the New York Times¹

¹<http://www.nytimes.com/interactive/2015/10/31/upshot/who-still-doesnt-have-health-insurance-obamacare.html>



IMPROVING TV MEDIA SPEND EFFICIENCY WITH THE 'CIVIS MEDIA OPTIMIZER'

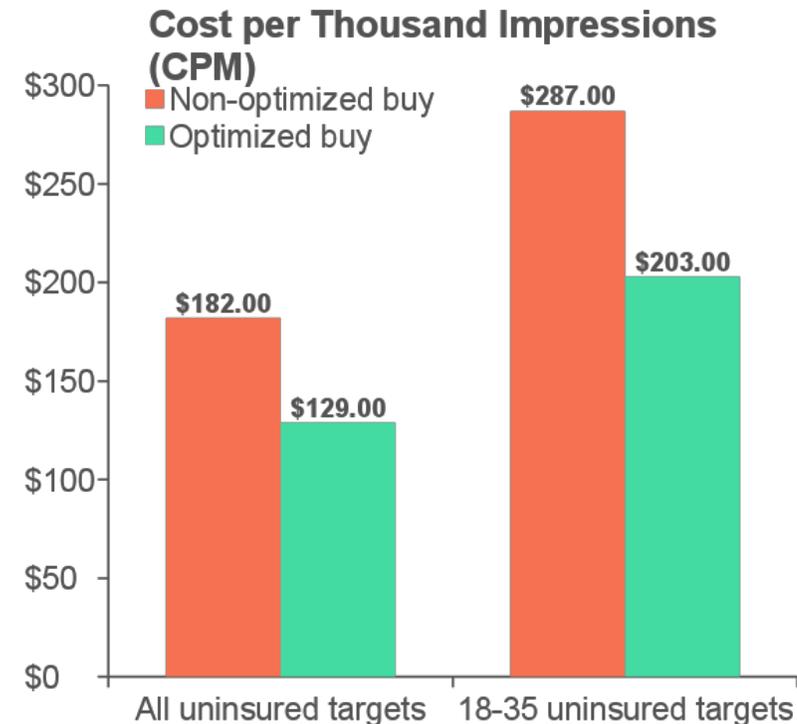
Example: Insurance Enrollment

Background

- ⦿ **The Problem:** There were 40 Million uninsured Americans, but the Affordable Care act now Mandates enrollment.
- ⦿ **The Goal:** Enroll as many uninsured Americans in to private exchanges and public assistance programs as possible.
- ⦿ **Our Challenges:** There is no list of uninsured Americas, different constituencies have different insurance likelihoods, and a lot of people really don't like the policy.
- ⦿ **Our Solution:** Civis built a person-level model that predicted the probability someone did or did not have healthcare coverage and helped optimize media to them

Efficiency Gains for Insurance Client

By merging the spot rates with viewership patterns for uninsured targets, we allocated \$6 million budget. The optimized media plan showed a 26% decrease in CPM compared to the ad agency's original media plan.

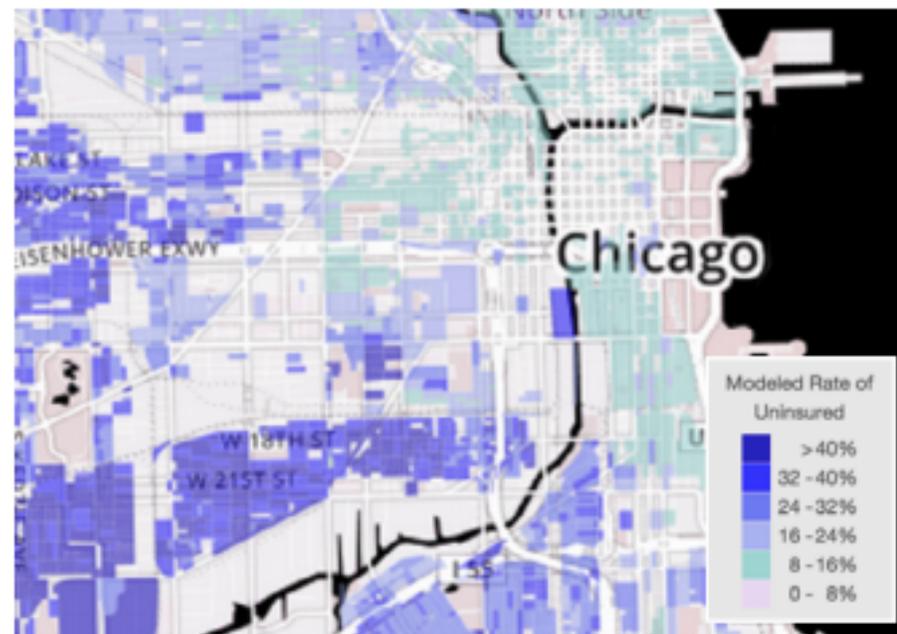


Enroll America's staff and volunteers needed a way to inform an outreach program and a way to prioritize who and where to spend resources

To inform its partnerships and large grassroots organizing efforts, Enroll America needed a way to speak directly with those least likely to have health insurance

Our Uninsured Model was a key component of Enroll America's strategy. Their efforts included:

- ⦿ Finding geographic “hotspots” of high concentrations of uninsured Americans
- ⦿ Visiting identified neighborhoods to establish partnerships with community leaders
- ⦿ Door-to-door canvassing high-uninsured city blocks

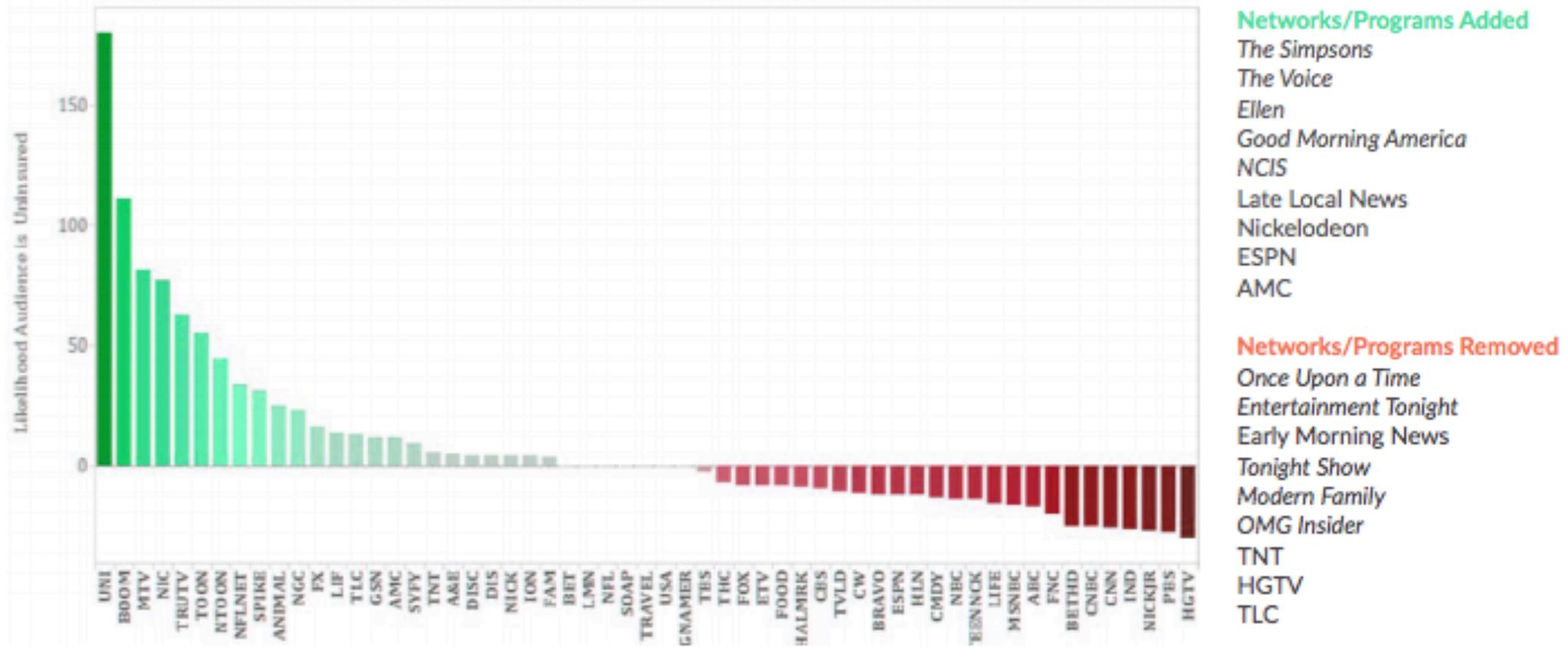


Source: Civis Uninsured Model 2013

An analysis proved Enroll America's conversations showed the model accurately predicted uninsured rates within 1 percentage point

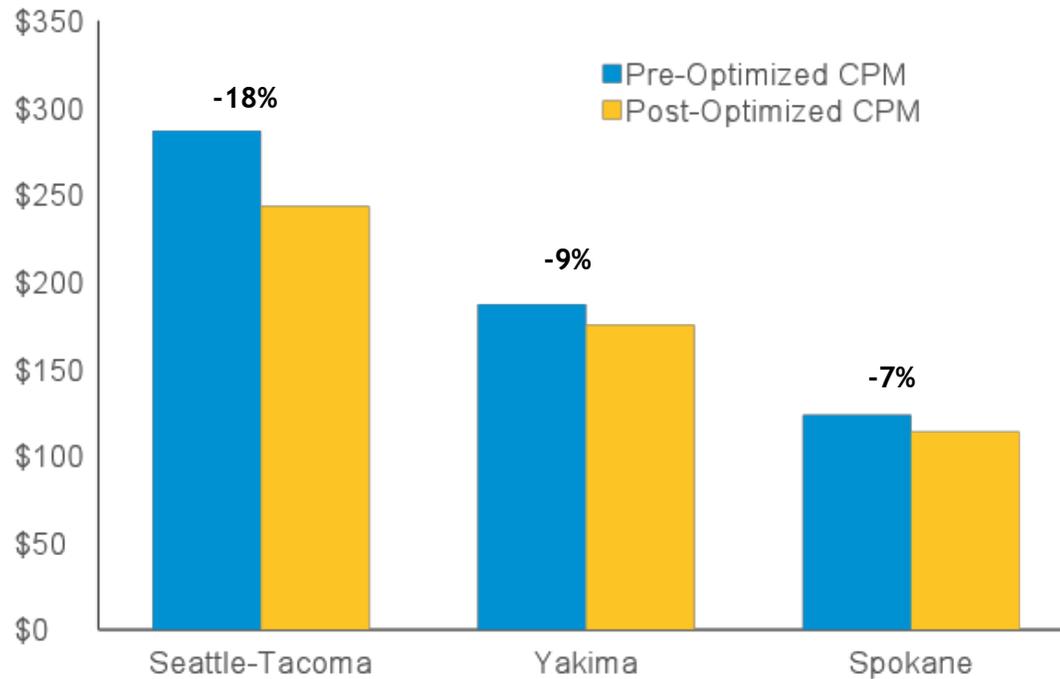


Additionally we worked with the Washington State Health Benefit Exchange and a media agency to optimize their media for the 2013 enrollment period



On television Civis Optimization netted 15.4% in additional impressions to our targeted uninsured audience

Uninsured People 18-35 CPM
(Cost Per Thousand Impressions)



Results

Each market saw its targeted CPM decrease once the uninsured media optimization recommendations were integrated into the media strategy

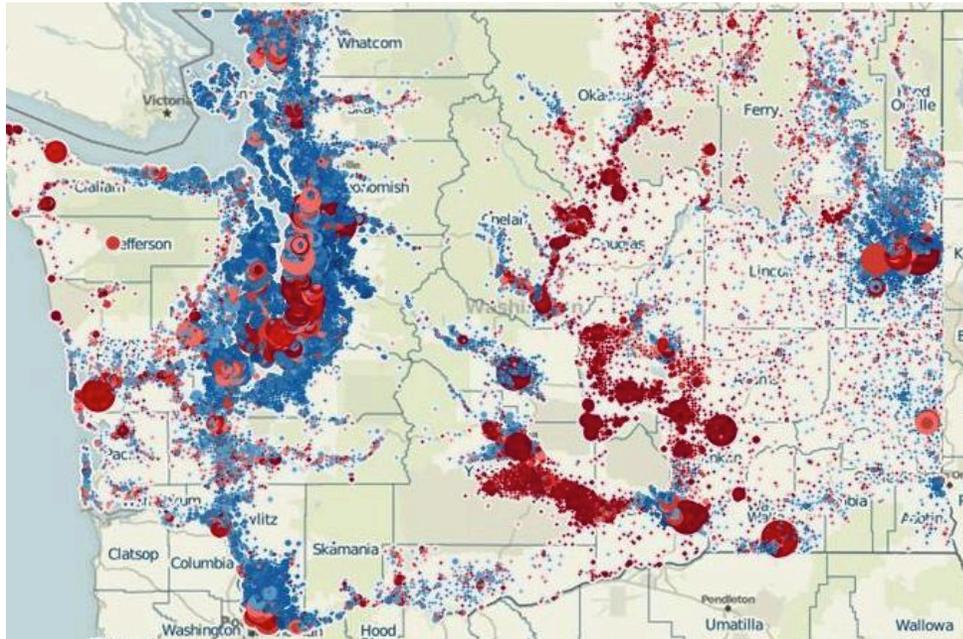
The broadcast choices made in Seattle generated the greatest value for the Washington State Exchange: there was a 26% decrease in CPMs for Uninsured 18-35 post-optimization

Media optimization also program made a big impact on cable choices in Yakima and Spokane with a 24% and 18% decrease in CPMs, respectively



For Out-of-Home placements, we created tools to help select better billboard placements and radio market choices

Uninsured Persons 18-64



Out of Home Signage

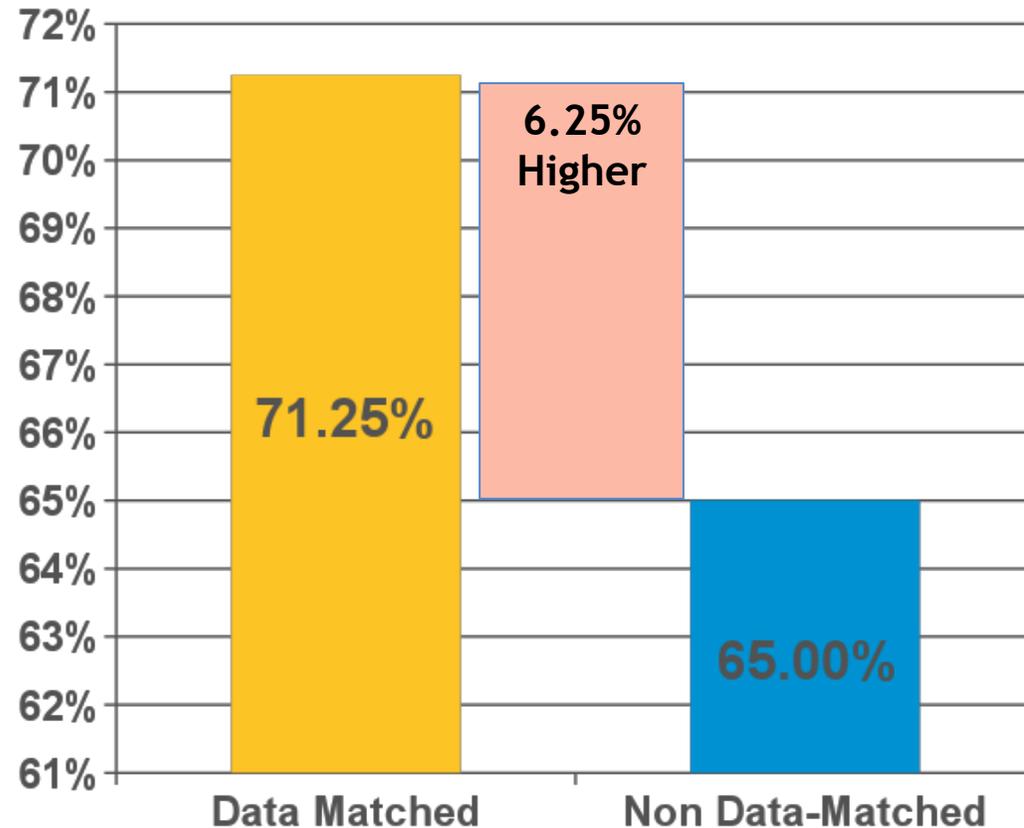


We picked the best locations for OOH - billboards, grocery cart and buses - that fit our creative, had the greatest opportunity for exposure among the target and created the most value per dollar



Digital Performance of modeled audience exceeded results from 3rd party data and demographic targeting

Online Video Performance



Results

The digital video buy proved that the ads matched to the modeled data were more relevant than using other data sources by almost 10%.

There were also click-thru-rates 9x higher through the modeled data digital video as compared to other 3rd party data available.

Both modeled data match and other 3rd Party digital video over-performed goal of 63%.



2018: THE ENTERPRISE

Bringing campaign-style data science to our enterprise partners



FIND TRUTH, TAKE ACTION, AND MAKE DATA SCIENCE INTEGRAL TO YOUR ORGANIZATION

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

- Founded in 2013, Civis Analytics is a 150-person data science technology and advisory company with offices in Chicago and Washington, D.C.
- We invented technology to operationalize data science for our consultants and our clients' internal data science teams.
- Today, Civis is the premier data science platform and provides world-renowned applied data science consulting.



DATA SCIENTISTS AND BUSINESS STAKEHOLDERS WILL LOOK AT THE SAME PROBLEM AND SEE DIFFERENT THINGS

Alignment starts with understanding the other point of view

Data scientists need to be comfortable with the trade-offs of business.

- ⦿ **Metrics** are often ill-defined or don't capture business value well. Don't be afraid to point this out, but understand if they're hard to change.
- ⦿ **Speed** is often better than perfection. An 80% solution today outweighs a 95% solution next month.
- ⦿ **Security and privacy** are paramount. The folks who control IT and data know this, and are incentivized accordingly.

Business stakeholders need to understand the challenges of working with data, as well as its potential.

- ⦿ **Data** is often messy and low-quality. It can take a long time to deal with that.
- ⦿ **Algorithms** lend themselves toward predictive power or interpretability, but not both.
- ⦿ **Deployment**, or making it so a model is being used by decision-makers, does not happen on its own.

AN EXAMPLE BUSINESS MANDATE:

How we articulate data science problems as business problems

WHO TO TARGET: Businesses want to understand their member/user journey, but it's measured by different departments/vendors in different systems.

WHAT TO SAY: Conventional opinion measurement is broken. Few companies are capable of providing accurate measurement technology with robust statistical controls, repeatable scale, and self-service.

WHO TO TALK TO: Because businesses don't have defined lists or measurement against those lists, they have no value-based means to prioritize their work, so they do so with intuition or randomly.

WHERE TO SAY IT: Without value-prioritized targeting, they outsource their outreach optimization to extraneous departments or bad external actors (like agencies).

DID IT WORK: There's no eventual feedback on what work did well or poorly, so the quality of work doesn't improve over time. This is the most critical deficiency.

FOUNDATIONAL TECHNOLOGY & CURATED DATA: Resolving these problems requires terabytes to low petabytes of storage/compute power & outside data, which most businesses lack.



WHAT A DATA SCIENTIST WOULD SEE IN THIS MANDATE

WHO TO TARGET: Businesses want to measure their customer/user journey, but it's measured by different departments/vendors in different systems.

data unification and identity resolution

WHAT TO SAY: Conventional opinion measurement companies are capable of providing accurate measurement technology with robust statistical controls, repeatable scale, and self-service.

causal inference on experimental data

WHO TO TALK TO: Because businesses don't have measurement against those lists, they have no value-based means to prioritize their work, so they do so with intuition or randomly.

machine learning models

WHERE TO SAY IT: Without value-based means, they source their outreach optimization to extraneous departments or bad external actors (like agencies).

algorithmic optimization

DID IT WORK: There's no eventual feedback or learning loop, so the quality of work doesn't improve over time. This is the most critical failure.

causal inference on observational data

FOUNDATIONAL TECHNOLOGY & CURATION: Solving these problems requires terabytes to low petabytes of storage/compute power & processes that lack.

software stack + data assets that support all this



The technical components of a data science pipeline map onto team roles

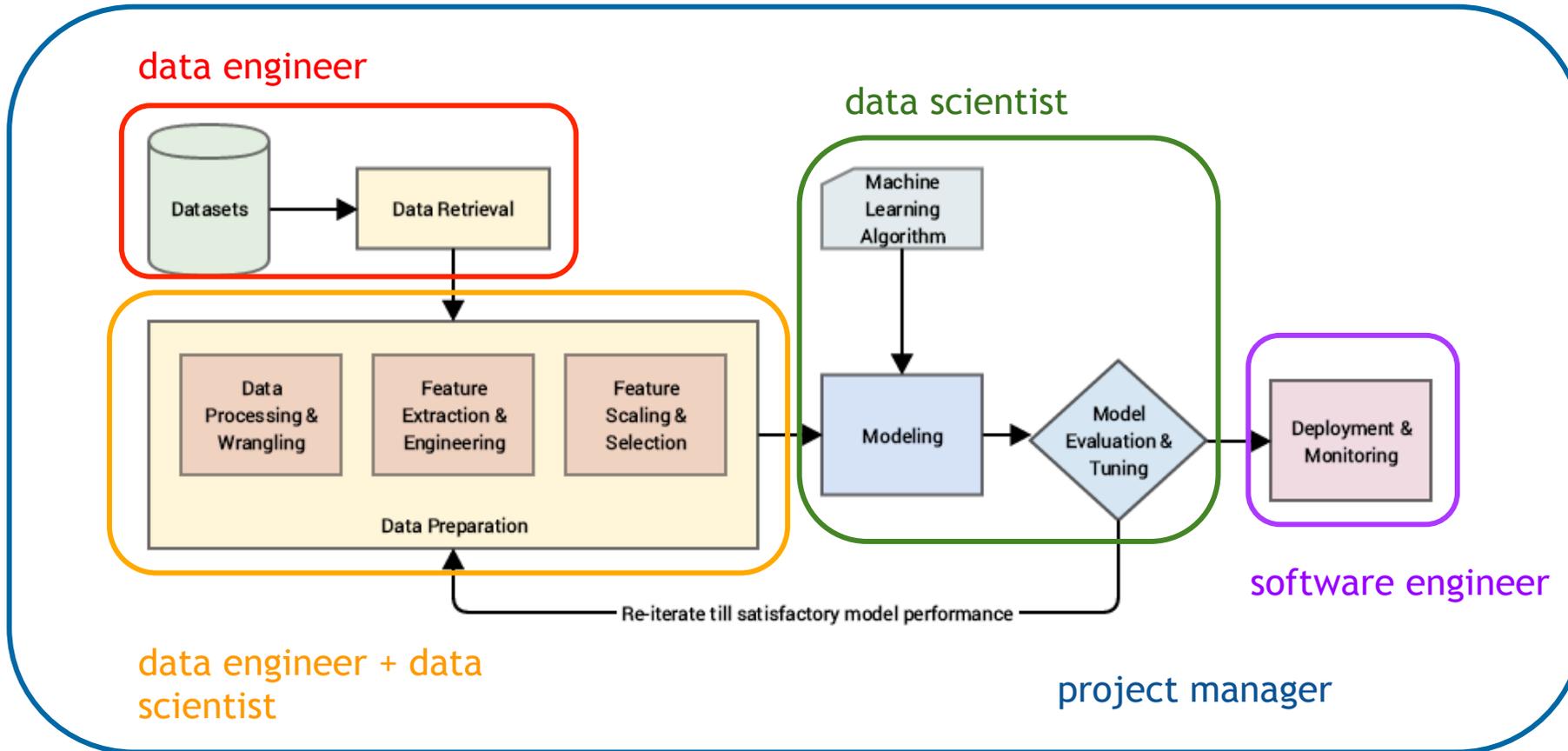


image source: Practical Machine Learning with Python, Apress/Springer)

The best data science teams are cross-functional: they combine data and software experts with people who have business context and scientific expertise.



Build out a team structure that supports the business objectives

Knowledge or Skill	Owner	Data Scientist	Data Engineer	Software Engineer	Project Manager	Executive Sponsor
Statistics and machine learning	valuable	critical	nice to have	nice to have	valuable	valuable
Practical data skills	nice to have	valuable	critical	valuable	nice to have	nice to have
Software engineering	valuable	nice to have	critical	critical	nice to have	nice to have
Collaboration and teamwork	critical	critical	critical	critical	critical	valuable
Communication	critical	critical	valuable	valuable	critical	critical
Project management	valuable	valuable	nice to have	nice to have	critical	nice to have



CLOSING THOUGHTS

- ⦿ We've come a long way since 2012
 - Methodologically
 - Technologically
 - Organizationally
- ⦿ Data science is professionalizing and business expectations are high
- ⦿ Enterprises are rapidly bringing themselves up to speed
 - Evolve to survive



THANK YOU

@multiarmbandit

cmalone@civisanalytics.com

lineardigressions.com

