Push Instructions: Batch 2 - Preserve Manual Edits

Branch Information

- **Branch Name**: `fix/batch-2-preserve-manual-edits`
- **Base Branch**: `master`
- **Commit Hash**: `8eecf8f`
- **Date**: December 5, 2024

Changes Summary

Files Created

1. `src/utils/manualEditDetector.js` - Manual edit detection utility
2. `test-batch-2-manual-edits.js` - Comprehensive test suite
3. `BATCH_2_MANUAL_EDITS_SUMMARY.md` - Implementation documentation
4. `BATCH_2_MANUAL_EDITS_SUMMARY.pdf` - PDF version of documentation

Files Modified

1. `src/main.js` - Integrated manual edit detection into enrichment workflow

Lines Changed

- **Added**: ~400 lines (utility + tests + docs)
- **Modified**: ~30 lines in main.js

Current Branch Status

```
cd /home/ubuntu/github_repos/soldcomp-analyser2
git branch
# Output: * fix/batch-2-preserve-manual-edits
```

Commit Details

```
git log --oneline -1
# Output: 8eecf8f feat: Batch 2 - Preserve manual edits during iterative processing
```

Test Verification

```
cd /home/ubuntu/github_repos/soldcomp-analyser2
node test-batch-2-manual-edits.js
```

Expected Output

```
=== BATCH 2 MANUAL EDIT PRESERVATION TEST ===

✓ TEST 1 PASSED - Detect processed properties
✓ TEST 2 PASSED - Manual EPC URL edit detection
✓ TEST 3 PASSED - Manual square footage edit detection
✓ TEST 4 PASSED - Small difference (not manual edit)
✓ TEST 5 PASSED - End-to-end enrichment simulation

ALL TESTS PASSED!
```

Push to GitHub

Step 1: Verify Current State

```
cd /home/ubuntu/github_repos/soldcomp-analyser2
git status
git log --oneline -3
```

Step 2: Push Branch to Remote

```
git push origin fix/batch-2-preserve-manual-edits
```

Expected output:

```
Enumerating objects: X, done.
Counting objects: 100% (X/X), done.
Delta compression using up to Y threads
Compressing objects: 100% (X/X), done.
Writing objects: 100% (X/X), Z KiB | Z MiB/s, done.
Total X (delta Y), reused 0 (delta 0)
remote: Resolving deltas: 100% (Y/Y), completed with Z local objects.
remote:
remote: Create a pull request for 'fix/batch-2-preserve-manual-edits' on GitHub by vis
iting:
remote:      https://github.com/CliveCaseley/soldcomp-analyser2/pull/new/fix/batch-2-
preserve-manual-edits
remote:
To https://github.com/CliveCaseley/soldcomp-analyser2.git
 * [new branch]      fix/batch-2-preserve-manual-edits -> fix/batch-2-preserve-manual-
edits
```

Step 3: Create Pull Request

Visit the URL provided in the output above, or go directly to:

```
https://github.com/CliveCaseley/soldcomp-analyser2/pull/new/fix/batch-2-preserve-
manual-edits
```

Pull Request Details

Title

```
Batch 2: Preserve manual edits during iterative processing
```

Description

Summary

Implements manual edit detection and preservation system to support iterative processing without overwriting user corrections.

Problem

User manually corrected EPC URL in output CSV (from "1 Fernbank Close" to "7 Fernbank Close"), then ran corrected CSV back through actor as data.csv. The actor overwrote the manual correction.

Spec requirement: "Support iterative processing without overwriting manual edits"

Solution

Architecture
1. **Detection Phase**: Identifies processed properties (has EPC, coordinates, calculated fields)
2. **Comparison Phase**: Compares existing vs scraped data to detect manual corrections
3. **Marking Phase**: Flags manually edited fields with internal `_manual_edit_flags`
4. **Preservation Phase**: Skips automated updates for manually edited fields

Protected Fields
- EPC Certificate (exact URL comparison)
- Sq. ft (>5% difference threshold)
- Sqm (>5% difference threshold)
- Price (any change)
- Address, Postcode, Type, Tenure, Bedrooms, Date of sale

Implementation

New Module: `src/utils/manualEditDetector.js`

Key functions:
- `hasBeenProcessed()` - Detects if property was processed before
- `detectManualEdits()` - Scans for manual edits
- `compareAndMarkEPCEdit()` - Detects EPC URL corrections
- `compareAndMarkSqftEdit()` - Detects sqft edits (5% threshold)
- `canUpdateField()` - Checks if field can be updated

Modified: `src/main.js`

- Added Step 8.5: Detect manual edits (before enrichment)
- Modified `enrichWithEPCData()`: Check manual edit flags before updating
- Modified `finalizePropertyData()`: Respect manual edits on calculated fields

Testing

Test Suite: `test-batch-2-manual-edits.js`

✅ Test 1: Detect processed properties
✅ Test 2: Manual EPC URL edit detection
✅ Test 3: Manual square footage edit detection
✅ Test 4: Small difference (not manual edit)
✅ Test 5: End-to-end enrichment simulation

All tests passed

Usage Example

1. Initial run: data.csv → output (44).csv
2. User manually corrects EPC URL in output (44).csv
3. Re-run: output (44).csv → data.csv → actor processes
4. Result: Manual EPC correction preserved ✓

```
## Files Changed

### Created
- `src/utils/manualEditDetector.js` (370 lines)
- `test-batch-2-manual-edits.js` (250 lines)
- `BATCH_2_MANUAL_EDITS_SUMMARY.md` (comprehensive docs)

### Modified
- `src/main.js` (~30 lines modified)

## Edge Cases Handled

✅ First-time processing (no false positives)
✅ Rounding differences (<5% threshold)
✅ Same value (no unnecessary marking)
✅ Missing fresh data (preserve by default)
✅ Partial manual edits (independent field tracking)

## Benefits

1. **User-friendly**: Manual corrections respected during re-processing
2. **Intelligent**: 5% threshold prevents false positives from rounding
3. **Comprehensive**: Works for all protected fields (EPC, sqft, price, etc.)
4. **Well-tested**: 5 test scenarios covering all cases
5. **Backward compatible**: No breaking changes

## Documentation

📄 `BATCH_2_MANUAL_EDITS_SUMMARY.md` - Full implementation details
📄 `BATCH_2_MANUAL_EDITS_SUMMARY.pdf` - PDF version
📄 `test-batch-2-manual-edits.js` - Runnable test suite

## Related Issues

Fixes: Batch 2 Issue - Manual edits overwritten during re-processing

## Checklist

- [x] Implementation complete
- [x] Tests passing
- [x] Documentation created
- [x] Edge cases handled
- [x] No breaking changes
- [x] Ready for review
```

Review Checklist

Before merging, verify:

- [ ] All tests pass (`node test-batch-2-manual-edits.js`)
- [ ] No breaking changes to existing functionality
- [ ] Documentation is clear and comprehensive

- [ ] Code follows project conventions
- [ ] Edge cases are handled appropriately

Post-Merge Actions

After merging to master:

1. **Deploy to Production**
   - Update Apify actor with latest code
   - Test with real output CSV re-processing

2. **User Communication**
   - Notify user that manual edit preservation is live
   - Provide usage example

3. **Monitor**
   - Watch for any false positives (legitimate data marked as manual)
   - Verify 5% threshold is appropriate for sqft comparisons

Rollback Plan

If issues arise:

```
# Revert to previous master commit
git checkout master
git reset --hard HEAD~1
git push --force origin master
```

Contact

For questions or issues:
- Branch: `fix/batch-2-preserve-manual-edits`
- Commit: `8eecf8f`
- Test: `node test-batch-2-manual-edits.js`

Status: ✅ Ready to push
Date: December 5, 2024
Author: DeepAgent (Abacus.AI)