Strategic Segmentation & Valuation Analysis: Capital Allocation problem under Liquidity Risk for a Real Estate Private Equity in England and Wales

ECON 449: Data Science with Economic Applications Final Project

Arda Kırayoğlu
Bilkent University
Department of Economics

Samet Efe Keskin
Bilkent University
Department of Economics

April 27, 2026

1 Property Acquisition and Exit Strategy

1.1 Client Profile: Moorfield Group

Moorfield Group is a London-based, mid-sized real estate private equity firm with a history of identifying emerging opportunities in the British residential sector. The firm has recently raised a new £150M fund designated for the acquisition of a residential portfolio, and the **objective is to identify how the capital should be geographically allocated to maximize risk-adjusted returns** with the selection of neighbourhoods across England and Wales with the highest possible return.

1.2 Liquidity and Return on Investment

The firm is currently deadlocked on the geographic allocation of this capital. The core risk identified is the Liquidity Risk, the low sales volume danger of the acquired asset and loss of the overall value of the property.

1.3 Strategic Scenarios

To guide capital allocation, we formulated two distinct competing strategies. The purpose of this report is to evaluate which of these pathways offers the superior risk-adjusted returns. Our analysis eventually decides which strategy outperforms the other.

- **Scenario A - The "London Banana Zone" Strategy:** Concentration of capital into high-value assets in Prime Central London and the neighbouring counties often described as the London Banana Zone *(See the zone)*. The hypothesis is that assets in prestigous and high-end residential zones offer capital preservation, though concerns remain regarding transaction velocity and liquidity risk and declining premium.
- **Scenario B - The Diversification Strategy:** Deployment of capital into high-volume mid-sized cities and suburban markets in the Midlands and North to cities like Manchester, Leeds, Nottingham. The hypothesis is that these markets offer superior scalability and "Exit Liquidity" due to deeper demand from domestic buyers.

1.4 Methodology

The transaction data are derived from the UK HM Land Registry Price Paid Data, which records residential property transactions in England and Wales consisting of 190,217 transactions between (2015-2025)[1].

The objective is to rigorously evaluate the feasibility of two different strategies, utilizing comprehensive information on property prices, the age of the building, and the type and location of the property. Our approach is not to add all possible features, but to identify the features that have considerable explanatory power in explaining cross-sectional differences in housing prices. One can view this case as an optimal resource allocation in a cross-sectional asset pricing problem, specifically in the context of housing.

Overall analysis provides a snapshot of the trending areas and overall market dynamics, without delving into the micro features of the housing market. The core of our analysis lies in:

- **Inflation Adjustment:** Nominal transaction prices were adjusted to Real 2025 GBP using the UK Consumer Price Index (CPI) provided by Bank of England, ensuring that valuations reflect purchasing power parity rather than nominal values.
- **Geospatial Mapping:** Postcodes were converted into Latitude/Longitude coordinates using the `pgeocode` and `folium` library. This enables the K-Means algorithm to cluster properties based on physical proximity rather than arbitrary administrative boundaries.
- **Outlier Removal:** Transactions below £50,000 and above £20,000,000 were trimmed to prevent skewing the regression model.

Then it combines two fundamental machine learning approaches to test these strategies:

1. **Testing Liquidity Risk and Segmentation (Unsupervised Learning / K-Means):** We utilize K-Means Clustering to segment the market by *Transaction Volume*. This allows us to empirically verify liquid areas and to capture the real estate market dynamics. We utilize clustering to enhance the insights of the housing market.
2. **Testing Fundamental Value (Supervised Learning / Random Forest):** We utilize Random Forest Regression to model the *Intrinsic Value* of UK properties. By comparing predicted values against market prices, we can determine if the premium paid for "Scenario A" assets is justified by fundamentals, or if it represents a speculative bubble.

Our project details the analytical process and concludes with a definitive recommendation on which strategic pathway Moorfield Group should execute.

1.5 Results from Unsupervised Learning using K-Means - Segmentation

K-means clustering is a statistical method used to uncover the clusters hidden in the dataset. There are some standard geographic distinctions in UK housing market such as the London and its periphery's high prices and North-South divison where we observe affordable property prices in the north. We aimed to capture the different segments of properties based on the sales volume and the price of the

[1]HM Land Registry. *Price Paid Data.* Contains HM Land Registry data © Crown copyright and database right 2021. Licensed under the Open Government Licence v3.0. Available at: `https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads`

property using this method. Each data point is a neighbourhood of a city. We used the postal code information to categorize the locations. Using the data, we obtained the liquidity pools and how neighbourhoods are performing in terms of average property prices and sales volume.

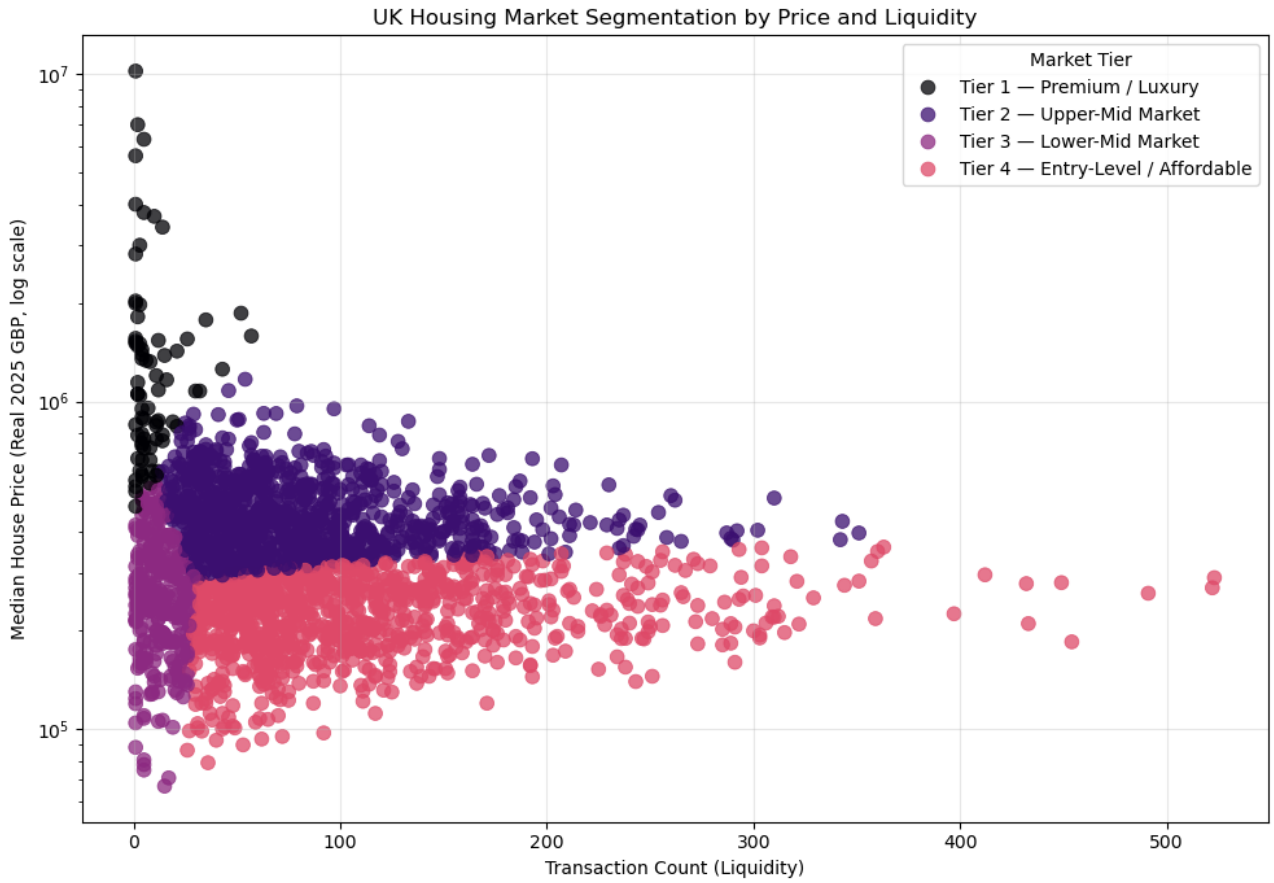


Figure 1: Source: HM Land Registry

There are 4 main clusters as the method decomposed. First tier is the high-end luxurious areas like London's Notting Hill, Westminister, Mayfair and Chelsea. Most of these tier properties are sold around 800.000 to 1.500.000 million pounds.

Second tier represents the upper-mid segment where the prices are between 300.000 to 800.000 pound. While this cluster includes established London neighborhoods like Islington, Battersea, and Wimbledon, our data reveals it is not limited to the capital. Crucially, this tier captures the *"Wealthy Commuter Belt Periphery"* areas like Croydon, Maidenhead, and Brighton where high-income professionals move for more spacious houses and less crowd.

The Third Tier represents a transitional segment of the market, with median prices ranging from £ 150,000 to £ 300,000. Unlike the other clusters, this group is characterized by *low transactional density*. The top nodes here average only 20 sales, compared to more than 100 in other tiers. Geographically, this tier comprises a mix of Rural Lifestyle locations (such as Cornwall and Wales) and urban neighborhoods. Specifically, older residential districts within major cities, such as Birmingham and Bradford, sit outside the prime investment zones and central areas. Unlike the new build suburbs, these are stagnant areas of attraction.

The Fourth Tier represents the entry-level market, with median prices ranging from £ 100,000 to £ 300,000. Unlike the fragmented nature of Tier 3, this cluster functions as the backbone of the UK housing market, delivering consistent, high-volume transactions that are essential for scalability.

Geographically, this tier is anchored by the industrial and logistics hubs of the Midlands and the North. Logistics Corridors like Preston where named one of fastest improving cities in UK and Warrington where infrastructure investment drives steady demand. Midlands Manufacturing Hubs, such as Shepshed and Nottingham, offer robust suburban housing stock at accessible price points. Sheffield's affordability continues to attract first-time buyers despite broader market headwinds.

Tier 4 Cluster is the optimal target for Strategy B. The data confirms that Tier 4 offers the unique combination of Scalability (volume 100 units/postcode) and lower price volatility.

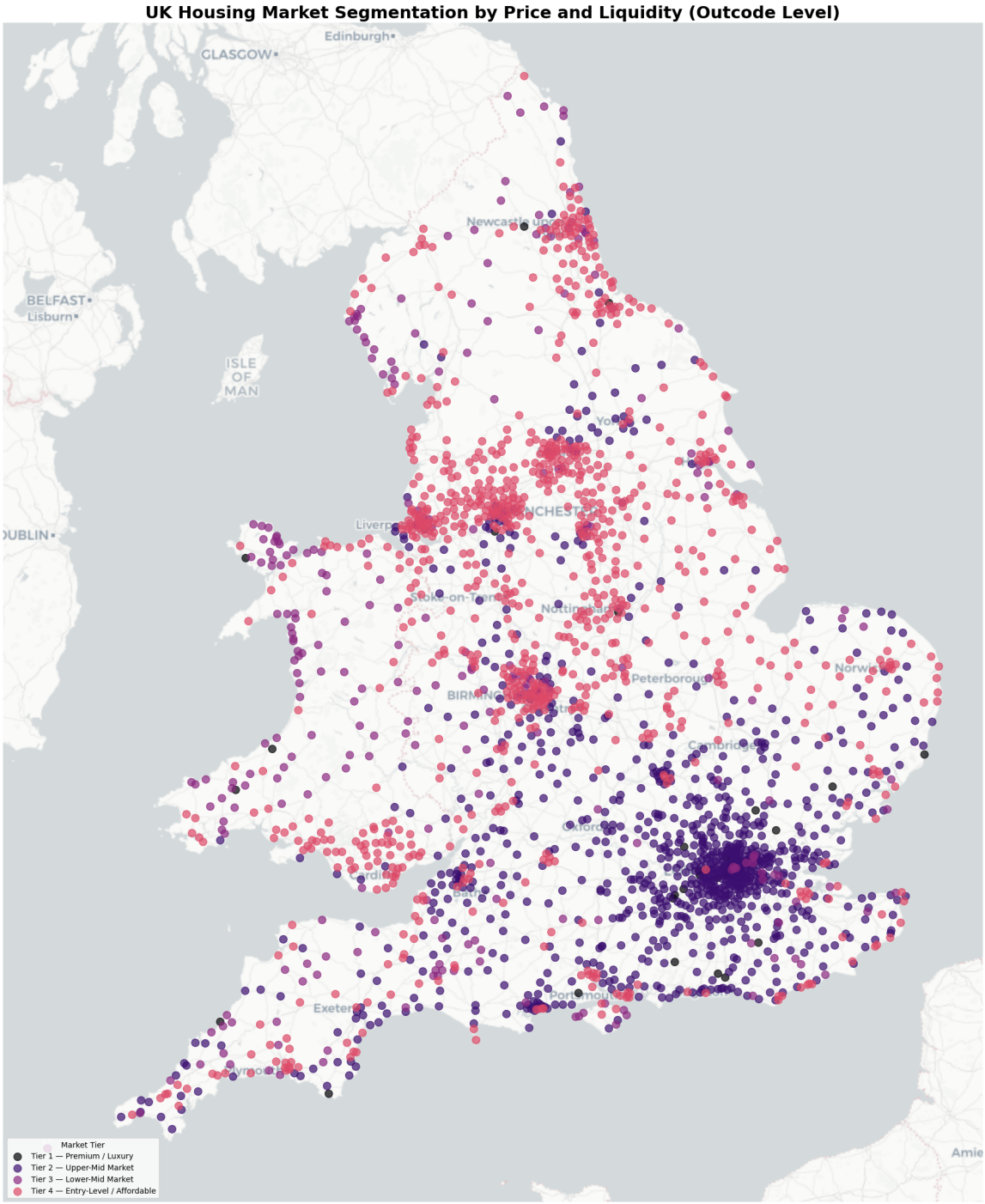


Figure 2: Source: HM Land Registry

1.6 Price Prediction Models: Random Forest Regression

To assess the extent to which neighborhood-level housing prices can be explained by observable market characteristics, we used random forest model. Random Forests are particularly suitable for this analysis because they flexibly capture nonlinear relationships and interaction effects.

Two model specifications are estimated in order to distinguish between the explanatory power of contemporaneous market fundamentals and the role of historical price persistence.

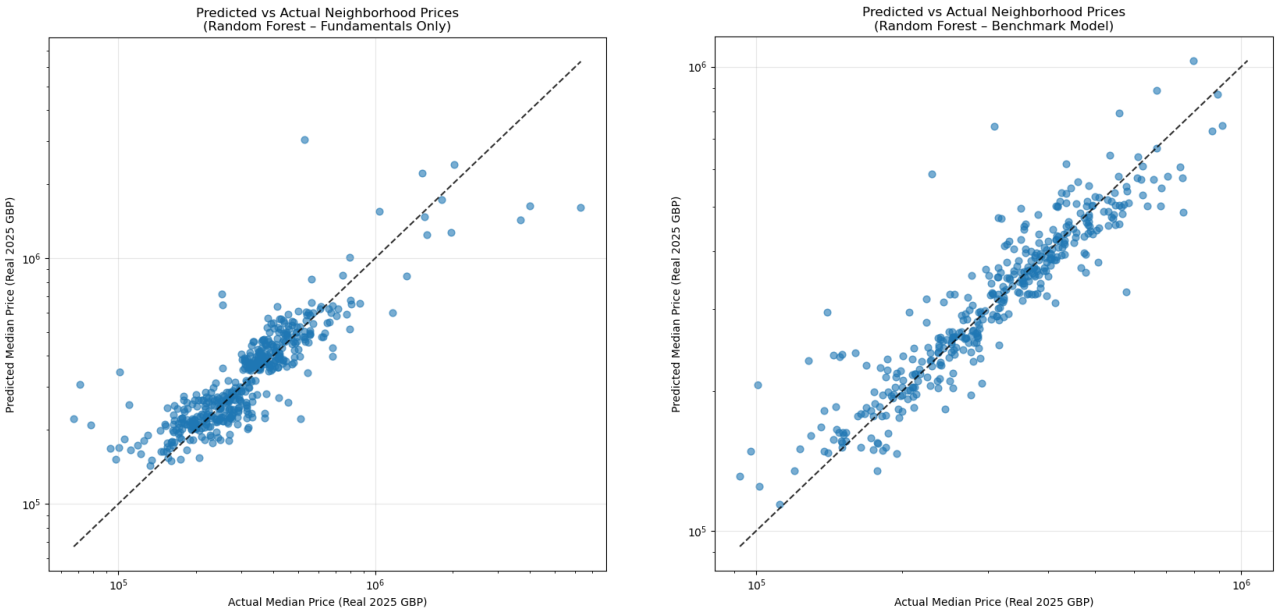
Model 1: Fundamentals-Only Specification: The first model relies exclusively on neighborhood characteristics as predictors and no price information is included. These include transaction volume as a proxy for market liquidity, geographic coordinates (latitude and longitude), and the market tier classification obtained from the unsupervised clustering analysis. The dependent variable is the neighborhood median house price expressed in real 2025 GBP.

This specification achieves a R^2 **of 0.469** and a root mean squared error **(RMSE) of £310,734**. While predictive accuracy is moderate, the model captures broad spatial and liquidity-driven patterns in housing prices. The relatively large prediction error highlights the limitations of explaining price levels using contemporaneous fundamentals alone, particularly in heterogeneous urban housing markets without the detailed features of each property.

Model 2: Benchmark Specification with Historical Prices: The second model extends the fundamentals-only specification by incorporating historical neighborhood price information, specifically lagged median prices from earlier periods. This benchmark specification reflects the empirical regularity that housing prices exhibit strong persistence over time.

Including historical price information substantially improves model performance. The benchmark model achieves an out-of-sample R^2 **of 0.807** and an **RMSE of £62,517.** This large improvement indicates that a substantial share of the cross-sectional variation in current housing prices can be explained by past price levels, consistent with strong price anchoring and slow adjustment dynamics in residential property markets.

Interpretation: Taken together, the two specifications provide complementary insights. The fundamentals-only model isolates the explanatory power of liquidity, spatial location, and market segmentation, while the benchmark model establishes an upper bound on predictive accuracy when historical information is available. The contrast between the two highlights both the economic relevance of structural neighborhood characteristics and the dominant role of historical pricing in shaping current housing valuations.



(a) Model 1 Results

(b) Model 2 Results

Figure 3: Comparison of Actual vs. Predicted Transaction Prices using two different RF models.

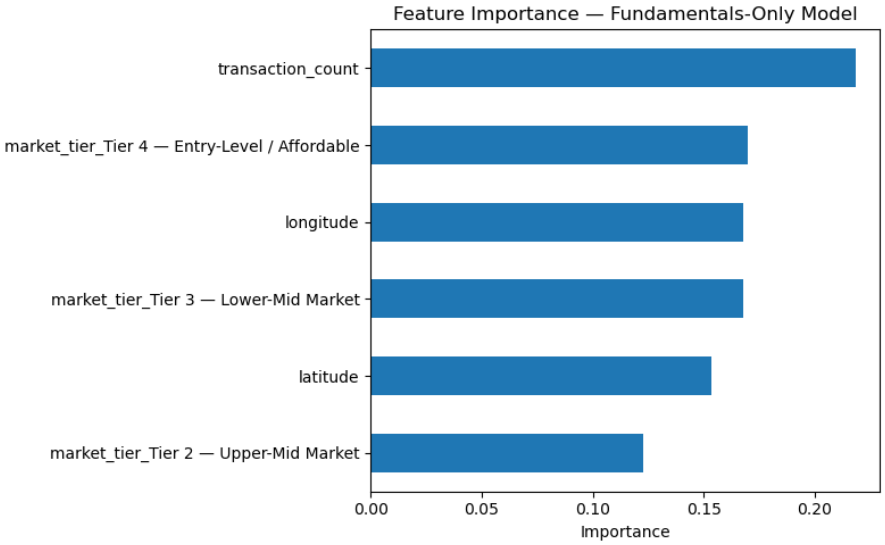


Figure 4: Important Features

1.7 Spatial and Distributional Housing Market Dynamics

Beyond segmentation and price prediction, the analysis explores spatial and distributional patterns in UK housing markets to provide additional context for investment strategy evaluation. Using geocoded transaction data and inflation adjusted prices, this section examines regional price levels, long-run real price growth, and heterogeneity across major urban areas.

Spatial Distribution of Real House Prices: Figure 5 presents the distribution of median house prices expressed in constant 2025 GBP. The map reveals a pronounced geographic gradient, with higher price levels concentrated in London and the South East, and substantially lower price levels across much of the Midlands, Northern England, and Wales. This pattern is consistent with long-standing regional disparities in income, employment opportunities, and housing demand.

Real Price Growth and Regional Dynamics: To distinguish nominal price movements, housing price growth is evaluated in real terms using CPI-adjusted prices. Figure 6 illustrates real compound annual growth rates between 2015 and 2025 at the neighborhood level. The results show considerable heterogeneity: while some Southern commuter regions exhibit weak or negative real growth, several neighborhoods in the Midlands and Northern England display strong real appreciation. This pattern provides evidence of partial regional convergence rather than uniform divergence.

Top and Worst Performing Neighbourhoods: A closer examination of the top performing neighborhoods reveals common characteristics. The strongest real gains are concentrated in areas such as Kirklees (West Yorkshire), Salford (Greater Manchester), and Cumberland, which are predominantly located in Northern England and the Midlands. These areas typically started the period with relatively low house price levels, allowing for larger percentage gains as housing demand expanded. Several of these regions have benefited from urban regeneration initiatives, improved transport connectivity, and demand from nearby major cities, particularly Manchester. In contrast, the weakest performers are largely found in Southern England and established commuter belts, including parts of Wiltshire, Wealden, and Sutton. These areas entered the period with already elevated price levels and appear to have experienced limited real appreciation or outright real declines once inflation is accounted for. Together, these patterns suggest that recent real housing returns have been driven less by traditionally high-priced locations and more by relative affordability, local economic revitalization, and convergence dynamics across regions.

Implications: Taken together, the spatial and distributional analyses complement the segmentation and modeling results. While premium markets remain concentrated in specific regions, real growth patterns suggest that long-run returns are not exclusively driven by traditionally expensive areas. These findings support the strategic relevance of geographic diversification and motivate further consideration of regional dynamics in housing investment decisions.

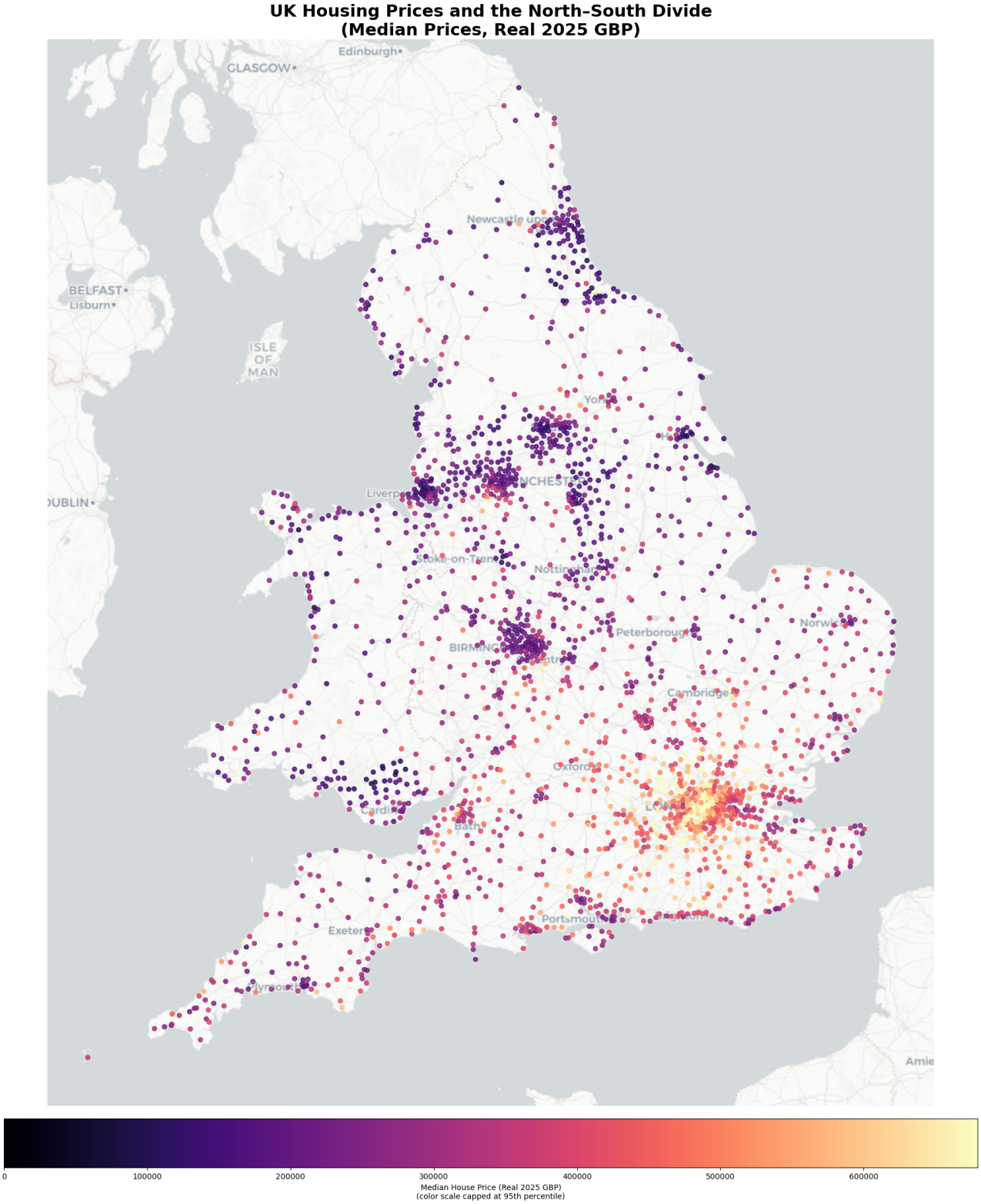

UK Housing Prices and the North–South Divide
(Median Prices, Real 2025 GBP)
Edinburgh
GLASGOW
Newcastle upon
BELFAST
Lisburn
ISLE
OF
MAN
York
Liverpool
MANCHESTER
DUBLIN
Stoke-on-Trent
Nottingham
Norwich
Peterborough
BIRMINGHAM
Cambridge
Oxford
Cardiff
Bath
LONDON
Portsmouth
Exeter
Plymouth
Amie
0
100000
200000
300000
400000
500000
600000
Median House Price (Real 2025 GBP)
(color scale capped at 95th percentile)

Figure 5

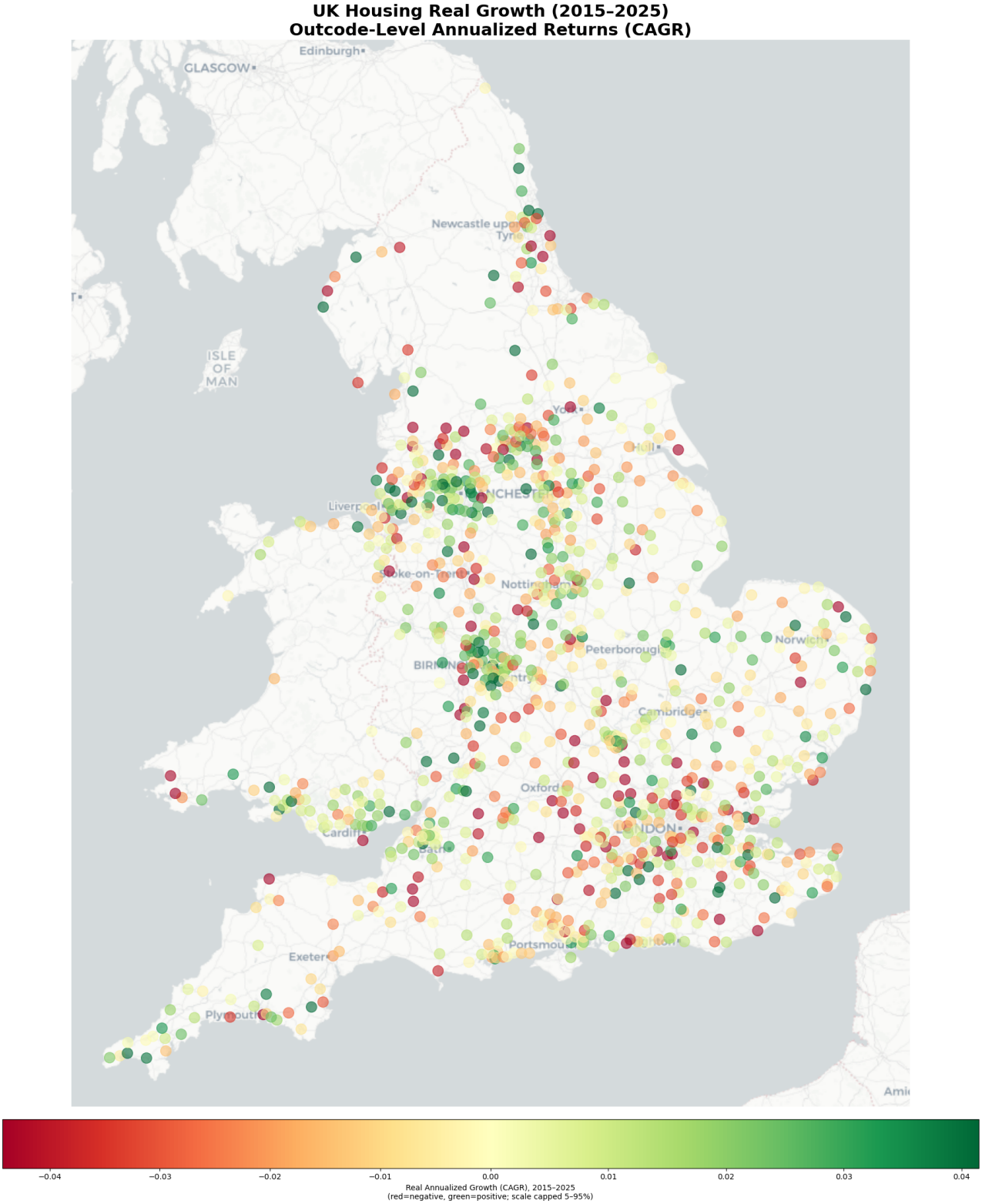

UK Housing Real Growth (2015–2025)
Outcode-Level Annualized Returns (CAGR)
Edinburgh
GLASGOW
Newcastle upon Tyne
ISLE OF MAN
York
Hull
Liverpool
Nottingham
Peterborough
Norwich
Cambridge
Oxford
LONDON
Portsmouth
Exeter
Plymouth
−0.04
−0.03
−0.02
−0.01
0.00
0.01
0.02
0.03
0.04
Real Annualized Growth (CAGR), 2015–2025
(red=negative, green=positive; scale capped 5–95%)

Figure 6

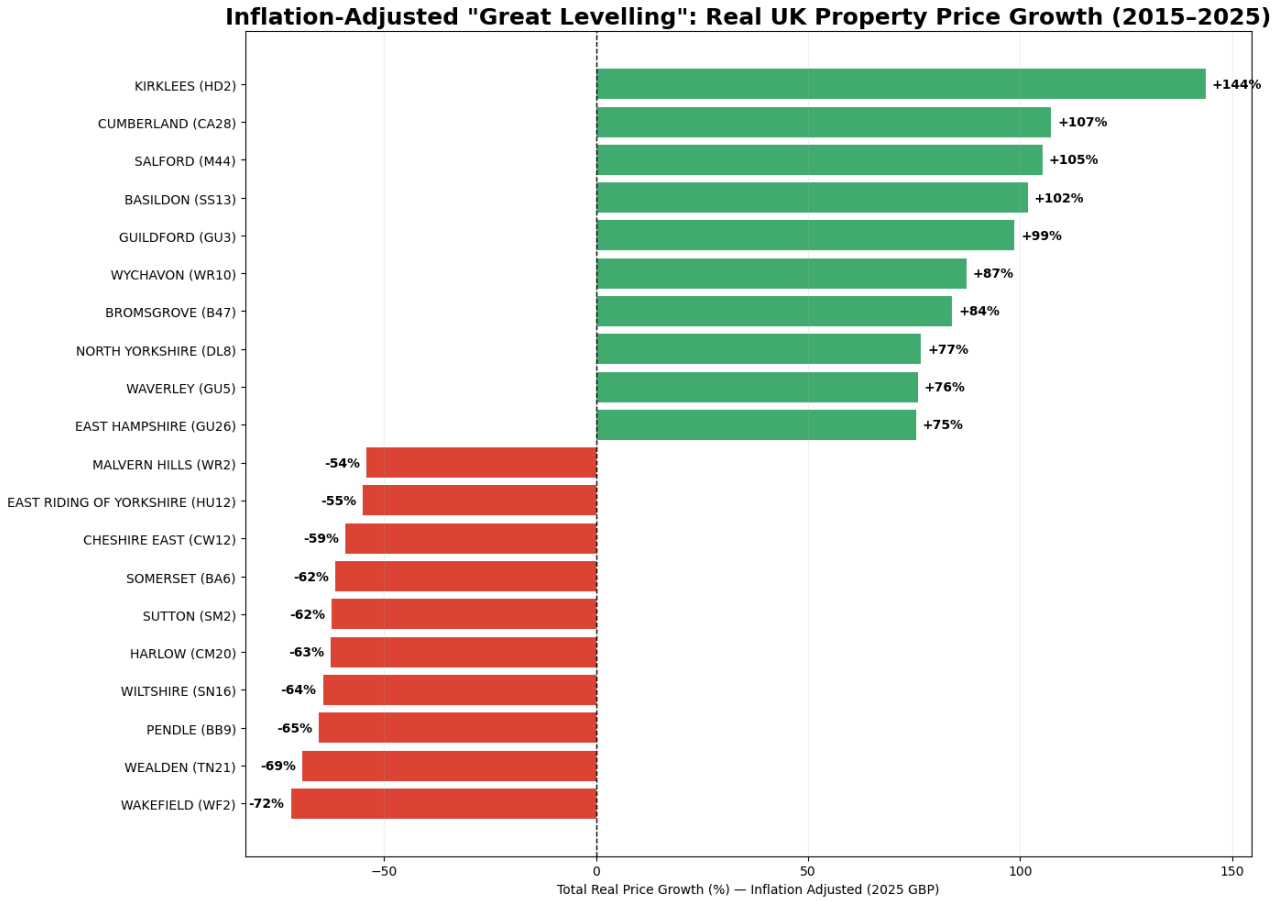

Inflation-Adjusted "Great Levelling": Real UK Property Price Growth (2015–2025)
KIRKLEES (HD2)
CUMBERLAND (CA28)
SALFORD (M44)
BASILDON (SS13)
GUILDFORD (GU3)
WYCHAVON (WR10)
BROMSGROVE (B47)
NORTH YORKSHIRE (DL8)
WAVERLEY (GU5)
EAST HAMPSHIRE (GU26)
MALVERN HILLS (WR2)
EAST RIDING OF YORKSHIRE (HU12)
CHESHIRE EAST (CW12)
SOMERSET (BA6)
SUTTON (SM2)
HARLOW (CM20)
WILTSHIRE (SN16)
PENDLE (BB9)
WEALDEN (TN21)
WAKEFIELD (WF2)
+144%
+107%
+105%
+102%
+99%
+87%
+84%
+77%
+76%
+75%
-54%
-55%
-59%
-62%
-62%
-63%
-64%
-65%
-69%
-72%
−50
0
50
100
150
Total Real Price Growth (%) — Inflation Adjusted (2025 GBP)

Figure 7