INSIGHTS DRIVING ACTIONS:

How to Get Real Results from your Data Science Program

wisdom 2018
new orleans

CLARKSTON CONSULTING

Agenda

01

TYPICAL ROADBLACKS TO SUCCESS

- Identifying the right use case
- Defining success criteria
- Getting the right data

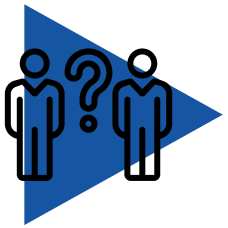
02

PROOF OF VALUE

03

HOW TO STRUCTURE + MANAGE A SUCCESSFUL PROJECT

Identifying the Right Use Case



ASKING THE RIGHT QUESTION CAN BE ONE OF THE HARDEST PARTS OF A DATA SCIENCE PROJECT

- Putting in time up front to ask the right question will save you time later down the line

MACHINE LEARNING ISN'T MAGIC

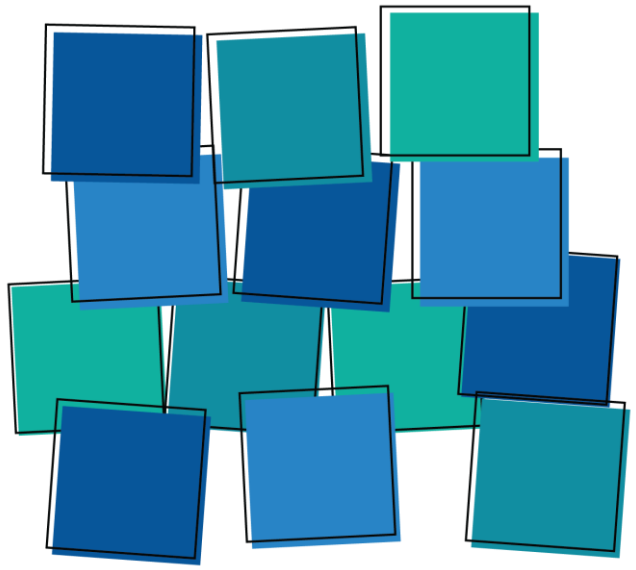
- Data Science Translator is a necessary skillset of the data science team to define good problem statements

WHAT MAKES A GOOD PROBLEM STATEMENT?

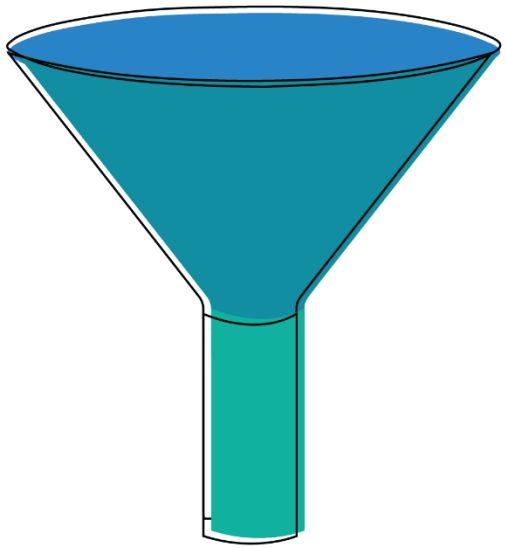
- "If this hypothesis is proven true, it will mean x for our business"

Design Thinking

Design thinking is an approach which allows you to set aside preconceived notions and think outside the box .



01



02

CP client transforming their business strategy to focus on a new customer segment

The questions that they asked us were:

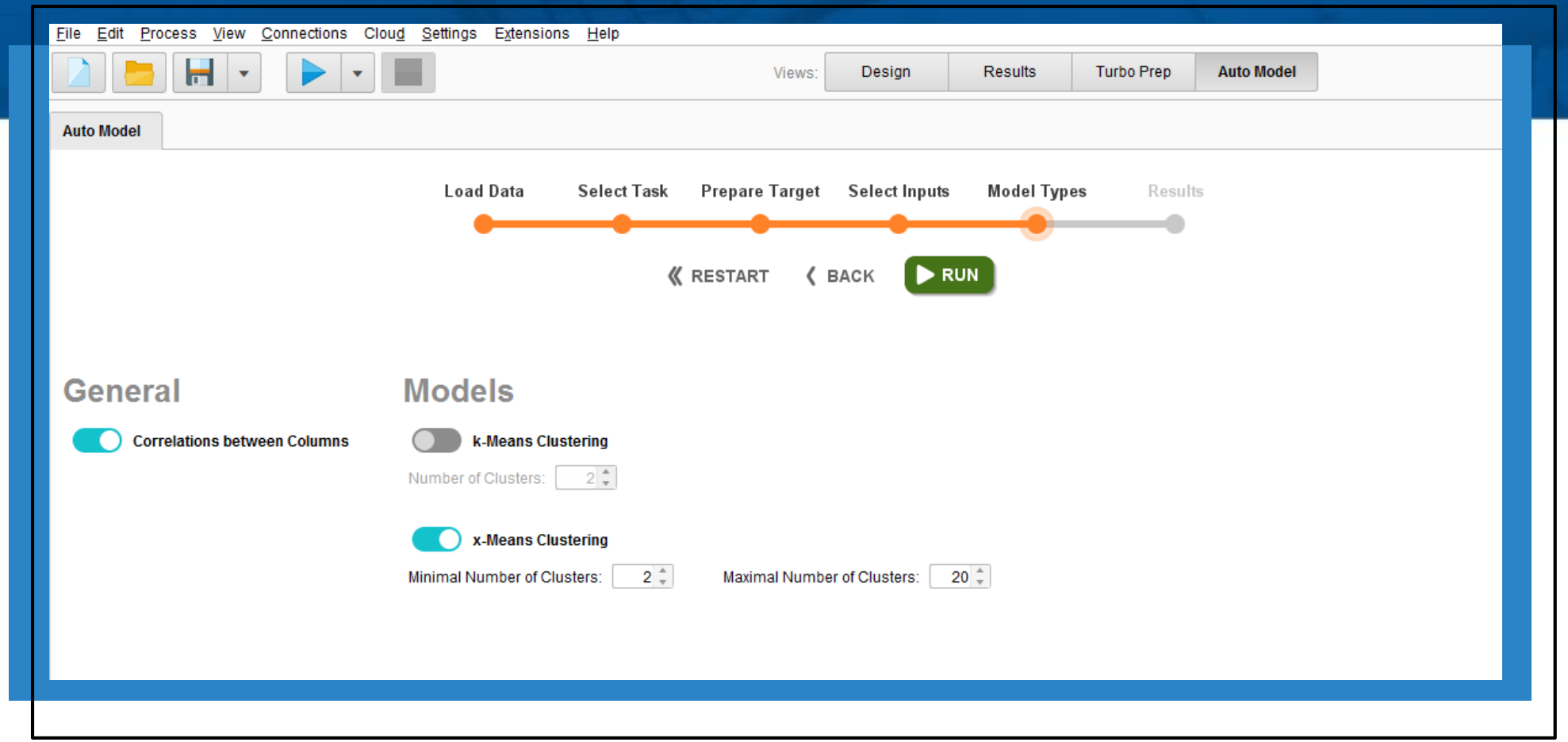
- Why aren't we more focused on our target customer's buying experience?
- Why won't our target customers buy online?
- Why do we believe that our target customer wants a relationship with us as a manufacturer?
- Why don't we incentivize retailers to share POS data?
- Why don't we have more products in the portfolio that our target customers need?

The problem statement we crafted was:

How can we use the data we have today to find out what the different customer segments are buying and identify our target customers in the data itself?

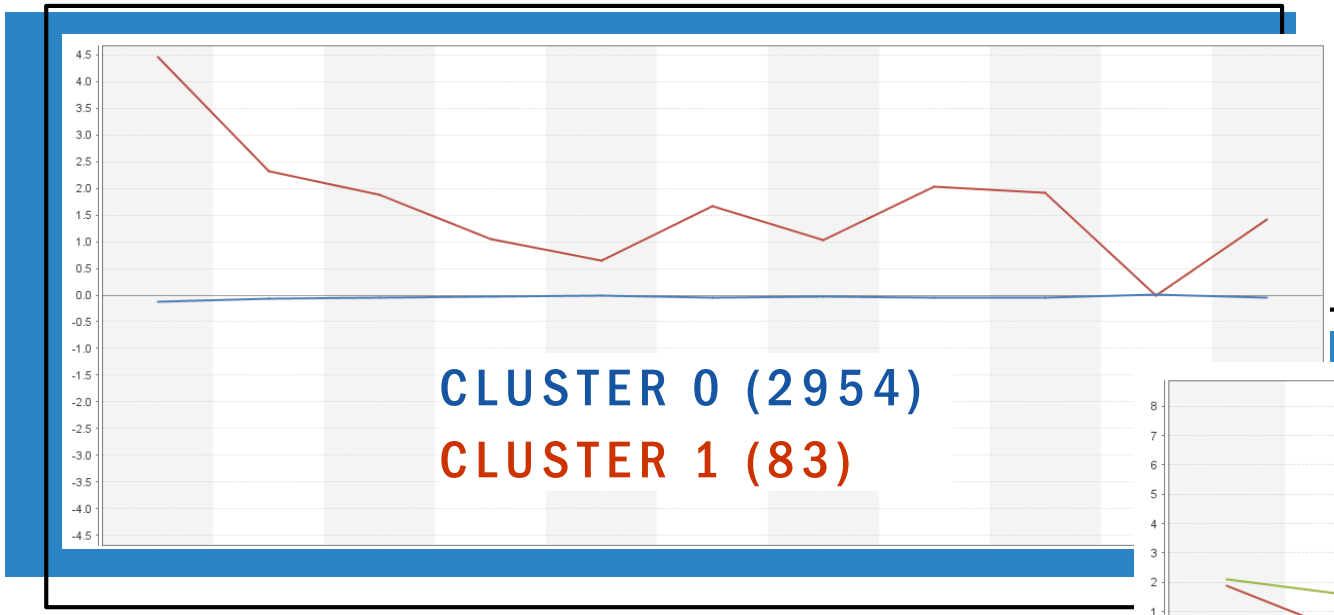
How we used RapidMiner for a bottom-up approach

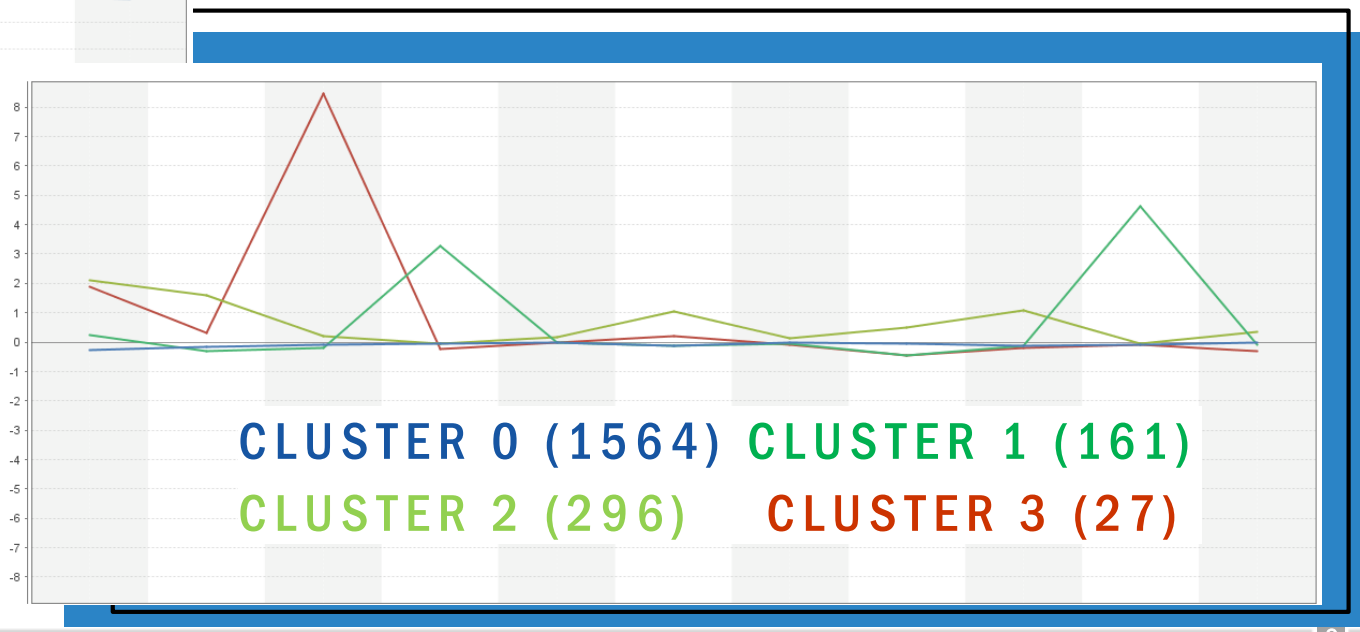
- Clustering to look for buying patterns in their POS data to separate out different customer segments
- Tip: Auto Model using x-Means made this process quick to iterate and decide which features we wanted to use



Tip for Clustering

- Peel the layers back approach: look at the initial clusters in your data, and if there is one or more clusters with a large portion of the data, separate that cluster off and re-cluster that group on its own to discover the next level of insights





Expectations vs. Reality

Though you do want to develop a problem statement/use case with the business, sometimes you still get ambiguous requests that you need to structure.

- It's about turning the question "what is interesting about this data" into specific questions and problems to solve

We had a food manufacturer give us their sales data and ask for interesting insights, but it was a very small piece of the overall trade promotions technology implementation. We needed to pare this request down to specific questions, and we did this through the crucial step of data profiling: looking at trends and spikes in the data, looking for patterns and what may drive them, and iteratively generating those basic questions we could go after with RapidMiner.

- We took the vague request and structured 2 specific problems from it.

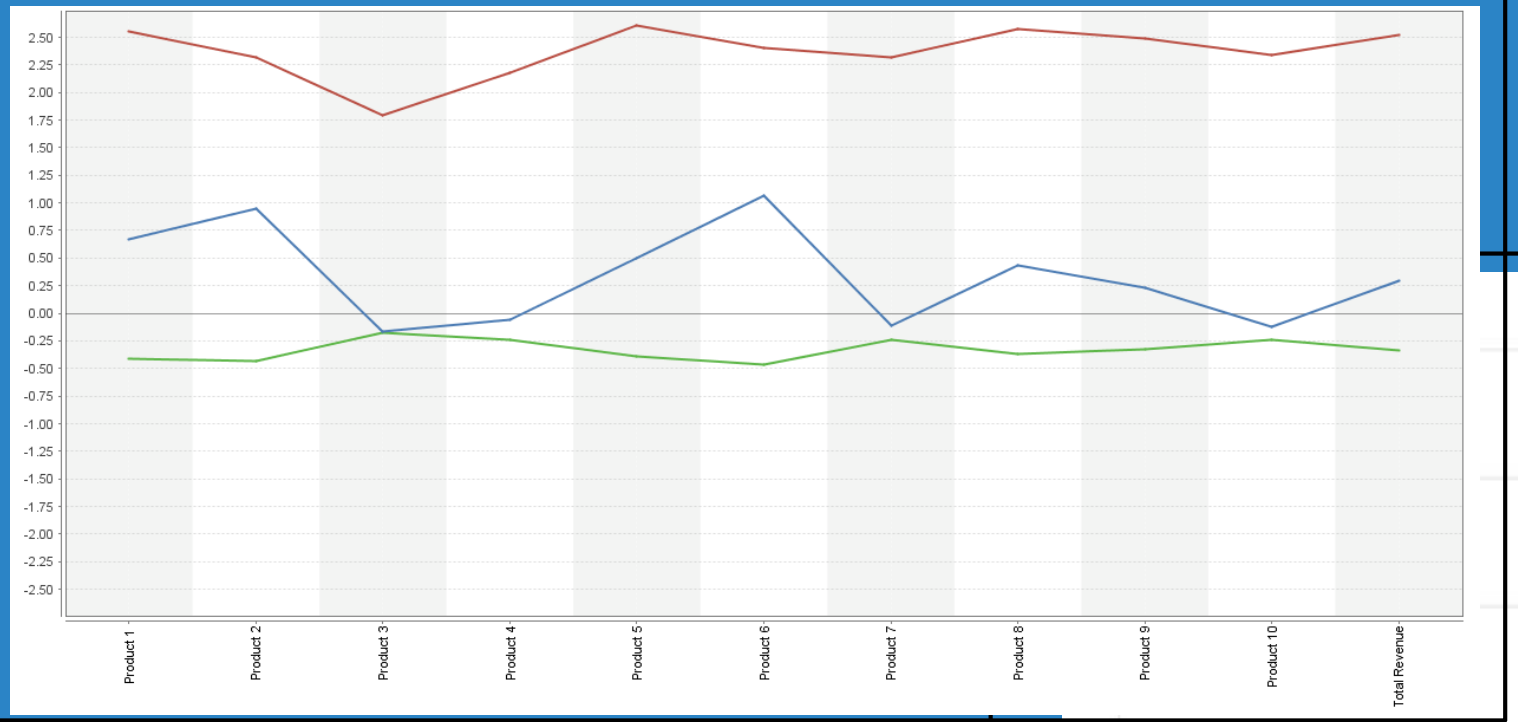
Can we identify buying patterns to optimize promotions?

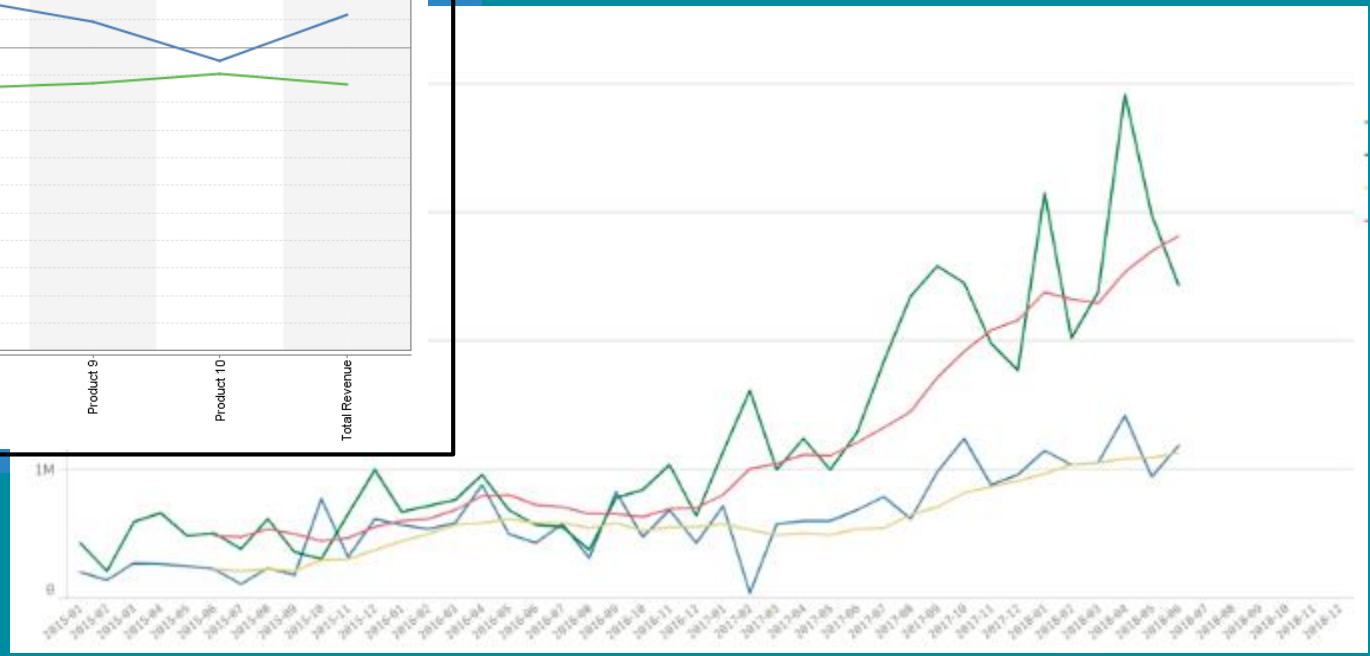
Forecast total revenue for the next 6 months.

Data profiling is a great exercise for learning and posing interesting questions about your data.

Data Profiling of Time Series

- Use clustering to identify different customer segments, plot actuals against moving average for those groups





Defining Success

QUANTIFYING ROI + BUSINESS IMPACTS

DEFINING AN OUTPUT: INSIGHTS TO ACTIONS

DEFINE WHAT IS GOOD ENOUGH TO DEPLOY

Defining Success

QUANTIFYING ROI + BUSINESS IMPACTS

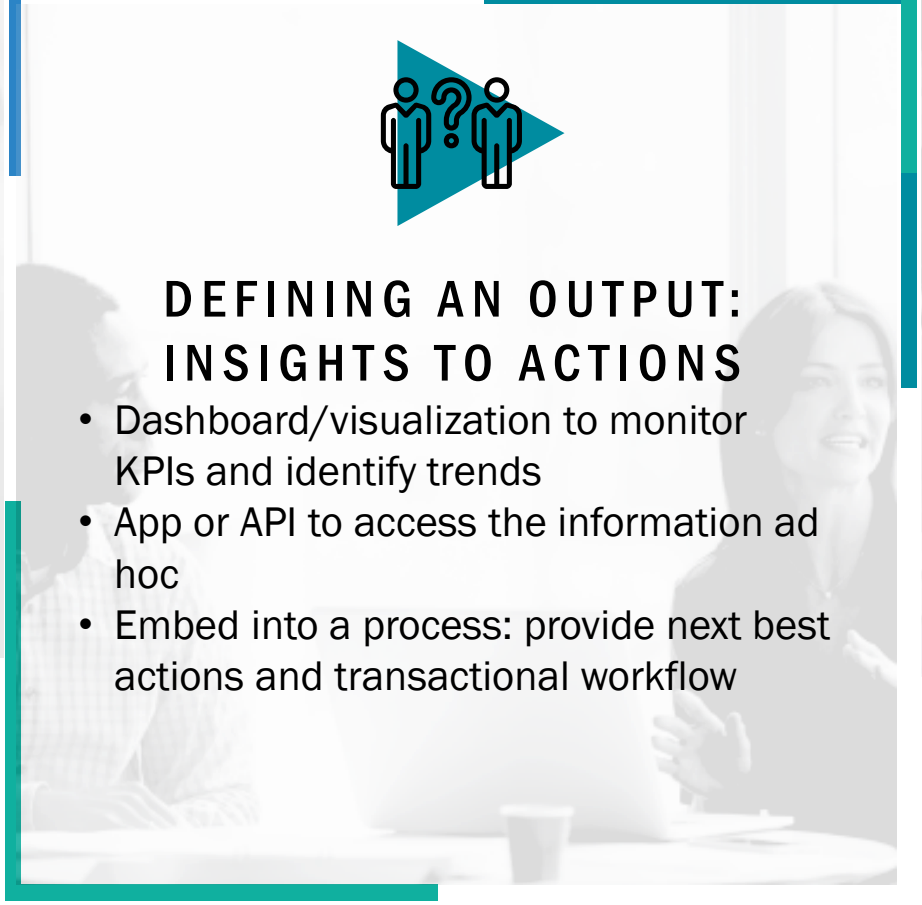
- Helpful in evangelizing data science in the organization
- Move away from slow moving, low-impact projects. Data science can sometimes drift into a curiosity support function

DEFINING AN OUTPUT: INSIGHTS TO ACTIONS

DEFINE WHAT IS GOOD ENOUGH TO DEPLOY

Defining Success

QUANTIFYING ROI + BUSINESS IMPACTS



DEFINING AN OUTPUT: INSIGHTS TO ACTIONS

- Dashboard/visualization to monitor KPIs and identify trends
- App or API to access the information ad hoc
- Embed into a process: provide next best actions and transactional workflow

DEFINE WHAT IS GOOD ENOUGH TO DEPLOY

Defining Success

QUANTIFYING ROI + BUSINESS IMPACTS

DEFINING AN OUTPUT: INSIGHTS TO ACTIONS

DEFINE WHAT IS GOOD ENOUGH TO DEPLOY

- Is the solution better than what you have today at 75% accurate? At 85%? Many data scientists get hung up on optimizing models until reaching that 99% accuracy
- Think of it as measuring lift of your analytics project, where your baseline is doing no analytics at all.
- “All models are wrong, but some are useful”

3 support pillars of successful data science projects

01 DATA GOVERNANCE

- Critical to trust the data quality that we're inputting to models and use a common language and metrics

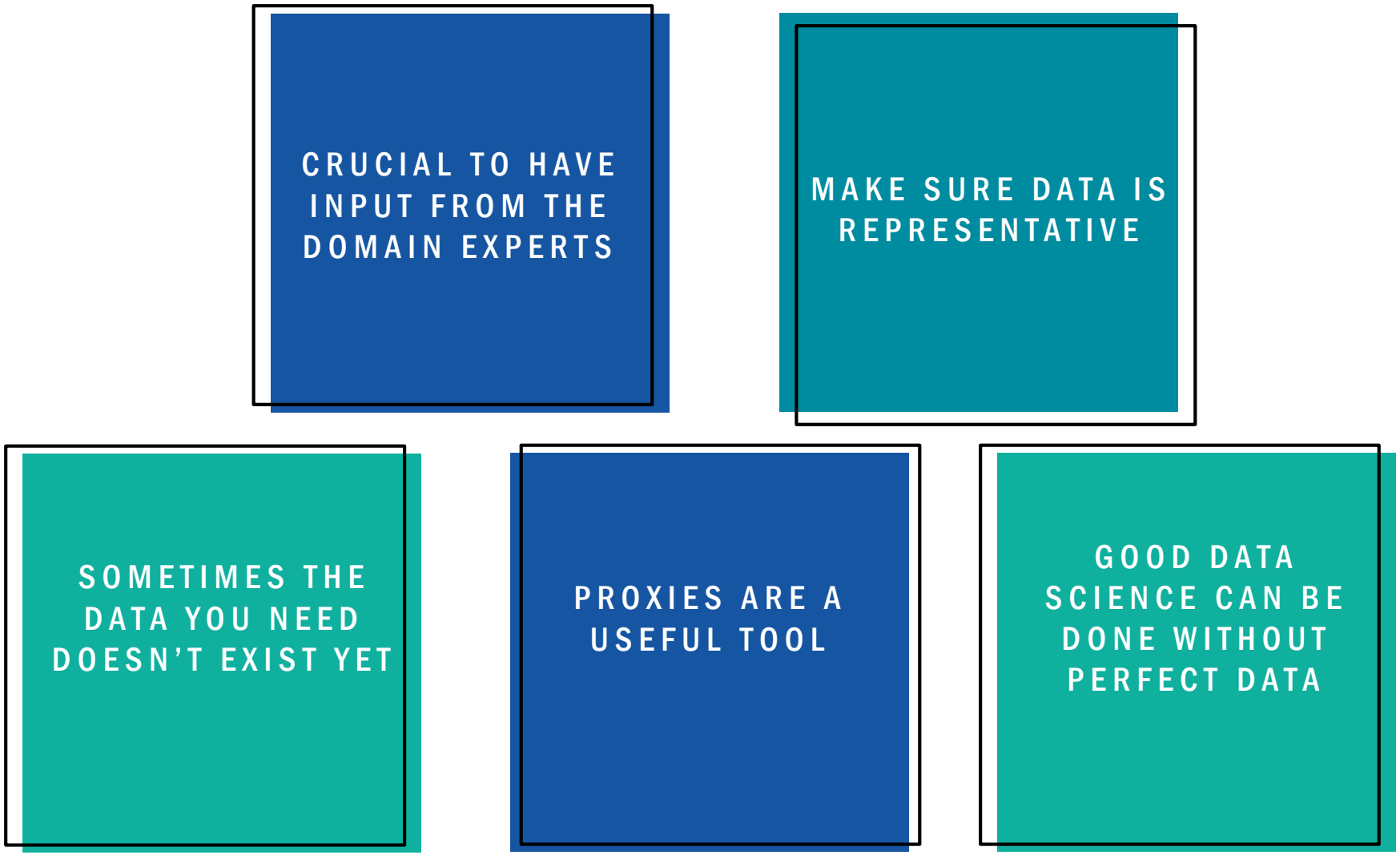
02 ORGANIZATIONAL ELEMENT

- There is a shift in how analysts work from reporting -> analysis -> predictive work and creating solutions

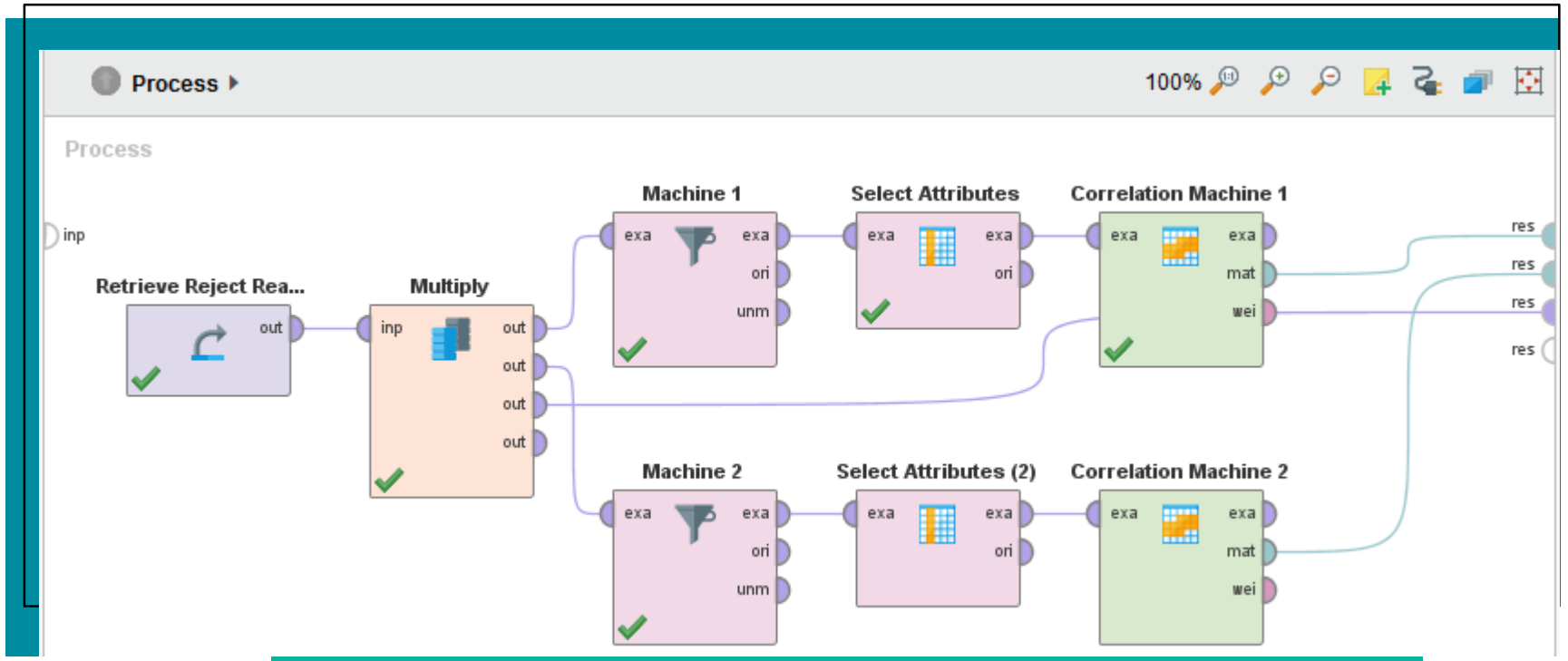
03 DECISION MAKING PROCESS

- How are we leveraging insights or not?
- Having executives and decision makers involved from the outset of a project is critical in activating data science projects and not just having them sit on the shelf
- Stop playing telephone: Getting analysts closer to decision makers is important

Getting the Right Data



Proxies and correlation



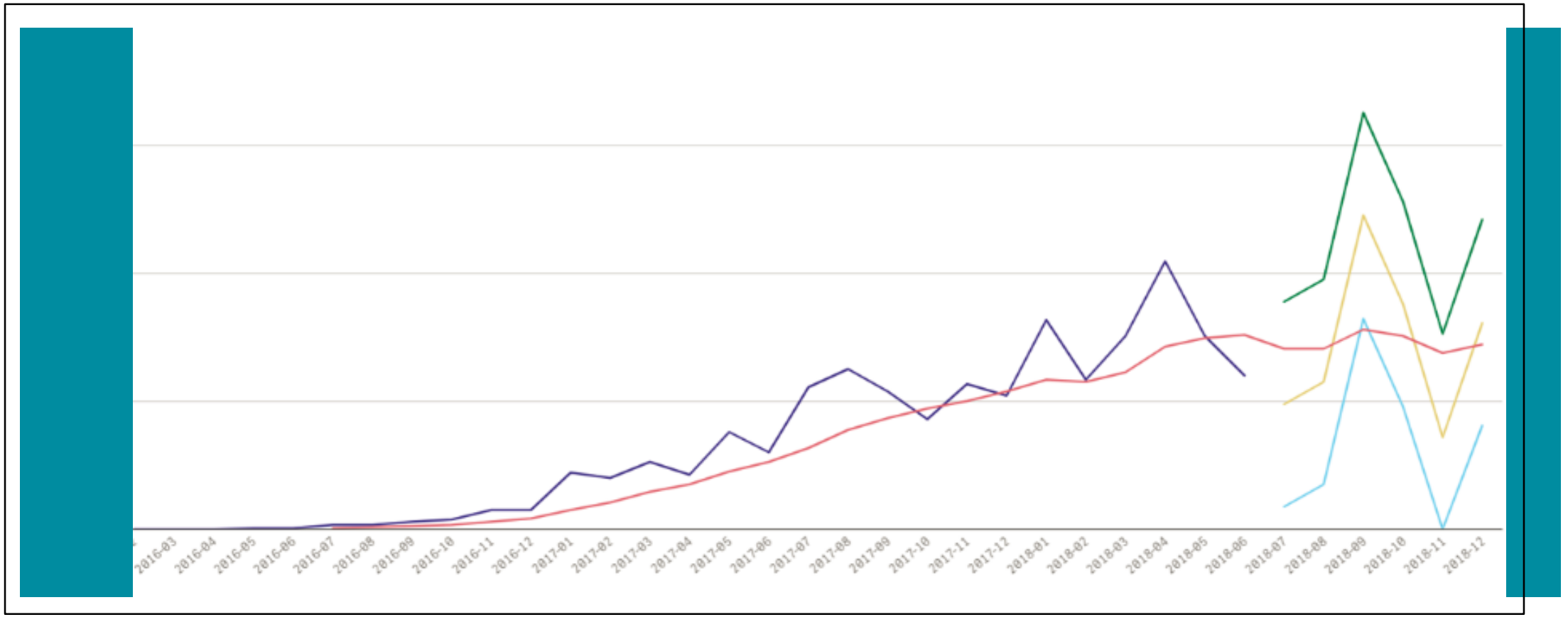
Attributes	Shift	Total Rejects	Stop Count	Stop Duration	Machine Speed
Shift	1	0.183	-0.057	-0.192	-0.187
Total Rejects	0.183	1	0.824	0.076	0.625
Stop Count	-0.057	0.824	1	0.317	0.704
Stop Duration	-0.192	0.076	0.317	1	0.019
Machine Speed	-0.187	0.625	0.704	0.019	1

EXAMPLE

- A large tobacco manufacturer sought a better understanding of their manufacturing efficiency, ultimately looking to establish a baseline yield and waste.
- We found correlation between their product rejections and machine speeds, a great opportunity for an optimization problem

Non ML uses for RapidMiner

- Using time series forecasting statistical methods like ARIMA and Holt-Winters



Proof of Value

You don't need a major overhaul or months-long analytics tool implementation to see results and develop an analytics roadmap and strategy.

DON'T SKIP STEPS!

- You still need the major steps of a successful analytics project: define the problem statement, profile the data, model, validate the model and the success criteria.
- Limit the scope of the data or the deployment

THE SOLUTION SHOULD BE SCALABLE

- Translate the outcomes of the POV into the business value to illustrate the potential
- As data scientists, we get excited about the outputs and results, but it's the business value that guides the decision to scale

Clarkston's Methodology: Rapid Analytics Delivery

01 INITIATE/DATA PLAN

- ALIGN TEAM
- CREATE A DATA PLAN

02 MODEL BUILD/TEST

- FEATURE SELECTION
- MODEL BUILD + TEST

03 SCALE + DEPLOY

- REFINE MODEL + OUTPUT
- DEPLOY
- CLOSEOUT

Clarkston's Methodology: Rapid Analytics Delivery

01 INITIATE/DATA PLAN

ALIGN TEAM

- Review the business problem
- **Develop/Define success criteria**
  - Accuracy & performance
  - Quantify business impact
  - Output Defined
- Establish roles and responsibilities
- Confirm technologies/methods
- Communicate project intent and plan to key stakeholders

CREATE A DATA PLAN

- Identify all data sources and access
- **Data profiling and understanding**
- Establish/build data transformation and cleansing processes
- Establish/build data pipelines
- Establish model progression and initial runs

02 MODEL BUILD/TEST

- FEATURE SELECTION
- MODEL BUILD + TEST

03 SCALE + DEPLOY

- REFINE MODEL + OUTPUT
- DEPLOY
- CLOSEOUT

Clarkston's Methodology: Rapid Analytics Delivery

01 INITIATE/DATA PLAN

- ALIGN TEAM
- CREATE A DATA PLAN

02 MODEL BUILD/TEST

FEATURE SELECTION

- Define features to acquire vs. generate
- Establish/build data blending and feature engineering processes
- Correlation analysis
- Verify selections with the business

MODEL BUILD + TEST

- Prototype model and iterate
  - Identify accuracy metrics
  - Preprocessing and normalization
  - **Verify initial results with business**
- Test and Tune the model
- Optimize parameters
  - Validate and fine tune accuracy
- **Share initial findings with stakeholders**

03 SCALE + DEPLOY

- REFINE MODEL + OUTPUT
- DEPLOY
- CLOSEOUT

Clarkston's Methodology: Rapid Analytics Delivery

01 INITIATE/DATA PLAN

- ALIGN TEAM
- CREATE A DATA PLAN

02 MODEL BUILD/TEST

- FEATURE SELECTION
- MODEL BUILD + TEST

03 SCALE + DEPLOY

REFINE MODEL + OUTPUT

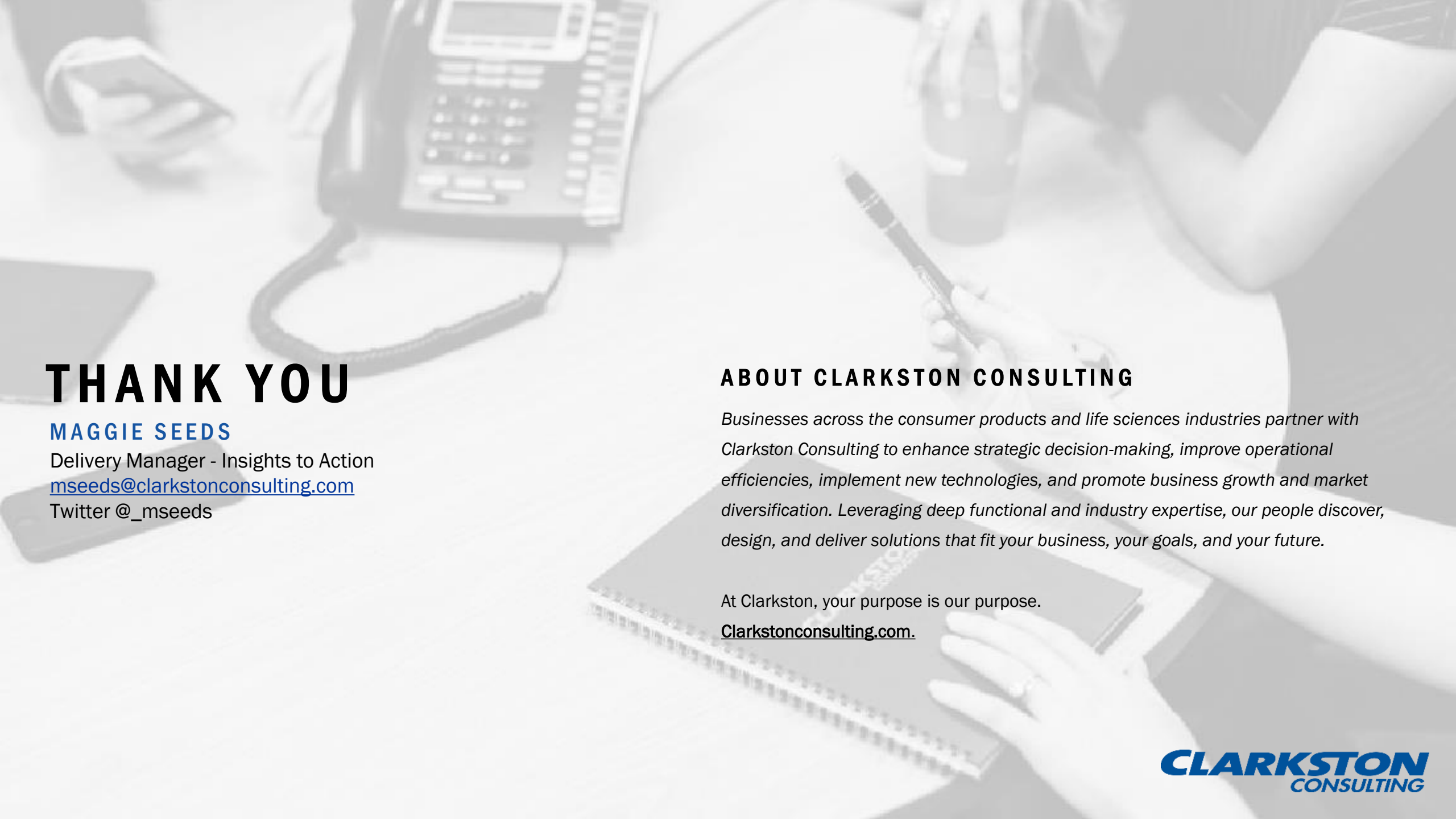
- Develop output
- Finalize production pipelines
- Fully scale and refine if necessary
- **Verify results with business**

DEPLOY

- Finalize output
- Validate success criteria and measure ROI
- **Verify results with business**
- **Production deployments**

CLOSEOUT

- Provide recommendations
- Report findings
- Knowledge transfer
- Production Tuning
- **High-Fives**



THANK YOU

MAGGIE SEEDS

Delivery Manager - Insights to Action
mseeds@clarkstonconsulting.com
Twitter @_mseeds

ABOUT CLARKSTON CONSULTING

Businesses across the consumer products and life sciences industries partner with Clarkston Consulting to enhance strategic decision-making, improve operational efficiencies, implement new technologies, and promote business growth and market diversification. Leveraging deep functional and industry expertise, our people discover, design, and deliver solutions that fit your business, your goals, and your future.

At Clarkston, your purpose is our purpose.
Clarkstonconsulting.com.

CLARKSTON CONSULTING