University of St Andrews

Exploring Time Series Analysis techniques to understand the impact of Low Traffic Neighbourhoods in Oxfordshire County

Department of Mathematics and Statistics

University of St Andrews

2023

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and appreciation to the following individuals for their invaluable contributions and support throughout the process of conducting this dissertation:

- Ben Swallow, my advisor from the St Andrews Department of Mathematics and Statistics. Your guidance, expertise, and continuous encouragement have been instrumental in shaping the direction and quality of this research.
- Peter Lindgren, the Head of Digitization at the Transport Research Lab. Thank you for your assistance and collaboration in providing valuable insights and resources for this study.
- Ruth Anderson, who works in Innovation and Development at Oxfordshire County Council. Your support and input have been invaluable in navigating the intricacies of the research subject and accessing relevant information.
- Aaron Croucher from Alchera Technologies. I extend my appreciation for your assistance in providing data and your valuable contributions in setting up the Alchera platform, which significantly enhanced the research process.
- My family, particularly my father, who holds a master's degree and has been a constant source of motivation and inspiration throughout my academic journey. Your unwavering support and belief in my abilities have been crucial in my pursuit of a postgraduate degree.

I am grateful for the contributions of these individuals, as their support has been vital in the successful completion of this dissertation.

ABSTRACT

"Exploring Time Series Analysis Techniques to Understand the Impact of Low Traffic Neighbourhoods in Oxfordshire County" investigates the effects of Low Traffic Neighbourhoods (LTNs) implemented as part of Oxfordshire County Council's Active Travel Strategy. Collaborating with Oxfordshire County Council, Transport Research Lab, and the Department of Mathematics and Statistics at the University of St Andrews, this research aims to comprehend the influence of LTNs on pedestrian, cyclist, and car counts in Oxfordshire. The study employs two primary methods for analysis. Firstly, Time Series forecasting using Autoregressive Integrated Moving Average (ARIMA) models predicts future values based on historical data. The ARIMA model comprises Autoregressive (AR), Integrated (I), and Moving Average (MA) components. Secondly, the research employs Causal Impact, also known as Bayesian Structural Time Series (BSTS) modeling, to estimate the causal impact of LTN implementation on the time series data. The exploratory data analysis reveals a decrease in car volume within LTN areas, and an increase in cyclist and pedestrian volumes post-LTN implementation. ARIMA modeling attempts to forecasts pedestrian and car counts, while cyclist count predictions show limitations. Causal Impact analysis confirms positive influences of LTNs on pedestrian and cyclist volumes within LTN areas, with mixed results on car counts at boundary roads. In conclusion, LTNs contribute to reduced car traffic within LTN areas and increased pedestrian and cyclist volumes, promoting active and sustainable travel. However, the effectiveness of LTNs varies at different boundary roads, warranting continuous evaluation. Limitations include assumptions violations in ARIMA modeling, real-life data complexity, the univariate nature of time series analysis, and the need for cautious comparison of information criteria and log-likelihoods. This research enhances understanding of LTN impacts, providing insights for policymakers, planners, and researchers in promoting safer and more sustainable transportation options. Further research can address the identified limitations and refine the models to better capture the complexities of real-world traffic dynamics.

Table of Contents

ACKNOWLEDGEMENTS *2*

ABSTRACT *3*

Chapter 1: INTRODUCTION *6*

1.1 Project Background 6

1.2 General Outline of Thesis 6

Chapter 2: BACKGROUND THEORY *7*

2. 1 Oxfordshire County Council 7

2.2 Active Travel Strategy 7

2.3 Low Traffic Neighbourhoods 7

2.4 East Oxford LTN Evaluation Snapshot Report June 2023 Overview 8

2.5 Prior Research: LTN, Air Quality, and impacts to health 10

2.6 Prior Research: Causal Impact Analysis 11

2.7 Prior Research: ARIMA 12

Chapter 3: Data *12*

3.1 About the Data 12

3.2 Data Contents and Processing 13

3.3 Exploratory Data Analysis 13

3.3.1 Pedestrian Traffic Counts for LTN areas and boundary roads 13
3.3.2 Cyclist Traffic Counts for LTN areas and boundary roads 16
3.3.3 Car Traffic Counts for LTN areas and boundary roads 18

Chapter 4: Methodology *21*

4.1 Methods: Time Series Forecasting with Arima 21

4.2 Assumptions: Time Series Forecasting 24

4.3 Methods: Causal Impact / Bayesian Structural Time Series (BSTS) Modelling 27

4.4 Assumptions: Causal Impact / Bayesian Structural Time Series (BSTS) modelling 28

Chapter 5: Results *28*

5.1 ARIMA 28

5.1.1 Forecasting 2023-2024 traffic counts in LTN Area 28
5.1.2 Forecasting 2023 -2024 traffic counts at LTN boundary roads 31

5.2 Causal Impact Analysis 40

5.2.1 LTN Area 40
5.2.2 LTN Boundary Roads 43

Chapter 6: Discussion *50*

6.1 Conclusion 50

6.2 Limitations and areas of improvement 51

References *53*

Appendix *54*
vivacityimport.R 54
exploratoryVIVA.R 55
forecastingLTNArea.R 57
forecastingLTNBoundary.R..58
causalImpactLTNArea.R 60
causalImpactBoundary.R 61

Chapter 1: INTRODUCTION

1.1 Project Background

The effort of this dissertation is shared between Oxfordshire County Council (OCC), Transport Research Lab (TRL), and the Department Mathematics and Statistics at University of St Andrews. Transport Research Lab is a team of expert scientists, engineers, and specialists working together with clients and partners to create the future of transport [https://www.trl.co.uk/about-us/who-we-are]. TRL publishes various transportation related software, conducts leading edge research into infrastructure, conducts detailed incident investigation, and collaborates with universities and partners to invest in basic and applied research that will support future needs. Oxfordshire County Council is a governing body in the Oxfordshire County. OCC's innovation and research department, in collaboration with the transport department, works to enhance aspects of community transport in line of their visions for active travel. The Department of Mathematics and Statistics at the University of St Andrews provides opportunities for external organizations to pursue research, motivated by students pursuing postgraduate degrees These three entities discussed the opportunity to conduct research on the topic of safe active travel. These three entities have discussed the opportunity to conduct research on the topic of safe active travel. The primary focus of this project is to investigate the impact of specific transport strategies implemented as a part of Oxfordshire County Council's Active Travel Strategy. Our research explores the effects of Low-traffic neighbourhoods (LTNS) in Oxfordshire County. LTNs are designated areas that promote safe travel for pedestrians and cyclists. We examine the changes in pedestrian and cyclist counts before and after the implementation of LTNs and attempt to model the impact of the implementation.

1.2 General Outline of Thesis

Chapter 2 provides background theory on Oxfordshire County Council as well as its Active Travel Strategy. It will explain the importance of low traffic neighbourhoods in context to the strategy and touch upon air quality and prior research into traffic strategy and infrastructure. The chapter concludes with background on methodologies used in analysis.

Chapter 3 gives an overview of the data, the data contents, and how the data was processed for this research. This section also provides exploratory data analysis on traffic volume data collected for certain LTN sensor locations in East Oxford.

Chapter 4 discusses the methodology this research employs. In this section, assumptions are checked before explaining the various methods used. The two main statistical modelling techniques used in this research are time series forecasting and causal impact analysis.

Chapter 5 examines the results from the chosen methodologies. In this section, we examine plots, summary output, and accuracy metrics for ARIMA forecasting on the LTN area and various LTN boundary road locations. For the causal impact analysis, we display and interpret plots for LTN Area and LTN boundary roads.

Chapter 6 is the discussion of the results and the overall aims of the research. This section is split into a few subsections: conclusion, areas of improvement, and limitations.

Chapter 2: BACKGROUND THEORY

This section will provide background theory on travel strategy in Oxfordshire County Council. Low traffic neighbourhoods are traffic strategy implementation motivated by the county council's Active Travel Strategy. This part will also explain what LTNs are and their importance, air quality in Oxfordshire, and summarize the June 2023 East Oxford LTN Evaluation Snapshot Report.

Commented [BS1]: Should be included as a reference in the bibliography.

2. 1 Oxfordshire County Council

Oxfordshire County Council is the county council (upper-tier local authority) for the non-metropolitan county of Oxfordshire in the Southeast of England. Established in 1889, it is an elected body responsible for the most strategic local government services in the county. Oxfordshire County Council provides a wide range of services, including education (schools, libraries, and youth services), social services, public health, highway maintenance, waste disposal, emergency planning, consumer protection and town and country planning for matters to do with minerals, waste, highways and education.

2.2 Active Travel Strategy

The motivation for research into sustainable and safe travel is from the Oxfordshire County Council's Active Travel Strategy. The strategy supports Oxfordshire's Local Transport and Connectivity Plan (LTCP) in its vision to create an inclusive and safe net-zero for Oxfordshire's transport system. The strategy focuses on active travel modes (walking, wheeling, and cycling), which are key to delivering the County Council's policies and plans for the next 10 years and to mitigating some of the biggest challenges we face: climate emergency, public health, congestion, air quality and social inequality (Oxfordshire County Council 2022). council suggests that it can have a significant impact on public health priorities which include improving health and wellbeing, tackling the overweight and obesity crisis and reduce health inequalities. To achieve these goals, the strategy considers the importance of traffic management techniques such as Low-traffic neighbourhoods (LTNs) as they are seen as integral to successfully promote walking and cycling (Oxfordshire County Council 2022).

2.3 Low Traffic Neighbourhoods

A LTN is an area where motorized traffic is prevented from taking shortcuts through a residential area. They are implementations of Oxfordshire's Active Travel Strategy and aim to create quieter and safer streets where residents may feel more comfortable when making local journeys by cycling, wheeling or on foot. All roads remain accessible, but drivers may have to find alternative routes instead of cutting through some streets. The LTN schemes in Cowley and east Oxford have been implemented through experimental traffic regulation orders (ETRO), following public consultation. ETROs are used when it is difficult to assess the impacts of a scheme beforehand, but the cost of implementation is low. The experimental phase allows residents, businesses, and other interested parties to see for themselves the impact of the scheme in their neighbourhood, and the council to monitor traffic flows, air quality, changes in people's perception, and shifts in their mode of travel. The schemes also allow the council to trial several types of measures for preventing shortcuts for motorised vehicles, including planters and bollards. On 19 July 2022, the county council's cabinet made the decision to keep the Cowley low-traffic neighbourhoods after an 18-month trial. They plan to undertake additional community and stakeholder engagement to further refine the scheme and continue to experiment these implementations in east Oxford (Oxfordshire County Council 2022).

Oxfordshire County Council are committed to become carbon neutral by 2030. They believe by leading the way and seeking to become net zero significantly ahead of the national 2050 target, they hope to show what can be achieved an inspire residents and local businesses to join the journey. LTNs not only encourage safe travel for pedestrians and cyclists, but they naturally improve the air quality in residential areas as less motorized traffic travel through those areas. Residents in the area can benefit from better air quality while improving their physical health because of the increase in walking and cycling behaviour. In this research, we plan to analyse those benefits for residents. Research into this area can encourage the implementation of more low-traffic neighbourhoods and more safe travel strategies to push for their carbon neutral and safe travel goals.

2.4 East Oxford LTN Evaluation Snapshot Report June 2023 Overview

Oxfordshire County Council published a snapshot report that provides a partial evaluation of the effects of East Oxford LTNs since their implementation on the 20th of May 2022. This report focuses on Traffic Volume and Air Quality analysis. The evaluation explains three metrics used to calculate the impact of LTN traffic filters on traffic volumes across different modes of transport: Simple Difference, Impact Difference, and Adjusted Impact Estimate.

The simple difference metric provides a simplified output of the difference to understand the effects of LTN filters on traffic volumes by comparing average daily counts by transport mode. However, it does not account for the wider trends in traffic volume across Oxford. The Impact Estimate metric uses a difference-in-differences (DiD) statistical technique to estimate the causal effect of the LTN filters by comparing changes in outcomes in East Oxford LTN areas and a control area before and after the LTNs were implemented (Reference LTN report). Lastly, the adjusted impact estimate, known as the difference-in-difference-in-differences statistical technique (DiDiD), is another metric that attempts to mitigate potential bias in the findings by incorporating data from the last 'typical year' (pre-COVID 19) (Oxfordshire County Council 2023).

The report highlights the findings of each metric for LTN roads and LTN boundary roads. The three LTN areas in East Oxford are the Divinity Road LTN, St Clements Road LTN, and ST Marys LTN. At each boundary road, the daily mean count is calculated for the pre-implementation period (20th Nov 2021 - 19th May 2022) and the post-implementation period (20th May 2022 - 10th April 2023). The report finds a decrease in the number of pedestrians on most boundary roads, except for a significant increase (+47%) in pedestrians on Morrell Avenue. It should be noted that parking spaces were removed along Morrell Avenue, leading people to walk further to reach locations on this road (Reference LTN report).

Regarding cycle counts, the report finds an increase in cycling volumes on all boundary roads, with Cowley Road reporting the largest increases (+32% and +16%). The increases on Morrell Avenue can be attributed to the cycle lanes that were installed during the evaluation periods (Oxfordshire County Council 2023).. Car volumes increased during the observation period at most monitoring locations. The report mentions a distinct difference in pre-existing volumes on Iffley Road and Cowley Road between the more northern and eastern locations, with lower volumes near the city centre and higher volumes at the outer sensors. These same sets of sensors see an increase in traffic post LTN implementation where the volume was lower, and a decrease in traffic where the volume was higher (Oxfordshire County Council 2023)..

On LTN roads, VivaCity count lines were only available to monitor traffic modes in the in-LTN areas. Data is not available from 2019 as sensors were not yet installed. Overall, the report finds that from the available data, car volumes have decreased at two LTN locations, pedestrian flows have slightly decreased (an average of 7%), and there are mixed changes in cycling volumes.

The report also highlights counts from automatic traffic count sensors (ATC). ATC can categorize 10 vehicle types: Cycle, Motorcycle, Car, LGV, Bus, and several categories of HGV. The daily average car count saw a general decrease across most ATCs, with a total average decrease of 61%. Five out of five ATCs that recorded higher car flows in 2021 all saw significant decreases of at least 50% (Oxfordshire County Council 2023). Motorcycle counts showed an increase in difference between 24% and 54% across all locations, due to the combination of reduced car traffic in LTNs and general increases in traffic on the boundary roads (Oxfordshire County Council 2023). A sensor on Bullingdon Road saw a significant impact increase of 70% in pedestrians. Of the two ATC locations that can record pedestrian counts, both experienced an 8% decrease. The daily average cycle volumes increased for both locations, with a similar trend of a 10% increase on Bullingdon Road (Oxfordshire County Council 2023). The snapshot report also evaluates air quality in East Oxford LTNs. Air Quality is sampled through diffusive samplers, which are widely used for indicative monitoring of ambient nitrogen dioxide (NO_2) (Oxfordshire County Council 2023). Of the 16 sensors with comparative 2021/2022 data, 11 show either no change, an insignificant decrease, or a significant decrease in NO_2 levels, while 5 show an insignificant or significant increase in NO_2 levels. Sensors placed inside LTNs show a decrease in NO_2 levels, while on the boundary roads, the picture is more mixed, with sensors on Morrell Ave showing reductions and sensors on St Clements showing a significant and consistent increase. In this research we will only consider cars and cyclists as our vehicular data. Both sets of counts are tracked by Vivacity sensors.

2.5 Prior Research: LTN, Air Quality, and impacts to health

Researchers have investigated the impacts of low traffic neighbourhoods (LTNs) on air quality and traffic. In a study on LTNs and their impact on NO2 and traffic, researchers found that LTNs have the potential to reduce air pollution and traffic in target areas without increasing air pollution or traffic volumes in surrounding streets (Yang et al., 2022). The study focused on Islington, North London, specifically three monitoring locations: St. Peter's, Canonbury East, and Clerkenwell. At each location, average NO2 levels and average traffic volume were measured outside, inside, at the LTN boundary pre-LTN and post-LTN. The study concludes that NO2 concentration reduced at the LTN boundary and internal sites for St. Peter's but decreased at the Canonbury East and Clerkenwell internal sites without significant changes at the boundary sites. Average traffic volume significantly reduced at the internal sites and decreased at the boundary sites compared to the external sites.

Other studies have explored the impact of LTNs on the community. In a study investigating LTNs, car use, and active travel, researchers found larger effects of the implementation, including decreased car ownership and use, as well as increased active travel. They suggest that LTNs are an important part of interventions aimed at reducing car use and promoting active travel (Aldred, R. and Goodman, A, 2020). Aldred and Goodman's People and Places study is a longitudinal study that treats the mini-Hollands program as a 'natural experiment'. The study involved participants providing demographic and socioeconomic information, along with a past-week travel diary documenting minutes of walking, cycling, and car use. The study concluded that there is a consistent trend towards reduced car use among LTN residents, and point estimates for active travel were the largest in the LTN areas (Aldred, R. and Goodman, A, 2020). Research shows that low traffic neighbourhoods (LTNs) are improving the quality of life for residents, particularly in terms of air quality improvement. The aim of this research is to understand the individual benefits for residents in LTN areas, which can be mental, physical, or social in nature. Another significant benefit of LTNs is the increased uptake of walking and cycling. In a study titled "Environmental Correlates of Walking and Cycling: Findings from Transportation, Urban Design, and Planning Literatures" by Saelens et al., researchers investigated the associations between physical environment variables and individuals' walking and cycling for transport. The study suggests that residents living in communities with higher density, greater connectivity, and more diverse land use report higher rates of walking and cycling for utilitarian purposes compared to low-density, poorly connected, and single land use neighbourhoods. This research supports the need for safer active travel options in urban cities due to the higher demand compared to rural areas.

Previous research in transportation and planning has also explored the association between the physical environment and physical activity, particularly walking and cycling. A notable example is a comparison between the United Kingdom and the United States regarding the safety of walking and cycling in urban areas. In the United States, residential environments are associated with low levels of walking for transport, primarily due to land use and design that hampers accessibility for these activities. On the other hand, residential areas in the United Kingdom have a higher rate of physical activity as the land design promotes accessibility for walking and cycling.

Research into transport infrastructure has shown that the provision of new sustainable transport infrastructure effectively promotes an increase in active commuting (Panter, Jenner et al. 2016). In this paper, we will explore whether Oxfordshire's efforts to promote safe active travel through the implementation of LTNs align with prior research on sustainable transport infrastructure. Previous research has established a connection between physical activity and health status. Warburton et al. conducted a study that demonstrated a linear relationship between physical activity and fitness (Warburton et al. 2006). Their review further highlights the role of physical inactivity in the development of chronic diseases and premature death. Oxfordshire County Council is dedicated to creating safe environments for residents while promoting healthy lifestyle habits, such as engaging in physical activity. The Oxfordshire Cycling Design Standards (Summer 2017) and Oxfordshire Walking Design Standards articulate the council's vision for improved active travel infrastructure. These standards align with their goals of making cycling and walking integral to transportation, planning, health, and clean air strategies. The national attention given to creating healthy environments that support increased walking and cycling is reflected in the government's publication of the Cycling and Walking Investment Strategy (Warburton et al. 2006). Oxfordshire's efforts to encourage walking and cycling stem from the numerous benefits residents can enjoy from increased physical activity. Cycling is widely recognized as an excellent form of exercise that helps individuals meet the recommended physical activity levels outlined by the NHS. It has positive effects on mental health and well-being and reduces the risk of life-threatening conditions such as cancer, type 2 diabetes, heart disease, and obesity. Promoting cycling can also contribute to addressing the health challenges that place significant strain on healthcare services. Studies investigating the benefits of cycling have shown a lower likelihood of death from cardiovascular disease (-24%) and cancer (-16%) compared to commuting by car (Cherry 2022). Walking, often overlooked as a form of exercise, is another fantastic way to improve physical health. Many people engage in walking daily without fully realizing the physical benefits it offers. Walking is simple, free, and one of the easiest ways to be active and improve overall health. It also has positive effects on mental well-being, reducing stress, improving sleep, and boosting energy levels.

2.6 Prior Research: Causal Impact Analysis

One of the methodologies that will be employed in this research is causal impact using Bayesian Structural Time-Series Models. Causal Impact analysis is a statistical approach used to estimate the causal effect of an intervention or treatment on a target time series, using Bayesian Structural Time Series (BSTS) models. This methodology allows us to determine whether the intervention had a significant impact on the outcome variable and quantify the magnitude of that impact. BSTS models are a class of Bayesian time series models that can capture the underlying structure and dynamic of a time series. They are composed of several components: trend, seasonality, autoregression, and any other known or potential causal factors (covariates). The key idea behind Causal Impact Analysis using BSTS models is to model the outcome variable (response) as a function of the intervention variable (treatment) and other potential causal factors. The intervention is usually a binary variable indicating the presence or absence of the treatment. In the paper, “Inferring Causal Impact using Bayesian Structural Time-Series Models,” Brodersen et al uses this methodology to investigate an important problem in econometrics and marketing. Their goal is to infer the causal impact that a designed market intervention has exerted on an outcome metric over time (Broderson, K.H et al 2015). Structural time-series models are state-space models for time-series data. The Bayesian structural time series methodology will be explained more in Chapter 4.

Commented [BS2]: I would add a sentence here to say these will be explained in more detail in Chapter 4

2.7 Prior Research: ARIMA

The other methodology employed in this research is Autoregressive Integrated Moving Average (ARIMA). ARIMA models predict future values based on past values. ARIMA makes use of lagged moving averages to smooth time series data. ARIMA has been employed in various contexts. Price forecasting is becoming increasingly relevant to producers and consumers in the electricity market. In the paper "ARIMA models to predict next-day electricity prices ", Contreras et al. attempt to predict next-day electricity prices based on ARIMA methodology (Contreras et al. 2003). Overall, ARIMA models are method to predict short run and long run trends based on prior time series data. In a different context, research using ARIMA methods have explored forecasting Egyptian GDP. Abonazel et al utilize the Box-Jenkins approach has been used to build the appropriate Autoregressive-Integrated Moving-Average (ARIMA) model for the Egyptian GDP data and fitted the model to forecast the GDP of Egypt for the next ten years (Abonazel et al. 2019). ARIMA methodology is one of the most popular methods for time series forecasting. Our research will utilize this methodology to understand the impact of Low Traffic Neighbourhoods in Oxfordshire County.

Chapter 3: Data

3.1 About the Data

Traffic volume data is provided by Oxfordshire County Council on behalf of Alchera Technologies, a leading authority in enterprise-grade mobility and infrastructure systems. Alchera specializes in delivering software tools and machine learning applications to enhance data-driven infrastructure. Through a close collaboration between Alchera and Oxfordshire County Council, a comprehensive platform has been developed, encompassing all transport-related data. Within the boundaries of Low Traffic Neighbourhoods (LTNs) and specific roads within these areas, advanced machine vision techniques are employed by the VivaCity object identification sensors to capture and categorize various road users, including cars, pedestrians, and cyclists. This study specifically focuses on the following locations: In East Oxford, the LTN areas consist of four count lines, three of which are situated in Divinity Road near the junction with Cowley Road, and one in Leopold Street, also near the Cowley Road junction. The LTN Boundary Roads in East Oxford comprise seven count lines, including Iffley Road, St Clements Street, two at Cowley Road, and three at the Morrell and Divinity intersection. It is important to highlight that no personal data is collected, and the raw camera images remain unused. The datasets utilized in this research cover the period from January 1, 2019, to July 14, 2023. Through this exploratory analysis, our objective is to understand the impact of COVID and the implementation of LTNs on traffic volume. The transportation sector and active travel continue to experience the effects of COVID. Examining traffic volume prior to the pandemic allows us to draw conclusions about how the implementation of LTNs has influenced traffic counts. For our analysis, we consider the pre-COVID period to be the entire year of 2019. The UK began its phased exit from lockdown in March 2021. The pre-implementation count period for LTNs is considered from November 2021 until May 2022. The post-implementation period for LTNs is considered from May 20, 2022, until the present. 2019 and 2020 data are not available for location at Morrell Boundary Road.

3.2 Data Contents and Processing

This section will describe the contents of the data used in this research. Categorised flow counts are extracted from the Viva City API aggregated hourly. Covariates include:

- Time from
- Time to
- Location
- Direction
- Count line ID
- Count line Name
- Available Directions
- Cyclist
- Car
- Pedestrian
- other vehicle covariates: motorbike, taxi, van, minibus, bus, rigid, truck, emergency car, emergency van, fire engine
- Aggregated record count
- Data Quality Warning
- ID
- Ingest Time (UTC)

The original data observations were contained in a CSV file. This data was the read into R for pre-processing and wrangling. The data was subset to relevant columns: Time from, Count line Name, Cyclist, Car, and Pedestrian. This subset data contained over 800,000 observations which spanned from January 1, 2019, to July 14, 2023. For certain tasks like exploration and modelling, this subset was summarised to monthly and hourly averages to reduce computational complexity. All missing values were dropped.

3.3 Exploratory Data Analysis

3.3.1 Pedestrian Traffic Counts for LTN areas and boundary roads

First, we conduct exploratory data analysis on the traffic volume data. To begin, will investigate the impact LTNs have on the pedestrian counts at LTN areas. *Figures 1 and 2* is the comparison of hourly and monthly pedestrian counts for all Viva City sensors in LTN areas. Immediately it can be observed that COVID significantly reduced the pedestrian volume. In 2019, pedestrian volume was much higher on average compared to 2020, 2021, and 2022. The increase between 2020 and 2021 represents the lifting of COVID restrictions and the return to normal life, however pedestrian volume did not return to the level it was pre-covid.

In 2023, we observe a large increase in pedestrian volume. This increase is likely due to the LTN implementation as its purpose is to encourage safe active travel.

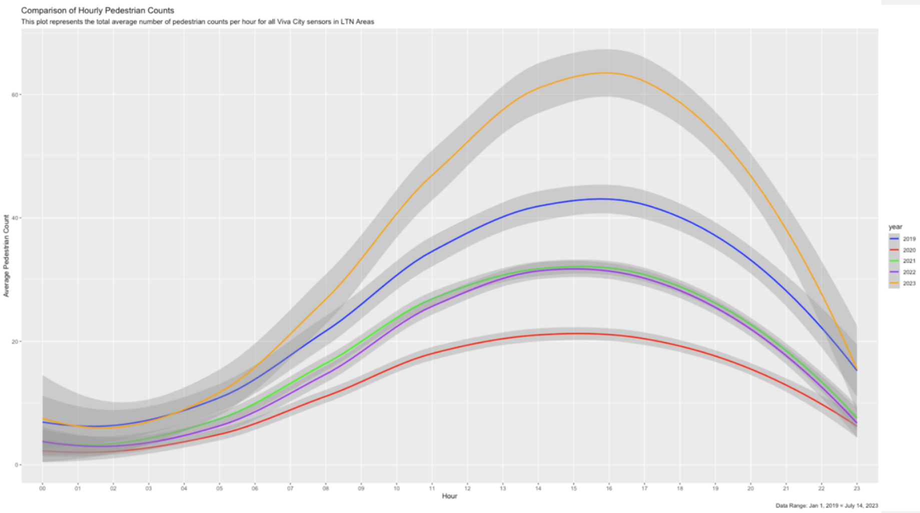

Figure 1: Comparison of Hourly Pedestrian Counts in LTN Areas (junction of Divinity Rd and Cowley Rd)

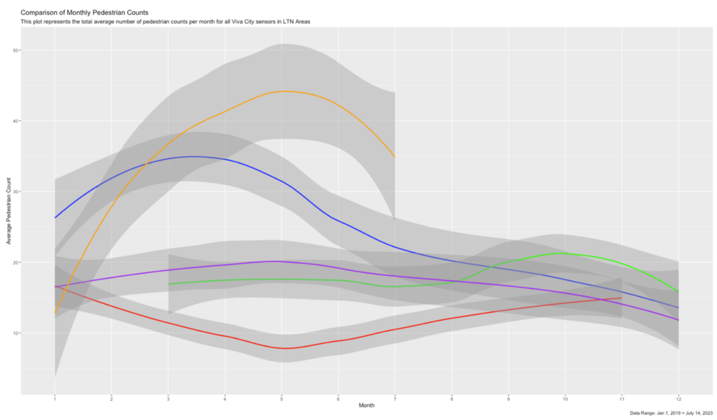

Figure 2: Comparison of Monthly Pedestrian Counts in LTN Areas (junction of Divinity Rd and Cowley Rd)

In the monthly comparison of pedestrian counts shown in *Figure 2*, we observe similar trends that were observed in *Figure 1*. As a reminder, the LTN implementation period started May 2022. We see that in 2023, the monthly averages from March to July are much higher than they were pre covid further proving the impact of the implementation. Overall, it is observed that pedestrian volume in the LTN area has increased after the LTN implementation and COVID.

Next, we focus on pedestrian volume at LTN boundary roads. *Figures 3 and 4 are* comparisons of hourly and monthly pedestrian volume by region. At the Cowley Road East count line, we see that pedestrian volume is much higher in 2023 compared to every other year. Pre-covid and pre-implementation we see pedestrian levels are similar. In 2022, pedestrian levels were higher which could be caused by LTN implementation and return to normal lifestyle post COVID. At Iffley, pedestrian levels have not returned or exceed to where they were pre-COVID.

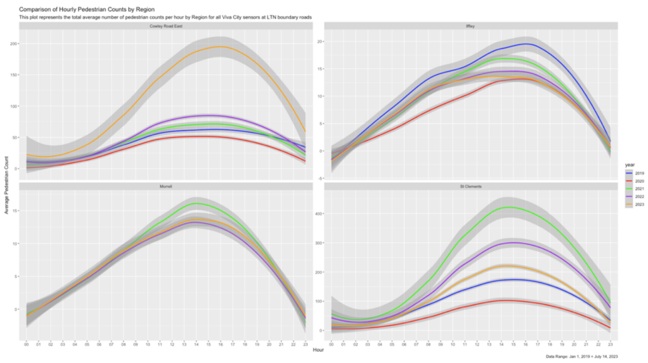

Figure 3: Comparison of Hourly Pedestrian Counts at LTN boundary roads

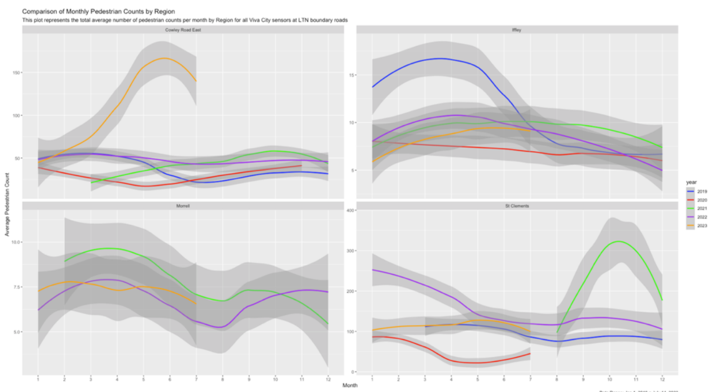

Figure 4: Comparison of Monthly Pedestrian Counts at LTN boundary roads

The monthly comparison in *Figure 4* shows us similar results. In 2023, pedestrian volume was the largest at Cowley Road East. This sensor is in the same area as the LTN area sensor, so we expect there to be a larger number of pedestrians included at the boundary sensor compared to other boundary count lines.

In examination of pedestrian volume at LTN area and LTN boundary road sensors, we observe that the implementation of LTNs has led to increases in pedestrian volume as noticed by LTN area count lines and the Cowley Road East LTN boundary county line. Through *Figures 1 and 2* we clearly see the affect COVID had on pedestrian volume at the junction near Divinity Road and Cowley Road. 2023 hourly and monthly hourly averages have surpassed the pre covid averages which can be partly attributed to the implementation of the LTN and the encouraged safe active travel.

3.3.2 Cyclist Traffic Counts for LTN areas and boundary roads

Next, we explore cyclist counts for LTN areas and boundary roads. *Figures 5 and 6* are comparisons of hourly and monthly cyclist counts for the count lines in the LTN areas. In the LTN area, we notice a lower level of cyclist counts compared to pedestrian counts just by the scale of *Figure 5*. Each year, cyclist volume appears to follow a similar trend throughout the day. In 2020, cyclist volume drops compared to 2019. This decrease can be attributed the nationwide COVID lockdowns. In 2021 and 2022, cyclist levels are similar and at a higher level than in 2022. In 2023, we see an increase in cyclist volume and hourly averages are much higher from noon to midnight compared to the pre-COVID period. When looking at the monthly averages in *Figure 6*, it can be observed that monthly averages are the highest in 2023 and the lowest in 2020.

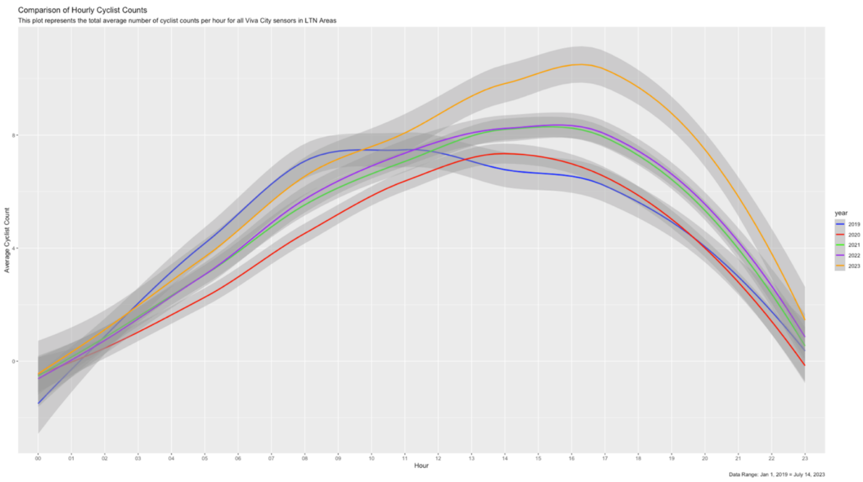

Figure 5: Comparison of Hourly Cyclist Counts in LTN Areas (junction of Divinity Rd and Cowley Rd)

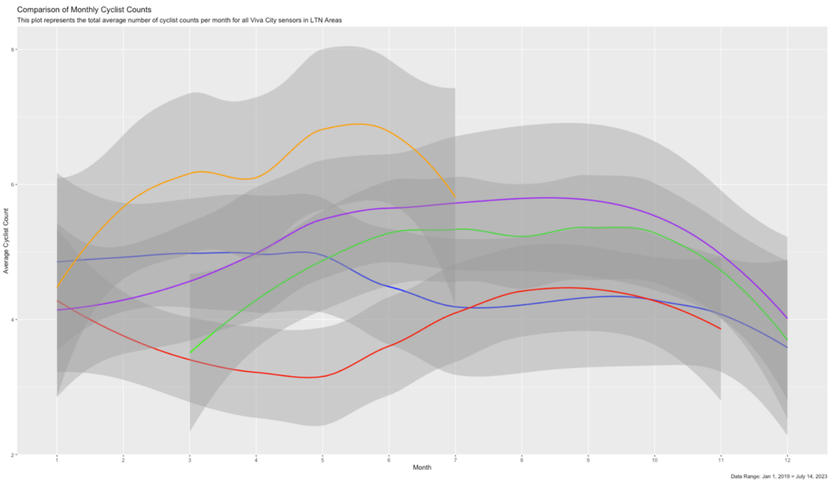

Figure 6: Comparison of Monthly Cyclist Counts in LTN Area (junction of Divinity Rd and Cowley Rd)

Next, we explore cyclist volume at the LTN boundary count lines. *Figures 7 and 8* provide comparisons of hourly and monthly cyclist counts for the various count lines at LTN boundary roads.

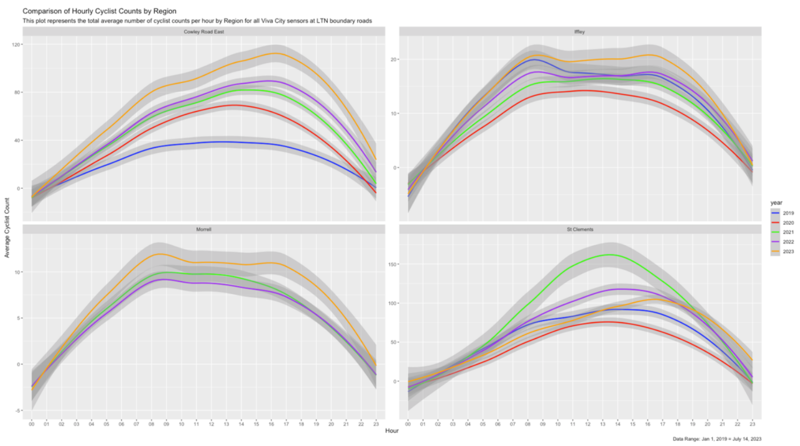

Figure 7: Comparison of Hourly Cyclist Counts at LTN boundary roads.

At the Cowley Road East count line, we notice that cyclist levels are the largest post LTN implementation. Cyclist levels were at their lowest pre-COVID and have increased every year since. At Iffley, we see that cyclist volume in 2023 is higher than other years. In 2022, we see the lowest cyclist volume which can be attributed to COVID. Since then, cyclist volume has increased and has now surpassed pre-COVID levels. At Morrell, cyclist volume also increased on in 2023. However, at the St Clements count line cyclist volume was at its highest in 2021, the year where COVID restrictions were lifted and the pre-LTN implementation period. At St Clements it appears that the LTN implementation has reduced cyclist volume throughout most of the day.

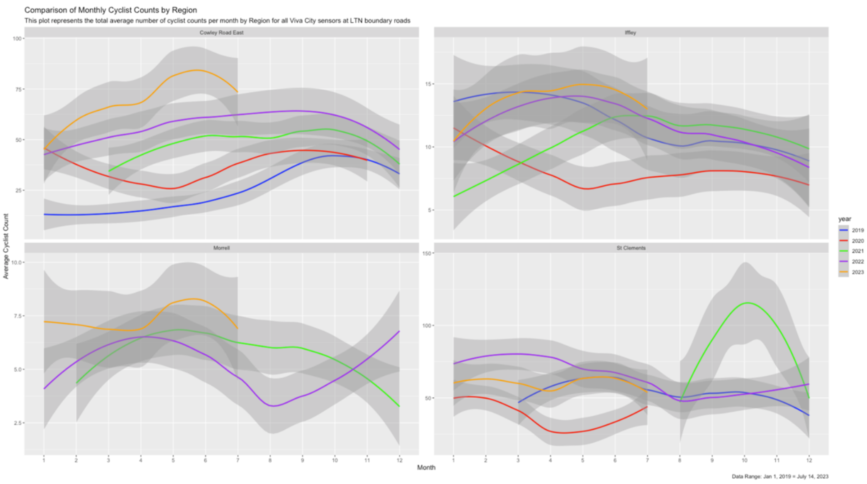

Figure 8: Comparison of Monthly Cyclist Counts at LTN boundary roads.

Through this exploration of cyclist volume at LTN areas and LTN boundary roads, we see that the implementation of the LTN has led to an increase in cyclists in the LTN areas. This is confirmed by observing hourly and monthly cyclist volume at the LTN area show in *Figure 5 and 6* and the Cowley Road East count line in *Figures 7 and 8*. We observed the effect of the COVID lockdowns where 2020 cyclist averages are much lower throughout the day and year. This is evident at the Iffley and St Clements count lines.

3.3.3 Car Traffic Counts for LTN areas and boundary roads

Finally, we explore car volume at LTN areas and LTN boundary roads. *Figures 9 and 10* are comparisons of hourly and monthly car counts in LTN areas. Inside the LTN area, we notice in the pre-COVID period, car volume was at its highest. During COVID, car volume dropped off significantly. Post COVID, car saw its largest counts in 2021, but then decreased again in 2022. This decrease can be attributed to the implementation of the LTN. As these sensors are inside the LTN areas we expect there to fewer cars passing throughout the day. In 2023, car volume is lower compared to every other year.

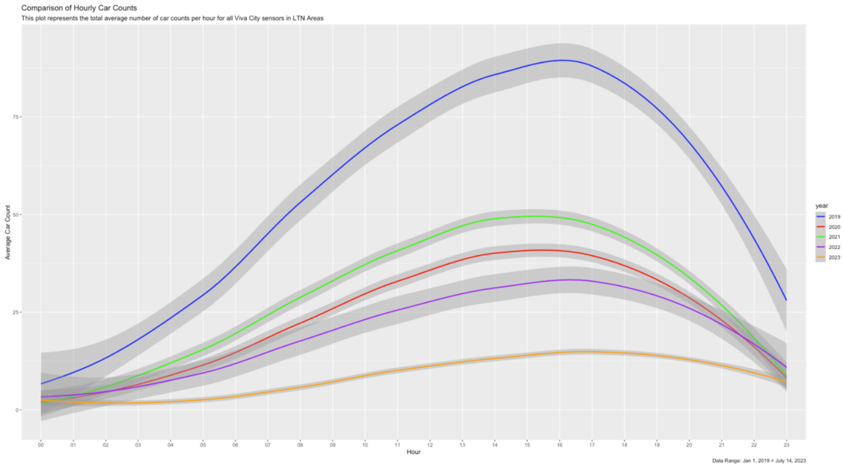

Figure 9: Comparison of Hourly Car Counts at LTN area (junction of Divinity Rd and Cowley Rd)

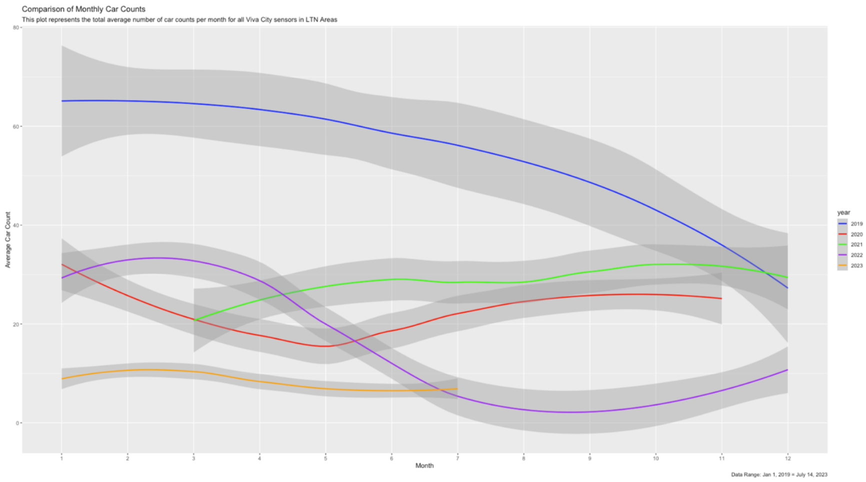

Figure 10: Comparison of Monthly Car Counts at LTN area (junction of Divinity Rd and Cowley Rd)

A comparison of hourly and monthly averages at LTN boundary count lines can be observed in *Figures 11 and 12*. At the Cowley Road East count line, we see that car volume has dropped off from 2019 to 2023. Car volume at this count line was at its peak pre-COVID. Post COVID, we see the hourly averages increase but then decrease again for 2022. This decrease can be attributed to the LTN implementation as we expect less cars and more cyclist and pedestrians. At Iffley, we see that car counts are more similar for all years. Again, car volume was at its peak pre-COVID, and it appears the implementation of the LTN has not led to more cars at this count line as 2023 levels are lower than 2019, 2021, and 2022. At St Clements, we notice that car volume was at its lowest pre-COVID and during the lockdown year. Hourly car counts increase significantly in 2021, reduce in 2022 and 2023, but are still higher than hourly car counts pre COVID.

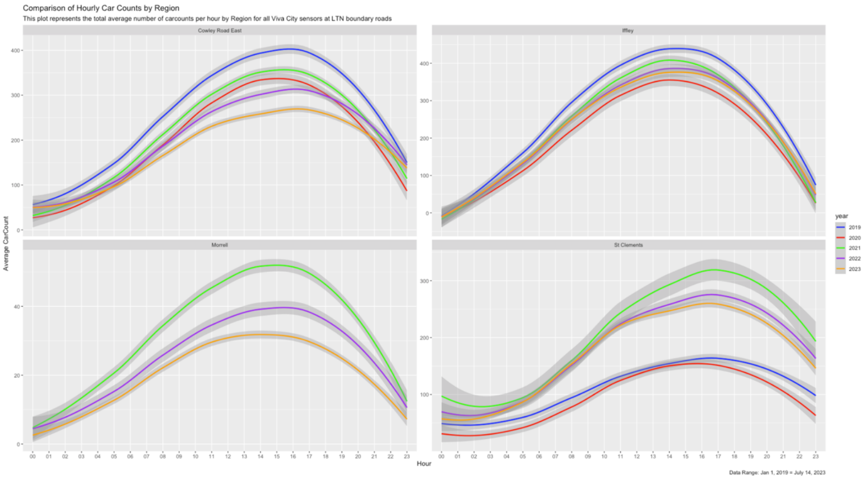

Figure 11: Comparison of Hourly Car Counts at LTN boundary roads

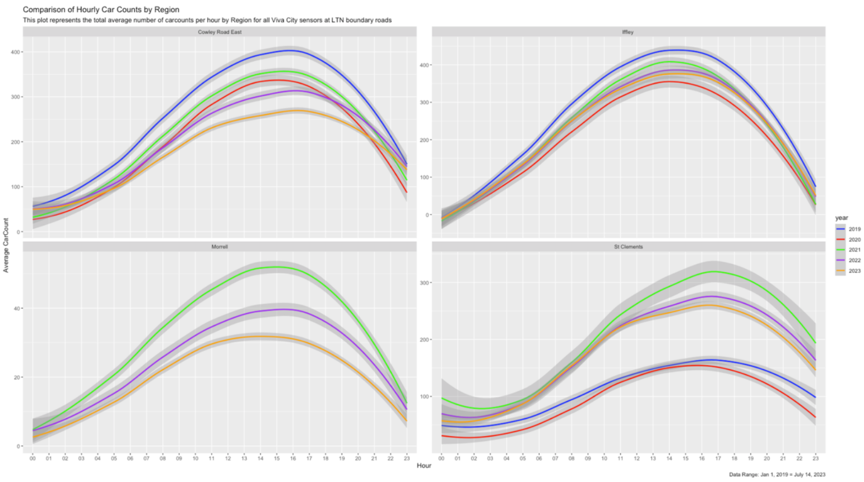

Figure 12: Comparison of Monthly Car Counts at LTN boundary roads

Overall, in exploring car volume in LTN areas and at LTN boundary roads, we can examine how counts have changed pre-COVID, post-COVID, pre-LTN, and post-LTN. We observed car volume decrease in LTN areas post-LTN implementation in *Figures 9 and 10*. However at the boundary roads, we get a mixed picture as car volume has been increasing at the St Clements count line post COVID.

Chapter 4: Methodology

4.1 Methods: Time Series Forecasting with Arima

Time series forecasting with ARIMA (Autoregressive Integrated Moving Average) is a popular method used to predict future values of a time series based on its historical data. ARIMA models are widely employed in various fields, including economics, finance, and engineering. For example, Benevenuto et al proposed a simple econometric model that could be useful to predict the spread of COVID-2019. They performed Auto Regressive Integrated Moving Average (ARIMA) model prediction on the Johns Hopkins epidemiological data to predict the epidemiological trend of the prevalence and incidence of COVID-2019 (Benevenuto et al 2020). Nevertheless, The ARIMA model is composed of three components: Autoregressive (AR), Integrated (I), and Moving Average (MA).

Commented [BS3]: It would be good if you could include a reference on where these have been used in a couple of different applications.

Autoregressive (AR) Component: The (AR) component captures the relationship between the current value of the time series and its past values. It assumes that the current value is influenced by its own previous values with a lag. The (AR) component is denoted by AR(p), where 'p' represents the order of the autoregressive process. The AR(p) equation can be written as:

(4.1) $X_t = c + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + \varepsilon_t$ where:

- X_t is the value of the time series at time ‘t’.
- c is a constant term.
- φ_1, φ_2, …., φ_3 are the autoregressive coefficients.
- ε_t is the white noise error term at time ‘t’.

Integrated (I) Component: The (I) component represents the differencing of the time series to make it stationary. Stationarity is a crucial assumption for ARIMA modelling, as it ensures the statistical properties of the time series do not change over time. Differencing is performed to remove trends and seasonality from the data. The differencing equation can be written as:

(4.2) $Y_t = (1 – B)^d X_t$ where:

- Y_t is the differenced time series.
- B is the backwards shift operator (e.g., $BX_t = X_{t-1}$)
- ‘d’ is the order of differencing.

Moving Average (MA) Component: The (MA) component represents the relationship between the current value of the time series and its past white noise error terms. It assumes that the current value is influenced by the weighted sum of the past error terms. The (MA) component is denoted by MA(q) where 'q' represents the order of the moving average process. The (MA) equation can be written as:

(4.3) $X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$ where:

- μ is the mean of the time series.
- $\theta_1, \theta_2, \ldots, \theta_q$ are the moving average coefficients.
- ε_t is the white noise error term at time 't'.

The general form of an ARIMA model is denoted as ARIMA(p, d, q) where 'p' is the order of the autoregressive component, 'd' is the order of differencing, 'q' is the order of the moving average component. The ARIMA(p, d, q) combines these three components to model time series data and make forecasts for future observations. The parameters (p,d,q) are typically selected using various model selection techniques, such as Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC), to find the best-fitting model for the given data.

In the context of our research, we will apply time series forecasting with ARIMA to data from LTN areas and boundary roads for pedestrian, cyclist, and car counts. First, we fit an ARIMA model to a univariate time series for each type of count. In, R, we use the function auto.arima() which automatically gives the best ARIMA model for the specified time series input. We specify two parameters: Stepwise = FALSE and Parallel = TRUE. By setting stepwise to false, the ARIMA model will consider all combinations of ARIMA orders (p, d, q) and seasonal orders (P, D, Q, s) without performing a stepwise search. By setting parallel to true, we allow the ARIMA function to utilize multiple cores or threads for parallel computation when evaluating different models. Parallelization can speed up the model selection process, especially when we have a larger number of potential models to evaluate. By using this combination of parameters, we are instructing the function to perform an exhaustive search of ARIMA models without using the stepwise algorithm and to use parallel processing for faster evaluation of models. This method will select the best fitted ARIMA model based on the chosen model section criteria.

First, we will use all data from Jan 2019 to July 2023 and forecast for the rest of 2023 and 2024. Next, we will use data from the pre implementation period to forecast the post implementation period and then compare the forecast to the original data. The difference from between the observed and forecasted values will give us insight on the impact of the LTN implementation. We will also employ out of sample cross validation as well as assess accuracy metric measure the performance of the model. Below are the accuracy metrics, their definitions, and equations:

Mean Error (ME) - represents the average difference between forecasted values and the actual observed values. It can be positive or negative which indicates whether forecasts on average are overestimating or underestimating the actual values.

Commented [BS4]: Could you add a mathematical equation for each of these metrics?

ME Formula: ME = Σ (predicted value - actual value) / n where:

- Σ denotes the sum of the differences for all data points.
- "Predicted value" is the value predicted by the model.
- "Actual value" is the true value observed in the data.
- "n" is the number of data points.

Root Mean Squared Error (RMSE) – measure of the typical magnitude of forecast errors. It represents the square root of the average of the squared differences between forecasted values and the actual values. Small values for RMSE indicate better forecasting accuracy.

RMSE = √ [Σ (predicted value - actual value) ² / n]

Mean Absolute Error (MAE) – measure of the average absolute difference between the forecasted values and the actual observed values. It represents the average of the absolute values of the forecast errors. Lower MAE values indicate better accuracy.

MAE = Σ |predicted value - actual value| / n

Mean Percentage Error (MPE): calculates the average percentage difference between the forecasted values and the actual observed values. It is expressed as a percentage and indicates the average percentage by which the forecasts are off from the actual values. Positive MPE indicates an overall overestimation, while a negative MPE indicates an overall underestimation.

MPE = (Σ [(predicted value - actual value) / actual value] / n) * 100

Mean Absolute Percentage Error (MAPE): like MPE but takes the absolute values of the percentage errors before averaging. It represents the average absolute percentage difference between the forecasted and actual values. Lower MAPE values indicate better forecasting accuracy.

MAPE = (Σ | (predicted value - actual value) / actual value| / n) * 100

Mean Absolute Scaled Error (MASE): measure of forecasting accuracy that compares the forecast's performance to that of a naïve forecast. It is calculated as the mean absolute error of the forecast divided by the mean absolute error of the naive forecast. MASE values close to 1 indicate the forecast's performance is like or better than naïve forecast.

MASE = (MAE / MAE naïve)
where:

- MAE represents the Mean Absolute Error of the forecasted values.
- MAE naïve represents the Mean Absolute Error of a naïve forecast.

Mean Cross-Validation Error (Mean CV Error): This is the average error calculated during the process of cross-validation. Cross validation is a technique used to evaluate the performance of a predictive model on an independent dataset. The Mean CV error indicates the average prediction error across multiple folds or partitions of the data during cross-validation. It gives an estimate of how well the model is likely to perform on unseen data.

Mean CV Error = Σ (CV value) / n

4.2 Assumptions: Time Series Forecasting

Before employing these techniques, it is protocol to check if assumptions are valid or to at least acknowledge assumptions that might be violated. Checking assumptions adds credibility to the methods employed. For time series forecasting we must check the following assumptions: Stationarity, Independence, Constant Variance, No Seasonal Patterns, Normality. First, we want to ensure our time series is stationary. Stationarity means that the mean, variance, and autocorrelation structure should remain constant overtime. Stationarity ensures that the statistical properties of the time series do not change. To evaluate stationarity, we will observe the ACF of type of count for LTN areas and LTN boundary roads.

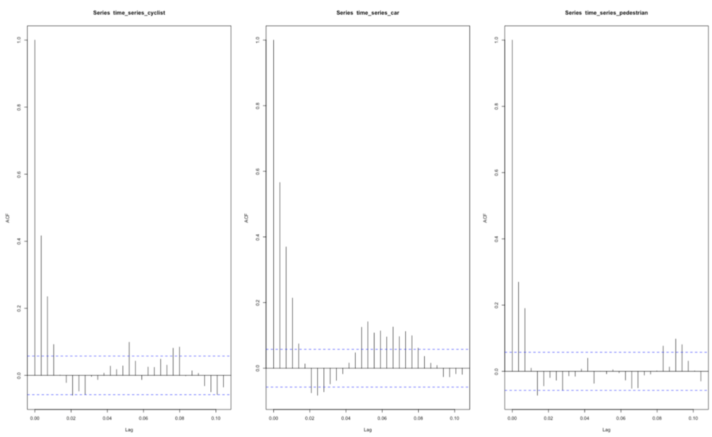

Figure 13: ACF plots for Time Series objects for Car, Cyclist, and Pedestrian for LTN boundary road data

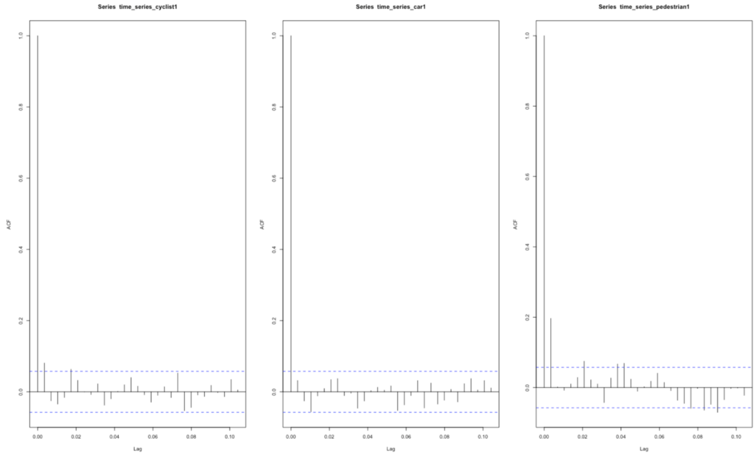

Figure 14: ACF plots for Time Series objects for Car, Cyclist, and Pedestrian for LTN boundary road data

Based on the ACF plots provided in *Figures 13 and 14*, we can determine that there is a lot of auto correlation throughout the data especially as time increases for both LTN boundary roads and LTN area data. We can conclude that our data is not stationary from observing ACF plots. These ACF plots also tell us there is a seasonality in our data. Though we have violated another assumption, we must make sure that our time series model accounts for seasonality. The independence assumption is also violated here. Since we have traffic count data at specific locations, each count observation could influence another account. For example, a larger number of counts at a certain time could be due to rush hour or groups of individuals travelling for similar purposes. Next, we must check the constant variance assumption. Here, we examine time series plots for each variable displayed in *Figure 15 and 16*. Looking at *Figure 15 and 16*, we can identify cycle trends and patterns for each time series. This is enough to prove that we do not have constant variance in the data. Lastly, we check normality by performing the Shapiro-Wilk Test for each time series. We choose our significance level to be 0.05.

Time Series	Car (LTN area)	Cyclist (LTN Area)	Pedestrian (LTN Area	Car (LTN Boundary)	Cyclist LTN boundary	Pedestrian LTN boundary
W	0.52927	0.42911	0.59335	0.78761	0.76622	0.88424
p-value	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16

Table 1: Shapiro Wilks Test statistics and p-value for each Time Series object

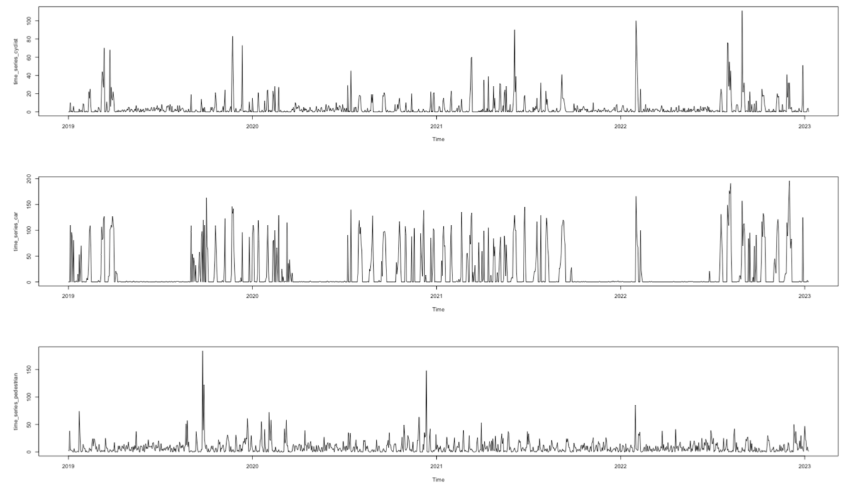

Figure 15: Time Series plot for Car, Cyclist, and Pedestrian for LTN boundary road data

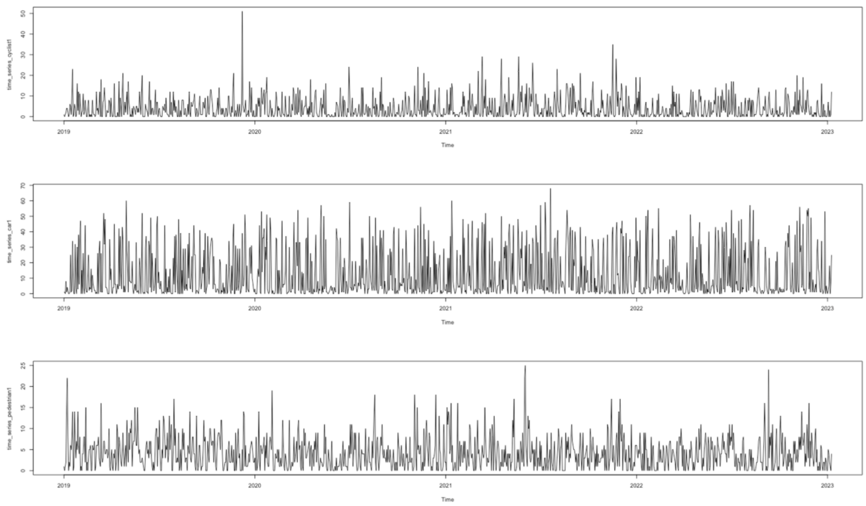

Figure 16: Time Series plot for Car, Cyclist, and Pedestrian for LTN Area data

The null hypothesis for Shapiro-Wilk Test is that the sample has been generated from a normal distribution. Based on the p-values in *Table 1*, we can observe that every time series objects rejects that null hypothesis giving evidence that the data is not normally distributed. Overall, many assumptions are violated for our time series forecasting method. Despite these violations, we can still proceed with caution. We must acknowledge the violation of these assumptions as limitations to our research. Since we are observing a large dataset with many observations collected hourly and daily, we must be careful when making forecast predictions. We are however using data from the population, so we can draw cautious conclusions from the results.

4.3 Methods: Causal Impact / Bayesian Structural Time Series (BSTS) Modelling

Causal Impact, also known as Bayesian Structural Time Series (BSTS) modelling, is a statistical method used to estimate the causal impact of an intervention on a time series. It is particularly useful when attempting to understand the effect of a specific event or treatment on the time series. The BSTS model is based on Bayesian statistics, which allows us to incorporate prior knowledge and uncertainty into the model. The main components of the BSTS model include:

State Space Model: The BSTS model represents the time series as a state space model, which consists of two components: the state equation and the observation equation. The state equation describes how the underlying states (latent variables) of the system evolve over time and is written as the following:

(4.3.1) $\theta_t = \boldsymbol{G} * \theta_{t-1} + w_t$ where:

- θ_t is a vector of underlying states at time 't'.
- $\boldsymbol{G}$ is the state transition matrix that describes how the states evolve over time.
- w_t is a vector of process errors (white noise) at time 't' and are assumed to follow a multivariate normal distribution with mean zero and covariance matrix W.

The observation equation relates the observed data to the underlying states and is written as the following:

(4.3.2) $y_t = \boldsymbol{F} * \theta_t + v_t$ where:

- Y_t is a vector of observed data at time 't'.
- $\boldsymbol{F}$ is the observation matrix that relates the observed data to the underlying states.
- v_t is a vector of observation errors at time 't' and are assumed to follow a multivariate normal distribution with mean zero and covariance matrix V.

Intervention Effect: The BSTS model includes an intervention effect, which is used to model the impact of the intervention on the time series. This effect is often represented as a step function, where the value changes at the intervention date. The intervention effect equation can be written as follows:

(4.3.3) Intervention Effect = $\boldsymbol{X} * \boldsymbol{\delta}$ where:

- $\boldsymbol{X}$ is a matrix representing the intervention design matrix
- $\boldsymbol{\delta}$ is a vector of parameters representing the effect of the intervention

4.4 Assumptions: Causal Impact / Bayesian Structural Time Series (BSTS) modelling

To employ BSTS modelling methodology, we must verify the following assumptions: No omitted variables, data must have a stable relationship, common trends between pre intervention data and post intervention data, independence of observations, and normality. The first assumption is easy to verify as no covariates were omitted. We are examining univariate traffic counts against time, so our only variables of interest are the specific traffic count and time. To verify the stable relationship assumption, we must observe the relationship between counts and time as shown in *Figures 15 and 16*. The plot displays that there is no clear change in the relationship between each count type and time from pre-intervention and during intervention. Our independence assumption check carries from the previous section. Since we have traffic count data from specific locations, each count observation could influence another account. For example, a larger number of counts at a certain time period could be due to rush hour or groups of individuals travelling for similar purposes. As extra confirmation let's examine the autocorrelation function of the residuals. The normality assumption can be checked by referring to *Table 1* with the Shapiro Wilks test for normality. From Section 4.2, we determined that the normality assumption is violated as determined by the p-values from the Shapiro Wilks tests.

Chapter 5: Results

In this section, we will examine the results of the chosen methodologies. We employ time series forecasting with ARIMA to model what could happen for the rest of 2023 and 2024 for traffic counts inside the LTN area at the various LTN boundaries. Next, we employ Bayesian Structural Time Series, or Causal Impact Analysis to the traffic data for the LTN area and LTN boundaries. With BSTS, we aim to understand the impact of the intervention by modelling the counterfactual, or what we would expect counts to be like if LTNs were not implemented. Overall, we employ these two time series methodologies to understand more about the impact of the LTN implementation on traffic counts in Oxfordshire City Council.

5.1 ARIMA

5.1.1 Forecasting 2023-2024 traffic counts in LTN Area

First, we examine the ARIMA model summary output for each type of traffic count in the LTN Area. *Figure 17* is a plot of the forecast values for pedestrians, cyclists, and cars for the LTN Area. The observed data ranges from Jan 1, 2019, to July 14, 2023, and is represented by the black solid line. The forecast is set for 535 days which is to the end of 2024, and the forecasted values are represented by the blue solid lines. The shaded area around the forecasted values represents the 80% and 95% confidence intervals. *Table 2 is a table of the ARIMA model summary output for each type of traffic count.* The table displays the model and its parameters, estimated coefficients, standard errors, variance, log-likelihood and information criteria. *Table 2* is a comparison of accuracy metrics for the ARIMA models for each type of traffic count.

Count Type	ARIMA Model	Estimated Coefficients	Estimated Standard Errors	Variance	Log-Likelihood	Information Criteria
Car	ARIMA (1,0,1) (0,1,0) [288]	AR(1) 0.5834 MA(1)): 0.2680	AR(1): 0.0367 MA(1): 0.0428	274.1	-3680	AIC: 7366 AICc 7366.02 BIC:7380.31
Cyclist	ARIMA (1,0,3) (0,1,0) [288]	AR(1): 0.9644 MA(1): -0.323 MA(2): -0.3232 MA(3): -0.208	AR(1): 0.0165 MA(1): 0.037 MA(2): 0.036 MA(3): 0.0323	12.53	-2335.1	AIC: 4680.2 AICc: 4680.27 BIC: 4704.05
Pedestrian	ARIMA (1,0,0) (0,1,0) [288] with drift	AR(1): 0.6817 Drift: 0.0038	AR(1): 0.248 Drift: 0.0038	108.7	-3277.25	AIC: 6560.5 AICc: 6560.53 BIC: 6574.81

Table 2: ARIMA model summary output

Each summary output begins with the specification of the ARIMA model used for forecasting. The numbers inside the parentheses represent the order of autoregressive (AR), differencing (I), and moving average (MA) terms, respectively. For example, ARIMA(1,0,1)(0,1,0)[288] includes an AR(1) term, a differencing of order 1, an MA(1) term, and a seasonal period of 288 time points. The presence of AR and MA terms indicates that the model is capturing dependencies between past observations and residual errors. The estimated coefficients represent the values of the autoregressive (AR) and moving average (MA) terms in the ARIMA model. For example, in the Car Counts ARIMA model, the AR(1) coefficient (ar1) is 0.5834, and the MA(1) coefficient (ma1) is 0.2680. These coefficients indicate the strength and direction of the relationship between current observations and past observations or past forecast errors. The standard errors provide an indication of the uncertainty associated with the estimated coefficients. For instance, in the Cyclist Counts ARIMA model, the standard error for AR(1) is 0.0164, and for MA(1), it is 0.037. Smaller standard errors imply greater confidence in the accuracy of the coefficient estimates. Variance (sigma^2): The estimated variance (sigma^2) represents the variability of the error term in the ARIMA model. In the Car Counts ARIMA model, the variance is 274.1, while in the Pedestrian Counts ARIMA model, it is 108.7. Lower values of variance indicate that the model is better at explaining the variability in the data. The log likelihood is a measure of how well the ARIMA model fits the data. A higher log likelihood indicates a better fit. For example, in the Cyclist Counts ARIMA model, the log likelihood is -2335.1, which suggests a better fit compared to the Car Counts ARIMA model with a log likelihood of -3680. Information criteria, such as AIC, AICc, and BIC, are used for model selection. Lower values of these criteria indicate a better model fit while considering the complexity of the model. In all three models, the AIC, AICc, and BIC values for the Cyclist Counts ARIMA model are the lowest, suggesting that it provides the best trade-off between goodness-of-fit and model complexity. Overall, the results show the specifications and performance of the ARIMA models applied to the respective traffic data (Car Counts, Cyclist Counts, and Pedestrian Counts). The Cyclist Counts ARIMA model seems to be the most appropriate and well-fitted among the three models based on the information criteria and log likelihood values. We also must be cautious of comparing information criteria and log-likelihoods fitted to different data. This is discussed further in the limitations section, but since the data we are modelling with is of the same nature and distribution, comparing information criteria and log likelihood can be informative. It is essential to consider these results in the context of the research objectives and the practical implications of the models' performance.

Commented [BS5]: Does it make sense to compare information criteria/log-likelihoods fitted to different data?

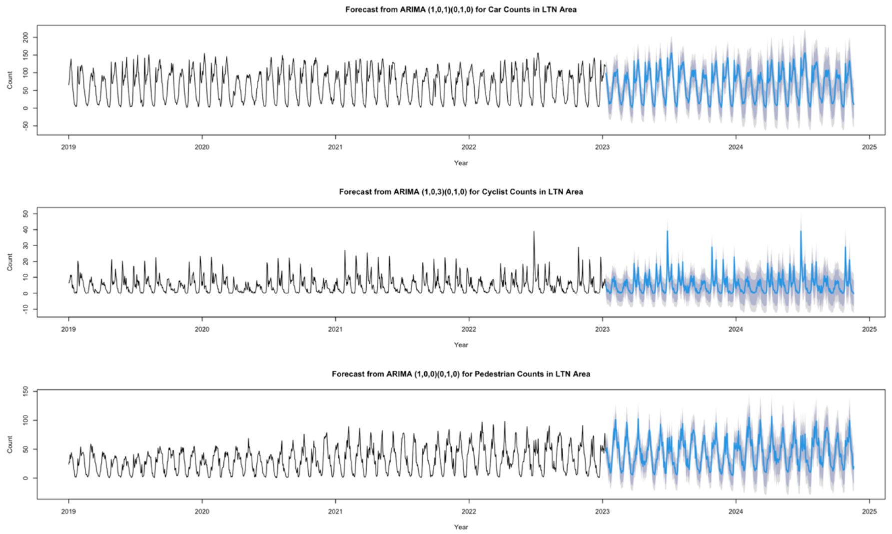

Figure 17: Forecasting after post-LTN implementation period (end 2023 – end of 2024) from ARIMA for Pedestrian, Cyclist, and Car Counts in LTN Area

In *Figure 17*, we observe the plot of observed and forecasted values for each traffic count in the LTN area. Each forecast seems to replicate the observed values, but for the cyclist forecasted values we notice two peaks in the forecast that can indicate an increase in cyclist counts at certain points in the forecast horizon. In pedestrian plot we notice that observed values from 2019 to about 2021 are relatively lower than the rest of the observed values. The forecasted pedestrian values are much larger than those observed in 2019 and 2021.

	ME	RMSE	MAE	MPE	MAPE	MASE	Mean CV Error
Pedestrian	0.05018721	5.323114	3.036398	-1.413978	24.7592	0.3160235	117.8976
Cyclist	0.008387	3.06125	1.6978	NaN	Inf	0.64844	4.33607
Car	-0.00075526	6.634951	3.843602	-Inf	-Inf	0.1586374	122.7247

Table 3: Comparison of accuracy metrics for ARIMA models for traffic counts in LTN areas

Table 3 is comparison of metrics that quantify the accuracy of the forecasts. By evaluating accuracy, we can understand how well the forecast captures the actual behaviour of the time series. For pedestrian counts, we get an RMSE of 9.0295 and an ME of -0.00462 which indicates that on average, the ARIMA model is slightly underestimating the pedestrian counts during the forecast period. Mean CV error is -2735.382 which suggest on average, the ARIMA model underestimates the pedestrian counts during cross-validation. This value indicates bias in the model's predictions during the cross-validation process. For cyclist counts, ME is 0.008387 which indicates that on average, the ARIMA model is slightly overestimating the cyclist counts during the forecast period. RMSE is 3.06125 which is smaller than RMSE for pedestrian counts. This value is smaller than for pedestrian counts which means the ARIMA model is forecasting better for cyclist counts than pedestrian counts. The mean CV error is -328.3852 which suggests that the ARIMA model underestimates the cyclist counts during cross-validation. This value indicates bias in the model's predictions during the cross-validation process. For car counts, ME is 0.3154018 which indicates that on average the ARIMA model slightly overestimates the car counts during the forecast period. RMSE is 14.33649 which is much higher compared to the RMSE for pedestrian and car counts. This indicates that the ARIMA model is not forecasting as accurate for cars. Mean CV error is -763.522 which suggest that on average the ARIMA model underestimates the car counts during cross-validation. This value indicates bias in the model's predictions during the cross-validation process.

5.1.2 Forecasting 2023 -2024 traffic counts at LTN boundary roads

Here, we examine the results from forecasting 2023 and 2024 counts with ARIMA at LTN boundary roads. This section will be split up by LTN boundary locations. For each location we discuss summary output, display plots with the observed data and forecasted values for pedestrians, cyclists, and cars. The observed data ranges from Jan 1, 2019, to July 14, 2023, and is represented by the black solid line. The forecast is set for 535 days which is to the end of 2024, and the forecasted values are represented by the blue solid lines. The shaded area around the forecasted values represents the 80% and 95% confidence intervals. We also examine accuracy metrics for the models to understand how well the models are fitting the data.

Iffley Road LTN Boundary

Count Type	ARIMA Model	Estimated Coefficients	Estimated Standard Errors	Variance	Log-Likelihood	Information Criteria
Pedestrian	ARIMA (3,1,2) (0,1,0) [288]	AR(1): 1.8050 AR(2):-0.9279 AR(3): 0.0630 MA(1): -1.9222 MA(2): 0.9351	AR(1): 0.0373 AR(2):-0.0650 AR(3): 0.0362 MA(1): 0.0156 MA(2): 0.0159	4.554	-1896.12	AIC: 3804.24 AICc: 3804.34 BIC: 3832.85
Car	ARIMA (1,0,0) (0,1,0) [288]	AR(1): 0.3980	AR(1): 0.0311	329.4	-3759.44	AIC: 7522.87 AICc: 7522.88 BIC: 7532.41

Table 4: ARIMA model summary output

Table 4 is the ARIMA model summary output for traffic counts at the Iffley Road LTN Boundary. For pedestrian counts, the ARIMA (3,1,2) (0,1,0) [288] model is utilized. The estimated coefficients for the autoregressive (AR) and moving average (MA) terms are shown, including AR(1): 1.8050, AR(2): -0.9279, AR(3): 0.0630, MA(1): -1.9222, and MA(2): 0.9351.

The standard errors for these coefficients are also given. The estimated variance is 4.554. The log-likelihood value is -1896.12, indicating how well the model fits the data. The information criteria are presented as AIC: 3804.24, AICc: 3804.34, and BIC: 3832.85, which are used for model selection. The ARIMA (3,1,2) (0,1,0) model suggests that the pedestrian count data exhibits some degree of seasonality (since it includes differencing) and requires autoregressive terms up to order 3 and moving average terms up to order 2 to capture its dynamics. The relatively complex model structure could indicate that pedestrian movement is influenced by various factors, including seasonality, day-to-day variations, and potentially external events. For car counts, the ARIMA (1,0,0) (0,1,0) [288] model is applied. The car count data is modeled with an autoregressive term of order 1 and no moving average terms. This suggests that the data might exhibit a short-term dependency or trend, but not strong evidence of moving average effects. The simpler ARIMA model for car count data at the Iffley Road LTN boundary might suggest that car counts exhibit similar characteristics across different LTN boundaries. The simpler model structure could indicate that car movements might be influenced by more straightforward trends or relatively short-term patterns. The estimated coefficient for the autoregressive (AR) term is AR(1): 0.3980. The standard error for this coefficient is AR(1): 0.0311. The estimated variance (sigma^2) is 329.4. The log-likelihood value is -3759.44. The information criteria are presented as AIC: 7522.87, AICc: 7522.88, and BIC: 7532.41, used for model selection. These results provide valuable insights into the forecasting models' performance for pedestrian and car counts at the Iffley Road LTN boundary. The ARIMA (3,1,2) (0,1,0) [288] model seems to fit the pedestrian counts data well, as indicated by the lower information criteria values and the variance estimate. On the other hand, the ARIMA (1,0,0) (0,1,0) [288] model is used for car counts but shows relatively higher information criteria values. *Figure 18* is a plot of the forecasted values for 2023 and 2024 for pedestrian and car counts at the Iffley Road LTN Boundary. From the plot, we observe that the forecast has captured the seasonality from the observed values, however it appears that both forecasts for cars and pedestrians appear to overfit, or exactly replicate the observed values.

	ME	RMSE	MAE	MPE	MAPE	MASE
Pedestrian	0.005978739	1.843631	1.048049	6.159181	34.45717	0.3501385
Car	0.01114654	15.71501	9.426565	0.4135243	8.877896	0.1600954

Table 5: Comparison of accuracy metrics for ARIMA models for traffic counts at Iffley Road LTN Boundary

Table 5 displays the accuracy metrics for ARIMA models for traffic counts at the Iffley Road LTN Boundary. For the pedestrian model, the ME is approximately 0.006, indicating a slight overestimation of pedestrian counts on average. For the car mode, the ME is approximately 0.011, showing a slight overestimation of car counts on average. For pedestrians, the RMSE is around 1.844, indicating that, on average, the forecasts have an error of approximately 1.844 in predicting pedestrian counts. For cars, he RMSE is approximately 15.715, suggesting that the car count forecasts have an error of around 15.715 on average. The MAE for the pedestrian model is about 1.048, meaning that, on average, the forecasts have an absolute error of approximately 1.048 in predicting pedestrian counts. The MAE for the car model is approximately 9.427, indicating an absolute error of around 9.427 in predicting car counts on average. Both MPE values for the pedestrian and car models are positive indicating overestimation. For pedestrians this value is 6.16%, and for cars, this value is 0.41%. For Pedestrian: The MPE is approximately 6.16%, suggesting a slight overestimation of pedestrian counts on average. For the pedestrian model, the MAPE is around 34.46%, indicating that, on average, the forecasts have an absolute percentage error of approximately 34.46% for pedestrian counts. For cars, the MAPE is approximately 8.88%, showing an absolute percentage error of around 8.88% for car counts. Overall, it appears that the ARIMA models have relatively low accuracy for forecasting pedestrian counts (higher errors), but they perform reasonably well for car counts, with significantly lower errors compared to a naïve forecast.

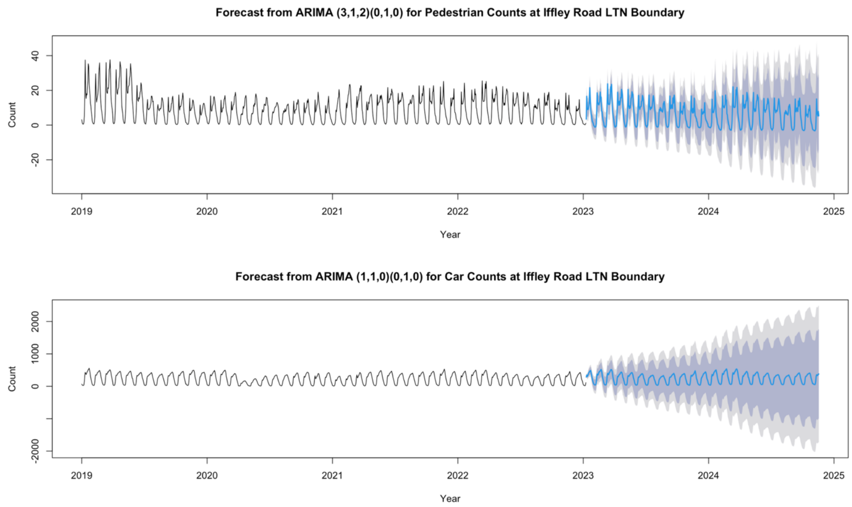

Figure 18: Forecasting after post-LTN implementation period (end 2023 – end of 2024) from ARIMA for Pedestrian and Car Counts at the Iffley Road LTN Boundary

In *Figure 18,* we observe the plot of observed and forecasted values for pedestrian and car counts at the Iffley Road LTN boundary. For pedestrians, we observe that in parts of 2019, the observed counts were much higher than the rest of the observed counts. The pedestrian forecast values appear to capture the seasonality of the observed data, but it appears that the forecasted values are decreasing closer to the end of 2024. We also notice for pedestrian counts that the confidence intervals are much larger towards the end of the 2024 compared to the earlier points in the forecast. For car counts, the forecasted values appear to exactly replicate the observed values, but the confidence intervals for the forecasted values are continuously increasing throughout the forecast horizon.

Morrell Avenue LTN Boundary

Count Type	ARIMA Model	Estimated Coefficients	Estimated Standard Errors	Variance	Log-Likelihood	Information Criteria
Car	ARIMA (1,1,2) (0,1,0) [288]	AR(1): 0.7956 MA(1): -0.7279 MA(2): -0.2217	AR(1): 0.0367 MA(1): 0.0428 MA(2): 0.373	15.53	-2424.77	AIC: 4857.55 AICc 4857.59 BIC: 4876.62
Cyclist	ARIMA (0,1,3) (0,1,0) [288]	MA(1): -0.1859 MA(2): -0.4945 MA(3): -0.1957	MA(1): 0.0332 MA(2): 0.0311 MA(3): 0.0367	6.533	-2053.66	AIC: 4115.32 AICc: 4115.37 BIC: 4134.39
Pedestrian	ARIMA (3,1,2) (0,1,0) [288]	AR(1): 0.5992 AR(2): 0.4352 AR(3): 0.2568 MA(1): 0.0077 MA(2): -.9689	AR(1): 0.0344 AR(2): 0.0392 AR(3): 0.0342 MA(1): 0.0110 MA(2): 0.0106	8.794	-2183.16	AIC: 4378.31 AICc: 4378.41 BIC: 4406.92

Table 6: ARIMA Summary Output

Table 6 presents the results of the ARIMA models used to forecast traffic counts for car, cyclist, and pedestrian data at the Morrell Avenue LTN (Low Traffic Neighbourhood) boundary. For car counts, the ARIMA(1,1,2)(0,1,0)[288] model is applied. The car count data is modelled with an ARIMA model including an autoregressive term of order 1 and two moving average terms. This suggests that there might be short-term dependencies and some moving average effects in car count data. The estimated coefficients for the moving average (MA) terms are shown, including MA(1): -0.7279 and MA(2): -0.2217.The estimated variance is 15.53. The log-likelihood value is -2424.77, indicating how well the model fits the data. The information criteria are presented as AIC: 4857.55, AICc: 4857.59, and BIC: 4876.62. For cyclist counts, The ARIMA(0,1,3)(0,1,0)[288] model is utilized. The cyclist count data is modeled with an ARIMA model including three moving average terms but no autoregressive terms. This suggests that the cyclist count data might be influenced mainly by recent past values and moving average effects. The estimated coefficients for the moving average (MA) terms are shown, including MA(1): -0.1859, MA(2): -0.4945, and MA(3): -0.1957. The estimated variance (is 6.533. The log-likelihood value is -2053.66. The information criteria are presented as AIC: 4115.32, AICc: 4115.37, and BIC: 4134.39. For pedestrian counts, the ARIMA(3,1,2)(0,1,0)[288] model is applied. Similar to the pedestrian model for the Iffley Road LTN Boundary, the pedestrian count data still requires a relatively complex ARIMA model with differencing and multiple autoregressive and moving average terms. The estimated coefficients for the autoregressive (AR) and moving average (MA) terms are shown, including AR(1): 0.5992, AR(2): 0.4352, AR(3): 0.2568, MA(1): 0.0077, and MA(2): -0.9689. The estimated variance is 8.794. The log-likelihood value is -2183.16. The information criteria are presented as AIC: 4378.31, AICc: 4378.41, and BIC: 4406.92, which are used for model selection. These results provide valuable insights into the forecasting models' performance for car, cyclist, and pedestrian counts at the Morrell Avenue LTN boundary. Each ARIMA model captures the underlying patterns in the respective traffic data, and the information criteria help select the most appropriate models for future traffic predictions. These findings contribute to understanding traffic patterns in the area and can assist in making informed decisions for traffic management and planning.

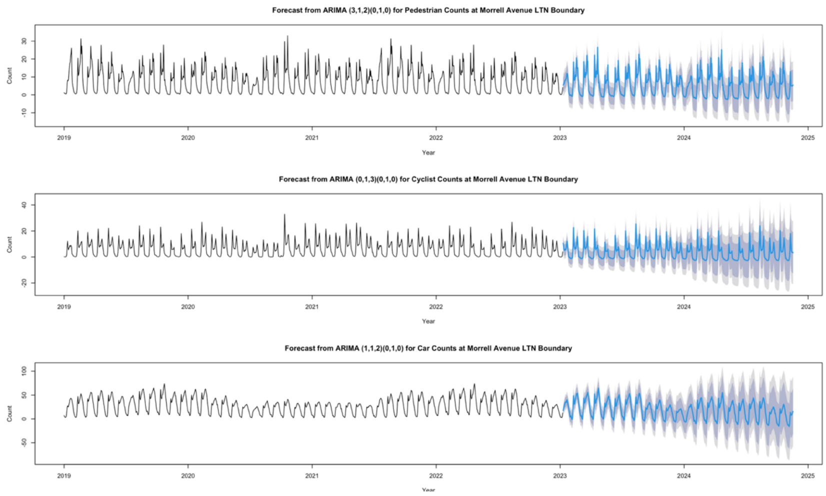

Figure 19: Forecasting after post-LTN implementation period (end 2023 – end of 2024) from ARIMA for Pedestrian, Cyclist, and Car Counts at the Morrell Avenue LTN Boundary

In *Figure 19,* we observed the observed and forecasted values for traffic counts at the Morrell Avenue LTN boundary. For the forecasted pedestrian counts, it appears these values are overall slightly lower than the observed values. In parts of the observed values, we detect much higher counts than in the forecasted values. For cyclist counts, the forecasted values appear to project an overall decrease in volume compared to the observed counts. The confidence intervals for car counts increase substantially after 2024. For car counts, the forecasted values also appear to decrease throughout the horizon. Like the confidence intervals of cyclist counts, the confidence intervals for car counts widen through the forecast horizon.

	ME	RMSE	MAE	MPE	MAPE	MASE
Pedestrian	-0.01595545	2.561938	1.273147	-0.7537186	39.85099	0.5945288
Car	-0.06425454	3.388157	2.076851	0.1361701	11.78934	0.1886435
Cyclist	0.1886435	2.210627	2.210627	42.55571	105.4998	0.4581109

Table 7: Comparison of accuracy metrics for ARIMA models for traffic counts at Morrell Avenue LTN Boundary

Table 7 is the comparison of accuracy metrics for ARIMA models for each type of traffic counts at the Morrell Avenue LTN Boundary. For pedestrians, the ME is approximately 0.016, indicating a slight underestimation of pedestrian counts on average. For cars, the ME is approximately -0.064, showing a moderate underestimation of car counts on average. For cyclists, the ME is approximately 0.189, indicating a slight overestimation of cyclist counts on average. The RMSE for the pedestrian model is 2.562, indicating that, on average, the forecasts have an error of approximately 2.562 in predicting pedestrian counts. For cars and cyclists this value is 3.388 and 2.211 respectively. For pedestrians, the MAE is about 1.273, meaning that, on average, the forecasts have an absolute error of approximately 1.273 in predicting pedestrian counts. For the car model, the MAE is approximately 2.077, indicating an absolute error of around 2.077 in predicting car counts on average. For the car model, the MAE is approximately 2.211, indicating an absolute error of around 2.211 in predicting cyclist counts on average. For pedestrians, the MPE is approximately -0.754%, suggesting a slight underestimation of pedestrian counts on average. For cars, the MPE is about 0.136%, indicating a slight overestimation of car counts on average. For cyclists, the MPE is approximately 42.556%, indicating a substantial overestimation of cyclist counts on average. For pedestrians, the MAPE is around 39.851%, indicating that, on average, the forecasts have an absolute percentage error of approximately 39.851% for pedestrian counts. For cars, the MAPE is approximately 11.789%, showing an absolute percentage error of around 11.789% for car counts. For cyclists, the MAPE is approximately 105.500%, indicating an extremely high absolute percentage error of around 105.500% for cyclist counts. Overall, the results suggest that the ARIMA model provides relatively accurate forecasts for pedestrian and car counts, with moderate errors and low MASE values. However, for cyclist counts, the model shows higher errors and a high MAPE, indicating that the forecast for cyclists may not be as accurate.

Cowley Road LTN Boundary:

Count Type	ARIMA Model	Estimated Coefficients	Estimated Standard Errors	Variance	Log-Likelihood	Information Criteria
Car	ARIMA (3,1,2) (0,1,0) [288]	AR(1): 0.9175 AR(2): -0.2523 AR(3): 0.1539 MA(1): -0.9151 MA(2): -0.0524	AR(1): 0.1637 AR(2): 0.1561 AR(3): 0.0445 MA(1): 0.1639 MA(2): 0.1558	741.6	-4110.76	AIC: 8233.52 AICc 8233.61 BIC: 8282.13
Cyclist	ARIMA (1,0,3) (0,1,0) [288] with drift	AR(1): 0.9223 MA(1): 0.1172 MA(2): -0.2966 MA(3): -0.2112 Drift: 0.0357	AR(1): 0.0242 MA(1): 0.0431 MA(2): 0.0459 MA(3): 0.0425 Drift: 0.0122	179.3	-3493.95	AIC: 6999.91 AICc: 7000.01 BIC: 7028.53
Pedestrian	ARIMA (0,1,5)	MA(1): 0.5161 MA(2): 0.3013 MA(3): 0.2617 MA(4): 0.3324 MA(5): 0.322	MA(1): 0.0281 MA(2): 0.0305 MA(3): 0.0317 MA(4): 0.0311 MA(5): 0.028	95.87	-4283.05	AIC: 8578.11 AICc: 8578.18 BIC: 8608.43

Table 8: Arima Summary Output

Table 8 is the ARIMA summary output for the Cowley Road LTN Boundary. For car counts, the car ARIMA model is (ARIMA (3,1,2) (0,1,0)). The car count data is again modelled using an ARIMA model that includes autoregressive terms up to order 3 and moving average terms up to order 2. This suggests that the car count data might exhibit short-term dependencies and some moving average effects. The estimated coefficients for the autoregressive and moving average terms provide insights into the relationships between past and current car counts. The log-likelihood value is relatively high, indicating a good fit. For cyclist counts, the cyclist ARIMA model is (ARIMA (1,0,3) (0,1,0) [288] with drift). The cyclist count data is modeled with an ARIMA model including an autoregressive term of order 1 and three moving average terms, along with a drift term. The inclusion of a drift term indicates that there might be a linear trend in the data. The estimated coefficients for the autoregressive and moving average terms, as well as the drift term, provide insights into how past values and trends affect cyclist counts. For pedestrian counts, the pedestrian ARIMA Model is (ARIMA (0,1,5)). The pedestrian count data is modelled with an ARIMA model including five moving average terms. This suggests that pedestrian count data might be influenced mainly by recent past values and moving average effects. In terms of implications on the data, the chosen ARIMA model for car data indicates short-term dependencies and some moving average effects. This suggests that car counts might be influenced by factors that have an immediate impact on the count. The inclusion of a drift term in the cyclist ARIMA model suggests the presence of a linear trend in cyclist count data. The choice of ARIMA model indicates that cyclist counts might be influenced by both short-term dependencies and a linear trend. The inclusion of several moving average terms in the pedestrian ARIMA model suggests that pedestrian counts are influenced mainly by recent past values and moving average effects. The absence of autoregressive terms might indicate that longer-range dependencies might not be as prominent.

Figure 20 displays the forecasted traffic counts for 2023 and 2024 for the Cowley Road LTN boundary. Like the 2023/2024 forecasted traffic counts for the other LTN boundaries, we observe that the forecast has captured the seasonality from the observed values, however it appears that forecasts for cyclist and car counts seem too overfit. For pedestrian counts we notice that the forecast converged to a constant and is not capturing any seasonality, but this could be due to how the observed values are very low and don't express seasonality trends. Again, we must be careful in making conclusions of these forecasts as the stationarity assumption is violated.

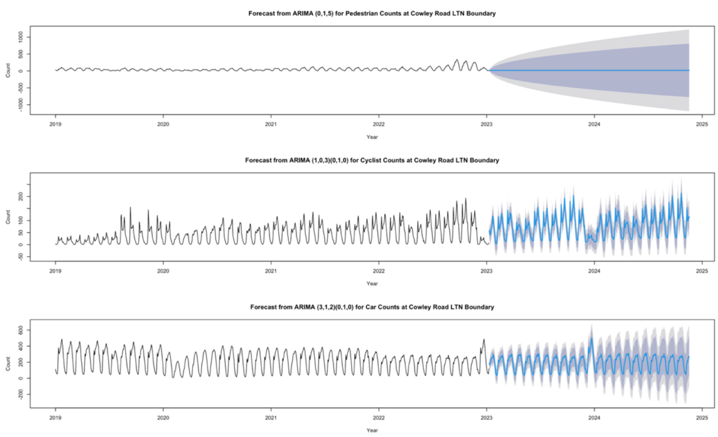

Figure 20: Forecasting after post-LTN implementation period (end 2023 – end of 2024) from ARIMA for Pedestrian, Cyclist, and Car Counts at the Morrell Avenue LTN Boundary

	ME	RMSE	MAE	MPE	MAPE	MASE
Pedestrian	-0.002066193	9.765937	6.811718	2.580067	25.46197	0.2420371
Car	0.3200012	23.52584	14.4423	1.40405	12.41542	0.2759235
Cyclist	0.01355657	11.57457	6.496579	-22.02687	36.26245	0.3315531

Table 9: Comparison of accuracy metrics for ARIMA models for traffic counts at Cowley Road LTN Boundary

Table 9 is a comparison of accuracy metrics for ARIMA models for traffic counts at Cowley Road LTN boundary. Overall, the results suggest that the ARIMA models provide reasonably accurate forecasts for pedestrian and car counts, with moderate errors and MASE values less than 1. However, for cyclist counts, the model shows higher errors and a negative MPE, indicating that the forecast for cyclists may not be as accurate, and the model tends to underestimate cyclist counts.

ST Clements LTN Boundary:

Count Type	ARIMA Model	Estimated Coefficients	Estimated Standard Errors	Variance	Log-Likelihood	Information Criteria
Pedestrain	ARIMA (3,1,2) (0,1,0) [288]	AR(1): 1.3484 AR(2): -1.1032 AR(3): 0.4627 MA(1): -0.9832 MA(2): 0.7479	AR(1): 0.0685 AR(2): 0.0671 AR(3): 0.0348 MA(1): 0.0722 MA(2): 0.0514	475.6	-3917.65	AIC: 7847.3 AICc 7847.4 BIC: 7875.91
Cyclist	ARIMA (1,1,3) (0,1,0) [288]	AR(1): 0.7293 MA(1): -0.4181 MA(2): -0.4592 MA(3): -0.0734	AR(1): 0.0406 MA(1): 0.0532 MA(2): 0.0320 MA(3): 0.0454	262.8	-3660.18	AIC: 7330.37 AICc: 7330.44 BIC: 7354.21
Car	ARIMA (1,1,0) (0,1,0) [288]	AR(1): 0.378	AR(1): 0.0313	251.9	-3642.75	AIC: 7289.5 AICc: 7289.51 BIC: 7299.04

Commented [BS6]: Can you make any general statements about the type of ARIMA models that are selected for different transport types or in/out of the LTNs? If so, what does this mean about the data?

Table 10: Arima Summary Output

Table 10 provides ARIMA summary output for the St Clements LTN Boundary. For pedestrians, the ARIMA model employed is ARIMA (3,1,2) (0,1,0). This model suggests that the pedestrian count data exhibits some degree of seasonality (since it includes differencing) and requires autoregressive terms up to order 3 and moving average terms up to order 2 to capture its dynamics. For cyclist counts, the cyclist ARIMA model is (ARIMA (1,1,3) (0,1,0)).

Like the pedestrian model, this model captures a different transport type, cyclist count data, with autoregressive terms up to order 1 and moving average terms up to order 3. For the car counts, the car ARIMA model is (ARIMA (1,1,0) (0,1,0)). The car count data is modelled with just an autoregressive term of order 1. This suggests a certain level of short-term dependency or trend in the data. The model is relatively simpler compared to the pedestrian and cyclist models, potentially indicating fewer complex dynamics in car count data. In terms of implications about the data, for the pedestrian data, the higher-order autoregressive and moving average terms in the pedestrian model might suggest more complex dynamics or longer-range dependencies in pedestrian count data. This could be due to factors like day-to-day variations, seasonality, or even external events affecting pedestrian movement. For cyclist data, the inclusion of multiple moving average terms in the cyclist model could indicate that cyclist count data is influenced by recent past values and possibly some longer-range dependencies. For car data, the car model being simpler might suggest that car counts exhibit less intricate dynamics compared to pedestrians and cyclists. Car count variations could be driven by more straightforward factors or trends.

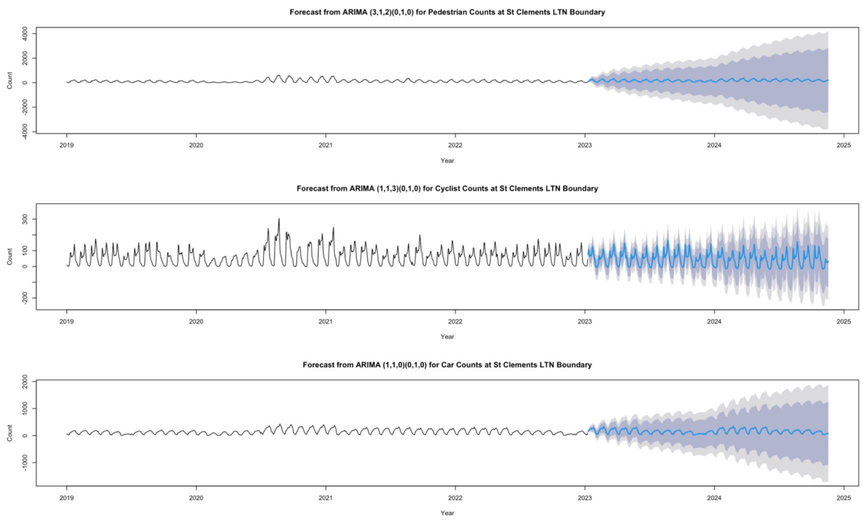

Figure 21: Forecasting after post-LTN implementation period (end 2023 – end of 2024) from ARIMA for Pedestrian, Cyclist, and Car Counts at the Morrell Avenue LTN Boundary

In *Figure 21*, the plot displays the observed and forecasted values for each type of traffic count for the St Clements LTN Boundary. For the pedestrian model, we observe that the forecasted values replicate the observed values and the confidence intervals for the forecasted values continuously increases throughout the forecast horizon. For cyclist counts, the forecast values appear to be less than the observed. After 2024, the confidence interval for the forecasted cyclist counts widens throughout the forecast horizon. For car counts, we the plot is like the pedestrian counts, and the confidence intervals are continuously widening throughout the forecast horizon.

	ME	RMSE	MAE	MPE	MAPE	MASE
Pedestrian	0.01864797	18.84097	11.91326	-0.2082206	30.61148	0.133332
Cyclist	-0.008781367	14.01265	7.754821	-3.956708	41.38517	0.2662033
Car	0.01302544	13.74254	8.807787	1.156467	11.75305	0.134768

Table 11: Comparison of accuracy metrics for ARIMA models for traffic counts at Cowley Road LTN Boundary

Table 11 provides accuracy metrics for the ARIMA models employed on the traffic counts at the Cowley Road LTN Boundary. Overall, the results suggest that the ARIMA models provide reasonably accurate forecasts for pedestrian and car counts, with moderate errors and MASE values less than 1. However, for cyclist counts, the model shows higher errors and a negative MPE, indicating that the forecast for cyclists may not be as accurate, and the model tends to underestimate cyclist counts.

5.2 Causal Impact Analysis

5.2.1 LTN Area

In this section, we examine the results from Causal Impact analysis on traffic counts in the LTN area. *Figures 26-28* display plots for Causal Impact analysis for each type of traffic count. Each plot contains three sections: original, pointwise, and cumulative. The original data refers to the actual values of the target metric over time. It is represented by the black solid line and is a time series plot that shows the fluctuations and trends of the metric before and after the intervention. This component provides a visual understanding of how the target metric behaves naturally, without any intervention or treatment. The pointwise comparisons are represented as shaded regions or confidence intervals around the predicted counterfactual data. These comparisons are made on a day-by-day basis to assess the impact of the intervention at each time point. The shaded regions indicate the uncertainty associated with the causal effect estimate at each specific time point. The cumulative comparisons are represented as cumulative sums of the causal impact over time. This component shows the overall impact of the intervention throughout the entire period of observation. It is helpful for assessing the longer-term or cumulative effect of the intervention on the target metric.

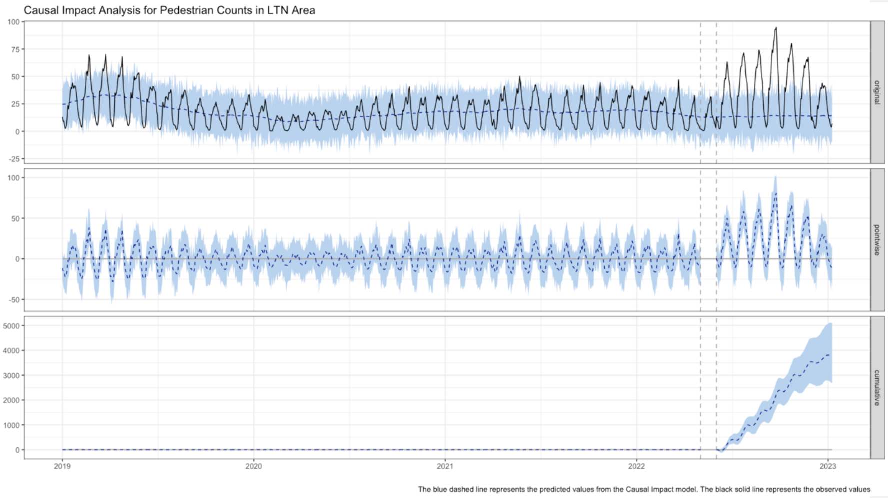

Figure 22: Causal Impact for Pedestrian Counts in LTN Area

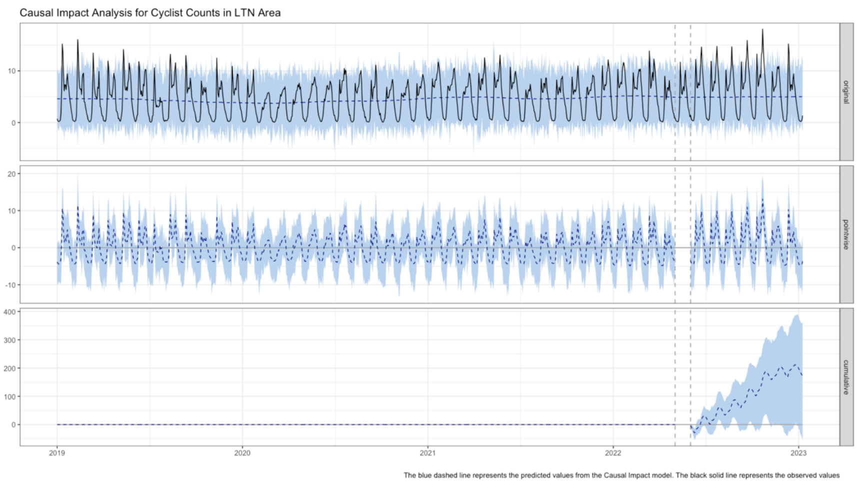

Figure 23: Causal Impact for Cyclist Counts in LTN Area.

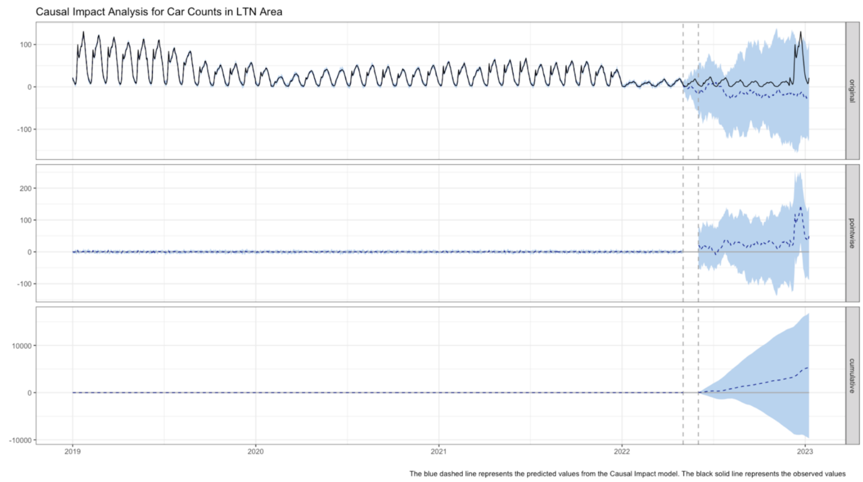

Figure 24: Causal Impact for Car Counts in LTN Area.

First, we examine results for pedestrian counts in the post intervention period. From *Figure 22*, we see that the observed values exhibit higher counts compared to the counterfactual. The response variable (pedestrian counts) had an average value of approximately 35.24. In the absence of the intervention, the expected average counts would have been 13.70. The 95% interval of this counterfactual prediction (expected response) is [6.02, 19.97]. The difference between the observed response and the counterfactual prediction yields the estimated causal effect of the intervention on the response variable. The causal effect is estimated to be 21.54, with a 95% interval of [15.27, 29.22]. This means the intervention resulted in an increase of 21.54 pedestrian counts on average during the post-intervention period. *Figure 23* is the causal impact plot for cyclist counts. Again, we observe that the observed counts exhibit relatively higher counts compared to the response. For the post-intervention period, the response variable (cyclist counts) had an average value of approximately 5.92. In the absence of the intervention, the expected average response would have been 4.94. The 95% interval of this counterfactual prediction (expected response) is [3.85, 6.23]. The difference between the observed response and the counterfactual prediction yields the estimated causal effect of the intervention on the response variable. The causal effect is estimated to be 0.98, with a 95% interval of [-0.30, 2.08]. This means the intervention resulted in an estimated increase of 0.98 cyclist counts on average during the post-intervention period. Lastly, *Figure 24* is the causal impact plot for cyclist counts. For the post-intervention period, the response variable (car counts) had an average value of approximately 16.40. In the absence of the intervention, the expected average response would have been -14.13. The 95% interval of this counterfactual prediction (expected response) is [-79.97, 71.74]. The difference between the observed response and the counterfactual prediction yields the estimated causal effect of the intervention on the response variable. The causal effect is estimated to be 30.53, with a 95% interval of [-55.34, 96.37]. This means the intervention resulted in an estimated decrease of 30.53 car counts on average during the post-intervention period.

5.2.2 LTN Boundary Roads

Iffley Road LTN Boundary:

In this section, we examine causal impact analysis on traffic counts at the various LTN boundary locations. First, we look at the Iffley Road LTN Boundary. *Figures 29-31* display causal impact plots for pedestrian, cyclist, and car counts respectively.

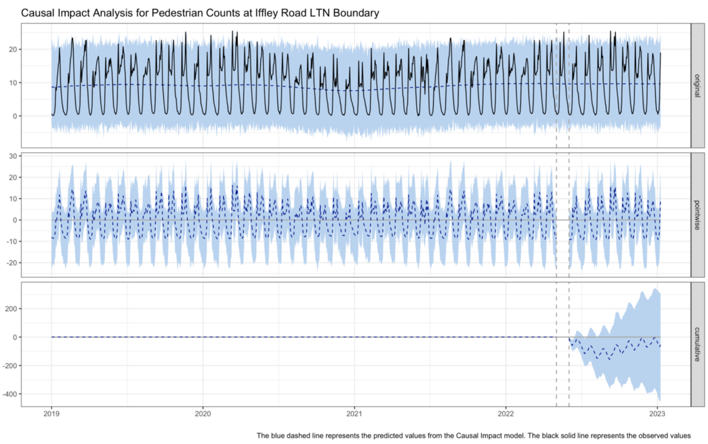

Figure 25: Causal Impact Plot for Pedestrian Counts at Iffley Road LTN Boundary

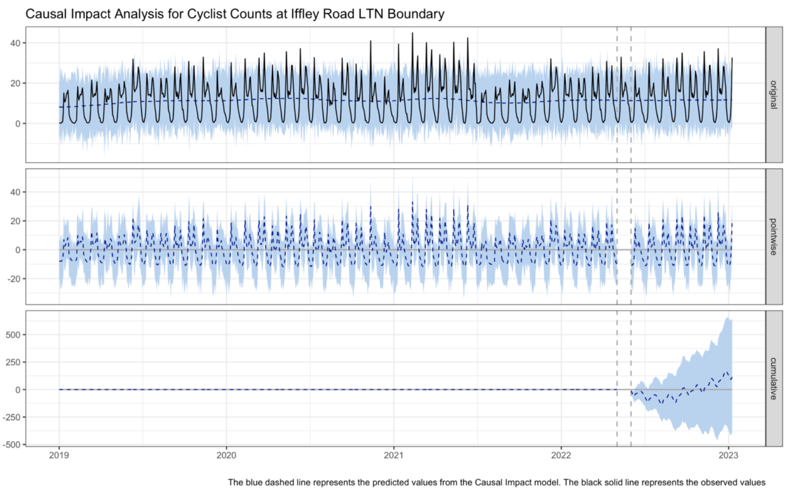

Figure 26: Causal Impact Plot for Cyclist counts at Iffley Road LTN Boundary

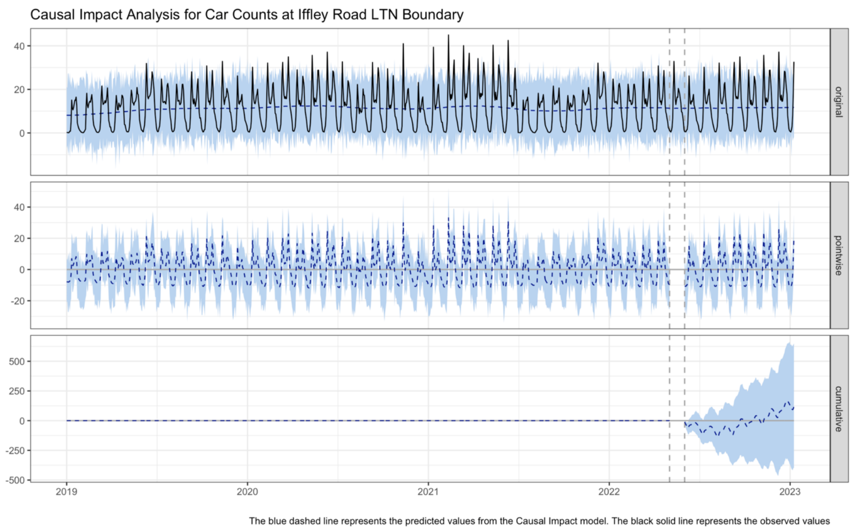

Figure 27: Causal Impact Plot for Car Counts at Iffley Road LTN Boundary

For pedestrian counts, we observe a similar pattern to the counterfactual as observed in *Figure 25*. During the post-intervention period, the response variable (pedestrian counts) had an average value of approximately 9.32. In the absence of the intervention, the expected average response would have been 9.63. The 95% interval of this counterfactual prediction (expected response) is [7.50, 11.86]. The difference between the observed response and the counterfactual prediction yields the estimated causal effect of the intervention on the response variable. The causal effect is estimated to be -0.31, with a 95% interval of [-2.54, 1.82]. This means the intervention resulted in an estimated decrease of 0.31 units on average during the post-intervention period. For cyclist counts, we observe similar plotted results to pedestrian counts from *Figure 25.* During the post-intervention period, the response variable (cyclist counts) had an average value of approximately 12.22. In the absence of the intervention, the expected average response would have been 11.53. The 95% interval of this counterfactual prediction (expected response) is [8.50, 14.48]. The difference between the observed response and the counterfactual prediction yields the estimated causal effect of the intervention on the response variable. The causal effect is estimated to be 0.69, with a 95% interval of [-2.27, 3.71]. This means the intervention resulted in an estimated increase of 0.69 cyclist counts on average during the post-intervention period. Lastly, for car counts, we observed again that the causal impact models the counterfactual to converge to an average value. During the post-intervention period, the response variable (car counts) had an average value of approximately 234.12. In the absence of the intervention, the expected average response would have been 233.70. The 95% interval of this counterfactual prediction (expected response) is [186.80, 277.69]. The difference between the observed response and the counterfactual prediction yields the estimated causal effect of the intervention on the response variable. The causal effect is estimated to be 0.42, with a 95% interval of [-43.57, 47.32]. This means the intervention resulted in an estimated increase of 0.42 car counts on average during the post-intervention period.

Morrell Avenue LTN Boundary:

Here, we examine the causal impact analysis results at the Morrell Avenue LTN Boundary. *Figures 28-30* display causal impact plots for each type of traffic count.

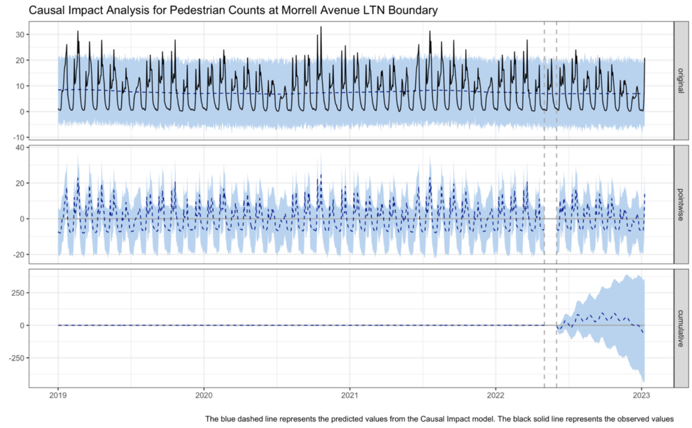

Figure 28: Causal Impact Plot for Pedestrian Counts at Morrell Avenue LTN Boundary

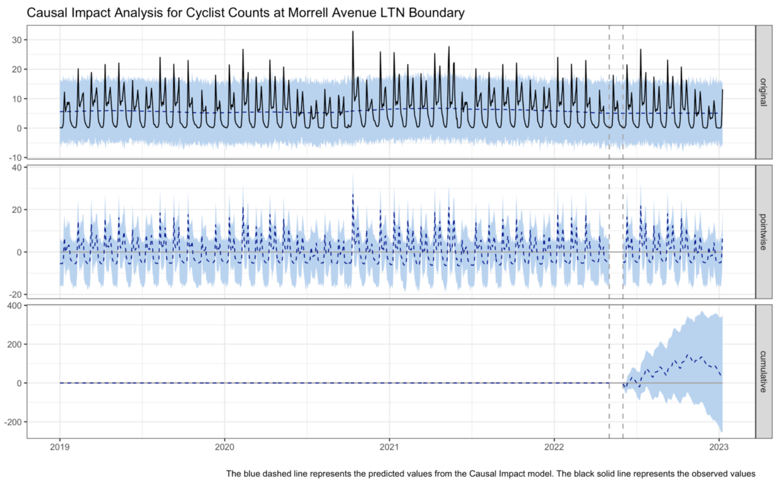

Figure 29: Causal Impact Plot for Cyclist Counts at Morrell Avenue LTN Boundary

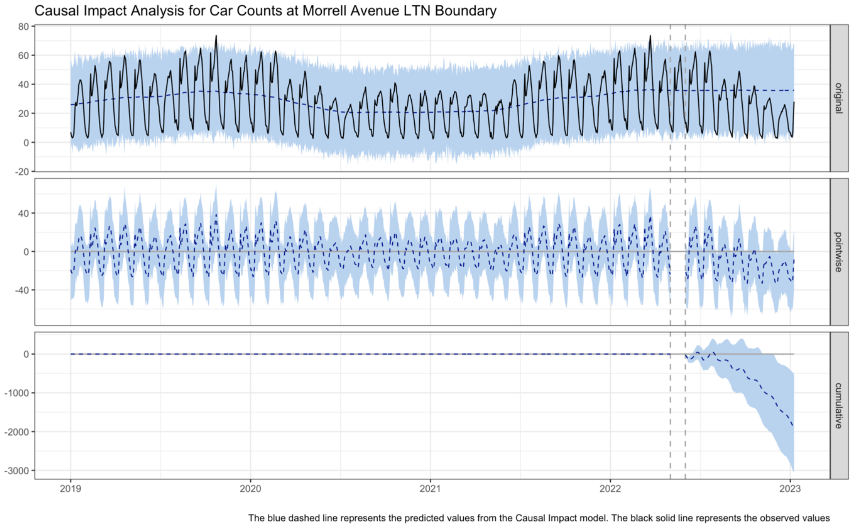

Figure 30: Causal Impact Plot for Car Counts at Morrell Avenue LTN Boundary

For pedestrian counts, we examine summary output of the causal impact model. During the post-intervention period, the response variable (pedestrian counts) had an average value of approximately 6.68. In the absence of the intervention, the expected average response would have been 6.96. The 95% interval of this counterfactual prediction (expected response) is [4.60, 9.14]. The difference between the observed response and the counterfactual prediction yields the estimated causal effect of the intervention on the response variable. The causal effect is estimated to be -0.27, with a 95% interval of [-2.46, 2.08]. This means the intervention resulted in an estimated decrease of 0.27 in pedestrian counts on average during the post-intervention period. For cyclist counts, we examine summary output and causal impact results. During the post-intervention period, the response variable (cyclist counts) had an average value of approximately 5.31. In the absence of the intervention, the expected average response would have been 5.06. The 95% interval of this counterfactual prediction (expected response) is [3.31, 6.74]. The difference between the observed response and the counterfactual prediction yields the estimated causal effect the intervention had on the response variable. The causal effect is estimated to be 0.25, with a 95% interval of [-1.43, 2.00]. This means the intervention resulted in an estimated increase of 0.25 in cyclist counts on average during the post-intervention period. Lastly for car counts, during the post-intervention period, the response variable (car counts) had an average value of approximately 25.04. In the absence of the intervention, the expected average response would have been 35.76. The 95% interval of this counterfactual prediction (expected response) is [27.97, 42.53]. The difference between the observed response and the counterfactual prediction yields the estimated causal effect the intervention had on the response variable. The causal effect is estimated to be -10.72, with a 95% interval of [-17.50, -2.93]. This means the intervention resulted in an estimated decrease of 10.72 in car counts on average during the post-intervention period.

Cowley Road LTN Boundary:

Here, we examine the causal. Impact analysis results at the Cowley Road LTN boundary. *Figures 31-33* display the causal impact plots for each type of traffic count.

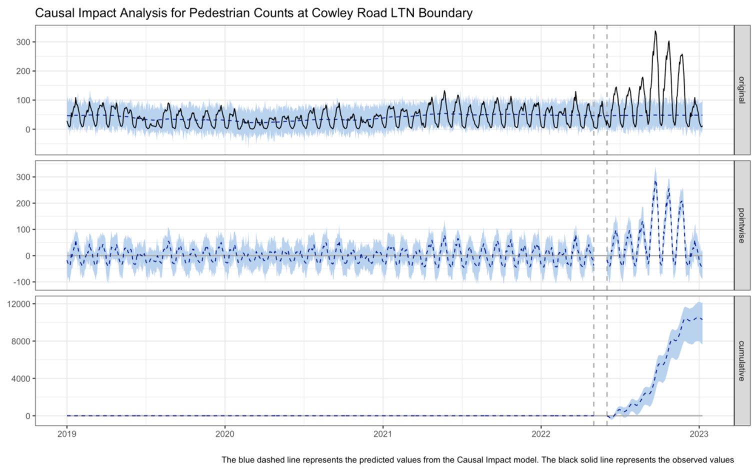

Figure 31: Causal Impact Plot for Pedestrian Counts at Cowley Road LTN Boundary

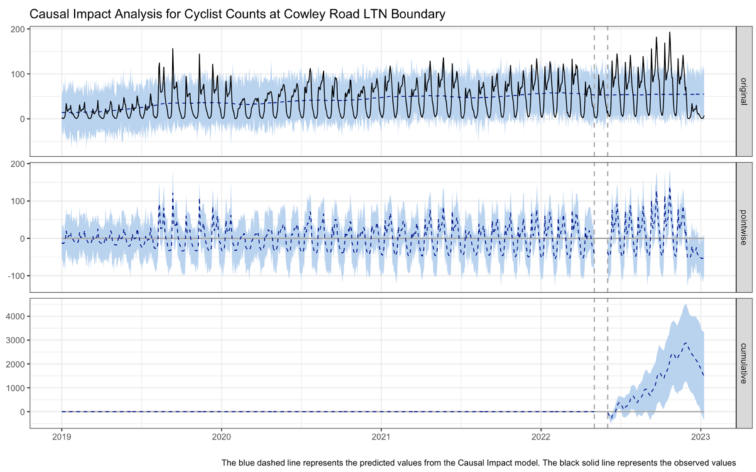

Figure 32: Causal Impact Plot for Cyclist Counts at Cowley Road LTN Boundary

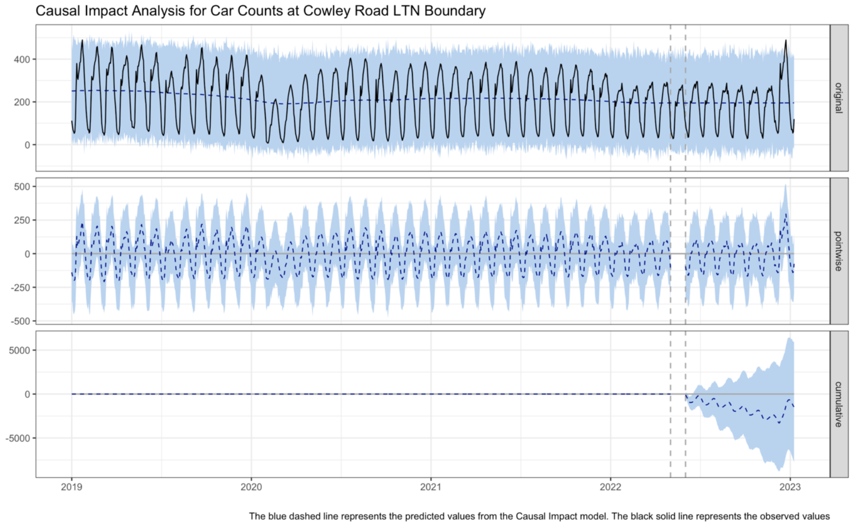

Figure 33: Causal Impact Plot for Car Counts at Cowley Road LTN Boundary

For pedestrian counts, we examine summary output of the causal impact model. During the post-intervention period, the response variable (pedestrian counts) had an average value of approximately 106.56. In the absence of the intervention, the expected response would have been 47.70. The 95% interval of this counterfactual prediction is [37.26, 62.77]. Subtracting this prediction from the observed response yields an estimate of the causal effect the intervention had on the response variable. This effect is 58.86 with a 95% interval of [43.79, 69.30]. This means the intervention resulted in an estimated decrease of 0.31 pedestrian counts on average during the post-intervention period. For cyclist counts, we also examine summary output. During the post-intervention period, the response variable (cyclist counts) had an average value of approximately 62.36. In the absence of the intervention, the expected average response would have been 53.90. The 95% interval of this counterfactual prediction (expected response) is [43.38, 64.39]. The difference between the observed response and the counterfactual prediction yields the estimated causal effect the intervention had on the response variable. The causal effect is estimated to be 8.46, with a 95% interval of [-2.03, 18.98]. This means the intervention resulted in an estimated increase of 8.46 in cyclist counts on average during the post-intervention period. For car counts, we also examine summary output on the causal impact model. During the post-intervention period, the response variable (car counts) had an average value of approximately 186.86. In the absence of the intervention, the expected average response would have been 195.15. The 95% interval of this counterfactual prediction (expected response) is [153.17, 231.21]. The difference between the observed response and the counterfactual prediction yields the estimated causal effect the intervention had on the response variable. The causal effect is estimated to be -8.29, with a 95% interval of [-44.35, 33.69]. This means the intervention resulted in an estimated decrease of 8.29 in car counts on average during the post-intervention period.

St Clements LTN Boundary:

Here, we examine the causal. Impact analysis results at the St Clements LTN boundary. *Figures 34-36* display the causal impact plots for each type of traffic count.

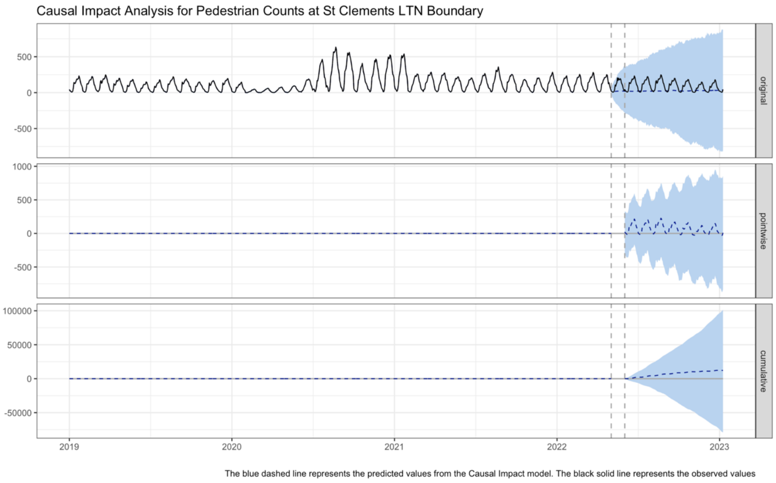

Figure 34: Causal impact Plot for Pedestrian Counts at St Clements LTN Boundary

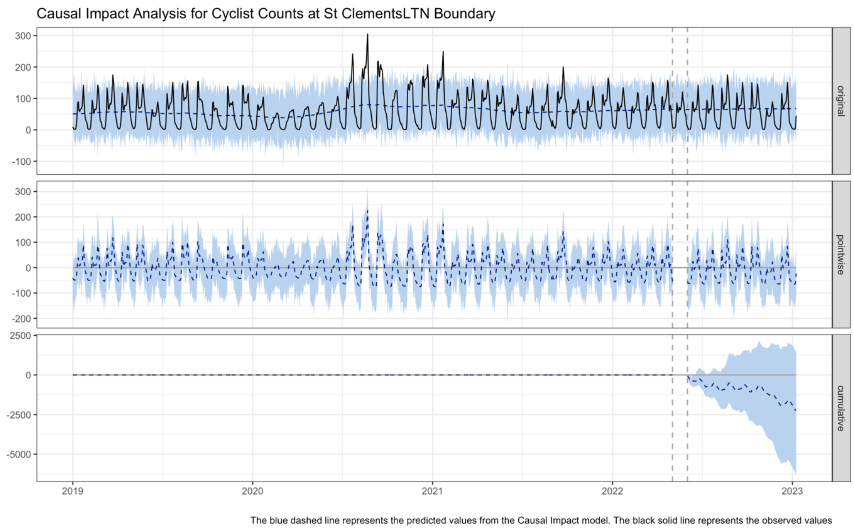

Figure 35: Causal Impact plot for Cyclist Counts at St Clements LTN Boundary

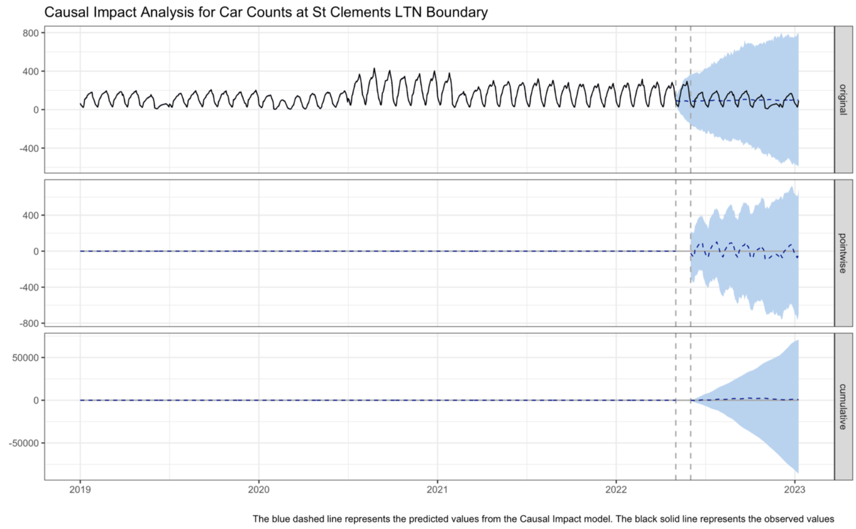

Figure 36: Causal Impact plot for Car Counts at St Clements LTN Boundary

For pedestrian counts, during the post-intervention period, the observed average pedestrian count was approximately 94.97. In the absence of the intervention, the counterfactual prediction suggests an average response of 25.36. The 95% interval of this counterfactual prediction is [-484.23, 546.05]. The causal effect is 69.69 with a confidence interval [-451.08, 579.20]. For cyclist counts, during the post-intervention period, the observed average cyclist count was approximately 53.37. In the absence of the intervention, the counterfactual prediction suggests an average response of 66.24. The 95% interval of this counterfactual prediction is [44.65, 89.51]. The causal effect of the intervention on cyclist counts is -12.87 with a 95% interval of [-36.14, 8.72]. Lastly for car counts, during the intervention period, the observed average car count was approximately 101.76. In the absence of the intervention, the counterfactual prediction suggests an average response of 97.13. The 95% interval of this counterfactual prediction is [-303.59, 592.95]. The causal effect of the intervention on car counts is 4.63 with a 95% interval of [-491.19, 405.35].

Chapter 6: Discussion

6.1 Conclusion

In conclusion, the exploratory data analysis indicates that the implementation of Low Traffic Neighbourhoods (LTNs) has led to a decrease in car volume within LTN areas. However, at the boundary roads, the impact on car volume appears mixed, with some areas showing an increase in car volume post-COVID. On the other hand, the implementation of LTNs has resulted in an increase in cyclist and pedestrian volumes within LTN areas, contributing to safer and more active travel options.

Regarding time series forecasting using ARIMA models, the results demonstrate relatively accurate forecasts for pedestrian and car counts at LTN boundaries, with moderate errors and low Mean Absolute Scaled Error (MASE) values. However, the forecasts for cyclist counts show higher errors and a negative Mean Percentage Error (MPE), suggesting some limitations in predicting cyclist counts accurately.

The causal impact analysis reveals that the implementation of LTNs has positively influenced pedestrian and cyclist volumes within LTN areas, resulting in increased counts during the post-intervention period. The effect on car counts at LTN boundaries show mixed results, with some areas experiencing a decrease, while others show an increase in car volumes during the post-intervention period.

Considering the overall findings, we expect LTNs to have a positive impact in the long run, leading to reduced car traffic within LTN areas and increased pedestrian and cyclist volumes, promoting safer and more sustainable modes of transportation. However, the effectiveness of LTNs may vary at different boundary roads, and continuous monitoring and evaluation are crucial to ensure their success in managing traffic and promoting active travel options.

6.2 Limitations and areas of improvement

Limitations:

1. Assumptions Violated: The application of ARIMA models assumes stationarity, linearity, and independence of the time series data. In real-life scenarios, these assumptions may be violated as traffic data can be influenced by external factors, trends, and seasonality, leading to non-stationary and nonlinear patterns.
2. Real-Life Data Complexity: Real-life traffic data may not comply with the data generating processes assumed by forecasting models implemented in forecasting support systems. This discrepancy can result in suboptimal model performance and forecasts that may not fully capture the intricacies of the data.
3. Univariate Time Series: The use of univariate time series analysis limits the incorporation of additional explanatory variables (covariates) that may influence traffic patterns. The absence of relevant covariates may lead to less comprehensive insights into the factors impacting traffic counts.
4. Comparing Information Criteria / Log – likelihoods: All ARIMA models in this research are fit to different data which means comparing information criteria and log – likelihood may not be appropriate. Information Criteria tends to favour more complex models when the sample size is large. In such cases, it's important to use the criteria in conjunction with other methods, like cross-validation, to avoid overfitting. In terms of comparing log-likelihoods fitted to different data, this can be done, but it's important to ensure that the data being compared are similar in terms of nature, distribution, and context. For instance, we might compare log-likelihoods for different subgroups of data, or log-likelihoods from different time periods if you're dealing with time series data. The goal is to understand how well a model fits the specific data it was designed for. In summary, comparing information criteria and log-likelihoods across different models or datasets can be

informative, but it requires careful consideration of model assumptions, sample size, interpretability, and purpose of comparison.

Areas of Improvement

1. Forecasting Accuracy: Enhancing forecasting accuracy is crucial, as even small improvements in forecast error can translate to significant gains in the utility of the forecasts. To achieve this, more sophisticated and advanced forecasting techniques could be explored, such as machine learning algorithms (e.g., LSTM, XGBoost) that can handle large datasets and capture complex relationships between variables.
2. Covariate Analysis: Incorporating covariates, such as weather conditions, public events, or roadwork schedules, into the forecasting models can provide a more holistic understanding of the factors influencing traffic counts. Including relevant covariates can improve the accuracy of the forecasts and allow for better policy recommendations and traffic management strategies.
5. Handling Large Datasets: The use of large datasets may pose challenges in terms of computational resources and processing time. Employing distributed computing and parallel processing techniques can help handle large datasets and enable the application of more complex forecasting methods.
6. Model Selection: Evaluating a broader range of forecasting models beyond ARIMA, such as seasonal decomposition of time series (STL) or state-space models, can help identify the most suitable approach for the specific traffic data and address the limitations associated with the ARIMA modelling.
7. Continuous Monitoring and Model Updating: Real-life traffic patterns may change over time due to various factors, including infrastructure developments or changes in transportation policies. Regularly updating the forecasting models and monitoring their performance can ensure that the forecasts remain relevant and accurate.

In conclusion, acknowledging and addressing the limitations related to assumptions, real-life data complexity, and univariate time series analysis can lead to improved forecasting accuracy and a better understanding of the underlying factors influencing traffic patterns. By exploring more advanced techniques, incorporating relevant covariates, and updating models as needed, the utility of traffic forecasts can be significantly enhanced, supporting informed.

References

Active Travel Strategy - Oxfordshire County Council, www.oxfordshire.gov.uk/sites/default/files/file/roads-and-transport-policies-and-plans/ActiveTravelStrategy.pdf. Accessed 8 Aug. 2023.

Aldred, Rachel, and Anna Goodman. *Low Traffic Neighbourhoods, Car Use, and Active Travel: Evidence from the People and Places Survey of Outer London Active Travel Interventions*, 2020, https://doi.org/10.31235/osf.io/ebj89.

Brodersen, Kay H., et al. "Inferring Causal Impact Using Bayesian Structural Time-Series Models." *The Annals of Applied Statistics*, vol. 9, no. 1, 2015, https://doi.org/10.1214/14-aoas788.

Benvenuto, Domenico, et al. "Application of the Arima Model on the COVID-2019 Epidemic Dataset." *Data in Brief*, vol. 29, 2020, p. 105340, https://doi.org/10.1016/j.dib.2020.105340.

"The Case for Cycling: Health." *The Case for Cycling: Health | Cycling UK*, www.cyclinguk.org/briefing/case-cycling-health. Accessed 8 Aug. 2023.

Innovation Hub (IHUB) Officers, Oxfordshire County Council, et al. "East Oxford LTN Evaluation Snapshot Report." *Oxfordshire County Council*, June 2023, www.oxfordshire.gov.uk/sites/default/files/file/roads-and-transport-major-projects/EastOxfordLTNsmonitoringandevaluationsnapshotreportJune2023.pdf.

Oxfordshire County Council, 01865 792422. "About Our Low Traffic Neighbourhoods (Ltns)." *Oxfordshire County Council*, www.oxfordshire.gov.uk/residents/roads-and-transport/connecting-oxfordshire/low-traffic-neighbourhoods/about-our-ltns. Accessed 8 Aug. 2023.

Panter, Jenna, et al. "Impact of New Transport Infrastructure on Walking, Cycling, and Physical Activity." *American Journal of Preventive Medicine*, vol. 50, no. 2, 2016, https://doi.org/10.1016/j.amepre.2015.09.021.

Saelens, Brian E., et al. "Environmental Correlates of Walking and Cycling: Findings from the Transportation, Urban Design, and Planning Literatures." *Annals of Behavioral Medicine*, vol. 25, no. 2, 2003, pp. 80–91, https://doi.org/10.1207/s15324796abm2502_03.

Warburton, Darren E., and Shannon S. Bredin. "Health Benefits of Physical Activity: A Strengths-Based Approach." *Journal of Clinical Medicine*, vol. 8, no. 12, 2019, p. 2044, https://doi.org/10.3390/jcm8122044.

Yang, Xiuleng, et al. "Evaluation of Low Traffic Neighbourhood (LTN) Impacts on No2 and Traffic." *Transportation Research Part D: Transport and Environment*, vol. 113, 2022, p. 103536, https://doi.org/10.1016/j.trd.2022.103536.

Contreras, J., et al. "Arima models to predict next-day electricity prices." *IEEE Transactions on Power Systems*, vol. 18, no. 3, 2003, pp. 1014–1020, https://doi.org/10.1109/tpwrs.2002.804943.

"Forecasting Egyptian GDP Using ARIMA Models." *Mohamed Reda Abonazel, Ahmed Ibrahim Abd-Elftah, Forecasting Egyptian GDP Using Arima Models*, doi.org/10.12988/ref.2019.81023. Accessed 18 Aug. 2023.

Appendix

vivacityimport.R

```r
#install packages
library(tidyverse)
library(janitor)
library(stringr)
library(lubridate)

## Lood Data
viva1923 <- read.csv("/Users/seandube/Desktop/TRL Disseration HUB/Data/Traffic Related/Viva CIty/viva1923.csv")

## WRANGLE
newViva1923 <- viva1923 %>%
  clean_names() %>%
  na.omit()

specific_char <- ","
char_position <- regexpr(specific_char, newViva1923$location)

finalViva <- newViva1923 %>%
  mutate(date = as_date(substring(time_from, 1, 10))) %>%
  mutate(time_from = substring(time_from, 12, 16)) %>%
  mutate(time_to = substring(time_to, 12, 16)) %>%
  mutate(location = str_extract(location, paste0("(?<=\\", "[", ").*?(?=\\", "]", ")"))) %>%
  mutate(location = gsub("\\[", "", location)) %>%
  mutate(latitude = sub(".*,", "", location)) %>%
  mutate(longitude = sub(",.*", "", location)) %>%
  select(date, time_from, time_to, countline_name, latitude, longitude, cyclist, motorbike, car, pedestrian, taxi, van, minibus, bus)
## save
save(finalViva, file = "/Users/seandube/Desktop/TRL Disseration HUB/Data/TRL R DATA FILES/finalViva.RData")

##------------------------------------------------------------
# filter for Viva City sensors locations for east oxford in LTN area and boundarys
##--------------------------------------------------------------

## MORRELL AND DIVINITY BOUNDARY ROADS

morrell1923 <- finalViva[str_detect(finalViva$countline_name, "morrell"),]
warneford1923 <- finalViva[str_detect(finalViva$countline_name, "warneford"),]

divinityBoundary1923 <- rbind(morrell1923, warneford1923) %>%
  mutate(LTNregion = "Morrell")

## ST CLEMENTS BOUNDARY ROAD

stclements1923 <- finalViva[str_detect(finalViva$countline_name, "Cowley_road_north"), ] %>%
  mutate(LTNregion = "St Clements")

## LEOPOLD X ST CLEMENTS

leopold1923 <- finalViva[str_detect(finalViva$countline_name, "OX44_Cowley_Rd_East"), ] %>%
  mutate(LTNregion = "Cowley Road East")

# IFFLEY BOUNDARY ROAD

iffley1923 <- finalViva[str_detect(finalViva$countline_name, "Iffley_Rd_North"), ] %>%
  mutate(LTNregion = "Iffley ")
```

```r
### CREATE BOUNDARY LTN DF AND SAVE

boundaryLTNs <- rbind(divinityBoundary1923, stclements1923, leopold1923, iffley1923)
save(boundaryLTNs, file = "/Users/seandube/Desktop/TRL Disseration HUB/Data/TRL R DATA FILES/boundaryLTNs.RData")

# SENSOR AT DIVINITY ROAD IN THE LTN AREA

LeopoldXDivinity <- finalViva[str_detect(finalViva$countline_name, "Leopold"), ]
divinity <- finalViva[str_detect(finalViva$countline_name, "Divinity"), ]

ltnArea <- rbind(LeopoldXDivinity, divinity)
save(ltnArea, file = "/Users/seandube/Desktop/TRL Disseration HUB/Data/TRL R DATA FILES/ltnArea.RData")

#finalViva <- load("/Users/seandube/Desktop/TRL Disseration HUB/Data/TRL R DATA FILES/finalViva.RData")
#boundaryLTNs <- load("/Users/seandube/Desktop/TRL Disseration HUB/Data/TRL R DATA FILES/boundaryLTNs.RData")
#ltnArea <- load("/Users/seandube/Desktop/TRL Disseration HUB/Data/TRL R DATA FILES/ltnArea.RData")
```

exploratoryVIVA.R

```r
## LOAD PACKAGES
library(tidyverse)
library(ggplot2)
library(changepoint)
library(hms)

### ------------------------------------------------------------------------
# Comparison of Hourly Pedestrian / Cyclist / Car Counts at the different LTN locations in East Oxford
### ------------------------------------------------------------------------

# Load Data

#boundaryLTNs <- load("/Users/seandube/Desktop/TRL Disseration HUB/Data/TRL R DATA FILES/boundaryLTNs.RData")
#ltnArea <- load("/Users/seandube/Desktop/TRL Disseration HUB/Data/TRL R DATA FILES/ltnArea.RData")

## prepare data
newBoundary <- boundaryLTNs %>%
  mutate(year = format(date, "%Y"),
         month = format(date, "%m"),
         hour = factor(substr(time_from, 1, 2), levels = sprintf("%02d", 0:23))) %>%
  mutate(month = as.numeric(month)) %>%
  select(date, year, month, hour, cyclist, pedestrian, cyclist, LTNregion)

hourlyLTNboundary <- boundaryLTNs %>%
  mutate(year = format(date, "%Y"),
         month = format(date, "%m"),
         hour = factor(substr(time_from, 1, 2), levels = sprintf("%02d", 0:23))) %>%
  mutate(month = as.numeric(month)) %>%
  group_by(year, month, hour, LTNregion) %>%
  summarise(avg_pedestrian_count = mean(pedestrian),
            avg_cyclist_count = mean(cyclist),
            avg_car_count = mean(car))

#plot pedestrians

hourlyLTNboundarypedestrian <- ggplot(hourlyLTNboundary, aes(x = hour, y = avg_pedestrian_count, color = year)) +
  geom_smooth(aes(group = year)) +
  facet_wrap(~ LTNregion, nrow = 2, scales = "free_y") +
  labs(x = "Hour", y = "Average Pedestrian Count", title = "Comparison of Hourly Pedestrian Counts by Region",
       subtitle = "This plot represents the total average number of pedestrian counts per hour by Region for all Viva City sensors at LTN boundary roads",
       caption = "Data Range: Jan 1, 2019 = July 14, 2023")  +
  scale_color_manual(values = c("blue", "red", "green", "purple",  "orange"))

# plot cyclists

hourlyLTNboundarycyclist <- ggplot(hourlyLTNboundary, aes(x = hour, y = avg_cyclist_count, color = year)) +
  geom_smooth(aes(group = year)) +
  facet_wrap(~ LTNregion, nrow = 2, scales = "free_y") +
  labs(x = "Hour", y = "Average Cyclist Count", title = "Comparison of Hourly Cyclist Counts by Region",
       subtitle = "This plot represents the total average number of cyclist counts per hour by Region for all Viva City sensors at LTN boundary roads",
       caption = "Data Range: Jan 1, 2019 = July 14, 2023")  +
  scale_color_manual(values = c("blue", "red", "green", "purple",  "orange"))
```

```r
# plot cars

hourlyLTNboundarycar<- ggplot(hourlyLTNboundary, aes(x = hour, y = avg_car_count, color = year)) +
  geom_smooth(aes(group = year)) +
  facet_wrap(~ LTNregion, nrow = 2, scales = "free_y") +
  labs(x = "Hour", y = "Average CarCount", title = "Comparison of Hourly Car Counts by Region",
       subtitle = "This plot represents the total average number of carcounts per hour by Region for all Viva City sensors at LTN boundary roads",
       caption = "Data Range: Jan 1, 2019 = July 14, 2023")  +
  scale_color_manual(values = c("blue", "red", "green", "purple",  "orange"))

### ------------------------------------------------------------------------
    # Comparison of Monthly Pedestrian / Cyclist / Car Counts at the different LTN locations in East Oxford
### ------------------------------------------------------------------------

# plot pedestrians

monthlyLTNboundarypedestrian <- ggplot(hourlyLTNboundary, aes(factor(month), y = avg_pedestrian_count, color = year)) +
  geom_smooth(aes(group = year)) +
  facet_wrap(~ LTNregion, nrow = 2, scales = "free_y") +
  labs(x = "Month", y = "Average Pedestrian Count", title = "Comparison of Monthly Pedestrian Counts by Region",
       subtitle = "This plot represents the total average number of pedestrian counts per month by Region for all Viva City sensors at LTN boundary roads",
       caption = "Data Range: Jan 1, 2019 = July 14, 2023")  +
  scale_color_manual(values = c("blue", "red", "green", "purple",  "orange"))

# plot cyclists

monthlyLTNboundarycyclist <- ggplot(hourlyLTNboundary, aes(factor(month), y = avg_cyclist_count, color = year)) +
  geom_smooth(aes(group = year)) +
  facet_wrap(~ LTNregion, nrow = 2, scales = "free_y") +
  labs(x = "Month", y = "Average Cyclist Count",
       title = "Comparison of Monthly Cyclist Counts by Region",
       subtitle = "This plot represents the total average number of cyclist counts per month by Region for all Viva City sensors at LTN boundary roads",
       caption = "Data Range: Jan 1, 2019 = July 14, 2023")  +
  scale_color_manual(values = c("blue", "red", "green", "purple",  "orange"))

# plot cars

monthlyLTNboundarycar <- ggplot(hourlyLTNboundary, aes(factor(month), y = avg_car_count, color = year)) +
  geom_smooth(aes(group = year)) +
  facet_wrap(~ LTNregion, nrow = 2, scales = "free_y") +
  labs(x = "Month", y = "Average Car Count", title = "Comparison of Monthly Car Counts by Region",
       subtitle = "This plot represents the total average number of car counts per month by Region for all Viva City sensors at LTN boundary roads",
       caption = "Data Range: Jan 1, 2019 = July 14, 2023")  +
  scale_color_manual(values = c("blue", "red", "green", "purple",  "orange"))
```

```r
### -----------------------------------------------------------------------
# Comparison of Hourly Pedestrian / Cyclist / Car Counts at LTN Areas in East Oxford
### -----------------------------------------------------------------------

# filter for LTN Area

hourlyLTNArea <- ltnArea %>%
  mutate(year = format(date, "%Y"),
         month = format(date, "%m"),
         hour = factor(substr(time_from, 1, 2), levels = sprintf("%02d", 0:23))) %>%
  mutate(month = as.numeric(month)) %>%
  group_by(year, month, hour) %>%
  summarise(avg_pedestrian_count = mean(pedestrian),
            avg_cyclist_count = mean(cyclist),
            avg_car_count = mean(car))

# plot pedestrian

hourlyLTNpedestrian <- ggplot(hourlyLTNArea, aes(x = hour, y = avg_pedestrian_count, color = year)) +
  geom_smooth(aes(group = year)) +
  labs(x = "Hour", y = "Average Pedestrian Count", title = "Comparison of Hourly Pedestrian Counts",
       subtitle = "This plot represents the total average number of pedestrian counts per hour for all Viva City sensors in LTN Areas",
       caption = "Data Range: Jan 1, 2019 = July 14, 2023")  +
  scale_color_manual(values = c("blue", "red", "green", "purple",  "orange"))

# plot cyclist

hourlyLTNcyclist <- ggplot(hourlyLTNArea, aes(x = hour, y = avg_cyclist_count, color = year)) +
  geom_smooth(aes(group = year)) +
  labs(x = "Hour", y = "Average Cyclist Count", title = "Comparison of Hourly Cyclist Counts",
       subtitle = "This plot represents the total average number of cyclist counts per hour for all Viva City sensors in LTN Areas",
       caption = "Data Range: Jan 1, 2019 = July 14, 2023")  +
  scale_color_manual(values = c("blue", "red", "green", "purple",  "orange"))

# plot car

hourlyLTNcar <- ggplot(hourlyLTNArea, aes(x = hour, y = avg_car_count, color = year)) +
  geom_smooth(aes(group = year)) +
  labs(x = "Hour", y = "Average Car Count", title = "Comparison of Hourly Car Counts",
       subtitle = "This plot represents the total average number of car counts per hour for all Viva City sensors in LTN Areas",
       caption = "Data Range: Jan 1, 2019 = July 14, 2023")  +
  scale_color_manual(values = c("blue", "red", "green", "purple",  "orange"))

### -----------------------------------------------------------------------
# Comparison of Monthly Pedestrian / Cyclist / Car Counts at LTN Areas in East Oxford
### -----------------------------------------------------------------------
```

```r
### -----------------------------------------------------------------------
# Comparison of Monthly Pedestrian / Cyclist / Car Counts at LTN Areas in East Oxford
### -----------------------------------------------------------------------

monthlyLTNpedestrian <- ggplot(hourlyLTNArea, aes(factor(month), y = avg_pedestrian_count, color = year)) +
  geom_smooth(aes(group = year)) +
  labs(x = "Month", y = "Average Pedestrian Count", title = "Comparison of Monthly Pedestrian Counts",
       subtitle = "This plot represents the total average number of pedestrian counts per month for all Viva City sensors in LTN Areas",
       caption = "Data Range: Jan 1, 2019 = July 14, 2023")  +
  scale_color_manual(values = c("blue", "red", "green", "purple",  "orange"))

monthlyLTNcyclist <- ggplot(hourlyLTNArea, aes(factor(month), y = avg_cyclist_count, color = year)) +
  geom_smooth(aes(group = year)) +
  labs(x = "Month", y = "Average Cyclist Count", title = "Comparison of Monthly Cyclist Counts",
       subtitle = "This plot represents the total average number of cyclist counts per month for all Viva City sensors in LTN Areas",
       caption = "Data Range: Jan 1, 2019 = July 14, 2023")  +
  scale_color_manual(values = c("blue", "red", "green", "purple",  "orange"))

monthlyLTNcar <- ggplot(hourlyLTNArea, aes(factor(month), y = avg_car_count, color = year)) +
  geom_smooth(aes(group = year)) +
  labs(x = "Month", y = "Average Car Count", title = "Comparison of Monthly Car Counts",
       subtitle = "This plot represents the total average number of car counts per month for all Viva City sensors in LTN Areas",
       caption = "Data Range: Jan 1, 2019 = July 14, 2023")  +
  scale_color_manual(values = c("blue", "red", "green", "purple",  "orange"))
```

forecastingLTNArea.R

```r
library(tidyverse)
library(forecast)
library(ggplot2)
library(rio)
library(tseries)
library(prophet)

# create ts objects for each type of count to predict what will happen in the rest of 2023/2024
head(hourlyLTNArea)
tsCarArea <- ts(hourlyLTNArea$avg_car_count, frequency = 24*12, start = c(2019, 1), end = c(2023,7))
tsPedArea <- ts(hourlyLTNArea$avg_pedestrian_count, frequency = 24*12, start = c(2019, 1), end = c(2023,7))
tsCycArea <- ts(hourlyLTNArea$avg_cyclist_count, frequency = 24*12, start = c(2019, 1), end = c(2023,7))

# create ts objects for each type of count to predict post LTN from pre LTN

preCarArea <- ts(hourlyLTNArea$avg_car_count, frequency = 24*12, start = c(2021, 11), end = c(2022,5))
prePedArea <- ts(hourlyLTNArea$avg_pedestrian_count, frequency = 24*12, start = c(2021, 11), end = c(2022,5))
preCycArea <- ts(hourlyLTNArea$avg_cyclist_count, frequency = 24*12, start = c(2021, 11), end = c(2022,5))

#create arima models
arimaCarArea <- auto.arima(tsCarArea, stepwise = FALSE, parallel = TRUE)
arimaPedArea <- auto.arima(tsPedArea, stepwise = FALSE, parallel = TRUE)
arimaCycArea <- auto.arima(tsCycArea, stepwise = FALSE, parallel = TRUE)

preArimaCar <- auto.arima(preCarArea, stepwise = FALSE, parallel = TRUE)
preArimaPed <- auto.arima(prePedArea, stepwise = FALSE, parallel = TRUE)
preArimaCyc <- auto.arima(preCycArea, stepwise = FALSE, parallel = TRUE)

#forecast

forecastCarArea <- forecast(arimaCarArea, h = 535)
forecastPedArea <- forecast(arimaPedArea, h = 535)
forecastCycArea <- forecast(arimaCycArea, h = 535)

postForecastCar <- forecast(preArimaCar, h = 335)
postForecastPed <- forecast(preArimaPed, h = 335)
postForecastCyc <- forecast(preArimaCyc, h = 335)

#residuals

carResid <- residuals(arimaCarArea)
cycResid <- residuals(arimaCycArea)
pedResid <- residuals(arimaPedArea)

## ACF

par(mfrow = c(3,1))
acf(carResid, main = "LTN Area Car ACF")
acf(cycResid, main = "LTN Area Cyclists ACF")
acf(pedResid, main = "LTN Area Pedestrians ACF")
```

```r
## ACF

par(mfrow = c(3,1))
acf(carResid, main = "LTN Area Car ACF")
acf(cycResid, main = "LTN Area Cyclists ACF")
acf(pedResid, main = "LTN Area Pedestrians ACF")

#plot 2023/2024 forecast

par(mfrow = c(3,1))
plot(forecastCarArea, main = "Forecast from ARIMA (1,0,1)(0,1,0) for Car Counts in LTN Area", ylab = "Count", xlab = "Year")
plot(forecastCycArea, main = "Forecast from ARIMA (1,0,3)(0,1,0) for Cyclist Counts in LTN Area", ylab = "Count", xlab = "Year")
plot(forecastPedArea, main = "Forecast from ARIMA (1,0,0)(0,1,0) for Pedestrian Counts in LTN Area", ylab = "Count", xlab = "Year")

par(mfrow = c(3,1))
plot(postForecastCar, main = "Forecast from ARIMA (1,0,1)(0,1,0) for Car Counts in LTN Area", ylab = "Count", xlab = "Year")
plot(postForecastPed, main = "Forecast from ARIMA (1,0,3)(0,1,0) for Cyclist Counts in LTN Area", ylab = "Count", xlab = "Year")
plot(postForecastCar, main = "Forecast from ARIMA (1,0,0)(0,1,0) for Pedestrian Counts in LTN Area", ylab = "Count", xlab = "Year")


accuracy(forecastCarArea)
accuracy(forecastCycArea)
accuracy(forecastPedArea)

### CROSS VALIDATOIN

# Perform cross-validation for ARIMA models
num_folds <- 5  # Specify the number of folds for cross-validation

# Cross-validation for Car Area
cv_results_car <- tsCV(tsCarArea, forecastfunction = forecast, h = 336, method = "arima", cv = num_folds)

# Cross-validation for Pedestrian Area
cv_results_ped <- tsCV(tsPedArea, forecastfunction = forecast, h = 336, method = "arima", cv = num_folds)

# Cross-validation for Cyclist Area
cv_results_cyc <- tsCV(tsCycArea, forecastfunction = forecast, h = 336, method = "arima", cv = num_folds)

# Calculate mean cross-validation errors for each model
mean_cv_error_car <- mean(cv_results_car, na.rm = TRUE)
mean_cv_error_ped <- mean(cv_results_ped, na.rm = TRUE)
mean_cv_error_cyc <- mean(cv_results_cyc, na.rm = TRUE)

# Calculate the mean of the cross-validated forecast errors for Car Area
mean_cv_error_car <- mean(cv_results_car, na.rm = TRUE)

# Print the mean cross-validation errors
cat("Mean CV Error for Car Area:", mean_cv_error_car, "\n")
cat("Mean CV Error for Pedestrian Area:", mean_cv_error_ped, "\n")
cat("Mean CV Error for Cyclist Area:", mean_cv_error_cyc, "\n")
```

forecastingLTNBoundary.R

```r
library(tidyverse)
library(forecast)
library(ggplot2)
library(rio)
library(tseries)

#filter data for each region

iffleyLTN <- boundaryLTNs %>%
  filter(LTNregion == "Iffley ") %>%
  mutate(year = format(date, "%Y"),
         month = format(date, "%m"),
         hour = factor(substr(time_from, 1, 2), levels = sprintf("%02d", 0:23))) %>%
  mutate(month = as.numeric(month)) %>%
  mutate(hour = as.numeric(hour)) %>%
  group_by(year, month, hour) %>%
  summarise(avg_pedestrian_count = mean(pedestrian),
            avg_cyclist_count = mean(cyclist),
            avg_car_count = mean(car))

morrellLTN <- boundaryLTNs %>%
  filter(LTNregion == "Morrell") %>%
  mutate(year = format(date, "%Y"),
         month = format(date, "%m"),
         hour = factor(substr(time_from, 1, 2), levels = sprintf("%02d", 0:23))) %>%
  mutate(month = as.numeric(month)) %>%
  mutate(hour = as.numeric(hour)) %>%
  group_by(year, month, hour) %>%
  summarise(avg_pedestrian_count = mean(pedestrian),
            avg_cyclist_count = mean(cyclist),
            avg_car_count = mean(car))

cowleyLTN <- boundaryLTNs %>%
  filter(LTNregion == "Cowley Road East") %>%
  mutate(year = format(date, "%Y"),
         month = format(date, "%m"),
         hour = factor(substr(time_from, 1, 2), levels = sprintf("%02d", 0:23))) %>%
  mutate(month = as.numeric(month)) %>%
  mutate(hour = as.numeric(hour)) %>%
  group_by(year, month, hour) %>%
  summarise(avg_pedestrian_count = mean(pedestrian),
            avg_cyclist_count = mean(cyclist),
            avg_car_count = mean(car))

stclementsLTN <- boundaryLTNs %>%
  filter(LTNregion == "St Clements") %>%
  mutate(year = format(date, "%Y"),
         month = format(date, "%m"),
         hour = factor(substr(time_from, 1, 2), levels = sprintf("%02d", 0:23))) %>%
  mutate(month = as.numeric(month)) %>%
  mutate(hour = as.numeric(hour)) %>%
  group_by(year, month, hour) %>%
  summarise(avg_pedestrian_count = mean(pedestrian),
            avg_cyclist_count = mean(cyclist),
            avg_car_count = mean(car))
```

```r
## create TS for each counts for each region 23/24 forecasts

iffley_pedesTS <- ts(iffleyLTN$avg_pedestrian_count, frequency = 24*12, start = c(2019, 1), end = c(2023,7))
iffley_cycTS <- ts(iffleyLTN$avg_cyclist_count, frequency = 24*12, start = c(2019, 1), end = c(2023,7))
iffley_carTS <- ts(iffleyLTN$avg_car_count, frequency = 24*12, start = c(2019, 1), end = c(2023,7))

cowley_pedesTS <- ts(cowleyLTN$avg_pedestrian_count, frequency = 24*12, start = c(2019, 1), end = c(2023,7))
cowley_cycTS <- ts(cowleyLTN$avg_cyclist_count, frequency = 24*12, start = c(2019, 1), end = c(2023,7))
cowley_carTS <- ts(cowleyLTN$avg_car_count, frequency = 24*12, start = c(2019, 1), end = c(2023,7))

morrell_pedesTS <- ts(morrellLTN$avg_pedestrian_count, frequency = 24*12, start = c(2019, 1), end = c(2023,7))
morrell_cycTS <- ts(morrellLTN$avg_cyclist_count, frequency = 24*12, start = c(2019, 1), end = c(2023,7))
morrell_carTS <- ts(morrellLTN$avg_car_count, frequency = 24*12, start = c(2019, 1), end = c(2023,7))

stclements_pedesTS <- ts(stclementsLTN$avg_pedestrian_count, frequency = 24*12, start = c(2019, 1), end = c(2023,7))
stclements_cycTS <- ts(stclementsLTN$avg_cyclist_count, frequency = 24*12, start = c(2019, 1), end = c(2023,7))
stclements_carTS <- ts(stclementsLTN$avg_car_count, frequency = 24*12, start = c(2019, 1), end = c(2023,7))

## create arima models

arimaIffleyPedes <- auto.arima(iffley_pedesTS, stepwise = FALSE, parallel = TRUE)
arimaIffleyCyc <- auto.arima(iffley_cycTS, stepwise = FALSE, parallel = TRUE)
arimaIffleyCar <- auto.arima(iffley_carTS, stepwise = FALSE, parallel = TRUE)
arimaMorrellPedes <- auto.arima(morrell_pedesTS, stepwise = FALSE, parallel = TRUE)
arimaMorrellCyc <- auto.arima(morrell_cycTS, stepwise = FALSE, parallel = TRUE)
arimaMorrellCar <- auto.arima(morrell_carTS, stepwise = FALSE, parallel = TRUE)
arimaCowleyPedes <- auto.arima(cowley_pedesTS, stepwise = FALSE, parallel = TRUE)
arimaCowleyCyc <- auto.arima(cowley_cycTS, stepwise = FALSE, parallel = TRUE)
arimaCowleyCar <- auto.arima(cowley_carTS, stepwise = FALSE, parallel = TRUE)
arimaSTCPedes <- auto.arima(stclements_pedesTS, stepwise = FALSE, parallel = TRUE)
arimaSTCCyc <- auto.arima(stclements_cycTS, stepwise = FALSE, parallel = TRUE)
arimaSTCCar <- auto.arima(stclements_carTS, stepwise = FALSE, parallel = TRUE)

summary(arimaCowleyCyc, "report")

## summary outputs
summary(arimaIffleyPedes)
summary(arimaIffleyCar)
summary(arimaMorrellPedes)
summary(arimaMorrellCar)
summary(arimaMorrellCyc)
summary(arimaCowleyPedes)
summary(arimaCowleyCyc)
summary(arimaCowleyCar)
summary(arimaSTCPedes)
summary(arimaSTCCyc)
summary(arimaSTCCar)
```

```r
# create forecasts

forecastIffleyPedes <- forecast(arimaIffleyPedes, h = 535)
#forecastIffleyCyc <- forecast(arimaIffleyCyc, h = 365)
forecastIffleyCar <- forecast(arimaIffleyCar, h = 535)
forecastMorrellPedes <- forecast(arimaMorrellPedes, h = 535)
forecastMorrellCyc <- forecast(arimaMorrellCyc, h = 535)
forecastMorrellCar <- forecast(arimaMorrellCar, h = 535)
forecastCowleyPedes <- forecast(arimaCowleyPedes, h = 535)
forecastCowleyCyc <- forecast(arimaCowleyCyc, h = 535)
forecastCowleyCar <- forecast(arimaCowleyCar, h = 535)
forecastSTCPedes <- forecast(arimaSTCPedes, h = 535)
forecastSTCCyc <- forecast(arimaSTCCyc, h = 535)
forecastSTCCar <- forecast(arimaSTCCar, h = 535)

forecast2IffleyPed <- forecast(arima2IffleyPed, h = 335)
forecast2IffleyCar <- forecast(arima2IffleyCar, h = 335)
forecast2IffleyCyc <- forecast(arima2IffleyCyc, h = 335)
forecast2MorrelPed <- forecast(arima2MorrelPed, h = 335)
forecast2MorrelCyc <- forecast(arima2MorrelCyc, h = 335)
forecast2MorrelCar <- forecast(arima2MorrelCar, h = 335)
forecast2CowleyCar <- forecast(arima2CowleyCar, h = 335)
forecast2CowleyPed <- forecast(arima2CowleyPed, h = 335)
forecast2CowleyCyc <- forecast(arima2CowleyCyc, h = 335)
forecast2STCPed <- forecast(arima2STCPed, h = 335)
forecast2STCCar <- forecast(arima2STCcar, h = 335)
forecast2STCCyc <- forecast(arima2STCCyc, h = 335)


# Plotting the 2023/2024 forecasts

par(mfrow = c(2,1))
plot(forecastIffleyPedes, main = "Forecast from ARIMA (3,1,2)(0,1,0) for Pedestrian Counts at Iffley Road LTN Boundary", ylab = "Count", xlab = "Year")
#plot(forecastIffleyCyc) ## MODEL NOT RUNNING TAKING TOO LONG
plot(forecastIffleyCar, main = "Forecast from ARIMA (1,1,0)(0,1,0) for Car Counts at Iffley Road LTN Boundary", ylab = "Count", xlab = "Year")

par(mfrow = c(3,1))
plot(forecastMorrellPedes, main = "Forecast from ARIMA (3,1,2)(0,1,0) for Pedestrian Counts at Morrell Avenue LTN Boundary", ylab = "Count", xlab = "Year")
plot(forecastMorrellCyc, main = "Forecast from ARIMA (0,1,3)(0,1,0) for Cyclist Counts at Morrell Avenue LTN Boundary", ylab = "Count", xlab = "Year")
plot(forecastMorrellCar, main = "Forecast from ARIMA (1,1,2)(0,1,0) for Car Counts at Morrell Avenue LTN Boundary", ylab = "Count", xlab = "Year")


par(mfrow = c(3,1))
plot(forecastCowleyPedes, main = "Forecast from ARIMA (0,1,5) for Pedestrian Counts at Cowley Road LTN Boundary", ylab = "Count", xlab = "Year")
plot(forecastCowleyCyc, main = "Forecast from ARIMA (1,0,3)(0,1,0) for Cyclist Counts at Cowley Road LTN Boundary", ylab = "Count", xlab = "Year")
plot(forecastCowleyCar, main = "Forecast from ARIMA (3,1,2)(0,1,0) for Car Counts at Cowley Road LTN Boundary", ylab = "Count", xlab = "Year")

par(mfrow = c(3,1))
plot(forecastSTCPedes, main = "Forecast from ARIMA (3,1,2)(0,1,0) for Pedestrian Counts at St Clements LTN Boundary", ylab = "Count", xlab = "Year")
plot(forecastSTCCyc, main = "Forecast from ARIMA (1,1,3)(0,1,0) for Cyclist Counts at St Clements LTN Boundary", ylab = "Count", xlab = "Year")
plot(forecastSTCCar, main = "Forecast from ARIMA (1,1,0)(0,1,0) for Car Counts at St Clements LTN Boundary", ylab = "Count", xlab = "Year")
```

```r
## plot post implementation forecasts

par(mfrow = c(3,1))
plot(forecast2IffleyPed, main = "Forecast from ARIMA (2,1,3) for Pedestrian Counts at Iffley Road LTN Boundary", ylab = "Count", xlab = "Year")
plot(forecast2IffleyCyc, main = "Forecast from ARIMA (2,0,2) for Cyclist Counts at Iffley Road LTN Boundary", ylab = "Count", xlab = "Year")
plot(forecast2IffleyCar, main = "Forecast from ARIMA (4,0,1) for Car Counts at Iffley Road LTN Boundary", ylab = "Count", xlab = "Year")

par(mfrow = c(3,1))
plot(forecast2MorrelPed, main = "Forecast from ARIMA (2,0,2) for Pedestrian Counts at Morrell Avenue LTN Boundary", ylab = "Count", xlab = "Year")
plot(forecast2MorrelCar, main = "Forecast from ARIMA (2,0,2) for Car Counts at Morrell Avenue LTN Boundary", ylab = "Count", xlab = "Year")
plot(forecast2MorrelCyc, main = "Forecast from ARIMA (2,0,3) for Cyclist Counts at Morrell Avenue LTN Boundary", ylab = "Count", xlab = "Year")

par(mfrow = c(3,1))
plot(forecast2CowleyPed, main = "Forecast from ARIMA (2,1,3) for Pedestrian Counts at Cowley Road LTN Boundary", ylab = "Count", xlab = "Year")
plot(forecast2CowleyCar, main = "Forecast from ARIMA (4,0,1) for Car Counts at Morrell Avenue LTN Boundary", ylab = "Count", xlab = "Year")
plot(forecast2CowleyCyc, main = "Forecast from ARIMA (1,1,3) for Cyclist Counts at Morrell Avenue LTN Boundary", ylab = "Count", xlab = "Year")

par(mfrow = c(3,1))
plot(forecast2STCPed, main = "Forecast from ARIMA (4,0,1) for Pedestrian Counts at St Clemenets LTN Boundary", ylab = "Count", xlab = "Year")
plot(forecast2STCCar, main = "Forecast from ARIMA (5,0,0) for Car Counts at St Clemenets LTN Boundary", ylab = "Count", xlab = "Year")
plot(forecast2STCCyc, main = "Forecast from ARIMA (4,0,1) for Cyclist Counts at St Clemenets LTN Boundary", ylab = "Count", xlab = "Year")
```

causalImpactLTNArea.R

```r
library(tidyverse)
library(forecast)
library(ggplot2)
library(lubridate)
library(janitor)
library(zoo)
library(CausalImpact)

# set pre and post period
pre.period <- c(as.numeric(as.yearmon("2019-01")), as.numeric(as.yearmon("2022-05")))
post.period <- c(as.numeric(as.yearmon("2022-06")), as.numeric(as.yearmon("2023-06")))

#create tS objects
pedesAreaTS <- ts(hourlyLTNArea$avg_pedestrian_count, start = c(2019, 1), end = c(2023, 7), frequency = 24*12)
cycAreaTS <- ts(hourlyLTNArea$avg_cyclist_count, start = c(2019, 1), end = c(2023, 7), frequency = 24*12)
carAreaTS <- ts(hourlyLTNArea$avg_car_count, start = c(2019, 1), end = c(2023, 7), frequency = 24*12)

#create CI model
modelPedArea <- CausalImpact(zoo(pedesAreaTS), pre.period, post.period)
modelCycArea <- CausalImpact(zoo(cycAreaTS), pre.period, post.period)
modelCarArea <- CausalImpact(zoo(carAreaTS), pre.period, post.period)

#make plots
plot(modelPedArea) +
  labs(title = "Causal Impact Analysis for Pedestrian Counts in LTN Area",
       caption = "The blue dashed line represents the predicted values from the Causal Impact model. The black solid line represents the observed values")
plot(modelCycArea)  +
  labs(title = "Causal Impact Analysis for Cyclist Counts in LTN Area",
       caption = "The blue dashed line represents the predicted values from the Causal Impact model. The black solid line represents the observed values")
plot(modelCarArea) +
  labs(title = "Causal Impact Analysis for Car Counts in LTN Area",
       caption = "The blue dashed line represents the predicted values from the Causal Impact model. The black solid line represents the observed values")
```

causalImpactBoundary.R

```r
## install packages
library(tidyverse)
library(forecast)
library(ggplot2)
library(lubridate)
library(janitor)
library(zoo)
library(CausalImpact)

#set pre and post period

pre.period <- c(as.numeric(as.yearmon("2019-01")), as.numeric(as.yearmon("2022-05")))
post.period <- c(as.numeric(as.yearmon("2022-06")), as.numeric(as.yearmon("2023-06")))

# filter for different location

iffley_CI <- hourlyLTNboundary %>% filter(LTNregion == "Iffley ")
morrel_CI <- hourlyLTNboundary %>%  filter(LTNregion == "Morrell")
cowley_CI <- hourlyLTNboundary %>%  filter(LTNregion == "Cowley Road East")
stclements_CI <- hourlyLTNboundary %>%  filter(LTNregion == "St Clements")

## IFFLEY

iffleyPedTS <- ts(iffley_CI$avg_pedestrian_count, start = c(2019, 1), end = c(2023, 7), frequency = 24*12)
iffleyCycTS <- ts(iffley_CI$avg_cyclist_count, start = c(2019, 1), end = c(2023, 7), frequency = 24*12)
iffleyCarTS <- ts(iffley_CI$avg_car_count, start = c(2019, 1), end = c(2023, 7), frequency = 24*12)

### create Impact model

iffleyPedImpact <- CausalImpact(zoo(iffleyPedTS), pre.period, post.period)
iffleyCycImpact <- CausalImpact(zoo(iffleyCycTS), pre.period, post.period)
iffleyCarImpact <- CausalImpact(zoo(iffleyCarTS), pre.period, post.period)

## plot for Iffley
iffley_pedes_plot <- plot(iffleyPedImpact) +
  labs(title = "Causal Impact Analysis for Pedestrian Counts at Iffley Road LTN Boundary",
       caption = "The blue dashed line represents the predicted values from the Causal Impact model. The black solid line represents the observed values")

iffley_cyclist_plot <- plot(iffleyCycImpact) +
  labs(title = "Causal Impact Analysis for Cyclist Counts at Iffley Road LTN Boundary",
       caption = "The blue dashed line represents the predicted values from the Causal Impact model. The black solid line represents the observed values")

iffley_car_plot <- plot(iffleyCycImpact) +
  labs(title = "Causal Impact Analysis for Car Counts at Iffley Road LTN Boundary",
       caption = "The blue dashed line represents the predicted values from the Causal Impact model. The black solid line represents the observed values")
```

```r
#summary

summary(iffleyPedImpact, "report")
summary(iffleyCycImpact, "report")
summary(iffleyCarImpact, "report")

## MORRELL

morrellPedTS <- ts(morrel_CI$avg_pedestrian_count, start = c(2019, 1), end = c(2023, 7), frequency = 24*12)
morrellCycTS <- ts(morrel_CI$avg_cyclist_count, start = c(2019, 1), end = c(2023, 7), frequency = 24*12)
morrellCarTS <- ts(morrel_CI$avg_car_count, start = c(2019, 1), end = c(2023, 7), frequency = 24*12)

## create Impact models

impactMorrellPed <- CausalImpact(zoo(morrellPedTS), pre.period, post.period)
impactMorrellCyc <- CausalImpact(zoo(morrellCycTS), pre.period, post.period)
impactMorrellCar <- CausalImpact(zoo(morrellCarTS), pre.period, post.period)

## plot
morrell_pedes_plot <- plot(impactMorrellPed) +
  labs(title = "Causal Impact Analysis for Pedestrian Counts at Morrell Avenue LTN Boundary",
       caption = "The blue dashed line represents the predicted values from the Causal Impact model. The black solid line represents the observed values")

morrell_cyclist_plot <- plot(impactMorrellCyc) +
  labs(title = "Causal Impact Analysis for Cyclist Counts at Morrell Avenue LTN Boundary",
       caption = "The blue dashed line represents the predicted values from the Causal Impact model. The black solid line represents the observed values")

morrell_car_plot <- plot(impactMorrellCar) +
  labs(title = "Causal Impact Analysis for Car Counts at Morrell Avenue LTN Boundary",
       caption = "The blue dashed line represents the predicted values from the Causal Impact model. The black solid line represents the observed values")

## summary

summary(impactMorrellPed, "report")
summary(impactMorrellCyc, "report")
summary(impactMorrellCar, "report")
```

```r
## COWLEY

cowleyPedTS <- ts(cowley_CI$avg_pedestrian_count, start = c(2019, 1), end = c(2023, 7), frequency = 24*12)
cowleyCycTS <- ts(cowley_CI$avg_cyclist_count, start = c(2019, 1), end = c(2023, 7), frequency = 24*12)
cowleyCarTS <- ts(cowley_CI$avg_car_count, start = c(2019, 1), end = c(2023, 7), frequency = 24*12)

## create Impact Model

impactCowleyPed <- CausalImpact(zoo(cowleyPedTS), pre.period, post.period)
impactCowleyCyc <- CausalImpact(zoo(cowleyCycTS), pre.period, post.period)
impactCowleyCar <- CausalImpact(zoo(cowleyCarTS), pre.period, post.period)

##plot

cowley_ped_plot <- plot(impactCowleyPed) +
  labs(title = "Causal Impact Analysis for Pedestrian Counts at Cowley Road LTN Boundary",
       caption = "The blue dashed line represents the predicted values from the Causal Impact model. The black solid line represents the observed values")

cowley_cyc_plot <- plot(impactCowleyCyc) +
  labs(title = "Causal Impact Analysis for Cyclist Counts at Cowley Road LTN Boundary",
       caption = "The blue dashed line represents the predicted values from the Causal Impact model. The black solid line represents the observed values")

cowley_car_plot <- plot(impactCowleyCar) +
  labs(title = "Causal Impact Analysis for Car Counts at Cowley Road LTN Boundary",
       caption = "The blue dashed line represents the predicted values from the Causal Impact model. The black solid line represents the observed values")

#summary

summary(impactCowleyPed, "report")
summary(impactCowleyCyc, "report")
summary(impactCowleyCar, "report")

# ST CLEMENTS

stcPedTS <- ts(stclements_CI$avg_pedestrian_count, start = c(2019, 1), end = c(2023, 7), frequency = 24*12)
stcCycTS <- ts(stclements_CI$avg_cyclist_count, start = c(2019, 1), end = c(2023, 7), frequency = 24*12)
stcCarTS <- ts(stclements_CI$avg_car_count, start = c(2019, 1), end = c(2023, 7), frequency = 24*12)

## create impact model

impactSTCPed <- CausalImpact(zoo(stcPedTS), pre.period, post.period)
impactSTCCyc <- CausalImpact(zoo(stcCycTS), pre.period, post.period)
impactSTCCar <- CausalImpact(zoo(stcCarTS), pre.period, post.period)
```

```r
#plot

stc_ped_plot <- plot(impactSTCPed) +
  labs(title = "Causal Impact Analysis for Pedestrian Counts at St Clements LTN Boundary",
       caption = "The blue dashed line represents the predicted values from the Causal Impact model. The black solid line represents the observed values")

stc_cyc_plot <- plot(impactSTCCyc) +
  labs(title = "Causal Impact Analysis for Cyclist Counts at St ClementsLTN Boundary",
       caption = "The blue dashed line represents the predicted values from the Causal Impact model. The black solid line represents the observed values")

stc_car_plot <- plot(impactSTCCar) +
  labs(title = "Causal Impact Analysis for Car Counts at St Clements LTN Boundary",
       caption = "The blue dashed line represents the predicted values from the Causal Impact model. The black solid line represents the observed values")

#summary

summary(impactSTCPed, "report")
summary(impactSTCCyc, "report")
summary(impactSTCCar, "report")
```