

Site structure:

Home/Listings page (<https://listings.umn.edu/listing>): This page uses heavy JavaScript to load data, so a headless browser is necessary. At the top it shows filters for Beds, Baths and price (Per Unit vs. Per Person). Below the filters are neighbourhood tabs (e.g., Dinkytown, Marcy-Holmes, Como, Prospect Park) that limit results to each area. Selecting a tab triggers an AJAX call to update the listings grid.

Listings grid: Each listing is displayed as a card (<div>), with an image, the property's name, price range (e.g., "\$1250.00–\$3475.00"), bed range (e.g., "1–5 Bed"), and a brief address line (street + city). It also shows an "Available" date and a small "Walk Time To Campus" estimate. There is no traditional <a> link; clicking anywhere on the card triggers a JavaScript event to load the property details.

Property detail modal/page (?property=<id>): When you click a listing card, the site appends ?property=<id> to the URL and displays a modal overlay containing full property details. This modal includes:

Property name & full address (street address, city, state, ZIP).

Contact information (phone and email for the landlord).

Unit table: A <table> lists all available units. Columns include "Name" (unit identifier), "Beds", "Baths", "Price From", "Sq.Ft", "Available" (date), and sometimes "Tour" (indicates if virtual tour is available). Each row corresponds to a different floorplan. This table is scrollable if there are many units.

Other data: The modal may also display a bed-range and price-range summary at the top (e.g., "1–5 Bed \$1250.00–\$3475.00"), along with images and a Street-View tab.

Important notes for scraping:

Because the site relies on JavaScript, requests/BeautifulSoup won't work – use Selenium/Playwright to drive a real browser.

After the initial page load, you may need to scroll to load more listings (infinite scroll). Use a loop to scroll to the bottom and wait for new cards until no more appear.

Each card click opens a modal; you should wait for the modal to render fully before extracting data. Use `WebDriverWait` on the modal's unit table element.

Extract the property name, full address, bed range and price range from the modal header. Then parse each row of the unit table into a structured record (unit ID, beds, baths, price, square footage, availability date).

The site does not expose latitude/longitude, so to calculate `dist_to_campus_km` you'll need to geocode the address yourself (e.g., with Nominatim or Google Geocoding API) and then compute the haversine distance to the campus coordinates (lat = 44.9731, lon = -93.2359).

Close the modal after processing a property so you can click the next card. Use a short sleep or explicit wait to avoid overwhelming the site.

Save the collected data to a CSV or Excel file with columns: `listing_id` (use the property ID from the URL), `building_name`, `full_address`, `street`, `city`, `state`, `zip`, `beds`, `baths`, `size_sqft`, `rent_raw` (price from column), `rent_min/rent_max` (same as `rent_raw` because only a single price is given per unit), `available`, `unit_label`, `listing_url`, `landlord_phone`, etc.