

Marketing Analytics in the AI Era

Marketing Analytics in the AI Era

*Running the AI-Augmented Desk
from Warehouse to Boardroom*

Paul S. Richardson

Copyright © 2026 Cambria Press

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means without written permission from the publisher.

About this book: This book was created the way it teaches. It was drafted in collaboration with artificial intelligence (Anthropic's Claude); every number in its worked sessions came from executed code, and every element was directed, verified, edited, and approved by the author, who — per the rerun rule taught in Chapter 1 — signs it.

ISBN 978-1-63857-445-3 (PDF)

ISBN 978-1-63857-456-9 (paperback)

Printed in the United States of America

Contents

Preface	1
1 The Analyst's Job and the AI-Augmented Desk	3
2 Data Wrangling and Quality: Making the Warehouse Worth Trusting	15
3 Descriptive Analytics and the Executive Dashboard: What Is Actually Happening Here	29
4 Segmentation: Finding the Customers Inside "Our Customer"	41
5 Targeting and Propensity Models: Who Gets the Offer, and How Would You Know You're Right?	53
6 Pricing and Elasticity: The Most Expensive Number You Will Ever Estimate	65
7 Forecasting: Planning Against a Future You Cannot See	77
9 Experimentation: The Campaign Is Live. Did It Work?	87
10 Driver Analysis and Regression: What Actually Moves Revenue?	99
11 Text Analytics and the Voice of the Customer: Listening at Scale	111
12 Attribution: The Credit Fight	123
13 Churn, Survival, and Customer Lifetime Value: The Goodbye Analysis	135
14 The Audit: Algorithmic Fairness in Marketing Systems	145
15 Synthesis and the Briefing: One Year, One Argument	155
A The Course Roadmap: How the Chapters, the Labs, and the Dossier Fit Together	165

B	The Dossier and the Briefing: A Specification	169
C	The Four Warehouses: Extended Case Narratives	173
D	The Analyst's Prompt Pattern Library	177
E	Careers in AI-Era Marketing Analytics: A Field Guide	181
F	Consolidated Course Glossary	185

Preface

This book is the analytics half of a story its companion volume, *Integrated Marketing Communications in the AI Era*, tells from the other side. In that course, students charter a one-person, AI-augmented agency and build a complete campaign for a realistic client. In this one, they take the desk down the hall: each student joins one of the same four companies as its marketing analyst, inherits its data warehouse, and spends fifteen weeks answering the questions every marketing organization actually asks — how many, how much, what worked, who is leaving, and what should we do about it.

The premise is the same one reshaping the industry the companion volume documents: fluent, confident, plausible analysis has become effectively free. The machine on the analyst's desk drafts working code in seconds, explains any method at any depth, and narrates any table it is shown — and it produces numbers by pattern rather than by lookup, with the same confidence either way. That combination changes what an analytics course owes its students. What they need is not another tour of methods with an AI sidebar, but practiced judgment at an AI-augmented desk: directing the machine, verifying its claims against data it cannot see, and signing only what a cell they executed actually produced. This book compresses that discipline into one law — the rerun rule: no number enters any deliverable unless it came from code you executed — and everything else here is commentary on it.

The method is the studio. Every chapter pairs with a lab in which the student analyzes their own warehouse — synthetic data generated from a personal seed, so that every warehouse shares the same structure and story but no two share the same numbers. Because the data is constructed, real effects are planted at known strength, and the traps working analysts fall into are planted alongside them: a leaky feature that makes a worthless model look miraculous, a vanity metric wearing a regression coefficient, a fluent collaborator scripted, once each week, to be helpfully and confidently wrong. Students who find the planted truth are right in a way real-world analysts rarely get to enjoy; students who accept the convenient finding discover, in the gentlest possible setting, exactly how that feels.

The arc runs from the warehouse to the boardroom: data quality, description, and segmentation; prediction, pricing, and forecasting; then causal measurement

— experiments, drivers, text, attribution, churn — and finally the audit and the briefing, where the semester’s skepticism is aimed at the analyst’s own work and a year of analysis becomes seven minutes of argument. Week eight is the midterm; there is no chapter eight, and the numbering honors the course calendar.

The clients here are the same four businesses the companion course serves, and the campaign this book’s students measure is the campaign those students built. The two courses are one universe, and the lesson travels in both directions: the maker learns that the work is graded by measurement, and the measurer learns that the numbers were made by someone. Both learn the only skill the machine cannot supply — the willingness to check, and to say plainly what checking found.

This volume is one of the *Marketing in the AI Era* series — four companion courses, one shared universe, each with its own studio. The complete series is at <https://www.marketingintheaiera.ai>.

Paul S. Richardson, PhD
Professor of Marketing
Holzschuh College of Business Administration
Niagara University

The Analyst's Job and the AI-Augmented Desk

Opening Scenario: Four Answers to One Question

It is 9:40 on your first Monday as the marketing analyst at Lakemont Roasting Co., and the CEO has asked you the simplest question in business: “How many subscribers do we have?”

You ask the systems. The e-commerce platform says 4,112. The email platform says 3,644. A spreadsheet the previous analyst left behind, last touched in March, says 4,020. And the AI assistant on your desk — asked casually, the way you’d ask a colleague — answers instantly and warmly: “Lakemont currently has approximately 4,300 active subscribers, reflecting strong recent growth.”

Four answers. Three of them came from systems that at least *touched* real data and disagree for knowable reasons — one counts anyone who ever subscribed, one counts only deliverable email addresses, one is stale. The fourth number came from nowhere at all. The assistant did not query a database. It produced a plausible-sounding figure the way it produces plausible-sounding everything, by pattern rather than by lookup, and it delivered the fabrication with the same confident fluency it uses for things that are true.

Here is what the analyst who keeps this job does next. She opens a notebook, writes four lines of code against the subscriptions table, defines “active” explicitly — a subscription spell with no end date as of today — and gets 3,806. Then she answers the CEO with a number *and its definition attached*: “3,806 active subscribers as of this morning, defining active as currently billing; the platform’s 4,112 includes 306 paused or expiring accounts, and I’d flag that our email list undercounts by about 160 addresses we should reconcile.” Ninety seconds of work. But notice everything inside it: a definition chosen and disclosed, a computation executed rather than recalled, the rival numbers explained rather than ignored, and a next action attached. That ninety seconds is this entire course in miniature — and the reason the first chapter of a book about AI-powered analytics begins with the machine getting a number *wrong*.

What a Marketing Analyst Actually Does

Strip away the job-posting language and a marketing analyst does one thing: **converts recorded behavior into defensible decisions**. Everything else — the dashboards, the models, the decks — is scaffolding for that conversion.

The conversion runs along a chain worth memorizing, because every failure you will ever diagnose lives at one of its links. **Data** — raw recorded events: an order row, an app open, a review — becomes **information** when organized into something readable: monthly revenue, churn by cohort, ratings over time. Information becomes **insight** when it explains something that matters: *why* churn spikes in January, not merely that it does. And insight becomes a **decision** when someone changes what the company does: promote the pause feature instead of discounting. Marketing analytics as an academic field studies this whole chain; marketing analysts as professionals are paid for the last link. Scholars who surveyed the field's explosion in data-rich environments reached the same conclusion from the research side: the binding constraint on analytics value is rarely the sophistication of the method — it is whether the analysis connects to a decision anyone can act on (Wedel and Kannan 2016). Firms that actually make decisions from data measurably outperform those that decorate decisions with it (Brynjolfsson and McElheran 2016).

Where does this person sit? Titles vary — marketing analyst, growth analyst, insights analyst, marketing data scientist — but the seat is remarkably consistent across companies: close enough to marketing to know what the campaign was *for*, close enough to the data to be trusted with it, and close enough to leadership to be asked the Monday-morning questions. In a large company you might be one of forty on an analytics team with data engineers upstream of you; at a company like Lakemont you are the entire department, which is precisely why this course embeds you at mid-sized clients — the solo seat is where every skill gets used weekly and nowhere to hide exists. A representative week: Monday, leadership wants last week's numbers contextualized before the ten o'clock; Tuesday, the email manager asks whether the subject-line test "worked" (it was underpowered; part of your job is saying so kindly); Wednesday, you are cleaning the data the new intern's export mangled; Thursday, the CFO wants next quarter's forecast defended cell by cell; Friday, you get two quiet hours for the churn analysis that might actually change the company — and the discipline to protect those hours is a career skill no one lists on the syllabus. Notice what dominates that week: not modeling. Definition, verification, translation, and judgment. The models appear for a few hours and borrow their value from everything around them.

One more feature of the seat deserves naming because students consistently underestimate it: the analyst is the company's **translator**. Upstream, you translate business anxiety into computable questions — "are we okay?" becomes "is

churn-adjusted revenue growth positive excluding gifting seasonality?” Downstream, you translate computation back into decisions — a hazard ratio becomes “promoting the pause feature should retain roughly one in three of the subscribers we currently lose to guilt-cancellation, and here is the test that would confirm it.” Both translations are graded in every lab in this course, because both are where analysts earn or lose their seat at the table. The best model in the company, described badly, loses to a bar chart described well — a sentence that should bother you into taking Chapter 3 seriously.

That is why this course grades everything you produce against **the decision standard**: an analysis that ends in a table is homework; an analysis that ends in a recommendation a non-technical decision-maker could act on — with its uncertainty honestly stated — is a job. Every lab deliverable in this course is a *decision memo*, not a printout, and the last section of every memo answers the only question executives actually have: *so what should we do?*

The work itself divides into three registers you will practice in sequence across the semester. **Descriptive analytics** establishes what is happening — the counts, trends, and distributions of Chapters 1 through 3, and the register where most real analyst hours are spent. **Predictive analytics** estimates what is likely — the segments, propensities, and forecasts of Chapters 4 through 7. **Prescriptive analytics** recommends what to do — the experiments, driver models, and optimization judgments of the back half of the course. The registers stack: a forecast built on miscounted subscribers predicts nonsense, and a recommendation built on a broken forecast prescribes it, which is why this book spends two full weeks on data itself before a single model appears.

Your Company, Your Warehouse, Your Seed

In this course you are not studying analytics; you are *employed in it*. In Week 1 you join one of four companies as its marketing analyst — Lakemont Roasting, Loop, Verdana, or Visit Chautauqua Trail — and every week your employer needs something: an orientation memo, a cleaned dataset, a segmentation, a price recommendation, a verdict on whether the big campaign worked. By Week 15 your accumulated memos, revised and bound, become your **Analyst's Dossier**, and you defend it in a live briefing to the client's leadership. If you have taken — or later take — this book's sibling course in marketing communications, you will recognize your employer: it is the same company, seen from inside. The campaign that course's students build is the campaign you will measure.

Your company's **data warehouse** is synthetic — constructed for this course — and that is not a limitation; it is the course's superpower, for three reasons you should understand on day one.

First, **your warehouse is yours alone**. It is generated from your personal seed — a code derived from your identity — so every student's data shares the

same structure and story but *different numbers*. Your true churn rate is not your neighbor's. This means your answers cannot be copied, a fact with a pleasant flip side: your findings are genuinely yours, discovered by you, in data no one else has ever analyzed.

Second, **the truth is in there**. Because the data was constructed, real effects were planted in it at known strength — a genuine driver of churn, a real segment hiding in the customer file, a true campaign lift. When you find them, you are right in a way real-world analysts rarely get to enjoy; when your method is sloppy, the data was built so that sloppiness finds something *specifically wrong*. Treat every convenient finding with suspicion. Some of them were left there for you.

Third, **the warehouse agrees with the company's official story**. Your client's brand brief — the strategy document with the company's stated figures — reconciles with your warehouse: the brief says roughly 3,800 subscribers and 9.5 percent monthly churn, and honest analysis of your data will find numbers near those. When your analysis and the official story diverge beyond tolerance, the first suspect is your query, not the world. Professional analysts call this **reconciliation**, and it is the fastest error-detector you own.

Your map to all of it is the **data dictionary** — a table, inside the warehouse itself, that documents every table and column. It is the single authoritative source for what exists, and reading it *first* — before profiling, before modeling, before asking the AI anything — is the professional habit Lab 1 installs. Read it the way a lawyer reads a contract: note the **grain** of each table (what one row *is* — a customer? an order? a subscription spell?), because grain errors are the parent of half of all analytical mistakes; note which columns are facts and which are definitions in disguise (an `active_current_month` flag embodies someone's definition of active — whose, and does it match yours?); and note what is *absent*, because the questions a warehouse cannot answer are as strategically important as the ones it can. The dictionary sounds bureaucratic until you meet the alternative, which is the subject of the next section.

A brief aside on the claim log, because it changes character in this course. In the sibling communications course, the log tracks claims as verified, refuted, or assumption-flagged — a judgment discipline. Here it becomes nearly mechanical, and more powerful for it: a claim about your data is **verified** when a notebook cell produced it this session, **corrected** when a cell contradicted what you or the machine expected, and an **assumption** when you chose a definition the data could not choose for you (“active means currently billing”). Every memo you submit carries its log, and the assumptions row is the one experienced readers check first — because in analytics, the assumptions are where the argument actually lives.

The AI-Augmented Desk

Now the second character in this book. Alongside the warehouse, your desk has an AI collaborator — in the Studio, the Coach — and this course's method is learning to run the two together: the machine that talks and the data that doesn't.

Be precise about what the machine is. A large language model is a system trained on text to produce fluent language — including fluent *code*, fluent *statistical narration*, and fluent *numbers*. Its fluency with the first two will transform your productivity this semester: it drafts working pandas code in seconds, explains any statistical concept at whatever depth you ask, critiques your analysis plan, and translates findings into executive prose. Field research on generative AI at work finds exactly this shape of gain, with the largest lifts going to the least experienced workers — the “floor rises” (Brynjolfsson, Li, and Raymond 2025). For a student analyst, the floor rising is the whole game: you arrive with modest coding skill, and the machine supplies the syntax while you supply the judgment.

But its fluency with *numbers* is a trap with a specific mechanism you must understand rather than merely fear. The model does not look numbers up; it *generates* them, token by token, the same way it generates adjectives. Asked about your data, it produces figures shaped like the truth — plausible magnitude, confident phrasing — with no connection to your warehouse whatsoever, exactly like the 4,300 in the opening scenario. It will also “remember” the structure of your data with the same false fluency: asked to describe your tables, it may confidently include a column that does not exist, because tables like yours *usually* have one. Researchers mapping where AI helps and where it silently fails call this boundary the **jagged frontier** — capability is uneven in ways that are invisible from the output's confidence (Dell'Acqua et al. 2023). On one side of the frontier, the machine writing your code is a genuine expert; on the other, the machine reciting your numbers is a fluent fabulist. The two sound identical.

Learn the machine's failure modes by name, because each has a different tell and a different countermeasure. **Fluent fabrication** is the opening scenario's 4,300: a specific number produced by pattern, not lookup — countered by the rerun rule absolutely. **False structure memory** is the phantom column: asked about your tables, the model describes tables *like* yours, including elements yours lacks — countered by the data dictionary, and drilled in Lab 1. **The silent assumption** is subtler: asked for churn, the machine's drafted code quietly picks a definition (perhaps counting paused subscribers as churned) without announcing the choice — the code runs, the number is real, and it is still wrong for your question; countered by reading every draft's definitional moves before running it, and by asking the machine directly “what did you assume?” — a question it answers honestly and never volunteers. **Agreeable statistics** is the most dangerous in the back half of this course: presented with your eager hypothesis

and an underpowered result, the model will find the encouraging interpretation, because agreement is fluent and challenge is not; countered by pre-registration and by explicitly briefing the machine to attack your finding — a red-team role it performs superbly when directed and abandons the moment you sound pleased. Four modes, one common thread: none of them announce themselves, and all of them survive on your inattention.

The course's answer is one law, and everything else in this book is commentary on it:

The rerun rule. *No number enters any deliverable unless it came from code you executed. Not from the model's memory. Not from your memory. From a cell, run by you, this session.*

The rule sounds austere and is actually liberating, because it assigns each party the work it is good at. The machine may *propose* any number it likes — as a hypothesis, an expectation, a sanity check. You may *report* only what ran. In practice the division of labor at an AI-augmented desk looks like this: you **direct** (state the business question and what would count as an answer), the machine **drafts** (the code, the approach, the caveats you didn't think of), you **run** (the Studio Notebook executes against your actual warehouse), you both **read** (paste the output back; the machine is excellent at interpreting results it can finally see), and you **decide** (the recommendation is yours, because the accountability is). Notice where the machine's memory appears in that loop: nowhere.

Three working practices operationalize the loop, and Lab 1 drills all three. **Brief like a director:** “analyze my data” gets you a wandering essay; “here are my tables per the data dictionary; I need active subscribers defined as currently billing; draft the pandas and tell me what edge cases my definition misses” gets you a colleague. **Predict, then run:** before executing anything, ask the machine — or yourself — what the result *should* be, roughly; the reconciliation between expectation and output is where errors announce themselves. **Log your claims:** this course's claim log tracks every factual assertion in your deliverable as verified by execution, corrected, or flagged as an assumption. In an analytics course the log has teeth the humanities never dreamed of: a claim either has a cell behind it or it does not.

The Desk in Action: One Question, Worked End to End

Theory earns its keep in the transcript. Here is the direct-draft-run-read-decide loop on a real question at the Lakemont desk — the analyst's actual chat and notebook, annotated. (The numbers are one analyst's seed; yours will differ, which is rather the point.)

Analyst: I need our monthly churn rate for the leadership deck. Per the data dictionary I have subscriptions (one row per subscription spell; end empty means active) and monthly_summary (month, active_start, churned, added). Before you write anything: what definitional choices am I facing, and what would you expect churn to be for a craft-coffee subscription business?

Coach: Three choices to make explicitly. Denominator: churned divided by starting actives is standard; average-of-period is defensible; ending actives flatters. Window: a single month is noisy — a trailing twelve-month mean with the range disclosed is deck-appropriate. Paused subscribers: your dictionary shows a pause feature; decide whether paused counts as churned (I'd say no — paused is retained, and the distinction may matter strategically here). Expectation: subscription D2C coffee typically churns mid-single to low-double digits monthly; your brand brief states 9.5 percent, so if your data says 4 or says 18, suspect the query before the world.

☛ *Everything good about the exchange is in the analyst's prompt: tables named with their grain, the dictionary cited, and — the professional move — asking for definitional choices and an expectation before any code. The Coach's brief-anchored prediction (9.5 percent) is the reconciliation target; note it may state that figure because it comes from the brief, not from the data it cannot see.*

The notebook, next:

```
last12 = monthly_summary.tail(12)
churn_monthly = last12["churned"] / last12["active_start"]
print(f"Trailing-12 mean: {churn_monthly.mean():.1%}")
print(f"Range: {churn_monthly.min():.1%} to {churn_monthly.max():.1%}")
print(last12.set_index("month")["churned"].to_string())
Trailing-12 mean: 9.6%
Range: 8.1% to 13.2%
2025-09      341
2025-10      355
...
2026-01      492
...
2026-07      478
2026-08      360
```

☛ *Three lines of executed truth. The mean reconciles with the brief (9.6 against 9.5 — inspect nothing), and the monthly detail has already yielded a bonus finding: January and July spike visibly. The analyst pastes this output back into the chat — the machine cannot see the notebook, and interpreting results it has finally been shown is where it shines.*

Analyst: Output pasted above. Mean 9.6 percent, but January hit 13.2 and July 12.1. Reading?

Coach: Reconciles with the brief — report it with confidence. The spikes are your lead, not your footnote: January suggests post-holiday budget cancellations (consistent

with the “guilt splurge” pattern in your brief’s exit surveys), July suggests travel-season pausing demand being met by cancellation instead — worth checking whether pause-feature usage is as low as the brief implies. For the deck: one line on the rate, one chart on the spikes, one sentence proposing the pause-promotion test. Want me to draft that slide text against your voice?

☛ *And there is the whole loop closing into the decision standard: a rate, its seasonal structure, and a testable recommendation — in one question’s worth of work. The analyst will verify the pause-usage claim with another cell before the deck ships, because the Coach’s “consistent with your brief” is a hypothesis wearing a helpful tone. Total elapsed time at a practiced desk: under ten minutes. Total numbers taken from anyone’s memory: zero.*

Ethics Sidebar: The Number You Report Is a Promise

Every chapter of this book carries an ethics sidebar, and the series begins with the foundation the rest stand on: **reporting a number is making a promise**. When your memo says churn is 9.7 percent, you are promising that a defensible definition, correctly executed, on trustworthy data, produced that figure — and that you would show your work to anyone who asked. Most analytical dishonesty in real organizations is not fabrication; it is promise-inflation. The number was a churn rate — under the flattering definition, on the convenient subset, from the one query that wasn’t checked. The habits this chapter installs — definitions disclosed, code executed, rivals reconciled, uncertainty stated — are not academic niceties. They are what makes your signature on a memo mean something, and in this course you sign everything. The machine, meanwhile, makes no promises at all: it is not lying when it says 4,300, because it is not asserting; it is completing a pattern. *You* are the only party at your desk capable of honesty — which is another way of saying you are the only one responsible for it.

Studio Lab Briefing — Lab 1: Meet the Warehouse

Open in Analytics Studio → Lab 1. Due before Session 2. Your workspace loads the Coach; below the chat, the Studio Notebook generates your personal warehouse in your browser — no installation, nothing to configure. (Your course also provides a Colab fallback notebook if your machine struggles.)

Part A — Calibration drills (graded from transcript and notebook). (1) *The tour*: read the `data_dictionary` first, then profile every table — rows, columns, date coverage, one surprise — with every claim in chat backed by a cell. (2) *The false memory*: ask the Coach to describe your tables from memory, then audit its answer against the dictionary and your profiling; somewhere in its confident recollection is at least one thing that does not exist. Find it, correct it with evidence,

and say what the catch teaches. (3) *The first number*: your manager wants the size of the business in one figure a stranger would understand; compute three candidates, choose, and defend the choice.

Part B — The deliverable. One page: your **analyst's charter** (your working principles as director, runner, and translator, plus one professional goal) and your **data-orientation memo** (the warehouse table by table with row counts from your cells, the date range, your chosen size-of-business number with one sentence of defense, and the two most valuable questions this data could answer for this company this year).

Start your working document now — plain text or markdown, kept beside the Studio all semester. Every lab's memo joins it; by Week 13 it is your Dossier draft.

Rubric emphasis (lab). Editorial Judgment 30 percent — the false-memory catch and the rerun rule are the points; a deliverable containing any number that appears nowhere in your notebook caps this dimension. Craft Quality 25 · AI Collaboration Skill 25 · Strategic Fit 20, with the last rewarding “valuable questions” that connect to *your* company's stated strategic problem rather than to any company's generic ones.

Review Questions

1. In the opening scenario, three systems and one AI produced four different subscriber counts. Classify each of the four failures by its cause, and explain why the AI's failure is categorically different from the other three.
2. State the rerun rule and explain what it permits the machine to do with numbers. Why is “propose but not report” the right boundary rather than banning machine numbers entirely?
3. Walk the data-information-insight-decision chain using January churn at a subscription coffee company. Give one example of a failure at each link.
4. Your warehouse is synthetic, seeded personally to you, and reconciled to your client's brief. Explain how each of those three properties changes what honest work means in this course.
5. A classmate says: “The AI wrote my analysis code, so the analysis isn't really mine.” Using the direct-draft-run-read-decide loop, argue where authorship actually lives — and name the one step where outsourcing *would* forfeit it.
6. In the worked session of Section 1.5, identify every definitional choice the analyst made or accepted, every reconciliation performed, and the one Coach claim that still requires a verifying cell before the deck ships. What does the session's final annotation mean by “a hypothesis wearing a helpful tone”?

Key Terms

- **Claim log** — The course's running record of every factual assertion in a deliverable: verified by execution, corrected, or flagged as an assumption.
- **Agreeable statistics** — The AI failure mode of finding the encouraging interpretation of weak evidence because agreement is more fluent than challenge; countered by pre-registration and directed red-teaming.
- **Data dictionary** — The authoritative in-warehouse documentation of every table and column; the standard against which all recollection, human or machine, is checked.
- **Data-information-insight-decision chain** — The conversion an analyst is paid for: recorded events, organized into readable form, explaining something that matters, changing what the company does.
- **Decision standard** — The course's requirement that every analysis end in a recommendation a non-technical decision-maker could act on, with uncertainty stated.
- **Descriptive, predictive, and prescriptive analytics** — The three registers of the field: what is happening, what is likely, and what to do about it.
- **Direct-draft-run-read-decide loop** — The working rhythm of the AI-augmented desk: the analyst directs, the machine drafts, the analyst runs, both read the output, the analyst decides.
- **Embedded analyst** — This course's operating role: the student as the marketing analyst inside a client company, producing the memos the company needs, week by week.
- **Fluent fabrication** — A language model's production of plausible, confident numbers or facts by pattern-completion rather than lookup; the failure mode the rerun rule exists to contain.
- **Grain** — What one row of a table represents; the first thing to establish about any dataset, because grain errors parent half of all analytical mistakes.
- **Jagged frontier** — The uneven, invisible boundary between tasks where AI performs expertly and tasks where it fails confidently.
- **Reconciliation** — Checking an analytical result against an independent source — the brief, another system, a prior period — and treating divergence as a signal to inspect the query first.
- **Silent assumption** — The AI failure mode of drafted code embodying an unannounced definitional choice; the number produced is real and still wrong for the question.
- **Rerun rule** — The course's central law: no number enters a deliverable unless it came from code the analyst executed, this session, in the notebook.

- **Seed** — The personal code from which each student's warehouse is deterministically generated; the reason every student's numbers are their own.
- **Studio Notebook** — The in-browser Python environment where the warehouse generates and all analysis runs; the home of the rerun rule.
- **Translator role** — The analyst's two-way conversion: business anxiety into computable questions upstream, computation into actionable decisions downstream.
- **Synthetic data warehouse** — A constructed, internally consistent company dataset with effects planted at known strength, reconciled to the client's official brief.

References and Further Reading

Scholarly foundations

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. *The Quarterly Journal of Economics*, 140(2), 889–942. — *Large-scale field evidence that AI assistance lifts less-experienced workers most: the “floor rises” argument of Section 1.4.*

Brynjolfsson, E., & McElheran, K. (2016). The rapid adoption of data-driven decision-making. *American Economic Review*, 106(5), 133–139. — *The firm-level evidence that deciding from data, not merely having it, is what pays; the backbone of the decision standard.*

Dell’Acqua, F., McFowland, E., Mollick, E., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Kraymer, L., Candelon, F., & Lakhani, K. (2023). *Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality.* Harvard Business School Working Paper 24-013. — *The origin of Section 1.4’s jagged-frontier framing: capability uneven, confidence constant.*

Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing*, 80(6), 97–121. — *The field’s landmark synthesis: what marketing analytics comprises and why its value binds at the decision, not the method.*

Industry and professional sources

Davenport, T. H., & Harris, J. G. (2017). *Competing on analytics: The new science of winning* (updated ed.). Harvard Business Review Press. — *The managerial classic on analytics as competitive strategy; Chapter 1’s “conversion, not decoration” framing in book-length form.*

McKinsey & Company. (2025). *The state of AI (global survey).* — *Adoption context for the desk you are joining: the overwhelming majority of organizations now report using AI somewhere in the business; the analyst’s question is no longer whether but how well.*

