

AGENDA

- Building Data Science Capabilities
- The 3 Pillars of Data Science
 - Culture - Leading a Change
 - Talent Structure
 - Tech Infrastructure
- Case Discussion

MOTIVATION

- Graduating seasoned data scientists
 - DSI vs. MSDS
- 23.5% of our graduates have managerial / DS-Leadership roles
 - Either lead the effort
 - or be a function in the system
- Going to give you an outlook of the job and other parts related to it

MACHINES ARE TAKING OVER OUR ORGANIZATIONS

Today, most innovative companies rely on data science & machine learning to drive business process and user experience to add value



The future of business innovation has an Artificial Intelligence components in it very core structure



THE 3 PILLARS OF DATA SCIENCE



These organizations have:

- Strong data culture
- Strong teams
- Technology infrastructure





BUILDING DATA SCIENCE CAPABILITIES

Many organizations believe in the power
and potential of data science,

but are challenged in establishing a
sustainable data science capability.

Why?

How do organizations embed data science
across their enterprise so
that it can enhance performance
and return on investment?

Building a data science capability in any organization isn't easy—there's a lot to learn, with roadblocks and pitfalls at every turn.



3 PILLARS OF DATA SCIENCE



DATA CULTURE

HEALTHY DATA CULTURE



- Create an **environment** where data scientists can **thrive**
- Get into the habit of making **data-driven decisions**
- Data is **available** for **ANYONE!** (not just the analysts and Data Scientists)
- Prioritize **investment** in with the highest **Data ROI**
- Data governance is clear

A MODEL FOR LEADING A CHANGE

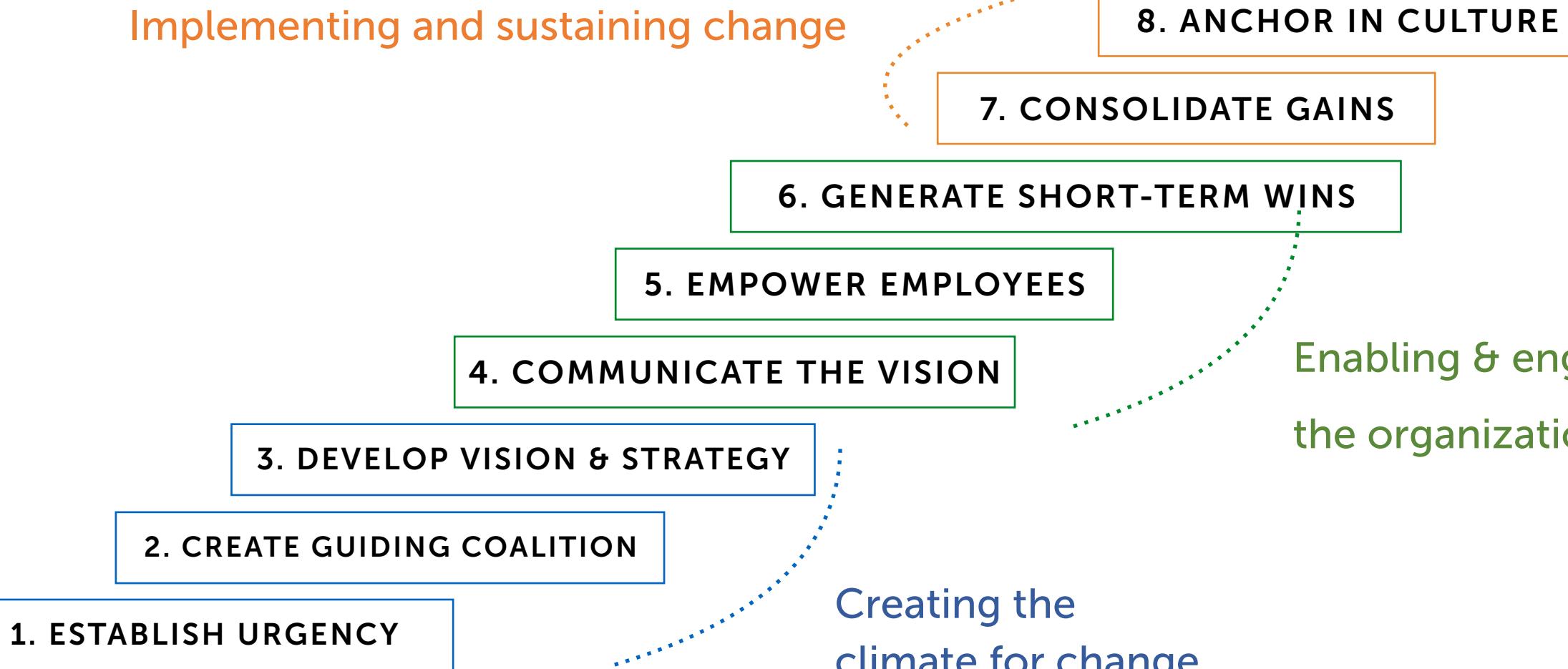


1. Establish a sense of **urgency**
2. Create a guiding **coalition**
3. Develop a **clear vision and a strategy**
4. **Communicate** the change vision
5. **Empowering employees** for broad based action
6. Generating short-term **wins**
7. Consolidating **gains** and producing more wins
8. Anchoring new approaches in **culture**

KOTTER FRAMEWORK FOR CHANGE



Implementing and sustaining change



Creating the
climate for change

Enabling & engaging
the organization

1. BUILD URGENCY

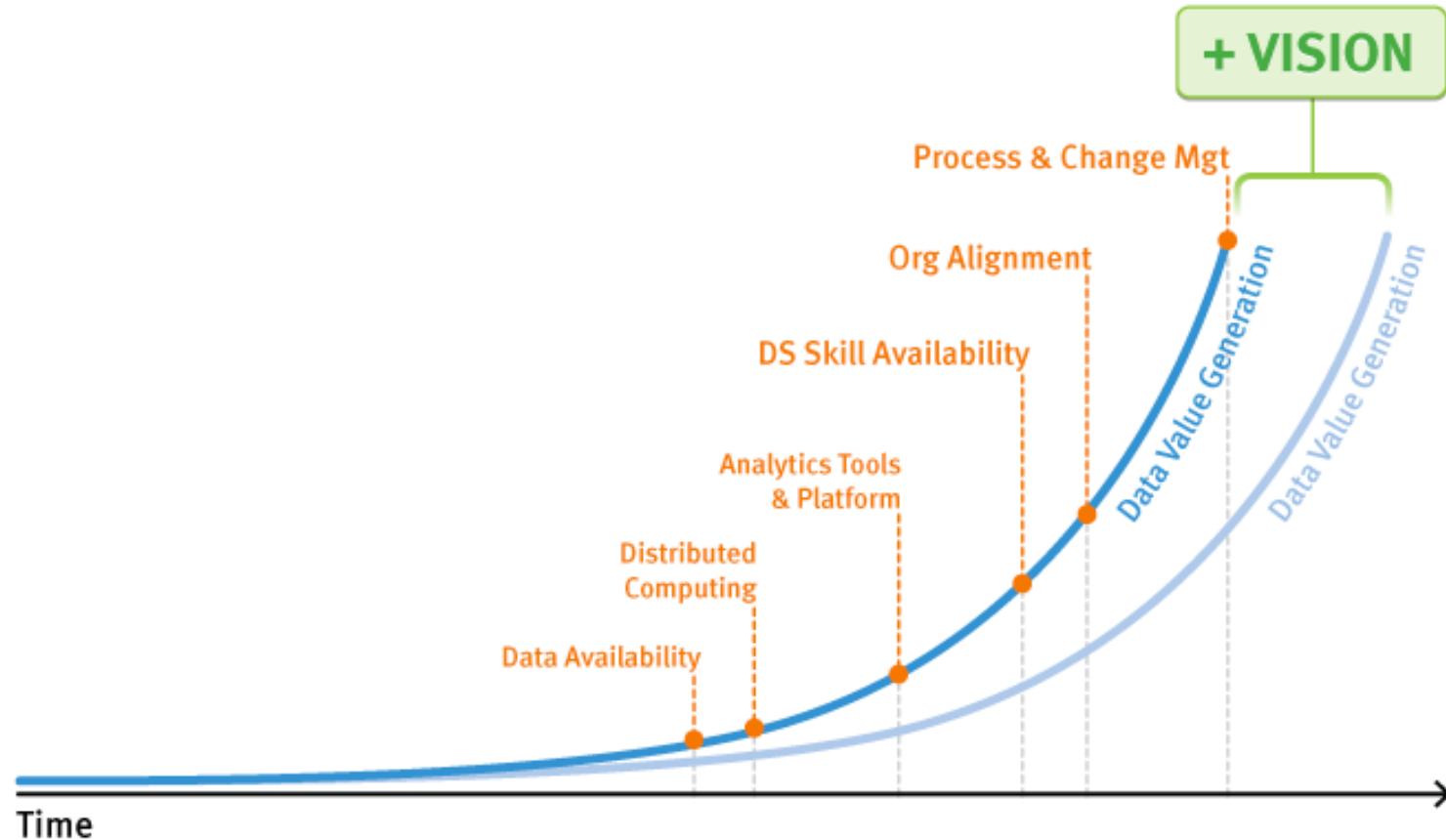


- Acknowledge the fact that **data science is here and now**
- Look to technology companies and **competitors**
- Identify opportunities for **data ROI**
- Make data part of the **mission**
- Make data part of the **daily routine**

2. COALITION BUILDING



3. VISION & STRATEGY - MAKES THE DIFFERENCE



4. COMMUNICATE THE TRANSFORMATION



- ▶ Showcase to stakeholders the value of your experience, plus the value of the data
- ▶ How can data help us achieve our goals?
- ▶ Data as a competitive advantage

5. EMPOWER DATA TEAMS



- **Data Scientists** are much more effective **working in teams**
- Create a **multi-disciplinary team** across all the Data Science capabilities (math, computer-science, domain experts).
- Some existing employees may be able to contribute immediately, but most will need to be **trained**

6. GET QUICK WINS



Technology Adoption Journey of a Major Healthcare Provider

Prove that better technology can speed up discovery

- Hackathon

Prove that better technology can improve model quality

- Length of Stay Modeling

Prove that technology is accessible to my clinicians and researchers

- Comorbidity Feature Generation App

Prove that data science can help in areas other than clinical analytics

- Fraud Detection for Accounts Payable

Prove that, once trained, our scientists can get to insights as quickly as the Pivotal DS team

- EDIP Modeling in 4 days

7. CONSOLIDATE & WIN BIGGER



- **Expand** from a 1-2 departments to 3-4 and so on
- Bring data **closer to semi-technical folks** & management
- Continue **investing** in technology
- **Share best practices** across functional teams

8. MAKE IT OFFICIAL

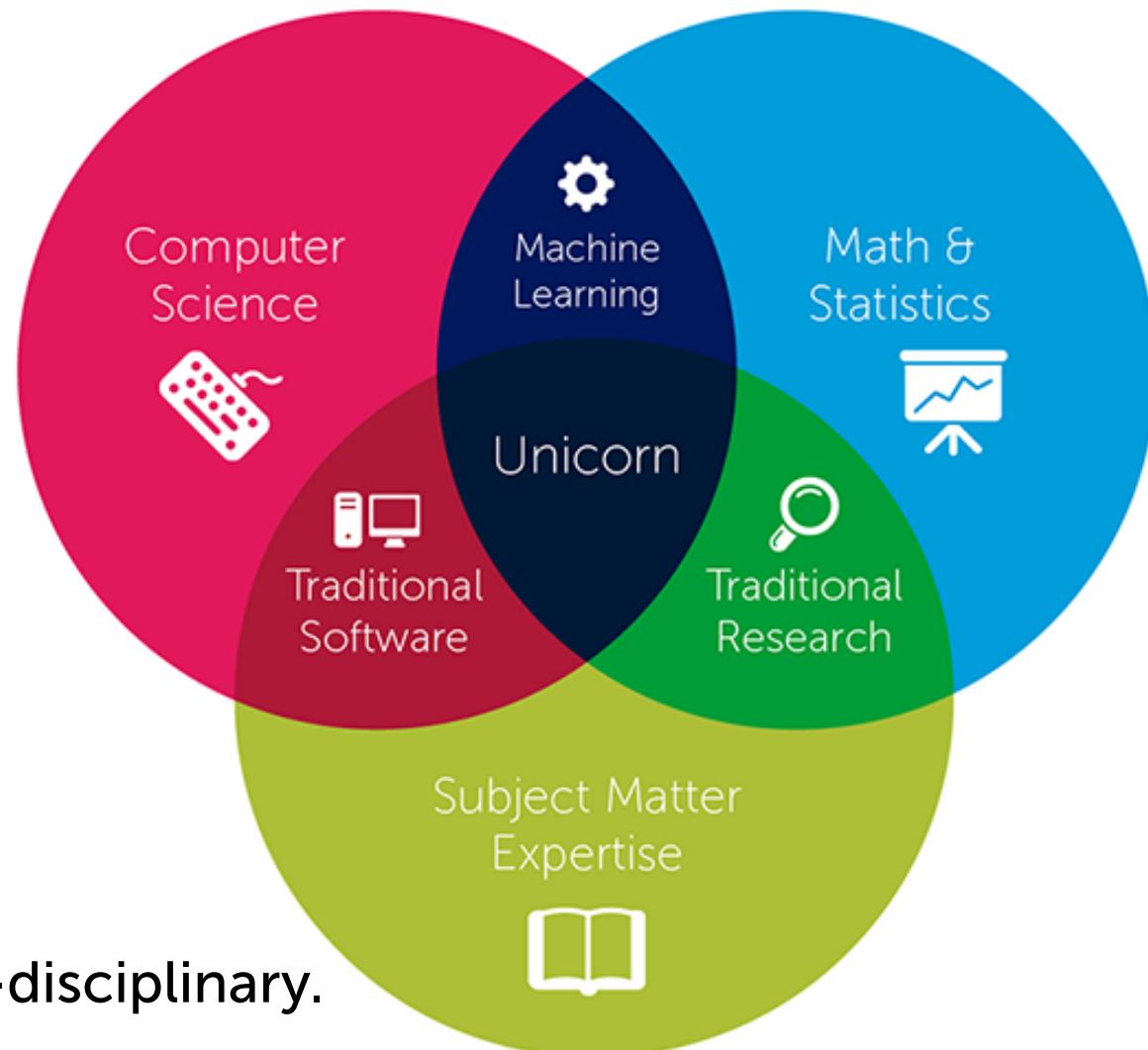


- Ensuring HR processes **reward**/favor new procedures in promotions hiring
- Establish **committees** and/or role for cross-functional collaboration
- **Evaluate processes** that should be removed or updated



DATA TALENT

DATA SCIENCE TALENT



Data science is inter-disciplinary.

TODAY'S DISCUSSION QUESTIONS

- When do you need to **build a team?**
 - How do you know? What are the needs?
 - What kind of people do you need on the team?
- **Organizational model?**
 - Where should the data science team live within an organization ?
 - Is it always necessary to build a team? When to build, buy or partner?



Data Scientist is a Tall Order

- Data Engineer
- Experimental Scientist
- Effective Advisor
- Technical Analyst
- Business Expert

YOU →





How to Grow Data Science Talent

Look for:

- ▶ Curiosity & growth mindset
- ▶ Interdisciplinary perspective & Skills
- ▶ Technical details and big picture questions



Does a Data Scientist need a PhD?

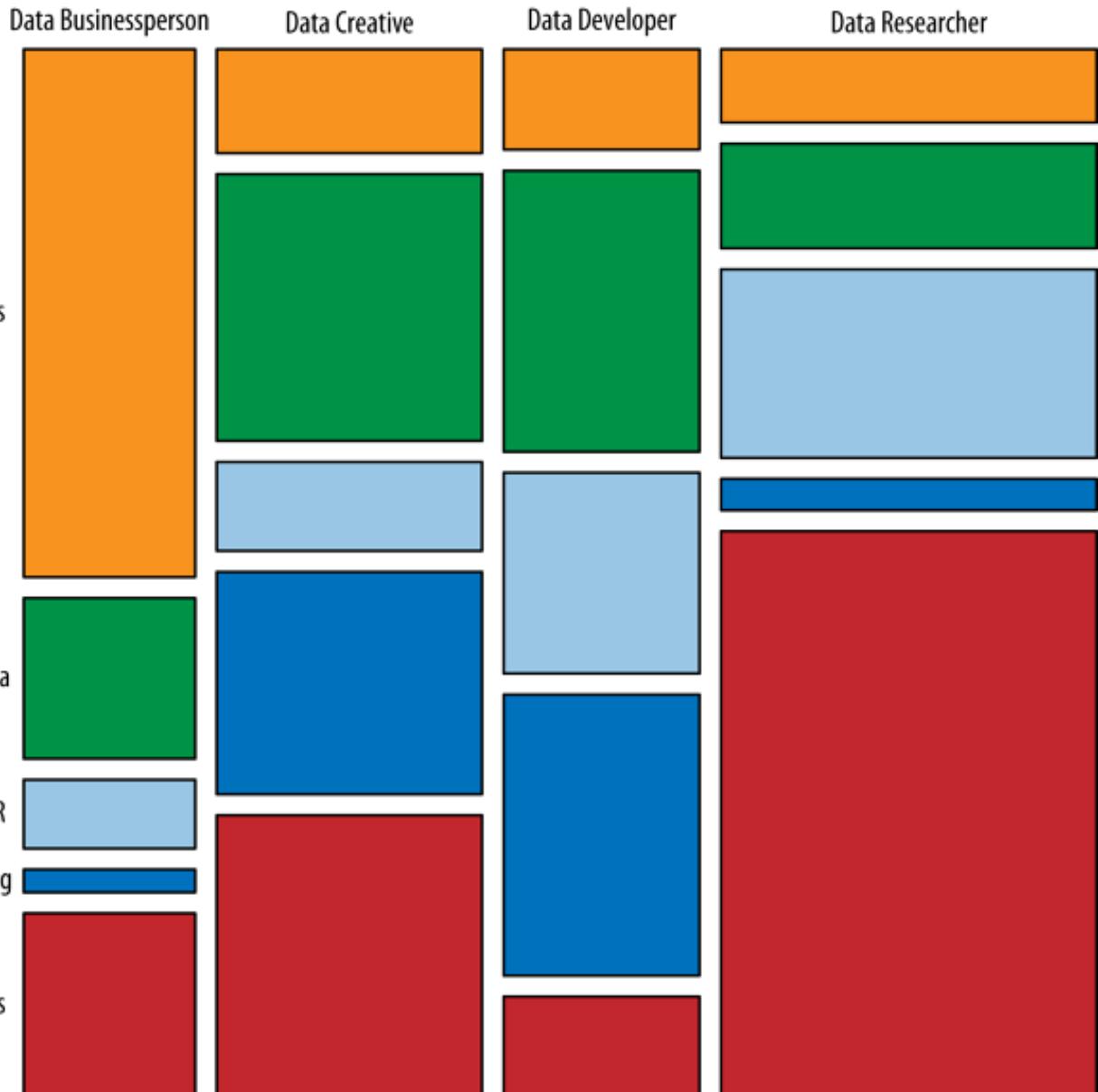
- ▶ Pros:
 - ▶ Math & Statistics depth
 - ▶ Experimental mindset
 - ▶ Highly self-directed
- ▶ Cons:
 - ▶ Rare, expensive
 - ▶ Require business training
 - ▶ Programming skills



Highly Performing Data Teams

- Understand data science is a team sport
- Have engineering resources
- Communicate closely with business decision makers

Skills and Self-ID Top Factors



Team Composition = Roles



*business process,
stakeholder*

*data prep, discovery,
modeling, etc.*

*software engineering,
automation*

*systems engineering,
availability*

introduced
capability

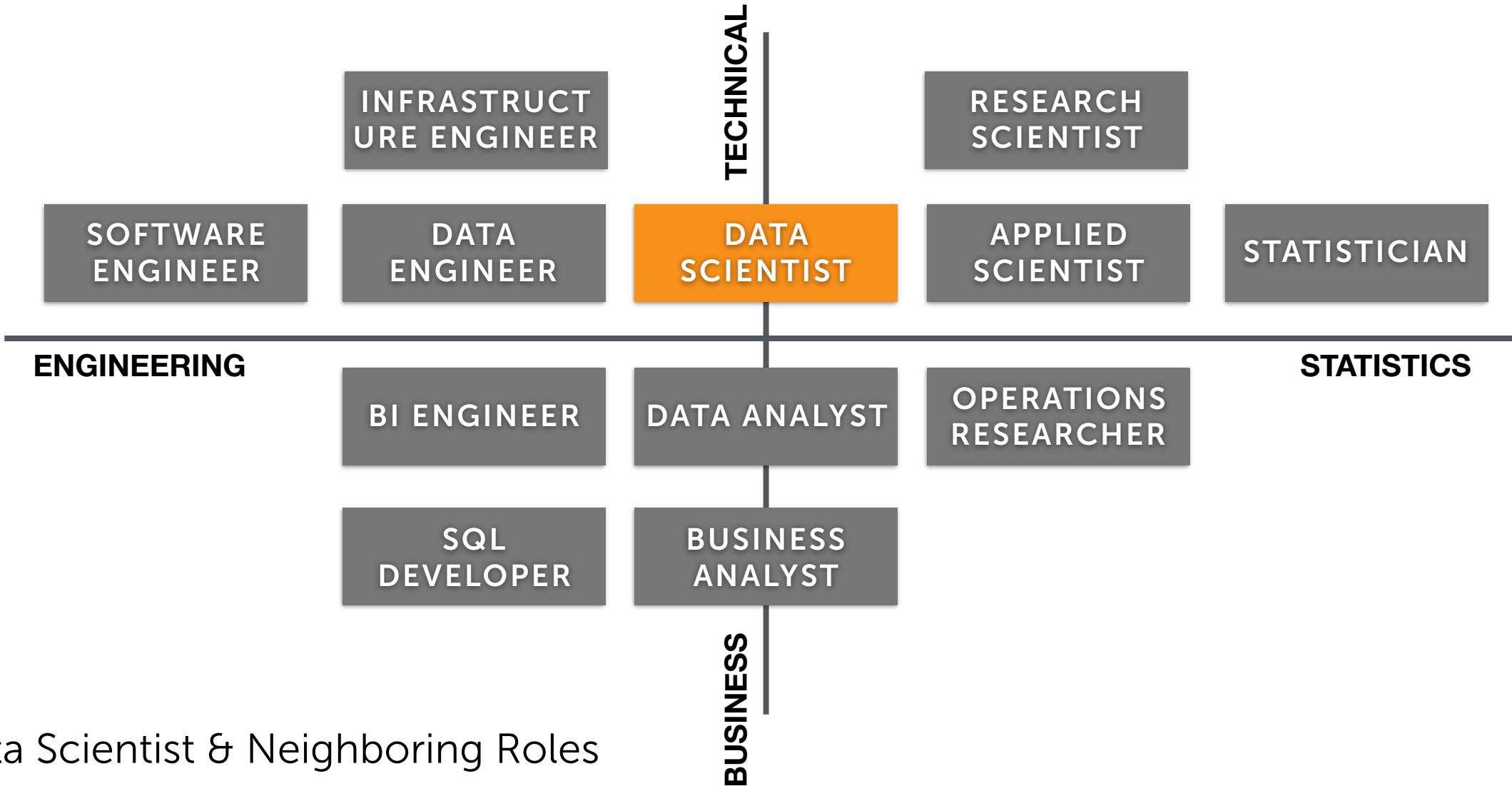
data
science



ABCs: 3 Kinds of Data Scientists

- **A**nalyst
 - Subject matter expert, partners closely with business
- **B**uilder
 - Integrates machine learning into production systems
- **C**onsulting
 - Focused on business strategy and decision science

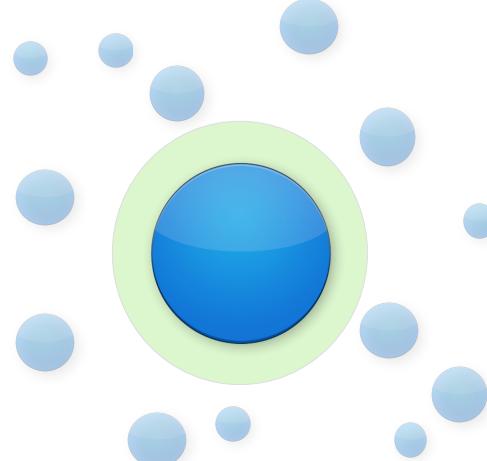
DATA SCIENCE TALENT MAP



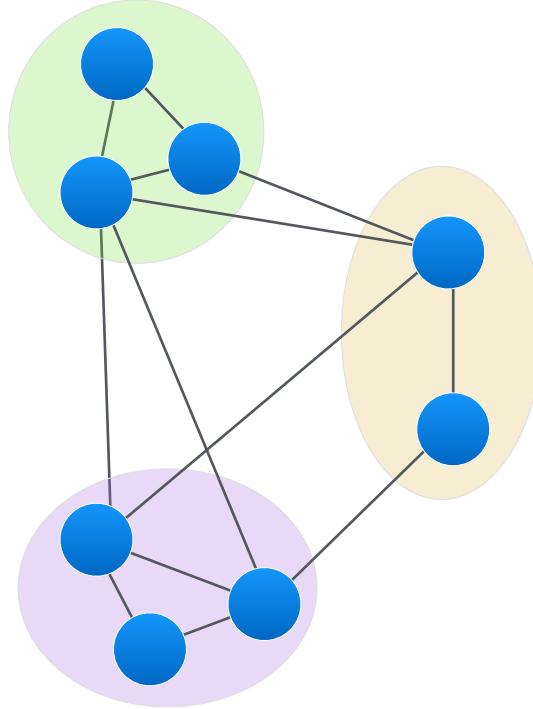


ORG STRUCTURE

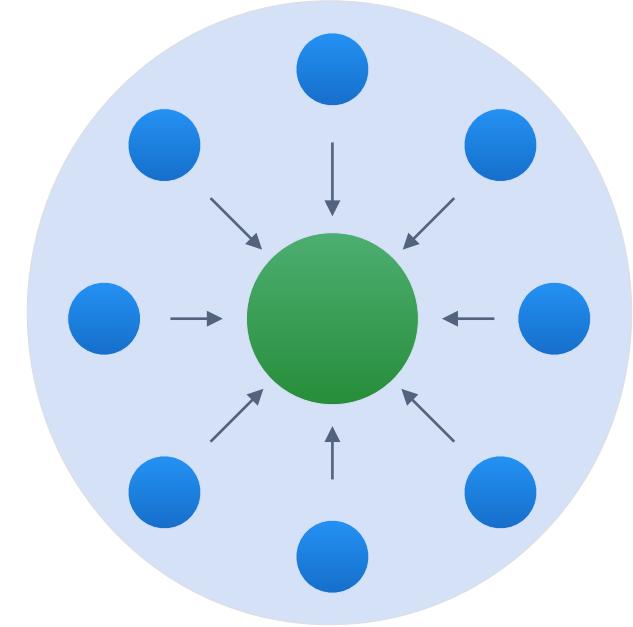
ORGANIZING DATA TEAMS



Centralized



Embedded



Hub & Spoke

3 Models of Data Science Teams

ORGANIZING DATA TEAMS



Centralized Team

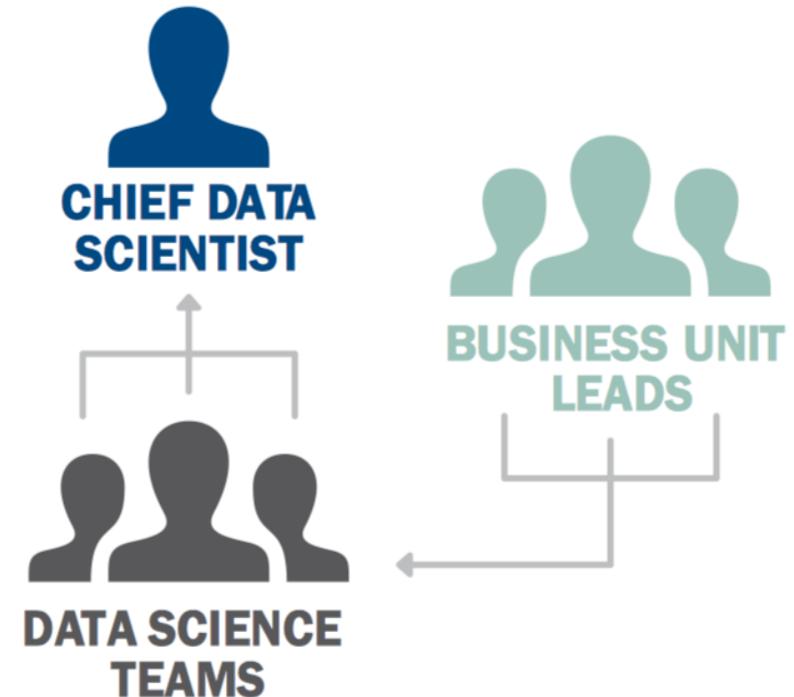


Data Scientists sit on a central team

Act as internal consultants for lines of business

Field and answer questions from multiple teams
within the organization

Define tools for the organization





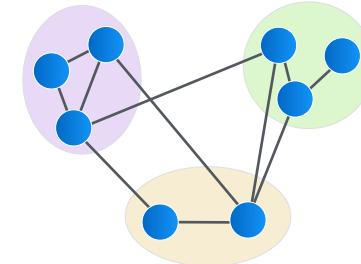
Centralized Model - discussion

- ▶ Write 3 advantages and challenges for the Centralized model
- ▶ What is this model especially good for?
 - ▶ Project Mgmt
 - ▶ Team Work
 - ▶ Education
 - ▶ “Selling Analytics”

ORGANIZING DATA TEAMS

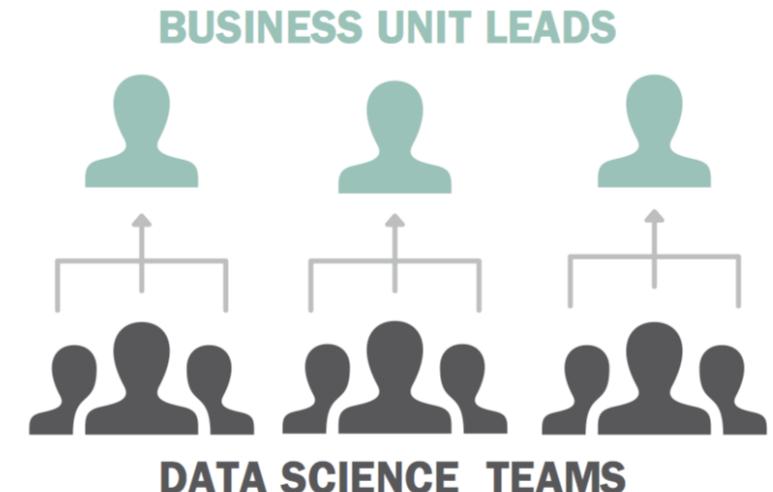


Embedded Team



Data Scientists are almost wholly embedded in a team and solve problems for that line of business, developing domain expertise

May meet with data scientists on different teams at the company for a lunch or study group





Diffused/Embedded Model - discussion

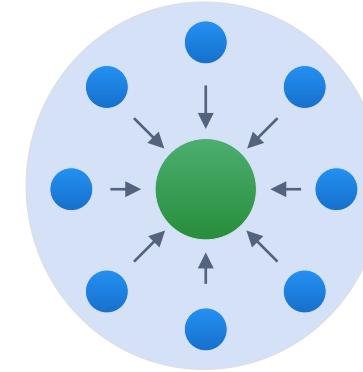
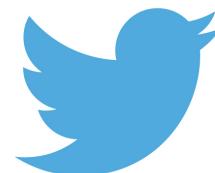
- ▶ Write 3 advantages and challenges for the Diffused model
- ▶ What is this model especially good for?
 - ▶ Governance
 - ▶ Peer Collaboration
 - ▶ Creative Outlets



Hub and Spoke (Hybrid)

Centralized team with decentralized tasks

- Data scientists sit together part-time, share career paths, and collaborate on problems
- Data scientists get deployed to work on specific problems within the organization or with specific teams in a rotational model





Deployed/HS Model - discussion

- ▶ Write 3 advantages and challenges for the Centralized model
- ▶ What is this model especially good for?
 - ▶ Conflict Management (Competing priorities)
 - ▶ Formal Performance Feedback
 - ▶ Rotation
 - ▶ Pipeline



Tradeoff: Data Access

Centralized: data scientists have a very high view of all available data, and may need to use all of it



Decentralized: data scientists may not be aware of existence of data outside their domain





Tradeoff: Prioritization

Centralized: Global data science priorities like tools, processes, knowledge sharing, and talent development can be prioritized

Decentralized: Priorities in decentralized model can be fractured, but function heads can prioritize projects



Tradeoff: Domain Expertise

Centralized: Difficult to have deep domain expertise and sensitivity to a particular area of the business challenging. More difficult fit for large, complex organizations

Decentralized: Data scientists do not benefit from working collaboratively and sharing solutions that may be common across different problems



Shifting Models at Airbnb

SUMMARY

Context: Began centralized and enjoyed the benefits of fast, shared infrastructure and team alignment.

Problem: A fully siloed structure made decision-making less seamless and the team became reactive to the business instead of proactive.

Solution: Transitioned to a hybrid model, where sub teams partner directly with business teams and sit together half-time. Data scientists are more effective as they build domain expertise and enjoy centralized career development

TAKEAWAYS

- Different structures work for different organizations depending on their goals and situation
- Steady-state can often be somewhere along the hybrid spectrum

CASE DISCUSSION



- ▶ Why are Data Scientists frustrated by DE? And vice-versa
- ▶ In what case you do not need a specialized engineering team?
- ▶ Who are the Thinkers, Doers, and Plumbers?
- ▶ How did Stitch Fix do it differently?
- ▶ Should data scientists own the analysis of the data and the outcome of the data all the way to production? What are the benefits?
- ▶ What is the role of an engineer in this new model? and the data scientist?

SUMMARY FOR DEVELOPING DATA SCIENCE TEAMS



Executive Engagement

Data-driven CEO

Chief Data Officer

Organizational Model

Centralized

Decentralized

Hybrid

Developing Data Science Capabilities

Transforming

Creating

As-a-Service

Crowdsourcing

Data Science Team

Data Scientist

BI Analyst

Project Sponsor

Project Manager

Business User

Data Engineer

DBA

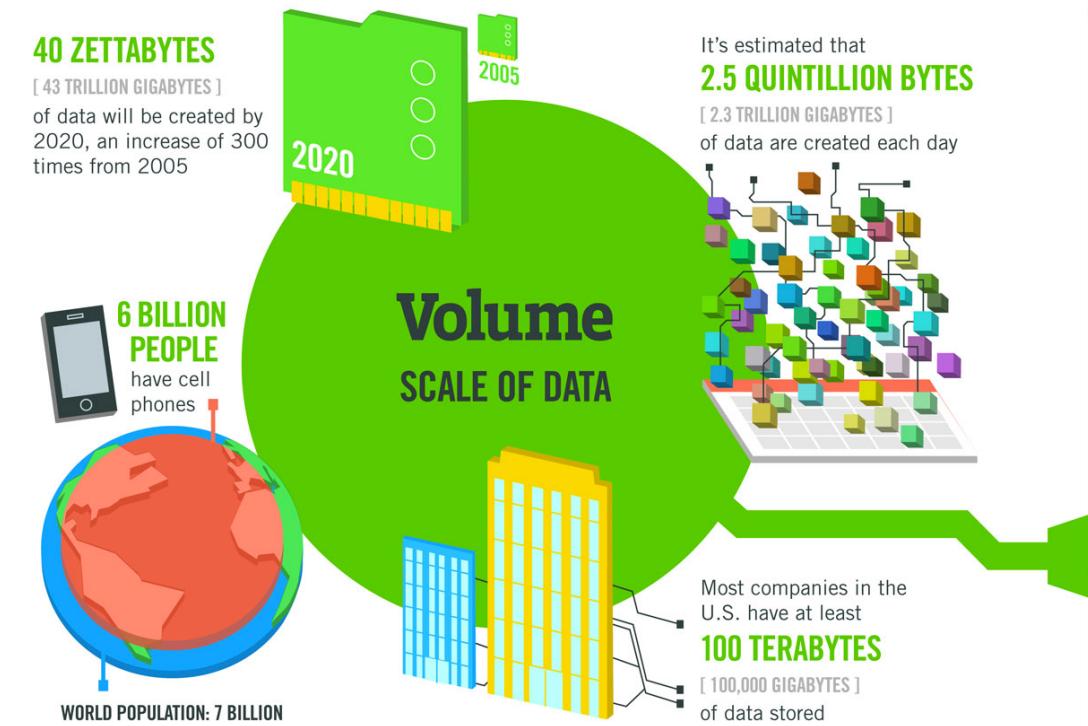


DATA TECH



Data is Changing

- ▶ 90% of data created in past few years
- ▶ Most US companies have 100+ TB
- ▶ Different types of data and volume





Key Data Technologies

- **Cloud** computing & cloud storage (Watson, AWS)
- Big data: **Distributed systems**, Hadoop, MapReduce, Spark
- **Databases**: SQL & NoSQL
- **Software engineering**: Version control (Github)



Key Data Technologies

- Technology is ready and available, low cost
- Well charted territory - competition is high
- New trend: Machine Learning as-a-service (IBM, Uber, Amazon)



Different technology for different tasks

Analytics	Find patterns in data
Statistics	Build predictive models
Machine Learning	Use algorithms to learn and predict
Artificial Intelligence	Studies how to create machine intelligence





DELTTA Method Assesses Maturity

Data, Enterprise, Leadership, Targets, Technology, Analysts

Published by Tom Davenport of Harvard, MIT, and Babson in *Analytics at Work* and *Big Data at Work*

Designed to inform a data strategy

Levels 1 to 5 for each part of the Assessment



DELTTA(C) Method

- **D**ata - the answers for better decisions
- **E**nterprise - degree of needed collaboration
- **L**eadership - most important
- **T**argets - questions
- **T**echnology
- **A**nalytics Talent
- **C**ulture



5 Levels Rubric: Data Example

Level	Description
<i>Impaired</i>	Fractured, poor quality, no governance, hard to query & integrate
<i>Localized</i>	Local, transactional systems
<i>Aspirational</i>	Some enterprise-wide targets & strategy, domain-specific data warehouses, short-term goals
<i>Analytical</i>	Cross-functional, enterprise data warehouse & data integration; senior leadership drives targets & strategy; has long-term vision
<i>Competitor</i>	Organization-wide data culture, strong governance, data provides differentiating advantage



UNDERSTANDING DATA ROI

DATA ROI

Media: Huffington Post

How leveraged data

- Huffington's leadership believes in **running the business based on data**.
- This includes improving the **user experience in real-time** from social trends, **recommendations**, **moderation**, and **personalization**
- They **optimize** the site many ways, and their analytics platform powers the entire analytical process.

Business Benefits (ROI)

Huffington Post grew last year into the number one online news site in the United States.

Increase overall traffic & engagement

DATA ROI

Airlines: Southwest

How leveraged data

- Southwest uses **speech analytics** to help improve the interactions between customers and personnel.
- Southwest uses big data to understand **online behaviors** and actions, improving offers for customers and leading to growth in loyalty year over year.

Business Benefits (ROI)

For companies focused on customer relationships, providing great service is top of mind via social channels and other interactions.

Increase in customers satisfaction and overall tickets sold.

DATA ROI

Media: FT.com

How leveraged data

- FT.com uses data to understand and serve the customer better, create **targeted advertising**, and design new products based on information collected.
- Their CEO claimed that big data transformed their business. The company uses many data points to **analyze customer content preferences, increase relevance in their communications, and personalize the content**—all to keep visitors and traffic.
- The data also helps the company understand **time of day consumption** based on both PC and mobile channels.

Business Benefits (ROI)

Increase keep visitors and traffic

Telecommunications: Sprint

How leveraged data

- Sprint spoke about using big data analytics to **improve quality and customer experience** while reducing network error rates and customer **churn**.
- They handle 10s of billions of transactions per day for 53 million users, and their **big data analytics put real-time intelligence** into the network.

Business Benefits (ROI)

Driving a 90% increase in network capacity.

The project helped identify service issues and avoid needless, costly repair work.

DATA ROI

Financial Services: AMEX

How leveraged data

- The American Express Company looked to shift traditional **business intelligence**-based hindsight reporting or trailing indicators of how business was doing to **predict loyalty**.
- Their sophisticated **predictive models** analyzed historical transactions with 115 variables to forecast potential **churn**.

Business Benefits (ROI)

In the Australian market, they now believe they can identify 24% of accounts that will close within four months.

Online Training and Gaming: Skillsoft

How leveraged data

- Skillsoft is using big data to **learn and apply knowledge** across 19 million users and 60,000 learning assets.
- Content has been **individualized** based on **direct email response behavior** and **surveys**.
- The company's leadership see that the **analysis of big data** has generated substantial results and trusted **advanced machine learning** and **optimization algorithms** to deliver.

Business Benefits (ROI)

Since applying big data approaches, there has been a 128% improvement in user engagement and recommendations have proven to be much more relevant and actionable.

SMALL GROUP CHALLENGE

- ▶ Using the company examples and the model definitions, identify 3 data science use cases that UNICEF should adopt?
- ▶ What is an immediate action that you as a manager can take to ensure successful execution?



FINANCE

FINANCIAL SERVICES: AIG

How leveraged data

- American International Group (AIG) uses big data and **data visualization** to help fight **fraud**.
- The system takes **structured and unstructured data** from claims databases and handwritten adjuster notes to **identify potential fraud**.
- Besides listing **priority claims** to investigate, charts and visualizations, like heat maps, inform teams of other **insights** and also help them make improvements to machine learning algorithms.

Business Benefits (ROI)

Avoid losing money due to fraudulent behavior



OPERATIONS

DATA ROI

RETAIL: WALGREENS

How leveraged data

- At Walgreens, big data is being used by clinicians at in-store health clinics.
- The company is delivering **advanced analytics** at the point of care to better assess patient conditions and provide **recommendations** that improve health overall and avoid future medical costs.
- For example, a current system can catch an unfilled prescription to help people stick to their healthcare plans and avoid further, unnecessary costs.

Business Benefits (ROI)

Over 7.5 billion medical events for 100 million people power the big data system with information like demographics, enrollment, diagnoses, procedures, and data from managed-care plans.

DATA ROI

AIRLINES: DELTA

How leveraged data

- Delta has used big data to help with one of the most uncomfortable travel situations that exists—lost baggage.
- With over 130 million bags checked per year, the company held a lot of **tracking data** about bags and became the first major airline to allow customers to track their bags from **mobile** devices.

Business Benefits (ROI)

To date, the app has been downloaded over 11 million times and gives customers much greater peace of mind while traveling while also differentiating Delta as a customer-centric company.

LOGISTICS: UPS

How leveraged data

- On a daily basis, UPS makes 16.9 package and document deliveries every day and over 4 billion items shipped per year through almost 100,000 vehicles.
- With this volume, there are numerous ways UPS uses big data, and one of the applications is for **fleet optimization**.
 - On-truck telematics and advanced algorithms help with **routes**, **engine idle time**, and **predictive maintenance**.

Business Benefits (ROI)

Since starting the program, the company has saved over 39 million gallons of fuel and avoided driving 364 million miles. The next steps include completion of the roll-out and applying the operational efficiency to their airplanes.

DATA ROI

HEALTHCARE: KAISER

How leveraged data

- At Kaiser Permanente, big data from **electronic health records** for 9 million people have been used to help **improve care and reduce costs, improve recommendations** for care, **support adherence to prescriptions**, and more.
- Recently, Kaiser used big data to study the incidence of blood clots within a group of women taking oral contraceptives.
 - The analysis revealed that one formula contained a drug that increased the threat of blood clots by 77%—understanding these types of patterns can help many people avoid visits to the doctor or emergency room.

Business Benefits (ROI)

Improve patient health, reduce costs, make recommendations



PRODUCT

DATA ROI

AUTOMOTIVE: TESLA

How leveraged data

- Tesla is the poster child for instrumenting vehicles with **sensors** and sending all the data back to the mother ship for **analysis**, using a big data technology (Apache Hadoop® cluster) to **collect the data**.
- For example, the company is **notified** if the car is not functioning properly and consumers can be **advised to get a service**.

Business Benefits (ROI)

The data is used to improve the company's R&D, car performance, car maintenance, and customer satisfaction.

These capabilities have helped Tesla create market share in a difficult environment where charging stations are not widely deployed.

DATA ROI

AUTOMOTIVE: FORD

How leveraged data

- Ford's CEO explained how the company is using **big data** and **transportation analytics** to become more of a technology company.
- They are looking to use big data to **address vehicle quality, insurance costs, transportation, vehicle intelligence, driving patterns**, and more.
- For example, they use **data to help lower insurance costs** for the driver and even has approximately 200 big data and analytics experts supporting major decisions throughout the company.
- In a marketing, the company **analyzes multiple data streams** on what was built, sold, in inventory at the time of sale, and what customers are searching for on websites along with economic data such as housing starts and employment rates—all of this is used to help sell more cars.

Business Benefits (ROI)

Technology company sift
better car experience

Sell more cars