MANCHESTER
1824
The University of Manchester

Analyzing public transport mode choice determinants in Greater Manchester—Integrated GIS-Machine Learning Approach

A dissertation submitted to The University of Manchester for the degree of

Master of Science in Data Science

in the Faculty of Social Statistics

Year of submission

2025

Student ID

11531463

School of Social Science

Contents

Word count: 7402

List of figures

List of tables

Abstract

This study investigates the spatial and socio-economic determinants of public transport use for commuting in Greater Manchester, with a particular focus on changes between 2011 and 2021 and the impacts of the COVID-19 pandemic. Using geospatial analysis, bus stop densities and nearest-neighbour accessibility measures were derived from OpenStreetMap data at the LSOA level, revealing strong central–periphery contrasts: Manchester city centre and inner urban areas such as Oldham exhibit high accessibility, while peripheral districts such as Wigan and Trafford remain poorly served. Geographically Weighted Regression demonstrates that deprivation alone does not explain spatial patterns of public transport use, highlighting the importance of accessibility and service provision. Behavioural modelling through decision trees and random forests further identifies key factors shaping modal choice, including deprivation (IMD), commuting distance, population density, household structure, age, health, and occupation. Comparative analysis shows that from 2011 to 2021, car use, public transit, and active travel all declined, while working mainly from home rose sharply, reflecting pandemic-induced structural changes. The findings align with Greater Manchester's Bee Network strategy, which emphasises integrated, sustainable, and equitable mobility, but also underscore the need for investment in peripheral areas where accessibility gaps persist. The study demonstrates the value of combining spatial and machine learning methods to inform transport policy and to support the transition towards sustainable commuting in metropolitan regions.

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Dr.Ransford, for his invaluable guidance, support, and encouragement throughout the course of this research. His insightful feedback and expertise were fundamental to shaping both the direction and the quality of this dissertation.

I am also grateful to all the tutors for MSc. Data Science courses and especially our pathway lead Nuno Pinto, whose input and assistance contributed significantly to this work. Finally, I would like to thank my family and friends for their continuous support and encouragement during my MSc studies. I also want to express thanks to everyone at Stockport County football club, providing me the opportunities to apply what I have learnt from university into business insights.

Declaration of originality

I hereby confirm that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright statement

1 Introduction

1.1 Background and motivation

Urban transport systems are central to the economic, social, and environmental sustainability of city-regions. In Greater Manchester, commuting patterns are shaped by the public transport availability, socio-economic disparities, and built-environment characteristics. Buses, trams, and trains form the backbone of the public transport network, their usage varies widely across space and demographic groups. Over the past decade, car dependence has remained dominant, while public transport and active travel face persistent barriers related to accessibility, affordability, and service quality.

The COVID-19 pandemic altered the travel behaviour significantly. Public transport demand fell sharply due to infection risk and service suspensions, while car use also declined as commuting volumes dropped. In contrast, remote working has increased dramatically, particularly among professionals and information workers. These changes not only reveal structural vulnerabilities in the transport system but also highlight the varying abilities of different groups to adapt to rapid shocks. Understanding these changes is crucial for developing a sustainable mobility strategy for Greater Manchester, particularly as policymakers promote decarbonization and sustainable transport.

1.2 Aims and objectives

This report investigates the determinants and dynamics of commuting mode choice in Greater Manchester, with a focus on spatial, socio-economic, and occupational influences before and during the COVID-19 pandemic. The specific objectives are:

- To identify spatial and temporal patterns of public transport usage (bus, tram, train) across the city-region.
- To examine changes in commuting mode choice between 2011 and 2021, with particular attention to the impact of the COVID-19 pandemic.
- To assess the influence of socio-economic status, occupation, commuting distance, and built-environment factors on public transport uptake.
- To apply spatial analysis (GWR) and machine learning methods (decision trees, random forest) to quantify and interpret determinants of mode choice.
- To draw policy-relevant insights that support sustainable mobility, enhanced accessibility, and reduced inequality in Greater Manchester.

1.3 Report structure

The remainder of this report is structured as follows. Section 2 reviews the academic literature on commuting mode choice, socio-economic determinants, and the impact of COVID-19 on urban transport. Sectio 3 and Section 4 outlines the datasets, spatial analysis, and machine learning techniques employed. Section 5 presents the geospatial analysis, commuter behaviour modelling, and comparative results between 2011 and 2021. Section 5.3 interprets the findings in light of existing evidence and discusses their policy implications.

2 Literature review

2.1 Introduction

Understanding what shapes people's choice of public transport is becoming more important for building sustainable and inclusive urban mobility. In Greater Manchester, a region with varied socio-economic backgrounds and a complex built environment, people can choose between buses, trams, and trains. Knowing why commuters prefer certain modes is vital for effective transport planning.

Research has shown that travel behaviour is influenced not only by the spatial layout of cities and land use, but also by personal factors such as attitudes, lifestyles, and demographic characteristics. The COVID-19 pandemic further changed these patterns, highlighting inequalities in access to transport and encouraging shifts toward active or individual travel.

Today, with detailed spatial and socio-economic data becoming more available, and with advances in machine learning, it is possible to model and predict mode choice with much greater accuracy. This review brings together key theories, methods, and findings on public transport choices, with a focus on how spatial analysis and machine learning can help explain the changing commuting patterns in Greater Manchester.

2.2 What is travel behaviour

Travel behaviour describes how people and households decide when and how to move through space to meet daily needs. This can mean going to work, attending school, or taking part in leisure activities. It involves choices about mode of transport, departure time, route, and the order of activities. Jones et al. (1987) argued that travel is a derived demand. People travel not for its own sake, but because they need to take part in activities spread across time and space. This view is especially relevant in Greater Manchester, where commuting is influenced by job locations, family responsibilities, and the transport system. Within households, decisions are also shaped by how family members coordinate schedules and share resources such as cars and time.

To capture these dynamics, Arentze et al. (2000) created microsimulation models that show how daily activity patterns shape travel choices. Such models are valuable in large urban areas like Greater Manchester, where work hours, access to transport, and household structures vary widely. Later, Buliung and Kanaroglou (2007) expanded this research by pointing out the limits of traditional trip-based models. They stressed the need for behaviourally realistic approaches that account for priorities, time-space limits, and household decision-making. These ideas are highly relevant in the Greater Manchester context, where understanding why commuters make certain choices, instead of only focusing on their journey origins and destinations, can support better planning and demand management.

Looking at travel as part of human activity and household life offers a strong base for studying commuting in Greater Manchester. It also helps shape models and policies that reflect how people actually plan and carry out their daily movements, which is key to creating transport systems that are both sustainable and responsive.

2.3 Factors that affect behaviour

Urban form and the distribution of activities and residences are well known to shape travel mode choice (Boarnet and Crane 2001). Boarnet and Crane (2001) points out that density, land use mix, and the layout of transport networks influence how practical and attractive different modes are. Compact, mixed-use areas often support walking and cycling because distances are shorter and streets are more connected. These areas also encourage higher levels of public transport use, since people can reach stops more easily and ridership can remain strong. In contrast, low-density suburban areas often push residents towards private cars. Longer distances and limited access to reliable transit make alternatives less feasible. Schwanen, Dijst, and Dieleman (2004) shows how planning policies can shift these choices. Measures such as transit-oriented development, parking limits, and pedestrianisation reduce car use and make public transport more appealing. Commuters are more likely to take buses or trains when infrastructure is designed to prioritise access and efficiency. Yet, Kitamura, Mokhtarian, and Laidet (1997) remind us that urban form alone does not explain all variation. With urban sprawl, differences in mode choice also reflect socio-demographic characteristics and personal attitudes. In such areas, strong car dependence can remain even when transit is available, showing that travel behaviour must be understood through both physical and social factors.

Recent studies highlight the need to combine these objective aspects with subjective ones such as values and preferences. Golob (2003) used structural equation modelling to show how lifestyle and attitudes influence decisions. They found that views on convenience, safety, freedom, and environmental issues strongly shape whether people walk, cycle, drive, or use transit. For instance, people concerned about sustainability may choose buses or trams even when cars are available. Simma and Axhausen (2001) add that car ownership and employment play a major role. Full-time workers often prefer predictable and time-saving modes, making cars attractive unless transit is highly efficient. Meanwhile, those without cars, especially in low-income households, often rely on buses despite long travel times. Their work highlights how social and economic constraints can limit choice and stresses the importance of improving service quality to reduce forced car dependence.

Building on this, Dieleman, Dijst, and Burghouwt (2002) argue that lifestyles are linked to life circumstances such as age, income, housing, and occupation. These factors shape expectations and routines, which in turn influence travel modes. Young adults and students may favour cycling or walking because these are cheap and flexible, while older adults often choose cars or buses for comfort and safety. Renters or people in temporary housing may also be more flexible and multimodal compared to long-term homeowners who develop car-based routines. De Vos, Singleton, and Gärling (2022) add the concept of a "mode choice circle," where satisfaction and convenience reinforce existing habits. Positive experiences with public transport, such as punctuality, comfort, safety, and affordability, encourage more use over time. But long distances or poor first-mile and last-mile links can push people towards cars or bikes instead. Taken together, these studies show that mode choice depends not just on access and infrastructure, but on the interaction of the built environment, socioeconomic conditions, and feedback from personal experience.

2.4 Machine learning methods on understanding behaviours

Machine learning has become an increasingly valuable approach for studying complex travel behaviours (Bishop 2006). Unlike traditional models that rely heavily on fixed assumptions, machine learning can adapt to large and diverse datasets. This makes it well suited to capture the many factors that influence mode choice. Hagenauer and Helbich (2017) compared a range of machine learning models and showed that predictive performance varies widely across methods. Their study makes clear that there is no single "best" algorithm. The choice depends on the specific data, the research questions, and the balance between accuracy and interpretability.

Several studies have demonstrated the potential of different algorithms. C. Xie, Lu, and Parkany (2003) tested both decision trees and convolutional neural networks (CNNs). Decision trees were faster to run and easier to explain, which helped reveal how commuters weigh different options. CNNs, on the other hand, achieved stronger predictive accuracy but were more difficult to interpret. This shows the trade-off between understanding the process and achieving the best prediction. Y. Zhang and Y. Xie (n.d.) examined support vector machines (SVMs) and compared them with traditional logit models. Their results showed that SVMs can provide very high accuracy while also handling large, high-dimensional datasets. This flexibility makes them especially useful when travel behaviour is shaped by many interacting factors that are not easily modelled with linear approaches.

Other research has shown how machine learning can incorporate new types of variables that traditional models struggle to include. Liu, Wang, and J. Zhang (2012) used multivariate classifiers that integrated environmental conditions such as weather and road quality. They found that bad weather often leads to higher use of public transport, showing how external conditions strongly affect individual choices. Hagenauer and Helbich (2017) also tested models that combined socio-demographic factors, household attributes, built environment features, and weather data. Their results highlighted the strong performance of tree-based methods overall, but also the specific strength of SVMs when weather factors were included.

These findings also point to the limitations of traditional models. Multinomial logit approaches often fail to capture the complexity of interactions across factors. Machine learning models, by contrast, can make use of richer datasets that include real-time information on weather, household dynamics, and local environments. This ability improves both prediction and reliability. Importantly, it also gives policymakers and planners more detailed insights into how people actually make travel decisions. By identifying patterns that are not visible in conventional models, machine learning methods can help design transport systems that are more responsive, sustainable, and effective.

2.5 Effects of COVID-19

In recent years, the COVID-19 pandemic has disrupted transportation systems worldwide and reshaped daily mobility patterns. Thombre and Agarwal (2021) examined its impact on travel behaviour in India. They found that strict measures such as lockdowns and movement restrictions caused a steep decline in overall travel activity. Many people were forced to stay at home, while others reduced the frequency and duration of their trips. Concerns over exposure to the virus in shared transit spaces also pushed travellers towards alternatives. Walking and cycling became more common, and many turned to private vehicles for safety and reliability. At the same time, public transport systems experienced dramatic drops in users, raising questions about their resilience in times of crisis.

The recovery of public transport depends on restoring trust and improving service standards. De Vos, Singleton, and Gärling (2022) emphasised that higher quality, better cleanliness, and reliable operations are necessary to rebuild confidence. They also argued that sustainable recovery requires investment in active travel infrastructure. Expanding cycling lanes and ensuring safe pedestrian routes can support the long-term shift towards healthier and low-carbon mobility. Bhaduri et al. (2020) brought attention to the role of socioeconomic factors. Their findings showed that income, employment type, and car ownership shaped how different groups responded. Higher-income households, particularly white-collar professionals, were more likely to adopt remote work. This reduced their need to travel, while lower-income groups often continued to depend on physical commuting. These unequal effects underline that the pandemic's impact was not uniform across society. They highlight the need for transport policies that consider disparities in resources and access, ensuring that future systems are more inclusive as well as sustainable.

2.6 Summary

The literature shows that public transport mode choice is shaped by many factors. These include spatial form, the built environment, socio-economic status, and individual preferences. Earlier studies focused on how urban form and planning can encourage or discourage the use of public transport. More recent work has added the role of personal values and satisfaction with services, which are important for understanding why people choose one mode over another. Structural equation models have been useful for showing how life circumstances influence travel indirectly through lifestyles and attitudes.

The growth of machine learning has added new ways to study these behaviours. These methods are strong in identifying complex and non-linear patterns, especially when combining data from different sources such as GIS indicators and census-based socio-demographic variables. At the same time, the COVID-19 pandemic has drawn attention to behavioural change and equity. Remote working, income level, and access to resources have all had clear effects on commuting demand.

To sum up, these studies highlight the value of using a GIS-integrated machine learning approach to study mode choice in Greater Manchester. This is particularly relevant given the region's changing spatial patterns and the challenges of post-pandemic recovery.

3 Exploratory Data

3.1 Introduction

This research uses a range of multi-dimensional datasets to build a detailed picture of commuter behaviour in Greater Manchester and the factors that shape it. The data comes from NOMIS and the Office for National Statistics (ONS), with a focus on the 2011 and 2021 Census years. Looking across this decade makes it possible to compare long-term structural changes with the more recent disruptions linked to the COVID-19 pandemic.

The datasets reflect the aim of connecting travel behaviour with its main drivers. Geographical information, including administrative boundaries, workplace locations, and residential patterns, supports spatial analysis of commuting flows. Socio-economic variables, such as population density, employment types, and household income, show how demographic and economic conditions affect mode choice. Travel mode data, covering modal shares and changes in commuting methods, captures the outcomes directly and highlights shifts towards or away from sustainable transport options.

3.2 Geographical Data

Geographical data forms the spatial foundation of this study, enabling the mapping and analysis of commuting flows across the Greater Manchester Area. This includes administrative boundaries at LSOA scale, spatial coordinates for workplace and residential locations, and the delineation of key transport corridors. Such spatial information allows for the examination of how commuting patterns vary across different parts of GMA and how proximity to employment centres, transport nodes, and infrastructure influences mode choice and journey length. By linking these spatial units with other datasets, the research can explore the geographic distribution of commuting behaviour.

3.3 Socioeconomic Data

Socioeconomic data provides the demographic and economic context necessary to interpret variations in commuting patterns. Variables such as population density, age distribution, household composition, employment sector, occupational type, and income levels were extracted from NOMIS and ONS datasets for both 2011 and 2021. These indicators are essential for understanding how social and economic structures shape travel demand, as different demographic groups and employment types often exhibit distinct commuting behaviours. The decade-long comparison also highlights how

Fig. 1. LSOA of Greater Manchester Area 2021, by area

economic restructuring, urban regeneration, and the rise of flexible working have altered the socioeconomic drivers of commuting in GMA.

3.4 Travel Mode Data

Travel mode data captures the behavioural dimension of commuting, detailing the distribution and changes in the modes of transport used by GMA's working population. Drawing from Census commuting statistics, this includes the proportions of commuters using private cars, public transport (bus, rail, tram), active travel modes (walking, cycling), and mixed-mode journeys. The inclusion of both 2011 and 2021 data allows for an assessment of modal shifts over the decade, including the influence of transport investments such as the Metrolink network expansion and the behavioural impacts of the COVID-19 pandemic. These data underpin the core analysis of this study, linking observed changes in travel behaviour to the spatial and socioeconomic contexts described in preceding sections.

3.5 Summary

By integrating these diverse datasets, the research aims to identify not only what changes have occurred in commuting patterns in GMA between 2011 and 2021, but also how these changes can be attributed to spatial, socioeconomic, and infrastructural factors. This approach supports the broader research goal of understanding commuter behaviour in a way that informs evidence-based urban

Table 1. Metadata for Socioeconomic and Travel Mode Datasets

Variable Name	Description	Source	Year(s)	Spatial Resolution	Unit
Socioeconomic Data					
Population	Number of commuters per square kilometre	ONS Census	2011, 2021	LSOA	count
Age Distribution	Percentage of population in defined age bands (e.g., 16–24, 25–44, 45–64, 65+)	ONS Census	2011, 2021	LSOA	count
Household Composition	Number of households by type (single-person, family, multi-person)	ONS Census	2011, 2021	LSOA	count
Employment by Sector	Number of employed residents by industry classification	NOMIS	2011, 2021	LSOA	count
Occupational Type	Number of residents by occupation categories	NOMIS	2011, 2021	LSOA	count
IMD	Index of Multiple Deprivation	ONS	2011, 2021	LSOA	ranking
Travel Mode Data					
Private Car	Number of commuters travelling by car or van (driver or passenger)	ONS Census (Travel to Work)	2011, 2021	LSOA	count
Public Transport	Number of commuters using bus, train, or tram	ONS Census (Travel to Work)	2011, 2021	LSOA	count
Active Travel	Number of commuters walking or cycling to work	ONS Census (Travel to Work)	2011, 2021	LSOA	count
Work From Home	Number of residents working mainly from home	ONS Census (Travel to Work)	2011, 2021	LSOA	count
Other Modes	Number of commuters using motorcycle or other means	ONS Census (Travel to Work)	2011, 2021	LSOA	count
Infrastructure Data					
Bus Stops	Number and location of bus stops	OpenStreetMap	2021	Point	count
Rail/Tram Stations	Number and location of rail or tram stations	OpenStreetMap	2021	Point	count
Rail/Tram Lines	Length of rail or tram lines	OpenStreetMap	2021	Line	km

transport planning and policy-making.

4 Methods

4.1 Introduction

This chapter presents the methodological framework used to study how spatial and socioeconomic factors shape commuting in Greater Manchester. The analysis combines geospatial techniques with machine learning methods to detect patterns, measure relationships, and predict outcomes using the datasets described in Section 3. Geospatial analysis adds spatial context and highlights locational dependencies. Machine learning models capture complex and non-linear interactions between different factors and commuting behaviours. Together, these methods provide both spatial insights and predictive tools for understanding travel behaviour.

4.2 Geospatial Analysis

4.2.1 Spatial Join

Spatial join techniques were applied to integrate travel mode, socioeconomic, and geographical datasets at the Lower Layer Super Output Area (LSOA) level. To maintain accuracy, all datasets were first projected into the same coordinate reference system. Point-based infrastructure data, such as the locations of bus stops, rail stations, and tram stops, were then linked to LSOA polygon boundaries. This step made it possible to calculate the total number of each type of facility in every LSOA. The land area of each LSOA (in km^2) was also extracted from its polygon geometry. Using these measures, facility density was derived as the number of bus stops or rail stations per square kilometre. These indicators were combined with travel behaviour and socioeconomic variables to form a single analytical dataset. In this way, the analysis was able to examine not only how many transport facilities were available in an area, but also how accessibility interacts with social and economic conditions to shape commuting patterns.

4.2.2 Nearest Neighbour Analysis

Nearest neighbour analysis (NNA) was used to evaluate both the accessibility and equity of public transport services in Greater Manchester. Earlier work by Shanmukhappa, Ho, and Tse (2018) showed how spatial methods based on network theory can provide useful insights for bus service planning. Building on this, the analysis in this study generated random points within each LSOA to represent potential trip origins. For each point, the Euclidean distance to the nearest bus stop or rail station was calculated. These distances were then summarised at the LSOA level using mean distance, variance, and range. The mean distance provides an indication of overall accessibility. Variance highlights how evenly services are distributed across the area. A high variance suggests that some residents may be well served while others live much further from facilities. The range helps capture extreme values, showing whether certain locations fall far outside typical access distances. Together, these measures offer a picture of both the level of provision and the fairness of distribution. They also give insight into how differences in accessibility within an area might influence decisions to use public transport or to rely on private and alternative modes.

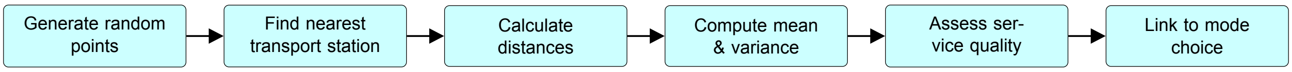

Fig. 2. Nearest neighbour analysis workflow used in this study

4.2.3 Geographically Weighted Regression

Geographically Weighted Regression (GWR) was employed to investigate spatial heterogeneity in the relationship between commuting behaviours and a set of explanatory factors, with particular attention to the role of local deprivation levels. Wheeler (2019) introduced the mathematics foundation of GWR and its usages. Mathematically, the GWR model can be expressed as:

$$y_i = \beta_{0i} + \sum_{k=1}^{p-1} \beta_{ki} x_{ki} + \varepsilon_i, \tag{1}$$

where y_i is the dependent variable (e.g., proportion of residents in LSOA i commuting by a given mode), x_{ki} is the value of the k-th explanatory variable for observation i, β_{0i} is the intercept for location i, β_{ki} is the local regression coefficient for the k-th variable at location i, and ε_i is the random error term.

In this study, GWR was implemented at the LSOA level, with travel mode shares (e.g., proportion of residents commuting by public transport, active travel, or private car) as dependent variables. Explanatory variables included population density, employment composition, median household income, proximity to public transport nodes, and the Index of Multiple Deprivation (IMD) as a measure of area-level socioeconomic disadvantage. Prior to modelling, all variables were standardised to enable direct comparison of coefficient magnitudes and to mitigate issues arising from differences in measurement units.

4.3 Machine Learning Methods

4.3.1 Data Preprocessing

Prior to modelling, the dataset underwent a structured preprocessing workflow to ensure data quality and maintain the interpretability of subsequent models. Missing values were imputed using the mean values from geographically adjacent LSOAs, under the assumption that spatial proximity is associated with similar socioeconomic and transport characteristics. This spatially informed imputation preserved the completeness of the dataset while minimising the introduction of bias.

Factor Analysis (FA) was then applied to explore the latent structure among variables, providing an initial understanding of how different indicators relate to each other. Multicollinearity was assessed through Variance Inflation Factors (VIF), and attributes with high collinearity were removed to improve model stability and interpretability. Principal Component Analysis (PCA) was deliberately not employed, as the primary research objective required retaining the original meaning of explanatory variables rather than transforming them into composite components.

Following the removal of highly collinear variables, travel behaviour indicators were pre-classified into thematic categories based on their relevance to different mode choices. An initial classification of travel modes was performed to identify broad behavioural patterns, and a subset of the most influential predictors was selected as the final analytical indicators. This preprocessing ensured that the dataset was complete, free from redundancy, and composed of interpretable variables with a clear theoretical link to commuting mode choice.

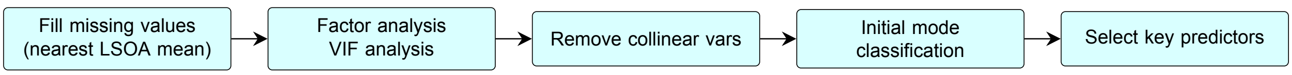

Fig. 3. Pre-processing workflow used in this study

4.3.2 Single Decision Tree

As Myles et al. (2004) introduced, decision trees are supervised learning algorithms that classify observations by recursively splitting the dataset into subsets based on the feature and threshold that maximise a chosen criterion, such as information gain or Gini impurity. At each internal node, the algorithm selects the most informative feature for the split, creating a hierarchy of decision rules until a stopping condition is met, such as a maximum tree depth or minimum number of samples per leaf.

Figure 4 illustrates a generic decision tree structure as introduced by Yan-yan and Ying (2015). In this diagram, the root node represents the initial decision based on feature X_1, with subsequent branches

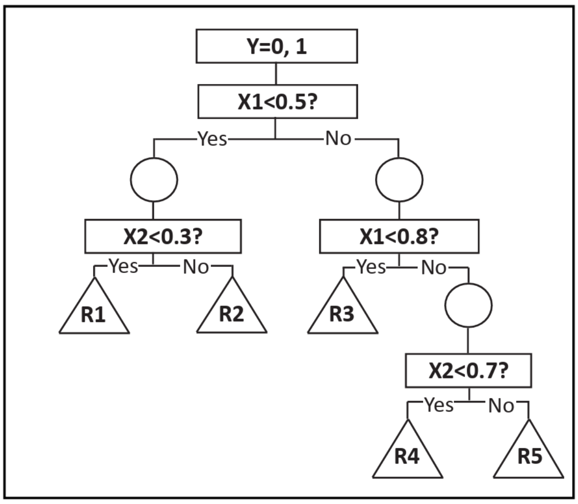

Fig. 4. Basic structure of a decision tree

representing further splits on other features (e.g., X_2). The terminal leaf nodes (R1–R5) represent the final predicted classes. In our application, these classes correspond to different modes of travel, such as private car, public transport, or active travel, allowing a direct interpretation of how the model arrives at each classification.

In this study, the decision tree model was used to classify the dominant commuting mode for each LSOA based on socioeconomic and accessibility predictors. The predictors included variables such as the Index of Multiple Deprivation (IMD) score, median income, employment composition, and public transport accessibility measures (e.g., density of stops/stations, mean and variance of nearest stop distance). The output leaf nodes represent predicted travel modes, while the internal nodes represent decision rules based on the predictor variables. This method offers high interpretability, enabling the identification of explicit rules that link socioeconomic disadvantage and accessibility to mode choice patterns.

4.3.3 Random Forest

Random forest regression, an ensemble method based on multiple decision trees, was applied to improve prediction accuracy and generalisation. Each tree was trained on a bootstrap sample of the dataset, and feature selection at each split ensured decorrelation among trees. The final prediction was obtained by averaging across all trees. Feature importance scores derived from the model were used to identify the most influential predictors of commuting behaviour. The scheme of random forest is introduced by Liu, Wang, and J. Zhang (2012) as shown in Figure 5, where multiple decision trees are trained in parallel on different subsets of the data, and their outputs are combined through majority voting or averaging to produce the final result.

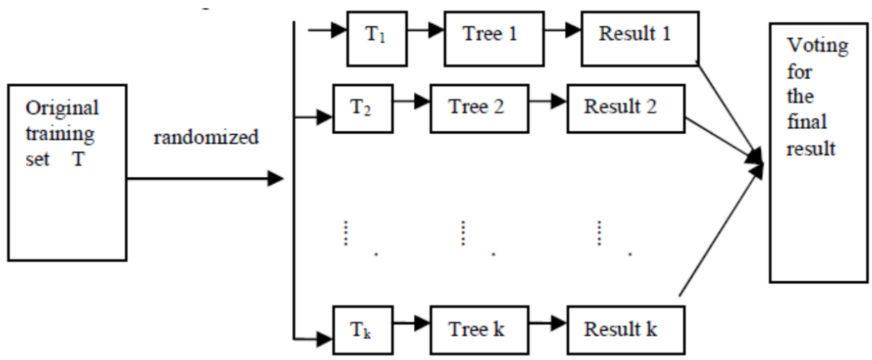

Fig. 5. Scheme of random forest

In the context of this study, random forest regression was employed to model the relationship between a range of socioeconomic, spatial, and accessibility variables and the observed commuting mode shares at the LSOA level. This approach allowed for capturing complex, non-linear interactions between predictors such as the Index of Multiple Deprivation (IMD) score, employment composition, and public transport accessibility measures (including the mean and variance of nearest-station distances). By aggregating the outputs of multiple decision trees, the model provided robust predictions while reducing overfitting risks. Moreover, the derived feature importance rankings offered valuable insights into which factors most strongly influence travel mode choice, helping to better understand commuting behaviours and potential inequalities in transport service provision.

4.3.4 Comparing to Deep Learning Methods

To benchmark performance, a set of traditional machine learning models was compared with a feed-forward deep neural network (DNN). The DNN was designed with several fully connected layers, each followed by a Rectified Linear Unit (ReLU) activation function. This structure introduced non-linearity and allowed the model to capture complex relationships between predictors and commuting behaviour. To reduce the risk of overfitting, dropout regularisation was applied between layers, which randomly deactivated a proportion of neurons during training. The model was trained with the Adam optimiser, which uses an adaptive learning rate to speed up convergence. Early stopping was also introduced, so that training ended once validation loss stopped improving.

Model performance was assessed using root mean square error (RMSE) and the coefficient of determination (R^2). These metrics provided a clear basis for comparing accuracy across all models, showing both the size of prediction errors and the explanatory power of each approach. The results, discussed further in Section 2.4, showed that the DNN was able to capture non-linear and complex patterns in the data more effectively than many baseline models. However, it also displayed signs of overfitting, which is likely linked to the relatively small sample size and the high dimensionality of the features. This highlights a common trade-off: deep learning models can provide powerful predictive capabilities, but they require larger and more balanced datasets to reach their full potential in transport behaviour research.

5 Results and discussion

5.1 Geospatial Analysis

5.1.1 Spatial Join

Spatial join techniques were applied to integrate public transport infrastructure data from OpenStreetMap with the 2021 LSOA boundaries for Greater Manchester. All bus stops, rail stations, and tram stations were represented as point features, while LSOA boundaries were polygon features. The join operation assigned each transport facility to the LSOA polygon in which it was located.

Following the join, the density of bus stops (per km^2) was calculated by dividing the total station count within each LSOA by its area. Figure 6 illustrates the spatial distribution of bus stop density across Greater Manchester. Areas shaded in darker red exhibit a higher density of bus stops, typically corresponding to urban centres and transport interchanges, while lighter shades indicate more sparsely served suburban or rural areas. The map clearly shows that Manchester city centre has the highest accessibility and service quality, followed by other urban areas such as Bolton and Oldham. In contrast, areas such as Wigan and Trafford display relatively lower accessibility, reflecting more limited bus stop coverage. This spatial representation provided a key accessibility measure for subsequent analysis of commuting behaviour and travel mode choice.

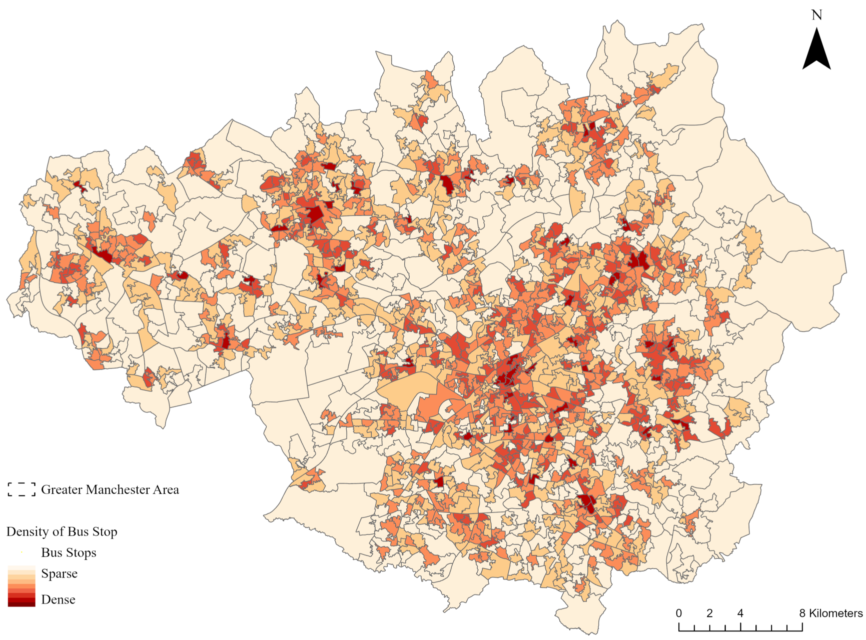

Fig. 6. Density Map of Bus Stops in GMA

5.1.2 Nearest Neighbour Analysis

To complement the density-based accessibility measure, a nearest neighbour analysis was conducted to calculate the mean distance from each LSOA centroid to the nearest bus stop. This metric provides an additional perspective on public transport accessibility, capturing not only the concentration of stops but also their spatial proximity to residents. Figure 7 presents the spatial variation in mean distance to the nearest stop across Greater Manchester. Areas shaded in lighter colours indicate shorter average walking distances, reflecting higher accessibility, whereas darker shades represent areas where residents must travel further to access bus services. The results reveal that Manchester city centre and surrounding inner-urban districts have the closest proximity to bus stops, followed by urban centres in Bolton and Oldham. Conversely, peripheral areas such as Wigan, Trafford, and parts of the northern fringe show significantly longer distances to the nearest stop, suggesting relatively poorer service coverage.

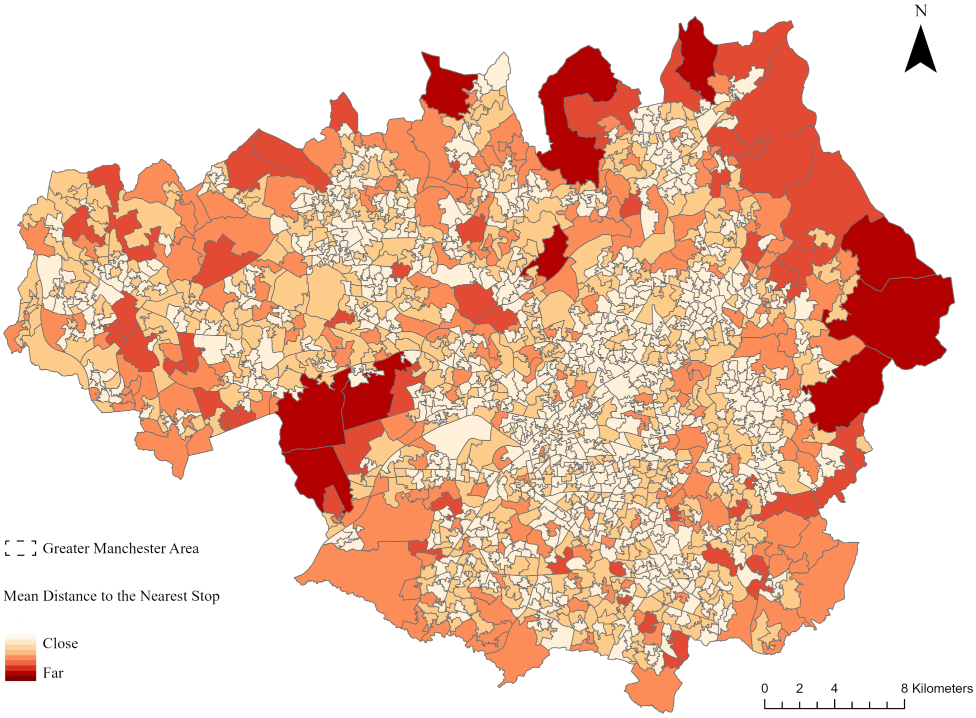

Fig. 7. Mean distance to the nearest public transport stop by LSOA in Greater Manchester

Table 2. Population, transport facilities, and proximity measures by LSOA (simplified)

LSOA21CD	Population	Rail Users	Bus Users	Mean Dist. Bus	Var Dist. Bus	Mean Dist. Rail	Var Dist. Rail
E01004766	677	9	22	0.067186	0.002326	1.405870	0.007757
E01004767	876	17	24	0.078055	0.006081	0.737835	0.085786
E01004768	842	4	5	0.280072	0.049933	2.610748	0.007388
E01004769	773	6	11	0.050218	0.001121	0.890742	0.016210
E01004770	638	7	11	0.149449	0.018753	0.768247	0.014055

5.1.3 Geographically Weighted Regression

To explore the spatially varying relationship between socio-economic deprivation and public transport usage, a Geographically Weighted Regression (GWR) was conducted using the Index of Multiple Deprivation (IMD) as the explanatory variable and the proportion of commuters using public transport as the dependent variable. Unlike Ordinary Least Squares (OLS), which assumes a constant relationship across space, GWR estimates location-specific coefficients, allowing for the identification of local variations in the IMD–public transport use relationship.

Figure 8 presents the standardized residuals from the GWR model. Residuals close to zero (white to light shading) indicate areas where the model fits well, whereas positive residuals (green shades) represent LSOAs where public transport use is higher than predicted by local deprivation levels, and negative residuals (purple shades) indicate areas with lower-than-expected public transport use.

The spatial distribution shows that central Manchester and parts of Oldham and Rochdale exhibit higher-than-expected bus usage given their deprivation levels, possibly reflecting the availability and convenience of transit services in these urban centres. In contrast, areas in Trafford, Wigan, and the outer periphery tend to have lower-than-expected public transport usage, which may be linked to lower service frequency or a stronger reliance on private vehicles. These findings highlight that while deprivation is a significant factor influencing public transport uptake, local service provision and accessibility play an equally critical role in shaping travel behaviour.

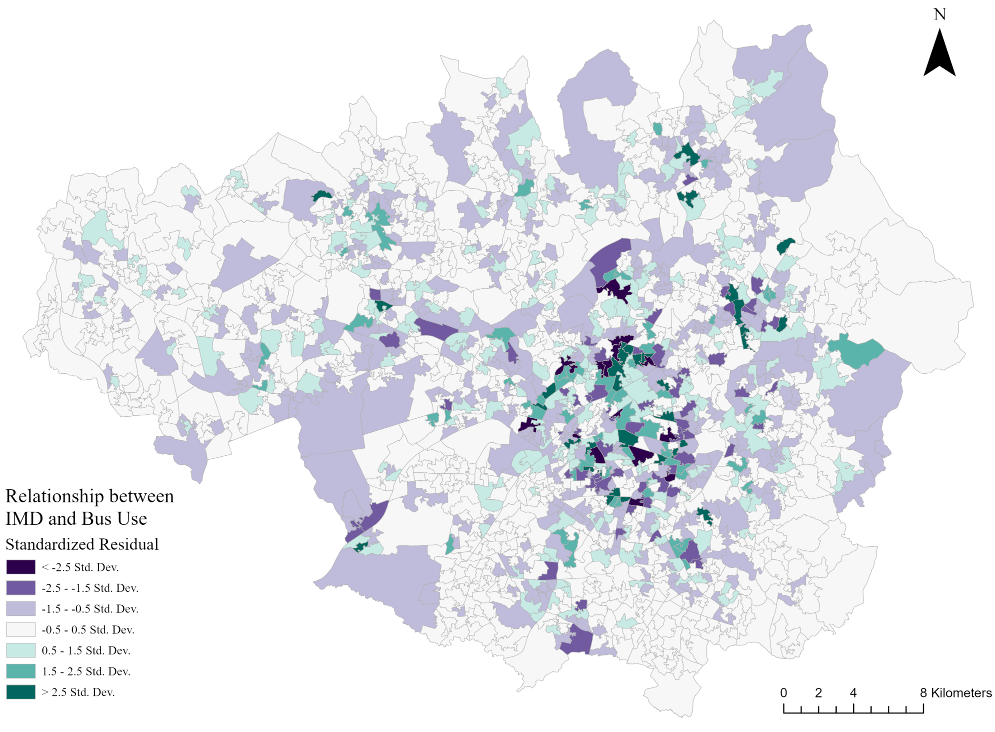

Fig. 8. Standardized residuals from GWR model relating IMD to public transport use

5.2 Commuter Behaviour Analysis

5.2.1 Multi-collinearity Analysis

Table 3 reports the VIF results for selected variables after screening from over one hundred candidate attributes. The table is ordered by VIF in descending order, where variables at the lower end indicate weaker multicollinearity and therefore greater explanatory importance in subsequent modelling.

The analysis demonstrates that variables with lower VIFs like industry and population density are relatively independent and thus provide stronger interpretive power in explaining travel mode choice. By contrast, variables with higher VIFs show stronger collinearity with other demographic indicators and should be treated with caution. Importantly, the results highlight that commuting distance, occupation, economic deprivation, and population density emerge as robust, non-redundant predictors of mode choice.

In the results, a large proportion of variables exhibited excessively high VIF values, indicating severe multicollinearity and limiting their explanatory power. For the subsequent behavioural analysis, only variables with comparatively low VIF values were retained as explanatory indices. These include the IMD as a proxy for socioeconomic status, industry categories, commuting distance bands, specific household structures, and the proportion of elderly residents. This ensures that the retained variables are relatively independent, thus providing more robust and interpretable insights into the determinants of travel mode choice.

Table 3. Variance Inflation Factor (VIF) results for selected variables (ordered by importance)

Feature	VIF	Rank (Importance)
Aged 65 years and over	98.88	14
8 or more people in household	69.89	13
L Real estate activities	54.68	12
7 people in household	47.38	11
Other method of travel to work	35.04	10
60km and over	32.17	9
Index of Multiple Deprivation (IMD)	31.07	8
30km to less than 40km	27.37	7
40km to less than 60km	18.85	6
E Water supply; sewerage, etc.	14.87	5
D Electricity, gas, etc.	12.12	4
population density	7.60	3
A Agriculture, forestry and fishing	2.86	2
B Mining and quarrying	1.68	1

5.2.2 Single Decision Tree

The decision tree (Figure 9) illustrates the key socio-economic and spatial factors shaping public transit use. The *Index of Multiple Deprivation (IMD)* emerges as the primary split variable, highlighting that regional deprivation levels strongly explain public transit dependence. Areas with lower IMD scores (less deprived) tend to show higher reliance on public transport, while more deprived areas demonstrate lower levels of usage.

Following IMD, *commuting distance* becomes a decisive factor, with long-distance commuters exhibiting higher probabilities of public transit use. *Population density* also appears repeatedly as a splitting variable, indicating that residents in high-density areas are more likely to adopt public transit due to greater accessibility and service provision.

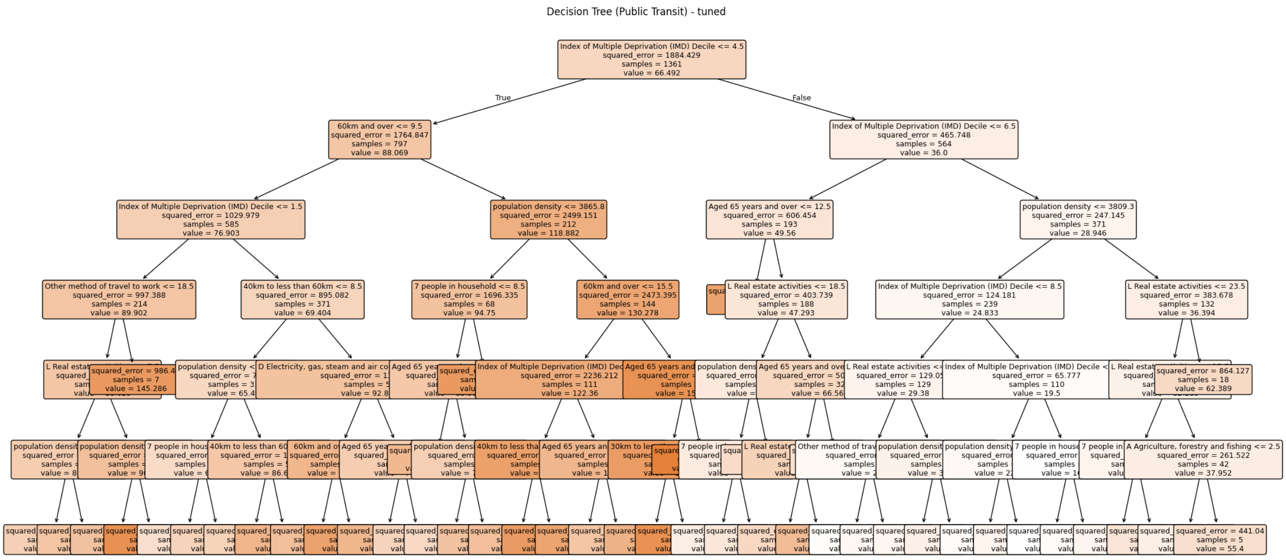

Fig. 9. Single Decision Tree (Public Transit as an example, 2021)

At deeper levels of the tree, additional conditional factors emerge, including *household size* , the proportion of the population *aged 65 and over*, and the share of employment in certain *industries*. These factors capture marginal differences, suggesting that multi-person households and older populations are more likely to depend on collective modes under certain conditions.

Overall, the decision tree confirms three major findings from 2021: (1) socio-economic status (IMD) is the dominant explanatory factor; (2) commuting distance and population density jointly structure mode choice; (3) household composition and demographic attributes exert additional influence on public transit use in specific contexts.

To illustrate the marginal effect of a single variable on mode choice, a univariate decision tree was trained using only *commuting distance*(banded) to predict the dominant travel mode at the LSOA level for 2011(Figure 10). The tree partitions the population with very low node impurity, showing that distance alone already explains much of the variation in chosen modes.

Overall, the univariate tree confirms a strong, monotonic association between distance and mode in 2011: bus/minibus/coach dominates across most distance bands, active or rail modes appear only in small terminal nodes, and taxi emerges as a highly localised option around 5–10 km. Because the model is intentionally restricted to a single predictor, these patterns represent the *direct* distance–mode relationship without confounding from socioeconomic or accessibility covariates; later multivariate models show how deprivation, population density and industry composition further modulate these distance effects.

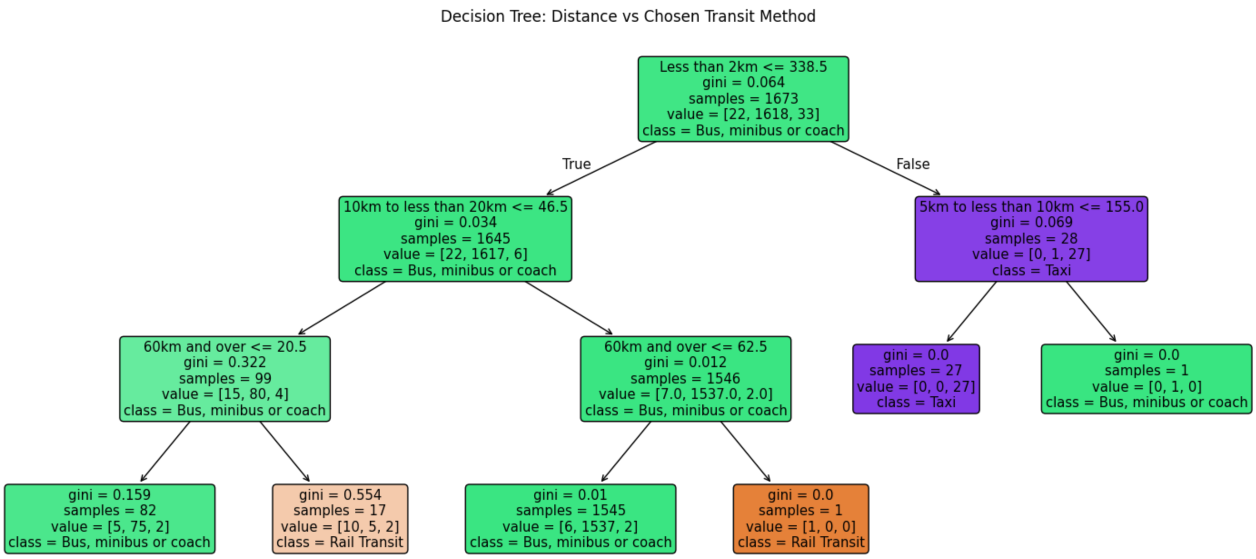

Fig. 10. Univariate decision tree: commuting distance as example, 2011

5.2.3 Random Forest

Figure 11 visualises the dependency structure between household characteristics and transit methods using a covariance-based network. Edge thickness and colour indicate the strength and direction of association. Households without cars are strongly linked to public and active transit, while multi-car households are most closely associated with private car use.

Regarding commuting distance, short trips are connected with active travel, medium distances correspond to car use, and long-distance commuting displays more dispersed patterns with weaker ties across modes. This graph provides an intuitive representation of how car ownership and distance jointly influence travel mode choice.

To exemplify the explanatory power of machine learning in travel behaviour research, a Random Forest model was trained on the 2021 dataset, as shown in Table 4. This approach enables not

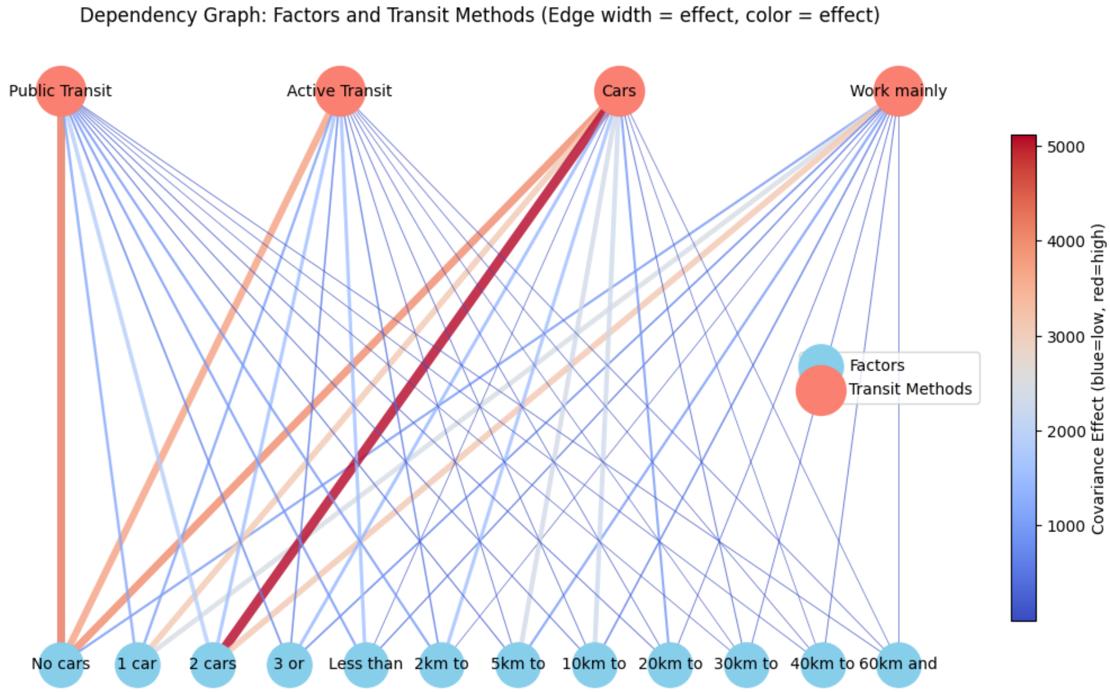

Fig. 11. Dependency graph between household and commuting factors with transit methods

only robust prediction of mode choice but also quantification of the relative contribution of socio-demographic and health variables through feature importance measures. The results demonstrate that age, gender, and health status are decisive factors in shaping modal preferences. Male respondents and those around 50 years of age show strong associations with public transit and private car use, while younger cohorts (16–25) are distinctly aligned with active transit methods.

Table 4. Top 5 Feature Importances for Each Travel Mode

Feature	Public Transit	Active Transit	Car	Work Mainly
Male	0.32	–	0.05	–
Aged 50	0.22	0.06	0.47	0.07
Aged 25	0.08	0.22	–	0.41
Bad health	0.07	0.08	–	0.11
Aged 16	0.06	0.34	0.18	–
Good health	–	–	0.06	–
Aged 35	–	–	0.05	0.22
Very good health	–	0.06	–	–
Fair health	–	–	–	0.07

Figure 12 illustrates how socio-demographic and health factors shape public transit usage. Younger groups (16–34) are the most frequent users, while reliance declines sharply among those aged 65 and over. Health status also matters: individuals in good or very good health show greater engagement, whereas those in poor health rely significantly less on public transit. These results confirm that age and health are decisive in shaping transit demand, underscoring the importance of enhancing accessibility

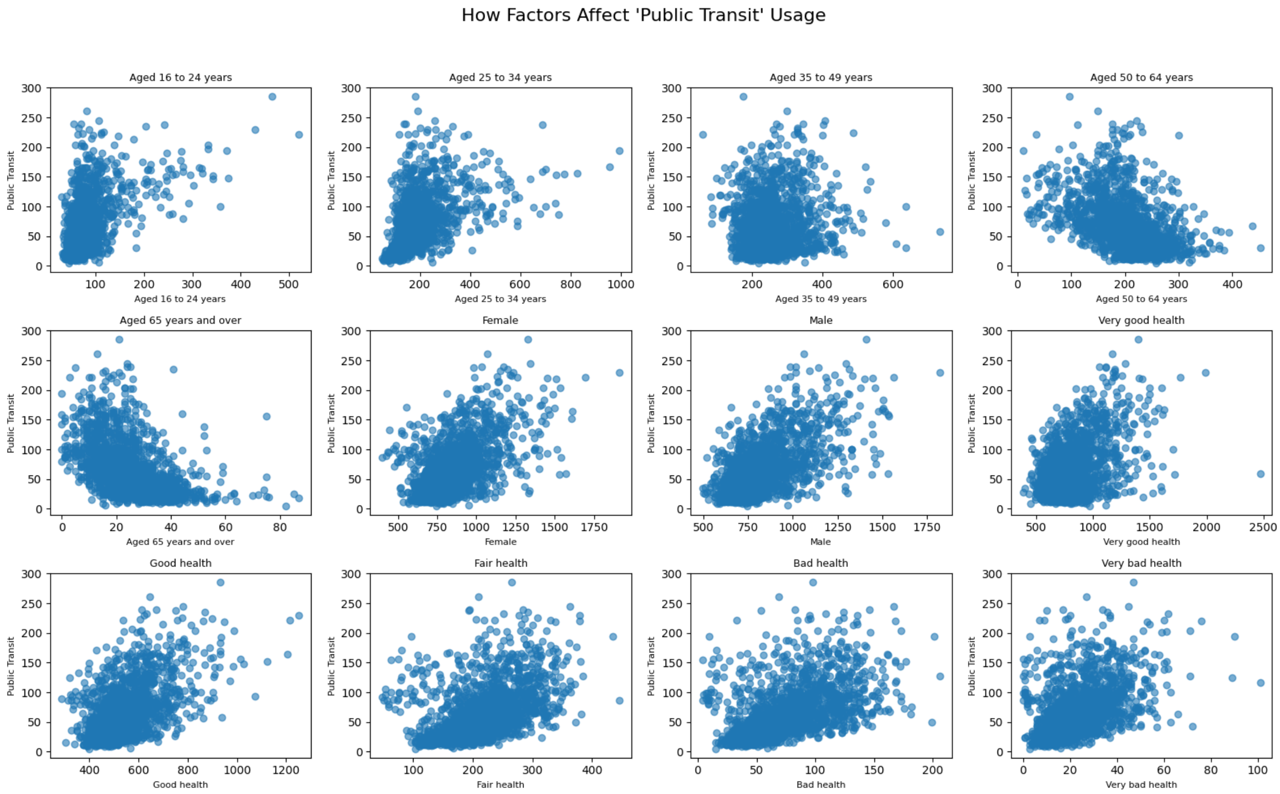

Fig. 12. How chosen factors effect the choice of using public transit

for older and health-disadvantaged groups.

Table 5. Top 5 Occupational Influences on Travel Mode Choice (2021, Random Forest Results)

Rank	**Public Transit**	**Active Transit**	**Car Use**	**Work Mainly at Home**
1	Accommodation	Accommodation	Construction	Professional
2	Administrative	Education	Wholesale	Information
3	Construction	Professional	Manufacturing	Public
4	Transport	Information	Human	Education
5	Public	Human	Transport	Financial

The analysis of occupational categories shown in Table 5 highlights distinct patterns in the determinants of travel mode choice. For public transit, employees in the accommodation sector overwhelmingly dominate, suggesting a strong reliance on collective transport among workers engaged in service and hospitality roles. Administrative and transport-related occupations also exhibit moderate importance, reflecting their structural dependency on scheduled commuting. By contrast, active transit is most influenced by accommodation, education, and professional services, with the prominence of education and professional roles pointing to the role of urban centrality and shorter commutes that can feasibly be undertaken by walking or cycling.

Car use, in turn, is most strongly associated with the construction sector, far exceeding any other occupational influence. This reflects the dispersed and site-specific nature of construction activities, where flexibility and geographic reach make private vehicle use a necessity. Finally, working mainly from home is dominated by professional, information, and financial sectors, aligning with their higher digitalisation and capacity for remote work. Together, these results confirm that occupational structure acts as a critical stratifier of transport mode choices, with clear sectoral distinctions between collective, individual, and remote forms of mobility.

5.3 Comparison between 2011 and 2021

According to Figure 13 and Figure 14, between 2011 and 2021, car use fell from 64.5% to 52.2 (-12.3 pp), public transit from 15.0% to 9.1 (-5.9 pp), and active travel from 12.7% to 9.8 (-2.9 pp). Meanwhile, work mainly at or from home rose sharply from 7.8% to 28.9 ($+21.1$ pp). Decision trees highlight deprivation (IMD), commuting distance, and density as core drivers of mode choice.

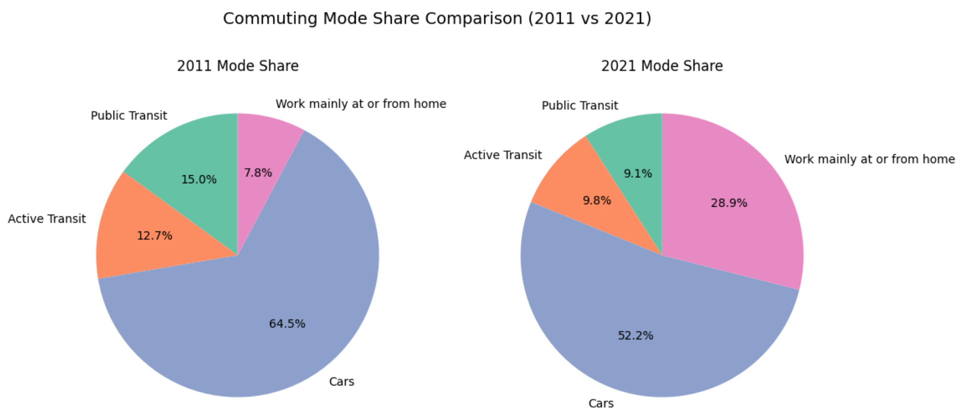

Fig. 13. Mode Choice comparison in 2011 and 2021

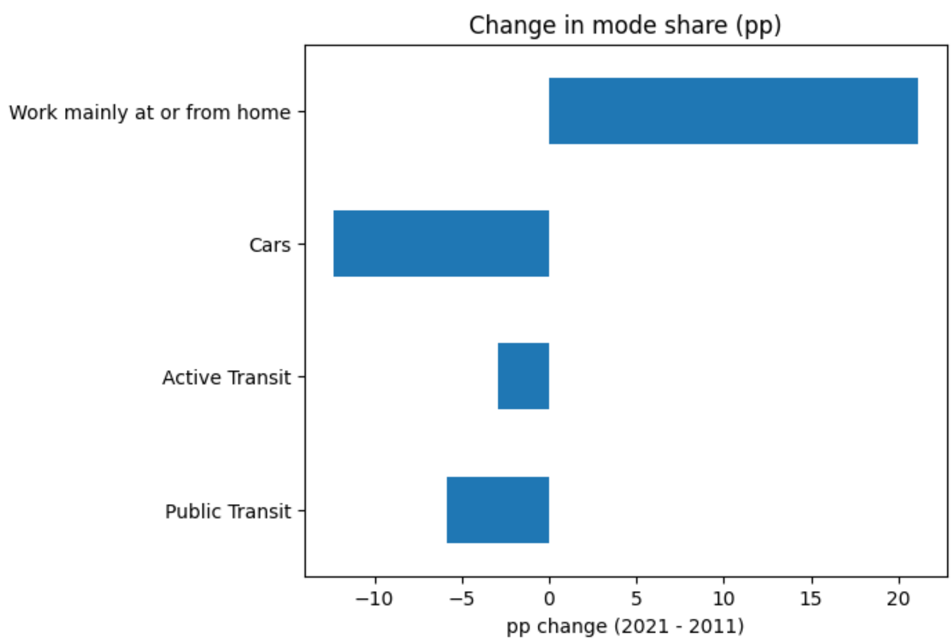

Fig. 14. Mode Choice comparison in 2011 and 2021 in change of percentage

Random forest results in Section 5.2.3 using 2021 data add nuance: **public transit** is linked to service jobs and middle ages; **active travel** to younger groups and education/professional roles; **car use** to older cohorts and construction/manufacturing; and **home working** to professional/information sectors. These patterns align with our distance splits and VIF predictors. Bus stop density and access remain uneven: central Manchester and Oldham are well served, while Wigan and Trafford are sparse. GWR shows that deprivation alone does not explain bus use—accessibility and centrality matter strongly.

As for the rise of home working during the COVID-19 pandemic, is closely aligned with occupational patterns, as professional and information workers were able to shift heavily to remote work. This structural change reduced overall commuting demand across all modes. Public transit experienced the sharpest decline, not only due to reduced commuting volumes but also because of service suspensions and heightened infection risk. Car use also fell, as the reduction in daily commuting outweighed any potential substitution effect. Meanwhile, active travel did not expand significantly, since commuting distances and the suburban form of Greater Manchester constrained its feasibility despite increased public awareness of health and outdoor mobility during the pandemic.

5.4 Summary and discussion

The findings reveal clear spatial and temporal patterns of public transport use in Greater Manchester. Bus stop density and nearest-stop distance show a strong urban–suburban divide: central Manchester and Oldham are well served with dense, proximate networks, whereas Wigan and Trafford display sparse coverage and poorer accessibility. These spatial inequalities translate directly into mode choice, where residents of high-density and central areas are more likely to use public transit, while those in peripheral districts rely more on cars.

Between 2011 and 2021, travel behaviours shifted substantially. Car, public transit, and active travel all declined, while home working rose sharply, reflecting the impact of the COVID-19 pandemic. Decision tree and random forest models highlight that socio-economic deprivation (IMD), commuting distance, and population density remain the dominant structural factors, with occupational roles further stratifying choices: service and hospitality workers are tied to public transit, construction to car use, and professional/information sectors to remote work. The pandemic amplified these divides, with infection risks and service suspensions accelerating declines in bus and rail, while suburban distances limited any expansion of active travel.

To make the statistical results more tangible, this section illustrates how individual attributes, sector of employment, location, and commuting distance interact to shape travel mode choice. These examples are derived from the GWR, decision tree, and random forest analyses, and are consistent with the spatial accessibility patterns revealed in bus stop density and nearest-neighbour analysis. Rather than treating factors in isolation, the following examples demonstrate how socio-economic and spatial contexts combine to produce distinct mode preferences.

For instance, service workers located in central or sub-centre areas with dense bus networks and short average distances to the nearest stop are strongly inclined towards public transport. By contrast, construction workers in peripheral districts such as Wigan or Trafford, where accessibility is comparatively weaker, are more dependent on private cars, particularly when commuting distances extend beyond 10 km. Younger cohorts (16–25), especially students and early professionals, exhibit higher reliance on active travel and public transport when living within short-distance, high-accessibility zones. Meanwhile, professional and information workers have the greatest propensity to work from home, a trend accentuated post-pandemic. For older cohorts, age and health factors become critical: car use remains dominant in suburban or peripheral areas, but public transport becomes feasible in central districts where service density offsets individual constraints. Finally, deprivation (IMD) interacts with spatial context: deprived central areas with strong transport supply show higher-than-expected transit use, whereas deprived peripheral areas with poor service provision are associated with suppressed transit uptake and stronger reliance on cars.

Table 6. Case-based factor-to-mode predictions

Profile	Sector	Location	Distance band	Likely mode
Male, 30, good health	Accommodation services	Central/Oldham core	2–10 km	Public transport
Female, 23, student	Education/professional	Central	$<$5 km	Active $\rightarrow$ PT
Male, 52	Construction	Peripheral (Wigan/Trafford)	5–20 km	Car
Female, 34	Professional/Information	Inner urban	0 km (WFH)	Work from home
Household with no car	Services	Central/sub-centres	$<$5 km	PT/Active
Aged 68, fair health	Mixed	Peripheral	5–15 km	Car $\rightarrow$ PT (if central)
High IMD (central)	Services	Central	2–8 km	Public transport
High IMD (peripheral)	Mixed	Peripheral	5–15 km	Car

Overall, the results demonstrate that sustainable mobility in Greater Manchester depends on more than socio-economic context alone. Accessibility, service provision, and occupational structures all shape commuting behaviour, while external shocks such as pandemics can rapidly disrupt established patterns. Insights from geospatial and machine learning analysis therefore highlight the need for policies that reduce spatial inequalities, strengthen public transport resilience, and expand active travel infrastructure to build a more adaptable and sustainable transport system.

6 Conclusions and future work

6.1 Conclusions

This study reveals persistent spatial disparities in public transport accessibility across Greater Manchester. Central areas such as Manchester city centre and Oldham benefit from dense bus networks and short walking distances to stops, while peripheral areas—most notably Wigan and Trafford—remain under-served (see Figures 6, 7). These inequalities translate into commuting behaviour: high-density, central zones show greater public transport use, whereas peripheral zones lean more heavily on car travel.

Between 2011 and 2021, a marked reconfiguration of commuting occurred, largely shaped by the COVID-19 pandemic. While car, public transit, and active travel all declined, home working surged from under 8% to nearly 29%. Machine learning models corroborate this change: public transit use remains anchored in service and hospitality sectors and middle-aged groups; active travel aligns with younger cohorts and professional/education sectors; car use is concentrated among older workers in construction/manufacturing; and remote work is dominated by professional and information sectors (Tables 4, 5). Spatial regression results further show that transport-deprived central areas outperform expectations due to better infrastructure, while deprived peripheries fall below predictions (Figure 8).

These findings carry direct relevance for the Greater Manchester Transport Strategy 2040 and the evolving Bee Network. The ambition for 50% of journeys to be by walking, cycling, or public transport by 2040 Greater Manchester (n.d.) aligns with our findings: central areas are already moving in this direction, but peripheries require investment. The ongoing re-regulation of the bus network, fare capping, and expansion of the integrated Bee Network (including active travel components), offer structural remedies to the service inequalities we identified. However, successful policy must also account for socio-economic and occupational dynamics—for instance, supporting remote working in professional sectors while investing in transit for car-dependent occupations.

6.2 Future work

Future research should extend the analysis beyond 2021 to capture mid-pandemic and recovery phases, allowing assessment of resilience and behavioural inertia over time. In parallel, quasi-experimental approaches could be employed to evaluate the impacts of Bee Network policies such as fare caps, franchising, and the introduction of electric fleets on mode shift and equity. A more comprehensive perspective on multimodal accessibility is also needed, incorporating walking, cycling, tram, and rail travel into a composite index that reflects the integrated mobility vision of the Bee Network.

Furthermore, the use of fine-grained data sources, such as mobile phone or smart card records, could enable detailed disaggregation of commuter profiles, trip purposes, and mode choices, particularly in peripheral areas. Finally, scenario-based modelling of proposed interventions, including the underground rail network and orbital bus routes, would provide valuable insights into how different socio-economic and climate policy trajectories may reshape future travel behaviour in Greater Manchester.

In summary, this research demonstrates how combined spatial and machine learning methods can deliver actionable insights for sustainable transport policy in Greater Manchester. By integrating infrastructure, socio-economic, and mode choice dynamics, it supports targeted interventions that promote equity and resilience in urban mobility.

Word count: 7402

References

Arentze, T. et al. (2000). “ALBATROSS: Multiagent, Rule-Based Model of Activity Pattern Decisions”. In: *Transportation Research Record* 1706.1, pp. 136–144. DOI: 10.3141/1706-16.

Bhaduri, E. et al. (2020). “Modelling the effects of COVID-19 on travel mode choice behaviour in India”. In: *Transportation Research Interdisciplinary Perspectives* 8, p. 100273. DOI: 10.1016/j.trip.2020.100273.

Bishop (2006). *Pattern Recognition and Machine Learning*. urlhttps://link.springer.com/book/9780387310732, accessed 22 June 2025.

Boarnet, M. G. and R. Crane (2001). *Travel by Design: The Influence of Urban Form on Travel*. Oxford University Press.

Buliung, R. N. and P. S. Kanaroglou (2007). “Activity–Travel Behaviour Research: Conceptual Issues, State of the Art, and Emerging Perspectives on Behavioural Analysis and Simulation Modelling”. In: *Transport Reviews* 27.2, pp. 151–187. DOI: 10.1080/01441640600858649.

De Vos, J., Patrick A. Singleton, and T. Gärling (2022). “From attitude to satisfaction: introducing the travel mode choice cycle”. In: *Transport Reviews* 42.2, pp. 204–221. DOI: 10.1080/01441647.2021.1958952.

Dieleman, F. M., M. Dijst, and G. Burghouwt (2002). “Urban Form and Travel Behaviour: Micro-level Household Attributes and Residential Context”. In: *Urban Studies* 39.3, pp. 507–527. DOI: 10.1080/00420980220112801.

Golob, T. F. (2003). “Structural equation modeling for travel behavior research”. In: *Transportation Research Part B: Methodological* 37.1, pp. 1–25. DOI: 10.1016/S0191-2615(01)00046-7.

Greater Manchester, Transport for (n.d.). *Greater Manchester Transport Strategy 2040*. Greater Manchester Transport Strategy 2040 | Bee Network | Powered by TfGM. Accessed 20 August 2025. URL: https://tfgm.com/2040-transport-strategy.

Hagenauer, J. and M. Helbich (2017). “A comparative study of machine learning classifiers for modeling travel mode choice”. In: *Expert Systems with Applications* 78, pp. 273–282. DOI: 10.1016/j.eswa.2017.01.057.

Jones, P. M. et al. (1987). *Understanding Travel Behaviour*. 2a. reimp. Aldershot, UK: Gower.

Kitamura, R., P. L. Mokhtarian, and L. Laidet (1997). “A micro-analysis of land use and travel in five neighborhoods in the San Francisco Bay Area”. In: *Transportation* 24.2, pp. 125–158. DOI: 10.1023/A:1017959825565.

Liu, Y., Y. Wang, and J. Zhang (2012). “New Machine Learning Algorithm: Random Forest”. In: *Information Computing and Applications*. Ed. by B. Liu, M. Ma, and J. Chang. Berlin, Heidelberg: Springer, pp. 246–252. DOI: 10.1007/978-3-642-34062-8_32.

Myles, A. J. et al. (2004). “An introduction to decision tree modeling”. In: *Journal of Chemometrics* 18.6, pp. 275–285. DOI: 10.1002/cem.873.

Schwanen, T., M. Dijst, and F. M. Dieleman (2004). “Policies for Urban Form and their Impact on Travel: The Netherlands Experience”. In: *Urban Studies* 41.3, pp. 579–603. DOI: 10.1080/004209804200017869

Shanmukhappa, T., I. W.-H. Ho, and C. K. Tse (2018). “Spatial analysis of bus transport networks using network theory”. In: *Physica A: Statistical Mechanics and Its Applications* 502, pp. 295–314. DOI: 10.1016/j.physa.2018.02.111.

Simma, A. and K. W. Axhausen (2001). “Structures of commitment in mode use: a comparison of Switzerland, Germany and Great Britain”. In: *Transport Policy* 8.4, pp. 279–288. DOI: 10.1016/S0967-070X(01)00023-3.

Thombre, A. and A. Agarwal (2021). “A paradigm shift in urban mobility: Policy insights from travel before and after COVID-19 to seize the opportunity”. In: *Transport Policy* 110, pp. 335–353. DOI: 10.1016/j.tranpol.2021.06.010.

Wheeler, D. C. (2019). “Geographically Weighted Regression”. In: *Handbook of Regional Science*. Springer, Berlin, Heidelberg, pp. 1–27. DOI: 10.1007/978-3-642-36203-3_77-1.

Xie, C., J. Lu, and E. Parkany (2003). “Work Travel Mode Choice Modeling with Data Mining: Decision Trees and Neural Networks”. In: *Transportation Research Record* 1854.1, pp. 50–61. DOI: 10.3141/1854-06.

Yan-yan, S. and L. Ying (2015). *Decision tree methods: applications for classification and prediction - Shanghai Archives of Psychiatry*. Accessed 15 August 2025. URL: https://shanghaiarchivesofpsychiatry.org/en/215044.html.

Zhang, Yunlong and Yuanchang Xie (n.d.). *Travel Mode Choice Modeling with Support Vector Machines*. https://journals.sagepub.com/doi/abs/10.3141/2076-16, accessed 22 June 2025.

Appendices

A Project Repository

For all the codes, figure outputs and tables please see in the repository on Github.

Click the link: UA_11531463_Repository

Or paste into your browser: `https://github.com/gggaiii/UA_11531463_Repository_UoM_MSc_DS`