

Seven Exploratory Vignettes for the SOAR Urban Data Model: Illustrative Examples to Encourage Broader Urban Data Applications

Gareth Simons^{a,*}, Kayvan Karimi^a, Sepehr Zhand^a

^a*Space Syntax Laboratory, UCL Bartlett School of Architecture, 22 Gordon Street, London, United Kingdom*

Abstract

Large-scale urban datasets offer substantial opportunities for comparative research across cities and regions. This paper presents exploratory worked examples using the SOAR urban data model—a pan-European dataset covering 699 cities with over 100 metrics per street network node. We provide seven vignette templates demonstrating how workflows can leverage the data for data quality filtering, multi-scale analysis, access gap identification, predictive modelling, city benchmarking, typology classification, and site selection. Each vignette provides a reproducible workflow with code, results, and suggested extensions. These are illustrative starting points rather than definitive findings—templates to help researchers get started with the data and build more rigorous investigations. The purpose is to lower the barrier to entry for researchers exploring pan-European urban analysis.

Keywords: urban data models, comparative urban analysis, walkability, data quality assessment, accessibility metrics, European cities, reproducible research, POI saturation

1. Introduction

Large-scale urban datasets offer substantial opportunities for comparative research, but realising their potential requires understanding what questions the data can address and how analytical workflows might be structured. This paper provides worked examples using the SOAR (Scalable, Open, Automated, and Reproducible) urban data model, demonstrating the range of questions the data can support and lowering the barrier to entry for researchers exploring pan-European urban analysis.

This paper presents exploratory vignettes—worked examples that combine motivation, methodology, analysis, and interpretation. These are explicitly *not* definitive research findings. Each vignette is a starting point: a demonstration of what the data contains and how it might be used as part of urban analytical workflows. The vignettes are intentionally simple, using relatively standard Python workflows and related methods (correlations, Random Forests, clustering). Researchers pursuing rigorous investigations should treat these as templates to get started with the data and on which more rigorous work can be developed. Each vignette includes:

*Corresponding author
Email address: gareth.simons@ucl.ac.uk (Gareth Simons)

- Research motivation
- SOAR metrics utilised
- Analytical workflow and code
- Results
- Possible extensions

The remainder of this paper is structured as follows: Section 2 introduces the SOAR data model; Sections 3–9 present seven vignettes covering data quality, multi-scale analysis, access gaps, predictive modelling, benchmarking, typology classification, and site selection; Section 10 discusses cross-cutting themes and limitations; Section 11 concludes.

2. Introduction to SOAR

SOAR is a pan-European urban data model providing pre-computed metrics for 699 urban centres defined by Eurostat Urban Audit boundaries [1]. The dataset integrates multiple open data sources: Eurostat census grids (population, demographics), Copernicus Urban Atlas (land cover, green space, building footprints), and Overture Maps (POI locations, transport infrastructure, street networks). For each city, metrics are computed at the street network node level, yielding over 100 variables per node across multiple spatial scales (200–4,800m). The data paper [TODO: citation] provides full technical documentation of the dataset structure, processing pipeline, and metric definitions.

The metric categories include: (1) **network centrality**—closeness and betweenness at multiple radii capturing pedestrian and neighbourhood-scale accessibility; (2) **land-use accessibility**—network distances to nearest POI in 11 categories plus diversity indices; (3) **building morphology**—density, height, footprint area, and form complexity; (4) **green space proximity**—distances to parks and tree canopy; and (5) **census demographics**—population density and age structure interpolated to network nodes.

This consistent feature set across 699 cities enables diverse analytical applications:

- **Data quality filtering:** Identifying cities with saturated POI coverage for downstream analysis
- **Multi-scale analysis:** Comparing within-city patterns to between-city relationships
- **Access gap identification:** Locating underserved areas using distance-to-nearest metrics
- **Predictive modelling:** Training generalisable models across multiple urban contexts
- **Benchmarking:** Ranking cities against policy targets or peer comparisons
- **Typology classification:** Identifying recurring neighbourhood types across cities
- **Site selection:** Filtering locations by multiple criteria for planning or intervention

The vignettes that follow provide reproducible worked examples for each application type.

3. Vignette 1: Data Quality Filtering

3.1. Motivation

POI completeness varies geographically, with some regions exhibiting systematic undersaturation. This vignette applies multi-scale regression of POI counts against population densities to identify cities where crowdsourced data may be too sparse for reliable analysis, allowing researchers to filter or weight observations accordingly.

3.2. SOAR Metrics Utilised

- **POI counts:** 11 land-use categories (accommodation, active_life, arts_and_entertainment, attractions_and_activities, business_and_services, eat_and_drink, education, health_and_medical, public_services, religious, retail)
- **Census demographics:** Population counts at 1 km² grid resolution
- **Multi-scale neighbourhoods:** Local (2 km), intermediate (5 km), and large (10 km) radii

3.3. Methodology

We develop a grid-based multi-scale regression approach to assess POI data saturation across cities, comparing observed POI densities against population-based expectations to identify undersaturated areas that may indicate data incompleteness. This method provides a quantitative foundation for evaluating data quality prior to comparative urban analysis.

3.3.1. Multi-Scale Regression Workflow

The saturation assessment workflow (`paper_research/code/eg1_data_quality/`) operates at the 1 km² census grid level, enabling fine-grained spatial analysis:

1. **Grid-level aggregation:** POI counts are computed within each census grid cell. Multi-scale population neighborhoods are calculated at local, intermediate, and large radii to capture hierarchical catchment effects.
2. **Random Forest regression:** For each land-use category k , a Random Forest model is fitted in log-space:

$$\log(\text{POI}_k + 1) = f(\log(\text{pop}_{\text{local}}), \log(\text{pop}_{\text{intermediate}}), \log(\text{pop}_{\text{large}})) + \epsilon \quad (1)$$

Log transformation linearizes the power-law relationship between population and POI counts ($\text{POI} \propto \text{pop}^\beta$), yielding more normally distributed residuals suitable for z-score computation.

3. **Z-score computation:** Standardized residuals quantify deviation from expected POI counts. Negative z-scores indicate undersaturation (fewer POIs than expected); positive z-scores indicate saturation.
4. **City-level aggregation:** Grid z-scores are aggregated per city, computing mean (overall saturation level) and standard deviation (spatial variability within city).

5. **Quadrant classification:** Cities are classified by mean z-score \times variability into four quadrants: consistently undersaturated, variable undersaturated, consistently saturated, and variable saturated.

3.3.2. Quadrant Interpretation

The quadrant classification provides actionable guidance for data usage:

- **Consistently Undersaturated** (low mean, low std): Systematic data gaps; use with caution across all analyses
- **Variable Undersaturated** (low mean, high std): Partial coverage; some grid cells may be reliable
- **Consistently Saturated** (high mean, low std): Complete coverage; suitable for all analyses
- **Variable Saturated** (high mean, high std): Good overall coverage with spatial heterogeneity

3.4. Results

Analysis of 699 European urban centres reveals a core-periphery pattern in POI data saturation. Central and Western European cities (Germany, Netherlands, France, Belgium) achieve mean z-scores near zero with low spatial variability, indicating a higher likelihood of reliable data. Peripheral European regions show more prevalent undersaturation: Spanish cities (particularly Madrid satellites) average -0.6 to -1.1 , with similar patterns in Romania, Bulgaria, Poland, and southern Italy.

This pattern likely reflects differential OpenStreetMap contributor activity, varying commercial formalisation practices, and regional differences in POI aggregator coverage. The effect is pronounced for business services and retail ($R^2=0.73, 0.70$), while accommodation shows weakest predictability ($R^2=0.56$), suggesting tourism infrastructure follows different spatial logic.

Table 1 summarises Random Forest model performance by POI category. R^2 values range from 0.56 (accommodation) to 0.73 (business services), with local population scale consistently the strongest predictor for everyday amenities (retail, eat_and_drink, health_and_medical) while intermediate-scale population better predicts destination categories (attractions_and_activities).

3.5. Implications

Researchers comparing POI-derived metrics across European cities should account for systematic data quality variation. Options include: (1) restricting analyses to consistently saturated cities; (2) stratifying by saturation quadrant; or (3) applying z-score corrections in undersaturated regions.

3.6. Extensions

Potential directions: temporal trends in POI completeness; category-specific quality metrics; validation against municipal records; correlations between data quality and urban characteristics; saturation vectors as city-level features.

Table 1: Random Forest regression performance by POI category. Local, intermediate, and large columns show relative feature importance for each population scale.

Category	R ²	Local	Intermed.	Large
Business And Services	0.73	0.76	0.14	0.10
Education	0.73	0.72	0.16	0.12
Eat And Drink	0.72	0.72	0.15	0.12
Retail	0.70	0.75	0.14	0.12
Health And Medical	0.69	0.72	0.14	0.14
Public Services	0.69	0.65	0.20	0.15
Active Life	0.65	0.64	0.21	0.15
Arts And Entertainment	0.63	0.48	0.33	0.19
Attractions And Activities	0.60	0.28	0.52	0.21
Religious	0.59	0.56	0.23	0.21
Accommodation	0.56	0.41	0.33	0.26

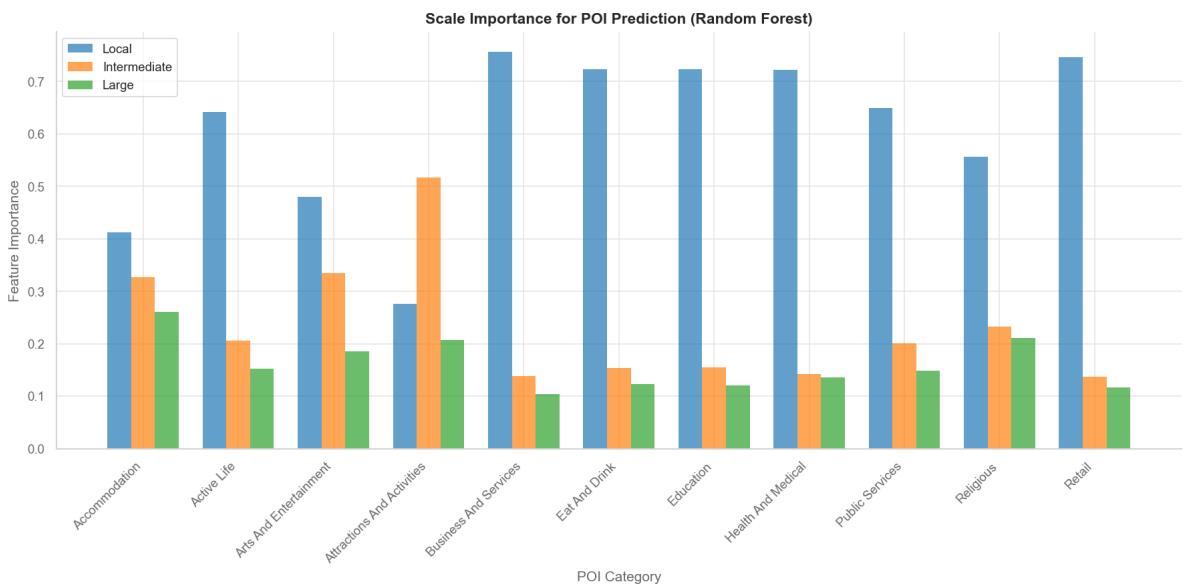


Figure 1: Feature importance showing which population scale (local, intermediate, large) best predicts POI distribution for each category.

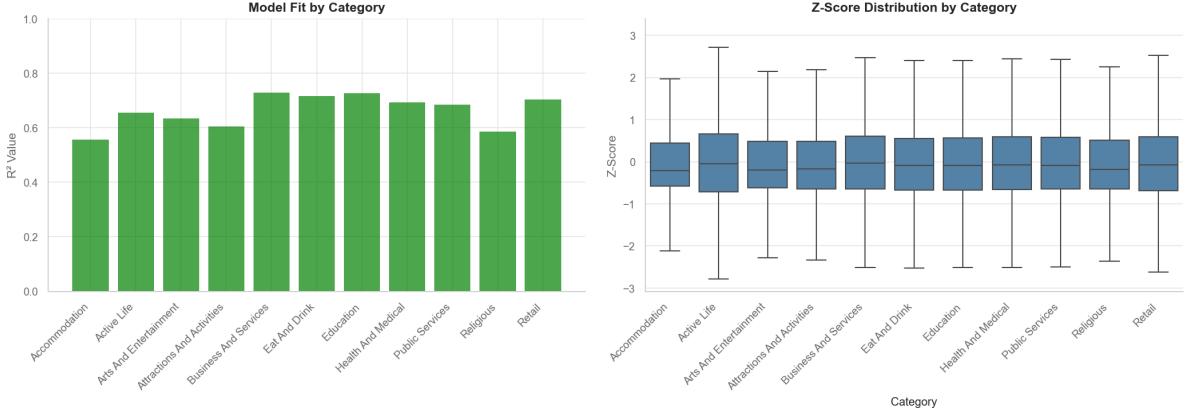


Figure 2: Exploratory data analysis. Left: Random Forest model fit (R^2) by POI category. Right: distribution of z-scores across grid cells per category.

3.7. Reproducibility

Code, outputs, and documentation: https://github.com/UCL/t2e-soar/tree/main/paper_research/code/eg1_data_quality/

4. Vignette 2: Multi-Scale Analysis

4.1. Motivation

Relationships observed within cities may differ from those observed between cities. SOAR enables both within-city analysis and city-level aggregations for cross-city comparison, and the two perspectives can yield different conclusions. This vignette explores the topic in the context of access to green spaces, examining within-city correlations between population density and green space proximity.

4.2. SOAR Metrics Utilised

- **Green space accessibility:** Network distance to nearest green block (1,600m catchment)
- **Tree canopy accessibility:** Network distance to nearest tree canopy (1,600m catchment)
- **Population density:** Persons per km² (interpolated from Eurostat 1km grid)

4.3. Methodology

For each city with ≥ 100 street network nodes, we compute Spearman rank correlations between population density and distance to green space/tree canopy as well as area of green space/tree canopy. Negative correlations indicate compact urban cores with proximate green access (“dense-and-green”), while positive correlations suggest peripheral green amenities with undersupplied centres (“dense-but-grey”). Results are visualised as diverging bar charts sorted by correlation strength, with cities categorised by the direction and magnitude of their density-green relationship.

Multiple Regression Diagnostics: Predicted vs Observed

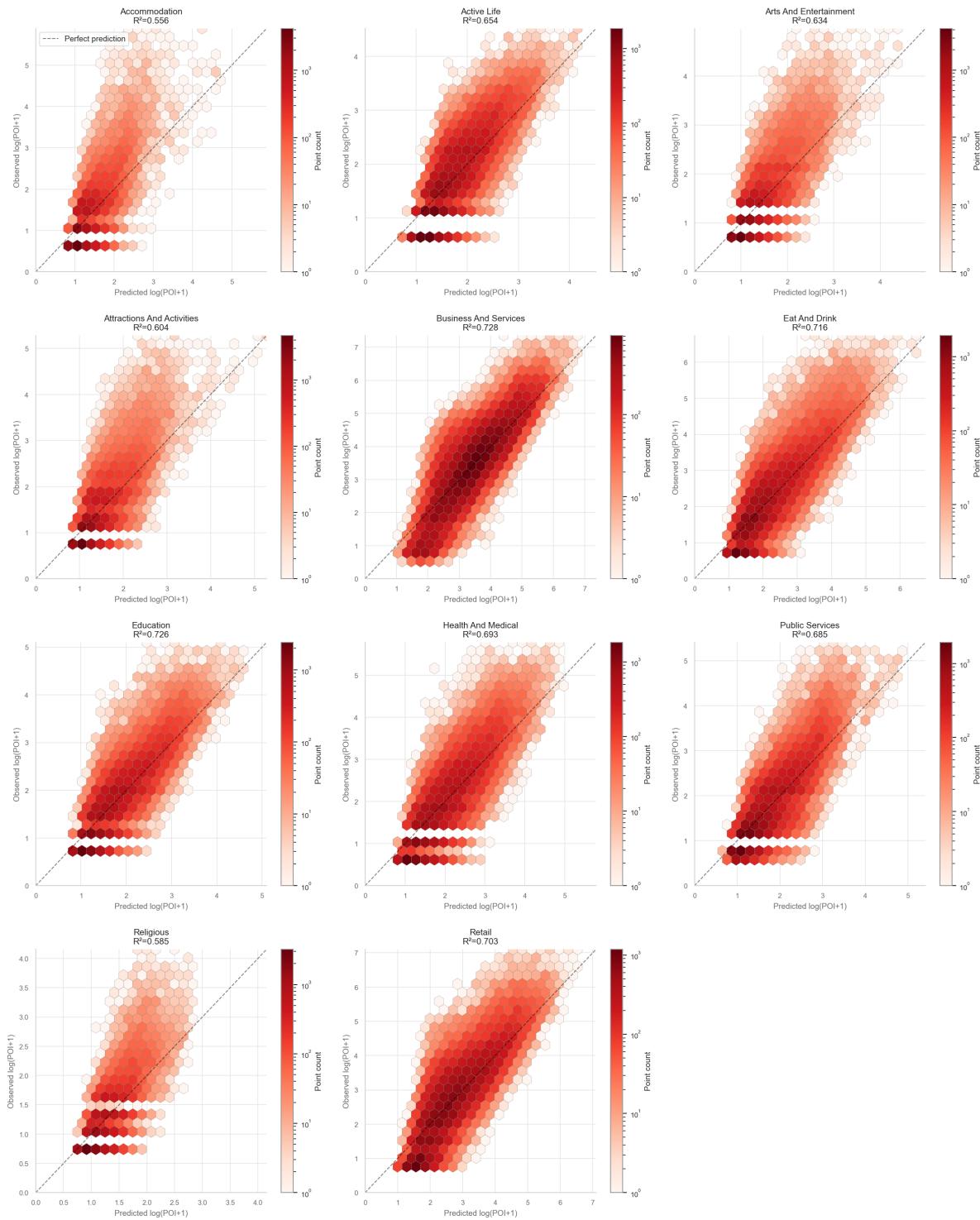


Figure 3: Regression diagnostics: predicted vs. observed POI counts (log scale) for each category.



Figure 4: City quadrant analysis. X-axis: mean z-score (negative = undersaturation). Y-axis: standard deviation (within-city variability). Quadrant colours: red = consistently undersaturated; green = consistently saturated; orange = variable undersaturated; blue = variable saturated.

4.4. Results

Analysis of 672 cities across 16.6 million street network nodes suggests a consistent within-city pattern for green blocks alongside contrasting behavior for tree canopy:

Green space (parks): 666 cities (99%) exhibit positive correlations, where denser areas face longer walks to parks. Median distance is 93.5m, with 91.0% of nodes within a 5-minute walk (400m). The strongest positive correlation (Verviers, Belgium: $\rho = 0.81$) exemplifies peripheral park placement, while rare negative outliers like Meiderich/Beeck, Germany ($\rho = -0.07$) and Płock, Poland ($\rho = -0.05$) demonstrate integrated green infrastructure in high-density zones.

Tree canopy: 569 cities (85%) show negative correlations, indicating that denser neighbourhoods have *better* tree canopy access. Median distance is 72.2m, with 91.2% within 400m. Strong negative correlations (e.g., Valdemoro, Spain: $\rho = -0.67$; Coslada, Spain: $\rho = -0.60$) suggest street tree programmes concentrated in urban cores, likely reflecting municipal maintenance priorities and sidewalk infrastructure availability.

4.5. Discussion

The central finding illustrates the importance of multi-scale analysis: within-city and between-city patterns can diverge. **Within cities**, denser areas consistently have worse park access (positive correlations with distance, negative with area). However, **between cities**, there is no systematic relationship—denser cities are not inherently worse for green access than sparser cities. A compact city can have excellent overall green provision even though its densest neighbourhoods are farther from parks than its peripheries.

This divergence matters for policy interpretation. Within-city correlations might suggest that density inherently conflicts with green access, but cross-city comparisons show this is not the case. Dense urban forms can coexist with good green infrastructure; the within-city gradient reflects land economics and planning choices rather than an insurmountable trade-off.

The correlation signs flip between distance and area metrics due to their opposite directionality (nearer is better for distance; larger is better for area), but the substantive pattern is consistent: denser areas within cities have worse park access by both measures. The tree canopy pattern reverses, though this is partly a data artefact: the Copernicus Street Tree Layer specifically captures urban tree canopy and does not extend to peripheral areas, so denser zones naturally show better coverage.

4.6. Extensions

Potential directions: green space quality metrics; breakdown by green space types; behavioural validation; green space typology effects; policy mechanism studies.

4.7. Reproducibility

Code, outputs, and documentation: https://github.com/UCL/t2e-soar/tree/main/paper_research/code/eg2_multiscale/

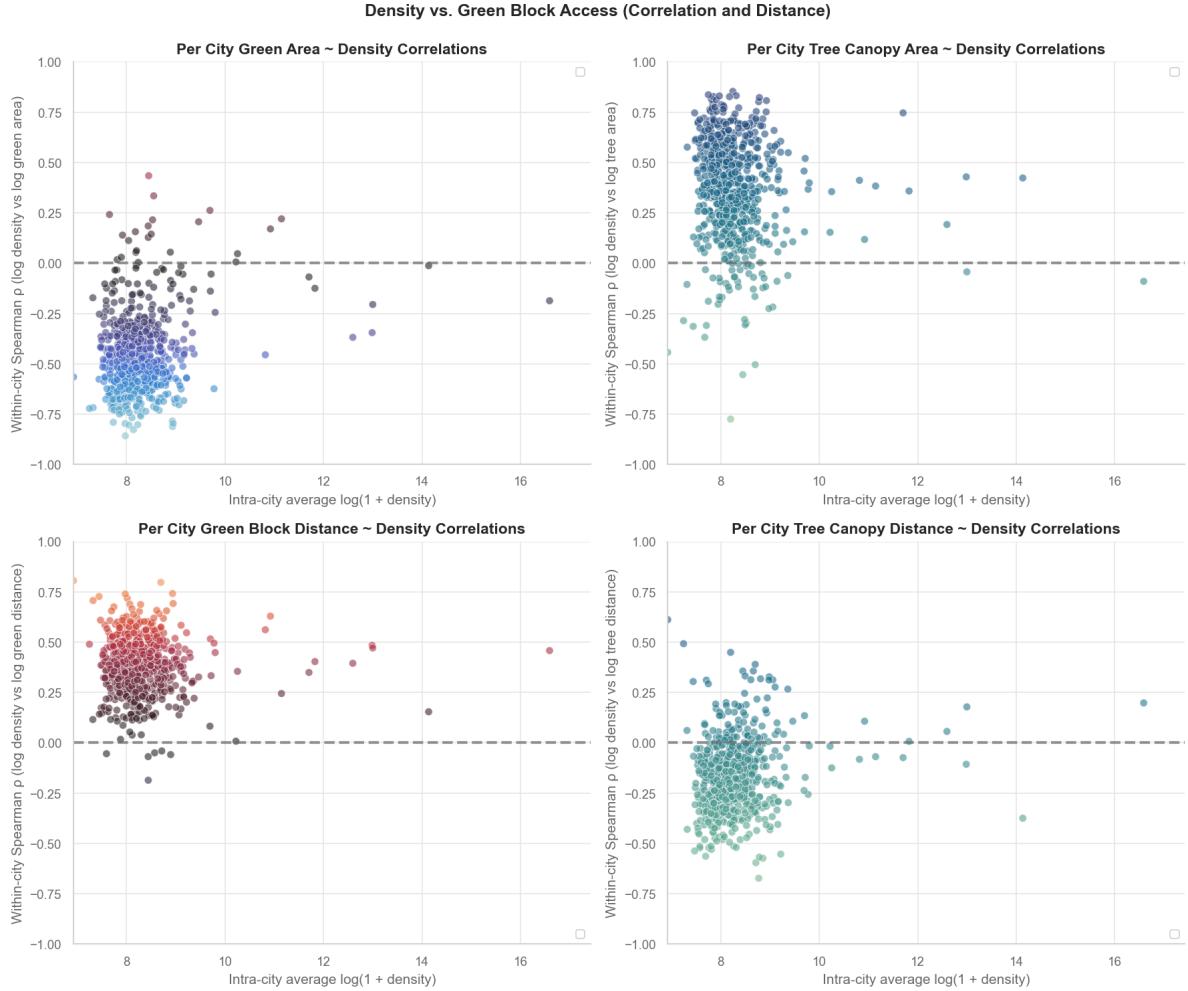


Figure 5: **Green space accessibility and tree canopy versus population density.** 2×2 grid comparing distance metrics (top row) and correlation analysis (bottom row) across 491 European cities. Left column: green blocks (parks). Right column: tree canopy. Points colored by Spearman correlation strength (blue=negative, red=positive). Top panels show no systematic relationship between city-level density and mean green distance; bottom panels confirm the absence of cross-city patterns for density-access correlations.

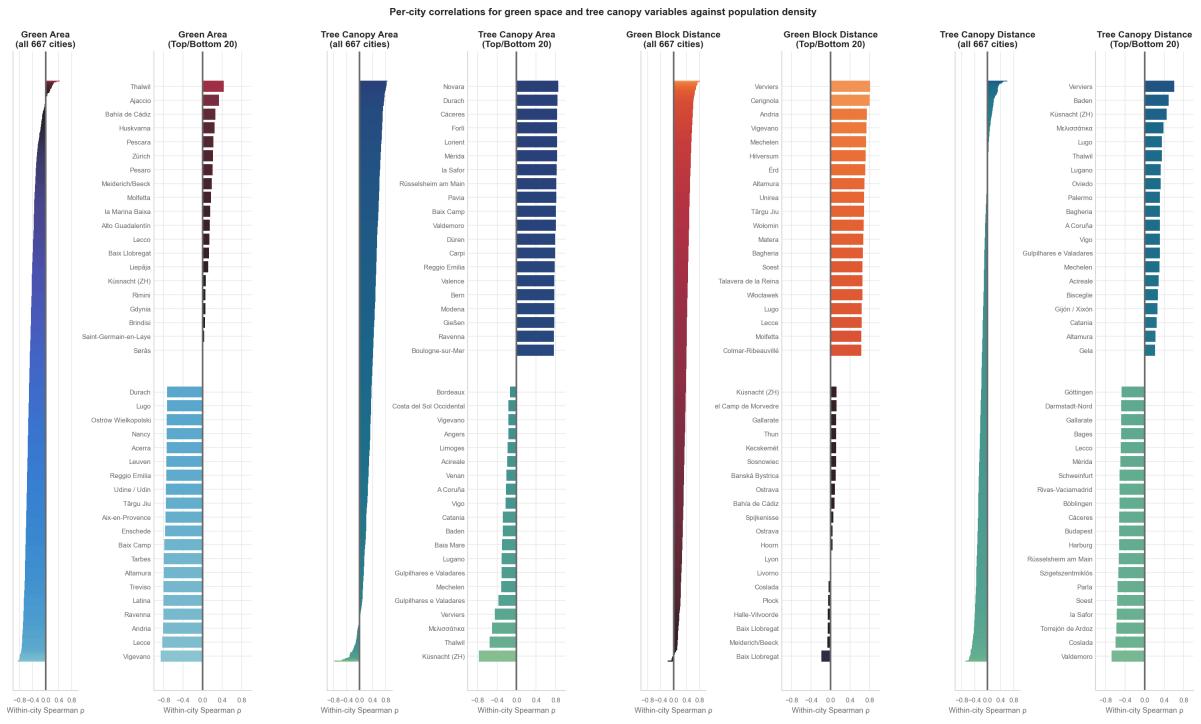


Figure 6: **Per-city correlation patterns for green space accessibility.** Diverging bar chart ranking cities by density-green correlation. For green blocks: only 4 cities show negative correlations (dense neighborhoods closer to parks); 487 cities show positive correlations (parks in peripheries). For tree canopy: 478 cities show negative correlations (street trees in urban cores); only 13 show positive correlations.

5. Vignette 3: Access Gap Identification

5.1. Motivation

Distance-to-nearest metrics can reveal locations where distances to amenities or services are greater than average or exceed targeted thresholds. This vignette demonstrates two basic methods for identifying access gaps: education (spatial equity and absolute access) and transport (demand-supply mismatches). These methods provide basic prototypes for access analysis, though more refined options may be used for detailed analysis depending on specific research questions and policy guidelines.

5.2. SOAR Metrics Utilised

- `cc_education_nearest_max_1600`: Network distance to nearest education POI (m)
- `cc_transport_nearest_max_1600`: Network distance to nearest transport stop (m)
- `cc_beta_800`: Local network centrality (demand proxy)
- `density`: Population density (demand proxy)

5.3. Methodology

We conduct two complementary analyses:

Education access: For cities with **Consistently Saturated** education POI coverage (from Vignette 1), we compute mean and median network distances to the nearest school, along with the proportion of nodes within 400m and 800m walking distance. To capture spatial equity, we calculate the P75/P25 ratio and the percentage of nodes with access worse than twice the city mean.

Transport gaps: For cities with reliable transport POI coverage, we identify locations where high demand (based on network centrality and population density) coincides with poor transport supply (long distances to nearest stop). Gap areas are classified using percentile thresholds: high-demand nodes (top 30%) with low transport supply (bottom 30%) are flagged as gap areas; those with critically low supply (bottom 15%) are flagged as critical gaps.

5.4. Results

Education access varies substantially across Europe. Cities like Küsnacht (CH) and Płock (PL) have over 65% of nodes within a 5-minute walk of a school, while cities like Como (IT) and Iserlohn (DE) have mean distances exceeding 600m. Table 2 highlights the top and bottom performers.

Equity is not guaranteed by abundance. Even in cities with good average access, pockets of disadvantage persist. The P75/P25 ratio ranges from approximately 2.4 to over 6, and in the least equitable cities, a substantial fraction of nodes are severely underserved (Table 3).

Table 2: Education access: top and bottom cities by mean distance and % within 400m.

City	Country	Mean Dist. (m)	% within 400m
Küschnacht (ZH)	CH	328	69.8
Płock	PL	376	69.4
Hoorn	NL	388	64.3
A Coruña	ES	395	61.8
Leiden	NL	396	65.0
...			
Douai	FR	614	37.1
Rönninge By	SE	616	38.1
Vereinbarte Verwaltung	DE	624	37.4
Iserlohn	DE	635	37.0
Como	IT	644	34.4

Table 3: Most and least equitable cities by P75/P25 ratio.

City	Country	P75/P25 Ratio	% Severely Underserved
Nieuwegein	NL	2.4	9.8
Almere	NL	2.5	10.9
Almelo	NL	2.5	10.1
...			
Ξάνθη	GR	6.2	18.5
Toledo	ES	6.0	19.4
Hoya de Huesca / Plana de Uesca	ES	5.7	18.3

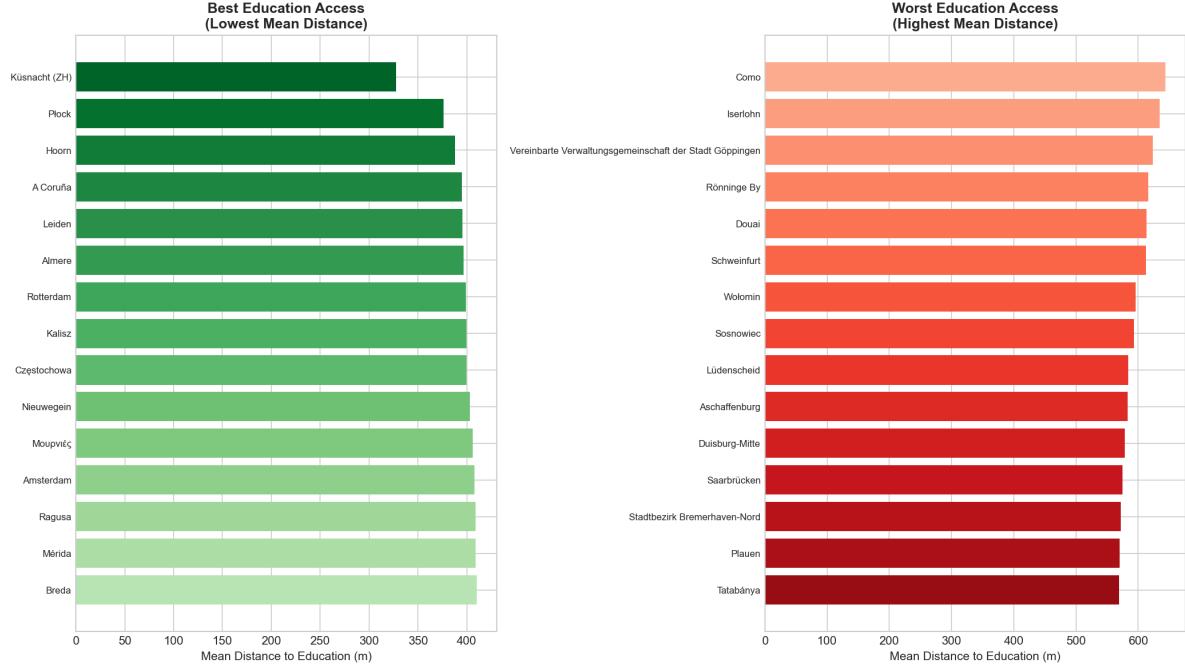


Figure 7: **Education access rankings.** Left: Top 15 cities with shortest mean distances to education facilities. Right: Bottom 15 cities with longest distances. Colour intensity reflects distance magnitude.

5.5. Discussion

Educational access varies substantially across Europe, with mean distances ranging from approximately 330m to over 640m across the analysed cities. The P75/P25 ratio—ranging from approximately 2.4 to over 6—distinguishes cities with equitable distribution from those with concentrated provision. Notably, cities with good average access may still have substantial pockets of disadvantage: even well-performing cities show 8–20% of nodes classified as severely underserved. This highlights why average access metrics can mask substantial within-city disparities.

The methodology outlined for transport gap analysis demonstrates how demand-supply mismatches can be identified by combining centrality (as a proxy for activity potential) with accessibility metrics. Locations where high network centrality and population density coincide with poor transport access represent candidates for further investigation. These are descriptive findings that could inform targeted infrastructure planning, not causal claims about infrastructure adequacy.

5.6. Extensions

Potential directions: school capacity and enrolment data; socioeconomic correlates of underserved areas; scenario modelling for new transport stops; temporal trend analysis; case studies of identified gap areas; validation against actual travel behaviour.

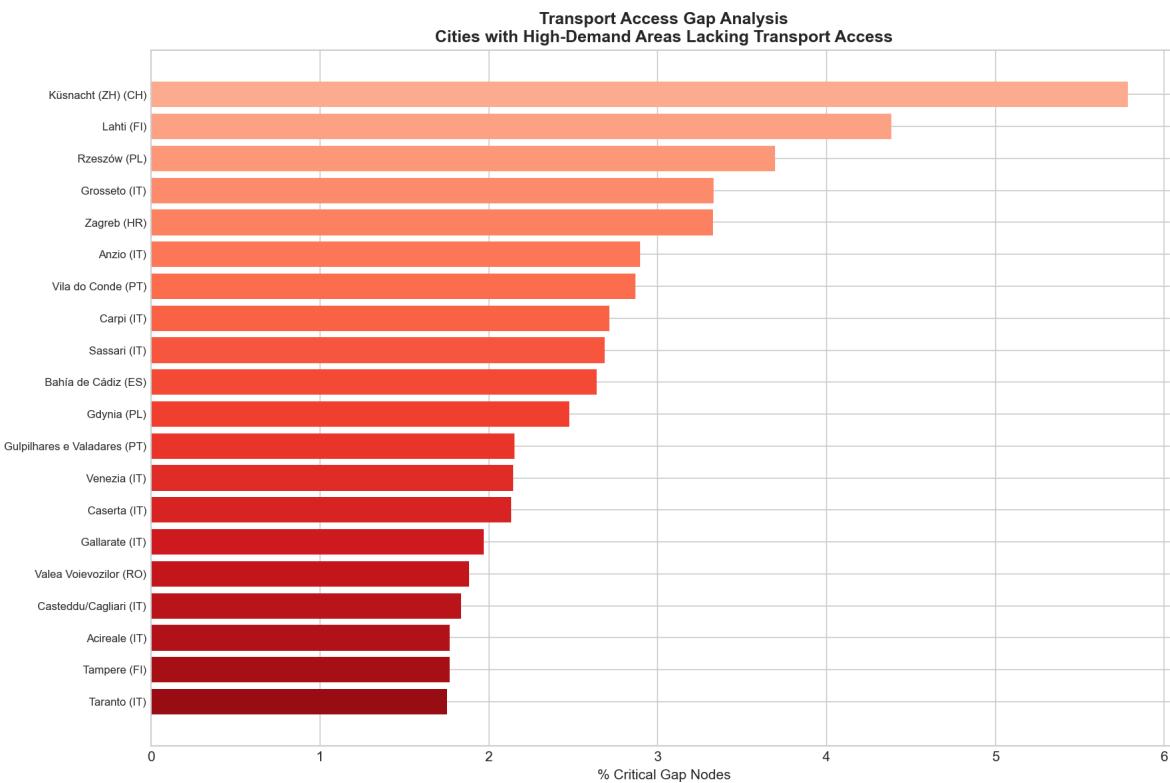


Figure 8: **Transport gap rankings.** Top 20 cities by proportion of nodes classified as critical transport gaps (high demand combined with critically low transport supply). These represent locations where network centrality and population density suggest high activity potential, yet transport infrastructure remains distant.

5.7. Reproducibility

Code, outputs, and documentation: https://github.com/UCL/t2e-soar/tree/main/paper_research/code/eg3_access_gaps/

6. Vignette 4: Predictive Modelling

6.1. Motivation

Large-scale datasets from multiple cities can enable training of generalisable models. This vignette uses network centrality and population density to predict levels of eating and drinking establishments as well as business and services intensities. The consistent feature set across cities can support transfer learning or pooled models.

6.2. SOAR Metrics Utilised

- **Network centrality:** Closeness centrality at 400m, 800m, 1,200m, 1,600m, and 4,800m radii
- **POI counts:** Eat & drink establishments (400m); Business & services establishments (400m)
- **Census variables:** Population density, age structure (under 15, 15–64, 65+), and employment ratio (interpolated from Eurostat 1km grid)
- **Saturation classification:** From Vignette 1 (cities classified as **Consistently Saturated**)

6.3. Methodology

We develop an Extra Trees regression approach to predict node-level POI counts based on multi-scale network centrality and census demographics. Analysis is restricted to cities with **Consistently Saturated** POI coverage for both eat & drink and business & services categories (as identified in Vignette 1), yielding 27 cities and 694,527 street network nodes.

6.3.1. Model Training Workflow

The workflow (`paper_research/code/eg4_prediction/`) operates at the street network node level:

1. **Data preparation:** Extract closeness centrality metrics at five spatial scales (400–4,800m) and POI counts for eat & drink and business & services categories. Filter to cities with consistent saturation to avoid training on incomplete data.
2. **Feature engineering:** Combine 5 centrality features with 5 census features (density, age groups, employment). Log-transform all features and targets: $\log(x + 1)$.
3. **Train-test split:** Randomly split nodes into 90% training and 10% testing sets, stratified by city to ensure geographic representation.
4. **Extra Trees training:** Fit separate models for eat & drink and business & services. Hyperparameters: 100 estimators, maximum depth of 20.

- 5. Per-city evaluation:** Compute R^2 , MAE, and RMSE separately for each city to assess how well the centrality-amenity relationship generalises across urban contexts.

6.4. Results

Extra Trees models achieve strong predictive performance on held-out test data: $R^2 = 0.731$ for both eat & drink and business & services. Per-city R^2 values reveal how consistently the centrality-amenity relationship holds across different urban forms.

Eat & drink: Median city $R^2 = 0.709$, with 96.3% of cities achieving $R^2 > 0.5$. Italian and Spanish cities dominate the best-predicted list (Table 4), suggesting that network structure strongly determines hospitality location in these contexts. Heerlen ($R^2 = 0.44$) shows poorest fit, potentially indicating amenity distributions driven by factors beyond network accessibility.

Business & services: Median city $R^2 = 0.721$, with 100% of cities exceeding $R^2 > 0.5$. The pattern mirrors eat & drink, with Italian cities showing strongest network-amenity alignment (Table 5).

Table 4: Top 10 cities by R^2 for Eat & Drink (400m).

City	R^2	MAE	RMSE	Nodes
Bari	0.857	0.430	0.584	14,855
la Safor	0.821	0.425	0.603	5,435
Ragusa	0.819	0.375	0.512	6,693
la Plana Alta	0.816	0.448	0.628	10,265
Alessandria	0.809	0.458	0.590	5,417
...				
Heerlen	0.440	0.488	0.611	33,426
Pordenone / Pordenon	0.585	0.498	0.655	10,823
Gallarate	0.586	0.479	0.606	17,720
Bergamo	0.590	0.496	0.638	41,833

Feature importance reveals that intermediate-scale closeness centrality (1,200–1,600m) dominates predictions for both categories (Table 6), consistent with pedestrian catchment theory—amenities locate where they can serve walkable neighbourhoods rather than immediate adjacency (400m) or regional accessibility (4,800m). Census features contribute substantially, with employment ratio and population density ranking among the top predictors.

6.5. Discussion

Models achieve $R^2 > 0.7$ on held-out data, indicating that network centrality and census features correlate with amenity counts. Intermediate-scale centrality (1,200–1,600m) dominates feature importance.

Table 5: Top 10 cities by R^2 for Business & Services (400m).

City	R^2	MAE	RMSE	Nodes
Ragusa	0.865	0.405	0.571	6,693
Cremona	0.816	0.497	0.642	6,138
Bari	0.809	0.549	0.725	14,855
Ia Safor	0.804	0.550	0.759	5,435
Alessandria	0.798	0.535	0.711	5,417
...				
Modena	0.564	0.679	0.892	14,782
Heerlen	0.594	0.566	0.726	33,426
Pordenone / Pordenon	0.622	0.601	0.771	10,823
Prato	0.637	0.639	0.796	20,100

Table 6: Feature importance for Extra Trees regression models.

Eat & Drink		Business & Services	
Feature	Importance	Feature	Importance
Cc Beta 1200	0.173	Cc Beta 1200	0.217
Cc Beta 1600	0.150	Cc Beta 1600	0.170
Y 1564	0.126	Cc Beta 800	0.146
Cc Beta 800	0.117	Y 1564	0.097
Y Ge65	0.092	Y Lt15	0.068
Density	0.082	Emp	0.065
Emp	0.080	Y Ge65	0.064
Y Lt15	0.072	Density	0.056
Cc Beta 4800	0.049	Cc Beta 4800	0.044
Cc Beta 400	0.032	Cc Beta 400	0.040

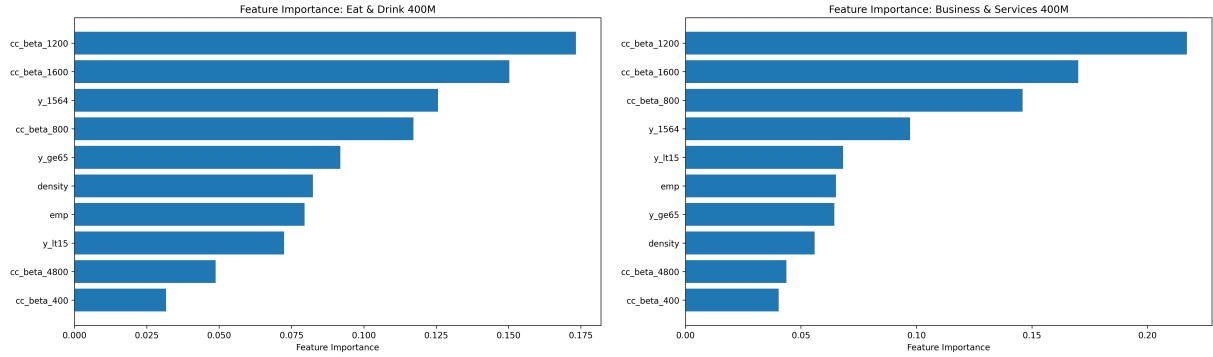


Figure 9: **Feature importance for amenity prediction models.** Left: Eat & drink. Right: Business & services. Intermediate-scale closeness centrality (1,200–1,600m) dominates both models, consistent with pedestrian catchment theory.

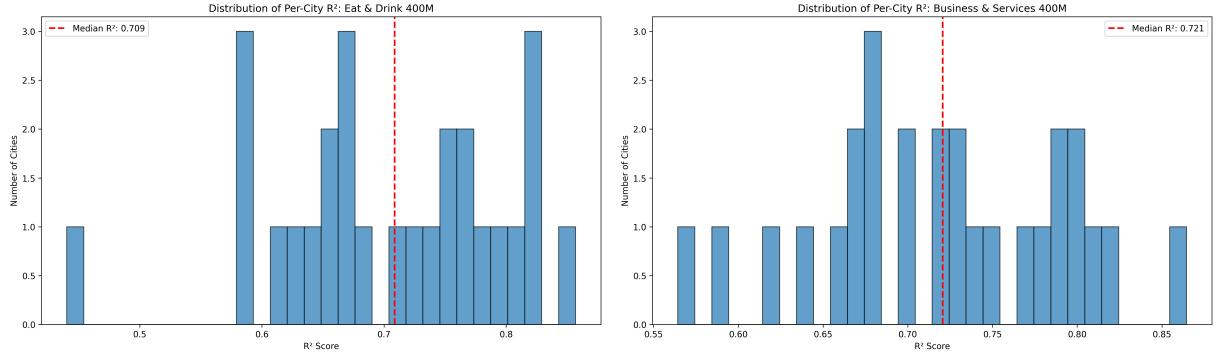


Figure 10: **Distribution of per-city R² scores.** Left: Eat & drink. Right: Business & services. Most cities cluster at high R² values, with a few outliers showing weaker network-amenity relationships.

These are correlational findings. The models do not establish that centrality causes commercial location; both could be driven by underlying factors (land values, zoning, historical development). Per-city variation in model performance suggests the relationship differs across urban contexts.

6.6. Extensions

Potential directions: additional explanatory variables (land values, zoning); residual analysis to identify underserved areas; temporal stability; cross-category comparisons; city-specific models.

6.7. Reproducibility

Code, outputs, and documentation: https://github.com/UCL/t2e-soar/tree/main/paper_research/code/eg4_prediction/

7. Vignette 5: Benchmarking

7.1. Motivation

Cities can be ranked on standardised metrics enabling comparative assessment against peer cities or policy targets. This vignette ranks cities by walkable access to amenities and services, measuring the proportion of locations with access to all of SOAR's major land-use categories within an 800m (approximately 10-minute) pedestrian-scale walking distance.

7.2. SOAR Metrics Utilised

- **POI network distances:** cc_{category}_nearest_max_1600 for 10 essential categories
- **POI categories assessed:** active_life, arts_and_entertainment, attractions_and_activities, business_and_services, eat_and_drink, education, health_and_medical, public_services, religious, retail (accommodation excluded as non-essential for daily access)
- **Saturation classification:** From Vignette 1 (cities with Consistently Saturated or Variable Saturated combined POI coverage)

7.3. Methodology

We develop a node-level completeness scoring approach to benchmark cities against walkable city ideals. Analysis is restricted to cities with reliable POI coverage across multiple categories (as identified in Vignette 1’s between-category quadrant classification).

7.3.1. Completeness Scoring Workflow

The workflow ([paper_research/code/eg5_benchmarking/](#)) operates at the street network node level:

1. **City filtering:** Select cities classified as **Consistently Saturated** or **Variable Saturated** in the between-category quadrant analysis (Vignette 1), ensuring reliable POI data across multiple service types.
2. **Distance extraction:** For each node, extract network distances to nearest POI in each of 10 categories using SOAR’s pre-computed `cc_{category}_nearest_max_1600` metrics.
3. **Per-node completeness:** Count how many of the 10 categories are accessible within 800m (approximately 10 minutes at 80m/min walking speed). A node with “full access” reaches all 10 categories within this threshold.
4. **City-level aggregation:** Compute the percentage of nodes with full access (all 10 categories within 800m), mean completeness score (average number of accessible categories divided by 10), and per-category access rates.
5. **Bottleneck identification:** Identify which POI categories most frequently limit full access, revealing systematic infrastructure gaps.

7.3.2. Threshold Rationale

The 800m threshold operationalises a 10-minute walk assuming approximately 80m/min walking speed, consistent with pedestrian planning standards for local accessibility. This threshold captures immediate neighbourhood-scale daily walking catchments and represents a more stringent standard than traditional 15-minute city concepts, better distinguishing genuinely walkable local centres from areas requiring longer trips.

7.4. Results

Analysis of cities with reliable POI coverage suggests substantial variation in 10-minute walkable access completeness across European urban centres.

Most cities achieve substantial local completeness, though universal coverage within 800m remains challenging. The median city has nodes with access to all 10 POI categories within 800m, though the precise percentages vary by urban form and service distribution patterns. The gap between median

and best performers indicates room for improvement—even cities with good overall accessibility may have neighbourhoods lacking coverage across all service types at this more stringent distance threshold.

Top-performing cities in this analysis demonstrate that compact urban form and mixed-use development can deliver high levels of 10-minute access. Table 7 shows cities with highest percentages of fully-accessible nodes. These cities share characteristics of pedestrian-oriented development, fine-grained land-use mixing, and comprehensive local service provision.

Table 7: Top 10 cities by 10-minute walkable access completeness.

City	Country	% Full Access	Mean Completeness
Venezia	IT	94.5	0.991
Bahía de Cádiz	ES	93.5	0.990
Siracusa	IT	92.9	0.983
Trapani	IT	90.7	0.983
Amersfoort	NL	90.1	0.984
Pesaro	IT	89.2	0.980
Verona	IT	88.6	0.979
Gouda	NL	88.5	0.978
Firenze	IT	88.5	0.978
Ancona	IT	88.1	0.977

Bottom-performing cities in this analysis exhibit either sprawling urban form, car-oriented development patterns, or systematic gaps in specific service categories. Table 8 highlights cities with lowest completeness scores.

Table 8: Bottom 10 cities by 10-minute walkable access completeness.

City	Country	% Full Access	Mean Completeness
Gulpilhares e Valadares	PT	48.5	0.918
Gulpilhares e Valadares	PT	57.3	0.932
Gulpilhares e Valadares	PT	58.6	0.939
Como	IT	59.1	0.929
Hradec Králové	CZ	61.4	0.950
None	None	61.9	0.943
Roosendaal	NL	62.0	0.952
Roanne	FR	62.1	0.950
Thalheim bei Wels	AT	62.2	0.939
Toulon	FR	62.6	0.943

National differences are also evident. Table 9 aggregates city scores by country, showing that Southern European countries (Spain, Romania, Croatia, Italy) tend to have higher proportions of fully accessible

nodes compared to Nordic countries (Sweden, Denmark, Norway, Finland), likely reflecting differences in urban density and zoning traditions.

Table 9: Country rankings by mean 10-minute walkable access completeness score.

Country	Cities	Mean % Full Access	Median % Full Access
ES	11	82.9	81.9
RO	7	82.7	83.0
HR	5	82.4	82.5
IT	62	81.7	82.8
GR	4	79.9	81.1
...			
SE	7	72.4	71.5
DK	4	71.9	70.6
NO	3	70.8	72.6
PT	8	70.5	77.0
FI	4	66.9	65.4

Bottleneck categories reveal which services most frequently limit full 10-minute access. Table 10 ranks POI categories by mean access rate, identifying systematic gaps that planners could target for intervention.

Table 10: POI categories ranked by mean 10-minute access rate (bottleneck analysis).

Category	Mean Access Rate (%)
Religious	87.4
Arts and Entertainment	92.5
Attractions and Activities	94.2
Education	97.3
Health and Medical	97.7
Public Services	97.9
Active Life	98.2
Eat and Drink	98.8
Retail	99.5
Business and Services	99.9

7.5. Discussion

The analysis reveals that full 10-minute access is typically limited by one or two categories rather than general deficits across all services. Certain categories (such as religious facilities and arts & entertainment venues) consistently emerge as bottlenecks, while others (business & services, retail) achieve near-universal

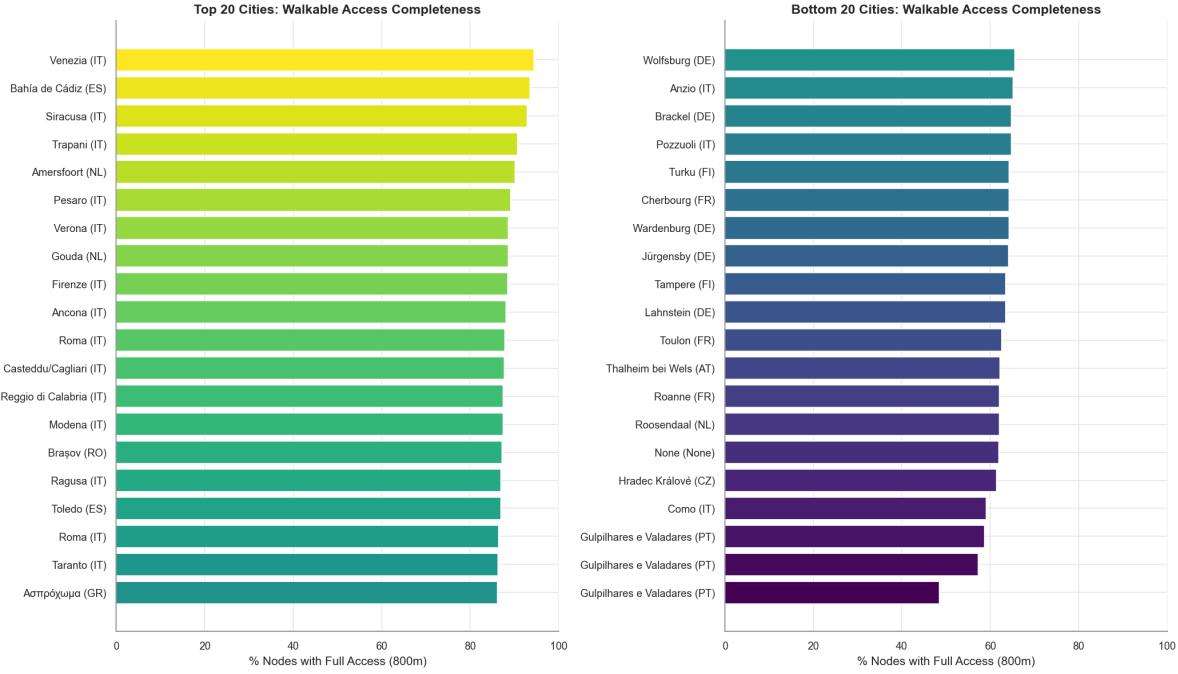


Figure 11: **10-minute walkable access completeness ranking.** Left: Top 20 cities with highest proportion of nodes achieving full 800m access. Right: Bottom 20 cities. Colour intensity reflects completeness score.

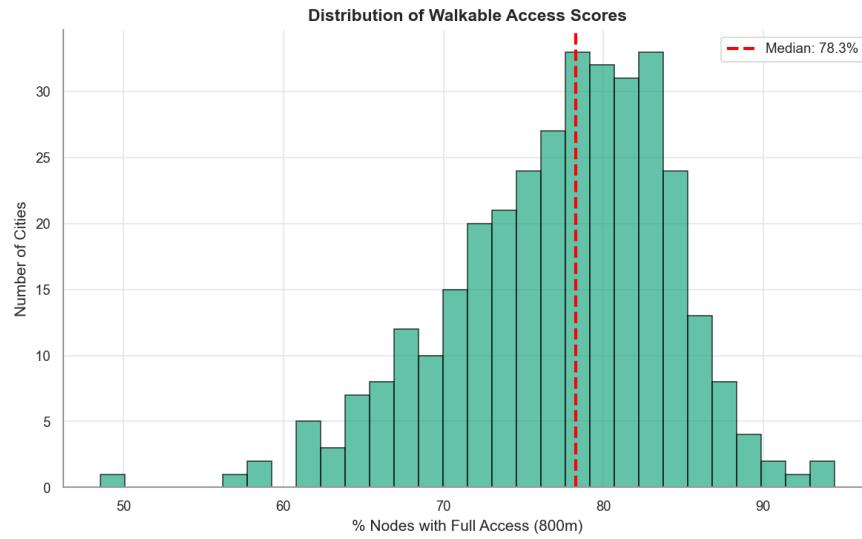


Figure 12: **Distribution of 10-minute walkable access scores across European cities.** Histogram showing the percentage of nodes with full access to all 10 POI categories within 800m. Red dashed line indicates median.

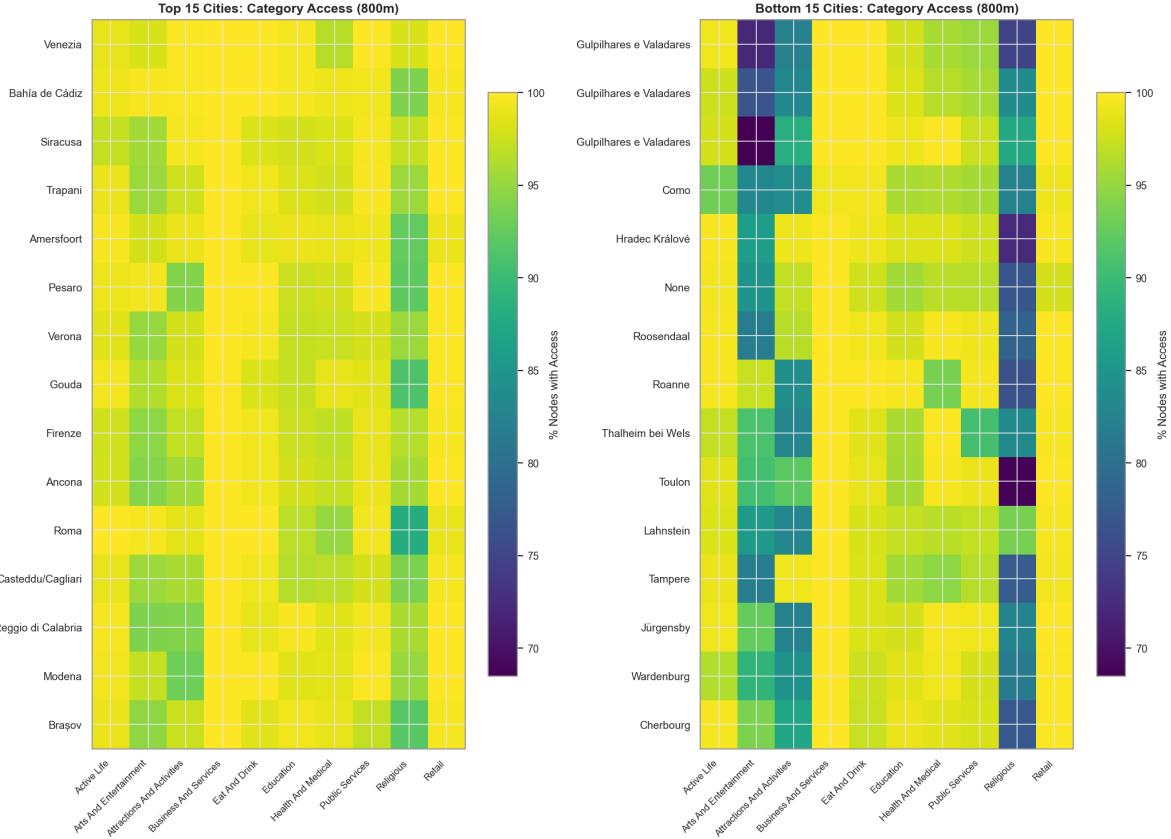


Figure 13: **Per-category access rates for top and bottom cities.** Heatmap showing which POI categories are most/least accessible within 800m. Columns represent categories; rows represent cities. Darker colours indicate lower access rates, revealing category-specific bottlenecks.

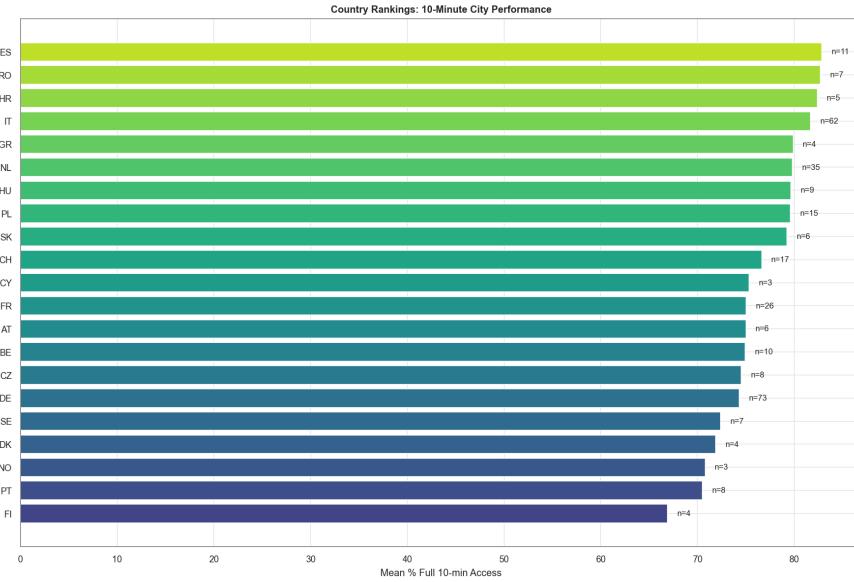


Figure 14: **Country rankings by 10-minute walkable access performance.** Bar chart showing mean completeness scores aggregated by country. Southern European countries (Spain, Italy, Croatia) show higher average scores than Nordic countries, likely reflecting differences in urban density and planning traditions.

coverage. This pattern suggests that commercial activities locate to maximise accessibility, while specialised cultural and religious facilities serve larger catchments.

Top-performing cities share characteristics of compact historic cores with fine-grained land-use mixing. Bottom performers tend to be smaller cities or suburban agglomerations where dispersed development patterns limit walkable service provision.

National patterns are evident: Southern European countries (Italy, Spain, Croatia) show higher median completeness than Nordic countries (Sweden, Denmark, Finland), likely reflecting historical urban density, zoning traditions, and car-oriented versus pedestrian-oriented development. However, these patterns should be interpreted cautiously—POI data completeness (Vignette 1) may also vary systematically by region.

This metric is sensitive to category definitions and distance thresholds; different operationalisations would yield different rankings. The 800m threshold operationalises one interpretation of walkable local access, but actual travel times depend on terrain, pedestrian infrastructure quality, and individual mobility.

7.6. Extensions

Potential directions: category weighting by use frequency or essentiality; actual routing travel times accounting for pedestrian infrastructure; opening hours analysis to capture temporal accessibility; cycling vs walking comparisons; socioeconomic correlates of access gaps; scenario modelling for targeted infrastructure investment.

7.7. Reproducibility

Code, outputs, and documentation: https://github.com/UCL/t2e-soar/tree/main/paper_research/code/eg5_benchmarking/

8. Vignette 6: Typology Classification

8.1. Motivation

Clustering algorithms applied to street-level features can identify recurring neighbourhood types that transcend administrative boundaries, revealing morphological similarities across different urban contexts. This vignette applies clustering to identify urban morphological forms and characterises cities by their mix of types.

8.2. SOAR Metrics Utilised

Morphology features (8 variables at 200m scale, aggregated to nearest adjacent street segments):

- **Density:** Building count, Block count
- **Verticality:** Mean building height (median), Height variation (MAD)
- **Scale:** Building footprint area (median)
- **Form complexity:** Fractal dimension (median)
- **Aggregation:** Block coverage ratio (median), Shared walls ratio (median)

External characterisation variables:

- Population density (persons/km²)
- Street network density (street segment count at 1,200m)
- Land-use diversity (Hill number $q = 0$ at 200m)

8.3. Methodology

We develop a node-level clustering approach that identifies morphological neighbourhood types across all European cities, then profiles each city and country by their distribution across these types.

8.3.1. Clustering Workflow

The workflow (`paper_research/code/eg6_typology/`) operates at the street network node level:

1. **Data preparation:** Extract 8 morphology features at 200m scale for all nodes. Apply log transformation to normalise skewed distributions, then standardise (z-score) across all nodes.
2. **BIRCH clustering:** Apply Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) to identify $k = 8$ morphological neighbourhood types. BIRCH provides O(n) complexity suitable for millions of nodes, building a CF-Tree structure with final agglomerative clustering (Ward linkage) on subclusters.

3. **External characterisation:** For each cluster, compute mean population density, street network density, and land-use diversity to interpret what each morphological type represents in functional terms.
4. **City/country profiling:** Compute the proportion of nodes in each cluster for every city and country, creating compositional vectors that characterise urban form distributions.
5. **Contrasting cluster visualisation:** Identify clusters with highest/lowest density and mixed-use values; plot cities by their proportions in these contrasting types to reveal national patterns.

8.3.2. Feature Selection Rationale

The 8 features capture complementary dimensions of urban form: density metrics (building and block counts) describe how much is built; verticality metrics (height median and variation) describe the skyline profile; scale (building area) captures footprint size; form complexity (fractal dimension) distinguishes regular from irregular building shapes; and aggregation metrics (block coverage, shared walls) describe how buildings relate to their parcels and neighbours. This parsimonious set avoids redundancy while covering the key morphological dimensions identified in urban morphometrics literature.

8.4. Results

Analysis of street network nodes across European cities identifies 8 distinct morphological neighbourhood types with interpretable characteristics.

The resulting cluster profiles suggest distinct urban fabrics. Figure 15 shows standardised feature profiles for each cluster, and Table 11 summarises their key characteristics. The clusters span a range of urban forms: high-density clusters (Clusters 1, 4, 6, and 8, all exceeding 12,000 persons/km²) represent dense urban cores with varying morphological characteristics, while low-density clusters (Clusters 2, 3, 5, and 7, ranging from 4,151 to 7,475 persons/km²) capture suburban and peripheral development patterns. Clusters 3 (30.8% of nodes) and 5 (29.2%) dominate the sampled network, representing the most common morphological forms across European cities.

Table 11: Characteristics of the 8 morphological clusters. Pop Density in persons/km², Network Density in nodes/km².

Cluster	Nodes	% Total	Pop Density	Network Density	Mixed Use
1	15,652	7.8	13776	1372.9	7.00
2	21,297	10.6	5764	858.5	2.48
3	61,610	30.8	5461	1025.7	3.46
4	2,845	1.4	12988	1056.7	3.41
5	58,474	29.2	4151	799.4	2.11
6	17,993	9.0	12118	1261.2	4.99
7	11,020	5.5	7475	1074.5	4.06
8	11,109	5.6	13160	1058.3	4.61

External metrics appear to support cluster interpretations. Ranking clusters by population density, street network density, and land-use diversity (Figure 16) confirms that morphological clustering captures functional urban differences. High-density clusters (Clusters 1, 4, 6, 8, ranging from 12,118 to 13,776 persons/km²) exhibit approximately 3× higher population density than the lowest-density cluster (Cluster 5 at 4,151 persons/km²), with corresponding differences in street network connectivity (ranging from 799 to 1,373 nodes/km²) and mixed-use intensity (Hill $q = 0$ from 2.11 to 7.00).

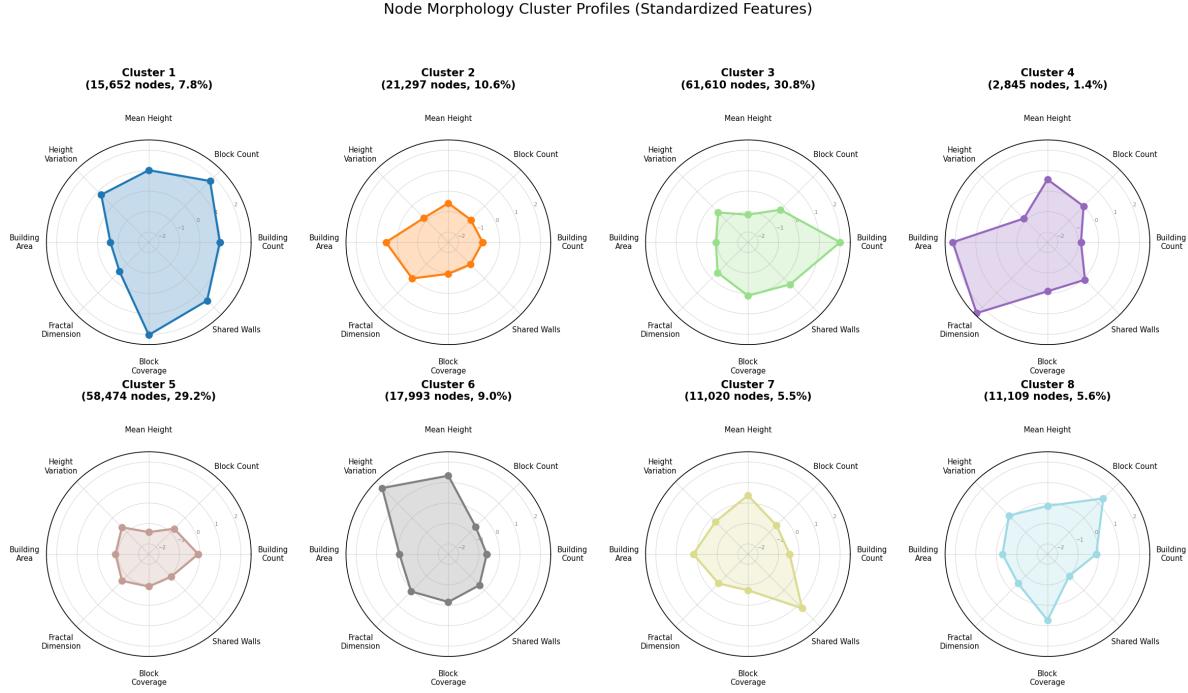


Figure 15: **Morphological cluster profiles.** Radar plots showing standardised feature values for each of 8 clusters. Axes represent: Building Count, Block Count, Mean Height, Height Variation, Building Area, Fractal Dimension, Block Coverage, and Shared Walls. Values are z-scores (0 = mean across all clusters).

National patterns appear to emerge in city compositions. When cities are plotted by their proportions in contrasting cluster types (Figure 17), clear geographic patterns appear. Northwestern European cities show strong concentrations in specific cluster types—Netherlands (73.0% in Cluster 3), Belgium (65.3% in Cluster 3), and Ireland (64.3% in Cluster 3)—indicating morphological homogeneity within these countries. Southern European cities show more varied distributions across cluster types. Nordic countries (Denmark, Finland) concentrate in Cluster 5 (51–61% of nodes), while Greek cities show high proportions in Cluster 1 (35.5%).

Countries cluster by morphological similarity. Hierarchical clustering of country composition vectors (Figure 18) reveals regional groupings: Western European countries (Netherlands, Belgium, Germany) share similar morphological profiles distinct from Southern European (Italy, Spain) and Eastern European (Romania, Poland, Bulgaria) groupings. This suggests that planning traditions, historical development pat-

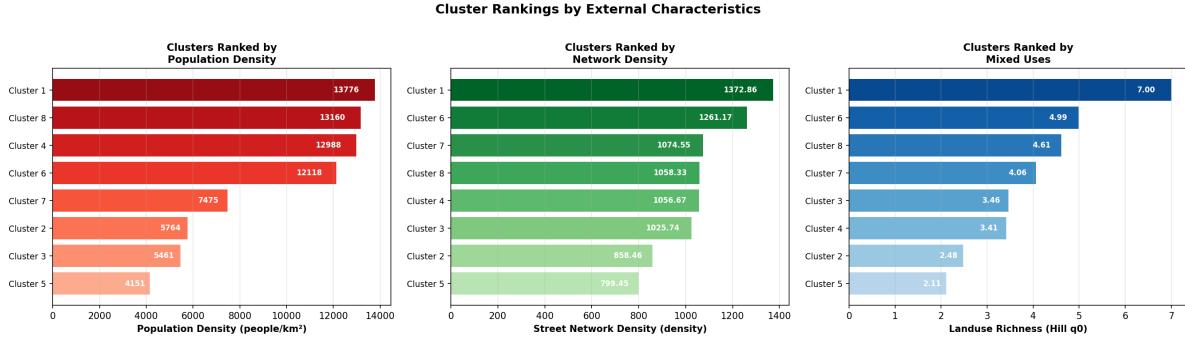


Figure 16: **Clusters ranked by external characteristics.** Horizontal bar charts ranking the 8 morphological clusters by mean population density (left), street network density (centre), and land-use diversity (right). Different clusters rank highest for different metrics: Cluster 1 has highest network density (1,373) and mixed-use diversity (7.00), while Cluster 8 has highest population density (13,160).

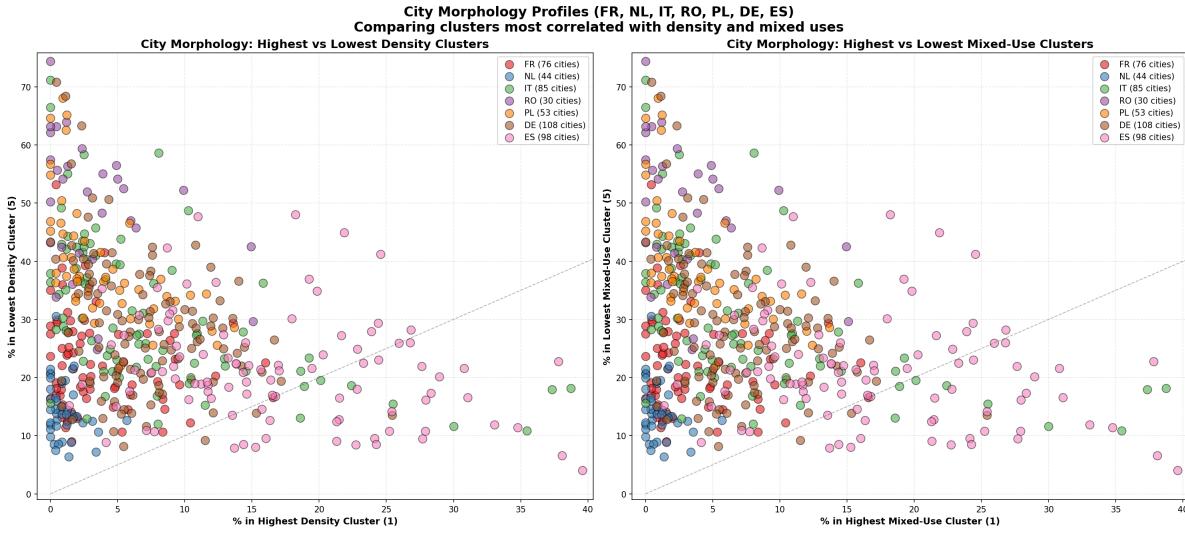


Figure 17: **City morphology profiles by contrasting cluster proportions.** Left: Cities plotted by percentage of nodes in highest-density vs. lowest-density clusters. Right: Cities plotted by percentage in highest vs. lowest mixed-use clusters. Points coloured by country. Italian cities (green) cluster toward dense/mixed forms; Eastern European cities spread toward lower-density types.

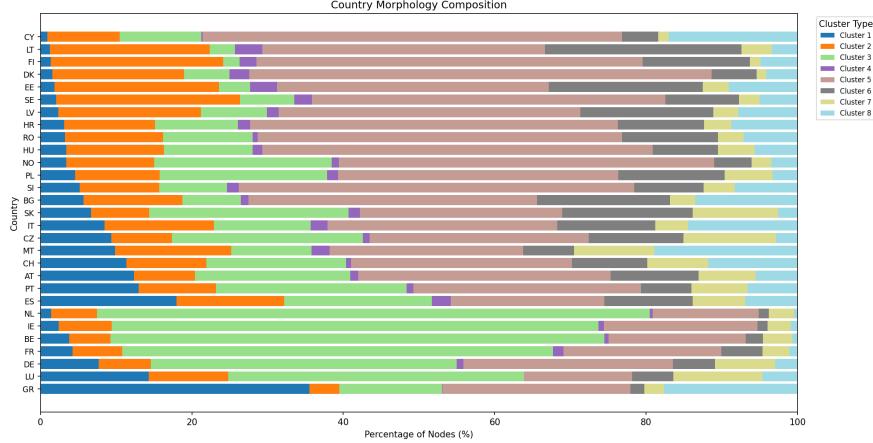


Figure 18: **Country morphology compositions.** Stacked bar chart showing the proportion of nodes in each morphological cluster by country. Countries are ordered by hierarchical clustering of their composition vectors, grouping nations with similar urban form distributions.

terns, and regulatory frameworks leave detectable signatures in aggregate urban form.

8.5. Discussion

The analysis suggests that European cities may share common neighbourhood building blocks despite diverse planning traditions. The 8 clusters span a spectrum from lower-density suburban development (Cluster 5: 4,151 persons/km²) to dense urban cores (Clusters 1, 6, 8: 13,776, 12,118, and 13,160 persons/km² respectively). The most common cluster types are Cluster 3 (30.8% of sampled nodes) and Cluster 5 (29.2%), representing moderate-density suburban-to-urban transitional forms that dominate European urban fabric.

The external characterisation reveals that morphological clustering captures meaningful functional differences: population density varies by a factor of 3.3 across clusters (from 4,151 to 13,776 persons/km²), network density by 1.7 (from 799 to 1,373), and land-use diversity (Hill $q = 0$) by 3.3 (from 2.11 to 7.00). The clusters represent recognisable urban fabrics that recur across national contexts, but the mix varies systematically by country—Greek cities show high concentrations in Cluster 1 (35.5%), while Dutch, Belgian, and Irish cities concentrate heavily in Cluster 3 (73.0%, 65.3%, and 64.3% respectively). Nordic countries (Denmark, Finland) show strong presence in Cluster 5 (51–61% of nodes).

Most cities contain multiple morphological types—this heterogeneity would be invisible in city-level aggregates but is captured by the compositional approach. Country clustering suggests that shared planning histories, housing policies, and development economics leave detectable signatures in aggregate urban form. The distinction between Northwestern European countries (Netherlands, Belgium, Ireland: Cluster 3 dominant with 64–73%), Nordic countries (Denmark, Finland: Cluster 5 dominant with 51–61%), and Southern European countries (more varied distributions) aligns with known differences in planning traditions and urban development patterns.



Figure 19: **Representative satellite imagery for morphological clusters.** Each panel shows a 5×5 tile grid (zoom level 18) of a representative location for each cluster type. Top row (left to right): Clusters 1–4. Bottom row: Clusters 5–8. Visual inspection confirms that morphological metrics capture meaningful differences in building density, arrangement, and urban fabric. Satellite imagery: ESRI World Imagery (Maxar, Earthstar Geographics, and the GIS User Community).

8.6. Extensions

Potential directions: temporal evolution of compositions using historical imagery; correlations with urban outcomes (energy use, health, economic activity); transitional neighbourhood identification at cluster boundaries; morphological vs functional similarity comparisons; satellite imagery validation of cluster assignments.

8.7. Reproducibility

Code, outputs, and documentation: https://github.com/UCL/t2e-soar/tree/main/paper_research/code/eg6_typology/

9. Vignette 7: Site Selection

9.1. Motivation

Large-scale datasets can be filtered by multiple criteria to identify candidate locations for new facilities, housing, or infrastructure investments. This vignette identifies locations with high centrality, mixed uses, and transport access, but lower population density as potential candidates for development.

9.2. SOAR Metrics Utilised

- **Diversity indices:** Hill numbers ($q = 0, q = 1, q = 2$) at 400m measuring land-use mix
- **Network centrality:** `cc_beta_1600` (20-minute catchment)

- **Transport access:** `cc_transport_nearest_max_1600` (distance to nearest stop)
- **Population density:** `density` (persons/km²)
- **Saturation classification:** From Vignette 1 (cities with Consistently Saturated or Variable Saturated POI coverage)

9.3. Methodology

We classify street network nodes into typologies based on combinations of centrality, diversity, transport access, and density. Nodes with high centrality, high diversity, and good transport access but lower population density represent potential development opportunities—locations where urban infrastructure supports intensification but current utilisation is low.

9.3.1. Classification Workflow

The workflow (`paper_research/code/eg7_site_selection/`) operates at the street network node level:

1. **City filtering:** Select cities with reliable POI coverage (as identified in Vignette 1).
2. **Metric extraction:** For each node, extract diversity indices, centrality, transport distance, and population density.
3. **Threshold classification:** Define high/low thresholds using within-city percentiles (70th percentile for high, 30th for low). Classify nodes as:
 - Mixed-use dense: High diversity and high density
 - Mixed-use opportunity: High diversity, high centrality, good transport access, but low density
 - Single-use dense: Low diversity but high density
 - Peripheral: Low centrality and low diversity
4. **City profiling:** Compute the percentage of nodes in each typology per city.
5. **Ranking:** Rank cities by proportion of mixed-use nodes and by proportion of opportunity nodes.

9.4. Results

Analysis of cities with reliable POI coverage suggests variation in urban form typologies and development potential.

Mixed-use cities have high proportions of nodes with diverse land-use accessibility, indicating fine-grained mixing of residential, commercial, and service functions. Table 12 shows cities with highest mixed-use proportions.

Opportunity cities have substantial proportions of nodes that combine high connectivity and diversity with lower current population density, suggesting potential for sustainable densification. Table 13 highlights cities with highest opportunity proportions.

Table 12: Cities with highest proportion of mixed-use nodes.

City	Country	Mixed Score	% Mixed Dense	% Mixed Opp.
Den Haag	NL	0.42	30.0	0.0
Leiden	NL	0.42	30.0	0.0
Venezia	IT	0.41	8.3	3.4
Νέα Αλικαρνασσός	GR	0.40	21.5	1.7
La Rochelle	FR	0.40	30.0	0.0
Grasse	FR	0.40	30.0	0.0
Bayonne	FR	0.39	30.0	0.0
Bahía de Cádiz	ES	0.39	7.6	5.1
el Baix Segura / La	ES	0.38	20.2	2.7
Bari	IT	0.38	19.3	1.1

Table 13: Cities with highest proportion of development opportunity nodes.

City	Country	% Opportunity	Centrality	Transport
Grasse	FR	21.3	0.37	0.50
Draguignan	FR	19.8	0.39	0.50
Grasse	FR	19.3	0.37	0.50
Bayonne	FR	18.7	0.49	0.50
Cherbourg	FR	18.5	0.50	0.50
Vannes	FR	17.6	0.56	0.50
Søholt	DK	17.1	0.48	0.50
Rætebøl	DK	17.0	0.48	0.50
Toulon	FR	16.9	0.44	0.50
La Rochelle	FR	16.3	0.51	0.50

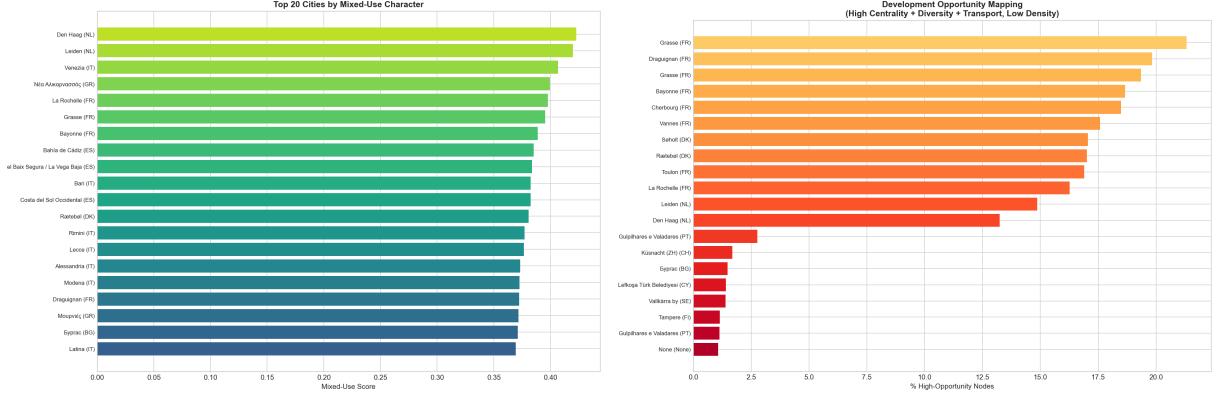


Figure 20: **City rankings by urban form typology.** Left: Cities ranked by proportion of mixed-use nodes. Right: Cities ranked by proportion of development opportunity nodes (high connectivity and diversity with lower population density).

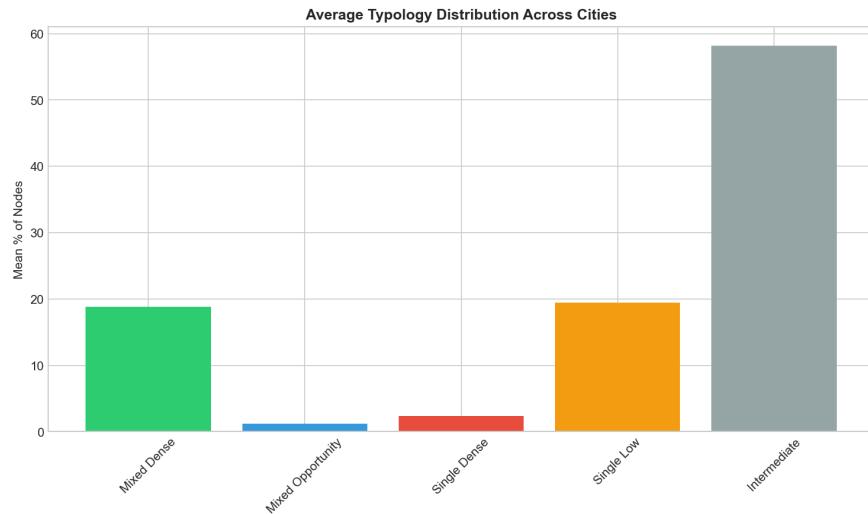


Figure 21: **Distribution of urban form typologies across analysed cities.** Histogram showing the proportion of nodes classified into each typology category (mixed-use dense, mixed-use opportunity, single-use dense, peripheral).

9.5. Discussion

The typology classification illustrates how to identify locations where infrastructure (connectivity, land-use mix, transport access) exceeds current utilisation (population density). These are not recommendations for development—such decisions require local planning knowledge, market analysis, and community input—but rather a filtering mechanism to identify areas warranting further investigation.

The threshold-based classification is sensitive to parameter choices; different percentile cutoffs would identify different opportunity areas.

9.6. Extensions

Potential directions: additional criteria (land ownership, zoning, building age); scenario modelling for new infrastructure; correlation with property values or vacancy rates; longitudinal analysis of densification patterns; case studies of identified opportunity areas.

9.7. Reproducibility

Code, outputs, and documentation: https://github.com/UCL/t2e-soar/tree/main/paper_research/code/eg7_site_selection/

10. Discussion

10.1. Cross-Cutting Themes

Across all seven vignettes, several themes emerge: (1) data quality assessment provides a foundation for comparative analysis; (2) multi-scale metrics capture neighbourhood effects at varying radii; (3) node-level granularity identifies within-city inequities missed by coarse zonal aggregations; and (4) reproducible workflows using standardised metrics enable researchers new to spatial network analysis.

10.2. Limitations

Each question provides sufficient methodological detail to enable replication. Researchers can expand these analyses with:

- **Domain-specific theoretical frameworks:** Grounding analyses in urban planning theory, geography, sociology, economics, or other relevant disciplines
- **Additional validation:** Incorporating field observations, administrative data, surveys, or behavioral data to test whether patterns hold beyond the available metrics
- **Sensitivity analyses:** Examining how results change with different parameter choices, spatial scales, or methodological approaches
- **Longitudinal perspectives:** Adding temporal dimensions to understand how patterns evolve
- **Contextual depth:** Conducting detailed case studies of specific cities or regions to understand local mechanisms

- **Cross-dataset integration:** Combining SOAR with other data sources (mobility data, economic indicators, policy records) for richer analyses

Additional limitations include: (1) POI data quality variations across regions (addressed in Vignette 1); (2) temporal constraints (SOAR represents a snapshot); (3) lack of behavioural validation (network distances are proxies for actual travel behaviour); (4) computational requirements; and (5) the inherent limitations of any single dataset in capturing urban complexity.

10.3. Adapting These Analyses

Researchers can adapt these analyses by:

- **Parameter tuning:** The spatial scales, distance thresholds, and statistical cutoffs used here are starting points; sensitivity testing may reveal more appropriate values for specific contexts
- **Local data integration:** Combining SOAR with municipal datasets, regional surveys, or national statistics can provide validation and additional explanatory power
- **Methodological alternatives:** The analytical approaches demonstrated here (Random Forests, correlations, descriptive statistics) are illustrative; researchers should explore alternative methods (hierarchical models, spatial econometrics, machine learning ensembles) as appropriate
- **Geographic focus:** While we analyze 699 cities, in-depth investigations of subsets (single countries, specific typologies, matched pairs) may yield richer insights
- **Stakeholder engagement:** Collaborating with planners, policymakers, or community organizations can ensure that analyses address real-world priorities and benefit from local knowledge
- **Computational considerations:** Some analyses may benefit from high-performance computing resources, spatial databases, or cloud platforms

11. Conclusion

This paper demonstrates the analytical potential of the SOAR urban data model through seven exploratory vignettes spanning data quality assessment, multi-scale analysis, access gap identification, predictive modelling, benchmarking, typology classification, and site selection. Each vignette provides reproducible code and clear workflows that researchers can adapt for their own investigations.

These vignettes are starting points designed to help researchers get started with the data. Building on these templates, researchers can:

- Ground analyses in domain-specific theory
- Validate patterns against local data sources
- Conduct sensitivity analyses on parameters and thresholds
- Explore causal mechanisms underlying observed patterns

The contribution is practical: by providing worked examples across diverse application types, we aim to lower the barrier to entry for pan-European urban research and enable researchers to assess whether SOAR suits their needs.

Acknowledgements

This work was supported by the European Union’s Horizon programme under the TWIN2EXPAND project. We acknowledge the data providers: Overture Maps Foundation, Copernicus Land Monitoring Service, and Eurostat.

Funding

This work was supported by the European Union’s Horizon programme under grant agreement No. [TODO: no.] (TWIN2EXPAND project).

References

- [1] G. Simons, Others, Soar: A scalable, open, automated, and reproducible urban data model for the eu, Data in BriefIn preparation (2025).