Power

PHASE 1 OF 6

A 6-PART SERIES

How Modern B2B Companies Build a Total Addressable Market (TAM) That Actually Works

Client Data Extraction & Enrichment

How to extract and enrich customer data so your TAM reflects real buying behavior, not guesswork.

Power

Your TAM Is Built on Bad Data. Here's What to Do About It.

How to extract and enrich customer data so your TAM reflects real buying behavior, not guesswork.

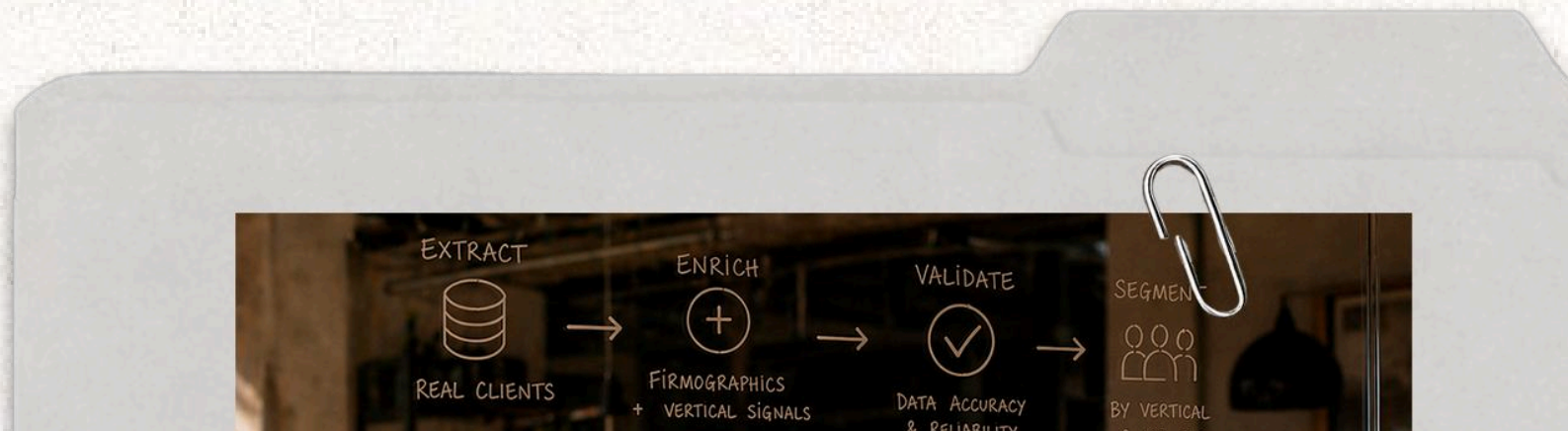
The "Who Said It Was Good?" Problem

Ask most marketing teams how they built their TAM and you'll get some version of the same story: someone pulled a list from ZoomInfo or Apollo, applied a few filters, revenue over $10M, 50+ employees, North America, and called it a day.

Six months later, pipeline quality is inconsistent and sales is complaining about lead fit. Someone asks: **"Why did this company make the list?"** And the answer is always the same: **"Someone said it was good."** Who? When? Based on what? Nobody remembers. The TAM has become an opinion dressed up as analysis.

This is how the majority of B2B TAMs are built. Not with bad intentions, with bad inputs. And once the inputs are wrong, everything downstream breaks: targeting, messaging, media spend, pipeline quality, close rates.

This is the first in a six-part series on building a TAM that actually works. We're starting with the data foundation, because this is where most teams get it wrong.



Power

“You’re making qualification decisions before you’ve defined what qualification means.”

Power

Start with Customers, Not the Market

The instinct to size the market first feels logical. But when you apply filters inside a third-party database before understanding what a qualified opportunity actually looks like, you're making qualification decisions before you've defined what qualification means. Different providers categorize industries differently. Revenue methodologies vary. Two teams using different databases with identical filters will generate completely different TAMs.

A better starting point is the dataset you already have: **your closed-won deals and current clients.** These companies bought. They stayed. They represent validated product-market fit. They're not theoretical, they're real.

"But we already have customer data in Salesforce." Not quite. CRM systems manage pipelines, not market intelligence. Over time, SDRs manually enter revenue from a Google search, reps update industry classifications mid-deal, automation rules overwrite fields with conflicting data, and managers edit tiers on gut calls. Multiply that across thousands of records and years, and **annual revenue, employee size, and industry classification, the three fields teams rely on most, are the three most likely to be wrong.**

The distortion only becomes visible once you introduce external validation. Until then, you're building ICP models on top of input bias.

Power

What Enrichment Actually Means

Enrichment is not “add more columns.” It’s validating internal records against external sources and layering signals your CRM never captured. It starts with a baseline: **annual revenue** (the most obvious scale indicator but most inconsistent across providers), **employee headcount** (tends to be the most reliable), and website traffic (a proxy for digital maturity).

But the real insight is this: **enrichment must be segment-specific.** The signals that indicate a strong-fit client in B2B SaaS look nothing like the signals that matter in professional services. There is no universal “good client” formula, even within B2B.

B2B SaaS / Technology	B2B Professional Services
Funding stage and total raised	Client roster size and concentration
Monthly site traffic and growth rate	Geographic coverage / office footprint
Product usage signals (G2, Capterra)	Industry specialization depth
Tech stack and integrations	Partnership and referral network
Hiring velocity in GTM roles	Headcount growth rate vs. revenue

A Series B SaaS company with aggressive GTM hiring tells you something fundamentally different about growth trajectory than a 500-person professional services firm with deep compliance specialization. Both might be great clients. The data that proves it is completely different. If your enrichment strategy applies identical signals across every vertical, you’ll miss strong-fit companies and include ones that fail on the signals that actually matter.

Power

The Process in Six Steps

The 6-step enrichment process

1

Pull Seed List
Closed-won deals & current clients from CRM

2

Clean the Data
Flag churned or poor-fit accounts separately

3

Segment by Vertical
SaaS, prof. services, fintech: one tab each

4

Baseline Firmographics
Revenue, headcount, web traffic: all segments

5

Layer Vertical Signals
Funding, G2, client concentration, footprint

6

Validate
Spot-check 10-15 records; >20% off = replace source

1. Pull your seed list from real deals.

Export closed-won opportunities or current clients from your CRM. You need company name, domain, and whatever segmentation fields you use internally. These companies have already proven they can buy and succeed.

Power

2. Clean before you enrich.

Remove or flag accounts that churned quickly, deals that were clearly poor fits, or engagements that failed for non-targeting reasons. Segment poor-fit accounts separately, they often reveal disqualification signals just as valuable as patterns in your best clients.

3. Segment by vertical.

SaaS in one tab. Professional services in another. Fintech in another. Each segment starts with three fields: company name, vertical, domain. Everything else gets layered through enrichment, and the enrichment strategy differs per segment.

4. Add baseline firmographics across all segments.

Revenue, headcount, and website traffic go on every segment. These are the floor of your enrichment, not the ceiling.

Tool spotlight: Clay

The tool most teams use for this is **Clay**. Clay connects to 150+ third-party data providers (clay.com/integrations), so you can access firmographic data from whatever providers you already subscribe to. Connect your existing subscription via API key and it runs through your credentials at no extra Clay credits. No subscription? You can still access the same providers through Clay and pay credits per row instead.

Clay also has its own pre-built enrichments through **Claygents**: Enrich Company, Work Email, Find Mobile Number & Professional URLs, and others. These run on Clay credits and don't require a third-party subscription, alongside formatting and cleaning utilities that handle data prep inside Clay before you enrich.

Power

How the enrichment layers work in Clay:

Enrichment Layer	What it surfaces and how
Third-party provider enrichments	All firmographic data: annual revenue, employee headcount, company domain, LinkedIn profile, website traffic, tech stack, funding data, job postings, and more. Access any of Clay's 150+ integrated providers directly. Have your own Apollo, ZoomInfo, or Clearbit subscription? Connect via API key and run through your existing credentials at no extra Clay credits. No subscription? Clay still gives you access and charges credits per row instead. Full list at clay.com/integrations.
Clay native enrichments (pre-built Claygents)	Clay's own set of pre-built Claygents: Enrich Company, Work Email, Find Mobile Number & Professional URLs, and others. Run on Clay credits with no third-party subscription required. These sit alongside formatting and cleaning utilities (normalize company names, scrape website, format dates) built into Clay for data prep.
Claygent: custom web research	Any question a standard enrichment cannot answer. Examples: Is this company product-led or sales-led? Do they have a dedicated demand gen function? Are they actively hiring in GTM roles? Claygent accesses the internet and returns structured answers per row. These cost Clay credits, so pressure-test on 10 rows before running at scale.

5. Layer in vertical-specific signals.

Funding stage and G2 reviews for SaaS. Client concentration and geographic footprint for professional services. Regulatory indicators for fintech. The goal is signals that historically correlate with successful engagements in that specific vertical.

Power

Standard enrichments only go so far. They won't tell you whether a company is product-led or purely sales-driven, whether they have a dedicated demand gen function, or whether their headcount is actually growing. A 200-person company actively hiring in GTM roles and one that hasn't posted a marketing job in a year look identical in a database. They are not the same prospect.

That's where custom Claygents come in. A Claygent is an AI agent inside Clay that accesses the internet to answer a specific question for each row in your dataset. You give it a direction, it creates its own step-by-step prompt, references the company name and domain already in your table, and returns structured answers at scale.

The question that unlocks the right Claygents is not "what data can we get?" It's **"what would a great rep notice in the first five minutes of researching this account?"** Those observations are what you're turning into enrichments.

Three rules for Claygents

Pressure test first. Run 10 rows before scaling. Claygents cost credits, and bad instructions are expensive at thousands of rows.

Use them for hard questions only. If a standard enrichment can answer it, use that. Claygents earn their cost on questions that require actual web research.

Start wide: 65 to 75 enrichments. Most won't survive the accuracy and coverage check. That's the point. You want the data to surface what matters, not assume it.

Power

What this looks like in practice varies completely by segment. The signals that predict fit in one market have almost no overlap with the signals that matter in another:

Segment	Creative signals worth enriching	Why it changes the picture
Fintech	Funding raised, funding stage, active user count or app downloads, business model (DTC vs. employer-sponsored vs. freemium), regulatory/licensing status, mobile app store rating and review volume, primary consumer demographic	A Series B fintech with 2M active users, a 4.7 app store rating, and a freemium-to-premium model tells you something completely different about growth stage and marketing maturity than one with the same revenue but no consumer product.
Hospitality	Property count, ownership model (owned vs. franchised vs. managed), average daily rate, booking channel mix (direct vs. OTA-heavy), guest capacity, brand positioning, Instagram follower count, review volume across Google and TripAdvisor	A boutique brand doing 60% direct bookings signals a very different marketing sophistication than one that is 80% OTA-dependent. One needs CRO and retention. The other needs a direct channel strategy first.
Franchise brands	Total franchise count, corporate vs. franchisee-operated location split, open territories, parent company or PE portfolio ownership, franchise investment range, SBA loan approval status, consumer review volume as a proxy for foot traffic	A 400-unit franchise brand backed by a PE roll-up with open territories is a fundamentally different conversation than a 40-unit owner-operator brand, even if their top-line revenue looks similar.
Healthcare & mental health	Care delivery model (teletherapy, in-person, medication management), licensing coverage, payer relationships (insurance accepted vs. DTC-only), patient volume or session count, therapist or provider network size, certifications and clinical affiliations	A mental health platform covered by Aetna and Cigna with 500K monthly sessions has completely different acquisition economics than a DTC-only coaching app.

The point is not that your signals will look like these. It is that your signals will be just as specific, and just as different from whatever generic filters you started with. Revenue and headcount are the floor. Everything above that is where the real differentiation lives.

Power

6. Validate. Don't trust blindly

Different providers routinely return different values for the same company. Spot-check 10 to 15 records against actual LinkedIn profiles and company websites. If more than 20% of values are materially wrong, that source needs to be replaced or weighted differently.

What This Changes

Your assumptions get corrected. One of the most common discoveries is that revenue thresholds aren't what teams assumed. Companies with lower annual revenue than previously thought can still be strong clients. Teams that set hard floors based on CRM data alone often find they'd been excluding viable accounts for years.

Subjectivity gets replaced with logic. Instead of one person deciding which companies belong in TAM, and that decision shifting every quarter based on who's in the room, you have a structured rubric. The same data, the same logic, the same result every time. If the rules need to change, you change them explicitly and reapply them across the entire dataset. No more "who said it was good."

Next: Turning Patterns Into Guardrails

You now have a fully enriched dataset: baseline firmographics validated against external sources, creative vertical-specific signals built through Claygents, and a clean separation between your best clients and your worst ones.

In Phase 2, we cover what to do with it. How to run the pattern analysis that surfaces which signals actually predict fit, how to define the guardrails that separate a qualified opportunity from one that wastes everyone's time, and how to use AI to find the thresholds your gut feel never would have landed on.The data is built. Now we make it do something.

Next: Phase 2, Pattern Analysis & Guardrail Definition