Judah Phillips

Foreword by Dr. Thomas Davenport

ECOMMERCE ANALYTICS

Analyze and Improve the Impact of Your Digital Strategy

Praise for *Ecommerce Analytics*

"Leveraging analytics to improve business results requires first knowing the questions that need to be answered. *Ecommerce Analytics* is the book to read if you are looking to use data to improve your online performance."

—**Josh James**, Founder and CEO, Domo; Cofounder and Former CEO, Omniture

"With U.S. ecommerce sales set to cross $450 billion in 2017, the opportunity in front of you is immense. Yet, it is likely your conversion rates are stuck at 2%. I'm excited about Judah's comprehensive tome because it is just what the doc prescribed to help unstick your ecommerce strategies. From cart abandonment to multichannel attribution to lifetime value… You'll get precise guidance to win big!"

—**Avinash Kaushik**, Digital Marketing Evangelist, Google;
author, *Web Analytics 2.0, Web Analytics: An Hour a Day*

"Are you a digital analyst working for an ecommerce company? Or are you an ecommerce marketer and your boss just asked you to provide more data about your digital efforts? If you are, then you should check out *Ecommerce Analytics*!

"Judah does a great job of making digital analytics for ecommerce logical and easy to understand. If you're a digital analyst familiar with topics like attribution modeling, you'll love how Judah discusses the specifics for an ecommerce business.

"If you're new to digital analytics you'll find the content easy to approach and very actionable. But that doesn't mean that he cuts corners! Judah is really, really thorough! He takes the time to dive into all the different metrics and analysis techniques that you can perform on your ecommerce business.

"I put *Ecommerce Analytics* on my bookshelf and plan to use it whenever I work with an ecommerce company—you should, too!"

—**Justin Cutroni**, Analytics Evangelist, Google;
author of *Google Analytics* and *Performance Marketing with Google Analytics*

"This important book is required reading for anyone who wants to understand how to deliver successful ecommerce analysis and data science. It's instructive and helpful, unifying the subject matter in way that is actionable for leadership, managers, technologists, and analysts."

—**Raj Aggarwal**, Cofounder and CEO, Localytics

"Judah has created a must-read book for all digital analysts. It's clearly framed and combines a comprehensive understanding of the topic with a practical flavor only the author can bring through decades of experience. This should be a hit in any college analytics class and will be on my graduate analytic course reading list in the future."

—**Rand Schulman**, Managing Partner, Efectyv Digital;
Cofounder, DealSignal; Cofounder, Digital Analytics Association

"In *Ecommerce Analytics*, Judah has delivered a comprehensive survey of the field, covering a broad array of topics important to implementers, analysts, and executives. This book contains a wealth of information that will be valuable in successfully executing an ecommerce strategy."

—**Bob Page**, Director Emeritus, Digital Analytics Association; Internet entrepreneur

business concerns. It is common for a "relational data model" to be chosen for ecommerce. The most common relational model for analytical processing is a dimensional data model. Of course, in today's world of big data, there are "post-relational" data models (such as graphs and multivalue stores) that may be used. But this book is not meant to and will not explore the reasons for implementation choice behind data models. Instead, readers need to understand the dimensional data model and its logical representation. It is from the dimensional data model that analytics can be extracted, new data can be joined, and detailed analysis can be done. The ecommerce data model focuses on facts/measures and dimensions. A fact can be a business measurement, a quantity (an amount), or an event that is part of a business process. A fact is typically a continuous numeric value that is additive, but it does not have to be. Facts can be additive (summable), semi-additive (summable under certain conditions), and non-additive. Facts are used to answer questions such as "What were sales last month?" Facts are identified in a construct named a "fact table." Without getting into data model design, it's important for an ecommerce analyst to know that measures are calculated values that are based on facts. A measure points to a fact and also specifies the aggregation on the value (such as a sum or median). For example, you may have a fact named "Order Transaction Fact Table." In that table, you may have some measures, such as sales price, sales quantity, cost (as well as foreign keys to dimensional tables, such as time/date, product ID, category ID, and promotion ID). A measure is usually an additive numeric value for representing a business metric. You may have multiple measures associated with a fact (Chen 1976). For example, if your fact table is used to store data related to international purchases, you may have measures for each type of currency based off of one fact (i.e., the international purchase). For ecommerce you might also have measures related to product cost, list price, sale price, average sale price, and so on.

Dimensions provide context about a fact, such as what or to whom the fact would apply, when it can be used, under what conditions the fact can be measured, and where. Dimensions correspond to people and objects. Facts correspond to events and numbers. Dimensions are typically discrete variables that can represent a character, string, or number value. Dimensions that do not change are called static

dimensions; those that change are called slowly changing dimensions. For example, you may have a dimension named "brand" that represents the brands you sell. In many businesses the brands do not change over time, but in others they can change slowly over time. Slowly changing dimensions have special methodologies, referred to as Type 0 to 6. In ecommerce analysis, dimensions are used to identify and group data, such as customers, locations, marketing campaigns, and categories. You may use a dimension to answer the question "What products did we sell last holiday season?" For example, you may have a dimension named "Customer Dimension" that references the fact "Customer ID" (as the primary key). In this Customer Dimension, you might want to track the customer name, address, and other attributes of the customer profile. Other types of dimensions, beyond the scope of this book, include conforming dimensions, junk dimensions, degenerate dimensions, and role-playing dimensions.

You may be asking yourself, "What is a primary key?" A primary key is a database construct that is identified in a dimensional table, and is thus part of a dimension. It is a column or set of columns in a table where the values uniquely identify a row in the table. The primary key is unique for each row in the dimension table. The value of the primary key is what's stored in the fact table. Within a fact table, you will have another type of key named a "dimension foreign key," which is the key whose values correspond to the values of the primary key. Primary keys are related to dimension foreign keys in the sense that the foreign key enables the relationship between the fact and the dimension by enforcing referential integrity between the two tables (Agrawal 2001). The primary key can be a "surrogate key" that joins the dimension to the fact.

When you are doing ecommerce analytics, it is important to understand the facts, measures, dimensions, and keys in your database (if they are made available) in order to query the database. In some tools, esoteric details related to the dimensional model and the query language can be abstracted and hidden from the user via drag-and-drop interfaces or interfaces in which you visually select a dimension or measure as, respectively, the rows and columns in a report. Google Analytics' "custom reporting" feature is an example of this type of visual query abstraction.

As an analyst, you likely will have to query databases or express, in somewhat technical terms, the data you are looking to analyze. Understanding facts, measures, dimensions, and the types of keys will help you navigate through the data and get your bearings in order to get the data you need for analysis.

Explaining a Sample Ecommerce Data Model

Now that you have a general understanding of dimensional data models and the concepts applied to them, it is possible to define a sample ecommerce data model. Your ecommerce data model will be custom to your company. Analysts work with technologists to define the conceptual data model and then implement it logically (with keys) and physically (in a database so it can be with data from site activity). Although the model will vary by company, it can have facts similar to those in the following list. Each of these facts will have dimensions and measures and at least one primary key for joining data across facts to create new combined data:

- **Inventory fact** contains data about the products within your inventory wherever it exists, such as in your warehouse or, in the case of drop-shipping, at the manufacturer. The information contained in this fact will include dimensions about the inventory, how long it has been at the company, and other information about cost and sales.
- **Products fact** represents information specific to particular, unique products, such as the product name, the SKU, the retail price, the brand, the category, the department name, and other data and metadata related to specific products.
- **Order fact** indicates data related to the cost and profit of orders, the number of orders, the number of items in an order, the date the order was created, and more.
- **Order Items fact** contains data about the item's brand, the identifiers, the sale price, the data purchased, and information related to whether the item was returned and details about the return.

- **Customer fact** identifies data about customers, including the customer ID, the name, the address, the e-mail address, and other personally identifiable and mostly anonymous data (such as gender).
- **Customer Order fact** represents data about customers after they place an order. It can contain information about how long the user has been a customer, when she first and last ordered an item, whether she is a repeat customer, and other derivative information. Some data modelers may include this data in the customer fact and not create a separate fact.

In this simplified data model we have users who order products that are in inventory. As you can see, this sample model is elegant in its simplicity (Occam's razor, anyone?) but may not be sufficient in all business cases. These facts will contain data that can be represented as dimensions and measures, which are used for analysis. When working with a technical team to model this data, you must identify the data you need and relationships necessary to perform analysis that answers your business questions. The following sections expand on these facts and define the dimensions and measures you may want to create within them, or use as a basis for understanding or evolving your company's data model.

Understanding the Inventory Fact

"Inventory" is a collection of goods that an ecommerce company retains for the purpose of reselling them for profit. An Inventory fact will have different dimensions and measures. An Inventory Identifier will always be created to uniquely specify the inventory item. Dimensions such as Cost of the inventory as paid by the ecommerce company are included. The Date Created and Sold Date dimensions are used to specify the calendar date when the inventory was first available (in the warehouse or the site) and then sold. Days in Inventory and Days Since Arrival may be used. Measures in an Inventory fact could include cost metrics, such as Average Cost and Total Cost. The Number on Hand and Number Sold are useful Inventory measures, as is the Sold Percent.

Understanding the Product Fact

A Product is a specific inventory item that is sold on the ecommerce site. Products are managed and marketed. Promotions and discounts are offered on products. Products are, of course, very important to track and understand. Thus, having a comprehensive fact is helpful. A Product Fact will have a Product Identifier to uniquely specify it and also a Product Name. The Shop Keeper's Unit (SKU) may be assigned. Price-based dimensions such as Wholesale Price, Retail Price, Sale Price, Gross Margin, and Discount Percentage might be used. The product's Brand, Category, and Department can be dimensions, as might be the Sale Rank representing the popularity of the product.

The Product fact can have measures, most commonly, counts of dimensions, like Total Count, Brand Count, and price and discount-based measures, like Average Sales Price and Average Promotional Discount. If you sell services online, then you would have a Services Fact, which will have similar and probably slightly different dimensions, like Provider.

Understanding the Order Fact

The Order is a request for one or more products (or services) that occurs as the result of the purchasing process. Order analysis is a core activity in ecommerce; thus, the Order fact has many possible dimensions. An Order Identifier, Order Created Date, and the Order Status are all necessary. Financially derived dimensions, like the Order Cost, Order Total Amount, Order Gross Margin, and Order Profit are helpful. Dimensions related to marketing programs, like Campaign Source and Traffic Source, could be used. Dimensions that flag specific concepts may be created, such as First Purchase Flag, Days Since Last Order, and Repeat Order Within *X* Days (when order data is joined with customer data).

The Order fact has measures such as the Count of Orders, the Purchase Date, the Average Order Amount, Average Profit, and Total Profit. If orders can be "pending," then measures to determine the Maximum Days Pending or Average Days Pending could

be calculated, as could the Repeat Purchase Rate for identical orders. Measures related to customers and their orders can be created as Order measures, like the Count Percent Change (the change in the count of orders compared across periods) and the Count Percent Total (the total percentage of orders). New Customer Revenue, New Customer Orders, Repeat Customer Revenue, and Repeat Customer Orders may be modeled.

Understanding the Order Item Fact

The Order Item fact is related to the Order in that the Order contains the Order Item. It's awkward to write, but simple to understand. When you place an order on a site, you buy things. The item is the thing you buy. You may buy many items in an order. The Order Item has dimensions, such as the Order Item Identifier and many financially based dimensions, such as the Wholesale Price, Retail Price, Sales Price, Gross Margin Total, Gross Margin Percentage, Gross Margin Tier, Contribution Margin Total, Contribution Margin Percentage, and Contribution Margin Tier.

The Order Item fact also has measures, such as the Count of Order Items (people order more than one individual item in the order). Financially based metrics should be included, such as Average Gross Margin, Total Gross Margin, Average Contribution Margin, and Total Contribution Margin (and Percentages). Of course, you can also include Average Sales Price and Total Sales Price.

Understanding the Customers Fact

A Customer is a user who has ordered one or more products from your ecommerce site's inventory. The Customer fact has many dimensions, which should seem familiar to you. The Customer Identifier uniquely identifies the customer. Dimensions are included such as Name, Email, Address, City, State, Zip Code, and Country. The Gender of the customer may be a dimension. Marketing-focused dimensions in the Customers fact could include the Campaign (attributed to acquiring the customer), Marketing Channel (attributed, again). Value-based dimensions include Number of Orders, Lifetime Value,