CSC8499 Individual Project: Metro Data Insights Website

Aditi Jain

MSc in Advanced Computer Science,
School of Computing Science, Newcastle University.
a.jain9@ncl.ac.uk

Abstract. Metro is a service which people of Newcastle use on a regular basis. Analysis done based on surveys and interviews for this project, showed that there is a lot of confusion related to metros, which can be resolved by having a dedicated metro insights website. User research informed the prototype of a new website which aims to provide information about the metro like the schedule, metro staff availability, frequency of metro, crowd etc. All these insights are generated by fetching significant features from historical metro data and training those features to become meaningful insights using machine learning. This study represents a significant step towards leveraging technology to improve public transportation services and sets a precedent for similar initiatives in other cities. A key component of this study is the focus on Human-Computer Interaction (HCI). The insights are designed with user-centered principles to ensure ease of use and accessibility. Through user testing and feedback from users, the interface has been optimized to provide a seamless experience. This includes clear navigation, responsive design, and interactive features that allow users to personalize their experience.

1 Introduction

Public transport plays an important role in in transportation system in Newcastle, owing to an ample number of international people requiring an economical and environment friendly alternative to private vehicles. The Newcastle metro is an integral part of the transportation infrastructure, serving a large number of commuters daily. “In 2018–19, an estimated 36.4 million passenger journeys were made on the Metro, making it the third-most used light rail network in the United Kingdom after London's Docklands Light Railway (121.8 million passenger journeys) and Manchester’s Metrolink (43.7 million passenger journeys)” [1].
Despite its significance and availability, a lot of commuters often face challenges in getting access to accurate and timely information [2].

With growing dependency on digital systems – mobile applications and websites, there is an opportunity to enhance metro journey experience by leveraging technology. The digital platform, designed to provide real-time, data-driven insights, will significantly enhance the user travel experience. By leveraging data to offer meaningful, intuitive, and interactive features, the platform ensures that users receive accurate, personalized information tailored to their specific needs [3]. This data-driven approach allows for the continuous optimization of the service, making travel planning more efficient and responsive to real-time conditions, ultimately leading to a smoother and more satisfying journey for all users [4].

Furthermore, having a user-centered design (UCD) has helped bridge the gap between complex and vast metro data and the user experience [5]. By focusing on HCI principles, the platform was designed to ensure that the interaction between users and the system is intuitive, efficient, and enjoyable [6]. HCI emphasizes the importance of designing interfaces that are not only functional but also easy to navigate, reducing the cognitive load on users [7]. This approach makes planning journeys ahead of time, purchasing the correct ticket, finding the right train and platform, and accessing all necessary information significantly easier, even in the absence of metro staff to assist users. By prioritizing both UCD and HCI, the platform is able to create a seamless and user-friendly experience that meets the diverse needs of all metro users.

1.1 Motivation

The need for reliable and easily accessible information in public transportation is critical. Current information spreading methods for Newcastle's Metro service are uneven and often outdated, leading to user dissatisfaction. The development of a centralized platform that offers thorough and up-to-date information can greatly enhance the commuter experience, making daily travel more seamless, reliable, and less stressful.

Moreover, historical data may be transformed into useful insights as a result of advancements in machine learning and data analytics. Predictive models and historical data analysis allow for the prediction of service conditions and the delivery of precise, up-to-date information to users. This project is motivated by the desire to harness these technological advancements to address the identified gaps and enhance the overall efficiency of the Metro service.

1.2 Aim

The aim of this project is to create a user-friendly website that consolidates essential metro service information, provides real-time updates, and ultimately enhances the overall user experience.

1.3 Objectives

Below Objectives were taken in account while developing this project:

- *Conduct User Research*: Understand problems faced by users, and gain feedback for the same before starting the development so that the website design is user centred.
- *Data Pipeline Creation:* Develop a robust data pipeline that includes stages for data pre-processing, transformation, training, and visualization. Also, ensure the pipeline is efficient and capable of handling real-time data updates.
- *Meaningful Visualizations Generation:* Generate meaningful visuals and tables providing information about the metro.
- *Leverage Machine Learning for Predictive Insights:* Use historical Metro data to generate real-time insights and predictions through machine learning techniques.
- *Develop a Dedicated Metro Insights Website Prototype*: Develop a centralized platform prototype that provides comprehensive information, insights and predictions of crowd levels and delays in regard to the Metro service.
- *Focus on Human-Computer Interaction (HCI):* Design the website with user-centred principles to ensure ease of use and accessibility. The intractability and detailed information access of the website provide users a wholesome experience.
- *Enhance User Experience:* Provide accurate, timely, and personalized information to reduce confusion and improve the commuting experience. Additionally, users get alerts and notifications for real-time updates to their trip.
- *Evaluation:* Test the prototype and efficiency of the visualizations on the website. To achieve this, a persona was set up and a user journey map was created to help fellow students step into the intended user's shoes and provide valuable feedback. Furthermore, in-person usability testing was performed, and results were recorded for the same.

2 Background

To understand the current state of Newcastle metro information systems and to identify gaps that the project aims to fill, review of existing researches and systems was done. Upon reviewing existing researches, systems and getting feedback from people about metros and public transportation system, it can be said that in spite of millions of passenger travelling in Newcastle Metro annually, there is still a huge opportunity to improve the information dissemination process.

2.1 Initial Research

Reviewing various research papers on existing systems for enhancing the metro journey experience presented an opportunity to learn several key insights, validating that a user-centric, data-driven approach is an effective way to develop this project. One noteworthy finding was how the ‘spatial characteristics’ of underground metro stations influence passengers’ comfort and safety, as shown by research in Warsaw which

identified key architectural features impacting user experience [8]. Additionally, study on travel mode choices in cities like Doha presented the factors affecting commuters' decisions, such as socioeconomic characteristics and trip conditions, which is derived using advanced statistical and machine learning models [9]. Likewise, research on optimizing metro operations through timetable scheduling and passenger flow control, as conducted in Beijing, highlighted the importance of data management to reduce congestion and improve service quality [10]. Moreover, research on transit performance measures using data-driven methods for reliability improvement, as demonstrated in study from Minnesota, highlighted the importance of addressing traffic delays and boosting operational reliability [11]. After reviewing existing systems and research, it was deemed pertinent to develop a robust data-driven website aimed at enhancing metro operations and user satisfaction.

2.2 Public Transportation Systems

Thorough reading into various studies show the positive impact of real-time information systems on public transport. As seen in a study, conducted on the OneBusAway real-time transit information system in Seattle showed that providing real-time arrival information reduced passenger unease and increased satisfaction with the transport service [12]. Another research on New York City's Bus Time system showed that real-time information lowered perceived wait times and improved passenger satisfaction [13]. Similarly, a study on the effect of real-time bus information on the perceptions and behaviour of transit riders in Seattle, found that real-time information systems can increase the prospect of using public transportation by improving the reliability and convenience of the service [14]. These studies clearly highlight the advantage of a real-time information systems in enhancing the user experience in public transportation systems.

2.3 Existing Metro Information Systems

Several cities have implemented advanced metro information systems that provide real-time updates and predictive insights. For instance, the Seoul Metro uses a system known as the Seoul Metropolitan Subway Operation Information System (TOPIS). To give passengers accurate and timely information, this system integrates live data from many sources, such as train schedules, passenger counts, and service updates. Through the reduction of wait times and provision of precise train timetables, the installation of TOPIS resulted in a considerable improvement in passenger satisfaction and operational efficiency [15].

The New York City Subway uses an app called MTA Subway Time, which forecasts congestion levels based on past data and offers real-time service updates. The travel experience has been improved by this program, which offers trustworthy information on anticipated crowd sizes, service modifications, and train arrivals. Passengers are now able to organise their trips more efficiently and conveniently thanks to this [16].

In case of Newcastle, Metro information platforms like Traveline.info and Nexus.org.uk are currently available. While they offer valuable information for

Newcastle's transport system, there is an opportunity to enhance the user experience by focusing specifically on metro services. Traveline.info provides basic features such as ticket booking and metro timetables [17], but can be further enhanced by adding more comprehensive details. Nexus.org.uk offers a broader range of information, yet it could benefit from the inclusion of real-time updates and predictive insights. The metro insights project aims to fill these gaps by providing a dedicated platform for metro services that includes real-time data, predictive analytics, and a more user-friendly interface.

2.4 Human-Computer Interaction (HCI)

Creating user-centric applications requires an understanding of HCI principles. An interface that has been well-designed should be easy to use, accessible, and intuitive [18]. Studies have indicated that integrating HCI concepts can greatly improve information systems' usability and efficacy. One study shows that by using usability engineering, which involves systematically identifying and addressing usability issues early in the design process, leads to improved user experience [19]. Furthermore, creating interfaces with user goals and habits in mind might result in more productive and fulfilling interactions, per a study by Shneiderman and Plaisant [20]. A different study stressed the value of user feedback and iterative design in developing interfaces that are both practical and pleasurable to use [21]. These studies demonstrate the critical role of HCI principles in enhancing the usability and effectiveness of information systems.

3 Survey and Interview Findings

To incorporate user feedback and understand the Metro staff's perspective on operations, surveys were conducted with international Metro users in Newcastle and interviews were held with Metro staff. The survey was anonymous, and the responses have been utilized to guide the development of the project to be focused on user research and feedback.

3.1 Survey Methodology

The targeted demographic for the survey were the international passengers. They were specifically chosen because they are new to using the Newcastle metro and have experience with metro services in other countries, making them well-suited to identify challenges in the metro information system. The survey questions were designed keeping user touchpoints in mind, from locating the metro station to arrival at destination [22]. The questions capture detailed information on the challenges faced by these users during each step of their travels and their specific needs for metro-related information.

3.2 Key Findings from the Surveys

Out of 20 participants, only one had difficulty finding the closest metro station, likely due to the availability of focused navigation tools like Google Maps. However, once at the station, 57.9% of passengers struggled to find the correct platform to board their metro. Additionally, 42.1% of respondents felt confused about which zone or type of ticket to purchase, citing poor signage and display boards as the main reasons for their confusion. Once inside the station, 42.2% of participants faced challenges in locating the correct metro or determining the right platform and time for their departure.

The responses indicated that a dedicated metro insights website would be highly beneficial for passengers to plan their travels more efficiently. Participants shared various thoughts, such as: “It could make it easier to find information,” and “I’ve had difficulty finding out about delays or cancellations in the past; an app or website could help with this.” Another respondent noted, “Getting reception in the underground would make a huge difference, and apps could guide us to the station and show where to get off.” Others mentioned the challenges of different routes running on the same platform, with one person saying, “It can be confusing to know if the train on the platform is going to the destination you need. Apps like Google Maps could provide details on when the train will reach each station and the exact route it will follow.” Additional suggestions included “Updating the app during strikes” and “Creating a live interface to help track and direct us to the platforms easily”.

When asked about negative metro experiences, participants frequently mentioned issues with cleanliness, unclear directions, poor security, and lack of network connectivity. While most passengers did not experience trouble while deboarding, 10.6% encountered a situation where the door did not open, preventing them from getting off the train.

The final question was open-ended, asking participants what they thought the metro could improve to enhance their journey experience. The overall feedback indicated a strong need for clearer and more accurate information dissemination to ensure a smoother and more efficient journey. Some suggestions included “TV screens showing where the metro is going”, “clearer signs, better directions, more staff”, and “easier access to metro schedules and data”.

3.3 Interview Methodology

Interviews were held with two Newcastle metro staff to understand the system they currently follow and what according to them might be helpful for the passengers. The interview questions were designed to capture the crucial metro information and understand the user complaints and feedback they have received. The interviewees are kept anonymous and answers to the interviews were recorded using pen and paper as opposed to audio recording.

3.4 Key Findings from Interviews

After talking with metro staff, several important points came up that show where passengers struggle and how the metro service could be better. A big issue is that passengers often don't get real-time information about when trains will arrive, if there are delays, or if services are disrupted. Even though updates are shared through announcements, display boards, and online, they don't always reach everyone effectively. Passengers also find the current website and app hard to use, and many wish they were easier to navigate.

The staff pointed out that having real-time data on how crowded trains are and when they'll arrive would be helpful for passengers to plan their trips better. During service disruptions, it's crucial to manage the flow of people and provide clear alternative routes, which could be improved with better real-time communication.

On the staff side, they sometimes struggle to quickly get and share important information, especially during busy times. They suggested that simpler internal systems that highlight the most important information would help. Overall, these conversations made it clear that there's a need for a more user-friendly and integrated metro information system that focuses on providing timely updates, useful predictions, and easier access for everyone.

3.5 Summary of the Findings

From the interviews and surveys, it can be concluded that there is a huge requirement of a portal which contains aggregated meaningful information for the passengers to be able to travel efficiently. Through the surveys, the challenges of passengers can be seen. Many of the passengers focused on a need of staff to be present due to a lack on correct information availability which can be easily fixed by the presence of an app with consolidated data.

Staff members mostly focused on the importance of broadcasting real-time updates to the crowd for effective passenger movement management. It can be inferred from the interviews that there is a clear need for an improved system to provide information to the Newcastle Metro commuters.

Based on the comprehensive survey and interview findings, several key functional and non-functional requirements for the Metro Insights project were identified [23]. These requirements are essential for ensuring the system addresses the pain-points of Newcastle Metro passengers and improves overall operational efficiency. The table below outlines these requirements, which were derived from user feedback and the insights provided by metro staff.

Table 1. Functional and Non-functional Requirements.

Category	Requirement	Description
Functional	Real-Time Updates	The system must provide accurate real-time updates on train schedules, delays, and service disruptions.

Functional	Predictive Insights	The system should offer predictive insights on crowd levels and potential service disruptions to help passengers plan their journeys better.
Functional	User-Friendly Interface	The website must have an intuitive and easy-to-navigate interface, accommodating users of all tech-savvy levels.
Functional	Interactive Maps and Directions	Include interactive maps that guide passengers to the correct platform and provide directions within the station.
Functional	Ticket Information and Purchase	Display clear information about ticket and zone types, and integrate a purchasing option within the app.
Functional	Real-Time Communication	Enable real-time communication for service updates, especially during disruptions.
Functional	Customization Options	Allow users to customize notifications and alerts based on their preferences and routes.
Non-Functional	Performance	The system should handle high volumes of real-time data efficiently with minimal latency.
Non-Functional	Scalability	The system architecture must be scalable to accommodate future expansions, such as integrating other transportation systems.
Non-Functional	Reliability	Ensure the system is reliable, with high uptime and accurate data processing to provide consistent user experience.
Non-Functional	Security	Protect user data and ensure secure transactions, particularly in ticket purchases and personal information handling.
Non-Functional	Maintainability	The system should be easy to maintain and update, with clear documentation and modular components [24].
Non-Functional	Usability	Focus on usability to ensure that even first-time users can navigate and use the system effectively.

4 System Design

The system design consists of a data pipeline, machine learning models, and a front-end website prototype. Python, its libraries, and Jupyter Notebook are used for data pipeline development, data analysis and model development, while Figma is employed for designing the website prototype.

4.1 System Architecture

The architecture of the Metro Insights system is designed to be flexible and scalable, ensuring that it can handle vast volumes of data and provide real-time updates. The key components include:

- *Data Collection:* Historical data combined with dummy data inspired from the London metro [25] has been collected to be utilized for the project.
- *Preprocessing:* After the data is gathered, it is preprocessed. This entails cleaning out unnecessary and improperly prepared data, dealing with missing values, and formatting the data into a format that is appropriate for analysis.
- *Machine Learning Models:* Few machine learning models have been developed to analyze data and generate predictive insights.
- *Data Visualization:* The data after preprocessing is used to generate insights in form of visualizations and tables using interactive charts and graphs. Python libraries Altair and Plotly are used for this purpose [26] [27].
- *Website Prototype*: The front-end prototype is developed using Figma, focusing on a user-centric design that ensures ease of use and accessibility.

4.2 Data Pipeline

The data pipeline is a step by step flow of data from raw data to meaningful insights. The data pipeline is created on Jupyter Notebook using python, pandas and machine learning [28]. It consists of the following steps:

- *Data Ingestion:* Data is first loaded into the pipeline in raw form and formatted such that it can be used for further processing. The datasets are converted to pandas ‘dataframe’ and ready for further processing.
- *Data Cleaning and Transformation*: The raw data ingested is cleaned to remove unnecessary details, remove the null values, and transformed into a structured form which is suitable for analysis. This step is also important to reduce the processing time moving forward in the code by reducing the data and keeping only what’s required. For Example, some data sets had ‘trainId’ starting with alphabets like ‘TDA1246’ and some didn’t. So, to make it uniform, the unwanted alphabets were removed.
- *Feature Engineering*: Important pieces of information, or "features," are picked out and refined from the data to help the machine learning models work

better. For instance, created a ‘delayMinutes’ column from train times, train arrival time and train departure time.

- *Insightful Visualization Generation***:** Using Altair and Plotly, insightful and interactive visuals are created based on the research, surveys and interviews performed.
- *Model Training and Validation:* The transformed data is then divided into an 80:20 ratio. 80% of the data is used to train the machine learning models and 20% is used to test the model to ensure their accuracy and reliability. This step is performed to generate predictive insights and analysis based on historical data.
- *Deployment and Testing*: The outputs are deployed over the website prototype and that is tested with the international people in Newcastle for feedback.

4.3 Front-End Website Prototype

The front-end website prototype was made with metro travellers' ease of use in mind. Ensuring visual attractiveness, accessibility, and ease of use were the main priorities during the design process. An important aspect of this design process was figuring out what average users need and their behaviour.

A user journey map (figure 1) was built to make sure that the Metro Insights website satisfies the expectations of its users. This map illustrates the several stages that a typical user, like Sarah, experiences when using the website. The website was designed with her obstacles and overall experience in mind, taking into account her goals, thoughts, and actions at each level.

Newcastle Metro User : Sarah

Actor:	*Scenario:*	*Goals & Expectations:*
Sarah Sarah uses the Newcastle Metro 1-2 times a week. She needs to travel to different places in Newcastle.	**Planning her journey** She requires information like, metro plaform, timings and ticket information to plan her journey and has to leave home a lot ahead of time to be able to figure out all these.	**Goal:** To be able to plan her journey easily and travel efficiently and comfortably.

Phase Name Phase Goal Phase duration	Action	Feeling
Awareness Learn about the metro insights website 2-4 months	• **Scrolls** the internet • **Sees a** social media advertisment • **Searches information** about the metro schedules online •**Talk to friends** about their metro journey experience • **Receive advice** from friends to use this website	**Excitement:** Chance to learn more information about **Apprehension**: So many unknowns and lack of trust on the information
Visit Website Visit the website 2 weeks	• **Clicks on the link** to visit the metro insights website • **Arrives at** homepage • **Notices** the introductory banner, User Interface and readability in the website	**Satisfaction**: So much information about the metro **Confusion**: Lots of information and routes to process **Fear**: Lack of trust on the information
Explore Features Go through the website and explore features it provides 1 day	• **Scan** for relevant information • **Observes** the insights and metro information. • **Uses** the search bar, navigation menu, introductory banner and other features	**Overwhelmed**: A lot of moving factors and features **Conflicted**: Is this a trust worthy website? **Motivated**: If it works out, it will save a lot of time
Detailed Information Access Get access to detailed information in form of tables and graphs 1 day	• **Navigates** through route maps • **Uses** search bar to get specific route information • **Goes through** specific route schedule detailed information • **Explores** FAQs and about sections	**Hopeful**: Hoping the train details are correct **Ready:** To make the commitment **Impatient:** Want to save as much time **Out-of-control:** Factors beyond her control
Real-time Updates Receive real time updates on change in schedule, alerts and notifications, crowd levels. 1 month	• **Schedules notifications** for upcoming trips • **Receives alerts** for last minute schedule changes • **Sets up notifcations** to learn about crowd levels • **Gets updated** on better routes or any delays along the way	**Happy:** She got the update before reaching and can change her plan timely **Overwhlemed:** Too many alerts and notifications **Proud:** Spending less time and reaching everywhwere on time
Personalization The website and data insights is interactive and user-centered 1 day	• **Interacts** with the tables and graph • **Uses** option to sort, filter or change view • Only **sees** the information required for the travel	• **Excitement:** Able to focus on critical aspects • **Confusion:** Gets confused with so many options
Feedback and Interaction Provide feedback on any issues they face and access to Frequently asked questions(FAQs) 1 day	• **Provides feedback** on incorrect information or bug in the website • **Asks questions** for any confusion on the website • **Reads through** FAQs to see if anyone else faced similar issue	• **Calm:** Shares any problem faced and gets solution • **Belongingness:** Able to relate and learn from FAQs
Regular Use and Revisit Use it for daily commute regularly. ~	• **Finds** the website helpful for daily travels • **Uses it** for planning commutes to a new destination • **Checks** metro schedules and live updates beforehand to save time	• **Satisfaction:** Regular use of a webiste with so many features and information makes planning trips easier • **Efficient:** Using the website saves a lot of time and heps reach the destination efficiently

Figure1. User journey map and a detailed persona of a Newcastle Metro user.

The user journey map played a key role in pinpointing important touchpoints where users could run into problems or where the design could be improved to better suit their needs. Important design choices including the arrangement of navigational components, the information flow, and the kinds of visual aids employed were influenced by this. Furthermore, "Sarah," a persona who utilises the Newcastle Metro 1-2 times a week, offered a targeted illustration of the kind of user the website is intended to support. Features were developed that align closely with user demands

while remaining functional, thanks to a clear understanding of the user's goals and challenges.
The prototype effectively strikes a balance between functionality and user experience thanks to this user-centred design process, making it both aesthetically pleasing and simple to use.

4.4 Technology Stack

The technology stack used for the development of the Metro Insights website includes:

- *Python and Pandas:* For data analysis and data pipeline development.
- *Altair and Plotly:* Python libraries used for creating visuals.
- *Jupyter:* For collaborative coding and model training.
- *Figma:* For designing the website prototype and display front end.

5 Implementation

5.1 Data Preprocessing

The following steps were performed to get the data ready for processing:

- The historical data for this project was huge and was stored in a format not fixable manually.

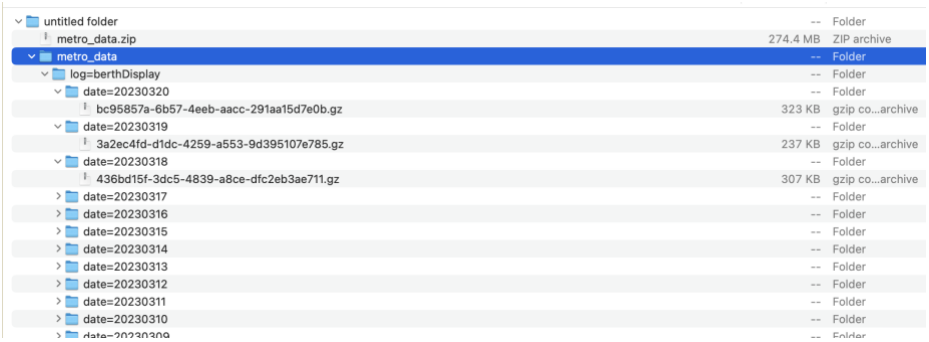


Figure 2. Folder Structure of raw data.

Figure 3. Number of contains large amounts of sub-folders.

- Using python all the zip files inside the sub-folders were unzipped, converted to .csv files, combined under one folder each and then saved into dataframes for further preprocessing.
- Next step was to delete all the null and duplicate values from the datasets. This step is necessary as the data is large and would otherwise take a lot of time in processing further.
- There is a 'Berthmap' dataset and 'train_details' data set that contains all the meta data about the berths and trains respectively. The last step for preprocessing was to join the other datasets to these datasets to get the berth, train and platform information and only kept the columns required for generating visuals.

5.2 Visualizations Generation

The visualizations offer a station-level perspective, allowing users to explore patterns across different times of the day, days of the week, and other contextual factors. A universal station drop-down enhances the interactivity, enabling focused analysis on specific stations.

5.2.1 Crowd Levels by Day of the Week and Time of Day

This bar graph shows the number of people at different metro stations during the week and at different periods of the day.
The interactive legend gives users the option to filter the crowd level data by choosing particular stations, giving them a customized perspective of the crowd trends for that station. Additionally, the tooltips allow users to see more information like the name of the station and the precise crowd level by hovering over the bars.
The graph clearly shows the busiest periods of the day at various stations, making it simple to determine when and where the largest crowds are present. The interactive feature of the visualizations allows users to toggle between stations, enabling them to focus on and analyze data specific to their station of interest in a more detailed and personalized manner.

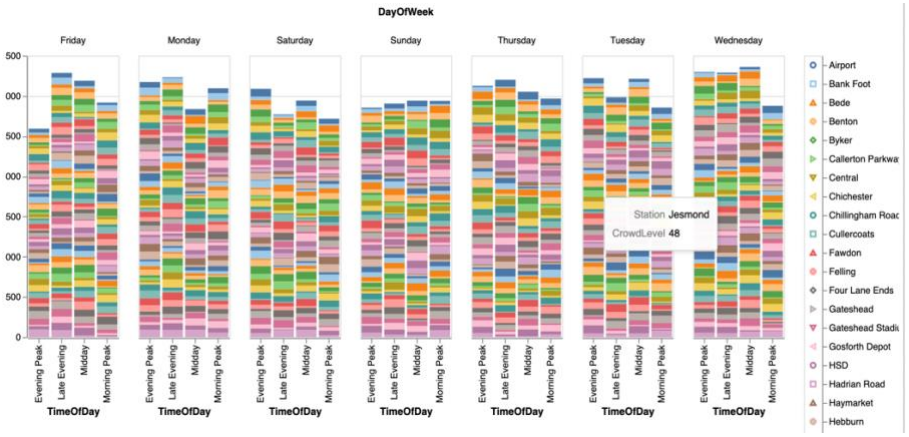


Figure 4. Graph depicting crowd levels by day of the week and time of day.

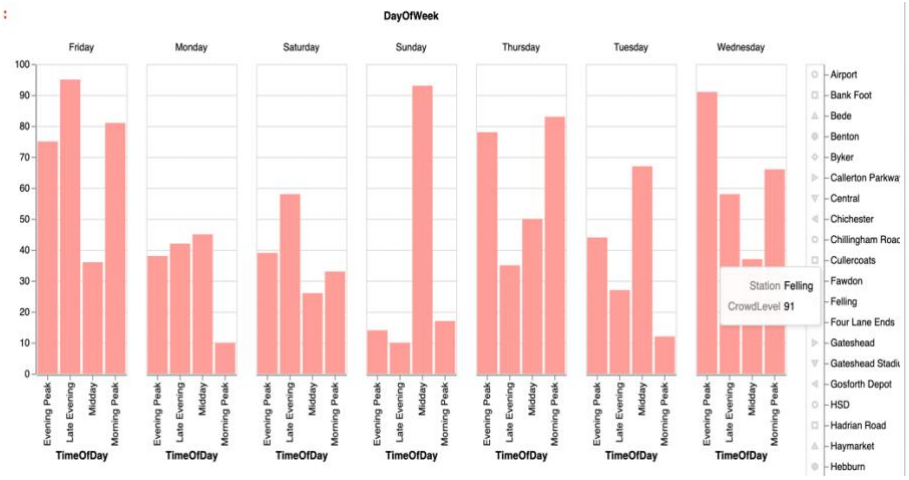


Figure 5. Crowd levels by day of the week and time of day for a particular station.

5.2.2 Crowd Levels by Event Nearby and Day of the Week

This bar graph examines the correlation between crowd sizes at metro stations during the week and neighboring events (such as concerts, festivals, and sporting events). From the interactive event legend, users can choose an event to view how adjacent events impact the number of people at that specific station on different days. Additionally, the bars are color coded according to the kind of events. The graphic makes it evident how events affect the number of people at certain stations, enabling users to predict peak periods based on surrounding activities.

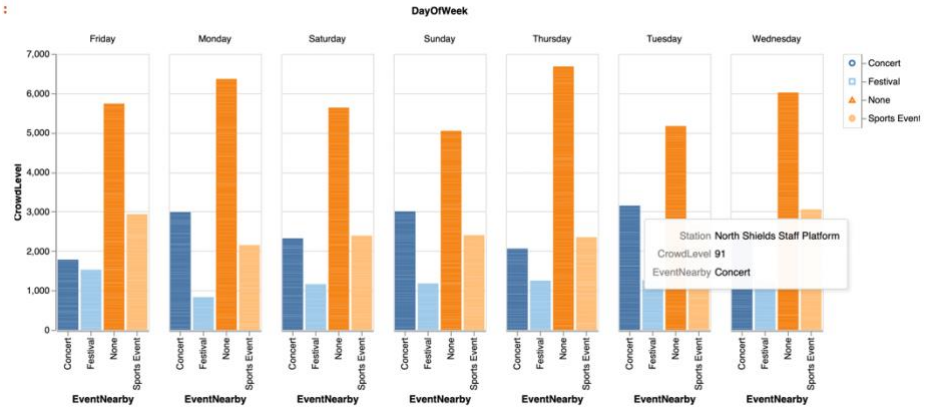


Figure 6. Graph depicting crowd levels by event nearby and day of the week.

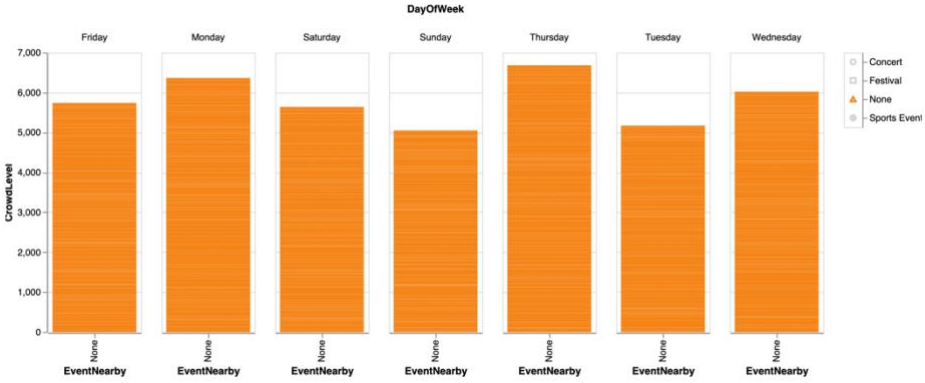


Figure 7. Crowd levels by event nearby and day of the week for no event.

5.2.3 Average Train Delay by Station and Weather Condition

The average train delays at several metro stations under various weather conditions (cloudy, rainy, snowy, sunny, windy) are displayed in a bar chart. A targeted study of train delays under various weather circumstances is made possible by the interactive weather legend, which lets users filter the chart to show data for specific weather conditions. Hovering over the bars also displays the tooltips with comprehensive information on average delay times and associated weather conditions. With the help of this visualization, users can see how various weather conditions affect train delays at various stations, providing important information for operational and planning choices.

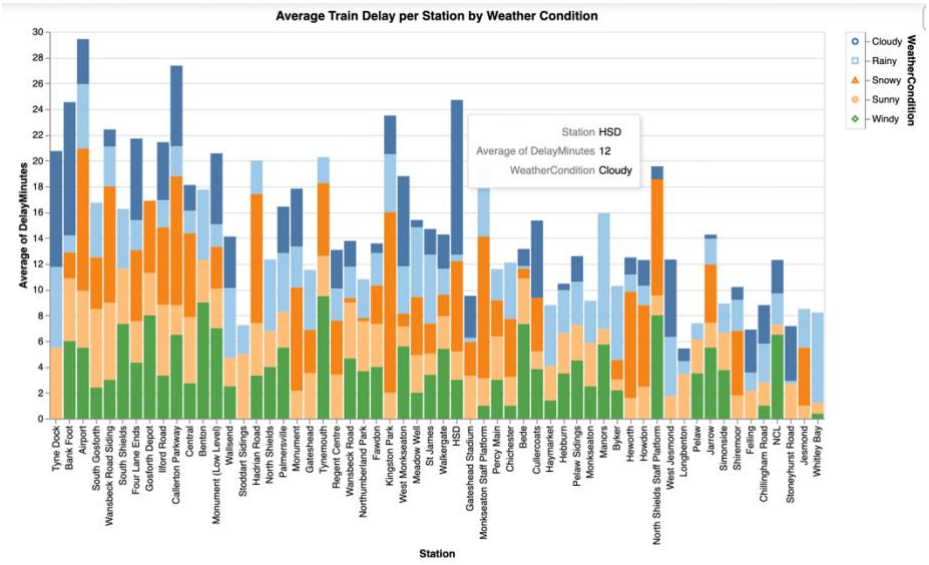


Figure 8. Graph Depicting average train delay by station and weather conditions.

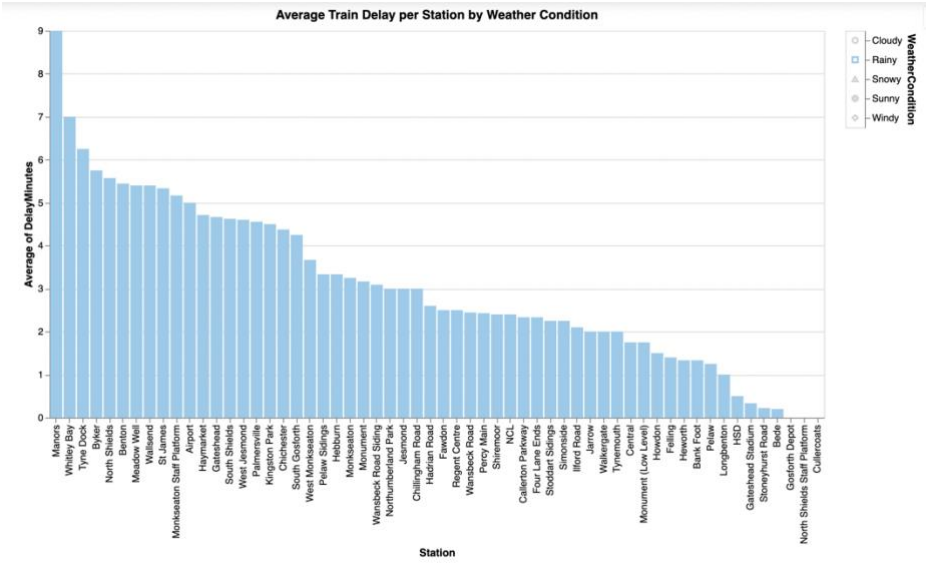


Figure 9. Average train delay by station and weather conditions for rainy weather.

5.2.4 Three graphs with Universal Station Dropdown Menu

The charts are interactive, with the ability to select stations via a drop-down menu, offering a focused view for all three graphs. User will not have to toggle legend for the

graphs separately, they can just select one station from the drop down and all three graphs will display focused views.

- *Average Train Frequency by Time of Day:* A line with points chart that visualizes the average train frequency (in minutes) at different times of the day across various stations. Users can easily compare train frequency patterns across different times of the day, aiding in the identification of periods with higher or lower service availability.
- *Crowd Levels by Station and Time of Day:* This heatmap provides an overview of crowd levels at different stations across various times of the day. The heatmap offers a comprehensive overview of crowd patterns across the metro system, enabling users to identify critical times and stations where crowd management strategies might be necessary.

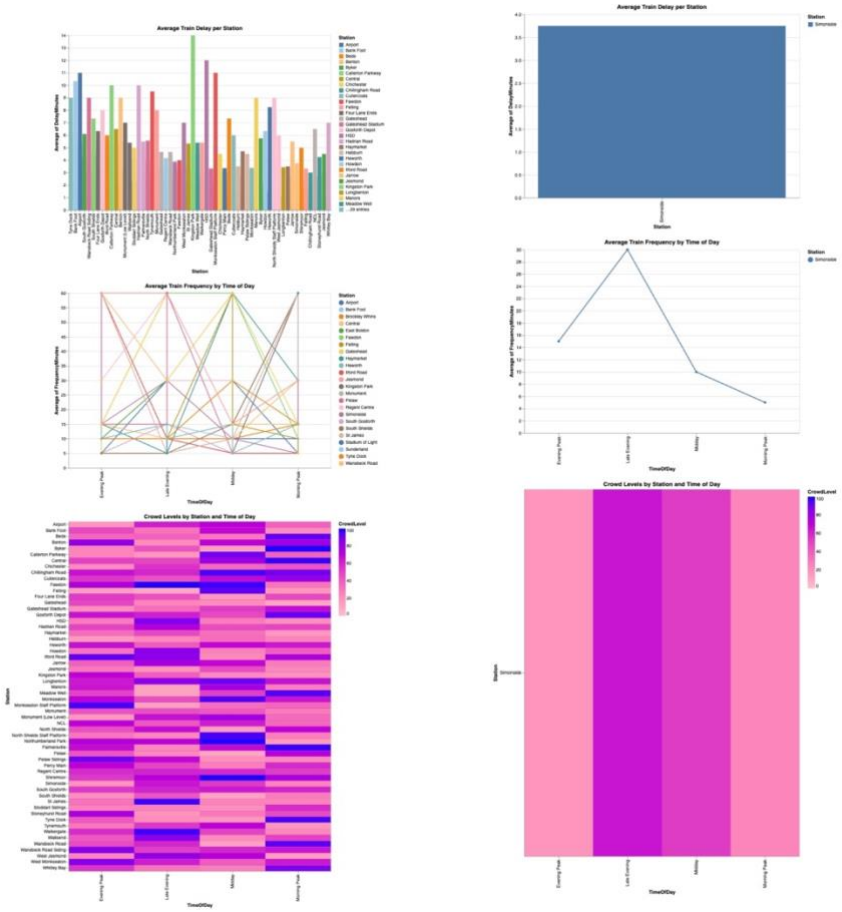


Figure 10. On left the three graphs mentioned in section 6.4 vs on the left filtered by station 'Simonside'.

- *Average Train Delay per Station:* This simple bar chart displays the average train delay (in minutes) at each metro station. Each bar represents the average delay time for a specific station, providing a straightforward comparison of delays across the metro system.

5.3 Machine Learning Model Development

Leveraging historical metro data predictive models are developed to analyse expected crowd levels and service delays. After evaluating various machine learning algorithms, the ones with highest accuracy have been implemented.

5.3.1 Data Analysis

A thorough data analysis was conducted to identify underlying patterns and linkages in the data before creating the machine learning models for forecasting crowd sizes and train delays. The dataset contained a variety of factors that were utilised as predictors for delays and crowd sizes, including station names, time of day, day of the week, weather, and neighbouring events. The goal variables included train delays and crowd densities at various times and locations.
Crowd Levels: Information on crowd densities was analysed to determine peak hours and factors that lead to higher crowd densities. Crowd levels have been found to be significantly impacted by variables such as the time of day and adjacent activities.
Train Delays: Understanding how outside variables like weather and events affected train punctuality was the main goal of delay prediction. Patterns and trends were found using historical data on delays.

5.3.2 Model Selection

Several machine learning algorithms were taken into consideration for forecasting both crowd levels and delays when the data was understood. The models listed below were assessed:

- *Random Forest Regressor:* Valued for its robustness against overfitting and its capacity to handle both numerical and categorical features.
- *Gradient Boosting Regressor* was selected because to its capacity to enhance prediction accuracy by rectifying faults in earlier models.
- *XGBoost:* A sophisticated boosting technique that is used for both regression and classification applications and is renowned for its excellent performance and efficiency.
- *Support Vector Machines(SVMs):* assessed for categorisation tasks, such as anticipating the existence of delays.
- *Decision Trees:* Due to their interpretability, simpler models were taken into consideration for first investigations.

Regression models such as Random Forest and Gradient Boosting were given priority for crowd level prediction. A two-step method was used to forecast delays: first, a classifier was used to predict whether a delay would occur, and then a regression model was used to predict how much of a delay would occur.

5.3.3 Model Training and Validation

After selecting the models, the following steps were undertaken:

- *Data Preprocessing:* Categorical variables were one-hot encoded, and the data was split into training and testing sets.
- *Hyperparameter Tuning:* GridSearchCV was used to tune hyperparameters such as n_estimators, learning_rate, and max_depth, optimizing model performance.
- *Model Training:* Models were trained using the training set, with cross-validation ensuring generalizability.
- *Model Validation:* Performance was validated using the test set, with Mean Squared Error (MSE) and accuracy as key metrics.

5.3.4 Results

Based on the following indicators, the machine learning models' performance was assessed:

- *Crowd Density Forecast:* The mean squared error (MSE) for the Gradient Boosting Regressor was 779.19. The algorithm can identify some patterns, but there is still a large margin of error in forecasting the precise crowd sizes, as indicated by the comparatively high mean square error. The complexity of crowd behavior, which may not be fully captured by the features available, or noise in the data could cause this high MSE.
- *Forecasting Delays:* With 64.6% accuracy, the *classification model* was able to anticipate when delays would occur. Although there is need for development, the model can distinguish between delayed and on-time trains to some extent, as evidenced by its moderate accuracy. A dataset imbalance (more on-time trains than delayed ones) or a lack of strong enough characteristics to discern between the two groups could be the cause of the decreased accuracy.
  The Mean Squared Error (MSE) of the *regression model* used to forecast the length of delays was 57.83. Although the model can forecast delays to some extent, as indicated by this MSE, the error reveals that predictions are not very precise. The model's incapacity to capture intricate correlations between characteristics or variability in the factors driving delays could be the cause of the high MSE.

5.4 Website Prototype

A prototype of the website was developed using Figma. The design focused on user-centred principles, ensuring ease of use and accessibility. Key features include interactive visualizations, real-time updates, and personalized user settings.

5.4.1 Prototype Design

The website design follows key HCI principles to ensure it is intuitive and user-friendly. Keeping the HCI principles and user feedback in mind, the prototype consists of nine pages. All the pages have a consistent layout and users also have the option to Sign In or Register to save their journey plans and get personalised alerts and notifications.

Figure 11. Website Header containing option to navigate to different pages.

5.4.2 Home Page:

Home page (Figure 5) shows option to see the live updates, planned repairs, planned diversions or any changes in the timetable. It also consists an about section for the Newcastle metro and the website and a map of the metro routes.

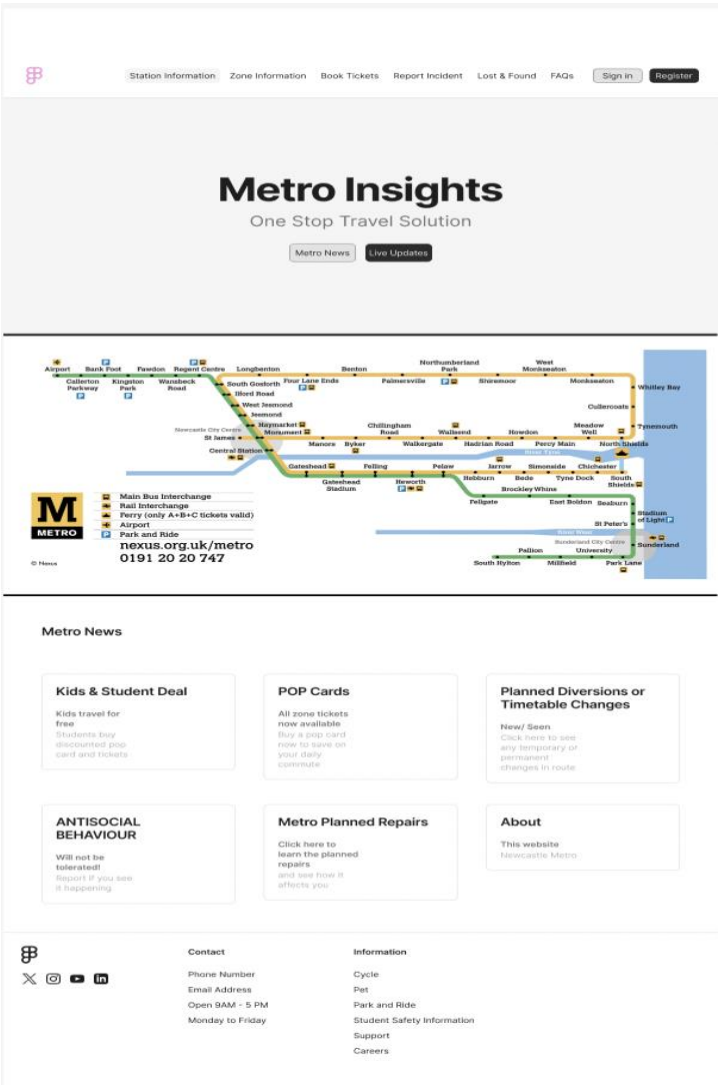
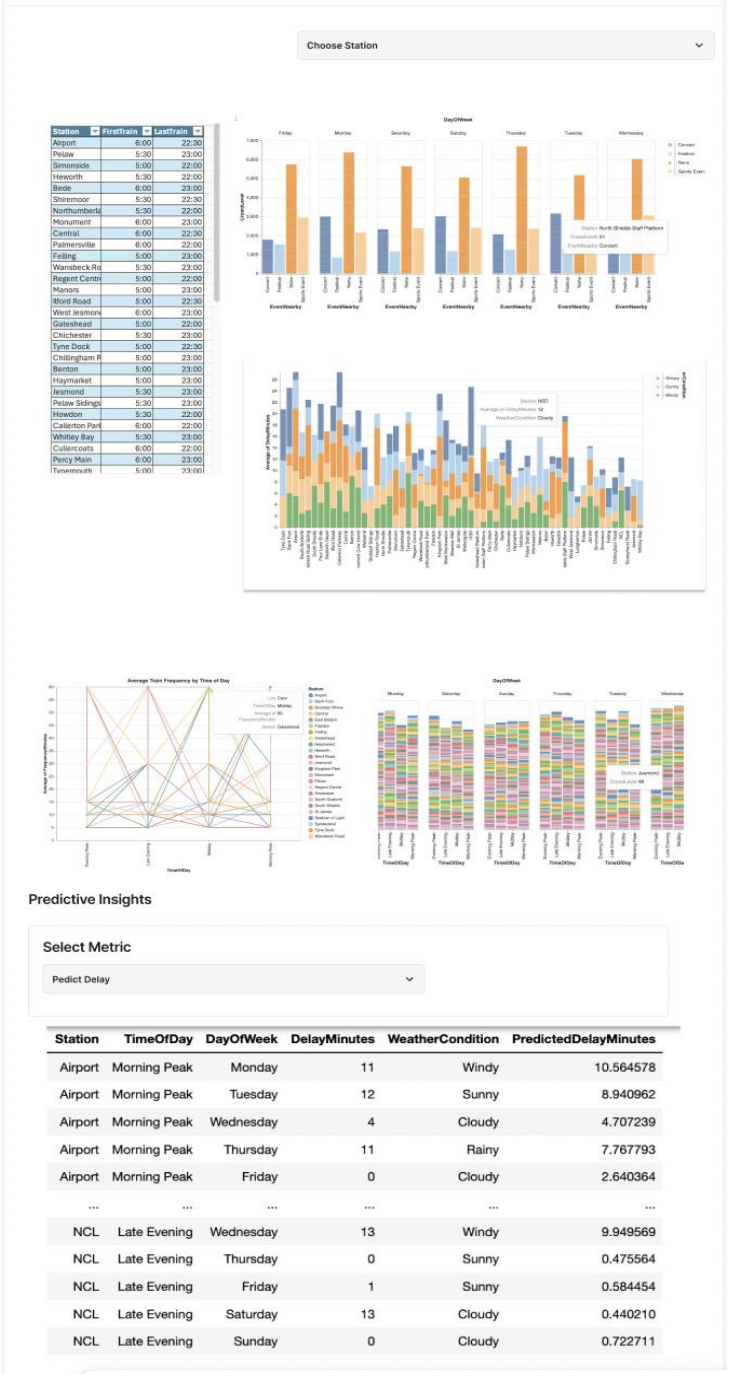


Figure 12. Screenshot of the Home Page.

- *Station Information*: This page contains all detailed information about the metro, visualizations and tables. Passengers can see data according to the station they want to travel to and from. Passengers can also save their journey plan after logging in and see predictive insights.



Station	TimeOfDay	DayOfWeek	DelayMinutes	WeatherCondition	PredictedDelayMinutes
Airport	Morning Peak	Monday	11	Windy	10.564578
Airport	Morning Peak	Tuesday	12	Sunny	8.940962
Airport	Morning Peak	Wednesday	4	Cloudy	4.707239
Airport	Morning Peak	Thursday	11	Rainy	7.767793
Airport	Morning Peak	Friday	0	Cloudy	2.640364
...	...	...	...	...	...
NCL	Late Evening	Wednesday	13	Windy	9.949569
NCL	Late Evening	Thursday	0	Sunny	0.475564
NCL	Late Evening	Friday	1	Sunny	0.584454
NCL	Late Evening	Saturday	13	Cloudy	0.440210
NCL	Late Evening	Sunday	0	Cloudy	0.722711

Figure 13. Screenshot of the Station Information Page except header and footer.

- *Real Time Updates*: Real time updates page contains any kind of real time updates, timetable changes, defects or delays.

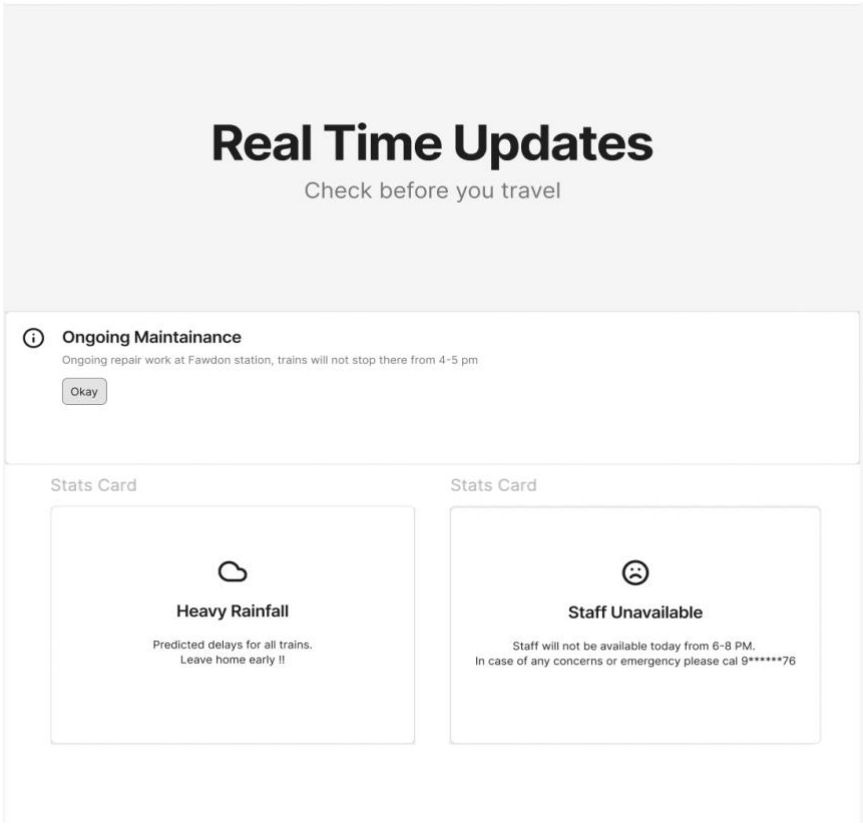


Figure 15. Screenshot of the Real Time Updates page without header and footer.

- *Book Tickets*: Here users can buy tickets according to the zone and type of ticket they would need or they can buy 'pop cards' which are rechargeable cards and can be used multiple times to commute.

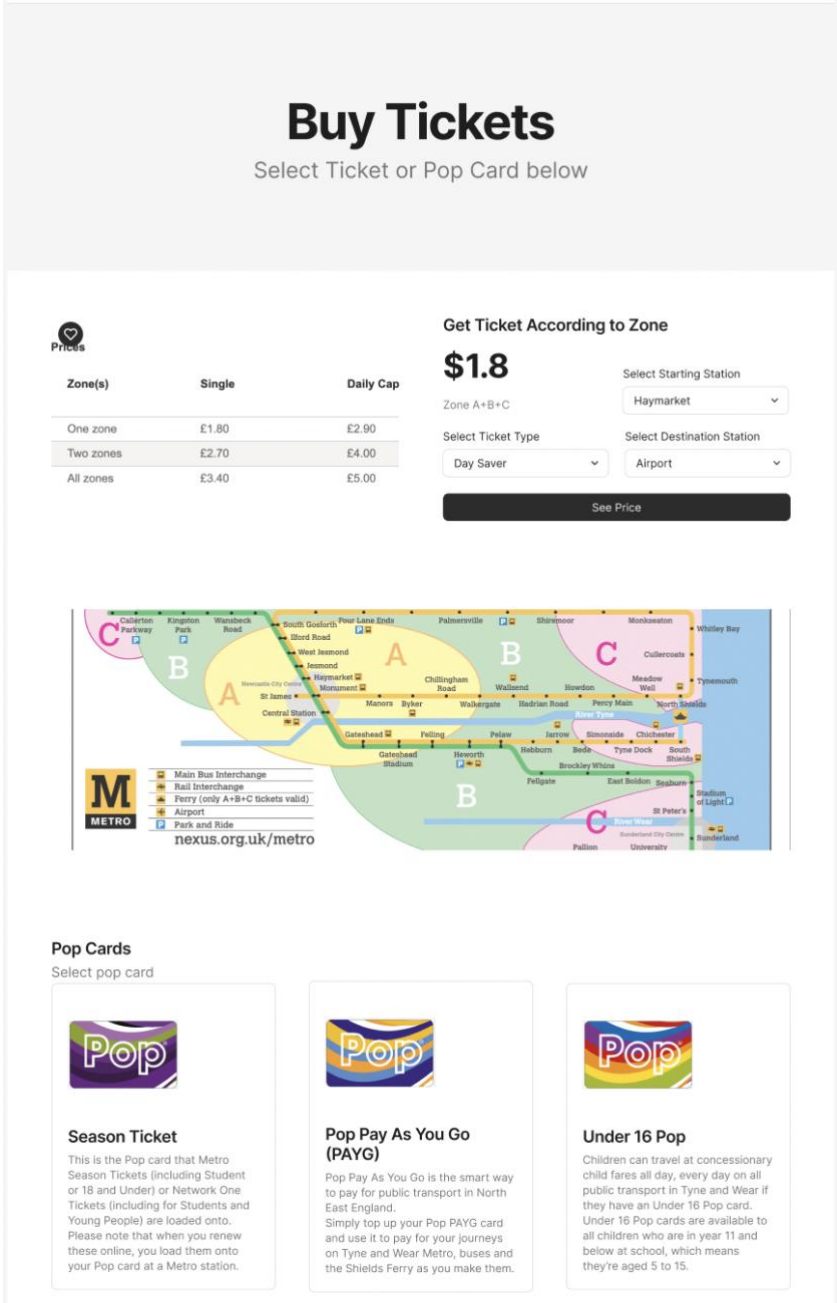


Figure 16. Screenshot of Book Tickets page excluding header and footer. The metro route map is sourced from the Nexus.org.uk website [29].

- *Report Incidents*: Any faults or problems with the metro or the platform faced by the passengers can be reported here and will be shared with the staff to fix. This way we can incorporate passenger feedback continuously and will make passenger journey experience better in the long run for passengers and easier for the staff to identify defects. Any cases of anti-social behaviour can also be reported here. This page also improves the security aspect of travelling in the metro which a lot of survey participants mentioned is lacking.

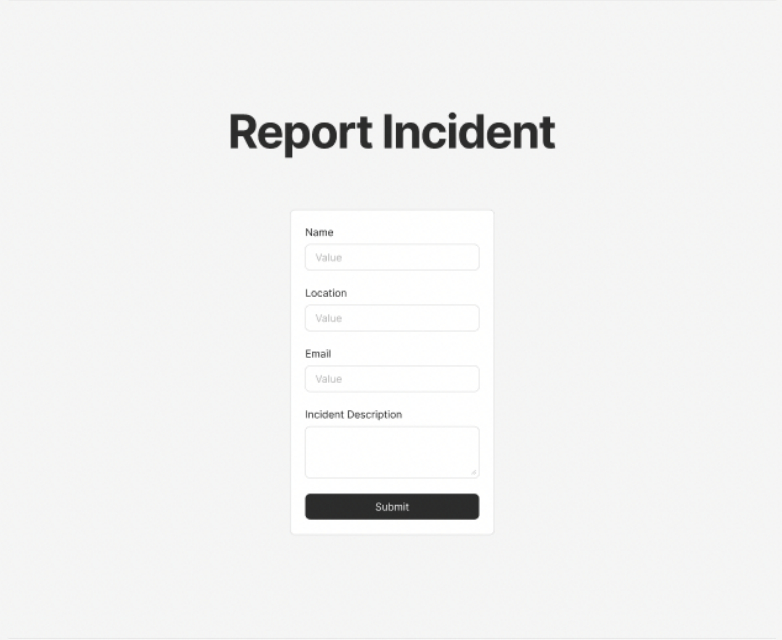


Figure 17. Screenshot of Report Incident page excluding header and footer.

- *Lost & Found***:** In case the staff finds an item on the metro which a passenger left behind they can post it here with a picture. If a passenger left something on the metro they can come to this page and see if their item has been posted about. Passengers can also declare any lost item as soon as possible so as to make the staff aware of the same.

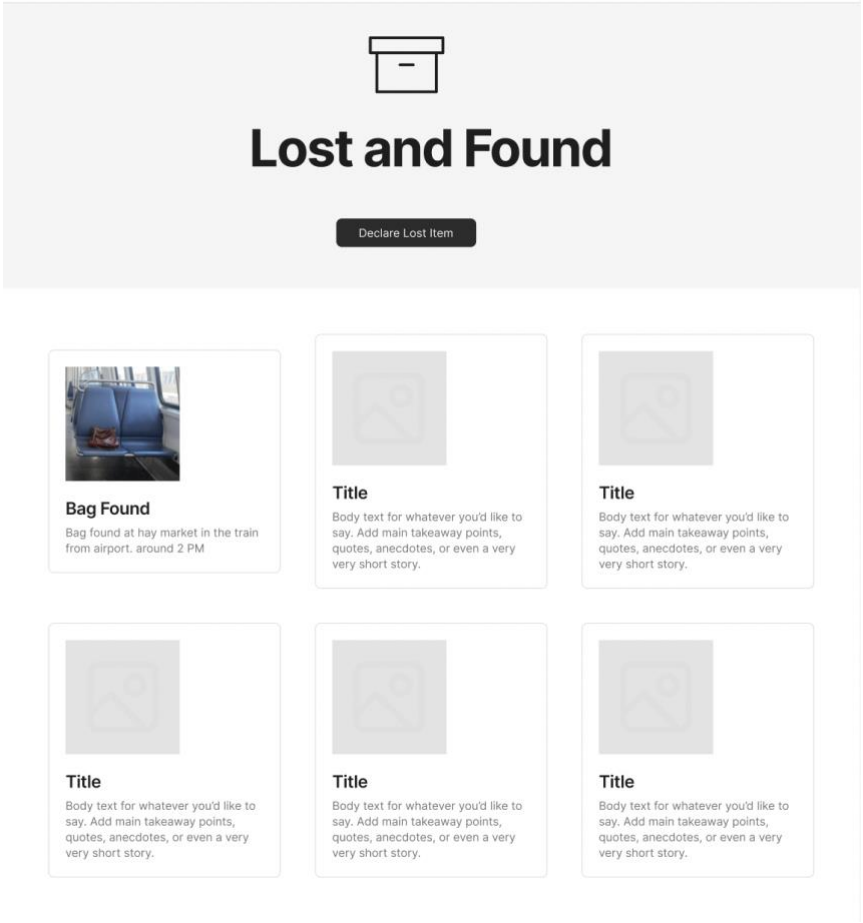


Figure 18. Screenshot of Lost and Found Page without header and footer.

- *FAQS:* If a user is having trouble finding some information regarding the website or metro travel, they can go through the FAQs before contacting the metro staff.

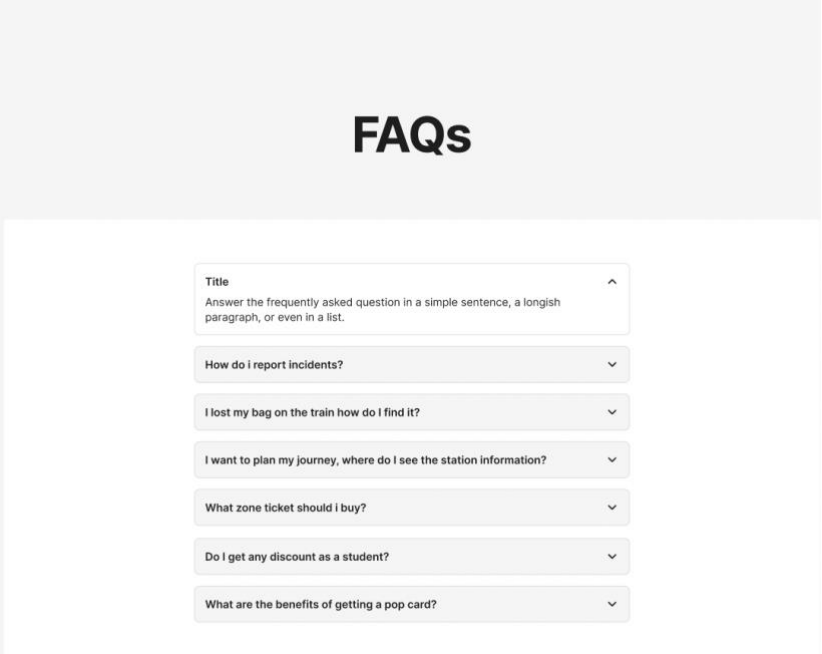


Figure 19. Screenshot of the FAQs Page without header and footer.

5.4.3 Prototype Features

The prototype includes several key features:

- *Real-Time Updates*: Providing real-time updates on train schedules and service disruptions.
- *Predictive Insights*: Offering predictive insights on crowd levels and potential delays.
- *Interactive Visualizations*: Displaying data through interactive charts and graphs.
- *Personalized Settings*: Allowing users to customize their preferences and receive personalized alerts.

5.4.4 User Testing

The prototype was tested with a few of the survey participants to gather feedback and identify areas for improvement. The feedback was used to refine the design and ensure it meets the needs of the users. Few of the feedbacks received were about the user interface of the website and how it can be more accessible and easier to understand. Some users provided feedback on the visuals which were incorporated and users also

provided advice on new features that can be added, for instance, FAQs page and Lost & Found Page.

5.4.5 Website Usability Testing

The usability of the website prototype was evaluated through user testing sessions with the survey participants to make sure that the website provides solution to the problems they voiced.

5.4.6 Methodology

- *Participants:* Five international users were recruited as participants for the usability testing [30]. The method used was in-person testing.
- *Tasks:* Participants were asked to complete a series of tasks using the website, such as checking real-time train schedules, viewing predictive crowd levels, and setting personalized alerts.
- *Metrics:* Usability metrics such as task completion rate, time taken to complete tasks, and user satisfaction were recorded [31].

5.4.7 Results

- *Task Completion Rate*: The average task completion rate was 92%, indicating that most participants were able to successfully complete the tasks.
- *Time Taken*: The average time taken to complete tasks was 3 minutes, which is within an acceptable range for a user-friendly interface.
- *User Satisfaction*: User satisfaction was measured using a Likert scale [32], with an average rating of 4.5 out of 5, indicating high levels of satisfaction with the website.

5.4.8 Feedback and Improvements

Feedback from the usability testing sessions was collected and analysed to identify areas for improvement. Key feedback points included:

- *Navigation*: Some users found the navigation menu to be cluttered. The design was refined to provide a more streamlined and intuitive navigation experience.
- *Visualizations*: Users appreciated the interactive visualizations but suggested adding more detailed explanations for certain charts. Additional tooltips and legends were incorporated to address this feedback.
- *Real-Time Updates*: Users emphasized the importance of real-time updates. The system's data pipeline was optimized to ensure more frequent updates and reduce latency.

- *Website Features*: Users pointed out a need for FAQs and Lost and Found page.
- *Machine Learning Accuracy*: More pertinent information, including real-time statistics, holidays, and special events, could be added to increase accuracy. Improving the model could entail adding more predictive features (such real-time traffic data or more detailed weather information) and balancing the dataset using methods like SMOTE.

5.4.9 Impact on Passenger Experience

- *Enhanced Information Access*: The website provided passengers with easy access to accurate and timely information, improving their overall travel experience.
- *Reduced Uncertainty*: Real-time updates and predictive insights reduced uncertainty and helped passengers plan their journeys more effectively.

6 Conclusion

This project succeeded in developing a prototype website that, by utilising the power of historical data and machine learning algorithms, provides Newcastle metro users with insightful information about their commute. The website tackles some of the most urgent issues that metro users deal with, like the requirement for precise real-time updates on train timetables and the capacity to forecast crowd sizes and possible service interruptions. The solution enhances passengers' overall travel experience by enabling them to make well-informed decisions through the provision of these insights.

One of the critical accomplishments was the integration of HCI principles as it made the website highly accessible and functional. User experience and feedback was given the highest priority in the design process due to which passengers can navigate the website seamlessly and easily obtain the information that they are seeking. The implementation of a user-centric approach is crucial for the website’s success among a diverse user community.

Furthermore, a notable advantage of the system is its modular architecture, which provides scalability and flexibility for future developments. The step-by-step pipeline architecture facilitates the easy incorporation of new features and functionalities in response to evolving user requests or the availability of fresh data. The design facilitates continual refinement and adaption, whether it is adding real-time data streams, improving predictive analytics with more complex algorithms, or reaching additional transit networks with the system.

This initiative not only provides for the immediate needs of passengers on Newcastle Metro, but it also institutes a step toward the use of data-driven solutions to public transport systems worldwide. The project's learnings can guide future initiatives to improve public transport services, making them more dependable, efficient, and user-friendly.

Overall, the project demonstrates a successful application of technology, data-driven solutions, user-centric design to solve real-world problems, whilst paving the way for future advancements that can further improve metro commute experience.

6.1 Summary of Achievements

- *Development of Machine Learning Models:* Successful development and deployment of machine learning models for time series forecasting, classification, and clustering.
- *User-Centric Website Design:* Creation of a user-friendly website prototype using Figma, incorporating feedback from usability testing.
- *Real-Time Data Integration:* Implementation of a robust data pipeline to handle real-time data updates and provide accurate insights to users.

6.2 Future Work

- *Integration of Real-Time Data:* Expanding the system to integrate real-time data from additional sources, such as IoT sensors, mobile applications, CCTV footage.
- *Website Interface for the Staff:* Development of separate interface after logging in for the staff and provide access based on their job profiles.
- *Advanced Predictive Analytics:* Enhancing the predictive models with more advanced algorithms and incorporating additional features to improve accuracy.
- *User Personalization:* Developing more personalized features based on user preferences and travel patterns.
- *Scalability:* Scaling the system to handle larger datasets and support more users as the system gains adoption.

6.3 Implications for Public Transport Information Systems

The Metro Insights project's success illustrates how utilising cutting-edge technologies in public transit information systems may have a transformative effect. This study has shown how real-time updates and predictive insights may be integrated into such systems to improve operational efficiency and improve the overall passenger experience.

The capacity to give passengers timely and accurate information is one of the most important consequences since it is essential to their ability to make well-informed travel decisions. Passengers can lessen their stress and uncertainty by modifying their plans in response to real-time updates on train schedules, delays, and service interruptions. Passengers are empowered to choose alternate routes or avoid busy hours thanks to predictive data, which can include projected crowd sizes or possible service delays. This results in a more comfortable and effective travel experience.

The capacity to track and evaluate real-time data can help transport operators operate the transport network more effectively and allocate resources more wisely. As a result,

things may run more smoothly, there may be fewer delays, and overall service reliability may rise. Transport authorities can also improve service continuity by proactively mitigating disruptions by foreseeing problems before they become more serious.

The Metro Insights project also establishes a standard for related endeavours in other cities, acting as an example of how to successfully execute smart transportation systems. The system's flexible and modular architecture allows it to be tailored to various cities or transport networks, meeting particular local requirements without compromising its essential features. Given its versatility, it can be used in a variety of public transport systems, including bus services and metro networks.

Additionally, the initiative advances the concept of smart cities, which enhance urban residents' quality of life through interconnected, data-driven systems. The need for intelligent, adaptable, and user-centred transportation systems grows as cities expand and transportation needs rise. The Metro Insights project's success highlights the value of innovation in this field and promotes more research and funding for related technologies.

As a result, the project's effects go beyond what the passengers of Newcastle Metro will immediately gain from it. It illustrates how public transit networks around the world have the capacity to develop into increasingly sophisticated, effective, and user-friendly systems. By establishing a benchmark for the integration of cutting-edge technology, the Metro Insights project clears the path for subsequent advancements that will further enhance public transport networks worldwide.

6.4 Self-Reflection on Challenges Faced

There were a lot of challenges faced during the development of this project, overcoming them provided a lot of opportunities to learn.

- As a first term user of Figma for front-end prototyping, there were a lot of new features of the software to learn. This is a very helpful tool which can be used to get user feedback before the development of any digital platform.
- The size of the datasets available for this project is huge. Reading and further processing the data in Python proved to be a challenging task. This was overcome by removing unwanted data, null values and duplicates. It was essential to optimize the code at every step and working on sample size of the data during the testing phase of the development.
- It was a challenge to achieve high accuracy in the machine learning algorithms applied to fetch predictive insights. This iterative process was very enlightening.

Acknowledgments. I would like to express my sincere gratitude to my dissertation supervisor, Dr. Caroline Claisse, for her patience, invaluable guidance, and unwavering support during countless hours, ensuring the successful completion of this project. I also extend my thanks to my Degree Program Director, Dr. Jason Steggles, for his insightful advice and timely updates. Additionally, I am grateful to all the participants of the survey and interview for their time and contributions.

References

[1] R. McNally, "Light Rail and Tram Statistics: England 2019/20," 25 June 2020. [Online]. Available: https://assets.publishing.service.gov.uk/media/5ffa22f3d3bf7f65de8015d4/light-rail-and-tram-statistics-england-march-2020.pdf.

[2] F. C. Nuannuan Leng, "The role of information availability to passengers in public transport disruptions: An agent-based simulation approach," *Transportation Research Part A: Policy and Practice,* vol. 133, pp. 214-236, 2020.

[3] O. Karadeniz, "The Power of Data and Personalization for Winning Customer Experiences," NMQ Digital, 27 March 2024. [Online]. Available: https://nmqdigital.com/blog/the-power-of-data-personalization-for-winning-customer-experiences#:~:text=Impact%20Product%20Personalization%3F-,What%20is%20Personalization%3F,engagement%2C%20sales%2C%20and%20loyalty..

[4] Johnpetter, "Importance of Data Analytics in Optimizing Digital Marketing Agency in UAE," Medium, 25 September 2023. [Online]. Available: https://medium.com/@johnpetter3312/importance-of-data-analytics-in-optimizing-digital-marketing-agency-in-uae-452986e93963.

[5] C. Pryor, "What Is User-Centred Design and How Can it Boost Your Data's Impact?," Contino, 18 March 2021. [Online]. Available: https://www.contino.io/insights/user-centred-design.

[6] P. R, K. Sanjaya, S.Rathika, A. H. Alawadi, K. Makhzuna, S. Venkatesh and B. Rajalakshmi, "Human-Computer Interaction: Enhancing User Experience in Interactive Systems," *E3S Web of Conferences,* vol. 399, no. E3S Web Conf., 2023.

[7] N. Hollender, C. Hofmann, M. Deneke and B. Schmitz, "Integrating cognitive load theory and concepts of human–computer interaction," *Computers in Human Behavior,* vol. 26, no. 6, pp. 1278-1288, 2010.

[8] K. Jasińska and D. Kłosek-Kozłowska, "Passengers' experience in underground non-transfer metro stations: The impact of spatial characteristics," *Tunnelling and Underground Space Technology,* vol. 143, 2024.

[9] A. Abulibdeh, "Analysis of mode choice affects from the introduction of Doha Metro using machine learning and statistical analysis," *Transportation Research Interdisciplinary Perspectives,* vol. 20, 2023.

[10] Y. Hu, S. Li, Y. Wang, H. Zhang, Y. Wei and L. Yang, "Robust metro train scheduling integrated with skip-stop pattern and passenger flow control strategy under uncertain passenger demands," *Computers & Operations Research,* vol. 151, 2023.

[11] E. Wong and A. Khani, "Transit Delay Estimation Using Stop-Level Automated Passenger Count Data," *Journal of Transportation Engineering, Part A: Systems*, vol. 144, no. 3, 2018.

[12] B. Ferris, K. E. Watkins and A. Borning, "OneBusAway: Results from Providing Real-Time Arrival Information for Public Transit," in *Proceedings of the 28th International Conference on Human Factors in Computing Systems*, Atlanta, 2010.

[13] C. Brakewood, S. barbeau and K. Watkins, "An experiment evaluating the impacts of real-time transit information on bus riders in Tampa, Florida," *Transportation Research Part A*, vol. 69, 2014.

[14] A. Gooze, K. E. Watkins and A. Borning, "Benefits of Real-Time Transit Information and Impacts of Data Accuracy on Rider Experience," *Transportation Research Record Journal*, vol. 2351, no. 1, pp. 95-103, 2013.

[15] M. G. Cedillo-Campos, C. M. Pérez-González, J. Piña-Barcena and E. Moreno-Quintero, "Measurement of travel time reliability of road transportation using GPS data: A freight fluidity approach," *Elsevier*, vol. 130, pp. 240-288, 2019.

[16] M. T. A. (MTA), "MTA Subway Time: Real-Time Arrival Information for New York City Subways," 2019. [Online]. Available: https://new.mta.info/.

[17] traveline, "Plan Your Journey," traveline, [Online]. Available: https://www.traveline.info/.

[18] J. Nielsen, Usability Engineering, San Francisco: Morgan Kaufmann, 1993.

[19] Y. Rogers, H. Sharp and J. Preece, Interaction Design: Beyond Human-Computer Interaction, 6th Edition, John Wiley and Sons, 2023.

[20] B. Shneiderman and C. Plaisant, Designing the User Interface: Strategies for Effective Human-Computer Interaction-4th Edition, Addison-Wesley Computing, 2010.

[21] J. Nielsen, "Iterative User Interface Design," Nielsen Norman Group, 1 November 1993. [Online]. Available: https://www.nngroup.com/articles/iterative-design/.

[22] surveymonkey.com, "identify-customer-touchpoints," surveymonkey.com, 1999. [Online]. Available: https://www.surveymonkey.com/mp/identify-customer-touchpoints/.

[23] altexsoft, "Functional and Nonfunctional Requirements: Specification and Types," altexsoft, 30 November 2023. [Online]. Available: https://www.altexsoft.com/blog/functional-and-non-functional-requirements-specification-and-types/.

[24] Microsoft, "Engineering Fundamentals Playbook/Non-Functional Requirements/Maintainability," Microsoft, [Online]. Available: https://microsoft.github.io/code-with-engineering-playbook/non-functional-requirements/maintainability/.

[25] gov.uk, *Open data users*, London: Transport for London (TFL).

[26] J. a. G. B. a. H. J. a. M. D. a. W. K. a. S. A. a. L. E. a. T. I. a. W. B. a. S. S. VanderPlas, "Altair: Interactive statistical visualizations for python," *Journal of open source software,* vol. 3, no. 32, p. 1057, 2018.

[27] P. T. Inc., *Collaborative data science,* Montreal: Plotly Technologies Inc., 2015.

[28] T. p. d. Team, "pandas-dev/pandas: Pandas," Feb 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3509134.

[29] nexus.org.uk, "Tyne and Wear Metro," nexus.org.uk, [Online]. Available: https://www.nexus.org.uk/metro.

[30] optimizely.com, "Usability testing," optimizely, [Online]. Available: https://www.optimizely.com/optimization-glossary/usability-testing/#:~:text=The%20goal%20of%20usability%20testing,go%20to%20a%20competitor's%20site..

[31] C. Mirabal, "7 Key usability metrics to unlock user insights," maze.co, 09 june 2023. [Online]. Available: https://maze.co/blog/measure-usability-metrics/.

[32] G. M. &. A. A. R. J. Sullivan, "Analyzing and interpreting data from likert-type scales," *Journal of graduate medical education,* vol. 5, no. 4, p. 541–542, 2013.

Appendices

Appendix A: Prototype Link
https://www.figma.com/proto/YY5ae1KhAyQmqVCOlT1lTF/Metro-Insights?node-id=1-1083&t=EPrSQFTn2aJzVFFj-1&scaling=scale-down&content-scaling=fixed&page-id=1%3A6&starting-point-node-id=1%3A1083
Steps to test the prototype: Click on the navigation menu on the header, which will lead to the other pages. Buttons are clickable too. To navigate to the home screen, click on the logo on the left of the header.

Appendix B: Survey Questions

International Metro Users:

1. I consent for my response to be collected for the purpose of this project.
2. How often do you use the metro?
3. Please rate below importance of staff being present. (5 being the highest).
4. Select below if you faced these difficulties in locating the closest metro station to you?
    - GPS redirected you to the wrong location.
    - Couldn't find any station close to you online, when there was one available.
    - Wasn't able to find the right entrance for the platform you had to go to.
    - Other
5. There were clear directions for what zone or type of tickets to purchase. If no please explain the issue faced.
6. Once in the metro station, are you able to tell which train to take, which platform the train will arrive at and at what time?
7. Do you believe an app or website providing this information for Newcastle metros will help you plan your travels?
8. How do you think online apps and websites can help in improving your journey experience?
9. Please explain any bad journey experiences you have had in the metro, for example. cleanliness, less clarity of what station the train has reached etc.
10. Any issues you have faced while deboarding the metro, for example doors not opening, unable to find exit or the next train, etc.
11. Please explain any other things you feel, the metro needs improvement on to make the journey of passengers better.

Appendix C: Interview Questions

Metro Staff:

1. Can you describe the most common issues passengers face when using the Newcastle Metro?
2. How do you currently broadcast service disruptions or delays to passengers?
3. What types of information do passengers frequently ask for when they approach metro staff?
4. How often do service disruptions occur, and how are they typically managed?
5. What challenges do you face in providing real-time updates to passengers?
6. According to you, what kind of real-time data would be most useful for passengers (e.g., train arrival times, crowd levels, etc.)?
7. In your opinion, how user-friendly are the current systems (e.g., website, apps) for passengers, and what improvements could be made?
8. What current systems do you have to go through the information to make decisions, and how can it be made better?

Appendix D: Ethical Approval Documents

Project Title: Metro Insights Website

Approval Date: May 30, 2024

Approved By: Research Policy Intelligence and Ethics Team

Committee Reference Number: 47743/2023

Ethical Considerations:

- Ensuring the privacy and confidentiality of survey participants.
- Obtaining informed consent from all participants.
- Ensuring that the data collected is used solely for the purposes of this research project.

Appendix E: Interview Consent Form

METRO INSIGHTS WEBSITE

About the interview

This interview is being taken to understand the requirements and difficulties faced by the Newcastle Metro staff and passengers. Historical/dummy data from the metro will be leveraged to fulfil the requirements identified, and will be displayed on a website. This website is being developed as a part of researcher 'Aditi Jain's' dissertation for the course MSc. advanced computer science at Newcastle university.

Consent to take part in research

- I..... voluntarily agree to participate in this research study.
- I understand that even if I agree to participate now, I can withdraw at any time or refuse to answer any question without any consequences of any kind.
- I understand that I can withdraw permission to use data from my interview within two weeks after the interview, in which case the material will be deleted.
- I have had the purpose and nature of the study explained to me in writing and I have had the opportunity to ask questions about the study.
- I understand that participation involves answering questions about my experience with the metro facilities.
- I understand that I will not benefit directly from participating in this research.
- I understand that all information I provide for this study will be treated confidentially.
- I understand that in any report on the results of this research my identity will remain anonymous. This will be done by changing my name and disguising any details of my interview which may reveal my identity or the identity of people I speak about.
- I understand that disguised extracts from my interview may be quoted in the researcher's dissertation which will be accessible by university professors and the researcher.
- I understand that signed consent forms and answers notes will be retained in the researchers laptop and shared with her supervisor until completion of researcher's dissertation (Till August).
- I understand that under freedom of information legalisation I am entitled to access the information I have provided at any time while it is in storage as specified above.

• I understand that I am free to contact any of the people involved in the research to seek further clarification and information.

Researcher – Aditi Jain Email – a.jain9@ncl.ac.uk

Supervisor – Caroline Classie Email – caroline.classie@ncl.ac.uk

I am consenting to the above terms and understand that information provided by me will be used in the development of the Metro Insights Website.

---------- ----------

Signature of participant Date

.....

For the researcher –

I believe the participant is giving informed consent to participate in this study

---------- ----------

Signature of researcher Date

Appendix F: Information Sheet:

Title of Study: Metro Insights Website

Invitation and Brief Summary

You are being invited to take part in a research study. Before you decide whether or not you wish to take part it is important that you understand why the research is being done and what it will involve. Please read this information carefully and discuss it with others if you wish. Take time to decide whether or not you wish to take part. If you do decide to take part, you will be asked to sign a consent form. However, you are free to withdraw at any time, without giving any reason and without any penalty or loss of benefits.

What is the purpose of the research?

The purpose of this research is to understand the difficulties faced and requirements to improve the metro journey experience.

What does taking part involve?

The interview will take around 30 minutes of your time and should not require us to meet again. In case we have any further or follow up questions, we will reach out and it is your decision to consent to talk again or not. The questions will include difficulties faced by metro staff in the metro functionality and what all information can help them better manage the operations.

What information will be collected and who will have access to the information collected?

The information collected will not be personal in nature, it will be regarding the use of Newcastle metro. All identifiable information collected will be hidden and used anonymously and will only be accessible by the researcher, to reach out for any follow up questions.
We will use your name and contact details [email] to contact you about the research study. Individuals at Newcastle University may look at your research data to check the accuracy of the research study.

Why have I been invited to take part?

For in-depth understanding of the metro functionalities interviews will be held for Newcastle metro employees.

Who is the sponsor and data controller for this research?

Newcastle University is the sponsor for this study based in the United Kingdom. Newcastle University will be using information from you in order to undertake this study and will act as the data controller for this study. This means that Newcastle University is responsible for looking after your information and using it properly.

Has this study received ethical approval?

This study has received ethical approval from Newcastle University Research Policy Intelligence and Ethics Team on 30/05/2024.

Who should I contact for further information relating to the research?

Aditi Jain, +447438303342

Who should I contact in order to file a complaint?

Research Study: Caroline Classie, caroline.classie@ncl.ac.uk

Appendix G: Dataset Schemas

Table: berth_occupied

- datetime_arrival: timestamp
- berth: string

- train: string
- date: string
- datetime_departure: timestamp

*Table***:** berthhistory

- datetime: timestamp
- berth: string
- train: string
- state: varchar(9)

Table: berthmap

- berth list: string
- berth location: string
- direction: string
- branch: string
- route: string

Table: berths

- datetime: timestamp
- berth: string
- state: string
- train: string
- date: string

*Table***:** log_alarm

- datetime: string
- type: string
- alarm: string
- state: string
- location: string
- alarmtype: string
- equipment: string

*Table***:** log_berth

- datetime: string
- type: string
- berth: string
- train: string
- train_date: string
- train_time: string
- train_datetime: string
- delta: string

*Table***:** log_berthdisplay

- datetime: string
- type: string
- berth: string
- state: string
- train: string
- train_date: string
- train_time: string
- train_datetime: string
- delta: string

*Table***:** log_equipment

- datetime: string
- type: string
- id: string
- state: string
- equipmenttype: string

*Table***:** log_infrastructure

- datetime: string
- type: string
- id: string
- state: string

Table: log_mobile

- datetime: string
- type: string
- train: string
- location_type: string
- location_number: string
- speed: string
- door: string
- service_in_ready: string

Table: log_platform

- datetime: string
- type: string
- platform: string
- platform_direction: string
- platform_event: string
- train: string

- train_date: string
- train_time: string
- train_datetime: string
- delta: string

*Table***:** log_train

- datetime: string
- type: string
- train: string
- update: string
- train_date: string
- train_time: string
- train_datetime: string
- delta: string
- location_type: string
- location_number: string
- speed: string
- door: string
- service_in_ready: string
- driver: string
- location: string