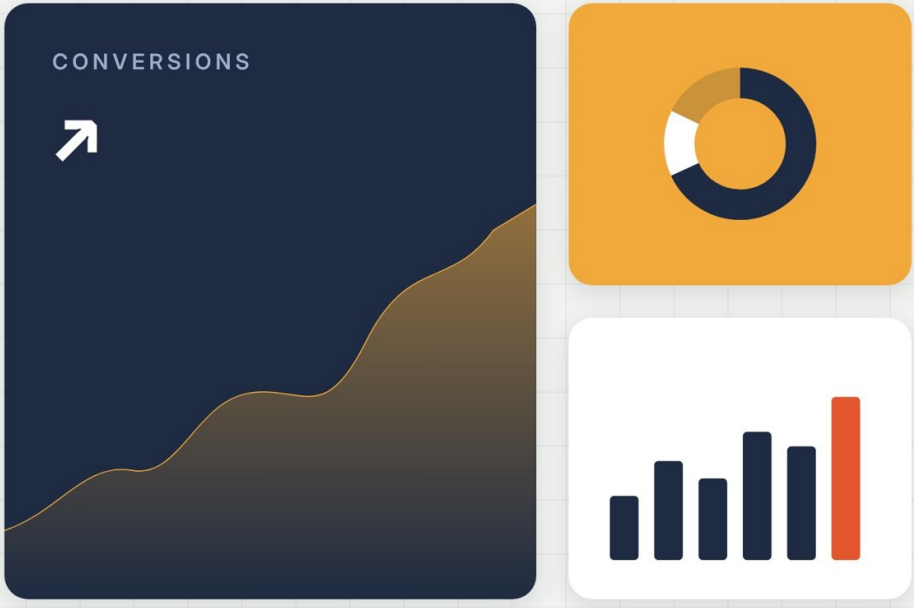
JULIAN DELPHIKI

UPDATED EDITION

Web Analytics & Big Data

Improve your eCommerce metrics
and online insights to sell more,
and explore the world of Big Data



Thinking School

CLEAR THINKING · CAREFULLY EDITED

WEB ANALYTICS & BIG DATA

UPDATED EDITION

Web Analytics & Big Data

Improve your eCommerce metrics and online insights to sell more, and explore the world of Big Data

JULIAN DELPHIKI

Thinking School

Thinking School

Published by Thinking School

Web Analytics & Big Data — Updated edition

© Julian Delphiki, 2025. All rights reserved.

No part of this publication may be reproduced, distributed, or transmitted in any form or by any means, including photocopying, recording, or other electronic or mechanical methods, without the prior written permission of the publisher, except in the case of brief quotations embodied in critical reviews and certain other noncommercial uses permitted by copyright law.

Product names, logos and brands mentioned in this book are the property of their respective owners and are used for identification purposes only. Information about tools and platforms was checked at the time of writing; features change frequently.

ISBN 979-8-6429-0767-2

Thinking School — books for readers who would rather understand than be told.

CONTENTS

PART I · FOUNDATIONS 1

1 The State of Web Analytics 3
- Meet Lumen & Leaf 3
- From the Truth to Indicative Metrics 5
- GA4, the Inevitable 7
- How This Book Is Organized 9

2 Measurement in a Privacy-First World 11
- The Rules of the Game 11
- The Cookie U-Turn 13
- Consent and the Consent Management Platform 14
- Consent Mode 14
- First-Party Data: Your Most Valuable Asset 17
- Lumen & Leaf's Privacy-First Plan 19

3 The Basics 21
- Web Analytics Components 21
- Goals, Measures and KPIs 25
- Putting It into Practice: A Measurement Plan for Lumen & Leaf 29
- Naming Events Consistently 31

4 Setting Up GA4 the Right Way 35
- How GA4 Is Organized 35
- Installing the Google Tag 36
- Events, Key Events and Values 37
- Custom Dimensions and Metrics 37
- Settings That Are Easy to Forget 37
- Tagging Your Campaigns 39
- Checking Your Setup 40
- A 20-Point GA4 Audit Checklist 42

PART II · FROM DATA TO DECISIONS 45

5 Inside GA4's Reports 47
- Report Collections: Business Objectives or Life Cycle 48
- Realtime Reports 48
- Acquisition: How Do People Find Us? 49
- Engagement: What Do Visitors Do? 52
- Monetization and Retention: Do They Buy, and Do They Come Back? 53
- User Reports: Who Are Our Visitors? 54
- Explorations: Your Analysis Workshop 54
- Know the Limits: Sampling, Thresholds and Modeled Data 56

6 eCommerce Analytics 59
- The eCommerce Events 59
- Reading the Monetization Reports 60

Funnel Arithmetic: Where Lumen & Leaf Loses Customers 61
Products and Promotions 63
Customer Value: From Orders to Relationships 64
RFM: A Simple Way to Segment Customers 65

7 Segmentation and Reporting 69
Building Segments in GA4 70
From Segments to Audiences 71
Turning Analysis into Reports 72
A One-Page Report That Gets Read 74

8 Dashboards 79
What Makes a Dashboard Work 79
Choosing the Right Chart 80
Building a Dashboard in Data Studio 81
A Dashboard Makeover 83

9 Attribution, Incrementality and Marketing Mix Modeling 87
Why Attribution Matters 87
Attribution Models in GA4 88
The Limits of Attribution 89
Incrementality: What Would Have Happened Anyway? 90
Marketing Mix Modeling 92
Triangulation: Using All Three Together 93

10 Testing and Optimization 97
Deciding What to Test 97
How Long Should a Test Run? 100
Common Testing Mistakes 102
Server-Side and Feature-Flag Testing 102
Lumen & Leaf's First Test 103
Testing Tools Today 104

11 AI and the Future of Analytics 107
Search Without Clicks 107
The AI Assistant Channel 108
What AI Already Does in Your Analytics 109
Where AI Answers Go Wrong 110
The Analyst's Job in the Age of AI 111

PART III · BIG DATA 113

12 Understanding Big Data 115
The Origins of Big Data 115
Uses of Big Data 118
The Generative AI Leap 120
Privacy, Security and the Future 122

13 The Big Data Project Roadmap 125
Capture, Process, Visualize, Store, Analyze, Apply and Revert 125

Scope and Risks 132
The Modern Data Stack 134
Working with the GA4 BigQuery Export 136
Data Governance and Quality 138

Glossary 143

Further Reading and Sources 151

About the Author 153

Also by Julian Delphiki 155

Index 157

PART I

Foundations

What analytics can and cannot tell you, how to measure responsibly, and how to set your tools up properly.

CHAPTER 1

The State of Web Analytics

Why today's analytics is a compass, not a map—and what privacy, GA4 and AI have changed.

WEB ANALYTICS HAS never been more powerful, and it has never been less complete. Your tools can now predict which visitors are likely to buy, tell you in plain language why revenue dropped last Tuesday and show you how many people arrived from an AI assistant. At the same time, a growing share of your visitors decline cookies, block tracking scripts or never reach your site at all because a search engine answered their question on the results page.

This edition is written for that world. It keeps the principles that made the first edition useful—start from your business objectives, measure only what you will act on, test before you change—and rebuilds the practical chapters around the way analysts actually work in 2026: with Google Analytics 4 (GA4), consent-based measurement, data warehouses, experiments and AI.

> FROM THE AUTHOR
>
> **A note on reading order**
>
> This chapter sets the scene and is the most advanced in Part I. If you are new to analytics, you may prefer to read Chapter 3, The Basics, first and come back here once the core vocabulary feels familiar. Chapters 2 and 4 are the most technical; marketers who will not configure tools themselves can skim them and keep them as a reference for conversations with developers.

Meet Lumen & Leaf

Theory only takes you so far. To show how data moves from a tag on a web page to a decision in a meeting, every chapter of this book follows the same online store: **Lumen & Leaf**, a fictional company based in Valencia that sells indoor

plants, ceramic pots, care kits and a monthly plant-food subscription to customers in Spain, Portugal, France, Italy and Germany.

Lumen & Leaf is small enough to be understandable and big enough to have real problems. Its founder, Marta Ruiz, runs the business. Dani Ortega leads marketing and spends about €28,000 a month across Google Ads, Meta, influencers and email. Iker Sanz looks after the website, and Sofía Molina answers customers. You have just joined as their first analyst. Here is the month you inherit.

Measure	March 2026
Sessions	120,000
Users	84,000 (61,000 of them new)
Purchases	2,160 (a session conversion rate of 1.8%)
Revenue	€133,900 (average order value €62)
Share of sessions on mobile	71%
Visitors who accept analytics cookies	62%
Monthly marketing spend	€28,000

Lumen & Leaf at a glance: March 2026

All figures for Lumen & Leaf are invented for teaching purposes, but they are realistic for a European online store of this size, and they stay consistent from chapter to chapter. When Chapter 6 calculates the store's checkout drop-off or Chapter 10 works out how long a test must run, it uses the same numbers you see here.

TRY IT

Open the GA4 demo account

Google offers a free demo account with real, anonymized data from the **Google Merchandise Store**, an online shop that sells Google-branded products, and from a mobile game called Flood-It!. Search for "Google Analytics demo account," sign in with any Google account and choose the Merchandise Store property. You will have viewer access: you can explore every report and build your own explorations, but you cannot change settings or export data. Every chapter in this book ends its practical sections with a short exercise you can do there.

From the Truth to Indicative Metrics

With the latest developments in data privacy, the implementation of consent modes, the widespread use of ad blockers, and the sampling and thresholding applied by tools such as GA4, digital marketers face a significant challenge. These changes have introduced a new landscape where data collection is not as comprehensive as it once was. Consequently, web analytics must now be viewed more as a compass guiding us through trends and patterns rather than a complete, detailed picture of user behavior.

Data privacy regulations, such as the GDPR and CCPA, have tightened the rules around how user data can be collected and used. Users are now more informed about their privacy rights and are exercising them, often opting out of tracking and data collection. This increased awareness and regulation mean that a large portion of user data is no longer available for analysis. At Lumen & Leaf, for example, only 62% of visitors accept analytics cookies; the behavior of the other 38% is either estimated or not recorded at all. Chapter 2 explains how consent works and how tools such as Consent Mode fill part of the gap.

Ad blockers add another layer of complexity to data collection. With the growing use of these tools, a significant segment of web traffic is effectively invisible to traditional analytics tools. Ad blockers prevent the tracking scripts from running, which means that the interactions of these users go unrecorded.

This results in a gap in the data, making it harder to get a complete view of user behavior and engagement.

Additionally, Google Analytics 4 (GA4) introduces its own approximations. Its standard reports are unsampled, but explorations and other ad-hoc analyses can be sampled once a query exceeds the property's quota, and GA4 applies thresholds that hide some data when user counts are small to protect privacy. Sampling can lead to approximations that, while useful, are not always entirely accurate. This introduces another limitation to the precision of the data collected, further emphasizing the need to view web analytics as a directional tool rather than a definitive source of truth.

A fourth gap has opened more recently: people who never arrive. When a search engine or an AI assistant answers a question directly, the visit that would once have appeared in your reports simply does not happen. You cannot measure a click that was never made, which is why the evidence on this shift, summarized below, matters to every analyst.

FROM THE RESEARCH

When the answer appears on the results page

In 2025, the Pew Research Center analyzed the real browsing activity of 900 US adults during March of that year—68,879 Google searches in total. About 18% of those searches produced an AI-generated summary at the top of the results. When a summary appeared, users clicked a traditional search result on 8% of visits, compared with 15% when there was no summary, and they clicked a link inside the summary itself on just 1% of visits. Users were also more likely to end their browsing session altogether after seeing a summary (26% of visits, against 16%).

What it means for you: a fall in organic traffic is no longer proof that your content or rankings got worse. Part of your audience is being served before it reaches you. Chapter 11 shows how to measure this shift—and the new traffic that AI assistants send.

Given these developments, web analytics can no longer be relied upon for a full, detailed view of what is happening on websites. Instead, it serves as a **directional tool**, providing insights and trends that can guide marketing

strategies. This compass-like role helps marketers understand general user behavior, identify broad trends, and make informed decisions based on the available data. However, it is crucial to recognize that this data is incomplete and should be interpreted with caution.

KEY CONCEPT

Directional data

Treat your analytics as a **compass**, not a map. It is reliable for spotting trends, comparing periods and ranking channels or pages against each other; it is not reliable for exact, auditable counts of every visitor.

In this new environment, the key is to adapt and find new ways to understand and engage with audiences. This might involve combining web analytics with other data sources, such as your order database, customer feedback, surveys and qualitative research, to fill in the gaps and create a more holistic view of user behavior. When Lumen & Leaf's GA4 property reports 2,160 purchases in a month, the store's order system is the authority on how many orders were actually placed; GA4's job is to explain where those customers came from and what they did on the way.

The shift towards privacy and the increasing use of ad blockers, along with the sampling and modeling built into GA4, are reshaping the landscape of digital marketing. While web analytics remains a valuable tool, its role has evolved. It is no longer the definitive source of truth but a guide that helps marketers navigate the complexities of user behavior in an era of incomplete data. Embracing this new perspective is essential for successfully adapting to the changing digital landscape.

GA4, the Inevitable

In October 2020, Google launched a new version of its web analytics service, Google Analytics 4 (GA4), which has become the default for Google Analytics (GA). This service collects data and analyzes web traffic for millions of businesses, helping them monitor marketing channels and measure Key

Performance Indicators (KPIs). GA4 introduces several features that distinguish it from its predecessor, Universal Analytics (UA), with the primary goal of analyzing user behavior and tracking their journey toward becoming customers. Universal Analytics stopped processing data on 1 July 2023 (1 July 2024 for the paid Analytics 360 version) and has since been switched off entirely, so GA4 is no longer a choice but the standard.

GA4 is built on the same platform as the "App + Web" system released by Google in 2019. This foundation allows marketers to track users across websites and apps in a single property, offering a holistic view of user interactions. This cross-platform tracking is critical as it enables businesses to understand how users transition between different platforms, which has historically been challenging.

The primary purpose of GA4, according to Google, is to offer a next-generation approach to AI-based predictive data, privacy-first tracking, and cross-platform measurement. Its machine learning (ML) models can predict user behavior and fill some of the gaps left when traditional data collection is limited by privacy choices and regulations like the California Consumer Privacy Act (CCPA) and the General Data Protection Regulation (GDPR). This makes GA4 particularly useful in an environment where website visitors are more likely to opt out of cookie usage and other data collection methods.

One of GA4's significant advantages is its integration with other Google tools, including YouTube, Google Ads and Search Console. This integration enhances the ability to evaluate advertising campaigns, measure view-through conversions and share audiences with your ad platforms. GA4 also offers free access to the BigQuery export, previously available only to paying Analytics 360 customers. This feature lets you copy your raw event data into Google's cloud data warehouse for deeper analysis—a capability Chapter 13 puts to work.

GA4 already offers predictive metrics such as purchase probability, churn probability and predicted revenue, which help businesses estimate the future value of their audiences and optimize their return on investment (ROI). Its enhanced measurement feature automatically tracks scrolls, outbound clicks, site search, file downloads, form interactions and engagement with embedded YouTube videos—no coding required—which makes it accessible to a much broader audience.