

Making Everything Easier!™

Alteryx Special Edition

Data Blending

FOR
DUMMIES[®]
A Wiley Brand

Learn to:

- Understand what data blending is and why it is important
- Combine all necessary data to get the answers your business needs
- Simplify common data tasks to focus on finding new business opportunities

Brought to you by

alteryx

Michael Wessler, OCP & CISSP



About Alteryx

Alteryx is a 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. More than 600 customers and thousands of data analysts worldwide rely on Alteryx daily.

Visit **www.alteryx.com** or call 1-888-836-4274.

Data Blending

FOR
DUMMIES[®]
A Wiley Brand

Alteryx Special Edition

by Michael Wessler, OCP & CISSP

FOR
DUMMIES[®]
A Wiley Brand

Data Blending For Dummies® Alteryx Special Edition

Published by
John Wiley & Sons, Inc.
111 River St.
Hoboken, NJ 07030-5774
www.wiley.com

Copyright © 2015 by John Wiley & Sons, Inc., Hoboken, New Jersey

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the Publisher. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permissions>.

Trademarks: Wiley, the Wiley logo, For Dummies, the Dummies Man logo, A Reference for the Rest of Us!, The Dummies Way, Dummies.com, Making Everything Easier, and related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries, and may not be used without written permission. Alteryx is a registered trademark of Alteryx, Inc. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc., is not associated with any product or vendor mentioned in this book.

LIMIT OF LIABILITY/DISCLAIMER OF WARRANTY: THE PUBLISHER AND THE AUTHOR MAKE NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE ACCURACY OR COMPLETENESS OF THE CONTENTS OF THIS WORK AND SPECIFICALLY DISCLAIM ALL WARRANTIES, INCLUDING WITHOUT LIMITATION WARRANTIES OF FITNESS FOR A PARTICULAR PURPOSE. NO WARRANTY MAY BE CREATED OR EXTENDED BY SALES OR PROMOTIONAL MATERIALS. THE ADVICE AND STRATEGIES CONTAINED HEREIN MAY NOT BE SUITABLE FOR EVERY SITUATION. THIS WORK IS SOLD WITH THE UNDERSTANDING THAT THE PUBLISHER AND AUTHOR ARE NOT ENGAGED IN RENDERING LEGAL, ACCOUNTING, OR OTHER PROFESSIONAL SERVICES. IF PROFESSIONAL ASSISTANCE IS REQUIRED, THE SERVICES OF A COMPETENT PROFESSIONAL PERSON SHOULD BE SOUGHT. NEITHER THE PUBLISHER NOR THE AUTHOR SHALL BE LIABLE FOR DAMAGES ARISING HEREFROM. THE FACT THAT AN ORGANIZATION OR WEBSITE IS REFERRED TO IN THIS WORK AS A CITATION AND/OR A POTENTIAL SOURCE OF FURTHER INFORMATION DOES NOT MEAN THAT THE AUTHOR OR THE PUBLISHER ENDORSES THE INFORMATION THE ORGANIZATION OR WEBSITE MAY PROVIDE OR RECOMMENDATIONS IT MAY MAKE. FURTHER, READERS SHOULD BE AWARE THAT INTERNET WEBSITES LISTED IN THIS WORK MAY HAVE CHANGED OR DISAPPEARED BETWEEN WHEN THIS WORK WAS WRITTEN AND WHEN IT IS READ.

For general information on our other products and services, or how to create a custom *For Dummies* book for your business or organization, please contact our Business Development Department in the U.S. at 877-409-4177, contact info@dummies.biz, or visit www.wiley.com/go/custompub. For information about licensing the *For Dummies* brand for products or services, contact BrandedRights&Licenses@Wiley.com.

ISBN 978-1-119-03766-8 (pbk); ISBN 978-1-11903764-4 (ebk)

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

Publisher's Acknowledgments

Some of the people who helped bring this book to market include the following:

Senior Project Editor: Zoë Wykes

Acquisitions Editor: Amy Fandrei

Editorial Manager: Rev Mengle

Business Development Representative:
Kimberly Schumacker

Project Coordinator: Melissa Cossell

Alteryx Contributors: Matt Madden,
Bob Laurent, Paul Ross

Table of Contents

.....

Introduction	1
About This Book	1
Foolish Assumptions	2
Icons Used in This Book.....	2
Beyond the Book.....	3
Where to Go from Here	3
Chapter 1: The Rise of the Data Analyst	5
Introducing Today’s Data Analyst	6
Supporting the Business Decision Maker	7
Data Analysis in the Data Blending World	8
Chapter 2: Understanding Data Blending	11
Exploring the Wide World of Data	12
Accessing Multiple Data Sources	14
Accessing the Relevant Data	16
Defining Data Blending	18
Differentiating Data Integration from Data Blending.....	20
Chapter 3: The Building Blocks of Effective Data Blending	21
Identifying the Components of Data Analysis	22
Moving through Data Blending	23
Data preparation and cleansing.....	23
Joining data	24
Automation and repeatability	25
Empowering the Analyst and Decision Makers.....	27
Chapter 4: Using Data Blending in the Real World . . .	29
Maximizing Business Agility with Better Tools.....	30
Intuitive workflows	30
Deeper business insight.....	31
Faster access to information.....	31
Implementing Data Blending in Business.....	32
Putting the Right Tools in the Hands of the Data Analyst.....	32
Using Alteryx Designer	34

**Chapter 5: Ten Things to Consider
with Data Blending.....39**

- Integrate Predictive Analytics into Your Business 39
- Gain Insights Rapidly..... 40
- Augment the Tools You Now Use 40
- Work behind the Scenes 41
- Consider Spatial Analysis 41
- Leverage Analytic Apps..... 42
- Incorporate Visualization 43
- Improve Business Analyst and IT Relations 43
- Drive Down IT Support Costs 44
- Feed the Downward Stream..... 44

Introduction

Data analysts support their organization's decision makers by providing the timely information and answers to key business questions needed to drive the business forward. Decisions are only as good as the information they are based on, so data analysts strive to use the best and most complete information possible. Unfortunately, as the amount of data in the world grows exponentially, the time available to identify and combine all the data sources where insight may exist is shrinking. Being successful in business depends on being agile and knowing more than your competitor does; thus, the speed of business is accelerating. Under this pressure and with more work and less time to do it in, the data analyst is indeed in a challenging position.

Fortunately, the data analyst *can* become faster at delivering better information and results — all with less effort and complexity. Giving the analyst tools to *blend data* from multiple data sources ensures that all the relevant data, no matter where it exists, is available to the analyst. Data blending reduces the time to prepare and join data and empowers data analysts to be self-reliant in their job.

About This Book

Data blending is important because it allows data analysts to access data from all the relevant data sources: Big Data, the cloud, social media, third-party data providers, in-house databases, department data stores, and more. Historically, the challenge of data analysts has been accessing this data and then cleansing and preparing the data for analysis. These stages of access, cleansing, and preparing data are complex and time intensive. Easy-to-use software tools that reduce the burden of this data preparation and turn data blending into an asset greatly empower the data analyst to become more effective and open new opportunities to the business.

The focus of this book is how data blending is used and what it can provide the data analyst working to support business decision makers. I identify what features to look for in data blending tools and how to successfully deploy these tools and data blending within your business.

Foolish Assumptions

Assumptions often cause problems, but still they have their place — and if they weren't useful, people wouldn't make them. I sometimes have to make assumptions too, and writing this book is a prime example. Mainly, I assume that you know something about data, perhaps have heard of Big Data and the cloud, and also know something about data analysis. Hopefully, you also know what data analysts are, the tools they sometimes use, and the important work they do. As such, this book is written primarily for data analysts and the business folks they support.

Icons Used in This Book

Throughout this book, you occasionally see special icons that call attention to important information. You won't find cute little emoticons, but you'll definitely want to take note. Here's what you can expect.



This icon points out things you'll be glad I mentioned later on. This is the stuff you want to remember when you start using the material on your own.



I try to keep the techie stuff to a minimum, but I am a techie person at heart and old habits die hard! These are technical tidbits that aren't essential, but they are nice to know.



This icon points out pieces of sage wisdom that I wish someone had told me when I was learning this subject.

Beyond the Book

Alteryx Analytics provides data analysts with an intuitive workflow for data blending and advanced analytics that leads to deeper business insights in hours rather than the days or weeks required of most analytics solutions. To learn more about Alteryx Analytics, visit www.alteryx.com.

Where to Go from Here

Blending data doesn't have anything to do with a kitchen blender, although I do use a cooking analogy later in this book. But, just like cooking, having all the right ingredients at the start is very important; no one likes to begin a project only to realize they forgot something. Furthermore, just having all the ingredients isn't enough; if you want to be successful, you need to have a good recipe to follow as well.

In this book, I give you everything you need to become good at data blending. Or, at least I show you how to start a data blending project with the confidence that you'll be successful at the end.

If you're trying to learn about data blending and the role of the data analyst for the first time, the task may seem daunting, but with *Data Blending For Dummies*, Alteryx Special Edition, you have help to guide you on this exciting adventure.

If you don't know much about data analysis or what a data analyst does, Chapter 1 would be a good place to start. If you're ready to jump into data blending, Chapter 2 is for you. However, if you see a particular topic that interests you, feel free to jump ahead to that chapter. Each chapter is written to stand on its own, so feel free to start reading anywhere or to skip around. Read this book in any order that suits you (although I wouldn't recommend upside down or backward). I promise that you'll put the book down thinking, "Wow, I didn't know this stuff could be so easy!"

Chapter 1

The Rise of the Data Analyst

.....

In This Chapter

- ▶ Identifying the requirements of a data analyst
 - ▶ Providing value to today's business decision makers
 - ▶ Building the case for data blending
-

The role of the data analyst has expanded as organizations have gone from the primitive days of data processing on mainframes into the fast-paced, cloud-driven world of today. Businesses have always had people filling the role of data analysts, but that role has grown with technology's increased capabilities. This growth is good because it has enhanced the capabilities of data analysts and made their contributions to a company's success even more important.

Today, the data analyst is no longer shackled to the IT department. Modern analytics tools that can be used without writing code or sitting through tons of training classes allow data analysts to access more data sources, join data, perform complex calculations, and quickly answer the tough business questions that are asked. These capabilities firmly put the data analyst in the critical path traveled by business decision makers who determine a company's direction.

This chapter looks at how the role of the data analyst has evolved and what factors have influenced that growth. It also sets the stage for continued evolution with data blending.

Introducing Today's Data Analyst

The data analyst's tools are always evolving, and the data analyst's responsibilities are growing. Data processing and analytics capabilities are ever increasing, so what was impossible only a few years ago is now commonplace. To learn more about how the data analyst has evolved, see the "Evolution of the data analyst" sidebar.

Today's data analysts are a force to be reckoned with; they know their companies' business rules, understand how to use technology, know what data they need, and are not afraid to venture out on their own or collaborate with dedicated IT resources to get the results required by the companies they work for. Data analysts typically support their organizations from within the business side (not the IT side) of their companies, such as:

- ✓ Sales
- ✓ Marketing
- ✓ Operations
- ✓ Product and service design
- ✓ Finance and accounting

Successful data analysts provide the best information for their companies by leveraging these skills:

- ✓ Working from within a business perspective
- ✓ Knowing business requirements better than anyone else
- ✓ Using a logical and analytical mindset to solve problems
- ✓ Accessing and analyzing data to provide a result or answer that satisfies an outstanding business question or solves a business problem

The value that a data analyst brings to a company is that of answering critical business questions based on calculations, analytics, and models performed on very specific datasets. Technology has evolved to give the data analyst greater, but not yet complete, ability to provide answers to straightforward questions in a timely manner, as I discuss later.



In previous decades, IT was more static and defined, which doesn't mean that IT was *easier* than it is today. In many ways, IT was more difficult in the past. Limitations in processing

Evolution of the data analyst

In the early days of data processing, the computing environment was conceptually simple in terms of what options were available — limited sources of structured data in a static environment with very few tools to process that data. This proved to be both a blessing and a curse:

- ✔ Fewer tools and sources of data allowed experts to dig deep into what features were available at the time. The relatively few people with the education and access to these tools became experts in their use.
- ✔ The limitation of tools and processing power hindered the ability to get information out of the data that was available. Additionally, only a few qualified people were able to perform the complex work.

As data grew in volume, variety, and velocity, the only way to get access to the data organizations needed to make decisions was to hire and train expensive data scientists. These individuals were skilled in data access languages such as SQL and Python, and the complex code they wrote served as the barrier to ensure they remained at the center of data analysis.

As new workflow-based analytics tools became available that could deliver the same insights faster and easier, data analysts in the lines of business had an alternative to waiting on IT and their data scientists. Data analysts themselves became empowered by these new tools and today can find the information they need without heavy IT department involvement.

power coupled with terse command-line interfaces dictated that IT was not for the faint of heart. As hardware and software have improved to make some aspects of computing easier, more computing options and industry turbulence have added complexity in new areas. Basically, for those in IT, one set of challenges has been replaced with a new set.

Supporting the Business Decision Maker

Questions from an organization's decision makers are what drive data analyst requirements. The business decision makers are the executives and senior managers who determine what products and services to develop and sell, which

markets to enter or avoid, and which business opportunities or challenges exist that require their action.

To be effective, business decision makers ask their data analysts tough questions. For example, they may ask, “Should I increase marketing in the northern Indianapolis suburbs where the average new sports car customer is in the age range of 30 to 40 and has been employed for more than one year?” This is a difficult question to answer without the right tools and data, but it *is* a question that could identify a new business opportunity and is therefore sent to data analysts to answer.

The needs of business decision makers center on:

- ✓ Getting the right data and answers to make key decisions in a timely manner
- ✓ Gaining deeper business insight to really understand the changing market environment
- ✓ Utilizing reports and analytics that present data in a business context rather than from an IT perspective
- ✓ Having the ability to share their insights with other business colleagues and run their queries again with different input values

The demands of business decision makers are high because fast-moving markets, fleeting opportunities, tight profit margins, and fierce competition don’t allow for delays or incorrect results. Data analysts must continue to evolve their tools and processes to meet these increasing demands.

Data Analysis in the Data Blending World

To understand why data blending (discussed in detail later in the book) is necessary, understanding the current data analysis environment is important. Technology has evolved to make the job of the data analyst better, but the full potential is not yet fully realized. Using traditional approaches such as spreadsheets, data analysts can do the following:

- ✔ Find answers to straightforward questions
- ✔ Work with simple algorithms and formulas
- ✔ Access readily available, prepared data from a few sources

The evolution of hardware and software has enabled data analysts to perform simple to low-complexity work. However, when the business questions become more difficult, requiring more complex computing algorithms or large amounts of data from different sources, data analysts often seek the help of IT experts.

In cases where the requirements exceed what data analysts can provide by themselves, a company's IT experts partner with data analysts to:

- ✔ Identify, cleanse, and prepare relevant data from a vast sea of frequently irrelevant data
- ✔ Join data from multiple input sources and channels
- ✔ Develop complex algorithms and programs to process data based on business criteria
- ✔ Perform the computational analysis
- ✔ Format and display the output results in a usable format
- ✔ Package and automate the preceding process so the same question can be asked again as needed

Working together, data analysts and IT do accomplish more, but this relationship does have some downsides:

- ✔ Reliance on IT for support introduces long delays for the data analyst because IT staff is seldom able to assist immediately.
- ✔ Maintaining the requisite large, highly skilled IT staff is expensive.
- ✔ Many analytic tools used by IT are expensive and complex, requiring customization and training before their use.
- ✔ IT staff knows technology, but their knowledge of business rules is relatively weak. This necessitates long and complex Scope of Work (SOW) documents, and if input from data analysts is lacking, mistakes can be made.

Although collaboration with IT using complex tools can provide results (if there's enough time), the process of data analysis is itself a complex operation with multiple challenges, as you can see in Figure 1-1.

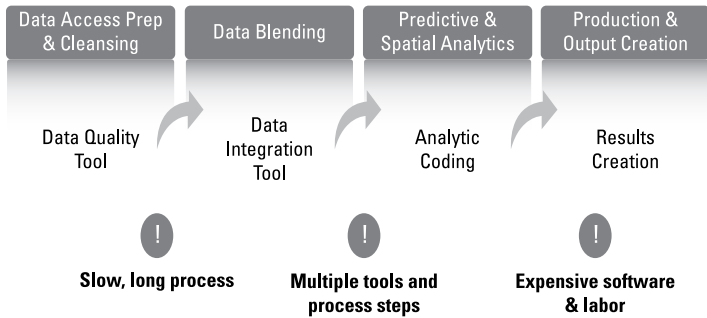


Figure 1-1: Traditional data analysis process and challenges.

Each process takes time and is complex, requiring multiple advanced tools. Each step is also expensive and reliant on IT or others within the organization. Unfortunately for the data analyst, the business environment is continuing to evolve with increased demands, such as:

- ✔ Greater focus on using complex analytical analysis
- ✔ Expansion of data in terms of size, complexity, and number of input sources
- ✔ Increased requirements for ad-hoc analysis and reporting
- ✔ Expectations of faster response times

Business requirements have increased the demands of the data analyst to the point where the current slow, methodical relationship with IT is no longer sufficient. What's needed are more tools that are data analyst friendly and perform the complex steps without the dependence on IT. Fortunately, tools that combine data blending with powerful, easy-to-use workflows, such as those provided by Alteryx, meet the needs to fully empower the data analyst.

Chapter 2

Understanding Data Blending

.....

In This Chapter

- ▶ Discussing the forms of data in the real world and how it naturally exists
 - ▶ Understanding the challenges of accessing data
 - ▶ Retrieving all necessary data
 - ▶ Explaining data blending in support of business analytics
 - ▶ Knowing the difference between data blending and data integration
-

Making the right business decisions is crucial to success for any profitable company. The best decisions are arrived at by identifying all the right data, applying the correct analytical methodology to that data, and correctly interpreting the results to form the decision. In business, this is a collaborative effort of data analysts and business decision makers working together.

Basing your decision on the right data is critical. Without the right data, you may have incomplete results on which to make your decisions. In a worst-case scenario where you have the wrong data, you get incorrect results that send you down the path of a disastrous decision. Knowing *where* the right data exists and *how* to bring all that data from different sources into your decision-making process is at the core of good decision making in business.

This chapter looks at what data blending is and how it is used in decision making.

Exploring the Wide World of Data

In the real world, data comes in all different shapes, sizes, and formats. Increasingly, data comes from everywhere, ranging from corporate databases that are decades old to information generated on smartphones during the time it takes to read this sentence.

Data exists in several different formats:

- ✔ **Structured:** Traditional data stored in a neat record format with well-defined data types such as fixed field numeric and alphanumeric characters. Structured data is the basis for most existing and legacy databases and is relatively easy to store and manage.
- ✔ **Semi-structured:** Unformatted or loosely formatted numbers or characters inside a field but with little or no structure within the field. A social media post such as a Twitter tweet is an example of semi-structured data. Semi-structured data is more complex to store and process than structured data.
- ✔ **Unstructured:** Data that is not text based, such as pictures, images, or sound files generated by devices or posted on social media. Unstructured data is a challenge to manage because it is large in size, difficult to catalog and index, and problematic to store in databases.

In addition to characterizing data based on its format, a useful way to view data is based on its nature. Data naturally falls into these three categories:

- ✔ **Traditional:** Data that lives in existing or legacy databases and is often in a well-structured format. Rows in an Excel spreadsheet, records in an accounting database table, or account information in an insurance mainframe database are examples of traditional data. More modern examples include information in data warehouses and cloud applications.
- ✔ **Enrichment:** Data that is industry specific or special purpose and used to supplement (or enrich) existing data. For example, spatial grid coordinates identifying where customers like to shop would enrich sales information, or

demographic information about a customer's background could help a retailer looking at traditional sales data.

- ✔ **Emerging:** Data that is often related to Big Data (see the sidebar, “Why Big Data is so big”) as well as other sources such as social media or marketing automation data are common examples of emerging data. This is newer, more valuable, and often the most difficult data to identify and leverage. This data helps sales and marketing staff follow their brand and identify useful patterns.

Why Big Data is so big

Big Data is, well, excuse the pun, a *big* thing in organizations today. Actually, Big Data isn't a thing at all but is rather a classification of data that is very large, generated very fast, and comes from a wide variety of sources, formats, and structures.

The IT industry research group Gartner defines Big Data as:

“Big Data are high-volume, high-velocity, and/or high-variety information assets that require new forms of processing to enable enhanced decision making, insight discovery, and process optimization.”

Data categories such as enrichment and emerging data are most likely to include sources of Big Data, but even some traditional data components, particularly when dealing with cloud technologies, could be Big Data related. Unstructured and semi-structured formats are most common within Big Data, but structured data can also be included in some cases.

Analysts will have to work with new types and sources of data, with Big Data being one such example. Here

are a few common examples of Big Data:

- ✔ Social media feeds such as Facebook and Twitter where a relentless stream of semi- and unstructured data include text, sound, and images
- ✔ Point-of-sale terminal information related to credit card purchases, customer identity and demographic information, and inventory control data
- ✔ Online sales and shopping information as well as consumer tracking, demographics, and website metrics
- ✔ Scientific, medical, and device (Internet of Things) data showing images, diagnostics, and telemetry of nearly any device imaginable

Big Data is important to business and data analysts because it represents much of the best, new data available for better decision making. Any tool that purports to support data blending needs to access Big Data quickly and intuitively.

Based on the business question being posed, various categories of data (traditional, enrichment, and emerging) are funneled into the analysis process. The output is a better, deeper understanding of the market environment to identify and capitalize on business opportunities. Understanding the data types and formats is important for the analyst. Knowing the characteristics of the data allows the analyst to better pick and choose which sources have the right data to be used to build the dataset that will feed the analysis engine during processing.

Accessing Multiple Data Sources

Bringing data together from different sources is how data adds value to the decision-making process. In the real world, the data you need usually won't sit neatly in a predefined database inside a fully prepared data table just waiting for you to access it. In most cases, data must be obtained from different sources to add the depth and wide scope necessary for the best possible analysis and decision making.

The simple fact is that the best decisions will be made only when all the relevant data is available for analysis. The key data almost always comes from multiple data sources and often comes in different formats.

Data analysts face the following challenges when attempting to access data from different data sources:

- ✔ Data outside users' workstations is typically owned and managed by the IT department. Requests for access must involve the IT staff.
- ✔ Access to the right data is a multistep process, often with multiple parties involved with lengthy documentation. This equates to long turnaround times for access.
- ✔ Data analysts have limited tools to access more complex data sources. While access to lower-end databases is possible, connectivity to more powerful databases and useful access to that data via SQL is often both manual and complex.

- ✔ IT staff can assist (once they have time), but IT staff doesn't know the core business processes and requirements as well as the data analyst. If data access is left solely to the IT department, the results could be slow and incorrect.
- ✔ Access to enrichment and emerging data can represent a paradigm shift for IT, especially if the staff is experienced only in traditional database technologies. The associated learning curve can extend the turnaround time for data access.
- ✔ Once data is obtained, differing data formats, characteristics, and amounts of irrelevant data greatly add to the time and complexity required to cleanse and prepare the data for analytical processing. This data cleanup and preparation is one of the most time-consuming and complex steps performed by data analysts.

Despite these challenges, the rewards of accessing multiple data sources far outweigh the challenges. Given these challenges, solutions to improve the situation for data analysts include the following:

- ✔ Tools that empower the data analyst to not require IT support at the technical level. IT involvement should be limited to granting security access to the data.
- ✔ Tools that have predefined access to multiple data sources, including traditional, enrichment, and emerging sources. Analysts should spend their time analyzing data, not struggling to access data.
- ✔ Access to the new paradigm of data such as Big Data, cloud data, and third-party data brokers. Access should be streamlined, intuitive, and easily implemented without IT involvement.
- ✔ Data cleansing and preparation tools that must be powerful yet intuitive and easily usable by the data analyst. Because these steps have historically been complex and time consuming, special emphasis should be placed to use tools that reduce the complexity while speeding up the process for the analyst.

The more that data analysts are empowered and become self-reliant to find, access, prepare, and ultimately analyze and process the relevant data, the more effective the data analysts

will be in their role. Tools that allow analysts to perform these tasks easily will encourage the analysts to seek out new, relevant data because the process isn't arduous, further improving the outcome for the business.



Getting a second opinion is always a wise practice with medical decisions, so why not get your data from multiple sources when making key business decisions? Many intricate relationships within data can be identified only after datasets from different sources are combined and analyzed. Only through data blending can this kind of analysis occur.

Accessing the Relevant Data

The nature of Big Data and IT is more data, bigger data, and faster incoming data. Furthermore, once data is created, it is rarely deleted. For a variety of reasons, companies tend to store data for years (or even decades) even if it probably won't be used. These factors create myriad sources that could *potentially* hold the data required by the analyst.

Despite the wide range of potential data sources, analysts likely know where the relevant data exists. Odds are, analysts know the business processes very well and understand where data comes to the business from, where and how (at a non-technical level) data is processed, and where that data goes once processing is complete.

At a minimum, analysts know what traditional data is needed for business processing to occur. If the analysts are business savvy and experienced, they have also used enrichment data to add value to business processes and analytics. In a best case, analysts are also now using emerging data to gain greater awareness of product performance and customer perception to really add more intelligent business planning.

The great struggle is, how do analysts access all the relevant data to build a usable dataset to support the necessary analytical processing? Analysts know what data to use, but not necessarily how to get to it. Fortunately, tools such as Alteryx have the capability to access these relevant, traditional enrichment and emerging data sources:

- ✔ Oracle, IBM, and Teradata databases
- ✔ Microsoft Access, SQL Server, SharePoint, and Excel products
- ✔ Amazon Web Services, Salesforce, SAS, and Google Analytics software and services
- ✔ MongoDB and Hadoop products
- ✔ Salesforce.com and Marketo cloud-based services
- ✔ Twitter, Facebook, Foursquare, and other social media services

That's MY data!

Organizations can become very territorial and protective of their data — often for good reason. The most sensitive details of a company, its customers and employees, and its sales information are within its data, typically in databases managed by the IT department. Legal, professional, and ethical requirements to safeguard specific data elements are very strong and must be adhered to. Understandably, access to data by anyone, especially those outside the immediate group of data owners and users, is scrutinized and often challenging to obtain. No company wants to make the news by having its customers' data stolen as has recently happened with several big-name companies. The damage to those companies' reputations and their legal liabilities are staggering.

Unfortunately, in some companies (and even worse in many government agencies), access to the data even for reasons that benefit the business can be difficult to obtain. Database administrators (DBAs) and

data owners justifiably take data security very seriously and sometimes go overboard protecting their data. I've witnessed large IT projects fail mainly due to data owners blocking access to key data from outside systems, which is truly unfortunate.

As a data analyst, what are you to do when faced with such adversity? First, resist any temptation to use a back-door method to access data you're not authorized for; those in IT call that *data theft* and *hacking*, and it will get you fired. Worse yet, if the data is sensitive, you can face legal ramifications. The proper response is to follow the procedures to request data access and limit it to the least privilege and amount of data you require, which will greatly appease the security team. Inform the data owners why you are requesting access and what initiative it supports. If you still face opposition, let your management chain know and unleash your project sponsor or key business decision maker on the opposing data owner, which often yields positive results.

The intuitive, easy-to-use interface of Alteryx allows direct access to traditional, highly structured databases and data sources within a company, ranging from larger databases to smaller micro applications hosted on power users' machines. Of even greater value is the data analyst's ability to use Alteryx to access all relevant data, from Big Data, enrichment data, and emerging data without dependency on IT, lengthy delays, or unnecessary complexity.



If the previous list of data sources seems bewildering, and you perhaps had to look up a few of the names, don't worry. The explosion of Big Data, cloud computing, and social media has brought many new players to the IT field and has caused long-standing pillars of the technology world, such as IBM, Oracle, and Microsoft, to evolve to meet business opportunities. To learn more about Alteryx and its technologies, check out www.alteryx.com/try.

Defining Data Blending

Once the right data sources are identified, and the access to those data sources is established, the next step is merging, sorting, joining, and otherwise combining all the useful data into a functional dataset while discarding the vast, loud "noise" of unnecessary data. This process is referred to as *data blending*.

Data blending is defined by Alteryx as "the process used by data analysts to combine data from multiple sources to reveal deeper intelligence that drives better decision making." What does that statement mean? Here's a breakdown:

- ✓ Data blending is a *process* (not just a one-time event), and the process can be repeated as necessary to add or remove more data sources.
- ✓ *Data analysts* in line-of-business departments are empowered by platforms such as Alteryx to perform data blending and don't have to rely on their IT departments.
- ✓ Data comes from *multiple sources* to include one or many sources, but usually many. The sources can be from anywhere to include traditional, enrichment, or emerging data types.
- ✓ The intent is to *reveal deeper intelligence*, which means that greater knowledge is obtained from multiple sources than could be obtained from one source.

- ✓ The ultimate purpose is to *drive better decision making* by senior leaders whose job is to guide the business toward promising opportunities to maximize organizational benefit.

Data blending is essentially using data from multiple sources so that you have the best mix of relevant data for your analytical toolsets. In Figure 2-1, you see how the analyst combines data into a unified, analytical decision-making process.

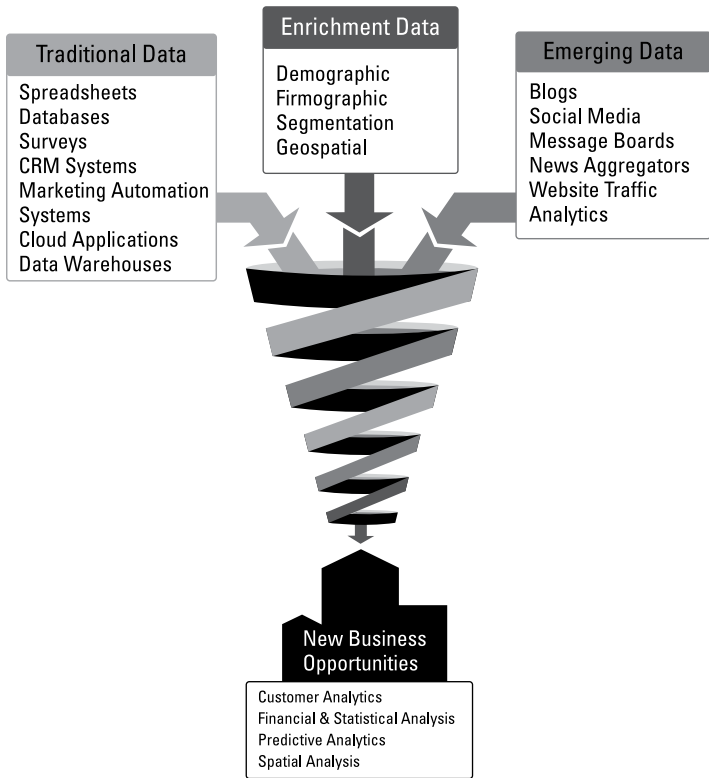


Figure 2-1: Multiple data approaches of data blending.

As Figure 2-1 shows, data blending allows different data sources, which by themselves can be difficult to use, to be combined into a single view that easily generates deep insights into new business opportunities.



Data blending is a complex technological process, but the data analyst should be shielded from the details by the toolset.

Differentiating Data Integration from Data Blending

Data integration is a common term in IT, so a great question would be, “How is data integration different from data blending?” The answer lies in the primary use of that data and how long the data is stored.

Data integration is a process in which multiple data sources are combined (or integrated) to create a single unified version of the data in databases, data warehouses, or data marts. These data stores are permanent entities with cleansed, normalized data and are managed by IT database administrators (DBAs). Access to these new databases is controlled by DBAs and business intelligence (BI) experts. Over time, these databases grow, are used to support applications, and themselves become data sources for other systems and workflows.

Data blending is a process conducted by a business or data analyst (not a DBA or BI expert) to build a dataset for use in analytic processing to answer a specific business question. The data for the dataset is created from one or more data sources. The *blending* occurs as the dataset is built from multiple data sources to capture only the relevant data. Analytical processing occurs on that purpose-built dataset to derive an answer for the question being posed.

The key differentiators are as follows:

- ✔ **Integration** results in a permanent database with the intent of storing a single copy of data and is managed by DBAs and BI experts.
- ✔ **Blending** results in a dataset with the purpose of supporting analysis for a specific business question and is created by business and data analysts.

Both data integration and data blending have a valid place in business, but given the increasing need for delivering better information and results — faster and with less effort and complexity — more and more organizations are turning to data analysts who use Alteryx to deliver these deeper business insights in hours, not weeks, as other analytics processes often require.

Chapter 3

The Building Blocks of Effective Data Blending

.....

In This Chapter

- ▶ Looking at the workings of data analysis
 - ▶ Detailing the phases in the data blending process
 - ▶ Enhancing the capabilities of data analysts and business decision makers
-

Once data analysts have identified and accessed the desired data sources, then their real work begins. The art and science of transforming myriad data into a usable, very specific dataset for use in analytical processing is next. This data blending process is where the data analysts really shine and show their value. However, this process has historically been a difficult and tiresome endeavor.

Fortunately, now that the data blending process is understood, new toolsets exist that greatly simplify and accelerate this difficult process. Tools that support data blending are empowering both the analyst and business decision maker, resulting in increased business agility.

This chapter looks at what the most critical steps are in data blending and how data blending tools help the analyst and ultimately the business.

Identifying the Components of Data Analysis

Analysts follow a structured methodology when performing their analysis. Using modern toolsets, the process of data blending is composed of three fundamental phases:

- ✔ **Access data:** Identify and obtain access to the data.
- ✔ **Prepare and cleanse data:** Reformat data into a usable format and correct or remove invalid data.
- ✔ **Join data:** Combine data for further analysis.

After performing these three phases of data blending, analysts often continue their analysis using an additional two phases for advanced analytics:

- ✔ **Apply advanced analytics:** Execute predictive and spatial analytic processes against the dataset to obtain results.
- ✔ **Output results:** Deliver insight in a readable, often graphical report or output data to feed another process.

Figure 3-1 shows the components and phases of data blending and advanced analytics. You can see that the data analyst performs a process-oriented methodology to connect to data sources, cleanse the data, blend the data into a usable dataset, apply analytics against the dataset, and finally prepare the results for interpretation and presentation.

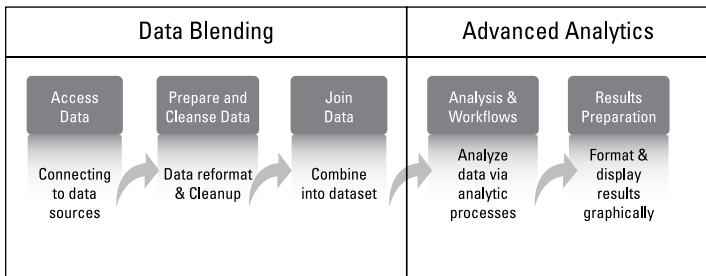


Figure 3-1: The five phases of data blending and advanced analytics.

Moving through Data Blending

In cooking, often the most difficult and time-consuming part is the preparation. Ingredients are gathered, cleaned, and sliced and diced while pots and pans are assembled and ovens and stovetops heat up. Experienced chefs know that the actual cooking is a relatively quick component in comparison to the total preparation process. Good chefs also know that if key parts of the preparation are missed (such as forgetting to add a main ingredient), the overall dish will suffer.

The same concept of preparation holds true for the data analyst; the cruxes of data blending are the data preparation/cleansing and the data joining; these are also the areas where analysts have typically spent most of their time.

Fortunately, just as in cooking where appliances such as electric blenders speed up the food preparation process, data analysts have automated tools such as Alteryx that greatly accelerate and improve the data processes. The next section discusses how Alteryx makes the preparation steps for the analyst much easier and faster.

Data preparation and cleansing

Data analysts are estimated to spend 60 to 80 percent of their time in the preparation and cleansing phase, and when you consider all the detailed work required, you can see why preparation and cleansing is time intensive.

During this phase, the analyst dives deep into the large amount of data provided by the data sources, determines exactly what data is and isn't required, and reviews the data for quality and completeness. The analyst commonly restructures and reformats the incoming data. Almost always, a large amount of extraneous data must be filtered out so that only relevant data remains; this is a particularly important part. Additionally, the analyst must determine how much of the available data to use — in some cases, the full amount of data isn't required and only a statistical sampling size is needed. These key stages are noted here:

- ✔ Restructure, reformat, and recondition data
- ✔ Identify and correct missing or incorrect data

- ✔ Remove or cull out vast amounts of extraneous data
- ✔ Determine full or partial dataset size

Alteryx tools assist the data analyst with these stages by:

- ✔ Providing drag-and-drop tools in the form of icons to manipulate and alter the data formats
- ✔ Using transformations to summarize and aggregate the data
- ✔ Addressing the issues with missing or incomplete data
- ✔ Automatically setting and assigning field types on data
- ✔ Creating and applying formulas to support if/then/else conditions against the data
- ✔ Filtering records that are specific to the data analysis
- ✔ Adding spatial attributes to the analysis as needed

The data cleansing and preparation phase is critical because the analyst must determine what data to use and then clean up the data so that only quality, relevant data is utilized.

Joining data

Joining data is the phase where the prepared data from various data sources is combined together to build the working dataset that will be the subject of the advanced analytics phase. At this point, the data is already prepared so that it is free from errors, is in usable formats, and is relevant for the specific analysis being performed; but the data is still in separate “buckets” from each data source. The analyst must join that data together to create a dataset so that analysis can begin.

Historically, the data scientist or data analyst used the following processes to join data into a workable dataset:

- ✔ Developed manual processes, formulas, and lookups to join spreadsheets
- ✔ Wrote SQL queries to link and join data based on specific conditions and characteristics
- ✔ Leveraged complex, purpose-built data integration tools

- ✓ Implemented simple, custom scripts and processes for small and well-defined data sources
- ✓ Pulled in special, outside tools and processes that included spatial context

These processes were functional, but they were seldom easy and took a long time to use. Fortunately, Alteryx improved the joining phase with tools to allow the analyst to:

- ✓ Join data of any data type
- ✓ Join data at the row level
- ✓ Join data based on multiple fields
- ✓ Implement fuzzy logic matching or nonidentical data
- ✓ Incorporate spatial context without specialized tools

Fuzzy logic matching is matching two datasets based on nonmatching data. Rather than using an exact match, fuzzy logic matches are based on approximate (or fuzzy) values. An example of fuzzy logic matching is creating a customer dataset based on lists composed of sales information and addresses where the names on the lists may not be exact. Fuzzy logic matching identifies close enough values where a listing of John, Jon, Jonny, or Johnathan at one address is likely the same person and would process any of those names as a match.

Alteryx allows the data analyst to join data easily and intuitively as part of an overall workflow for data blending and advanced analytics (see Figure 3-2). Its intuitive, graphical interface contains powerful tools that obscure the complexity from the analyst. You can see in Figure 3-3 how Alteryx supports data blending via an intuitive workflow process.

Automation and repeatability

Once the data analyst has built the workflow, the same analysis often needs to be performed daily, weekly, monthly, quarterly, yearly, or on an ad-hoc basis. So, having the capability to rerun the same analysis without repeating the entire process is a highly desirable option for the organization.

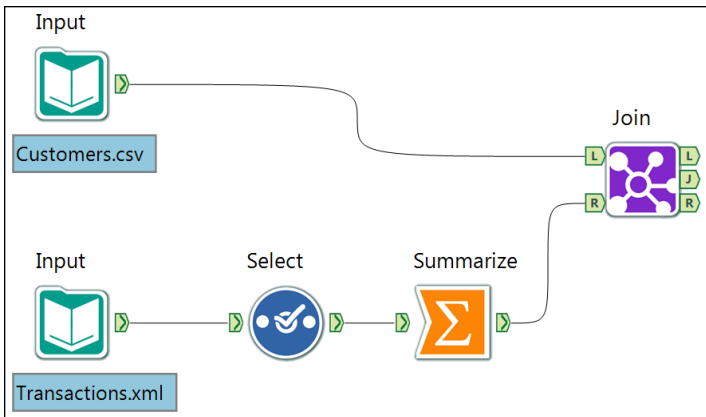


Figure 3-2: Access, prepare, and join data in an Alteryx workflow.

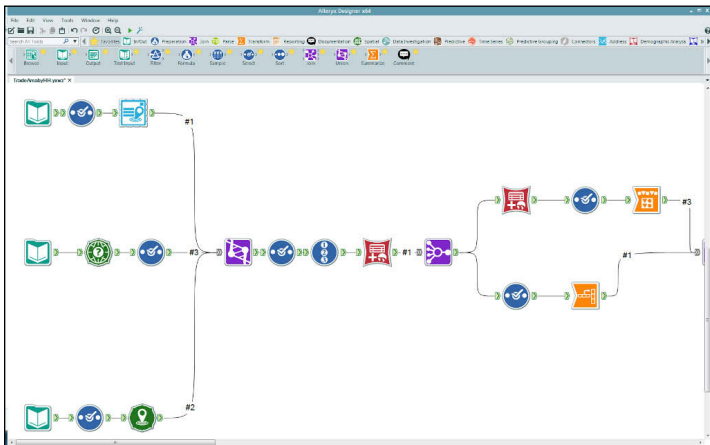


Figure 3-3: Alteryx workflow allows data blending steps to be automated.

Alternatively, the same basic analysis may need to be repeated but with some minor tweaks to account for new business conditions and what-if analysis. For example, the analyst may need to apply a set of predictive calculations for each city in which the company is considering opening a new franchise.

Alteryx supports automation and repeatability within its workflow processes via:

- ✓ Creating macros from analytic steps that are often repeated so they can be added to future workflows
- ✓ Allowing choices to be made by the analyst during an analytic process to support customization
- ✓ Scheduling existing processes to be run automatically with new data and data sources
- ✓ Saving processes into applications (analytic apps) that can be shared within the company or community for reuse via Alteryx Server or the Alteryx Analytics Gallery

Figure 3-4 shows how the Alteryx automated workflow process supports the creation and reuse of analytic apps, graphical visualization tools to assist decision makers, and custom reports that can be repeated on a regular basis.

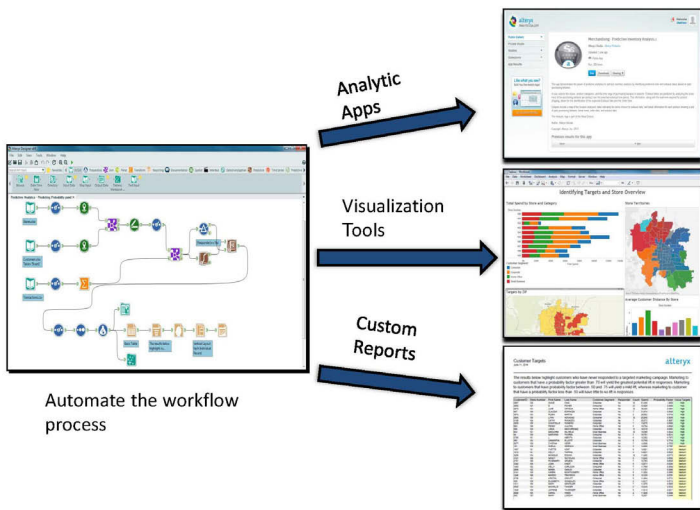


Figure 3-4: Workflow automation and resulting outputs.

Empowering the Analyst and Decision Makers

The greatest benefit of data blending is the capability to put accurate, actionable data in the hands of the analysts and decision makers in a timely manner. Diving deeper into the benefits, you can see that data blending:

- ✔ Puts the most critical data in front of analysts and decision makers at the right times
- ✔ Simplifies the tools used by analysts so that highly trained and expensive technical people are not needed every step of the way
- ✔ Integrates all data types from across all different sources in a rapid, straightforward manner
- ✔ Opens up the access to large amounts of data with minimal complexity to promote a more complete level of analysis with better data and results

Alteryx empowers line-of-business analysts with an intuitive workflow that can easily access, prepare, cleanse, and join all sources of data and deliver critical insight in the fastest time possible.

Enhancing decision making with predictive analytics

Predictive analytics is the process of using complex computational algorithms on a dataset to develop a model of likely behavior in the future. Predictive models and statistics provide a mathematical likelihood of an outcome based on the data in response to a business question.

The business value of predictive analytics is nearly unlimited. Ask a business leader if a crystal ball that could tell the future would be useful, and the answer would be “Yes!” The same concept applies to predictive analytics. Possible business scenarios include the following:

- ✔ What is the likelihood of insurance fraud in this group of insurance claims?

- ✔ What percentage of customers turns to a competitor if you increase the cost of a new product by 5 percent in a specific geographic location?

- ✔ What is the probability of single males in a specific suburban area between the ages of 25 and 30 of purchasing a new sports car in the next year if the car costs less than \$35,000?

Predictive analytics gives forward-thinking business leaders powerful insight into how to best plan for the future and take advantage of opportunities. Savvy data analysts will blend data that supports predictive analytics and use predictive macros against the dataset to identify these opportunities for the business.

Chapter 4

Using Data Blending in the Real World

.....

In This Chapter

- ▶ Understanding how data blending tools help the analyst
 - ▶ Knowing in what business situations data blending makes sense
 - ▶ Observing data blending tools in action
 - ▶ Examining Alteryx's real-world data blending tool
-

By understanding what data blending is and how it works, you can identify its use in real business situations. You know that data blending makes the analyst's job easier, but how does that translate into business value in the real world? To understand this, you need to see how data blending is deployed within businesses.

Examining the situations of where and how data blending is deployed in business helps to highlight its value. Identifying the characteristics of successful data blending tools allows you to accurately realize their benefit and select the right tool when the time comes to equip your analysts.

This chapter looks at how data blending is implemented in the real world and how it will enable your analysts to have a greater business impact.

Maximizing Business Agility with Better Tools

What is the purpose of better, faster tools in the workplace? Sure, they make people's jobs easier, but what is the underlying business reason for doing so? The answer is that by enabling employees to be more effective, your business becomes faster, smarter, and better able to take advantage of changing market conditions and new opportunities.

Success is closely tied to agility and the ability to read changing conditions faster than your competitors. Tools that your business uses should support the tenets of increased agility and greater market intelligence. By the same token, tools and processes that are slow, antiquated, or otherwise don't support business agility or market intelligence should be replaced.

In the realm of the data analyst, what are the key characteristics of tools that support business agility and market intelligence? The tools of the data analyst should support intuitive workflows, deeper business insight, and faster access to information.

Intuitive workflows

Data analysts are both creative and logical people; their tools should work in harmony with their thought processes. Using intuitive workflows that flow with the natural processes of the data analyst ensures that analysts' work and creativity are not disrupted. Essentially, you want tools that work *with* your analysts, not *against* them.

Workflows provide the benefit of documenting what is occurring during the data preparation; this functionality makes the job of the data analyst easier and more efficient. Unlike cumbersome spreadsheets that often lead to confusion and don't support retracing the steps through a process, analytic tools such as Alteryx give analysts the ability to easily understand processes using graphical workflows (see Figure 4-1).

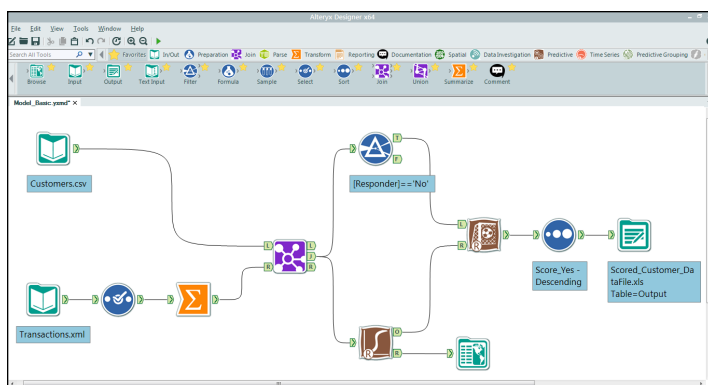


Figure 4-1: Alteryx's intuitive workflow for data blending and advanced analytics.

Deeper business insight

Answers to difficult business questions are out there, buried somewhere in the data; the challenge lies in whether the analyst can dig deep enough into the data to find those answers. Tools that bring the power of advanced analytical functions to the analyst and that can interrogate data from multiple sources are likely to find the answers sought by the analyst. Simple tools with limited capability will not suffice; tools that perform complex tasks in a simplified manner are what you need.

Faster access to information

In the military intelligence community, actionable data has a short shelf life before it becomes irrelevant; the same concept applies in business. Tools that allow the data analyst to identify market trends and new business opportunities before the competition identifies them have real value for the business. On the other hand, if a tool is so complex, large, or cumbersome that any answers it provides are yesterday's news, that tool (and its information) has little value.

The secrets to business success really aren't secrets, but companies often lose sight of what's necessary to be successful. The same pitfall holds true for data analysts and their tools;

analysts' tools must support business agility and market intelligence in a timely manner with the depth necessary to have value.

Implementing Data Blending in Business

Nearly all business departments can benefit from some degree of data blending, but some areas will see greater benefits than others. Generally, situations where the ability to search through larger amounts of data from different data sources to identify trends, patterns, and relationships will be the best environments to use data blending.

Here are some areas where data blending is commonly used:

- ✔ Sales and Marketing
- ✔ Finance Operations
- ✔ Site and Merchandising Operations

The key to knowing where to deploy data blending is in understanding where competitive advantage can be gained by accessing large amounts of data from different data sources. If, for their analysis, analysts need to access and incorporate a mix of data from traditional sources within the company, cloud and social media sources, Big Data, and external third-party data sources, odds are that data blending is the right solution.

Putting the Right Tools in the Hands of the Data Analyst

When selecting tools for the data analyst, you should develop a set of key criteria. Desired capabilities and features based on business requirements create the evaluation yardstick against which tools will be judged. You want to establish a set of evaluation criteria to:

- ✔ Ensure that all the essential capabilities are present in any tool selected and that no required capabilities are omitted

- ✔ Separate nice-to-have features from core requirements to reduce the risk of over-purchasing unnecessary components
- ✔ Reduce the likelihood of being swayed solely on a sales presentation or marketing materials
- ✔ Build a supportable justification for selecting a specific product that can withstand the scrutiny of management and contracting oversight

As you determine your set of core requirements that must be supported by a data analysis tool, make sure that the tool includes these capabilities:

- ✔ Easy-to-use User Interface (UI), intuitive design, and robust support and help documentation. The tool must be readily usable by the analysts; otherwise, it will be discarded or not fully implemented.
- ✔ Able to access all the relevant data regardless of its location or type: databases, local files, third-party data sources, Big Data, social media, and legacy systems.
- ✔ Data preparation, data cleansing, and data joining functions within the tool to build usable datasets with accurate, appropriately sized data; 60 to 80 percent of analysts' time is spent on the data preparation and joining steps, so tools to reduce that workload are important.
- ✔ Predefined tools for predictive analytics functions and spatial capabilities. Data analysts do not have time and training to write code to create these complex tools from scratch, so leveraging prebuilt functions is a must. The vendor should provide regular updates to these tools and functions to increase capabilities.
- ✔ Workflow processing and drag-and-drop capabilities to perform the tasks within the tool intuitively. Data analysts work in a creative and logical manner, so workflows that support those thought processes will increase the analysts' productivity.

Selecting the toolset to be used by data analysts is an important task, but it doesn't have to be overly difficult. Identify and document the core business-specific features required by the analysts and use the preceding list of capabilities to guide your product evaluation and selection process.



If the opportunity exists, be sure to take advantage of the software trial periods that many companies, including Alteryx, offer prospective customers.

Using Alteryx Designer

Alteryx Designer is the Alteryx flagship tool for data blending and advanced analytics. The Alteryx Designer tool is built for the needs of today's data analyst. The key capabilities of Alteryx Designer are categorized as:

- ✔ **Data blending:** Accessing and blending data from all the relevant sources. Easy-to-use workflows assist the analyst in data preparation and cleansing to reduce the time spent in preparation phases.
- ✔ **Predictive analytics:** Leveraging more than 40+ prebuilt analytical tools coupled with the capability to create custom predictive tools based on open-source R, a powerful predictive analytics language widely used throughout the world. Analysts are able to perform predictive analytics without the requirement to be a highly trained data scientist or programmer.
- ✔ **Spatial analytics:** Blending location data with advanced spatial analytics give the analyst insight into location-specific functions to better understand customers and answer business questions. Analysts use intuitive workflows to uncover geographic relationships and information that otherwise would be lost in a sea of data.
- ✔ **Sharing insights:** Displaying in an understandable, insightful format the analytical data results so that business decision makers see relationships and discover new opportunities. Analysts can easily and quickly send results into friendly output formats and tools such as automated custom reports, graphical displays, Microsoft Excel, XML, PDFs, Tableau and QlikView visualization files, and more.

The features supported by Alteryx Designer are rich and engineered to support the analyst every step of the way during the data analysis process.

Alteryx Designer provides easy-to-use workflows that enable the analyst to be self-sufficient and to accomplish more work in less time. Figure 4-2 shows you the Alteryx Designer capability to use a workflow to blend data.

What you see in Figure 4-2 is the Alteryx Designer workflow to blend data from Salesforce and Marketo into a dataset. In this case, the analyst is gathering data from a cloud CRM provider (Salesforce) and a marketing automation system (Marketo) in support of a sales and marketing campaign.

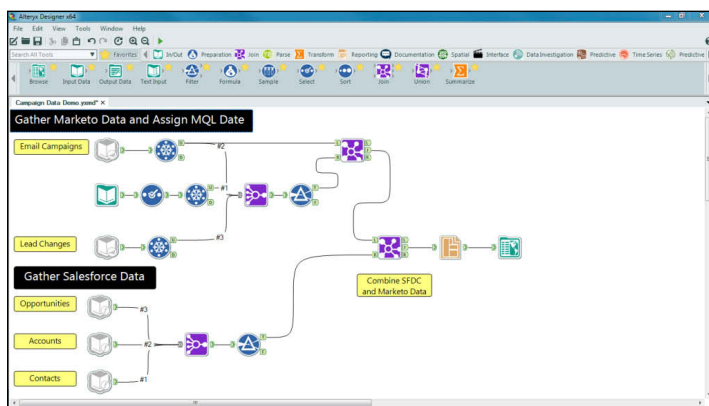


Figure 4-2: Easily blend Marketo data and Salesforce.com data using an Alteryx Designer workflow.

Once the analytical processing is completed, many different output presentation formations are available to the data analyst. As shown in Figure 4-3, Alteryx can also create the native file formats used by Tableau and QlikView, which are especially effective with senior business decision makers to perform data discovery.

In Figure 4-3, you see output in the form of a Tableau workbook displaying sales results by categories for stores in a city. The visualization aspect allows a clear understanding of sales data per store based on customer segment. Additionally, a spatial aspect is present with data relating to the customers' distance from their stores, zip codes, and territory assignment data.

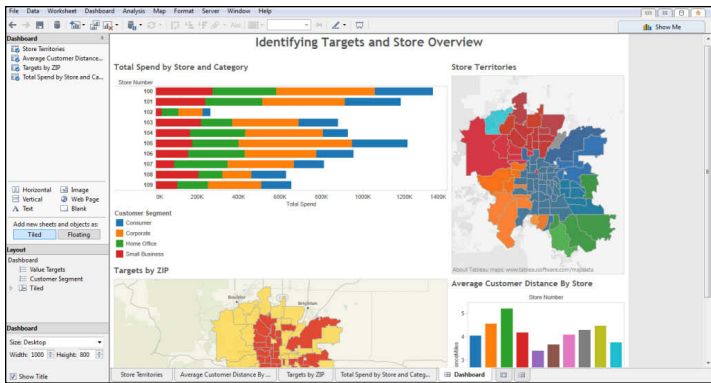


Figure 4-3: Output directly to a visualization tool. Alteryx supports direct output to Tableau and QlikView.

Alteryx Designer empowers the data analyst to be self-reliant to easily and rapidly perform data analysis against all the relevant data. Using intuitive workflows, prebuilt macros, and analytical tools, the analyst answers meaningful business questions and puts the results into many formats, including easy-to-understand visualization tools.

More precise than Excel

Microsoft Excel is a great tool, but the wise data analyst knows when to put Excel away and use a tool more suited to the task at hand. Platforms such as Alteryx match the simplicity of Excel but exceed its capabilities via:

- Drag-and-drop interface to easily and quickly create analytic workflows
- Rapid access to data from all data sources, including complex data sources from the cloud, Big Data, and social media

- Graphical tools that cleanse, prepare, and blend data from multiple sources without requiring IT involvement
- Leveraged automation, prebuilt macros, and spatial capabilities

More than 200,000 data analysts currently use Alteryx to outperform the basic data blending capabilities of Excel and use the same intuitive workflow for subsequent data analysis and reporting.

Case study: Data blending in business

DatabaseUSA is a leading provider of sales leads, business and personal mailing lists, email lists, and other related products. DatabaseUSA customers are the sales and marketing departments of large companies that need accurate, reliable sales and marketing data. In the customer data business, data accuracy and reliability is everything, and DatabaseUSA has built its reputation on data quality and depth.

The challenge faced by DatabaseUSA was building the high-quality, deep data their customers needed when faced with so much extraneous data. Somehow, DatabaseUSA had to sort through the data of 238 million people, 14 million businesses, and 164,000 households to build quality datasets for its customers. Reducing the time, effort, and cost to build the datasets were additional requirements.

Existing DatabaseUSA data blending methods were centered on custom scripting languages on their internal hosted databases. Using highly skilled IT data specialists, DatabaseUSA was able to blend data, but the process was slow, complex, and intensive.

With millions of rows of raw data being added monthly, the IT staff was struggling to keep up with the increased workload of bringing in so much data, cleansing and checking it for quality, and blending it into usable datasets. DatabaseUSA opted to

keep its internally hosted databases (housing its data) but set out to find a tool to better perform the data blending and ETL functions it required.

Already an Alteryx user, DatabaseUSA began to leverage the capabilities within Alteryx to:

- ✔ Input and update data while removing duplicate data
- ✔ Blend in data from all its data sources
- ✔ Clean and prepare data and test the data for accuracy
- ✔ Standardize data format, values, and fields

In particular, DatabaseUSA liked how Alteryx could standardize different data types and remove duplicate data prior to loading into the company's internal databases. The data integration capabilities of Alteryx ensured that only correct data in standard formats was being loaded. Additionally, because of data privacy and security concerns, Alteryx was deployed to the DatabaseUSA secure hosted server environment.

Alteryx helped DatabaseUSA achieve its requirement of loading clean, accurate, and standardized data into its databases on a regular basis amid many possible large external data sources. Manual, low-level data programming was replaced by Alteryx fuzzy logic matching algorithms

(continued)

(continued)

resulting in less workload but faster, more repeatable processes; database builds are now completed in a week instead of a month. Finally, reduced dependency on expensive custom IT support is driving down support costs while lowering the risk of IT dependency.

DatabaseUSA had a data blending problem with too much varied data and not enough time and resources to process it. Alteryx was able to blend that data using Designer features to solve DatabaseUSA's problem while reducing costs and improving data quality and processing speed.

Chapter 5

Ten Things to Consider with Data Blending

.....

In This Chapter

- ▶ Combining analytics with data blending for better decision making
 - ▶ Unlocking the power of data faster using macros and prebuilt analytic apps
 - ▶ Deploying spatial and analytical applications to gain real business value
-

Understanding the fundamentals of both a technology and a toolset is key to sustained success. But, once you have a working knowledge of the fundamentals, understanding the tips, tricks, and special features of the toolset acts as a success accelerator to allow you to further maximize benefits and accomplish your goals.

This chapter takes a look at some specific ways to optimize data blending, combined with additional analytical tools and techniques. The chapter also extends the discussion of Alteryx Designer (see Chapter 4) to examine how Alteryx helps data analysts and what it takes to get started using it.

Integrate Predictive Analytics into Your Business

Predictive analytics enables companies to apply statistical analysis functions against datasets to develop probabilities of a future event or change. In effect, predictive analytics allows companies to forecast the future. Used properly, predictive analytics identifies new opportunities for companies and helps reduce existing expenses.

In addition to the robust tools for data blending, Alteryx contains prebuilt predictive analytics macros and functions. Building the right datasets with relevant data blended in benefits predictive analytics by generating more accurate predictive models. Data blending also reduces the time spent in the data preparation stage in the predictive analytics cycle.

Data analysts should be encouraged to unleash their creativity to experiment with the predictive analytical functions in their daily jobs to help unleash this powerful capability for the business. Beyond the raw technical capabilities of a toolset, analysts must adapt their mindset to look to predictive analytics as part of their job.

Gain Insights Rapidly

A challenge with many technologies is the lengthy time it takes to start using a new product or technology once it is brought into the IT shop. Depending on the complexity of the technology and the existing knowledge and workload of the IT staff, it can take weeks or months before a new technology is in use. Furthermore, before a technology really starts providing true business value and making a positive return on investment (ROI), it can take many months or even years.

Many Alteryx clients are pleasantly surprised to find out just how fast they can start using Alteryx Designer to gain insights into their business. Easy-to-understand and easy-to-use graphical workflows and prebuilt macros allow clients to perform data blending and advanced analytics to gain business insight more quickly. In many cases, Alteryx customers report that they have found positive business insights within the first week after installation. Having a positive initial experience with a product greatly enhances the likelihood that a company and its users will embrace it.

Augment the Tools You Now Use

“Why reinvent the wheel?” is a cliché that has merit, especially in IT and technology-related fields. For decades, programmers have touted code reuse as a benefit of many programming languages and code repositories — and for good reason.

Incorporating into your code a tested and proven program function that someone else has written *isn't* lazy; it's a wise use of time, on your part and your company's.

The same concept of code reuse applies to data analysts. When it makes sense, continue to use tools such as Microsoft Excel but use Alteryx to augment the processes that require more powerful tools. Alteryx is very good at all three phases of data blending — data access, data preparation, and data joining — as well as advanced analytics, areas in which Excel is not as strong. So, using a tool such as Alteryx to augment Excel makes good sense.

Work behind the Scenes

Alteryx tools are designed with prebuilt macros and functions to make the analyst's job easier and more productive, but an entry point for developers to write more complex code exists. Developers can leverage the Alteryx Application Programming Interfaces (APIs) and Alteryx Software Development Kit (SDK) to write their own custom tools within Alteryx workflows.

Developers can write their own code and leverage .NET and C++ APIs to make calls to the Alteryx Engine. Using the SDK, IT developers are empowered to write their own custom code in C and C++ to further enhance Alteryx to support their company-specific requirements. Writing custom code to assist with data blending further enhances Alteryx's support of the analyst during the time-consuming data access, preparation, and cleansing stages. For example, you could write a separate application to leverage and incorporate the reports generated from Alteryx workflows.

Consider Spatial Analysis

Spatial analysis is the value-added element to business equations that can make a big difference. Being able to target a business question to a specific geographic location instead of across a broad category allows business decision makers to make key decisions to open new stores or launch campaigns in areas with favorable conditions. Spatial analysis makes information actionable, providing information to decision makers that they can act upon. Data analysts have long

wanted spatial analysis capabilities, but they did not want to code custom spatial features themselves or to enlist IT to do the programming. In other cases, data analysts simply did not have the spatial data available even if they did have access to spatially aware analytics.

Alteryx Designer brings spatial analysis to the data analyst in two ways. First, Alteryx enables easy access to external data sources such as Big Data, social media, and mobile devices that provide spatial data. Using data blending features in Alteryx Designer, the analyst blends the spatial elements needed into the larger dataset used for analysis. Second, Alteryx provides prebuilt macros that support spatial analysis so that the data analyst doesn't need to handcraft complex spatial functions.

Leverage Analytic Apps

One of the great benefits of the cloud is the capability to access the right application to perform the job you need, when you need it, and to do so quickly and securely. Using the Alteryx Analytics Gallery of apps, your data analyst has access to the Alteryx cloud of prebuilt analytical applications (apps). The analyst browses the Gallery in the cloud to find an app that does the job the analyst needs, and then the analyst simply downloads and executes that app. Because the app is downloaded to the analyst's workstation, no data is passed into the cloud, so the analyst knows that the process is secure. Analysts have the capability to customize or tweak the app as needed for the specific task at hand. Plus, analysts can even write their own apps to upload (publish) to the Alteryx Analytics Gallery to be securely shared within their company or industry.

You may also choose to deploy Alteryx Server as an effective way to share analytic apps with your employees. In addition to offering the data processing power of a server-based analytic architecture, Alteryx Server gives everyone in your organization the power and flexibility to run their own analytic apps and customize the output to their specific needs while keeping all components of the analytic process behind your corporate firewall.



Leveraging analytic apps is another smart way to avoid reinventing the wheel. Analysts can build on the efforts of others in the form of analytic apps so they can better focus on their data analysis duties.

Incorporate Visualization

Many analytical tools perform number crunching well, but the results are output as raw data or in a format that is difficult for nontechnical people to understand. Business decision makers are often visual people so the capability to display output in a graphical format is important. In the past, data analysts have imported that data into presentation tools such as Microsoft Excel to render more user-friendly graphics. This process works, but the time has come to evolve past this tedious process.

Alteryx Designer supports generating output into a wide variety of formats, including Excel, Extensible Markup Language (XML), and Portable Document Format (PDF) files. Furthermore, the data analyst can generate output to Tableau and QlikView visualization products.

Visualization products allow data to be more easily understood and relationships to be identified and explored, thus allowing business decision makers to more quickly identify new opportunities. However, visualization tools are not as strong with advanced data blending and preparation, and they lack predictive and spatial analytics capabilities. Therefore, using a combination of Alteryx to perform the data blending and advanced analytics with the output displayed within visualization tools leverages the individual strengths of the tools to deliver a successful combined outcome.

Improve Business Analyst and IT Relations

Data blending can change the relationship for the better between business analysts and IT. Rather than relying on IT for all the data, the analysts are more empowered to access more data. However, with that empowerment comes a greater collaboration between IT and analysts in support of data governance. Both parties must work to ensure that the data of the company and that of the customers is used in compliance with the data governance standards appropriate to their environment.

Drive Down IT Support Costs

It's no secret that hardware, software, and support infrastructures are complex and expensive. Additionally, tools supporting deep data analysis, reporting, and business intelligence (BI) are themselves highly specialized and expensive.

Compounding the cost of IT hardware and software is the cost of large staffs of highly trained people to manage IT. Often the IT staff is so busy simply supporting the existing infrastructure that they can't help the business and data analysts use the tools they have.

Alteryx's direction toward reducing IT costs focuses on empowering end users to be more self-reliant, moving away from custom and complex coding, automation and reuse of workflows, and analytic apps published to Alteryx Server or the Alteryx Analytics Gallery. These Alteryx features reduce the amount of IT support required (lowering personnel costs) and allow consolidation and elimination of pricy data analysis, reporting, and BI tools.

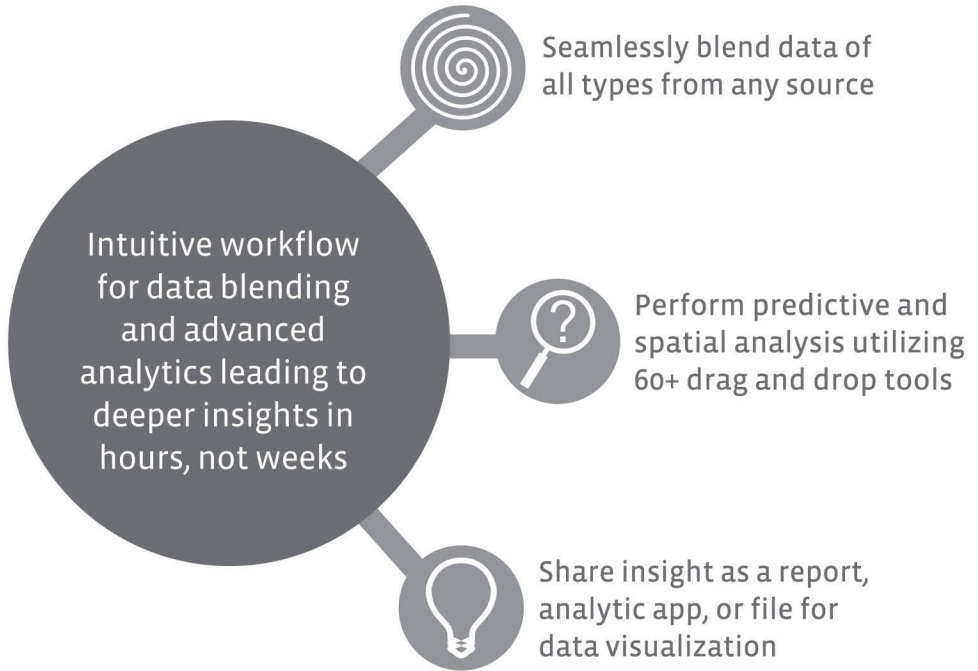
Beyond reducing IT costs, data blending also frees up IT to work on other important projects instead of directly supporting data analysts. Reducing the requirement to support data analysts allows highly skilled IT folks to work on the more complex technical tasks that the company needs.

Feed the Downward Stream

The most common output of data blending is either a report or a visualization; after all, odds are that the analyst is trying to support a key decision with this data. However, a data-blending effort may not feed directly into a processing cycle followed by a formatted report. Instead, data blending may result in a dataset that will be used as part of a larger analytic process with multiple steps, sub-steps, and processes.

Data blending can conceivably generate input to an existing predictive model or to data that is written to a data warehouse so that every data analyst has access to the same blended data. Thinking outside the box to consider how data blending can serve larger processes in addition to its more traditional roles leads the analyst to be more creative and more effective.

Get a Free Trial of Alteryx Designer



Over 600 customers and thousands of data analysts worldwide rely on Alteryx. Now, try it with your data.

Download Alteryx today: www.alteryx.com/try

Blend data to gain deeper business insights in hours, not weeks

As a data analyst, you know that one of the biggest challenges you face is bringing together the data you need for analytics to gain insight from files on your hard drive, the cloud, Big Data, social media, databases, and all the other areas of data storage. This book shows you how data blending is used to rapidly build the data set you need to meet your organization's analytical needs, without writing scripts or waiting on other departments.

- *Understand data blending* — learn how data blending works and why it is important for business advantages
- *Use data blending* — see how data blending helps your analysts dive deep into data to find key information and identify trends
- *Create powerful workflows* — use intuitive, graphical tools to blend data and perform critical business analysis faster than ever



Open the book and find:

- Why data blending matters to business
- How to gain deeper insight and better understanding amid a vast universe of data
- Tools to put the power of data blending in the hands of the analysts for faster analysis
- Which areas of the business see the greatest jump in agility thanks to data blending tools
- How to better support your business decision makers

Go to Dummies.com^{*}
for videos, step-by-step examples,
how-to articles or to shop!



WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley's ebook EULA.