

Project Work:

AI Agent – Lead Scraping for Development Opportunities

Objective

Develop an automated AI Agent to identify and collect potential single-family residential properties in Newton, MA that represent development or teardown opportunities.

Core Goals

- Scrape public real estate listings (MLS, Zillow, Redfin, Realtor, etc.) for active, pending, and recently sold single-family homes.
- Detect potential development leads such as large lots, older homes, or underbuilt parcels.
- Output structured data ready for analysis or CRM import.

Data Sources

- Zillow, Redfin, Realtor.com
- MLS feed (if accessible via API)
- City of Newton property database (parcel data, zoning, land size)
- Public records (Assessors database, GIS)

Data Points to Collect

Category	Example Fields
Property Info	Address, MLS ID, Listing URL, Asking Price, DOM, Status
Lot & Zoning	Lot Size (sqft), Zoning Code, FAR, Buildable Area, Frontage
Structure Info	Living Area, Year Built, Bedrooms, Condition keywords (e.g., 'needs work', 'as is')
Market Indicators	Price/SF, Assessed Value, Days on Market trend, Price Change
Ownership	Owner Name (if public), Last Sale Date, Last Sale Price
Listing Notes (NLP)	Detect phrases like 'tear down', 'builder', 'contractor special', 'development opportunity'
Comparable Potential	Identify underbuilt lots vs allowed FAR or zoning maximums

Functional Requirements

- Run daily or weekly automated scans.
- Filter by Newton, MA (optionally expandable to nearby towns).
- Rank properties by development potential score (based on lot size, zoning, age, and price).
- Export structured CSV or push to Zoho CRM / internal Google Sheet.
- Allow manual trigger refresh and review log of collected leads.

Technical Components

- Web scraping (BeautifulSoup / Scrapy / Playwright / API if available).
- NLP filtering to detect 'development-related' listing language.
- Geospatial logic for zoning/lots (using Newton GIS or public shapefile).
- Optional: integration with OpenAI API or custom prompt for text classification ('Is this a development opportunity?').

Output Format

- Excel / CSV file with all fields above.
- Optional: JSON feed for integration.
- Include timestamp, data source, and URL for traceability.

Optional Future Add-Ons

- Add map visualization (Google Maps or Leaflet).
- Include automated email/Slack alerts for new opportunities.
- Score leads by ROI potential or estimated buildable SF.