25 10 -10 0

Revenue Bands to Lead Scores

The four-number rubric for scoring company revenue in a B2B lead model — which band earns +25, which earns -10, why an unknown is a zero, and how to set the boundaries from your own closed-won data instead of someone else's benchmark.

ABOUT THIS GUIDE

Who wrote it, and how to use it

Ro Maldonado

Founder, gRO. Fifteen years building and running B2B and consumer acquisition programmes as an operator rather than an adviser — in the ad accounts, the CRM and the analytics, not presenting to the people who are.

Lead scoring is where more pipeline quietly dies than in any ad account, because a broken model is invisible: it keeps producing numbers.

What this guide will and will not do

It will give you a defensible rubric, the arithmetic behind it, a worksheet you can fill in, and a quarterly test that tells you whether your model is real. It will not give you a lead-scoring benchmark, because **no credible one exists.**

That is not a hedge. Collecting the published MQL-to-SQL “benchmarks” for this guide produced seven different answers spanning a four-fold range, none stating a methodology or a sample size. One widely repeated set of win-rate figures is attributed to a benchmark report that, when downloaded and read in full, does not contain win-rate data at all. Page 20 lists what we found and refused to print.

So the guide is built on the two things that survive checking: the acquisition economics of customer size, which are measured and cited here, and arithmetic, which is shown in full so you can check it yourself.

CONTENTS

IN BRIEF

The rubric, before the reasoning

+25

Inside your ICP revenue band. The companies the product and the price were built for.

+10

One band adjacent. Could buy, could grow in. Real fit, weaker economics.

-10

Clearly out of band. A no, not a maybe — the negative pulls them below the line.

0

Revenue unknown. Never negative. Absence of data is not evidence of bad fit.

- **Revenue is worth a quarter of a 100-point model.** Enough to matter, not enough to qualify a company on identity alone. At 40 or 50 points you route accounts to sales for what they are rather than what they did.
- **The bands come from your closed-won data, not from a benchmark.** There is no credible published benchmark for this, and the acquisition economics of customer size are not a straight line — so a rubric copied from anyone else is mis-scoring the middle of its own book.
- **An unknown scores zero, never a penalty.** Scoring missing revenue as negative buries every lead your enrichment provider missed, which is an enrichment problem wearing a scoring costume.

The rest of this guide is the reasoning, the evidence for the band shape, a worked 100-point model, the five ways these models break in production, a worksheet, and the quarterly test that tells you whether yours is forecasting or decorating.

Section One

The rubric

Four scores, one attribute. What each band means, and why the numbers are the size they are.

The four scores

In a 100-point lead scoring model, company revenue maps to four values and no more. A company inside your ICP revenue band scores **+25**. A company one band adjacent — small enough or large enough that it sits off your centre but could still buy — scores **+10**. A company clearly out of band, one you cannot serve profitably at any price, scores **−10**. A company whose revenue you do not know scores **0**. Not minus five. Not “pending”. Zero.

Four is not an arbitrary stopping point. It is the most granularity that closed-won data at normal B2B deal volumes can actually confirm. A model with nine revenue tiers implies a precision you cannot validate and adds maintenance without adding a single routing decision, because no sales team behaves differently toward tier six than toward tier five.

The two positive scores are a fit signal. The negative is a disqualifier, and it does work the absence of points cannot do: it actively pulls an account below the threshold rather than merely failing to lift it. That asymmetry is the point of the design, and page 15 shows the arithmetic.

Figure 1: Four scores, and what each one is actually asserting

Company revenue attribute in a 100-point B2B lead scoring model

REVENUE BAND	SCORE	WHAT THE SCORE ASSERTS
Inside your ICP band	+25	The product and the price were built for companies this size. Full firmographic credit.
One band adjacent	+10	Could buy, or could grow into the centre. Real fit, weaker economics.
Clearly out of band	−10	Cannot be served profitably at any price. Both ends of the range live here, for opposite reasons.
Unknown or unenriched	0	Nothing. The model declines to guess and the enrichment gap gets fixed instead.

Notes: The values are proportions, not constants. In a 50-point model, halve all four; the ratios are what carries.
Source: gRO scoring-model design, as built for B2B engagements.

Why revenue is worth 25, and not 50

Revenue band is the strongest pre-intent signal available. A lead can download everything you publish and never buy, because the company behind the address cannot afford the product or does not have the problem at their size. That is why firmographic fit deserves real weight.

But it is a ceiling, not a floor. Twenty-five points means an in-ICP company still has to *do* something before it crosses an MQL threshold: visit pricing, book a demo, come back a third time. Give revenue 40 or 50 points and the model routes companies to sales for what they are rather than what they did, and your sales team starts calling well-shaped strangers.

There is also an evidentiary reason for the ceiling, and it is the more honest one. The spread inside any single revenue band is wider than the gap between adjacent bands. In the acquisition data on page 10, companies in the $100K–$250K contract band show a 25th-to-75th percentile range of 20 to 37 months against a median of 24 — a 17-month spread within one band, against a median difference of zero between that band and the one below it.[1]

A signal with that much internal variance is worth tilting a decision. It is not worth making one. Twenty-five points out of a hundred is roughly the weight the evidence supports, and it is the number this guide defends.

Figure 2: Firmographics open the door; behaviour walks through it

Point allocation, 100-point model, MQL threshold at 60

COMPONENT	WEIGHT	EXAMPLE SIGNALS
Revenue band	**25**	The +25 / +10 / -10 / 0 rubric
Industry or vertical	**15**	Core vertical, adjacent, excluded
Role and seniority	**15**	Economic buyer, influencer, neither
Engagement behaviour	**25**	Return visits, content depth, email engagement
High-intent actions	**20**	Pricing page, demo request, audit signup

Note: Firmographics total 40 and person-level fit 15, so identity alone reaches 55 against a threshold of 60. The design intent is that no company can qualify without doing something.
Source: gRO scoring-model design.

Section Two

Setting the bands

The scores are the easy part. The boundaries are where models go wrong, and where the published advice is least reliable.

Bands come from your closed-won data

The numbers only work if the boundaries are honest, and the boundaries come from what you have actually sold, not from what you would like to sell.

The procedure is short. Pull your last twenty closed-won accounts and record the revenue of each. Add cycle length and first-year retention beside it. The band where deals close fast and retain well is your **+25**. One step outside it on either side is your **+10**. Everything beyond that is the **−10**.

Twenty is a floor, not a target. If you have two hundred, use them. If you have six, you do not have a scoring model yet; you have a hypothesis, and it should be labelled as one and given less weight until the book fills in.

Note what the procedure does not include: any number from this guide, or from any other. Your band boundaries are a property of your price, your delivery cost and your sales motion. Two companies selling adjacent products to the same market routinely have different bands, and both are right.

The rest of this section is about why copying someone else's boundaries is worse than it sounds — because the relationship between customer size and acquisition economics does not run in the direction most models assume.

THE PULL, IN ONE QUERY

For each of the last 20–200 closed-won accounts: **company revenue** (enriched, not self-reported), **days from first touch to signature**, **first-year retention or renewal**, and **gross margin on the account** if you can get it.

Sort by revenue. The +25 band is where cycle length and retention are both best — not where the deals are biggest. Those are frequently different bands, which is the subject of the next three pages.

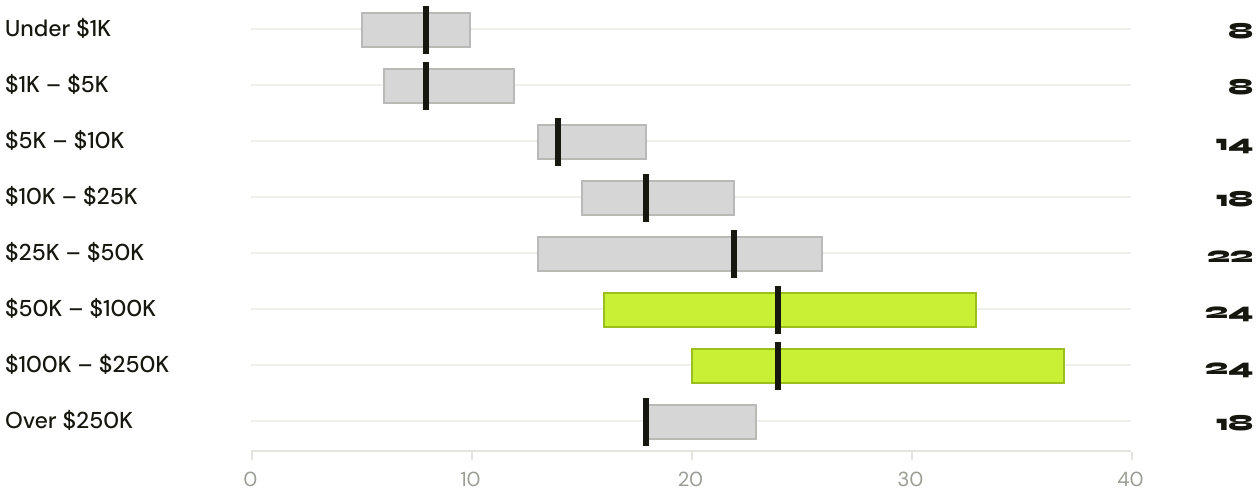
Acquisition cost by customer size

Most revenue-band rubrics encode one of two folk assumptions: that bigger customers are better and should score higher, or that enterprise is ruinously slow and expensive and should score lower. The only primary dataset available on the question says both are wrong.

Median CAC payback rises steadily with contract value — 8 months at the bottom, 24 at $50K–$250K — and then **falls back to 18 months above $250K.** The curve is an inverted U. The worst band is the middle, not the top.[1]

Figure 3: Acquisition payback is worst in the middle of the range, not at the top

CAC payback period in months, by average contract value. Bar = 25th to 75th percentile; rule = median



Notes: Values are months, read from the source chart. Axis is months. The two highlighted bands are the payback peak; note that the band above them recovers to the level of the $10K–$25K band. Benchmarkit's own commentary: larger ACV deals above $250K are "materially lower than solutions in the $50K – $100K range and even lower than in the $25K – $50K range", suggesting "an Enterprise solution that requires more time and resources to win may actually be more profitable over time"
Source: Benchmarkit, 2025 B2B SaaS Performance Metrics Benchmarks, p.24. N=148

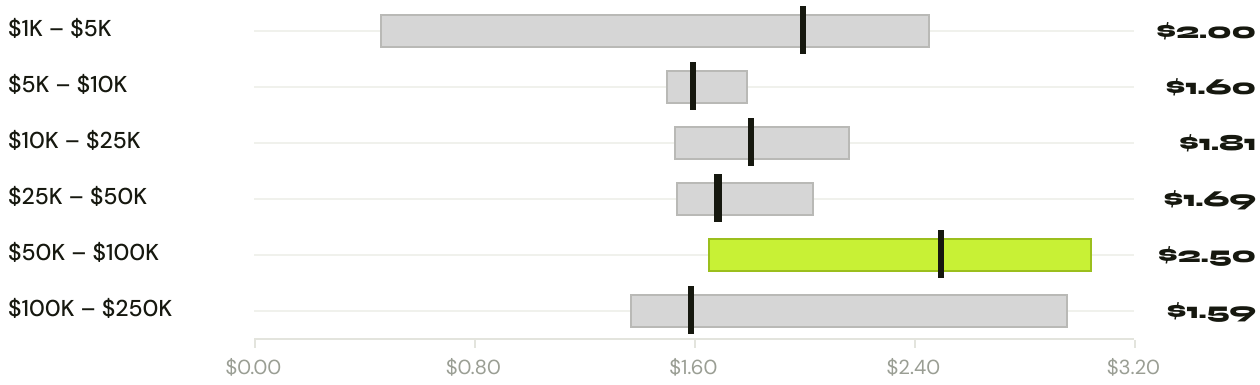
The middle is the expensive band

The payback curve is not an artefact of one metric. Sales-and-marketing cost per dollar of new recurring revenue traces the same shape from a separate sample: it peaks at **$2.50** in the $50K–$100K contract band and drops to **$1.59** in the band above it.[1]

Benchmarkit reports the same anomaly a third time, and notes it is not a one-year exception: solutions in the $10K–$50K range "are often more expensive to acquire" than those in the $50K–$100K range.[1] Three metrics, three samples, one shape.

Figure 4: Cost per dollar of new revenue peaks mid-market and falls above it

New customer CAC ratio, dollars of sales and marketing per $1 of new ARR, by average contract value



Notes: Lower is better. The highlighted band is the cost peak. The $1K–$5K band carries by far the widest spread, from $0.46 to $2.46, which is the signature of a segment where some companies have found a repeatable low-touch motion and others are selling small contracts by hand
Source: Benchmarkit, 2025 B2B SaaS Performance Metrics Benchmarks, p.19 and p.17. N=73 and N=43

What the curve does to your -10

Three consequences follow, and each one changes a rubric that was built on intuition.

First: a monotonic rubric is wrong by construction. If your scores rise in a straight line with company revenue, you are asserting a relationship the data does not show. The band that deserves your highest score is the one where your own payback and retention are best, and that is an empirical question with a non-obvious answer.

Second: "too big" is the least reliable -10 there is. The reflex to disqualify large accounts — long cycles, procurement committees, heavy delivery — describes real friction, and the friction is real. What the data disputes is the conclusion drawn from it. Above $250K those deals paid back faster than the two bands beneath them.[1] If you are scoring large companies negative, that has to be a finding from your book, not a feeling about enterprise sales.

Third: the bottom of the range is where a -10 is usually earned. A company that cannot fund the engagement is disqualified by arithmetic rather than by preference, and no amount of intent behaviour changes it. This is the one direction where the folk assumption and the evidence agree.

One caution on transferring any of this directly. The figures above are contract value, not customer revenue, and a $100K contract can come from a $5M company or a $5B one. They are evidence about the *shape* of the relationship between deal size and acquisition economics, not a lookup table for your bands. The shape is the transferable part: run the check on your own data and expect a curve, not a ramp.

Figure 5: Three assumptions the curve retires, and what replaces each

Consequences for a revenue-band rubric built on intuition

THE ASSUMPTION	WHY IT FAILS	WHAT TO DO INSTEAD
Scores rise with company size	Asserts a straight line where the data shows a curve	Score the band where your own payback and retention are best, wherever it sits.
"Too big" earns a -10	Above $250K, payback beat the two bands below it	Disqualify large accounts only on a finding from your own book.
"Too small" is a soft no	It is the one case where folk wisdom and the data agree	Keep the -10. Arithmetic, not preference, decides it.

Notes: Contract value is not customer revenue; these are evidence about the shape of the relationship, not a lookup table for your bands
Source: Derived from Figures 3 and 4.

Section Three

The model in context

A revenue score does nothing on its own. Here is the whole 100-point model it sits inside, and the arithmetic that makes the negative work.

A worked 100-point model

Firmographics carry 40 points: revenue band 25, industry or vertical 15. Person-level fit carries 15, for role and seniority. Behaviour carries 45: engagement 25, high-intent actions 20. MQL sits at 60.

Run the arithmetic and the design intent falls out of it. An in-ICP company with a decision-maker on the form sits at 40 before anyone has clicked anything — two-thirds of the way to the threshold, and unable to cross it without doing something.

Figure 6: Identity gets you to 55 of the 60 you need; the last five have to be earned

Points available by source, 100-point model, MQL threshold 60

SOURCE OF POINTS	MAX	WHAT IT CAN AND CANNOT DO ALONE
Revenue band	25	Cannot qualify a lead. Can disqualify one.
Industry or vertical	15	Cannot qualify a lead.
Role and seniority	15	Cannot qualify a lead.
Identity subtotal	**55**	**Five points short of the threshold. By design.**
Engagement behaviour	25	Return visits, depth, email engagement.
High-intent actions	20	Pricing, demo request, audit signup.
Total available	**100**	MQL at 60.

Note: The identity subtotal is set five points below the threshold deliberately. If it lands above the threshold, the model qualifies companies for existing and the sales team learns to ignore the score
Source: gRO scoring-model design.

Scale the whole table if your model is not out of 100. What matters is the relationship between the identity subtotal and the threshold, not the absolute numbers.

The arithmetic of the negative

A zero and a -10 feel similar and behave nothing alike. The zero is neutral: it leaves an account needing 60 points from everything else. The -10 is a handicap: it leaves them needing 70 from a board that only holds 75 outside the revenue line, and 45 of that 75 requires sustained behaviour.

That is the whole function of the negative score. It makes it arithmetically impossible for an out-of-band company to reach sales on curiosity alone, without hard-blocking them — so an out-of-band account that genuinely behaves like a buyer can still get through, which is what you want, because occasionally your band is wrong.

Figure 7: The negative does not block an account; it raises what the account must prove

Points still required to reach an MQL threshold of 60, by revenue score

REVENUE SCORE	STILL NEEDED	AVAILABLE ELSEWHERE	WHAT THAT MEANS IN PRACTICE
+25 in band	**35**	**75**	Decision-maker plus one real intent action clears it.
+10 adjacent	**50**	**75**	Needs genuine engagement, not one visit.
0 unknown	**60**	**75**	Can still qualify on behaviour. The enrichment gap costs them nothing they earned.
-10 out of band	**70**	**75**	Needs almost the entire behavioural board. Effectively reserved for accounts that are proving your band wrong.

Note: Available elsewhere is 100 minus the 25 allocated to revenue. Read the bottom row as the design goal: not a block, but a bar high enough that only a genuine anomaly clears it
Source: Arithmetic of the model on page 14.

This is also the answer to the most common objection to negative scoring, which is that it hides good leads. It does not hide them. It prices them. An out-of-band account that clears 70 behavioural points is the most interesting lead in your database, and it should trigger a review of the band rather than a routing exception.

Section Four

Failure and validation

How these models break in production, a worksheet for building yours, and the quarterly test that separates a forecast from a decoration.

Five ways the model breaks

These are the failure modes that show up in nearly every scoring model worth auditing. Four of the five are invisible from inside the model, because a broken scoring model does not error — it keeps producing numbers.

01 Punishing unknowns

Scoring missing revenue as a negative buries every lead the enrichment provider missed. That population is not small and it is not random — it skews toward newer companies and toward the ones without a public revenue estimate, which in many books is the growth segment. The fix is to score the unknown at zero and treat the gap as a data problem, which it is. Redman puts the cost of bad data at 15 to 25 percent of revenue for the average company, though he presents it as an estimate rather than a measurement.[2]

02 Too many bands

Nine revenue tiers imply a precision that closed-won data at normal deal volumes cannot confirm, and they add maintenance without adding a routing decision. If no one behaves differently toward tier six than tier five, the two tiers are one tier with extra bookkeeping.

03 A monotonic rubric

Scores that rise in a straight line with company size assert a relationship the acquisition data does not show. Both ends of the range can be a - 10, for opposite reasons, and the strongest band is frequently not the largest one. Page 10 is the evidence; your own book is the arbiter.

04 Set and forget

Bands drift as pricing and the product move. If accounts you scored - 10 keep closing and retaining, the bands are wrong, not the buyers. Re-validate quarterly, and treat a run of out-of-band wins as data rather than as luck.

05 Trusting the form

Self-reported revenue is answered optimistically often enough that a serious model treats an enriched figure and a form answer as different fields with different confidence, and prefers the enriched one. If you only have the form answer, that is closer to an unknown than to a fact.

The scoring worksheet

Print this page. Fill it in from your CRM, not from memory. If a row cannot be completed from data you already hold, that row is the first thing to fix.

STEP 1 — THE CLOSED-WON PULL

BAND (COMPANY REVENUE)	DEALS WON	MEDIAN CYCLE	YR-1 RETENTION	SCORE YOU ASSIGN

STEP 2 — ASSIGN THE FOUR SCORES

RUBRIC POSITION	SCORE	YOUR BAND BOUNDARY
Inside ICP band	+25	
One band adjacent	+10	
Clearly out of band	-10	
Unknown / unenriched	0	— not applicable, by design

STEP 3 — THE TWO CHECKS BEFORE YOU SHIP IT

Identity check. Add every point an account can score without doing anything. If that subtotal is at or above your MQL threshold, the model qualifies companies for existing. Lower the firmographic weights until it sits just below.

Unknown check. Count the share of your database with no enriched revenue. If it is above a few percent and your model scores unknowns negative, you are suppressing that entire population for a reason that has nothing to do with them.

Grading the model against revenue

A scoring model is a forecast, and forecasts get graded. Once a quarter, pull score-at-MQL against close rate in three buckets: leads that crossed at 60 to 75, leads that crossed above 75, and — the revealing one — everything sales worked that never reached 60 at all.

Two results falsify the model. If the 75-plus group does not close meaningfully better than the 60-to-75 group, the weights are decoration. If sub-60 leads close at the same rate as scored MQLs, the model is theatre, and the sales team already knows.

Figure 8: Three buckets, and what each result obliges you to do

Quarterly score-to-close validation

BUCKET	RESULT THAT CONFIRMS	RESULT THAT FALSIFIES, AND THE RESPONSE
Above 75	Closes clearly better than 60–75	If it does not, the weights are not separating anything. Re-derive them from closed-won rather than adjusting by feel.
60 to 75	Closes better than sub-60	If it does not, the threshold is in the wrong place. It is an output of this test, not a setting you pick once.
Below 60, worked anyway	Closes materially worse	If it closes the same, the model is not informing routing. Sales has already routed around it.

Note: Run it on a full quarter of closed outcomes, not on open pipeline, and hold the MQL definition constant across the window. Forrester notes that when teams do not trust a scoring model they insist on access to all inquiries, calling a broader pool with fewer attempts per lead — which is what the third row detects
Sources: gRO validation practice; Forrester, "The AQL: A Missed Opportunity With Lead Scoring?"

The third row is the one worth running first. It is the cheapest diagnostic in the list, it needs no change to the model to perform, and a failure there makes every other question moot.[3]

METHOD

What we refused to print

Every figure in this guide was checked against a primary source that was actually opened. These did not survive that check, and they are all in circulation.

Figure 9: Four claims in wide circulation that did not survive a source check

Lead-scoring figures checked against the primary they are credited to

THE CLAIM	CREDITED TO	WHAT WE FOUND
Win rates by deal size: 35–45% under $50K, 15–25% above $100K	A named 2025 SaaS benchmark report	Downloaded and searched all 72 pages. The report contains no win-rate, MQL, SQL or lead-qualification data of any kind. **The figures are not in the source they are credited to.**
MQL-to-SQL conversion benchmarks	Seven secondary sources	13%, 12–18%, 26–51%, 39%, 40%, 45%, 55% — a four-fold spread. None stated a sample size, a methodology, or a fixed definition of MQL. There is no benchmark here to quote.
Respond in 5 minutes: 100× to connect, 21× to qualify	Harvard Business Review	The HBR article is real and its authorship was verified.[4] These figures are not from it; they come from a separate study. The HBR text is paywalled and was not read.
Firmographic accuracy “averages 50%”; match rates 40–62% vs 80–98%	Data vendors	Every instance traced to a vendor’s own marketing, describing a category that vendor sells into. No independent test.

Notes: An MQL is a threshold each company sets for itself, which is why the second row can never resolve: the sources are not measuring the same thing. Measure your own fill rate and your own conversion instead — they are numbers you already hold, and the only ones that govern your model
Source: content-strategy/lead-magnet-scoring-2026/DATA-SOURCES.md, which records each check and its date.

HOW GRO BUILDS THESE

Scoring models, built to be graded

- **Bands from closed-won, not from vibes.** The rubric is calibrated against your actual won-deal revenue distribution, with cycle length and retention beside it, before a single point is assigned.
- **Wired to routing.** A score that does not change who sales calls first is decoration. The model ships connected to CRM routing and the SLA that follows it, or it does not ship.
- **Graded quarterly.** Score-to-close correlation is checked every quarter against the three buckets on page 19, and the weights move when the data says so rather than when someone senior has a hunch.

The Funnel Audit

Lead qualification is one section of a written diagnosis of the whole acquisition funnel: cost per lead by channel, conversion at every stage, offer clarity, audience fit, and the scoring and routing layer this guide covers. Prioritised fixes, ranked by pipeline impact.

If there is no clear scaling opportunity, we say so and the engagement does not proceed.

applygro.com/audit

Ro Maldonado · Growth Revenue Optimization

Thirty-minute intro call.
If we are not the right fit, we will say so.

NOTES AND SOURCES

Every figure in this guide is traceable to a source below that was opened and read. Sample sizes are stated on the exhibits. Where a source presents an estimate rather than a measurement, the guide says so at the point of use. The full working file, including what was rejected and why, is at **content-strategy/lead-magnet-scoring-2026/DATA-SOURCES.md**.

1. Benchmarkit, *2025 B2B SaaS Performance Metrics Benchmarks*, May 2025, produced with Pavilion. Sample size varies by metric and is stated on each exhibit; the CAC Payback by ACV chart is N=148, New Customer CAC Ratio by ACV is N=73, Blended CAC Ratio by ACV is N=43 and CLTV:CAC by ARR is N=101. benchmarkit.ai
2. Redman, Thomas C., "Seizing Opportunity in Data Quality," *MIT Sloan Management Review*, 27 November 2017. Redman presents the 15–25% figure as an estimate, not a measurement, and credits it to research by Experian plc and to consultants James Price of Experience Matters and Martin Spratt of Clear Strategic IT Partners. sloanreview.mit.edu
3. Forrester, "The AQL: A Missed Opportunity With Lead Scoring?" Forrester Research blog. forrester.com
4. Oldroyd, James B., Kristina McElheran and David Elkington, "The Short Life of Online Sales Leads," *Harvard Business Review*, March 2011. Cited here only for its existence and authorship, both verified directly. Its findings sit behind a paywall and were not read, so no figure from it appears in this guide — see page 20.

