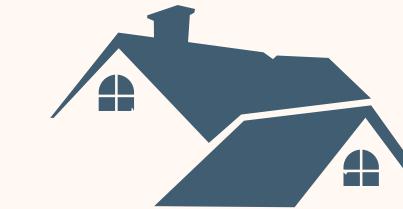


AFFORDABILITY IN MOTION

How Walkability, Transit, and Competition Shape Urban Housing Costs

By Insharah Irfan Nazir





PROBLEM STATEMENT

What factors really drive housing affordability across major U.S. cities, and do those drivers differ by city?



Zillow



Walk Score



US Census

DATA COLLECTION

- APIs: [zillow56](#) (unofficial), [Walk Score](#) (official), [Census](#) (official)
- Neighborhood-level housing prices, walkability & transit scores, demographics
- Queried APIs and merged datasets by city/neighborhood
- Used only publicly available, aggregated data
- Data is already public; used only for research purposes, not commercial gain, and accessed moderately
- Goal: ensure analysis highlights structural patterns, not individuals

address	Full street address of the property	String
city	City where the property is located	String
state	State abbreviation (US postal code format)	String
zip_code	5-digit ZIP Code	String/Int

DICTIONARY

price	Monthly rent price (USD)	Float
bedrooms	Number of bedrooms	Float
bathrooms	Number of bathrooms	Float
square_feet	Living space area (sq. ft.)	Float
population	Total population of the census area	Float

listing_date	Listing date (YYYY-MM-DD; may default if unavailable)	Date
latitude	Latitude coordinate of property	Float
longitude	Longitude coordinate of property	Float
url	Zillow property listing URL	String (URL)
walkscore	Walkability score (0–100; higher = more walkable)	Float
transit_score	Public transit accessibility score (0–100)	Float
bike_score	Bike-friendliness score (0–100)	Float
walkscore_error	Missing/error flag for Walk Score (null if no issue)	String/Null
median_income	Median household income in the property's census area (USD)	Float

white_pop	White population count	Float
black_pop	Black/African-American population count	Float
asian_pop	Asian population count	Float
hispanic_pop	Hispanic/Latino population count	Float
competition_density	Market competition proxy (integer category; higher = more competition)	Int
population_density_proxy	Approximate population density (population ÷ area estimate)	Float
diversity_index	Measure of demographic diversity (0–1, higher = more diverse)	Float
affordability_index	Affordability score (lower = more affordable relative to income)	Float
price_per_sqft	Rent price divided by square footage (USD/sq. ft.)	Float
affordability_ratio	Rent-to-income ratio (rent ÷ income, unitless)	Float
rent_premium	Premium/discount vs. expected rent (USD, can be negative)	Float



METHODOLOGY

CLEANING

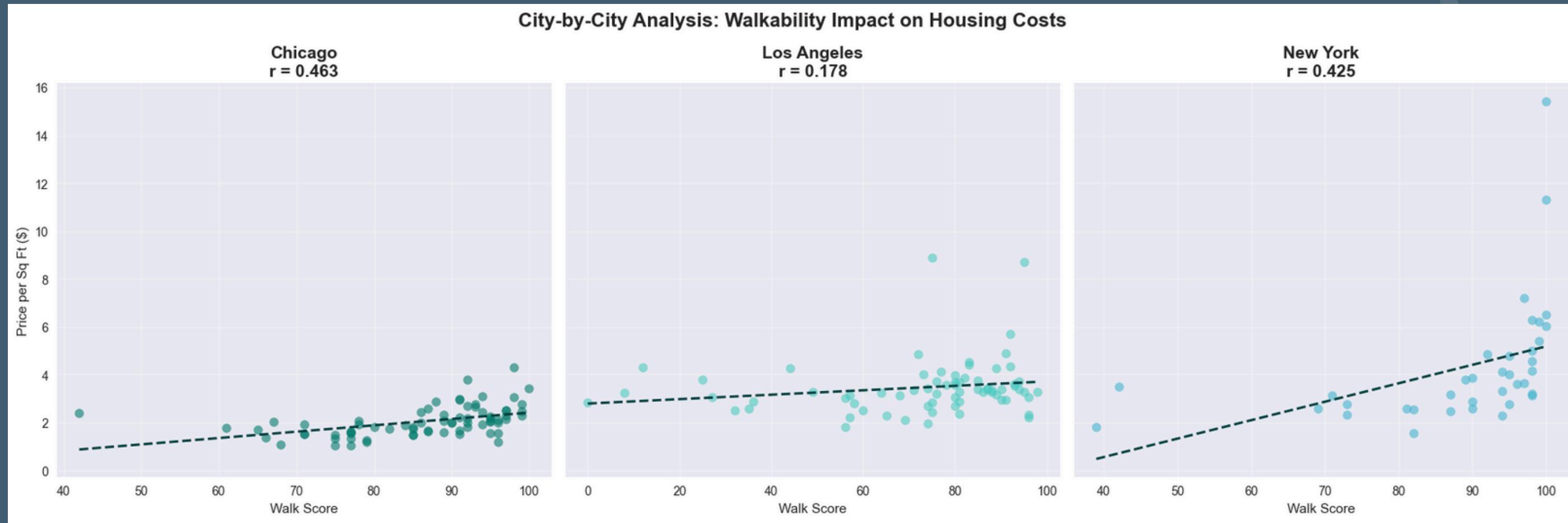
- Removed duplicate listings (using unique URL)
- Standardized ZIP codes, city, and state formatting
- Converted numeric/text/date fields into consistent formats
- Dropped rows with missing key values (price, ZIP, or failed geocoding)
- Fixed missing coordinates with geocoding (Nominatim)
- Filtered out extreme outliers in price, size, bedrooms, and bathrooms

ENRICHMENT

- Added competition density (listings per ZIP code)
- Created population density proxy ($\frac{\text{population}}{\text{competition}}$)
- Calculated diversity index (demographic mix measure)
- Derived affordability index ($\frac{\text{income}}{\text{price}}$)
- Engineered new features:
- Price per sq. ft.
- Affordability ratio ($\frac{\text{annual rent}}{\text{income}}$)
- Rent premium (vs. median rent)

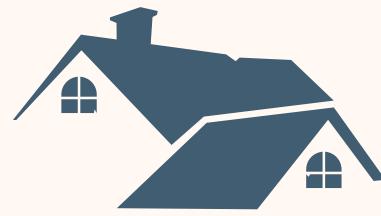
ANALYSIS

- Exploratory Data Analysis (EDA): Looked at distributions, outliers, and demographic patterns
- Correlation Analysis: Measured relationships between rent prices and predictors (income, population, walkability, etc.)
- Used AI tools to speed up correlation analysis by highlighting strongest predictors and interpreting coefficients.
- Helped prioritize features
- Provided statistical summaries



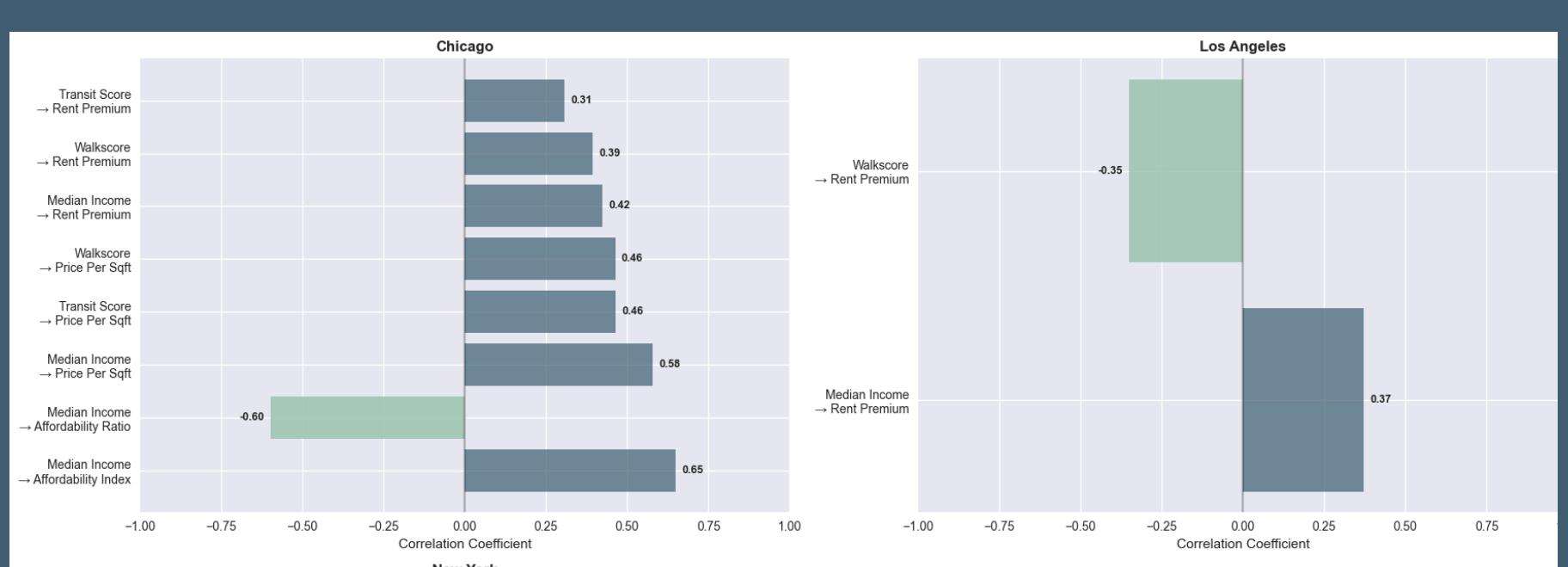
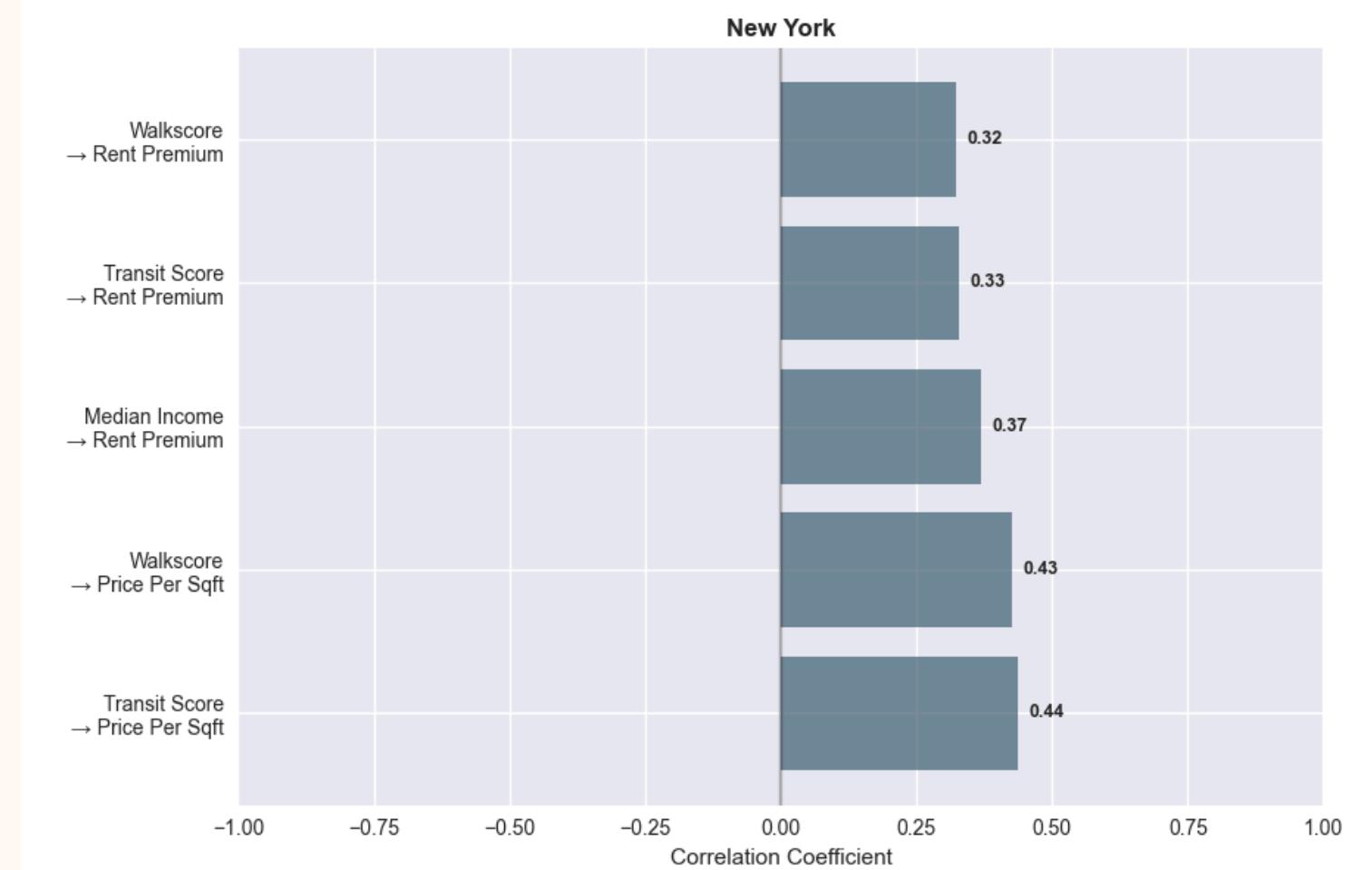
- Chicago shows a positive correlation ($r=0.463$) between Walk Score and Price Per Sq Ft, with points clustered around lower scores, indicating walkability modestly increases costs in a spread-out urban layout
- Los Angeles exhibits a weak correlation ($r=0.178$), suggesting walkability has minimal influence on housing prices due to car-dependent sprawl
- New York displays the steepest slope and strong correlation ($r=0.425$), trending upward, highlighting how high walkability in dense areas significantly drives up per-square-foot costs
- Overall, walkability's impact varies by city density: stronger in compact NYC, weaker in expansive LA

EDA: Walkability Impact on Housing Costs



INCOME & AFFORDABILITY

- Chicago, Los Angeles, and New York have very different housing markets
- City-specific trends instead of assuming all urban areas behave the same way
- Strongest correlations with affordability metrics by city



Chicago

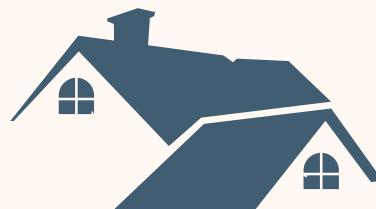
- Higher income/ratios → lower affordability
- Price per sq. ft. ↑ with affordability

Los Angeles

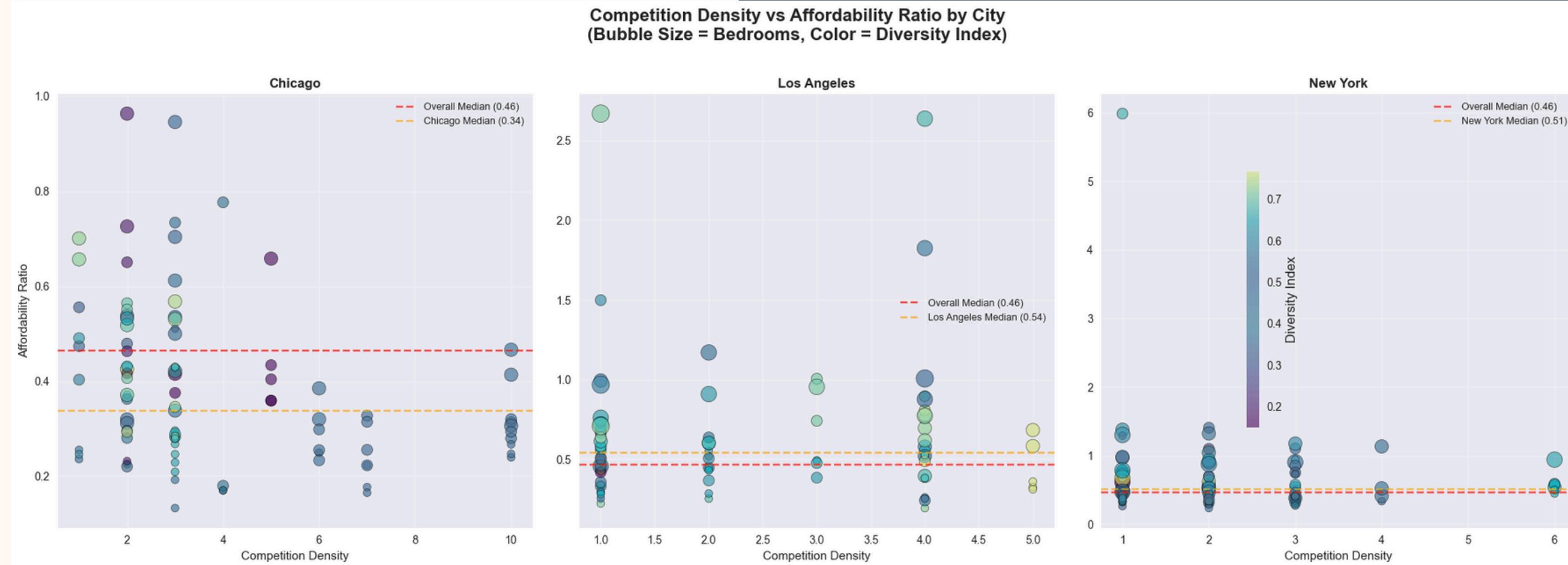
- Walkscore ↑ → rent premium ↑
- Income ↑ → offsets premium in low-walk areas

New York

- Walkability + income → push affordability ↓
- City Differences
- Chicago: income gaps
- LA: rent premiums
- NYC: spatial premiums (walkability)



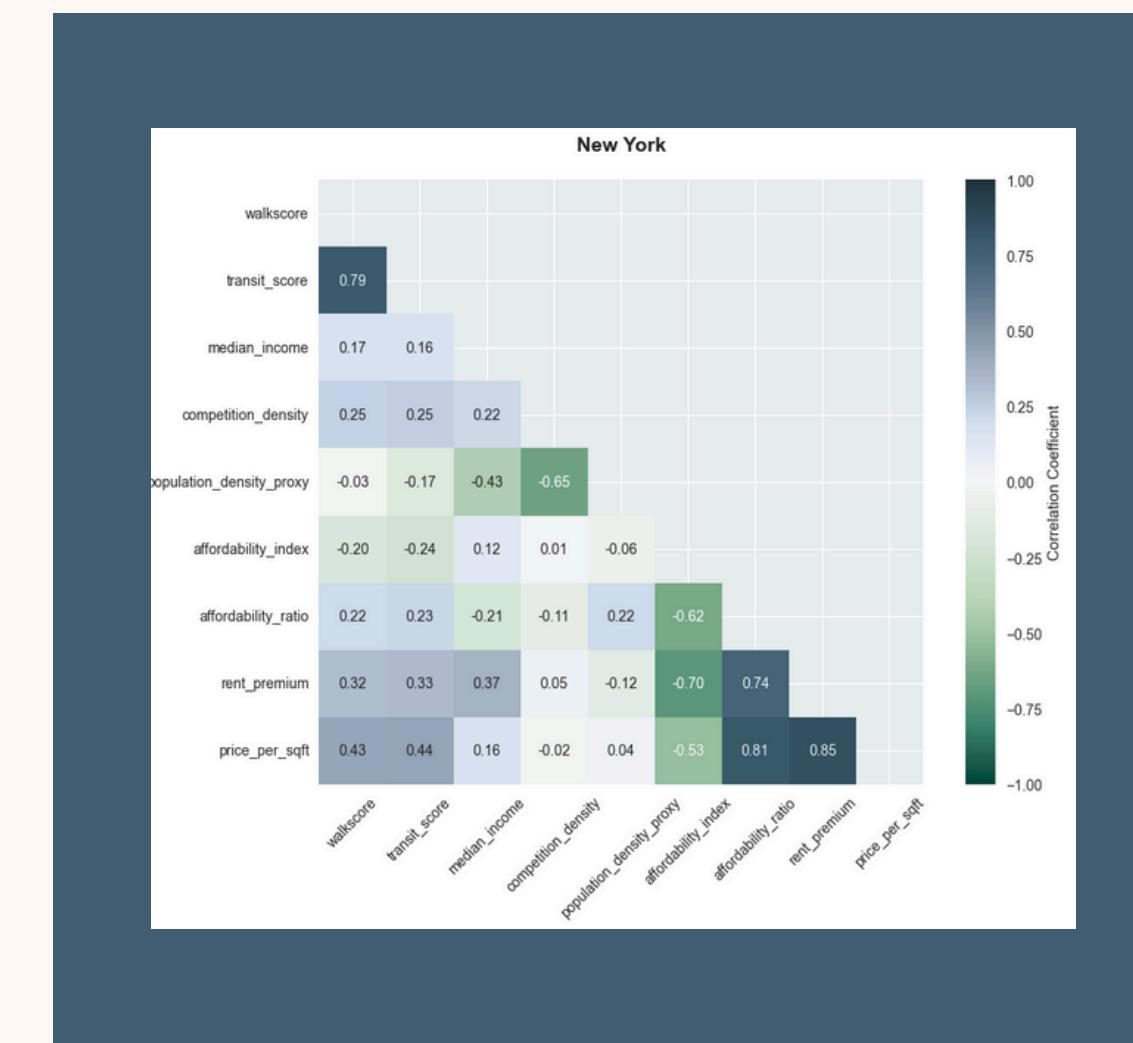
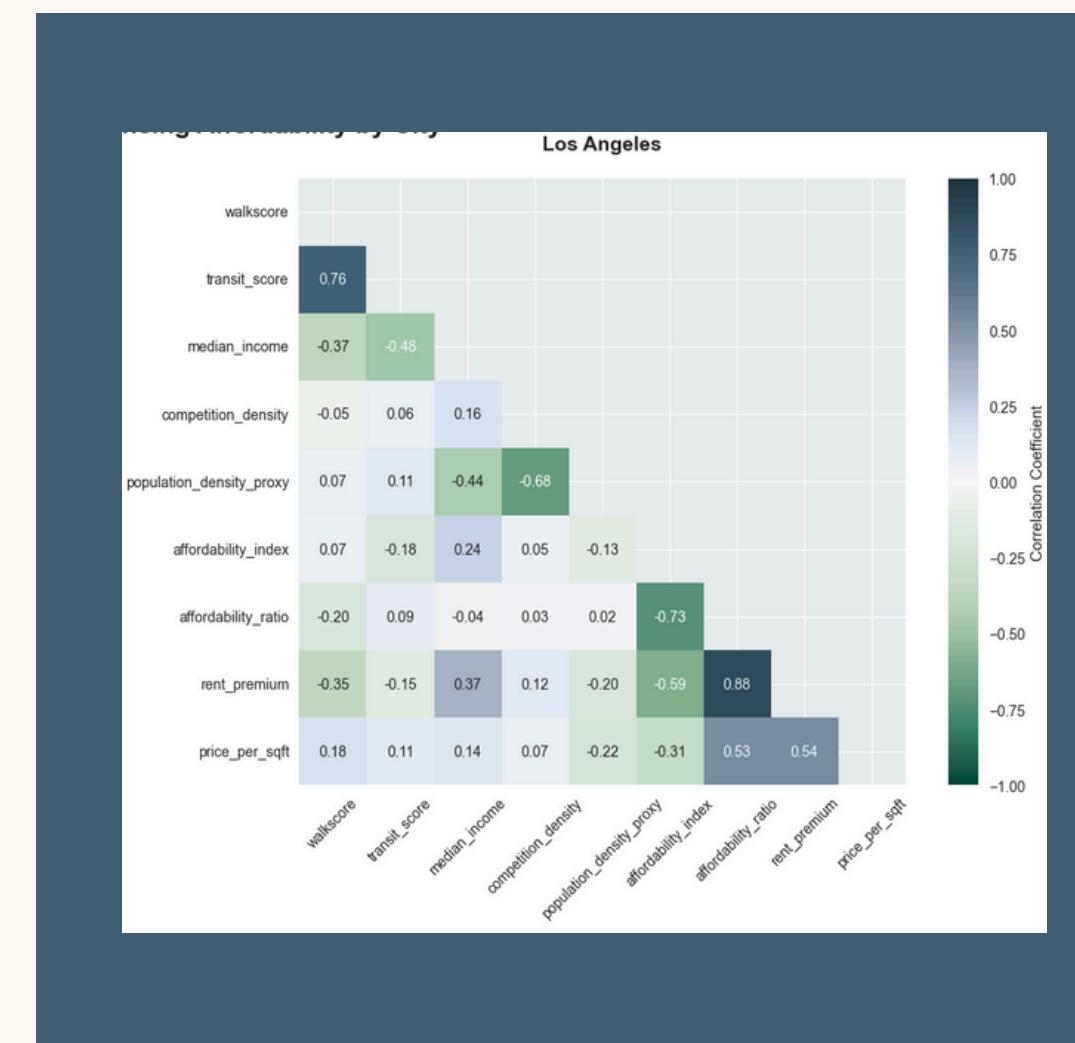
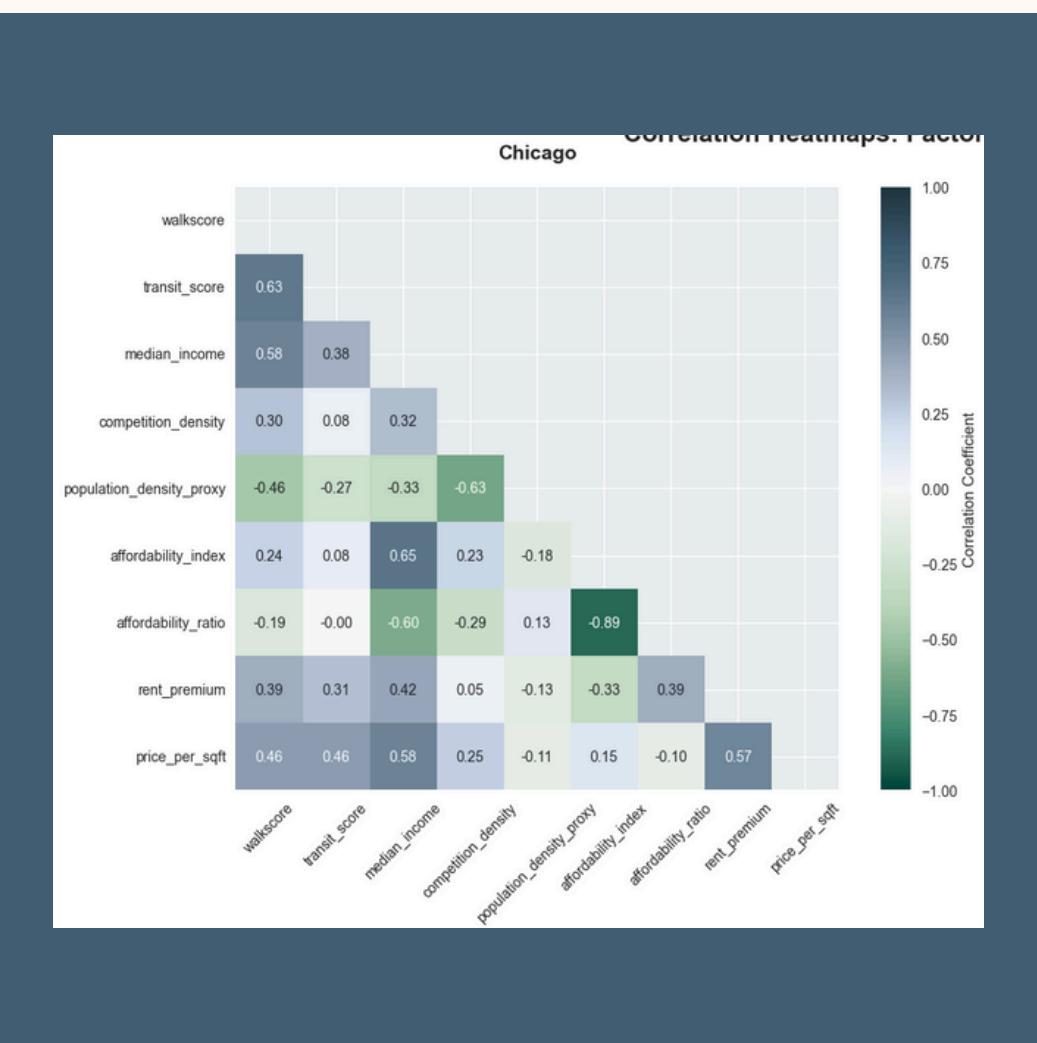
COMPETITION EFFECTS

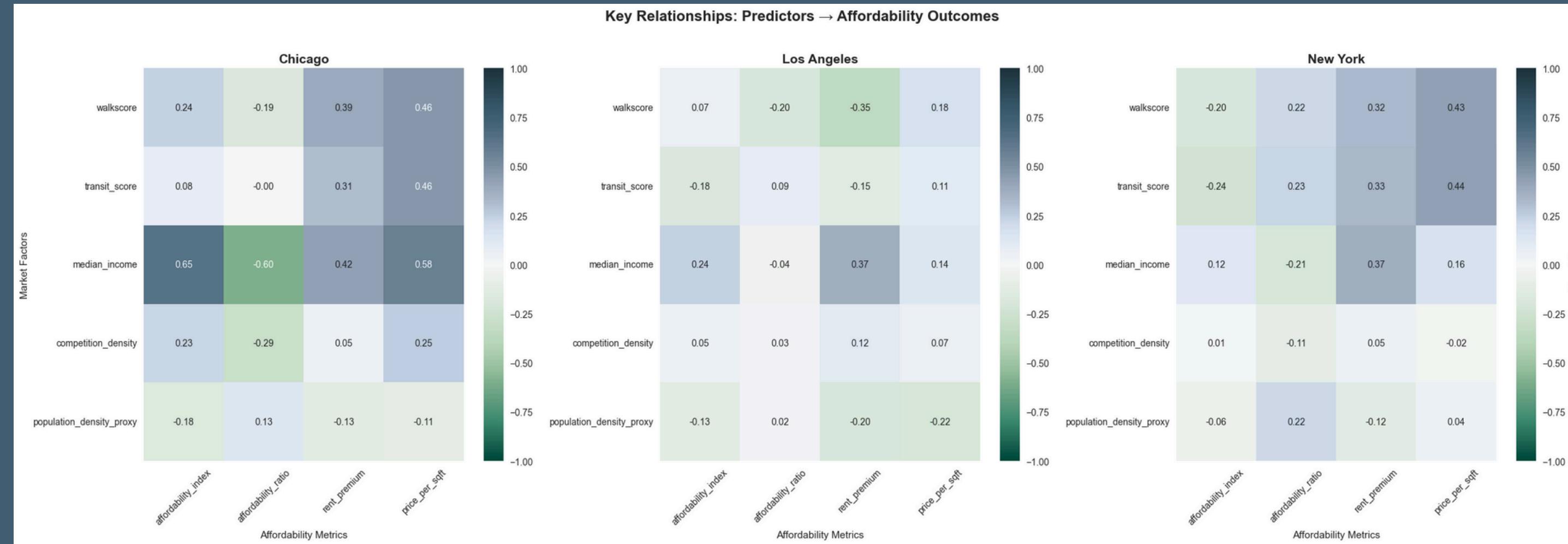


- Chicago: Low competition in affordable areas; dense zones show more diversity across affordability levels
- Los Angeles: Competition clusters in multi-bedroom markets – higher density, but affordability doesn't improve
- New York: Slight affordability gains with density, but niche pockets show intense competition
- Insight: Competition density plays out differently by city – in LA it intensifies costs, while in New York it only helps in select neighborhoods
- Pattern: Higher competition density generally correlates with worse affordability, varying by city scale—LA faces more intense multi-unit competition

CORRELATIONS: FACTORS INFLUENCING AFFORDABILITY

- Chicago: Clear trade-off – higher density reduces affordability despite strong walk–transit alignment
- Los Angeles: Affordability less tied to income; dominated by rent premiums and weaker correlations elsewhere
- New York: Strongest interconnections – density and premiums tightly linked, amplifying affordability loss
- Insight: Each city has a distinct affordability driver – density in Chicago, premiums in LA, and spatial amenity costs in New York





Predictors play different roles by city

Chicago: income dominates

Los Angeles: income buffers affordability

New York: density + premiums amplified

Overall Insights

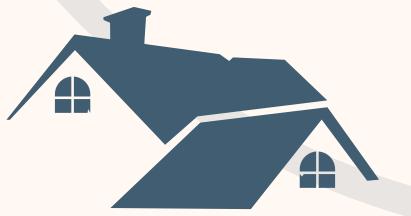
Walkability premiums strongest in Chicago & NYC

Income = biggest affordability driver

Competition lowers costs in Chicago, not LA

Transit access pushes prices up → risk to affordability

Most important: city-specific dynamics → no one-size-fits-all policies



RECOMMENDATIONS

How do we improve housing affordability?



- Tailor strategies city by city — affordability drivers differ
- **New York City Planners & Transit Agencies:** Expand transit/walkability in low-score neighborhoods
- **Chicago Policymakers & Housing Advocates:** Target income-driven affordability gaps
- **Los Angeles Developers & City Council:** Encourage multi-unit housing where competition is fiercest

LIMITATIONS

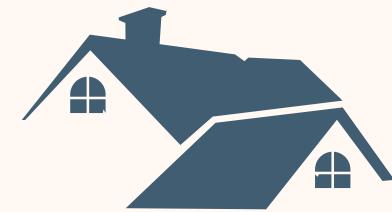


- Data Access Constraints: Relied on free APIs (limited coverage)
- Correlation ≠ Causation: Findings show associations, not proof of direct cause-effect
- Patterns highlight meaningful signals for policy and investment decisions despite limits

NEXT STEPS

- Explore rent trends over time (seasonality, year-over-year changes)
- Analyze neighborhood-level effects beyond ZIP codes
- Collect historical rent data for forecasting potential





THANK YOU

Patterns differ by city
Affordability is multidimensional

Data helps planners, policymakers, and investors target smarter
solutions



<https://github.com/insharahn/ds-challenge>