The Journal of Transport and Land Use jtlu.org
Vol. 17 No. 1 [2024] pp. 731–750

Association between land use features and changes in walking patterns from pre-pandemic to post-pandemic: A case study of city of Sydney (2013–2023)

Fatemeh Nourmohammadi
University of New South Wales
f.nourmohammadi@unsw.edu.au

Zahra Nourmohammadi
University of New South Wales
z.nourmohammadi@unsw.edu.au

Tanapon Lilasathapornkit
University of New South Wales
t.lilasathapornkit@unsw.edu.au

Meead Saberi
University of New South Wales
meead.saberi@unsw.edu.au

Abstract: While the impact of the pandemic on active mobility patterns is widely studied in several cities, the underlying characteristics that describe the heterogeneity in changes in active mobility are less understood. This is particularly important for post-pandemic active mobility planning. This study aims to investigate and describe the association between urban population and land-use features, as well as changes in the spatio-temporal patterns of walking from pre-pandemic to post-pandemic through a case study of the city of Sydney, Australia, using 11 years of pedestrian count data from 2013 to 2023. The findings indicate that during the pandemic, the average daily pedestrian traffic in Sydney decreased significantly compared to the pre-pandemic period. However, since experiencing the lowest pedestrian traffic in 2020, activities in the study area have shown signs of partial recovery, with a 51% increase observed in 2023. The observed changes in pedestrian activities are, however, spatially heterogeneous. Modeling results reveal that areas with greater commercial land use, more points of interest (POIs), higher population density, and higher network connectivity experienced a significant negative change in the number of walking trips from the pre-pandemic to the pandemic period. Areas with higher percentages of educational and residential use and with higher personal income experienced smaller changes in pedestrian activities during the pandemic compared to the pre-pandemic period. During the post-pandemic recovery, the influential features remain mostly unchanged; however, the association direction is the opposite.

Keywords: Pedestrian, walking, pandemic, land use, Sydney

Article history:
Received: April 12, 2024
Received in revised form: July 2, 2024
Accepted: August 31, 2024
Available online: October 25, 2024

1 Introduction

Active mobility patterns in cities have been significantly affected by COVID-19 pandemic (Angel et al., 2023; Enoch et al. 2022; Möllers et al., 2022; Pareek et al.,

https://doi.org/10.5198/jtlu.2024.2511
ISSN: 1938-7849

The *Journal of Transport and Land Use* is the official journal of the World Society for Transport and Land Use (WSTLUR) and is published and sponsored by the University of Minnesota Center for Transportation Studies.

2020;). People's travel behavior has noticeably changed over the past few years (Currie et al., 2021). Extended and heavily enforced lockdowns during the pandemic forced a large proportion of the population to study and work from home when possible. The pandemic restrictions on mobility have put the significant role of walkability and walkable neighborhoods in the spotlight (Kang et al., 2020). Some studies revealed that physical activity has declined dramatically due to stay-at-home policies during the pandemic, although it did not meet the recommended levels even before the pandemic (Guthold et al., 2018; Tison et al. 2020). Understanding the changes in walking patterns during pre- and post-pandemic periods helps transportation planners and management authorities make more informed decisions and policies about active transportation.

There have been numerous studies on the impact of COVID-19 on overall population mobility at different scales and with different modes (Borkowski et al., 2021; Gkiotsalitis & Cats, 2021; Hintermann et al. 2023; Li & Xu, 2021). Few studies have also explored the changes in walking patterns during the pandemic (Delclòs-Alió et al., 2022; Doubleday et al., 2021; Hunter et al., 2021; Obuchi et al., 2021; Power et al., 2023). However, evidence on how walking patterns have changed before and after the pandemic and whether the decline in pedestrian travel demand has fully recovered has remained inconclusive. This paper aims to provide insights and new evidence into the changes in the spatial and temporal patterns of pedestrian activities affected by COVID-19 in the city of Sydney, Australia as a case study using empirical walking counts over 11 years from 2013 to 2023.

Walkability has been a subject of research from different perspectives including sustainable development, public health, and urban planning (Gao et al., 2020; Loo, 2021; Mavoa et al., 2018; Mendiate et al., 2022; Neves et al., 2021; Tribby et al., 2016; Yin et al., 2023). Despite having a few different definitions, walkability can be broadly described as the degree to which an environment, especially the built environment, is pedestrian-friendly and makes walking easy or pleasant (Hall & Ram, 2019). Research on understanding walkability and walking patterns has made significant progress lately, including the identification of functional urban zones and the detection of urban anomalies in human activity (Stier et al., 2020; Wu et al., 2020). Several experimental investigations using statistical analysis have been carried out to understand walking behavior and estimate pedestrian safety under various walking conditions (Fujita et al., 2019). However, understanding walking behavior requires more than descriptive analysis. A more quantitative methodology such as clustering has been used by Humagain and Singleton (2021) to examine the connections between walking patterns and spatial urban characteristics. Li and Xu (2021) applied clustering to pedestrian hourly count time series to reveal the changing patterns in walking before, during, and after COVID-19 restrictions in Beijing, China.

Another study by Falchetta and Noussan (2020) found that active mobility decreased during the initial stages of restrictions in Australia, with survey respondents indicating their intention to further reduce their active mobility in the future. Following the easing of restrictions, the researchers re-surveyed the participants about their use of active transportation modes and found that they had indeed reduced their usage as confirmed by the respondents. In this study, we focus on pedestrian activities and walking as a less studied and explored mode of transportation. To study the impact of COVID-19 on walking patterns in the city of Sydney, we use 10 years' worth of empirical hourly pedestrian count data collected across more than 100 locations in Sydney's central area from 2013 to 2023. The data includes three periods: pre-pandemic (2013 to 2019), during the pandemic (2020 and 2021), and post-pandemic (2022 and 2023). Subsequently, we conduct spatial clustering and apply different regression analyses to examine and describe the changes in pedestrian activities from one period to another.

The remainder of the paper is organized as follows. The next section presents the methodology and data followed by the results section that presents the time series analysis, spatial clustering, and linear and random forest regression modeling. The final section provides the study conclusion, limitations, and a discussion on policy implications.

2 Methods and data

In this section, we first provide a brief description of the data and the study location, followed by an initial time series analysis of the walking patterns in the city of Sydney's central area. To further understand the spatial patterns of pedestrian activities, we then employ a time series clustering to classify the pedestrian count sites by distinguished peak period patterns of walking trips. Next, we apply Lasso regression to identify the most statistically significant variables for constructing regression models (Fonti & Belitser, 2017). Additionally, we use the Variance Inflation Factor (VIF) to detect multicollinearity among the variables (Montgomery et al., 2021), ensuring it is addressed in the analysis. Then, linear regression, and negative binomial regression models are used to evaluate how different variables contribute to the changes in the walking patterns (Nelder & Wedderburn, 1972). We also develop non-linear random forest models to quantify the extent to which population and land use features influenced changes in the walking patterns over time from the pre-pandemic to the post-pandemic period.

2.1 Data description

We obtained hourly pedestrian count data from the city of Sydney from 2013 to 2023 except for 2018. The data contains the number of pedestrians across a total of 115 sites. The pedestrian counts were collected two times each year (October and March) from 6 am to 11 pm. The pedestrian count data were gathered on typical weekdays (and weekends), ensuring that the counts represent typical pedestrian flows devoid of holiday-related spikes or seasonality influences. While recognizing the significance of weather and holiday variations in pedestrian studies, our analysis remains constrained by the data's temporal granularity, which limits our ability to directly assess these factors' impacts on pedestrian flow within the study periods.

To balance the number of years across distinct periods for modeling purposes and to ensure a meaningful analysis of temporal trends in walking, we use modified time windows: pre-pandemic (2017 and 2019), during the pandemic (2020 and 2021), and post-pandemic (2022 and 2023), with each distinct period including two years. In preparation for the regression modeling, we create a 300 meters buffer around each site. Subsequently, we incorporate land use data such as industrial, hospital/medical, parkland, commercial, residential, educational, transportation, and water obtained from the Australian Bureau of Statistics (ABS) Census 2021. We also obtain the number of Points of Interest (POI) and network connectivity measured through street intersection count and number of links to number of nodes ratio from OpenStreetMap (OSM), average personal weekly income, and population information from ABS Census 2021 at each site. Note that significant changes in land use typically occur gradually over long periods due to extensive planning and regulatory processes (Calderón-Loor et al, 2021). Although individual POIs may emerge or close, the overall distribution and function of POIs within urban settings exhibit a high degree of stability. Blundell et al. (2017) discusses the stability of income distributions over time, noting that while there are annual fluctuations, significant shifts are infrequent over a span of several years. Figure 1 shows the spatial distribution of pedestrian count sites, corridors, and villages in the city of Sydney.

2.2 Temporal walking patterns

To understand the temporal changes in the walking patterns, we compare the average hourly pedestrian counts from 2013 to 2023, implementing a comparative analysis by time of day, day of the week, and seasons across all sites in the study area (See Figure 2). We then use the mean daily pedestrian count across years considering both weekdays and weekends, plus the percentage change in average daily count at each site to understand the differences between weekdays and weekend patterns over the years. Changes in the peak period of walking patterns on weekdays and weekends across villages and corridors are studied through time series clustering. Note that the first confirmed COVID-19 case in the state of New South Wales (NSW) was identified on 19 January 2020 in Sydney. Following the outbreak of COVID-19, the first lockdown for NSW began on 31 March 2020.

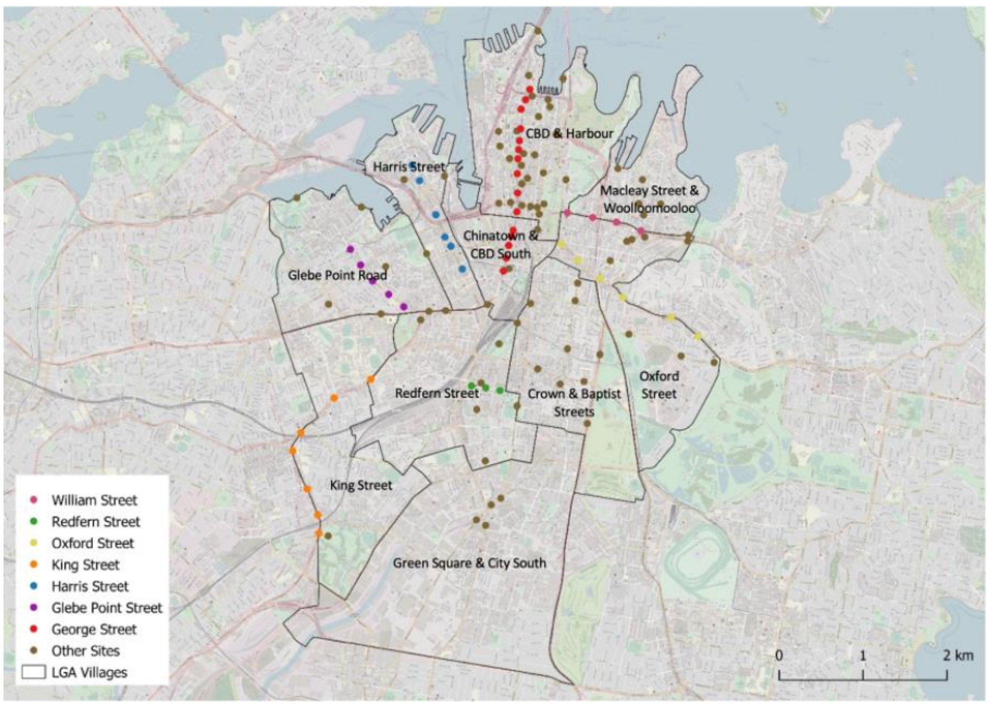


Figure 1. Spatial distribution of the pedestrian count sites, corridors, and villages in the city of Sydney

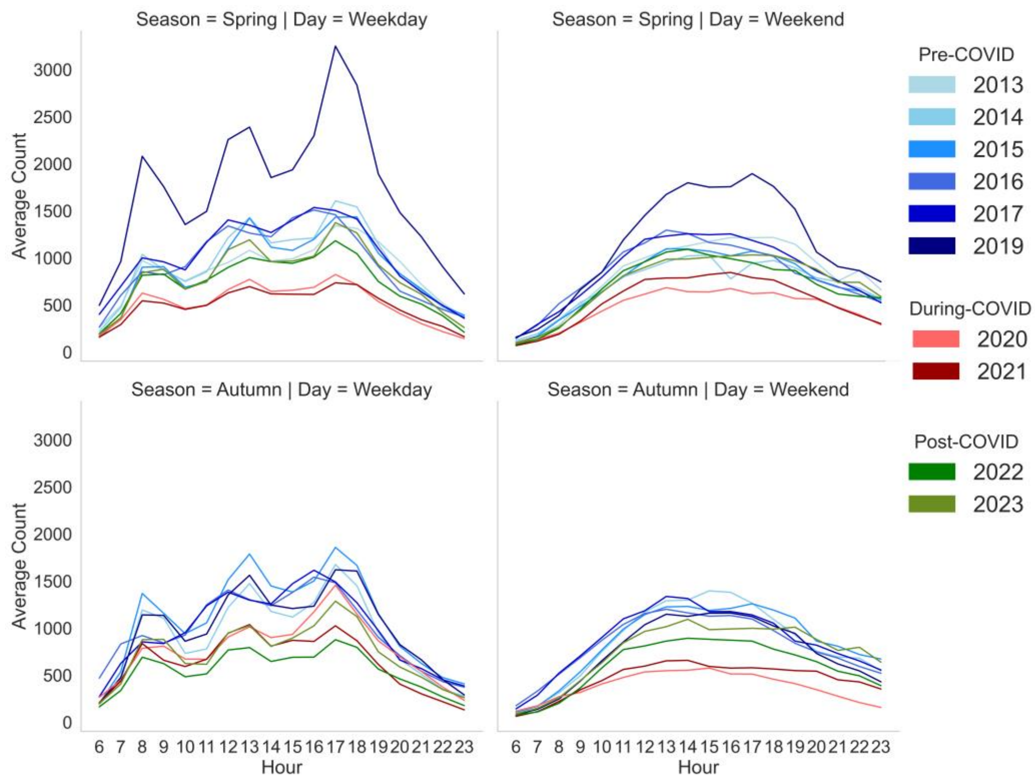


Figure 2. Temporal patterns in pedestrian activities (average pedestrian count across all sites) in the city of Sydney at different times of day, seasons, and years

2.3 Time-series clustering

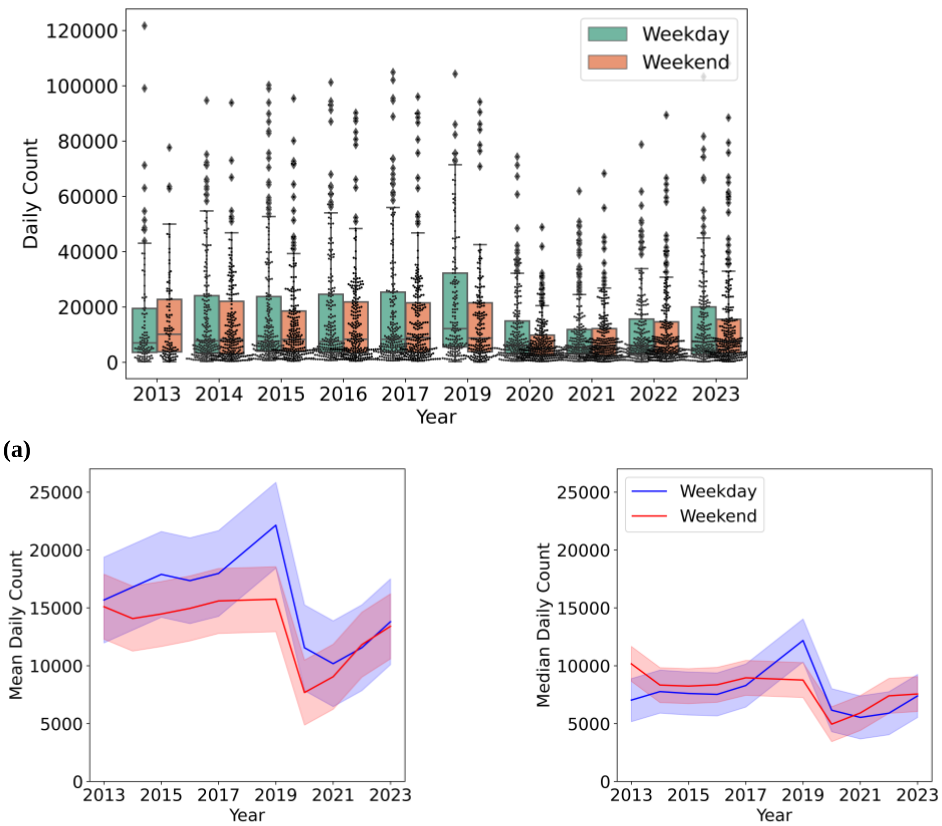
The wide adoption of "work from home" during the pandemic led to the reduction of commuting trips to work, especially during peak periods (Wu et al., 2020). Understanding how the walking peak periods have changed at the pre-, during, and post-pandemic periods can have significant implications on local transportation planning and practices. In this study, we identify pedestrian activity clusters and group the pedestrian count sites according to the hourly average pedestrian count over the study's three time periods. We apply min-max scalar to the data (Tian & Chen, 2021), as well as a fast Fourier transform (Cochran et al., 1967) to smoothen and normalize the pedestrian count time series. This process helps to better retrieve the structure of the series according to peak hour rather than capturing the noise and fluctuations in the time series. We split the time series based on pre-, during, and post-pandemic periods, as well as weekdays or weekends into six categories. This helps us identify specific trends and behaviors for each group more clearly and can help in capturing the temporal patterns effectively (Wang et al., 2013). To further refine our data and address the dimensionality concerns, we apply a Principal Component Analysis (PCA) (Aghabozorgi et al., 2015). PCA aids in transforming the original features into a new set of uncorrelated features, which are the principal components of the original features. This not only reduces the dimensionality of the data but also extracts the most significant features, enhancing the efficiency and effectiveness of the subsequent k-means clustering. Then a k-means clustering (Bishop & Nasrabadi, 2006) is applied to classify the time series as a semi-supervised machine learning method. We select the number of clusters using a plot of average Silhouette width against the number of clusters, for between 2 and 10 clusters. The Silhouette width decreases significantly at the optimal number of clusters. The numerical results are presented later in section 3.1.

2.4 Regression modeling

We apply a Least Absolute Shrinkage and Selection Operator (LASSO) analysis (Fonti & Belitser, 2017) to select important features and develop both linear and non-linear regression models to assess the impact of different selected input features (e.g., population and land use characteristics) at each count site on the changes in the number of walking trips, examining the coefficients to describe and compare the changes in the pedestrian activities for pre- to during pandemic and during to post-pandemic periods. To achieve this, we employ a linear regression model and a generalized linear model, namely negative binomial regression, to more precisely assess the influence of various factors (Hilbe, 2011; Kutner et al., 2005). The coefficients of the linear regression model and the negative binomial regression model are not directly comparable. Firstly, the assumptions underlying linear regression differ from those of generalized regression models. Linear regression is designed for continuous data that can take both positive and negative values, whereas negative binomial regression is intended for count data, which are always positive. Therefore, comparing the signs and values of the coefficients between these methods is not meaningful. Instead, we compare the overall performance of the models to determine and discuss the contributing factors to changes in pedestrian activities (Cameron & Trivedi, 2013; Neter et al., 1983). Finally, a random forest model is also estimated. (Menze et al., 2009) that is an ensemble learning technique that measures the importance score of non-linear relationships between land use and network features and pedestrian counts. By constructing multiple decision trees, each trained independently on its corresponding bootstrap sample, random forest identifies the most influential features.

3 Results

Our empirical analysis suggests that COVID-19 restrictions in 2020-2021 significantly dropped the number of pedestrians both on weekdays and weekends. The average daily pedestrian count during the pandemic (2020-2021) on weekdays was 41% smaller than the pre-pandemic average (2013-2019). The average daily pedestrian count on weekends during the pandemic was also 47% smaller than the pre-pandemic average. Although COVID-19 response measures varied between metropolitan areas, in most Australian cities including Sydney, lockdown and social distancing started after the declaration of the NSW emergency. The scope of the lockdown order prohibited the people in the city of Sydney from leaving their homes except for necessities like grocery shopping, work, or school for those who are unable to work or learn from home, compassionate reasons, or any other vital services. While the pedestrian activities across the study area have been recovering since 2020 with a 27% increase in the weekday average daily pedestrian count in 2023 compared to 2020, the weekday average daily pedestrian count was still 11% lower in 2023 than its highest record in 2019. See Figure 3.



(a) **(b)** **(c)**

Figure 3. (a) Distribution of daily pedestrian counts, (b) mean, and (c) median daily pedestrian count across all sites over the study's 11-year period for weekdays (blue line) and weekends (red line) (The shaded areas represent standard deviation from the mean and median daily count)

3.1 Clustering analysis

To further understand the impact of the pandemic response measures on pedestrian activities and its post-pandemic recovery across different sites, we investigate the spatial changes in the identified time-series clusters for both weekdays and weekends, following the methodology described earlier in section 2.3. See Figure 4. The main idea is to find similarities between different time series and to pair them under the same identified cluster such that the time series in the same cluster follows a similar pattern. Overall, we have identified seven meaningful clusters in the data:

- Morning peak only,
- Morning and afternoon peak,
- Morning and noon peak,
- Morning, noon, and afternoon peak,
- Noon peak only,
- Noon and afternoon peak, and
- Afternoon peak only.

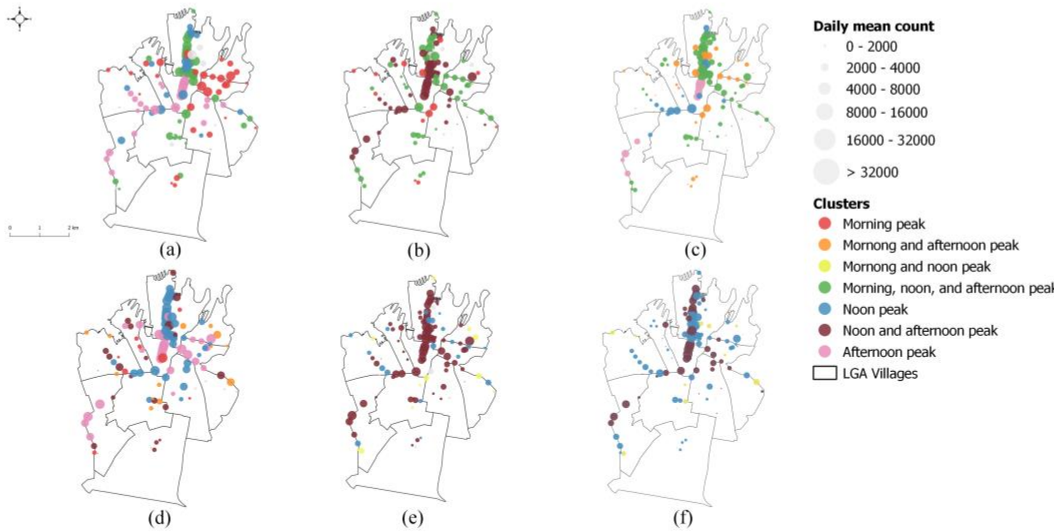


Figure 4. The spatial distribution of the identified clusters and pedestrian count sites across the study area: (a) pre-pandemic (2013-2019) on weekdays, (b) during pandemic (2020-2021) on weekdays, (c) post-pandemic (2022-2023) on weekdays, (d) pre-pandemic (2013-2019) on weekends, (e) during pandemic (2020-2021) on weekends, and (f) post-pandemic (2022-2023) on weekends. Colors represent different clusters (The size of the circles represents the average daily pedestrian count per site over the period)

Most locations in the CBD & Harbor Village display three peaks during the pre-pandemic period (2013-2017) on weekdays, indicating a significant presence of both work (commuting) trips and non-work trips in the area. Over the same period, most of the sites along Glebe Point Road, King Street, Chinatown, and CBD South villages show an afternoon peak only that suggests considerable pedestrian activities related to both leisure and work-related journeys. In contrast, most locations throughout the whole study area during weekends, both pre- and during pandemic periods, only show peaks at noon and in the afternoon, indicating an anticipated recreational and shopping walking pattern associated with weekends. During the post-pandemic period, many of the sites in the CBD & Harbor village fell back into the three peak clusters (morning, noon, and afternoon peak) and the noon and afternoon peak cluster as both work and non-work-related pedestrian activities started to recover.

3.2 LASSO regression and variance inflation factor (VIF) analysis

In this section, we apply a LASSO regression analysis to describe the effects of different populations, land use features, and network connectivity on the observed changes in the average daily pedestrian count from the pre-pandemic to pandemic period (impact) as well as from the during pandemic to the post-pandemic period (recovery). The change in the average daily pedestrian counts within the pre- to during pandemic and during to post-pandemic periods is treated as the dependent variables (Cameron & Trivedi 2013). See Figure 5 for the distributions. The independent variables include land use characteristics such as the percentage of industrial, hospital/medical, water, transportation, commercial, parkland, residential, and educational land use, network connectivity features measured by the street intersection count, number of links per number of nodes ratio, and the average node degree. In addition to population, weekly average personal income, and number of Points of Interest (POI). The estimated LASSO regression coefficients provide insights into the importance of each feature and its impact on the changes in pedestrian activities. See Figure 6. A positive coefficient suggests a positive association, where an increase in the independent variable corresponds to an increase in the target variable. Negative coefficients indicate a negative relationship, where an increase in the independent variable corresponds to a decrease in the target variable. The magnitude of the coefficient reflects the strength of the association between the independent variable and the target variable.

The results suggest that the number of POIs, population, and intersection count, followed by the percentage of commercial land use and the number of links per number of nodes ratio had the largest association with the changes in the pedestrian activities from the pre-pandemic to during the pandemic period. This is consistent with the existing empirical evidence in the literature suggesting that the “work from home” phenomenon has had a significant impact on urban travel patterns including walking trips. The number of POIs, population, intersection counts, and percentage of commercial land use had a negative association with the changes in pedestrian activities, suggesting a large drop in the number of walking trips in areas with a higher number of POIs, larger population, and greater commercial land use. However, the percentage of residential, weekly person income, the number of links over nodes ratio, and the percentage of education land use are found to be positively associated with the changes in the average number of pedestrians. Results suggest that areas with higher personal income and a higher percentage of non-work-related land use experienced a smaller change in the number of walking trips.

A similar association pattern also describes the recovery pattern of pedestrian activities from the during pandemic to the post-pandemic period with the percentage of commercial land use, population, intersection count, and number of POIs having the largest and positive association with the change in the average daily pedestrian count. While the percentage of transportation, water, and residential land uses, and the number of links over nodes ratio had a negative association. Also, the impact of the weekly person income on the changes in pedestrian activities was found to be greater for the during to pre-pandemic period compared to the post- to during the pandemic period.

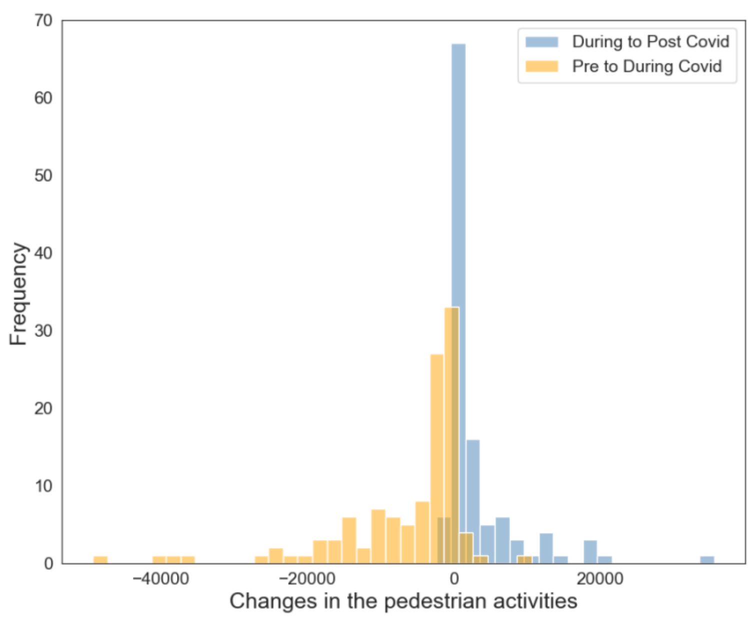


Figure 5. Distribution of the changes in the pedestrian activities across all sites in the study area: pre- to during the pandemic period and during pandemic to post-pandemic period

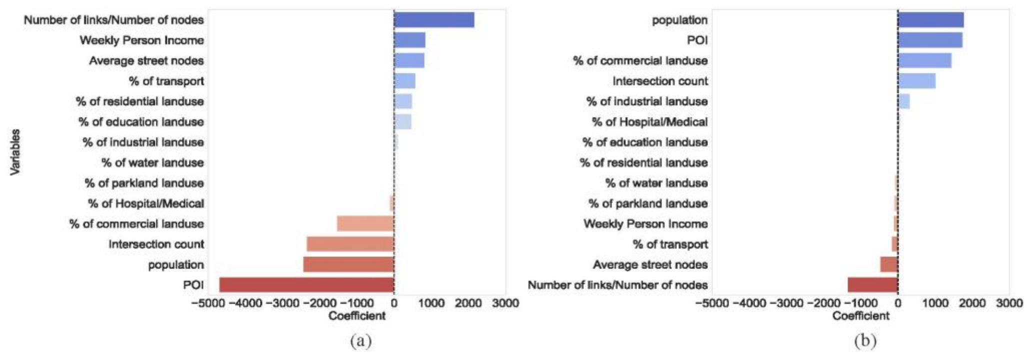


Figure 6. Estimated LASSO regression coefficients for (a) pre- to during the pandemic period and (b) during to post-pandemic period

We also conduct a Variance Inflation Factor (VIF) analysis (Neter et al., 1983) to assess multicollinearity among the predictor variables in our spatial regression models described in the next sections. The VIF measures reveal insights into the degree of correlation among the variables. The recommended threshold for the Variance Inflation Factor (VIF) is 10 (Hair, 2009). A VIF of 1 indicates no multicollinearity, values between 1 and 10 suggest moderate multicollinearity and values above 10 indicate high multicollinearity. In this study, we set values greater than 20 as indicative of infinite multicollinearity. Notably, variables such as percentage of commercial land use, percentage of education land use, percentage of parkland land use, and percentage of residential land use displayed extremely high VIF values, indicating significant multicollinearity. There is high multicollinearity between the number of links over nodes ratio and the average node degree in the street network. Weekly person income, population, number of POIs, and intersection count indicate insignificant (or very weak) multicollinearity with any of the variables. The findings are intended to guide our variable selection process and enhance the reliability of our regression modeling analysis in the following sections. Table 1 provides a summary of the VIF results of different sets of variables used in this research for reference.

Table 1. Variance Inflation Factor (VIF) analysis results for the predictor variables

Variable	**VIF (all variables)**	**VIF (Model 1)**	**VIF (Model 2)**	**VIF (Model 3)**
Population	1.68	1.355	1.48	1.57
Number of POIs	1.87	1.366	1.74	1.74
Weekly person income	2.04	1.417	1.54	1.91
% of commercial land use	inf	1.43	2.53	3.26
Intersection count	2.45	-	2.37	2.28
Number of links/number of nodes ratio	16.46	-	1.49	1.56
% of residential land use	inf	-	1.88	2.32
% of education land use	inf	-	-	1.53
% of industrial land use	1.33	-	-	1.28
Average node degree	15.11	-	-	-
% of water land use	1.08	-	-	-
% of transportation land use	inf	-	-	-
% of hospital/medical land use	inf	-	-	-
% of parkland land use	inf	-	-	-

3.3 Linear regression

To quantify the linear importance of each predictor feature on the changes in pedestrian activities, we utilize the selected features from the Lasso analysis and incorporate them into a set of Linear Least Squares (LLS) regression models. Table 2 and 3 provide a summary of the model estimates for the pre- to during the pandemic period and the during to post-pandemic period, respectively. We compare the results of different models and durations using Root Mean Square Error (RMSE), Log-Likelihood, and Akaike Information Criterion (AIC). Additionally, we applied the F-test to evaluate how the inclusion of variables affects the model's performance. Results further confirm the significant impact of the number of POIs, population, and percentage of commercial land use on the changes in pedestrian activities for both the pre- to during pandemic and during to post-pandemic periods. However, none of the remainder variables, including other types of land use, and weekly person income were found to be statistically insignificant. Network connectivity features, despite being identified as important in the Lasso analysis, were found to be only weakly significant. Although they exhibited favorable p-values, their inclusion in the prediction model reduced the f-statistic. This indicates that adding these features to the model does not enhance the model's overall performance.

Table 2. Estimated linear regression models for the pre- to during pandemic period

Variables	Base Model	Model 1	Model 2	Model 3
	Estimated coefficients			
Constant	-6482.14**	-6482.14**	-6482.14**	-6482.14**
Number of POIs	-6664.48**	-5280.50**	-4830.85**	-4812.63**
Population	-	-2570.55**	-2532.65**	-2479.77**
% of commercial land use	-	-2609.61**	-1726.34*	-1536.70
Weekly Person Income	-	-	696.93	833.56
Intersection Count	-	-	-2319.02*	-2295.49*
Number of links/nodes ratio	-	-	1362.72*	1309.46*
% of residential land use	-	-	321.93	461.77
% of education land use	-	-	-	391.79
% of industrial land use	-	-	-	46.75
Model goodness of fit measures				
R^2	0.5	0.62	0.67	0.67
RMSE	6,621	5,814	5,346	5,337
F-test	-	59.59	27.34	21.82
AIC	2,354	2,328	2,320	2,322
Log-Likelihood	-1,175	-1,160	-1,150	-1,150

** p-value < 0.01, * p-value < 0.05, with no asterisk p-value > 0.05

Table 3. Estimated linear regression models for the during to post-pandemic period

Variables	Base Model	Model 1	Model 2	Model 3
	Estimated coefficients			
Constant	2897.03**	2897.03**	2897.03**	2897.03**
Number of POIs	2801.16**	1898.51**	1745.98**	1764.07**
Population	-	1760.06**	1869.81**	1815.92**
% of commercial land use	-	1698.98**	1328.8*	1573.21*
Weekly Person Income	-	-	-36.59	-110.98
Intersection Count	-	-	1003.01**	1010.85**
Number of links/nodes ratio	-	-	-806.09*	-875.43*
% of residential land use	-	-	-65.50	127.94
% of education land use	-	-	-	151.43
% of industrial land use	-	-	-	342.91
Model goodness of fit measures				
R^2	0.27	0.42	0.46	0.46
RMSE	4,640	4,125	4037	4025
F-test	-	26.87	12.26	9.48
AIC	2,272	2,249	2,252	2,255
Log-Likelihood	-1,134	-1,120	-1,118	-1,117

** p-value < 0.01, * p-value < 0.05, with no asterisk p-value > 0.05

3.4 Negative binomial regression

Given the pedestrian activities data are technically count data, negative binomial and Poisson regression models are expected to fit the data better than a linear regression model (Targa & Clifton, 2005). Negative binomial regression is found to be more appropriate than Poisson regression because the variance in the dependent variable is larger than the average. In this study, the mean and variance of the dependent variable (change in the walking counts) were 6,820 and 83,751,366 for the pre to during pandemic period, and 3,077 and 28,305,311 for the during to post-pandemic period suggesting high levels of overdispersion. The overdispersion is usually assessed using measures like the

Pearson and chi-squared statistics divided by the number of observations. If the overdispersion metric is greater than 1.0, it indicates that the data has more variability than expected, making the negative binomial model a more suitable choice (Hilbe, 2011).

In this section, we estimate a series of negative binomial regression models following the same structure as the linear regression models described earlier in section 3.3, including a base model and three extended models. See Table 4 and 5. The RMSE, loglikelihood, and AIC values across all estimated models show significant improvement as a larger number of variables are included in the extended models compared to the base model. However, the F-test results indicate that adding additional variables does not significantly improve the model. Despite network connectivity measures being statistically significant based on p-values, their inclusion does not lead to a substantial enhancement in model performance.

Overall, results suggest that both linear and negative binomial regression models seem to perform relatively well in capturing patterns in the data. Both modeling approaches can describe the association between the predictor variables and changes in pedestrian activities across both periods reasonably well. However, by comparing RMSE, log-likelihood, and AIC — where lower values indicate a better model — and considering that we are dealing with count data instead of continuous data, we conclude that negative binomial regression outperforms linear regression. This suggests the presence of strong non-linear effects between the predictor variables and pedestrian counts.

The results from the negative binomial regression mirrored those of the linear regression in determining the statistical significance of features, using measures such as the p-values and F-test. Results highlight that the number of POIs, population, and commercial land use strongly influenced changes in pedestrian activity during both the pre- to during-pandemic and during to post-pandemic. However, other factors, such as other types of land use and weekly personal income, were not significant. Although network connectivity features seemed important initially, they did not add real value to the predictive power of the model when included.

Table 4. Estimated Negative binomial regression models for the "pre- to during pandemic" period

Variables	**Base Model**	**Model 1**	**Model 2**	**Model 3**
	Estimated coefficients			
Constant	8.35**	8.3**	8.26**	8.25**
POI	1.07**	0.72**	0.57**	0.57**
Population	-	0.23**	0.24*	0.21
% of commercial land use	-	0.43**	0.27	0.26
Weekly Person Income	-	-	-0.04	-0.11
Intersection Count	-	-	0.38**	0.37*
Number of links/Number of nodes	-	-	-0.13	-0.10
% of residential land use	-	-	-0.12	-0.08
% of education land use	-	-	-	-0.07
% of industrial land use	-	-	-	0.03
Model goodness of fit measures				
Pseudo R^2	0.61	0.65	0.67	0.68
RMSE	51,980	8,223	5,347	5,375
F-test	-	12.67	0.59	0.05
AIC	2,155	2,147	2,146	2149
Log-Likelihood	-1,076	-1,069	-1,065	-1,064

Table 5. Estimated Negative binomial regression models for the "post- to during pandemic" period

Variables	Base Model	Model 1	Model 2	Model 3
	Estimated coefficients			
Constant	7.56**	7.42**	7.38**	7.37**
POI	1.02**	0.47**	0.45**	0.44**
Population	-	0.26**	0.38**	0.35**
% of commercial land use	-	0.69**	0.35*	0.44*
Weekly Person Income	-	-	0.09	0.04
Intersection Count	-	-	0.37*	0.36
Number of links/Number of nodes	-	-	-0.18	-0.18
% of residential land use	-	-	-0.26	-0.20
% of education land use	-	-	-	0.01
% of industrial land use	-	-	-	0.11
Model goodness of fit measures				
Pseudo R^2	0.61	0.70	0.73	0.73
RMSE	19,619	3,788	3858	3,802
F-test	-	27.11	3.85	1.12
AIC	1,973	1,946	1,944	1947
Log-Likelihood	-984	-969	-964	-963

3.5 Random forest regression

We also apply a random forest regression to determine the importance of variables that exhibit non-linear relationships with changes in pedestrian activities. The feature importance scores are calculated using the Mean Decrease Impurity (MDI) from the random forest model. Figure 7 illustrates the feature importance scores for both the pre- to during pandemic and the during to post-pandemic periods. The error bars represent the standard deviation of feature importance scores for each variable, computed across the individual decision trees in the random forest ensemble. The results further confirm the significant association between the number of POIs and population, intersection count, and commercial land use on the changes in pedestrian activities in both time windows, compared with the other features.

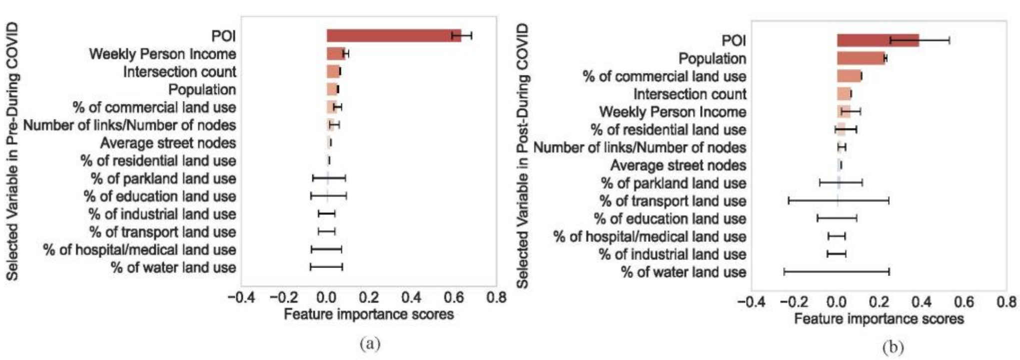


Figure 7. Random forest importance scores for the selected features for (a) pre- to during the pandemic period (b) and during to post-pandemic period

4 Discussion

The study investigates pedestrian traffic patterns and their associated variables before, during, and after the COVID-19 pandemic, revealing a significant and persistent decline in pedestrian numbers post-pandemic compared to pre-pandemic levels. The analysis utilized data from multiple walking count sites and employed time-series and regression methods to assess changes in pedestrian flow. This finding aligns with broader trends in urban mobility where pedestrian behavior has been notably changed by pandemic-related shifts in work, travel, and lifestyle habits. The observed decline may be attributed to the increased prevalence of remote work, as studies have documented a substantial rise in work-from-home arrangements, reducing the frequency of commuting and, consequently, pedestrian traffic in business districts (Noland et al., 2023). Additionally, changes in public behavior, such as increased reliance on private vehicles and continued adherence to social distancing practices, likely contribute to the decreased pedestrian activity (Du et al., 2024). Our further regression analysis on the factors affecting change in the number of walking trips before, during, and after the pandemic suggests that the most significant changes are associated with areas with larger population, larger number of POIs, and commercial land use, followed by urban form measured by the intersection counts (Li & Xu, 2021; Yang et al., 2023). Although other land uses, such as residential and educational, and network connectivity features (e.g., the number of links per node), also affected the change in the number of walking trips, their effects were not as significant as the others.

The sustained reduction in pedestrian numbers has significant implications for urban planning and policy making. It suggests that urban centers may need to adapt to a new normal where traditional peak pedestrian flows are altered. Economic factors, including job losses and reduced disposable income, may have diminished pedestrian activities related to shopping and leisure, further contributing to the decline (Cheng, 2024). Future research should investigate these factors in detail to understand their individual and combined impacts on pedestrian patterns. Qualitative methods, such as surveys and interviews with urban residents, could provide deeper insights into the motivations behind changes in pedestrian behavior, complementing the quantitative analysis and offering a more comprehensive understanding of the evolving urban mobility landscape (Martell et al., 2024).

The findings from the presented analysis of pedestrian activities in the city of Sydney highlight a significant drop in pedestrian counts during the pandemic, particularly in areas with high commercial land use and larger number of POIs. To mitigate the reduction in pedestrian activities observed during the pandemic, urban planners can prioritize enhancing infrastructure in areas with diverse land uses. Investments in pedestrian-friendly infrastructure in commercial zones, such as wider sidewalks, pedestrian plazas, and traffic calming measures, can encourage walking by creating a more inviting environment. Moreover, integrating mixed-use developments that include residential, educational, and commercial facilities can buffer against such reductions by providing destinations that serve multiple purposes, thereby promoting steady pedestrian traffic regardless of external disruptions.

Areas with higher commercial activity and population density showed a greater rebound in pedestrian counts post-pandemic. This suggests that policies should focus on equitable recovery strategies that address the disparities in pedestrian activity resurgence. For instance, implementing targeted pedestrian infrastructure improvements in residential areas and neighborhoods with higher percentages of low-income populations can enhance accessibility and encourage walking. Additionally, creating incentives for local businesses and community services in these areas can increase pedestrian traffic and support local economies, fostering a more balanced recovery across the city.

5 Conclusions

The impact of COVID-19 on urban travel behavior has been substantial. While there have been numerous studies in the literature exploring how the pandemic affected mobility patterns, very little effort has gone into understanding its impact on pedestrian activities or walking over time and space. This study focused on analyzing walking patterns in the city of Sydney across an 11-year period: pre-, during, and post-pandemic. A comprehensive set of statistical analyses including time-series clustering as well as linear and non-linear regression was applied to understand the walking patterns before, during, and post-pandemic and to quantify the importance of various population and land use features affecting the changes in the pedestrian activities.

Results revealed that the average daily pedestrian count across the city of Sydney dropped by 41% and 47% on weekdays and weekends, respectively, during the pandemic compared to the pre-pandemic period. While the pedestrian activities across the study area have been recovering and increased by 51% in 2023 compared to the lowest record in 2020, the weekly average daily pedestrian count across the city of Sydney is still 11% lower than the observed record in 2019.

The observed reduction in pedestrian activities from pre- to during the pandemic period, as well as the recovery in pedestrian activities from during to the post-pandemic period, was not spatially homogenous. A LASSO and a Random Forest regression analysis as well as a Linear Least Squares modeling were applied to understand the factors that had greater association with the changes in the pedestrian activities. All modeling methods provided consistent outcomes in which the number of POIs, population, and percentage of commercial land use were found to be the most significant features with the strongest association with changes in the number of walking trips. Areas with more commercial land use, number of POIs, higher population, and intersection counts experienced a significant negative change in the number of walking trips from the pre-pandemic to the during pandemic period. While areas with higher percentages of education, residential, and commercial and with higher personal income experienced smaller changes in pedestrian activities during the pandemic compared to the pre-pandemic period. During the recovery phase, the influential features remained mostly unchanged; however, the association direction was the opposite. Areas with a larger number of POIs and greater commercial land use and population and intersection count have experienced a greater positive change in the number of walking trips. While areas with higher percentages of residential, educational, and weekly person income experienced a smaller change in pedestrian activities from the during pandemic to the post-pandemic period.

The data used in this study had some limitations. No information on the travel purpose was available. Not all sites have been surveyed consistently over the study's 11-year period. No data from 2018 was available. Overall, the revealed insights are expected to help urban and transportation planners to better understand the impact of COVID-19 on walking trips and how pedestrian activities have been changing over a decade before the pandemic, through the pandemic, and during the post-pandemic recovery.

Acknowledgments

This research is partly supported through funding from the city of Sydney local government. Authors thank Annie Matan, Sarah Brickhill and Bhavesh Rathod. Meead Saberi is supported by an Australian Research Council Discovery Grant (DP220102382).

Author contribution

The authors confirm contribution to the paper as follows: Z. Nourmohammadi, T. Lilasathapornkit, and M. Saberi: study conception and design; Z. Nourmohammadi, T. Lilasathapornkit, and M. Saberi: data collection; Z. Nourmohammadi, F. Nourmohammadi, and T. Lilasathapornkit: analysis and interpretation of results: Z. Nourmohammadi, F. Nourmohammadi, T. Lilasathapornkit, and M. Saberi: draft manuscript preparation. All authors reviewed the results and approved the final version of the manuscript.

References

Aghabozorgi, S., Shirkhorshidi, A. S., & Wah, T. Y. (2015). Time-series clustering–a decade review. *Information systems*, *53*, 16–38.

Angel, A., Cohen, A., Dalyot, S., & Plaut, P. (2023). Impact of COVID-19 policies on pedestrian traffic and walking patterns. *Environment and Planning B: Urban Analytics and City Science, 50*(5), 1178–1193.

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*. New York: Springer.

Blundell, R., Joyce, R., Keiller, A. N., & Ziliak, J. P. (2018). Income inequality and the labor market in Britain and the US. *Journal of Public Economics*, *162*, 48–62.

Borkowski, P., Jażdżewska-Gutta, M., & Szmelter-Jarosz, A. (2021). Lockdowned: Everyday mobility changes in response to COVID-19. *Journal of Transport Geography*, *90,* 102906.

Calderón-Loor, M., Hadjikakou, M., & Bryan, B. A. (2021). High-resolution wall-to-wall land-cover mapping and land change assessment for Australia from 1985 to 2015. *Remote Sensing of Environment*, *252*, 112148.

Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data*. Cambridge, England: Cambridge University Press.

Cheng, S. (2024). Reshaping the concept of leisure and labor in the era of post-COVID-19. In *Leisure and work in China* (pp. 94–106). Abingdon, Oxfordshire, England: Routledge.

Cochran, W. T., Cooley, J. W., Favin, D. L., Helms, H. D., Kaenel, R. A., Lang, W. W., … & Welch, P. D. (1967). What is the fast Fourier transform? *Proceedings of the IEEE, 55*(10), 1664–1674.

Currie, G., Jain, T., & Aston, L. (2021). Evidence of a post-COVID change in travel behavior–Self-reported expectations of commuting in Melbourne. *Transportation Research Part A: Policy and Practice*, *153*, 218–234.

Delclòs-Alió, X., Gutiérrez, A., Tomàs-Porres, J., Vich, G., & Miravet, D. (2022). Walking through a pandemic: How did utilitarian walking change during COVID-19? *International Journal of Sustainable Transportation, 17*(10), 1155–1170.

Doubleday, A., Choe, Y., Busch Isaksen, T., Miles, S., & Errett, N. A. (2021). How did outdoor biking and walking change during COVID-19?: A case study of three US cities. *PLOS One*, *16*(1), e0245514.

Du, B., Zhang, C., Qu, T., Wang, Q., Zhou, Q. S., Cui, T., … & Astell-Burt, T. (2024). Impact of the COVID-19 pandemic on daily travel: Findings from New South Wales, Australia. *Travel Behavior and Society*, *36*, 100798.

Enoch, M., Monsuur, F., Palaiologou, G, Quddus, M. A., Ellis-Chadwick, F., Morton, C., & Rayner, R. (2022). When COVID-19 came to town: Measuring the impact of the coronavirus pandemic on footfall on six high streets in England. *Environment and Planning B: Urban Analytics and City Science*, *49*(3), 1091–1111.

Falchetta, G., & Noussan, M. (2020). The impact of COVID-19 on transport demand, modal choices, and sectoral energy consumption in Europe. In *IAEE Energy Forum* 2020, 48–50. https://www.iaee.org/documents/2020EnergyForumSI.pdf

Fonti, V., & Belitser, E. (2017). Feature selection using lasso. *VU Amsterdam Research Paper in Business Analytics*, *30*, 1–25.

Fujita, A., Feliciani, C., Yanagisawa, D., & Nishinari, K. (2019). Traffic flow in a crowd of pedestrians walking at different speeds. *Physical Review E, 99*(6), 062307.

Gao, J., Kamphuis, C. B., Helbich, M., & Ettema, D. (2020). What is 'neighborhood walkability'? How the built environment differently correlates with walking for

different purposes and with walking on weekdays and weekends. *Journal of Transport Geography, 88,* 102860.
Gkiotsalitis, K., & Cats, O. (2021). Public transport planning adaption under the COVID-19 pandemic crisis: Literature review of research needs and directions. *Transport Reviews, 41*(3), 374–392.
Guthold, R., Stevens, G. A., Riley, L. M., & Bull, F. C. (2018). Worldwide trends in insufficient physical activity from 2001 to 2016: A pooled analysis of 358 population-based surveys with 1–9 million participants. *The Lancet Global Health, 6*(10), e1077–e1086.
Hair, J. F., Black, W. C., Babin, B. J. & Anderson, R. E. (2009). *Multivariate data analysis,* 7th ed. Pearson.
Hall, C. M., & Ram, Y. (2019). Measuring the relationship between tourism and walkability? Walk score and English tourist attractions. *Journal of Sustainable Tourism, 27*(2), 223–240.
Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge, England: Cambridge University Press.
Hintermann, B., Schoeman, B., Molloy, J., Schatzmann, T., Tchervenkov, C., & Axhausen, K. W. (2023). The impact of COVID-19 on mobility choices in Switzerland. *Transportation Research Part A: Policy and Practice, 169,* 103582.
Humagain, P., & Singleton, P. A. (2021). Exploring satisfaction with travel time profiles towards understanding intrinsic utilities of travel time. *Travel Behavior and Society, 24*, 22–33.
Hunter, R. F., Garcia, L., de Sa, T. H., Zapata-Diomedi, B., Millett, C., Woodcock, J., … & Moro, E. (2021). Effect of COVID-19 response policies on walking behavior in US cities. *Nature communications, 12*(1), 3652.
Kang, M., Choi, Y., Kim, J., Lee, K. O., Lee, S., Park, I. K., … & Seo, I. (2020). COVID-19 impact on city and region: What's next after lockdown? *International Journal of Urban Sciences, 24*(3), 297–315.
Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models.* New York: McGraw-Hill.
Li, Y., & Xu, L. (2021). The impact of COVID-19 on pedestrian flow patterns in urban Pois—An example from Beijing. *ISPRS International Journal of Geo-Information, 10*(7), 479.
Loo, B. P. (2021). Walking towards a happy city. *Journal of Transport Geography, 93,* 103078.
Martell, M., Terry, N., Sengupta, R., Salazar, C., Errett, N. A., Miles, S. B., … & Choe, Y. (2024). Open-source data pipeline for street-view images: A case study on community mobility during COVID-19 pandemic. *PLOS One, 19*(5), e0303180.
Mavoa, S., Boulangé, C., Eagleson, S., Stewart, J., Badland, H. M., Giles-Corti, B., & Gunn, L. (2018). Identifying appropriate land-use mix measures for use in a national walkability index. *Journal of Transport and Land Use, 11*(1), 681–700.
Mendiate, C. J., Nkurunziza, A., Machanguana, C. A., & Bernardo, R. (2022). Pedestrian travel behaviour and urban form: Comparing two small Mozambican cities. *Journal of Transport Geography, 98,* 103245.
Menze, B. H., Kelm, B. M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., & Hamprecht, F. A. (2009). A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics, 10,* 1–16.
Möllers, A., Specht, S., & Wessel, J. (2022). The impact of the COVID-19 pandemic and government intervention on active mobility. *Transportation Research Part A: Policy and Practice, 165,* 356–375.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to Linear Regression Analysis.* Hobojen, NJ: John Wiley & Sons.

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *135*(3), 370–384.

Neter, J., Wasserman, W., & Kutner, M. H. (1983). *Applied linear regression models.* Homewood, IL: Richard D. Irwin.

Neves, C. E. T., Da Silva, A. R., & de Arruda, F. S. (2021). Exploring the link between built environment and walking choice in São Paulo city, Brazil. *Journal of Transport Geography, 93,* 103064.

Noland, R. B., Iacobucci, E., & Zhang, W. (2023). Public views on the reallocation of street space due to COVID-19. *Journal of the American Planning Association, 89*(1), 93–106.

Obuchi, S. P., Kawai, H., Ejiri, M., Ito, K., & Murakawa, K. (2021). Change in outdoor walking behavior during the coronavirus disease pandemic in Japan: A longitudinal study. *Gait and Posture, 88,* 42–46.

Pareek, M., Bangash, M. N., Pareek, N., Pan, D., Sze, S., Minhas, J. S., …, & Khunti, K. (2020). Ethnicity and COVID-19: An urgent public health research priority. *The Lancet, 395*(10234), 1421–1422.

Power, D., Lambe, B., & Murphy, N. (2023). Trends in recreational walking trail usage in Ireland during the COVID-19 pandemic: Implications for practice. *Journal of Outdoor Recreation and Tourism*, *41*, 100477.

Stier, A. J., Berman, M. G., & Bettencourt, L. (2020). *COVID-19 attack rate increases with city size* (research paper 19). Chicago: Mansueto Institute for Urban Innovation. https://ssrn.com/abstract=3564464.

Targa, F., & Clifton, K. (2005). The built environment and trip generation for non-motorized travel. *Journal of Transportation and Statistics*, *8*(3), 55–70.

Tian, X., & Chen, M. (2021). Descriptor selection for predicting interfacial thermal resistance by machine learning methods. *Scientific Reports, 11*(1), 739.

Tison, G. H., Avram, R., Kuhar, P., Abreau, S., Marcus, G. M., Pletcher, M. J., & Olgin, J. E. (2020). Worldwide effect of COVID-19 on physical activity: A descriptive study. *Annals of Internal Medicine, 173*(9), 767–770.

Tribby, C. P., Miller, H. J., Brown, B. B., Werner, C. M., & Smith, K. R. (2016). Assessing built environment walkability using activity-space summary measures. *Journal of Transport and Land Use, 9*(1), 187.

Wang, X., Mueen, A., Ding, H., Trajcevski, G., Scheuermann, P., & Keogh, E. (2013). Experimental comparison of representation methods and distance measures for time series data. *Data Mining and Knowledge Discovery, 26,* 275–309.

Wu, L., Cheng, X., Kang, C., Zhu, D., Huang, Z., & Liu, Y. (2020). A framework for mixed-use decomposition based on temporal activity signatures extracted from big geo-data. *International Journal of Digital Earth, 13*(6), 708–726.

Yang, X., Zheng, X., Cao, Y., Chen, H., Tang, L., & Yang, H. (2023). Connectivity analysis in pedestrian networks: A case study in Wuhan, China. *Applied Geography*, *151*, 102843.

Yin, C., Cao, J., Sun, B., & Liu, J. (2023). Exploring built environment correlates of walking for different purposes: Evidence for substitution. *Journal of Transport Geography, 106,* 103505.