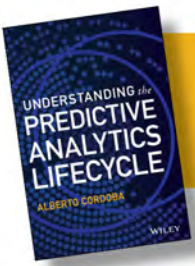




UNDERSTANDING *the*  
**PREDICTIVE  
ANALYTICS  
LIFECYCLE**

**ALBERTO CORDOBA**

**WILEY**



From *Understanding the Predictive Analytics Lifecycle*.  
Full book available for purchase [here](#).

---

# Contents

**Foreword xi**

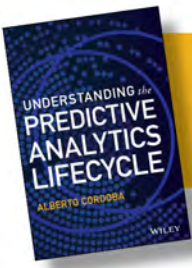
**Preface xiii**

**Acknowledgments xv**

Chapter 1	<b>Problem Identification and Definition</b>	<b>1</b>
	Importance of Clear Business Objectives	4
	Office Politics	8
	Note	13
Chapter 2	<b>Design and Build</b>	<b>15</b>
	Managing Phase	16
	Planning Phase	18
	Delivery Phase	19
	Notes	32
Chapter 3	<b>Data Acquisition</b>	<b>33</b>
	Data: The Fuel for Analytics	36
	A Data Scientist's Job	41
	Notes	53
Chapter 4	<b>Exploration and Reporting</b>	<b>55</b>
	Visualization	57
	Cloud Reporting	61
Chapter 5	<b>Modeling</b>	<b>69</b>
	Churn Model	71
	Risk Scoring Model	77
	Notes	99
Chapter 6	<b>Actionable Analytics</b>	<b>101</b>
	Digital Asset Management	104
	Social Media	104

Chapter 7 Feedback	129
What the Different Software Components Should Do	132
Note	148
<b>Conclusion</b>	<b>149</b>
<b>Appendix: Useful Questions</b>	<b>155</b>
<b>Bibliography</b>	<b>209</b>
<b>About the Author</b>	<b>211</b>
<b>Index</b>	<b>213</b>

From *Understanding the Predictive Analytics Lifecycle* by Alberto Cordoba. Copyright © 2014, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.

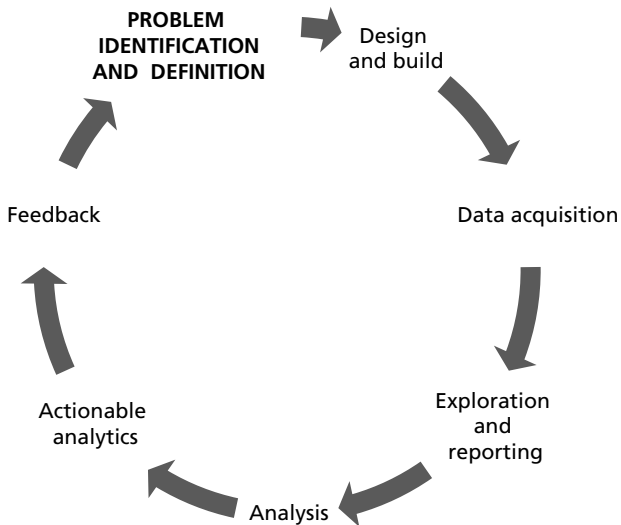


From *Understanding the Predictive Analytics Lifecycle*.  
Full book available for purchase [here](#).

# CHAPTER 1

## Problem Identification and Definition

*How executives focus resources and assess an organization's readiness for meeting the challenges posed by new business realities*



Recently I met with a pair of business executives at the Gaylord Convention Center near Washington, DC.

Two analysts glided their way toward me. I smiled and went in for handshakes, exclaiming “Hello there!” Their names were Zizi and Javier. Both worked for a big corporation right outside of the Beltway in Maryland. I quickly launched into a flurry of business jargon, briskly walking toward the coffee kiosk, mouth running at a hundred million miles per minute. The executives shuffled after me, saying “We are very interested in finding out more about developing a modern analytical system.”

I bought a soy latte with an extra espresso shot. As the caffeine kicked in, I started by asking, “What is your firm’s level of analytical maturity?”

Javier looked at me and said, “Before we get started, do we have an NDA in place?” A nondisclosure agreement is a document signed to protect both parties. (A sample agreement is presented at the end of this chapter.) “We sure do,” I answered. “Great! So let’s continue.”

Javier stammered, “I-I don’t know. I believe that analysis is a portion of the transformation cycle from data to knowledge to wisdom. So, probably the analytical maturity of an enterprise would tell how well it can leverage analysis and close the information gap. I am not sure where I would say our company is exactly.”

My eyes met his as I popped a huge sparkly smile. “Everybody knows the four key levels of an analytical framework are. . . .”

I waited for a response. Zizi replied, “Infrastructure, functionality, organization, and business, and these levels can be translated into an information evolution model for analytical applications.”<sup>1</sup>

Javier piped up, “What is the importance of this?”

I answered, “Those organizations that try simply to define and implement an advanced analytical solution in one step may end up taking far too long to finish building it and reap its benefits.”

Zizi lowered her glasses and continued my thought seamlessly. “And then, most likely, the analytical solution delivered will not meet needs because requirements usually change after an initiative is initiated or because the technology has already changed. We’ve been through that before.”

“Exactly!” I added, “There is an overarching need to build flexibility into contemporary analytical systems. Particularly now that data are growing exponentially and we are faced with big data everywhere. I believe enterprises need to assess the overall maturity of their analytical initiative and aim to add value incrementally rather than use an all-at-once approach. This is very important with the big data challenges. Results and challenges differ depending on the level of analytical maturity. I think the assessment of needs for an analytic platform or workbench should include choosing an appropriate software architecture for analysis and reporting, a hardware environment, a big data integration approach, and, of course, a data model for their structured data, among other things.”

They wondered, “Is that enough to ascertain success?”

I told it to them straight. “Hey, it’s anybody’s guess, but it increases the probability of success significantly!”

Results usually are measured in terms of effective usage of information technology (IT) investments and improved operational efficiency. Challenges primarily occur with IT infrastructure, culture, software technology, and functionality.

They looked at each other warily. I tried to reassure them a little bit. “Improved results usually are associated initially with having one version of analysis-derived information, the so-called truth, which improves the management of multiple departments. Some of the organizational challenges begin to take more focus and skills from the project team. Good results are associated with improved and faster decision-making processes than the competition.”

I decided not to mention the challenges that often occur at the business level, such as shifting business processes and methodologies to leverage new analytical capabilities for corporate performance management. Or changing business goals or objectives, based on insight gained. They were too apprehensive. Therefore, I wanted to stick with the most basic and positive aspects of reworking their business objectives.

I continued, “As your consultant, I have to ask you: Where is your firm going? Is the gut feeling still driving decision making? A successful analytical initiative needs good strategic business objectives.”

They winced at that statement. They knew I was right. Javier shook his head back and forth and sighed.

## IMPORTANCE OF CLEAR BUSINESS OBJECTIVES

I patted them on the backs. “Business objectives must drive analytical initiatives and investments. The success of an analytical initiative should be measured by how it affects strategic and operational business objectives—not how many rows of big structured and unstructured data can be loaded into a data framework in six hours or the complexity of a model developed. This is particularly true when we consider the vast amounts of data that most organizations have accumulated and that continue growing.”

Obviously, the lack of clearly defined business objectives would make assessing the success or value impact of an analytical initiative impossible.

“Do you think that you can use an analytical framework to align IT system initiatives with business objectives and make strategic choices?” I asked the executives. “It could be the best thing that ever happened to you.”

I recommended that they conduct a business value evaluation prior to investing in an analytical initiative. “This evaluation will provide a quick and low-cost validation of an analytical project’s proposed direction and deliverables. This evaluation will also bring focus and attention to an analytical initiative within your IT organization and the potential business stakeholders. It can also pinpoint weaknesses and threats to the future project.”

Zizi asked, “Do you think the evaluation will start a dialogue between the internal groups like the IT organization and business users that will identify the business objectives?”

I smiled and said, “Yes. It will identify how analysis can contribute to the success of business objectives. It will set the scope and size of the project and determine the appropriate investment levels. This is just the beginning. Even if your analytical initiative is already under way, I think that if you take a step back to assess the initiative, you may discover new areas of additional leverage or new risks.”

I continued to urge them to face the bitter truths of today's analytic realities. "It is important to ask questions to better understand which other business opportunities and objectives should be addressed and funded within the analytical initiative. It may be just as important to identify areas to keep outside of the project or that should come along in a second wave."

I took out my phone, glancing at the time. "Why is your organization embarking in analytical applications and big data insight anyway?"

Zizi said, "Al, we need to stay competitive, and this is really exciting. We also have a new executive team with the right approach to data and what we can do with it. Just think where this could lead with your help."

I appreciated that comment. "Thank you. I think we are on the right path. Business value in the real world can be achieved only when you leverage data that are relevant, accurate, timely, consistent, and, most of all, accessible. Most organizations that I have worked with start an advanced analytical project in an effort to drive revenue, increase profitability, optimize certain processes, decrease cost, make better decisions, manage the objectives, minimize risk, and/or improve infrastructure functionality. Does any one of those goals sound like your objectives?" They eagerly nodded.

I continued, "It sounds like you are planning to use analytical applications to gain a competitive edge in a highly competitive market. If so, are your specific business objectives well articulated? Do you already have your performance measures defined? How well and often are your key performance indicators (KPIs) measured and analyzed? Do the appropriate internal and external users have access to relevant data and analysis? Can you look at the KPIs and easily drill down for additional data? What would be the impact of new insights derived from increased or improved data access or analysis? What would be the impact of more real-time data and/or advanced predictive analysis? Is executive sponsorship and funding available?"

Their heads were spinning so I recommended an easy first step. "Analyze the strategic and tactical business objectives that will drive this analytical initiative and its funding. These objectives ultimately will define your project success."



We looked at each other across the table in the atrium of Gaylord's. An hour had gone by, and they confessed they were nervous. I could not blame them. The assignment to develop an analytical application initiative first requires a readiness test.

As obvious as it seems, an assessment of the IT organization and business user skills, levels of analytical activity, and culture will help the enterprise determine the probability of an analytical initiative's success—before it makes any significant investments.

I flashed them a million-dollar grin and encouraged them to feel excited about the upcoming changes. “Before embarking on a big data acquisition adventure and its complementary analytical initiative, an enterprise like yours should complete a self-assessment to determine readiness. You must honestly evaluate your available skills, existing processes, and levels of analytical capabilities and culture so that, before spending considerable sums of money, you understand the challenges ahead and have a way to determine how to proceed and the likelihood of success.”

Zizi raised her hand. “To assess potential for analytical success, should we rate the level of engagement on analytics demonstrated by both our IT department and our business user community?”

I quickly replied, “Yes! First, you should rate the degree to which the following statements apply to your *technical* organization: Does IT understand the need for and potential of analytical applications? Does IT have the required skills and resources to support an analytical environment? Is IT taking responsibility for setting up an analytical infrastructure? Does IT act as a catalyst for technical improvements in the enterprise? Is IT respected within the enterprise? Does IT have a history of success?”

Javier lifted an eyebrow. “What about the business side? In your experience, typically, do business users understand the need for and potential of analytical applications? Do they have a history of funding and championing analytics initiatives? Are the business users the ones to drive IT to deploy new technology? Do they seek an active partnership from the IT organization? Should the business user community participate in the technology selection and adoption process?”

I responded, “Well, you are going to have to do the detail work of answering all of those great questions. I have seen a bunch of different combinations. However, the importance of a readiness assessment is

undeniably clear. Any successful enterprise needs a portfolio of analytical applications to address the needs of a broad range of user requirements. But before it can develop that portfolio, the enterprise must determine what appropriate technical infrastructure and development methodologies are already in place, including: a platform to source data (e.g., a data mart, data warehouse, operational data store, multidimensional cubes, massive parallel processing (MPP) databases, big data framework), available data models and business definitions, rules for metadata use and integration, support for real-time use, access to cloud computing resources, and, when appropriate, methodologies for development, deployment, and change management. In addition, software for data management, data exploration, advanced analytics, and campaign management are also typically required.”

They looked a little flustered with my tech talk, but I wanted to cover a few more points before lunch so I quickly continued, “Initially, you should make sure that functionality is sufficient to ensure that an analytical initiative could deliver value. Later on, among many other tasks, you or your consultant team will define user requirements, decide whether to build or buy analytic applications, determine enterprise security and user access levels, assess scalability, and ask your IT counterparts to establish standards that match user types to appropriate tools.”

Zizi asked, “Shouldn’t the data be quality data to ensure that the analytical initiative delivers the expected value?”

I said, “You know the concept ‘garbage in, garbage out’? Definitely, data governance and data consistency are a high priority. For example, you should inventory data sources and means of access, identify data stewards, identify data quality solutions, and define methods to extract and transform data efficiently and correctly.”

I paused to think. “Be aware of timing. Many new infrastructure and functionality requirements are identified approximately six months after the initial analytical deployment. This makes an effective implementation methodology critical to ensure all the respective resources and skills are available throughout the system development life cycle to address those new requirements.”

Javier asked, “Will this technology assessment help validate technical and cost assumptions?”

I clasped my hands together and nodded slowly. “It will identify whether any critical factors were overlooked. It will spot potential weaknesses in the implementation of a plan.”

Zizi gave me a hard, discerning look. “What advanced analytical functionality does our company need, and what is the difference between that functionality and the kind of functionality we are using today?”

I beamed at her. “The analytical function can be seen in four main areas: integrate, report, model, and enable. First, “integrate” refers to the ability to collect and organize diverse data and make it ready and accessible for advanced analytical applications. It includes structured data like that generated in operational systems. Typically, it comes from database management systems. It also increasingly includes what is called nonstructured data coming from Web records and social networks and typically is very big. Today this data integration area is called *information management*. Second, we see an area for data *exploration, visualization, and reporting*. Third, “model” refers to the actual *advanced quantitative modeling* that takes advantage of statistical or mathematical techniques to gather information out of the data. The fourth area is *execution*. The predictive analytics function is an enabler of other applications like customer service, financial intelligence, or marketing services by improving the communications efficiency. I like your question. It is looking toward the future of analytics at your company. I see we are making progress, and I am becoming more confident about your company’s potential for success.”

As you can gather from the previous ideas, it is important to keep in mind that most organizations can derive great benefits when they provide these four functionalities using software as part of an integrated system within the context of an analytical framework. Most traditional analytic software platforms provide extraction, transformation languages, SQL generation, standard reporting, visualization, what-if analysis, alerts, corporate dashboards, statistics, data mining, advanced analysis and forecasting, campaign management, and optimization.

## OFFICE POLITICS

Javier scribbled a few notes on his iPad. “Considering the types of insight required and the interaction with different types of users, how

will we determine what functionality we need from the software we choose? There are so many choices.”

I agreed. “It *is* very confusing. You need to use a methodology to sort through all the vendors and tools. It is critical to have a clear objective in mind. For an analytical initiative to succeed, different types of users—personas—will need different software tools. Providing casual users like business analysts with analytical tools primarily intended for power users—that is, statistical programmers—will overwhelm the analysts, who most likely do not have the skills or the time to learn about these advanced tools. Likewise, asking power users—that is, programmers—to use simple reporting tools for their analysis work would not work. An inventory of existing tools and user types and their competency levels with a particular software tool will help the organization when the time arrives to select vendors. Most analytical vendors are beginning to deliver enhanced or next-generation products that merge data management, visualization, reporting, analysis, and communications functionality. As a result, a wider range of user types will have access to a broader range of functionality from a single and integrated analytical environment. Don’t forget the access control requirements for users.”

Javier asked warily, “Well, that all sounds good, but there must be a catch. What are some of the hidden costs associated with these analytical initiatives? Is that what people call the total cost of ownership?”

I told him the truth: “That’s correct! Over time organizations have adopted a large number of disparate and unrelated analytical technologies, adding to tool fragmentation in their organizations. This situation creates problems of support; when something fails, vendors blame each other. It also creates training problems when diverse applications work in different ways for no reason. This situation has also been complicated by the mergers and acquisitions within the analytical vendor community.”

Javier perked up. “Yeah! Our organization has been frustrated in our ability to deploy analytical solutions effectively because of the overabundance of unrelated end user technologies from various vendors. Our end users and the IT organization have deployed various analytical tools without much (if any) thought about integration,

future needs, or issues. Unfortunately, they were reacting to the day-to-day pressures.”

I totally agreed with him. “A random portfolio of software tools in any organization typically contains products that are current and relevant, older but still-used products, and discontinued and unsupported fads. An organization may also have what is called shelfware, software that has been bought and nobody uses it.”

Zizi looked at her watch. “Look, we know that our organization needs to find the right number and mix of tools. To do this, we must stop the proliferation of analytical tools. That will ensure that we can centralize and provide a consistent and manageable analytical environment for our internal users.”

“To stop proliferation, you must enforce some standardization around analytical tools and governance for your data. This can be difficult because end users are usually partial to certain tools and resist changing the analytical tools they use and how they operate,” I added.

“Let’s don’t forget politics. Our analytical initiatives could span multiple business and functional groups in our organization. I could see that the politics associated with getting participation, data, and resources from the internal groups could introduce challenges and delays to an analytical initiative,” Zizi said.

Javier said thoughtfully, “Let me add something here. Our political and organizational challenges are unique. The politics of who has control related to visibility, information, resources, funding, and technology choices often leads to delays in IT initiatives. I think a cross-functional analytical initiative will fail quickly if it does not have a credible team leadership that anticipates and addresses these challenges.”

He obviously had extensive experience with office politics. I tried to sum things up as the minute hand clicked away on Zizi’s watch. “The readiness evaluation should include technology, people, process, and politics. It is a package deal.”

We shook hands and agreed to meet via Skype later on that week. I left them sitting at the table reevaluating their company’s subjective and objective levels of analytical well-being, excited about taking the plunge into a new and fascinating analytical mind-set for their company.



## NOTE FILE: SAMPLE MUTUAL CONFIDENTIALITY AGREEMENT

THIS CONFIDENTIALITY AGREEMENT (this "Agreement"), effective this DATE, is by and between PARTY1 NAME ("PARTY1") and PARTY2 NAME ("PARTY2").

Representatives of PARTY1 plan to meet with representatives of Potential Partner to consider establishing a business relationship or providing products or services (collectively, the "Transaction"). In connection with the discussions, each party might disclose certain confidential information to the other. As a condition of disclosing confidential information, the parties have agreed to treat such information as stated in this Agreement.

IN CONSIDERATION of the mutual obligations of the parties, the parties hereby agree as follows:

1. "Confidential Information" Defined. "Confidential Information" means all information disclosed by or on behalf of the disclosing party to or obtained by the receiving party concerning the disclosing party's business or any product or service developed (or proposed to be developed) by the disclosing party, and whether disclosed in writing, orally, or by inspection. Confidential Information may include, but is not limited to, developer information, pricing, customized products and services, designs, specifications, technical information, protocols, process information, code, software, financial data, business plans, marketing plans, trade secrets, processes, and techniques. Notwithstanding the foregoing, Confidential Information shall not include:
  - a. Information that at the time of disclosure is in the public domain or is otherwise available to the receiving party other than on a confidential basis;
  - b. Information that, after disclosure, becomes a part of the public domain by publication or otherwise through no fault of the receiving party or any third party under a confidential agreement with the disclosing party;
  - c. Information disclosed to the receiving party by a third party not under an obligation of confidentiality to the disclosing party; or
  - d. Information that is or has been developed by the receiving party (as evidenced by the receiving party's records) independent of the disclosures by the disclosing party.
2. Covenant of Confidentiality. The receiving party agrees to retain in confidence all Confidential Information. The receiving party further agrees

that it will not use or disclose to any third party, nor permit the use or disclosure to any third party of, any Confidential Information, except that the receiving party may make the Confidential Information available to its directors, officers, employees, and attorneys (collectively, its “Representatives”) who agree to be bound to the terms of this agreement and who reasonably need the information for the receiving party to evaluate the Transaction and, if the parties agree to undertake a Transaction, for the performance the receiving party’s duties in connection with such Transaction.

3. **Covenant to Return Confidential Information.** In the event the parties’ discussions terminate, or upon the earlier request of the disclosing party, the receiving party agrees to immediately return to the disclosing party all tangible and intangible documents and files obtained from the disclosing party containing Confidential Information and any materials created or derived from Confidential Information, by whomever or whenever made, without retaining any copies thereof. Once returned, the receiving party agrees to delete all electronic copies of the documents and files from the receiving party’s systems. The receiving party agrees to verify compliance in writing if requested by the disclosing party.
4. **Non-Solicitation and Hiring.** Neither party shall solicit, hire, or retain directly or indirectly any employee of the other party for a period of 12 months following the later of either the termination date of this agreement or the Transaction termination date.
5. **Survival of Confidentiality Obligation.** The confidentiality obligation contained herein shall survive the termination of such discussions and negotiations.
6. **Remedies for Violation.** The parties acknowledge that all Confidential Information disclosed by the disclosing party to the receiving party is significant, confidential, and materially affects the disclosing party’s business and goodwill. The receiving party expressly understands and acknowledges that a violation of this Agreement by the receiving party will cause irreparable injury to the disclosing party, which injury will not be fully compensable by money damages. The parties therefore agree that in the event the receiving party breaches or threatens to breach the covenants contained herein, the disclosing party shall be entitled as a matter of right to a restraining order, an injunction, a decree or decrees of specific performance, or other adequate relief from a court of competent jurisdiction. The provisions of this paragraph shall survive the termination of the parties’ discussions and negotiations pertaining to any Transaction.

In the event of the receiving party's breach, the receiving party agrees to indemnify the disclosing party for all costs of enforcement, including reasonable court costs and attorney fees, incurred while enforcing the disclosing party's rights under the agreement.

7. Severability. If any provision of this Agreement is held to be illegal, invalid, or unenforceable under present or future law effective during the term hereof, such provision shall be fully severable and this Agreement shall be construed and enforced as if such illegal, invalid, or unenforceable provision never comprised a part hereof, and the remaining provisions hereof shall remain in full force and effect and shall not be affected by the illegal, invalid, or unenforceable provision or by its severance.
8. Binding Effect. This Agreement shall be binding upon and shall inure to the benefit of the parties hereto and their respective successors and permitted assigns.

IN WITNESS WHEREOF, the parties hereto have executed this Agreement as of the date first above written.

#### SIGNATURES

##### **PARTY1**

By: \_\_\_\_\_

Title: \_\_\_\_\_

Chairman, President, or Vice President

##### **PARTY2**

By: \_\_\_\_\_

Title: \_\_\_\_\_

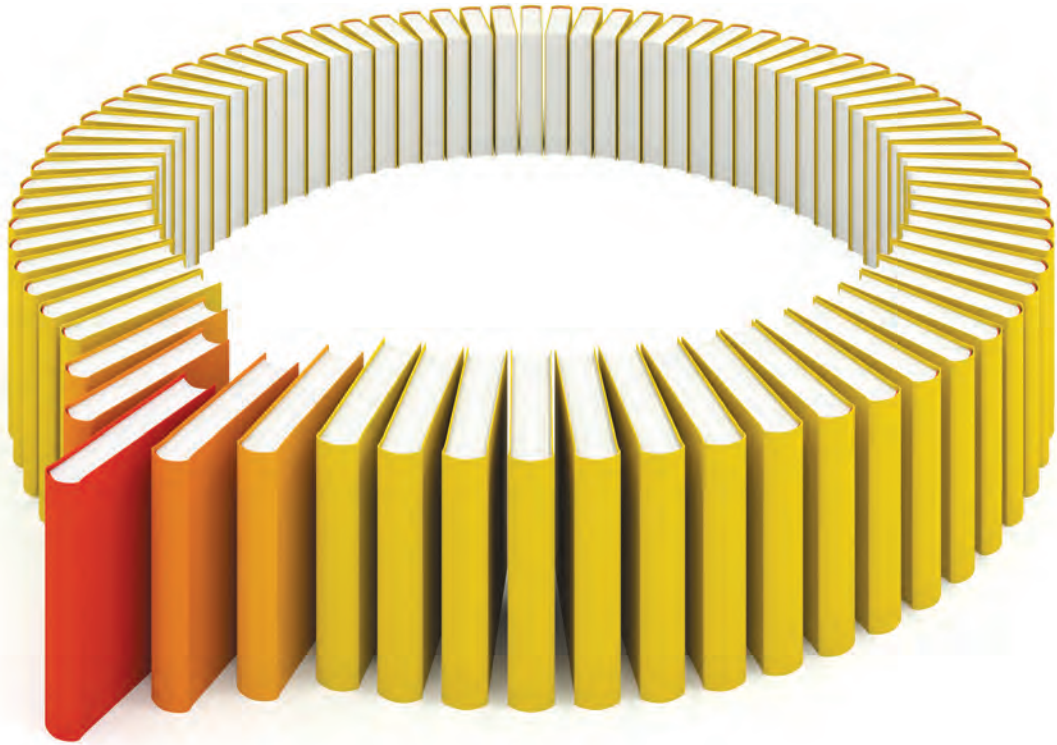
Chairman, President, or Vice President

#### **NOTE**

1. Jim Davis, Gloria J. Miller, and Allan Russell, *Information Revolution: Using the Information Evolution Model to Grow Your Business* (Hoboken, NJ: John Wiley & Sons, 2006).

From *Understanding the Predictive Analytics Lifecycle* by Alberto Cordoba. Copyright © 2014, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.





# Gain Greater Insight into Your SAS<sup>®</sup> Software with SAS Books.

Discover all that you need on your journey to knowledge and empowerment.

 [support.sas.com/bookstore](http://support.sas.com/bookstore)  
for additional books and resources.

  
THE POWER TO KNOW.®

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies. © 2013 SAS Institute Inc. All rights reserved. S107969US.0413