Target URL Overwrite Fix - Summary

Overview

This fix resolves the critical issue where target URLs were being overwritten with "View" in the CSV parser, causing the loss of actual Rightmove URLs in the output.

Problem Analysis

Issue Description

- **Input**: Target property with full Rightmove URL in "URL" column
- **Output**: Target property with "View" instead of the URL
- **Impact**: Loss of valuable property listing URLs in the output CSV

Root Cause

The issue occurred in the CSV parser's `normalizeData()` function:

1. **Header Mapping Conflict**: Both "URL" and "Link" columns were being mapped to the same "URL" property because "link" is listed as a variation of "URL" in `HEADER_VARIATIONS`

2. **Overwrite Behavior**: When processing each row:
   - Column 14 ("URL") → sets `normalizedRow.URL` = actual Rightmove URL
   - Column 15 ("Link") → **overwrites** `normalizedRow.URL` = "View"

3. **Result**: The second assignment from the "Link" column overwrites the first assignment from the "URL" column

Why PR #14 Didn't Work

PR #14 attempted to fix this by capturing URLs early and restoring them later. However, the URLs were already lost **during CSV parsing**, before the early capture logic could run. The restoration logic couldn't restore what was never captured in the first place.

Solution

Implementation

Modified `src/utils/csvParser.js` in the `normalizeData()` function to add URL preservation logic:

```
// CRITICAL FIX: Don't overwrite a valid URL with a non-URL value
// This happens when both "URL" and "Link" columns exist and both map to "URL"
// The "Link" column typically contains "View" or hyperlink formulas
if (standardHeader === 'URL' && normalizedRow.URL && isURL(normalizedRow.URL)) {
    // URL already set with a valid URL, don't overwrite with non-URL value
    log.info(` Preserving existing URL in row, not overwriting with: ${stringValue}`)
;
} else {
    normalizedRow[standardHeader] = stringValue;
}
```

How It Works

1. When processing each cell in a row, check if the standard header is "URL"
2. If a valid URL already exists in `normalizedRow.URL`, preserve it
3. Don't overwrite with non-URL values like "View" from the "Link" column

Benefits

- **Simple**: Minimal code change, easy to understand and maintain
- **Targeted**: Only affects URL column preservation, doesn't change other behavior
- **Safe**: Uses existing `isURL()` validation to ensure we're preserving actual URLs
- **Comprehensive**: Fixes the issue for all rows, not just the target

Testing

Test Files

1. **test-target-url-debug.js**: Basic debugging test to trace URL capture
2. **test-target-url-fix-comprehensive.js**: Full end-to-end validation test

Test Results

- ✅ CSV parsing preserves URLs correctly
- ✅ Target URL maintained:
  `https://www.rightmove.co.uk/properties/160516301#/?channel=RES_BUY`
- ✅ All 79 rows tested successfully
- ✅ No "View" overwrites of actual URLs

Test Input/Output

Input (`data (5).csv`):

```
Row 4: "317 Wharf Road, Ealand", URL="https://www.rightmove.co.uk/properties/
160516301#/?channel=RES_BUY", Link="View", isTarget=1
```

Expected Output:

```
Target URL: https://www.rightmove.co.uk/properties/160516301#/?channel=RES_BUY
```

Result: ✅ PASSED

Code Changes

Modified Files

- `src/utils/csvParser.js`: Added URL preservation logic (lines 361-370)

New Test Files

- `test-target-url-debug.js`: Debug test showing URL capture flow
- `test-target-url-fix-comprehensive.js`: Comprehensive validation test

Pull Request

Branch: `fix/target-url-only`
Commit: `cf19757`
PR URL: https://github.com/CliveCaseley/soldcomp-analyser2/pull/new/fix/target-url-only

PR Description

Title: Fix target URL overwrite issue in CSV parser

This PR fixes the critical bug where target URLs were being overwritten with "View" due to the "Link" column overwriting the "URL" column during CSV parsing.

Changes:
- Added URL preservation logic in `normalizeData()` function
- Prevents non-URL values from overwriting valid URLs
- Tested with real-world data (79 rows)

Impact:
- Target URLs now preserved correctly in output
- No more "View" replacing actual Rightmove URLs
- Fixes issue for all rows, not just target

Verification Steps

To verify the fix works:

```
# Run the comprehensive test
cd /home/ubuntu/github_repos/soldcomp-analyser2
node test-target-url-fix-comprehensive.js

# Expected output:
# ✅ TEST PASSED: Target URL is correctly preserved!
```

Next Steps

1. Review and merge the PR

2. Test with actual Apify actor run

3. Verify output CSV contains correct URLs

4. Close the issue once verified

Date: December 9, 2025
Branch: fix/target-url-only
Status: Ready for review