FP6-FutuRe
Cost effective regional lines

Grant Agreement Number: 101101962
Project Acronym: FP6 - FutuRe
Project title: Future of Regional Rail

DELIVERABLE 6.5

Specification of demand analysis algorithms

Project acronym:	FP6 - FutuRe
Starting date:	01/12/2022
Duration (in months):	48
Call (part) identifier:	Call: EU-RAIL JU Call Proposals 2022-01 (HORIZON-ER-JU-2022-01) Topic HORIZON-ER-JU-2022-FA6-01
Grant agreement no:	101101962
Grant Amendments:	NA
Due date of deliverable:	31-08-2024
Actual submission date:	03-09-2024
Coordinator:	Alessandro Mascis, Wabtec
Lead Beneficiary:	HACON
Version:	2.0
Type:	Report
Sensitivity or Dissemination level[1]:	PU
Taxonomy/keywords:	Mobility, Forecast demand, Occupancy, TMS integration

Europe's Rail

This project has received funding from the Europe's Rail Joint Undertaking (JU) under grant agreement 101015423. The JU receives support from the European Union's Horizon Europe research and innovation programme and the Europe's Rail JU members other than the Union.

[1] PU: Public; SEN: Sensitive, only for members of the consortium (including Commission Services)

Document history

Date	Name	Affiliation	Position/Project Role	Action/ Short Description
11/07/2024	Marco Ferreira	Hacon	Technical Manager/Task leader	Author. Final version for internal review of the deliverable reporting on Task 6.4 activities.
11/07/2024	Rui Eirinha	GTSP	Researcher/Task participant	Contributes with GTSP use case and related technical details
22/07/2024	Takamasa Suzuki	UIC	Reviewer	Review within WP6
05/08/2024	Zeno Pannunzio	Trenitalia	Reviewer	Review within WP6
23/08/2024	Marco Ferreira	Hacon	Technical Manager/Task leader	Updated document according to the reviewers' comments
23/08/2024	Fabrizio Burro	Wabtec	WP1 Leader	Quality Check
02/09/2024	Fabrizio Burro	Wabtec	WP1 Leader	Management of comments/remarks after Steering Committee review

Table of contents

Executive Summary

The EU-Rail FP6 Future project's Work Package 6 focuses on Regional Rail Services Requirements & Specifications, with the objective of developing and demonstrating highly accurate multimodal travel solutions for both on-board regional vehicles and at regional rail stations, for passengers and freight. Task 6.4 of WP6 focuses on providing short- and long-term travel demands using machine learning algorithms, enabling a more dynamic response to changing demand and allowing for the adjustment of planned rail services.

This deliverable, “Specification of demand analysis algorithms”, is a crucial component of the project, providing specifications for demand analysis algorithms under the specific view of regional lines. This deliverable will include use cases, system actors, capabilities and requirements, high-level architecture, exchange scenarios per use case, interfaces and standards, and algorithm descriptions.

The report's purpose is to provide a comprehensive understanding of the specifications of demand analysis algorithms, ensuring that they meet the project's objectives and requirements. The report details the scope of the work, including the partner’s developments involved, and the techniques used to develop the demand analysis algorithms. The report was developed considering also FP1 MOTIONAL project specifications on this topic, having a deep alignment on the designed solution.

The main findings and conclusions of the deliverable highlight the added value of the work, particularly in overcoming the current limitations of the public transport system, which often leads to costly operations and an inability to react to changing demand. The demand analysis algorithms developed in this deliverable will enable a more dynamic response to changing demand, reducing costs and improving the overall efficiency of regional rail services.

List of abbreviations, acronyms and definitions

Abbreviation / Acronym	Definition
CA	System Capability
CMS	Capacity Management System
DRT	Demand Responsive Transport
FP	Flagship Project
FRQ	Functional Requirement
IM	Infrastructure Manager
MaaS	Mobility-as-a-Service
NFRQ	Non-functional requirement
PIS	Passenger Information System
TMS	Traffic Management System
TRL	Technology Readiness Level
TSP	Transport Service Provider
UC	Use Case
WP	Work Package

List of figures

List of tables

1. Introduction

Imagine a future where regional rail services seamlessly connect European regions, ensuring that even the most remote lines play a crucial role in Europe's transportation network. This deliverable, D6.5, is a pivotal part of the FP6 FutuRe project, specifically within WP6, dedicated to Regional Rail Services Requirements & Specifications. The aim of this work package is to develop and demonstrate highly accurate multimodal travel solutions that provide service information for regional lines (lines with lower usage or secondary networks, that play a crucial role not only in serving European regions but also as feeders for passenger and freight traffic for the main/core network), for both passengers and freight. In addition, the work package aims to define and develop specifications documents for the corresponding work package, WP11, where the development and demonstration of the final releases are done aiming to achieve TRL 6.

In a world where travel demands are constantly shifting, the ability to adapt is crucial. So, this deliverable is developed to tackle this challenge as a part of Task 6.4 of WP6 focusing on the provision of short- and long-term travel demands, specifically the development of machine learning algorithms to calculate demand forecasts, enabling a more dynamic response to changing demand and allowing for the adjustment of planned rail services, or adjustment of offer from other modes. This task is essential in overcoming the current public transport system's limitations, which are based on fixed timetables and fixed numbers of provided seats, independent of the current demand. This often leads to costly operations and an inability to react to changing demand.
The deliverable will provide the specification of a demand analysis system and algorithms, focusing on long-term and short-term travel demand systems within the scope of regional lines.
While Chapter 2 describes the scope of the task responsible by the development of this deliverable, Chapter 3 will describe the aim of the deliverable. The outline of the main part of the deliverable is provided in Chapter 4, which describes the methodology applied to tackle the task of forecasting computing demand forecasts for regional lines. This methodology describes how Chapter 5 to 0 content is developed, including include use cases, system actors, system capabilities, requirements, high-level architecture, exchange scenarios, interfaces, and algorithm descriptions.
The use cases in Chapter 5 will provide an in-depth understanding of how the demand analysis algorithms will be used in practice, including the different system actors involved. In Chapter 0, the capabilities and requirements will outline the necessary features and functionalities of the demand analysis algorithms, while in Chapter 0 the high-level architecture will provide an overview of the system's design. The exchange scenarios per use case are described also in Chapter 0 how the different system actors will interact with the system and how the different systems involved interact with each other. Chapter 8 covers the interface specification that will ensure that the system is interoperable with other systems. Finally on Chapter 0, the algorithm descriptions will provide detailed information on how the demand analysis algorithms work.
. Conclusions are then drawn in Chapter 0.

Overall, the specification of demand analysis algorithms is a critical component of the FP6 FutuRe project, as it will enable a more dynamic response to changing demand, reduce costs, and improve the overall efficiency of the regional rail services. Additionally, these algorithms will enhance the passenger journey by optimizing travel schedules, reducing wait times, and ensuring a more reliable and comfortable travel experience. This deliverable will provide a comprehensive understanding of the demand analysis algorithms' specifications, ensuring that they meet the project's objectives and requirements.

2. Scope

The scope of the task 6.4 is to develop a comprehensive specification for algorithms that analyse short- and long-term travel demands specifically for regional rail services. The focus is on leveraging machine learning algorithms to forecast travel demand based on journey planning requests data and other available data sources (like vehicles or station counting data, local events or other relevant data), enabling a more dynamic response to changing demand and optimizing rail service operations as well as other transport modes connected to rail. This deliverable addresses the unique challenges faced by regional trains, which often operate on fixed timetables and/or seat allocations while having low frequency, leading to inefficiencies and high operational costs. By incorporating demand forecast data derived from journey planning requests, the system aims to provide information valuable to decision making on adjustments to planned services, thereby improving operational efficiency and reducing costs.

3. Objective/Aim

The objective of this deliverable is to develop a comprehensive specification for a demand analysis system and algorithms that enable a more dynamic response to changing demand in regional rail services. This specification will address both short-term and long-term travel demands. Short-term forecasts refer to the period from the current time up to two hours, while long-term forecasts cover the time frame from two hours to two weeks. This demand analysis is not limited to rail services alone but also encompasses the first/last mile transport modes.

Having accurate demand forecasts for regional rail lines and first/last mile transport modes offers several benefits. Firstly, it enables a more dynamic response to changing demand, allowing for the adjustment of planned rail services and alternative modes of transportation. This leads to increased operational efficiency and reduced costs. Secondly, it may facilitate a more optimized allocation of resources, ensuring that transportation services are aligned with the actual demand, thus improving overall service quality. Additionally, having demand forecasting information allows for better planning and decision-making processes in disruption situations. Ultimately, the demand forecasts can be seen as a tool for enhancing the overall user experience and satisfaction.

The methodology employed in this deliverable is based on the Arcadia methodology, which is a model-based engineering method for systems, which will be better described in the next chapter.

By developing these specifications, this deliverable aims to ensure that the subsequent development and demonstration in WP11 of the FP6 FutuRe project align with the project's objectives. The specifications will guide the development of a system that can effectively address the challenges faced by regional rail services and first/last mile transport modes.

4. Methodology

The purpose of this section is to describe the methodology used to develop the specifications of the system. The methodology employed is based on the Arcadia[2] methodology (model-based engineering method for systems), with a specific focus on the system and functional perspectives. The following subsections outline the key components and steps involved in this methodology.

1. Definition of Use Cases and Actors (cf. Chapter 5):
   The first step in the methodology involved defining the use cases and actors of the system. Use cases represent the various interactions or functionalities that the system needs to perform, while actors represent the entities that interact with the system. This step provided a clear understanding of the system's functional requirements and the stakeholders involved.
2. Capabilities and Requirements (cf. Chapter 0):
   Once the use cases and actors were defined, the next step was to identify the capabilities and requirements of the system. Capabilities represent the high-level functionalities that the system should possess, while requirements define the specific features and constraints that need to be met. This step helped in establishing a clear set of goals and objectives for the system.
3. Logical Architecture (cf. Section 7.1):
   With the capabilities and requirements in place, the methodology proceeded to define the logical architecture of the system. Logical architectures represent the structural organization of the system, including the different components and their interconnections. This step involved creating models to depict the system's structure and interfaces, enabling a better understanding of its overall design.
4. Components and Functions Descriptions (cf. Section 7.2):
   The next step involved describing the components and their associated functions within the system. Components represent the individual elements or modules that make up the system, while functions define the specific tasks or operations performed by these components. This step provides a detailed breakdown of the system's structure and functionality.
5. Exchange Scenarios (Per Use Case) (cf. Section 7.3):
   To further refine the system's functionality, exchange scenarios were developed for each use case. Exchange scenarios represent the sequence of interactions between the actors and the system components to achieve specific outcomes. These scenarios helped in capturing the dynamic behaviour of the system and validating its functionality.
6. Interfaces and Standards used (cf. Chapter 8):
   The interfaces and standards used within the system are also addressed. Interfaces represent the points of interaction between different systems, while standards define the protocols or guidelines that need to be followed for the interfaces implementation. This step ensures that the system's components and interfaces were compatible.
7. Algorithms description (cf. Chapter 0):
   Lastly, the methodology applied will describe the approaches taken for the development of the algorithms that will enable demand forecasting.

[2] https://www.sciencedirect.com/book/9781785481697/model-based-system-and-architecture-engineering-with-the-arcadia-method

This systematic approach ensured a comprehensive and well-defined specification for the system, enhancing its overall design and functionality.

5. Use Cases

This section provides the definition of use cases and actors within the context of the specification of demand analysis algorithms in the regional rail scope. The purpose of this section is to establish a clear understanding of the functional requirements and stakeholders involved in the development of the demand analysis algorithms.

In the regional rail scope, demand analysis algorithms can play a crucial role in understanding and predicting passenger demands, to optimizing resource allocation, and enhancing overall operational efficiency. To ensure the successful development of these algorithms, it is essential to define the use cases and actors that will interact with the system.

The actors are the entities or stakeholders (persons or external systems) that interact with the demand analysis system and algorithms. Defining the actors helps identify their roles, responsibilities, and perspectives, ensuring the algorithms meet their needs.

The use cases represent the various functionalities or interactions of the demand analysis system. They include information like the description of the use case, related project tasks/subtasks, involved actors, triggers, pre-conditions, input, result/requirement, sequence of steps, involved components and the responsible partner for the development of the use case.

All the detailed descriptions of the identified use cases and actors should be considered in the regional rail domain. Figure 1 identifies the use cases developed in Task 6.4, and related use cases from other tasks. This diagram visually represents the various actors involved, on the use cases and their interactions within the system. It also reflects the relation between the use cases highlighting possible dependencies between different use cases, offering a clear understanding of how the demand analysis system operates and how various stakeholders contribute to its functionality.

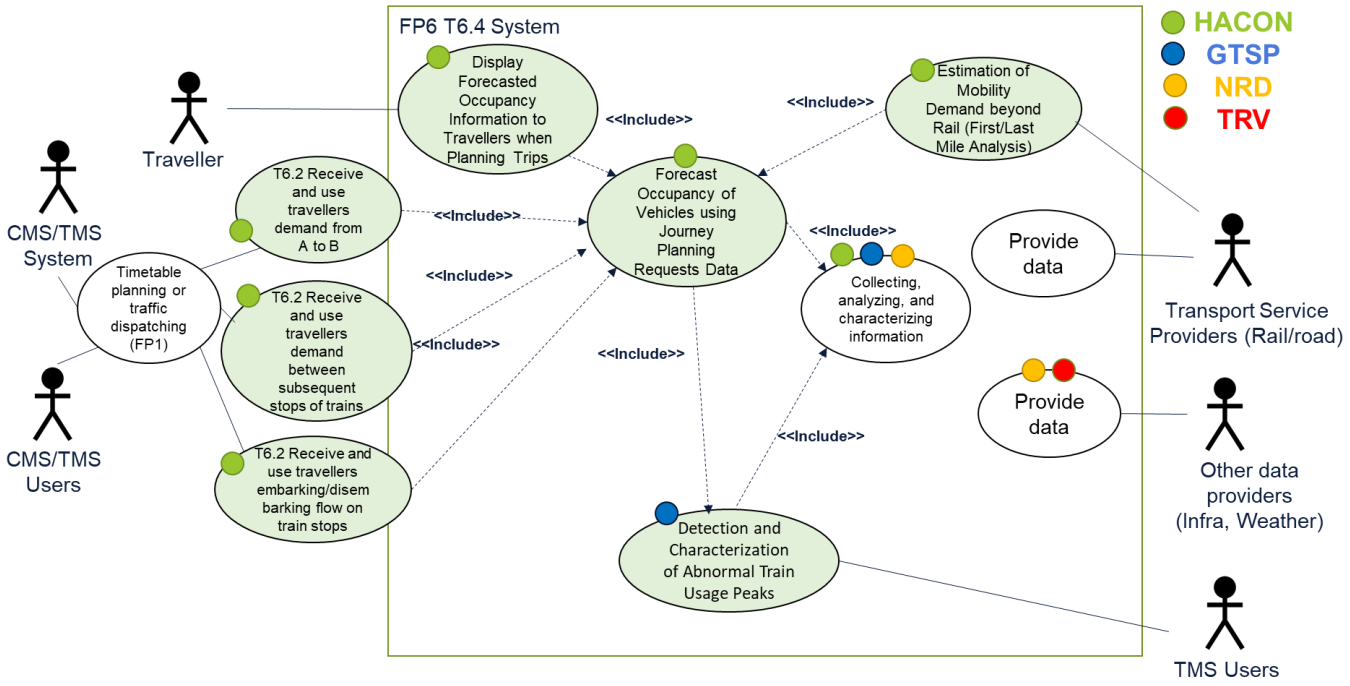

Figure 1 : Use cases and actors' identification

5.1. System Actors

In this section, we define the key actors involved in the use cases developed for this deliverable. Understanding the roles and responsibilities of these actors is crucial for comprehending how the demand analysis system operates and interacts with various stakeholders. Table 1: Actors definitions provides detailed descriptions of each actor, highlighting their functions and contributions within the system.

Actor	Description
CMS	A Capacity Management System (CMS) refers to a logical and integrated cycle of activities within a company or organization (IMs). Its purpose is to control and assure the competence of staff involved in rail operations. Essentially, it helps manage infrastructure capacity effectively, considering the needs of all users across various time horizons.
Transport Service Provider (TSP)	Organization providing both physical services and means of transport: trains, metros, coaches, buses, bike-sharing, car-sharing, DRT...
Traveller	The traveller is the person making or planning a travel.
MaaS platform	A MaaS (Mobility as a Service) platform is a digital platform that integrates various transportation services into a single, user-friendly interface. It aims to provide seamless and convenient travel experiences by offering a range of transport options, such as public transit, ridesharing, bike-sharing, car-sharing, and more, all in one place.
TMS	A TMS (Traffic Management System) system is a technology platform used in the rail industry to manage and control train operations. It serves as a centralized system that provides real-time monitoring, control, and coordination of trains, tracks, and related infrastructure.

Table 1: Actors definitions

Having defined the key actors involved in the demand analysis system, we now turn our attention to the specific use cases that illustrate how these actors interact within the system. The following section will detail various scenarios, demonstrating the practical application of the demand analysis algorithms and the roles played by each actor. This will provide a comprehensive understanding of the system's functionality and the collaborative efforts required to achieve the project's objectives.

5.2. UC-FP6-WP6-4.01 - Forecast Occupancy of Vehicles using Journey Planning Requests Data

Name	Forecast Occupancy of Vehicles using Journey Planning Requests Data
ID	UC-FP6-WP6-4.01
Description	This use case involves predicting the occupancy of transportation vehicles based on journey planning request data (or other sources), which includes information about the origin, destination, and expected time of travel for customers.
Related to task/subtask(s)	T6.4
Interactions SP/FP	Interaction with FP1 to align specification between main line and regional lines
Actor(s)	MaaS platform
Trigger	New journey planning requests data received
Pre-Condition(s)	Journey planning requests data is available, historical vehicle occupancy data is collected
Input	Journey planning request data (origin, destination, time of departure or arrival)
Result/Requirement	Predicted occupancy for specific routes and time slots
Sequence	List steps of the Use Case (to be filled during specification phase) 1. The MaaS platform process and provide new journey planning requests data, including origin, destination, and time of departure or arrival, during time interval (e.g. daily). 2. The data analytics platform gathers the new data 3. The data analytics platform considering historical vehicle occupancy data and journey planning request data, analyses the new data, training the model. 4. The system updates the trained occupancy model of the transport network. 5. Store the forecasted occupancy model for future reference.
Involved components (System)	Data Analytics Platform, Machine Learning Model
Responsible partner/person	Marco Ferreira (HACON)

Notes	The accuracy of the occupancy forecast may vary based on the quality and completeness of the journey planning request data.

5.3. UC-FP6-WP6-4.02 - Display Forecasted Occupancy Information to Travelers when Planning Trips

Name	Display Forecasted Occupancy Information to Travelers when Planning Trips
ID	UC-FP6-WP6-4.02
Description	This use case involves displaying forecasted vehicle occupancy information to travellers when they plan their trips through a trip planning interface.
Related to task/subtask(s)	T6.4
Interactions SP/FP	Possible interaction with FP1 to align specification between main line and regional lines
Actor(s)	Travellers
Trigger	Traveller initiates the trip planning process
Pre-Condition(s)	Traveller uses journey planning tool Forecasted vehicle occupancy data is available
Input	Journey planning request data (origin, destination, time of departure or arrival)
Result/Requirement	Display of forecasted vehicle occupancy information for the proposed journeys to travellers.
Sequence	List steps of the Use Case (to be filled during specification phase) 1. Traveller enters their origin, destination, and time of departure or arrival. 2. The trip planning interface displays several trip option and fetches forecasted vehicle occupancy information for the journeys. 3. Display the forecasted vehicle occupancy to the traveller on the trip planning interface.
Involved components (System)	Journey Planning app, Data Analytics Platform
Responsible partner/person	Marco Ferreira (HACON)
Notes	The displayed forecasted occupancy information is for planning purposes and may not reflect real-time changes in vehicle availability.

5.4. UC-FP6-WP6-4.03 - Estimation of Mobility Demand beyond Rail (First/Last Mile Analysis) (FP6)

Name	Estimation of Mobility Demand beyond Rail (First/Last Mile Analysis)
ID	UC-FP6-WP6-03
Description	This use case involves estimating mobility demand beyond rail transportation by conducting a first/last mile analysis. Transport Service Providers can analyse areas with high demand and low offering. On basis of this analysis demand gaps can be identified and reported to transport service providers. This allows them to create adapted offers for seamless transportation connections between rail stations and surrounding locations.
Related to task/subtask(s)	T6.4
Interactions SP/FP	Possible interaction with FP1 to align specification between main line and regional lines
Actor(s)	Transport Service Providers (Rail and others)
Trigger	Transport Service Provider requests a report about demand gaps
Pre-Condition(s)	Journey planning requests data are available, historical travel demand data is collected
Input	Journey planning requests data, historical travel demand data
Result/Requirement	Estimated mobility demand for first/last mile connections and identification of areas with high demand and low offering is provided to Transport service providers.
Sequence	1. Collect rail transportation data, including station locations, schedules. 2. Collect historic travel demand data based on journey planning requests data. 3. Train the machine learning model using historical travel demand data combining rail and other modes in the surrounding areas. 4. Estimate the mobility demand for first/last mile connections between rail stations and surrounding locations. 5. Identify areas with high demand and low offering based on the analysis results. 6. Provide data insights for transport service providers through the demand analytics dashboard.

Involved components (System)	Data Analytics Platform, Demand analytics dashboard
Responsible partner/person	Marco Ferreira (HACON)
Notes	The accuracy of the demand estimation depends on the quality and completeness of the available data.

5.5. UC-FP6-WP6-4.04 - Detection and Characterization of Abnormal Train Usage Peaks (FP6)

Name	Detection and Characterization of Abnormal Train Usage Peaks
ID	UC-FP6-WP6-4.04
Description	This use case focuses on identifying abnormal peaks in train usage by analysing historical train occupancy data alongside influencing factors such as environmental conditions and public/disruptive events. By using machine learning techniques designed for anomaly detection and prediction, it is expected to identify not only the regular peaks driven by daily commuter patterns, but also unusual peak times caused by external factors. The identified anomalies can then be provided to the TMS with all the necessary context to support informed decision-making for potential service adjustments.
Related to task/subtask(s)	T6.4
Interactions SP/FP	
Actor(s)	TMS
Trigger	Scheduled, based on weather forecast data frequency, public/disruptive events
Pre-Condition(s)	Availability of train schedule data. Access to weather data, public or disruptive events. Forecasted (or observed) vehicle occupancy data is available.
Input	Train schedule data. Weather condition data. Public/disruptive events information. Output from the vehicle occupancy forecast system (or historical occupancy data).
Result/Requirement	Detection of abnormal train usage peaks. Insights into how weather conditions, and public/disruptive events interact to provoke these anomalies.

Sequence	1. Weather forecast data (depending on data source release frequency), public and disruptive events, and forecast occupancy of vehicles (or observed) are collected and pre-processed. 2. Considering the historical and newly added data, the ML component performs a training process to adjust models' parameters (anomaly detection and anomaly prediction models). 3. The trained ML model for anomaly prediction is used to identify new anomalies based on upcoming weather and events. 4. Anomaly contextual information covering weather conditions, nearby events and historical comparisons with similar conditions are gathered and compiled. 5. The compiled information is sent to the Data Analytics Platform to be integrated into an anomaly structured message. 6. The anomaly structured message is delivered to the TMS for further analysis.
Involved components (System)	Data Analytics Platform Forecast Occupancy of Vehicles using Journey Planning Requests Data (T6.4) Anomaly Detection and Prediction Component
Responsible partner/person	GTSP
Notes	The accuracy of detecting train usage peaks depends heavily on the quality and completeness of the available data. The contextual data for anomalies can be further enhanced by incorporating, if available, additional complementary information such as train composition details, trending passenger feedback, and other relevant factors.

6. Capabilities and Requirements

In this section, the capabilities of the system are outlined, which are high-level functions required to be fulfilled by the Technical Enablers specified. Each capability is identified by a capability ID in the format of T6.4_CAxy.

The system is seen as the set of components (described in section 7.2) necessary to accomplish the task of generating demand forecasting information related to the usage of a transportation network.

The capabilities provide a concise overview of the high-level functions that the system must possess. By defining and understanding these capabilities, we ensure that the system meets the task objectives and that we define the requirements for each of the capabilities, facilitating the development of the solution.

This chapter will also include the requirements for each capability. These requirements will be defined using a template for technical requirement descriptions. The template consists of the following fields:

- Requirement ID: A unique identifier for the requirement in the format T6.x_UCx.y_RequirementType0z.
- Requirement Name: The name of the requirement.
- Use Case ID: A link to the corresponding use case.
- Category: The category of the requirement.
- Priority: The priority of the requirement, either "MUST" or "Nice-to-have with high priority."
- Main goal: A detailed description of the requirement, explaining what it is and why it is needed.
- Assumptions: Any assumptions made for the requirement.
- Specification: A detailed description of the requirement and how it should work.
- Additional Notes: Any additional notes about the requirement, such as risks, dependencies, or constraints.

By including the requirements for each capability, we ensure that the system is developed to meet the specific needs and objectives of the project. These requirements provide a detailed understanding of the functionality and performance expected from the system, guiding the development process effectively.

6.1. T6.4_CA01 Data Collection and Integration

The Data Collection and Integration capability, focuses on the system's ability to gather and integrate relevant data for occupancy forecast. This capability ensures that the system can collect journey planning requests, including origin, destination, and expected travel time, as well as optionally integrate vehicle occupancy sensor data.

T6.4_UC4.1_FRQ01

Requirement ID	T6.4_UC4.1_FRQ01
Requirement Name	The system must be able to gather data relevant for occupancy forecast
Use Case ID	UC-FP6-WP6-4.01
Category	Functional
Priority	MUST
Main goal	The system must be able to gather data relevant for occupancy forecast: • Journey planning requests (including origin, destination, and preferred time of travel • Vehicle occupancy sensor data (optional) • Other relevant data (optional)
Assumptions	• The system assumes that a journey planning system is available. • The system assumes that journey planning requests will be made by users and provide information such as origin, destination, and preferred time of travel. • The system assumes that the data received will be sufficient and accurate for demand forecasting purposes.
Specification	• The system should have the capability to receive and process journey planning requests data from users, including the origin, destination, and preferred time of travel. • If available, the system should be able to collect and utilize vehicle occupancy sensor data to further enhance the accuracy of the occupancy forecast. • The system should also be able to handle other relevant data (like historical data or ticketing data), if provided, in order to improve the accuracy of the occupancy forecast.
Additional Notes	Additional quality sources may help to improve the quality of the forecast.

T6.4_UC4.1_NFRQ01

Requirement ID	T6.4_UC4.1_NFRQ01
Requirement Name	The system must be able process the collected data in a reasonable time
Use Case ID	UC-FP6-WP6-4.01
Category	Non-Functional
Priority	MUST
Main goal	The system must be able to receive data (such as journey planning requests, vehicle occupancy, and other relevant data) from a Journey Planning System or other systems and process it in a reasonable time frame. The processed data will be consumed by demand forecast algorithms.
Assumptions	• The system assumes that data will be received from a Journey Planning System or other systems. • The availability and format of the received data will be compatible with the system's processing capabilities. • The definition of "reasonable time" may vary depending on the specific requirements and expectations of the final application. • Other data may be available (like disruptions or delays information, occupancy sensors on the train)
Specification	• The system should have the capability to receive data from a Journey Planning System or other systems. The data can include journey planning requests, vehicle occupancy information, and any other relevant data. • The received data should be processed efficiently and effectively within a reasonable time frame. • The system should have the capability to handle and process large volumes of data to ensure accurate demand forecasting. • The processed data should be made available for consumption by demand forecast algorithms. • The complete recalculation of the demand on the network is only performed daily • Journey planning requests should be processed to update the demand forecast information daily • Other data considered by the algorithms (disruptions, delays, counter), which have a higher frequency (real-time data), may lead to partial updates of the demand forecast on the network

Additional Notes	A reasonable time may vary and depend on the expectation for the demonstration. Since the amount of Journey Planning requests needs to reach sufficient numbers, the forecast may be calculated on a daily basis and needs to be available within the current day for the next, meaning that the “reasonable time” may still be a couple of hours. For data with higher frequency, this should be processed in few seconds, but will lead to small updates of the demand forecast model.

6.2. T6.4_CA02 Forecast vehicles occupancy

The Forecast vehicle occupancy capability focuses on the system's ability to predict vehicle occupancy based on historical data and journey planning requests. This capability involves training a machine learning model using journey planning request data and historical vehicle occupancy data. The system then processes the collected data and generates a vehicle occupancy model, which is used to forecast the travellers' demand. The system also informs the MaaS platform about the new occupancy model, ensuring accurate information is available for travellers or TSPs.

T6.4_UC4.1_FRQ02

Requirement ID	T6.4_UC4.1_FRQ02
Requirement Name	The system must be able to train a machine learning model using journey planning request data and historical vehicle occupancy data.
Use Case ID	UC-FP6-WP6-4.01
Category	Functional
Priority	MUST
Main goal	Train a machine learning model using journey planning request data and historical vehicle occupancy data
Assumptions	• The system assumes that journey planning request data and historical vehicle occupancy data will be available for training the machine learning model. • The availability and quality of the training data will impact the accuracy and effectiveness of the trained model. • The system assumes that the machine learning model will be trained using appropriate algorithms and techniques to achieve the desired forecasting outcomes.
Specification	• The received data should be pre-processed and transformed into a suitable format for training the model. • The system should utilize appropriate machine learning algorithms and techniques to train the model based on the provided data.

	• The trained model should be able to analyse and make accurate predictions based on new journey planning request data and current vehicle occupancy data. • The demand forecast precisions should achieve 50% precision in the average forecast 1 week in advance and achieve 65% precision in the forecast at 1 hour. • The trained machine learning model should be stored and made available for future demand forecasting purposes.
Additional Notes	• The availability and quality of the training data will significantly impact the accuracy and effectiveness of the trained machine learning model. • The system should provide mechanisms for monitoring and evaluating the performance of the trained model to ensure its effectiveness over time.

T6.4_UC4.1_FRQ03

Requirement ID	T6.4_UC4.1_FRQ03
Requirement Name	The system must be able to process new data and retrain the occupancy model considering historic data.
Use Case ID	UC-FP6-WP6-4.01
Category	Functional
Priority	MUST
Main goal	To keep the demand forecast model up to date, the system must be able to process the expected input of journey planning request data and historical vehicle occupancy data automatically.
Assumptions	• The system assumes that new data, including journey planning requests and vehicle occupancy information, will be available for updating the occupancy model. • The availability and quality of the new data will impact the accuracy and effectiveness of the updated model. • The system assumes that historical data will be used as a reference for updating the occupancy model.
Specification	• The system should have the capability to receive new data, such as journey planning requests and vehicle occupancy information, for updating the occupancy model. • The received new data should be pre-processed and integrated with the existing historical data to create an updated dataset.

	• The system should utilize appropriate techniques and algorithms to update the occupancy model based on the new and historical data. • The updated occupancy model should take into account the patterns and trends observed in the historical data, as well as the new data, to provide accurate predictions. • The updated occupancy model should be stored and made available for future demand forecasting and analysis. • The trained model should be regularly updated and refined to improve the accuracy of the vehicle occupancy predictions.
Additional Notes	• The availability and quality of the training data will significantly impact the accuracy and effectiveness of the trained machine learning model. • The system should provide mechanisms for monitoring and evaluating the performance of the trained model to ensure its effectiveness over time.

T6.4_UC4.1_FRQ04

Requirement ID	T6.4_UC4.1_FRQ04
Requirement Name	The system must be able to receive a new journey planning request and apply the trained model to predict vehicle occupancy for the requested journey.
Use Case ID	UC-FP6-WP6-4.01
Category	Functional
Priority	MUST
Main goal	Predict vehicle occupancy for the requested journey
Assumptions	• The system assumes that a trained model for predicting vehicle occupancy based on journey planning requests and historical data is available. • The accuracy of the trained model will impact the accuracy of the occupancy prediction for the requested journey. • The system assumes that the journey planning request will provide the necessary information for predicting vehicle occupancy, such as origin, destination, and preferred time of travel.
Specification	• The system should have the capability to receive a new journey planning request, which includes information such as origin, destination, and preferred time of travel.

	• The received journey planning request should be processed and utilized by the trained model to predict the vehicle occupancy for the requested journey. • The system should apply the trained model to analyse the journey planning request and provide an accurate prediction of the expected vehicle occupancy. • The prediction of vehicle occupancy should consider the patterns and trends observed in the historical data used to train the model. • The system should provide the predicted vehicle occupancy information in a format that can be easily communicated to the user or integrated with other systems.
Additional Notes	• The accuracy of the occupancy prediction will depend on the availability and quality of the trained model and the data used for training. • The system should provide mechanisms for monitoring and evaluating the performance of the predictions to ensure their effectiveness over time.

T6.4_UC4.1_FRQ05

Requirement ID	T6.4_UC4.1_FRQ05
Requirement Name	The system must be able to inform the Journey Planning System about a new occupancy forecast.
Use Case ID	UC-FP6-WP6-4.01
Category	Functional
Priority	MUST
Main goal	In regular time intervals, the forecasted occupancy information is extracted from the model and exported to the Journey Planning System. The Journey Planning System is then capable of using this information to enrich the journey planning response with the forecasted occupancy information.
Assumptions	• The system assumes that a trained occupancy model is available and regularly updated with new data. • The Journey Planning System can receive and process the forecasted occupancy information provided by the system. • The availability and accuracy of the forecasted occupancy information will impact the effectiveness of the Journey Planning System in enriching journey planning responses.

Specification	• The system should extract forecasted occupancy information from the trained occupancy model in regular time intervals. • The extracted forecasted occupancy information should be exported to the Journey Planning System in a format that is compatible and easily consumable by the platform. • The exported forecasted occupancy information should be integrated into the journey planning responses provided by the Journey Planning System. • The system should ensure the accuracy and timeliness of the forecasted occupancy information by regularly updating the occupancy model with new data.
Additional Notes	• The effectiveness of the Journey Planning System in enriching journey planning responses with forecasted occupancy information will depend on the accuracy and timeliness of the exported information. • The system should provide mechanisms for handling any errors or inconsistencies in the exported forecasted occupancy information and ensure proper communication with the Journey Planning System.

T6.4_UC4.1_NFRQ02

Requirement ID	T6.4_UC4.1_NFRQ02
Requirement Name	The system should keep the occupancy model as updated as possible
Use Case ID	UC-FP6-WP6-4.01
Category	Non-Functional
Priority	Nice-to-have
Main goal	The system should aim to keep the occupancy model updated to enable timely availability of updated and improved forecasts in the Journey Planning System, resulting in higher quality journey planning results.
Assumptions	The system assumes that the occupancy model can be updated with new data.
Specification	• The system should have mechanisms in place to regularly update the occupancy model with new data. • The system should prioritize the speed and efficiency of updating the model to ensure timely availability of updated forecasts.

	• The updated occupancy model should be seamlessly integrated with the Journey Planning System to provide improved journey planning results.
Additional Notes	The frequency and timing of the model updates should be balanced with the availability and quality of the new data.

6.3. T6.4_CA03 Forecasted Occupancy Retrieval

The Forecasted Occupancy Retrieval capability focuses on the system's ability to retrieve forecasted vehicle occupancy information from the model based on user input. This capability involves processing user input, including origin, destination, and preferred time of travel, and using this information to fetch the forecasted occupancy data for the proposed journeys.

T6.4_UC4.2_FRQ01

Requirement ID	T6.4_UC4.2_FRQ01
Requirement Name	The system must be able to process user input, including origin, destination, and preferred time of travel.
Use Case ID	UC-FP6-WP6-4.02
Category	Functional
Priority	MUST
Main goal	The system must be able to process user input, which includes origin, destination, and preferred time of travel.
Assumptions	• The system assumes that users will provide input in the form of origin, destination, and preferred time of travel. • The provided user input will be used for journey planning and demand forecasting purposes.
Specification	• The system should have the capability to collect and process user input, which includes origin, destination, and preferred time of travel. • The received user input should be validated to ensure it meets the required format and data constraints. • The system should utilize the processed user input for journey planning and demand forecasting purposes. • The user input should be incorporated into the data provided to algorithms and techniques used for demand forecasting.
Additional Notes	• The system should provide mechanisms for error handling and feedback to the user in case of invalid or incomplete input.

T6.4_UC4.2_FRQ02

Requirement ID	T6.4_UC4.2_FRQ02
Requirement Name	The Journey planning system must be able to fetch forecasted vehicle occupancy information for the proposed journeys based on the user's input.
Use Case ID	UC-FP6-WP6-4.02
Category	Functional
Priority	MUST
Main goal	Fetching Forecasted Vehicle Occupancy Information
Assumptions	The system assumes that there is a forecasted vehicle occupancy model available.
Specification	• The system should have the capability to fetch forecasted vehicle occupancy information for the proposed journeys based on the user's input. • The user's input, including origin, destination, and preferred time of travel, should be utilized to determine the relevant forecasted vehicle occupancy information. • The system should utilize the forecasted vehicle occupancy to generate accurate predictions for the proposed journeys.
Additional Notes	The accuracy of the fetched forecasted vehicle occupancy information will depend on the availability and quality of the forecasted vehicle occupancy model.

T6.4_UC4.2_NFRQ01

Requirement ID	T6.4_UC4.2_NFRQ01
Requirement Name	The trip planning interface should be user-friendly and intuitive, making it easy for travellers to enter their information
Use Case ID	UC-FP6-WP6-4.02
Category	Non-Functional
Priority	Nice to have
Main goal	Provide a User-Friendly and Intuitive Trip Planning Interface
Assumptions	• The system assumes that there is a trip planning interface through which users can enter their information. • The usability and intuitiveness of the interface will impact the user experience and ease of information entry.

Specification	• The trip planning interface should be designed in a user-friendly manner, with clear and intuitive navigation and controls. • The interface should provide clear instructions and guidance to users on how to enter their information, including origin, destination, and preferred time of travel. • The system should provide real-time feedback and validation to users to ensure the accuracy and completeness of the entered information. • The system should consider user preferences and customization options in the design of the trip planning interface.
Additional Notes	The user-friendliness and intuitiveness of the trip planning interface can enhance the overall user experience and encourage more users to utilize the system.

6.4. T6.4_CA04 Trip Option Generation and Display of forecasted occupancy information

The Trip Option Generation and Display of forecasted occupancy information capability focuses on the system's ability to generate multiple trip options based on user input and provide the available forecasted occupancy information for each trip. The interface is designed to be user-friendly and intuitive, ensuring travellers can easily understand the displayed occupancy information. Overall, this capability enhances the travel planning experience enabling travellers to make informed decisions.

T6.4_UC4.2_FRQ03

Requirement ID	T6.4_UC4.2_FRQ03
Requirement Name	The Journey Planning system must be able to consider forecasted occupancy information for the generation of trip options
Use Case ID	UC-FP6-WP6-4.02
Category	Functional
Priority	MUST
Main goal	The system must be able to generate several trip options based on the user's input and available forecasted occupancy information.
Assumptions	• The system assumes that it has access to the user's input, including origin, destination, and preferred time of travel. • The system also assumes that it has access to the available forecasted occupancy information for the relevant vehicles or modes of transportation.

Specification	• The system should utilize the user's input, including origin, destination, and preferred time of travel, to generate several trip options. • The trip options should consider the available forecasted occupancy information, considering the predicted occupancy levels for the relevant vehicles or modes of transportation. • The generated trip options should consider factors such as travel time, cost, and forecasted occupancy levels to provide a range of choices to the user. • The system should consider any additional preferences or constraints specified by the user in the input to refine the generated trip options.
Additional Notes	The accuracy and relevance of the generated trip options will depend on the availability and quality of the journey planning and forecasted occupancy information.

T6.4_UC4.2_FRQ04

Requirement ID	T6.4_UC4.2_FRQ04
Requirement Name	The Journey Planning system must be able to display the forecasted vehicle occupancy information to the traveller on the trip planning interface.
Use Case ID	UC-FP6-WP6-4.02
Category	Functional
Priority	MUST
Main goal	Displaying Forecasted Vehicle Occupancy Information
Assumptions	• The system assumes that there is forecasted vehicle occupancy information available for the relevant vehicles or modes of transportation. • The forecasted vehicle occupancy information will be relevant to the user's proposed journey based on their input.
Specification	• When planning a journey, the fetched forecasted vehicle occupancy information should be displayed to the traveller on the trip planning interface. • The system should present the generated trip options to the user in a clear and organized manner, highlighting key information such as estimated travel time, cost, and forecasted occupancy levels.

Additional Notes	• The accuracy and relevance of the displayed forecasted vehicle occupancy information will depend on the availability and quality of the forecasted occupancy model and data. • The display of the forecasted vehicle occupancy information can assist the traveller in making informed decisions about their journey and choosing the most suitable trip option.

T6.4_UC4.2_NFRQ02

Requirement ID	T6.4_UC4.2_NFRQ02
Requirement Name	The trip planning interface should be user-friendly and intuitive, making it easy for travellers to understand the displayed forecasted occupancy information.
Use Case ID	UC-FP6-WP6-4.02
Category	Non-Functional
Priority	Nice to have
Main goal	User-Friendly and Intuitive Trip Planning Interface
Assumptions	The system assumes that there is a trip planning interface through which travellers can access and view the forecasted occupancy information.
Specification	• The trip planning interface should be designed in a user-friendly manner, with clear and intuitive presentation of the forecasted occupancy information. • The displayed forecasted occupancy information should be easily understandable and relevant to the traveller's journey planning needs. • The interface should provide clear explanations and visual cues to help travellers interpret and comprehend the forecasted occupancy information.
Additional Notes	The user-friendliness and intuitiveness of the trip planning interface can enhance the overall user experience and facilitate the traveller's understanding of the forecasted occupancy information.

6.5. T6.4_CA05 Demand Estimation

The Demand Estimation capability focuses on the system's ability to estimate travel demand and provide reliable forecasted occupancy information. This capability includes training a machine learning model using historical travel demand data, combining various transportation modes in the surrounding areas. It should also include estimation of the demand for first/last mile connections between rail stations and nearby locations. It ensures reliability and availability for requesting forecasted occupancy information by the Transportation Service Provider (TSP). Overall, this capability provides travellers demand estimation that will support the TSP in making informed decisions and optimizing their transportation services offers.

T6.4_UC4.3_FRQ01

Requirement ID	T6.4_UC4.3_FRQ01
Requirement Name	The system must be able to train a machine learning model using historical travel demand data, combining rail and other modes in the surrounding areas.
Use Case ID	UC-FP6-WP6-4.03
Category	Functional
Priority	MUST
Main goal	The system must be able to train a machine learning model using historical travel demand data, combining rail and other modes (such as bus, trams, DRT or walking) in the surrounding areas, to estimate mobility demand beyond rail (first/last mile analysis).
Assumptions	• The system assumes that there is historical travel demand data available for the relevant areas, including rail and other modes of transportation. • The machine learning model will be trained using this historical data to estimate the mobility demand beyond rail, specifically for first/last mile analysis.
Specification	• The system should have the capability to access and utilize the historical travel demand data for the relevant areas. • The historical travel demand data should include information on journey request, and optionally other relevant factors for rail and other modes of transportation. • The machine learning model should be trained using the historical travel demand data, combining information from rail and other modes in the surrounding areas. • The training process should incorporate appropriate algorithms and techniques to analyse and learn from the historical data, capturing relevant patterns and trends specific to first/last mile mobility demand.

	• The trained machine learning model should be able to estimate the mobility demand beyond rail, providing insights and predictions for first/last mile analysis.
Additional Notes	

T6.4_UC4.3_FRQ02

Requirement ID	T6.4_UC4.3_FRQ02
Requirement Name	The system must be able to estimate the mobility demand for first/last mile connections between rail stations and surrounding locations.
Use Case ID	UC-FP6-WP6-4.03
Category	Functional
Priority	MUST
Main goal	Estimating Mobility Demand for First/Last Mile Connections
Assumptions	• The system assumes that it has access to relevant data such as rail station locations, surrounding locations, and journey planning requests. • The estimation of mobility demand will be based on this data to determine the level of demand for transportation options connecting rail stations and surrounding areas.
Specification	• The system should utilize available data on rail station locations and surrounding areas to estimate the mobility demand for first/last mile connections. • The estimation of mobility demand should consider factors such as travel patterns or other relevant variables if available (like population density or proximity to rail stations). • The system should incorporate appropriate algorithms and techniques to analyse the data and generate accurate estimates of mobility demand. • The estimation of mobility demand should provide insights into the expected volume of travellers requiring first/last mile transportation options, allowing stakeholders to understand the level of demand for first/last mile connections and make informed decisions. • The estimation of mobility demand should consider various scenarios and factors that may impact demand, such as time of day, day of the week, and special events.

Additional Notes	The accuracy and reliability of the estimated mobility demand will depend on the quality and availability of the data used in the estimation process.

T6.4_UC4.3_FRQ03

Requirement ID	T6.4_UC4.3_FRQ03
Requirement Name	The system must be able to export first/last mile demand forecast data.
Use Case ID	UC-FP6-WP6-4.03
Category	Functional
Priority	MUST
Main goal	Make first/last mile demand forecast data available so that it can be processed by other systems.
Assumptions	• The system has already derived first/last mile demand forecast data.
Specification	• The system must offer the user (TSP) the option to export first/last mile demand forecast data. • The system must enable the user (TSP) to specify the export, e.g. by a date range. • The export must happen in a typical and machine readable file format such as CSV or JSON.
Additional Notes	

T6.4_UC4.3_NFRQ01

Requirement ID	T6.4_UC4.3_NFRQ01
Requirement Name	The system should be reliable and available for request for forecasted occupancy information whenever required by the TSP.
Use Case ID	UC-FP6-WP6-4.03
Category	Non-Functional
Priority	Nice to have
Main goal	The system should be reliable and available for requests of forecasted occupancy information whenever required by the TSP
Assumptions	• The system assumes that there is a TSP who requires access to forecasted occupancy information. • The TSP should be able to request this information from the system whenever needed.

Specification	• The system should be designed and implemented to ensure high reliability, minimizing downtime and disruptions in providing forecasted occupancy information. • The system should have robust mechanisms for handling and processing these requests efficiently and in a timely manner. • The system should be able to handle multiple concurrent requests from different TSPs, ensuring fair and equitable access to the forecasted occupancy information.
Additional Notes	

6.6. T6.4_CA06 Demand analysis and identification

The Demand Analysis and Identification capability focuses on the system's ability to analyse the estimation results and identify areas with high demand and low offering. The system provides data insights to transport service providers through a demand analytics dashboard, enabling them to make data-driven decisions and optimize their services. It ensures the accuracy and completeness of available data to improve the accuracy of demand estimation. Overall, this capability enhances the understanding of demand patterns and supports transport service providers in meeting the demand effectively.

T6.4_UC4.3_FRQ04

Requirement ID	T6.4_UC4.3_FRQ04
Requirement Name	The system must be able to analyse the estimation results over time and identify areas with high demand and low offering.
Use Case ID	UC-FP6-WP6-4.03
Category	Functional
Priority	MUST
Main goal	The analysis of estimation results will be performed to identify areas where there is a high demand for transportation services but a low offering or availability, or vice versa.
Assumptions	• The system assumes that it has access to the estimation results, which include information on mobility demand and available transportation services.
Specification	• The system should analyse the estimation results to identify areas with high demand for transportation services. This analysis may consider other factors such as travel demand and population density. • The system should also assess the availability or offering of transportation services in these areas.

	• The analysis should compare the demand and offering in each area, allowing for the identification of areas with a significant disparity between demand and offering.
Additional Notes	• The analysis of estimation results is crucial in identifying areas where there is a mismatch between transportation demand and availability, helping to inform planning and decision-making processes. • It is important to consider the accuracy and reliability of the estimation results and ensure that the analysis takes into account any uncertainties or limitations in the data.

T6.4_UC4.3_FRQ05

Requirement ID	T6.4_UC4.3_FRQ05
Requirement Name	The system must provide data insights for transport service providers through a demand analytics dashboard.
Use Case ID	UC-FP6-WP6-4.03
Category	Functional
Priority	MUST
Main goal	Provision of Data Insights through a Demand Analytics Dashboard for TSPs
Assumptions	• The system assumes that there are transport service providers who require access to data insights for informed decision-making. • The data insights will be provided through a demand analytics dashboard.
Specification	• The system should have a demand analytics dashboard that presents data insights to TSPs. • The demand analytics dashboard should provide clear and understandable visualizations or reports that highlight areas with high demand and low offering of transportation services. • The data insights presented on the dashboard should be based on the analysis of estimation results, considering factors such as travel demand, population density, and other relevant variables. • The demand analytics dashboard should allow TSP to view and explore the data insights, enabling them to gain a comprehensive understanding of the demand and offering landscape.

	• The dashboard should provide interactive features, such as filters and drill-down capabilities, to allow users to customize their view and focus on specific areas or aspects of the data. • The system should present the results of the analysis in a clear and understandable manner, providing visualizations or reports that highlight areas with high demand and low offering.
Additional Notes	• The analysis of estimation results is crucial in identifying areas where there is a mismatch between transportation demand and availability, helping to inform planning and decision-making processes. • It is important to consider the accuracy and reliability of the estimation results and ensure that the analysis takes into account any uncertainties or limitations in the data. • The system should allow for a collaborative approach between transport service providers (multi tenants) in addressing areas with high demand and low offering.

T6.4_UC4.3_NFRQ02

Requirement ID	T6.4_UC4.3_NFRQ02
Requirement Name	The system must validate the accuracy and completeness of the available data to improve the accuracy of demand estimation.
Use Case ID	UC-FP6-WP6-4.03
Category	Non-Functional
Priority	MUST
Main goal	Accuracy and Completeness of Data for Improved Demand Estimation
Assumptions	• The system assumes that there is available data used for demand estimation. • Improving the accuracy of demand estimation relies on the accuracy and completeness of the sources of data.
Specification	• The system should implement mechanisms to validate the accuracy and reliability of the available data used for demand estimation. • This includes conducting data validation and verification processes to identify any inaccuracies or inconsistencies in the data.

	• The system should also incorporate data cleaning techniques to remove any irrelevant or erroneous data that may affect the accuracy of demand estimation.
Additional Notes	-

6.7. T6.4_CA07 Data Collection and Preprocessing

The Data Collection and Preprocessing capability focuses on the system's ability to gather and preprocess relevant data for anomaly detection and prediction regarding train usage peaks. This capability ensures that the system can collect all necessary information, including train occupancy data, weather conditions, and public/disruptive event data. Additionally, it performs preprocessing actions such as data cleaning and transformation to integrate the data into the Anomaly Detection and Prediction Component.

T6.4_UC4.4_FRQ01

Requirement ID	T6.4_UC4.4_FRQ01
Requirement Name	Data Collection and Preprocessing
Use Case ID	UC-FP6-WP6-4.04
Category	Functional
Priority	MUST
Main goal	The system must be able to collect and store the relevant data for anomaly detection and prediction
Assumptions	1) The system assumes that the following data is available: • Train Occupancy • Weather conditions • Public/disruptive events • Train Schedule • Other relevant data (optional) 2) The system assumes that the data received will be sufficient and accurate for anomaly detection and prediction
Specification	• The system should be capable of receiving data from different sources: train occupancy data (forecast or observed if available), weather data and event data. • The system should be capable of preprocessing data, including mechanisms for data cleaning and integration that combines data from different sources to provide a unified view. • The system should also be capable of handling other relevant data (e.g., train delays, historical maintenance, disruption logs, etc.) if available in order to improve the accuracy of the models.

Additional Notes	Additional quality sources may help to improve the quality of the detection and prediction.

T6.4_UC4.4_NFRQ01

Requirement ID	T6.4_UC4.4_NFRQ01
Requirement Name	Performance of data collection, preprocessing and storage of data for anomaly detection and prediction
Use Case ID	UC-FP6-WP6-4.04
Category	Non-Functional
Priority	MUST
Main goal	The system should be able to perform the collection, preprocessing and storage procedures in a reasonable time. The processed data will be consumed by anomaly detection and prediction models.
Assumptions	• The system assumes that data will be received from a PIS backend Platform (occupancy data) and other sources that cover weather data and public/disruptive event data. • The availability and format of the received data will be compatible with the system's processing capabilities. • The definition of "reasonable time" may vary depending on the specific requirements and expectations of the final application. • Other data may be available (e.g., train delays, historical maintenance, disruption logs, etc.)
Specification	• The system should have the capability to receive data from a PIS backend (occupancy data) and other sources that cover weather data and public/disruptive event data. • The received data should be processed efficiently and effectively within a reasonable time frame. • The system should have the capability to handle and process large volumes of data to ensure accurate anomaly detection and prediction. • The processed data should be made available for consumption by anomaly detection and prediction models.
Additional Notes	-

6.8. T6.4_CA08 Anomaly Detection in Train Occupancy

The Anomaly Detection in Train Occupancy capability focuses on the system's ability to detect unusual positive or negative spikes in train occupancy data based on patterns found within historical data. This capability involves fitting a model to the data to learn these underlying patterns and structures, aiming to produce a labelled dataset containing the identified anomalies, which will be included in the dataset used to train the anomaly prediction model.

T6.4_UC4.4_FRQ02

Requirement ID	T6.4_UC4.4_FRQ02
Requirement Name	Model training for anomaly detection
Use Case ID	UC-FP6-WP6-4.04
Category	Functional
Priority	MUST
Main goal	The system must be able to train the anomaly detection model using train occupancy data, weather data, and public/disruptive event data.
Assumptions	1. The system assumes that the following data is available for training the anomaly detection model: • Train Occupancy • Train Schedule 2. The availability and quality of the training data will impact the accuracy and effectiveness of the trained model. 3. The system assumes that the machine learning model will be trained using appropriate algorithms and techniques to achieve the desired outputs.
Specification	• The system should be capable of implementing mechanisms for data transformation (e.g., encoding) and feature engineering in order to prepare data to be used by a machine learning and/or statistical model for anomaly detection. • The system should utilize appropriate machine learning algorithms to train the model based on the provided data. • The trained model should be able to analyse and accurately detect anomalies based on the provided data. • The trained model should be able to produce/update a labelled dataset with a feature for the detected anomalies (e.g., a column of binary values for whether it is an anomaly).

Additional Notes	• The availability and quality of the training data will significantly impact the accuracy and effectiveness of the trained machine learning model. • In unsupervised learning techniques for anomaly detection, the term "training" is used differently compared to supervised learning. In this context, "training" often refers to the process of fitting the model to the data to learn the underlying patterns and structures.

T6.4_UC4.4_FRQ03

Requirement ID	T6.4_UC4.4_FRQ03
Requirement Name	Model update for anomaly detection
Use Case ID	UC-FP6-WP6-4.04
Category	Functional
Priority	MUST
Main goal	The system must be able to process the data and update the anomaly detection model considering new incoming data.
Assumptions	1. The system assumes that the following data is available for training the anomaly detection model: • Train Occupancy • Train Schedule 2. The availability and quality of the training data will impact the accuracy and effectiveness of the trained model. 3. The system assumes that historical data will be used as a reference for updating the anomaly detection model.
Specification	• The system should have the capability to receive new data, such as train occupancy data, for updating the anomaly detection model. • The received new data should be pre-processed and integrated with the existing historical data to create an updated dataset. • The system should utilize appropriate techniques and algorithms to update the anomaly detection model based on the new and historical data. • The updated anomaly detection model should take into account the patterns observed in the historical data, as well as the new data, to accurately detect anomalies.

	• The trained model should be regularly updated and refined to improve the accuracy of the detected anomalies.
Additional Notes	• The availability and quality of the training data will significantly impact the accuracy and effectiveness of the trained machine learning model.

T6.4_UC4.4_NFRQ02

Requirement ID	T6.4_UC4.4_NFRQ02
Requirement Name	Model update lifecycle for anomaly detection
Use Case ID	UC-FP6-WP6-4.04
Category	Non-Functional
Priority	Nice-to-have with high priority
Main goal	The system should keep the anomaly detection model as updated as possible.
Assumptions	The system assumes that the anomaly detection model can be updated with new data.
Specification	• The system should have mechanisms in place to regularly update the anomaly detection model with new data. • The system should prioritize the speed and efficiency of updating the model to ensure timely availability of updated dataset containing anomaly labelled data.
Additional Notes	The frequency and timing of the model updates should be balanced with the availability and quality of the new data.

T6.4_UC4.4_NFRQ03

Requirement ID	T6.4_UC4.4_NFRQ03
Requirement Name	Performance evaluation of anomaly detection model
Use Case ID	UC-FP6-WP6-4.04
Category	Non-Functional
Priority	Nice-to-have with high priority
Main goal	The anomaly detection model should achieve high accuracy, minimizing false positives and false negatives
Assumptions	The system should have mechanisms for evaluating the training process of the model, including accuracy metrics and validation.
Specification	• The system should be capable of conducting a series of accuracy metric tests (e.g., accuracy, precision) for the anomaly detection model.

	• The system should be capable of performing validation tests (e.g., Cross-Validation). • The system should be capable of fine-tuning the model parameters in order to improve its performance.
Additional Notes	-

6.9. T6.4_CA09 Predictive Modelling of Anomalies

The Predictive Modelling of Anomalies capability focuses on the system's ability to predict future anomalies related to unusual spikes in train occupancy data based on an analysis of patterns found within historical data, including the influence of external factors like weather conditions and public/disruptive events. Additionally, this capability ensures that the predicted anomalies are provided with contextual information regarding the influencing conditions under which they might occur in the future.

T6.4_UC4.4_FRQ04

Requirement ID	T6.4_UC4.4_FRQ04
Requirement Name	Model training for anomaly prediction
Use Case ID	UC-FP6-WP6-4.04
Category	Functional
Priority	MUST
Main goal	The system must be able to train the anomaly prediction model using anomaly data, weather data, and public/event data
Assumptions	1. The system assumes that the following data is available for training the anomaly detection model: • Weather conditions • Public/disruptive events • Anomaly labelled data (dataset generated by the anomaly detection model) • Train Schedule • Other relevant data (optional) 2. The availability and quality of the training data will impact the accuracy and effectiveness of the trained model. 3. The system assumes that the machine learning model will be trained using appropriate algorithms and techniques to achieve the desired outputs.
Specification	• The system should be capable implementing mechanisms for data transformation (e.g., encoding) and feature engineering in order to prepare data to be used by a

	machine learning and/or statistical model for anomaly prediction. • The system should utilize appropriate machine learning algorithms to train the model based on the provided data. • The trained model should be able to analyse and accurately predict anomalies based on the provided data, including weather conditions and public/disruptive events data.
Additional Notes	The availability and quality of the training data will significantly impact the accuracy and effectiveness of the trained machine learning model.

T6.4_UC4.4_FRQ05

Requirement ID	T6.4_UC4.4_FRQ05
Requirement Name	Model update for anomaly prediction
Use Case ID	UC-FP6-WP6-4.04
Category	Functional
Priority	MUST
Main goal	The system must be able to process the data and update the anomaly prediction model considering new incoming data.
Assumptions	1. The system assumes that the following data is available for training the anomaly detection model: • Weather conditions • Public/disruptive events • Anomaly labelled data (dataset generated by the anomaly detection model) • Train Schedule • Other relevant data (optional) 2. The availability and quality of the training data will impact the accuracy and effectiveness of the trained model. 3. The system assumes that historical data will be used as a reference for updating the anomaly prediction model.
Specification	• The system should have the capability to receive new data, such as train weather conditions and public/disruptive event data, for updating the anomaly prediction model. • The received new data should be pre-processed and integrated with the existing historical data to create an updated dataset.

	• The system should utilize appropriate techniques and algorithms to update the anomaly prediction model based on the new and historical data. • The updated anomaly prediction model should take into account the patterns observed in the historical data, as well as the new data, to accurately predict anomalies. • The updated anomaly prediction model should be stored and made available for future anomaly prediction and analysis. • The trained model should be regularly updated and refined to improve the accuracy of the predicted anomalies.
Additional Notes	The availability and quality of the training data will significantly impact the accuracy and effectiveness of the trained machine learning model.

T6.4_UC4.4_FRQ06

Requirement ID	T6.4_UC4.4_FRQ06
Requirement Name	Predict new anomalies for upcoming data covering weather conditions and public/disruptive events
Use Case ID	UC-FP6-WP6-4.04
Category	Functional
Priority	MUST
Main goal	The system must be able to receive a new data covering weather conditions and public/disruptive events and apply the trained model to predict new anomalies for the specified timeframe.
Assumptions	• The system assumes that a trained model for predicting anomalies based weather conditions, public/disruptive event data and historical data is available. • The accuracy of the trained model will impact the accuracy of the anomaly prediction for the input data. • The system assumes that the data received will provide the necessary information for predicting anomalies, such as weather conditions and public/disruptive events data.
Specification	• The system should have the capability to receive new data covering weather conditions and public/disruptive events data. • The received data should be processed and utilized by the trained model to predict the anomalies for the specified timeframe.

	• The system should apply the trained model to provide an accurate prediction of the anomalies in the specified timeframe. • The prediction of anomalies should consider the patterns and trends observed in the historical data used to train the model. • The system should perform a contextual analysis to provide the predicted anomalies with relevant information, including weather conditions, public/disruptive events and temporal features associated with each anomaly.
Additional Notes	• The accuracy of the anomaly prediction will depend on the availability and quality of the trained model and the data used for training.

T6.4_UC4.4_NFRQ04

Requirement ID	T6.4_UC4.4_NFRQ04
Requirement Name	Model update lifecycle for anomaly prediction
Use Case ID	UC-FP6-WP6-4.04
Category	Non-Functional
Priority	Nice-to-have with high priority
Main goal	The system should keep the anomaly prediction model as updated as possible.
Assumptions	The system assumes that the anomaly prediction model can be updated with new data.
Specification	• The system should have mechanisms in place to regularly update the anomaly prediction model with new data. • The system should prioritize the speed and efficiency of updating the model to ensure timely availability of updated predictions.
Additional Notes	The frequency and timing of the model updates should be balanced with the availability and quality of the new data.

T6.4_UC4.4_NFRQ05

Requirement ID	T6.4_UC4.4_NFRQ05
Requirement Name	Performance evaluation of anomaly prediction model
Use Case ID	UC-FP6-WP6-4.04

Category	Non-Functional
Priority	Nice-to-have with high priority
Main goal	The anomaly prediction model should achieve high accuracy, minimizing false positives and false negatives
Assumptions	The system should have mechanisms for evaluating the training process of the model, including accuracy metrics and validation.
Specification	• The system should be capable of conducting a series of accuracy metric tests (e.g., accuracy, precision) for the anomaly prediction model. • The system should be capable of performing validation tests (e.g., Cross-Validation). • The system should be capable of fine-tuning the model parameters in order to improve its performance.
Additional Notes	-

6.10. T6.4_CA10 Generation and Delivery of structured messages covering the contextual information of the identified anomalies

The Generation and Delivery of structured messages covering the contextual information of the identified anomalies capability focuses on the system's ability to construct template messages containing contextual information for each anomaly prediction, including weather conditions, public/disruptive events and temporal features, and deliver them to the TMS (via notification to a TMS Dashboard) in order to support informed decision-making for potential service adjustments.

T6.4_UC4.4_FRQ07

Requirement ID	T6.4_UC4.4_FRQ07
Requirement Name	Generation of structured message with contextual information for predicted anomalies
Use Case ID	UC-FP6-WP6-4.04
Category	Functional
Priority	MUST
Main goal	The system must gather contextual information on anomaly predictions covering weather conditions, public/disruptive events and temporal features, and generate structured messages for each predicted anomaly to be delivered to the TMS
Assumptions	• The system assumes that contextual information for predicted anomalies is available. • The system assumes that contextual information is in a format that can be easily integrated in other systems.

Specification	• The system should be capable of receiving the predicted anomalies containing the contextual information covering weather conditions, public/disruptive events and temporal features. • The system should be capable of compiling the contextual information and structure it in a known message delivery format (e.g., JSON). • The system should be capable of delivering the structured message to the TMS, which is expected to be received by means of a notification. • The system should be capable of integrating, if available, more contextual information like historical comparisons regarding similar conditions in order to provide context on the frequency of anomalies.
Additional Notes	-

T6.4_UC4.4_NFRQ06

Requirement ID	T6.4_UC4.4_NFRQ06
Requirement Name	Performance of structured message generation regarding contextual information for predicted anomalies
Use Case ID	UC-FP6-WP6-4.04
Category	Non-Functional
Priority	MUST
Main goal	The system should be able to generate structured messages for predicted anomalies within a reasonable time.
Assumptions	The system assumes that contextual information for predicted anomalies is available.
Specification	• The system should be capable of implementing mechanism to automatically generate a structured message based on the predicted anomalies, ensuring that this process is completed within a reasonable timeframe. • The system should be capable of implementing a mechanism to ensure that the contextual information for predicted anomalies is interpretable.
Additional Notes	-

T6.4_UC4.4_NFRQ07

Requirement ID	T6.4_UC4.4_NFRQ07
Requirement Name	Flexible design for structured message generation regarding contextual information for predicted anomalies
Use Case ID	UC-FP6-WP6-4.04
Category	Non-Functional
Priority	Nice-to-have with high priority
Main goal	The system should be designed for easy maintenance and updates, allowing for the addition of new message templates.
Assumptions	The system assumes that there is a configuration-driven design, allowing the utilization of configuration files to change/update message templates.
Specification	• The system should be capable of integrating and storing new message templates to be used for generating structured messages with contextual information regarding predicted anomalies. • The system should be capable of ensuring the interpretability of the message template. • The system should be capable of ensuring that the format and main structure (e.g., primary fields) of the message templates does not change throughout the different versions.
Additional Notes	-

7. Logical Architecture

7.1. High Level Architecture

The purpose of this section is to present the logical architecture diagrams of the system, as defined within the framework of the Arcadia methodology. The logical architecture diagrams provide a comprehensive overview of the system's components, functions, and relations. This section will aid in understanding the structure and organization of the system, highlighting the interdependencies and interactions among its various elements.

The logical architecture diagrams presented distinguish between blue components and white components. Blue components represent external elements that are not developed within the scope of this deliverable task. These components may include external systems, interfaces, or dependencies that the system under analysis relies upon. On the other hand, white components denote the elements that are developed specifically within the scope of this deliverable task. These components encompass the core functionalities and modules that are being analysed for demand analysis and will be described on the next section.

The diagrams provide a visual representation that aids in communication and facilitates discussions among project teams, allowing for a more holistic and efficient analysis.

At first Figure 2, present a high-level view of the main blocks developed with WP6. On this deliverable we are focused on the details of the Demand Forecast block and the interfaces with TMS and the Congestion Monitoring blocks. The Journey Planning System, DRT service provider and DRT simulation systems are specified in Task 6.1, while the Congestion Monitoring and Feedback App are specified in Task 6.5.

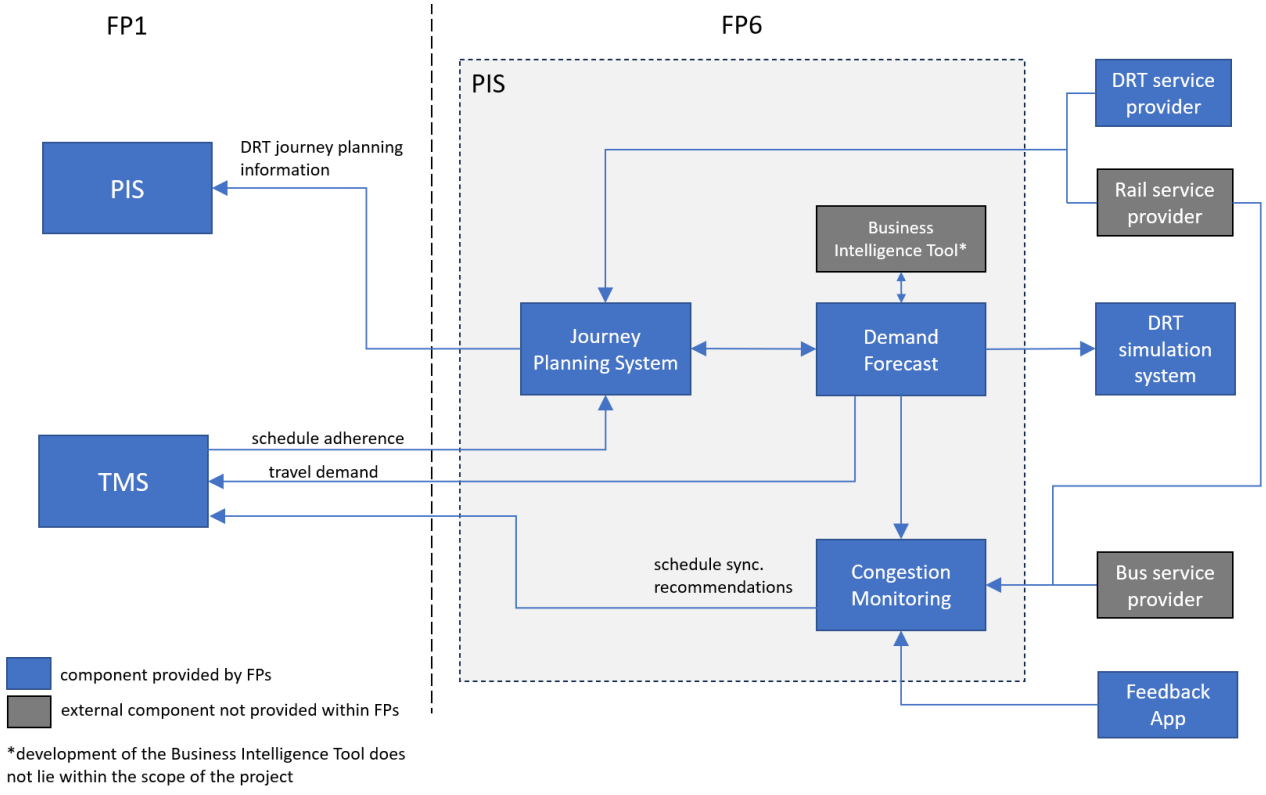

Figure 2 – High level view of WP6 system architecture

Now, on Figure 3 and Figure 3Figure 4 the logical architecture diagram of the components involved in short- and long-term demand forecast is described.

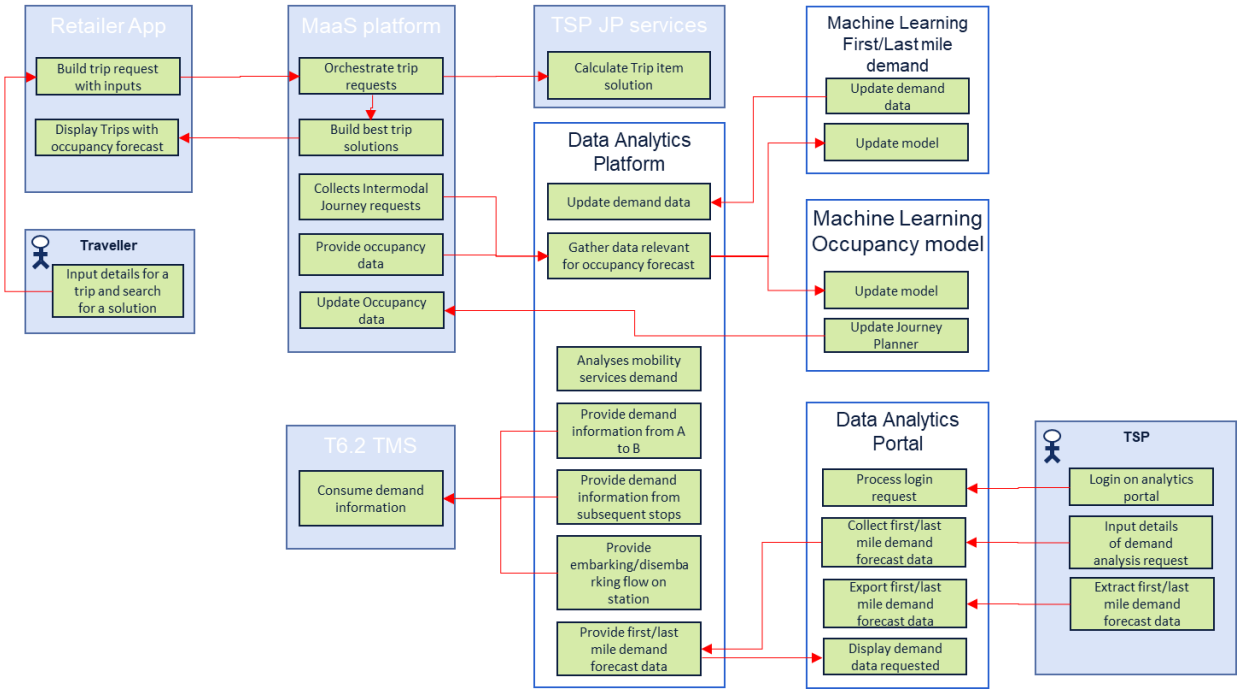

Figure 3: Architecture diagram for demand forecast

The diagram in Figure 3 presents a high-level view of the component and functions involved on the use cases related to demand forecast.
As input to the main system component “Data analytics platform” is the data generated by the MaaS platform, with focus on the journey planning request gathered through the retailer app. But if available also occupancy data collected from other sources can be used.
All the data gathered is used to train the vehicles occupancy model and the first/last mile demand model, which can be used to provide demand information back to the MaaS platform to inform traveller or to the data analytics portal to provide insights to the TSPs, or to provide demand information to traffic management systems (TMS).

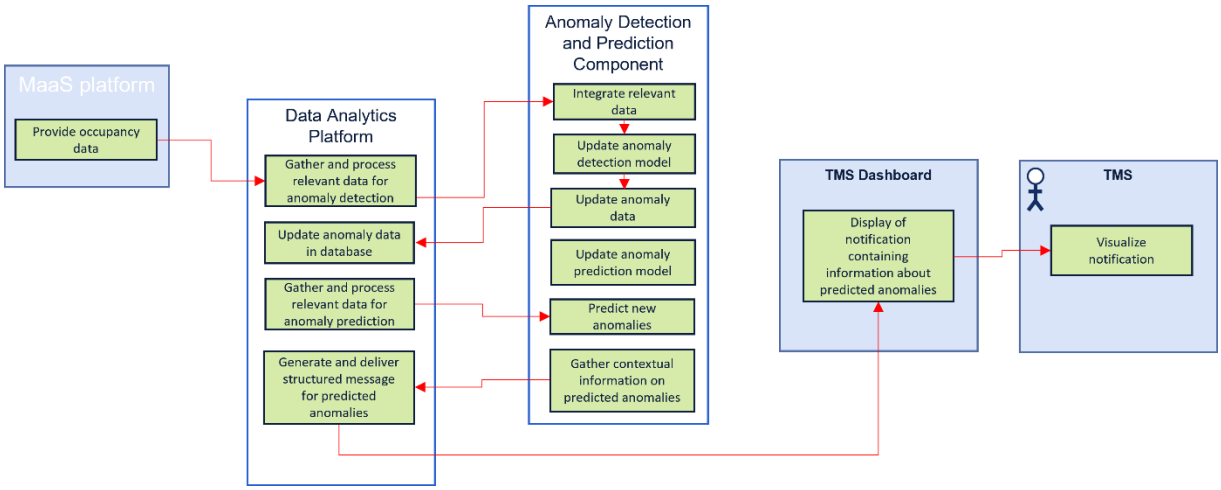

Figure 4: Architecture diagram for detection and characterization of abnormal train usage peaks

Figure 4 presents a high-level view of the components and functions involved in the anomaly detection and prediction related to unusual spikes in train usage. The Data Analytics Platform plays a central role in this system due to its capability of gathering relevant data to feed into the Anomaly Detection and Prediction Component. This data will be used by the anomaly detection and prediction models for training purposes and to predict future anomalies when upcoming events (weather and public/disruptive events) are received as data. Additionally, the Data Analytics Platform is responsible for producing and providing structured messages to the TMS containing contextual information regarding the predicted anomalies. The TMS operator is informed by these details via a notification message received in the TMS Dashboard.

7.2. Components and functions

This section includes a description of all components that will be developed in FP6 FutuRe project in Task 6.4. Each component is responsible for one or more functions. Components and functions are linked with functional and non-functional requirements respectively, introduced in Chapter 0.

7.2.1. Data Analytics Platform

Component Name	***Component Description***	***NFRQ ID***
Data Analytics Platform	The Data Analytics Platform repeatedly processes input like journey planning requests or other information sources if available to create occupancy forecasts.	- T6.4_UC4.1_NFRQ01 - T6.4_UC4.3_NFRQ01 - T6.4_UC4.3_NFRQ02
Function Name	***Function Description***	***FRQ ID***
Gather data relevant for occupancy forecast	As input for the forecast calculation, the Data Analytics Platform receives the journey planning requests from the MaaS Platform and optionally may request additional information from the MaaS Platform, such as counted passenger information from the vehicles, which contributes to higher quality forecasts. This data is then forwarded to the Machine Learning Occupancy Model and the Machine Learning first/last mile demand components.	- T6.4_UC4.1_FRQ01 - T6.4_UC4.1_FRQ03 - T6.4_UC4.1_FRQ04
Analyses mobility services demand	Journey requests to the MaaS Platform are forwarded to this function of the Data Analytics Platform to analyse the demand as the ground data for the demand analysis calculation.	- T6.4_UC4.1_FRQ02 - T6.4_UC4.3_FRQ02 - T6.4_UC4.3_FRQ03
Update demand data	Whenever the Machine Learning module for first/last mile demand gets data updates, this function will receive, process, and store them.	- T6.4_UC4.3_FRQ02
Provide first/last mile forecast demand	When requested to provide insights into first/last mile forecast demand, the Data Analytics Platform returns the calculated demand data.	- T6.4_UC4.3_FRQ05
Provide demand information from A to B	This function formats the demand data between two points in the transport network in a structured way to interface with the TMS model.	T6.2 requirements
Provide demand	This function formats the demand data from	T6.2 requirements

information from subsequent stops	subsequent stops in a structured way to interface with the TMS model.	
Provide embarking/disembarking flow on station	This function formats the estimated flow of people embarking and disembarking on a station, the data is provided in a structured way to interface with the TMS model.	T6.2 requirements

Table 2: Data Analytics Platform component

7.2.2. Data Analytics Portal

Component Name	*Component Description*	*NFRQ ID*
Data Analytics Portal	The Data Analytics Portal is the frontend component of the Data Analytics Platform and provides a user interface to the information provided by the Data analytics platform in various graphical forms to the data analyst.	- T6.4_UC4.3_NFRQ02
Function Name	***Function Description***	***FRQ ID***
Process login request	Checks authenticity and authorization of the representative of the TSP.	- T6.4_UC4.3_FRQ05
Collect first/last mile demand forecast data	Requests the Data Analytics Platform to provide the analysed demand for first/last mile.	- T6.4_UC4.3_FRQ05
Display demand data request	After collecting the demand forecast, this component process and display the results provided by the Data Analytics Platform.	- T6.4_UC4.3_FRQ05
Export first/last mile demand forecast data	Allows users to extract and export forecasted demand data related to first and last mile transportation connections.	- T6.4_UC4.3_FRQ03

Table 3: Data Analytics Portal component

7.2.3. Machine Learning Occupancy model

Component Name	*Component Description*	*NFRQ ID*
Machine Learning Occupancy model	The Machine Learning Occupancy Model is used to calculate and regularly update the occupancy forecasts used by other components, such as the MaaS platform or the Data Analytics Platform.	- T6.4_UC4.1_NFRQ02
Function Name	***Function Description***	***FRQ ID***
Update model	The Machine Learning Occupancy Model receives input data from the Data Analytics Platform and recalculates the occupancy forecast model regularly.	- T6.4_UC4.1_FRQ03 - T6.4_UC4.1_FRQ04
Update Journey planner	When the Machine Learning Occupancy Model recalculates its occupancy forecast, it provides this new model to the MaaS Platform, which incorporates it into its journey planning.	- T6.4_UC4.1_FRQ05 - T6.4_UC4.2_FRQ02

Table 4: Machine Learning Occupancy model component

7.2.4. Machine Learning First/Last mile demand

Component Name	*Component Description*	*NFRQ ID*
Machine Learning Occupancy model	The Machine Learning Occupancy Model is used to calculate and regularly update the occupancy forecasts used by other components, such as the MaaS platform or the Data Analytics Platform.	- T6.4_UC4.3_NFRQ01
Function Name	***Function Description***	***FRQ ID***
Update model	The Machine Learning for first/last demand calculations receives input data from the Data Analytics Platform and provides updated demand data to the platform.	- T6.4_UC4.3_FRQ01
Update demand data	When the Machine Learning Occupancy Model recalculates its demand forecast for first/last mile, it provides this new model to the MaaS Platform.	- T6.4_UC4.3_FRQ02

Table 5: Machine Learning Occupancy model component

7.2.5. MaaS platform

Component Name	*Component Description*	*NFRQ ID*
MaaS platform (Provided by Task 6.1)	The MaaS Platform orchestrates the services of multiple connected TSP or other MaaS platforms.	- T6.4_UC4.1_NFRQ01
Function Name	***Function Description***	***FRQ ID***
Orchestrate trip requests	Upon receiving a journey planning request, the MaaS platform identifies the relevant TSP's Journey Planning Services and orchestrates these external services to receive the results. Optionally, this orchestration may include calls toward other MaaS platforms as well.	- T6.4_UC4.2_FRQ01
Build best trip solutions	When the MaaS platform receives the results from the TSP Journey Planning service and MaaS platforms, it analyses and builds the best overall trip solutions. The results are provided to the Retailer App.	- T6.4_UC4.2_FRQ03
Collect intermodal Journey requests	The MaaS Platform forwards the journey requests to the Data Analytics platform.	- T6.4_UC4.1_FRQ01
Provide occupancy data	In addition to the journey planning requests, a MaaS platforms may offer additional occupancy data, such as counting data for the vehicles. This data is provided via this function to the Data Analytics Platform.	- T6.4_UC4.1_FRQ01
Update Occupancy data	Whenever the Machine Learning Occupancy Model updates its forecasted model for occupancy, this information is provided to the MaaS platform through this function, which	- T6.4_UC4.2_FRQ02

	incorporates the information into journey planning. This way, the occupancy information is provided because of the journey planning to the Retailer App.	

Table 6: MaaS platform component

7.2.6. Retailer app

Component Name	*Component Description*	*NFRQ ID*
Retailer app (Provided by Task 6.1)	The Retailer App is the traveller and purchaser facing component to interact with the MaaS Platform. The Retailer App is used for journey planning and occupancy use cases.	- T6.4_UC4.2_NFRQ01 - T6.4_UC4.2_NFRQ02
Function Name	***Function Description***	***FRQ ID***
Build trip request with inputs	Create a journey planning request from the input from the Traveller.	- T6.4_UC4.2_FRQ01
Display Trips with occupancy forecast	When the results of the journey planning request contain occupancy forecast information, it is displayed with the trips as well.	- T6.4_UC4.2_FRQ04

Table 7: Retailer app component

7.2.7. Anomaly Detection and Prediction Component

Component Name	*Component Description*	*NFRQ ID*
Anomaly Detection and Prediction Component	The Anomaly Detection and Prediction Component comprises the anomaly detection and prediction models and it is responsible for receiving and processing relevant data, including train occupancy, weather conditions and public/disruptive events to accurately detect and predict anomalies, i.e. unusual peaks in train usage. The predicted anomalies are used by other components, such as the Data Analytics Platform and the TMS.	- T6.4_UC4.4_NFRQ02 - T6.4_UC4.4_NFRQ03 - T6.4_UC4.4_NFRQ04 - T6.4_UC4.4_NFRQ05
Function Name	***Function Description***	***FRQ ID***
Integrate relevant data	The relevant data is received and prepared to properly feed the anomaly detection and prediction models for training purposes.	- T6.4_UC4.4_FRQ02 - T6.4_UC4.4_FRQ04
Update anomaly detection model	The anomaly detection model receives input data covering train occupancy, performs a fitting procedure to learn the underlying patterns and structures and produces a labelled dataset with the	- T6.4_UC4.4_FRQ03

	detected anomalies.	
Update anomaly data	The labelled dataset is updated when the anomaly detection model fits to new input data and it is provided to the Data Analytics Platform to store it.	- T6.4_UC4.4_FRQ03
Update anomaly prediction model	The anomaly prediction model receives input data covering weather conditions and public/disruptive events, performs a training procedure to learn the underlying patterns and structures and it is ready to receive new upcoming data to predict anomalies.	- T6.4_UC4.4_FRQ05
Predict new anomalies	The anomaly prediction model receives new data covering upcoming weather conditions and public/disruptive events and predicts new anomalies for a specified timeframe.	- T6.4_UC4.4_FRQ06
Gather contextual information on predicted anomalies	Alongside predicting new anomalies for a specified timeframe, contextual information, covering weather conditions and public/disruptive events, is gathered and handed to other components, such as Data Analytics Platform, to be further delivered to the TMS for supporting informed decision-making.	- T6.4_UC4.4_FRQ07

Table 8: Anomaly Detection and Prediction component

7.2.8. TMS Dashboard

Component Name	***Component Description***	***NFRQ ID***
TMS Dashboard	The TMS Dashboard is the component used by TMS operators to visualize data related to the predicted anomalies and its contextual information in order to support informed decision-making.	- T6.4_UC4.4_NFRQ06 - T6.4_UC4.4_NFRQ07
Function Name	***Function Description***	***FRQ ID***
Display of notification containing information about predicted anomalies	The TMS Dashboard allows TMS operators to receive notifications, where they can access messages regarding predicted anomalies and their contextual information, which can be useful to support informed decision-making.	- T6.4_UC4.4_FRQ07

Table 9: TMS Dashboard component

7.3. Exchange scenario (per use case)

This section focuses on describing the exchange scenarios for each use case at the logical architecture level, following the Arcadia methodology. These scenarios provide a detailed analysis of the information exchanges and interactions between components during the execution of each use case. Understanding the exchange scenarios is crucial for comprehending the flow of information and dependencies within the system. By examining these scenarios, stakeholders gain insights into how the system handles inputs, processes data, and generates outputs in response to user actions or external events. Those helped to identify bottlenecks, dependencies, and areas for improvement, ensuring that the system meets requirements and delivers desired outcomes.
Each use case is individually analysed, and the corresponding exchange scenarios are presented in a structured manner. This approach enables focused examination of interactions and information flow, providing a concise representation of the system's behaviour.

7.3.1. ES4.01 - Forecast Occupancy of Vehicles using Journey Planning Requests Data

The exchange scenario for use case UC-FP6-WP6-4.01 is shown in Figure 5. The MaaS Platform continuously forwards the journey planning requests to the Data Analytics Platform which optionally, in addition, may retrieve occupancy data from the MaaS Platform to gather relevant data for occupancy forecasting. This gathered data is forwarded to the Machine Learning Occupancy Model which updates itself accordingly. After the update is done, the new model is forwarded to the MaaS Platform which incorporates the information into its journey planning functionality.

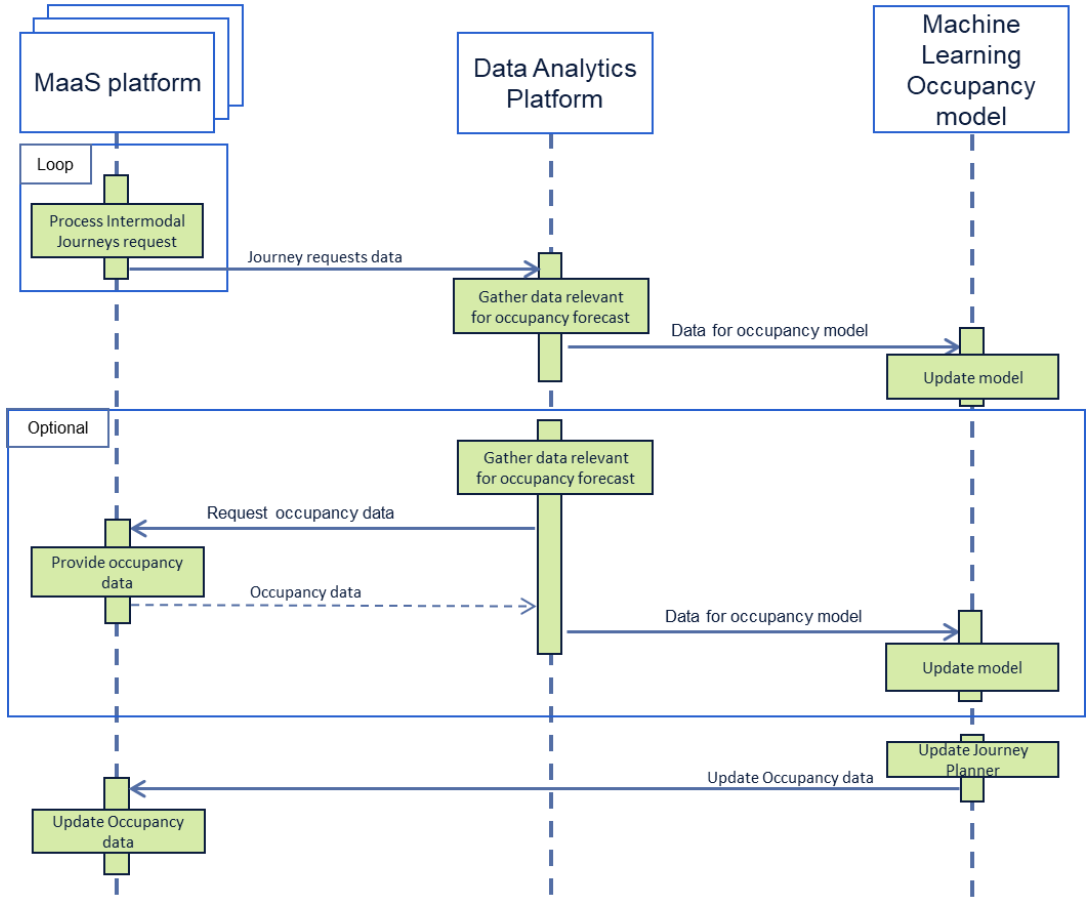

Figure 5: Exchange Scenario 4.01

7.3.2. ES4.02 - Display Forecasted Occupancy Information to Travelers when Planning Trips

Figure 6 depicts the exchange scenario for use case UC-FP6-WP6-4.02. When a traveller requests a trip via the Retailer App from the MaaS Platform, it orchestrates the journey plan across integrated TSP Journey Planning Services. The result may contain the occupancy information for each trip leg if the TSP Journey Planning Services and the MaaS Platform have incorporated the previously created occupancy data model.

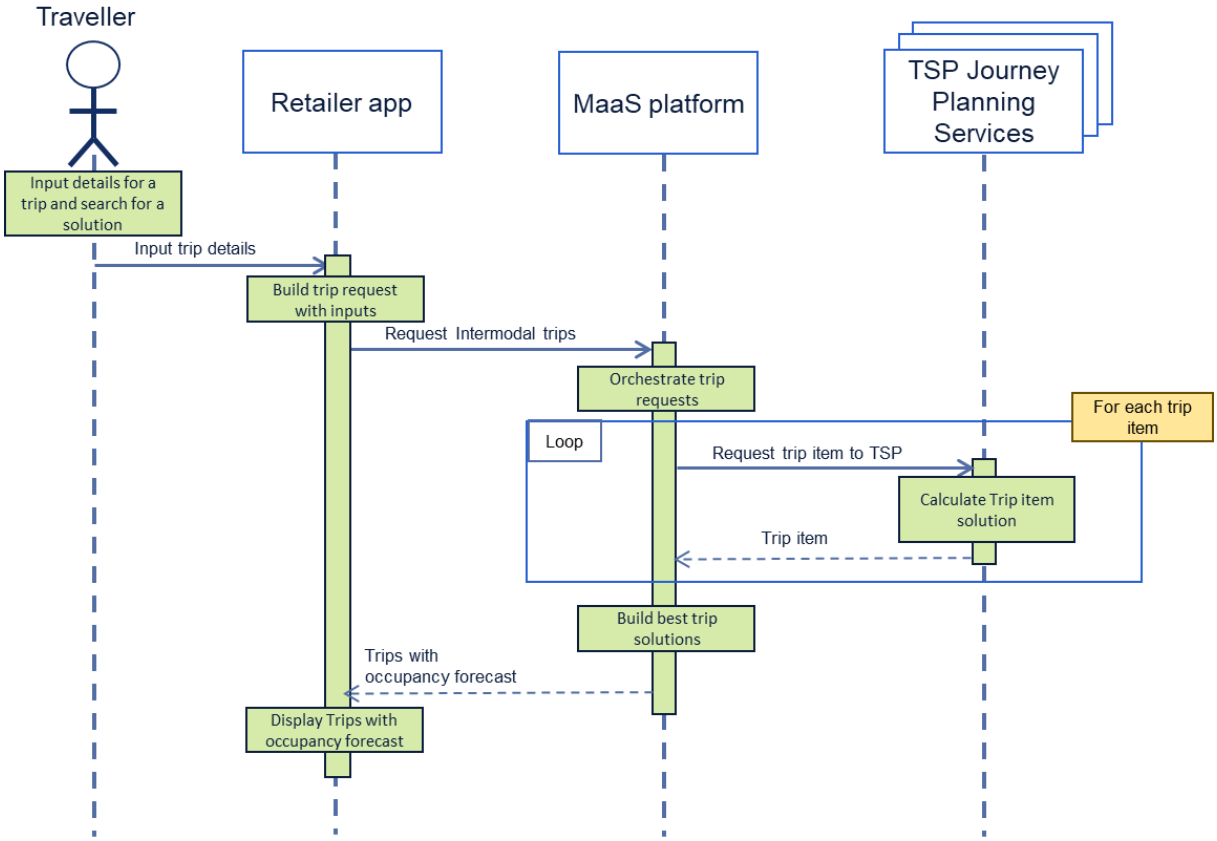

Figure 6: Exchange Scenario 4.02

7.3.3. ES4.03 - Estimation of Mobility Demand beyond Rail (First/Last Mile Analysis)

Figure 7 illustrates the sequence diagram for use case UC-FP6-WP6-4.03. As in the first use case, the MaaS Platform continuously forwards the journey planning requests to the Data Analytics Platform which this time, are forwarded to a different model that will focus on first/last mile connection demand, comparing the offer and demand of the traveller on different regions. The Machine Learning Model which updates itself accordingly with all the new data, will also update the MaaS Platform with the relevant information for first/last mile demand analysis.
The TSP playing the role of Data Analyst, can access the Data Analytics Portal to inspect the demand data. The Data Analytics Portal requests all the demand information to the Data Analytics Platform which may provide occupancy forecasts, and identify patterns and trends, which as specified on the use case will include demand analysis for first/last mile services. The Data Analytics Portal Dashboard displays the result to the Data Analyst.

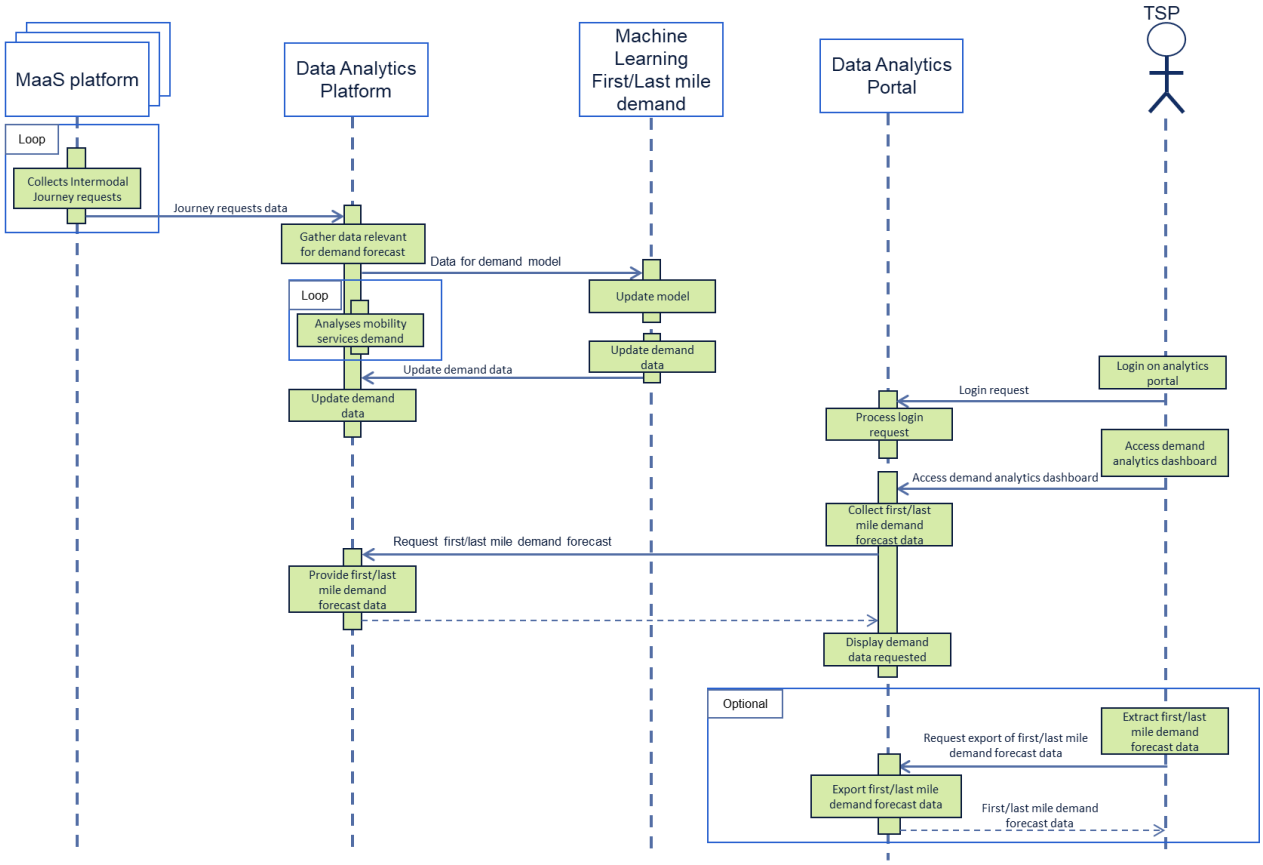

Figure 7: Exchange Scenario 4.03

7.3.4. ES4.04 – Detection and Characterization of Abnormal Train Usage Peaks

Components, functions and data flows required for use case UC-FP6-WP6-4.04 are shown in Figure 8. The Data Analytics Platform gathers and processes relevant data necessary for anomaly prediction. This data is then sent to the Anomaly Detection and Prediction Component, which uses it to predict new anomalies. The component gathers contextual information on these predicted anomalies and generates a structured message that is forwarded to the TMS Dashboard. The TMS Dashboard displays a notification containing the information about the predicted anomalies, allowing the TMS operator to visualize the notification.

Simultaneously, the Data Analytics Platform continues to gather and process data for anomaly detection and prediction models. This data is integrated and used to update the anomaly detection and prediction models. The anomaly data is also updated in the database, ensuring that the latest information is available for future predictions and notifications.

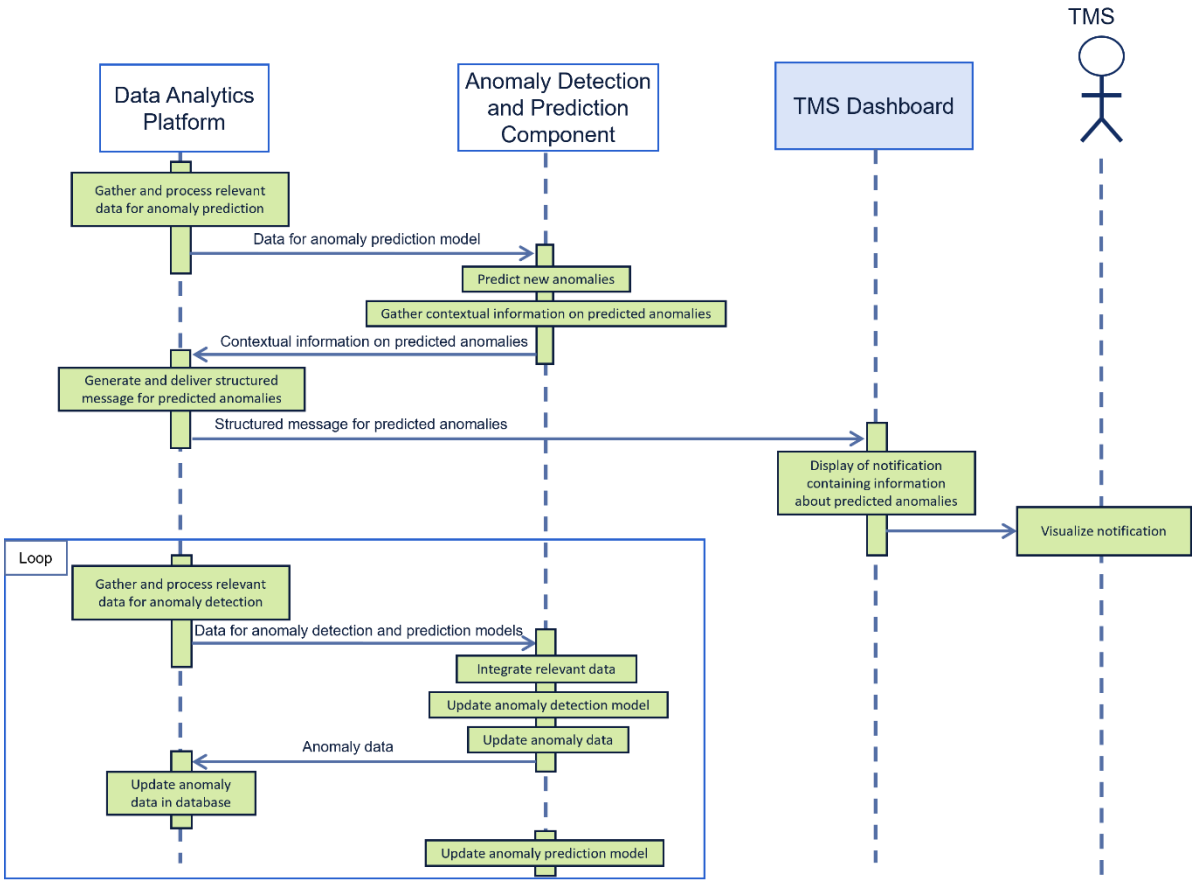

Figure 8: Exchange Scenario 4.04

8. Interfaces & standards

8.1. Interface between data analytics platform and Anomaly Detection and Prediction component

This interface facilitates the exchange of demand forecast data between the Data Analytics Platform and the Anomaly Detection and Prediction Component. It will enable anomaly detection and prediction under the scope of unusual spikes in train usage, contextual analysis for predicted anomalies, and structured message generation, ensuring understandable and accurate anomaly information, and continuous model updates.

Data elements:

Name	Attributes Type	Description
trip_id	Integer	Unique identifier for the trip.
sequence	Integer	The sequence number of the stop within the trip.
line	Integer	The line number or route number for the trip.
first_id	Integer	The identifier for the first stop of the trip.
dep_id	Integer	The identifier for the departure stop in the sequence.
arr_id	Integer	The identifier for the arrival stop in the sequence.
first_date_time	DateTime	Date and time when the trip starts at the first stop.
dep_date_time	DateTime	Date and time when the vehicle departs from the departure stop.
arr_date_time	DateTime	Date and time when the vehicle arrives at the arrival stop.
occupancy_cat	Integer	Occupancy category for the trip segment. 11 – Low Occupancy 12 – Medium Occupancy 13 – High Occupancy (In case of having data with enough precision, this attribute can be used to indicate the estimated number of passengers)

Data Format:

As an initial solution a transfer of CSV (Comma-Separated Values) files can be implemented as this is simple to read and widely supported.

For a final implementation, a structured data solution can be used, like JSON (JavaScript Object Notation), example provided below.

```
{
    "trip_id": 11053,
    "sequence": 1,
    "line": 442,
    "first_id": 2204066,
    "dep_id": 2204066,
    "arr_id": 2204068,
    "first_date_time": "2024-02-28T07:59:00",
    "dep_date_time": "2024-02-28T07:59:00",
    "arr_date_time": "2024-02-28T08:01:00",
    "occupancy_cat": 12
}
```

European standards should be used if they meet the requirement to transport the necessary information.

8.1.1. Standards

This information related to transport demand forecast can be transported using the Transmodel framework[3] and NeTEx XML format[4].

Transmodel relevant Entities and Relationships:

- **VehicleJourney:** Represents a specific trip made by a public transport vehicle.
- **JourneyPattern:** Defines the sequence of stops for a VehicleJourney.
- **StopPoint:** Represents a specific location where a vehicle stops to pick up or drop off passengers.
- **ScheduledStopPoint:** An instance of a StopPoint in a specific JourneyPattern.
- **PassengerDemand:** Represents the forecasted demand for a specific VehicleJourney or StopPoint.

Transmodel Mapping:

- **VehicleJourney**: Mapped to `trip_id`.
- **JourneyPattern**: Derived from `line`, `first_id`, `dep_id`, and `arr_id`.
- **ScheduledStopPoint**: Mapped to `sequence`, `dep_id`, and `arr_id`.
- **VehicleJourneyScheduledTime**: Mapped to `first_date_time`, `dep_date_time`, and `arr_date_time`.
- **PassengerDemand**: Mapped to `occupancy_cat`.

Here's an example snippet in NeTEx XML format for a single trip:

```
<VehicleJourney>
    <id>11053</id>
    <JourneyPatternRef ref="JP_442_2204066_2204068"/>
    <OperatingDay>2024-02-28</OperatingDay>
    <VehicleJourneyStopPoint>
        <ScheduledStopPointRef ref="SP_2204066"/>
        <ArrivalTime>07:59:00</ArrivalTime>
        <DepartureTime>07:59:00</DepartureTime>
        <Occupancy>
            <OccupancyCategoryRef ref="OC_12"/>
        </Occupancy>
    </VehicleJourneyStopPoint>
    <VehicleJourneyStopPoint>
        <ScheduledStopPointRef ref="SP_2204068"/>
        <ArrivalTime>08:01:00</ArrivalTime>
    </VehicleJourneyStopPoint>
</VehicleJourney>
```

[3] Transmodel – CEN Reference Data Model for Public Transport (transmodel-cen.eu)

[4] NeTEx (netex-cen.eu)

Explanation of the NeTEx XML Elements:

- **VehicleJourney**: Represents the specific trip made by the vehicle.
  - `<id>`: Unique identifier for the trip, corresponding to `trip_id`.
  - `<JourneyPatternRef>`: Reference to the journey pattern, which is derived from the `line`, `first_id`, `dep_id`, and `arr_id`.
  - `<OperatingDay>`: The date of the operation, extracted from `first_date_time`.
  - **VehicleJourneyStopPoint**: Represents each stop in the sequence of the trip.
    - `<ScheduledStopPointRef>`: Reference to the specific stop point, corresponding to `dep_id` and `arr_id`.
    - `<ArrivalTime>`: The arrival time at the stop.
    - `<DepartureTime>`: The departure time from the stop.
    - **Occupancy**: Represents the occupancy information.
      - `<OccupancyCategoryRef>`: Reference to the occupancy category, corresponding to `occupancy_cat`.

Here is a more comprehensive example on how to build a complete NeTEx message including multiple trip samples.
Complete NeTEx XML Example for Multiple Trips:

```
<PublicationDelivery>
    <dataObjects>
        <CompositeFrame>
            <frames>
                <ServiceFrame>
                    <vehicleJourneys>
                        <VehicleJourney>
                            Trip 1
                        </VehicleJourney>
                        <VehicleJourney>
                            Trip 2
                        </VehicleJourney>
                    </vehicleJourneys>
                </ServiceFrame>
            </frames>
        </CompositeFrame>
    </dataObjects>
</PublicationDelivery>
```

Explanation of the NeTEx XML Elements:

- **PublicationDelivery**: Root element for the NeTEx document.
  - **dataObjects**: Container for data objects.
    - **CompositeFrame**: Composite frame that groups together various frames.
      - **frames**: Container for different frames.
        - **ServiceFrame**: Frame for service-related data.
          - **vehicleJourneys**: Container for vehicle journeys.

8.2. PIS – TMS interface (T6.2)

The goal of this interface is to facilitate seamless communication between the PIS and TMS in the railway domain. By exchanging transport demand forecast information, we aim to enhance operational efficiency, optimize resource allocation, and improve passenger experience. To achieve this, it will be used information generated by the forecast demand algorithm described before in this document specifically focusing on the following information:

- **Travellers Demand from A to B**: The PIS can provide real-time information about passenger demand between specific origin (A) and destination (B) stations. It helps TMS operators to anticipate capacity requirements and allocate trains accordingly.
- **Travellers Demand Between Subsequent Stops**: The interface enables the PIS to share demand data for intermediate stops along a train route. The TMS might use this information to adjust schedules, manage platform capacities, and optimize train services.
- **Travellers Embarking/Disembarking Flow**: By tracking passenger flow at train stops, the interface assists TMS in predicting station congestion, ensuring timely boarding, and forecasting accurately train dwell times to be expected.

In summary, this interface bridges the gap between passenger demand insights and operational decision-making, ultimately enhancing rail service reliability and efficiency.

This interface will be developed in conjunction with Task 6.2 and will be documented in Deliverable 6.3.

9. Algorithms descriptions

This section will describe the approaches taken for the development of the algorithms that will enable the previously described use cases. Section 9.1 will focus on the first three use cases from Task 6.4 and section 9.2 will focus on the fourth use case.

9.1. Data analytics platform and occupancy model

A traveller who informs himself about the travel possibilities in a mobility app usually has a concrete travel need in the immediate and near future. The request to the app and the recommended routes are good indicators for the actual travel of this traveller in the future.

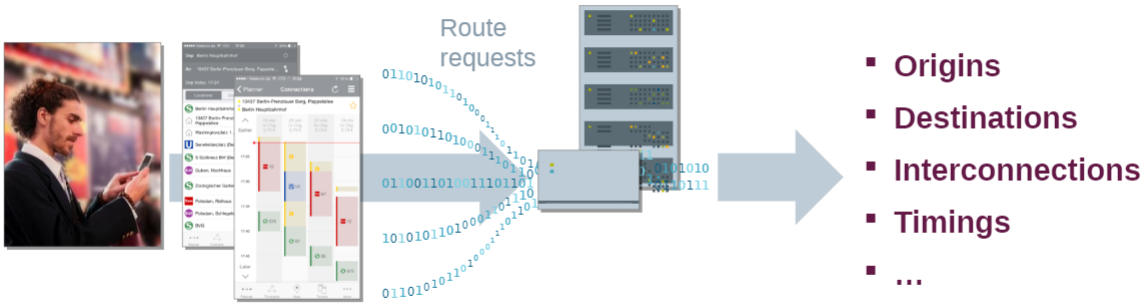

Figure 9 - Data from journey planning requests

Not every request in a mobility app - for example for a trip the next day - necessarily leads to an actual trip according to the route recommendations. ***For our purposes, however, it is not important to predict the travel behaviour of individual travellers, but only the travel behaviour in total, which we want to predict with a reasonable reliability***.
Examples of forecasts that we strive for and that are possible:

- Line Z will be about half-full between stops X and Y tomorrow at 08:30.
- The day after tomorrow, at 20:15, there will be twice as many people waiting at stop Z as usual.
- The average vehicle occupancy rate for a trip on the Z line, departing from the X stop, tomorrow at 11:05, will be less than 40%.
- How high is the occupancy rate of individual line segments with a destination stop 'hospital' tomorrow at 04:30?

These are examples of forecasts that we do not aim for and that cannot be reliably made:

- Mrs. Schmidt will leave tomorrow 13:14 to her destination.
- Mr. Xanten will leave for his destination in the morning the day after tomorrow.

For the forecasts we are aiming for, it is important that we know the travel requests of many travellers. A journey planning system for a medium sized customer has several million route requests per month and for these the journey planning backend generates many more route recommendations. This is a very high volume of data that is more than sufficient for the purpose

of capacity utilization forecasts.
As described above: A request in a travel information app indicates a travel intention for the near or medium future. To derive forecasts from the number of these individual requests, two things are needed, as described in the following sections:

a) Big data technologies, machine learning
b) Calibration using counting data

9.1.1. Big data technologies and machine learning

Big Data Technologies is a collective term for several technologies for working with large amounts of data. Against the background of large data volumes, it is all about performance, pattern recognition, focused evaluation and archiving. Machine learning is necessary to automatically generate insights and forecasts from the very large amounts of data. This involves technologies such as neural networks, random forest, Bayesian networks and extreme gradient boosting.
Examples of questions or insights that we can address with Machine Learning:

- Which segments show similar patterns in terms of enquiry numbers and capacity utilization?
- What is the relationship between specific points of interest and patterns that we see in the inquiry numbers?
- How do capacity utilization and request patterns change in the event of delays?

What in this form still sounds quite theoretical helps to answer relevant questions of our customers, e.g:

- What is the load factor on the Z line on working days vs. public holidays?
- At which stop do more than 100 passengers wait at the same time on weekdays?
- How many buses will the rail replacement service on the Z line need in the next 2 weeks?
- How does the capacity requirement for rail replacement services change when passengers start to switch to alternative routes?
- How many additional travellers will there be when/where due to a concert?

9.1.1.1. Process Model

Demand forecast, which encompasses a wide range of applications such as traveller information systems, adheres to the industry-standard process model known as CRISP-DM (Cross-Industry Standard Process for Data Mining), as illustrated in Figure 10.

The CRISP-DM methodology comprises six phases: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment. While Figure 10 depicts an idealized sequence of these phases, it is essential to note that the CRISP-DM process is inherently iterative. In practical implementations, phases may be revisited multiple times or executed in varying sequences. The figure does not attempt to represent all potential iterations. Instead, it visualizes the iterative nature of the process with a large encompassing circle, while the arrows between the phases indicate the most common iterative transitions.

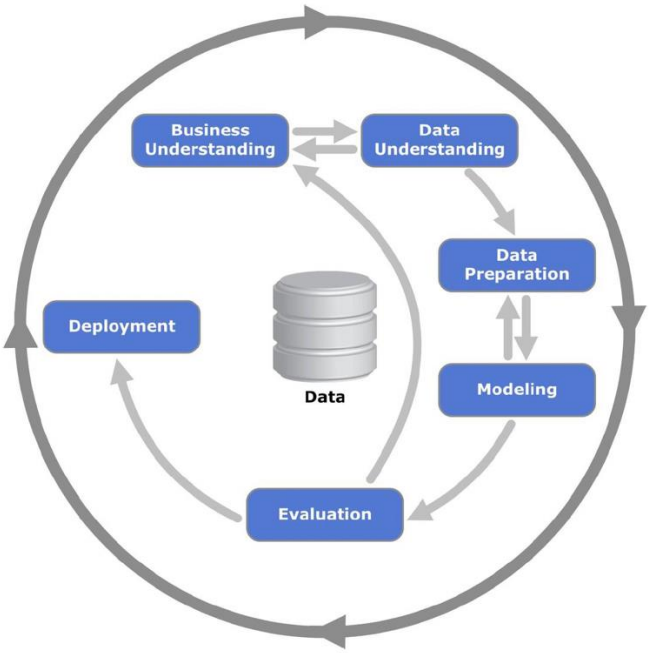

Figure 10: CRISP-DM Cycle[5]

9.1.1.2. Addressing Data Imbalance in Occupancy Prognosis with SMOTE

In the present use case, the dataset is anticipated to exhibit significant class imbalance, characterized by numerous segments with low occupancy. This imbalance arises due to coverage of remote regions and off-peak or nighttime services, presenting additional challenges for machine learning models.

To mitigate these challenges, we employed the Synthetic Minority Oversampling Technique (SMOTE) on the training dataset. SMOTE effectively generates synthetic samples of the minority class, thereby balancing the class distribution and enhancing model performance. The process of synthetic sample generation is illustrated in Figure 11.

[5] Source: https://statistik-dresden.de/archives/1128.

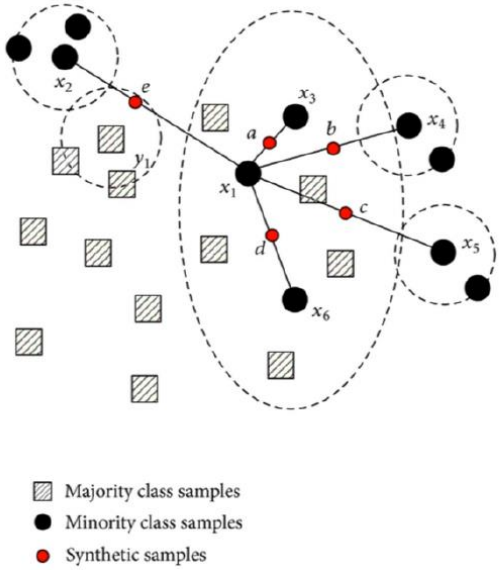

Figure 11: Synthetic Minority Oversampling Technique (SMOTE)

9.1.1.3. Modelling Approach

We continuously evaluate the performance of various machine learning models and adapt our algorithms based on their outcomes. Notably, we have achieved robust results using Supervised Machine Learning, particularly with Gradient Boosted Trees implemented via the XGBoost framework, for medium- to long-term forecasts, typically considering a time window of 10 to 14 days.

A primary limitation of simple decision trees is their lack of predictive power, despite their simplicity. To address this, ensemble models enhance a base learner (a shallow decision tree) by iteratively combining multiple base learners (weak learners) into a strong learner. This combination is achieved by generating different base learners for subsets of the dataset either in parallel or sequentially. When performed sequentially, this process is known as boosting, where individual tree functions are greedily added to a loss function l to minimize the regularization objective L.

In gradient-boosted trees, the steepest descent direction for minimization is estimated using a gradient descent function. This function aims to predict the optimal gradient for the additive model to achieve a local minimum of the loss function. The model is initially constructed with a set of weights, and the loss function is minimized by iteratively updating these weights. Specifically, XGBoost employs K additive functions to predict the target variable.

Given the imbalanced nature of the dataset, the results are heavily biased towards the low occupancy class, as most samples belong to this class. Consequently, the model's objective to minimize overall error is best achieved without distinguishing between the three classes. However, through hyperparameter tuning, we increased the model's complexity, risking overfitting but enabling the model to detect more intricate relationships. Additionally, we fine-tuned the parameters to improve classification performance for the high occupancy class, compelling the model to allocate more effort towards distinguishing this class.

9.1.1.4. Short-Term Forecasts

In case real-time passenger counting systems are deployed in any of the vehicles covered by our forecast system, we can use this real-time information to provide short-term forecast updates. Due to the usually very high number of segments which are supplied with a prognosis, the short-term forecasts are done very selectively as follows: If, after vehicle departure, a significant mismatch is detected between predicted and counted occupancy, an updated forecast is sent to the MaaS platform for the remaining vehicle run. Other types of short-term forecast updates consider real-time delay information (if available). We are currently working on an approach using Graph Neural Networks to respect the effect of delays on occupancy and update the prognoses accordingly if such delays are detected.

9.1.2. Calibration using counting data

As already mentioned above: the requests in the information app show an intention to travel, but it is not certain whether the trip will actually take place according to the route recommendation. In order to deduce actual passenger numbers from the query data, a calibration is usually required. For this purpose, counting data can be used, i.e. data that reliably represents the real number of passengers - e.g. on line Z on Mondays at 9 a.m. or at stop xy on Fridays at 12 a.m. This counting data is not required across the board, partial coverage is sufficient, e.g. only for some lines or for selected vehicles.
With the help of this counting data, the algorithms for the evaluations and forecasts are calibrated and as a result the quality, i.e. the accuracy of the results is improved.

9.2. Anomaly detection and prediction

Trains often experience fluctuating occupancy levels due to various factor including weather conditions, public events, and typical daily commuter patterns. Identifying and predicting abnormal spikes in train occupancy is what we intend to do using a two-fold procedure that falls into two learning categories: unsupervised and supervised learning.

9.2.1. Anomaly detection approach

This phase focuses on identifying unusual spikes in train occupancy levels. Depending on occupancy data precision and granularity, this data can take two shapes: categorical (e.g., "low occupancy," "medium occupancy," "high occupancy") or numerical (exact passenger counts or percentages). In this case, unlike numerical data, categorical data must undergo encoding techniques such as Label Encoding to ensure compatibility with machine learning models.

To detect anomalies, as an initial approach, Isolation Forest model will be employed. Isolation Forest is an ensemble learning method particularly suited for anomaly detection as it isolates observations by randomly selecting features and split values, which leads to the identification of outliers. This model can handle both categorical and numerical data after appropriate preprocessing. Moreover, by adding contextual features such as the day of the week and whether it is a weekend, it can better understand normal occupancy patterns, making it more effective at

spotting significant deviations.

In case numerical data is available, for comparative purposes, rolling statistics are also expected to be used. Rolling statistics use the rolling mean and standard deviation to identify anomalies. By calculating the rolling mean and standard deviation over a specified window, occupancy values that deviate from the mean by a defined threshold, such as three standard deviations, can be detected.

Finally, in addition to Isolation Forest and rolling statistics, other models like Hidden Markov Models (HMM) can improve the anomaly detection capabilities when time-dependency is relevant, which will be taken into account as part of the data analysis. HMM capture temporal dependencies and identify sequences of occupancy states that deviate from expected patterns, making them suitable for detecting anomalies in sequential occupancy data.

9.2.2. Anomaly prediction approach

After the anomaly detection phase, the objective of the following approach is to forecast future anomalies in train occupancy based on influencing factors such as weather conditions and public events.

To setup the anomaly prediction model, initially a dataset will be prepared, that includes contextual features covering weather and events data, and a labelled column with the identified anomalies. By leveraging supervised learning techniques, as a first approach, it is expected to evaluate two models: Logistic Regression and Random Forest.

Logistic Regression serves as a good baseline model. It is a simple statistical model, interpretable and suitable for binary classification tasks like anomaly detection. In this context of predicting anomalies in train occupancy, Logistic Regression will allow us to quantify the relationship between the probability of an anomaly and the influencing factors such as weather conditions, public events, and temporal features.

In other hand, Random Forest is an ensemble learning method that combines multiple decision trees to improve predictive accuracy and robustness. It is expected to be well-suited for this task as it can handle complex interactions between features. Additionally, Random Forest also provides feature importance scores, which can help us understand the relative impact of different factors on the occurrence of anomalies.

In case additional data and corresponding features are incorporated into the dataset, more complex models can be considered in order to capture the different relationships within the data. Gradient Boosting Machines (e.g., XGBoost) are examples of models that can handle complex interactions between multiple features, which could improve the predictive accuracy. Additionally, in case time patterns turns out to be relevant in this context, Long Short-Term Memory (LSTM) networks could be used to capture long-term dependencies and patterns over time, which could be useful to understand how sequence of events and conditions lead to anomalies.

10. Conclusions

The task 6.4 from the FP6 FutuRe project has made significant strides in the development and refinement of demand analysis algorithms. The objective and aim of this work, as presented throughout the document, have been to improve the accuracy and effectiveness of forecast demand in rail regional lines.

Our work contributes to new knowledge by updating an occupancy model based on new and historical data from journey planning requests. This approach considers observed patterns and trends, providing accurate predictions that are crucial for future demand forecasting and analysis.

The main points and results of our work have significant implications. The updated occupancy model, when effectively implemented, can inform planning tools and decision-making processes, addressing areas with high demand and low offering.

However, our work is not without its challenges. The availability and quality of the training data can significantly impact the accuracy and effectiveness of the trained machine learning model. Furthermore, the system should provide mechanisms for monitoring and evaluating performance, considering any uncertainties or limitations in the data.

In conclusion, our work underscores the importance of collaborative approaches between transport service providers in addressing demand analysis, to provide the best services adapted to the needs of their customers. The work specified under this deliverable will be continued in D11.4 where will be reported the details of the implemented solution for demand forecast and how it will be demonstrated.

End-of-document