

FROM DATA TO DECISION

A Handbook for the Modern Business Analyst

*Marco Vriens, Song Chen,
and Chad Vidden*



SNEAK PREVIEW

For more information on adopting this
title for your course, please contact us at:
adopt@cognella.com or 800-200-3908

FROM DATA TO DECISION

FROM DATA TO DECISION

A Handbook for the Modern Business Analyst

Marco Vriens, Song Chen, and Chad Vidden



Bassim Hamadeh, CEO and Publisher
Angela Schultz, Senior Field Acquisitions Editor
Michelle Piehl, Senior Project Editor
Abbey Hastings, Associate Production Editor
Jess Estrella, Senior Graphic Designer
Stephanie Kohl, Licensing Associate
Gustavo Youngberg, Interior Designer
Natalie Piccotti, Senior Marketing Manager
Kassie Graves, Vice President of Editorial
Jamie Giganti, Director of Academic Publishing

Copyright © 2019 by Cognella, Inc. All rights reserved. No part of this publication may be reprinted, reproduced, transmitted, or utilized in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information retrieval system without the written permission of Cognella, Inc. For inquiries regarding permissions, translations, foreign rights, audio rights, and any other forms of reproduction, please contact the Cognella Licensing Department at rights@cognella.com.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Cover image copyright © 2017 iStockphoto LP/MicroStockHub.

Printed in the United States of America.

ISBN: 978-1-5165-2063-3 (pbk) / 978-1-5165-2064-0 (br)



BRIEF CONTENTS

1	INTRODUCTION: THE BUSINESS OF ANALYTICS	1
2	DATA AND DATA STRATEGY	13
3	A FRAMEWORK AND PROCESS FOR ANALYTICS	29
4	FROM INSIGHTS TO IMPACT	53
5	BASIC ANALYSIS	69
6	LINEAR REGRESSION	103
7	LOGISTIC REGRESSION	127
8	DECISION TREES	149
9	MULTI-DIMENSIONAL SCALING (MDS)	171
10	PRINCIPAL COMPONENT ANALYSIS AND FACTOR ANALYSIS	193
11	CLUSTER ANALYSIS	217
12	TIME SERIES ANALYSIS	243
13	NEURAL NETWORKS AND MACHINE LEARNING	267
14	TEXT ANALYTICS	279
	INDEX	301

CONTENTS

1	INTRODUCTION: THE BUSINESS OF ANALYTICS	1
	What am I going to learn?	1
	The Rise and Scope of Analytics	1
	Defining Analytics	2
	<i>What and How</i>	2
	<i>General Purpose</i>	3
	<i>Data</i>	4
	The Analytical Challenge	4
	The Value of Analytics	5
	<i>Praactive Benefits</i>	6
	<i>Reactive Benefits</i>	6
	<i>Empirical Evidence</i>	7
	Quiz	9
	References	9
2	DATA AND DATA STRATEGY	13
	What am I going to learn?	13
	What Data Is Available?	13
	<i>External Data Versus Internal Data</i>	13
	<i>Levels of Aggregation</i>	14
	<i>Cross-Sectional Versus Longitudinal Versus Time Series Data</i>	14
	<i>Hard Data Versus Soft Data</i>	15
	<i>Scanner Data</i>	15
	<i>Structured Versus Unstructured Data</i>	16
	<i>Observational Data</i>	16
	<i>RFID Data</i>	17
	<i>MtM SIM Card Data (IoT)</i>	17
	<i>Spatial Data</i>	18
	<i>Path Data</i>	18
	<i>Big Data</i>	18
	Data Governance	20
	<i>Data Inventory</i>	20
	<i>Data Processes and Evaluation</i>	21
	<i>Data Integration</i>	23
	Privacy and Ethical Considerations	24
	Quiz and Exercises	26
	References	26
3	A FRAMEWORK AND PROCESS FOR ANALYTICS	29
	What am I going to learn?	29
	How it works	29
	Target 1: Market Dynamics	29
	<i>Market Definition and Sizing</i>	30
	Target 2: Market Share Dynamics	31
	<i>Aggregate-Level Market Share Models</i>	31
	<i>Survey-Based Disaggregate Brand Choice Models</i>	31
	<i>Predicting New (or Improved) Product Success</i>	32
	Target 3: Marketing Campaign Dynamics	32
	Target 4: Customer Dynamics	35
	Analytical Project Execution	37
	<i>Business Problem and Understanding</i>	38
	<i>Reviewing What We Already Know</i>	38
	<i>Data Review and Preparation</i>	38
	<i>Analysis and Modeling</i>	39
	<i>Interpretation and the So What</i>	41
	<i>Validation</i>	41
	<i>Deployment</i>	42
	Advanced Topics: Data-Driven Versus Theory-Driven	43
	<i>Better Research Design</i>	45
	<i>Guiding Analysis and Modeling</i>	45
	<i>Helping with the Interpretation and Validation</i>	46
	<i>A Data-Driven Approach</i>	47
	A Rapidly Expanding Field	47
	Quiz	48
	References	49
4	FROM INSIGHTS TO IMPACT	53
	What am I going to learn?	53
	Introduction	53
	Intention to Use	54
	<i>Insights Knowledge</i>	54
	<i>Insights Credibility</i>	55
	<i>Advisor Credibility and Trust</i>	56
	<i>Acceptability</i>	57

Driving Decisions Quality	59
<i>Outcome and Insights Uncertainty</i>	59
<i>Decision Failure Example</i>	60
Managing Analytical Insights	63
Quiz and Exercises	64
References	65

5 BASIC ANALYSIS 69

What am I going to learn?	69
How it works	69
Dataset	69
Types of Variables	70
Data Exploration and Visualization	72
<i>Categorical Variables and Cross Tabulation</i>	72
<i>Numeric Variables and Distributions</i>	73
Data Transformations	81
Missing Data	84
Bivariate Relationships	86
<i>Correlation</i>	86
Statistical Tests for Comparing Groups	
(χ^2 , T-test, and ANOVA)	89
χ^2 -Test	90
<i>Student's T-test</i>	92
<i>Analysis of Variance (ANOVA)</i>	94
Advanced Topics	97
<i>ANOVA Extensions</i>	97
<i>Multivariate Analyses</i>	99
Business Applications	99
<i>Assessing if a Medical Adherence</i>	
<i>Regimen Works</i>	99
<i>Assessing if Executives Should be Rewarded</i>	
<i>for the Performance Gains</i>	101
Quiz	101
References	102

6 LINEAR REGRESSION 103

What am I going to learn?	103
How it works	103
Linear Regression	104
Implementing Linear Regression	105
Multiple Linear Regression	106
Is My Result Usable?	107
<i>Prediction and Forecasting</i>	107
<i>Causality and Driver Analysis</i>	109
External Validation	112
Troubleshooting	112
Advanced Topics	116
Business Applications: Predicting the	
Success of Movies	119
Quiz and Exercises	121

References	125
------------	-----

7 LOGISTIC REGRESSION 127

What am I going to learn?	127
How it works	127
<i>Data Exploration</i>	128
<i>Why Linear Regression Fails</i>	128
<i>Logistic Regression</i>	129
Driver Analysis	132
<i>Model Features</i>	132
<i>Key Driver Strength</i>	133
<i>Key Driver Stability</i>	134
Is My Result Usable?	135
<i>Accuracy and the Confusion Matrix</i>	135
<i>Choosing the Best Probability Threshold</i>	136
<i>Decision Costs and Benefits</i>	138
A Business Example: Competitive Drivers	
and the Switchable Consumer	139
<i>The Switchable Consumer for Restaurants</i>	141
Advanced Topics	143
<i>Multi-class Logistic Regression</i>	143
Quiz and Exercises	144
<i>Multinomial Logistic Regression</i>	144
<i>Ordinal Classification</i>	144
References	148

8 DECISION TREES 149

What am I going to learn?	149
How it works	149
Decision Tree Algorithm	150
Model Interpretation	155
Under-Fitting and Over-Fitting	156
When to Use Decision Trees?	159
<i>Decision Tree Advantages</i>	159
<i>Decision Tree Disadvantages</i>	159
Advanced Topics	160
<i>Types of Decision Trees</i>	160
<i>Marketing Segmentation</i>	160
<i>Random Forests</i>	161
Business Applications	164
Quiz	165
References	169

9 MULTI-DIMENSIONAL SCALING (MDS) 171

What am I going to learn?	171
<i>Terminology</i>	171
How it works	172

<i>The Basic MDS Idea</i>	174
<i>Evaluating MDS Maps</i>	175
Further Discussion	179
Advanced Topics	180
<i>Variations of MDS</i>	180
<i>Other Dimension Reduction Tools</i>	181
Business Application	184
<i>Dataset</i>	185
Quiz and Exercises	187
References	191

10 PRINCIPAL COMPONENT ANALYSIS AND FACTOR ANALYSIS 193

What am I going to learn?	193
How it works	194
Scaling Data	196
PCA	196
<i>Basic Example</i>	196
<i>Selecting Principal Components</i>	198
<i>Visualizing Principal Components</i>	200
Exploratory Factor Analysis	202
<i>Rotation and Uniqueness</i>	202
PCA and Multicollinearity	208
Comparing PCA and FA	209
Advanced Topics	209
<i>Confirmatory Factor Analysis</i>	209
<i>Parallel Analysis</i>	210
Business Applications	210
Quiz	212
References	216

11 CLUSTER ANALYSIS 217

What am I going to learn?	217
How it works	218
<i>Select the Segmentation Variables</i>	219
<i>Decide Whether to Standardize</i>	219
<i>Decide How to Measure the Similarity</i>	
<i>Between Respondents</i>	219
<i>Selection of a Cluster Analysis Algorithm</i>	219
<i>K-means</i>	221
<i>Hierarchical Clustering</i>	224
<i>Hierarchical vs. K-means</i>	226
<i>Evaluation Clustering Solutions</i>	226
The Use of Core Members	229
Typing Tools	230
Advanced Topics	232
<i>K-modes</i>	232
<i>Latent Class Analysis</i>	232
<i>Additive and Extended Trees</i>	232

<i>Ensemble Method</i>	233
Practical Considerations	233
<i>Large Number of Variables</i>	233
A Business Application	234
<i>Analysis and Evaluation</i>	235
<i>Results</i>	235
Quiz and Exercises	237
References	239

12 TIME SERIES ANALYSIS 243

What am I going to learn?	243
How it works	243
What Is ARIMA?	244
<i>Auto-Regression</i>	244
<i>Stationary Series</i>	244
<i>Moving Average</i>	246
Data Exploration	246
Linear Regression	248
Decomposition	250
Stationary Series	252
ARIMA Model Order	253
Fit and Model Testing	256
Seasonality	260
Advanced Topics	260
<i>Exponential Smoothing State Smooth</i>	
<i>Time Series Model</i>	260
Business Application	261
<i>Marketing-Mix Models</i>	261
Quiz	264
References	264

13 NEURAL NETWORKS AND MACHINE LEARNING 267

What am I going to learn?	267
How it works	268
Terminology	268
What Is Machine Learning?	269
Neural Networks	270
<i>The basic model idea</i>	270
<i>Architecture and complexity</i>	271
Advanced Topics	275
Business Applications	277
Quiz	277
References	278

14 TEXT ANALYTICS 279

What am I going to learn?	279
How it works	280

Dataset	280
Text Analytics Process	281
<i>Text Cleaning</i>	281
<i>Word Frequency</i>	284
<i>Document Term Matrix</i>	285
<i>LDA Topic Models</i>	287
<i>Comments on LDA Models</i>	292
Advanced Topics	294
<i>Natural Language Processing</i>	294
Business Applications	295

Quiz	296
References	299
INDEX	301

INTRODUCTION: THE BUSINESS OF ANALYTICS

WHAT AM I GOING TO LEARN?

In this chapter, we first discuss the rise of analytics as an increasingly separate discipline and outline the breadth and depth analytics has within business decision making. Next, we define analytics and advanced analytics and present various types. Lastly, we outline the analytics value chain and present research findings and case studies that show the impact of analytics on business performance. The latter shows why analytics has become a key capability and skill that executives and managers need to have.

THE RISE AND SCOPE OF ANALYTICS

Analytics is not a new discipline. Some firms embraced it decades ago even if only on an ad hoc basis (e.g., Gensch, Aversa, & Moore, 1990). More recently, some firms have made it an essential part of their business model (e.g., Davenport & Harris, 2007), and the interest in analytics as a business tool has surged (Davenport & Harris, 2007; Mahler & Boudreau, 2017; Vriens, 2012). Several books on the topic (Davenport & Harris, 2007; Davenport, Harris, & Morison, 2010; Siegel, 2013; Vriens, 2012), and many published academic papers document how firms successfully adopted analytics (e.g., Germann, Lilien, Fiedler, & Kraus, 2014; Lilien, Roberts, & Shankar, 2013; Silver, 2012). Analytics has now been applied across virtually every business domain (and outside business for that matter, e.g., medicine, chemistry, legal disputes, and biology, for example) and advanced analytics has been especially applied to the following:

- Marketing topics such as understanding the drivers of market share analysis (Gensch et al., 1990), identifying optimal pricing (Natter, Reutterer, Mild, & Taudes, 2007), new product adoption (Bass, 1969), finding optimal products (Kohli & Krishnamurti, 1989; Sawtooth, 2003), market segmentation (Wedel & Kamakura, 2000), marketing efficiency (Wiesel, Pauwels, & Arts, 2011), etc. We discuss these areas and others in chapter 3
- Finance topics such as fraud detection (Ngai, Hu, Chen, & Sun, 2011), bankruptcy (Fletcher & Goss, 1993), credit scoring and risk (Baesens, Rösch, & Scheule, 2017), thrift failures (Salchenberger, Cinar, & Lash, 1992), and financial distress (Coates & Fant, 1993), etc.
- Legal topics such as trademark infringement, patent cases/potential damage matters, consumer fraud, licensing non-compliance, and anti-trust (Altuglu & Schwabe, 2018; Fair & O’Laughlin, 2018)

- Operations and manufacturing topics such as process and operational efficiency (Evans & Fisher, 1994), early warning detection (Halvorsen, 2006), predictive maintenance, and scheduling optimization. With the growth of the Internet of Things (IoT) (see also chapter 2), firms will have big data on manufacturing and machine performance (O'Donovan, Leahy, Bruton, & O'Sullivan, 2015). Mining and analyzing such data can yield significant cost savings.
- HR topics such as the relationship between HR processes, employee engagement, and firm performance. For example, Coco, Jamison & Black (2011), studying Lowe's, found that a highly engaged employee leads to a 4% higher average per customer sales per store.
- Government and non-profit organizations now are increasingly benefitting from analytics (Wamba, Akter, Edwards, & Chopin, 2015).

Also, the usefulness of analytics is not limited to any given industry or company size; consumer products (CPG), financial services, pharmaceuticals, retail, ecommerce, telecom, etc., as well as small and large firms have been shown to benefit from the use of analytics (Davenport & Harris, 2007; Vriens, 2012).

DEFINING ANALYTICS

Analytics and advanced analytics (e.g., predictive analytics) are not fully defined as disciplines. People who work in areas such as analytics, marketing research, and advanced modeling will give a variety of answers. Here, we define analytics along three dimensions: (a) what and how, (b) the general purpose, and (c) the type of data.

What and How

In our experience, most practitioners will recognize that analytics and advanced analytics will view:

- 1 Data as a strategic asset, and will use tools to manage and analyze the data (e.g., tools to summarize or visualize the data and make the data more easily accessible),
- 2 Use specific models to quantify multivariate relationships (e.g. Vriens, 2012) so more and better insights can be found,
- 3 Analyze and model in multiple stages. Analytics is an iterative process. Initial analyses may not give any interesting insights. It is common to analyze a data set in many different ways before one discovers the insights (e.g. Vriens & Kidd, 2014).
- 4 Analyze and model using multiple data sources. Sometimes there are various components to an advanced analysis plan that are executed as stand-alone analyses. Once completed, the various results are later integrated. Combining different data sources can often bring unique perspectives and uncover new insights. For example, we can combine attitudinal, financial, and transactional data, or we can combine micro and macro data.

Other definitions have been suggested; for example, Davenport and Harris (2007) mention elements such as “extensive use of data,” “predictive analytics,” and the possibility of “fully automated decisions.” Laursen & Thorlund (2010) consider analytics to be a sub-set of business analytics. See also Mortensen, Doherty, and Robinson (2015) for more discussion on definitions.

General Purpose

Another way to define analytics is by distinguishing between descriptive, exploratory, predictive, causal, and prescriptive analytics. We note that these purposes can exist alongside each other and are not always mutually exclusive (e.g., an explorative study may have a predictive component, etc.).

Descriptive analytics refers to summaries of data via tables, charts, and maps. Typically, there is only basic analysis involved (e.g., simple frequency counts, indexing, etc.).

Exploratory analytics is also descriptive in nature but its purpose is to gain deeper insight into a topic or use the results for further analysis. Techniques such as factor analysis, cluster analysis, and multidimensional scaling (see corresponding chapters) are often used. We distinguish between measurement, data reduction, and representation as types of exploratory analyses:

- *Measurement.* Factor analysis can be used to determine if a set of survey variables capture a concept of interest. Say, we believe the concept of “brand love” exists and have developed survey questions to measure it. In this case we can use factor analysis to determine if there is one underlying factor that explains the co-variation between these survey variables (the details are discussed in the factor analysis chapter).
- *Data reduction.* Factor analysis can also be used as a data reduction technique. For example, we have many variables (many of which are inter-correlated) and we want to reduce this large set to a smaller number of factors. If we have survey data where two brands are evaluated on 40 perception attributes (e.g., “This is a brand I trust”; “This brand offers good quality”, etc. [40 or more is not uncommon]), we may want to simplify visualizing differences between brands. By reducing the 40 attributes to say five factors, we can substantially simplify the reporting and potential further analysis.
- *Representation.* Cluster analysis can be used to identify if groups of respondents or customers exist that are similar to each other but different than those in other groups. Instead of looking at similarities between respondents or customers, we can also be interested in similarities between products and brands. Here, we might use a multi-dimensional scaling approach to map the brands or products in a two-dimensional space.

Predictive analytics aims to predict something: sales next month, the value of a house, who will buy a product, who will defect on their mortgage, etc. We will expand on different approaches to predictive analytics in chapter 3.

Related to predictive purposes is gaining an understanding of the possible causal variables that cause a relevant business metric (e.g., what is driving our market share; what is driving how satisfied our customers are; etc.). Techniques that are used for both predictive

and causal analytics include regression, CHAID (more generally tree analyses), conjoint, structural equation modeling, etc. To assess or support a causality claim, we often will need to do additional analyses.

Prescriptive analytics refers to situations where we aim our analytics to provide specific recommendations, for example, when Spotify makes a song recommendation for you, when Amazon makes book recommendations, or when we develop a decision-support system that recommends SKU prices based on an objective profit optimizing function (which is based on a price-demand model) (see Natter et al., 2007).

Data

We note that there are three functional areas that substantially overlap where the dominant data used for analytics differs:

- 1 Traditionally there is the area of market research. Market research utilizes a variety of methodologies to collect data on consumers and markets including surveys, focus groups, etc. However, such data, in most cases, needs to be analyzed, and analytical approaches, such as the ones discussed in this book, will be applied.
- 2 Increasingly, analytics has also become a separate function in most organizations; firms collect data on their own customers (e.g., what they buy, when, etc.) or from their own operations and it is typically this data that gets analyzed.
- 3 There is the area of big data, sometimes referred to as data science. This function is tasked with the collection of massive big data and/or unstructured data. It may be part of an analytics or market research team or it may be its own stand-alone function (more in chapter 2 about big data).

Each function aims to collect and analyze data from consumers and markets and aims to create insights and knowledge that can help the firm make better decisions.

THE ANALYTICAL CHALLENGE

Many firms “have leveraged analytical insights to improve operational efficiency, grow revenues, and improve market position” (e.g. Arthur, 2013; Lilien, Roberts, & Shankar, 2013; Vriens, 2012; Vriens & Kidd, 2014). Others are not, or are barely, using analytics at all. For example, Rusetski (2014) interviewed 116 brand managers and found that 69% relied on some heuristic to make their pricing decisions. Information Week (Sullivan, 2005) reported that most managers in retail who are responsible for price setting prefer gut feel, while only 5 to 6 percent reported using decision support systems to help them leverage insights to set prices. Another study found that 40% of major decisions are based not on facts or insights, but on the manager’s gut (Rich, McCarthy, & Harris, 2009; Vriens, 2012). If that is true, then the percentage of “minor” or “less major” decisions based on gut feeling is likely to be a lot higher.

Firms who are not using analytics may simply not be aware of what can be done with analytics, or they may simply not see the benefits. A 2009 McKinsey survey among C-level

executives revealed that only about 10% of firms regularly deploy analytics. Kucera & White (2012) found that among the 160 businesses they interviewed, only about 16% used predictive analytics.

For those who have unsuccessfully used analytics, it may have been because it was misapplied. This may have resulted in weak or even invalid insights. If the results are uninteresting, they will quickly dampen the energy to pursue analytics. If they lead to incorrect recommendations, they may create a situation where analytics are seen as risky.

A survey among 3,000 executive managers worldwide found that half of the sample believed improving information and analytics was a top priority in their firm (Lavalle, Hopkins, Lesser, Shockley, & Kruschwitz, 2010). Almost half the interviewed sample “indicated that a ‘lack of understanding of how to use analytics to improve business’ was an obstacle to the widespread adoption of analytics in the organization.” Even though this study was published in 2010, in our practice, we see it almost daily. Executives often still lack a basic understanding of what analytics is and what it can do for them.

There are three knowledge gaps firms need to be aware of. First, executives need to know what business problems can be tackled with analytics so they can utilize this “tool” when feasible. Second, when the results of data analytics are presented to them, they need to be familiar enough with this discipline so they can assess and accept the insights that are presented. Third, there are the internal intelligence gatherers need to know how to deploy analytics productively (Vriens, 2012).

These barriers, knowledge of the potential value and the skill to implement it correctly, make it difficult to build and sustain an analytical capability that can deliver insights consistently and with impact. Firms may simply not have the right *tools*, *processes*, and *people* to do this and to keep momentum to establish a function that has the potential to be a source of competitive advantage.

The next section details how firms can benefit from analytics and summarizes the evidence that analytics leads to improved business results.

THE VALUE OF ANALYTICS

Figure 1.1, derived and combining elements from Vriens (2012) and Corte-Real, Oliveira, & Ruivo (2017), loosely describes how analytics can provide value to a firm.

In the beginning of the chain, we have data and the analytical skills and resources (tools, processes, people) as the primary “production of insights” input and make these analytical insights available to decision makers in the firm (more on this in chapter 4). The benefits of having analytical insights can be both proactive and reactive. Proactive means analytics helps identify decisions that a firm should make but that weren’t identified or prioritized by the firm (yet). Reactive refers to decisions that a firm already has identified it should make, or decisions it is already making. The goal for analytics is then to make these decisions better or more efficiently.

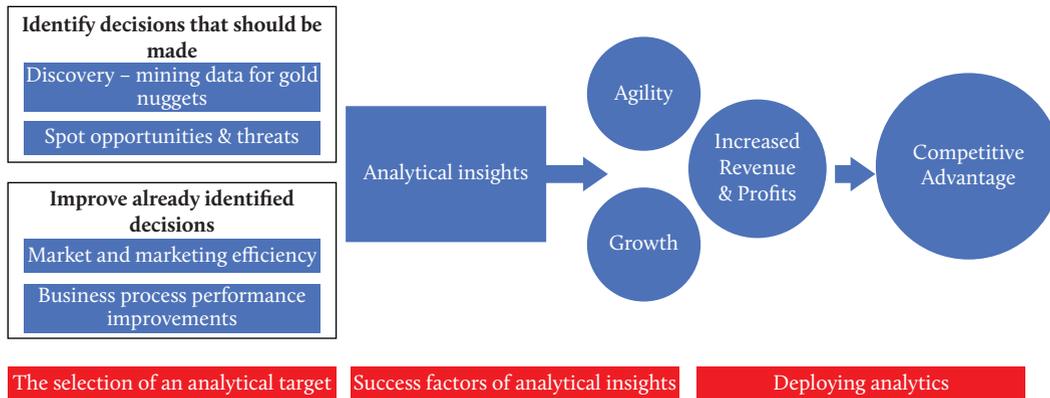


FIGURE 1.1 The analytics value chain adapted from Corte-Real. (2017).

Proactive Benefits

Having access to data enables data discovery activities without any specific decision in mind and possibly even without a decision maker asking for this discovery work. This may lead to identification of decisions that should be made. Such an analysis of the data might reveal actionable insights or an unexpected outcome. The firm may not have been aware of a trend, and may not have expected it, but once the data reveals it, it now presents a new choice for decision makers to consider or it may increase the urgency of an already identified decision. For example, Frito-Lay reviewed changes in preferences for its snack categories resulting from demographic, economic, and health trends to predict what categories would grow and shrink, and it helped the company focus on the growth categories (Girju, Adams, & Ratchford, 2009). Russo and Schoemaker (2002) describe how Encyclopedia Britannica saw their sales drop by as much as 53% because they insufficiently understood the growing appeal of the CD-ROM. Pro-active monitoring of the market and trendspotting could have avoided their demise.

Reactive Benefits

When decision makers do ask for data to be analyzed, we get reactive benefits. Reactive benefits include producing analytical insights to improve marketing efficiency or to improve process performance. Vriens (2012) discusses four types of analytical targets: market dynamics, marketing dynamics, campaign dynamics, and customer dynamics (see chapter 3 for an extensive discussion). Examples include selecting only the most promising new product concepts, thereby avoiding wasting time on concepts that would never succeed in the market or identifying customers in your database that will most likely respond to cross-selling efforts, etc.

Some examples of process improvement include the following:

- Using analytics to optimize inventory management.
- Using scorecards with a “single version of the truth” of key business performance metrics to avoid lengthy, recurrent, and unproductive discussions as to what the “numbers” should be.
- As an example of an improvement in the collection of required data itself: we now can leverage big data and social data (data pulled from discussions, posts, blogs) on the Internet. The availability of consumer-generated content on the Internet has enabled the creation of market structure and brand maps merely by using data that is already available instead of using the traditional way of collecting data via surveys (which may take months) (e.g., Vriens et al., 2017).

Proactive and reactive benefits can improve organizational agility and lead to growth. Organizational agility refers to the speed to which a firm can respond to new opportunities and threats (Corte-Real et al., 2017) and the speed at which it can make better marketing decisions. Growth refers to the ability to identify winning ideas (products, marketing campaigns, etc.). Some examples include the following:

- U-Haul recognized that consumers shop aggressively for the best daily rate, resulting in low margins. Once the customer signs a rental agreement, they are captive. This enables them to sell accessories at good margins. Recognizing this, U-Haul sought to buy the cheapest storage location in key locations, giving them a cost advantage and allowing them to keep the rates low (See Gadiesh & Gilbert, 1998).
- ABB Electric used analytical insights to identify which customers in their market were most likely to switch to their brand. This enabled them to target these “switchable consumers” with customized products. By doing this they efficiently grew their market share (Gensch et al., 1990).

Empirical Evidence

Ultimately, analytics needs to result in increased revenue, profits, and in some cases, competitive advantage. An impressive amount of evidence has been produced to support these claims. Evidence that data-driven and insight-driven decisions lead to better performance comes from two different sources: surveys and case studies:

Survey studies:

- a. A recent study among 3,000 executive managers (in the United States and the UK) showed that firms using business information and analytics to guide future strategies and analytical insights to guide day-to-day operations are twice as likely to be top performers (LaValle et al., 2010).
- b. Two studies done in New Zealand found similar results (Ganeshasundaram & Henley, 2007). One study showed that firms carrying out mostly decision research (defined as research explicitly done to support decisions) self-rated higher in terms of firm performance. In a second study, information use

correlated with more effective marketing planning, which in turn was positively related to firm performance.

- c. In one recent study (Brynjolfsson, Hitt, & Kim, 2011), CIOs and HR managers of 179 publicly traded firms were asked to what degree data-driven decision making was happening in their firms (e.g., one of the questions was “What is the typical basis for the decision on the creation of a new product?”). The response categories ranged from “1” (experience/expertise) to “5” (data). The analyses, using advanced econometric approaches, showed that data-driven decision making was positively related to sales, profitability, and the market value of the firm. Specifically, for productivity (sales), they find that data-driven decision making can account for a 5 to 6% increase.
- d. A Harvard Business Review article by Kashani (1989) reports his research on 17 cases. Of the cases where no formal research was involved, two-thirds failed, whereas only one third failed in cases where formal research had been done.
- e. Germann and colleagues (2014) interviewed 418 senior managers in retail (e.g., CMOs, CEOs, head of sales, etc.) across a number of different industries (e.g., retail, energy, insurance, telecom, etc.). Across all industries the use of customer analytics and analytics-based decision making was positively and significantly related to business performance such as stated sales growth and profits.
- f. After reviewing over 400 client engagements, McKinsey (Bhandari et al., 2014) found that integrated analytics can free up 15% to 20% of marketing spending.
- g. In a multi-country study in Europe, Corte-Real and colleagues (2017) tested and confirmed some of the relationships outlined in Figure 1.1. Using a structural equation approach, they confirmed links between (a) agility and competitive advantage, and (b) agility and process performance. The link between process performance and competitive advantage was not confirmed.

Case studies:

- a. Several books have described cases briefly (see Davenport & Harris, 2007; Clancy & Krieg, 2000); specifically, decisions that companies have made and have been successful or failed. Davenport & Harris (2007) illustrate how by deploying analytics, firms can make better decisions, and Clancy & Krieg (2000) illustrate the mistakes firms make by failing to gather analytical insights or by failing to use them.
- b. Many papers have been published (see chapter 3) that describe specific details as to what type of data, methodology, and analytics, what and how analytics were applied, and what specific financial impact these analytical efforts had (see Lilien et al., 2013; Vriens, 2012). The list of reputable brands and firms that have successfully applied analytics is long: ABB Electric, Baumax, CVS, Electronic Arts, Inofec, Jetstar, Kroger, Marriott, Microsoft, Netflix, Harrah’s Entertainment, J. Deere, Olive Garden, Rhenania, UBS, and Wal-Mart.

QUIZ

- 1 How can we define analytics?
- 2 What are reasons why firms might not have adopted analytics yet?
- 3 What is the difference between predictive and causal analytics? What do they have in common?
- 4 How can analytics add value to a firm?
- 5 Give an example of how a firm can benefit from descriptive analytics.
- 6 What are three reasons for using exploratory analytics?
- 7 What are proactive and reactive benefits of analytics? Give an example of each.
- 8 Which is more important: proactive benefits or reactive benefits?

REFERENCES

- Altuglu, V., & Schwabe, R. (2018). Machine learning in litigation. In N. Mizik, N. & D. Hanssens, (Eds.), *Handbook of Marketing Analytics* (pp. 661–669) Northampton, MA: Edward Elgar Publishing.
- Arthur, L. (2013), *Big Data Marketing*. J. Wiley & Sons, Hoboken, NJ. USA.
- Baesens, B., Rösch, D., & Scheule, H. (2017). *Credit risk analytics: Measurement techniques, applications, and examples in SAS*. Wiley India Pvt. Ltd., India.
- Bass, F. M. (1969). A new product growth model for consumer durables. *Management Science*, 15(5), 215–227.
- Bhandari, R., Sinfer, M., & van der Scheer, H. (2014). Using marketing analytics to drive superior growth. *McKinsey Report*. <https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/using-marketing-analytics-to-drive-superior-growth>.
- Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in numbers: How does data-driven decision-making affect firm performance. *Thirty Second International Conference on Information Systems*, Shanghai, China, pp. 1–18.
- Clancy, K., & Krieg, P. C. (2000). *Counter-intuitive marketing*. New York, NY: Free Press.
- Coates, P. K., & Fant, L. F. (1993). Recognizing financial distress patterns using a neural network tool. *Financial Management*, 22(3), 142–155.
- Coco, C.T., Jamison, F., & Black, H. (2011), “Connecting people’s investments and business outcomes at Lowe’s”, *People & Strategy*, Vol. 34, No. 2, pp. 28–33.
- Corte-Real, N., Oliveira, T., & Ruivo, P. (2017). Assessing business value of big data analytics in European firms. *Journal of Business Research*, 70, 379–390.
- Davenport, T. H., & Harris J. G. (2007). *Competing on analytics: The new science of winning*. Boston, MA: Harvard Business School Publishing.
- Davenport, T. H., Harris, J. G., & Morison, R. (2010). *Analytics at work: Smarter decisions, better results*. Boston, MA: Harvard Business School Publishing.
- Evans, B., & Fisher, D. (1994). Overcoming process delays with decision tree induction. *IEEE Expert*, 9(1), 60–66.
- Fair, R. K., & O’Laughlin, L. (2018). Avoiding bias: Ensuring validity and admissibility of survey evidence in litigations. In N. Mizik & D. Hanssens (Eds.), *Handbook of marketing analytics* (pp. 549–560). Northampton, MA: Edward Elgar Publishing.

- Fletcher, D. & Goss, E. (1993). Forecasting with neural networks: An application using bankruptcy data. *Information and Management*, 24, 159–167.
- Gadiesh, G. & Gilbert, J.L. (1998) Profit pools: A fresh look at strategy. *Harvard Business Review*, 139–147.
- Ganeshasundaram, R., & Henley, N. (2007). “Decision research” correlates directly with better business performance. *Marketing Intelligence and Planning*, 25(1), 45–65.
- Gensch, D. H., Aversa, N., & Moore, S. P. (1990). A choice-modeling market information system that enabled ABB Electric to expand its market share,” *Interfaces*, 20(1), 6–25.
- Germann, F., Lilien, G., Fiedler, L., & Kraus, M. (2014). Do retailers benefit from deploying analytics? *Journal of Retailing*, 90(4), 587–593.
- Girju, M., Adams, M., & Ratchford, B. M. (2010). Demo impact: Modeling, forecasting and managing the impact of major U. S. socio–demographic trends on multi–category snack consumption. *Review of Marketing Science*, 8(1), 2–18.
- Halvorsen, L. A. (2006). Comprehensive model for manufacturing analytics. WBF North American Conference (Atlanta, GA). http://www.controlglobal.com/assets/Media/MediaManager/wp_06_041_wbf_halvorsen.pdf.
- Kashani, K. (1989, September/October). Beware the pitfalls of global marketing. *Harvard Business Review*, pp. 91–98.
- Kohli, R., & Krishnamurti, R. (1989). Optimal product design using conjoint analysis: Computational complexity and algorithms. *European Journal of Operations Research*, 40, 186–195.
- Kucera, T., & White, D. (2012). Predictive analytics for sales and marketing: Seeing around corners. *Aberdeen Group Research Brief*. Retrieved from www.aberdeen.com
- Lavalle, S., M., Hopkins, E., Lesser, E., Shockley, R., & Kruschwitz, N. (2010). Analytics: The new path to value: How the smartest companies are embedding analytics to transform insights into action. *MIT Sloan Management Review* [Research report in collaboration with the IBM Institute for Business Value].
- Laursen, G. H. N., & Thorlund, J. (2017). *Business analytics for managers: Taking business intelligence beyond reporting*. Hoboken, NJ: Wiley.
- Lilien, G. L., Roberts, J. H., & Shankar, V. (2013). Effective marketing science applications: Insights from the ISMS-MSI Practice prize finalist papers and projects. *Marketing Science*, 32(2), 229–245.
- Mahler, J. H. & Boudreau, J. W. (2017). An evidence-based review of HR analytics. *International Journal of Human Resource Management*, 28(1), 3–26.
- Mortensen, M. J., Doherty, N. F., & Robinson, S. (2015). Operational research: From Taylorism to terabytes: A research agenda for the analytics age. *European Journal of Operations Research*, 241, 583–595.
- Natter, M., Reutterer, T., Mild, A., & Taudes, A. (2007). An assortment-wide decision-support system for dynamic pricing and promotion in DIY retailing. *Marketing Science* 26(4), 576–583.
- Ngai, E. W. T., Hu, Y., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of the literature. *Decision Support Systems*, 50, 559–569.
- O’Donovan, P., Leahy, K., Bruton, K., & O’Sullivan, D. J. T. (2015). Big data in manufacturing: A systematic mapping study. *Journal of Big Data*, 2(20), 1–22.

- Rich, D., McCarthy, B., & Harris, J. (2009). Getting serious about analytics to drive outcomes. *Accenture Report*. Retrieved from: http://www.umsl.edu/~sauterv/DSS/Accenture_Getting_Serious_About_Analytics.pdf
- Rusetski, A. (2014). Pricing by intuition: Managerial choices with limited information. *Journal of Business Research*, 67(8), 1733–1743.
- Russo, J. E., & Schoemaker, P. J. H. (2002). *Winning decisions: Getting it right the first time*. New York, NY: Doubleday.
- Salchenberger, L. M., Cinar, E. M., & Lash, N. A. (1992). Neural networks: A new tool to predict thrift failures. *Decision Sciences*, 23, 899–916.
- Sawtooth. (2018). *Advanced simulation module (ASM) for product optimization v1.5 technical paper*. Retrieved from: <https://www.sawtoothsoftware.com/support/technical-papers/sawtooth-software-products/advanced-simulation-module-asm-technical-paper-2003>.
- Siegel, E. (2013). *Predictive analytics: The power to predict who will click, buy, lie or die*. Hoboken, NJ: John Wiley & Sons.
- Silver, N. (2012). *The signal and the noise: Why some predictions fail and some don't*. New York, NY: Penguin.
- Sullivan, L. (2005, August). Fine-tuned pricing. *Information Week*. Retrieved from: <https://www.informationweek.com/fine-tunedpricing/d/d-id/1035175>.
- Vriens, M. (2012). *The insights advantage: Knowing how to win*. Bloomington, IN: i-Universe.
- Vriens, M. & Kidd, P. (2014), “The big data shift: What every marketer needs to know about advanced analytics”, *Marketing Insights*, 23–27.
- Vriens, M., Chen, S., Kaularz, S., & Vidden, C. (2017). Assessing the impact of a brand crisis using big data: The case of the VW Diesel Emission Crisis. *DMA Annual Analytics Journal*, 107–118.
- Wamba, S. F., Akter, S., Edwards, A., & Chopin, G. (2015). How big data can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, 234–246.
- Wedel, M., & Kamukura, W. A. (2000). *Market segmentation: Methodological and conceptual foundations*. New York, NY: Springer.
- Wiesel, T., Pauwels, K., & Arts, J. (2011). Marketing's profit impact: Quantifying online and off-line funnel progression. *Marketing Science*, 30(4), 604–611.