PUBLIC TRANSPORT DELAYS

Big Data and Business Intelligence

Karan Desai
S3529053

Teesside University
Middlesbrough, England,
United Kingdom
January, 06 2026

Module Leaders: Prof. Annalisa Occhipinti, Dr. Chaimaa Tarzi

TABLE OF CONTENTS

1. Executive Summary

Public transport systems operate in complex environments where service reliability is affected by multiple factors such as traffic congestion, peak-hour demand, and weather conditions. Delays in public transport services negatively impact passenger satisfaction, operational efficiency, and overall trust in the transport network. As transport systems generate large volumes of operational data, Business Intelligence (BI) tools can be used to transform this data into meaningful insights that support better decision-making.

This project focuses on analysing public transport delay data using Microsoft Power BI. The dataset contains trip-level information, including transport mode, delay status, traffic congestion levels, weather conditions, and peak-hour classification. The main objective of the analysis is to identify patterns in delays and understand the key factors that contribute to service unreliability.

An interactive Power BI dashboard was developed to visualise the data and present key performance indicators. The analysis shows that delays occur more frequently during peak hours, particularly when traffic congestion levels are high. Bus and Train services experience the highest number of delays, while adverse weather conditions such as rain and fog further increase the likelihood of delays.

Based on the findings, the report recommends increasing service frequency during peak periods, prioritising reliability improvements for the most affected transport modes, and using traffic and weather data to support operational planning. Overall, this project demonstrates how Business Intelligence dashboards can support data-driven decisions and help improve public transport service reliability.

2. Introduction

Public transport systems play a vital role in supporting daily commuting and reducing traffic congestion in urban areas. Reliable and punctual services are essential for maintaining passenger trust and ensuring efficient transport operations. However, delays remain a persistent challenge for transport providers and can lead to dissatisfaction, operational inefficiencies, and increased costs.

Public transport delays can occur due to a variety of factors, including traffic congestion, adverse weather conditions, peak travel demand, and operational limitations. Understanding how these factors contribute to delays is important for identifying problem areas and improving overall service performance. Without effective analysis, transport operators may struggle to prioritise interventions and allocate resources efficiently.

This project focuses on analysing public transport delays using Business Intelligence techniques. The primary objective is to examine delay patterns and identify how different operational factors influence service reliability. The analysis aims to support better decision-making by providing clear insights into when delays are most likely to occur and which conditions contribute most significantly to service disruption.

The dataset used in this project consists of **2,001 records and 24 attributes** related to public transport operations. It includes information such as transport mode, weather condition, peak hour indicator, traffic congestion index, and delay status. The dataset is appropriate for Business Intelligence analysis as it captures multiple factors that can influence transport performance.

Power BI was selected as the primary analysis tool due to its strong data visualisation and interactive reporting capabilities. By transforming raw data into an interactive dashboard, Power BI enables users to explore delay trends, compare operational conditions, and gain insights that support performance monitoring and operational improvement.

From a business perspective, public transport operators aim to improve service reliability while managing operational costs and passenger expectations. Delays can lead to customer dissatisfaction, increased operational pressure, and reputational impact. By analysing operational delay data, transport decision-makers can better understand performance issues and implement targeted improvements that enhance overall service quality.

2.1 Problem Definition

Public transport systems play a critical role in supporting daily mobility within urban areas. However, delays in public transport services can negatively impact passenger satisfaction, operational efficiency, and overall system reliability. Transport authorities need clear, data-driven insights to understand when, where, and why delays occur in order to improve service performance and planning decisions.

The primary problem addressed in this study is the lack of visibility into the key factors contributing to public transport delays. Delays may be influenced by multiple operational and external conditions such as weather, peak travel hours, traffic congestion, and transport mode. Without effective analysis, it becomes difficult for decision-makers to identify patterns and prioritise corrective actions.

This project aims to analyse public transport delay data using Business Intelligence techniques to uncover meaningful patterns and trends. By applying data visualisation and dashboard design, the study seeks to answer key operational questions such as which conditions lead to higher delays and how different factors interact with one another.

The insights generated from this analysis are intended to support operational performance improvement. The final dashboard provides an interactive platform that enables stakeholders to explore delays dynamically and supports evidence-based decision-making for improving service reliability.

2.2 Data Source and Description

The dataset used in this project is provided in CSV format and contains **2,001 records with 24 attributes** related to public transport operations. Each record represents a transport trip and includes both operational and contextual information relevant to delay analysis.

Key attributes in the dataset include transport mode (Bus, Metro, Train, Tram), weather conditions (Clear, Cloudy, Fog, Rain), peak hour indicators, traffic congestion index, scheduled and actual arrival or departure times, and delay status. These attributes allow for a comprehensive examination of factors affecting transport delays from multiple perspectives.

The dataset is well-suited for Business Intelligence analysis due to its structured format and balanced size. With a sufficient number of records and relevant variables, it enables effective aggregation, filtering, and visual exploration within Power BI without being overly complex.

Overall, this dataset provides a strong foundation for analysing public transport delays and supports the creation of meaningful visualisations and dashboards that align with real-world operational challenges.

3. Data Pre-processing

Data pre-processing is a critical stage in Business Intelligence projects, as the quality of insights depends heavily on the quality of the underlying data. Effective data preparation ensures that visualisations and analytical results are accurate, consistent, and reliable.

In this project, data pre-processing was carried out using **Power BI Power Query**. The aim was to prepare a clean and structured dataset that could support meaningful analysis of public transport delays. Several steps were applied, including data cleaning, feature engineering, and data normalisation.

This section explains the key pre-processing steps performed on the dataset and justifies how each step contributed to improving data reliability and analytical accuracy.

3.1 Data Cleaning

The first step in the pre-processing phase involved cleaning the dataset to remove errors and inconsistencies. This ensured that the data used for analysis was accurate and suitable for generating reliable insights.

Duplicate records were checked and removed to prevent double-counting of trips. Missing or null values were reviewed, and records with incomplete critical information (such as missing delay indicators or transport mode) were handled appropriately to maintain data integrity.

Column names were also reviewed and renamed where necessary to improve clarity and readability within Power BI. Additionally, incorrect data types were corrected, for example ensuring that numeric fields such as delays, congestion index, and temperature values were stored as numerical data rather than text.

3.2 Feature Engineering and Data Transformation

After cleaning, feature engineering was applied to enhance the dataset and support more meaningful analysis. New derived fields were created from existing attributes to allow better categorisation and aggregation of data.

One important transformation involved creating or confirming a **Peak Hour indicator**, which categorises trips into peak and off-peak periods. This enabled the analysis of how congestion during peak hours impacts delays. Similarly, delay indicators were standardised to ensure consistency across all visualisations.

Additional transformations included formatting date and time fields to support time-based analysis and ensuring categorical fields such as weather condition and transport mode were correctly labelled and grouped.

These transformations improved the analytical depth of the dataset and allowed more insightful comparisons across different operational conditions.

3.3 Data Normalisation

Data normalisation was applied to ensure consistency across numerical attributes and to support accurate comparison within visualisations. Fields such as traffic congestion index and delay values were reviewed to confirm they followed a consistent scale and unit of measurement.

Normalisation also helped reduce the impact of extreme values on visual outputs, particularly in charts comparing traffic congestion against total delays. By maintaining consistent data ranges and formats, the dashboard visuals became easier to interpret and more reliable for decision-making.

This step ensured that all numerical values contributed fairly to the analysis and supported meaningful visual comparison across different conditions and transport modes.

4. Data Visualisation

Data visualisation plays a key role in Business Intelligence as it transforms raw data into meaningful insights that are easy to understand and interpret. In this project, visualisation techniques were used to explore patterns, trends, and relationships within the public transport delay dataset before designing the final interactive dashboard.

Initially, exploratory visualisations were created to understand how delays vary across different conditions such as weather, peak hours, transport modes, and traffic congestion levels. Bar charts were used to compare total delays across categorical variables, including weather conditions and transport types, as they provide a clear and effective way to identify differences between groups. These visuals helped highlight which factors contribute most to delays.

Line and column charts were also utilised to analyse time-based patterns, such as the impact of peak and off-peak hours on delays. This made it easier to observe fluctuations in delays during different periods of the day. Additionally, traffic congestion was visualised using binned values to better understand how increasing congestion levels influence total delays.

The insights gained from this visual exploration informed the design of the final Power BI dashboard. Each visual was selected to clearly answer key business questions and support decision-making. By using appropriate chart types and consistent visual formatting, the analysis ensures that trends and problem areas within public transport operations can be quickly identified by stakeholders.

These exploratory visualisations helped identify key trends and relationships in the data and directly informed the design of the final interactive Power BI dashboard.

5. Dashboard Design and Explanation

The dashboard was designed to support both operational analysis and managerial decision-making. Visuals were selected to directly address the defined business questions, while maintaining a clear and simple layout suitable for non-technical users. The inclusion of key performance indicators and interactive filters allows stakeholders to quickly assess performance and explore specific operational scenarios.

The dashboard integrates multiple visuals on a single page to allow users to analyse delays from different perspectives while using filters and slicers to focus on specific scenarios. Key performance indicators (KPIs) such as total trips analysed and total delayed trips are included to give a quick summary of overall system performance. Each visual is explained in the following subsections.

5.1 Delays by Weather Conditions

This Visual displays the total number of delayed trips under different weather conditions such as clear, cloudy, fog, and rain. A bar chart was selected to allow easy comparison between weather categories.

The visual helps identify how environmental conditions affect transport reliability. From the analysis, adverse weather conditions such as rain and fog tend to show higher delay values compared to clear weather. This indicates that weather plays a significant role in operational performance and should be considered when planning schedules and allocating resources.

5.2 Impact of Peak Hours on Delays

This visual illustrates the relationship between peak hours and total delays by comparing peak and off-peak periods. Peak hour is represented as a binary indicator (0 = off-peak, 1 = peak).

The chart clearly shows that delays are higher during peak hours. This reflects increased passenger demand and network congestion during busy times of the day. The insight highlights the need for better capacity planning and service optimisation during peak periods to reduce delays and improve punctuality.

5.3 Delays by Transport Mode

This visual analyses total delays across different transport modes, including bus, metro, train, and tram. A bar chart was used to compare delay levels across these categories.

The results indicate that certain transport modes experience more delays than others, with buses and trains generally showing higher delay counts. This may be due to factors such as road congestion for buses and scheduling complexity for trains. The visual helps identify which transport services require closer operational monitoring and targeted improvements.

5.4 Traffic Congestion vs Total Delays

This visual explores the relationship between traffic congestion levels and total delays using binned traffic congestion index values. A column chart was chosen to observe how delays change as congestion increases.

The analysis shows a clear trend where higher congestion levels are associated with increased delays. This confirms that traffic congestion is a major contributor to service disruption, particularly for surface transport modes. The insight supports the need for congestion management strategies and real-time traffic monitoring.

5.5 Dashboard Interactivity (Filters and Slicers)

The dashboard includes interactive slicers for transport mode, weather condition, and peak hour. These slicers allow users to dynamically filter the data and analyse delays under specific conditions.

For example, users can examine delays for a particular transport mode during peak hours or assess how delays change under specific weather conditions. This interactivity enhances the usability of the dashboard and enables stakeholders to perform scenario-based analysis, making the dashboard a practical tool for operational decision-making.

6. Analysis and Key Insights

The analysis of the dashboard reveals several important patterns related to public transport delays. One of the most significant findings is the impact of peak hours on delay frequency. Delays are noticeably higher during peak periods compared to off-peak hours, indicating that increased passenger demand and network pressure contribute to reduced service reliability.

Weather conditions also play an important role in operational performance. Adverse weather such as rain and fog is associated with a higher number of delays when compared to clear weather. This suggests that environmental factors affect travel speed, safety, and scheduling efficiency, leading to service disruptions.

Transport mode analysis shows that buses and trains experience more delays than other modes such as metro and tram. This may be due to greater exposure to traffic congestion for buses and operational complexity for trains. These findings highlight the need for targeted performance improvements based on transport type.

Finally, the relationship between traffic congestion and total delays demonstrates a clear positive trend, where higher congestion levels lead to increased delays. This confirms that congestion is a key driver of delays and reinforces the importance of traffic management and data-driven scheduling decisions.

7. Recommendations

Based on the analysis and insights obtained from the Power BI dashboard, several practical recommendations can be proposed to improve public transport operational performance and reduce delays.

Firstly, additional capacity and service frequency should be considered during peak hours. Since delays are significantly higher during peak periods, increasing the number of services or optimising schedules during these times can help reduce congestion-related delays and improve punctuality.

Secondly, targeted improvements should be prioritised for transport modes that experience higher delays, particularly buses and trains. For buses, measures such as dedicated bus lanes and traffic signal prioritisation could help reduce the impact of road congestion. For trains, improved timetable management and operational monitoring may enhance service reliability.

Thirdly, traffic congestion data should be actively used in operational planning. As higher congestion levels are strongly linked to increased delays, integrating real-time traffic monitoring into scheduling decisions could support more responsive and adaptive transport operations.

Finally, weather conditions should be considered in planning and resource allocation. During adverse weather such as rain or fog, additional operational measures, such as adjusted schedules or increased staffing, may help minimise disruptions and maintain service reliability.

8. Conclusion

This project applied Business Intelligence techniques to analyse public transport delays and identify key factors affecting service reliability. Using a structured dataset and an interactive Power BI dashboard, the study explored how delays vary across weather conditions, peak hours, transport modes, and traffic congestion levels.

The analysis demonstrated that delays are more frequent during peak periods, increase with higher traffic congestion, and differ across transport modes. Weather conditions were also shown to influence delay patterns, particularly during adverse conditions such as rain and fog. These findings highlight the importance of considering multiple operational and external factors when evaluating transport performance.

The Power BI dashboard developed in this project provides a clear and interactive platform for analysing delay patterns and monitoring performance. By combining key performance indicators, visual comparisons, and interactive filters, the dashboard supports effective exploration of data and facilitates data-driven decision-making.

This project also demonstrates the effectiveness of Business Intelligence tools such as Power BI in transforming operational transport data into actionable insights. By combining data preparation, visualisation, and interactivity, the dashboard supports continuous monitoring and evidence-based decision-making within public transport operations.

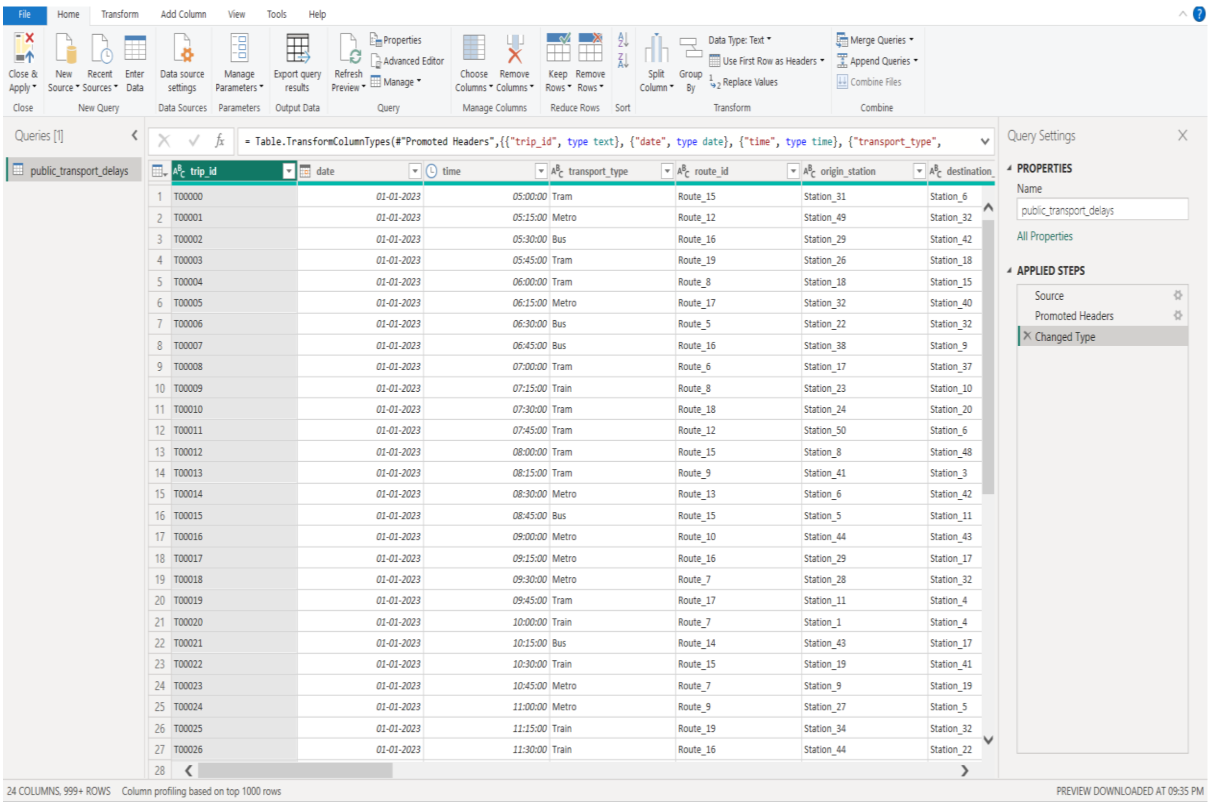
9. Appendix

The appendix provides supporting evidence for the analysis carried out in this project. It includes screenshots related to data pre-processing and dashboard design, which demonstrate how the dataset was prepared and how the final Power BI dashboard was developed. The appendix is intended to support the main report without interrupting its flow.

9.1 Data Pre-processing Screenshots

This section contains screenshots taken from the Power Query Editor in Power BI to demonstrate the data pre-processing steps applied to the dataset. The screenshots show the dataset after cleaning and preparation, including promoted headers, corrected data types, and applied transformation steps.

These screenshots provide evidence that the raw CSV data was successfully transformed into a structured and analysis-ready format. No step-by-step clicks are shown, as the focus is on demonstrating the final cleaned dataset used for analysis.



(Figure 9.1: Cleaned dataset in Power Query Editor after applying data preparation steps)

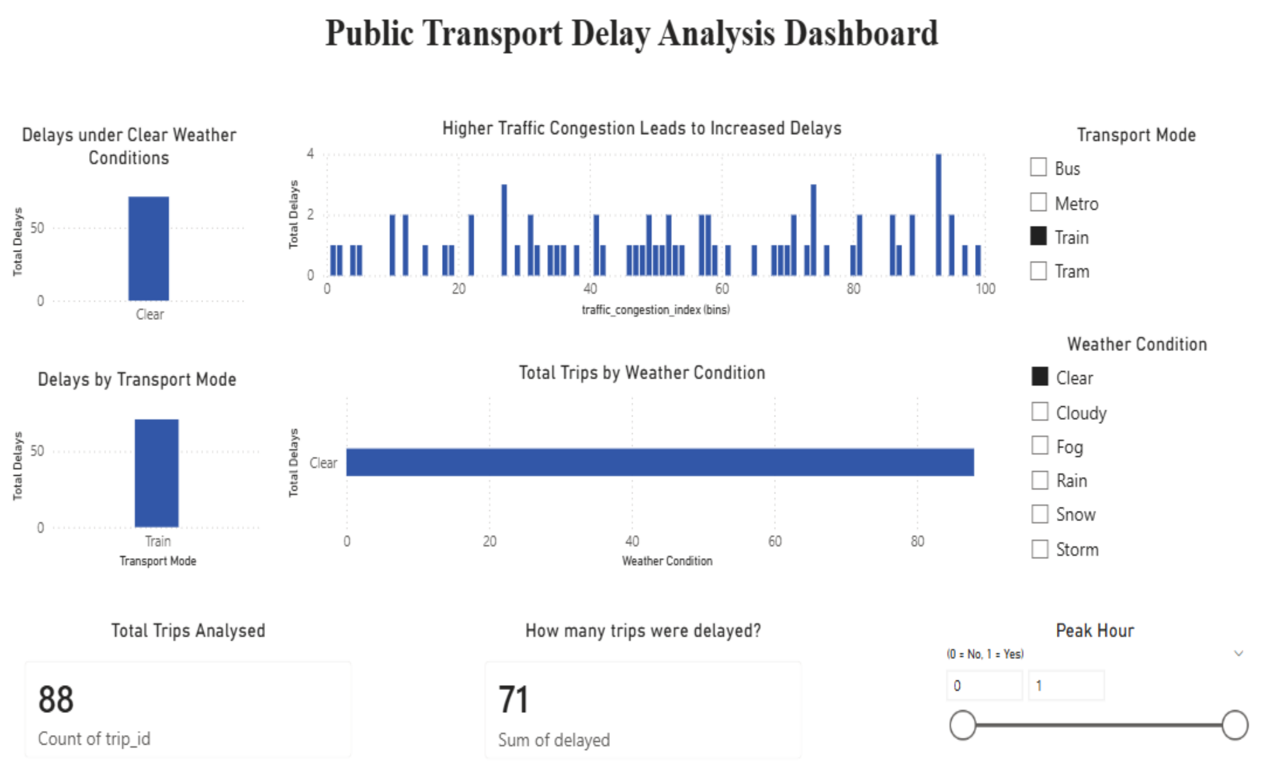
9.2 Dashboard Screenshots

This section presents screenshots of the final Power BI dashboard developed to analyse public transport delays. The dashboard consists of **two pages**, each serving a different analytical purpose.

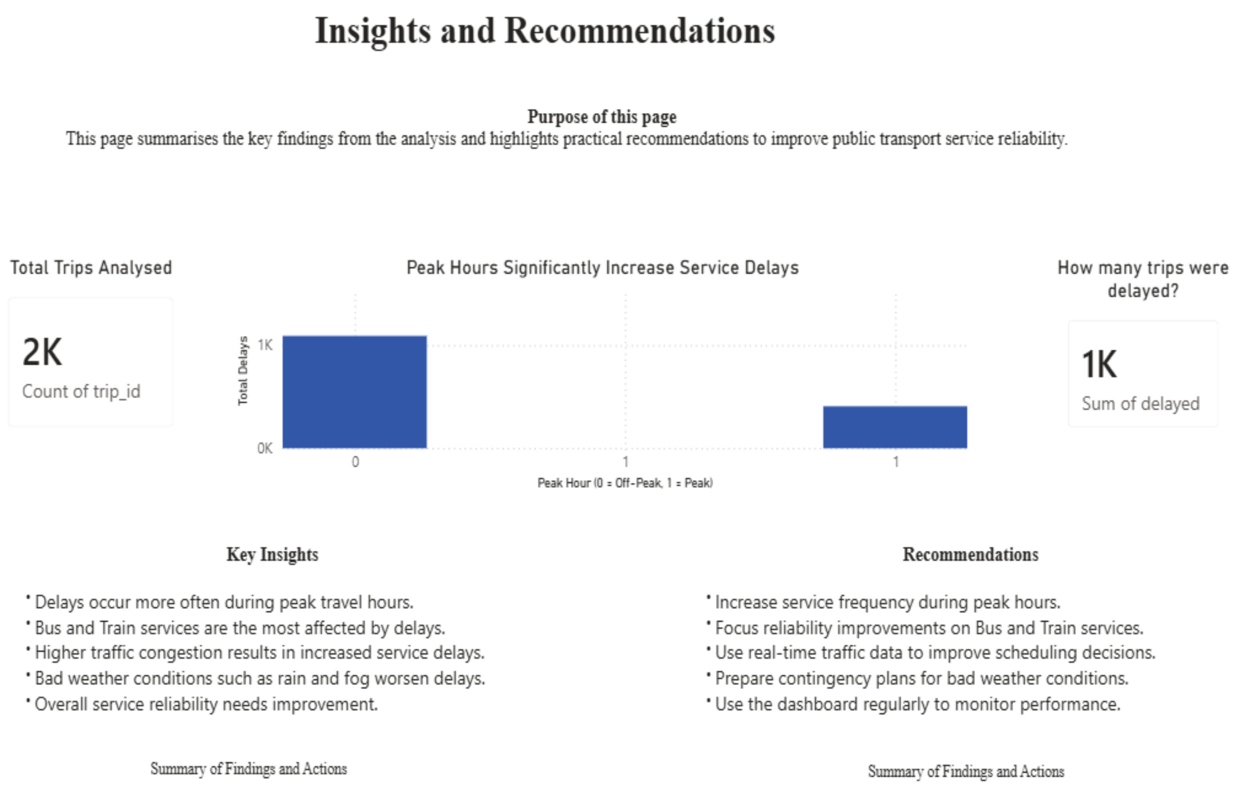
Dashboard Page 1 focuses on detailed analysis and displays key performance indicators showing **Total Trips Analysed (88)** and **Total Delayed Trips (71)** based on the applied filters. It includes visualisations of delays by transport mode, weather conditions, traffic congestion, and peak hours, supported by interactive slicers for data exploration.

Dashboard Page 2 provides a summary view, showing overall KPIs of **Total Trips Analysed (2K)** and **Total Delayed Trips (1K)** for the full dataset. This page highlights the impact of peak hours on delays and presents consolidated insights and recommendations to improve service reliability.

Overall, the dashboard demonstrates effective use of Power BI visuals, KPIs, and interactivity to support both detailed analysis and high-level decision-making.



(Figure 9.2.1: Public Transport Delay Analysis Dashboard (Main Dashboard))



(Figure 9.2.2: Insights and Recommendations Dashboard)

9.3 Additional Supporting Information

This section includes any additional material that supports the analysis presented in the report. This may include brief notes on calculated measures, data field descriptions, or formatting decisions used within Power BI.

The information provided in this section is supplementary and is included to enhance clarity and transparency, while ensuring that the main report remains concise and focused.

10. Self-Assessment

NO.	Report Section	Description	Grade (0–100)
1	Report Structure	The report follows the ICA specification, is clearly structured, and includes all required sections in a logical order.	92
2	Data Pre-processing and Data Modelling	Data cleaning and transformation were performed using Power Query to prepare the dataset for analysis and dashboard development.	92
3	DAX and M Language	DAX measures and M language transformations were applied to support KPI calculations and data preparation.	90
4	Dashboard Design	The dashboard includes two interactive pages with relevant visuals, filters, and slicers aligned to the business problem.	90
5	Average	Overall average score across all assessed sections.	91