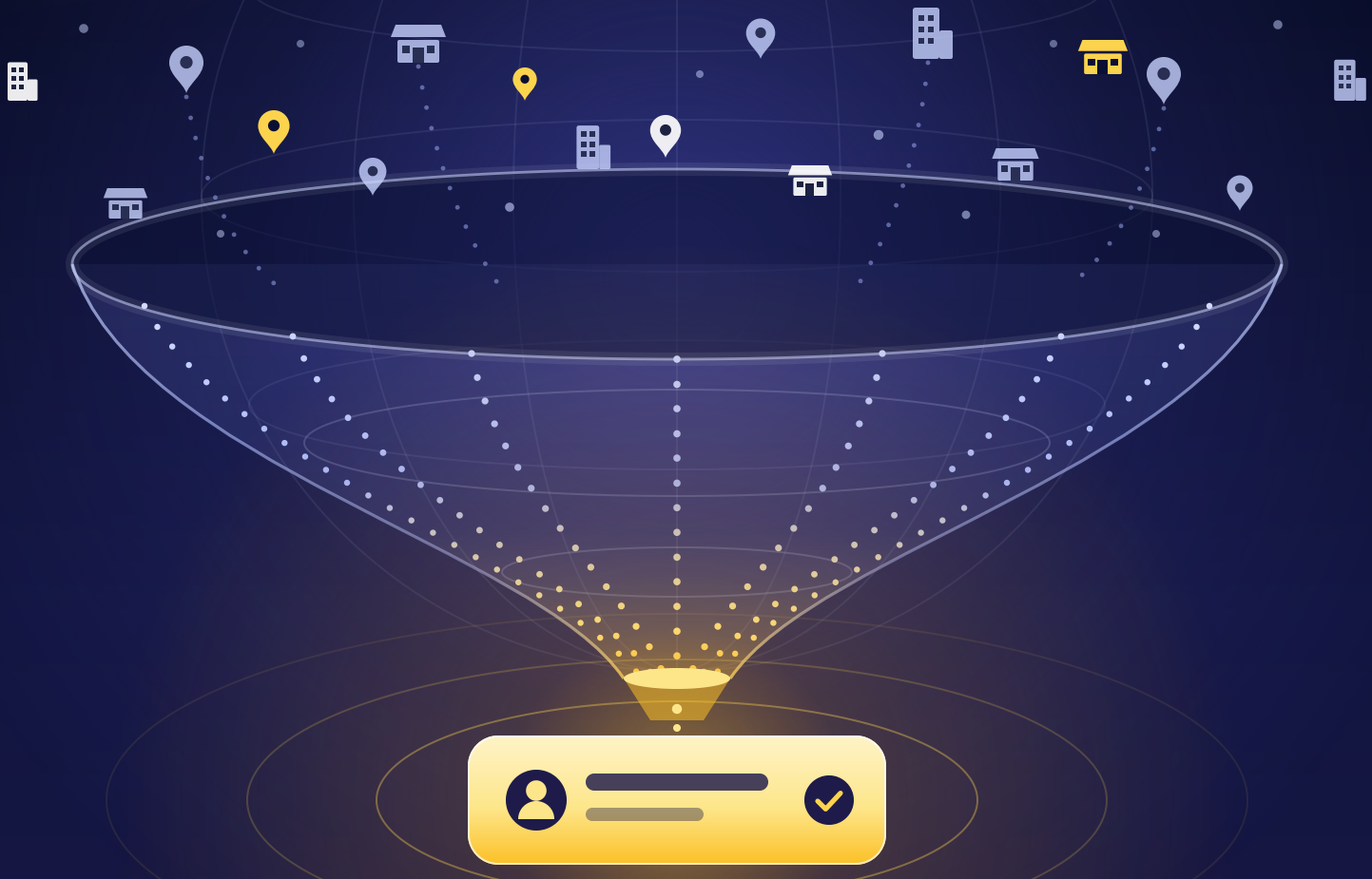
ProxyHorizon

2026 EDITION

A PROXYHORIZON PLAYBOOK

Lead Generation with Public Web Data

Find, Qualify and Reach the Right Prospects



- Prospects from maps, reviews, websites and public signals
- Verification, scoring and CRM rules that keep lists clean
- Outreach that earns replies and stays compliant

BEFORE YOU BEGIN

About This Book

The businesses you want to reach publish what you need to know every day: listings, websites, job posts, reviews and news. This book shows you how to turn that public data into a lead engine you can run every week.

Who It's For

Founders doing their own sales

Every prospecting hour has to count. You need a short list and a reason to write to each one.

START WITH chapters 1, 2, 6 and 9

Agencies running outbound for clients

You build lists for several clients and need a process you can repeat, price and defend.

START WITH chapters 3, 7, 8 and 10

SDRs, sales ops and growth marketers

You own the pipeline numbers and want fresher data, better timing and a CRM you trust.

START WITH chapters 6, 7 and 8

Agencies selling to local businesses

You sell websites, SEO or marketing to local firms and want the ones that visibly need you.

START WITH chapters 4, 5 and 9

How to Read It

Read chapters 1 and 2 first, then follow your card above. Every chapter ends with key takeaways, and the appendix holds a 30-day launch plan and printable checklists. Watch for these signposts:

Pro tip Saves time or money.

Note Context worth knowing.

Watch out A common mistake to avoid.

Risk Legal or ethical exposure.

Pro move What experienced teams do.

Free tool A free ProxyHorizon tool.

2026 Edition. Provider figures come from the ProxyHorizon directory as of November 2026. Prices move often, so check proxyhorizon.com/proxies for live numbers. Businesses, people, phone numbers and `.example` domains in the examples are fictional. Nothing here is legal advice.

© 2026 ProxyHorizon. Share this PDF freely, unchanged and with credit.

YOUR MAP

Contents

	The List Is the Strategy Why built lists beat bought ones, and how to use this book responsibly.	04
01	**The Public-Data Lead Engine** Why public data wins, the engine end to end and the numbers to watch.	05
02	**Define Your ICP and Buying Signals** Turn your best customers into a filter you can run against data.	09
03	**Where Leads Live: A Map of Public Sources** Twelve public sources, what each one gives you and how hard it is.	13
04	**Local Business Leads From Maps and Reviews** Cover a city, remove duplicates and score every business by opportunity.	18
05	**B2B Leads From Company Websites** Which pages to read, what to extract and how to crawl politely.	23
06	**Trigger Signals: Timing Beats Volume** A signal library with weights, shelf lives and a monitoring rhythm.	28
07	**Enrichment, Verification and Deliverability** Find the right person, verify the address and land in the inbox.	32
08	**Scoring, Routing and CRM Hygiene** Fit and intent scores, A, B and C tiers, and records you can trust.	38
09	**Outreach That Earns Replies** Relevant personalization, a multichannel cadence and the compliance rules.	43
10	**Infrastructure: Proxies, Automation and Cost per Lead** Proxies by source, nine providers, no-code workflows and real costs.	50
A	**Appendix** A 30-day launch plan, checklists, a glossary and free resources.	55

INTRODUCTION

The List Is the Strategy

Most outbound fails before anyone writes a word. The list was too broad, too stale or built on guesses, so even a sharp message lands on the wrong desk at the wrong moment. And the usual fix, sending more, only makes it worse.

Bought lists deepen the problem. They're usually compiled once and sold to many buyers, they start going stale the day they're built, and they tell you who exists rather than who needs you now. Meanwhile the businesses you want are publishing exactly what you need to know: a Maps listing with a phone number and a run of unanswered reviews, a careers page hiring a first marketer, a website that falls apart on a phone.

> **Stop buying lists of everyone. Build a list of the right businesses, at the right moment, for a reason you can name.**

This book shows you how to build that list, and the engine that keeps it fresh, mostly with free and low-cost tools and very little code. You'll see where prospects leave public traces, how to tell a good fit from a time-waster, how to verify contacts without burning your domain, and how to write messages people actually answer.

By the End, You'll Be Able To

- Turn your best customers into an ICP you can filter real data against.
- Pull prospects from maps, reviews, company websites and other public sources.
- Spot the trigger signals that say a business is ready to talk.
- Verify contacts and set up sending so your emails reach the inbox.
- Score, route and store leads in a CRM your team trusts.
- Write outreach that earns replies, and know the rules in each market.
- Work out what each qualified lead really costs you.

Use This Responsibly

This book is about public business information and B2B outreach. Don't scrape behind logins, respect each site's terms and robots.txt, collect the least personal data that does the job, honor every opt-out quickly, and follow GDPR, CAN-SPAM, CASL and UK PECR wherever your recipients are. Chapter 9 explains the rules in plain English. It's general information, not legal advice.

01



The Public-Data Lead Engine

A bought list tells you who exists. Public web data tells you who fits, what's changing and when to reach out. This chapter maps the engine you'll build in the rest of the book, and the four numbers that tell you it's working.

IN THIS CHAPTER

1.1 Why public data beats bought lists

1.2 The lead engine, end to end

1.3 The numbers that matter

1.1 Why Public Data Beats Bought Lists

Contact databases are convenient: pick an industry and a headcount band, export a few thousand rows, start sending. The trouble is that anyone can export the same rows, the data was gathered months ago, and nothing in it says why a business would want to hear from you this week.

Public web data flips that. You collect it when you need it, from places businesses keep current because their own customers rely on them, and it arrives with context: what they sell, what they run, what they're struggling with and what just changed.

	Bought list	Built from public data
Freshness	Compiled months ago, decaying since	Collected when you need it, refreshed on a schedule
Relevance	Industry codes and headcount bands	What a business actually sells, runs and needs
Timing	A static snapshot	Live signals: hiring, reviews, launches, funding
Exclusivity	Sold to many buyers, so prospects are over-contacted	Your own angle on your own list
Provenance	Often unclear where a contact came from	You can name the source of every record
Cost	Per record or per seat, paid upfront	Time and tools, falling as you automate

Three advantages that compound

Freshness

A business that moved or closed last month is already wrong in a static list. Fresh pulls mean fewer bounces and wasted calls.

Relevance

You filter on what a business does, not how a database coded it: "dentists with no online booking", not "healthcare, 10 to 50 staff".

Signals

Job posts, reviews and news tell you when to write and what to say. Good timing turns a cold email into a relevant one.

When bought data still earns its place

Purchased data isn't useless. It's handy for sizing a market before you commit, or as a skeleton of company names you then enrich, verify and qualify with public data. What doesn't work is sending to it raw.

Public doesn't mean anything goes

Business facts like a company's category, address and website are fair game for research. Personal details carry obligations even when they're published. The rule in this book: collect business facts freely, personal data sparingly, and always record where each one came from.

1.2 The Lead Engine, End to End

Every reliable outbound program runs the same loop, whether it's a founder with a spreadsheet or an agency with ten clients. There are nine stages, each with one job and one output that feeds the next. The rest of this book takes them one at a time.

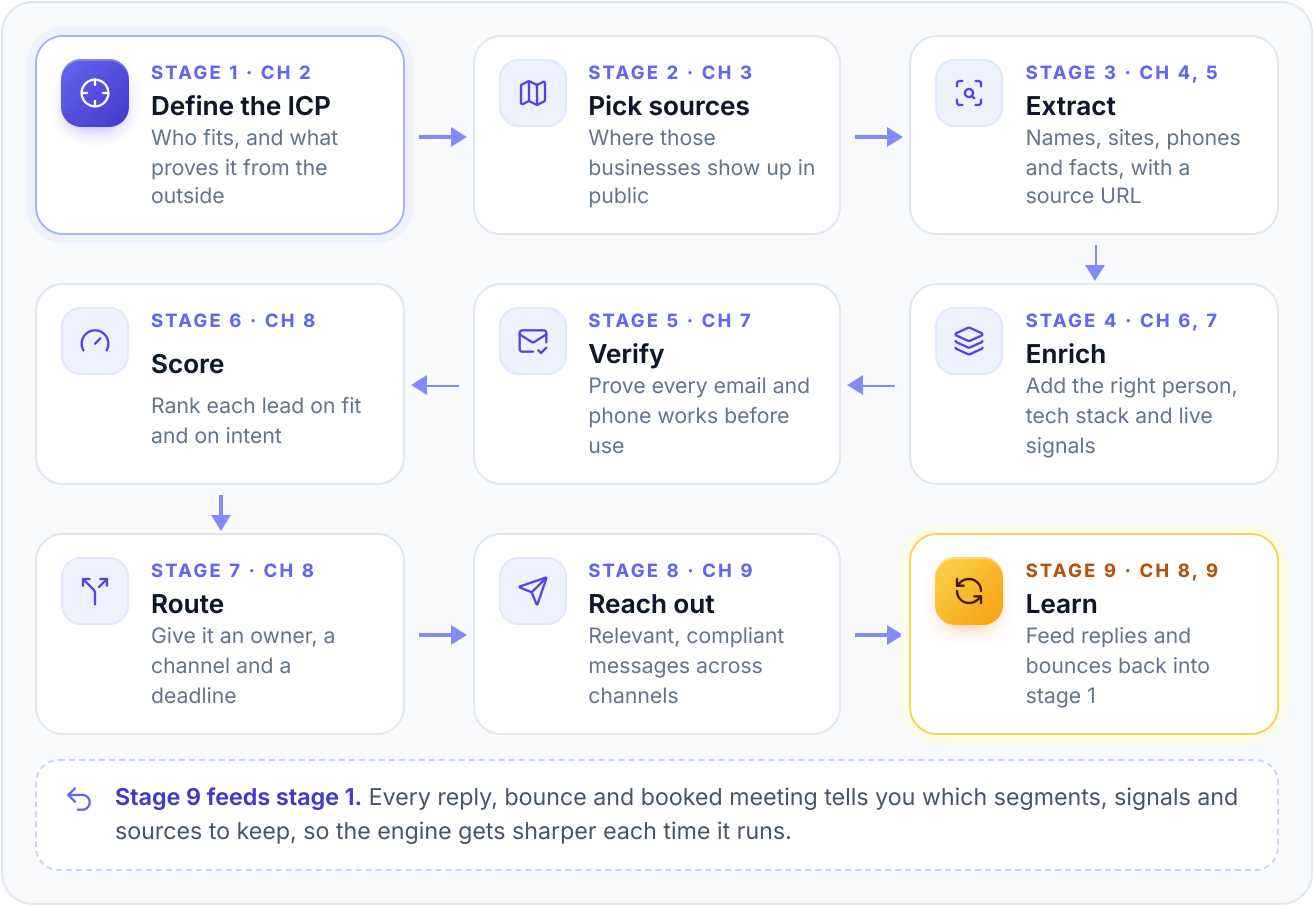


Figure 1.2 The public-data lead engine. Stages 1 to 5 build the list and stages 6 to 9 work it. Each card names the chapter that covers it.

What each stage hands to the next

Write the output of every stage down before you build anything. The ICP hands over a filter you can apply to a spreadsheet row. Sources hand over a short, ranked list of places to pull from. Extraction hands over raw rows, each with a source URL and a date. Enrichment adds the person, the tech and the signals; verification marks each contact valid, risky or invalid; scoring adds two numbers; routing adds a name and a deadline.

Run the loop by hand once

Before you automate anything, push 50 prospects through all nine stages in a spreadsheet. You'll find the weak stage in an afternoon, and you'll know exactly which step is worth automating first.

1.3 The Numbers That Matter

Rows collected and emails sent feel like progress, but they're inputs. These four outcome numbers tell you whether the engine works, and each one points at a different stage when it doesn't.

Cost per qualified lead

```
monthly cost ÷ leads that pass ICP and verification
```

The number to budget on and to compare sources with.

Bounce rate

```
hard bounces ÷ emails sent
```

Rising bounces mean verification (stage 5) is failing.

Reply rate

```
replies ÷ prospects contacted
```

Track positive replies separately from "not interested".

Meetings booked

```
meetings ÷ prospects contacted
```

The number your pipeline actually lives on.

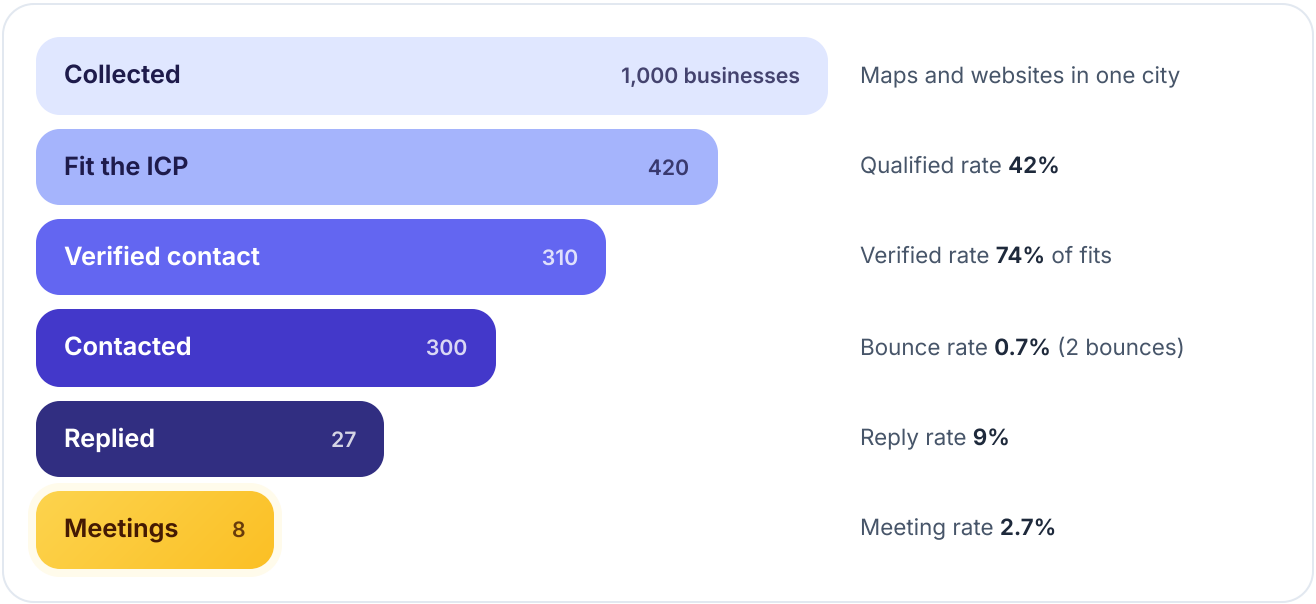


Figure 1.3 One month of a worked example, not a benchmark. Bars aren't to scale, and your ratios will depend on your market, offer and list quality.

KEY TAKEAWAYS

- Public data is fresher, more relevant and full of timing signals. Bought lists are none of those.
- The engine runs in nine stages from ICP to learning, and each hands a clear output to the next.
- Judge it on cost per qualified lead, bounce rate, reply rate and meetings booked.

Go deeper: Proxies for Lead Generation · Proxies for Market Research

02



Define Your ICP and Buying Signals

An ideal customer profile isn't a persona slide. It's a filter you can run against data: fields you can collect, values you can check and signals that say "now". This chapter turns your best customers into that filter.

IN THIS CHAPTER

2.1 Firmographics: who fits

2.2 Technographics and buying signals

2.3 Red flags and your ICP worksheet

2.1 Firmographics: Who Fits

Start with your best five to ten customers, not your imagination. What do they have in common that you could check from the outside, without asking them? That's your ICP. Anything you can't observe in public data is a hope, not a filter.

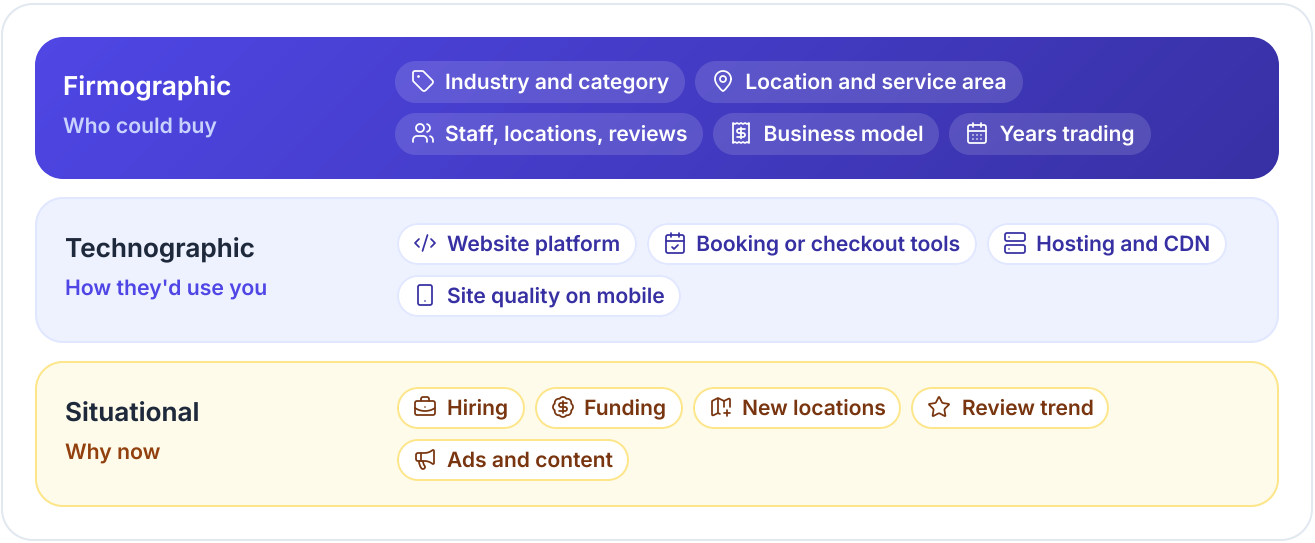


Figure 2.1 Three layers of fit. Firmographics say who could buy, technographics say how they'd use you, and situational signals say when to get in touch.

Firmographic fields you can actually collect

Field	Where it shows up in public	A filter that works
Industry	Maps category, directory section, the site's own copy	"Orthodontists", not "healthcare"
Location	Address, service-area pages, Maps listing	Within 40 km of Leeds
Size	Team page, employee counts, number of locations, review volume	Two or more locations
Business model	Pricing page, checkout, booking widget	Takes bookings online
Maturity	Business registry, review history, copyright year	Trading for three years or more

Use review volume as a size signal

Local businesses rarely publish headcount or revenue, but review counts track customer volume. A clinic with 400 reviews serves far more patients than one with 12, and usually has more budget to match.

2.2 Technographics and Buying Signals

Technographics tell you how a business would use you, and sometimes they're the whole ICP. A Shopify app can only sell to Shopify stores. An agency that rebuilds WordPress sites wants WordPress sites, ideally old ones. The free Website Technology Checker shows a site's CMS, framework, CDN and hosting, plus the bot protection in front of it.

Situational signals are the other half. They don't change whether a business fits, they change whether it's ready. A dental group that just opened a second clinic has a reason to care about local search this month that it didn't have last month.

Match signals to what you sell

If you sell	Strongest fit signal	Strongest timing signal
Web design	A dated site that breaks on phones	A new location, rebrand or service line
Local SEO	Ranks outside the top 10 in Maps for its main service	A competitor opens nearby
Reputation management	Rating under 4.0, or few owner replies	A run of one- and two-star reviews
Marketing services	Runs ads to a weak landing page	Hiring its first marketer
Recruiting or HR software	Twenty or more staff	Several roles open at once
B2B sales tools	A sales team of five or more	A new sales leader, or fresh funding

Fit decides who, signals decide when

Fit without a signal

A perfect-fit business with nothing changing is a slow burn. Keep it on the list, contact it lightly, and watch for a signal before you push.

A signal without fit

A hot signal at the wrong kind of business is a distraction. Hiring ten engineers means nothing to a florist's web designer. Fit filters first, every time.

Chapter 6 turns signals into a weighted library with shelf lives and a monitoring rhythm. For now, pick the two or three that matter most for your offer and write them down.

Free tools: Website Technology Checker · LinkedIn Company Employees for team size and recent hires

2.3 Red Flags and Your ICP Worksheet

Negative signals save as much time as positive ones. Write your disqualifiers down and filter on them before anyone spends a minute on research. These six cover most offers:

- Closed, or "temporarily closed", on Maps
- A "no solicitations" notice on the website
- Outside your service area, language or hours
- A franchise location where head office buys
- Already a customer or an open deal in your CRM
- Just announced a deal with a direct competitor

Your ICP worksheet

Fill in the right-hand column from your best customers. The middle column shows a worked example for an SEO agency that sells to dental clinics.

Criterion	Example: local SEO for dentists	Your ICP
Industry	General and cosmetic dentists, orthodontists	
Location	Metro areas in Texas, within one hour's drive	
Size	1 to 5 locations, 40+ Google reviews	
Business model	Private practice that takes new patients	
Tech	Owns a website; any platform	
Fit signals	Maps rank 8 to 40 for "dentist in [city]"	
Timing signals	New location, review slowdown, new competitor	
Red flags	Group-owned chain, rating under 3.5	
Who to contact	Practice owner or practice manager	
Sources	Google Maps, clinic websites, reviews	

KEY TAKEAWAYS

- Build your ICP from your best customers, using only traits you can observe in public data.
- Firmographics say who fits, technographics say how they'd use you, signals say when.
- Write your disqualifiers down and filter on them before any research starts.

03



Where Leads Live: A Map of Public Sources

Your prospects leave public traces in a dozen places: maps, review sites, directories, their own websites, job boards, video channels and business registries. This chapter maps each source to what you get from it, how hard it is to collect and whether you'll need proxies.

IN THIS CHAPTER

3.1 The source map

3.2 Maps, reviews and directories

3.3 Companies, teams and creators

3.4 Marketplaces, events and registries

3.1 The Source Map

Each source answers a different question: maps say who exists in a place, websites say what a business does, job boards and reviews say what's changing. The difficulty dots and proxy column are ProxyHorizon's editorial ratings for collecting at volume. One-off lookups rarely need proxies.

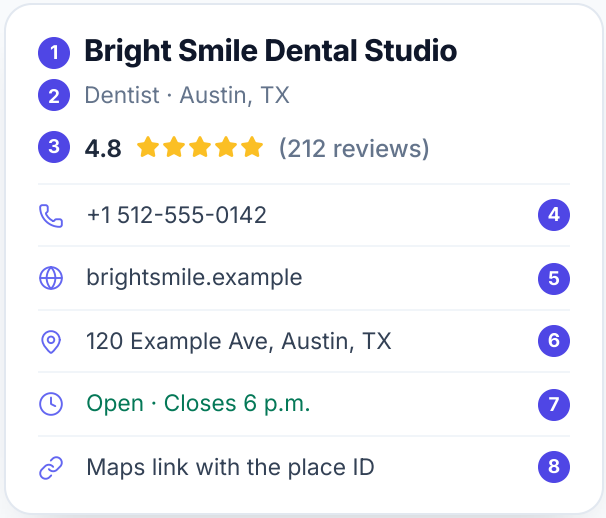
Source	What you get	Best for	Difficulty	Proxies?
Google Maps and local listings	Name, category, rating, reviews, phone, website, address	Local service businesses	●●●○○	Residential, city
Review sites	Ratings, review text, recency, owner replies	Quality and pain points	●●●○○	Residential
Industry directories	Members, categories, websites, partner tiers	Niche B2B verticals	●●○○○	Rarely
Company websites	Contacts, services, team, pricing, tech, schema	Almost every B2B ICP	●●○○○	Datacenter at volume
LinkedIn company pages	Industry, size band, HQ, staff by title	The right role, team size	●●●●●	Use a tool
Job boards and careers pages	Open roles, titles, locations, tools named in posts	Hiring signals, tech stack	●●●○○	Residential
YouTube channels	Audience, upload rhythm, published business emails	Creators and media brands	●●○○○	Use a tool
Social profiles	Followers, posting activity, bio links and emails	Consumer brands, local shops	●●●●○	Residential or mobile
Marketplaces	Sellers, products, sales rank, reviews	E-commerce services	●●●●○	Residential
App stores	Developer, website, support email, ratings, updates	Mobile app services	●●○○○	Rarely
Event exhibitor lists	Exhibitors, sponsors, categories, booth numbers	Timely pre-show outreach	●○○○○	No
Business registries	Legal name, status, incorporation date, address	New businesses, checks	●○○○○	Official API

Start with two sources, not twelve

Pick the source where your ICP is most complete, usually Maps for local businesses and company websites for B2B, plus one that adds timing, such as job posts or reviews. Add a third only when the first two are running every week.

3.2 Maps, Reviews and Directories

Google Maps is the most complete directory of local businesses there is, and owners and customers keep it fresh for you. One listing carries most of what you need to qualify a local prospect, and the rest is one click away on its website.



1. **Name.** For the greeting, and with the address, a dedup key.
2. **Category.** Your first ICP filter.
3. **Rating and review count.** Quality, and a stand-in for size.
4. **Phone.** A call channel and a strong dedup key.
5. **Website.** The door to Chapter 5. No website is a lead in itself.
6. **Address.** Service area, territory and routing.
7. **Status and hours.** Filter out anything marked closed.
8. **Place ID.** The most reliable way to spot duplicates.

Figure 3.2 One Maps listing, eight fields worth keeping. The business, phone number and domain are fictional.

Reviews tell you what's going on inside

Google reviews cover local businesses, Trustpilot covers many online shops, and G2 and Capterra list software vendors with their reviews. Beyond the star rating, look at how recent the reviews are, what customers complain about and whether the owner replies. Chapter 4 turns those into an opportunity score.

Directories that pre-qualify for you

Trade associations, chambers of commerce and certification bodies publish member lists. The best hidden source is the partner locator: manufacturers and software vendors list their certified installers, resellers and agencies. If you sell to installers, the brand has already qualified the list for you. Use the public pages only, and skip anything behind a members' login.

Read the terms before you scale

Google's Maps terms prohibit mass downloading its content. For research-sized pulls, a purpose-built tool run at modest volume is the practical route. For production-scale lists, Google's official Places API is the licensed alternative, with its own limits on what you may store.

Free tools: Google Maps Scraper · Google Reviews Analyzer · **Read:** Best Proxies for Google Maps Scraping

3.3 Companies, Teams and Creators

Once you know which companies fit, you need three more things: how to reach the business, who to talk to and what they're up to. These free ProxyHorizon tools each answer one of those questions from public pages, so you don't have to build a scraper for every source.

Contact Information Finder
Emails, phones and socials a site publishes, contact page included.

Social Media Links Finder
Every social profile a site links to, across 35+ networks.

LinkedIn Company Employees
Current staff by title, team and country, plus recent hires.

YouTube Channel Analyzer
Subscribers, median views, engagement and upload rhythm.

YouTube Email Finder
Business emails a creator publishes, and where each was found.

Instagram Profile Analyzer
Followers, engagement rate and how often a profile posts.

LinkedIn: research it, don't scrape it

A LinkedIn company page shows industry, size band and headquarters, and its employee list shows who holds which role. That's exactly what you need to pick the right contact and spot a hiring spree. The LinkedIn Company Employees tool reads public profiles through a data provider, never through your account. Don't point scrapers at your own logged-in profile: LinkedIn's terms forbid it, and restricted accounts are the usual result.

Creators are businesses too

If you sell editing, sponsorships, merchandise or software to creators, YouTube is a public directory of businesses with audiences. Creators who want deals publish a business email in their channel description or on their website. The YouTube Email Finder reads those places and shows where each address came from. It never guesses, and it leaves alone the email YouTube keeps behind its sign-in and CAPTCHA.

Employee lists are personal data

Names, titles and photos identify people, so GDPR and similar laws apply even on public profiles. Keep the name and title of the one or two people you'll actually contact. Don't copy whole org charts into your CRM.

3.4 Marketplaces, Events and Registries

Four more sources are niche, but for the right ICP they're gold, and two of them are among the easiest data you'll ever collect.

Marketplaces

Amazon storefronts show who sells what. The Amazon Sales Estimator reads sales rank history and Buy Box share, so you can spot sellers growing fast.

App stores

Listings name the developer, link a website and show the last update. A good app with no update in a year is a mobile agency's prospect.

Event exhibitor lists

Trade shows publish exhibitors months ahead. They're paying to meet buyers, and "see you at booth 412" is a natural opener.

Business registries

Registries such as the UK's Companies House offer official APIs and bulk data. New companies need banks, accountants and websites.

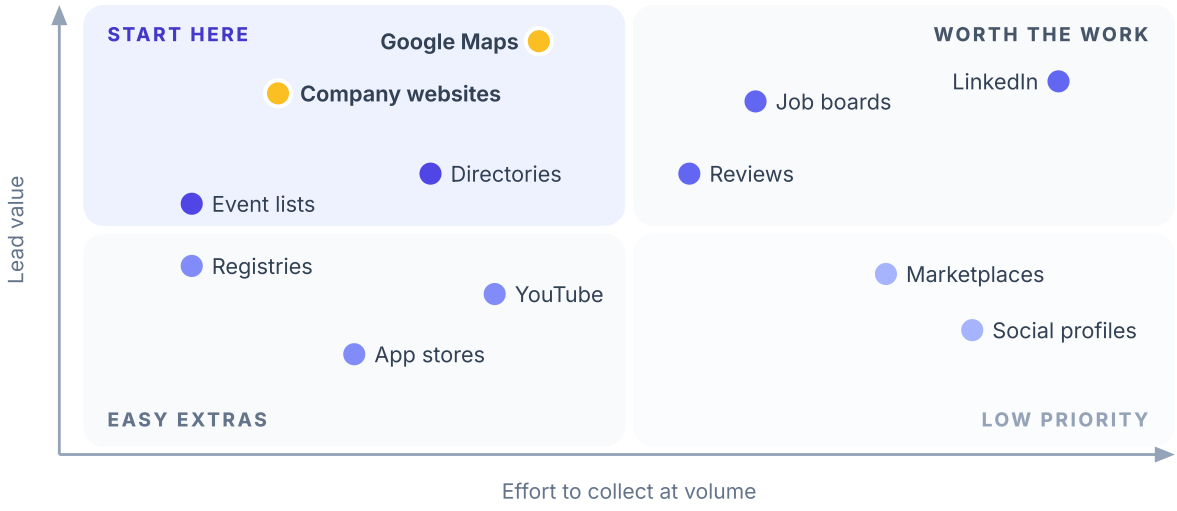


Figure 3.4 Where to start: ProxyHorizon's editorial view for a typical B2B or local-services seller, with the usual first two sources in amber. Your ICP moves the dots: for an Amazon agency, marketplaces jump to the top.

KEY TAKEAWAYS

- Each source answers a different question: who exists, what they do, or what's changing.
- Start with two sources: one where your ICP is complete and one that adds timing.
- Prefer official APIs, bulk data and purpose-built tools, and never scrape behind a login.

04



Local Business Leads From Maps and Reviews

If you sell to local businesses, Maps tells you who exists and reviews and websites tell you who needs help. This chapter shows how to cover a whole city, collapse duplicates into one clean record each, and score every business by how much you could do for it.

IN THIS CHAPTER

4.1 Grid searches: city × category

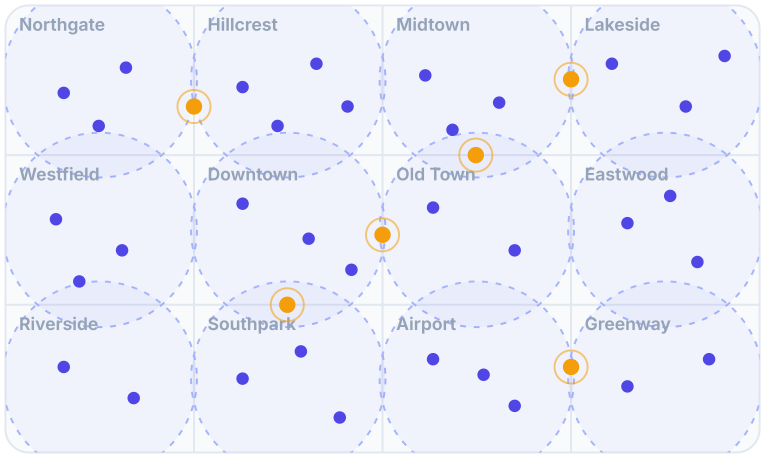
4.2 One business, one record

4.3 Scoring opportunity

4.4 Local rank and a worked example

4.1 Grid Searches: City × Category

One Maps search never shows you every business in a city. The ProxyHorizon Google Maps Scraper returns up to 40 places per run, and Maps itself favors the places closest to the area you name. To cover a market, split the search into small pieces: every area times every way people describe the business.



CATEGORY TERMS

- dentist
- dental clinic
- cosmetic dentist
- orthodontist
- emergency dentist

12 areas × 5 terms = 60 searches

Up to 40 places each, before duplicates

Figure 4.1 A grid search. Each area gets its own search for each term. Search areas overlap at the edges, so businesses near a border (amber) turn up twice, which is why section 4.2 exists.

The recipe

1. **List the terms people actually search,** plus the category names Maps uses: "dentist", "dental clinic", "orthodontist".
2. **Split the market** into neighborhoods, postcodes or nearby towns, small enough that each search returns fewer than 40 places.
3. **Run one search per term and area,** such as "orthodontist in Hillcrest, Austin", and tag every row with the query that found it.
4. **Stop when you saturate.** Once a new search returns almost nothing but duplicates, that slice of the market is covered.

Running searches yourself? Match the exit to the city

Maps results depend on where the searcher appears to be, and repeated automated searches from one IP get challenged fast. Use rotating residential proxies targeted to the city you're searching, keep one sticky session per scrolling session, and keep volumes modest.

4.2 One Business, One Record

Grid searches overlap by design, so the same clinic shows up under "dentist" and "dental clinic", and in two neighboring areas. Merge duplicates before you do anything else, checking these keys in order and stopping at the first match.

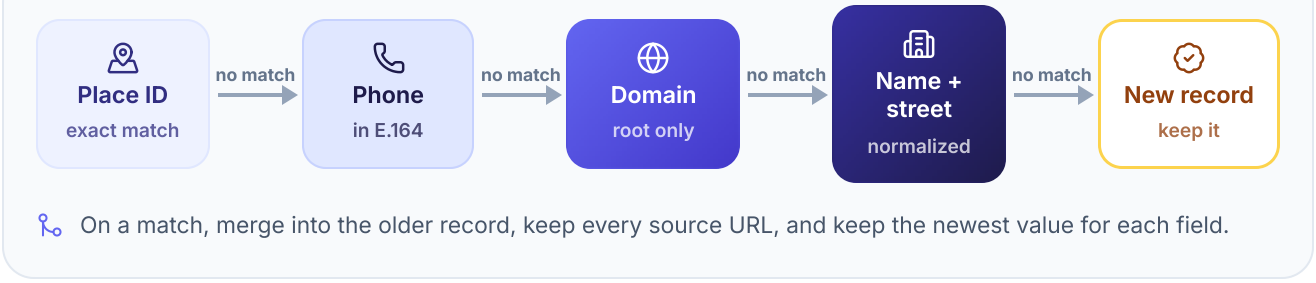


Figure 4.2 The dedup waterfall. The place ID catches most duplicates from Maps; phone and domain catch the same business found through other sources.

Normalize before you compare

Field	Raw values you'll see	Normalized key
Phone	`(512) 555-0142` and `+1 512.555.0142`	`+15125550142`
Domain	`https://www.BrightSmile.example/contact?ref=maps`	`brightsmile.example`
Name	Bright Smile Dental Studio, PLLC	`bright smile dental studio`
Street	120 Example Avenue, Suite 4	`120 example ave`

Shared websites aren't identities

Plenty of small businesses list a Facebook page, an Instagram profile or a link-in-bio page as their website. Thousands of businesses share those domains, so never dedupe on them. Fall back to the place ID or phone number instead.

Chains: one record per location, grouped by brand

Five clinics sharing one domain aren't duplicates. Keep a record per location, because each has its own reviews and rankings, and link them to one parent account so nobody emails the same owner five times.

Spot-check 20 merges after every run

Open 20 random merges and 20 new records side by side. Two different businesses merged into one means a rule is too loose; obvious twins left apart mean you need the next key in the waterfall.

4.3 Scoring Opportunity

For an agency that sells to local businesses, the best prospect isn't the biggest one. It's a good business with a gap you can visibly fix: a site that breaks on phones, reviews nobody answers, a ranking stuck on page two. These checks turn those gaps into points out of 100.

Signal	Points	How to check it
WEBSITE, UP TO 35		
No website at all	**+35**	Maps listing has no site, or only a social page
Fails mobile checks	**+15**	Mobile Website Screenshot: tap targets, text size, overlays
On-page SEO score under 60	**+10**	Website SEO Analyzer
Missing title or description	**+5**	Meta Tag Checker
No LocalBusiness schema	**+5**	Schema Markup Checker
REVIEWS, UP TO 35		
Fewer than 50 reviews	**+15**	Google Reviews Analyzer
Owner replies to under 20%	**+10**	Google Reviews Analyzer, reply rate
No new review in 60 days	**+10**	Date of the newest review
VISIBILITY, UP TO 30		
Maps rank 11 to 40	**+30**	Google Maps Rank Checker
Maps rank 4 to 10	**+15**	Outside the three-pack, within reach of it
Not in the top 40	**+10**	Often too far from the searched area

Disqualify before scoring: marked closed, rated under 3.5, or a group-owned chain. Weights are a starting point for a web and local SEO offer; flip them for yours.

Google Reviews Analyzer
Rating breakdown, topics and owner reply rate.

Mobile Website Screenshot
The site on a phone, with usability checks.

Website SEO Analyzer
An on-page audit with a score out of 100.

Meta Tag Checker
Titles, descriptions and social tags.

Schema Markup Checker
Finds and validates JSON-LD schema.

Google Maps Rank Checker
Where a business ranks in the top 40.

4.4 Local Rank and a Worked Example

The Google Maps Rank Checker searches the phrase a customer would type, like "dentist in Austin, TX", and finds your prospect in Google's own order. It searches the phrase, not a GPS point, so run it once per neighborhood for a rough visibility map.



Figure 4.4 A fictional clinic's rank for "dentist in [neighborhood], Austin", one check per phrase. Strong at home, invisible two miles away: that's the opening line of an SEO pitch.

Six businesses, scored

Business (fictional)	Reviews	Website	Maps rank	Score	Tier
Lakeview Smiles	38, few replies	Fails mobile, SEO 52	#14	**80**	A
Elm Street Dental Care	9, no replies, none recent	None	Not in top 40	**80**	A
Southpark Smiles	47, 22% replies	Fails mobile, SEO 58	#9	**55**	B
Riverbend Dentistry	120, 12% replies	No schema, SEO 71	#7	**30**	C
Northside Family Dental	212, 61% replies	Strong, SEO 84	#3	**0**	C
Hillcrest Dental Studio	Rated 3.2	Basic	#25	**n/a**	Skip

Tiers: A is 60 or more, B is 35 to 59, C is under 35. Chapter 8 adds timing signals on top of this fit score.

KEY TAKEAWAYS

- Cover a city with a grid: every area times every category term, until results saturate.
- Dedupe on place ID, then phone, then domain, then name and street, and never on social URLs.
- Score the gap, not the size: thin reviews, weak sites and page-two rankings are your openings.

Read: Best Proxies for Google Maps Scraping · Proxy Geo-Targeting Explained

05



B2B Leads From Company Websites

A company's website is its most honest profile: what it sells, who it serves, how it prices, what it runs on and how to reach it. This chapter shows which pages to read, what to pull from each one, and how to collect it without being a nuisance.

IN THIS CHAPTER

5.1 Which pages to read

5.2 What to extract

5.3 Contacts with code (optional)

5.4 Crawl politely

5.1 Which Pages to Read

You rarely need more than six to ten pages of any company's site. Find them through the homepage's menu and footer, or the sitemap, rather than guessing URLs. A sitemap's last-modified dates double as a change signal: a pricing page updated last week is worth a fresh look.

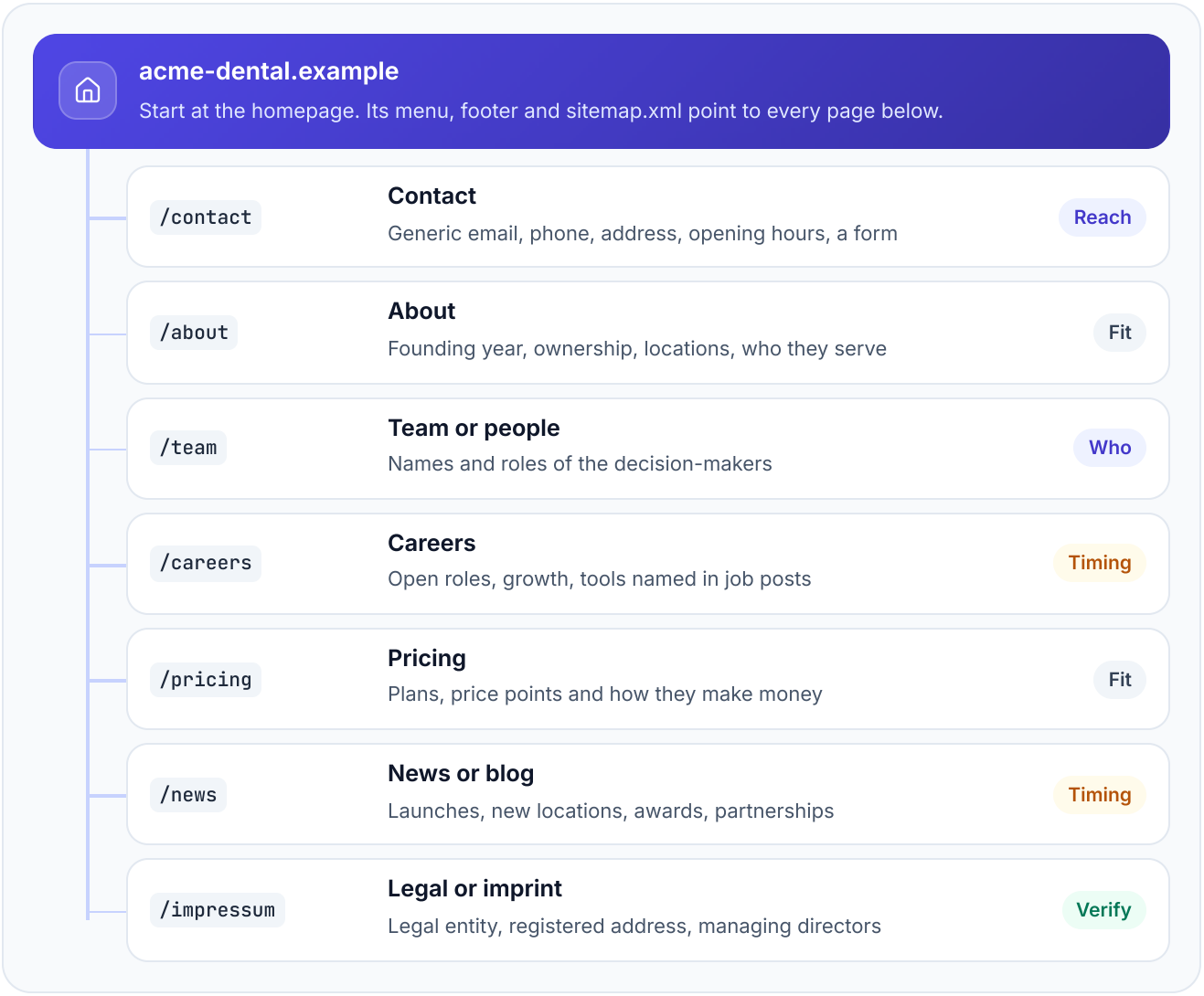


Figure 5.1 The pages worth reading on a typical company site, and what each one tells you. Paths vary from site to site, so follow the links rather than guessing them.

The imprint page is a gift in Germany and Austria

Businesses there must publish an imprint (Impressum) with their legal name, registered address and responsible managers. It's the most reliable place to confirm who you're dealing with, and it's meant to be read.

5.2 What to Extract

Collect fields you'll actually use to qualify, route or write to a prospect, and record the page each one came from. Free ProxyHorizon tools cover most of the list without code.

Field	Where it lives	Tool that pulls it
Generic emails	Contact page, footer, mailto links	Contact Information Finder
Phone numbers	Header, contact page, tel links	Contact Information Finder
Address and hours	Footer, contact page, structured data	Schema Extractor
Social profiles	Header and footer icons, schema sameAs	Social Media Links Finder
Platform and tech	Page code, headers, cookies	Website Technology Checker
Plans and prices	Pricing page, often rendered by JavaScript	Website Pricing Extractor
Decision-makers	Team and about pages	By hand, or LinkedIn Company Employees
Open roles	Careers page and job boards	By hand, or a job board feed

Structured data: the cleanest fields on the page

Many sites describe themselves to search engines in JSON-LD: an `Organization` or `LocalBusiness` block with the official name, phone, address and social links, already labeled. When it's there, trust it over scraped text. The Schema Extractor pulls every JSON-LD, microdata and RDFa item from a rendered page into one table.

LocalBusiness JSON-LD, as a dental clinic might publish it

```
{
  "@context": "https://schema.org",
  "@type": "Dentist",
  "name": "Bright Smile Dental Studio",
  "telephone": "+1-512-555-0142",
  "email": "hello@brightsmile.example",
  "address": {
    "@type": "PostalAddress",
    "streetAddress": "120 Example Ave",
    "addressLocality": "Austin",
    "addressRegion": "TX",
    "postalCode": "78701"
  },
  "sameAs": ["https://social.example/brightsmile"]
}
```

5.3 Contacts With Code (Optional)

Prefer no code? The Contact Information Finder does all of this on any site, contact page included, and exports CSV. For developers, here's the core of it in Python with requests, BeautifulSoup and phonenumbers.

Python · find and normalize contacts on one page

```python
import re
import requests
import phonenumbers
from bs4 import BeautifulSoup

HEADERS = {"User-Agent": "AcmeLeadsBot/1.0 (+https://acme.example/bot)"}
EMAIL = re.compile(r"[\w.+-]+@[\w-]+(?:\.[\w-]+)+")
AT = re.compile(r"\s*[\[(]\s*at\s*[\])]\s*", re.I)            # "jane [at] firm"
DOT = re.compile(r"\s*[\[(]\s*dot\s*[\])]\s*", re.I)          # "firm [dot] com"
IMAGES = (".png", ".jpg", ".jpeg", ".gif", ".webp", ".svg")   # logo@2x.png isn't an email

def find_contacts(url, region="US"):
    html = requests.get(url, headers=HEADERS, timeout=20).text
    soup = BeautifulSoup(html, "html.parser")
    text = DOT.sub(".", AT.sub("@", soup.get_text(" ")))
    hrefs = [a["href"].strip() for a in soup.select("a[href]")]
    mailtos = [h[7:].split("?")[0] for h in hrefs if h.lower().startswith("mailto:")]
    tels = [h[4:] for h in hrefs if h.lower().startswith("tel:")]

    found = EMAIL.findall(text) + mailtos
    emails = {e.lower() for e in found if not e.lower().endswith(IMAGES)}
    phones = {
        phonenumbers.format_number(m.number, phonenumbers.PhoneNumberFormat.E164)
        for m in phonenumbers.PhoneNumberMatcher(text + " " + " ".join(tels), region)
    }
    return {"source": url, "emails": sorted(emails), "phones": sorted(phones)}
```

What the snippet takes care of

- **Obfuscated addresses.** "jane [at] firm [dot] com" becomes a real address before matching.
- **False positives.** Retina image names like `logo@2x.png` are filtered out.
- **One phone format.** Every number comes out in E.164, ready to dedupe.
- **Provenance.** The source URL travels with every result.

> ⚠ **An obfuscated address is a hint**
>
> Someone who writes "jane [at] firm [dot] com" expects humans, not harvesters. Decoding it for one relevant, personal message is fair. Feeding it into bulk sequences isn't.

5.4 Crawl Politely

Most small-business sites run on modest hosting, and your crawler is a guest. Polite crawling isn't only ethical, it's practical: sites that never notice you never block you. Check each site's **robots.txt** before you fetch anything. It's a public standard (RFC 9309) that says which paths crawlers may visit, and ProxyHorizon's free Robots.txt Tester shows the rule that applies to any URL.

Setting	Polite default	Why
Requests at once, per site	One	Small sites run on small servers
Delay between pages	2 to 5 seconds, randomized	Spreads load the way a reader would
Pages per company	Ten or fewer	Contact, about, team, careers and pricing are enough
User-Agent	Your bot's name and a contact URL	Owners can reach you instead of blocking you
robots.txt	Fetch it first and obey Disallow rules	It's the site's stated wish
On 403 or 429	Stop that site for a day; honor Retry-After	Pushing harder never pays
Timeouts	10 s to connect, 30 s to read	Dead sites shouldn't stall the whole run

Keep the receipt for every field

Store the source URL and the date next to each value you collect. You'll need the source when a prospect asks where you found their details, which GDPR lets them do, and the date tells you when a value is old enough to re-check.

Public pages only

Never crawl behind a login, past a paywall or through a CAPTCHA, and never use proxies to dodge a block a site has deliberately put on you. If a site doesn't want to be crawled, take the hint and move on.

KEY TAKEAWAYS

- Six to ten pages tell you almost everything: contact, about, team, careers, pricing, news and legal.
- Trust structured data over scraped text, and keep the source URL and date for every field.
- Crawl one page at a time per site, obey robots.txt, and stop at the first 403 or 429.

Read: Web Scraping Legality and robots.txt · How to Clean and Store Scraped Data

06



Trigger Signals: Timing Beats Volume

The same email lands very differently the week a company raises money, posts five sales jobs or collects a string of one-star reviews. This chapter shows where those signals appear, how much each one is worth, how fast it goes stale and how often to look.

IN THIS CHAPTER

6.1 Where signals show up

6.2 The signal library

6.3 Shelf life and a monitoring rhythm

6.1 Where Signals Show Up

A trigger gives you a reason to write that's about them, not you. It also tells you when a budget conversation is probably already happening inside the company. Sending fewer emails at better moments beats sending more at random, and every signal below can be watched from public pages.

Hiring

Careers pages and job boards show what a company is about to invest in. A first marketing hire, three sales roles or a "head of operations" each say something different.

Watch with: LinkedIn Company Employees, which shows the share of staff who joined in the last year

News and funding

Funding rounds, acquisitions, new offices and awards all reach the news. So do layoffs, which is a signal to wait.

Watch with: Google News Checker, live results for any company in 15 countries

Pricing and page changes

A new plan, a price rise or a fresh landing page means new messaging, new pages and often a new problem to solve.

Watch with: Website Pricing Extractor monthly, and Website Page Comparison for two live pages side by side

Review spikes and slumps

A run of complaints about the same thing, a sudden burst of praise after a launch, or months of silence all say something about the business.

Watch with: Google Reviews Analyzer, sorted by newest, with the owner reply rate

Tech changes

A move to a new platform or CMS is a project with loose ends: redirects, tracking, speed and content that didn't migrate cleanly.

Watch with: Website Technology Checker, run every quarter

Ads and social activity

A business that starts advertising has budget. One that suddenly posts every day after months of silence has a new person, or a new push.

Watch with: Instagram Profile Analyzer for posting rhythm, plus public ad libraries

Compare against your prospect's rival

Website Page Comparison lines up any two live pages across 26 checks. Put a prospect's service page next to the competitor page that outranks it, and you have a specific, useful opening line instead of a generic audit offer.

6.2 The Signal Library

Not every signal deserves the same response. Weight each one by how strongly it predicts a need for what you sell, and note how long it stays useful. This library is a starting point; your own reply data should adjust it within a couple of months.

Signal	Weight	Goes stale after	Opening angle
Funding announced	●●●●●	60 to 90 days	Teams spend on growth right after a round
Hiring for a role you support	●●●●○	While the post is live, then 30 days	Help the new hire get results faster
New leader in your buyer's seat	●●●●○	The leader's first 90 days	New leaders set priorities early
New location or expansion	●●●●○	60 days	Get the new site found in its own area
Platform or tech change	●●●○○	90 days	Migrations leave loose ends to tidy up
Pricing or packaging change	●●●○○	30 days	New plans need new pages and messages
Review slump	●●●○○	14 to 30 days	A fix for the complaint customers repeat
Started advertising	●●○○○	30 days	Ads only work if the landing page does
Content or social burst	●●○○○	14 days	Turn the new momentum into enquiries

Weights and shelf lives are ProxyHorizon's suggested starting values, not measured benchmarks.

Weights depend on what you sell

A reputation management agency should rate a review slump a five, because it's the whole reason to call. A recruiting platform cares about hiring above everything else and can ignore pricing changes. Rewrite the weight column for your offer before you score anything, and keep the shelf lives unless your data says otherwise.

> ⚠ **Some signals mean "not now"**
>
> Layoffs, a recent acquisition by a larger group, or a public crisis usually freeze new spending. Treat them as negative signals: pause the account and look again in a quarter.

6.3 Shelf Life and a Monitoring Rhythm

A signal's value drops fast. The week a company announces a new location, every vendor in town notices; two months later the decisions have been made. Act within days, and let stale signals fall out of your scores automatically (Chapter 8 shows how).

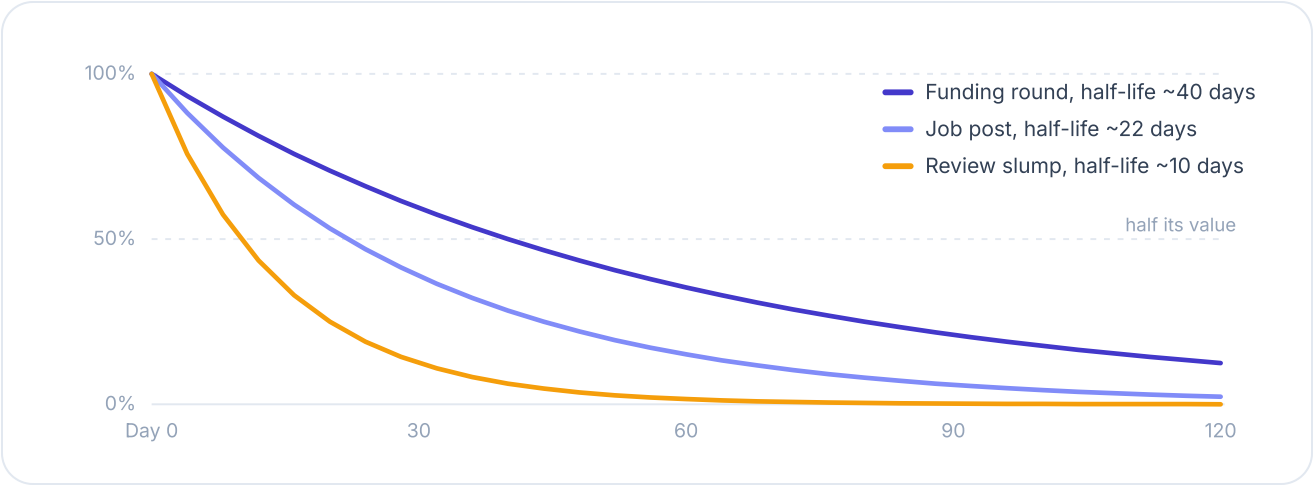


Figure 6.3 Illustrative half-lives, not measurements. The point is the shape: most of a signal's value is gone within a few weeks, so the check rhythm below matters more than the exact numbers.

A monitoring rhythm you can keep

DAILY
- News on your top 20 accounts
- Replies and opt-outs, before anything else

WEEKLY
- Job posts for your ICP
- New reviews on A-tier prospects
- New Maps listings in your areas

MONTHLY
- Pricing pages and sitemaps
- Maps ranks for open deals
- Social posting rhythm

QUARTERLY
- Tech stack re-checks
- Full list refresh and re-verify
- Review weights and shelf lives

KEY TAKEAWAYS

- A signal gives you a reason to write that's about them, and a moment when budget is in play.
- Weight signals for your own offer, and treat layoffs and acquisitions as "not now".
- Signals go stale in weeks, so check on a fixed rhythm and act within days.

07



Enrichment, Verification and Deliverability

A list is only as good as the addresses on it. This chapter finds the right person, verifies their address without burning your domain, cleans phones and records, and sets up the sending basics Gmail, Yahoo and Outlook now expect from anyone who emails at volume.

IN THIS CHAPTER

7.1 Find the right person

7.2 Verify before you send

7.3 Bounces, phones and hygiene

7.4 Deliverability basics

7.5 What the inbox providers require

7.1 Find the Right Person

In a five-person business, the owner is the buyer. In a 300-person company, the owner of the problem you solve is usually a manager two levels down, with a director who signs off. Match the contact to the size of the company before you look for anyone's email.

Company size	Contact first	Where to find them
1 to 10 staff	The owner or founder	About page, owner replies on reviews, registry records
10 to 50	The owner, or the manager of the function you help	Team page, LinkedIn Company Employees
50 to 200	The head of that function	LinkedIn Company Employees, press releases
200 and up	The manager who owns the problem, plus their director	LinkedIn Company Employees, event speaker lists

Email patterns

Most companies use one format for every staff address. For Jane Doe at `acme.example`, it's usually one of these:

`jane@`	`jane.doe@`	`jdoe@`	`janed@`	`jane_doe@`
first	first.last	initial + last	first + initial	first_last

To pick the right one, find any address the company has published under its own domain, on a press page, a blog author box, a PDF or the imprint, and copy its pattern. A guessed address is only a hypothesis until section 7.2 checks it.

A role inbox is often the right first choice

For micro-businesses, `hello@` or `info@` usually lands with the owner anyway. A generic address that doesn't name a person usually isn't personal data at all, which makes it the lower-risk option. At larger companies it reaches a queue, so find the person instead.

A guessed address is still personal data

Building jane.doe@ from a pattern means you're processing her personal data under GDPR, even though nobody published that exact address. Have a legitimate reason, tell her where her details came from at first contact, and delete them if she objects (Chapter 9).

7.2 Verify Before You Send

Every bounce tells the inbox providers you don't know who you're writing to. Verify each address up a ladder of checks, cheapest first, and stop as soon as one fails.

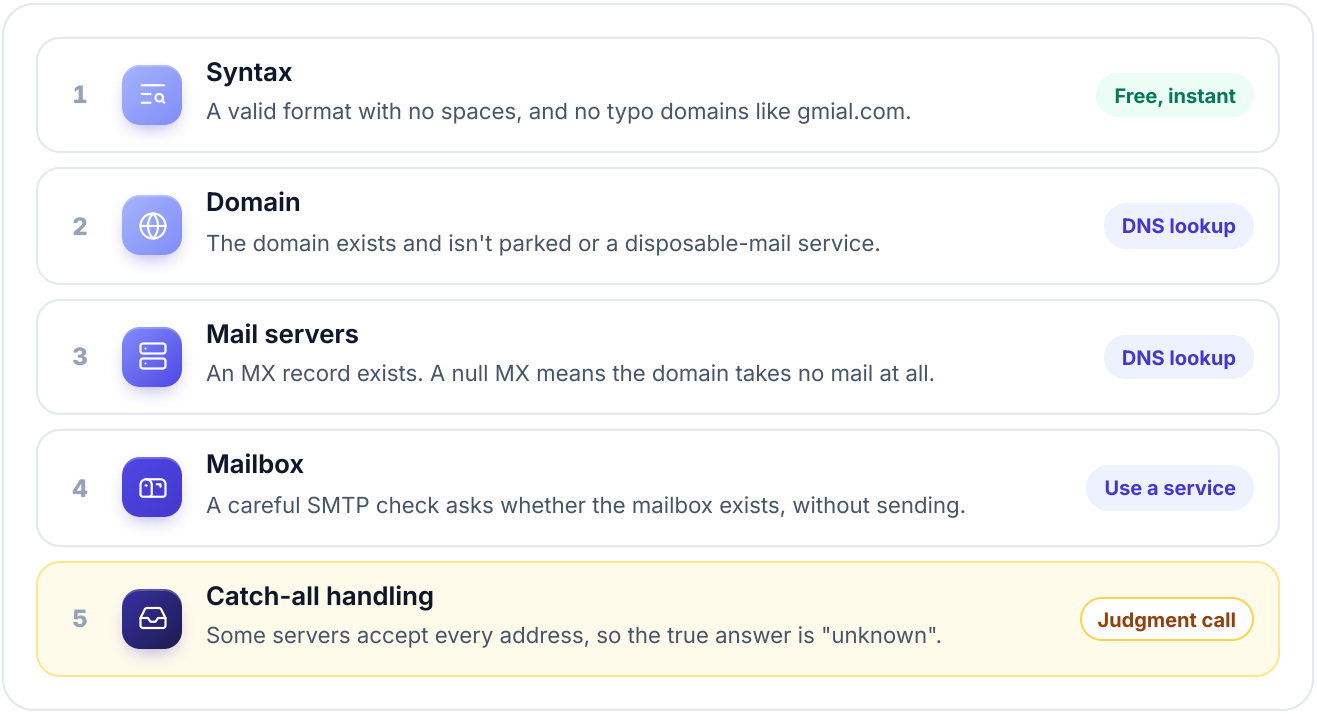


Figure 7.2 The verification ladder. Steps 1 to 3 are free and safe to run yourself. Leave step 4 to a reputable verification service, away from your sending IPs: mail servers treat bulk mailbox probing as abuse and blocklist the source.

Python · dnspython · check a domain's mail servers

```python
import dns.resolver

def mx_check(domain):
    """Return a verdict and the domain's mail servers, best first."""
    try:
        answer = dns.resolver.resolve(domain, "MX", lifetime=5)
    except dns.resolver.NXDOMAIN:
        return "no-such-domain", []
    except (dns.resolver.NoAnswer, dns.resolver.NoNameservers):
        return "no-mx", []              # mail may still go to the A record: treat as risky
    except dns.exception.Timeout:
        return "retry-later", []
    hosts = [str(r.exchange).rstrip(".") for r in sorted(answer, key=lambda r: r.preference)]
    hosts = [h for h in hosts if h]     # "0 ." is a null MX: the domain takes no mail
    return ("ok" if hosts else "null-mx"), hosts

print(mx_check("example.com"))          # ('null-mx', []): example.com accepts no email
```

7.3 Bounces, Phones and Hygiene

Verification gives every address a verdict. Each verdict has exactly one right action, and writing them down stops anyone from sending to "unknown" on a busy Friday.

Result	What it means	What to do
Valid	The mailbox exists	Send
Invalid	No such mailbox or domain	Remove, and add to your suppression list
Catch-all	The server accepts any address	Only A-tier leads, in small batches; watch bounces
Unknown	The server didn't answer in time	Retry in a day, then treat as catch-all
Role address	A shared inbox such as info@	Fine for small firms; find a person at larger ones
Disposable	A throwaway mail service	Remove

Bounce rate: our rule of thumb

Under 1%
Healthy. Keep verifying every batch the same way.

1 to 2%
Tighten up: re-verify old records and cut catch-alls.

Over 2%
Stop sending and clean the list before the next batch.

Phones: store them in E.164

E.164 is the international format: a plus sign, the country code and the number, up to 15 digits with no spaces. `(512) 555-0142` becomes `+15125550142` and `020 7946 0018` becomes `+442079460018`. Keep the original next to it, note mobile or landline when you know it, and remember calls and texts have their own rules (Chapter 9).

Hygiene rules worth automating

- [ ] Source URL and date stored for every field
- [ ] Suppression list checked before every send
- [ ] Names trimmed and properly capitalized
- [ ] Hard bounces suppressed automatically
- [ ] One record per business, deduped as in 4.2
- [ ] Re-verify addresses older than 90 days
- [ ] Personal data you don't use, deleted
- [ ] Opt-outs kept forever, even after deletion

7.4 Deliverability Basics

Inbox providers decide where your email lands before anyone reads it, and authentication is the first thing they check. Three DNS records prove your mail is really yours. Your email provider gives you the exact values; your job is to publish them and confirm they pass.

RECORD	WHAT YOU PUBLISH IN DNS (EXAMPLE)	WHAT THE RECEIVER ASKS
SPF	`v=spf1 include:_spf.mailhost.example ~all`	Is this server allowed to send mail for the domain?
DKIM	`s1._domainkey v=DKIM1; k=rsa; p=MIIB...`	Did the domain sign this message, and is it unchanged?
DMARC	`_dmarc v=DMARC1; p=none; rua=mailto:dmarc@acme.example`	Does the visible From domain match SPF or DKIM, and what if it doesn't?

Pass and align: you're eligible for the inbox. Fail: expect the spam folder, or an outright rejection.

Figure 7.4 SPF, DKIM and DMARC in one view. Records are examples with placeholder hosts; copy the real values from your email provider.

Use a separate domain for cold outreach

Your main domain carries invoices, support replies and password resets. Send prospecting email from a second domain, such as `tryacme.example` next to `acme.example`, so a rough campaign can't hurt the mail your business runs on. Give it its own SPF, DKIM and DMARC, point it at your website, and use real inboxes that someone reads. It protects your main domain; it doesn't excuse bad sending.

Warm up and stay modest

Start small

Begin a new mailbox at 10 to 20 emails a day and raise it gradually over three to four weeks.

Cap each inbox

Our rule of thumb is about 50 cold emails a day per mailbox. Add mailboxes, not volume.

Earn replies

Real conversations build reputation faster than any warm-up tool. Write to people who'll answer.

Test it before you trust it

Send a test to a Gmail account and choose "Show original" from the message menu. It lists SPF, DKIM and DMARC as PASS or FAIL: the quickest proof your records work.

7.5 What the Inbox Providers Require

Checked in November 2026 against Google's, Yahoo's and Microsoft's published sender guidelines. These rules change, so check them again before a large send.

Requirement	Gmail	Yahoo	Outlook.com
Who it covers	Every sender; more rules at about 5,000+ a day to Gmail. Bulk status is permanent.	Bulk senders	Domains sending 5,000+ a day, enforced since May 2025
SPF and DKIM	SPF or DKIM for everyone; both for bulk	Both	Both must pass
DMARC	Required for bulk; `p=none` is enough	`p=none` or stricter, and it must pass	Required; `p=none` at minimum
From-domain alignment	With SPF or DKIM, for bulk	With SPF or DKIM; relaxed is fine	With SPF or DKIM
One-click unsubscribe	Marketing mail: RFC 8058 header and a visible link	Header plus a visible link	Recommended
Honor unsubscribes	Within 48 hours	Within 2 days	Promptly
Spam complaint rate	Below 0.3% in Postmaster Tools; aim for under 0.1%	Below 0.3%	Keep it low
Also required	Forward and reverse DNS, TLS, RFC 5322	Valid forward and reverse DNS	Non-compliant mail goes to Junk, then may be rejected

Sources: Google's sender guidelines and FAQ, Yahoo Sender Hub, Microsoft's Outlook sender announcement.

Before every campaign

- [ ] SPF, DKIM and DMARC published and passing
- [ ] From domain aligned with SPF or DKIM
- [ ] One-click unsubscribe header and a visible opt-out line
- [ ] Unsubscribes honored within two days
- [ ] Complaints watched in Google Postmaster Tools
- [ ] Last batch under 2% hard bounces
- [ ] Every address verified in the last 30 days
- [ ] Your postal address in the footer

KEY TAKEAWAYS

- Match the contact to company size, and copy a published email pattern rather than guessing one.
- Verify up the ladder, cheapest check first, and leave mailbox checks to a service.
- Authenticate a separate sending domain, warm it up slowly and keep bounces under 2%.

08



Scoring, Routing and CRM Hygiene

Not every verified lead deserves the same effort. This chapter scores each one on fit and on intent, sorts them into A, B and C tiers, sends each to the right next step, and keeps your CRM clean enough that the team trusts what it says.

IN THIS CHAPTER

8.1 A fit × intent model

8.2 A lead scorecard and three tiers

8.3 Routing rules

8.4 CRM hygiene and data decay

8.1 A Fit × Intent Model

One score hides too much. A perfect-fit company with nothing happening and an average one that just raised money can both land on 60, and they need very different treatment. Score every lead twice: fit from your ICP, intent from live signals.

Fit, up to 50

Industry or category matches	+15
Size inside your band	+10
Inside your territory	+10
Tech or business model matches	+10
A visible gap you can fix	+5

Intent, up to 50

Strongest live signal, weight × 6	up to +30
Each extra live signal	+10, max +20
Signal in the first half of its shelf life	× 1
In the second half	× 0.5
Past its shelf life	× 0

LEAD SCORE, OUT OF 100

```
fit (0-50) + intent (0-50, each signal multiplied by its freshness)
```

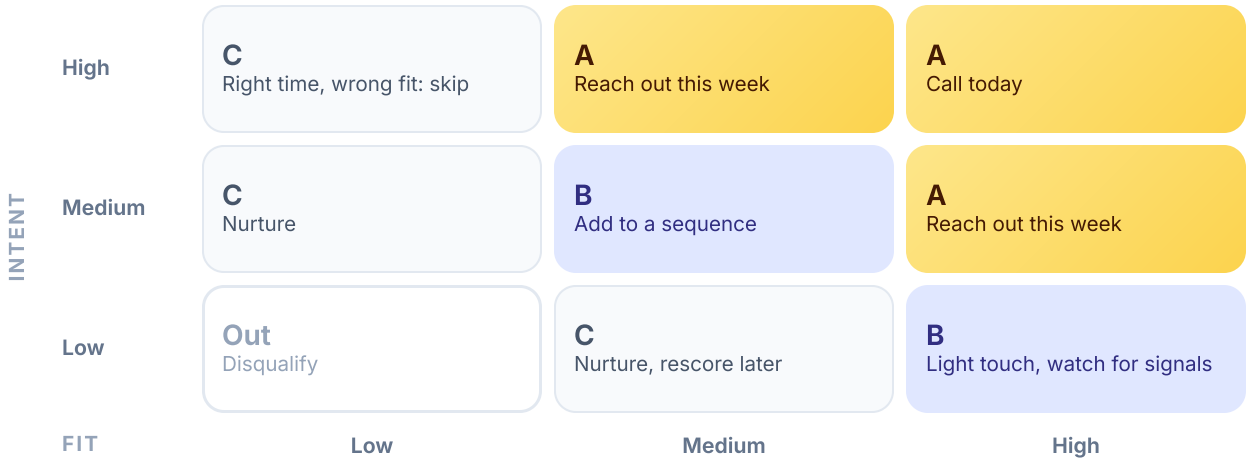


Figure 8.1 Fit decides whether a lead belongs on your list at all; intent decides how fast to move. Wrong-fit leads stay out of the A tier however hot their signals are.

For local prospects, halve the opportunity score from Chapter 4 and use it as fit. Start with these weights, then move points toward whatever your won deals had in common.

8.2 A Lead Scorecard and Three Tiers

Put both scores on the record with the reason behind every point. A rep who can see why a lead is an A writes a better first line, and a manager can check the scoring in seconds.

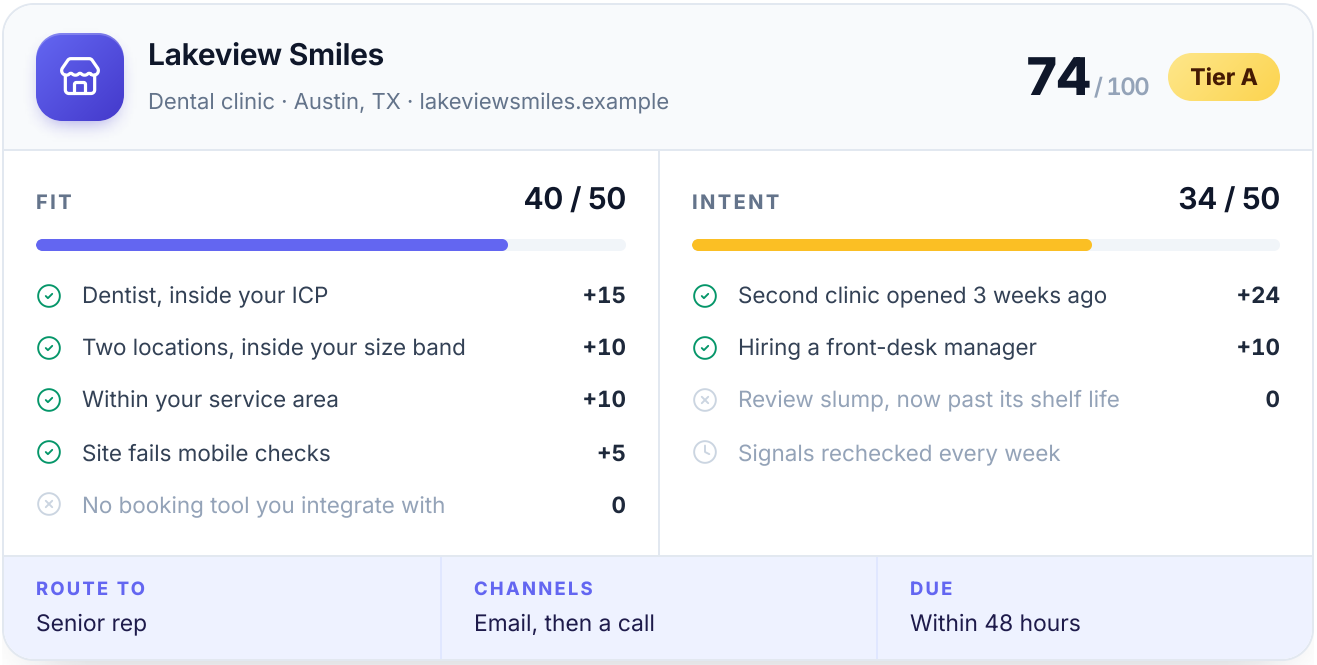


Figure 8.2 A fictional lead record. Fit 40 plus intent 34 makes 74: a Tier A lead with an obvious reason to call, the new clinic.

Three tiers, three levels of effort

Tier	Score	Who works it	Channels and speed	Personalization
A	70 and up	Founder or a senior rep	Email, LinkedIn and phone, within 48 hours	Two or three specific observations
B	45 to 69	An SDR, or a sequence	Email and LinkedIn, within a week	One observation plus a segment angle
C	Under 45	Nobody, yet	Recheck signals in 90 days	Segment level only

Keep the A tier scarce

If more than about a quarter of your list scores A, your weights are too generous and the tier stops meaning anything. Raise the threshold until A leads are the ones you'd be annoyed to miss.

8.3 Routing Rules

Routing decides who does what next, and by when. Run every new or rescored lead through the same questions in the same order. The first two protect you; the last two decide where the effort goes.

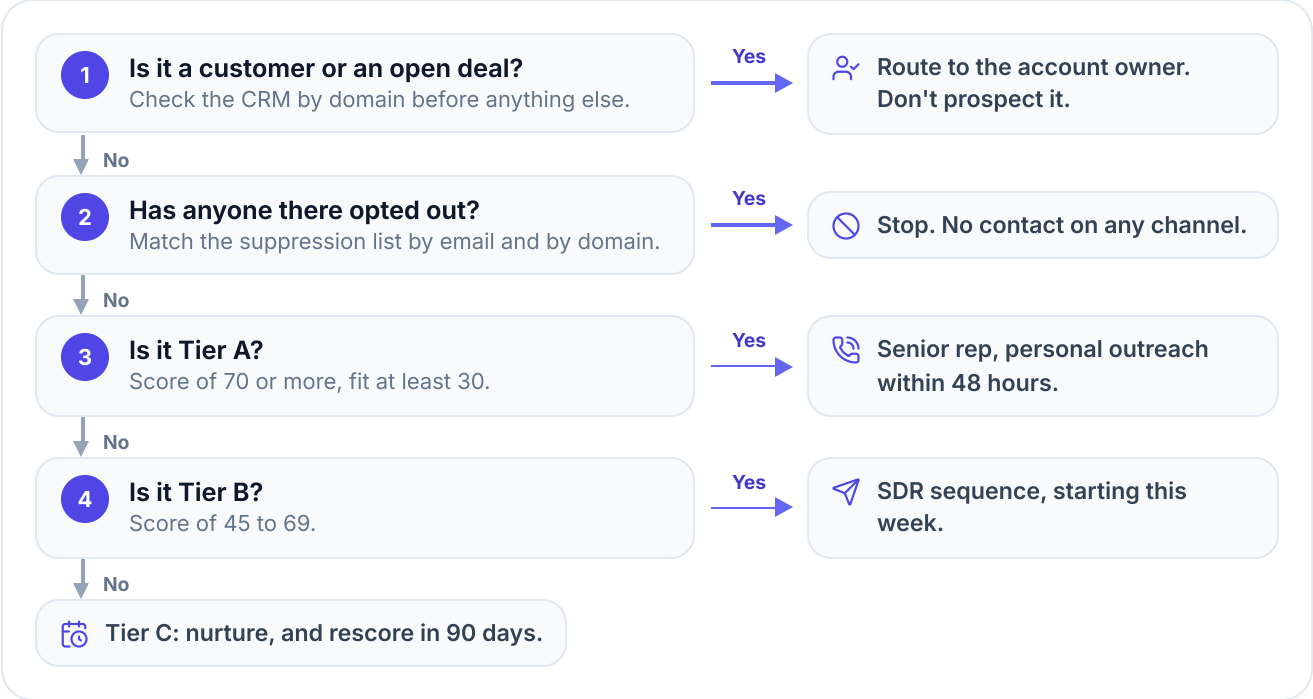


Figure 8.3 Four questions, in order. Customers and opt-outs are caught before anyone looks at a score.

Speed wins A leads

Signal-driven leads cool in days. Keep the time from turning A to first touch under two working days.

Territory or rotation

Route local businesses by territory and the rest in rotation. Never let two people work one domain.

Always set a date

A lead without a due date is one nobody owns. Routing sets the owner and deadline together.

Turn the rules into automations

When this happens	Your CRM or workflow tool does this
A lead becomes Tier A	Assign an owner by territory, create a task due in two working days, post an alert
An opt-out arrives on any channel	Set the flag, stop every sequence, add the email and domain to the suppression list
A signal expires	Rescore the lead; if it drops a tier, move it to the right queue
An owned lead goes 30 days untouched	Return it to the pool and tell the manager

8.4 CRM Hygiene and Data Decay

Business data decays every month: people change jobs, businesses move, merge or close, and phone numbers get reassigned. A CRM nobody trusts gets worked around, so treat these fields as required and these rules as automatic.

Required field	Why it's required
Company name and root domain	The domain is the unique key that prevents duplicate accounts
Source URL and date collected	Answers "where did you get my details?" and shows the record's age
Fit score, intent score, tier	Drives routing; rescored whenever a signal changes
Owner and next-step date	Someone is accountable, by a date
Lawful basis or consent status	Required for EU, UK and Canadian contacts (Chapter 9)
Last verified date	Tells you when the email needs checking again
Opt-out flag	Kept forever, on every channel, even after the record is deleted

Merge rules

- Match companies on root domain or place ID, people on email.
- The newest verified value wins each field.
- Never overwrite a manual edit with scraped data.
- Keep all activity history from both records.

Refresh rules

- Re-verify emails older than 90 days before any send.
- Recheck signals weekly for A and monthly for B.
- Rescore every lead when a signal is added or expires.
- Delete unqualified, never-contacted records after 12 months.

Old data is a liability, not an asset

GDPR's storage-limitation principle says personal data shouldn't be kept longer than you need it. A scheduled clean-out lowers your legal exposure and your bounce rate in the same move.

KEY TAKEAWAYS

- Score fit and intent separately, and let stale signals drop to zero automatically.
- Check customers and opt-outs before any score, then route by tier with a deadline.
- Require source, dates and an opt-out flag on every record, and clean out old data on schedule.

09



Outreach That Earns Replies

Good data earns you the right to send one relevant message. This chapter shows how to personalize from public signals without being creepy, structure an email people answer, run a multichannel cadence, draft faster with AI, and stay inside the rules in each market.

IN THIS CHAPTER

9.1 Relevant, never creepy

9.2 An email that earns a reply

9.3 A multichannel cadence

9.4 AI drafts, human judgment

9.5 The rules by market

9.6 GDPR in practice

9.1 Relevant, Never Creepy

Personalization works when it proves you understand someone's business. It backfires the moment it proves you've been watching them. The line between the two is simple: use business facts they published for business reasons, and leave everything else alone.

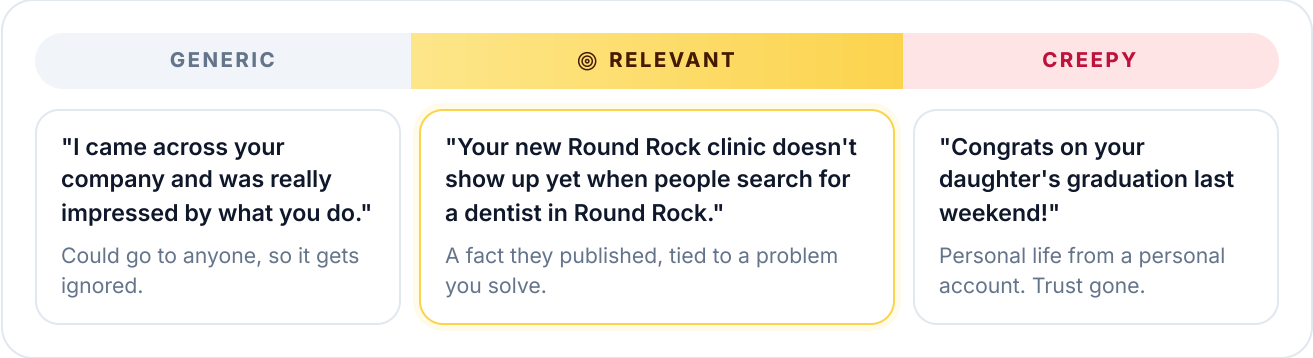


Figure 9.1 The relevance spectrum. Aim for the middle: specific to their business, silent about their private life.

Signals into first lines

Signal	A first line that works
Maps rank	"When I search 'dentist in Hyde Park', you're 14th, behind three clinics with fewer reviews than you."
Website gap	"On a phone, the Book button on your homepage sits behind the cookie banner."
Hiring	"Saw you're hiring your first marketing manager. Most teams want the site fixed before that person starts."
Review pattern	"Four of your last ten reviews mention waiting on hold. That's a fixable problem."
News	"Congrats on the new office in Leeds. Expansion is usually when outbound gets rebuilt."

Fair to use

Their website, listings, reviews, job posts, press and company social accounts, all used for what they're about: the business.

Leave alone

Personal social accounts, family, health, home life, and anything that reveals tracking, like "I saw you on our pricing page three times".

The trade-show test

Would the line sound normal if you said it to them face to face at their booth? If it would feel awkward out loud, it's awkward in an inbox. Cut it.

9.2 An Email That Earns a Reply

Cold emails that get answered tend to share one shape: a signal, the problem it usually means, a short reason to believe you and one small ask. Here it is with every part labeled.

FROM Sam Lee <sam@tryacme.example>
SUBJECT **Round Rock clinic on Maps** 1

Hi Maria, 2

Congrats on opening the Round Rock clinic. When I search "dentist in Round Rock", it isn't in the top 20 yet, while your Austin clinic ranks second. 3

New locations usually need their own listing, a local page and a few early reviews before Maps trusts them. 4 We've done exactly that for two other Austin practices and I'm happy to share what worked. 5

Worth 15 minutes next week? 6

Sam Lee, Acme Local · 100 Example St, Austin, TX 78701
I found your details on your clinic's website. Not the right person, or rather not hear from us? Reply "no" and we won't email again. 7

1. **Subject.** Short and specific, like a colleague's.
2. **Name.** A real person, spelled right.
3. **Signal.** A published fact you checked.
4. **Problem.** What that signal usually means.
5. **Proof.** A short, true reason to trust you.
6. **One ask.** Small and easy to say yes to.
7. **Source, address, opt-out.** Where you got their details, who you are and an easy way out.

Figure 9.2 A fictional first email, about 90 words. It names one signal, makes one ask and covers the legal basics in two lines.

Three frameworks worth keeping

Signal, problem, ask
The default above. Works whenever you have a fresh, specific signal.

Give first
Lead with something useful: the rank map, a mobile screenshot, three quick fixes. The ask comes second.

Quick question
"Are you the right person for local search, or is that someone else?" Best for routing at larger firms.

House rules: under about 120 words, plain text, one link at most, no attachments in a first email.

Subject lines that work
- Round Rock clinic on Maps
- Question about your booking page
- Hyde Park search results

Subject lines to skip
- Quick question!!! (shouting)
- Re: our call (a fake reply, and deceptive)
- Triple your revenue (hype)

9.3 A Multichannel Cadence

One email is rarely enough, and five identical follow-ups are worse than none. A good cadence mixes channels, adds something new at each step and ends politely. This one runs three weeks for an A or B lead.

Day 1

Email: the signal-led first touch

The email from 9.2. One signal, one ask.

EMAIL

Day 3

LinkedIn: view and connect, no pitch

Done by hand. A connection request with no note, or a one-line note that mentions the email.

LINKEDIN

Day 6

Email: give something useful

Send the rank map, the mobile screenshot or two quick fixes. Value, not "just checking in".

EMAIL

Day 9

Phone: a short call to the business line

For local businesses, avoid opening time and the lunch rush. Ask for two minutes, not a meeting.

PHONE

Day 14

Email: proof

A relevant example from a similar business, in three sentences.

EMAIL

Day 21

Email: the polite close

"Should I close the file on this?" Then stop, and let the signal monitor decide when to try again.

EMAIL

Any reply stops everything

A reply, an opt-out or a bounce ends the whole sequence on every channel at once. Automate that before you automate anything else.

Calls and texts have their own rules

In the UK, screen numbers against the TPS and CTPS before live sales calls. In the US, never text or autodial a mobile number without prior consent.

9.4 AI Drafts, Human Judgment

AI is good at first drafts, subject-line variations and tone. It's bad at facts it can't see and judgment it doesn't have. Let it draft, then put a person between the draft and the send button, every time.

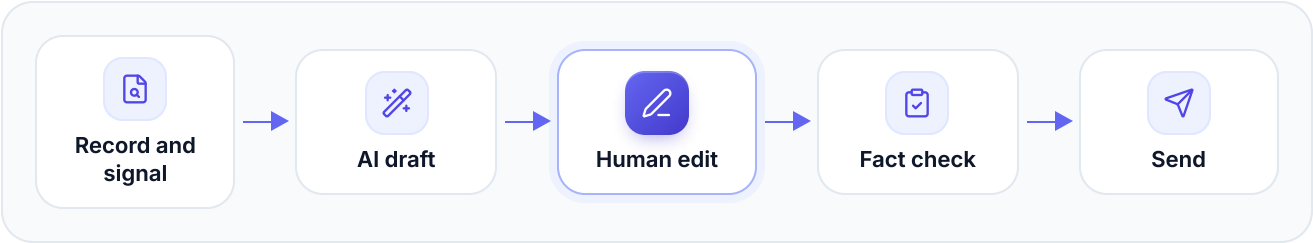


Figure 9.4 Give the AI the facts, never ask it to find them. The human step adds the one line only you could write and removes anything you can't stand behind.

AI Email Generator
Cold emails and follow-ups, three subject lines per version.

AI LinkedIn Post Generator
Posts that give prospects a reason to trust your profile.

Google Reviews Analyzer
The exact topics customers raise, to quote accurately.

The AI Email Generator writes cold outreach and follow-ups in the tone you choose and leaves placeholders where it lacks details, which is exactly where your research goes. Prospects also check your LinkedIn profile before they reply, so a few useful posts from the AI LinkedIn Post Generator, edited in your own voice, make the outreach land better.

Before an AI draft goes out

- [ ] Every fact matches its source
- [ ] Placeholders replaced with real details
- [ ] Sounds like you on a good day
- [ ] No invented compliments, numbers or clients
- [ ] One ask, under about 120 words
- [ ] Source line, address and opt-out in place

Feed AI tools the minimum

Paste in the business facts the draft needs, not whole CRM records. Check how your AI provider stores and uses inputs before any personal data goes near it.

9.5 The Rules by Market

General information, not legal advice. Laws differ by country and change often; for anything high-stakes, talk to a lawyer where your recipients are.

The law that applies depends on where your recipients are, not where you are. These four cover most B2B outreach from and into North America, the UK and the EU.

Law	Cold B2B email	You must	Opt-outs
GDPR and UK GDPR	Usually possible on legitimate interests after a balancing test, for named people in the EU or UK; national e-privacy rules add conditions	Document the balancing test; tell people where you got their data within a month or in your first message (Art. 14)	Stop marketing to anyone who objects, at once (Art. 21)
CAN-SPAM (US)	Allowed without prior consent; covers B2B email too	Honest headers and subject lines, say it's an ad, include your physical postal address and a working opt-out	Honor within 10 business days; keep it working 30 days after sending
CASL (Canada)	Needs consent. Implied if the person or their employer conspicuously published the address, with no "no unsolicited messages" note, and you write about their role	Identify yourself with contact details, include an unsubscribe mechanism, keep proof of consent	Within 10 business days
UK PECR	Allowed to corporate subscribers such as limited companies and LLPs; sole traders and some partnerships need consent	Say who you are and give a valid address to opt out; UK GDPR still covers named people	Promptly; keep a do-not-email list

$53,088
maximum penalty per violating email under CAN-SPAM, per the FTC's compliance guide

C$10M
maximum CASL penalty per violation for a business, per the CRTC

€20M or 4%
of worldwide annual turnover, whichever is higher, under GDPR's top tier

Inside the EU, check the country too

National e-privacy laws sit on top of GDPR. France's regulator, the CNIL, accepts cold email to professionals when it's about their job and opting out is easy. Germany generally expects prior consent, even for business addresses. In the UK, the Data (Use and Access) Act 2025 is updating PECR, so check the ICO's current guidance.

9.6 GDPR in Practice

For B2B outreach to named people in the EU or UK, legitimate interests is usually the workable basis: GDPR itself says direct marketing may be a legitimate interest. It isn't automatic, though. You have to weigh your interest against theirs, write the result down and be open about what you're doing.

The three-part balancing test

1 Purpose

Is there a real interest? Offering a relevant B2B service to a business usually qualifies.

2 Necessity

Do you need this person's data for it? Would the generic inbox do? Collect the minimum.

3 Balance

Would they reasonably expect it? It should be about their job, easy to refuse and never intrusive.

Run the test once per campaign or segment, not per email, and keep it with your records. If the balance tips against you, for example a personal address or an offer unrelated to their role, don't send.

Article 14 in one line, plus a link

When you didn't collect someone's data from them directly, you must tell them where it came from and how you use it, at the latest in your first message. In practice: one plain line in the email, and a link to a privacy notice covering the rest (who you are, why, the legal basis, how long you keep data and their rights, including complaining to a regulator).

> **"I found your details on your clinic's website. Here's how we handle them (link). Reply 'no' and we'll delete them."**

When someone objects

- Stop all marketing to them immediately, on every channel.
- Delete what you don't need, but keep a minimal suppression entry so they're never re-added from a fresh scrape.
- Answer access and deletion requests within a month, and log what you did.

KEY TAKEAWAYS

- Personalize with business facts they published, never with their private life.
- Signal, problem, proof, one ask: under 120 words, with a source line, address and opt-out.
- Know the rules where your recipients are, and treat every opt-out as permanent.

10



Infrastructure: Proxies, Automation and Cost per Lead

The engine needs plumbing: the right proxies for each source, a few automations that run it every week, and a cost model that tells you what each qualified lead really costs. This chapter covers all three, and when to build versus buy.

IN THIS CHAPTER

10.1 Proxies by source

10.2 Providers worth testing

10.3 No-code automation

10.4 Cost per lead, and build or buy

10.1 Proxies by Source

You don't need proxies to look up a few dozen businesses, and the ProxyHorizon tools handle their own. You need them when you run searches at volume, when results depend on location, or when a site sits behind bot protection. Match the proxy to the source, not to the project.

Source	Proxy type	Rotation	Targeting
Google Maps and reviews	Residential	Sticky per scrolling session	The city you're searching
Company websites	Datacenter; residential for protected sites	Per request	Country, if content varies
News and search results	Residential	Per request	The market's country
Social profiles, logged out	Residential or mobile	Sticky per session	The profile's country
Marketplaces	Residential	Per request	The marketplace's country
LinkedIn	Don't scrape it. Research by hand or use the LinkedIn Company Employees tool.		
Business registries	No proxies. Use the official API or bulk download.		

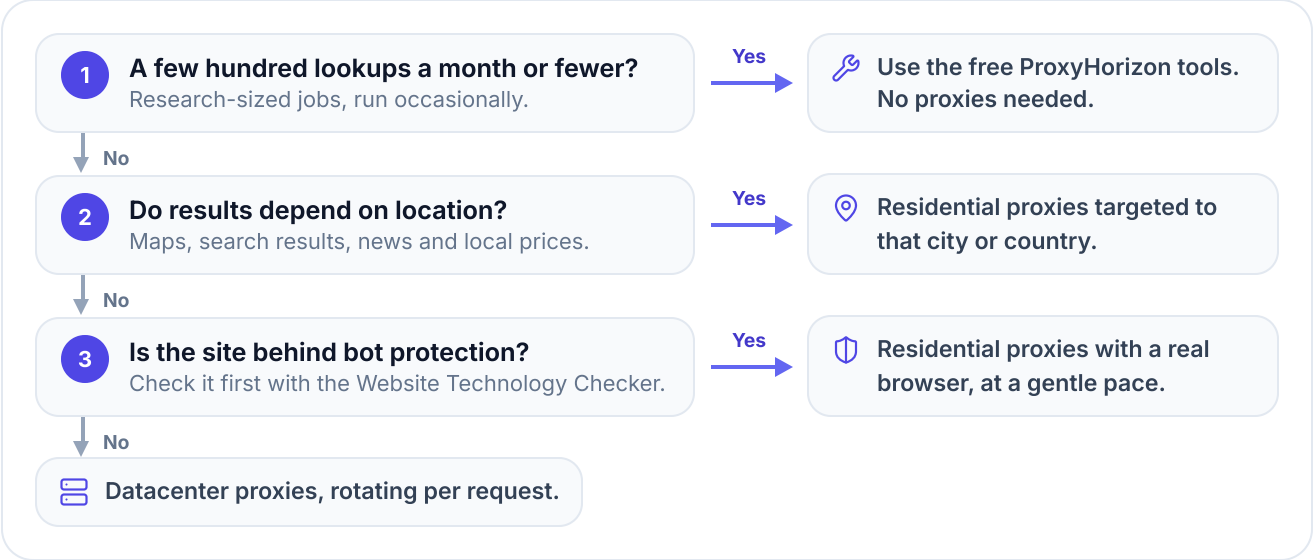


Figure 10.1 Pick the cheapest setup that works. Most company websites never get past question 3.

Read: What Is a Residential Proxy? · Proxy Geo-Targeting · Best Proxies for Market Research

10.2 Providers Worth Testing

Every network below is a credible choice. Start with the pick that matches your main job, then trial a second one on your hardest source before you commit.

SELLING TO LOCAL BUSINESSES
City-targeted residential: SOAX or Decodo

CRAWLING COMPANY WEBSITES
A small datacenter plan: Webshare

RUNNING IN BURSTS
Traffic that doesn't expire: IPRoyal

Bright Data

Proxies, scraper APIs and ready-made datasets, if you'd rather buy the data.

Residential · ISP · Datacenter · Mobile

RESIDENTIAL	COUNTRIES
$5.04/GB	195

Decodo

Formerly Smartproxy. Every proxy type and an easy dashboard.

Residential · ISP · Datacenter · Mobile

RESIDENTIAL	COUNTRIES
$8.50/GB	195

Geonode

Unlimited-bandwidth residential plans for heavy browser runs.

Residential

RESIDENTIAL	COUNTRIES
$4/GB	190

IPRoyal

Residential traffic that never expires, for bursty campaigns.

Residential · ISP · Datacenter · Mobile

RESIDENTIAL	COUNTRIES
$7/GB	195

NodeMaven

Quality-filtered IPs and long sticky sessions for scrolling.

Residential · Mobile

RESIDENTIAL	COUNTRIES
$3.50/GB	195

Oxylabs

A premium network plus scraper APIs for steady volume.

Residential · ISP · Datacenter · Mobile

RESIDENTIAL	COUNTRIES
$8/GB	195

PYProxy

A low entry price per GB, plus SOCKS5 per-IP plans.

Residential · ISP · Datacenter

RESIDENTIAL	COUNTRIES
$3.30/GB	190

SOAX

Clean pools with city and ISP targeting for local accuracy.

Residential · Mobile

RESIDENTIAL	COUNTRIES
$6.60/GB	195

Webshare

Affordable self-serve datacenter proxies for crawling sites.

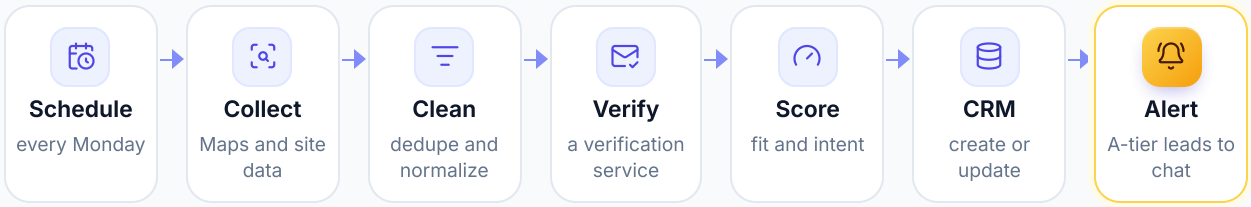
Datacenter · Residential

DATACENTER	COUNTRIES
$2.99/mo	50

Entry prices, product types and coverage from the ProxyHorizon directory, November 2026. Prices change often: see proxyhorizon.com/proxies for live numbers and proxyhorizon.com/compare to put any two side by side.

10.3 No-Code Automation

Once you've run the engine by hand (section 1.2), automate the boring middle. A weekly workflow in n8n, Make or Zapier can collect, clean, verify, score and file new leads, and only wake you up for the A tier.



Before the CRM step, check the suppression list. An automation that re-adds an opted-out contact is the costliest bug in the whole engine.

Figure 10.3 A weekly lead workflow. Each box is one or two nodes in n8n, or a module in Make or Zapier.

	n8n	Make	Zapier
Best for	Technical founders who want full control	Visual, multi-step scenarios	Non-technical teams who want speed
Hosting	Self-host, or n8n's cloud	Cloud	Cloud
Proxies	The HTTP Request node has a proxy option	Easiest through a scraping API	Easiest through a scraping API
Cost shape	Free to self-host, or paid cloud plans	Free tier, then usage-based plans	Free tier, then plans priced by tasks

Three workflows worth building first

New this week

Re-run your grid search weekly and send only listings you haven't seen before to the CRM. New businesses need everything.

Signal watcher

Check news and job posts for A and B accounts, add signals with dates, and rescore automatically.

Opt-out sync

Push every unsubscribe and "no" reply to one suppression list that every tool reads before sending.

Read: n8n vs Zapier · Best n8n Nodes for Proxy Data Collection · Best Proxy Providers for n8n · Best Proxies for Zapier

10.4 Cost per Lead, and Build or Buy

Say an agency produces 600 qualified, verified local leads a month with the setup from this book. Here's where the money goes, with the assumptions written out so you can swap in your own.

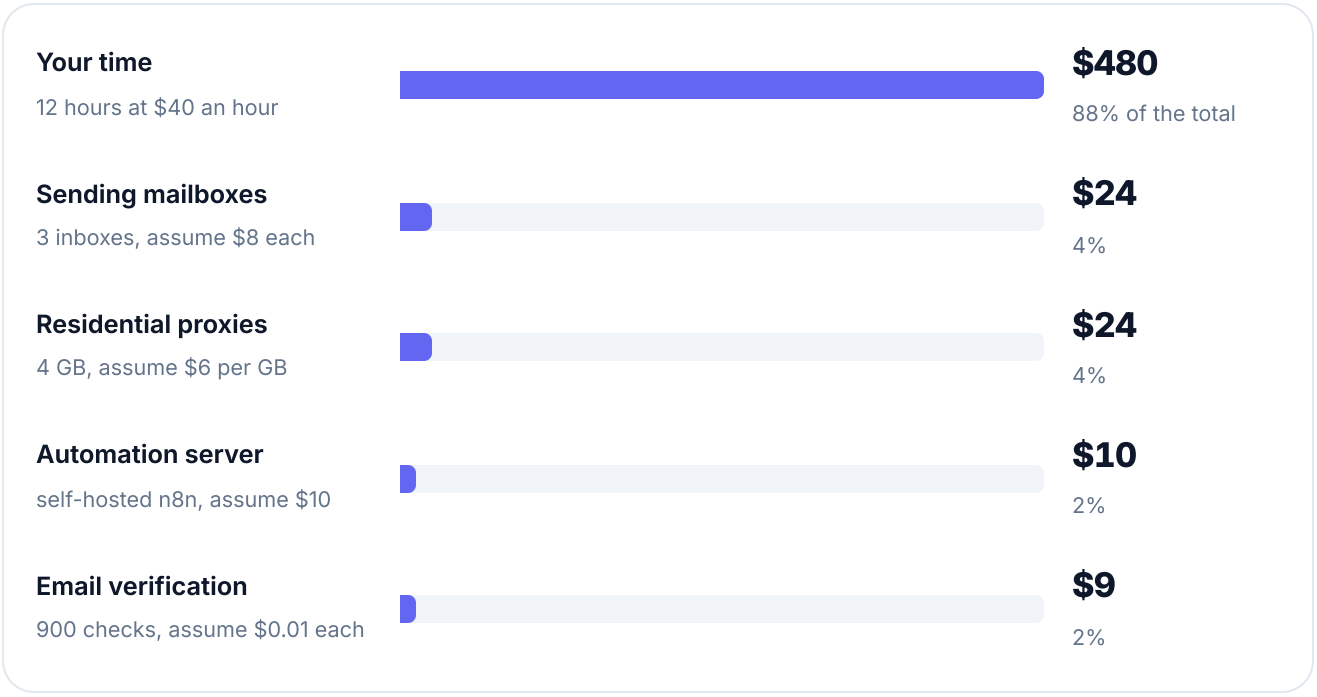


Figure 10.4 A worked example with assumed prices: $547 a month, or about $0.91 per qualified lead. If the month books 20 meetings, that's about $27 a meeting. Time is the cost to automate first.

Build or buy

Option	Best when	Watch for
Free tools, by hand	Small volumes, or testing a new ICP	Your hours; it doesn't scale
Automated build	Steady volume, someone technical	Upkeep when a source changes
Official APIs	A licensed feed, like the Places API	Per-call pricing, storage limits
Scraper APIs, datasets	Scale fast, or hard sources	Cost per record, freshness, sourcing
Contact databases	Sizing a market quickly	Decay, shared lists, unclear sourcing

KEY TAKEAWAYS

- Match proxies to the source: residential and city-targeted for Maps, datacenter for most company sites.
- Automate the hours-heavy middle of the engine first, with the suppression check built in.
- Track cost per qualified lead and per meeting. Time, not tooling, is usually the biggest line.

APPENDIX A

Your 30-Day Launch Plan

Here's how to turn the book into a working engine in a month, one focused week at a time, with a review on day 30 before you scale anything.

Week 1

Define the target, pick two sources

Fill in the ICP worksheet, list your disqualifiers and pick two sources: one complete, one that adds timing. Push 50 prospects through all nine stages by hand.

CHAPTERS 1 TO 3

Week 2

Collect, clean and set up sending

Run your grid search or site crawl, dedupe, normalize and score fit. Register a separate sending domain, publish SPF, DKIM and DMARC, and start warming it gently.

CHAPTERS 4, 5 AND 7

Week 3

Verify, write and start small

Verify every address, write the first email and the cadence, document the balancing test, publish your privacy notice and start at about 20 emails a day per mailbox.

CHAPTERS 7 AND 9

Week 4

Add signals, measure, automate one step

Start the signal rhythm, score intent and route by tier. Track cost per qualified lead, bounces, replies and meetings, then automate the step that ate the most hours.

CHAPTERS 6, 8 AND 10

Day 30

Review before you scale

Compare the four numbers with week 1. Move points toward what your replies had in common, drop the weaker source, and only then add volume.

CHAPTER 1

The best lead lists aren't bought or blasted. They're built from what businesses tell the world, timed to what's changing for them, and sent by someone who did their homework.

Lead generation | Free tools | Proxy directory | Guides

APPENDIX B

Checklists

Print these, or copy them into your project tool. Each one takes a couple of minutes and saves you from the mistakes that cost the most.

Before you add a source

- [ ] Public pages only: no login, paywall or CAPTCHA
- [ ] robots.txt allows the pages you need
- [ ] Terms checked; official API or bulk data preferred
- [ ] The exact fields you need, and why, written down
- [ ] Personal data limited to the people you'll contact
- [ ] Source URL and date stored with every field
- [ ] Pacing set: one request at a time per site
- [ ] Proxy type matched to the source (10.1)
- [ ] Dedup keys decided before the first run (4.2)

Before you send

- [ ] Recipient's country checked: GDPR, CASL, PECR
- [ ] Balancing test documented for EU and UK contacts
- [ ] Source line and privacy notice link in email one
- [ ] Real name, company and postal address included
- [ ] One-click unsubscribe and a plain opt-out line
- [ ] Suppression list checked by email and domain
- [ ] Honest subject line; no fake "Re:" or hype
- [ ] SPF, DKIM and DMARC passing on the sending domain
- [ ] Every address verified in the last 30 days

Every week

- [ ] Bounce rate under 2% on every batch
- [ ] Opt-outs processed within two days
- [ ] Every A-tier lead touched within 48 hours
- [ ] Calls screened against TPS and CTPS (UK)
- [ ] Spam complaints watched in Postmaster Tools
- [ ] Replies answered or routed within a day
- [ ] Stale signals expired and leads rescored
- [ ] The four numbers updated: cost per qualified lead, bounce, reply and meeting rates

This isn't legal advice

These checklists cover the common ground for B2B outreach. They don't replace advice from a lawyer who knows the countries you sell into, especially before large campaigns or anything involving consumers.

APPENDIX C

Glossary

A, B and C tiers

Priority levels set by lead score, each with its own owner, channels and response time.

Bounce

An email that couldn't be delivered. Suppress hard bounces at once; soft bounces can be retried.

CAN-SPAM

The US law for commercial email: honest headers, a postal address and an opt-out honored within 10 business days.

CASL

Canada's anti-spam law. Commercial messages need express or implied consent.

Catch-all domain

A mail server that accepts every address, so verification can't confirm a mailbox exists.

DKIM

A signature proving a domain sent a message and that it wasn't changed in transit.

DMARC

A DNS policy telling receivers what to do with mail that fails SPF or DKIM alignment.

E.164

The international phone format: a plus sign, country code and number, up to 15 digits.

Firmographics

Company facts such as industry, location, size and business model.

Fit and intent scores

Fit measures how well a lead matches your ICP; intent measures how ready it looks right now.

ICP

Ideal customer profile: the observable traits of companies most likely to buy and stay.

JSON-LD

Structured data embedded for search engines; often a business's cleanest details.

Legitimate interests

A GDPR lawful basis that needs a documented balancing test.

MX record

The DNS record naming a domain's mail servers. A null MX means it accepts no email.

PECR

The UK's electronic marketing rules, which treat companies differently from sole traders.

Place ID

Google's unique identifier for a Maps listing, and the best dedup key for local businesses.

Residential proxy

An IP from a real home connection, which websites treat as an ordinary visitor.

robots.txt

A file at a site's root telling crawlers which paths they may visit, standardized as RFC 9309.

SPF

A DNS record listing the servers allowed to send email for a domain.

Suppression list

Every address and domain that must never be contacted again, checked before each send.

Trigger signal

A public event, like funding or a new location, that suggests a company needs you now.

More terms, with longer explanations, in the ProxyHorizon Glossary at proxyhorizon.com/glossary.

APPENDIX D

Free Tools and Further Reading

Every tool below has a public page on ProxyHorizon; most need a free account to run.

Tools used in this book

Google Maps Scraper	Local businesses, phones, sites	**Google Maps Rank Checker**	Where a business ranks
Google Reviews Analyzer	Ratings, topics, reply rate	**Contact Information Finder**	Published emails and phones
Social Media Links Finder	Profiles a site links to	**LinkedIn Company Employees**	Staff by title and team
YouTube Email Finder	Creators' business emails	**Website Technology Checker**	Platform and bot protection
Website Pricing Extractor	Every plan in one table	**Schema Extractor**	All structured data on a page
Google News Checker	Live news for any company	**AI Email Generator**	Cold emails and follow-ups

Directories and use cases

Lead generation	Proxies for lead gen	**Market research**	Proxies for research
Proxy providers	Tested, rated, compared	**Comparison tool**	Any two, side by side

Guides worth reading next

- **Best Residential Proxies**
- **Best Social Media Scraping Tools**
- **Best Data Extraction APIs**
- **n8n Review**
- **n8n Use Cases That Save Time**
- **Proxies for LinkedIn Outreach Tools**
- **Managing Several LinkedIn Accounts**
- **Web Scraping Legality and robots.txt**

Running LinkedIn outreach tools? Read LinkedIn's User Agreement first: it restricts automated activity, and accounts that break it get restricted.

ProxyHorizon

Compare confidently. Implement correctly.

ProxyHorizon is the independent directory and resource hub for proxies, VPNs and antidetect browsers: expert-tested comparisons, free tools and practical guides to choose and use them.

- Compare 120+ proxy providers
- Free web data and SEO tools
- In-depth, fact-checked guides



MORE FREE EBOOKS FROM PROXYHORIZON

- **The Ultimate Proxy Handbook** — Free PDF
- **The Scraping Blueprint** — Free PDF
- **The Antidetect Browser Handbook** — Free PDF
- **Proxies for AI Agents and LLM Data** — Free PDF

© 2026 PROXYHORIZON · LEAD GENERATION WITH PUBLIC WEB DATA