Mini Project Idea 1: Real Estate Pricing in Hanoi

Research Question:

Which district-level and location characteristics predict listing price per m² in Hanoi's 12 central districts, and are those effects different for locations nearer the city center/main roads versus farther away? Secondarily, did the average price per m² change between Jul–Oct 2022 and the same months in 2023?

Motivation:

Hanoi's rapid growth produces sharp neighborhood price differences that matter for households, developers, and city planners. Understanding how access (distance to CBD and main roads), built environment (road width), and density relate to prices can guide zoning, transport priorities, and equitable development—especially when downtown districts (e.g., Hoàn Kiếm, Hai Bà Trưng) behave differently from edge districts.

Proposed Data Sources

- Real-estate listings (Jul–Oct 2022 and Jul–Oct 2023) scraped from major Vietnam property sites. Fields: listing price (VND), property area (m²), address/ward, road width (if stated), text description.
- Administrative boundaries for districts/wards (Hanoi open data or GADM).
- Points of interest (POIs): lakes, tourist sites, universities (OpenStreetMap).
- Road network (OpenStreetMap) to compute distance to nearest primary/secondary road; road-class as proxy for "main road."
- District characteristics: area (km²), population, and population density (General Statistics Office, Hanoi Statistical Yearbook).

Methodology

I will make a clean, de-duplicated dataset of listings in the 12 central districts, calculate price per square meter (m²), and put each listing on a map so it's assigned to the right district/ward. For each listing, I'll measure how far it is from downtown, from major roads, and from lakes. I will also add simple district facts like size, population, and how crowded it is (density) and mark whether a place is in a tourist area, near a university, or close to a lake. I'll tag listings so we can compare the same months in 2022 and 2023. For visuals, I'll make a map of average price per m² with a density layer, show basic price distributions by district, compare district averages across 2022 vs. 2023, and include a small correlation summary. I expect homes that are closer to downtown/major roads and in denser areas to be pricier. After accounting for access and density, raw district size or population probably won't matter much, and the "near lake/tourist/university" labels may have only small effects.

Challenges

- Listing bias & duplication: Premium areas are over-advertised; deduplicate aggressively.
- Measurement error: "Road width" in text can be inconsistent; treat as noisy or proxy with road class from OSM.
- Geocoding accuracy: Vietnamese addresses vary; manual checks on outliers.
- Omitted variables: Building quality/age, school zones; mitigate with district FE and rich proximity controls.
- Spatial autocorrelation: Nearby listings have correlated errors; consider clustering by ward or using spatial HAC (optional).
- Temporal comparability: Only two short windows (Jul–Oct); avoid broad claims about long-run trends.