alteryx

The Definitive Guide to Data Blending
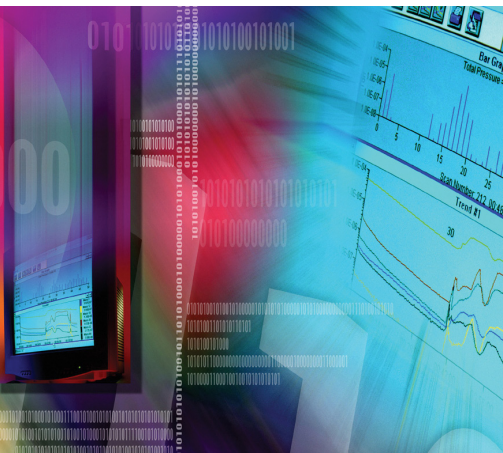
Introduction

Data is at the heart of today's interconnected world. It is captured in almost every aspect of our lives—from 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 fulfill their requirements to get the results they want — whether it pertains to sales, marketing, operations or even finance in many organizations. It is up to them to use this knowledge and business insight to make sense of all of this 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 are having difficulty with this new data challenge because their traditional tools and approaches are not capable enough to deal with this environment. Utilizing spreadsheets, manual processes or custom scripting or relying on data scientists to build analytical datasets just doesn't work or is too time consuming to handle the amount of ad-hoc requests that analysts deal with.

Today's analysts need to live at the forefront of innovation to keep their organizations competitive. For them, the real game is about extending 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.


```
CREATE TABLE department
(
 DepartmentID INT,
 DepartmentName VARCHAR (20)
);

CREATE TABLE employee
(
   LastName VARCHAR (20),
   DepartmentID INT
);

INSERT INTO department V
INSERT INTO department V
INSERT INTO department V
INSERT INTO department V

INSERT INTO department VALUES ('Rafferty', 31);
INSERT INTO department VALUES ('Jones', 33);
INSERT INTO department VALUES ('Heisenberg', 33);
INSERT INTO department VALUES ('Robinson', 34);
INSERT INTO department VALUES ('Smith', 34);
INSERT INTO department VALUES ('John' NULL);
```

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

1

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).
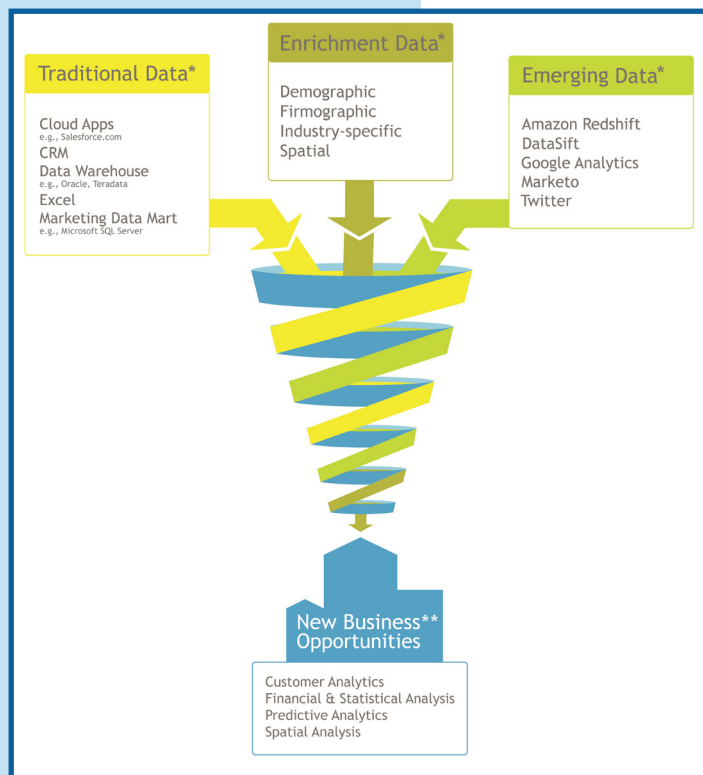
Data blending is needed when an organization's data management processes and infrastructure are insufficient for bringing together analytic or specific datasets required by line of business groups. It can, for example, readily bring together disparate data like customer information from a cloud sales automation system (like Salesforce.com) with click stream 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 data warehouses, data marts or other systems of record within an organization—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.

Traditional Data*

Cloud Apps
e.g., Salesforce.com
CRM
Data Warehouse
e.g., Oracle, Teradata
Excel
Marketing Data Mart
e.g., Microsoft SQL Server

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

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 many different cases that data analysts are involved in. Below are a few examples where data blending can help.

Sales and Marketing

For every organization, growing revenue typically coincides with targeting prospects who are ready to buy. Marketing departments spend lots of time trying to identify these prospects so they can focus campaigns on them, 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 like Salesforce.com—while marketing prospect data is stored in a separate system like Marketo. And, information about customer and prospect activity on their website is captured by web analytics technology like Google Analytics. Historically, finding the relevancy of this data might require generating spreadsheet-based reports from both the marketing automation and CRM systems. From there an analyst might need to combine these into one spreadsheet with multiple tabs and construct formulas using VLOOKUP functions to reference information that is relevant. Or, they may just combine the two sheets 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 it is located in. All you need is the right credentials to access

the data, pull the data from the right systems and start combining the data on common fields, blending in the specific information you are looking for. You can combine data on customer ID for instance, or understand what products or services are having the biggest impact not only on sales, but also what is driving 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, the importance data plays in approving people for loans or a credit card can make the difference on whether they are likely to commit fraud and reduce the financial risk for their organization. It can be as simple as building out a customer model and then track and trend detailed client information over several years. Typically this means a lot of data from several sources, such as web logs which can be unstructured or need to be cleaned up, or even several different 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

To truly have a successful store, understanding your potential customers and prospects is crucial. To find this out, you may need to look at customer spend levels, purchase history and path-to-purchase. 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 inventory data. But what if you don't have enough of this data, or what if you're trying to better determine where to locate that store and what merchandise to support it, to be successful?

One way to do this is by analyzing 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, and make sure you are opening stores in the right location and putting the right types of merchandise in each store to make it successful.

Fulfilling the Requirements of Data Blending

Data blending empowers a data analyst to incorporate data of any type or any 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 with no "black boxes", and quickly identify where issues may lie. This drag and drop technology allows an analyst to focus more on the data and less on the technology by eliminating the need for coding or programming.
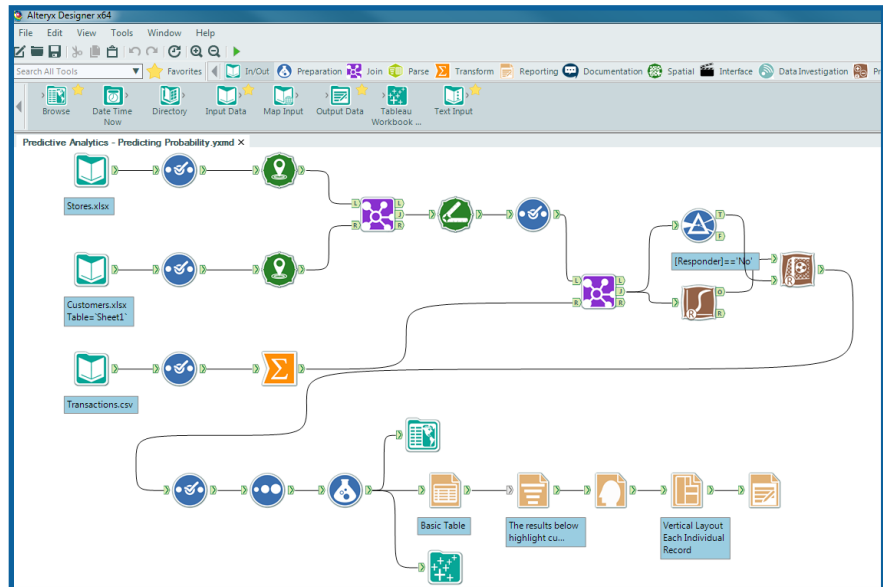
Enable direct access to data— Alteryx gives analysts direct access to data of any type or any source to help deliver a more complete view of the insight they need to make more informed decisions. No longer do they need to rely on others to get the data they need. Removing IT or data specialists from the equation, they can access the full fidelity of:

• 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— It's been estimated 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 data 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.
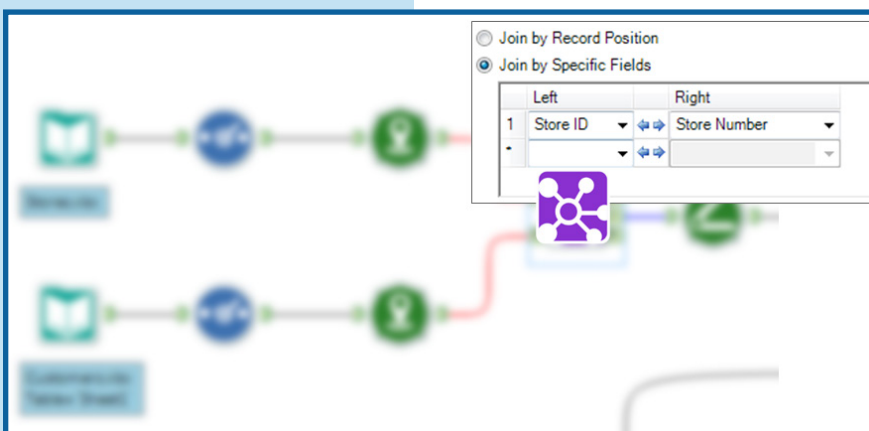
Simplify blending of data— Alteryx give users complete flexibility when it comes to joining multiple datasets through an array of tools to help with virtually any data situation. Joining data in Alteryx is not limited to just one field or column; Alteryx allows data of any type or any level to be brought together. This means that it can be joined at both the record and field level and it can even be expanded to include multiple key fields. Alteryx does not discriminate; it is flexible enough to join data on non-identical fields as well as incorporate spatial characteristics such as customer points

to be combined into your data. Other tools like 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.

Automate and repeat processes— With the amount of ad-hoc analysis being required by today's analysts, what if there was a way to make this process easier, faster, and repeatable? 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 done; it is now time to 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 PhD 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. Since the cards, so to speak, weren't fully in the analyst's hands, the results they sought were 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 organization really needs. 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 like Alteryx, today's analysts can now do what previous generations of analysts could only dream of. These analysts are able to perform data blending to create the analytic dataset they need to deliver the deeper business insight they require.

To learn more about Alteryx and see first-hand how you can use its data blending, processing, analytics and reporting capabilities to your advantage, visit www.alteryx.com.

alteryx