

Premium B2B Multi-Niche Lead Generation Pipeline

An async-first, production-grade Google Maps scraping system delivering 186,662 enriched B2B leads across 12 high-value commercial niches — from corporate law to industrial cleaning — covering all 50 US states and 5,011 cities.

Scope

50 States · 5,011 Cities · 12 Niches

Data Volume

186,662 Records · 27 Enriched Fields

Architecture

Async Playwright + Aiohttp + DNS

Date

August 2026

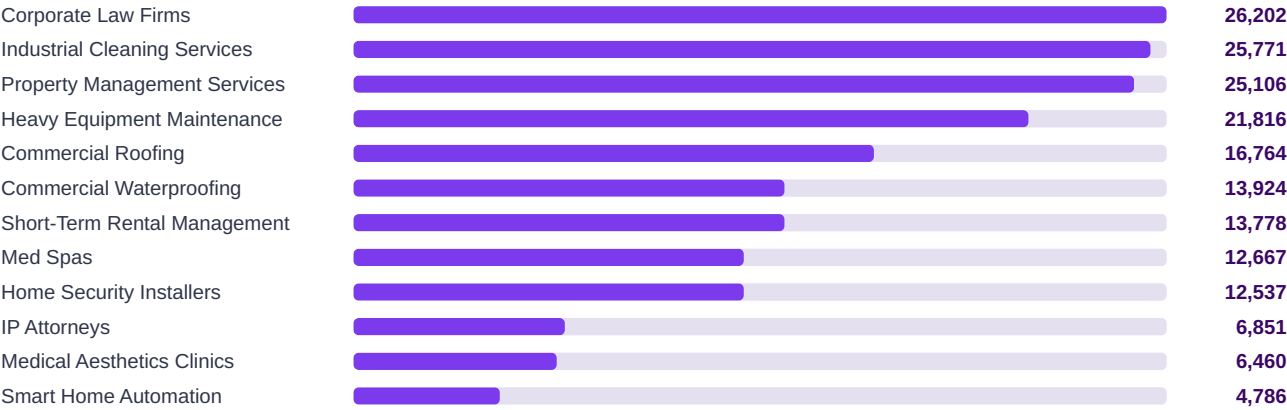
1. Executive Summary

This project delivers an enterprise-grade B2B lead enrichment engine that scrapes, crawls, and validates 12 distinct commercial niches at national scale. Built on an async Python architecture — combining Playwright for browser discovery, aiohttp for high-throughput website crawling, and DNS async resolution for email verification — the system ingests Google Maps search results and produces 27 fully enriched fields per business record.



2. The 12 Target Niches

Unlike single-industry scrapers, this system targets a diverse portfolio of high-value B2B verticals — each selected for lead quality and commercial potential:



3. Technical Architecture

1

Async Discovery

Playwright searches Google Maps; async scrolls sidebar feed to collect place URLs per niche+city.

2

Hybrid Crawl

aiohttp fetches websites at 15-concurrency semaphore; browser fallback for JS-rendered pages.

3

Deep Enrichment

Email extraction + DNS MX verification; social URLs; CMS detection; GA4, FB Pixel, Schema, SSL, DMARC, SPF checks.

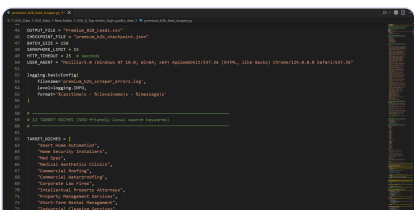
4

Blocklist Filter

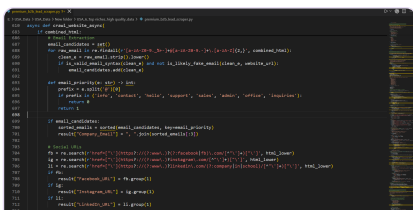
100+ mega-corp blocklist excludes Walmart, Home Depot, Marriott, etc. from results.

Async at scale: Unlike sequential scrapers, this system uses Python asyncio with aiohttp sessions and async DNS resolution. The 15-concurrency semaphore balances throughput with rate-limiting. A 100+ entry mega-corp blocklist filters out national chains. Batch processing (150 queries per batch) prevents memory leaks, with JSON checkpointing for crash recovery.

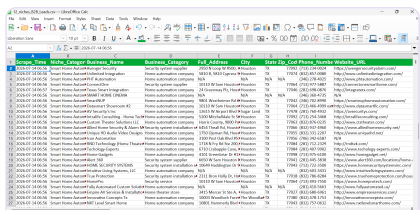
4. Screenshots — Architecture & Data



12 niches & async scraper config



Email + Social URL enrichment logic



Output: 186K+ records, 27 fields

5. Enrichment Depth

48.0%

EMAIL CAPTURED

42.3%

FACEBOOK

28.7%

INSTAGRAM

18.2%

LINKEDIN

Technical Stack Detection

SIGNAL			SIGNAL		
Mobile Responsive	128,052	68.6%	SSL Valid	117,007	62.7%
SPF Configured	99,418	53.3%	GA4 Installed	60,654	32.5%
Booking Widget	53,117	28.5%	DMARC	61,233	32.8%
Local Schema	43,013	23.0%	FB Pixel	16,442	8.8%

CMS Platform Distribution



6. Data Quality & Geographic Coverage

Rating Distribution



Top 8 States

STATE		STATE	
California	25,565	Illinois	7,554
Texas	18,962	New York	6,580
Florida	15,636	Ohio	6,317
N/A (unmapped)	17,809	Michigan	6,205

7. Business Impact

Sales & Outreach

- 89,635 verified business emails across 12 niches
- 42.3% with Facebook — social selling channels
- 18.2% with LinkedIn — executive-level targeting
- 28.5% have booking widgets — identify sales-ready businesses
- 100% GMB-claimed: every lead actively manages its Google presence

Tech Stack Intelligence

- 68.6% mobile-responsive — digital maturity indicator
- 62.7% SSL-valid — security posture signal
- CMS detection pinpoints technology gaps for SaaS/agency sales
- GA4 + FB Pixel presence flags data-driven prospects
- DMARC/SPF data enables email deliverability campaigns

8. Technical Stack

COMPONENT	TECHNOLOGY	ROLE
Browser Automation	async_playwright (Chromium headless)	Google Maps search, scroll, listing extraction
Async HTTP	aiohttp (15-conn semaphore)	High-throughput website crawling
HTML Parsing	BeautifulSoup4	CMS detection, schema extraction, content scan
Email Verification	dnspython (async MX + A fallback)	DNS-level domain verification
Data Pipeline	CSV append + JSON checkpoint	Crash recovery, batch-based output
Blocklist	100+ mega-corp patterns	Filters national chains from niche results

9. Deliverable

The final output — **12_niches_B2B_Leads.xlsx** (27.8 MB, 186,662 records) — provides 27 enriched fields per row, including business email, social URLs, CMS platform, and 8 technical stack signals. Every record links to its source Google Maps entry. The dataset imports directly into any CRM or analytics platform.

