A Data-Driven Strategy for Alleviating Traffic Congestion in London

Application of EDA and Intelligent Transport Systems

Tobias Zeier, the 14th of July 2025
University of Essex Online

1 Introduction

The London region faces escalating challenges in traffic congestion, road safety, and environmental objectives due to its dense population and extensive transportation network. This report critically evaluates extensive traffic statistics and proposes a comprehensive Enterprise Data Analysis (EDA) strategy, exploring advanced data architectures like lakehouses and cloud technologies, including storage solutions to enhance traffic management through data-driven decision-making, particularly for congestion.

2 Traffic in the London region

London's road network consistently experiences substantial congestion, particularly during peak commuting hours, with Inner London acutely affected by its high population density and restricted road space. Key issues include a rising volume of private vehicle journeys, unreliable bus services in congested areas, and increased pressure from Heavy Goods Vehicles (HGVs) and delivery vans, largely due to the growth of e-commerce (Weltevreden, 2024).

Between 2018 and 2019, average speed on London's Strategic Road Network decreased by 0.07% and average delays surged by 12%, positioning the city as the slowest and most delayed region in the United Kingdom. While average traffic in 2023 was higher than 2022, it remained below pre-pandemic levels. A critical limitation for effective management is the continued reliance on manual measurements and the underutilisation of real-time traffic data, which severely restricts the city's capacity for dynamic, data-driven responses to evolving congestion patterns (Department for Transportation, 2023).

3 EDA Model: Storage-Centric Architecture

A typical storage configuration for a lakehouse model, exemplified by its proposed application in the London traffic management strategy, fundamentally integrates cloud infrastructure to harmonise the functionalities of traditional data lakes and data warehouses (Armbrust et al., 2021). This architecture uses a medallion approach with three layers. Bronze for immutable raw data (e.g. Parquet, Delta Lake); Silver for cleansed, standardised, and validated data, creating a single source of truth; and Gold for refined, aggregated data, often in star schemas, optimised for business intelligence and machine learning workloads (Sirbu, Taleanu and Pop, 2024). The underlying storage is typically provided by scalable, durable, and cost-effective cloud object storage services (e.g. AWS S3, Azure Blob Storage, GCP Cloud Storage), which form the backbone of the data lake component. Figure 1 visualises the proposed architecture.

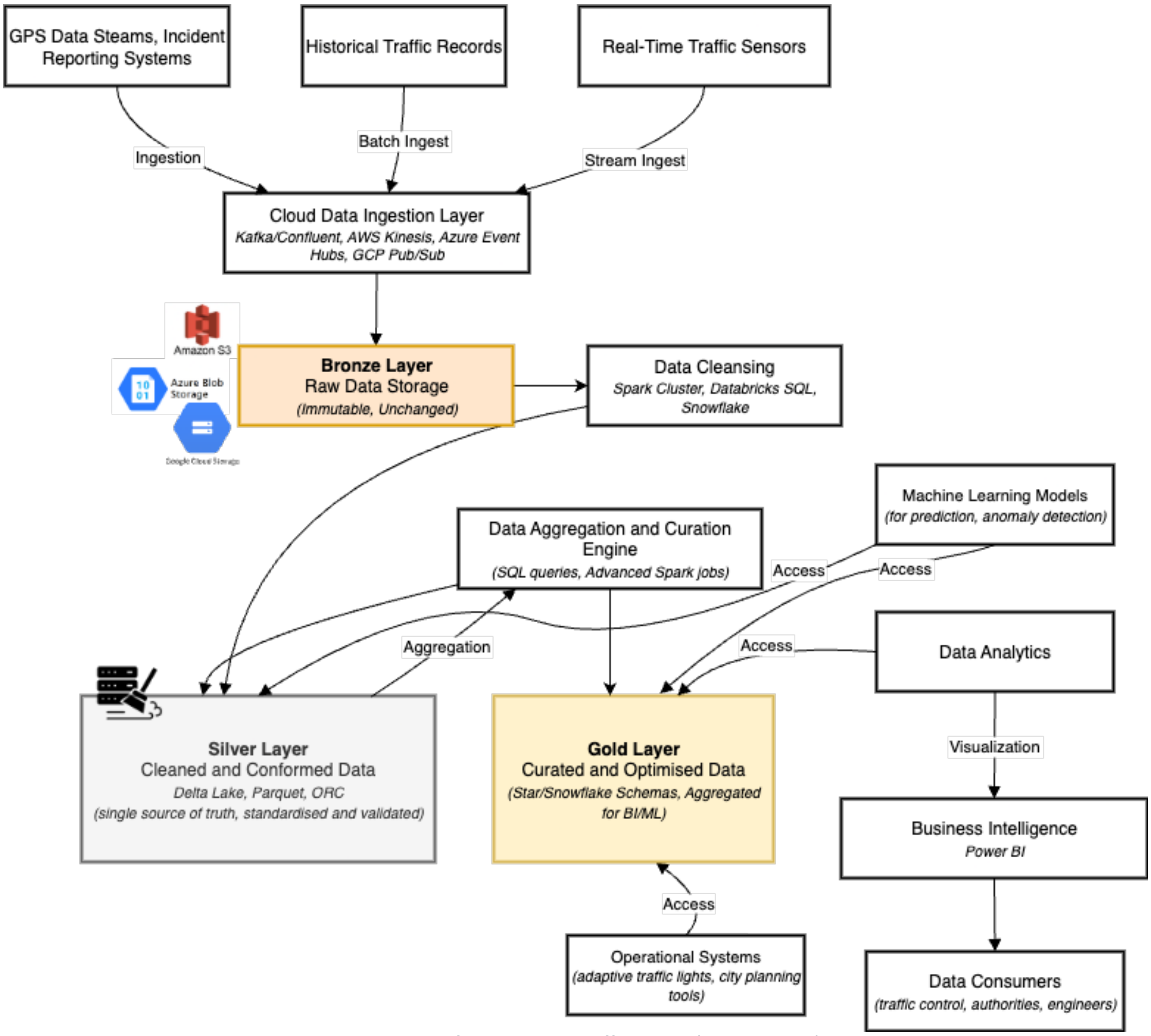

Figure 1: Proposed Lakehouse Architecture for London Traffic Data (Zeier, 2025)

Operating over these low-cost, high-latency cloud object stores presents inherent design challenges and trade-offs. Managed data services built upon this storage facilitate critical features like transactional capabilities, schema enforcement, data versioning, and integration with various data tools. However, systems must navigate choices in transaction coordination (relying on the object store for atomicity or using external services like Apache Hive MetaStore), metadata management for fast query access, and data update strategies (balancing read/write performance with Copy-On-Write or Merge-On-Read). Despite these, the design offers infinite scalability and robust data governance, including fine-grained access control and audit trails (Jain et al., 2023).

4 Data Analysis Method

The analytical methodology applied in this report is based on the Team Data Science Process (TDSP), chosen for its strong alignment with real-time analytics, cloud-native infrastructure, and adaptive modelling. Unlike more traditional approaches such as CRISP-DM, which are highly structured and governance-oriented, TDSP supports modular development, continuous iteration, and integration with modern tools like Azure Machine Learning or Databricks, all of which are vital in the context of smart city systems.

While CRISP-DM provides strong conceptual clarity and auditability, especially valued in public-sector projects, its linear structure and lack of direct tooling integration render it less suitable for real-time interventions required in managing urban traffic. TDSP's lifecycle model spans from data acquisition and cleansing to model deployment and monitoring, making it highly relevant for applications involving dynamic congestion forecasts and adaptive signal control (Hotz, 2025).

The project initially planned a linear analytical approach, but was re-evaluated due to urgent congestion hotspots, particularly in Inner London, led to a revised phased parallel approach. This ensured that real-time solutions could be developed without delay.

The revised design adopts a phased parallel approach. While foundational descriptive and diagnostic analytics continue as planned to develop dashboards and trend analyses, predictive and prescriptive analytics are developed in tandem. This facilitates early identification of patterns and the creation of real-time intervention models using optimisation algorithms and machine learning. A case in point is prescriptive analytics which are used to recommend adaptive traffic signal schedules based on predicted congestion levels derived from machine learning forecasts. These models could be tested using simulation environments and could deliver actionable outputs to traffic planners within weeks of initial data processing (Skali Lami, 2022). The revised workflow, enriched with the medallion layers, is laid out in Figure 2.

This adaptive design shift not only improves the responsiveness of the data strategy but also ensured that model development aligned more closely with real-world operational needs. It demonstrates the value of flexibility in data science methodologies when applied to dynamic urban systems like London's transport infrastructure.

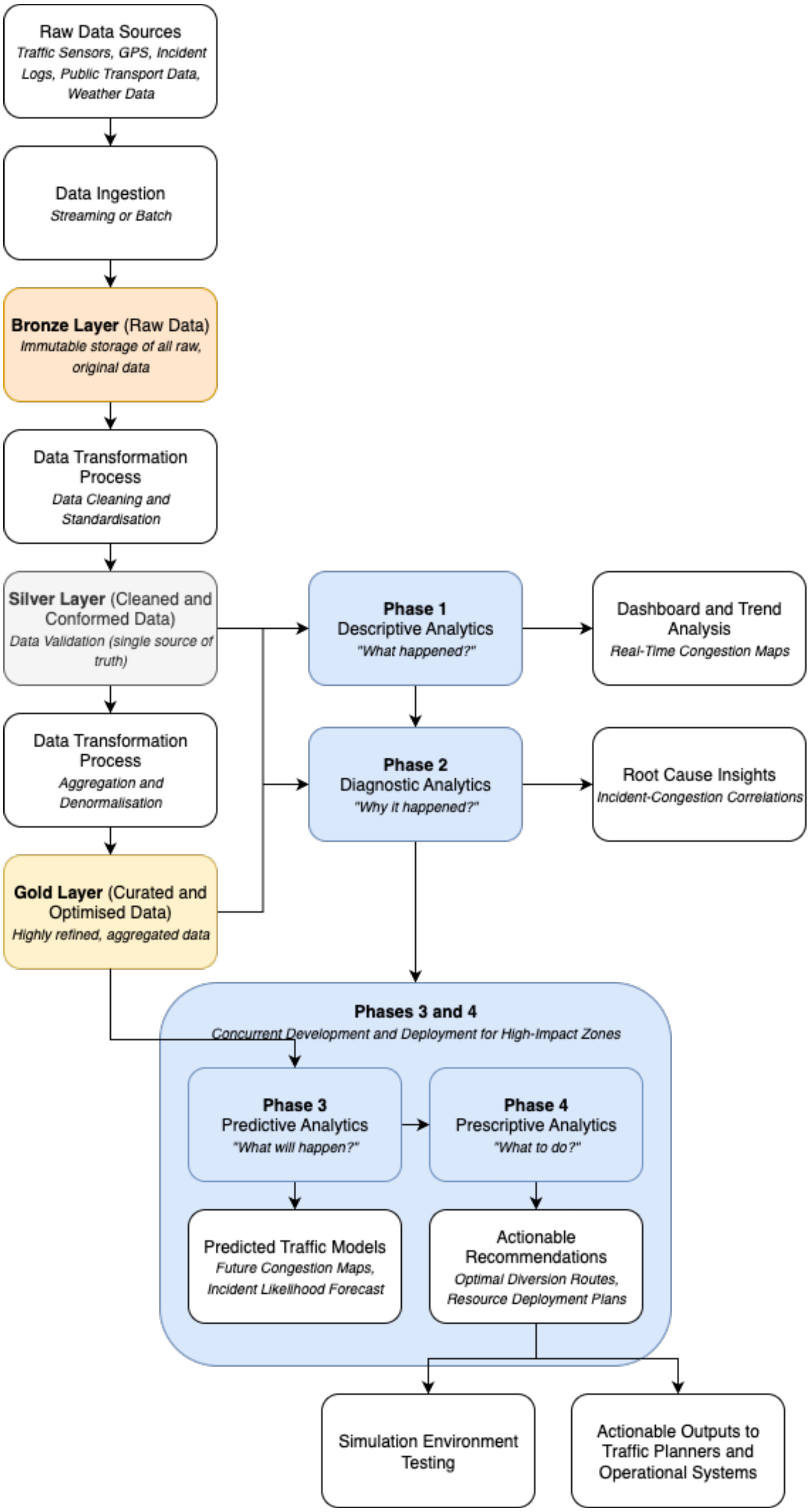

Figure 2: Adaptive Data Analysis Workflow for Real-Time Traffic Interventions in London (Zeier, 2025)

5 Data Analysis Results as Models

The data analysis encompassed descriptive, diagnostic, predictive, and prescriptive models. While currently in conceptual design and simulation phases rather than full implementation, these models show significant potential impact.

Descriptive analytics would enhance traffic monitoring by using near real-time time-series analyses to identify congestion patterns. Dashboards (e.g. in Power BI) would visualise fluctuations across times and locations, helping decision-makers understand bottleneck occurrences and inform interventions. Figure 3 shows a dashboard prototype that visualises vehicle flow trends across boroughs. Similar implementations in Power BI could assist Transportation for London (TfL) in identifying bottlenecks and responding to traffic delays with targeted signal adjustments.

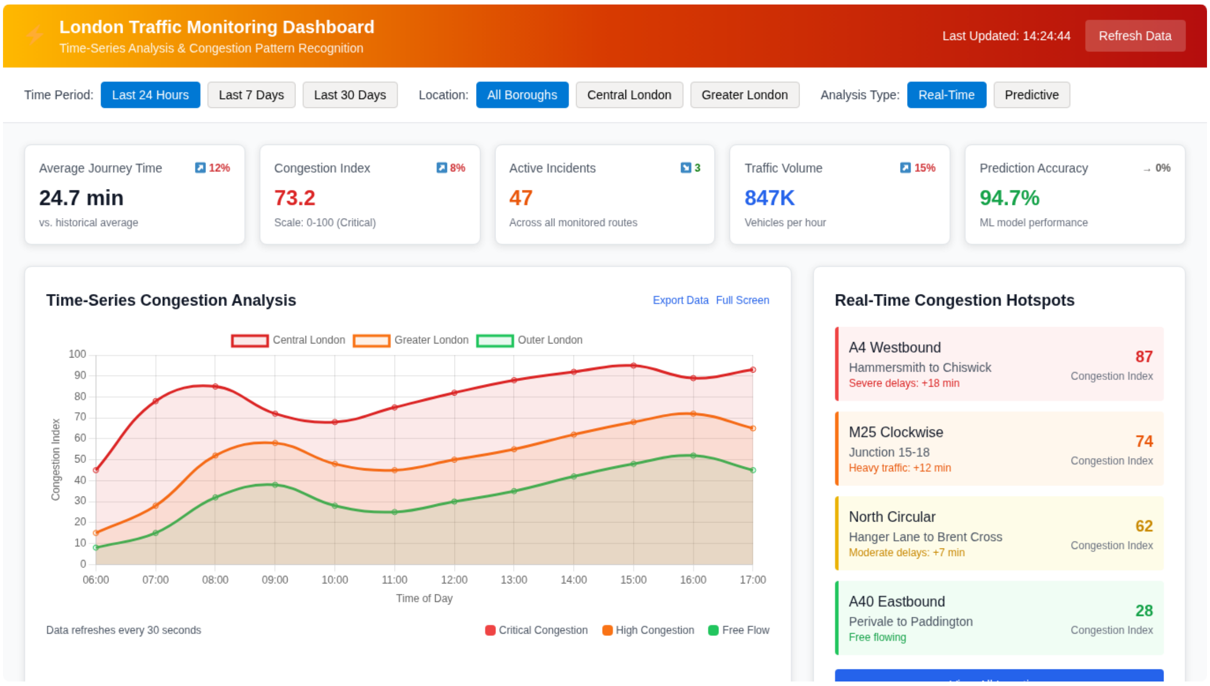

Figure 3: Time-Series Dashboard for Monitoring Congestion in London Boroughs (Canva AI, 2025)

Diagnostic analytics would identify root causes of congestion using statistical correlation and clustering methods, linking it to factors like weather, time of day, or events. For instance, consistently linking heavy rainfall to slower traffic in specific boroughs would enable contingency planning, such as early warnings or adjusted public transport services.

Predictive analytics would provide a forward-looking view, utilising machine learning models (e.g. Random Forest, ARIMA) to forecast future congestion levels with high accuracy. This foresight would enable pre-emptive actions like altering traffic signal schedules or diverting vehicles to minimise delays, especially during major events or emergencies.

As the most advanced phase, prescriptive analytics would propose specific actions based on predicted outcomes. Optimisation algorithms could suggest efficient signal timing to minimise vehicle idle time. At scale, these models could provide real-time guidance to traffic control systems and directly to drivers via navigation apps (Roy et al., 2022).

Collectively, these models, when implemented on a scalable, well-governed data platform, could transform London's traffic management from reactive to predictive and preventative, significantly reducing congestion and enhancing road safety across the region.

6 Interpretation of Results and Identified Issues

The data models developed through this EDA strategy would deliver clear insights directly addressing key issues in London’s traffic:

Underused Real-Time Data & Manual Methods:
Descriptive analytics would expose the lack of dynamic insight by contrasting static manual measurements with real-time visualisations of traffic flow and congestion. It would highlight which live data sources are most valuable, supporting investment in automation and data integration.

Congestion & Private Vehicle Growth:
Diagnostic analytics would link congestion to causes such as events, roadworks, or weather conditions. By showing correlations – such as spikes in traffic during football matches—it would highlight the need for proactive, event-specific measures rather than reactive responses.

Bus Reliability & HGV Impact:
Models could pinpoint where and when bus journeys are most delayed due to traffic, and quantify the impact of delivery HGVs, especially from e-commerce. These insights would guide targeted improvements to public transport and logistics.

Falling Speeds & Growing Delays:
Predictive analytics would forecast future slowdowns. An illustration of this might be a projected 5% drop in speeds in Inner London, making the scale and urgency of intervention clear and measurable.

Lack of Dynamic Response:
Prescriptive analytics would show how automated actions, like optimised signal timings or diversions, can ease congestion more effectively than delayed manual decisions, proving the value of real-time, data-driven control.

7 Solutions and Associated Limitations or Risks

There are a number of solutions that can be proposed, based on the findings in this EDA report. However, two specific problems are chosen to showcase solutions which could have a significant positive impact on the current traffic issues around London.

7.1 Intelligent Traffic Signal Systems (ITS)

One of the most pressing issues is the regular congestion experienced in certain areas. While it is less complicated to find solutions for issues that can be predicted based on the collected data, it is substantially more complex to find quick solutions for areas that are suddenly congested. To address this issue, an intelligent traffic signal system (ITS) should be implemented.

A study by Wu (2024) has found that intelligent traffic signal systems can significantly improve urban traffic flow in large cities. The study, which was conducted in Beijing, China, a city which has a population exceeding 21 million, found that such systems can reduce average travel times by over 30%, minimise delays and vehicle stops, and improve travel speeds. They adapt in real-time to changing traffic conditions using data and predictive algorithms, resulting in smoother journeys and shorter queues. From an environmental standpoint, these measures contribute to reduced fuel consumption, decreased carbon emissions, and enhanced air quality. The following fictional dashboard illustrates how TfL could use real-time congestion and signal status data to optimise junction coordination, reduce delays, and respond swiftly to incidents or unexpected traffic patterns in high-density areas like Westminster and Camden.

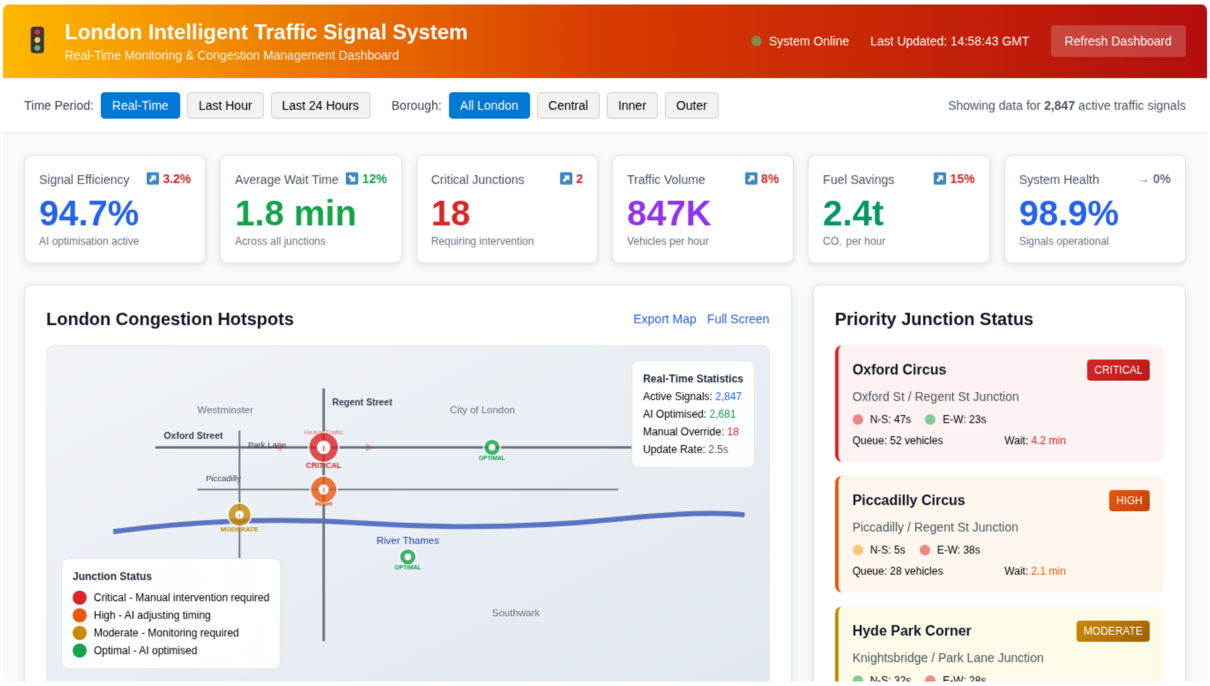

Figure 4: Conceptual ITS Dashboard for Adaptive Signal Control in Central London (Canva AI, 2025)

In contrast, Waqar et al. (2023) asserted that the effective deployment of Intelligent Transportation Systems (ITS) for urban smart mobility is substantially hampered by a range of impediments. These encompass technical issues, including reliability and safety; resource constraints, such as inadequate infrastructure; and interoperability concerns, particularly regarding compatibility and standardisation. Furthermore, significant hurdles arise from management challenges in data handling and security, personal considerations like privacy and employment implications, and the substantial economic outlays associated with implementation and maintenance.

7.2 Encourage Public Transport Usage

Optimised public transportation is crucial for mitigating urban congestion and fostering sustainable mobility. Akram et al. (2025) have conducted a study in Manchester, which shows its multi-faceted approach. For instance, it seeks to reduce private car dependency by expanding capacity and integrating its extensive MetroLink tram system. This includes funding for improved buses and implementing priority measures to enhance reliability and streamline journeys. Furthermore, Manchester leverages smart technologies and big data to boost public transport convenience and efficiency, treating connectivity as a service and enhancing digital infrastructure to provide reliable, real-time information. The aim is to create a positive feedback loop: increased public transport patronage reduces private car usage, thereby lessening overall traffic and further enhancing the attractiveness and reliability of public services.

However, Manchester appears to face specific limitations in this endeavour. The city experiences difficulties due to past infrastructure legacies and a historically increasing use of private cars, which pose challenges to encouraging a greater modal shift. Traffic congestion remains somewhat unpredictable, and ongoing problems persist in reducing transport emissions due to the persistently high number of private cars. Despite investments, achieving a significant modal shift remains a challenge where developed areas have been designed around car infrastructure. Additionally, while not explicitly detailed for Manchester, general limitations such as substantial investment, disruption during development, and public acceptance of measures like reallocating road space for bus lanes are inherent to such large-scale transformations.

8 Conclusion

The proposed Enterprise Data Analysis (EDA) strategy, featuring a storage-centric lakehouse architecture and a phased parallel analytical methodology, offers clear, actionable insights to address London's critical traffic challenges. The conceptual design and simulated impact of descriptive, diagnostic, predictive, and prescriptive models indicate a significant potential to transform London's reactive traffic management into a predictive and preventative system. This framework, combined with solutions like Intelligent Traffic Signal Systems and optimised public transport, directly addresses issues such as underutilised real-time data, pervasive congestion, poor bus reliability, and declining average speeds. Although the proposed solutions present implementation and historical infrastructure challenges, this approach promises to significantly reduce congestion and enhance road safety across the region.

Word Count: 2,030

References:

Weltevreden, J. (2024) *European E-Commerce Report 2024.* Amsterdam: University of Applied Sciences & Ecommerce Europe. Available at: https://ecommerce-europe.eu/wp-content/uploads/2024/10/CMI2024_Complete_light_v1.pdf (Accessed 18 June 2025).

Department for Transport (2023) *Road traffic statistics*. Available at: https://roadtraffic.dft.gov.uk/regions/6 (Accessed 18 June 2025).

Armbrust, M., Ghodsi, A., Xin, R. and Zaharia, M. (2021) ‘Lakehouse: a new generation of open platforms that unify data warehousing and advanced analytics’, *Conference on Innovative Data Systems Research (CIDR)*. Virtual Event, 11-15 January. Available at: https://www.cidrdb.org/cidr2021/papers/cidr2021_paper17.pdf (Accessed: 19 June 2025).

Sirbu, D.-I., Taleanu, A.-T., and Pop, F. (2024) ‘Replication as Lineage Mechanism for Materialized Views in Lakehouse Architectures’, *2024 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*. Craiova, Romania, 4-6 September. IEEE. 1-7. Available at: https://doi.org/10.1109/INISTA62901.2024.10683854

Jain, P., Kraft, P., Power, C., Das, T., Stoica, I., and Zaharia, M. (2023) ‘Analyzing and Comparing Lakehouse Storage Systems’, *Conference on Innovative Data Systems Research (CIDR)*. Amsterdam, Netherlands, 8-11 January. CIDR. Available at: https://mail.vldb.org/cidrdb/papers/2023/p92-jain.pdf (Accessed 19 June 2025).

Hotz, N. (2025) What is TDSP? Available at: https://www.datascience-pm.com/tdsp/ (Accessed 19 June 2025).

Skali Lami, O. (2022) *Predictive and Prescriptive Analytics in Operations Management*. PhD Thesis. Massachusetts Institute of Technology. Available at: https://dspace.mit.edu/handle/1721.1/144567 (Accessed 19 June 2025).

Zeier, T. (2025) Enterprise Data Executive Summary and Implementation Report. *IDS_PCOM7E April 2025*. Essay submitted to the University of Essex Online.

Canva AI (2025) Canva AI response to Tobias Zeier. 6 July.

Roy, D., Srivastava, R., Jat, M. and Karaca, M.S. (2022) 'A Complete Overview of Analytics Techniques: Descriptive, Predictive, and Prescriptive', In: Jeyanthi, P.M., Choudhury, T., Hack-Polay, D., Singh, T.P. and Abujar, S. (eds) *Decision Intelligence Analytics and the Implementation of Strategic Business Management*. Cham, Springer, pp. 15-30. Available at: https://doi.org/10.1007/978-3-030-82763-2_2

Wu, J. (2024) 'Analysis of the Impact of Intelligent Traffic Signals on Reducing Urban Traffic Congestion', *Frontiers in Science and Engineering*, 4(12), pp. 73-80. Available at: https://doi.org/10.54691/43k7t665

Waqar, A., Alshehri, A. H., Alanazi, F., Alotaibi, S., and Almujibah, H. R. (2023) 'Evaluation of challenges to the adoption of intelligent transportation system for urban smart mobility', *Research in Transportation Business & Management*, *51*, 101060. Available at: https://doi.org/10.1016/j.rtbm.2023.101060

Akram, N., Chen, W., Tariq, K., Hou Linjun, and Zhaosheng, L. (2025) 'Role of Public Transportation in Reducing Traffic Congestion, Enhancing Connectivity, And Promoting Sustainable Urban Development in Manchester And Shenzhen', *The International Journal of Social Sciences and Humanities Invention*, *12*(01), pp. 8401-8408. Available at: https://doi.org/10.18535/ijsshi/v12i01.01