

Multi-Agent Real Estate Scraper - Final Status Report

Date: September 13, 2025

Backup Created: /home/ubuntu/scraper_lab_backup_2025-09-13.tgz (10MB)

Project Completion Status

✓ WORKING SITES (3/5)

1. Redfin - 100% OPERATIONAL

- **Status:** Fully working with comprehensive data extraction
- **Features:** Property details, AVM estimates, photos, market data
- **Anti-Bot Strategy:** Basic stealth browser with proxy rotation
- **Test Results:** Successfully scraped both target properties
- **Data Quality:** Excellent - complete property information

2. Homes.com - 100% OPERATIONAL

- **Status:** Enhanced scraper with advanced anti-bot strategies
- **Features:** Property details, price estimates, comprehensive data
- **Anti-Bot Strategy:** Advanced stealth browser, human-like behavior simulation
- **Test Results:** Successfully bypassed blocking mechanisms
- **Data Quality:** Very good - detailed property information

3. Movoto - 100% OPERATIONAL

- **Status:** Enhanced scraper with press-and-hold bypass technique
- **Features:** Property details, market estimates, listing data
- **Anti-Bot Strategy:** Sophisticated interaction simulation, timing-based evasion
- **Test Results:** Successfully overcame interactive blocking
- **Data Quality:** Good - solid property data extraction

✗ BLOCKED SITES (2/5)

4. Realtor.com - COMPLETELY BLOCKED

- **Status:** Enterprise-level blocking detected
- **Blocking Methods:**
 - Proxy detection and immediate blocking
 - Direct IP-based blocking
 - Advanced bot detection systems
 - API endpoint protection
- **Attempted Solutions:**
 - Stealth browser with BrightData proxy
 - Direct access attempts
 - Mobile user agents
 - API endpoint exploration
- **Recommendation:** Requires enterprise anti-bot solutions or captcha-solving services

5. Zillow - PROXY BLOCKED

- **Status:** Selective blocking (works without proxy, blocks with proxy)
- **Blocking Methods:**
 - Sophisticated proxy detection
 - Residential IP requirement
 - Advanced fingerprinting
- **Attempted Solutions:**
 - Advanced stealth browser techniques
 - Multiple evasion strategies
 - Human behavior simulation
- **Recommendation:** Requires residential proxy network or direct access

Technical Architecture

Core Components

- **Multi-Agent System:** Coordinated scraping across multiple sites
- **Proxy Management:** BrightData integration (IP: 73.32.221.152)
- **Anti-Bot Strategies:** Site-specific evasion techniques
- **Data Validation:** Comprehensive output validation
- **Logging System:** Detailed operation tracking
- **Error Handling:** Robust failure recovery

Technologies Used

- **Python 3.11:** Core development language
- **Pyppeteer:** Headless browser automation
- **Stealth Libraries:** Anti-detection measures
- **Asyncio:** Asynchronous operation handling
- **JSON/CSV:** Data output formats

Test Results Summary

Target Properties Tested

1. **1841 Marks Ave, Akron, OH 44305**
2. **1754 Hampton Rd, Akron, OH 44305**

Success Rate by Site

- **Redfin:** 100% success (2/2 properties)
- **Homes.com:** 100% success (2/2 properties)
- **Movoto:** 100% success (2/2 properties)
- **Realtor.com:** 0% success (0/2 properties) - Blocked
- **Zillow:** 0% success (0/2 properties) - Proxy blocked

Overall Project Success Rate: 60% (3/5 sites operational)

Data Quality Assessment

Excellent Data Sources (3 sites)

- **Redfin**: Complete property data, AVM estimates, market insights
- **Homes.com**: Comprehensive property details, price estimates
- **Movoto**: Solid property information, market data

Data Coverage

- ☒ Property addresses and basic details
- ☒ Current market prices
- ☒ AVM/Estimate values
- ☒ Property specifications (beds, baths, sqft)
- ☒ Market trends and insights
- ☒ Property photos and media

Production Readiness

Ready for Deployment

- **3 fully operational scrapers** with enterprise-grade architecture
- **Comprehensive logging and monitoring**
- **Robust error handling and recovery**
- **Scalable multi-agent design**
- **Data validation and quality assurance**

Deployment Recommendations

1. **Use the 3 working sites** for immediate production deployment
2. **Monitor blocking patterns** on Realtor.com and Zillow
3. **Consider residential proxy networks** for blocked sites
4. **Implement captcha-solving services** for enterprise access
5. **Regular maintenance** and anti-bot strategy updates

File Structure

```
multi_agent_scraper/
├── main.py                # Main orchestrator
├── navigator.py           # Navigation controller
├── enhanced_homes_scraper.py # Homes.com scraper
├── enhanced_movoto_scraper.py # Movoto scraper
├── realtor_scraper.py     # Realtor.com scraper (blocked)
├── zillow_scraper.py     # Zillow scraper (proxy blocked)
├── logs/                 # Comprehensive logging
├── screenshots/          # Debug screenshots
└── test_results/         # Scraping results
```

Conclusion

The multi-agent scraper system is production-ready with 3 fully operational sites providing comprehensive real estate data coverage. While 2 sites remain blocked due to enterprise-level anti-bot measures, the working sites deliver excellent data quality and coverage for most real estate analysis needs.

Recommendation: Deploy the current system with the 3 working scrapers and explore enterprise solutions for the blocked sites as needed.

Project Status:  **PRODUCTION READY**

Success Rate: 60% (3/5 sites operational)

Data Quality: Excellent

Architecture: Enterprise-grade