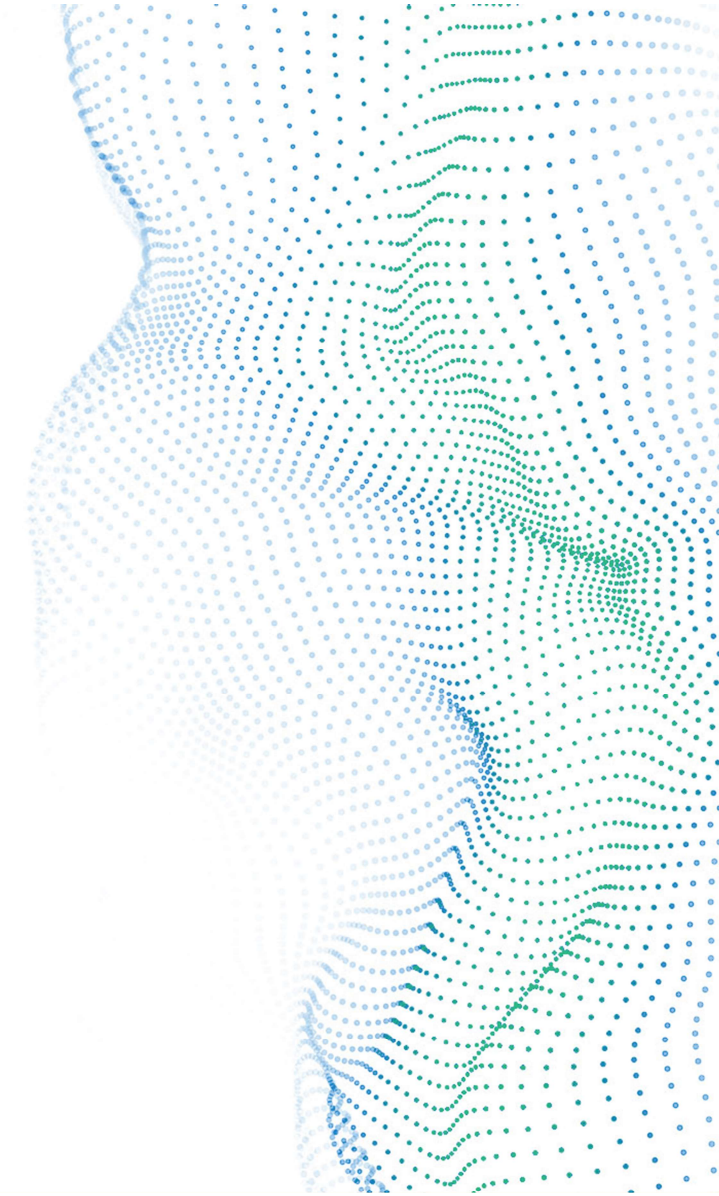
alteryx

Business Analytics: Getting Back to Basics

Dr. Dan Putler | Chief Data Scientist

May 2022

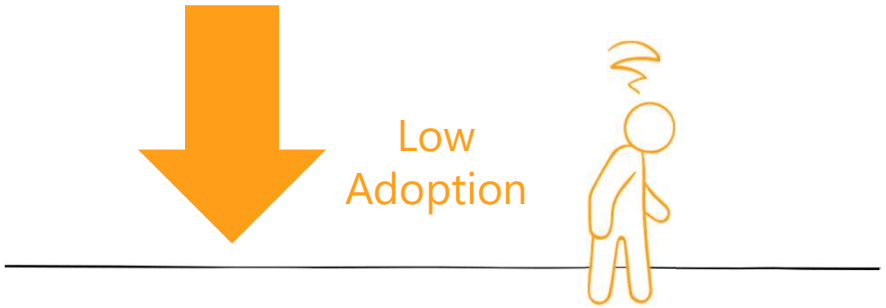


Agenda

1. The current state of analytic maturity for predictive analytics and ML
2. Developing a successful analytics project with 7 questions
3. The team skill set needed to answer the 7 questions
4. Applying the framework: Supply chain lead-time forecasting example

alteryx

Current Analytic Maturity for Predictive Analytics and ML



Common Challenges

- The shortage of trained data scientists
- Challenges in data infrastructure and integration
- Outcomes associated with predictive and ML projects often miss the mark

It is better to be vaguely right

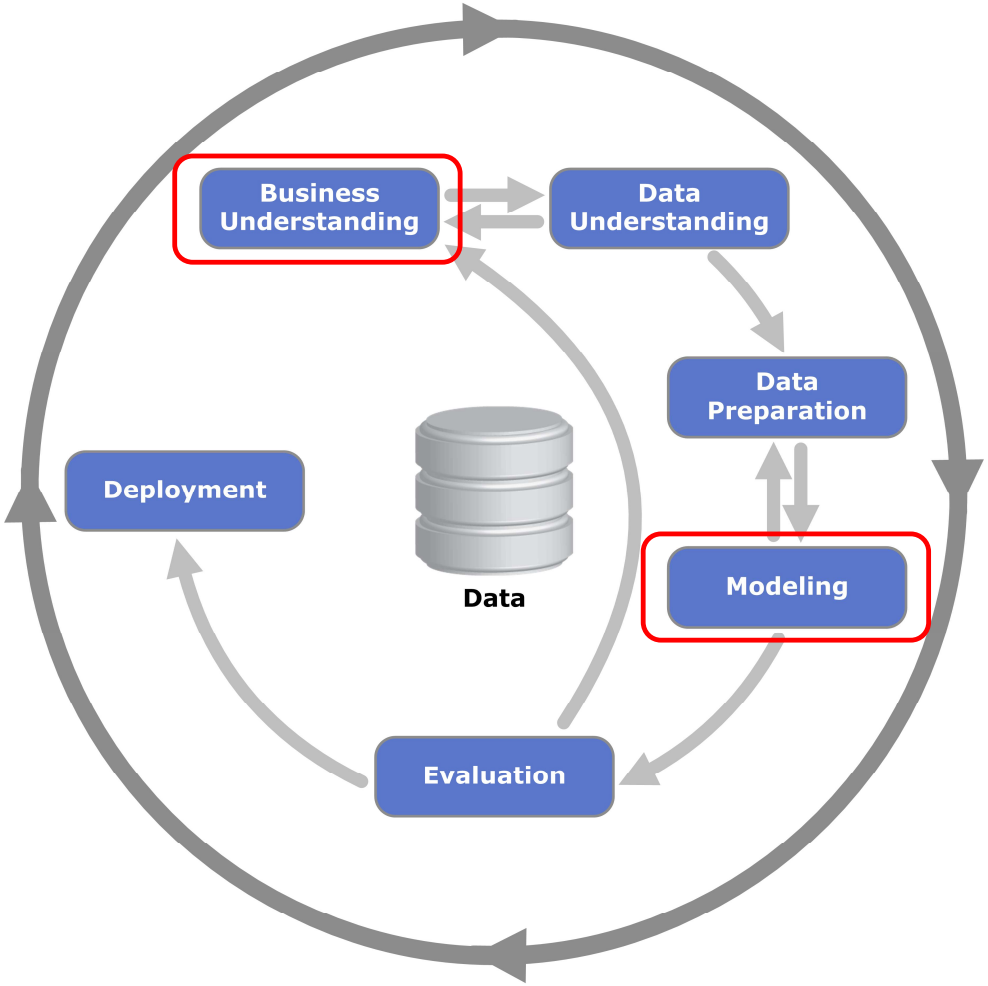
than precisely wrong!

alteryx

Why is the output of so many projects “precisely wrong”?

alteryx

The CRISP-DM Process Model



alteryx

Why the Bias for Models Over Business Understanding?

The focus on algorithms, techniques, and tooling

- University programs are often developed by faculty whose research focuses on the development of new algorithms and techniques
- Approaches developed in university programs diffuse into commercial training programs
- Software companies (and the press) are focused on product features as the basis of competition, and incorporating new algorithms and techniques into their products is central to this

The Kaggle problem

- In Kaggle competitions there is no business understanding phase in the process, there is only data
- Rankings in Kaggle competitions are not based on business value, but on how precisely participants are able to predict the "unseen" sample, and the differences in precision between competitors is very small
- The combination of these things contribute to a "script kiddie" culture in data science

If you have a hammer…

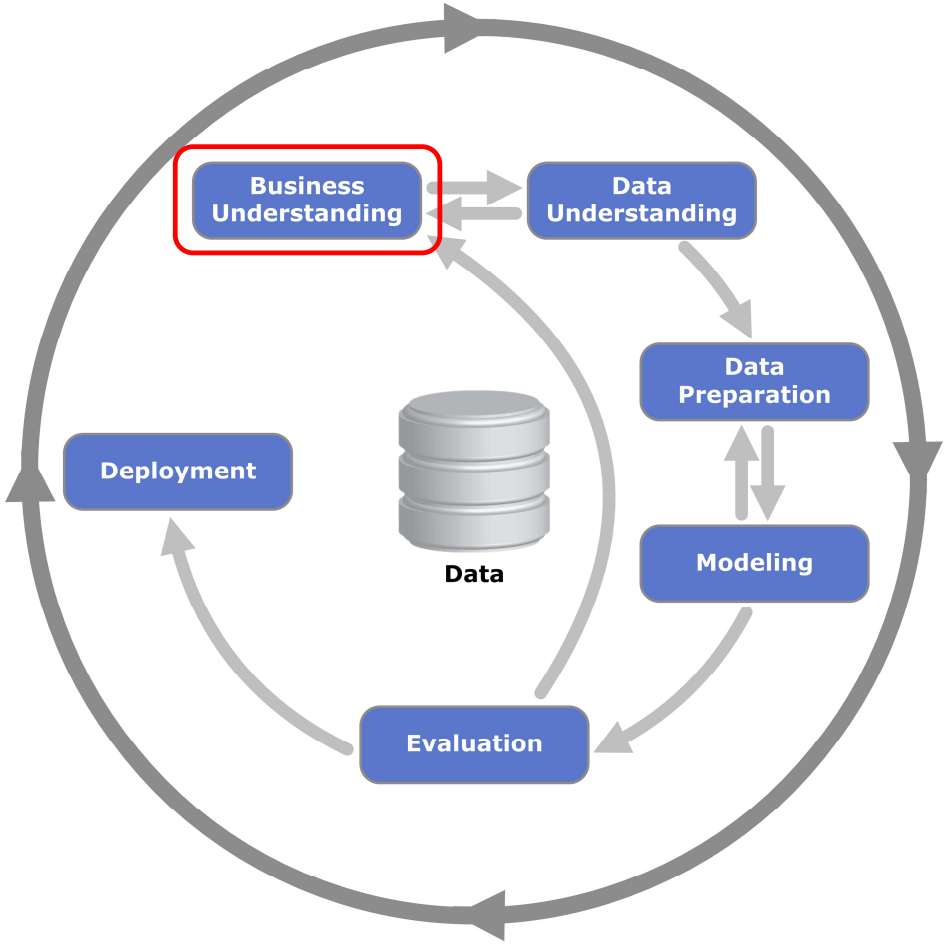
everything looks like a nail

alteryx

What is current state of analytic maturity for predictive analytics and data science?

- Not where it should be
- There is a need to get back to basics

Getting Back to Basics…
is Getting Back to Business Understanding



alteryx

How Can We Create Business Understanding?

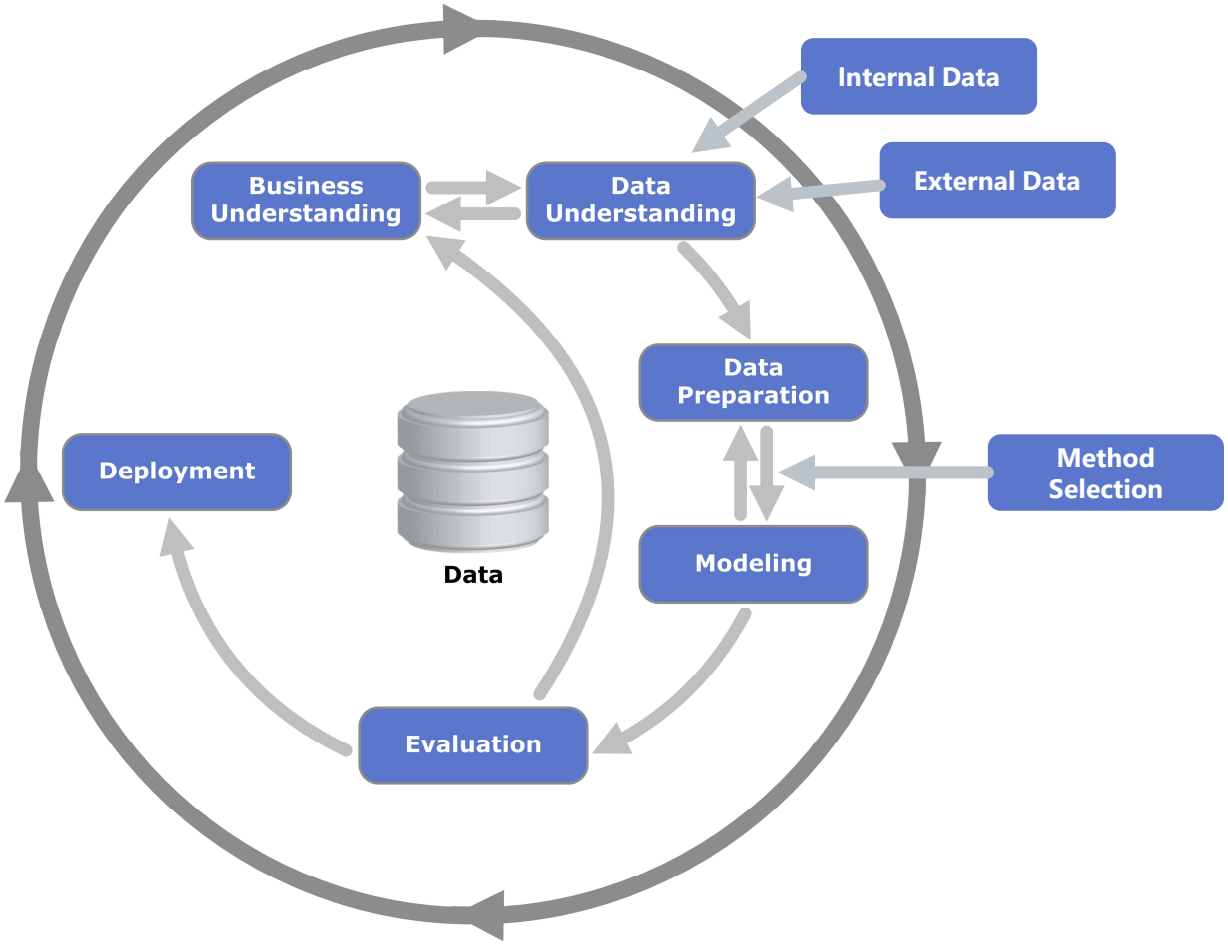
alteryx

7 Question Approach to Analytics Projects

1. Who are the relevant decision makers involved and who are their direct stakeholders?
2. What is the specific business decision that needs to be informed by analytics?
3. What are the business objectives that this decision needs to address?
4. What is the exact information needed to inform the business decision in a way that meets the objectives of the decision?
5. What is the state and availability of the local (internally generated) data resources that could potentially provide portions (perhaps all) of the needed information?
6. What (if any) third-party (external) data resources are needed (and exist) to complete the data resources required to generate the needed information?
7. What predictive analytics and ML techniques can provide the needed information?
    a. What BI interactive dashboards can provide the needed information?
    b. What static reporting can provide the needed information?

alteryx

Augmenting the CRISP-DM Process Model



alteryx

The Skills Needed to Answer the 7 Questions

- Interviewing Skills: The qualitative marketing research skills of depth and mini focus group interviewing methods
- Business acumen
- A knowledge of available internal data sources and how to get access to them
- A knowledge external data sources that are available, and the ability to find new external data sources
- A deep knowledge of a range of statistical and machine learning methods, and the ability to quickly come up to speed with other methods with which the analyst in not already familiar

alteryx

The Skills Needed to Answer the 7 Questions

Interviewing Skills	Business Acumen	Knowledge of Internal Data Sources	Knowledge of External Data Sources	Statistical and ML Methods
1. Who are the relevant decision makers and their direct stakeholders? 2. What is the specific business decision that needs to be informed by analytics? 3. What are the business objectives? 4. What exact information is needed to inform the business decision in a way that meets the objectives of the decision? 5. What is the state and availability of the local data resources? 6. What third-party data resources are needed to complete the data resources required to generate the needed information?	1. Who are the relevant decision makers and their direct stakeholders? 2. What is the specific business decision that needs to be informed by analytics? 3. What are the business objectives? 4. What exact information is needed to inform the business decision in a way that meets the objectives of the decision?	5. What is the state and availability of the local (internally generated) data resources that could potentially provide portions (perhaps all) of the needed information?	6. What (if any) third-party (external) data resources are needed (and exist) to complete the data resources required to generate the needed information?	7. What predictive analytics and ML techniques can provide the needed information?

alteryx

How do we find the perfect data scientist? … or How do we replace Jason Giambi?

- Finding people with all five needed skills to answer all seven of the questions to create successful analytics projects is hard!
- However, that isn't necessary if you take a "moneyball" approach
  - Form teams of several people that can work well together that span the needed skills
  - This approach has strong implications for organizations that have decided they will deal with the dearth of high-end data science talent by creating COEs staffed only by high-end data scientists
    - Don't do that!
    - Instead, staff COEs with the full set of skills needed to answer the "7 Questions"

alteryx

Appling the 7 Questions to Supply-Chain Lead-Times

Project Overview

- They could not manage product backorders in their European distribution center since they did not know when new inventories of materials would arrive.
- The system that was developed is currently in production at Cardinal Health
- The system was developed in Alteryx Desktop Designer and uses Alteryx Server for production
- This work was the subject of two different presentations at last year's Alteryx Inspire users conference

Multinational health care company specializing in the distribution of pharmaceuticals and medical products, and manufacturing of medical and surgical products such as PPE.

Cardinal Health helps pharmacies, hospitals and ambulatory care sites focus on patient care while reducing costs, improving efficiency and quality, and increasing profitability.

alteryx

Applying the 7 Questions Framework

1. DECISION MAKERS & STAKEHOLDERS

- Decision makers are the product allocation managers in the European distribution center and
- The direct stakeholders are the US and European supply chain teams and the European sales and marketing teams

2. BUSINESS DECISIONS TO BE INFORMED BY ANALYTICS

- To which customers should materials in short supply be allocated
- What to tell customers about when they can expect delivery of products that are currently unavailable

3. BUSINESS OBJECTIVES TO ADDRESS

- Equitably allocating scarce materials to customers in a way that maximizes public welfare
- Honestly managing customer delivery expectations and maintaining customer trust

4. INFORMATION NEEDED TO INFORM BUSINESS DECISION IN A WAY THAT MEETS THE OBJECTIVES

- For every stock-keeping unit (SKU) provide the expected arrival dates and inbound quantities for all orders of that SKU in Cardinal Health's ERP system

5. STATE & AVAILABILITY OF LOCAL DATA RESOURCES THAT COULD PROVIDE NEEDED INFORMATION

- Internal ERP system contains the needed information concerning important dates in the history of an order, details on how the order was carried out, and the point of origination of the materials. Access to that system was readily available to the supply chain management team.

6. THIRD-PARTY DATA RESOURCES NEEDED TO GENERATE REQUIRED INFORMATION

- The shipping agent provides the information associated with the mode of transportation, delayed departures of airplanes or ships, the existence of issues with customs paperwork, arrival (air)port congestion, and other factors that can result in shipment delays
- Access to that system was readily available to the Cardinal Health supply chain management team

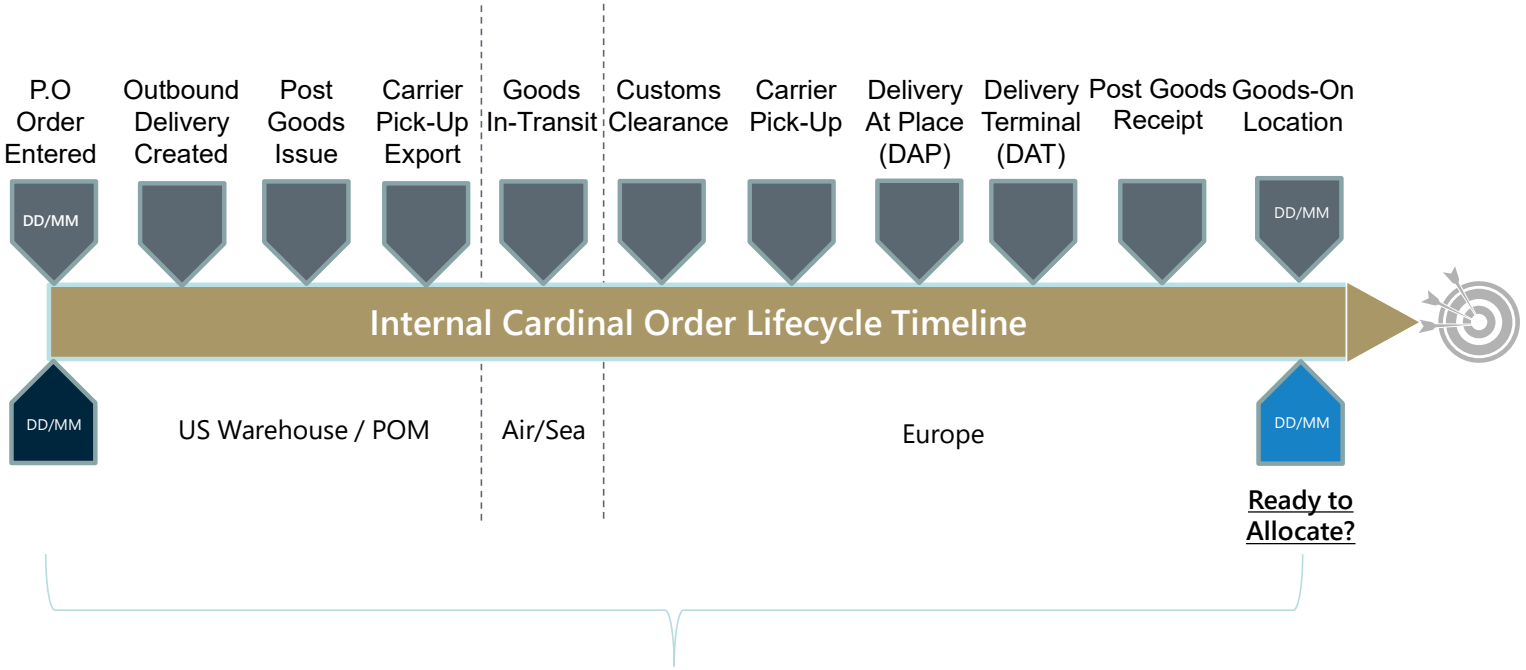
7. WHAT PREDICTIVE ANALYTICS AND ML TECHNIQUES CAN PROVIDE THE NEEDED INFORMATION

- The relevant time metric to the Cardinal Health team was days for each stage in the shipping process
- This results in the need for the use of regression-oriented methods that account for the "count" (integer) nature of the length of time each phase in the shipping process takes to complete

alteryx

The Business Problem

Trying to figure out when materials will become available for allocation…



Cardinal Products Available For Customer Allocation To Resolve Backorders

alteryx

What is Observed

- Dates associated with
  - The entry of the internal purchase order into the system
  - When a line number in a PO is “issued” from either a US based warehouse or from a point of manufacture
  - When the issued portion of a PO line number leaves the outbound airport/port
  - When the issued portion of a PO line number arrives at the European airport/port
  - When the issued portion of a PO line number arrives either directly to the customer or the loading dock of the European distribution center
- Information from the shipping company on in-transport issues
  - Incomplete or wrong paperwork
  - Late arrival of the airplane or cargo ship, the land transportation company, etc.
  - Insufficient warehouse space in the European distribution center
- The goods point of origination

alteryx

What Needs to be Predicted

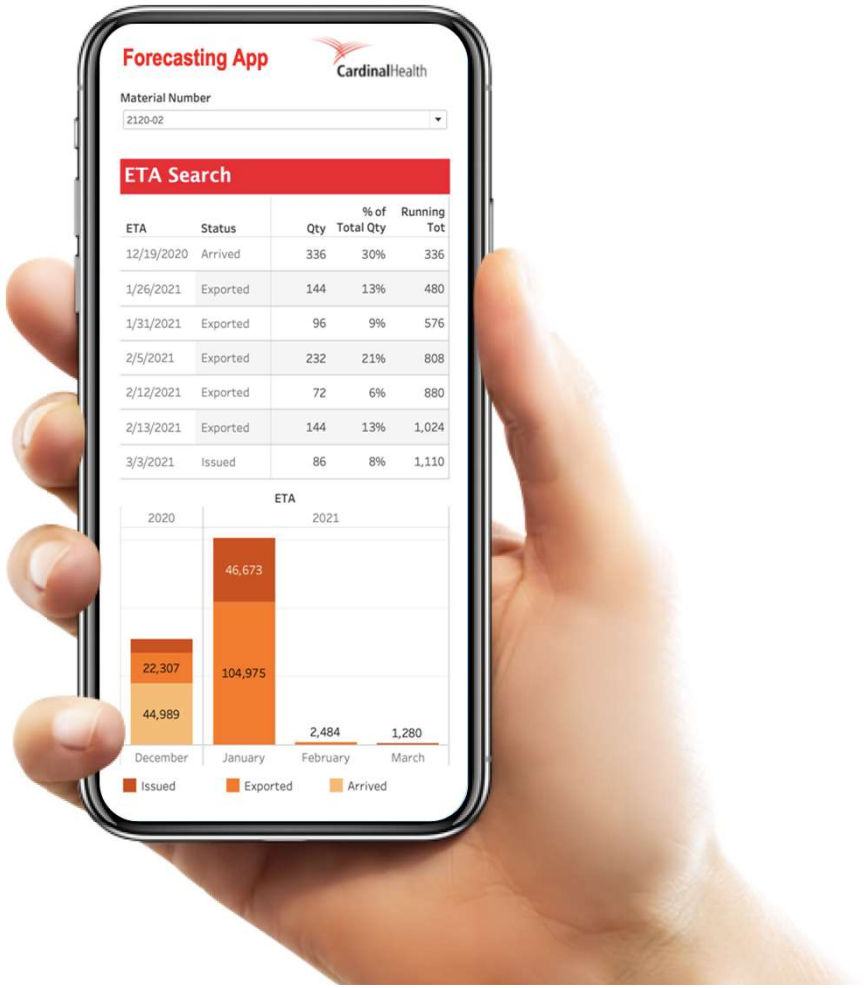
- The days between the PO being entered to part of a PO line number being issued
- The days between when a “batch” of a PO line number is issued to that material being exported from the origination point
- The days between when a “batch” is exported to its arrival at the destination point
- The days between what a “batch” has made it to the distribution center in Europe
- The last three of these will depend on whether the product is going by sea or air

The Automated Solution Architecture

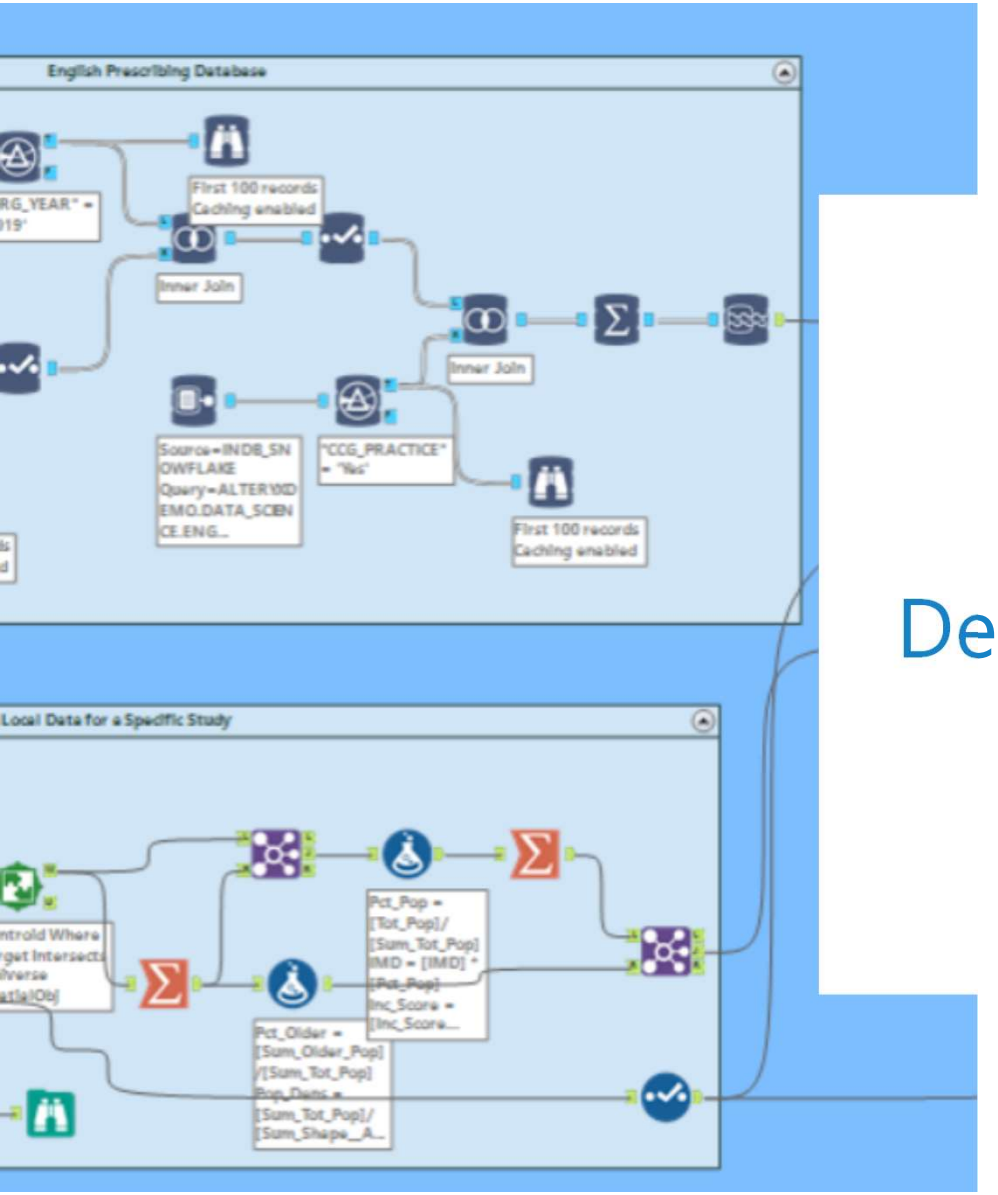
- The automated solutions consists of two workflows
  - The “estimation” workflow is run on a weekly basis using the last 600 days of data in a rolling window
  - The “scoring” workflow is run daily in order to incorporate all available information
  - Both workflows address seven different models for different stages of the process, broken down by mode or transportation (air or sea)
  - For scoring, imputation is used for unknown future variables, and are then replaced with actual values over time.
- The output of the system is presented to distribution center staff and managers via dashboards in order to manage inventories and product backorders
  - The dashboards are based on querying the predicted arrival times for all orders in-process for a particular SKU
  - Both desktop/laptop and mobile versions of the dashboard are provided

alteryx

An End Product: The Mobile Dashboard



alteryx



alteryx

Demo Time!

alteryx

Thank you

Questions?

Dr. Dan Putler | Chief Data Scientist
650.375.5651
dputler@alteryx.com

alteryx.com