Appendix B

Common Marketing Metrics

Dr. Jose Mendoza, Academic Director and Clinical Associate Professor
Version 1.1 · August 2026

Appendix Information

PURPOSE

This appendix is the guide's metric dictionary. It collects, in one place and in one format, every quantity the chapters and the two projects compute, so that a metric is fully defined once rather than defined differently in several places. Each entry states the business question the metric answers, its formula, its numerator and denominator, its window, its filters, its grain, its unit, what a higher or lower value means, the mistake most likely to be made with it, and a check it must survive.

USE THIS APPENDIX WHEN

Consult this appendix when you are about to compute a named metric, when two figures bearing the same name disagree, when you need to state a definition in a notebook, a dashboard footer, or a project methods section, or when you are checking whether a number someone has handed you means what its label says.

THIS APPENDIX DOES NOT

This appendix is a definition catalog. It does not teach the methods that produce the measures, does not attempt to catalog every metric used in marketing practice, and does not treat every metric as a key performance indicator. It catalogs only what the guide, the labs, the exercises, and the two projects actually use; Section B.14 records what was considered and left out, and why. Where a metric raises a question of method rather than definition, the entry names the chapter that owns the decision and stops there.

VERSION AND DATE

Version 1.1 · August 2026 · Language: English (United States)

SUGGESTED CITATION

Mendoza, J. (2026). Common marketing metrics. In *Applied business analytics for marketing decision-making: Business analytics and data visualization* (Appendix B, Version 1.1) [Open educational resource]. CC BY 4.0.

LICENSE AND RIGHTS

for identification only and do not imply endorsement. StyleCraft Collective is a fictional company created for instruction.

GENERATIVE AI USE

Generative artificial intelligence and other AI-assisted tools were used in the research, writing, revision, and production of this appendix, including outlining, preliminary drafts, revision of prose, support for code and analytical examples, and document formatting. These tools were used under the author's direction and are not credited as authors, researchers, or sources. The author determined the appendix's scope, boundaries, and content, and reviewed and approved all AI-assisted material: factual claims and citations were checked against the underlying sources rather than accepted from AI-generated summaries, and every formula and worked example was independently recomputed using either the appendix practice file or the published figures of the chapter that owns the metric, with each stated result confirmed. Responsibility for the accuracy, originality, and final form of this appendix rests entirely with the author. A fuller statement appears in the front matter of the complete guide.

COMPANION FILES

A machine-readable metric dictionary carrying every field of every entry below, a printable metric-definition form, tested Python functions for the most repeated metrics, a changelog, and the practice file these worked examples run on are in the Applied Business Analytics companion repository.

How to Read an Entry

Entries in this appendix are organized by metric family, not alphabetically, because the mistake this appendix exists to prevent is confusing two metrics that sit near each other rather than failing to find one. Section B.14 supplies the alphabetical index and names the metrics that were considered and deliberately excluded.

Every entry uses the same fields, in the same order. Table B.1 lists them. The first six are the elements that decide what number comes out; the last five are what a reader needs in order to use the number without misreading it. A field that a source genuinely does not specify is not left blank and is not guessed at; it is the analyst's to decide and to record.

Table B.1

The fields every metric entry carries, and what each one settles

Field	What it settles
Business question	The decision or comparison the metric supports. A metric with no decision behind it cannot be judged fit for anything.
Formula	Plain language first, then the code or symbolic form where precision requires it.
Numerator	The exact event, amount, or count included.
Denominator	The exact eligible base. For a total rather than a rate the field reads "none", which is itself information.
Window	The period and the boundary rule. A trailing window needs a start and an as-of date.

Field	What it settles
Filters	Channel, customer, campaign, product, and eligibility rules, applied identically to numerator and denominator.
Grain	What one computed observation represents. This is the fifth element the four-element discipline of Section 3.5.1 leaves implicit, and the one Appendix A promises is stated here.
Unit	Dollars, count, proportion, percentage, percentage points, days, or ratio. Percentage and percentage points are different units.
Required fields	The minimum data elements the computation needs.
Interpretation	What a higher or lower value means, and what the metric does not establish.
Common failure	The definitional or interpretive mistake most likely to be made.
Verification	A hand check, a reconciliation, or a bound the metric must satisfy.
Common variants	Present only where practice legitimately differs, so that a defensible alternative is named rather than silently overruled.
Owned by	The chapter that introduces and teaches the measure.

CONCEPT

One Definition, In One Place

A metric exists only as its definition. Two analysts who apply different definitions to the same data are not making an arithmetic error; they are computing different quantities, and no amount of recomputation will reconcile them. This appendix is the guide's single copy of each definition, and code, dashboards, and prose implement the entry rather than inventing a second version of it.

That rule has a consequence worth stating plainly. When a chapter, a lab, or a project needs a metric this appendix does not carry, the answer is not to invent one locally. It is to write the entry, using the form in Section B.1, and record it where the people who will disagree can see it.

Source: Course concept developed for this guide, informed by Farris et al. (2010) and Kimball and Ross (2013).

Worked examples in this appendix run on the same practice file that ships with Appendix A: twenty order lines, twelve orders, eight customers, and $1,312.00 of revenue, shaped exactly like StyleCraft's transactions table and small enough to check with a calculator. Where a family has no natural miniature in that file — media, funnel, classification, forecasting, and experiments — the example reuses the figures the owning chapter already publishes, so that a number quoted here can be traced back to the page it came from. No number in this appendix answers a graded exercise.

B.1 How to Use This Metric Dictionary

Chapter 3 established that a number is a claim: every reported figure is shorthand for a longer sentence naming a definition, a unit of analysis, a window, and a set of filters. This appendix writes that longer sentence down, once, for every metric the guide uses. This section explains

how to read the catalog, how to cite an entry, and what to do when a project needs a metric the catalog does not carry.

B.1.1 A metric is not a KPI

A metric is a quantity computed from data by an explicit rule, at a stated grain, for description or comparison. A key performance indicator is a metric an organization has elevated: attached to a target, assigned an owner, reviewed on a cadence, and connected to a decision (Parmenter, 2015). The distinction is not pedantry. An organization can compute hundreds of metrics and attend to only a handful, and calling every metric a KPI inflates a dashboard until nothing on it is key. Elevation also carries a cost this appendix cannot remove: once a measure becomes a target, behavior reorganizes around the measure and the measure degrades as evidence (Campbell, 1979). And because data quality is fitness for a use rather than a property of a file, a metric with no decision behind it cannot be judged fit for anything (Wang & Strong, 1996).

Every entry in Sections B.2 through B.12 is a metric. Section B.13 is different in kind: it catalogs the measures that exist only because a decision elevated them — a coverage threshold, a fixed-cost line, a forecast interval expressed in store-equivalents — and it explains the apparatus by which any entry in this appendix becomes a dashboard KPI. It creates no new formulas.

B.1.2 Five elements, not four

Section 3.5.1 requires that a ratio metric name its numerator, its denominator, its time window, and its filters. This appendix adds a fifth element to every entry, ratio or not: the grain. The addition is not a refinement of Chapter 3's rule but a consequence of it, and Appendix A commits to it by name — its entry for average order value points here precisely because the grain is stated here.

The reason is that a denominator often conceals a grain rather than revealing it. Average order value divides revenue by orders, and the metric's grain is the order; return on ad spend divides revenue by dollars of spend, and its grain is the campaign, not the dollar. Where the two come apart, an analyst who has checked only the four elements can still compute a defensible-looking number at the wrong grain. Chapter 12 reports the consequence in its most expensive form: an average line value of $50.41 and a true average order value of $108.00, on the same file, with no error message and no visual difference between the correct chart and the incorrect one.

B.1.3 Why two correct formulas disagree

The commonest failure this appendix is designed to prevent is not miscalculation. It is two people computing different quantities under one name and then arguing about arithmetic. Chapter 3 stages the case: three StyleCraft teams report repeat purchase rates of 60, 43, and 20 percent, and none of them has made a mistake. Table B.2 reproduces the three definitions from the eighteen-order preview the chapter computes them on, and shows where they part company.

Table B.2

Three correct repeat purchase rates, and the elements that separate them

Element	Lifetime (Retail)	Trailing 12 months (CRM)	Store only (Finance)
Numerator	Customers with 2+ orders ever	Customers with 2+ orders in the window	Customers with 2+ store orders in the window
Denominator	All customers with 1+ completed order in the available history	Customers with 1+ order in the window	Customers with 1+ store order in the window
Window	Entire history to the analysis date	Trailing 12 months to the analysis date	Trailing 12 months to the analysis date
Filters	All channels	All channels	Store channel only
Grain	Customer	Customer	Customer
Numerator count	6	3	1
Denominator count	10	7	5
Value	0.60	0.43	0.20

Read across the value row and the dispute looks like a data-quality problem. Read down the denominator row and it stops being mysterious: the three rates are computed over ten, seven, and five customers respectively, because each definition admits a different population. The organization does not need to decide which number is true. It needs to decide which definition serves the decision, record it, and report the others beside it under their own names. Section B.4's entry for repeat purchase rate carries the outcome of that decision, with the two alternatives named as variants.

B.1.4 Formula and notation standards

Every entry follows the same conventions, and a metric defined outside this appendix should follow them too.

Plain language leads. Each formula is stated in words first and in symbols or code only where precision requires it. A formula a stakeholder cannot read aloud is a formula nobody will audit.

The unit is named, and percentage points are not percentages. A rate moving from 20 to 24 percent rose by four percentage points and by twenty percent. Both are true and they are not the same claim; Chapter 13 names the practice of choosing between them for effect as framing.

Rates are recomputed from pooled totals, never averaged. The mean of group means is not the overall mean unless the groups are the same size. Where the unweighted average is genuinely wanted, it is labeled as one and its difference from the pooled figure is reported.

A rate over an empty base is undefined, not zero and not infinite. A campaign with no clicks has no cost per click. A table of infinities is a denominator problem wearing a printing error.

Rounding is a display decision and never a computational one. Formulas carry no implied rounding; figures are rounded when they are shown, and the rounding is not carried back into a subsequent calculation.

A classification metric travels with its base rate; a growth metric travels with its comparison period; an attributed metric travels with its attribution window. Each of these is the element without which the number cannot be read at all.

Causal verbs are earned, not chosen. Entries state what a measure does not establish. An attributed return is not an incremental one, a predicted probability is not a treatment effect, and an adjusted difference is not a causal effect.

B.1.5 Citing a definition, and adding one

A metric quoted outside this appendix carries a pointer to it: in a notebook, a comment naming the entry; in a dashboard, a footer naming the source, window, filters, and grain; in a project methods section, the entry key and the version of this appendix. The companion dictionary ships the same fields in machine-readable form, which is what an AI assistant should be given as context rather than being asked to supply a definition of its own.

The discipline is not peculiar to this guide. A public company presenting a metric is expected to define it, to say why it is useful and how management uses it, and to disclose the effects of any change in how it is calculated (U.S. Securities and Exchange Commission, 2020). A project deliverable owes its reader the same three things.

When a project needs a metric this appendix does not carry, do not modify an entry. Write a new one on the form in Table B.3, record it in the project's own dictionary, and note which existing entry it resembles and how it differs. A project-specific metric that quietly redefines a cataloged one is the Chapter 3 dispute reproduced at project scale.

Table B.3

The metric definition form, for a metric this appendix does not carry

Field	Your entry
Metric name	
Business question	
Formula	
Numerator	
Denominator	
Window	
Filters	
Grain	
Unit	
Required fields	
Interpretation	

Field	Your entry
Common failure	
Verification	
Owner and decision use	
Nearest cataloged entry, and how this differs	

B.1.6 The practice file the examples run on

Worked examples in Sections B.2 through B.4 and B.8 run on the practice file described above, rolled up from twenty order lines to the twelve orders in Table B.4. Every figure in those examples can be recovered from this table with a calculator, which is the point of keeping it small. The digital orders carry the reserved store identifier ONLINE and therefore have no store type, which is why they appear as a third group rather than being dropped.

Table B.4

The practice file at order grain, with its certified totals

Order	Customer	Store type	Revenue	Cost	Units	Lines
O10001	C001	Urban	96.00	48.00	2	2
O10002	C003	Digital	117.00	63.00	3	2
O10003	C002	Suburban	248.00	124.00	3	3
O10004	C004	Digital	84.00	42.00	2	1
O10005	C001	Urban	78.00	39.00	2	1
O10006	C005	Digital	76.00	44.50	2	2
O10007	C002	Suburban	192.00	96.00	3	3
O10008	C006	Digital	68.00	34.00	2	1
O10009	C001	Urban	60.00	30.00	2	1
O10010	C003	Digital	112.00	64.00	2	2
O10011	C007	Digital	88.00	44.00	2	1
O10012	C008	Digital	93.00	46.50	3	1
Total	**8 customers**		**1,312.00**	**675.00**	**28**	**20**

B.2 Revenue, Margin, and Volume Metrics

Every metric in this family is an amount or a count rather than a rate, which makes them the easiest to compute and among the easiest to misreport. Their failure mode is not a wrong denominator but an unstated one: a revenue total whose window, filters, and returns treatment nobody wrote down cannot be reconciled against a second total that used different ones.

B.2.1 Line revenue

Question. What did one order line contribute to net sales? **Formula.** quantity × unit_price × (1 − discount_pct); stored as line_revenue and recomputed rather than trusted.

Numerator. The extended, discounted value of a single order line. **Denominator.** None — it is an amount, not a rate. **Window.** The order's order_date; no window of its own. **Filters.** Certified lines only; a line whose quantity, unit_price, or discount_pct is unresolved has no line revenue and must stay missing rather than becoming zero. **Grain.** One order line. **Unit.** Dollars. **Required fields.** quantity, unit_price, discount_pct.

Interpretation. The building block every revenue metric in this appendix aggregates. It is net of line-level discount and gross of returns, credits, and refunds. **Common failure.** Summing a column in which unresolved lines were filled with 0.00. pandas sum() skips missing values by default, so a total computed over eleven lines of which one is unresolved is the revenue of ten lines wearing the label of eleven. **Verification.** Recompute from quantity, unit_price, and discount_pct and compare with the stored column; then confirm the count of unresolved lines is zero before quoting any total.

Common variants. Where returns are netted at line level, the same name denotes a different quantity. State whether returns and adjustments are inside the number.

Owned by. Chapter 4, Section 4.8.

B.2.2 Gross sales

Question. What would the period have billed at catalog prices, before discount? **Formula.** Sum of quantity × unit_price across the included lines.

Numerator. Extended catalog value of every included line. **Denominator.** None — it is a total. **Window.** Stated period, both ends inclusive. **Filters.** Stated channel, store, and product filters. **Grain.** Whatever population is stated; commonly chain, store, or campaign. **Unit.** Dollars. **Required fields.** quantity, unit_price.

Interpretation. Read only beside net revenue. Alone it overstates what the business earned by exactly the discount given. **Common failure.** Reporting gross sales as 'revenue'. The gap between the two names is the whole promotional story, and collapsing it hides the discount dependence Chapter 6 measures. **Verification.** Gross sales less discount dollars must equal net revenue exactly.

Owned by. Appendix B; the book uses net revenue throughout.

B.2.3 Discount dollars

Question. How much price was given away to produce this revenue? **Formula.** Gross sales − net revenue; equivalently sum of quantity × unit_price × discount_pct.

Numerator. Catalog value forgone on the included lines. **Denominator.** None — it is a total. **Window.** Stated period. **Filters.** Stated filters; lines with discount_pct of 0 contribute nothing. **Grain.** Stated population. **Unit.** Dollars. **Required fields.** quantity, unit_price, discount_pct.

Interpretation. Rising discount dollars against flat units is a margin story, not a demand story. **Common failure.** Confusing discount dollars with markdown at retail, which is an inventory-accounting quantity computed against a different base. **Verification.** Must be non-negative, and must equal gross sales less net revenue.

Owned by. Appendix B.

B.2.4 Net revenue

Question. What did the business actually earn from these orders? **Formula.** Sum of line_revenue across the included lines.

Numerator. Sum of line_revenue. **Denominator.** None — it is a total. **Window.** Stated period, both ends inclusive; a trailing window needs a start and an as-of date. **Filters.** Stated channel, store, customer, and campaign filters. **Grain.** Stated population — chain, store, metro, customer, campaign, or day. **Unit.** Dollars. **Required fields.** line_revenue, order_date, plus whatever the filters name.

Interpretation. The book's default meaning of the bare word revenue. It is net of line discount and does not net returns. **Common failure.** Quoting a revenue total whose window, filters, and returns treatment are unstated, so two correct totals disagree and neither can be reconciled. **Verification.** Reconcile to the certified source total to the cent. An aggregation redistributes rows and must never create or destroy dollars.

Common variants. Recognized revenue as finance defines it may net returns, cancellations, and gift-card breakage. Where the dashboard sits beside a finance report, say which is being shown.

Owned by. Chapter 4, Section 4.8.

B.2.5 Line cost

Question. What did the goods on one order line cost to sell? **Formula.** quantity × unit_cost.

Numerator. Extended variable cost of a single order line. **Denominator.** None — it is an amount. **Window.** The order's order_date. **Filters.** Certified lines only. **Grain.** One order line. **Unit.** Dollars. **Required fields.** quantity, unit_cost.

Interpretation. What line_cost carries decides whether the difference below it is contribution or gross profit. In the certified extract it carries merchandise cost, inbound freight, payment processing, and pick-and-ship. **Common failure.** Assuming the cost column is fully loaded when it carries merchandise cost alone. If it does, every quantity below is gross profit and must be relabeled. **Verification.** Confirm with finance what line_cost contains before naming anything computed from it.

Owned by. Chapter 13, Section 13.12.1.

B.2.6 Contribution

Question. What is left after the variable cost of selling, before the fixed footprint? **Formula.** Sum of (line_revenue − line_cost).

Numerator. Net revenue less variable cost on the included lines. **Denominator.** None — it is a total. **Window.** Stated period. **Filters.** Stated population; the go/no-go dashboard computes it for suburban and resort stores only. **Grain.** Stated population; the dashboard uses one store-week. **Unit.** Dollars. **Required fields.** line_revenue, line_cost.

Interpretation. Revenue after variable cost and before the fixed footprint. It is the quantity a store must generate enough of to cover its own rent. **Common failure.** Calling it contribution when line_cost carries merchandise cost alone. That quantity is gross profit, and every sentence using the word contribution is then wrong. **Verification.** Must be less than net revenue on the same population. Recompute on one store-week by hand from the underlying lines.

Common variants. Contribution after occupancy, and contribution after allocated overhead, are different and larger deductions. Name the deduction, not just the word.

Owned by. Chapter 13, Section 13.12.1.

B.2.7 Contribution margin

Question. What share of each revenue dollar survives variable cost? **Formula.** Sum of (line_revenue − line_cost) / Sum of line_revenue, on the same cost definition.

Numerator. Contribution on the stated population. **Denominator.** Net revenue on the same population. **Window.** Stated period. **Filters.** Same population for numerator and denominator, without exception. **Grain.** One population, not one line — it is a pooled ratio. **Unit.** Percentage. **Required fields.** line_revenue, line_cost.

Interpretation. A within-population conversion factor. Chapter 13 computes 46 percent for the suburban and resort stores, which is not the chain's margin and may not be used as one. **Common failure.** Applying a margin computed on one population to a quantity measured on another. The chain's contribution and the suburban stores' margin do not belong in the same fraction. **Verification.** Bounded above by 1. Recompute from the two totals rather than averaging line-level margins, which weights a $24 line like a $248 one.

Common variants. Gross margin uses merchandise cost alone in the numerator and is systematically larger. The two are routinely swapped in decks.

Owned by. Chapter 13, Section 13.12.1.

B.2.8 Net contribution after advertising spend

Question. Did this campaign's revenue survive both product cost and the advertising that produced it? **Formula.** revenue × margin_rate − ad_spend.

Numerator. Gross margin dollars on campaign revenue, less advertising spend. **Denominator.** None — it is an amount. **Window.** The campaign window. **Filters.** One campaign; advertising spend only, excluding creative production and agency fees. **Grain.** One marketing campaign. **Unit.** Dollars. **Required fields.** revenue, margin_rate, ad_spend.

Interpretation. The correction to return on ad spend that margin supplies. A campaign can lead on return on ad spend and still destroy contribution. **Common failure.** Calling it contribution margin. It is not: it excludes every marketing cost other than advertising, and the

name is chosen so the two cannot be confused. **Verification.** Recompute one campaign by hand. Confirm the sign: a negative value is a finding, not an error.

Owned by. Chapter 1, Lab 1.2.

B.2.9 Units sold

Question. How much merchandise moved, independent of price? **Formula.** Sum of quantity across the included lines.

Numerator. Sum of quantity. **Denominator.** None — it is a count. **Window.** Stated period. **Filters.** Stated filters. **Grain.** Stated population. **Unit.** Count of units. **Required fields.** quantity.

Interpretation. The volume half of every revenue movement. Revenue up with units flat is a price or mix story. **Common failure.** Counting order lines instead of units. A line carrying a quantity of three is one line and three units. **Verification.** Units must be at least as large as the line count on a file with no zero-quantity lines.

Owned by. Chapter 5, Section 5.7.

B.2.10 Number of orders

Question. How many purchase occasions occurred? **Formula.** Count of distinct order_id.

Numerator. Distinct order_id values. **Denominator.** None — it is a count. **Window.** Stated period. **Filters.** Stated filters. **Grain.** Stated population. **Unit.** Count of orders. **Required fields.** order_id.

Interpretation. The denominator of average order value and the volume term in the customer-value identity. **Common failure.** Using the row count of a line-grain file. A distinct count is not a row count; nunique on order_id counts orders and size counts order lines. **Verification.** Distinct orders must be no greater than the line count, and must equal the row count of any order-grain roll-up of the same file.

Owned by. Chapter 4, Section 4.8.

B.2.11 Number of customers

Question. How many distinct people bought? **Formula.** Count of distinct customer_id.

Numerator. Distinct customer_id values. **Denominator.** None — it is a count. **Window.** Stated period. **Filters.** Stated filters; lines carrying no customer_id cannot be counted and must be resolved, not dropped silently. **Grain.** Stated population. **Unit.** Count of customers. **Required fields.** customer_id.

Interpretation. The denominator of every customer-grain metric in Section B.3. **Common failure.** Letting a group-by drop rows whose customer_id is missing. pandas excludes missing group keys by default, so groups vanish without comment and the denominator shrinks unannounced. **Verification.** The customer count of a roll-up must equal the count of distinct non-null customer_id values in the lines it was built from.

Owned by. Chapter 4, Section 4.8.

B.2.12 Revenue growth rate

Question. How much did revenue change against its comparison period? **Formula.** (Current-period revenue − comparison-period revenue) / comparison-period revenue.

Numerator. Change in revenue between the two periods. **Denominator.** Comparison-period revenue. A comparison period of zero leaves the growth rate undefined; report the absolute change instead. **Window.** Two periods of equal length, both named; the comparison period is part of the metric, not context for it. **Filters.** Identical filters in both periods, including store roster. **Grain.** Stated population. **Unit.** Percentage. **Required fields.** line_revenue, order_date.

Interpretation. Meaningless without its comparison period. Year over year, quarter over quarter, and against plan are three different numbers. **Common failure.** Comparing periods whose store rosters differ, so an opening or closing is read as growth. Chapter 2's resort stores grew 40 percent partly because the prior year held a partial season. **Verification.** Confirm both periods carry the same length, the same filters, and a stated store roster. State the comparison period in the same sentence as the number.

Owned by. Chapter 2, Section 2.7.

B.2.13 Variance from target or baseline

Question. How far is the outcome from the number it was measured against? **Formula.** Actual − baseline for a dollar variance; (actual − baseline) / baseline for a percentage variance.

Numerator. Actual less the declared baseline. **Denominator.** The baseline, for the percentage form only. A baseline of zero leaves the percentage variance undefined; report the dollar variance instead. **Window.** The baseline's own period. **Filters.** The baseline's own filters. **Grain.** Stated population. **Unit.** Dollars for the difference; percentage for the ratio; never both under one label. **Required fields.** The actual, and a baseline declared before results were inspected.

Interpretation. The variance is only as meaningful as the baseline. A target set after the results are known is a description, not a benchmark. **Common failure.** Reporting a percentage variance and a dollar variance under one word and letting the reader assume which. State the unit. **Verification.** Confirm the baseline was declared in advance and record where. Recompute both forms and check they tell the same story.

Owned by. Chapter 2, Section 2.7.

B.2.14 Percentage change and percentage-point change

Question. Did the quantity move by a share of itself, or by an amount on its own scale? **Formula.** Percentage change = (new − old) / old. Percentage-point change = new − old, when both are already percentages.

Numerator. For percentage change, the difference; for percentage-point change, the difference itself is the answer. **Denominator.** For percentage change, the old value — an old value of zero leaves it undefined, and the change should be reported in levels; for percentage-point change, none. **Window.** Two stated periods. **Filters.** Identical in both periods. **Grain.**

Stated population. **Unit.** Percentage, or percentage points — these are different units and are never interchangeable. **Required fields.** The two values being compared.

Interpretation. A repeat-order share moving from 20 to 24 percent rose by 4 percentage points and by 20 percent. Both are true; they are not the same claim. **Common failure.** Reporting the larger of the two because it is more persuasive. Chapter 13 names this as framing: presenting a quantity in the form that produces the intended reaction where the alternative framing is equally true. **Verification.** Say the unit aloud. If the quantity being compared is itself a percentage, the difference is in points and the ratio is in percent.

Owned by. Chapter 13, Section 13.15.

B.2.15 Worked example

Take the practice file's twenty order lines. Gross sales — quantity times catalog unit price, before discount — come to $1,350.00. Three lines carry a discount, giving up $38.00, so net revenue is $1,312.00. That is the figure every other example in this appendix reconciles to.

In the practice file unit cost is exactly half of catalog unit price, so cost is $675.00 and contribution is $637.00, a contribution margin of 48.6%. The discount lines are where the margin story lives: a full-price line contributes exactly 50 percent, a line discounted 20 percent contributes 37.5 percent, and a line discounted 25 percent contributes 33.3 percent — because the discount comes off revenue and not off cost. Note that this is the practice file's own margin and not the 46 percent Chapter 13 computes for the suburban and resort stores; a margin belongs to the population it was computed on.

B.3 Basket and Customer Value Metrics

This family answers the question the expansion case turns on: what is a purchase occasion worth, what is a customer worth, and are those the same question. They are not.

B.3.1 Average order value (AOV)

Question. What does a typical purchase occasion come to? **Formula.** Sum of line_revenue / count of distinct order_id. In Tableau, SUM([line_revenue]) / COUNTD([order_id]).

Numerator. Net revenue on the included lines. **Denominator.** Distinct orders on the same included lines. **Window.** Stated period, both ends inclusive. **Filters.** Stated channel, store, metro, and customer filters, applied identically to numerator and denominator. **Grain.** One order. It is an order-grain mean: it is ordinarily computed from a line-grain file, but as revenue over distinct orders rather than as an average of line values. **Unit.** Dollars. **Required fields.** line_revenue, order_id.

Interpretation. Basket size in dollars. It knows nothing about frequency, so a store can lead on average order value and trail badly on revenue per customer. **Common failure.** Computing AVG(line_revenue) on a line-grain file. There is no error message and no visual difference between the correct chart and the incorrect one; Chapter 12 reports $50.41 for the line-grain average against $108.00 for the true chain figure. **Verification.** Recompute as total revenue over

distinct orders and confirm the two agree. On the practice file: $1,312.00 over 12 orders is $109.33.

Common variants. Some organizations compute AOV net of returns, and some at gross sales. Both are defensible; state which.

Owned by. Chapter 5, Section 5.7.

B.3.2 Average of group average order values (do not use)

Question. None. This entry exists so the quantity can be recognized and rejected. **Formula.** Unweighted mean of each group's AOV.

Numerator. Sum of the group AOVs. **Denominator.** Count of groups. **Window.** As stated. **Filters.** As stated. **Grain.** One group, equally weighted. **Unit.** Dollars. **Required fields.** Group-level AOV.

Interpretation. Reconciles to no total. A group of 40 orders and a group of 40,000 receive equal votes. **Common failure.** Reporting it as the overall figure. On the practice file the metro averages give $129.71 against a true pooled $109.33; Chapter 5's miniature gives $105.00 against a true $82.50. **Verification.** The pooled figure and the weighted mean must agree exactly. If a third number appears, it is this one.

Owned by. Chapter 5, Section 5.3.

B.3.3 Units per order (basket size)

Question. How many items are in a typical basket? **Formula.** Sum of quantity / count of distinct order_id.

Numerator. Units on the included lines. **Denominator.** Distinct orders on the same lines. **Window.** Stated period. **Filters.** Stated filters. **Grain.** One order. **Unit.** Units per order. **Required fields.** quantity, order_id.

Interpretation. The volume factor of the average-order-value identity. Chapter 5's designed finding is that the suburban gap is basket size and mix, not unit price. **Common failure.** Dividing by lines rather than orders, which understates basket size wherever orders carry more than one line. **Verification.** Must be at least 1 on a file with no zero-quantity orders. Multiply by revenue per unit and confirm the product is average order value exactly.

Owned by. Chapter 5, Section 5.7.

B.3.4 Revenue per unit (realized price per unit)

Question. What did an item actually sell for, after discount and mix? **Formula.** Sum of line_revenue / sum of quantity.

Numerator. Net revenue on the included lines. **Denominator.** Units on the same lines. **Window.** Stated period. **Filters.** Stated filters. **Grain.** One unit. The denominator is units — not lines and not orders. **Unit.** Dollars per unit. **Required fields.** line_revenue, quantity.

Interpretation. The price factor of the average-order-value identity. It moves with category mix, discount depth, and promotional participation, not only with list price. **Common failure.**

Reading it as a list price. StyleCraft's catalog prices do not differ by store; realized revenue per unit does, because mix does. **Verification.** Must fall inside the catalog band once discounts are accounted for. Recompute at a common price set to separate mix from price.

Owned by. Chapter 5, Section 5.7.

B.3.5 Average order value decomposition

Question. Is a basket-size gap made of more items or dearer items? **Formula.** AOV = units per order × revenue per unit. Exactly, by construction: revenue/orders = units/orders × revenue/units.

Numerator. Not applicable — a multiplicative identity. **Denominator.** Not applicable. **Window.** Stated period, identical for both factors. **Filters.** Identical for both factors. **Grain.** Order for the composite; order and unit for the two factors. **Unit.** Ratio factors, not additive shares. **Required fields.** line_revenue, quantity, order_id.

Interpretation. The identity yields factors that multiply, not percentage shares that add. On the practice file a 2.82x AOV gap between suburban and urban stores factors into a 1.50x basket factor and a 1.88x price factor. **Common failure.** Writing that 'basket size accounts for sixty percent of the gap.' The identity supplies no such allocation; an additive attribution needs a logarithmic decomposition and a stated convention. **Verification.** The two factors must multiply back to the composite exactly. Assert it rather than eyeballing it.

Owned by. Chapter 5, Section 5.7.

B.3.6 Revenue per customer

Question. What is one customer worth over the window? **Formula.** Sum of line_revenue / count of distinct customer_id.

Numerator. Net revenue attributed to the counted customers. **Denominator.** Distinct customers in the stated population. **Window.** A stated window. Without one this is a lifetime total wearing a period label. **Filters.** Stated filters; state whether non-purchasers are inside the denominator. **Grain.** One customer. **Unit.** Dollars per customer. **Required fields.** line_revenue, customer_id.

Interpretation. The customer-value counterpart of average order value. Revenue per customer equals AOV times orders per customer, which is where a weak repeat rate taxes a strong basket. **Common failure.** Interchanging it with average order value. Per order and per customer differ by frequency, and the expansion case must be argued on the customer figure. **Verification.** Confirm the identity: AOV × orders per customer = revenue per customer. On the practice file, $109.33 × 1.5 = $164.00.

Common variants. Computed over purchasers only, or over the whole customer file including non-purchasers. The two differ materially; name the base.

Owned by. Chapter 5, Section 5.7.

B.3.7 Orders per customer (purchase frequency)

Question. How often does a customer buy in the window? **Formula.** Count of distinct order_id / count of distinct customer_id.

Numerator. Distinct orders. **Denominator.** Distinct customers in the stated population. **Window.** A stated window; the figure is not comparable across windows of different length. **Filters.** Stated filters. **Grain.** One customer. **Unit.** Orders per customer. **Required fields.** order_id, customer_id.

Interpretation. The frequency factor between average order value and revenue per customer. **Common failure.** Comparing frequency across customers observed for different lengths of time. A customer who signed up nine weeks ago cannot have accumulated two years of orders; see exposure months. **Verification.** Must be at least 1 among purchasers. Recompute as the mean of the customer-level order counts and confirm agreement.

Owned by. Chapter 6, Section 6.5.

B.3.8 Frequency (RFM)

Question. How many purchase occasions has this customer had in the window? **Formula.** Count of distinct order_id for the customer in the window.

Numerator. Distinct orders placed by the customer. **Denominator.** None — it is a count. **Window.** The declared analysis window, open at the start and closed at the end, measured to a declared analysis date. **Filters.** Customers with at least one order in the window. **Grain.** One customer. **Unit.** Count of orders. **Required fields.** order_id, customer_id, order_date.

Interpretation. Counts orders, not lines: an order of five lines is one act of purchasing, not five. **Common failure.** Counting at line grain, which inflates frequency for multi-line baskets and quietly rewards the suburban shopper the analysis is trying to measure. **Verification.** The sum of customer frequencies must equal the distinct order count of the window.

Owned by. Chapter 6, Section 6.5.

B.3.9 Monetary value (RFM)

Question. How much has this customer spent in the window? **Formula.** Sum of line_revenue for the customer in the window.

Numerator. Net revenue on the customer's lines. **Denominator.** None — it is a total. **Window.** The declared analysis window. **Filters.** Customers with at least one order in the window. **Grain.** One customer. **Unit.** Dollars. **Required fields.** line_revenue, customer_id, order_date.

Interpretation. Strongly right-skewed at StyleCraft, and bounded by how long the customer has been observed. **Common failure.** Comparing monetary totals across customers of very different tenure, which restates the tenure question rather than answering the performance one. **Verification.** The sum of customer monetary values must equal the window's certified revenue total.

Common variants. A common variant uses average order value in place of total revenue. Total rewards frequent modest baskets; average rewards rare large ones. The choice is declared, not defaulted.

Owned by. Chapter 6, Section 6.5.

B.3.10 Recency

Question. How long has it been since this customer last bought? **Formula.** Analysis date − the customer's most recent order_date, in days.

Numerator. Elapsed days since the most recent purchase. **Denominator.** None — it is a duration. **Window.** Measured back from a declared analysis date; undefined for a customer who has never purchased. **Filters.** Purchasing population; every order_date must fall on or before the analysis date. **Grain.** One customer. **Unit.** Days. **Required fields.** order_date, customer_id, a declared analysis date.

Interpretation. Lower is better, which is why the recency score reverses the raw days. **Common failure.** A negative recency, which is a date defect rather than a customer behavior and means an order is dated after the analysis date. **Verification.** Assert that every recency is at or above zero, and that the minimum is consistent with the snapshot the features were built at.

Owned by. Chapter 6, Section 6.5.

B.3.11 Exposure months

Question. How long has this customer actually been observable inside the window? **Formula.** (Analysis date − the later of signup_date and the window start) in days, divided by 30.44, floored at 1.0 month.

Numerator. Observed days inside the window. **Denominator.** 30.44 days per month, then floored at 1.0. **Window.** The declared analysis window. **Filters.** Purchasing population. **Grain.** One customer. **Unit.** Months. **Required fields.** signup_date, a declared analysis date and window start.

Interpretation. A denominator only. It exists so that a customer observed for nine days does not appear to have a stable monthly rate. **Common failure.** Flooring the day count rather than the months, which produces a minimum of 0.986 rather than 1.0 and lets a near-zero denominator through. **Verification.** Assert the minimum is exactly 1.0.

Owned by. Chapter 6, Section 6.6.

B.3.12 Orders per month

Question. How often does this customer buy, adjusted for how long we have watched? **Formula.** frequency / exposure_months.

Numerator. Distinct orders in the window. **Denominator.** Exposure months. **Window.** The declared analysis window. **Filters.** Purchasing population. **Grain.** One customer. **Unit.** Orders per month. **Required fields.** frequency, exposure_months.

Interpretation. Reduces the mechanical effect of tenure on frequency. It does not eliminate it from a specification that still contains the raw totals. **Common failure.** Treating it as independent of frequency. It is frequency divided by a constant-per-customer denominator, and standardization does not remove that redundancy. **Verification.** Least stable for customers observed shortest; report the exposure distribution beside it.

Owned by. Chapter 6, Section 6.6.

B.3.13 Revenue per exposure-month

Question. What does this customer spend per month of observed relationship? **Formula.** monetary / exposure_months.

Numerator. Net revenue in the window. **Denominator.** Exposure months. **Window.** The declared analysis window. **Filters.** Purchasing population. **Grain.** One customer. **Unit.** Dollars per month. **Required fields.** monetary, exposure_months.

Interpretation. The measure on which a short-tenured segment stops looking weak. A segment can trail badly on lifetime revenue and lead on revenue per month of exposure. **Common failure.** Aggregating it to segment level by averaging the customer-level ratios, which gives a $40 customer and a $1,000 customer equal votes. Pool the numerators and denominators instead, or say plainly that the figure is a mean of ratios. **Verification.** State which aggregation was used, and use the same one for every ratio in the same exhibit.

Owned by. Chapter 6, Section 6.6.

B.3.14 Discount share

Question. How much of this customer's spending came on discounted merchandise? **Formula.** Revenue on lines with discount_pct greater than 0 / monetary.

Numerator. Net revenue on the customer's discounted lines. **Denominator.** The customer's total net revenue in the window. **Window.** The declared analysis window. **Filters.** Discounted lines are those with discount_pct greater than 0; the roll-up is at line grain, because an order mixing a discounted dress with a full-price scarf is one order and two different facts. **Grain.** One customer. **Unit.** Proportion between 0 and 1. **Required fields.** line_revenue, discount_pct, customer_id.

Interpretation. A margin proxy at customer level. A customer with no discounted lines has a share of exactly 0, which is data rather than missingness. **Common failure.** Computing it at order grain, which assigns a whole mixed basket to one side. **Verification.** Bounded in [0, 1] with a small tolerance, since shares summed from rounded line revenue can land a fraction above one in floating point.

Owned by. Chapter 6, Section 6.6.

B.3.15 Store share

Question. How much of this customer's buying happens in physical stores? **Formula.** Distinct orders with channel of Store / frequency.

Numerator. The customer's distinct store-channel orders. **Denominator.** The customer's distinct orders in the window. **Window.** The declared analysis window. **Filters.** Channel is read at order grain; the build asserts one channel per order. **Grain.** One customer. **Unit.** Proportion between 0 and 1. **Required fields.** channel, order_id, customer_id.

Interpretation. Separates the store-only occasion shopper from the digital-first core. **Common failure.** Taking the first channel of a multi-channel order, which manufactures a consistent order out of an inconsistent one. **Verification.** Bounded in [0, 1]. Assert that no order carries two channels before computing it.

Owned by. Chapter 6, Section 6.6.

B.3.16 Occasionwear share

Question. How much of this population's buying is occasionwear? **Formula.** Occasionwear units / total units for the unit-weighted form; occasionwear revenue / total revenue for the revenue-weighted form.

Numerator. Units, or revenue, on lines flagged is_occasionwear. **Denominator.** Total units, or total revenue, on the same lines. **Window.** Stated period. **Filters.** Requires a validated many-to-one join to products; an unmatched line has no flag and must not be silently counted as not occasionwear. **Grain.** Stated population — customer, metro, or store. **Unit.** Proportion. **Required fields.** quantity or line_revenue, is_occasionwear, product_id.

Interpretation. The mix term behind the suburban price factor. On the practice file the suburban store share is 0.500 against 0.000 urban. **Common failure.** Quoting 'occasionwear is 60 percent of suburban sales' without saying whether sales means units or revenue. The two differ whenever occasionwear prices differ from the rest. **Verification.** Bounded in [0, 1]. Confirm the join produced no unmatched lines before trusting the denominator.

Owned by. Chapter 5, Section 5.7.

B.3.17 Revenue share by decile

Question. How concentrated is revenue across customers? **Formula.** Decile revenue / total revenue in the window, over customers ranked by revenue and cut into ten equal-count groups.

Numerator. Revenue of the customers in the decile. **Denominator.** Total revenue in the window, across the ranked population. **Window.** A stated window — the interval is open at the start and closed at the end; concentration is window-dependent and the window belongs in the sentence. **Filters.** Customers with a resolved customer_id who were active in the window; ties broken deterministically so the cut is reproducible. **Grain.** One decile of customers. **Unit.** Proportion of revenue. **Required fields.** line_revenue, customer_id, order_date.

Interpretation. Decile 1 is the highest-revenue tenth. A top-decile share well above ten percent is the ordinary finding, not a surprising one. **Common failure.** Quoting a top-decile share from a short window. Measured concentration is inflated by observation error over short windows, so a 30-day decile share is not comparable with a trailing-twelve-month one.

Verification. The deciles must partition: customer counts sum to the ranked population and decile revenues sum to the window total.

Common variants. The Pareto shorthand — the top 20 percent account for X percent — is the same construction at a different cut. The true split is rarely the folkloric 80/20, and it varies meaningfully by category and window (Schmittlein et al., 1993).

Owned by. Chapter 5, Section 5.9.

B.3.18 Worked example

The practice file's twelve orders and $1,312.00 give an average order value of $109.33. Computing the same name at line grain gives $65.60, and averaging the three store-type figures gives $129.71. All three are arithmetically correct; only the first answers the question. Revenue per customer is a different quantity again: $1,312.00 over eight customers is $164.00, which is the average order value times 1.5 orders per customer.

Table B.5 decomposes the store-type gap. Suburban orders average $220.00 against $78.00 urban, a factor of 2.82. That factors exactly into a 1.50 basket factor and a 1.88 realized-price factor, and the two multiply back to the composite. The price factor is a mix effect rather than a pricing one: the unit-weighted occasionwear share — occasionwear units over total units, not over revenue — is 50.0% of suburban units against 0.0% of urban units, and StyleCraft's catalog prices do not differ by store.

Table B.5

The average-order-value decomposition on the practice file

Store type	Orders	Revenue	Units	AOV	Units per order	Revenue per unit	Product of the two factors
Suburban	2	440.00	6	220.00	3.00	73.33	220.00
Urban	3	234.00	6	78.00	2.00	39.00	78.00

The last column is the verification the identity entry asks for: units per order times revenue per unit returns average order value exactly, on every row. What the identity does not supply is an allocation. It is correct to say the gap factors into basket and price terms; it is not correct to say that basket size accounts for some percentage of it.

B.4 Retention and Customer Relationship Metrics

Retention measures are where a metric name most often outlives the definition it was given, because "came back" admits so many defensible operational readings. Every entry in this family therefore carries an unusually explicit window and eligibility rule, and the family's first entry carries two named variants rather than one silent choice.

B.4.1 Repeat purchase rate

Question. What share of customers came back? **Formula.** Customers with two or more orders in the window / customers eligible in the window.

Numerator. Customers with two or more completed orders in the window. **Denominator.** Customers with at least one order in the window. **Window.** Trailing twelve months to a declared analysis date, bounded at both ends. A one-sided filter makes the reported rate depend on the day the notebook happens to run. **Filters.** All channels. The headline definition applies no channel filter. **Grain.** One customer. **Unit.** Proportion. **Required fields.** order_id, customer_id, order_date.

Interpretation. StyleCraft's headline retention measure and the baseline of the expansion review, owned by the CRM manager. Customers acquired late in the window have had little time to repurchase, a mechanical depressant a same-age cohort comparison is designed to remove. **Common failure.** Quoting the name without the definition. Three defensible definitions of this phrase circulate at StyleCraft and produce 0.60, 0.43, and 0.20 on the same eighteen orders. **Verification.** Bounded in [0, 1]. Hand-count the numerator and denominator on a small file: on Chapter 3's preview, three repeaters over seven active customers is 0.43.

Common variants. Lifetime: two or more orders ever, over all customers with at least one completed order in the available history — 0.60 on the preview, and unresponsive to current behavior. Store-only: the same window restricted to the store channel in both numerator and denominator — 0.20 on the preview, and the right measure for the economics of the physical footprint. Report the store version beside the headline rather than instead of it.

Owned by. Chapter 3, Section 3.10.3.

B.4.2 Repeat-order share

Question. What share of orders were placed by customers who had bought before? **Formula.** Orders placed after the customer's first purchase date / distinct orders.

Numerator. Orders whose order_date is later than the customer's first purchase date. **Denominator.** Distinct orders in the population. **Window.** Stated period; the first purchase date is computed on full history, before any date filter, so that trimming a partial month does not manufacture first purchases. **Filters.** The go/no-go dashboard excludes the reserved ONLINE store, because the comparison is across physical store types. **Grain.** One order. **Unit.** Percentage. **Required fields.** order_id, customer_id, order_date.

Interpretation. An order-grain measure. On the practice file it is 4 of 12, or 33.3 percent, against a customer-grain repeat purchase rate of 37.5 percent on the same file. **Common failure.** Labelling it 'returning-customer share'. It is not the share of customers who returned, and the obvious name invites a customer-level reading the metric does not support. Because order_date carries a date rather than a timestamp, every order placed on a customer's first purchase date is classified as a first-day order. **Verification.** Bounded in [0, 1]. Confirm the chain figure is computed on the pooled denominator; Chapter 12 reports a chain share of 47 percent against an unweighted mean of the four store-type rates of 34 percent.

Owned by. Chapter 12, Section 12.13.2.

B.4.3 Churn rate (descriptive)

Question. What share of a stated customer population went quiet? **Formula.** Customers inactive by a stated rule / the stated customer population.

Numerator. Customers meeting the stated inactivity rule. **Denominator.** The stated eligible population. **Window.** A stated observation window. **Filters.** A stated eligibility rule. **Grain.** One customer. **Unit.** Proportion. **Required fields.** order_id, customer_id, order_date.

Interpretation. Descriptive bookkeeping. Churn is not naturally observed in a non-contractual business; it is operationalized, and different defensible rules produce different populations. **Common failure.** Presenting a descriptive churn rate as though it measured departure. It measures silence under one rule. **Verification.** Bounded in [0, 1]. State the rule and the window in the same sentence as the number.

Owned by. Chapter 5, Section 5.7.

B.4.4 Churn label

Question. Which customers are to be treated as having churned, for modeling? **Formula.** churned = 1 for an eligible customer with zero completed purchase orders in the outcome window.

Numerator. Customers with no outcome-window purchase order. **Denominator.** None — it is a per-customer flag, not a rate. **Window.** A six-month outcome window following a declared snapshot date; the feature window ends at the snapshot and both ends of each window are enforced. **Filters.** Both conditions: signup_date on or before the snapshot, and at least one purchase inside the feature window, so that every labeled customer has a history to be scored on. **Grain.** One customer. **Unit.** Binary flag. **Required fields.** order_id, customer_id, order_date, signup_date.

Interpretation. A measurement decision before it is a modeling one. Shortening the outcome window inflates the base rate with mid-cadence occasion shoppers; lengthening it purifies the label but treats customers too late to save. **Common failure.** Defining churn as zero net revenue rather than zero purchase activity. An order can net to zero after a discount and a returned line, so the two events are not the same and a revenue-based label mislabels them. **Verification.** Reconcile against the shipped answer key. Where activity and revenue definitions disagree, the disagreement is a finding about the revenue column rather than a bug in the label.

Common variants. A 90-day or 12-month outcome window, and activity redefined to include an app open or an email click, are each defensible and each produce a different base rate and a different treated population.

Owned by. Chapter 9, Section 9.3.

B.4.5 Cohort retention rate (period-active)

Question. Of the customers who started together, how many were active at a given age? **Formula.** Customers in the cohort active in period p / cohort size.

Numerator. Distinct customers in the cohort with at least one order in that calendar period. **Denominator.** Cohort size — the customers whose first order fell in the cohort's own period. **Window.** Periods measured from each cohort's own start, with an observability mask so that unreached periods stay missing rather than becoming zero. **Filters.** Deduplicated on customer and period, so a customer with three orders in a month counts once. **Grain.** One cohort in one period since its start. **Unit.** Percentage. **Required fields.** order_id, customer_id, order_date.

Interpretation. A period-active rate can rise as well as fall, because a customer who skipped a month can return. **Common failure.** Reading a blank cell as zero. Elapsed but inactive is 0.0; not yet reached is missing. A table that does not distinguish the two cannot be read down a column. **Verification.** Period zero must be 1.0 for every cohort; cohort sizes must sum to the distinct customer count; no cell may exceed 1.0.

Common variants. A survival retention rate holds the share retained continuously and is non-increasing by construction. A table that does not say which form it is cannot be read.

Owned by. Chapter 5, Section 5.8.

B.4.6 Cohort size

Question. How many customers started together? **Formula.** Count of customers whose first order fell in the cohort period.

Numerator. Distinct customers with a first order in the period. **Denominator.** None — it is a count. **Window.** The cohort's own defining period. **Filters.** Stated acquisition filters. **Grain.** One cohort. **Unit.** Count of customers. **Required fields.** customer_id, order_date.

Interpretation. Read down the left edge of every cohort table. A triumphant retention story over shrinking cohorts is two findings, not one. **Common failure.** Presenting retention percentages without the sizes that produced them. **Verification.** Cohort sizes must sum to the distinct customer count of the population.

Owned by. Chapter 5, Section 5.8.

B.4.7 Tenure

Question. How long has this person been a customer? **Formula.** Analysis date − signup_date, in days.

Numerator. Elapsed days since signup. **Denominator.** None — it is a duration. **Window.** Measured to a declared analysis or snapshot date. **Filters.** None beyond the stated population. **Grain.** One customer. **Unit.** Days. **Required fields.** signup_date, a declared analysis date.

Interpretation. The variable behind most apparent differences between new and established markets. Chapter 9's fairness screen exists because a new store's customers look unresponsive largely because they are new. **Common failure.** Reading a tenure artifact as a value verdict. A

lifecycle position is not a quality judgment. **Verification.** Tenure must be at least recency for every customer; a violation indicates a broken join rather than an unusual customer.

Owned by. Chapter 6, Section 6.6.

B.4.8 Time to second purchase

Question. How long does it take a new customer to come back? **Formula.** Days between a customer's first and second order_date.

Numerator. Elapsed days between the two orders. **Denominator.** None — it is a duration. **Window.** Measured from each customer's own first order, to a declared analysis date. **Filters.** Customers with at least two orders; customers with only one are censored, not missing, and must be reported as such. **Grain.** One customer. **Unit.** Days. **Required fields.** order_id, customer_id, order_date.

Interpretation. Averaging only over customers who did come back understates the true wait, because the customers who never returned are the slowest of all. **Common failure.** Reporting a mean over repeaters as though it described the base. State the share of customers with no second purchase in the same sentence. **Verification.** Report the censored share beside the statistic.

Owned by. Chapter 5, Section 5.7.

B.4.9 Worked example

The practice file's eight customers placed twelve orders. Three customers placed two or more, so the repeat purchase rate over the file's own window is 0.375. The repeat-order share is a different quantity computed on the same twelve orders: 4 of them were placed after the customer's first purchase date, so the share is 0.333. The two numbers are close, they are not the same metric, and the difference is the denominator — customers in the first, orders in the second.

Section B.1.3 supplies the larger demonstration. On Chapter 3's eighteen-order preview, three defensible definitions of the first metric give 0.60, 0.43, and 0.20. Each is correct; each answers a different question; and the middle one is the definition StyleCraft adopted as its headline, with the store version reported beside it rather than instead of it.

B.5 Acquisition, Media, and Campaign Metrics

Acquisition and media metrics arrive from platforms that computed them under their own conventions, which is what makes them dangerous. The denominators are set by someone else, the attribution window is a configuration value with a budget attached, and the revenue in the numerator is attributed rather than incremental.

B.5.1 Impressions

Question. How many times was the advertising served? **Formula.** Count of served impressions, as reported by the platform.

Numerator. Impressions recorded by the serving platform. **Denominator.** None — it is a count. **Window.** The campaign window. **Filters.** One campaign, or a stated roll-up of campaigns. **Grain.** One campaign. **Unit.** Count. **Required fields.** impressions.

Interpretation. A delivery count, not a count of people. Impressions are not reach. **Common failure.** Using impressions as the denominator of a conversion rate, which mixes a served-advertising count with a customer count and produces a number that is not a funnel stage. **Verification.** Must be at least the click count on the same campaign.

Owned by. Chapter 1, Section 1.10.

B.5.2 Clicks

Question. How many times did someone click through? **Formula.** Count of recorded clicks.

Numerator. Clicks recorded by the serving platform. **Denominator.** None — it is a count. **Window.** The campaign window. **Filters.** One campaign, or a stated roll-up. **Grain.** One campaign. **Unit.** Count. **Required fields.** clicks.

Interpretation. Serves as numerator of click-through rate and denominator of post-click conversion rate, which is why its definition has to travel with both. **Common failure.** Treating clicks as unique people. Unless the platform de-duplicates, a click count is an event count. **Verification.** Must be no greater than impressions. Under a one-conversion-per-click convention clicks must also be at least the conversion count; otherwise compare the two only after confirming that clicks and conversions use compatible event definitions, since one click can carry credit for several conversion events.

Owned by. Chapter 1, Section 1.10.

B.5.3 Click-through rate (CTR)

Question. How well did the creative earn attention from the advertising served? **Formula.** clicks / impressions.

Numerator. Clicks. **Denominator.** Impressions. **Window.** The campaign window. **Filters.** One campaign, or a stated pooled roll-up; a campaign with zero impressions has an undefined rate, not an infinite one. **Grain.** One campaign. **Unit.** Proportion. **Required fields.** clicks, impressions.

Interpretation. A creative and targeting measure. It says nothing about what happened after the click. **Common failure.** Averaging campaign click-through rates to get a portfolio figure. Recompute from pooled totals: on Chapter 1's four campaigns the pooled rate is 0.0500. **Verification.** Bounded in [0, 1]. A rate above 1 means clicks exceed impressions somewhere and the numerator and denominator have been reversed or the join has duplicated.

Owned by. Chapter 1, Section 1.10.

B.5.4 Post-click conversion rate

Question. Of the units that clicked or entered through the campaign, what share converted under the declared counting convention? **Formula.** conversions / clicks.

Numerator. Attributed conversion events on the campaign, counted on the same basis as the denominator. **Denominator.** Click events on the same campaign, counted on the same basis as the numerator. **Window.** The campaign window, plus the platform's conversion window. **Filters.** One campaign; the column is named for its denominator so that the definition travels with the number. Numerator and denominator must count the same unit: unique clickers against unique converters, sessions against converted sessions, or click events against conversion events under a stated one-conversion-per-click convention. **Grain.** One campaign. **Unit.** Proportion. **Required fields.** conversions, clicks.

Interpretation. One of at least four quantities the phrase conversion rate can denote. Conversions may be divided by clicks, sessions, visitors, leads, or recipients, and the results differ. **Common failure.** Reporting it as 'conversion rate' with no denominator named, or averaging the four campaign rates instead of pooling. Chapter 1's campaigns give 0.1161 pooled against 0.1179 averaged. **Verification.** Bounded in [0, 1] only when numerator and denominator count the same unit on the same basis; a value above 1 means they do not, most often because one converter was credited with several conversions against a single click. Confirm the conversion window is the same across any campaigns being compared.

Common variants. Conversion over impressions, over sessions, and over recipients are all in circulation. They are different metrics, not different estimates of one.

Owned by. Chapter 1, Section 1.10.

B.5.5 Cost per click (CPC)

Question. What did each click cost? **Formula.** spend / clicks.

Numerator. Media spend on the campaign. **Denominator.** Clicks on the same campaign. **Window.** The campaign window. **Filters.** One campaign; a campaign with zero clicks has an undefined cost per click. **Grain.** One campaign. **Unit.** Dollars per click. **Required fields.** spend, clicks.

Interpretation. An auction-efficiency measure. Low cost per click with low conversion buys traffic, not customers. **Common failure.** Letting a zero denominator produce an infinite value. A rate over nothing is undefined, not infinite, and a table full of inf is a denominator problem wearing a printing error. **Verification.** Recompute from pooled totals when aggregating; never average campaign-level rates.

Owned by. Chapter 5, Section 5.7.

B.5.6 Cost per thousand impressions (CPM)

Question. What did it cost to reach the audience a thousand times? **Formula.** spend / impressions × 1,000.

Numerator. Media spend. **Denominator.** Impressions, scaled to thousands. **Window.** The campaign window. **Filters.** One campaign. **Grain.** One campaign. **Unit.** Dollars per thousand impressions. **Required fields.** spend, impressions.

Interpretation. A buying-price measure that says nothing about response. Chapter 1's Clearance campaign runs at $139.71 against New Arrivals at $34.62. **Common failure.** Comparing cost per thousand across channels whose impression definitions differ. A video impression and a display impression are not the same event. **Verification.** Confirm the factor of one thousand is applied once. Recompute from pooled totals when aggregating.

Owned by. Appendix B; the book computes spend and impressions but does not name this ratio.

B.5.7 Cost per conversion

Question. What did each conversion cost in advertising? **Formula.** ad_spend / conversions.

Numerator. Advertising spend on the campaign. **Denominator.** Conversions attributed to the same campaign. **Window.** The campaign window plus the conversion window. **Filters.** One campaign; advertising spend only, excluding creative production and agency fees. A campaign with zero conversions has an undefined cost per conversion and is reported separately. **Grain.** One campaign. **Unit.** Dollars per conversion. **Required fields.** ad_spend, conversions.

Interpretation. A campaign-level efficiency measure. It is not customer acquisition cost, because a conversion is not necessarily a new customer. **Common failure.** Reading it as acquisition cost. Repeat purchases convert too. **Verification.** Recompute from pooled totals when aggregating. Confirm the spend column excludes non-media cost, or rename it.

Owned by. Chapter 1, Exercise 1.4.

B.5.8 Customer acquisition cost (CAC)

Question. What did it cost to acquire one new customer? **Formula.** Acquisition spend / customers acquired.

Numerator. Acquisition spend, on a stated cost basis. **Denominator.** Customers acquired, on a stated definition of acquired. **Window.** A stated window, with the window and the definition of acquired doing most of the work. **Filters.** Stated channel and campaign attribution rules. **Grain.** One acquired customer, reported for a cohort. **Unit.** Dollars per customer. **Required fields.** spend, customer_id, signup_date or first order_date.

Interpretation. The most reliably contested marketing metric, because two defensible cost bases produce two very different numbers and each answers a different question. **Common failure.** Comparing a campaign-attributed figure against a fully loaded one. Marketing's number — campaign spend over attributed new customers — is usually and sometimes substantially smaller than finance's, which divides total sales and marketing expense by every new customer, attributed or not. **Verification.** State the cost basis, the attribution rule, and the definition of acquired in the same place as the number. If the organization cannot state all three, that is the finding.

Common variants. Campaign cost per acquisition against fully loaded customer acquisition cost. Neither is wrong; they are different metrics and must not be compared.

Owned by. Chapter 5, Section 5.7; the dispute is set out in Chapter 3, Section 3.12.2.

B.5.9 Return on ad spend (ROAS)

Question. How much revenue came back for each advertising dollar? **Formula.** revenue / ad_spend.

Numerator. Revenue attributed to the campaign. **Denominator.** Advertising spend on the campaign, excluding creative production and agency fees. **Window.** The campaign window plus the attribution window. **Filters.** One campaign; a campaign with zero advertising spend has an undefined return on ad spend and is reported separately rather than assigned an infinite value. **Grain.** One campaign. **Unit.** Ratio. **Required fields.** revenue, ad_spend.

Interpretation. Attributed, not incremental. Return on ad spend answers what revenue was credited to the campaign, not what revenue the campaign caused. **Common failure.** Two distinct failures. First, ignoring margin: Chapter 1's Clearance campaign returns 2.46 on revenue and destroys $5,288 of contribution. Second, reading it causally: an attributed figure becomes an incremental one only through an experiment or a designed holdout. **Verification.** Recompute from pooled totals when aggregating; the four campaigns pool to 5.79 against an average of 12.33. Then compute net contribution after advertising spend beside it.

Common variants. Incremental return on ad spend, measured against a randomized or geographic holdout, is a different and stronger quantity. Use the word incremental only when the evidence supports it.

Owned by. Chapter 1, Section 1.10.

B.5.10 Attributed orders

Question. How many orders were credited to a campaign? **Formula.** Count of distinct order_id among lines carrying the campaign_id.

Numerator. Distinct orders carrying the campaign identifier. **Denominator.** None — it is a count. **Window.** Stated period. **Filters.** Lines with a non-null campaign_id; a null campaign_id may mean unattributed or may mean attribution failed, and the two look identical in the file. **Grain.** One campaign. **Unit.** Count of orders. **Required fields.** order_id, campaign_id.

Interpretation. The product of an attribution rule, not an observation. An attributed order may be credited on a view rather than a click. **Common failure.** Dropping null campaign_id rows before computing campaign performance, which assumes every null is a legitimate non-attribution. **Verification.** Attributed orders summed across campaigns must not exceed the distinct order count of the file. An order carrying two campaign identifiers would be counted once under each while that reconciliation still passed.

Owned by. Chapter 5, Section 5.8.

B.5.11 Attribution (conversion) window

Question. How long after an interaction may a conversion still be credited to it? **Formula.** A declared number of days following a click, view, or engaged view.

Numerator. Not applicable — it is a parameter of other metrics. **Denominator.** Not applicable. **Window.** It is the window. **Filters.** Set per conversion action and per platform.

Grain. Applies to every attributed metric downstream. **Unit.** Days. **Required fields.** A platform setting, recorded alongside every attributed figure.

Interpretation. A configuration value with a budget attached. Shortening it reduces the conversions the account records; the customers do not change. **Common failure.** Comparing channels measured on different windows. Every platform ships its own defaults, so this happens easily and the comparison is not close to fair. **Verification.** Record the window beside every attributed figure and refuse to compare channels whose windows differ.

Owned by. Chapter 3, Section 3.12.3.

B.5.12 Worked example

Chapter 1's four campaigns are the guide's canonical media example, and Table B.6 computes the family on them. Read the return on ad spend column first and the Loyalty Offer is the obvious winner at 28.00. Read the last column and the Clearance campaign, which returned $2.46 for every advertising dollar, destroyed $5,288 of contribution — because its margin rate is 0.18 and return on ad spend does not know that.

Table B.6

The acquisition and media family, computed on Chapter 1's four campaigns

Campaign	CTR	Post-click conv.	CPC	CPM	Cost per conv.	ROAS	Net contribution after ad spend
New Arrivals	0.0550	0.1084	0.63	34.62	5.81	15.50	13,545
Loyalty Offer	0.0620	0.1654	0.47	29.27	2.86	28.00	9,552
Clearance	0.0500	0.1147	2.79	139.71	24.36	2.46	-5,288
Welcome Series	0.0400	0.0833	1.73	69.33	20.80	3.37	3,550

The table also shows why rates are recomputed rather than averaged. Pooling the four campaigns gives a click-through rate of 0.0500 and a post-click conversion rate of 0.1161; averaging the four campaign figures gives 0.1179 for the second, which reconciles to no total. The distortion is far worse on return on ad spend, where the pooled figure is 5.79 and the average of the four campaign ratios is 12.33 — more than twice as large, because the average gives the tiny, efficient Loyalty Offer the same weight as the large, inefficient Clearance campaign.

B.6 CRM and Engagement Metrics

Engagement metrics share one problem and it is the denominator. Sent, delivered, and unique-opened bases all circulate under the same metric names, and they produce different numbers from identical behavior.

B.6.1 Delivery rate

Question. What share of messages reached an inbox? **Formula.** delivered / sent.

Numerator. Messages accepted by the receiving server. **Denominator.** Messages sent. **Window.** The send, or a stated period of sends. **Filters.** One send, or a stated roll-up. **Grain.** One send. **Unit.** Percentage. **Required fields.** sent, delivered.

Interpretation. The denominator every downstream engagement rate should probably use. **Common failure.** Ignoring it, and then computing open rate on sends without saying so. **Verification.** Delivery rate and bounce rate sum to 1 only once every message in the send has reached a final delivered-or-bounced status and the denominator excludes suppressed, pending, and otherwise unresolved records. State that condition, or state the share still unresolved.

Owned by. Appendix B; the book names deliverability but catalogs no formula.

B.6.2 Bounce rate

Question. What share of messages failed to reach an inbox? **Formula.** bounced / sent.

Numerator. Messages rejected, hard or soft. **Denominator.** Messages sent. **Window.** The send. **Filters.** One send; state whether hard and soft bounces are combined. **Grain.** One send. **Unit.** Percentage. **Required fields.** sent, bounced.

Interpretation. A list-health measure. A rising bounce rate is a data-quality finding before it is a marketing one. **Common failure.** Combining hard and soft bounces silently, which hides a list-decay signal inside a transient one. **Verification.** Bounce rate and delivery rate sum to 1 only under the condition stated in the delivery-rate entry: every message resolved, and no suppressed or pending records in the denominator.

Owned by. Appendix B.

B.6.3 Open rate

Question. What share of recipients opened the message? **Formula.** Unique opens / delivered. The headline definition counts unique opens; total opens is a different numerator and a different metric.

Numerator. Unique opens — one per recipient who opened at least once. **Denominator.** Delivered messages, or sent messages — these are different denominators. **Window.** The send, plus a stated attribution period for late opens. **Filters.** One send, or a stated roll-up. **Grain.** One send. **Unit.** Percentage. **Required fields.** unique_opens, delivered, sent.

Interpretation. A weak behavioral measure. Technical privacy protections pre-fetch images on the recipient's behalf, so an open may record no human attention at all. **Common failure.** Circulating three denominators under one name. On the worked example, 2,688 opens is 28.0 percent of delivered and 26.9 percent of sent — both correct, and not the same metric. **Verification.** Bounded in [0, 1] for the unique-open form only. A total-open rate is not bounded by 1, because one recipient can open repeatedly, and it should be reported as opens per delivered message rather than as a percentage. State the base — sent or delivered — in the same sentence,

and never compare an open rate across a period in which the platform changed its open measurement.

Common variants. Unique opens over delivered is the headline. Unique opens over sent is defensible and systematically smaller. Total opens over delivered is a different quantity altogether: it counts events rather than recipients, has no upper bound of 1, and must not be labeled a rate alongside the other two.

Owned by. Chapter 5, Section 5.7.

B.6.4 Click rate

Question. What share of recipients clicked? **Formula.** Unique clicks / delivered. The headline definition counts unique clicks; total clicks is a different numerator and is not bounded by 1.

Numerator. Unique clicks — one per recipient who clicked at least once. **Denominator.** Delivered, or sent. **Window.** The send, plus a stated attribution period. **Filters.** One send, or a stated roll-up. **Grain.** One send. **Unit.** Percentage. **Required fields.** unique_clicks, delivered, sent.

Interpretation. A stronger behavioral signal than open rate, because it is harder to trigger without intent. **Common failure.** Comparing a click rate computed on delivered against one computed on sent. On the worked example, 384 clicks is 4.00 percent of delivered and 3.84 percent of sent. **Verification.** Bounded in [0, 1] for the unique-click form only; a total-click count per delivered message has no upper bound of 1. Do not assume the click rate cannot exceed the open rate: a click can be recorded when the open-tracking pixel was blocked or never fired, so the ordering is an expectation rather than a guarantee, and a reversal is a tracking finding rather than an arithmetic error.

Owned by. Chapter 5, Section 5.7.

B.6.5 Click-to-open rate (CTOR)

Question. Of the recipients who opened, how many went on to click? **Formula.** Unique recipients with both a recorded open and a recorded click / unique opens.

Numerator. Unique recipients who both opened and clicked. This is not the same as all unique clickers, because a click can be recorded when no open was. **Denominator.** Unique opens. **Window.** The send. **Filters.** One send; numerator and denominator drawn from the same recipient population. **Grain.** One send. **Unit.** Percentage. **Required fields.** unique_opens, unique_clicks, and the recipient-level join that identifies which clickers also opened.

Interpretation. A content and offer measure, holding the subject line constant. On the worked example every clicker also registered an open, so 384 over 2,688 is 14.3 percent. **Common failure.** Taking all unique clickers as the numerator. Because the click-rate entry establishes that a click can be recorded when the open pixel never fired, that numerator is not a subset of the denominator and the result can exceed 100 percent. **Verification.** Bounded in [0, 1] by construction, because the numerator is a subset of the denominator. Under the all-clickers variant the bound does not hold, and a value above 1 is a tracking finding — open measurement

is incomplete — rather than an arithmetic error. Because opens are inflated by automatic pre-fetching, this rate is deflated by the same mechanism.

Common variants. Some platforms report all unique clickers over unique opens, which is easier to compute and is not bounded by 1. Where that form is used, say so and treat any figure above 100 percent as evidence about the tracking rather than about the creative.

Owned by. Appendix B.

B.6.6 Unsubscribe rate

Question. What did this message cost in permission? **Formula.** Unsubscribes / delivered.

Numerator. Unsubscribe events attributed to the send. **Denominator.** Delivered messages. **Window.** The send, plus a stated attribution period. **Filters.** One send. **Grain.** One send. **Unit.** Percentage. **Required fields.** unsubscribes, delivered.

Interpretation. A guardrail rather than a performance measure. A campaign that wins on click rate and loses permission has not won. **Common failure.** Omitting it from a campaign readout because it is small. The cost is cumulative and irreversible. **Verification.** Bounded in [0, 1]. Where the field is not instrumented, say so: a preregistered guardrail that cannot be calculated is a design defect, not a minor gap.

Owned by. Chapter 11, Table 11.3, where it is named as an instrumentation obligation.

B.6.7 Opt-in rate

Question. What share of customers have consented to be contacted on this channel? **Formula.** Customers with the channel's consent flag set / customers in the stated population.

Numerator. Customers with email_opt_in true, or with sms_opt_in true — one channel at a time, never combined. **Denominator.** Customers in the stated population. **Window.** A point in time, as of a stated date. **Filters.** Stated population; one named channel. **Grain.** One customer, reported for a population. **Unit.** Percentage. **Required fields.** email_opt_in or sms_opt_in, customer_id.

Interpretation. A permission measure. Because a Boolean is stored as 1 and 0, its mean is a legitimate and useful quantity. **Common failure.** Folding app_user into this rate. Installing an application is not consent to be contacted, and the two belong in separate entries; see the app-user share below. A second failure is reading opt-in as interest, when it may measure the aggressiveness of the signup flow in force when the customer joined. **Verification.** Bounded in [0, 1]. Confirm the flag is genuinely Boolean and that missing values have not been coerced to false. Report one channel per figure.

Owned by. Chapter 3, Section 3.4.1.

B.6.8 App-user share

Question. What share of customers have the mobile application, as of today? **Formula.** Customers with app_user true / customers in the stated population.

Numerator. Customers with app_user true. **Denominator.** Customers in the stated population. **Window.** A point in time, as of a stated date. This is a stock, not a flow: it is not an adoption or activation rate over a window. **Filters.** Stated population. **Grain.** One customer, reported for a population. **Unit.** Percentage. **Required fields.** app_user, customer_id.

Interpretation. A reachability measure and a strong behavioral correlate — Chapter 7 spends a whole section on why that correlation is not an effect. **Common failure.** Treating it as consent, or as an adoption rate. app_user records installation at a point in time; it carries no permission and no date, so a change between two extracts cannot be attributed to a period. **Verification.** Bounded in [0, 1]. State the as-of date, since the field is overwritten rather than versioned.

Owned by. Appendix B; the guide uses app_user as a feature and a holdout but catalogs no share.

B.6.9 Worked example

One send of 10,000 messages: 400 bounced, so 9,600 were delivered. 2,688 unique opens and 384 unique clicks followed — every clicker in this send also registered an open, which is what lets the click-to-open rate be read as a share — with 48 unsubscribes. Table B.7 computes the family twice, once on each base, and the point of the table is the gap between the two columns.

Table B.7

One email send, with every rate computed on both circulating bases

Metric	On delivered	On sent	Which base the entry recommends
Delivery rate	—	96.0%	Sent, always
Bounce rate	—	4.0%	Sent, always
Open rate	28.0%	26.9%	Delivered, stated in the same sentence
Click rate	4.0%	3.8%	Delivered, stated in the same sentence
Click-to-open rate	14.3%	—	Opens; the numerator counts only clickers who also opened
Unsubscribe rate	0.50%	—	Delivered

Neither column is wrong. An open rate of 28.0 percent and one of 26.9 percent describe identical behavior, and a team that reports one while its benchmark uses the other will conclude something false about a difference that does not exist. This is the Chapter 3 dispute in its smallest form, and the remedy is the same: name the base beside the number.

B.7 Funnel Metrics

A funnel is a sequence of stages counting the same unit, in one window, under an event-order requirement. Where any of those three conditions fails, the arithmetic still runs and the result is not a funnel.

B.7.1 Stage count

Question. How many units reached this stage? **Formula.** Count of units meeting the stage's membership rule.

Numerator. Units at the stage. **Denominator.** None — it is a count. **Window.** A single stated window, identical across every stage. **Filters.** A stated membership rule per stage, and a stated event-order requirement. **Grain.** The funnel's unit — session, visitor, customer, or lead. One unit for the whole funnel. **Unit.** Count. **Required fields.** Whatever the stage rules require, plus a stage timestamp.

Interpretation. Every stage of a funnel must count the same kind of thing. Sessions, clicks, and orders are three different units, not three stages. **Common failure.** Building a funnel whose stages count different units and calling the last figure a conversion rate. **Verification.** Each stage's count must be no greater than the stage above it, once the event-order requirement is enforced.

Owned by. Chapter 5, Section 5.8.

B.7.2 Stage-to-stage conversion rate

Question. Of the units that reached this stage, how many reached the next? **Formula.** Count at this stage / count at the previous stage.

Numerator. Units at this stage. **Denominator.** Units at the immediately preceding stage. **Window.** One window, shared by both stages. **Filters.** Shared membership rules; both stages must count the same unit. **Grain.** The funnel's unit. **Unit.** Percentage. **Required fields.** Stage counts.

Interpretation. Localizes the loss. On the worked funnel, carts to orders converts at 40.0 percent while sessions to product views converts at 30.0 percent. **Common failure.** Quoting a step rate and a cumulative rate under one word. A reader told one while imagining the other is misinformed by true numbers. **Verification.** Bounded in [0, 1]. The product of the step rates must equal the cumulative rate at the bottom: $0.300 \times 0.250 \times 0.400 = 0.030$.

Owned by. Chapter 5, Section 5.8.

B.7.3 Cumulative (overall) conversion rate

Question. Of everyone who entered, how many finished? **Formula.** Count at this stage / count at the top of the funnel.

Numerator. Units at this stage. **Denominator.** Units at the funnel's top stage. **Window.** One window. **Filters.** Shared membership rules. **Grain.** The funnel's unit. **Unit.** Percentage. **Required fields.** Stage counts.

Interpretation. The end-to-end figure. On the worked funnel it is 3.0 percent from session to order. **Common failure.** Computing it across stages drawn from different populations or periods, or dividing orders by impressions and calling the result a funnel conversion rate. Impressions are advertising delivery, not funnel entrants. **Verification.** Must equal the product of the step rates. If it does not, a stage is counting a different unit.

Owned by. Chapter 5, Section 5.8.

B.7.4 Drop-off rate

Question. What share of units were lost at this step? **Formula.** 1 − stage-to-stage conversion rate.

Numerator. Units that reached the previous stage and not this one. **Denominator.** Units at the previous stage. **Window.** One window. **Filters.** Shared membership rules. **Grain.** The funnel's unit. **Unit.** Percentage. **Required fields.** Stage counts.

Interpretation. The complement of the step rate, and often the more actionable framing. **Common failure.** Adding drop-off rates across steps. They are computed on different denominators and do not sum. **Verification.** Drop-off and step conversion must sum to 1 at every step.

Owned by. Appendix B; the book computes step rates but does not name the complement.

B.7.5 Worked example

A four-stage funnel counting sessions throughout: sessions 40,000, product views 12,000, carts 3,000, orders 1,200. Table B.8 gives the step and cumulative forms.

Table B.8

One funnel, read two ways

Stage	Count	Step conversion	Drop-off	Cumulative conversion
Sessions	40,000	—	—	100.0%
Product views	12,000	30.0%	70.0%	30.0%
Carts	3,000	25.0%	75.0%	7.5%
Orders	1,200	40.0%	60.0%	3.0%

The two rates describe the same funnel and answer different questions. Carts convert to orders at 40.0 percent, the strongest step in the funnel; sessions convert to orders at 3.0 percent, which is the number an executive means by conversion rate. The verification the entries require is that the step rates multiply to the cumulative rate: $0.300 \times 0.250 \times 0.400 = 0.030$. If they do not, a stage is counting something other than sessions, and the exhibit is not a funnel.

B.8 Segmentation and Targeting Measures

Segmentation measures describe groups the analyst constructed, which means they carry a risk the other families do not: they can look like findings about customers when they are findings about the specification. Several entries below are therefore validity measures rather than business measures, and each states what it does and does not establish.

B.8.1 Segment size

Question. How many customers are in this segment? **Formula.** Count of customers assigned to the segment.

Numerator. Customers in the segment. **Denominator.** None — it is a count. **Window.** The window in which the segmenting features were built. **Filters.** The declared segmentation population — at StyleCraft, customers with at least one order in the window. **Grain.** One segment. **Unit.** Count of customers. **Required fields.** customer_id, the segment assignment.

Interpretation. Sizes are findings, not footnotes: size is the first row of the profile. **Common failure.** Treating size alone as substantiality. Whether a segment is worth serving is a threshold question about whether differentiated execution pays, not a headcount test. **Verification.** Segment sizes must sum to the population that was segmented.

Owned by. Chapter 6, Section 6.10.

B.8.2 Segment share of customers

Question. What portion of the base does this segment represent? **Formula.** Customers in the segment / customers in the segmented population.

Numerator. Customers in the segment. **Denominator.** Customers in the segmented population, which is not the same as the customer file. **Window.** The feature window. **Filters.** The declared segmentation population. **Grain.** One segment. **Unit.** Proportion. **Required fields.** customer_id, the segment assignment.

Interpretation. On the practice file, NYC Suburban holds 37.5 percent of customers and 45.8 percent of revenue. **Common failure.** Quoting the share against the whole customer file when the segmentation excluded non-purchasers. **Verification.** Shares must sum to 1.

Owned by. Chapter 6, Section 6.11.

B.8.3 Segment share of revenue

Question. What portion of revenue does this segment produce? **Formula.** Segment revenue / total revenue in the window.

Numerator. Revenue of the segment's customers. **Denominator.** Total revenue of the segmented population. **Window.** The feature window. **Filters.** The declared segmentation population. **Grain.** One segment. **Unit.** Proportion. **Required fields.** line_revenue, customer_id, the segment assignment.

Interpretation. Read beside the customer share. A segment holding a fifth of customers and a half of revenue is a different proposition from one holding both. **Common failure.** Reporting a segment revenue share that does not add back to the certified window total, which means customers were lost between the feature table and the summary. **Verification.** Shares must sum to 1, and segment revenues must sum to the certified window revenue.

Owned by. Chapter 6, Section 6.11.

B.8.4 RFM scores (R, F, M)

Question. Where does this customer rank on recency, frequency, and spending? **Formula.** Quintile scores from 1 to 5 on recency and monetary value, with recency reverse-scored; frequency scored in declared bands where the data will not support five equal-count bins.

Numerator. Rank position within the scored population. **Denominator.** Five bins, equal-count for quintiles and declared-threshold for bands. **Window.** The declared analysis window. **Filters.** The scored population; every customer with the same frequency must receive the same frequency score. **Grain.** One customer. **Unit.** Ordinal score, unitless. **Required fields.** recency_days, frequency, monetary.

Interpretation. Scores preserve rank, not level. The top monetary quintile exists even when spending has fallen for every customer in the base. **Common failure.** Forcing five equal-count frequency bins on a lumpy count. Where a large share of customers have exactly one order, quintiles either fail outright or split identical customers on row order, which carries no behavioral content. **Verification.** Assert that customers sharing a frequency share a score. Report the cut points beside the scores, since the scores are not comparable across periods or populations without them.

Common variants. Monetary value may be scored on total revenue or on average order value. Total rewards frequent modest baskets; average rewards rare large ones.

Owned by. Chapter 6, Section 6.5.

B.8.5 RFM grid region

Question. Which named behavioral region does this customer fall in? **Formula.** A published lookup from the (R, F) score pair to a named region.

Numerator. Not applicable — a lookup, not a ratio. **Denominator.** Not applicable. **Window.** The declared analysis window. **Filters.** The scored population. **Grain.** One customer. **Unit.** Categorical label. **Required fields.** R and F scores.

Interpretation. The grid sees exactly three behaviors and draws its regions on two of them, so the largest-spending customer can land in an at-risk region. **Common failure.** Publishing a grid with gaps or overlaps, which lets two analysts apply the same rules and reach different assignments. **Verification.** Assert the grid is exhaustive and mutually exclusive over all twenty-five score pairs.

Owned by. Chapter 6, Section 6.5.

B.8.6 Silhouette score

Question. How cleanly separated is a clustering solution? **Formula.** For each customer, the comparison of average distance to its own cluster against average distance to the nearest other cluster, averaged over customers.

Numerator. Not applicable — a scaled comparison. **Denominator.** Not applicable. **Window.** The feature window. **Filters.** The standardized clustering inputs. **Grain.** One

customer, then averaged across the solution. **Unit.** Unitless score between -1 and +1. **Required fields.** The standardized feature matrix and the cluster assignment.

Interpretation. Near +1 means snugly placed, near 0 means on a boundary, negative means arguably misassigned. **Common failure.** Obeying it. The silhouette often crowns the coarsest cut, because the coarsest cut is the cleanest, even when finer structure is real and strategically necessary. **Verification.** Report it beside the inertia curve and the business constraint on the number of treatments. It grades geometry only: a high score says the partition is clean, not that its groups are marketing segments.

Owned by. Chapter 6, Section 6.9.

B.8.7 Adjusted Rand index (ARI)

Question. Do two clustering solutions agree about which customers belong together? **Formula.** Over all pairs of customers, how often two solutions agree about whether the pair belongs together, corrected for the agreement expected by chance.

Numerator. Not applicable — a chance-corrected pairwise agreement. **Denominator.** Not applicable. **Window.** The feature window. **Filters.** The customers both solutions saw. **Grain.** One pair of customers, summarized per pair of partitions. **Unit.** Index, unitless; 1.0 for identical partitions, near 0 for chance-level agreement, and negative for less agreement than chance, with a lower bound of −0.5. **Required fields.** Two cluster assignments over the same customers.

Interpretation. Invariant to relabeling, which is why it replaces cross-tabulating dominant cells. Used for three distinct questions: stability across seeds, sensitivity to the feature set, and recovery of a designed answer key. **Common failure.** Reading reproducibility as meaning. High agreement establishes that the partition is reproducible; it does not establish that the groups are marketing segments. **Verification.** A solution compared with itself must return exactly 1.0; assert it as a control before trusting any other value.

Owned by. Chapter 6, Section 6.10.

B.8.8 Inertia (within-cluster sum of squares)

Question. How tightly do customers sit around their assigned cluster centers? **Formula.** Total of squared distances from customers to their assigned centroids.

Numerator. Not applicable — a sum of squared distances. **Denominator.** Not applicable. **Window.** The feature window. **Filters.** The standardized clustering inputs. **Grain.** One clustering solution at a given number of clusters. **Unit.** Squared standardized distance, unitless. **Required fields.** The standardized feature matrix and the fitted model.

Interpretation. Its level is uninformative; its shape across candidate cluster counts is the signal. It reaches zero at one cluster per customer. **Common failure.** Citing low inertia as validation. It is a fit statistic, not evidence that the partition means anything. **Verification.** Expect it to fall and flatten. A rise at a higher cluster count is a warning to inspect initialization and convergence rather than a finding.

Owned by. Chapter 6, Section 6.8.

B.8.9 Cluster centroid

Question. What does the typical member of this cluster look like? **Formula.** The mean of the cluster's members on every clustering feature, read in standardized units and then in native units.

Numerator. Sum of feature values within the cluster. **Denominator.** Customers in the cluster. **Window.** The feature window. **Filters.** The clustering features only; profiling attributes are read after assignment, never used as inputs. **Grain.** One cluster on one feature. **Unit.** Standard deviations, then the feature's own units. **Required fields.** The standardized feature matrix and the cluster assignment.

Interpretation. A centroid of -0.8 on recency means the cluster buys more recently than average by four fifths of a standard deviation. **Common failure.** Reading a centroid as a description of every member. Cluster means are means: a monetary centroid can be propped by a few large customers, so spreads and medians travel with it. **Verification.** Recompute the centroids as group means of the standardized features and confirm they match the fitted model's. A cluster whose mean contradicts its own median is not one story.

Owned by. Chapter 6, Section 6.10.

B.8.10 Worked example

The practice file's eight customers split into two home metros. Table B.9 gives the family's descriptive measures. The suburban group holds a smaller share of customers than of revenue, which is the signature the segmentation chapter is looking for and the reason both shares are reported rather than one.

Table B.9

Segment measures on the practice file's eight customers

Segment	Customers	Share of customers	Revenue	Share of revenue	Revenue per customer
NYC Suburban	3	37.5%	601.00	45.8%	200.33
NYC Urban	5	62.5%	711.00	54.2%	142.20

Both share columns sum to 1, which is the verification those entries require. What the table cannot show at this scale is the validity question the rest of the family exists to answer: whether a partition of eight customers into two groups is a finding about customers or an artifact of the split. That is what the adjusted Rand index, the stability check, and the holdout profiling in Section B.8 are for, and none of them can be run on eight rows.

B.9 Regression and Predictive-Model Evaluation Measures

Although this appendix catalogs marketing metrics, the guide relies repeatedly on measures of analytical performance, and mixing them with operating measures is how a fit statistic ends up

on an executive scorecard. They are collected here and in Sections B.10 and B.11, clearly separated from the operating families above.

B.9.1 Pearson correlation coefficient (r)

Question. How strongly, and in which direction, do two quantities move together? **Formula.** Sample covariance of the two variables / the product of their two sample standard deviations.

Numerator. Sample covariance. **Denominator.** Product of the two sample standard deviations. **Window.** The window in which both variables were measured. **Filters.** Complete pairs. Software computes each cell of a correlation matrix over the rows complete for that pair, so one partly missing column produces a matrix whose cells describe different populations while presenting them as one exhibit. **Grain.** The grain at which both variables were measured — state it, because r computed over stores and r computed over customers are different quantities. **Unit.** Unitless, between -1 and +1. **Required fields.** Two numeric variables measured for the same observational cases and at the same grain.

Interpretation. A linear, symmetric, pairwise association. It is not directional, so 'drivers' language does not follow from it. **Common failure.** Reading a correlation near 0.9 on individual behavior as a triumph. Near-perfect r usually means the two columns are the same fact wearing two names. **Verification.** Undefined when either variable has zero variance; software returns a missing value rather than zero. Always plot the scatter: identical statistics do not mean identical evidence.

Common variants. Population covariance divides by n and sample covariance by n − 1. Correlation is invariant to the choice; covariance is not.

Owned by. Chapter 7, Section 7.3.

B.9.2 R-squared

Question. How much of the outcome's variation does this model account for on this data? **Formula.** 1 − (residual variation / total variation), where total variation is the sum of squared deviations of the outcome around its own mean.

Numerator. Residual sum of squares, subtracted from 1 as a ratio. **Denominator.** Total sum of squares around the outcome's mean. **Window.** The model's own estimation sample. **Filters.** In-sample, with an intercept, on the same observations, for nested models. **Grain.** One model. **Unit.** Proportion under the stated conditions; it is not bounded below by zero out of sample. **Required fields.** Fitted values and actual outcomes.

Interpretation. A measure of fit to this data. It is not a percentage of predictions that were right, and it is not accuracy. **Common failure.** Reading a rising R-squared as a better model. Because it cannot decrease when a variable is added, a rise is not by itself evidence of anything. An R-squared above roughly 0.9 on individual customer behavior should be treated as a defect report rather than a result. **Verification.** Confirm the model carries an intercept, since the uncentered version computed without one is not comparable and is routinely larger. Out of sample the statistic can go negative, which is a sign that it is not a percentage of anything.

Owned by. Chapter 7, Section 7.7.

B.9.3 Regression coefficient

Question. How much does the outcome differ, per unit of this predictor, holding the model's other predictors fixed? **Formula.** The fitted slope from ordinary least squares; for a dummy, the difference against the reference category.

Numerator. Not applicable — a fitted parameter. **Denominator.** Not applicable. **Window.** The estimation window. **Filters.** The estimation sample; adding or removing a predictor changes every other coefficient's meaning. **Grain.** One predictor, in outcome units per predictor unit. **Unit.** Outcome units per predictor unit; for a dummy, outcome units relative to the reference. **Required fields.** The design matrix and the outcome.

Interpretation. A coefficient is not read until its units are said aloud. Ceteris paribus means other included things equal, and a perfectly adjusted difference is still an association. **Common failure.** Reading a dummy coefficient as a level. The wrong sentence is 'the model shows Suburban Occasion customers spend $X'; the coefficient is dollars relative to the reference category, not dollars absolutely. **Verification.** Check sign, magnitude, units, and reference category against the descriptive figures the same data produced. Report the coefficient and its interval from the same fit.

Owned by. Chapter 7, Section 7.8.

B.9.4 Confidence interval for a coefficient

Question. How precisely is this association estimated? **Formula.** The fitted coefficient plus and minus a multiple of its standard error, at a stated confidence level.

Numerator. Not applicable. **Denominator.** Not applicable. **Window.** The estimation window. **Filters.** The estimation sample; robust and non-robust standard errors give different intervals from identical coefficients. **Grain.** One coefficient. **Unit.** Outcome units per predictor unit. **Required fields.** The coefficient and its standard error.

Interpretation. At large sample sizes nearly everything clears conventional significance, so the interval's width separates a great deal that the p-value separates hardly at all. **Common failure.** Quoting the coefficient from one fit and the p-value from another. Robust standard errors change the interval and leave the coefficient and the fitted values unchanged; mixing them answers two questions at once. **Verification.** The point estimate must lie inside its own interval. State the confidence level and the standard-error type.

Owned by. Chapter 7, Section 7.10.

B.9.5 p-value

Question. How surprising is this estimate, if the stated null hypothesis and model are correct? **Formula.** The probability, assuming the null and the model, of a test statistic at least as extreme as the one observed.

Numerator. Not applicable. **Denominator.** Not applicable. **Window.** The estimation window. **Filters.** The stated null hypothesis, which for a coefficient is that it equals zero given the model's other variables. **Grain.** One hypothesis test. **Unit.** Probability between 0 and 1. **Required fields.** The test statistic and its reference distribution.

Interpretation. A measure of surprise under the null. It is not the probability that a hypothesis is true, statistical significance is not practical significance, and no single p-value should by itself determine a conclusion. **Common failure.** Reading small as important. On eight thousand customers, near-trivial associations clear the conventional threshold easily. **Verification.** Report it beside the effect size and the interval, never alone.

Owned by. Chapter 7, Section 7.10.

B.9.6 Mean absolute error (MAE)

Question. By how much does a numeric prediction typically miss? **Formula.** The mean of the absolute differences between actual and predicted values.

Numerator. Sum of absolute errors. **Denominator.** Number of evaluated cases. **Window.** The outcome window of the evaluated cases. **Filters.** Evaluated out of sample; on the test set for the final grade, and on training folds during development. **Grain.** One prediction, averaged over the evaluated population. **Unit.** The target's own units — dollars for a spend target. **Required fields.** Actual and predicted values.

Interpretation. The typical miss, with every error weighted equally. On the worked example the model misses by $14.00 against a mean baseline that misses by $36.88. **Common failure.** Reporting a distance metric alone for a ranking decision, which quietly substitutes a question the analyst can answer for the one the business asked. **Verification.** Lower is better, so a value far below the baselines is the suspicious result and far above is the broken one. Compare against the strongest applicable baseline on the same folds.

Owned by. Chapter 8, Section 8.6.

B.9.7 Root mean squared error (RMSE)

Question. By how much does a numeric prediction miss, weighting large misses more heavily? **Formula.** The square root of the mean squared difference between actual and predicted values.

Numerator. Sum of squared errors, square-rooted after averaging. **Denominator.** Number of evaluated cases. **Window.** The outcome window. **Filters.** Evaluated out of sample. **Grain.** One prediction, averaged over the evaluated population. **Unit.** The target's own units. **Required fields.** Actual and predicted values.

Interpretation. The squaring is the policy: it weights a single $300 miss like nine separate $100 misses. RMSE materially above MAE announces that error is concentrated in a few large misses. On the worked example, $17.36 against an MAE of $14.00. **Common failure.** Switching to RMSE after losing on the declared metric. A candidate that wins on RMSE after losing on the declared capture and MAE has not won; the goalposts have moved. **Verification.** RMSE is always at least MAE on the same predictions. For nested ordinary-least-squares models fitted on

the same training observations, training RMSE cannot rise as terms are added, because least squares minimizes exactly this quantity; the guarantee does not extend to regularized or otherwise constrained models, and training MAE carries no corresponding guarantee under ordinary least squares.

Owned by. Chapter 8, Section 8.6.

B.9.8 Mean squared error (MSE)

Question. By how much does a numeric prediction miss, in squared units? **Formula.** The mean of the squared differences between actual and predicted values; the square of the RMSE on the same predictions.

Numerator. Sum of squared errors. **Denominator.** Number of evaluated cases. **Window.** The outcome window. **Filters.** Evaluated out of sample, on the same cases as any metric it is compared with. **Grain.** One prediction, averaged over the evaluated population. **Unit.** The square of the target's units — squared dollars for a spend target, which is why the guide reports RMSE instead. **Required fields.** Actual and predicted values.

Interpretation. Cataloged because the syllabus names it among the session outcomes, not because the guide reports it. It ranks candidate models identically to RMSE, since the square root is monotone. **Common failure.** Quoting it in a deliverable. Squared dollars are not a quantity any stakeholder can price, and the same information in readable units is one square root away. **Verification.** MSE must equal the square of the RMSE on the same predictions. Where an assignment asks for MSE, report RMSE beside it.

Owned by. Appendix B; the syllabus asks for it and no chapter defines it.

B.9.9 Baseline score

Question. What does the simplest defensible rule achieve under identical conditions? **Formula.** The mean baseline predicts the training mean for everyone; the majority-class baseline predicts the most common class; the last-value baseline predicts that the previous window repeats.

Numerator. Not applicable. **Denominator.** Not applicable. **Window.** The candidate's own evaluation window. **Filters.** Exactly the candidate's conditions: same folds, same test set, same metric. The mean must be computed from training labels, never test labels. **Grain.** Matches the candidate. **Unit.** Matches the candidate's metric. **Required fields.** Training labels, and for the last-value rule the prior window's outcome.

Interpretation. The last-value rule is not a strawman; it is the incumbent, and the sorted spreadsheet costs nothing. **Common failure.** Choosing the baseline after the results, which allows a baseline to be chosen to lose. **Verification.** Declare the baseline before fitting. Computing the mean baseline from test labels hands it a summary of the answers.

Owned by. Chapter 8, Section 8.7.

B.9.10 Margin over baseline

Question. What did the model add over the rule it must beat? **Formula.** A signed difference from baseline, stated in the direction in which improvement is positive. For loss metrics such as MAE and RMSE, baseline score − candidate score; for gain metrics such as capture, candidate score − baseline score. Computed on the same folds and the same test set, and preferably as a paired per-fold difference reported with its spread and its count of folds won.

Numerator. The improvement on the shared metric, signed so that a positive value always means the candidate is better. **Denominator.** Not applicable. **Window.** Shared evaluation window. **Filters.** Identical population and folds. **Grain.** One candidate against one baseline. **Unit.** The metric's own units. **Required fields.** Both scores, computed identically.

Interpretation. A model's demonstrated value is its margin over the strongest applicable baseline, not its raw score. An MAE of $54 is good if the best naive rule scores $80 and embarrassing if it scores $55. **Common failure.** Subtracting in the wrong direction. Because half the metrics in this appendix improve downward and half improve upward, a single subtraction order cannot serve both, and a margin reported without its direction is unreadable. Separately: a margin smaller than the fold-to-fold spread of the metric it is measured on is not yet a finding. **Verification.** State which direction represents improvement in the same sentence as the number, and confirm the sign: a model that beats its baseline must produce a positive margin under either convention. Pair fold by fold: folds differ in difficulty, so subtracting the baseline within each fold cancels that and the spread of the differences shows whether the candidate reliably wins or merely averages well.

Owned by. Chapter 8, Section 8.9.

B.9.11 Capture at a selection share

Question. Of the outcome the population produced, how much did the customers we could afford to treat account for? **Formula.** Actual outcome of the top-k customers ranked by predicted score / actual outcome of the whole evaluated population, where k is the capacity divided by the population size.

Numerator. Actual outcome-window value of the top-k ranked customers. **Denominator.** Actual outcome-window value of the whole evaluated population. **Window.** The outcome window for the actuals; the snapshot for the ranking. **Filters.** The evaluated population; ties broken deterministically so the marginal name does not change between the analysis and the campaign build. **Grain.** One ranked list, reported at the population level. **Unit.** Percentage of outcome. **Required fields.** Predicted scores, actual outcomes, a customer key, and a declared capacity.

Interpretation. The decision metric when the program can treat only a fixed number of customers. Higher is better, which is the opposite direction from MAE and RMSE. **Common failure.** Reading a capture gap as revenue the program will generate. It says the customers the model selected turned out to spend more than the customers the spreadsheet selected; it says nothing about what either group would have spent untreated. **Verification.** A constant score

carries no ranking information, so its capture is decided entirely by row order — which is why the mean baseline must break ties on a declared random draw. Compute the selection share from capacity rather than typing a proportion.

Owned by. Chapter 8, Section 8.6.

B.9.12 Selection share

Question. What fraction of the scorable population can the program actually treat? **Formula.** Program capacity / the population the model can honestly score.

Numerator. Funded treatment slots. **Denominator.** The eligible, scorable population. **Window.** As of the scoring run. **Filters.** The declared eligibility rule. **Grain.** One program, one population. **Unit.** Proportion. **Required fields.** A declared capacity and the eligible population count.

Interpretation. Capacity is a count, not a proportion: it does not grow if the eligible population turns out to be larger. The share is computed rather than typed. **Common failure.** Typing a proportion, which nobody re-derives when the population changes. **Verification.** Recompute whenever the eligibility rule changes. It is also the capture floor that random ordering achieves in expectation; any single realized random list scatters around it.

Owned by. Chapter 8, Section 8.6.

B.9.13 Worked example

Take eight customers whose actual six-month spend is known and a model that predicted it. The model's mean absolute error is $14.00 and its root mean squared error is $17.36. That RMSE sits above the MAE, which is the signal the entry describes: the error is not spread evenly but concentrated in a few larger misses.

Neither figure means anything alone. The mean baseline — predict $108.75 for everyone, computed from the training labels — posts a mean absolute error of $36.88. The model's margin is therefore $36.88 less $14.00, or $22.88, and that is the number the deliverable reports. Note the subtraction order. Mean absolute error improves downward, so its margin is baseline minus candidate; capture improves upward, so its margin is candidate minus baseline. The entry states both, because a margin reported without its direction cannot be read. A mean absolute error of $14 is a good result against a baseline of $37 and an embarrassing one against a baseline of $15.

B.10 Classification and Ranked-Targeting Measures

Classification measures are arithmetic summaries of four counts at one threshold, and almost every failure in this family comes from quoting one of them without the other three, without the threshold, or without the base rate. The family closes with the two ranked-targeting measures the retention program is actually budgeted from.

B.10.1 Base rate

Question. How common is the outcome we are trying to find? **Formula.** Positive cases / evaluated cases.

Numerator. Cases in the positive class. **Denominator.** All evaluated cases. **Window.** The outcome window. **Filters.** The declared eligible population. **Grain.** One case, reported for a population. **Unit.** Proportion. **Required fields.** The label.

Interpretation. The single most important number in a classification project and the one most reliably missing from celebrated results. It is a constructed number: a fact about the population under this label definition. **Common failure.** Reporting any classification metric without it. Accuracy, precision, and lift are uninterpretable in its absence. **Verification.** No classification result leaves the workspace without its base rate in the same exhibit. A drifting base rate silently re-prices the threshold even when the ordering holds.

Owned by. Chapter 9, Section 9.2.

B.10.2 Confusion-matrix counts (TP, FP, FN, TN)

Question. At this threshold, who did the model reach, miss, and correctly leave alone? **Formula.** Cross the actual class against the predicted class at a stated threshold; every evaluated case lands in exactly one of four cells.

Numerator. Not applicable — four counts. **Denominator.** Not applicable. **Window.** The outcome window. **Filters.** The evaluated set, at one stated threshold. **Grain.** One case; the four counts describe the evaluated set. **Unit.** Counts. **Required fields.** Predicted class at a stated threshold, and the actual label.

Interpretation. Every threshold-dependent classification metric is an arithmetic summary of these four counts, and changing the threshold changes the matrix. In campaign language: offer sent and customer was leaving; offer sent and customer was staying; no offer and customer left; no offer and customer stayed. **Common failure.** Pasting a matrix into a deck without re-adding it. A matrix that does not sum is a pasted matrix, and pasted matrices are how wrong numbers reach decks. **Verification.** Assert that the four counts sum to the evaluated set, then recompute precision and recall from the raw counts.

Owned by. Chapter 9, Section 9.6.

B.10.3 Accuracy

Question. What share of cases did the model classify correctly? **Formula.** (True positives + true negatives) / all evaluated cases.

Numerator. Correctly classified cases. **Denominator.** All evaluated cases. **Window.** The outcome window. **Filters.** The evaluated set, at one stated threshold. **Grain.** One case, reported for the evaluated set. **Unit.** Proportion. **Required fields.** Predicted class and actual label.

Interpretation. Under imbalance it is dominated by the model's performance on the class nobody is asking about. **Common failure.** Using it as a headline. On the worked ten-customer file, accuracy is 0.70 at both the 0.50 and the 0.30 threshold, while expected net value moves

from $32 to $48 — accuracy has failed to detect a difference the program will pay for. **Verification.** Report it only beside the base rate and the majority-class baseline, and compute the margin between them.

Owned by. Chapter 9, Section 9.6.

B.10.4 Precision

Question. Of the customers we treated, how many were the real thing? **Formula.** True positives / (true positives + false positives).

Numerator. True positives. **Denominator.** All predicted positives — that is, everyone treated. **Window.** The outcome window. **Filters.** The evaluated set, at one stated threshold. **Grain.** One case, reported for the evaluated set. **Unit.** Proportion. **Required fields.** Predicted class and actual label.

Interpretation. The purity of the target list. On the worked file, precision of 0.50 means half the retention budget landed on customers who were staying anyway. **Common failure.** Quoting precision and recall achieved at different thresholds, or quoting either without its threshold. Precision is undefined when nothing is treated, and must be reported as undefined rather than as zero. **Verification.** Bounded in [0, 1]. Recompute from the raw counts. A low precision at a deliberately low threshold is by design, not a failure.

Owned by. Chapter 9, Section 9.7.

B.10.5 Recall (true positive rate)

Question. Of the customers who really were leaving, how many did we reach? **Formula.** True positives / (true positives + false negatives).

Numerator. True positives. **Denominator.** All actual positives. **Window.** The outcome window. **Filters.** The evaluated set, at one stated threshold. **Grain.** One case, reported for the evaluated set. **Unit.** Proportion. **Required fields.** Predicted class and actual label.

Interpretation. Coverage of the at-risk class. On the worked file, recall of 0.67 at the 0.50 threshold means a third of the departing customers walked out untreated. **Common failure.** Reading it without precision. The two move in opposite directions as the threshold falls, so either alone can be made to look good. **Verification.** Bounded in [0, 1]. Recall and the false negative rate must sum to 1.

Owned by. Chapter 9, Section 9.7.

B.10.6 F1 score

Question. How do precision and recall trade off, weighted equally? **Formula.** 2 × precision × recall / (precision + recall) — the harmonic mean.

Numerator. Twice the product of precision and recall. **Denominator.** Their sum. **Window.** The outcome window. **Filters.** The evaluated set, at one stated threshold. **Grain.** One evaluated set. **Unit.** Unitless, between 0 and 1. **Required fields.** Precision and recall.

Interpretation. Harmonic rather than arithmetic, so it punishes imbalance between the two: precision 0.9 with recall 0.1 posts an F1 near 0.18, not a flattering 0.5. **Common failure.** Optimizing it and calling the result optimal. It weights the two errors equally, and the entire threshold apparatus exists because the two errors are not equally priced. **Verification.** Bounded in [0, 1]. Use it only where costs are genuinely unpriced, and treat that as a temporary condition.

Owned by. Chapter 9, Section 9.7.

B.10.7 False positive rate and false negative rate

Question. How are the model's errors distributed across the two classes, and across groups? **Formula.** False positive rate = false positives / all actual negatives. False negative rate = false negatives / all actual positives.

Numerator. False positives, or false negatives. **Denominator.** All actual negatives, or all actual positives. **Window.** The outcome window. **Filters.** The evaluated set at the threshold the program will actually use; reported per group in a fairness screen. **Grain.** One evaluated set, or one group within it. **Unit.** Proportion. **Required fields.** Predicted class, actual label, and the grouping attribute.

Interpretation. The pair a fairness screen compares across groups, at the operating threshold rather than on average scores. **Common failure.** Computing them on small groups. A fairness appendix that reports a 0.00 false negative rate on a group of nine has produced noise with a decimal point. **Verification.** Report the group size beside every rate. This guide screens out groups below roughly fifty cases, which is a course convention rather than a statistical boundary: the precision of a group's error rates depends on how many positives and negatives it holds, not on its total size alone.

Owned by. Chapter 9, Section 9.12.10.

B.10.8 ROC AUC

Question. How well does the model order cases, independent of any threshold? **Formula.** The area under the curve traced by the true positive rate against the false positive rate across the full threshold sweep; equivalently, the probability that a randomly chosen positive outranks a randomly chosen negative, with ties counted as half.

Numerator. Not applicable — an area. **Denominator.** Not applicable. **Window.** The outcome window. **Filters.** Cross-validated on training folds during development; once on the sealed test set for the grade. **Grain.** One model on one population. **Unit.** Unitless, between 0 and 1; 0.5 is chance. **Required fields.** Continuous scores and the actual label.

Interpretation. A ranking measure and a poor sole criterion. A model with the highest overall AUC can be inferior in the top two deciles, inferior near the operating threshold, worse calibrated, and worth less money. **Common failure.** Comparing a binary flag's AUC with a continuous score's. A flag has one nontrivial operating point and the two figures are not comparable. Separately: an AUC far above what the label definition and honest baselines would support is a leak until proven otherwise. **Verification.** Assert that a constant-score classifier

returns 0.5. Multiplying every score by a constant leaves AUC unchanged while breaking every economic calculation built on those scores.

Owned by. Chapter 9, Section 9.8.

B.10.9 Calibration (reliability) check

Question. Do the model's predicted probabilities match observed event frequencies? **Formula.** Within score bins, compare the mean predicted probability against the observed positive rate; report the difference and the count in each bin.

Numerator. Observed positives within the bin; dividing by the bin's case count gives the observed positive rate. **Denominator.** Cases in the bin. **Window.** The outcome window. **Filters.** Out-of-fold predictions on training customers only. Calibration is repaired using training evidence, never by adjusting a model repeatedly against the test set. **Grain.** One score bin. **Unit.** Probability points. **Required fields.** Predicted probabilities, actual labels, and a declared binning.

Interpretation. Distinct from discrimination, and neither implies the other. Calibration is the property that entitles a threshold derived from expected costs to be applied to raw scores. **Common failure.** Applying an economically derived threshold to uncalibrated scores. That is not a small technical error; it prices a decision in a currency the model does not print. **Verification.** Print the count per bin, since a bin holding eleven customers cannot support a calibration verdict. Read the two bins straddling the operating threshold first: a model beautifully calibrated at 0.05 and off by fifteen points at 0.30 is miscalibrated where it matters.

Owned by. Chapter 9, Section 9.8.

B.10.10 Decision threshold

Question. Above what probability is treating a customer worth the cost? **Formula.** Cost of treating / expected benefit of treating a true positive. At StyleCraft, $12 / $40 = 0.30.

Numerator. The certain cost per treated customer. **Denominator.** The expected benefit per treated true positive, itself the save rate times the value of a save. **Window.** The program's planning horizon. **Filters.** Applied to calibrated probabilities. **Grain.** One policy. **Unit.** Probability. **Required fields.** A costed treatment, a save-rate estimate, and a value-of-a-save estimate.

Interpretation. Not a property of the model but a policy choice. The software default of 0.5 embodies the assumption that the two errors cost the same and inherits its authority from nothing else. **Common failure.** Accepting the default. Four conditions ride on the derived threshold: the scores must be calibrated, the cost inputs are estimates, the save rate is a causal claim that observational machinery cannot certify, and the arithmetic assumes every customer shares one benefit figure. **Verification.** Re-derive it from the stated costs and publish the sensitivity: at a $30 benefit the cut moves to 0.40, at $60 to 0.20. A threshold that swings widely under plausible inputs is telling the program that its economics are the binding uncertainty.

Owned by. Chapter 9, Section 9.9.

B.10.11 Treatment rate (treated share)

Question. What share of the population, or of a group, does this policy treat? **Formula.** Cases scoring at or above the threshold / cases in the population or group.

Numerator. Cases at or above the threshold. **Denominator.** Cases in the population or group. **Window.** As of the scoring run. **Filters.** The stated group, at the threshold the program will actually use. **Grain.** One population, or one group within it. **Unit.** Proportion. **Required fields.** Scores, the threshold, and the grouping attribute.

Interpretation. The economic threshold answers whom it is worth treating; the budget answers how many treatments are affordable, and nothing forces the two answers to agree. **Common failure.** Setting the treated-list size first — a round number, a platform tier limit, last year's list — and never reconciling it against the gains curve. **Verification.** Report it beside the capacity. Where the threshold marks more customers than the budget covers, either cut by rank or take the gap to the meeting as a priced argument.

Owned by. Chapter 9, Section 9.9.

B.10.12 Expected net value

Question. What is this policy worth against its cost? **Formula.** True positives × benefit − treated × cost; per thousand customers, scaled by 1,000 / evaluated cases.

Numerator. Benefit earned on true positives, less cost spent on everyone treated. **Denominator.** For the per-thousand form, the evaluated case count. **Window.** The outcome window. **Filters.** One policy at one threshold. **Grain.** One policy, reported per population or per thousand customers. **Unit.** Dollars, or dollars per thousand customers. **Required fields.** The confusion-matrix counts, a costed treatment, and a benefit estimate.

Interpretation. On the worked file: treat nobody, $0; the model at 0.50, +$32; the model at 0.30, +$48; the ninety-day incumbent rule, -$8; treat everyone, exactly $0. The threshold, not the model, moved the program from $32 to $48. **Common failure.** Computing it from raw scores that were never checked for calibration. **Verification.** Per-thousand scaling keeps folds and populations of different sizes comparable. Confirm the blanket policy's arithmetic from the base rate alone as a check.

Owned by. Chapter 9, Section 9.9.

B.10.13 Lift by decile

Question. How much richer in the target class is each slice of the ranked file than a random one? **Formula.** The decile's positive rate / the base rate of the whole evaluated file.

Numerator. Positive rate within the decile. **Denominator.** The base rate of the evaluated file. **Window.** The outcome window. **Filters.** The evaluated file, ranked by score and cut into ten equal-count groups with ties broken deterministically. **Grain.** One decile of cases. **Unit.** Ratio, expressed as a multiple of the base rate. **Required fields.** Scores, labels, and a deterministic ranking.

Interpretation. A number with its baseline built in, which is why it survives in industry vocabulary. Lift should be strongest in the early deciles and weaken as the list deepens. **Common failure.** Hunting a non-monotone middle. Perfect monotonic decline is not guaranteed in a finite sample; a non-monotone middle is a finding about where the ordering blurs. What would be an error is a top decile whose lift sits near 1.0. **Verification.** Rank before cutting, so the ten groups are equal in size regardless of how many cases share a score. That is what makes lift comparable across deciles.

Owned by. Chapter 9, Section 9.10.

B.10.14 Cumulative gains

Question. Treating down to this depth of the list, what share of all positives do we reach? **Formula.** Running cumulative positives through this decile / total positives in the evaluated file.

Numerator. Cumulative count of actual positives through the decile. **Denominator.** Total positives in the evaluated file. **Window.** The outcome window. **Filters.** The evaluated file, score-ranked. **Grain.** One cumulative depth of the ranked file. **Unit.** Proportion of all positives. **Required fields.** Scores, labels, and a deterministic ranking.

Interpretation. Read against the diagonal that random targeting would trace. It reaches 1.0 at the whole file by definition. **Common failure.** Reading the curve correctly and then computing the program's return from raw scores nobody checked for calibration. **Verification.** Assert the last value is exactly 1.0. Plot the incumbent rule as a single point, not a curve: a fixed policy has one operating point.

Owned by. Chapter 9, Section 9.10.

B.10.15 Odds ratio

Question. By what factor do the odds of the outcome change per unit of this predictor? **Formula.** The exponentiated logistic coefficient.

Numerator. Not applicable — a multiplicative factor. **Denominator.** Not applicable. **Window.** The estimation window. **Filters.** The interpretive model, fitted without regularization and never used to treat anyone. **Grain.** One predictor. **Unit.** Multiplicative factor on the odds. **Required fields.** A fitted logistic coefficient.

Interpretation. Because the logistic curve is steepest at even odds, the same odds ratio moves a mid-risk customer's probability substantially and a low-risk customer's barely at all. Statements about percentage-point changes in probability are honest only with a stated starting point. **Common failure.** Reading these as causal. They are associations wearing precise multiplicative clothing: 'app users churn less' earns its verb while 'the app retains customers' does not. **Verification.** Report the interval on the odds-ratio scale. An interval straddling 1.0 is honest evidence of a weak association and should be reported as such rather than by quoting the point estimate alone.

Owned by. Chapter 9, Section 9.4.

B.10.16 Worked example

Ten scored customers, three of whom churned, give a base rate of 0.30. The retention program costs $12 per treated customer and a treated true churner is worth $40 in expectation, so the threshold is 12 divided by 40, or 0.30. Table B.10 prices the same model at that threshold and at the software default.

Table B.10

One model, two thresholds, and the measures that do and do not notice the difference

Measure	Threshold 0.50	Threshold 0.30
True positives	2	3
False positives	2	3
False negatives	1	0
True negatives	5	4
Customers treated	4	6
Accuracy	0.70	0.70
Precision	0.50	0.50
Recall	0.67	1.00
F1	0.57	0.67
Expected net value	$32	$48

Accuracy is 0.70 at both thresholds, and the majority-class baseline — predict that nobody churns — also scores 0.70. On this file accuracy is simultaneously unable to distinguish the two policies and unable to beat doing nothing, while expected net value moves from $32 to $48. The threshold, not the model, produced that gain. For comparison, the incumbent ninety-day lapse rule treats 4 customers, catches 1, and returns $-8.

Ranked down the file, the top two of the ten customers are both churners, so the first decile posts a lift of 3.33 against the base rate and the top two deciles capture 66.7% of all churners while treating 20% of the file. That is what the gains curve measures, and the budget stopping rule reads down it until the incremental capture in a decile is worth less than the decile costs to treat.

B.11 Forecast Evaluation Measures

Forecast evaluation has one measure that travels between series of different sizes and one catastrophic failure mode attached to it. The entries below keep the scale-free measure and its guard together, and add the signed measure the chapters use in code but define nowhere else. The scale-free measure and its failure at small denominators follow Hyndman and Koehler (2006).

B.11.1 Forecast error (signed)

Question. By how much, and in which direction, did the forecast miss? **Formula.** Actual − forecast, for each evaluated period.

Numerator. Actual less forecast. **Denominator.** None — it is a difference. **Window.** Each evaluated period inside the forecast horizon. **Filters.** Periods with both an actual and a forecast present. **Grain.** One period — for StyleCraft, one day. **Unit.** The series' own units, in dollars. **Required fields.** Actual and forecast series aligned on the same index.

Interpretation. Positive means the forecast undershot. The sign convention must be stated, because the season-total miss is conventionally computed the other way round. **Common failure.** Reporting only absolute error, which hides a persistent one-directional miss. **Verification.** Confirm the sign convention in the same sentence as the number.

Owned by. Chapter 10, Section 10.7.

B.11.2 Forecast bias (mean signed error)

Question. Does this method systematically run high or low? **Formula.** The mean of the signed errors across the evaluated periods.

Numerator. Sum of signed errors. **Denominator.** Number of evaluated periods. **Window.** Rolling origins, or a sealed holdout. **Filters.** Evaluated periods only. **Grain.** One method on one series. **Unit.** The series' own units, signed. **Required fields.** Actual and forecast series.

Interpretation. Where a trending, expanding series announces whether a method is systematically low. Every candidate that cannot climb is biased low on an expanding series, which is a fact about the business rather than about the software. **Common failure.** Treating a run of same-signed errors as noise. Seven positive errors in a row is a persistent undershoot, not bad luck. **Verification.** Report it beside the absolute measures; mature demand-planning organizations track signed error as seriously as absolute error.

Owned by. Chapter 10; operationalized in code, and cataloged here because the chapter defines it nowhere else.

B.11.3 Mean absolute percentage error (MAPE)

Question. How large are the forecast’s misses relative to the size of what actually happened? **Formula.** For each period, the absolute error divided by the actual, averaged across evaluated periods and multiplied by 100.

Numerator. Absolute error in each period. **Denominator.** That period’s actual value. **Window.** The evaluated periods. **Filters.** Requires strictly positive actuals; disqualified for any series that touches or approaches zero. **Grain.** One period, then averaged. **Unit.** Percentage. **Required fields.** Actual and forecast series.

Interpretation. Scale-free, which makes it planning’s shared currency for forecast accuracy. **Common failure.** Reporting it over a near-zero denominator. On the worked week a single storm-closed day of $400 turns a forecast whose typical daily miss is under three percent into a reported MAPE of 259.6 percent — one denominator, not seven bad forecasts, produced the

number, and six of the seven days were still forecast within ten percent. **Verification.** Assert that every denominator is strictly positive and refuse to return a number otherwise. Print the minimum denominator and a count of thin days beside every MAPE displayed. The sentence that must accompany it: the denominator, not the forecast, produced the number.

Owned by. Chapter 10, Section 10.7.

B.11.4 Accuracy by horizon

Question. How does forecast error grow as the forecast reaches further out? **Formula.** MAE and MAPE computed separately within each block of the horizon, pooled across forecast origins.

Numerator. Errors within the horizon block. **Denominator.** Periods in the block times the number of origins. **Window.** Equal blocks of the forecast horizon, from each origin. **Filters.** The selection window's origins only. **Grain.** One horizon block, pooled across origins. **Unit.** The series' units, and percentage. **Required fields.** Actual and forecast series from multiple origins.

Interpretation. Expected error rises with horizon; realized error need not rise at every step. A flat or falling curve is a signal to inspect rather than a diagnosis: it can mean the origins sit too close together or the horizon is too short, and it can mean the series is genuinely no harder to forecast at range. **Common failure.** Quoting the count of daily errors as though they were independent observations, which overstates a weekly row's stability sevenfold. Report the number of independent origins separately from the number of daily errors. **Verification.** Report the spread across origins at each horizon beside the mean, and print whether the curve rises monotonically rather than assuming it.

Owned by. Chapter 10, Section 10.7.

B.11.5 Naive and seasonal-naive baselines

Question. What accuracy is available for free, before any model is fitted? **Formula.** The naive rule extends the last observed value flat across the horizon. The seasonal-naive rule repeats the last full seasonal period.

Numerator. Not applicable. **Denominator.** Not applicable. **Window.** The same evaluation window as the candidate. **Filters.** A declared seasonal period — seven for daily data with a weekly rhythm. **Grain.** The series' own grain. **Unit.** The series' units. **Required fields.** The historical series and a declared period.

Interpretation. At a seven-day lag the seasonal-naive rule repeats last week, not last December. It carries the weekly rhythm for free and cannot climb a trend. **Common failure.** Reporting a fitted model's accuracy without its margin over the seasonal-naive rule. A forecaster who cannot state that margin does not know whether the model is contributing anything. **Verification.** Compute both on identical origins. A margin smaller than the noise band is not yet a finding.

Owned by. Chapter 10, Section 10.5.

B.11.6 Prediction interval and coverage

Question. What range of outcomes does the model consider plausible, and did reality land inside it? **Formula.** Nominal coverage is the interval's stated probability. Realized coverage is the share of evaluated periods that landed inside the band.

Numerator. Evaluated periods falling inside the band. **Denominator.** Number of evaluated periods. **Window.** The evaluated horizon, from a stated number of forecast origins. **Filters.** Evaluated periods only. **Grain.** One period; also reported by horizon block where the periods permit it. **Unit.** Percentage for coverage; the series' units for width. **Required fields.** Forecast bounds and actuals.

Interpretation. Coverage alone is badly incomplete, because an arbitrarily wide interval achieves perfect coverage while saying nothing. Publish the width beside the coverage so that no one can buy the second by inflating the first. **Common failure.** Adding up the edges of interval bounds to get an interval for a total. A quantile of a sum is not the sum of the quantiles: aggregate the simulated paths first and take quantiles second. **Verification.** Report nominal coverage, realized coverage, mean width, the number of evaluated periods, and the number of independent origins together. Coverage of 94 percent on sixty-one dependent daily observations is a reason to investigate, not a verdict.

Owned by. Chapter 10, Section 10.8.

B.11.7 Worked example

Seven days of chain revenue against a forecast. In an ordinary week the mean absolute error is $314.29 and the mean absolute percentage error is 2.75 percent, with a minimum denominator of $9,400. Now close a store on the Tuesday: actual revenue for that day falls to $400 while the forecast still reads $7,600.

The mean absolute error moves honestly, to $1,314.29. The mean absolute percentage error explodes to 259.59 percent — because that single day's absolute percentage error is 7,200 divided by 400, or 1,800 percent, one term of a seven-term average. 6 of the 7 days were still forecast within ten percent. One denominator, not seven bad forecasts, produced the number, and that is the sentence that must accompany every mean absolute percentage error this appendix's entry permits.

B.12 Experiment and Causal-Evidence Measures

Experimental measures are the only ones in this appendix that license a causal verb, and they license it narrowly: among the tested population, in the tested window, at the tested dose. The family includes the observational comparison these measures replace, so that the two can be told apart on sight. The reporting conventions follow Kohavi et al. (2020).

B.12.1 Treatment and control conversion rate

Question. What share of each randomized arm converted? **Formula.** Assigned customers with at least one qualifying order in the outcome window / assigned customers in the arm.

Numerator. Assigned customers with at least one qualifying order in the window. **Denominator.** All assigned customers in the arm, whether or not the message reached them. **Window.** A stated number of days following each customer's assignment. **Filters.** Intention to treat: every randomized customer with an observed outcome stays in the arm they were assigned to. A delivery failure is operational evidence and is not by itself grounds for removing anyone. **Grain.** One customer, the unit of randomization. **Unit.** Proportion, reported as a percentage. **Required fields.** variant, converted, assign_date, customer_id.

Interpretation. A customer who converts three times in the window is one conversion in a rate whose denominator is customers. **Common failure.** Switching the numerator to orders, which changes the question without saying so. Separately, filtering on anything measured after assignment breaks the randomization that the whole design bought. **Verification.** Bounded in [0, 1]. Reconcile the assignment flow: randomized, outcome observed, and included in the primary analysis, with any shortfall named.

Owned by. Chapter 11, Section 11.6.

B.12.2 Absolute lift

Question. How much did the treatment change the outcome, on the metric's own scale? **Formula.** Treatment rate − control rate.

Numerator. The difference between the two arm rates. **Denominator.** None — it is a difference. **Window.** The stated outcome window. **Filters.** Intention to treat, both arms. **Grain.** One customer, reported for the experiment. **Unit.** Percentage points, when the metric is a rate. **Required fields.** Both arm rates and both arm sizes.

Interpretation. The quantity that scales to business volume: multiplying it by the number of customers reached gives the incremental count directly. StyleCraft's test recovers +1.80 points against a 6.00 percent control rate. **Common failure.** Reporting it without the control rate beside it, or reporting only the relative figure. Thirty percent of an unstated base is not a quantity. **Verification.** Report the absolute lift first in percentage points with the control rate stated beside it, then the relative lift, then the translation into incremental conversions and dollars.

Owned by. Chapter 11, Section 11.6.

B.12.3 Relative lift

Question. How large was the change against its own baseline? **Formula.** Absolute lift / control rate.

Numerator. The absolute lift. **Denominator.** The control rate. **Window.** The stated outcome window. **Filters.** Intention to treat, both arms. **Grain.** One experiment. **Unit.** Percentage change — not percentage points. **Required fields.** Both arm rates.

Interpretation. Systematically the larger and more persuasive of the two whenever the baseline rate is small. StyleCraft's 1.80-point lift is a 30 percent relative lift. **Common failure.** Treating transportability as a statistical fact. A relative lift transfers to a different baseline only if the effect really is multiplicative in the baseline, which is an empirical hope rather than a property of the statistic. **Verification.** Compute its interval as a ratio, never by rescaling the difference interval by the observed control rate; that treats an estimate as a known constant and produces an interval that is too narrow and wrongly centered.

Owned by. Chapter 11, Section 11.6.

B.12.4 Confidence interval on a difference in proportions

Question. What range of true effects is consistent with what we observed? **Formula.** estimate ± $z^* \times$ SE, with $z^* = 1.96$ for a conventional 95 percent interval, where the standard error of a difference in two proportions is built from each arm's own rate and size.

Numerator. Not applicable. **Denominator.** Not applicable. **Window.** The stated outcome window. **Filters.** Intention to treat, both arms. **Grain.** One experiment. **Unit.** Percentage points. **Required fields.** Both arm conversion counts and both arm sizes.

Interpretation. StyleCraft's interval runs from 0.81 to 2.79 percentage points — a range of more than three to one in business terms, from a point estimate that looks precise. **Common failure.** Reporting the point estimate alone, or presenting an interval far narrower than the arm sizes support, which usually means the analysis slipped to a finer grain than the randomization. **Verification.** The unpooled standard error uses each arm's own rate and is the one behind the interval; the pooled version belongs to the test statistic. On StyleCraft's numbers they differ in the fourth decimal, which is why nobody notices when the wrong one is used.

Owned by. Chapter 11, Section 11.6.

B.12.5 Treatment dose

Question. How many times, and when, does a treated unit receive the intervention? **Formula.** A declared count and timing of exposures, stated as part of the intervention's definition.

Numerator. Not applicable — a design parameter. **Denominator.** Not applicable. **Window.** The assignment and outcome windows. **Filters.** Treated arm. **Grain.** One treated customer. **Unit.** Count of exposures. **Required fields.** A registration entry; it cannot be recovered from the outcome data.

Interpretation. The dose fixes the cost per assigned customer and therefore the break-even lift. An experiment run at one alert per drop licenses a rollout at one alert per drop. **Common failure.** Rolling out at a different frequency. That is a different intervention whose economics and whose response must both be re-derived; at two alerts StyleCraft's break-even doubles to 0.86 points and the same measured interval no longer clears it at its lower end. **Verification.** Declare it before the test runs, and quote it in the same sentence as any economic conclusion.

Owned by. Chapter 11, Section 11.7.

B.12.6 Break-even lift

Question. What lift would this program need just to pay for itself? **Formula.** Cost per assigned customer / contribution per incremental conversion, where cost per assigned customer is cost per exposure times the dose.

Numerator. Cost per assigned customer. **Denominator.** Contribution per incremental conversion. **Window.** The program's own window. **Filters.** Assigned customers, at the tested dose. **Grain.** One program at one dose. **Unit.** Percentage points. **Required fields.** A costed exposure, a declared dose, and a contribution estimate.

Interpretation. StyleCraft's Drop Alert costs $0.12 per assigned customer against $28 of contribution per incremental conversion, so the break-even lift is 0.43 percentage points. **Common failure.** Comparing it against the point estimate. The comparison is made against the whole interval: if the lower bound clears the line the program pays even in the pessimistic world the data still permits; if the line falls inside the interval, the honest report is that the evidence does not settle the question. **Verification.** Undefined until the dose is stated. Publish a sensitivity grid across plausible contribution figures and doses, flagging any cell that falls inside the confidence interval.

Owned by. Chapter 11, Section 11.7.

B.12.7 Incremental conversions and net contribution

Question. What is the program worth per thousand customers at the tested dose? **Formula.** Incremental conversions = absolute lift expressed as a proportion × 1,000; equivalently, absolute lift expressed in percentage points × 10. Net contribution = incremental conversions × contribution per conversion − 1,000 × cost per assigned customer.

Numerator. Contribution earned on incremental conversions, less program cost. **Denominator.** Not applicable; the per-thousand scaling is the reporting convention. **Window.** The program's own window. **Filters.** Enrolled customers at the tested dose. **Grain.** One thousand customers enrolled. **Unit.** Dollars per thousand customers. **Required fields.** The lift and its interval, a contribution estimate, and a costed dose.

Interpretation. On StyleCraft's numbers: at the point estimate, 18.0 incremental conversions worth $504.00 against $120.00 of cost, a net of $384.00; at the interval's pessimistic edge, $106.80; at its optimistic edge, $661.20. The decision does not hinge on which end proves true. **Common failure.** Attaching the phrase 'at any scale'. The arithmetic assumes cost is linear in customers reached, contribution per conversion is stable, no fixed implementation cost amortizes, no capacity or inventory constraint binds, unsubscribe and fatigue costs are zero, and response does not diminish as the program scales. Several of those are known to weaken at rollout. **Verification.** Compute at all three points of the interval, not only the estimate. Scrutinize the contribution figure: it is an average across observed orders, and the orders an incremental converter places need not carry the same mix or margin.

Owned by. Chapter 11, Section 11.7.

B.12.8 Observational difference (targeted versus untargeted)

Question. How did the customers who received the program differ from those who did not? **Formula.** Rate among those who received − rate among those who did not.

Numerator. The difference between the two observed group rates. **Denominator.** None — it is a difference. **Window.** Stated period. **Filters.** Groups formed by whatever mechanism assigned the treatment in the record — which is precisely what is unknown. **Grain.** One customer, reported for the two groups. **Unit.** Percentage points. **Required fields.** The outcome and a treatment indicator.

Interpretation. A true fact about the records and a false answer to the causal question. StyleCraft's observational difference of 5.30 points decomposes exactly into a 1.80-point within-stratum difference and a 3.50-point composition component — two thirds of the headline was who got the treatment. **Common failure.** Reporting it with a causal verb. Before quoting any comparison as evidence about a decision, state how each side was selected and what would have made the two sides differ even if the treatment had never existed. **Verification.** Where a confounder happens to be recorded, standardize on the treated group's own mix and report the composition component beside the adjusted difference. The two components must sum exactly to the naive difference.

Common variants. An adjusted difference, standardized on measured covariates, is stronger evidence and is still not a causal effect: adjustment corrects only the imbalances the analyst can name.

Owned by. Chapter 11, Section 11.3.

B.12.9 Statistical power and minimum detectable effect

Question. Is this test large enough to find an effect worth acting on? **Formula.** Power is the probability of detecting an effect of a stated size at a stated significance level and sample size; the minimum detectable effect is the smallest effect the design can reliably detect.

Numerator. Not applicable. **Denominator.** Not applicable. **Window.** The planned test duration. **Filters.** Per arm, two-sided, at a declared significance level. **Grain.** One customer per arm. **Unit.** Probability for power; percentage points for the minimum detectable effect. **Required fields.** A baseline rate from history, a declared minimum detectable effect, and a target power.

Interpretation. The minimum detectable effect is a business judgment, not a statistic. StyleCraft registers 1.5 points, needs 4,379 per arm, and runs 5,000. **Common failure.** Running underpowered and reading a null as no effect. Required sample size grows roughly with the inverse square of the smallest effect worth detecting — at 0.5 points the same design would need 36,778 per arm — and an underpowered test's significant results systematically overstate the effect's size. **Verification.** Register the baseline, the minimum detectable effect, the significance level, and the power before the test runs.

Owned by. Chapter 11, Section 11.5.

B.12.10 Difference-in-differences

Question. What changed in the treated group beyond what changed anyway? **Formula.** The treated group's change from before to after, less the comparison group's change over the same periods.

Numerator. Treated change less comparison change. **Denominator.** None — it is a difference of differences. **Window.** Equal pre and post periods for both groups. **Filters.** A comparison group chosen before the outcome is seen. **Grain.** The unit at which the treatment was applied — market, store, or customer. **Unit.** The outcome's own units. **Required fields.** Outcomes for both groups in both periods.

Interpretation. Rests on parallel trends, which cannot be verified, only supported. A pre-post report with no comparison group would have credited the campaign with the whole treated change. **Common failure.** Reporting the pre-post difference alone. It is a causal claim with no comparison group, whose implicit counterfactual is that nothing else would have changed. **Verification.** Plot both series over a long pre-period and inspect whether they moved together before the intervention. License only the sentence that begins 'under the stated parallel-trends assumption'.

Owned by. Chapter 11, Section 11.9.

B.12.11 Worked example

StyleCraft's Drop Alert test randomized 5,000 customers to each arm. The control arm converted at 6.00% and the treatment arm at 7.80%, an absolute lift of 1.80 percentage points and a relative lift of 30%. The standard error of the difference is 0.5066 percentage points, giving a 95 percent confidence interval on the difference of 0.81 to 2.79 percentage points.

The decision is made against the whole interval rather than the point estimate. At one alert per customer the program costs $0.12 per assigned customer against $28 of contribution per incremental conversion, so the break-even lift is 0.43 percentage points. The entire interval sits above that line, so the program pays even in the pessimistic world the data still permits. Per thousand customers enrolled, the point estimate is 18.0 incremental conversions worth $384.00 net; the interval's edges give $106.80 and $661.20.

Change the dose and the conclusion changes on identical evidence. At two alerts the break-even lift doubles to 0.86 percentage points, and the interval's lower bound of 0.81 no longer clears it. This is why the treatment dose is an entry in this family rather than a footnote to one.

B.13 Dashboard and Decision Measures

This section creates no new formulas. It catalogs the measures that exist because a decision elevated them, and it records the apparatus by which any entry above becomes a dashboard KPI.

B.13.1 Weekly contribution

Question. How much contribution does this store generate each week toward covering its fixed footprint? **Formula.** Sum of (line_revenue − line_cost) within store and store-age week.

Numerator. Contribution on the store's lines in the week. **Denominator.** None — it is a total. **Window.** One store-age week, on a scaffold that carries every open observable week. **Filters.** The declared store population; a week that is open and observable but carried no orders is zero, not missing. **Grain.** One store-week. **Unit.** Dollars per week. **Required fields.** line_revenue, line_cost, store_id, order_date, opening_date.

Interpretation. The unit in which a new store's ramp is read against the fixed cost it must eventually carry. **Common failure.** Dropping the zero-order weeks. That removes the weakest observations from any benchmark computed over them and biases the benchmark upward, in the direction that makes a new store's ramp look worse than it is. **Verification.** Reconcile the sum across store-weeks to the certified revenue and cost totals for the same population.

Owned by. Chapter 13, Section 13.12.1.

B.13.2 Four-week moving average of weekly contribution

Question. What is this store's underlying weekly contribution, with single-week noise damped? **Formula.** The mean of weekly contribution over the current store-age week and the previous three, computed within store, ordered by store age, and returning missing until four observable weeks exist.

Numerator. Contribution over the trailing four store-weeks. **Denominator.** Four store-weeks. **Window.** Store-age weeks, from each store's own opening. **Filters.** Grouped by store; ordered by store age ascending. **Grain.** One scaffold store-week. **Unit.** Dollars per week. **Required fields.** Weekly contribution on a complete store-week scaffold.

Interpretation. Every one of the five properties — the grouping, the ordering, the measure, the window, and the behavior before the window fills — has to be stated, because a moving calculation with any of them left to a default computes a different number and the difference is invisible in the output. **Common failure.** Computing it on a series with missing weeks, which averages the wrong four weeks. **Verification.** Confirm it returns missing for the first three weeks of each store rather than averaging a partial window.

Owned by. Chapter 13, Section 13.12.2.

B.13.3 Weekly fixed footprint (the coverage threshold)

Question. How much weekly contribution does a store need to carry its own fixed cost? **Formula.** The declared annual fixed footprint / 52. At StyleCraft, $92,000 / 52 = $1,769.23.

Numerator. The annual fixed footprint, a finance planning assumption. **Denominator.** 52 weeks. **Window.** Annual, expressed weekly. **Filters.** The population the finance assumption covers — new suburban and resort stores, with no figure supplied for the flagship format. **Grain.** One store-week; drawn as a constant reference line. **Unit.** Dollars per week. **Required fields.** A finance assumption, attributed to finance and declared before results are inspected.

Interpretation. Must be attributed to finance rather than to the data. A constant is correct here and a computed reference would be wrong. **Common failure.** Choosing it after the crossing weeks are known. A threshold chosen after the fact is not a benchmark but a description. **Verification.** Label the line with what it is, not with its value: weekly contribution carrying the $92,000 annual footprint, finance assumption, suburban and resort.

Owned by. Chapter 13, Section 13.12.1.

B.13.4 Weekly revenue at coverage

Question. How much weekly revenue does the footprint threshold correspond to? **Formula.** Weekly fixed footprint / contribution margin, on the same population. At StyleCraft, $1,769.23 / 0.46 = $3,846.15.

Numerator. The weekly fixed footprint. **Denominator.** The contribution margin of the same population. **Window.** Weekly. **Filters.** Suburban and resort stores, for both the threshold and the margin. **Grain.** One store-week. **Unit.** Dollars of revenue per week. **Required fields.** The threshold and a population-matched contribution margin.

Interpretation. A within-population conversion, and the only place that population's margin may be used. **Common failure.** Applying this population's margin to a chain-level quantity, or the chain's margin to this one. **Verification.** If the contribution margin moves by more than a point, the revenue threshold changes and every figure quoting it must be rebuilt before any title uses it.

Owned by. Chapter 13, Section 13.1.

B.13.5 First observed weekly coverage

Question. When did this store first cover its fixed footprint in a single week? **Formula.** The earliest store-age week in which weekly contribution reaches the declared weekly threshold.

Numerator. Not applicable — a week index. **Denominator.** Not applicable. **Window.** Store-age weeks from the store's own opening. **Filters.** Stores with enough observed history for the question to be meaningful. **Grain.** One store. **Unit.** Store-age week number. **Required fields.** Weekly contribution and the declared threshold.

Interpretation. A coverage milestone, not a payback period. It returns the same answer with or without the scaffold, because a week with no orders cannot be the first week above a positive threshold. **Common failure.** Reading it off a plotted line rather than computing it. A first-coverage week of 0 or 1 usually means the threshold was applied to revenue rather than to contribution. **Verification.** Compute it; do not read it. Confirm it is no later than the four-week-average coverage week for the same store.

Owned by. Chapter 13, Section 13.12.1.

B.13.6 First four-week-average coverage

Question. When did this store first reach its footprint on a smoothed basis? **Formula.** The earliest store-age week in which the four-week moving average of weekly contribution reaches the declared weekly threshold.

Numerator. Not applicable — a week index. **Denominator.** Not applicable. **Window.** Store-age weeks; the average returns missing until four observable weeks exist. **Filters.** The mature store population. **Grain.** One store. **Unit.** Store-age week number. **Required fields.** The four-week moving average and the declared threshold.

Interpretation. The go/no-go dashboard's decision measure, preferred to the raw crossing because it is harder to produce by accident. **Common failure.** Describing a smoothed first crossing in the language of the stricter one. It is not a claim that the store held coverage, stayed above the line, or established durable coverage. **Verification.** A four-week-average coverage week earlier than the same store's observed week is arithmetically impossible; assert it. The gap between the two is itself readable: a store whose weeks are close crossed on strength its neighboring weeks shared.

Owned by. Chapter 13, Section 13.12.1.

B.13.7 Predecessor range

Question. What range did comparable stores occupy at this same store age? **Formula.** The minimum and maximum of the four-week moving average across the mature comparable stores, computed separately at each store-age week.

Numerator. Not applicable — a pair of bounds. **Denominator.** Not applicable. **Window.** The store-age weeks the mature stores cover. **Filters.** The mature comparable stores only, on the smoothed measure. **Grain.** One store-age week, as a pair of bounds. **Unit.** Dollars per week. **Required fields.** The four-week moving average for each comparable store, by store age.

Interpretation. The observed minimum-to-maximum spread of a handful of stores, not a statistical prediction interval. **Common failure.** Comparing a smoothed band against a raw series. Identical axes do not make differently smoothed quantities comparable. A band that is flat across store age means the bounds were computed at table scope rather than per week. **Verification.** A minimum computed across stores at each week is biased upward wherever one store's zero weeks were dropped. Confirm the scaffold is complete before computing it.

Owned by. Chapter 13, Section 13.12.2.

B.13.8 Mature-store annual contribution

Question. What does a comparable store contribute in a full year? **Formula.** Trailing 52-week contribution for each mature comparable store over the year ending on the analysis date, then averaged across those stores.

Numerator. Trailing 52-week contribution per store. **Denominator.** The count of mature comparable stores. **Window.** The 52 weeks ending on the declared analysis date. **Filters.** The

mature comparable stores only. **Grain.** One store-year, averaged. **Unit.** Dollars per year. **Required fields.** Weekly contribution over a full trailing year.

Interpretation. StyleCraft carries approximately $137,000, and uses it as the denominator that turns a contribution interval into store-equivalents. **Common failure.** Annualizing a weekly mean, which inherits whatever seasonality the sampled weeks happened to carry. **Verification.** Compute from a genuine trailing 52 weeks per store rather than from a weekly average multiplied by 52.

Owned by. Chapter 13, Section 13.9.

B.13.9 Interval expressed in store-equivalents

Question. How large is the remaining forecast uncertainty in a unit the reader already understands? **Formula.** The width of the open forecast range / mature-store annual contribution, end to end; half that around the estimate.

Numerator. The end-to-end width of the interval, in dollars of contribution. **Denominator.** Mature-store annual contribution, in dollars of contribution per year. **Window.** The still-open portion of the forecast horizon. **Filters.** Chain in the numerator; the mature comparable stores in the denominator — a mismatch that must be named on the canvas. **Grain.** One horizon total. **Unit.** Mature-store-equivalents. **Required fields.** The forecast interval bounds and the mature-store annual contribution.

Interpretation. StyleCraft's $240,000 open range is about 1.8 stores end to end, or roughly plus or minus 0.9 stores around the estimate. **Common failure.** Two wrong conversions are available and both look plausible: multiplying the interval by the contribution margin, and dividing it by the annual fixed footprint. Neither belongs on a canvas. There is no margin in the correct arithmetic, because both quantities are already contribution. **Verification.** Show the conversion in steps on the canvas rather than asserting the result, and name the denominator as well as using it. Where a conversion requires an assumption that cannot be established, change what is forecast rather than disclosing the assumption and proceeding.

Owned by. Chapter 13, Section 13.9.

B.13.10 Worked example

The go/no-go dashboard's threshold is a worked example of the whole section. Finance carries an annual fixed footprint of $92,000 for a new suburban or resort store. Divided by 52 that is $1,769.23 of weekly contribution, which is the constant the primary view draws. Divided again by the population's 46% contribution margin it is $3,846.15 of weekly revenue — a within-population conversion, and the only place that margin may be used.

The forecast interval is the harder example. The still-open portion of the fiscal-year outlook spans $240,000 of contribution end to end. Divided by a mature store's annual contribution of $137,000, that is 1.8 store-equivalents end to end, or roughly plus or minus 0.9 stores around the estimate. Note what is absent from that arithmetic: there is no margin in it, because both quantities are already contribution. Two wrong conversions are available and both look plausible

— multiplying the range by the contribution margin, or dividing it by the fixed footprint — and neither belongs on a canvas.

Table B.11 records the apparatus by which a metric from any earlier section becomes a measure on that dashboard. It is worth noting what the apparatus does not include: the chapters name no per-metric owner. Ownership in Chapter 13 attaches to the decision, to the analyst who signs the artifact, and to the artifact's refresh process — which is a deliberate narrowing, not an omission.

Table B.11

How a metric becomes a measure on a decision artifact

Element	What it requires	Where the guide sets it
Tier	Primary, supporting, or diagnostic, by proximity to the decision — and expressed in size and position, because a hierarchy stated in the analyst's head is not communicated at all	Section 13.4
Reader's question	The question this view answers for this reader, written out	Table 13.2
Measure and grain	The measure, and what one mark represents	Table 13.8
Reference carried	The baseline, target, comparison group, or benchmark drawn on the view. If the comparison is against a target, draw the target.	Section 12.9
Decision threshold	An externally supplied cut, declared before the results are inspected, because a threshold chosen afterward is not a benchmark but a description	Sections 12.9 and 13.12.1
Review cadence	When the artifact refreshes and when the decision is revisited	Section 13.13
Reversal condition	Two or three observable events that would change the recommendation, each with a date or a threshold attached	Section 13.12.7

B.14 Metric Index and Exclusion Register

The catalog above is ordered by family. This section supplies the two indexes a reader needs when the family is not obvious: an alphabetical list of every cataloged metric with the section that defines it and the chapter that owns it, and a register of the metrics that were considered and deliberately left out. The second list is the more important of the two. A dictionary that silently omits a metric looks identical to one that never considered it, and a reader who cannot find a definition here needs to know which of those two situations they are in.

B.14.1 Alphabetical index of cataloged metrics

Table B.12 lists every entry in Sections B.2 through B.13. Where the owning chapter column reads Appendix B, the measure is used by the guide or required by a project but is defined nowhere else, and this appendix is its only source.

Table B.12

Every cataloged metric, its entry, and its owning chapter

Metric	Entry	Owned by	Metric	Entry	Owned by
Absolute lift	B.12.2	Ch. 11 §11.6	Mean absolute error (MAE)	B.9.6	Ch. 8 §8.6
Accuracy	B.10.3	Ch. 9 §9.6	Mean absolute percentage error (MAPE)	B.11.3	Ch. 10 §10.7
Accuracy by horizon	B.11.4	Ch. 10 §10.7	Mean squared error (MSE)	B.9.8	Appendix B
Adjusted Rand index (ARI)	B.8.7	Ch. 6 §6.10	Monetary value (RFM)	B.3.9	Ch. 6 §6.5
App-user share	B.6.8	Appendix B	Naive and seasonal-naive baselines	B.11.5	Ch. 10 §10.5
Attributed orders	B.5.10	Ch. 5 §5.8	Net contribution after advertising spend	B.2.8	Ch. 1, Lab 1.2
Attribution (conversion) window	B.5.11	Ch. 3 §3.12.3	Net revenue	B.2.4	Ch. 4 §4.8
Average of group average order values (do not use)	B.3.2	Ch. 5 §5.3	Number of customers	B.2.11	Ch. 4 §4.8
Average order value (AOV)	B.3.1	Ch. 5 §5.7	Number of orders	B.2.10	Ch. 4 §4.8
Average order value decomposition	B.3.5	Ch. 5 §5.7	Observational difference (targeted versus untargeted)	B.12.8	Ch. 11 §11.3
Baseline score	B.9.9	Ch. 8 §8.7	Occasionwear share	B.3.16	Ch. 5 §5.7
Base rate	B.10.1	Ch. 9 §9.2	Odds ratio	B.10.15	Ch. 9 §9.4
Bounce rate	B.6.2	Appendix B	Open rate	B.6.3	Ch. 5 §5.7
Break-even lift	B.12.6	Ch. 11 §11.7	Opt-in rate	B.6.7	Ch. 3 §3.4.1
Calibration (reliability) check	B.10.9	Ch. 9 §9.8	Orders per customer (purchase frequency)	B.3.7	Ch. 6 §6.5
Capture at a selection share	B.9.11	Ch. 8 §8.6	Orders per month	B.3.12	Ch. 6 §6.6
Churn label	B.4.4	Ch. 9 §9.3	Pearson correlation coefficient (r)	B.9.1	Ch. 7 §7.3

Metric	Entry	Owned by	Metric	Entry	Owned by
Churn rate (descriptive)	B.4.3	Ch. 5 §5.7	Percentage change and percentage-point change	B.2.14	Ch. 13 §13.15
Click rate	B.6.4	Ch. 5 §5.7	Post-click conversion rate	B.5.4	Ch. 1 §1.10
Clicks	B.5.2	Ch. 1 §1.10	Precision	B.10.4	Ch. 9 §9.7
Click-through rate (CTR)	B.5.3	Ch. 1 §1.10	Predecessor range	B.13.7	Ch. 13 §13.12.2
Click-to-open rate (CTOR)	B.6.5	Appendix B	Prediction interval and coverage	B.11.6	Ch. 10 §10.8
Cluster centroid	B.8.9	Ch. 6 §6.10	p-value	B.9.5	Ch. 7 §7.10
Cohort retention rate (period-active)	B.4.5	Ch. 5 §5.8	Recall (true positive rate)	B.10.5	Ch. 9 §9.7
Cohort size	B.4.6	Ch. 5 §5.8	Recency	B.3.10	Ch. 6 §6.5
Confidence interval for a coefficient	B.9.4	Ch. 7 §7.10	Regression coefficient	B.9.3	Ch. 7 §7.8
Confidence interval on a difference in proportions	B.12.4	Ch. 11 §11.6	Relative lift	B.12.3	Ch. 11 §11.6
Confusion-matrix counts (TP, FP, FN, TN)	B.10.2	Ch. 9 §9.6	Repeat-order share	B.4.2	Ch. 12 §12.13.2
Contribution	B.2.6	Ch. 13 §13.12.1	Repeat purchase rate	B.4.1	Ch. 3 §3.10.3
Contribution margin	B.2.7	Ch. 13 §13.12.1	Return on ad spend (ROAS)	B.5.9	Ch. 1 §1.10
Cost per click (CPC)	B.5.5	Ch. 5 §5.7	Revenue growth rate	B.2.12	Ch. 2 §2.7
Cost per conversion	B.5.7	Ch. 1, Exercise 1.4	Revenue per customer	B.3.6	Ch. 5 §5.7
Cost per thousand impressions (CPM)	B.5.6	Appendix B	Revenue per exposure-month	B.3.13	Ch. 6 §6.6
Cumulative gains	B.10.14	Ch. 9 §9.10	Revenue per unit (realized price per unit)	B.3.4	Ch. 5 §5.7
Cumulative (overall) conversion rate	B.7.3	Ch. 5 §5.8	Revenue share by decile	B.3.17	Ch. 5 §5.9
Customer acquisition cost (CAC)	B.5.8	Ch. 5 §5.7	RFM grid region	B.8.5	Ch. 6 §6.5

Metric	Entry	Owned by	Metric	Entry	Owned by
Decision threshold	B.10.10	Ch. 9 §9.9	RFM scores (R, F, M)	B.8.4	Ch. 6 §6.5
Delivery rate	B.6.1	Appendix B	ROC AUC	B.10.8	Ch. 9 §9.8
Difference-in-differences	B.12.10	Ch. 11 §11.9	Root mean squared error (RMSE)	B.9.7	Ch. 8 §8.6
Discount dollars	B.2.3	Appendix B	R-squared	B.9.2	Ch. 7 §7.7
Discount share	B.3.14	Ch. 6 §6.6	Segment share of customers	B.8.2	Ch. 6 §6.11
Drop-off rate	B.7.4	Appendix B	Segment share of revenue	B.8.3	Ch. 6 §6.11
Expected net value	B.10.12	Ch. 9 §9.9	Segment size	B.8.1	Ch. 6 §6.10
Exposure months	B.3.11	Ch. 6 §6.6	Selection share	B.9.12	Ch. 8 §8.6
F1 score	B.10.6	Ch. 9 §9.7	Silhouette score	B.8.6	Ch. 6 §6.9
False positive rate and false negative rate	B.10.7	Ch. 9 §9.12.10	Stage count	B.7.1	Ch. 5 §5.8
First four-week-average coverage	B.13.6	Ch. 13 §13.12.1	Stage-to-stage conversion rate	B.7.2	Ch. 5 §5.8
First observed weekly coverage	B.13.5	Ch. 13 §13.12.1	Statistical power and minimum detectable effect	B.12.9	Ch. 11 §11.5
Forecast bias (mean signed error)	B.11.2	Ch. 10	Store share	B.3.15	Ch. 6 §6.6
Forecast error (signed)	B.11.1	Ch. 10 §10.7	Tenure	B.4.7	Ch. 6 §6.6
Four-week moving average of weekly contribution	B.13.2	Ch. 13 §13.12.2	Time to second purchase	B.4.8	Ch. 5 §5.7
Frequency (RFM)	B.3.8	Ch. 6 §6.5	Treatment and control conversion rate	B.12.1	Ch. 11 §11.6
Gross sales	B.2.2	Appendix B	Treatment dose	B.12.5	Ch. 11 §11.7
Impressions	B.5.1	Ch. 1 §1.10	Treatment rate (treated share)	B.10.11	Ch. 9 §9.9
Incremental conversions and net contribution	B.12.7	Ch. 11 §11.7	Units per order (basket size)	B.3.3	Ch. 5 §5.7
Inertia (within-cluster sum of squares)	B.8.8	Ch. 6 §6.8	Units sold	B.2.9	Ch. 5 §5.7

Metric	Entry	Owned by	Metric	Entry	Owned by
Interval expressed in store-equivalents	B.13.9	Ch. 13 §13.9	Unsubscribe rate	B.6.6	Ch. 11, Table 11.3, where it is named as an instrumentation obligation
Lift by decile	B.10.13	Ch. 9 §9.10	Variance from target or baseline	B.2.13	Ch. 2 §2.7
Line cost	B.2.5	Ch. 13 §13.12.1	Weekly contribution	B.13.1	Ch. 13 §13.12.1
Line revenue	B.2.1	Ch. 4 §4.8	Weekly fixed footprint (the coverage threshold)	B.13.3	Ch. 13 §13.12.1
Margin over baseline	B.9.10	Ch. 8 §8.9	Weekly revenue at coverage	B.13.4	Ch. 13 §13.1
Mature-store annual contribution	B.13.8	Ch. 13 §13.9			

B.14.2 Metrics considered and not cataloged

The inclusion rule for this appendix is strict: a metric belongs here when it is used in the guide, in the labs and exercises, or in one of the two projects. A useful but unused metric is not added to make the catalog appear comprehensive, because every entry is a commitment to maintain a definition. Table B.13 records what was considered and excluded, and why. Several of these entries are ones a reader will expect to find, which is precisely why the register exists.

Table B.13

The exclusion register

Measure	Why it is not cataloged
Reach and media frequency	Neither appears in Chapters 1-13 or in either project. StyleCraft's campaigns table records impressions and clicks but no de-duplicated audience, so reach cannot be computed and frequency cannot be derived. Note the name collision: the book's frequency is the RFM count of orders per customer, not exposures per person.
Segment index versus the total population	Absent from Chapter 6 and from every other chapter. The role an index would play is occupied by segment share of customers, segment share of revenue, and standardized centroids read in standard-deviation units. Adding an index here would introduce a measure the chapters do not teach.
Lift versus base rate as a segmentation measure	The book's only lift is Chapter 9's lift by decile, which is cataloged in Section B.10. Chapter 6 contains no lift measure of any kind; the word appears there only inside the list of causal verbs the chapter forbids.
Customer lifetime value (CLV)	Named in Chapter 1's exercises and listed among the columns customer_features ships, but no chapter defines a formula, a horizon, or a discount rate for it. Cataloging one here would invent a definition the book

Measure	Why it is not cataloged
	does not teach. Where a project requires it, the student defines it and records the definition using the form in Section B.1.
App adoption or activation rate over a window	app_user is a stored customer attribute overwritten in place, not an event series, so a rate of adoption across a period cannot be computed from the certified data. The point-in-time stock is cataloged as app-user share in Section B.6.
Engagement rate as a single measure	Excluded deliberately. The book uses channel-specific rates — open, click, click-to-open — and a combined engagement rate would obscure the denominators those entries exist to make explicit.
Net promoter score, satisfaction, and brand-awareness measures	No survey instrument ships with StyleCraft and no chapter computes one. Chapter 3 discusses construct validity for such measures without cataloging them.
Brier score, expected calibration error, and log loss	No numeric calibration summary statistic appears anywhere in Chapter 9. The chapter's calibration evidence is the per-bin difference and its worst value near the operating threshold, which is cataloged in Section B.10.
Weighted MAPE, symmetric MAPE, and scaled error measures	Chapter 10 evaluates on MAE, MAPE, and signed error only. The alternatives exist to repair MAPE's near-zero failure, and the chapter's chosen repair is to disqualify MAPE for series that approach zero rather than to substitute a different statistic.
Cost per thousand impressions in the chapters	Included in Section B.5 as an Appendix B entry rather than a chapter-owned one. The book computes spend and impressions on the same campaigns but never names the ratio; it is cataloged because both projects may quote it and a student computing it needs the denominator convention.
Return rate, cancellation rate, and net-of-returns revenue	StyleCraft ships no returns table. Several entries in Section B.2 note where returns would change the number; none can be computed from the certified data as it stands.

Two of these exclusions carry a standing obligation. Customer lifetime value and the survey-based measures are both quantities a project may legitimately need; when a project defines one, the definition belongs in that project's own dictionary on the form in Table B.3, not in a local variation of an entry here. And because the guide's completion rule requires that this appendix be updated whenever a chapter uses a metric it does not already carry, this register is the place a future revision records its reasoning rather than leaving a silent gap.

References

The works below are the sources for the definitional conventions this appendix adopts, rather than for the individual formulas, which are the guide's own and are cited to their owning chapters in each entry.

Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning, 2*(1), 67–90. https://doi.org/10.1016/0149-7189(79)90048-X

Farris, P. W., Bendle, N. T., Pfeifer, P. E., & Reibstein, D. J. (2010). *Marketing metrics: The definitive guide to measuring marketing performance* (2nd ed.). Pearson Education.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Kimball, R., & Ross, M. (2013). *The data warehouse toolkit: The definitive guide to dimensional modeling* (3rd ed.). Wiley.

Kohavi, R., Tang, D., & Xu, Y. (2020). *Trustworthy online controlled experiments: A practical guide to A/B testing*. Cambridge University Press.

Parmenter, D. (2015). *Key performance indicators: Developing, implementing, and using winning KPIs* (3rd ed.). Wiley.

Schmittlein, D. C., Cooper, L. G., & Morrison, D. G. (1993). Truth in concentration in the land of (80/20) laws. *Marketing Science, 12*(2), 167–183. https://doi.org/10.1287/mksc.12.2.167

U.S. Securities and Exchange Commission. (2020). *Commission guidance on management's discussion and analysis of financial condition and results of operations* (Release Nos. 33-10751; 34-88094). https://www.sec.gov/rules/interp/2020/33-10751.pdf

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems, 12*(4), 5–33. https://doi.org/10.1080/07421222.1996.11518099