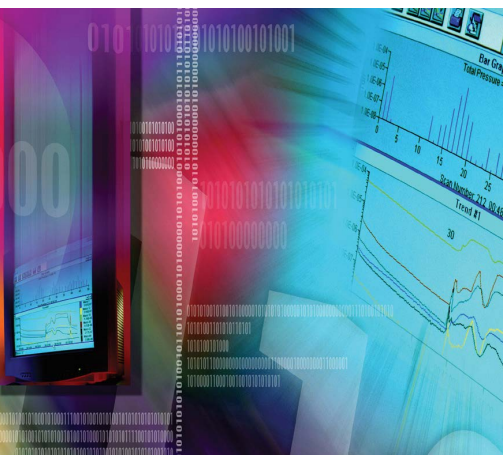


alteryx



# The Definitive Guide to Data Blending

## Leveraging Alteryx Analytics for data blending, you can:

- Gather and blend data from virtually any data source, including local, third-party, and cloud/social data, using an intuitive workflow that replaces complicated coding
- Make better, more informed business decisions, faster and more accurately
- Re-use and iterate workflows to answer the line-of-business questions you need answered in days rather than weeks

## Introduction

Data is at the heart of today's interconnected world. It is captured in almost every aspect of our lives—television shows we watch, websites we visit, groceries we purchase, and opinions we share. As the growth of data continues to expand so does the use of this data by organizations to better understand their customers, optimize their supply chain, and much more. Data analysts in the line of business have become the main driving force tasked with utilizing this data to answer more complex business questions.

These analysts know their business and understand what data is necessary to get the results they want—whether it pertains to sales, marketing, operations, or even finance. It is up to them to use this knowledge and business insight to make sense of all of the disparate data that resides in a data warehouse or data mart, or perhaps within a CRM system or a marketing automation system. It may even be social media data or spatial data—data that is becoming more prevalent and necessary to fulfill organizations' business requirements.

Many analysts find it difficult to address this new data challenge because their traditional tools and approaches are not robust enough to handle this environment. Utilizing spreadsheets, manual processes, and custom scripting or relying on data scientists to build analytical datasets are all too time-consuming, expensive, and complicated in the face of the number of ad-hoc requests that analysts receive each day.

Today's analysts need to live at the forefront of innovation to keep their organizations competitive. They must extend their natural capability and creativity with genuine business insight. They must focus their strengths on high-level business questions rather than the minutiae of spreadsheets and low-level SQL queries. Data blending helps today's analysts take full advantage of their expanding roles, as well as the expansion of data needed to make those critical business decisions.



The screenshot shows a Microsoft Excel spreadsheet with a SQL code editor window overlaid. The SQL code defines two tables: 'department' and 'employee'. The 'department' table has columns for 'DepartmentID' (INT) and 'DepartmentName' (VARCHAR(20)). The 'employee' table has columns for 'LastName' (VARCHAR(20)) and 'DepartmentID' (INT). The code includes several INSERT statements for the 'department' table, such as 'VALUES (35, 'Marketing')' and 'VALUES ('Rafferty', 31)'. The spreadsheet shows a VLOOKUP formula in cell C2: '=VLOOKUP(A2,\$B\$2:\$B\$4,1,FALSE)'. The formula bar also shows '=ISNA(VLOOKUP(REF1,REF1,1,FALSE))'. The spreadsheet data includes columns for 'Feb Products', 'Mar Products', and 'Discontinued'.

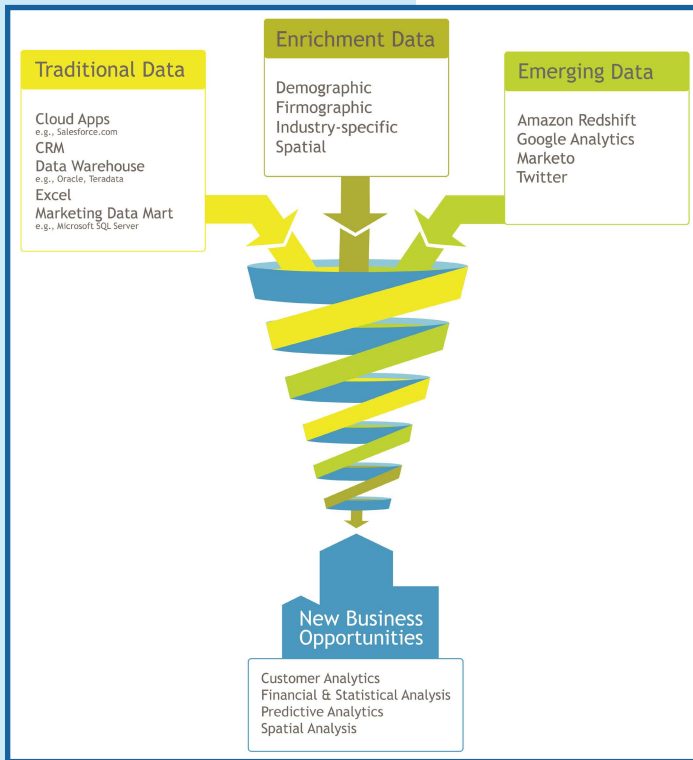
	A	B	C	D	E	F
1	Feb Products	Mar Products	Discontinued			
2	1234	1236	#N/A			
3	1235	1237	#N/A			

	A	B	C	D	E	F
1	Feb Products	Mar Products	Discontinued			
2	1234	1236	FALSE			
3	1235	1237	TRUE			
4	1236	1238	FALSE			

**Traditionally, analysts have relied on VLOOKUPS, scripting, and multiple spreadsheets for constructing datasets—but this can be clunky and time consuming.**

## Data Blending Defined

Data blending is the process of combining data from multiple sources to create an actionable analytic dataset for business decision-making (such as retail site selection or multichannel profiling) or for driving a specific business process (such as packaging data for sale by data aggregators).



Data blending is needed when an organization's data management processes and infrastructure are insufficient for bringing together analytic or specific datasets required by line-of-business groups. It can, for example, readily bring together disparate data, such as customer information from a cloud sales automation system (e.g., Salesforce.com) with clickstream web data stored in a Hadoop file system and segmentation models from Microsoft Excel. This is important, because while organizations aspire to have a completely integrated data management system, the majority of data required to make strategic business decisions still resides outside their IT-controlled data environment.

Data blending differs from data integration and data warehousing in that its primary use is not to create the single unified version of the truth that is stored in a data warehouse, data mart, or other system of record within an organization—and is typically conducted by a data warehousing or business intelligence professional. Instead, this process is conducted by a business or data analyst with the goal of building an analytic dataset to assist in answering a specific business question.

## Common Use Cases

Implementing data blending into the line of business can deliver greater benefits and deeper insight in hours—significantly faster than the weeks typically required for manual processes and traditional IT approaches. This time savings can be realized in the myriad business situations in which data analysts find themselves. Let's look at a few examples where data blending can positively impact business decision-making.

### Sales and Marketing

For every organization, growing revenue typically coincides with targeting prospects who are ready to buy. Marketing departments spend a lot of time trying to identify these prospects so they can focus campaigns, and allow salespeople to concentrate their cross-sell and up-sell efforts most efficiently.

Most customer data is stored in a CRM system—either in a database or possibly in a cloud solution, such as Salesforce.com—while marketing prospect data is stored in a separate system, such as Marketo. And information about customer and website prospect activity is captured by web analytics technology, such as Google Analytics. Historically, finding the relevancy of this data could require generating spreadsheet-based reports from both the marketing automation and CRM systems. From there, an analyst might need to combine these into a single spreadsheet with multiple tabs and construct formulas using VLOOKUP functions to reference relevant information. Or, they may just combine the two spreadsheets into one and manually look for duplications. What's more, the web analytics may not be something that can be incorporated at all without some sort of custom work by IT staff.

The beauty of data blending is that an analyst can access this data directly from the environment in which it is located. All they need is the right credentials to access the data. Then, they can pull the data from the right systems and start combining the data on common fields, blending in the specific information for which they are looking. They can combine data on customer ID, for instance, and discover what products or services not only have the biggest impact on sales, but also which of these drive the interest of prospective buyers.

### **Financial Operations**

Analysts within the realm of finance understand how critical it is to get the right information to deliver the right results. For instance, data plays an important role in the loan and credit card approval processes, making the difference between approving individuals with a low probability of default or fraud and those with a high probability, impacting the financial risk for the organization. It can be as simple as building out a customer model and then tracking and trending detailed client information over several years. Typically, this includes combining a lot of data from several sources, including web logs, which can be unstructured or need to be cleaned up, and even multiple third-party databases that contain information on past and current customers.

Data blending reduces the time to insight from weeks to hours, allowing analysts to work with the data directly to improve its quality and cleanliness, and combine it into a usable format that can be fed directly into existing models.

### **Site and Merchandising Optimization**

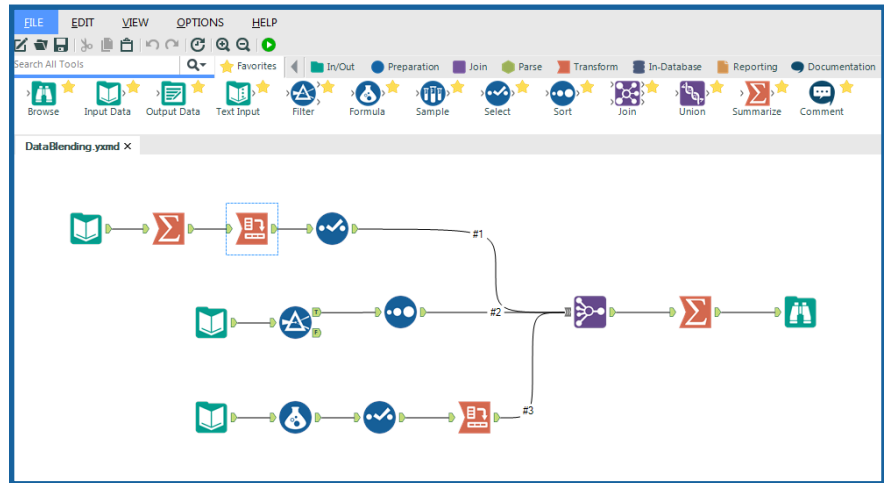
In order to have a successful store, understanding your potential customers and prospects is crucial. You may need to look at customer spend levels, purchase history, and path-to-purchase to discover these customers. Once you understand that, you can then use that insight for targeting, media planning, and other multi-channel initiatives. This might mean taking data from an existing CRM system, looking at loyalty card data, or even reviewing inventory data. But what if you don't have enough of this data to make a well-informed decision about where to locate a new store or what merchandise you need to stock in order to make that store successful?

One way to do this is to analyze data from third-party providers, such as Experian, Dun and Bradstreet, or the US Census Bureau, and combine it with internal customer data to identify the factors that indicate the highest propensity to buy. Examples include ethnicity mix, age, and consumer spending on similar goods and services. By determining these market factors up front, you can optimize your real estate investment by opening stores in the right location and putting the right types of merchandise in each store to drive profitability.

### **Fulfilling the Requirements of Data Blending**

Data blending empowers data analysts to incorporate data of any type or source into their analysis for faster, deeper business insight, but how do organizations enable a data analyst to perform data blending? Many line-of-business analysts have abandoned spreadsheets and custom work projects in favor of *Alteryx Analytics* because it fulfills today's data blending requirements.

**Understand the progression of data.** The drag-and-drop workflow environment in *Alteryx* allows analysts to build out analytic datasets the way they think. It lets the analyst understand how data progresses through the process without any "black boxes" and quickly identify where issues may lie. This drag-and-drop technology allows analysts to focus more on the data and less on the technology by eliminating the need for coding or programming.



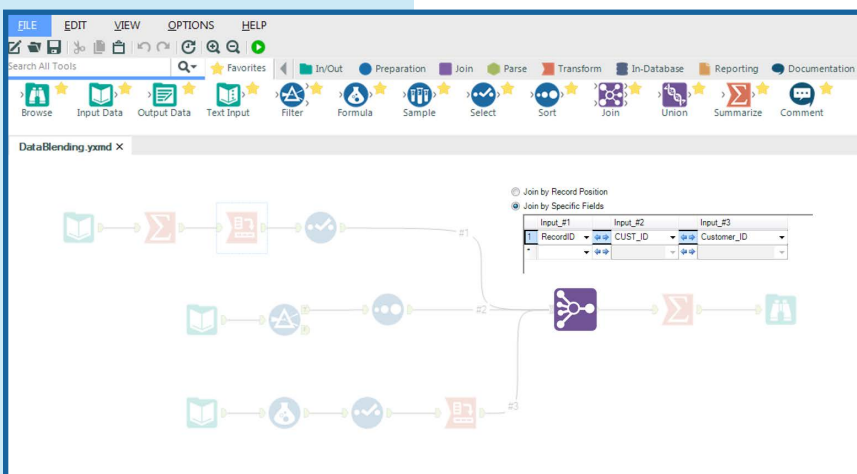
**Enable direct access to data.** Alteryx gives analysts direct access to data of any type or source to help deliver a more complete view of the insight they need to make more informed decisions. Because they no longer need to rely on overworked IT staff or data specialists, analysts can access all the information they need to make informed business decisions, including:

- Local data (spreadsheets, user device generated data, enterprise data warehouses, etc.)
- Third-party data (Dun & Bradstreet, Experian, Tom Tom, US Census, etc.)
- Cloud/social data (Twitter, Facebook, Salesforce.com, Marketo, etc.)
- Other analytics platforms (SPSS, SAS, Google Analytics, etc.)

**Expedite data cleansing and preparation.** Studies estimate that 60% to 80% of an analyst’s time is spent preparing data for analysis. Alteryx offers extensive tools for data preparation and data cleansing to speed up the time to create the right dataset, without having to rely on outside intervention. With options for restructuring, reformatting, and filling in missing or incomplete data, Alteryx ensures that data quality, integration, and transformations are done by the people who know the data and understand the business best, leading to the right dataset in the least amount of time.

**Simplify blending of data.** Alteryx gives users complete flexibility in joining multiple datasets thanks to an array of tools that can address virtually any data situation. Joining data in Alteryx is not limited to just one field or column; Alteryx allows data of any type or level to be brought together. This means that

data can be joined at both the record and field levels, and it can even be expanded to include multiple key fields. What’s more, Alteryx is flexible enough to join data from non-identical fields as well as incorporate spatial characteristics, such as customer points, into the dataset. Other tools, such as Fuzzy Matching, give users the ability to match two datasets based on non-matching data—typically names and addresses. In addition, tools such as Append Fields, Find Replace, and Make Group allow users to do even more to effectively blend or refine their resulting dataset.



## About Alteryx

Alteryx is the leader in data blending and advanced analytics software. *Alteryx Analytics* provides analysts with an intuitive workflow for data blending and advanced analytics that leads to deeper insights in hours, not the weeks typical of traditional approaches. Analysts love the Alteryx analytics platform because they can deliver deeper insights by seamlessly blending internal, third-party, and cloud data, and then analyze it using spatial and predictive drag-and-drop tools. This is all done in a single workflow, with no programming required. Thousands of customers, including Experian, Kaiser, Ford, and McDonald's, and 200,000+ users worldwide rely on Alteryx daily. Visit [www.alteryx.com](http://www.alteryx.com) or call 1-888-836-4274.

# alteryx

230 Commerce, Ste. 250, Irvine, CA 92602  
+1 714 516 2400  
[www.alteryx.com](http://www.alteryx.com)

**Automate and repeat processes.** With the amount of ad-hoc analysis required by today's analysts, what if there were a way to make this process easier, faster, and repeatable? With Alteryx, there is. Alteryx workflows can easily be saved and repeated for further data blending, processing, updates, and analysis. Updating the analysis or report is as simple as updating the data input(s).

**Output data easily.** Once the heavy lifting of data blending is completed, analysts can implement this data into the right processes of the business. This means that resulting outputs can then be pushed back into a database, incorporated into an operational process, analyzed further using statistical, spatial, or predictive methods, or pumped into visualization software, such as QlikView or Tableau.

## Conclusion

Traditionally, data was the domain of IT and data scientists—doling out access to a select few via careful SQL queries, heavily structured reports, BI dashboards, and, maybe, programmatic access. With first-generation tools, the process to generate results was long, expensive, and difficult. Highly skilled and expensive data scientists would work with Ph.D.-level statisticians and IT professionals to obtain and massage data, develop complex analytic models and, ultimately, generate analytic results. Analysts were left at the door waiting for results that would then have to be extensively reviewed, tested, and re-adjusted to fit their original business cases. The result? Data was often neither timely nor adequate to answer their questions.

While traditional data analysts use traditional IT tools to generate reports on historic data, today's analysts must extend that capability with their business insight and natural creativity to find information their organizations really need. With improvements in information technology and the constant influx of Big Data, a flood of new opportunities for business insight has appeared.

Empowered by next-generation tools such as Alteryx, today's analysts can now do what previous generations of analysts could only dream of doing. These analysts are able to perform data blending to create the analytic dataset they need to deliver the deeper business insights they require.

To learn more about Alteryx and see firsthand how you can use its data blending, processing, analytics, and reporting capabilities to your advantage, visit [www.alteryx.com](http://www.alteryx.com).