Push Instructions: Enhancements A-D

Date: December 5, 2025
Branch: `feat/enhancements-a-to-d`
Status: ✅ Successfully Pushed to GitHub

Summary

All 4 requested enhancements have been successfully implemented, tested, committed, and pushed to GitHub.

Enhancements Implemented

✅ **Enhancement A:** Fetch EPC certificate for target property (currently missing)
✅ **Enhancement B:** Add hyperlinked URL columns for all URLs
✅ **Enhancement C:** Geocode missing lat/long and calculate all distances from target
✅ **Enhancement D:** Scrape floor area from EPC certificate pages where Sq. ft is missing

Git Information

Repository: https://github.com/CliveCaseley/soldcomp-analyser2
Branch: `feat/enhancements-a-to-d`
Commit: `d4a0eae`
Remote: Successfully pushed to origin

Commit Message

```
feat: Implement enhancements A-D

Enhancement A: Fetch EPC certificate for target property
- Added explicit EPC enrichment step for target in main.js
- Target now receives EPC rating and certificate URL

Enhancement B: Add hyperlinked URL columns for all URLs
- Enhanced excelHelper.js to add EPC Certificate Link column
- Provides clickable 'View EPC' links in Excel output

Enhancement C: Geocode missing lat/long and calculate distances
- Modified geocodeAndCalculateDistances to skip geocoding if coordinates exist
- Always calculate distances when coordinates are available
- Reduces unnecessary API calls

Enhancement D: Scrape floor area from EPC certificates
- Added scrapeFloorAreaFromCertificate() function to epcHandler.js
- Multiple parsing methods for robust floor area extraction
- Integrates with enrichWithEPCData for automatic floor area population
- Converts sqm to sq ft automatically

Test Results:
- All enhancements tested with test-enhancements-a-to-d.js
- Floor area scraping verified (59 sqm / 635 sq ft)
- Hyperlink generation validated
- Distance calculation verified

Files Modified:
- src/main.js
- src/utils/excelHelper.js
- src/utils/epcHandler.js
- .gitignore (added)
- test-enhancements-a-to-d.js (new)
- ENHANCEMENTS_A_TO_D_SUMMARY.md (new)
```

Files Changed

Modified Files (3)

1. **src/main.js**
   - Added target property EPC enrichment
   - Enhanced geocoding to handle existing coordinates
   - Added floor area fallback scraping

2. **src/utils/excelHelper.js**
   - Added EPC Certificate Link generation

3. **src/utils/epcHandler.js**
   - Added scrapeFloorAreaFromCertificate() function
   - Integrated floor area scraping into scrapeEPCData()

New Files (4)

1. **.gitignore** - Standard Node.js gitignore (prevents node_modules from being committed)

2. **test-enhancements-a-to-d.js** - Comprehensive test suite
3. **ENHANCEMENTS_A_TO_D_SUMMARY.md** - Detailed implementation documentation
4. **ENHANCEMENTS_A_TO_D_SUMMARY.pdf** - PDF version of summary

---

Create Pull Request

To merge these changes into the master branch, create a pull request:

Pull Request URL:

https://github.com/CliveCaseley/soldcomp-analyser2/pull/new/feat/enhancements-a-to-d

Suggested PR Title

```
feat: Implement enhancements A-D (Target EPC, Hyperlinks, Geocoding, Floor Area)
```

Suggested PR Description

```
## Summary
Implements all 4 requested enhancements to improve data completeness and user
experience.

## Changes

### Enhancement A: Target Property EPC Certificate
- Target property now receives EPC rating and certificate URL
- Previously only comparables were enriched with EPC data

### Enhancement B: Hyperlinked URL Columns
- Added "EPC Certificate Link" column with clickable Excel formulas
- All URL types now have corresponding hyperlink columns

### Enhancement C: Smart Geocoding
- Reuses existing coordinates when available (reduces API calls)
- Always calculates distances consistently
- Handles both missing and existing coordinate scenarios

### Enhancement D: Floor Area Scraping
- Scrapes floor area from EPC certificate pages when missing
- Multiple parsing methods for high success rate
- Automatic sqm to sq ft conversion

## Testing
- ✅ All enhancements tested with test-enhancements-a-to-d.js
- ✅ Floor area scraping verified (59 sqm / 635 sq ft extracted)
- ✅ Hyperlink formulas validated
- ✅ Distance calculations confirmed

## Documentation
- Comprehensive implementation guide in ENHANCEMENTS_A_TO_D_SUMMARY.md
- Test suite included for validation

## Impact
- Improved data completeness for target properties
- Better user experience with clickable links
- More efficient API usage
- Higher floor area coverage

## Files Changed
- src/main.js
- src/utils/excelHelper.js
- src/utils/epcHandler.js
- .gitignore (new)
- test-enhancements-a-to-d.js (new)
- ENHANCEMENTS_A_TO_D_SUMMARY.md (new)
```

Verification Commands

To verify the changes locally:

```
# Switch to the branch
git checkout feat/enhancements-a-to-d

# Check commit log
git log -1 --stat

# Run tests
node test-enhancements-a-to-d.js

# View implementation summary
cat ENHANCEMENTS_A_TO_D_SUMMARY.md
```

Next Steps

1. **Create Pull Request** - Use the URL above to create a PR on GitHub
2. **Review Changes** - Review the implementation summary and test results
3. **Merge** - Merge the PR into master branch when ready
4. **Deploy** - Deploy the updated actor to Apify platform

Technical Notes

Dependencies

No new dependencies were added. All enhancements use existing packages:
- axios (HTTP requests)
- cheerio (HTML parsing)
- apify (logging and actor framework)

Backward Compatibility

All changes are backward compatible:
- Existing functionality is preserved
- New features add columns but don't remove or break existing ones
- Error handling ensures graceful fallbacks

Performance Impact

- **Positive:** Reduced geocoding API calls by reusing existing coordinates
- **Neutral:** Floor area scraping adds ~1-2 seconds per property (only when needed)
- **Optimization:** Caching prevents duplicate operations

Support

For questions or issues with these enhancements:
1. Review ENHANCEMENTS_A_TO_D_SUMMARY.md for detailed documentation
2. Run test-enhancements-a-to-d.js to verify functionality
3. Check commit history: `git log feat/enhancements-a-to-d`

Status: ✅ Complete and Ready for Review