

## Data Requirement

- **Deal basics:** Address, units, unit mix, square footage, construction type, purchase price, soft/hard costs, contingency, developer fee, timeline.
  - Why: These define total project cost, schedule, and revenue potential; they anchor IRR, LTV, DSCR, and cash flow forecasts.
  - Free sources & mechanism:
    - **User-entered** in the app with validations.
    - **Building permits and property records:** Local city open data portals (e.g., NYC Open Data), county assessor sites, PDFs/CSVs download; scrape with rate limits and caching.
- **Revenue assumptions:** Market rents by unit type, rent growth, vacancy, lease-up curve, AMI targeting and income-qualified rent caps (affordable units).
  - Why: Drives NOI and feasibility; affordability constraints cap rent even if market rates are higher.
  - Free sources & mechanism:
    - **HUD Income Limits & AMI:** CSV/Excel from HUD; scheduled ETL with yearly refresh.
    - **Census ACS (incomes, demographics):** API; pull block group/tract-level metrics and cache.
    - **Zillow/Redfin public datasets (asking rents, time series):** CSV downloads; monthly ETL, align to metro/county.
- **Operating expenses:** Taxes, insurance, utilities, repairs, management fees, reserves.
  - Why: NOI accuracy depends on realistic OpEx; DSCR and IRR are sensitive to these.
  - Free sources & mechanism:
    - **Property tax rates:** County assessor/treasurer sites; scrape mill rates and exemptions; normalize by jurisdiction.
    - **Insurance and utilities benchmarks:** State insurance commissioners' reports and utility tariff PDFs; manual seed + periodic scrape.
    - **Census CTPP/ACS proxies for household costs:** API to estimate utilities where tariffs are sparse.
- **Financing terms:** Loan amount, LTV, interest rate, amortization, covenants, construction loan draw schedule, permanent takeout, hedging.
  - Why: Drives leverage, DSCR, interest carry, and stabilization timing; affects IRR/Equity Multiple.
  - Free sources & mechanism:
    - **Rates & indices:** FRED for SOFR, Treasury yields; API scheduled pulls.
    - **FHFA Multifamily rates, Mortgage indices:** CSV/API; monthly ETL.

## Market, comps, and performance trends

- **Rent, price, absorption trends:** Time series by neighborhood/metro.
  - Why: Underwrite rent growth and lease-up risk; validate pro forma vs market reality.
  - Free sources & mechanism:
    - **Zillow/Redfin data:** CSV; join to ZIP/county; compute rolling medians.
    - **FHFA HPI (home price index):** CSV/API; metro/census division-level growth.
- **Employment and macro drivers:** Unemployment, payroll growth, CPI.
  - Why: Macro resilience impacts vacancy, concessions, and risk scores.
  - Free sources & mechanism:
    - **BLS (LAUS/CES, CPI):** APIs; monthly ETL.
    - **FRED mirror of macro series:** API; backfill and cross-check.
- **Population and household formation:** Demographics, income distribution, renters vs owners.
  - Why: Demand modeling and affordability targeting; opportunity sizing.
  - Free sources & mechanism:
    - **Census ACS (5-year):** API; tract-level, cached; yearly refresh.

## Regulatory monitor

- **Zoning and land use:** Base zoning, FAR/height, parking, setbacks, allowable density, overlays.
  - Why: Determines buildable area, density bonuses, feasibility constraints; central to entitlement risk.
  - Free sources & mechanism:
    - **Municipal zoning code PDFs + GIS layers:** City planning portals; shapefiles/GeoJSON via download; parse with GIS ETL (e.g., ogr2ogr), text extraction for code.
- **Inclusionary/affordable policies:** Set-aside requirements, income bands, rent caps, fee waivers.
  - Why: Directly modifies economics; affects AMI targeting and unit mix economics.
  - Free sources & mechanism:
    - **City/State housing department pages:** HTML/PDF scrape; maintain a policy registry with effective/sunset dates.
- **Permitting timelines and fees:** Process steps, typical durations, impact fees.
  - Why: Impacts carry and schedule risk; affects contingency.
  - Free sources & mechanism:
    - **Municipal permitting dashboards/open data:** CSV/PDF; scrape permit timelines by project type; compute median durations.
- **Risk scoring inputs:** Historic appeals, moratoria, tenant protections, rent control status.
  - Why: Converts policy complexity into a numeric risk score for decision-making.
  - Free sources & mechanism:
    - **City council records, planning commission minutes:** RSS/HTML crawl; keyword tagging; store with timestamps.

## Grants and tax credits module

- **Program catalog:** LIHTC, HOME, CDBG, Opportunity Zones, state/local grants, tax abatements.
  - Why: Non-dilutive capital and credits materially change IRR, Equity Multiple, and feasibility.
  - Free sources & mechanism:
    - **HUD (HOME, CDBG) program pages:** Bulk PDFs/HTML; scrape amounts, eligibility, deadlines; normalize.
    - **IRS/DoT for Opportunity Zones:** CSV shapefiles of OZs; geospatial join to site coordinates.
    - **State Housing Finance Agencies (LIHTC/QAPs):** PDF scraping for deadlines, scoring criteria; maintain per-state registry.
    - **City portals for local grants/fee waivers:** HTML/PDF; cron scraping; alerting on changes.
- **Eligibility logic:** AMI bands, location overlays (OZ, TOD), project type.
  - Why: Auto-matches deals to programs; drives “Create Alert” and recommendations.
  - Free sources & mechanism:
    - **HUD Income Limits + OZ maps:** API/CSV; on-the-fly geospatial and income band checks.

## GP portfolio and benchmarking

- **GP performance:** Historical deals, equity contributions, cash flows, distributions, realized/unrealized IRR, multiples.
  - Why: Relationship management and capital allocation decisions; benchmark vs industry.
  - Free sources & mechanism:
    - **Internal fund CRM/ledger:** Secure upload/API; required for real performance.
    - **Public filings (if applicable):** SEC EDGAR (forms, fund docs) via EDGAR API; parse for portfolio disclosures.
- **Industry benchmarks:** Typical IRR/multiple ranges by strategy and vintage.
  - Why: Context for “Top Performer / Needs Attention”; helps normalize risk-adjusted outcomes.
  - Free sources & mechanism:
    - **Academic/industry reports with open access:** Download PDFs; seed benchmark ranges; update annually.

## Fund returns and capital deployed

- **Capital account data:** Contributions, NAV, valuations, write-ups/downs, capital deployed by strategy.
  - Why: Drives Net IRR, Net Multiple, deployment charts, and remaining capital.
  - Free sources & mechanism:
    - **Internal fund accounting exports:** CSV from accounting systems; periodic ETL.
- **Quarterly performance:** Deal-level and roll-up time series.
  - Why: Populate “Performance Trends” and Net IRR charting over quarters.
  - Free sources & mechanism:
    - **Internal valuations:** CSV; standardized schemas and audit fields.

## Geographic investment map and geospatial layers

- **Geocoded deals:** Latitude/longitude, neighborhood, city, county, census tract.
  - Why: Map visualization, overlay with OZ and zoning, and unit counts by state/market.
  - Free sources & mechanism:
    - **OpenStreetMap Nominatim / Census Geocoder:** Free geocoding APIs; batch with rate limits; cache results.
- **Boundaries and overlays:** Tracts, ZIPs, OZs, zoning overlays.
  - Why: Spatial eligibility checks, market classification (S/M/L), and risk mapping.
  - Free sources & mechanism:
    - **Census TIGER/Line shapefiles:** Bulk download; pre-process to GeoJSON/tiles.
    - **IRS OZ polygons:** CSV/shapefile; join to deals.

## Financial model exports

- **Cash flow engine inputs:** Rent roll, expense schedule, capex, financing, draw schedules, exit assumptions.
  - Why: Generates 10-year IRR, Equity Multiple, LTV, DSCR, sensitivity tables, and annual cash flows.
  - Free sources & mechanism:
    - **Internal deal inputs + macro indices (FRED, BLS):** API-driven parameters; deterministic model with scenario toggles.
- **Sensitivity data:** Shock scenarios on rent, vacancy, interest rate, cost overruns.
  - Why: Stress testing; informs risk appetite and approval gates.
  - Free sources & mechanism:
    - **Macro series (FRED/BLS) + historical vol:** API; compute percentile shocks; store presets.

## Open, free sources and how to get them

- **HUD Income Limits, HOME, CDBG:** CSV/Excel/PDF; direct downloads; annual refresh; store versioned files.
- **IRS/Treasury Opportunity Zones:** Shapefiles/CSV; bulk download; geospatial join to deal points.
- **Census ACS & TIGER/Line:** APIs for ACS; FTP/bulk for TIGER; cache tract-level metrics; normalize geographies.
- **BLS (employment, CPI):** APIs; monthly pulls; keep series IDs and metadata.
- **FRED (rates, macro indices):** API; tag each series; backfill missing observations.
- **FHFA HPI & mortgage indices:** CSV/API; quarterly/monthly refresh; align to metros.
- **Zillow/Redfin public datasets:** CSV; monthly ETL; schema alignment; compute rolling stats.
- **Municipal open data (permits, zoning GIS, regulations):** Portals with CSV/GeoJSON/PDF; scrape with sitemaps and change detection; store provenance.
- **OpenStreetMap/Census Geocoder:** API for geocoding; batch jobs with retry and local caching.

Source	Exact Data You Get	Why It's Needed	How to Get It
HUD Income Limits, HOME, CDBG	AMI bands, income limits by household size, program allocations, grant amounts	Determines rent caps for affordable units, eligibility for HOME/CDBG funding	CSV/Excel/PDF downloads from HUD.gov; annual refresh
IRS/Treasury Opportunity Zones	Shapefiles/CSV of designated OZ census tracts	Identifies tax incentive eligibility for deals in OZs	Bulk download from IRS/Treasury portals; geospatial join to deal coordinates
Census ACS & TIGER/Line	Demographics (income, household size, tenure), tract boundaries, shapefiles	Market demand analysis, affordability targeting, geospatial overlays	ACS API for demographics; TIGER FTP for shapefiles; cache and normalize
BLS (employment, CPI)	Local unemployment rates, payroll data, CPI inflation indices	Macro drivers for vacancy, rent growth, and expense escalation	BLS APIs (LAUS, CES, CPI); monthly pulls with series IDs
FRED (rates, macro indices)	Treasury yields, SOFR, Fed Funds, macroeconomic indicators	Financing assumptions, discount rates, stress testing	FRED API; tag series IDs; backfill missing observations
FHFA HPI & mortgage indices	Home Price Index by metro/division, mortgage rate indices	Market growth benchmarking, cap rate trends	CSV/API from FHFA; quarterly/monthly refresh
Zillow/Redfin public datasets	Asking rents, home values, absorption, time series by ZIP/county	Rent comps, growth assumptions, lease-up risk	CSV downloads; monthly ETL; schema alignment
Municipal open data (permits, zoning GIS, regulations)	Permit records, zoning shapefiles, regulatory updates	Entitlement risk scoring, feasibility constraints	City open data portals; CSV/GeoJSON/PDF scrape; change detection

OpenStreetMap/Census Geocoder	Geocoded lat/lon for addresses, tract/ZIP joins	Map visualization, spatial overlays (OZ, zoning, AMI)	Free API calls; batch jobs with retry and caching
-------------------------------	---	---	---

## Data ingestion and refresh mechanism

- **Pipelines:** Scheduled ETL (daily/weekly/monthly based on source cadence) with idempotent jobs and schema validation.
- **Versioning:** Keep source timestamps, effective/sunset dates, and provenance for auditability.
- **Geospatial processing:** Normalize CRS, snap points to polygons, precompute joins for OZ, zoning, and tracts.
- **API layer:** Expose normalized endpoints (e.g., /markets/{geo}/rents, /grants/{geo}/eligibility) for the dashboard.
- **Change alerts:** Watch regulatory pages and grant portals for updates; trigger “Create Alert” notifications.

## Minimal data model (to implement what’s visible in the mockups)

- **Entities:** Deals, Units, Markets, Policies, Programs, GPs, Fund, CashFlows.
- **Key fields:**
  - **Deals:** id, address, lat/lon, unit\_mix, costs, timeline, financing\_terms, gp\_id, market\_id.
  - **Markets:** geo\_ids, rent\_series, vacancy, employment, CPI.
  - **Policies:** jurisdiction, type, description, effective\_date, sunset\_date, score\_components.
  - **Programs:** name, level (federal/state/local), type (credit/grant), amount, deadline, eligibility\_rules, links.
  - **GPs:** id, name, deals[], realized\_irr, unrealized\_irr, multiples.
  - **Fund:** capital\_committed, capital\_deployed, nav, net\_irr, net\_multiple.
  - **CashFlows:** deal\_id, date, inflow, outflow, tags (contribution/distribution/interest/capex).