RETAILMYMEDS

Intelligence Hub

Data & Methodology

How 33,185 pharmacies were sourced, verified, scored, and made actionable

PREPARED BY **Matthew Scott**

February 2026 | Louisville, KY

The Data

The RetailMyMeds Intelligence Hub is built on a database of **33,185 independent community pharmacies** across 51 states. Every record traces back to a federal government source. No purchased lists, no third-party data brokers, no estimates on identity or contact information.

Where the Data Comes From

The foundation is the CMS National Plan and Provider Enumeration System (NPPES) — the federal registry of every healthcare provider in the United States. The February 9, 2026 data dissemination file (approximately 9 million records) was downloaded and filtered to taxonomy code 3336C0003X (Community/Retail Pharmacy), Entity Type 2 (organizations only), active status, and US states plus DC.

This initial extraction identified **44,157** entities registered as independent community pharmacies. Chain pharmacies were excluded using 171 regex patterns covering CVS, Walgreens, Walmart, Kroger, Costco, Publix, H-E-B, Rite Aid, Target, all regional chains, grocery store pharmacies, mail-order operations, and hospital-affiliated pharmacies.

Verification

Every NPI was individually queried against the CMS NPI Registry API. Results:

- 41,763 of 41,775 found in registry (99.97%)
- 41,763 confirmed as Active by CMS

- 40,157 with primary Community/Retail Pharmacy taxonomy
- 39,611 after excluding mail-order operations

Address and phone data matched CMS records at 99.98% and 99.97% respectively. Owner information was extracted from the NPPES Authorized Official fields — the person legally responsible for the NPI.

Deduplication

The NPPES registry assigns NPIs to entities, not physical locations. A single pharmacy can have multiple NPIs from ownership changes, reincorporations, or specialty service registrations. Two different legal entities can operate at the same street address. Sending a sales team a list with duplicates erodes trust immediately.

A five-stage pipeline reduced the dataset from 39,611 to **33,185** verified, unique pharmacies — a 16.2% reduction.

Stage 1: Address Deduplication

−4,979 records

Every street address was normalized — stripping suite numbers, unit designators, apartment numbers, and punctuation variations. Records were grouped by normalized address + city + state + ZIP. Any group with multiple NPIs was resolved by keeping one winner per physical location, preferring records with owner information, longer pharmacy names, and older NPIs.

Stage 2: Institutional Removal

−711 records

Keyword matching removed entities that passed initial filters but are institutional: hospitals, VA facilities, FQHCs, health systems, university teaching pharmacies, and correctional facilities. These are not RMM prospects.

Stage 3: Specialty Taxonomy Removal

−464 records

Secondary NPPES taxonomy codes identified entities that are not standard retail: compounding pharmacies, nuclear pharmacies, specialty pharmacies (PBM-affiliated biologics), home infusion, and long-term care.

Stage 4: Chain Slip-Through Removal

−89 records

A cleanup pass caught chain pharmacies that escaped the initial 171-pattern filter through variant naming (e.g., "CVSPHARMACY without a space, Walgreens subsidiaries under local names).

Stage 5: Non-Pharmacy Clinic Removal

−183 records

Medical practices that register pharmacy NPIs for in-house dispensing were removed: clinics, urgent care centers, pediatric practices, dermatology offices, and behavioral health practices with dispensing sidelines.

PIPELINE DESIGN

The stages are ordered by confidence and impact: address dedup first (highest volume, lowest risk), institutional removal second (clearly wrong category), specialty taxonomy third (requires secondary NPPES fields), chain slip-throughs fourth (cleanup), and clinic removal last (most judgment-dependent). Each stage feeds its output into the next.

Enrichment

Starting point: 33,185 pharmacies with NPI, name, owner, address, phone, and state. The goal was to attach market context to every row so each pharmacy could be scored and qualified without manual research. Six federal data sources were merged.

CMS Medicare Part D Spending

State-level GLP-1 spending and claims volume per pharmacy. CMS publishes annual Part D spending by drug category and state. The GLP-1 category (semaglutide, tirzepatide, liraglutide, dulaglutide) shows total spending and claims per state. Dividing by community pharmacies per state gives an estimated per-pharmacy figure. This is the strongest enrichment column — the closest proxy to “how much GLP-1 volume flows through pharmacies in this market.”

Limitation: State average. A pharmacy in rural eastern Kentucky and one in downtown Louisville get the same figure.

CDC PLACES

ZIP-level diabetes prevalence and obesity prevalence. Modeled health estimates at the census tract and ZIP code level. For each pharmacy's ZIP:

- `zip_diabetes_pct` — % of adults with diagnosed diabetes
- `zip_obesity_pct` — % of adults with BMI ≥ 30

These are demand signals. GLP-1s are prescribed for diabetes management and weight loss. A pharmacy in a ZIP with 18% diabetes and 42% obesity is in a high-demand market whether or not they realize it.

Census American Community Survey

ZIP-level demographics — age distribution, median household income, and population.

- `zip_pct_65_plus` — Medicare eligibility proxy. More seniors = more Part D volume = more GLP-1 reimbursement exposure.
- `zip_median_income` — Lower income areas correlate with tighter margins and higher Medicaid mix.
- `zip_population` — Smaller markets mean fewer competitors and more vulnerability per lost fill.

HRSA Health Professional Shortage Areas

Whether the pharmacy is in a federally designated HPSA, and the HPSA severity score (0–25). **91.8%** of the 33,185 pharmacies are in HPSA-designated areas — often the only pharmacy option in their community.

USDA Rural-Urban Continuum Codes

Two-step lookup: Census ZCTA-to-County crosswalk maps each ZIP to its dominant county, then USDA 2023 RUCC codes classify each county as Metro (codes 1–3), Rural-Adjacent (4, 6, 8), or Rural-Remote (5, 7, 9). Output: county FIPS, county name, RUCC code, rural classification.

NADAC

National Average Drug Acquisition Cost from CMS. The reference point for the $37–39.50/fill loss estimate on GLP-1s — the gap between pharmacy acquisition cost and PBM reimbursement after DIR fee clawbacks. This is the basis for the annual loss calculation in every pharmacy record.

DATA INTEGRITY

All joins are left joins. If a ZIP has no CDC data (rare), the pharmacy gets null values that default to the 50th percentile in scoring — neither helped nor hurt. Every enrichment column traces to a free, public federal dataset.

Scoring Model

The scoring model answers one question: *which pharmacies would benefit most from RMM's software, and which would be most receptive to a conversation about it?*

Two scoring systems serve different purposes.

Intel Hub Scoring (7-Factor Model)

Used for the 33,185-pharmacy database. Ranks every pharmacy relative to the population using percentile-based weighted scoring.

Factor	Weight	Direction
State GLP-1 cost per pharmacy	25%	Higher = better
ZIP diabetes prevalence	20%	Higher = better
ZIP % age 65+	15%	Higher = better
ZIP obesity prevalence	10%	Higher = better
HPSA designation	10%	Binary (yes=100, no=0)
ZIP median income	10%	*Inverted*
ZIP population	10%	*Inverted*

Each factor is converted to a percentile rank (0–100) using `pandas.rank(pct=True)`. Inverted factors score lower values higher (lower income = tighter margins = more receptive; smaller population = fewer competitors). The weighted sum produces the composite RMM Score.

Grade cutoffs are cumulative, not fixed thresholds:

- **Grade A** (top 15%): 4,978 pharmacies — Immediate Outreach

- **Grade B** (next 25%): 8,296 pharmacies — High Priority
- **Grade C** (next 30%): 9,956 pharmacies — Standard
- **Grade D** (remaining 30%): 9,955 pharmacies — Monitor

Because cutoffs are percentile-based, grades redistribute automatically if the dataset changes.

Full Scorecard (3-Dimension Model)

Used for individual pharmacy qualification. Takes 12 inputs and scores against absolute thresholds, not the population.

Dimension 1: Financial Fit (45% weight)

Monthly Rx volume, GLP-1 fills, estimated monthly loss, and government payer mix. Brackets volume and fills into score bands (e.g., 500+ GLP-1 fills = 100, under 100 = 20). The heaviest dimension because it determines whether the ROI math works for this specific pharmacy.

Dimension 2: Operational Readiness (30% weight)

PMS system (PioneerRx scores highest due to clean API integration), staffing capacity, owner engagement, willingness to train a tech on routing, and prior routing experience. This dimension measures how quickly the pharmacy can start using the software.

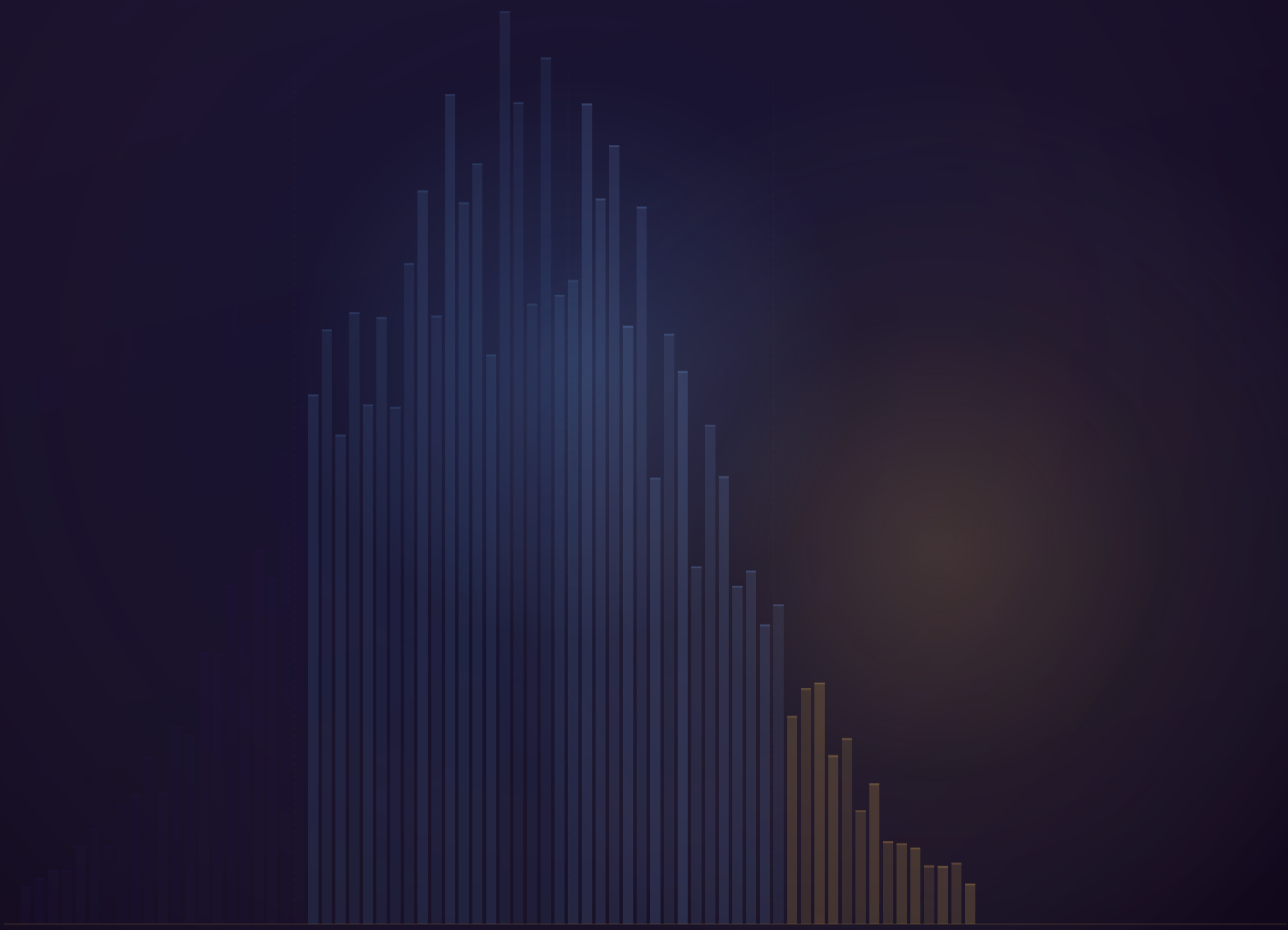
Dimension 3: Market Urgency (25% weight)

MFP drug exposure, DIR fee pressure, closure risk, competitive pressure, and problem awareness. This dimension measures how receptive the pharmacy is to the conversation right now.

Grade mapping: A (80–100), B (70–79), C (60–69), D (50–59), F (below 50). These are absolute thresholds — a pharmacy's grade reflects its own qualification, not its rank against others.

TWO MODELS, TWO PURPOSES

The 7-factor model uses publicly available data to rank 33,185 pharmacies. The 3-dimension model uses pharmacy-specific inputs to qualify individual prospects. One finds them. The other qualifies them.



The Tools

Three tools turn the data into action. Each serves a different user at a different stage of the sales process.

Intelligence Hub

`rmm-intel-hub.onrender.com`

Audience: Sales team (internal)

Purpose: Find and qualify pharmacy prospects from the 33,185-pharmacy database.

The dashboard shows the national grade distribution, ranks states by Grade A concentration (percentage of pharmacies that scored highest), and provides search across NPI, pharmacy name, owner name, city, and ZIP. Clicking a pharmacy generates a full intelligence report with demographics, GLP-1 exposure, estimated losses, and auto-generated outreach talking points.

State drill-down shows the top 50 pharmacies by score, filterable by grade. Any view can be exported as a CSV for direct import into Mailchimp, HubSpot, Salesforce, or any CRM.

Architecture: Python Flask serving a single HTML file. CSV loaded into memory at startup. No database, no build step, two dependencies (`flask`, `gunicorn`). Search runs in milliseconds against in-memory indexes.

Pharmacy Forecasting Tool

`rmm-pharmacy-tool.vercel.app`

Audience: Pharmacy owners (external)

Purpose: Show a pharmacy what they're losing on GLP-1s and what RMM would save them.

Four fields: pharmacy name, state, monthly Rx volume range, GLP-1 fills range. Returns estimated monthly and annual loss, savings at 5% routing, breakeven fills, and ROI verdict (Strong / Moderate / Marginal).

This is a conversation starter. A pharmacy owner sees their own numbers in 10 seconds. It does not touch the 33,185-pharmacy database — it is pure formula from user inputs.

Architecture: Static HTML on Vercel. Calls the Texume API on Render for computation. No framework, no build step.

Full Scorecard PDF

`POST texume-api.onrender.com/scorecard`

Audience: Qualified prospects

Purpose: Generate a branded, print-quality PDF scorecard for a specific pharmacy after qualification.

Takes 12 inputs (the full 3-dimension model), runs all three scoring dimensions, calculates ROI scenarios at 5%, 10%, and 15% routing, and compiles a multi-page PDF via LaTeX. The PDF includes dimension breakdowns, bar charts, financial projections, and a market context narrative.

Architecture: FastAPI endpoint. Jinja2 renders a LaTeX template with the pharmacy's data. An external LaTeX compilation service returns a PDF. Optionally sends the PDF via email (infrastructure ready, activation pending).

Data Integrity

Trust is earned through transparency about what the data does and does not claim.

What the Data Does

- Identifies every registered independent community pharmacy in the US with CMS-verified identity and contact information
- Attaches market context (disease burden, demographics, shortage status, rural classification) from six federal sources
- Ranks pharmacies relative to each other on factors that correlate with GLP-1 loss exposure and receptivity to routing solutions
- Provides outreach teams with qualified, prioritized lists segmented by state and grade

What the Data Does Not Claim

- No individual pharmacy financial data. CMS does not publish per-pharmacy fill counts, revenue, or payer mix. All financial estimates are derived from state averages and NCPA survey benchmarks.
- The $37/fill loss figure is an NCPA survey benchmark, not a per-pharmacy calculation.
- GLP-1 fills per pharmacy are state averages from CMS Part D, allocated evenly. Not adjusted for pharmacy size, location, or payer mix.
- The model ranks pharmacies relative to each other. It does not predict whether any specific pharmacy will convert to an RMM customer.

- No scraped data. No Google reviews, Yelp, or social media signals.

The NCPA Gap

NCPA reports approximately 18,960 independent pharmacies. This database contains 33,185. The 1.75x gap is definitional:

- NCPA counts self-identified members and survey respondents (storefronts)
- NPPES counts every registered entity with a community/retail pharmacy taxonomy
- Some NPPES entities are real pharmacies that do not participate in NCPA surveys
- Some may have low volume or niche operations

The dedup pipeline removes obvious non-independents but does not artificially shrink the count. The scoring model handles prioritization — pharmacies with weak market conditions score low regardless of their inclusion in the dataset.

> REPRODUCIBILITY
> Every data point traces back to a government source. The pipeline scripts are deterministic — same input produces same output. NPI verification can be re-run in under one hour. Part D data updates annually, demographics update yearly. The database is a living asset, not a static snapshot.

Key Numbers

Metric	Value
Total NPIs queried	41,775
CMS-confirmed active	39,611
After 5-stage deduplication	33,185
Grade A (Immediate Outreach)	4,978 (15%)
Grade B (High Priority)	8,296 (25%)
Grade C (Standard)	9,956 (30%)
Grade D (Monitor)	9,955 (30%)
HPSA designation rate	91.8%
Federal data sources	6
Data columns per pharmacy	24
State outreach lists	51

This document describes how the Intelligence Hub was built, why each design decision was made, and what the data can and cannot tell you. The methodology is transparent because the data is only as trustworthy as the process that produced it.

Every pharmacy in this database is a real business with a real owner making real decisions about profitability. The tools built on this data exist to help those pharmacies find the help they need — and to help RetailMyMeds find them first.

ABOUT THIS DOCUMENT

This document describes the data sourcing, verification, enrichment, scoring, and tooling methodology behind the RetailMyMeds Intelligence Hub. All data traces to federal government sources (CMS, CDC, Census Bureau, HRSA, USDA). The pipeline is fully reproducible.

Author
Matthew Scott

Prepared For
Kevin McCarron & Arica Collins

Date
February 2026

Location
Louisville, KY