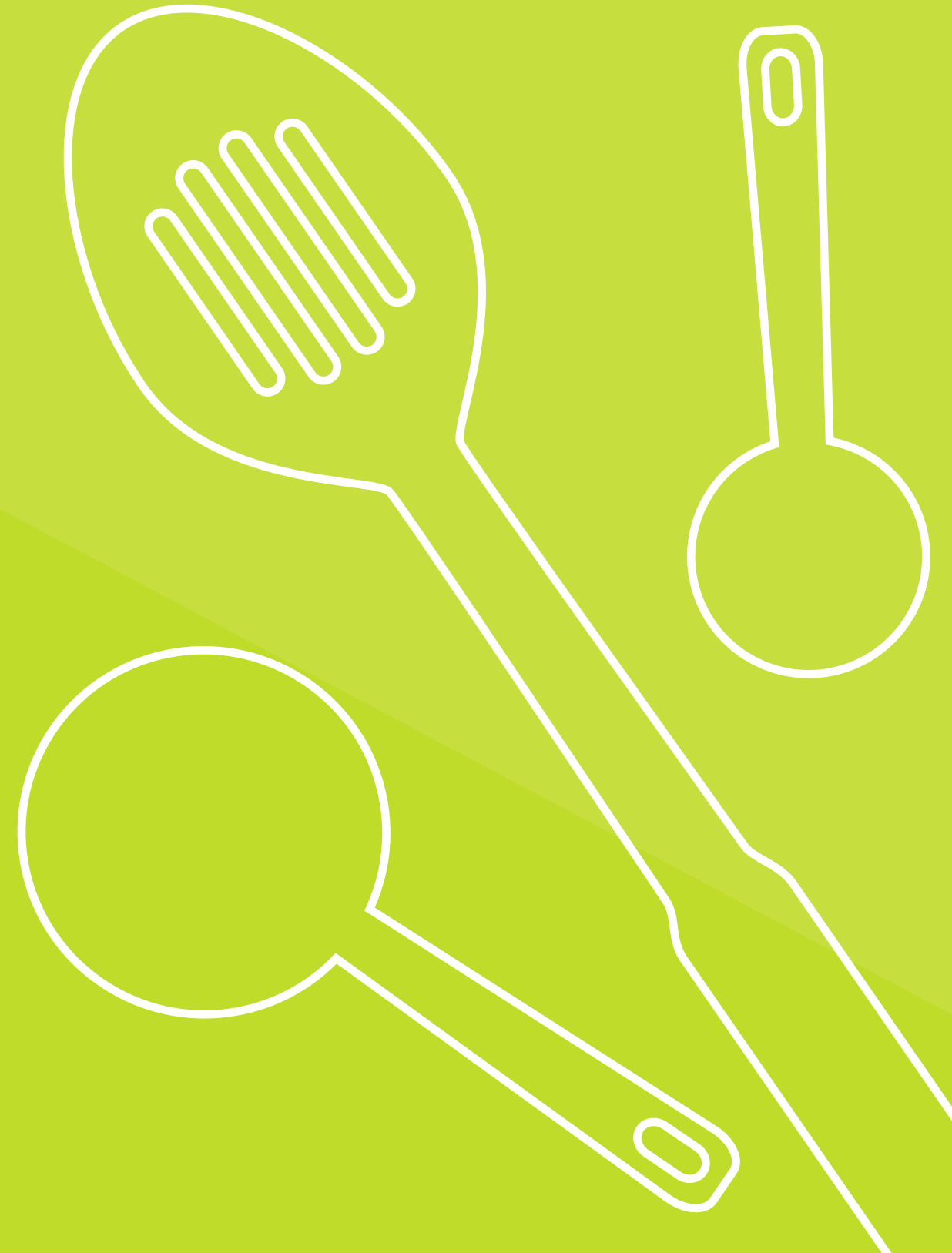# alteryx

COOKBOOK SERIES

7 Steps to Data Blending for Predictive Analytics

Evolution of Analytics

Just as the volume and variety of data has grown significantly in recent years, so too has the expectations for analytics. No longer is a standard dashboard depicting what happened last month acceptable for decision-makers.

To get ahead of the competition and improve their company's bottom line, data analysts need to understand what is likely to happen in the future so the organization can be more prepared to make impactful changes.

Predictive analytics allows organizations to gain an edge by analyzing current and historical data to better understand and predict the likelihood of future events.

The most challenging aspect of predictive analytics is getting the right data ready for analysis. This is where data blending can help.
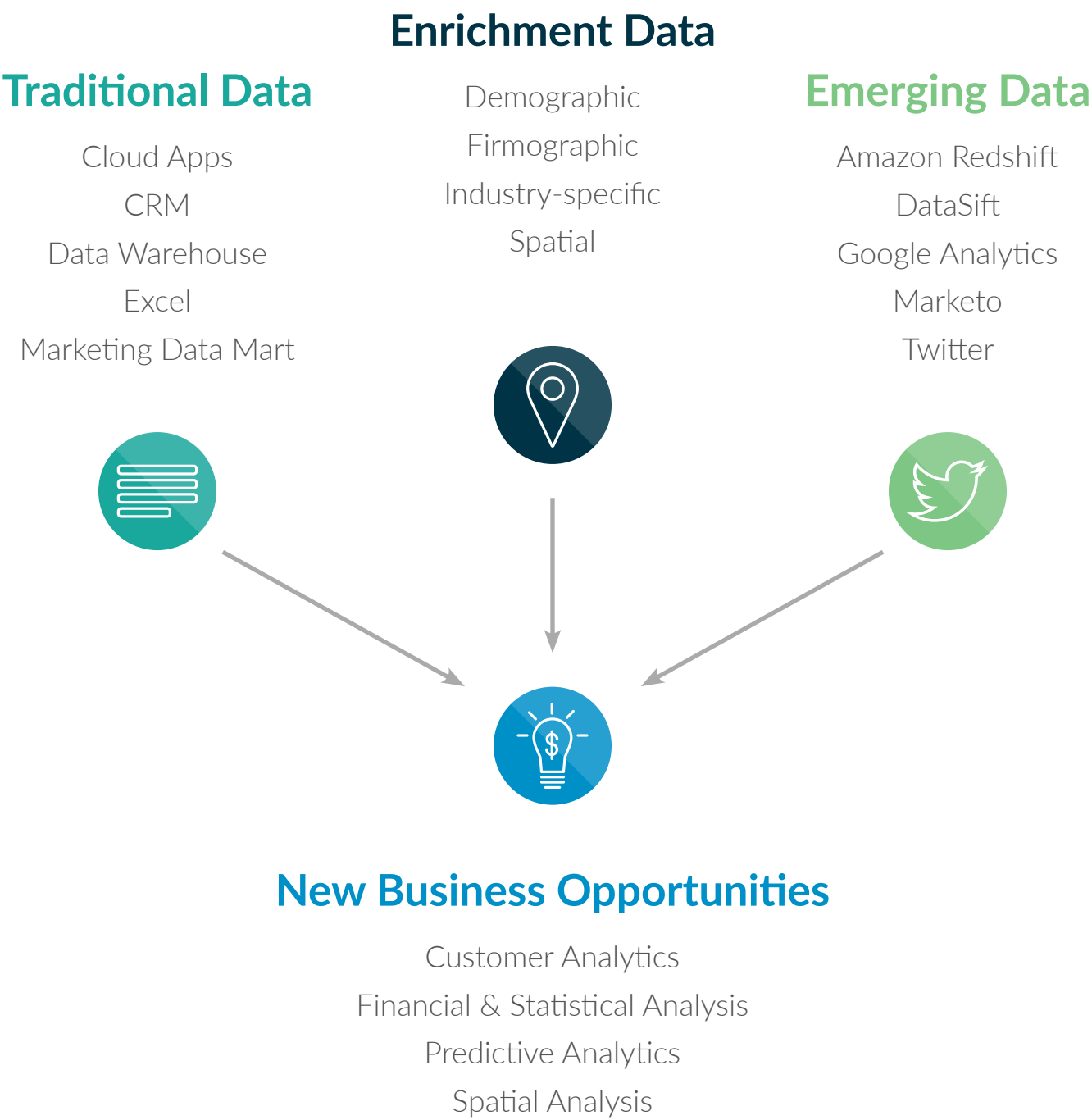
What is Data Blending?

What started out as a means to an end for a data analyst who dealt with a single source of data has led to the need to combine multiple sources of data.

Throughout this evolution, data blending has empowered those analysts in the line of business with the ability to access and combine data from multiple sources to reveal deeper intelligence that drives better business decision-making.

Analysts use data blending to build an analytic dataset to answer a specific business question or take advantage of opportunities, with insight into customer preferences, marketing campaign results, financial operations, site and merchandising optimization, and much more.

Traditional Data

Cloud Apps
CRM
Data Warehouse
Excel
Marketing Data Mart

Enrichment Data

Demographic
Firmographic
Industry-specific
Spatial

Emerging Data

Amazon Redshift
DataSift
Google Analytics
Marketo
Twitter

New Business Opportunities

Customer Analytics
Financial & Statistical Analysis
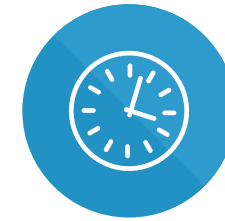Predictive Analytics
Spatial Analysis

What is Data Blending for Predictive Analytics?

In order to improve predictive analysis/models/outcomes, you need to make sure you are working with the right data. But there are more sources and formats of data than ever before that you need to incorporate to make the most informed decisions.

Data blending for predictive analytics helps analysts like you spend more time on model creation, evaluation, and deployment and less time on preparing data for analysis.

Analysts trying to do predictive analytics struggle with:

- Accessing the right types and systems of data

- Preparing and cleansing data

- Joining multiple datasets

- Delivering a repeatable process for future analysis

- Relying on others to create the dataset they need

Relying on data scientists or IT to prepare your data can be time consuming and delay your outcomes

Many analysts don't have the right tools to blend data or perform predictive analytics

Analysts have to remember and repeat the same tedious steps whenever data changes

Recipe for faster data blending

1   Get access to all of the data you need

2   Understand your data

3   Prepare and cleanse data

4   Transform and restructure data

5   Join the data for analysis

6   Incorporate predictive analytics

7   Output to existing predictive models

Ingredients you need

- A copy of Alteryx
  alteryx.com/download

- A list of the data sources you want to blend

- Access to each database or source you are going to use

- A rough specification of the dataset you need

See demo videos on Data Blending at alteryx.com/solutions/data-blending

1 Get access to all of the data you need

Anyone can make decisions based on data, but in order to make *accurate* decisions about things that can have a long-term effect on your company, the more data you have to work with the better.

Getting the data you need for predictive analysis used to be a challenge. Alteryx gives you the ability to access data of any type from any source:

- Use an *Input* tool to access data directly from a relational database or directly from your desktop in formats such as Excel, SAS, SPSS, ERSI, MapInfo, and more

- Use direct connectors to access social media data, such as *Twitter* and *Foursquare*

- Use cloud connectors to access *Salesforce.com*, *Marketo*, *Google Analytics*, *Amazon S3*, and more

Input tool

Twitter tool

Salesforce.com Input tool

Recipe for analytic greatness

- Work with IT to ensure you have the right credentials to access your data sources

- Analyze all data types without data size limitations (survey data, log files, mobile data, XML)

- Alteryx provides third-party data from Experian, Dun & Bradstreet, and TomTom to enrich your current datasets

- You can even connect to SPSS (.sav) or SAS (.sas7bdat) files to incorporate data from existing predictive models
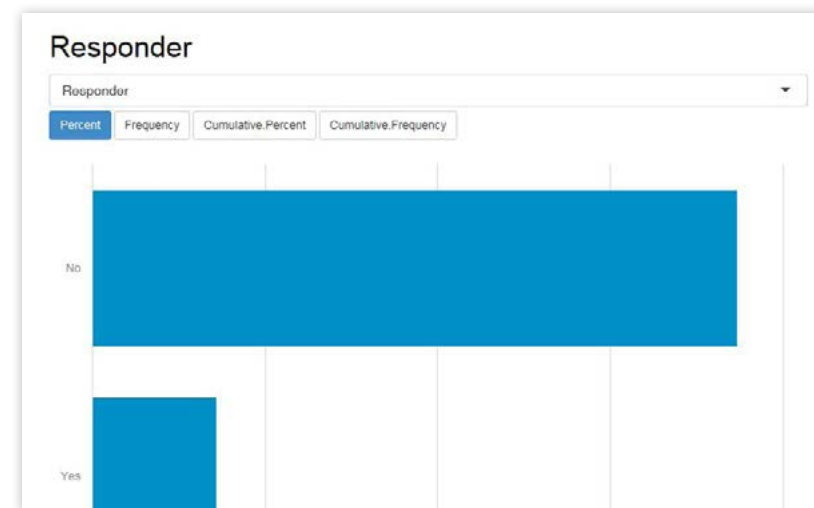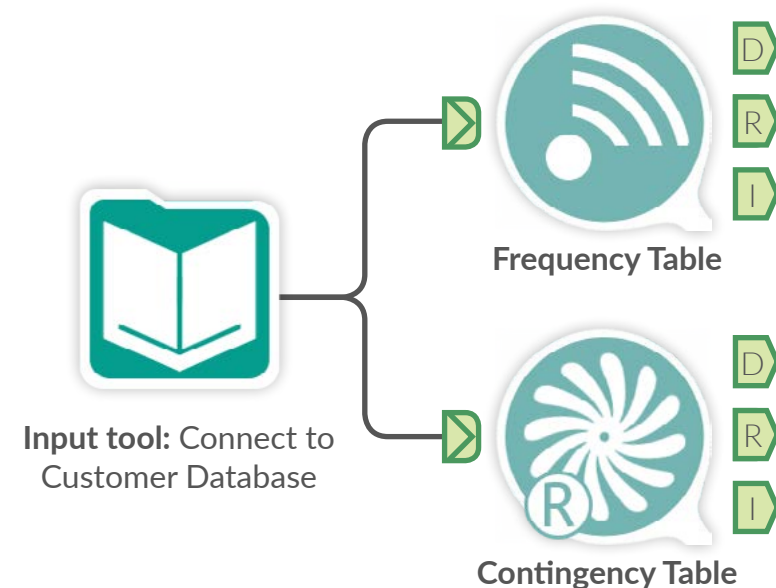
② Understand your data

Once you gain access to your data, data investigation can help you understand the important details of your data.

Alteryx has a wide range of tools to help you investigate your data:

- Understand counts and percentages of your data with a *Frequency Table*

- See combinations of field values and/or percentages with a *Contingency Table*

- View descriptive statistics for all of your selected variables with other tools, such as the *Field Summary* tool

Recipe for analytic greatness

- Interactive visualizations in Alteryx Analytics allow you to better understand and explore your data

- Take advantage of mining techniques, such as Association Analysis, to find out what fields are related

Frequency Table

Input tool: Connect to Customer Database

Contingency Table

Responder

Responder				▼
Percent	Frequency	Cumulative.Percent	Cumulative.Frequency	

Text

| Store.Number	Customer.Segment	Consumer	Corporate	Home Office	Small Business	Totals
100		73	121	85	78	357
101		67	116	74	58	315
102		15	39	21	24	99
103		45	90	55	60	250
104		53	123	62	70	308
105		71	112	87	82	352
106		59	108	79	72	318
107		46	87	67	64	264
108		58	75	46	50	229

3 Prepare and cleanse data

As an analyst, you probably wish you could spend more time analyzing data than preparing it for analysis. Alteryx helps cut data preparation time by up to 30%, giving you more time to test hypotheses and evaluate models.

Alteryx includes a variety of tools to ease the data preparation and cleansing process:

- Pick the *Auto Field* tool to automatically set the field type for each string field

- Use the *Select* tool to change field types or names

- Select the *Filter* tool to isolate a subset of the data in which you are interested

Recipe for analytic greatness

- Make product names, categories, and user IDs consistent as you bring multiple datasets together

- Read CSV or XML files in as strings or text fields

- Utilize parsing tools, such as *Regular Expression* and *XML Parse*, to convert, match, and replace data
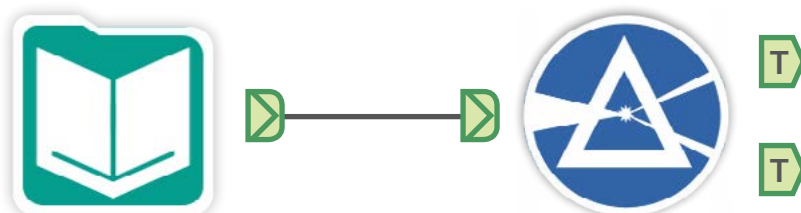
Input tool: Connect to Customer Database

Filter tool: Isolate a subset of data

Input tool: Connect to Product Sales

Auto Field tool: Automatically change field types

Select tool: Change field names

4 Transform and restructure data

Data comes in all shapes and forms, so getting the data in a format you need for analysis can oftentimes require transforming or restructuring it to fit your needs.

Alteryx delivers capabilities to simplify this process:

- Use the *Summarize* tool to aggregate data; e.g., calculate Total Spend and Number of Transactions by Customer ID. (Note: The Summarize tool does the work that pivot tables do in Excel)

- Select the *Transpose* tool or *Crosstabs* tool to rearrange fields in your data

- Utilize the *Formula* tool to create expressions, which are similar to formulas you can create in Excel

Recipe for analytic greatness

- Group and count records passing through a data stream

- Determine min and max values, plus keep a running total

- Crosstab and transpose on a full dataset rather than working in separate spreadsheets

Summarize tool: Count the number of transactions, and add the value of the transactions

Field	Action	Output Field Name
Customer_ID	GroupBy	Customer_ID
Sales	Count	Count
Sales	Sum	Spend

Transpose tool: Move data from columns to rows (or vice versa)

Formula tool: Create a new field based on IF THEN statements

# 5 Join the data for analysis

Once you have cleansed and shaped multiple sources of data, bringing that data together to create one dataset that can be used for predictive analytics is the next critical step.

Alteryx can join data in many ways:

- Combine two or more data streams with similar structures into a single stream using the **Union** tool

- Use the **Join** tool to combine multiple fields without having to pre-process and concatenate key fields together

- Perform inner and outer joins without the multiple VLOOKUP expressions that make combining data in Microsoft Excel hazardous

- Incorporate Fuzzy Matching to handle non-exact matching of customer names or addresses (e.g., John, Jon, Johnny, or Jonathan)
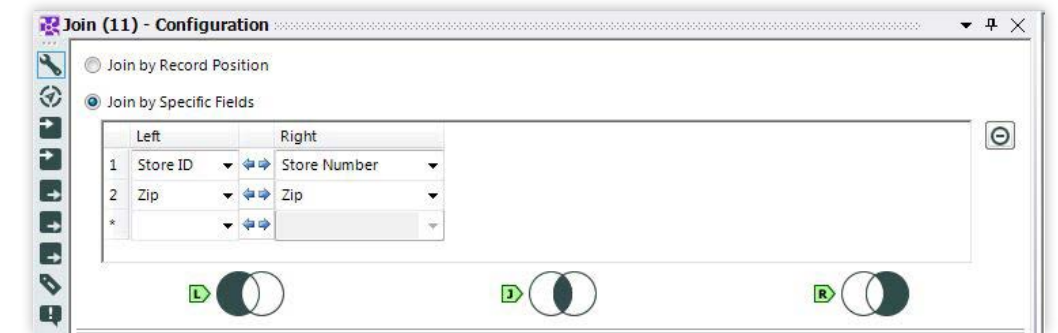
Join tool: Combine data streams on common fields
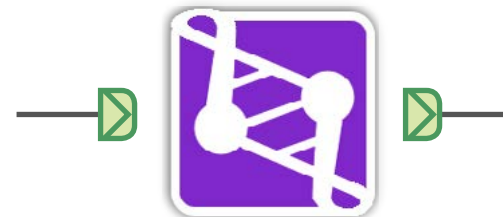
Union tool: Combine each data source into a single stream

Recipe for analytic greatness

- Bring together fields that don't match exactly using the *Fuzzy Match* tool

- Create a macro that brings in all spreadsheet tabs in a single click

6 Incorporate predictive analytics

Incorporating predictive analytics in Alteryx is as easy as dragging a tool onto the same canvas used for data blending. Alteryx removes the complexity of predictive analytics by eliminating the need for coding, yet it is also flexible enough for more advanced users.

Alteryx has over 30 drag-and-drop tools for modeling, grouping, and time series, all built on the R programming language:

- Use the *Logistic Regression* tool to determine if someone will respond to an ad or campaign

- Utilize the *Boosted Model* tool to understand the importance of predictor fields

- Cluster or group members quickly by using a *K-Centroid Analysis* tool

- Use the *R* tool to allow R programmers to bring in their own code or scripts

Logistic Regression tool   **Boosted Model tool**   **K-Centroids Analysis tool**   **R tool**

Recipe for analytic greatness

- Improve forecasting with *Time Series* tools

- Easily access the R code behind the Alteryx tools to better understand it or even make alterations

- Incorporate your own custom R models utilizing the *R* tool

7 Output to existing predictive models

Many organizations, and the analysts within them, spend a significant number of hours at an exorbitant cost to build predictive models and deploy them throughout their organization. Alteryx doesn't force you to replace everything you have built, but instead enables you to take advantage of these existing models.

- Incorporate your own pre-built R models utilizing the *R* tool

- Output to SPSS (.sav) or SAS (.sas7bdat) files to take advantage of models already in place in an organization

"Over the past five years we developed over a dozen models, but had outgrown the current architecture we were using. Primarily most of our bottlenecking was in the data preparation and data processing stage of the modeling.

I downloaded a trial of Alteryx and a process that would have taken 25 minutes to do under our old architecture took less than 5 seconds in Alteryx, and I figured I was onto something."

– Michael Barone
Predictive Modeling Analyst, Paychex

7 Output to existing predictive models

Customers with transactions.sav

Customers with transactions.sas7bdat
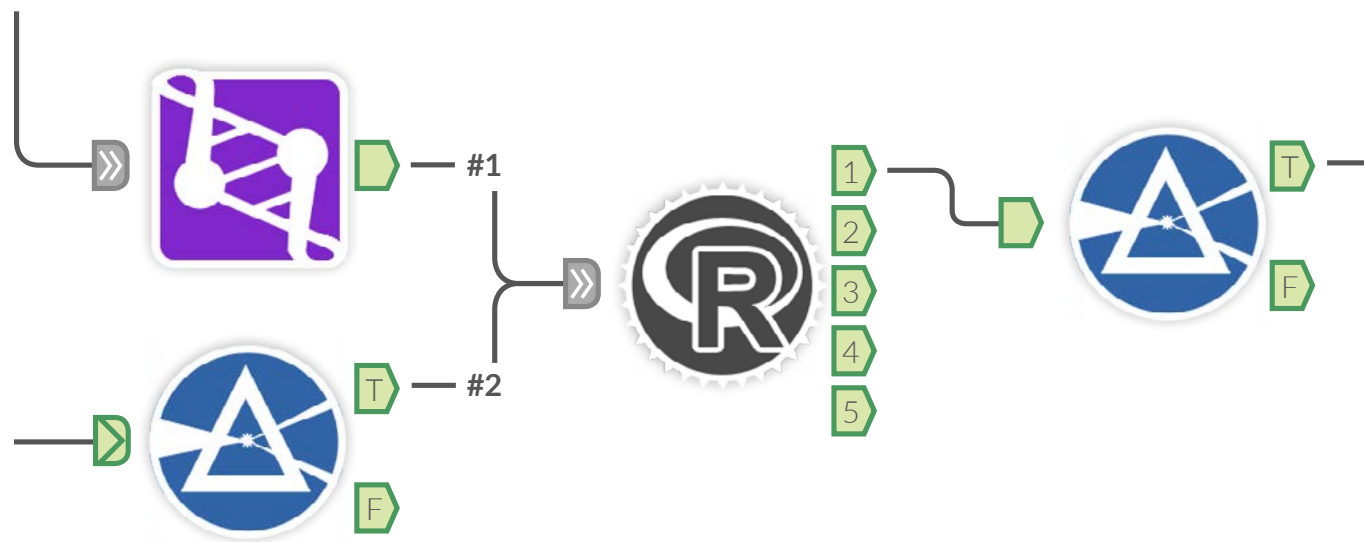
- Use the **R** tool to write your own R script or bring in a pre-built R model into an Alteryx workflow
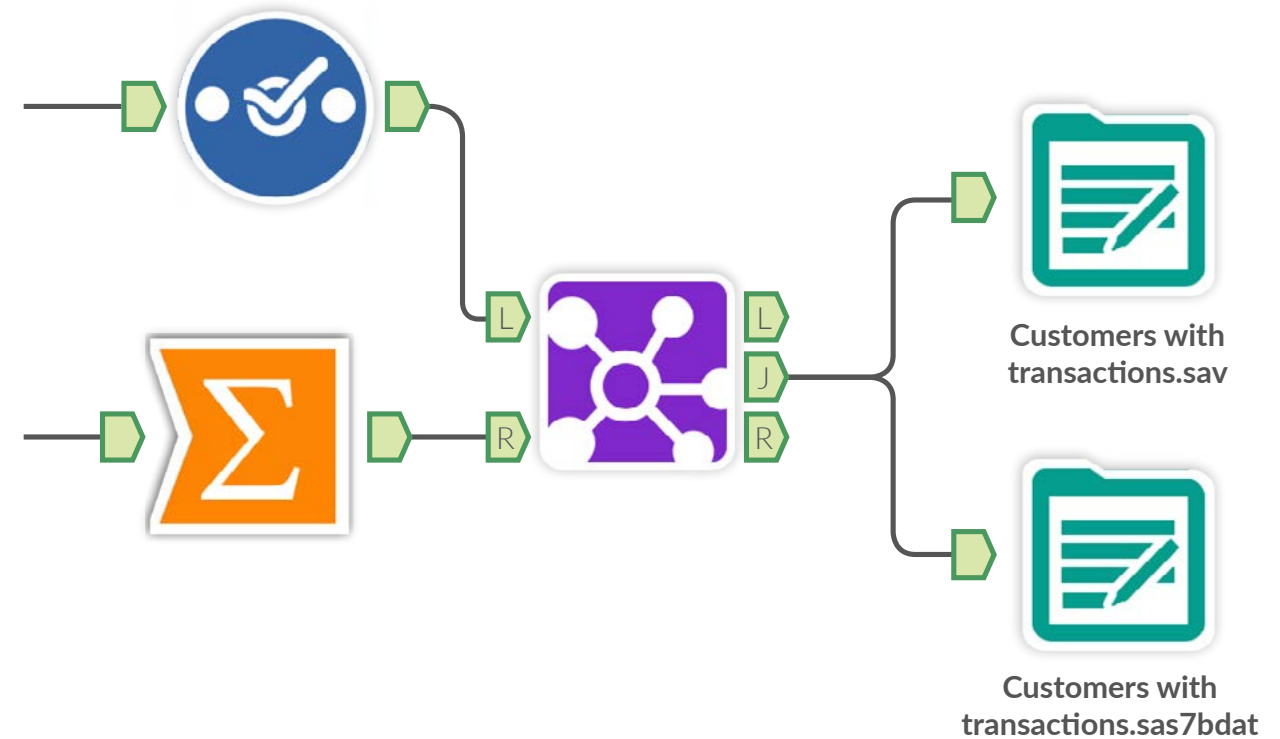
- Output to SPSS or SAS and take advantage of existing models within your organization
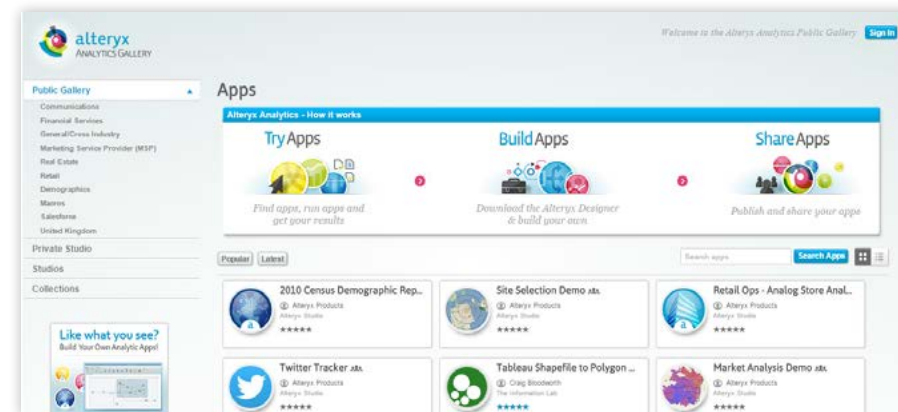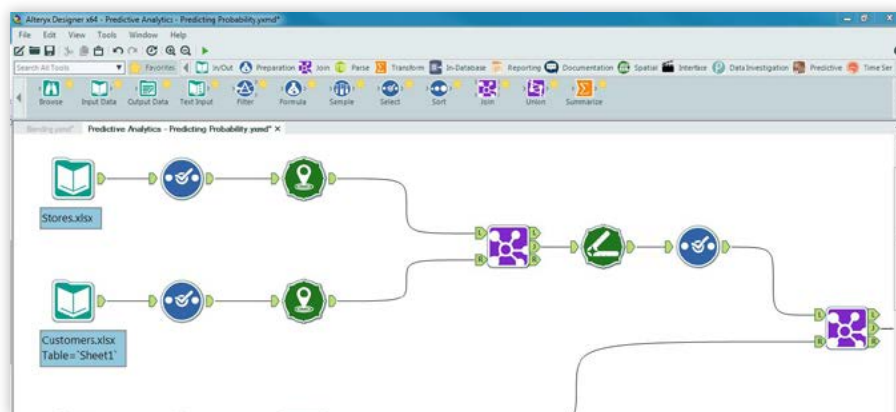
Bonus! Predictive Analytics for Decision- Makers

The Alteryx Analytics Gallery removes the need for decision-makers to build analytic workflows, giving them an engaging experience and the results they need to make better business decisions.

- Save and automate your workflow to run the same processes with updated data and eliminate the potential for error

- Package your workflow into an analytic application that can run in a browser-based environment

- Give decision-makers the ability to customize and run analytic applications without having to build workflows

Recipe for analytic greatness

- Create a workflow once and reuse it for future processes or outputs

- Create a macro to ease repeatable processes

- Set permissions on workflows and macros so only the correct people can access them

Predictive Analytics in Alteryx Enables Reduced Marketing Spend

SOUTHERN STATES®

To optimize its marketing efforts, Southern States Cooperative relies heavily on the predictive analytics tools within Alteryx.

Deeper Insights

The company was able to access and blend all relevant marketing and customer data from multiple sources, as well as perform predictive and spatial analytics to identify high-potential customers for targeted mailings.

Hours vs. Weeks

By reducing time-to-insight from weeks to mere hours and increasing the volume of insights gathered, the company was able to improve campaign response rates on an average by approximately 200 percent.

Intuitive Workflow

Dozens of analytic applications were deployed across the company—all developed without any coding—performing tasks like data extraction, cleansing, exploration, and modeling in a single workflow.

> "Previously, to do statistical analysis or predictive modeling, I would have to export the data into another software solution. With Alteryx, I am able to keep all of the analysis as part of the same workflow. In fact, I can do all of my data extraction, cleansing, exploration, and modeling in one module— seamlessly."
>
> **– Greg Bucko**
> **Manager of Customer Insights**

Alteryx Delivers on the three things analysts need most

Empowered analysts who can access all the data they need, when they need it

Single intuitive workflow for a complete data blending & advanced analytics process

Deeper business insight that can be achieved without relying on others

Next Steps

Try data blending in Alteryx
alteryx.com/download

Download the data blending kit
bit.ly/datablendingkit

View Customer videos
alteryx.com/customers

alteryx

Data Blending for Predictive Analytics

More than 1000 customers and thousands of
data analysts worldwide rely on Alteryx daily.
alteryx.com/solutions/data-blending