

Mapping Real Estate Visuals Across America

🔍 Goal:

Collect accurate pricing data for **real estate photography and videography services** for all **35,705+ municipalities across the United States**, from **credible and market-specific sources**, to power a national pricing database.

□ What Exactly Needs to Be Scraped

We are targeting **business listings and websites** of service providers who **specialize in real estate media**, including the following services:

🎯 Mandatory Services to Extract Pricing For:

1. Photography
2. Videography
3. Drone Photography
4. Drone Video
5. 3D Virtual Tour
6. Floor Plans
7. Virtual Staging
8. Twilight Photography
9. Agent Intro/Outro
10. Voiceover

✂ Data Fields to Scrape (Per City)

For each service provider, extract:

- **Service Name** (must match one of the 10 required)
- **Price** (as a numeric value only)
- **City & State** (to geo-tag the result)
- **Website or Source URL** (for traceability during QA)

□ Smart Scraping Features

To ensure relevance and data quality, scrapers will include:

- ✓ **Keyword filters** (e.g. must contain "real estate", "listing", "property", etc.)
- ✓ **Geo-filtering** (scrape only for one city/state at a time to maintain mapping)
- ✓ **Confidence scoring** (optional: flag scraped prices based on source confidence or presence of direct mention)
- ✓ **Duplicate detection** (avoid scraping the same business from multiple platforms)

Summary Description

The scraper will programmatically gather **real estate media service pricing** from **verified local business websites, directories, and Google listings**, across **all 35,705 US municipalities**. The data will be extracted for exactly **10 predefined service categories** and compiled into a **structured JSON hierarchy** with price values and interpolation flags. Irrelevant service markets (e.g. weddings, events) will be automatically filtered out using smart keywords and domain logic to ensure **data accuracy, market relevance, and coverage completeness**.

City Database: <https://simplemaps.com/data/us-cities> [1]

Questions:

Q1. Should we use a data source (website) of our own choice, or must we strictly follow the one provided by the user?

Q2. Is an API key mandatory for scraping data from each individual website, or are there exceptions?

Q3. According to the U.S. Census, there are 31,254 locations in the United States [1], as verified by SimpleMaps. Should we consider all of these locations in our data coverage?

Q4. If the data values are identical across all required fields for a particular location or service, how should we handle this scenario?

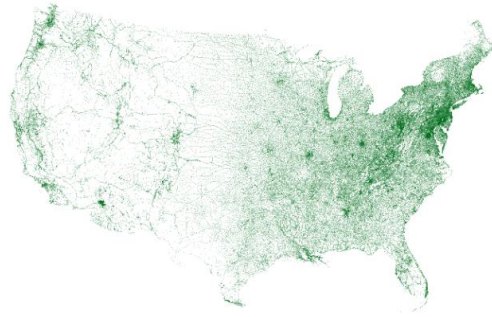


Figure 1: Here is a visualization of the data in our paid databases for the continental United States [1].

Reference:

[1] SimpleMaps, “United States Cities Database,” SimpleMaps.com, updated June 9, 2025. [Online]. Available: <https://simplemaps.com/data/us-cities>