

PropertyData Sq. ft Extraction Fix

Issue

- **Symptom:** Sq. ft column was 100% empty (0/30 rows had data)
- **Root Cause:** Regex pattern only matched “sq ft” but PropertyData uses “sq **feet**”

The Bug

```
// OLD REGEX (broken)
/sq\.\?\s*ft/i // Matches: "sq ft", "sqft"
// Does NOT match: "sq feet" ✗
```

PropertyData HTML structure:

```
<tr>
  <td>Internal area</td>
  <td>1,119 sq feet</td> <!-- Note: "feet" not "ft" -->
</tr>
```

The Fix

Changed all regex patterns to match both variants:

```
// NEW REGEX (fixed)
/sq\.\?\s*f(?:ee)?t/i // Matches: "sq ft", "sqft", "sq feet" ✓
```

Files Modified

- `src/scrapers/propertyDataScraper.js`
- Line 109: £/sqft extraction pattern
- Line 122-123: Main floor area extraction (added “internal area” pattern)
- Line 141: Fallback £/sqft pattern
- Line 157: Fallback floor area pattern

Test Results

Before Fix

```
{
  "Sq. ft": undefined,
  "£/sqft": "£93",
  "Sqm": undefined
}
```

After Fix

```
{  
  "Sq. ft": 1119,  
  "£/sqft": "£93",  
  "Sqm": 104  
}
```

Verified URLs

- ✓ <https://propertydata.co.uk/transaction/36A61A94-9FAE-DEF2-E063-4704A8C046AE> (1,119 sq ft)
- ✓ <https://propertydata.co.uk/transaction/3DCCB7CA-6C3C-5B9D-E063-4704A8C0331E> (840 sq ft)

Impact

- Sq. ft column now populates correctly (was 100% empty)
- Sqm column auto-calculates from Sq. ft
- Ranking algorithm can now use floor area (40% weight in comparables)

Commit

```
commit add08ba  
Fix: PropertyData scraper now extracts 'Internal area' (sq feet)
```