International Journal of Interdisciplinary Innovative Research & Development (IJIIRD)
ISSN: 2456-236X
Vol. 11 Special Issue 03 | 2026

Design and Implementation of Ridership, Revenue, and Energy Analytics Dashboard for public transport system using Power BI

Kushal Nehete, Department of Computer Engineering (M.E.) PHCET, Rasayani Affiliated to University of Mumbai, Mumbai, India

Dr. Rajashree Gadhave, Department of Computer Engineering, PHCET, RasayaniAffiliated to University of Mumbai Mumbai, India

DOI: 10.5281/zenodo.20631283

ABSTRACT

Urban metro rail systems generate large volumes of operational and financial data across multiple subsystems, including fare collection, train operations, and energy management. The lack of integrated analytics platforms often limits the effective utilization of this data for real-time monitoring and strategic decision-making. This paper proposes a data-driven analytics framework for metro rail operations using Power BI. The proposed framework integrates ridership, revenue, train operation, and energy consumption data through secure API-based data acquisition and centralized data staging. Interactive dashboards are developed to provide station-wise performance metrics, trend analysis, and demand forecasting. The proposed system aims to enhance data visibility, improve operational efficiency, and support evidence-based decision-making in metro rail management. The work demonstrates the practical application of business intelligence techniques in the context of smart and sustainable urban transportation systems.

Keywords—Metro Rail Systems, Business Intelligence, Power BI, Ridership Analytics, Revenue Analytics, Energy Monitoring, API Integration, Smart Transportation.

1. INTRODUCTION

1.1 Background

Rapid urbanization and population growth have significantly increased the demand for efficient, reliable, and sustainable public transportation systems. Metro rail networks have emerged as a critical mode of mass transit in metropolitan regions due to their high passenger-carrying capacity, reduced road congestion, and lower environmental impact compared to private transportation. As metro systems expand in scale and complexity, effective operational management has become increasingly dependent on timely and accurate data analysis. Modern metro rail operations generate large volumes of data from multiple subsystems, including Automatic Fare Collection (AFC), Train Control and Management Systems (TCMS), Energy Management Systems (EMS), and Operations Control Centers (OCC). These subsystems capture detailed information related to passenger movement, revenue collection, train scheduling, and energy consumption. Despite the availability of such data, operational insights are often derived from fragmented and periodic reports, which provide limited real-time visibility and analytical depth. The lack of integrated analytics platforms poses challenges in monitoring station-wise performance, identifying demand variations, optimizing resource utilization, and supporting proactive decision-making. Traditional reporting mechanisms are largely static, require extensive manual intervention, and lack predictive capabilities. As a result, the potential of operational data to support strategic planning, cost optimization, and service reliability remains underexploited. Business Intelligence (BI) platforms such as Power BI provide advanced capabilities for data integration, interactive visualization, and analytical reporting. When combined with secure API-based data acquisition, BI solutions enable real-time monitoring, trend analysis, and forecasting. Such data-driven approaches align with the principles of Intelligent Transportation Systems (ITS) and smart city initiatives, which emphasize efficiency, sustainability, and evidence-based decision-making.This background establishes the need for a centralized, data-driven analytics framework capable of transforming metro operational data into actionable insights. By leveraging modern BI tools and standardized data integration mechanisms, metro authorities can enhance operational visibility, improve decision-making, and support sustainable urban transportation management.

1.2 Motivation

The growing scale and complexity of urban metro rail systems demand advanced analytical capabilities to support efficient operations and informed decision-making. As passenger volumes increase and service expectations rise, metro authorities face challenges in effectively monitoring ridership patterns, revenue performance, operational efficiency, and energy utilization using traditional reporting mechanisms. Operational data generated by metro subsystems is often distributed across multiple platforms, resulting in fragmented visibility and delayed insights. Manual data consolidation processes are time-consuming, error-prone, and unsuitable for real-time operational monitoring. Moreover, static reports lack the analytical depth required to identify trends, detect anomalies, and support proactive planning. Energy consumption represents a significant portion of metro operational costs, and inefficient utilization directly impacts financial sustainability. Similarly, variations in passenger demand across stations, time periods, and service patterns necessitate continuous monitoring and predictive analysis to optimize resource allocation and service planning. Without an integrated analytics framework, it becomes difficult to correlate ridership demand with revenue generation and energy usage. Advancements in Business Intelligence technologies, data integration mechanisms, and analytical modeling provide an opportunity to address these challenges. Tools such as Power BI enable interactive visualization, real-time analytics, and trend-based forecasting when supported by structured data integration through APIs. Leveraging these technologies can transform raw operational data into actionable intelligence. The motivation for this project is to develop a centralized, data-driven analytics framework that consolidates metro operational data and delivers meaningful insights to stakeholders. Such a framework is expected to enhance operational transparency, support evidence-based decision-making, and contribute to the efficient and sustainable management of urban metro rail systems.

1.3 Problem Statement

There is a need to design and implement a centralized analytics system capable of integrating data from multiple metro operational subsystems through secure API-based data acquisition. The system should effectively analyze key operational parameters such as ridership, revenue, and energy consumption, and provide insights at various levels including station-wise, daily, monthly, and yearly perspectives. Additionally, the system should support trend analysis and forecasting to identify patterns and assist in future planning. The analytics framework must also enable interactive and intuitive visualization using Power BI, thereby facilitating informed decision-making for operational and strategic stakeholders.

2. LITERATURE REVIEW

2.1 Metro Station Ridership Influencing Factors (He, Zhao, Tsui – 2018)

He, Zhao, and Tsui (2018) conducted an in-depth study to identify the factors influencing metro station ridership using the Taipei Metro as a case study. The authors focused on station-level ridership behavior and examined how travel demand varies across weekdays and weekends. To achieve this, the study employed Ordinary Least Squares (OLS) regression combined with a backward stepwise variable selection approach. This methodology enabled the identification of statistically significant factors while eliminating redundant variables. A key contribution of this study was the incorporation of concepts from complex network theory, specifically degree centrality and betweenness centrality, to quantitatively represent metro network structure. Unlike earlier studies that treated transfer stations or terminal stations as simple dummy variables, this approach provided a more rigorous and measurable understanding of a station's importance within the network. Additionally, the authors performed outlier detection and introduced a dummy variable for transportation hubs to improve model goodness-of-fit. The findings revealed that ridership determinants differ significantly between weekdays and weekends. Weekday ridership was largely influenced by commuting-related factors, whereas weekend ridership was driven more by commercial land use and leisure activities. Land use, accessibility, proximity to the city center, feeder bus connectivity, and station age were identified as major contributors to ridership demand. While the study provides strong theoretical insights for transit-oriented development (TOD) and metro planning, it relies on static and short-term datasets and is primarily focused on planning-stage analysis. The absence of real-time operational monitoring and system-level integration limits its applicability for daily metro operations and decision support.

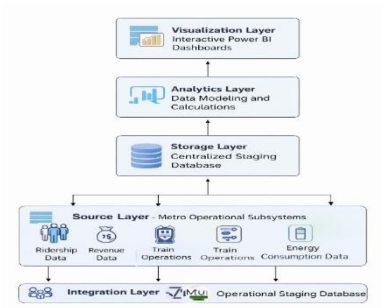
2.2 Logic Fare Revenue Forecasting in Public Transport (Krembsler et al. – 2024)

[3] Krembsler et al. (2024) presented a comparative case study on fare revenue forecasting for a large public transport operator in Berlin. The primary objective of the study was to evaluate the suitability of different predictive models for revenue forecasting at both product-group and total-market levels. The authors employed time-series forecasting techniques, regression models, and hybrid models incorporating exogenous variables to assess predictive performance under varying demand conditions. The study demonstrated that for fare products exhibiting stable trends and seasonality, simpler time-series models without exogenous variables were sufficient to achieve accurate predictions. However, for total market revenue, which is influenced by multiple external factors, the inclusion of exogenous variables significantly improved forecasting accuracy. The authors highlighted the importance of incorporating external influences such as mobility disruptions, policy changes, and evolving fare structures, particularly in the context of unprecedented events like the COVID-19 pandemic. While the study contributes valuable insights into revenue forecasting methodologies and supports management-level financial planning, its scope is limited to revenue data alone. The analysis does not incorporate ridership

patterns, operational performance, or energy consumption, nor does it address real-time data ingestion or interactive visualization. Furthermore, the forecasting framework is largely offline and analytical in nature, lacking an operational analytics or dashboard-based decision support system.

2.3 Advanced Technologies in Public Transportation (Krishnamoorthy & Venugopal – 2025)
Krishnamoorthy and Venugopal (2025) presented a comprehensive review of advanced technologies applied in public transportation systems, focusing on the integration of Intelligent Transportation Systems (ITS). The study examined the role of big data analytics, machine learning, Internet of Things (IoT), cloud computing, and advanced communication frameworks in enhancing transportation efficiency, reliability, and sustainability.
The authors discussed how real-time sensor data, GPS tracking, and data-driven algorithms can improve demand estimation, congestion mitigation, intelligent scheduling, and route optimization. The review also highlighted modern fare collection mechanisms such as biometric ticketing, smart cards, QR-code-based systems, and Automated Passenger Counting (APC) technologies, which contribute to reduced fare evasion and improved revenue assurance. Additionally, the study explored demand-responsive transit systems for low-demand areas, leveraging machine learning and optimization algorithms to enhance service coverage and passenger satisfaction.
Although the paper provides a broad and insightful overview of emerging technologies in public transportation, it remains conceptual and review-oriented. It does not present a concrete implementation framework, metro-specific analytics architecture, or operational dashboards. The lack of practical deployment scenarios and integrated performance analytics limits its direct applicability for real-world metro operational management.

3. METHODOLOGY



3.1 Proposed Architecture
The proposed architecture is designed to support centralized, scalable, and real-time analytics for metro ridership, revenue, and energy data. Since metro operational data is generated by multiple independent systems, a layered architectural approach has been adopted to ensure efficient data flow, security, and flexibility.
The architecture is divided into five logical layers, each performing a distinct function.

3.2 Source Layer:
The source layer consists of various operational subsystems deployed within the metro ecosystem. These systems continuously generate raw data related to daily metro operations. Key data sources include:
Automatic Fare Collection (AFC) System – Generates passenger entry–exit data and revenue transactions
Train Operation Systems (TCMS / OCC) – Provides scheduled and actual train movement data
Energy Management System (EMS) – Supplies energy consumption data for traction and auxiliary services
Each subsystem operates independently and maintains its own data repository.

3.3 Integration Layer:
The integration layer is responsible for secure and controlled data extraction from metro subsystems. This layer uses RESTful APIs to fetch structured data in near real-time.
Key responsibilities of this layer include:
Secure authentication and authorization
Data transfer using standardized formats (JSON/XML)
Controlled access to operational data without impacting source systems
This layer ensures that data is retrieved reliably while maintaining system integrity and cybersecurity compliance.

3.4 Storage Layer:
The storage layer acts as an intermediate staging and consolidation repository. Data received from APIs is stored in a centralized relational database such as SQL Server or Oracle.
Functions of this layer include:
Temporary and historical data storage

Data cleansing and normalization
Schema alignment across multiple data sources
Ensuring data consistency and integrity
This layer enables efficient data modeling and supports historical trend analysis.

3.5 Analytics Layer:
The analytics layer is responsible for transforming raw data into meaningful insights. This layer is implemented using Power BI's data model and DAX calculations.
Major functions include:
Defining relationships between datasets
Creating calculated measures such as cumulative ridership, average passengers per day, and revenue indicators
Performing comparative analysis (scheduled vs actual operations)
Implementing trend analysis and forecasting logic
This layer converts data into analytical metrics that can be interpreted by decision-makers.

3.6 Visualization Layer:
The visualization layer represents the presentation and user interaction component of the system. Power BI dashboards and reports are developed to display insights in an intuitive and interactive manner.
Key features include:
Station-wise and time-based dashboards
Interactive filters and drill-down options
Graphical representation of trends and forecasts
Management-level KPI views
This layer enables stakeholders to explore data dynamically and supports informed operational and strategic decisions.

3.7 Proposed Implementation Snapshots
The proposed system follows a structured and modular methodology to design and implement an integrated analytics framework for metro operations. The methodology ensures reliable data acquisition, accurate analysis, and meaningful visualization to support operational and strategic decision-making.

3.8 Identification of Key Performance Indicators (KPIs):
The first step involved identifying critical performance indicators required for monitoring metro operations. These KPIs were finalized based on operational relevance and management reporting needs. The identified metrics include:
Station-wise and cumulative ridership
Average daily passenger count
Maximum and minimum daily ridership
Total and average revenue per day
Station-wise revenue contribution
Mode of payment distribution
Scheduled versus actual train services
Train availability and punctuality
Energy consumption trends
Month-on-Month (MoM) and Year-on-Year (YoY) growth
Ridership trend and forecast indicators
These KPIs form the analytical foundation of the dashboards presented in the proposed solution.

3.9 Data Source Identification and API-Based Integration:
Operational data is generated by multiple metro subsystems such as fare collection systems, train operations, and energy management systems. To ensure automated and secure data retrieval, RESTful APIs are used as the primary data integration mechanism.
The integration process includes:
Establishing secure API endpoints for each subsystem
Fetching structured data in defined intervals
Validating data completeness and consistency
This approach eliminates manual data extraction and supports near real-time analytics.

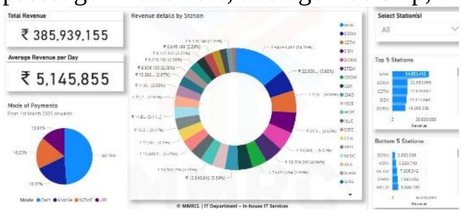
3.10 Data Staging and Storage Layer:
Fetched data is stored in a centralized staging database designed to support analytical processing. The staging layer acts as an intermediate repository between source systems and the analytics platform.

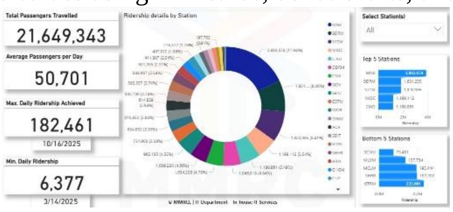
Key activities in this layer include:
Temporary storage of raw data Schema alignment across multiple data sources
Timestamp-based versioning for historical analysis
This design ensures scalability and allows future expansion to additional subsystems.

3.11 Data Cleaning, Transformation, and Normalization:
Before analytics and visualization, the data undergoes preprocessing to ensure accuracy and consistency. This step is critical due to variations in data formats and reporting intervals across subsystems.
The preprocessing process includes:
Removal of duplicate and incomplete records
Standardization of station codes and date formats
Aggregation of transactional data into daily, monthly, and yearly summaries
Validation through reconciliation with source reports
This ensures that calculated metrics such as total ridership, revenue, and averages are reliable.

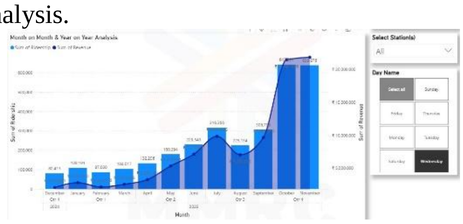
3.12 Data Modeling and Analytical Layer:
A relational data model is created within Power BI to support efficient analytics. Fact tables (ridership, revenue, train operations, energy consumption) are linked with dimension tables (date, station, payment mode).
Analytical measures are developed using DAX (Data Analysis Expressions), including:
Cumulative ridership and revenue
Average passengers and revenue per day
Maximum and minimum daily values
MoM and YoY growth calculations
Performance ratios for scheduled vs actual operations
This modeling layer enables dynamic filtering and drill-down analysis across dashboards.

3.13 Dashboard Development and Visualization
Interactive dashboards are developed using Power BI to present analytical insights in an intuitive and user-friendly manner. The dashboards include: Daily and Cumulative Ridership Dashboards: Display station-wise passenger movement, average ridership, and extreme values using KPI cards, donut charts, and barch arts.

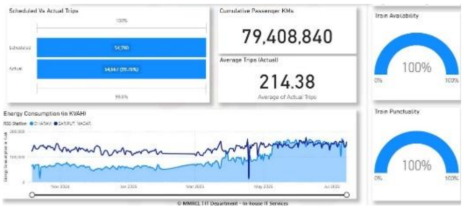




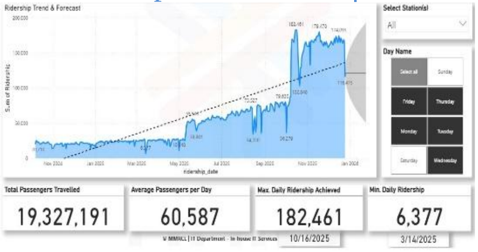
Revenue Analytics Dashboard: Shows total revenue, average revenue per day, station-wise revenue distribution, and payment mode analysis.



MoM and YoY Analysis Dashboard:
Visualizes month-wise and year-wise growth trends for ridership and revenue using combined bar and line charts.



Train and Energy Consumption Dashboard: Compares scheduled versus actual trips, train availability, punctuality, and energy consumption trends over time



Trend Analysis and Forecast Dashboard: Illustrates historical ridership trends and future projections using time-series forecasting models.Interactive slicers such as date range, station selection, and day type enable customized analysis for different stakeholders.

Trend Analysis and Forecasting:

Trend analysis is performed using historical ridership data to identify growth patterns and seasonal variations. Built-in Power BI forecasting features and time-series models are applied to estimate future ridership demand. The forecasting component supports:

Capacity planning
Service optimization
Demand-driven operational planning

This predictive capability enhances the decision-support value of the system.

Validation and Performance Evaluation:

The final stage involves validating dashboard outputs against source system reports to ensure accuracy. Key metrics such as total passengers, revenue figures, and operational counts are cross-verified.

Performance evaluation focuses on:

Reduction in manual reporting effort
Improvement in data visibility
Responsiveness of dashboards under multiple filters

Successful validation confirms the reliability and practical applicability of the proposed analytics framework.

4. LIMITATION

Practical Limitations:

Dependency on Source System Data Quality: The effectiveness of the analytics framework depends heavily on the accuracy, completeness, and consistency of data generated by metro operational subsystems. Any errors, delays, or inconsistencies in source systems such as fare collection, train operations, or energy monitoring directly impact the reliability of analytical outputs. API Availability and Performance Constraints: The proposed system relies on API-based data integration. Variations in API response time, temporary unavailability, or access restrictions can affect data refresh cycles. Additionally, high-frequency API calls may be subject to rate limits or throttling, which can restrict near real-time analytics. Limited Real-Time Capabilities: While the framework supports near real-time data updates, true real-time analytics may not be achievable due to latency in source systems, network delays, and scheduled data synchronization intervals. This may affect instantaneous monitoring during peak operational events. Infrastructure and Resource Constraints: The performance of the analytics system is influenced by the underlying infrastructure, including database capacity, network bandwidth, and Power BI service limitations. In large-scale deployments, increased data volume may require additional computational and storage resources. Security and Access Control Considerations: Access to operational and financial data requires strict security controls. Implementation of strong authentication, authorization, and encryption mechanisms may limit ease of integration and increase system complexity. Change Management and Operational Adoption: Successful utilization of the analytics platform depends on user adoption and organizational readiness. Resistance to change, lack of analytical skills, or insufficient training may limit effective use of the system by stakeholders.

Theoretical Limitations:

Assumptions in Data Aggregation: The analytical models assume that aggregated data accurately represents operational behavior. Aggregation at daily or monthly levels may obscure short-term fluctuations or anomalies that could be relevant for detailed operational analysis.

Simplified Analytical Relationships: The framework primarily relies on descriptive and trend-based analytics. Relationships between ridership, revenue, operations, and energy consumption are treated as correlated indicators rather than causally linked variables, which may limit deeper explanatory insights.

Forecasting Model Assumptions: Trend analysis and forecasting methods assume continuity in historical patterns. Unexpected events such as policy changes, service disruptions, or external socio-economic factors may significantly affect future outcomes, reducing forecast accuracy.

Limited Behavioral Representation: Passenger behavior is analyzed using aggregated historical data and does not explicitly account for behavioral factors such as traveler preferences, special events, or sudden demand shifts. This limits the ability to model complex human decision-making dynamics.

Scope Restriction to Quantitative Metrics: The proposed framework focuses on quantitative operational metrics and does not incorporate qualitative factors such as passenger satisfaction, service perception, or staff performance, which also influence overall system effectiveness.

5. CONCLUSION & FUTURE WORK

This study presents the design and implementation of a data-driven analytics framework for monitoring and analyzing metro rail operations using a business intelligence platform. The proposed system successfully integrates data from multiple operational subsystems through secure API-based mechanisms and transforms heterogeneous datasets into structured analytical models. By leveraging interactive dashboards, the framework enables comprehensive analysis of ridership patterns, revenue performance, train operations, and energy consumption at multiple levels of granularity. The developed solution provides both descriptive and trend-based analytical capabilities, supporting daily monitoring, station-wise performance comparison, and temporal analysis through Month-on-Month and Year-on-Year evaluations. The inclusion of trend analysis and forecasting enhances the system's ability to support proactive planning and informed decision-making. Overall, the framework demonstrates how operational data can be effectively utilized to improve visibility, efficiency, and strategic oversight in modern urban rail systems. Future extensions of this framework may incorporate advanced time-series and machine learning–based predictive models to improve the accuracy of ridership and revenue forecasting. Real-time analytics can be enabled through streaming data ingestion to support dynamic operational monitoring and rapid decision-making. Additionally, anomaly detection and predictive maintenance techniques may be integrated to identify irregular patterns in operations and energy usage. Cloud-native deployment and integration of external contextual datasets can further enhance scalability, resilience, and analytical depth of the systems.

Acknowledgment

The author is also thankful to the faculty members of the Department of Computer Engineering for their academic support and for providing the necessary infrastructure and learning environment to carry out this research. Appreciation is extended to the academic institution for facilitating access to computational resources and analytical tools required for the successful implementation of the proposed analytics framework. Additionally, the author acknowledges the developers and open-source communities behind data analytics and business intelligence platforms that supported this work, including Power BI and related data processing tools. Their contributions played a significant role in enabling the development and visualization of the analytical models presented in this project. Finally, heartfelt thanks are extended to family members and peers for their constant motivation and support during the course of this academic endeavor.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

6. REFERENCES

[1] Y. He, Y. Zhao, and K. L. Tsui, "An analysis of factors influencing metro station ridership: Insights from Taipei Metro," in Proc. 21st Int. Conf. Intelligent Transportation Systems (ITSC), Maui, HI, USA, Nov. 4–7, 2018, pp. 1598–1603, doi: https://www.researchgate.net/publication/329584082

[2] J. Krembsler, S. Spiegelberg, R. Hasenfelder, N. L. Kämpf, T. Winter, N. Winter, and R. Knapp, "Fare revenue forecast in public transport: A comparative case study," Research in Transportation Economics, vol. 102, Feb. 2024, Art. no. 101293, doi: https://www.researchgate.net/publication/381870203

[3] P. S. Krishnamoorthy and P. Venugopal, "Enhancing public transportation through advanced technologies: A comprehensive review," IEEE Access, vol. 13, pp. 185533–185557, Oct. 2025, doi: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=11219083

[4] L. Zhu, F. R. Yu, Y. Wang, and T. Tang, "Big data analytics in intelligent transportation systems: A survey," *IEEE Access*, vol. 6, pp. 366–3567, 2018. https://www.researchgate.net/publication/324712011_Big_Data_Analytics_in_Intelligent_Transportation_Systems_A_Survey https://www.researchgate.net/publication/324712011_Big_Data_Analytics_in_Intelligent_Transportation_Systems_A_Survey

[5] P. McMahon, T. Zhang, and R. Dwight, "Incorporation of big data analytics into railway asset management — A review," *IEEE Access*, 2020.

https://www.researchgate.net/publication/345780688_Incorporation_of_Big_Data_Analytics_into_Railway_Asset_Management_A_Review

[6] Y. He, Y. Zhao, and K. L. Tsui, "An adapted geographically weighted lasso (Ada-GWL) model for estimating metro ridership," *IEEE Access*, vol. 7, pp. 16552–16567, 2019. https://www.researchgate.net/publication/332168783_An_Adapted_Geographically_Weighted_LassoAda-GWL_model_for_estimating_metro_ridership

[7] N. Cherrier, J. Dubois, and M. Berbineau, "Context-aware automated passenger counting data denoising," *IEEE Access*, vol. 12, pp. 24511–24524, 2024. https://arxiv.org/pdf/2402.08688

[8] R. Sun, S. Schaefer, and C. F. Daganzo, "Transit-Gym: A simulation and evaluation engine for analysis of public transit systems," *IEEE Access*, vol. 9, pp. 103421–103434, 2021. https://www.researchgate.net/publication/355310040_TRANSIT-GYM_A_Simulation_and_Evaluation_Engine_for_Analysis_of_Bus_Transit_Systems

[9] L. Liu, Y. Wang, Z. Li, and J. Ye, "Online metro origin–destination prediction via heterogeneous information aggregation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 8145–8156, 2022. https://www.researchgate.net/publication/360986427_Online_Metro_Origin-Destination_Prediction_via_Heterogeneous_Information_Aggregation

[10] IEEE Intelligent Transportation Systems Society, "Special Issue on Big Data Technology and Applications in Intelligent Transportation," *IEEE Transactions on Intelligent Transportation Systems*, 2021. https://www.researchgate.net/publication/378224755_Special_Issue_on_Transportation_Big_Data_and_Its_Applications

[11] V. M. Urbano, M. Arena, and G. Azzone, "Big data for decision-making in public transport management," *Research in Transportation Business & Management*, vol. 48, 2025. https://www.researchgate.net/publication/388707837_Big_data_for_decision-making_in_public_transport_management_A_comparison_of_different_data_sources

[12] H. Ayan, "Investigating the factors that influence the ridership of urban rail systems," *Urban Science*, vol. 9, no. 1, pp. 1–18, 2025. https://www.researchgate.net/publication/389440039_Investigating_the_Factors_That_Influence_the_Ridership_of_Light_Rail_Transit_Systems_Using_Thematic_Analysis_of_Academic_Literature

[13] M. Wasesa, S. Handayani, and R. Hidayat, "Electronic ticket machine data analytics for public transport planning," in *Proc. IEEE Int. Conf. on Smart Transportation*, 2024 https://icid.sunankalijaga.org/downloads/Proceeding%20ICID%202018.pdf