TITLE OF THE MRP IN CAPITAL LETTERS

by

Andrey Zhuravlev

Supervisor: Pawel Pralat

Student Number: 501338749

Revised Abstract

Literature Review

The relationship between transportation infrastructure and real estate value is a cornerstone of urban economics, with a rich body of research demonstrating that enhanced accessibility is capitalized into property prices. Early models often relied on simple proximity-based metrics, such as distance to the nearest station. However, the academic literature has evolved significantly, now employing more sophisticated techniques to capture the complex, heterogeneous, and dynamic nature of this relationship. This review synthesizes key methodological trends and identifies a crucial gap in the academic analysis of London's housing market, which this research proposes to address.

A foundational approach in modern analysis involves modeling transit systems as networks and applying graph theory to quantify a location's importance. Scholars have consistently found that centrality indices are significant predictors of housing and land values across diverse urban contexts. Studies in Kolkata[1], Bangkok[2], numerous Chinese cities[3], and a bachelor's thesis on Munich[26] have shown that measures like closeness, eigenvector, and betweenness centrality serve as powerful proxies for the locational advantages conferred by a well-connected position in the transport network. For instance, Chakrabarti et al. (2022) found that higher closeness and eigenvector centrality were associated with price premiums of 9-40% in Kolkata[1], while Han and Wu (2023) demonstrated that a 10% rise in harmonic centrality (a variant of closeness) correlated with an 11% increase in land prices in China[3]. These studies validate the core premise of using graph-based metrics to analyze real estate markets.

While centrality metrics establish that a relationship exists, a significant body of research focuses on its spatial variability. The impact of transit access is rarely uniform across a city. To address this, many studies employ spatial econometric models like Geographically Weighted Regression (GWR) to account for heterogeneity. He (2020) used a Multiscale GWR (MGWR) to reveal that capitalization effects in Hong Kong varied significantly by sub-region[4], while others used GWR to map localized price premiums in Xiamen and Beijing[5], [6]. Research has also advanced beyond simple linear assumptions; Chen et al. (2022), for instance, used spline functions within a GWR framework to model the non-linear, "hump-shaped" effect of metro accessibility, where the price premium diminishes or even reverses at very close proximity to a station[7]. This precedent strongly supports this MRP's borough-level aggregation strategy, which is designed to uncover how the impacts of network changes vary across London's distinct sub-markets.

Moving beyond static, cross-sectional analyses, another research stream investigates the temporal and causal dimensions of transit investment. Quasi-experimental methods, particularly the Difference-in-Differences (DiD) framework, are crucial for isolating the impact of infrastructure changes from broader market trends. Comber and Arribas-Bel (2017) expertly applied a spatial DiD model to Ealing, demonstrating that the mere announcement of Crossrail in 2008 created a measurable, anticipatory uplift in house prices, long before the line became operational[19]. A more recent study on Shanghai's subway expansion used a Matching DiD design to track price dynamics before and after station openings, finding that price increases began one month prior to launch and persisted for six months[8]. These studies underscore the importance of a long-term temporal analysis, as proposed in this MRP, to capture both the anticipation and realization of transit benefits.

The methodological frontier in this domain is increasingly characterized by the application of Graph Neural Networks (GNNs), a topic covered in depth by foundational surveys[11], [15]. GNNs can learn complex, non-linear relationships directly from graph-structured data, generating rich embeddings of nodes that encode both network topology and intrinsic features. Their versatility is demonstrated by applications ranging from predicting urban polycentricity from street networks[23] to serving as computationally efficient surrogate models for traditional four-step transport demand forecasting[14]. In the real estate domain, recent studies have successfully used GNNs to predict housing prices in Scotland[12], Santiago[13], and Melbourne[16], consistently outperforming conventional models[13]. The proposed GNN models in these papers, such as PD-TGCN with its attention mechanism[13] and the multipartite graph structure of GSNE[16], offer a sophisticated framework for integrating the diverse data sources available for this project.

A key advantage of recent GNN applications is the focus on explainability. The work by Karamanou et al. (2024), along with related conference proceedings[18], is particularly relevant, as it applies explainability techniques to a GNN model for house price prediction in Scotland. By using methods like GNNExplainer, the authors identify not only the most influential features but also the specific neighboring regions (subgraphs) that drive a price prediction for a given area[12]. This provides a powerful tool for moving beyond correlation to offer interpretable, policy-relevant insights, a goal central to this MRP.

It is important to note that a number of papers reviewed here were published in Sustainability, an open-access journal by MDPI[2], [6], [7]. While widely cited, the journal has faced scrutiny regarding its academic rigor. Notably, it was removed from the official lists of approved academic journals by the Norwegian and Finnish national publication committees in 2023. Concerns have been raised about its rapid peer-review timelines and a business model reliant on article processing charges. While this does not invalidate the findings of individual papers, it highlights the need to critically assess their methodologies, a principle applied throughout this review.

Furthermore, several studies referenced in this review, particularly those focused on inference rather than prediction, do not employ a traditional training-testing data split. Instead, models are applied to the full dataset to explore spatial patterns and estimate the significance of relationships. This is a well-established practice in spatial econometrics where the primary goal is understanding coefficients across the entire study area, rather than optimizing a model for out-of-sample prediction. This methodological precedent aligns with the exploratory and inferential goals of this MRP.

While the global literature provides a robust methodological toolkit, a critical analysis reveals a significant gap concerning the London market. Most of the rigorous, peer-reviewed academic studies are focused on other global cities. Conversely, the most detailed, long-term analyses of London's transit-property value link come from reports commissioned by public or private sector entities. These include economic impact assessments of TfL's supply chain spending[22], analyses of future funding scenarios[21], and reports on specific infrastructure projects.

These reports offer invaluable data and context, confirming significant property value uplift around the Elizabeth Line[10], Crossrail[9], and the London Overground[20]. For instance, the TfL-commissioned report on Crossrail found a 7-10% price uplift near stations even before opening[9], and CBRE's analysis showed an 80% price increase near Elizabeth Line stations between 2008 and 2023[10]. However, as these studies are commissioned by organizations with a vested interest in demonstrating positive outcomes, they may carry an implicit promotional bias and typically lack the methodological transparency and peer-reviewed validation of academic research. They often rely on descriptive comparisons or simpler hedonic models rather than the advanced spatial, causal, or graph-based techniques found in scholarly journals.

This identifies a clear and compelling research opportunity: there is a lack of independent, academic research that applies state-of-the-art graph mining and spatial-temporal modeling techniques to analyze the impact of London's transit network evolution on its housing market. While Zhang et al. (2021) provided a landmark academic study using London's NUMBAT data to classify station roles through community detection[17], its focus was not on housing market impacts. The novelty of this MRP, therefore, lies in bridging this gap. It will synthesize advanced, academically-rigorous methods—validated by the international literature and drawing on concepts from transport geography and graph mining[25]—and apply them to a unique, 25-year dataset for London. By modeling the transit network based on observed passenger flows from RODS and NUMBAT, a practice supported by recent mobility data benchmarks[24], this project will capture a more functionally accurate measure of accessibility, allowing for a deeper and more nuanced understanding of how network structure shapes urban economies.

Data Description and Exploratory Data Analysis

Datasets

(i) UK House Price Index - Monthly average house prices for each London borough from 1995 to the present, including sales counts and property type breakdowns.

Dataset Description and Link: https://data.london.gov.uk/dataset/uk-house-price-index

The dataset is available under [UK Open Government License (v3)]()

It contains 5 sheets:

- Metadata
- By Type

  Monthly house prices and corresponding index values for London and the United Kingdom, broken down by property type (Detached, Semi Detached, Terraced, and Flat). The data is organized with a multi-level header: the first two columns indicate year and month, followed by eight columns for London (four for prices and four for indices) and eight for the UK (also split into prices and indices by property type). Each row represents a specific month, starting in 1995.

- Average price

  Monthly average house prices across all 33 London boroughs, grouped English regions, and national aggregates from January 1995 onward. The top row shows borough and region names (e.g., Camden, Westminster, South East, England), and the second row includes their official administrative codes (e.g., E09000007 for Camden). Each subsequent row begins with a date (e.g., Jan-95) followed by numeric values representing average sale prices in GBP for that month and geography. Data is granular at the borough level for London and aggregated by region elsewhere in England, with Inner/Outer London summaries and a national average ("England") included.

- Index Price

  This sheet follows the same structure as the previous one but contains index prices instead of average sale prices

- Sales Volume

  The structure is the same as the one of the Average Price and Index Price sheets, but it presents Sales Volume data, indicating the number of residential property transactions recorded each month in a specific area (33 London boroughs, Inner and Outer London, English regions, and the entire country).

Prices are nominal and not adjusted for inflation. Excludes non-market sales (e.g., right-to-buy, gifts, leases under 7 years). As of February 2025, the index was re-referenced to January 2023, and historical data was revised. A 3-month moving average is applied at the borough level.

(ii) TfL Rolling Origin-Destination Survey (RODS) - Annual survey capturing typical weekday entries, exits, and OD flows for London Underground stations, covering 2000–2017. Provides station-to-station trip matrices, boarding and alighting patterns, and station-level demand profiles.

Dataset Description (including License information):
https://data.london.gov.uk/dataset/tfl-rolling-origin-and-destination-survey

Dataset Link:
https://tfl.gov.uk/corporate/transparency/freedom-of-information/foi-request-detail?referenceId=FOI-1386-2021

The dataset is available under [UK Open Government License (v2)]()

There is a separate .xls file for each year and each file consists of 2 sheets:

- Matrix

    Origin-Destination (OD) flow matrix, showing weekday passenger journey counts between London Underground stations, broken down by time bands. Each row records the number of trips from a specific origin station (Station Name in column 2, with National Location Code in column 1) to a specific destination station (Station Name in column 4, with its own NLC in column 3). The remaining columns capture journey volumes across six time periods: Early (before 7am), AM Peak (7am–10am), Midday (10am-4pm), PM Peak (4pm-7pm), Evening (7pm-10pm), and Late (after 10pm), along with a final Weekday Total column summing across these bands. This format enables temporal analysis of commuter flows, revealing not only station-to-station connections but also the distribution of travel activity throughout the day.

- Zone

    Presents a zone-to-zone OD flow matrix from the RODS report, summarizing estimated journey volumes between London fare zones based on survey data reconciled to Autumn counts. The table is structured into multiple matrices for the same 6 time bands, with each matrix showing trip counts and percentage distributions across zones 1 through 7. For each time band, the rows represent origin zones and the columns represent destination zones, with totals and proportional flows included. Below each matrix, a compressed summary matrix aggregates flows into four broader groups: Zone 1, Zone 2, Zone 3, and combined outer zones (Zones 4–7).

(ii) TfL NUMBAT

Detailed smartcard-based estimates of rail demand across London Underground, Overground, DLR, and TfL Rail, covering 2016–2023. According to TfL, compared to RODS, NUMBAT is a much larger sample size as it uses ticketing data, oyster taps, train loadweigh and passenger counts, as opposed to the manual survey method of RODS which required a lot of scaling.

Dataset Link: https://crowding.data.tfl.gov.uk/

The dataset is available under [UK Open Government License (v3)]()

NUMBAT Origin-Destination Data consists of multiple files per year, each corresponding to a specific day type in the autumn period. For every year, separate files are provided for Fridays, Saturdays, and Sundays, while in more recent years (e.g., 2022), Monday flows are also split out from the midweek group (Tuesday to Thursday). Though filenames are inconsistent and not standardized, each one represents a typical autumn day for that day type, with major disruptions and anomalous days excluded to ensure data stability.

Each file contains a single sheet structured as an Origin-Destination (OD) matrix. Every row corresponds to a unique journey from an origin station (`mnlc_o`) to a destination station (`mnlc_d`), identified by their Master National Location Codes (NLCs), unique identifiers assigned to each station across the TfL network. The remaining columns (starting at column 3) represent the number of trips in each 15-minute interval throughout the traffic day, spanning from 05:00 to 04:59 the next morning. These intervals are indexed using quarter-hour slot numbers, starting at 21 (05:00-05:15) and cycling through to 96 (23:45-00:00), followed by 97-116 to capture post-midnight flows.

Trip counts in NUMBAT are typically non-integer values, reflecting modelled estimates derived from smartcard data and gateline counts, adjusted using timetable-based assignment. This contrasts with the RODS OD matrices, where all values are integers representing actual or scaled passenger counts from survey samples. It is designed for detailed analysis of travel demand, service provision, and customer experience, and for use in service planning and performance evaluation.

Unlike OD matrices in RODS, which are based on survey data, aggregated by fare zone and coarse time bands, NUMBAT provides station-to-station OD flows at 15-minute granularity across all TfL-operated services. This makes it a more precise and modern tool for analyzing how, when, and where passengers travel across the network.

NUMBAT Outputs are other measures representing typical flows, they are also given by the same 15 minute bends described earlier and there are also multiple files per year, with the same structure as NUMBAT Origin-Destination Data. Each file consists of 9 sheets:

- VersionControl, Note, Cover, Cover Page, _Cover: Metadata and documentation (title, publishing info, definitions), varies slightly in name year-to-year.
- Link Load: Number of passengers on the train between two different consecutive stations on a line e.g. from Lambeth North to Waterloo on the Bakerloo line.
- Link frequency (supply): Number of scheduled trains per quarter hour between two consecutive stations.
- Station Boarders and Station Alighters: Number of passengers boarding or alighting at a specific platform in a specific station.
- Line Boarders: Total boarders across a rail line, derived from summing station boarders. This gives a better picture than Entry/Exits for the utilisation of a line as it includes interchangers.
- Station Flow: Number of passenger movements inside a station between its entrances and its different platforms. This includes boarding/alighting and interchange flows.
- Station Entries and Station Exits: number of passengers entering/exiting the station through its gatelines.

Data Preprocessing and Cleaning

One of the first challenges that I faced is a lack of publicly available dataset linking each Underground station to its corresponding Borough. To obtain that information, I created *london_tube_station_borough_scraper.py* script designed to go to every Subcategory on "Category:Tube stations in London by borough" Wikipedia page, scape the corresponding Subcategory page for the list of all the Tube stations in any given borough, and the save that information in *london_tube_stations_by_borough.csv*. The challenge I encountered was the inconsistency in naming conventions across datasets obtained from different sources. The *UK_House_price_index.xlsx* file used one set of borough names, while the *london_tube_stations_by_borough.csv* file used another, leading to mismatches that prevented direct merging; additionally, tube station names in the *OD_matrix* files sometimes differed from those in the borough mapping file. It took some adjustment and cleaning for boroughs (for instance removing "London Borough of" or "Royal Borough of" from the beginning of borough names, or substituting "and" in boroughs' names for "&"), as well as station names (removing " tube station" from the end of the names, fixing minor punctuation differences like apostrophes, removing additional clarifying geographical information in parentheses (e.g., '(London)'), matching older terminal numbering to newer designations, etc.) to make all those names match. I used Venn diagrams to compare and visualize overlaps and mismatches in both borough and station names and listing the non-overlapping entries to guide manual mapping.

After all of this pre-processing, there were still some discrepancies between the datasets, though there are explanations for those:

The house price index contained six boroughs that were not found in the tube station dataset: Bexley, Bromley, Croydon, Kingston upon Thames, Lewisham, and Sutton. The reason for this discrepancy is that these are London boroughs that are not served by the London Underground network, relying on other transport systems like National Rail instead.

The comparison between the 2017 Origin-Destination (OD) Matrix and the scraped tube stations names showed mismatches in both directions:

Stations only in OD Matrix 2017 (16 total): Amersham, Buckhurst Hill, Chalfont & Latimer, Chesham, Chigwell, Chorleywood, Croxley, Debden, Epping, Grange Hill, Loughton, Moor Park, Rickmansworth, Roding Valley, Theydon Bois and Watford.

These stations exist in the operational OD matrix because they are part of the London Underground network, but they were not captured by the borough-based scrape as they are located outside the official Greater London boundaries.

Stations only in *london_tube_stations_by_borough* CSV (18 total):Battersea Power, Beckton, City Road, Emlyn Road, Heathfield Terrace, Hounslow Town, King William Street, London Victoria, Lothbury, Ludgate Circus, Mark Lane, Millwall, Nine Elms, Paddenswick Road, Rylett Road, St Katharine Docks, Surrey Docks North and The Grove.

This list is a composite of several categories. It includes stations that opened after 2017 (Battersea Power, Nine Elms), permanently closed "ghost" stations captured from historical Wikipedia lists (City Road, Mark Lane, King William Street, Hounslow Town), stations on other transit systems like the DLR (Beckton), and fully operational multimodal interchange stations that are not exclusively part of the Underground (London Victoria). It also includes a substantial number of authorised but never built Underground stations, which were included in historical or planning-related Wikipedia categories and thus scraped as valid entries despite never having been constructed. These are: Emlyn Road, Lothbury, Paddenswick Road, Heathfield Terrace, Rylett Road, The Grove, Surrey Docks North, St Katharine Docks, Millwall, and Ludgate Circus. These planned stations were associated with early Central London Railway proposals or unbuilt extensions of the Jubilee line, particularly the intended line to Woolwich Arsenal.

Since RODS was replaced with NUMBAT, analysis of the stations in NUMBAT files was also required. What made things even more challenging was that NUMBAT Origin-Destination matrix data only contained National Location Codes (NLCs), so the mapping of the stations to Location Codes was required. Originally I planned

415	NLCs in all three datasets

39	NLCs only in NLC Mapping
0	NLCs only in NUMBAT 2019
18	NLCs only in NUMBAT 2022
1	NLCs in NLC Mapping and NUMBAT 2019 only
0	NLCs in NLC Mapping and NUMBAT 2022 only
241	NLCs in NUMBAT 2019 and NUMBAT 2022 only

(Table 1)

To create a comprehensive mapping between station names and National Location Codes (NLCs), I created *create_station_nlc_mapping.py* that systematically processes all NUMBAT model output files from 2016 to 2023. Each of these files contains a Station_Entries sheet, which consistently stores the NLC in the first column and the station name in the third column. The script was designed to extract all unique pairs of NLCs and station names from each file, ensuring that any station present in any scenario or year would be included in the final mapping. After extracting the data from all 32 output files, the script combined and deduplicated the results, resulting in a master mapping that contains 472 unique station-NLC pairs. This mapping, saved as comprehensive_station_nlc_mapping.csv, provides a robust and up-to-date reference for all stations modeled in NUMBAT across the studied years.

However, a detailed investigation revealed an important distinction between different NUMBAT datasets. While the comprehensive mapping successfully captured all 472 stations listed in the NUMBAT output files' Station_Entries sheets, the 2022 NUMBAT OD matrix files contain 674 unique NLC codes, which is 202 more than the output files. This discrepancy highlights that the OD matrices include additional nodes beyond the passenger stations listed in the output files. Among these additional codes, three specific NLCs (6070, 6073, and 8204) were found to have significant passenger flows in the 2022 OD matrix but do not appear in any of the NUMBAT output files, or the *StationNodesDescription.xls* file provided by TfL themselves.

This finding suggests that these codes likely represent special network nodes such as interchange points, junction stations, or virtual nodes used for routing purposes rather than traditional passenger stations. For borough-based analysis, these codes can be safely excluded since they do not represent actual passenger stations with geographic locations that can be assigned to specific boroughs. Including them would introduce artificial nodes that could distort the spatial distribution of passenger flows and complicate the borough-level analysis without providing meaningful geographic insights.

Project Approach

A clearly defined research methodology
Justification of the selected approach based on prior studies
A graph or diagram illustrating the overall project methodology

GitHub Repository

https://github.com/azhuravlev1/MRP_LondonTransit_RealEstate

References

Relevant published sources
Consistent citation style (e.g., APA or IEEE)

1. Chakrabarti, S., Kushari, T., & Mazumder, T. (2022). *Does transportation network centrality determine housing price? Journal of Transport Geography, 103*, 103397. https://doi.org/10.1016/j.jtrangeo.2022.103397
2. Vichiensan, V., Wasuntarasook, V., Prakayaphun, T., et al. (2023). *Influence of Urban Railway Network Centrality on Residential Property Values in Bangkok. Sustainability, 15*(22), 16013. https://doi.org/10.3390/su152216013
3. Han, D., & Wu, S. (2023). *The capitalization and urbanization effect of subway stations: A network centrality perspective. Transportation Research Part A: Policy and Practice, 176*, 103815. https://doi.org/10.1016/j.tra.2023.103815
4. He, S. Y. (2020). *Regional impact of rail network accessibility on residential property price: Spatially heterogeneous capitalization effects in Hong Kong. Transportation Research Part A: Policy and Practice, 135*, 244–263. https://doi.org/10.1016/j.tra.2020.01.025]
5. Yang, L., Chu, X., Gou, Z., et al. (2020). *Accessibility and proximity effects of bus rapid transit on housing prices: Heterogeneity across price quantiles and space. Journal of Transport Geography, 88*, 102850. https://doi.org/10.1016/j.jtrangeo.2020.102850
6. Zhou, Y., Tian, Y., Jim, C. Y., et al. (2022). *Effects of Public Transport Accessibility and Property Attributes on Housing Prices in Polycentric Beijing. Sustainability, 14*(22), 14743. https://doi.org/10.3390/su142214743
7. Chen, K., Lin, H., Liao, L., et al. (2022). *Nonlinear Rail Accessibility and Road Spatial Pattern Effects on House Prices. Sustainability, 14*(8), 4700. https://doi.org/10.3390/su14084700
8. Wang, J., & Tang, H. (2025). *Subway Expansion and Housing Price Dynamics: Accessibility Improvements and Spatial Heterogeneity in Shanghai. SSRN Working Paper*. https://doi.org/10.2139/ssrn.5187633

9. Transport for London, & Department for Transport. (2022, April 27). *Crossrail baseline evaluation: Evaluation of Crossrail pre-opening property impacts*. https://content.tfl.gov.uk/property-impacts-report-acc.pdf
10. CBRE. (2024). *The Elizabeth Line: The impact on London's housing market*. Adaptive Spaces. https://www.cbre.co.uk/insights/reports/the-elizabeth-line-the-impact-on-london-s-housing-market
11. NYU Marron Institute of Urban Management. (2021). *Access to Opportunity: Housing Affordability and Transportation in New York City*. ArcGIS StoryMaps. https://storymaps.arcgis.com/stories/6c15fbaed4ff49e3bdaecddcfdbf4e9b
12. Karamanou, A., Brimos, P., Kalampokis, E., & Tarabanis, K. (2024). Explainable Graph Neural Networks: An Application to Open Statistics Knowledge Graphs for Estimating House Prices. *Technologies, 12*(8), 128. https://doi.org/10.3390/technologies12080128
13. Riveros, E., et al. (2023). *Scalable Property Valuation Models via Graph-based Deep Learning. Preprint under review.* https://arxiv.org/abs/2405.06553
14. Narayanan, S., Makarov, N., & Antoniou, C. (2024). *Graph neural networks as strategic transport modelling alternative – A proof of concept for a surrogate*. *IET Intelligent Transport Systems, 18*(11), 2059–2077. https://doi.org/10.1049/itr2.12551
15. Li, H., Zhao, Y., Mao, Z., Qin, Y., Xiao, Z., Feng, J., Gu, Y., Ju, W., Luo, X., & Zhang, M. (2024). *A Survey on Graph Neural Networks in Intelligent Transportation Systems*. arXiv preprint arXiv:2401.00713. https://arxiv.org/abs/2401.00713
16. Das, S. S. S., Ali, M. E., Li, Y.-F., Kang, Y.-B., & Sellis, T. (2021). *Boosting House Price Predictions using Geo-Spatial Network Embedding.* arXiv preprint arXiv:2009.00254. https://doi.org/10.48550/arXiv.2009.00254
17. Zhang, Y., Marshall, S., & Manley, E. (2021). Understanding the roles of rail stations: Insights from network approaches in the London metropolitan area. Journal of Transport Geography, 94, 103110. https://doi.org/10.1016/j.jtrangeo.2021.103110
18. Brimos, P., Karamanou, A., Kalampokis, E., Mamalis, M. E., & Tarabanis, K. (2023). *Explainable Graph Neural Networks on Linked Statistical Data for Predicting Scottish House Prices.* In *Proceedings of the 27th Pan-Hellenic Conference on Informatics (PCI 2023)* (pp. 48–56). ACM. https://doi.org/10.1145/3635059.3635065
19. Comber, S., & Arribas-Bel, D. (2017). *"Waiting on the train": The anticipatory (causal) effects of Crossrail in Ealing.*Journal of Transport Geography, 62, 92–104. https://doi.org/10.1016/j.jtrangeo.2017.08.004
20. Transport for London. (2012, February). *London Overground: Impact Study.* London: TfL. Retrieved from https://content.tfl.gov.uk/Item08-020212-Board-London-Overground-Impact-Study.pdf
21. Greater London Authority. (2022). *Economic impacts under future funding scenarios for TfL.* London: GLA Economics. Retrieved from https://www.london.gov.uk/business-and-economy-publications/economic-impacts-under-future-funding-scenarios-tfl

22. Hatch. (2024, November). *Transport for London Supply Chain: Economic Impact Assessment 2023/24*. Transport for London. Retrieved from https://content.tfl.gov.uk/tfl-supply-chain-economic-impact-assessment-2023-24.pdf
23. Ma, D., He, F., Yue, Y., Guo, R., Zhao, T., & Wang, M. (2024). *Graph convolutional networks for street network analysis with a case study of urban polycentricity in Chinese cities*. *International Journal of Geographical Information Science*. https://doi.org/10.1080/13658816.2024.2321229
24. Na, J., Nam, Y., Yoon, S., Song, H., Lee, B. S., & Lee, J.-G. (2025). Mobility networked time-series forecasting benchmark datasets. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 19, pp. 2539–2549). Association for the Advancement of Artificial Intelligence. https://doi.org/10.1609/icwsm.v19i1.35955
25. Beecham, R. (n.d.). *Class 5 — Commutes, connectivity and flows* [Web page]. Retrieved June 18, 2025, from https://www.roger-beecham.com/vis-for-gds/class/05-class/
26. Schwarzbart, J. (2021, May 27). *Analysis of public transport connectivity and different places of residence in Munich: Described by rental prices and the Isar* (Bachelor's thesis, Ludwig-Maximilians-Universität München). https://doi.org/10.5282/ubm/epub.77437