

東京大学大学院新領域創成科学研究科
社会文化環境学専攻

2022 年度
修 士 論 文

大きな GPS データを使用して鉄道交通を分析および視覚化する
インタラクティブシステム
An Interactive System to Analyze and Visualize Railway Traffic using
Big GPS Data

2022 年 7 月 15 日提出
指導教員 柴崎 亮介 教授

JEPH, Puneet
ジェフ プニート

THE UNIVERSITY OF TOKYO
Graduate School of Frontier Sciences
Department of Socio-Cultural Environmental Studies

**An Interactive System to Analyze and Visualize Railway Traffic using Big GPS
Data**

by JEPH Puneet

Abstract

The scheduled big events or unforeseen incidents like accidents cause substantial deviations in the city traffic, especially railway transportation which happens to be the dominant mode of transportation for most of the major cities. Analysis of such events and incidents' impact on railway transportation, therefore, is of great importance to urban planning and transport management, yet is quite challenging because of the lack of readily available data about railway passengers' citywide flow and event participants' choice of transportation mode. Previous works have mainly relied on precise but limited data like sensors, AFC (Automated Fare Collection), or smart-card get-on get-off data to estimate the railway passengers, and did not take a holistic multi-dimensional approach to analyze changes in railway traffic. To tackle these challenges, we propose a novel interactive Dashboard system that utilizes millions of smartphone GPS records across Japan. The dashboard can estimate and visualize railway traffic for the stations and links during the events or accidents and can also provide other relevant event participants' or railway commuter information. We also introduce a Congestion Index to measure the increase in congestion of stations during events. The dashboard can also assess the accidents by getting affected railway commuters' information and their commute time. This dashboard can be highly useful for event organizers, railway administrators, and city planners to assess and compare the impact of events and accidents on railway traffic.

Keywords: Railway Transportation, Visualization System, Big smartphone GPS data, Crowd estimation, Event analysis, Accident analysis

Acknowledgements

Foremost, I would like to express my sincere and deepest gratitude to my supervisor **Prof. Ryousuke SHIBASAKI**, who has continually supported and guided my research. He always had regular meetings with me and always encouraged me to strive for excellence with his valuable advises and constructive critique. He always listened patiently to my research presentations and gave valuable comments to improve my research. I would also like to give special thanks to my co-supervisor **Prof. Hill Hiroki KOBAYASHI**, who gave me valuable advice and guidance in my research and encouraged me to think from multi-dimensional perspectives.

I am very grateful to the researcher in Shibasaki Lab: **Dr. Renhe Jiang**, who was also my mentor. He helped me a lot during my initial learning as research student. He also helped me to get used to the daily life in Japan. He was always there to guide me and help me in my research, or personal problems whenever I approached him.

Special thanks to **Dr. Tianqi Xia**, who was my mentor and a friend. I enjoyed a lot doing research with him. He made me very comfortable and was never strict with me, but at the same time motivated me a lot to improve my research. He always answered all of my queries with lot of patience. I am very grateful to him.

In addition, I would also like to say thanks to my peer students: Chuyao Feng, Chuang Yang, Zekun Cai, Yanxiu Jin, who helped me get along and made me comfortable and feel like home in this two year journey.

Last but not the least, I would like to thank my loving wife Sukirti who always supported me, inspired me, and always cared for me. I would also like to pay my gratitude towards my parents and parents-in-law for inspiring me to pursue my dreams.

Table of Contents

Abstract	i
Acknowledgements	iii
Table of Contents	v
List of Figures	ix
Abbreviations	xi
1 Introduction	1
1.1 Research Background and Motivation	1
1.2 Challenges and Solutions	2
1.3 Contribution	4
2 Related Works	7
2.1 Traditional Data Sources based Research	7

2.2	Big GPS Mobility Data based Research	8
3	System Development	11
3.1	Data Sources	13
3.1.1	GPS Trajectory Data	13
3.1.2	Event and Accident Data	13
3.1.3	Railway Network Data	14
3.2	GPS Trajectory Data	15
3.2.1	Data Description	15
3.2.2	Data Rectification	15
3.3	Railway Traffic Estimation	16
3.3.1	Station-wise Railway Passenger Aggregation	16
3.3.2	Link-wise Railway Passenger Aggregation	17
3.4	Event Analysis Methodology	18
3.4.1	Local OD Density Estimation	18
3.4.2	Event Participants Features Extraction	18
3.4.3	Event Congestion Index	19
3.5	Accident Analysis Methodology	20
3.5.1	Effect on Railway Commuters	21

3.5.1.1	Identification of Railway Commuters	21
3.5.1.2	Commute Time Increase Factor	23
3.5.1.3	Affected Commuters Features	24
3.5.2	Line Load Drop Factor	25
4	Event Analysis: Dashboard Design and Case Studies	27
4.1	Layout introduction	27
4.2	Modules and User-Interaction Introduction	29
4.2.1	Railway Passenger Analysis Module	29
4.2.2	Event Comparison Module	32
4.2.3	Event-Participants Statistics Module	33
5	Accident Analysis: Dashboard Design and Case Studies	37
5.1	Layout introduction	37
5.2	Modules and User-Interaction Introduction	38
5.2.1	Station Analysis Module	38
5.2.2	Link Analysis Module	41
5.2.3	Commuter Analysis Module	43
5.2.4	Accident Overview Module	46
6	Conclusion	49

Bibliography

51

List of Figures

1.1	The System Framework of the Dashboard.	4
3.1	System Architecture	12
3.2	Boxplots of the entropy features for the four clusters of railway passengers	22
4.1	The Layout-design of the Dashboard.	28
4.2	Visualization of Get-on and Get-off numbers for the event of 27 th May, 2018 at Tokyo Race Course venue.	30
4.3	Visualization of density maps in the map panel	31
4.4	Visualizations in event-comparison module.	33
4.5	Event-statistics results.	35
5.1	The Layout-design of the Dashboard for Accident Analysis.	38
5.2	Visualization of Get-on and Get-off passenger numbers for the accident of 31 st January, 2020 between Oomori-Kamata stations on Keihin-Touhoku Line.	40
5.3	The Map panel in the Station-Analysis module.	40

5.4	Visualization of passenger load volume numbers for the accident of 31 st January, 2020 between Oomori-Kamata stations on Keihin-Touhoku Line.	42
5.5	The Map panel in the Link-Analysis module.	42
5.6	Commuter analysis and other accident statistics results.	44
5.7	Visualization of line/company share of affected commuters for the accident of 31 st January, 2020 between Oomori-Kamata stations on Keihin-Touhoku Line.	45
5.8	Share of home and work/destination cities for the affected railway commuters for the accident of 31 st January, 2020 between Oomori-Kamata stations on Keihin Line.	46
5.9	Accident-Overview Module for the accident of 31 st January, 2020 between Oomori-Kamata stations on Keihin Line.	47

Abbreviations

ICT	Information and Communication Technology
AFC	Automated Fare Collection
GPS	Global Positioning System
OSM	OpenStreetMap
MLIT	Ministry of Land, Infrastructure, Transport and Tourism
LBS	Location Based Services
OD	Origin Destination
NLNI	National Land Numerical Information

Chapter 1

Introduction

1.1 Research Background and Motivation

Rail transportation system plays an important role in urban development as it is energy-efficient, relatively time-saving, and a punctual mode of transportation for transporting passengers and freights. According to the survey of person trips conducted by the Japanese government [1], Tokyo has one of the highest rail transit usage ratio of up to 48%, showing that almost half of the Tokyo residents prefer to travel by train or metro services. Therefore, understanding railway traffic is of utmost importance for building an Intelligent Transportation System (ITS) in the development of a Smart City.

An important area of study in railway traffic analysis is the behavior during big events and other incidents, where railway traffic deviates from its regular pattern causing unforeseen and unusual congestion in some of the lines and stations. These can be of two types mainly – (1) Scheduled Events (2) Natural Disasters or Human Accidents.

In the case of scheduled big events, a big crowd converges in the venue of the event just before the start of the event and then diverges out of the venue after the end of the event. This causes unusual congestion in railway stations and lines, especially those nearby to the venue. This event-caused congestion is different from the usual congestion observed during peak morning or evening rush hours in terms of the features and exhibits different spatio-temporal characteristics [2]. Similarly, during natural disasters or accidents, the railway services get delayed or suspended for some duration, affecting a lot of railway passengers and their commute time. Therefore, it becomes important to study and analyze the railway traffic behavior during such kinds of events/incidents to understand their effect on railway passengers in particular and railway traffic in general. Moreover, for the scheduled events, it is also important for the event organizers and city planners to understand the features of the audience attending the event, for example, where did the audience come from, and which transportation mode did they use to arrive at the venue? Having this information can help them to better plan and manage future events. Similarly, learning about the impact of the suspension of services in some railway lines on the other nearby lines, and how the commute time gets affected for the regular train commuters, are also important for railway administrators for contingency planning.

1.2 Challenges and Solutions

Previous studies have mainly utilized sensors and AFC (Automated Fare Collection) based IC Card records for railway passenger analysis [3–5]. Although railway passenger numbers can be more precisely demonstrated by such IC Card or Smart Card data which have exact tap-in and tap-out timestamps, it can neither differentiate between event participants and other passengers nor analyze people's flow outside the origin and destination stations. In addition, the limitations with AFC and Smart

Card data are that they are difficult to obtain for the city-wide rail metro system over a longer time duration, and passenger's estimated route is mostly based on shortest path algorithm.

With the recent advent of GPS-enabled smartphones and the popularization of LBS (location-based services), huge amounts of GPS data has become available for human trajectory and traffic analysis. The human mobility data can be used as an effective data source to evaluate the effect and impact of the events. The traditional data like surveys, smart card data, etc. do not take into account the passengers' mobility pattern. The big smartphone GPS data can overcome these limitations of the traditional data. It can also help assess the mobility flow of event participants and their transportation modes. Though the mobility data is not as accurate as AFC and Smart Card Data and is not completely representative, it still can give a multi-faceted comprehensive analysis and has the potential to be a very useful tool. Over the past decade, a lot of research have focused on the analysis, prediction, and simulation of road traffic using GPS data [6, 7] but few have focused on railway traffic analysis and visualization using GPS data.

With the issues aforementioned, this study proposes an interactive dashboard system to effectively analyze and visualize the railway traffic along with other railway passengers' features during the events/accidents using big human GPS data. Figure 1.1 shows the system framework of this dashboard system. In this system, the inputs are the raw GPS data, rail-road network data and event/accident information data. In the data preprocessing step, the GPS data are processed to extract the stay points and user trajectories. These user trajectories are further labeled with their transportation mode (WALK/BIKE/CAR/TRAIN/STAY) and then map-matched. In the next step, we aggregate the railway passengers (i) station-wise, and (ii) link-wise, where the former provides entry/exit info for a station, while latter provides passenger volume for a given railway link. For the event analysis, we first aggre-

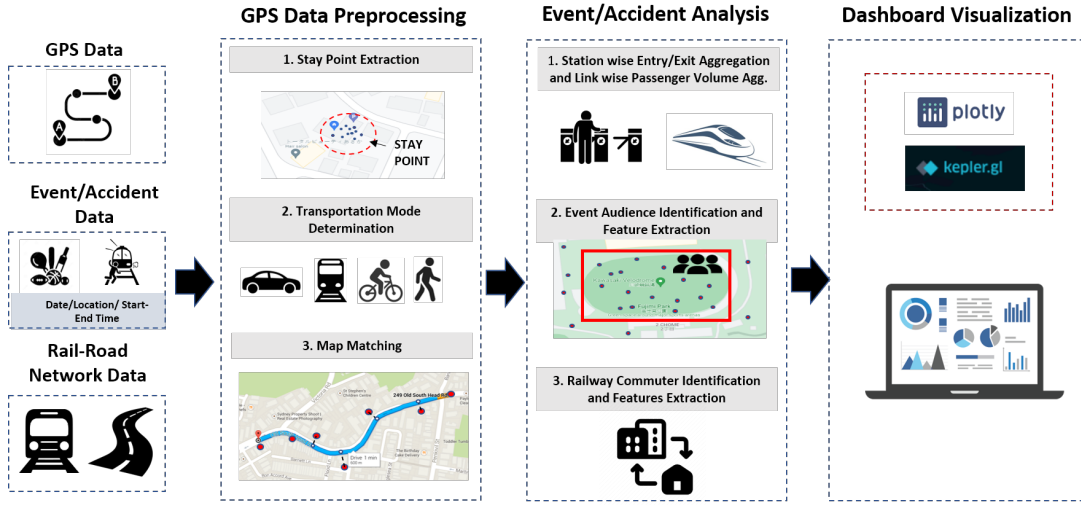


Figure 1.1: The System Framework of the Dashboard.

gate the identified railway passengers for selected stations nearby the venue (within a 1.5 km. radius buffer) and also compute the local OD (origin-destination) density of passengers for these stations, i.e. destination density for exiting passengers, and origin density for entering passengers. Then event participants are identified using appropriate spatial grids and time windows, along with their other characteristics like arrival/departure time, origin prefecture, transportation mode used to reach the event venue, etc. While for the accident analysis, the railway commuters are identified by those having low entropy in their home/work station and arrival/departure hours. In the final step, we integrate all the processed results and visualize them using figures, maps, and other statistics in an interactive dashboard browser application developed by a combination of Plotly Dash and Kepler.gl [8]. We also compute a score called Congestion Index and other indicators to help in the assessment and comparison of the event's impact on railway traffic and passengers.

1.3 Contribution

The main contributions of this research are listed as follows:

- We develop a novel dashboard by integrating both Plotly Dash and Kepler.gl platforms for comprehensive railway traffic analysis during events and accidents using big GPS data generated by millions of smartphone users from entire Japan.
- The dashboard can help visualize the change in congestion of the railway stations nearby the venue on the day of the event, as well as give information about event participants mobility flow like transportation mode used, time of arrival and departure, etc., which is generally not possible using traditional methods and data.
- The dashboard can help visualizing the change in railway traffic during an accident for nearby lines, and also give information about affected railway commuters' increased commute time.
- We also present case studies of several big events and accidents in Tokyo using our dashboard and demonstrate how those events affected the railway traffic.

The remaining part of this research is organized as follows: Chapter 2 introduces the past studies related to railway traffic analysis and the datasets they used. Chapter 3 explains system design, the data preprocessing, and the methodologies used for identifying railway passengers, event participants and railway commuters. Chapter 4 introduces the dashboard design and layout for event analysis and its different modules, while Chapter 5 similarly introduces the dashboard design for accident analysis. We also present case studies of some real-world events and accidents within these chapters to further demonstrate the functionalities of various modules of the dashboard, followed by conclusion and future works in chapter 6.

Chapter 2

Related Works

2.1 Traditional Data Sources based Research

The data used in past researches on Railway Transportation can be mainly categorized into two classes based on their origin. The first type is the information obtained by means of various sensors about railway trains like position, velocity, acceleration, weight, etc. This information can be used to detect the exact status of trains in real-time and can also help in their punctuality analysis. Massimo et al. [3] used sensor information for train identification, speed and acceleration detection, and dynamic load calculation.

The second type is the data primarily about railway passengers, like tap-in, tap-out time collected from AFC (Automated Fare Collection) or Smart Cards. Sun et al. Satoshi et al. [9] used OD-Matrix of Smart Card Ticket 'PASMO' origin-destination Data of Tokyo Metropolitan Area to examine and visualize railway passenger flow and demands during public events. Baichuan et al. [10] also used Smart Card Data collected from AFCs to distribute passengers over the rail network using the

Bayesian simulation-based optimization method. Tomoeda et al. [11] used daily ridership data to develop real-time train traffic simulator capable of detecting change in passenger flow pattern, but did not analyse the past accidents. Nonetheless, the limitation with AFC and Smart Card Data is that they can be difficult to obtain for entire city-wide rail metro systems consisting of various rail companies and lines, which is especially true for the Tokyo Metropolitan Area. Moreover, passengers' route choice have been mostly assumed to be based on the shortest path assumption and does not take other parameters like fare into account.

Some studies have tried to use other novel data sources to analyze railway traffic. Yuki et al. [12] used Bluetooth RSSI observed by passengers' mobile phones to estimate car-level train congestion using a Bayesian-based likelihood estimator. Kohei et al. [13] used the moving object detection method from web camera images to estimate rail traffic congestion, but passenger information was excluded. Masahiko et al. [14] used social media data from Twitter and Smart Card data in combination to analyze changes in railway passengers' flow during unusual phenomena in Tokyo Metro.

2.2 Big GPS Mobility Data based Research

With the advent of smartphone location services, GPS data has been widely used to analyze road traffic, but few studies have focused on railway traffic analysis using the GPS data. The earlier studies utilizing the GPS trajectory data were focused on extracting meaningful place of people [15], understanding people moving pattern [16], and predicting movement of people [17]. The recent studies have focused on analysis and prediction of human transportation. GPS log data of taxis was utilized [18] to analyze and detect the road traffic congestion. Xia et al. [19] used GPS

trajectory information to predict the kernel density in railway network using deep-learning based methodology LSTM (Long Short Term Memory). D'Andrea et al. [20] designed a system that detects real-time traffic congestion and incidents on roads based on analyzing GPS trajectories of moving vehicles.

Some studies have tried to develop dashboards for traffic visualization. Pathak et al. [21] utilized social network data like Twitter, to design a dashboard for obtaining a real-time view of the traffic data. Mingyu et al. [22] used Taxi GPS trajectory data to create a dashboard to find the cause of traffic congestion. Lwin et al. [23] used big data including GPS trajectory data to develop City Geospatial Dashboard as solution provider for disaster management. Li et al. [24] developed a dashboard to simulate railway passenger flow during train delays using origin-destination data.

Chapter 3

System Development

As introduced in the previous chapter, this study aims to develop a dashboard system for analyzing the influence of big events and other incidents on railway traffic via GPS data. The whole system is constructed based on the architecture illustrated in Figure 3.1. In this system, data collected from multiple data sources are preprocessed and stored in the PostgreSQL database on the server-side for ease of access and saving the memory. Then the statistic results based on selected events or accidents are further processed on the server-side and visualized on the user end in the browser application developed using Plotly Dash. The dashboard consists of several modules with different functionality, which can be visualized and controlled via different panels in the dashboard with either the response from server side, or the callback functions directly in the front-end. In the remainder of this chapter, we will introduce the panels and modules of the dashboard in detail. Users can interact with each module as per their choice and needs with customized settings.

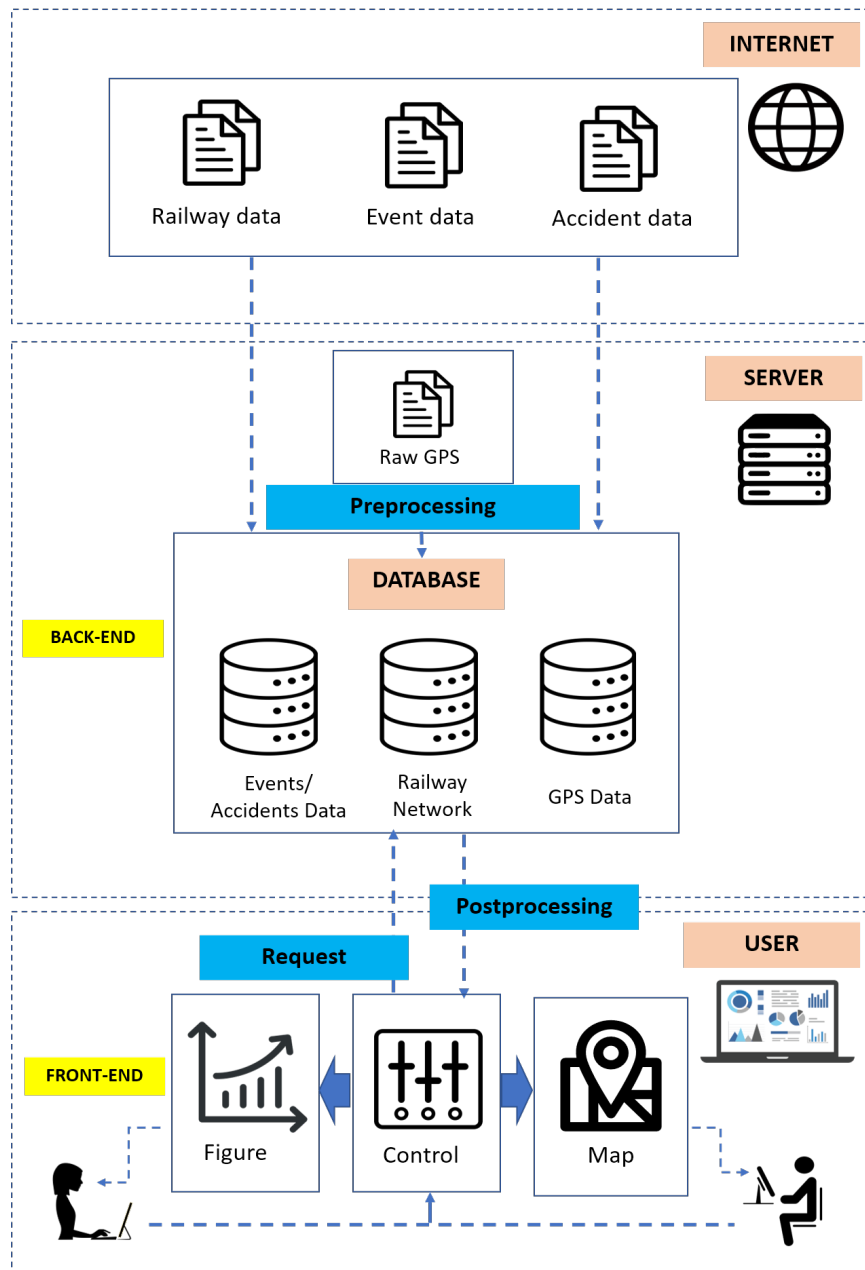


Figure 3.1: System Architecture

3.1 Data Sources

This system collects three types of data from different data sources, including big human mobility GPS data, event and accident data, and railway network data.

3.1.1 GPS Trajectory Data

For GPS trajectory data, we collaborated with Blogwatcher Inc. to get big human GPS trajectory data that covers almost 5 million people covering 47 prefectures of Japan. The location data are collected through smartphone applications that have a built-in location-sharing module provided by Blogwatcher Inc. with the user's consent. Any personal information, which could identify individuals, were not collected. Data attributes are anonymized ID, timestamp, longitude, and latitude. The time range of data is from 2018-01 to 2022-06 and entire Japan was focused upon.

3.1.2 Event and Accident Data

Event data is collected by the following steps: first, we choose several popular event venues and collect their location information from OpenStreetMap. Then for each venue, we collect event information from official websites and several ticket selling platforms, which includes the start time and the number of audiences. For each event venue, we extract the nearby stations by using buffer of 1.5 km radius, for each rail line, we choose the closest station as our research target.

For this study, three popular types of events in Japan are chosen – Horse Race Events, AAA Tokyo Dome Concerts, and Soccer Games. All of these events are quite popular among Japanese residents and regularly attract tens of thousands of audiences

in attendance. We collected the data from 81 horse race events organized at five different race-course venues and 9 Tokyo Dome events, spanning from 2018 to 2021. Similarly, 741 soccer games from 20 unique venues all across Japan were collected spanning from 2019-02 to 2022-05. These event data include various attributes about events like event name, day of the event, start time, end time, participants numbers, name and coordinates of the venue, etc.

Accident data is obtained in collaboration with JR-East (East Japan Railway Company) database. The data attributes are accident date, start time, end time, railway line code, railway link code, people affected. For each location of the accident, we extract all the nearby stations and links by using a buffer of 15 km square centered at the location of the accident as the area of study.

3.1.3 Railway Network Data

Railway network data is collected from the National Land Numerical Information (NLNI) of Japan and have been adapted by simplifying the network and by adding topology information of transfer stations. This data has a graph-like structure where nodes represent the stations, while edges represent the railway links that connect two stations. The total number of links in the network is 13472 with 3281 transfer lines, and the total number of railway stations is 10791.

3.2 GPS Trajectory Data

3.2.1 Data Description

GPS data generated by human mobility flow represent the spatio-temporal position of people. A GPS points trajectory can be denoted as $P = (p_1, p_2, \dots, p_n)$, where $p = (id, timestamp, longitude, latitude)$ and $n = \text{total no. of points for that user}$. To identify the railway passengers and their trajectories, it is important to first process the raw GPS data to segment them by their transportation modes, followed by map-matching. Please note that we use the pre-developed in-house tool for the transportation mode detection and map-matching, which is based on the previous works [25, 26] and is not part of our contribution. The human GPS trajectory data was already separated by stay points and labeled with the travel modes (TRAIN, CAR, BIKE, WALK, STAY), and also map-matched to the target road or railway line. The final data attributes are anonymous ID, trajectory number, positioning point number, travel mode, timestamp, longitude, latitude, target station ID/target road ID.

3.2.2 Data Rectification

When analysed, The map-matched GPS trajectories labeled as railway transportation modes were found to have some map-matching issues. For example, since the railway network data is represented in graph-like structure where stations can be considered as nodes, while links can be considered as edges, so sometimes the start or end of the railway mode trajectories are wrongly mapped to transfer links. The transfer links are those links which do not connect two separate stations physically by track, but act as means to represent the transfer stations where passengers move

from one station to other station to change the railway line. Therefore, sometimes map-matching wrongly maps the entry and exit stations to transfer links on basis of nearest station node. This problem was found to exist in almost 30% of the TRAIN mode GPS trajectories, and same was rectified by removing the transfer links from entry/exit nodes.

In addition, almost 1% of the TRAIN mode trajectories had the issue of having consecutive transfer links. The same was rectified by removing the consecutive transfer links and keeping only first and last nodes of the consecutive transfer links.

3.3 Railway Traffic Estimation

Using the rectified map-matched railway mode trajectories, we can estimate the time and location of get-on get-off for each identified railway passenger, with their entire route including transfer stations along with timestamps. To estimate the railway traffic, we can aggregated the railway passengers in two ways based on the unit of the aggregation. We propose the aggregation based on (i) station, and (ii) link. For this study, we mainly focus on estimating the railway traffic near the venue of the events or accidents to understand the congestion and deviations.

3.3.1 Station-wise Railway Passenger Aggregation

If we focus on station as the unit of aggregation, then we are essentially estimating railway traffic in terms of get-in and get-off numbers for that particular station. This type of aggregation can give us the measure of congestion for a station. Therefore, to estimate the congestion at stations, we prepare two kinds of railway passenger aggregation in a given time period. We count number of railway passengers getting-

on and getting-off separately for each railway station during 1-hour time period. Thus, we obtain two aggregations for each station - passengers entering the station to start the journey, and passengers exiting the station after completing the journey. It is to be noted that we don't include get-on/get-off numbers at transfer stations for aggregation since the passenger is still en-route and the journey is not yet finished. In this study, we take 1-hour time duration as the temporal unit of the aggregation, considering that it is not so short as to have a very low passengers count for small stations, and not so large as to ignore any significant short temporal changes.

3.3.2 Link-wise Railway Passenger Aggregation

A Railway-Link is the the link connecting two railway stations on each end. When a link is focused on as the unit of the aggregation, then it essentially means estimating the railway traffic in terms of number of passenger travelling within that link for a given time-period. This type of aggregation gives us the measure of passenger load volume for that particular link. Therefore, to estimate the passenger load volume in a link, we count number of passengers travelling in both *UP* and *DOWN* directions separately during 30-minute time interval. This aggregations is the estimate of passengers en-route and the traffic volume for a given link in both directions, and thus can be considered as the measure of congestion within the trains for particular links or railway lines. It can be noted that this aggregation conveys different representation of railway traffic compared to station-wise aggregation. While station-wise aggregation focuses on entry/exit numbers for the station, while link-wise aggregation represents the number of people travelling in trains.

3.4 Event Analysis Methodology

In this section, we define the various indicators and features used to analyse the effect of big events on railway traffic. We also explain the methodology used to extract the features and to compute the event congestion index.

3.4.1 Local OD Density Estimation

For big events, crowd management plays a very important role in the smooth organization of the event. This is not only true inside the event venue, but also for the surrounding area of the venue. Therefore, it becomes important for event organizers and city administrators to have an estimation of nearby areas the audience may visit before and after the event for an effective crowd management.

We try to solve this problem using the processed GPS trajectory data. We extract the previous and next trajectory OD information for the passengers entering and exiting the station of interest. The local origin/destination density is defined as the grid-density of the origin/destinations for the passengers entering/exiting the station, respectively. In simpler terms, it is simply the grid-based density of where the passengers are going to just after exiting the station, and where the passengers are coming from before entering the station.

3.4.2 Event Participants Features Extraction

Using GPS trajectories and transportation information, we try to identify the event participants and also other relevant information like origin prefecture, transportation mode used to arrive at the venue, arrival/departure time, time spent at the

venue, etc.

To identify the event participants, we assume that the user will have at least one GPS trajectory labeled as *STAY Mode* in the event venue area. So we extract all the *STAY* type GPS points within the event venue area during the one-hour time period from the start of the event and identify them as event participants.

After extracting the event participants, we find out their arrival time, and departure time by tracing their GPS trajectories. Similarly, we find out the transportation mode they used to arrive at the event venue by tracing their last *MOVE* type trajectory. We also find their origin prefecture by extracting the location of the first *STAY* point on the day of the event.

3.4.3 Event Congestion Index

To analyze the railway traffic congestion caused by the big events, we develop a score index to help in the assessment and comparison of the impact on railway traffic. We name this index as Congestion Index. This index can help in assessing the relative scale of congestion in railway traffic compared to past average congestion. Usually the events cause big increase in exit numbers and entry numbers of the nearby stations before the start of the event, and just after the end of the event, respectively. Therefore we define the congestion index separately for entry and exit numbers of the stations. They are defined as the following:

$$C_{Entry} = W_i \sum_{i=1}^N \left(\frac{I_{peak}^o}{\frac{1}{5} \sum_{m=1}^5 I_{peak}^m} \right) \quad (3.1)$$

$$C_{Exit} = W_i \sum_{i=1}^N \left(\frac{O_{peak}^o}{\frac{1}{5} \sum_{m=1}^5 O_{peak}^m} \right) \quad (3.2)$$

where C_{Entry} and C_{Exit} denote the Congestion Index for entry and exit numbers respectively, W_i is the weight for i^{th} station, I, O denote get-on and get-off numbers, respectively, N denotes the number of stations selected for that venue, and m denotes the number of past days taken for comparison. We first compute ratio of peak value on the day of the event to the average peak value observed in the last m days, for hourly get-on and get-off numbers separately, and then take the weighted average of these ratios station-wise. Equation 3.1 and Eq. 3.2 compute a weighted average of station-wise computed values which are defined as the Entry/Exit Congestion Index. The weights are determined by the share of the passengers for each station. We take value of m as five days for the purpose of computational efficiency. Moreover, these past five days are chosen in a manner to have same weekday type as the event-day to avoid any inherent weekday bias in railway traffic. A congestion index value of 1 will indicate no change in railway congestion compared with past five days, while a value of more than 1 will show an increase in the congestion.

3.5 Accident Analysis Methodology

In this section, we focus on the various indicators and features used to analyse the effect of accidents on railway traffic. We also explain the methodology used to identify the railway commuters and their commute time.

3.5.1 Effect on Railway Commuters

To assess the impact on commute time, it is important to identify the daily railway commuters and their average commute time beforehand in order to analyse what effect a particular accident had on their commute time compared to ordinary days. However, the same kind of comparison cannot be done for non-commuters since their commute time is not fixed and they do not travel regularly.

3.5.1.1 Identification of Railway Commuters

A commuter can be defined as a regular passenger who periodically travels between home and destination, and has high homogeneity in his travel destination and departure/arrival hours. This homogeneity can be measured by measuring entropy. The Entropy is defined as a measure of disorder in a given distribution, and can be computed as shown in the Eq. 3.3.

$$E = (-1) \sum \frac{P_i \log_2 P_i}{\log_2 I} \quad (3.3)$$

where E is the entropy value, P_i is the proportion of i^{th} label in the distribution, and I is the total number of items. If the distribution is random, then entropy will be higher, and if distribution is ordered and regular, then its entropy will be lower. So using this principle, railway commuters are identified based on their entropy of distribution of home/destination stations and departure/arrival hours.

The entropy is computed in following steps for the railway passengers:

- Identify home/destination stations, and departure/arrival hours for each railway user

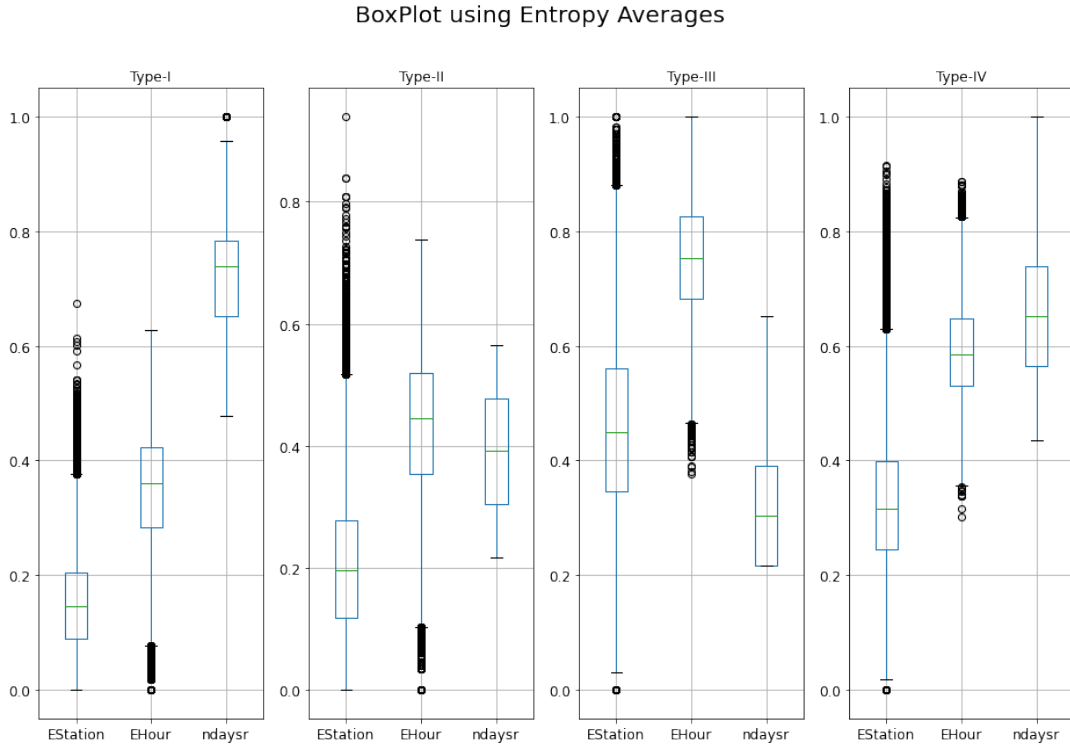


Figure 3.2: Boxplots of the entropy features for the four clusters of railway passengers

- Compute the Entropy of distribution of home/destination stations and departure/arrival hours for each railway user, denoted respectively as $EStation$ and $EHour$
- K-means clustering based on multi-dimensional features of $EStation$, $EHour$, and $ndaysr$ (normalized no. of days of railway travel in a month)
- Label the clusters having low level of entropy as potential railway commuters

The value of k (number of clusters) is kept as 4 based on the knee-point detection method. In the figure 3.2, Boxplots for these four clusters are shown. The figure shows that, type-1 and type-2 clusters have low level of entropy, and high no. of travel days, while type-3 and type-4 clusters have high level of entropy. The only difference in type-1 and type-2 is that type-2 users travel less frequently compared to type-1. Therefore, it is evident that type-1 and type-2 users have low level of entropy, meaning that they have fixed set of home/destination station and rather

uniform distribution of arrival/departure hours, thereby they are the candidates for being classified as railway commuters. We categorize these type-1 and type-2 users as commuters and are used in further study of accidents, while type-3 and type-4 having high level of entropy, are considered as non-commuters.

3.5.1.2 Commute Time Increase Factor

When a railway line's operation gets suspended or delayed due to any accident or incident, the commuters who use those lines, get affected. They will need to use some alternate railway line or mode of transportation to reach their destination, resulting in increase of their commute time. Here we try to find the increase in commute time for the affected railway commuters on the day of the accident compared to ordinary days.

Following steps are followed to find the average increase in commute time:

- A time window of 6 hours (3 hours before the start, and 3 hours after the end of the accident) is selected for commute time determination.
- The last five ordinary days are chosen for comparison (excluding the holidays or weekends)
- All the identified railway commuters, who used the affected railway line at least 3 times in last 5 ordinary days within the specified time-window, are selected. They are called 'Potential Affected Commuters', and considered to be regular users of the affected line.
- Out of Potential Affected Users list, only those users are extracted who travel on the day of accident day within specified time-window. They are called 'Affected Commuters'.

- The average railway commute time for all Affected Commuters is computed using their GPS trajectory data for last five ordinary days, and for the day of the accident.

Finally, we define the **Commuter Time Increase Factor** as following:

$$CTIF = \frac{\text{mean commute time}_{(accident\ day)}}{\text{mean commute time}_{(ordinary\ day)}} \quad (3.4)$$

3.5.1.3 Affected Commuters Features

Using the GPS mobility data, it is possible to also analyze how the behavior of railway commuters changed during a particular incident. We provide following additional statistics about Affected Commuters in order to understand effect of the accident:

- The share of railways lines and railway companies used during last five ordinary days by Affected Commuters
- The share of railways lines and railway companies used on accident day by Affected Commuters
- The share of new railway lines and railway companies used on the accident day by Affected Commuters, which were not taken up in last five ordinary days
- The share of home and destination cities for the Affected Commuters

3.5.2 Line Load Drop Factor

In last subsection, we discussed how we can identify railway commuters and their commute time to analyse the effect of the accident on the commuters. However, some accidents can occur during weekends or outside the rush hours, when regular commuter are not likely to travel. In these cases, the Commute Time Increase Factor will not be useful. Therefore, we need to find a different indicator which can evaluate an accident severity irrespective of time of occurrence. For this purpose, we choose the parameter of passenger load volume in the affected line. If the train services get affected, then the result would show in the drop of the passenger load volume in the affected line. Then the comparison of this load volume drop with last five days can give us the measure of accident severity.

We compute the average passenger load volume for all links of the affected line present within the buffer area of 15 km square centered at the accident location during the accident time-window, on the day of the accident, and also for last five ordinary days (same type of days).

The **Line Load Drop Factor** is defined as following:

$$LLDF = 1 - \frac{\text{mean passenger load volume}_{(\text{accident day})}}{\text{mean passenger load volume}_{(\text{ordinary day})}} \quad (3.5)$$

Chapter 4

Event Analysis: Dashboard Design and Case Studies

4.1 Layout introduction

Figure 4.1 shows the front-end design layout of the browser-side application. The layout is inspired by the official map example provided by Plotly Dash. The dashboard primarily consists of three panels which are -

1. Control Panel,
2. Map Panel,
3. Figure Panel

Control Panel: Control panel is mainly for selecting and setting the parameters and to generate desired statistics results and visualizations inside the figure and map panel. Users can select the events and stations that they want to analyze and visual-

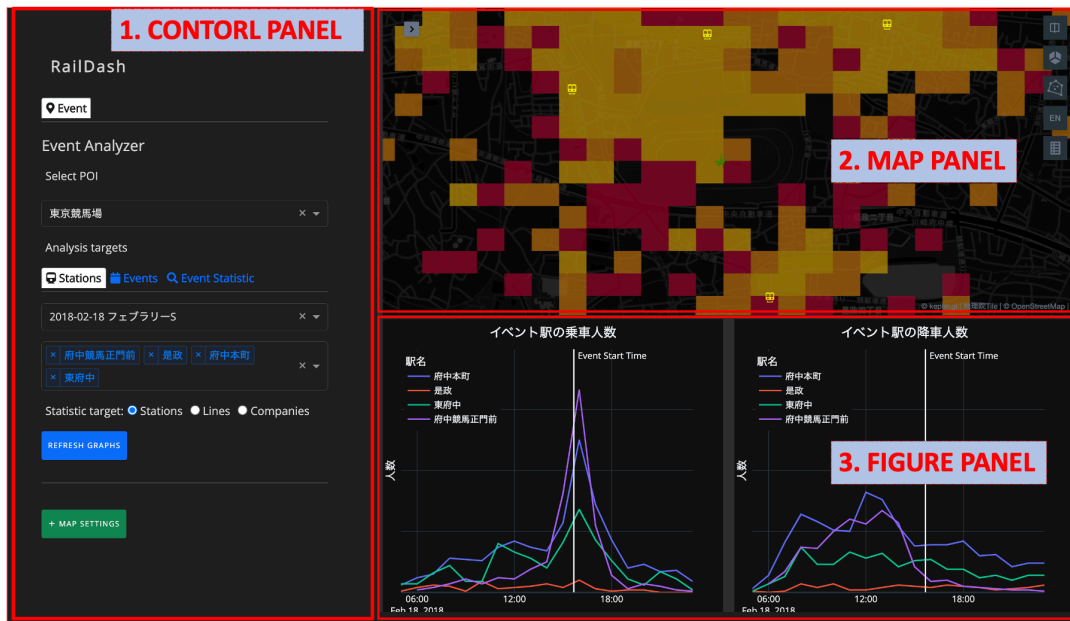


Figure 4.1: The Layout-design of the Dashboard.

ize from the control panel. In addition, a collapsible component called Map Setting is provided at the bottom of the control panel for more customizable interaction with the map visualization panel.

Map Panel: Map panel uses an extended Kepler.gl package ¹ for map visualization. Compared to other map visualization tools, Kepler.gl is much better suited for big data visualization and map animation, and thus is more powerful and faster in visualizing passenger density and human mobility using big GPS trajectory data. However, Kepler.gl cannot interact with Plotly Dash in its default form. Moreover, the controller in Kepler.gl can be difficult for people who are not familiar with GIS to manipulate. To solve these limitations, this study improves upon the communication part of Kepler.gl and makes it intractable with Plotly Dash in the front-end part by rewriting some parts of the source code.

Figure Panel: Figure panel consists of the figures created by Plotly Dash and is visible in all modules. The figures are refreshed based on the conditions selected by

¹ <https://natsuapo.github.io/keplerjis/>

users from the control panel. And just like all other Plotly Dash applications, the figures are interactive to set the filters and zoom levels directly by the user. It is to be noted that due to data security issues, we do not visualize the ticks of y-axis data during the demonstration of the dashboard.

4.2 Modules and User-Interaction Introduction

As mentioned in the system introduction, this application provides three separate modules for users to interact. The user can first select the venue of the events from a drop-down menu. Once a venue is selected, the dashboard loads the data of the events and stations belonging to that venue. Then the user can interact with the system using three separate modules where each module has separate functionality. These three modules are Railway Passenger Analysis, Events Comparison, and Event-Participants Statistics, respectively. The functionalities of these modules are now explained in detail along with some real-world case-studies for better understanding.

4.2.1 Railway Passenger Analysis Module

We utilize computed get-on and get-off numbers of stations for analyzing and visualizing railway passenger volume during the big events. In this module, users can select one event from the selected venue and can choose multiple stations which are to be analyzed and visualized. The aggregated get-on and get-off passenger numbers from the selected stations on the event date are visualized in the figure panel, while the passenger's local origin and destination location density are encoded to the standard grids of 125-meter Japanese industrial standardized (JIS) mesh code ²

² https://nlftp.mlit.go.jp/ksj/old/old_data_mesh.html



(a) Station-wise Get-on and Get-off numbers figure.

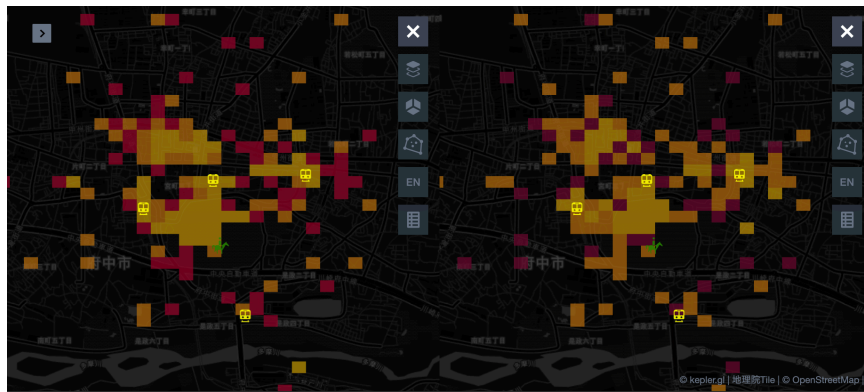


(b) Railway Company wise Get-on Get-off numbers figure.

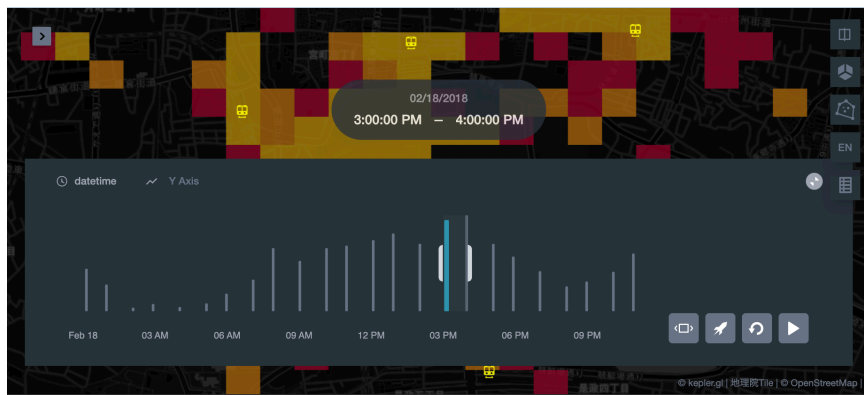
Figure 4.2: Visualization of Get-on and Get-off numbers for the event of 27th May, 2018 at Tokyo Race Course venue.

and visualized via the extended JIS mesh code layer in the map panel. Users can also choose to visualize either origin or destination density of the selected stations from the control panel. In addition, we also provide the user the option of choosing aggregated get-on and get-off numbers using railway lines or railway companies in place of stations.

Case-Study: To further demonstrate the functionality of this module, we choose a very popular horse race event as the case study. The event took place on 27th May, 2018 at Tokyo Race Course venue and an estimated 126,767 people attended the event as per the official website. This event is known as Japanese Derby and is an international Grade-1 flat horse race in Japan. There are four stations which fall inside the 1.5 km buffer area around the venue, which are - Fuchukeibaseimon-mae, Higashi-fuchu, Fuchuhommachi, Koremasa. The start time of the event was 15:40 hrs.



(a) OD density maps in dual-maps view: the left and the right maps are respectively the origin and destination density of all railway passengers from selected stations



(b) Origin density visualization with one hour filter and animation bar

Figure 4.3: Visualization of density maps in the map panel

Figure 4.2a shows the plots for get-on and get-off numbers for all these four stations separately. The peak observed in the get-on numbers around and after the event is quite evident from the figure, which shows that the departure pattern of most of the audience is concentrated just after the end of the event. Similarly, in the get-off numbers plot, the distributed peaks before the start of the event indicate the distributed arrival patterns of the audience and are not concentrated around a particular time. The user can also see which station gets the most passengers and can compare the congestion of four stations temporally. Figure 4.2b, on the other hand, shows the get-on get-off numbers using railway company-wise aggregation instead of station-wise aggregation.

Figure 4.3 shows the local OD density estimation of railway passengers in the map

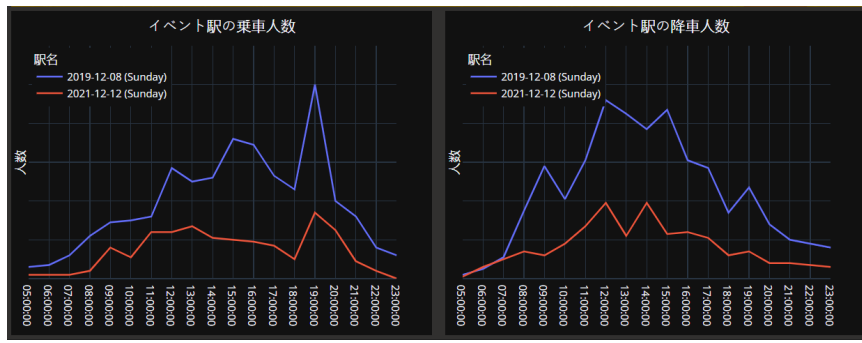
panel. This map utilizes three layers to respectively visualize railway stations and horse racing fields in icon layers, and OD density in a Japanese Standard Mesh layer. By looking at the color-graded density map, the user can estimate the crowd condition at various places nearby the venue. Moreover, the user can also change density map visualization via the control panel by adding temporal filters, by changing the density type (origin density by default), or visualize both origin and destination density via a dual-map view.

4.2.2 Event Comparison Module

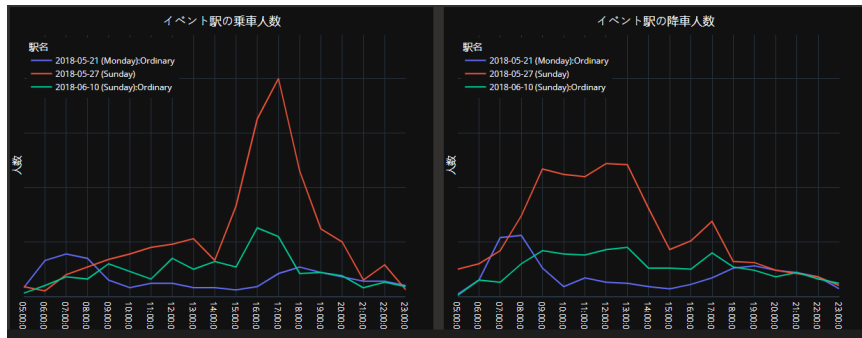
Event comparison module gives the user option of comparing railway passengers volume for different events on different dates which were organized at the same venue. The user can choose multiple events and the station to be analysed from the drop-down menu in this module. The figure panel will show the comparison of get-on and get-off numbers for different events for the selected station. We also give the user the option of comparing the get-on get-off numbers on the day of the event with a regular non-event day. The non-event days are chosen in a manner to have both weekday and weekend dates.

Although this module is quite similar to the previous Railway Passengers Analysis module in terms of visualization, it is more useful in analyzing and comparing the temporal changes in passenger volume for different events on different dates for a particular station of interest.

Case-Study: We choose Tokyo Dome live events as the case study to demonstrate the functionality of this module. Here we try to demonstrate two different scenarios - comparison of the same type of event organized in pre-COVID and post-COVID times, and comparison of an event with an ordinary non-event day. For



(a) Comparison of two different events.



(b) Comparison of an event with ordinary day.

Figure 4.4: Visualizations in event-comparison module.

the first scenario, we choose two live concert events performed by the same band at the same time of the day, organized on 8th December, 2019, and on 12th December, 2021, respectively. The selected station is Korakuen station. Figure 4.4a shows the comparison of get-on and get-off numbers for Korakuen station for the two events. It is evident that the station was more crowded during the pre-covid time event. On the other hand, fig. 4.4b shows the comparison of the Tokyo Horse Race event of 18th May, 2018 with ordinary days of 2018-05-21 and 2018-06-10 for the Fuchukeibaseimon-mae station. We can clearly see that the event caused quite large peaks compared to ordinary day numbers.

4.2.3 Event-Participants Statistics Module

This module focuses on the analysis and visualization of event participants' information as a whole instead of just railway passengers. In this module, the extracted

event participants' information is further aggregated and processed to provide a more comprehensive statistical result about the event. For each event, various statistical results about event participants are displayed in the figure panel. These general event statistics help users to understand event participants' information as a whole, for example, the event's start time, visiting numbers from different regions, the transportation mode ratio, railway usage ratio, etc. The map panel is not required in this module.

The details of the available statistics are as follows:

Congestion Index: It is the relative scale of the congestion caused by the event in get-on and get-off numbers compared to past 5 days of same weekday type. A value of higher than 1 will show the relative increase in the congestion. For example, a value of 2 means 100% increase in the congestion. This index assigns a single value to the event for its effect on railway traffic, therefore, being valuable in assessing and comparing different events.

Railway Usage Statistic: It shows the ratio of the railway stations used by those event participants who used railway service to reach the event venue, in the form of a pie-chart. This figure provides three further options to the user to display the usage ratio result - station wise, line wise, company wise.

Transportation Mode Ratio: Here we display the ratio of transportation mode used to reach the venue by the event participants in form of a pie chart. The transportation modes available are- *WALK, BIKE, CAR, TRAIN*.

Origin Prefecture Ratio: Origin prefecture is the prefecture where the event participant comes from, to attend the event. This is again displayed in the form of a pie chart. For better visualization performance, we only show the top five prefectures, while sum up the remaining ratio as the ratio from other prefectures (others).

Duration of Visit: This is a histogram figure that shows distribution of the time spent at the venue by the event participants.

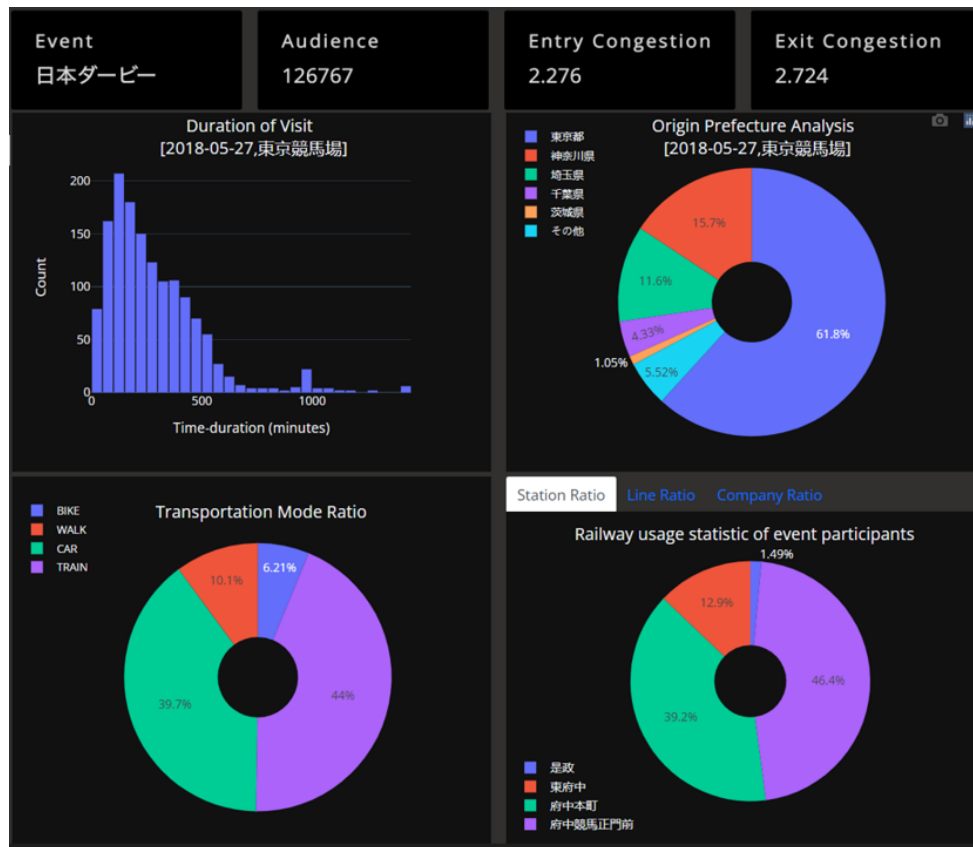


Figure 4.5: Event-statistics results.

Case-Study: For the case-study of this module, we once again choose the big event of the Tokyo Horse Race on 27th May, 2018. Figure 4.5 shows the query result of event participant statistic module. The query result includes two part, the upper part includes four cards which are respectively the event name, the audience number which collected from the internet, the entry congestion index and the exit congestion index. The lower part includes various of statistics figures. As is shown in the card part, the entry and exit congestion index were found to be 2.276 and 2.724, respectively, showing an increase of almost 150% in the congestion. From the figure part, We can also infer that almost 44% of identified event participants used railway transportation to reach the venue. Among the participants, who used railway transportation, almost 47% got off at Fuchukeibaseimon-mae station to reach the venue.

Almost 62% of the participants were from Tokyo prefecture, while Kanagawa and Saitama prefectures were distant second and third. Another interesting statistic is that most of the participants spent up to 5-6 hours at the venue, which indicates that the participants did not only enjoy Japan Derby, but also participated in the previous and post horse racing games held on the same day.

Chapter 5

Accident Analysis: Dashboard Design and Case Studies

5.1 Layout introduction

Figure 5.1 shows the front-end design layout of the browser-side dashboard application for accident analysis. The layout of this dashboard is quite similar to the event analysis dashboard as discussed in section 4.1. This accident analysis dashboard also consists of three panels which are Control panel, Map panel, and Figure panel, respectively. The functionality of these panels is quite similar as already discussed in section 4.1, but differ in some features which will be explained later in this chapter.

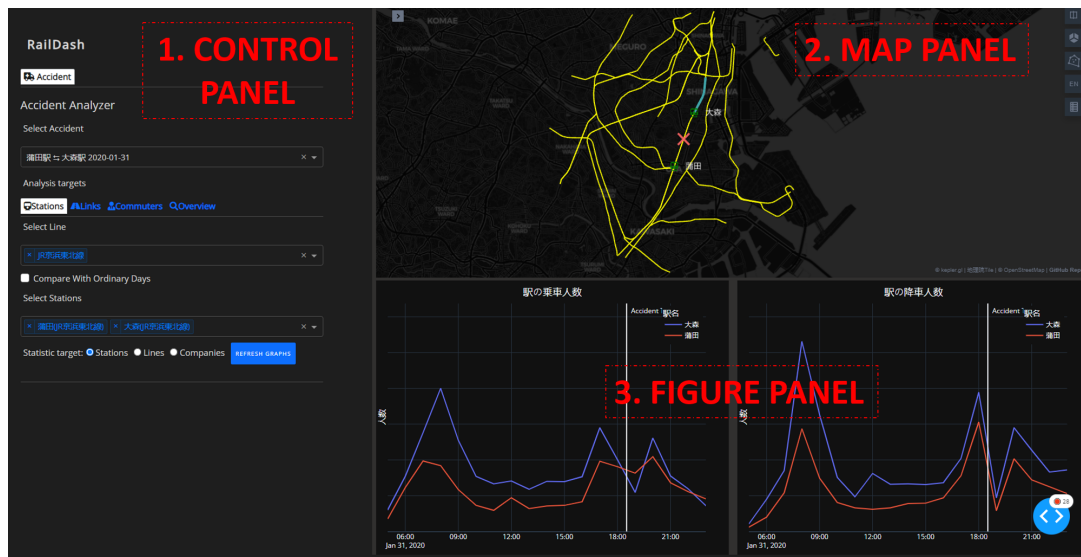


Figure 5.1: The Layout-design of the Dashboard for Accident Analysis.

5.2 Modules and User-Interaction Introduction

This application provides four separate modules for users to interact. The user can first select the accident of interest from a drop-down menu from the control panel. Once an accident is selected, the dashboard loads the data of the accident and the information about railway station and railway links within the 15 km. square buffer centered at the location of the accident. Then the user can interact with the system using four separate modules where each module has unique functionality. These four modules are (i) Station Analysis, (ii) Link Analysis, (iii) Commuters Analysis, and (iv) Accident Overview, respectively. The functionalities of these modules are now explained in detail along with some real-world case-studies for better understanding.

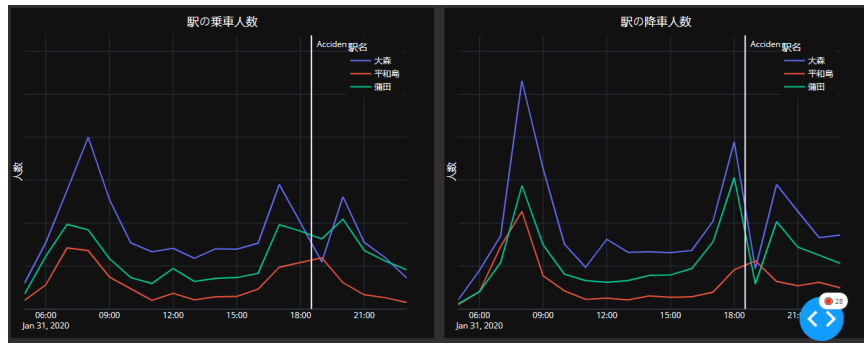
5.2.1 Station Analysis Module

We utilize computed get-on and get-off numbers of stations for analyzing and visualizing railway passenger entry/exit numbers. This module is similar to the module

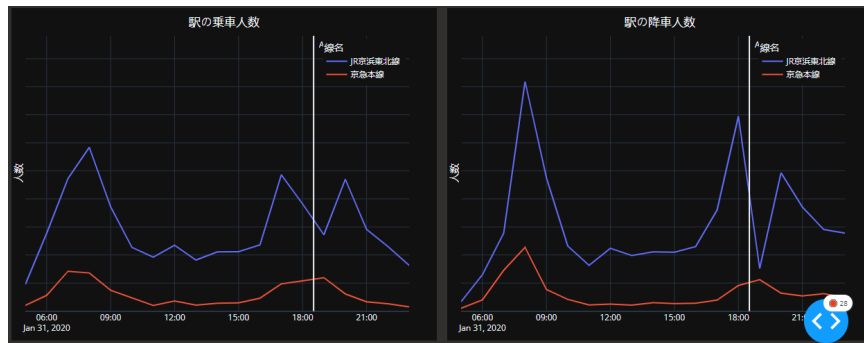
already discussed the subsection 4.2.1 of Event-Dashboard and plots the aggregated get-on and get-off passenger numbers for the selected stations in the figure panel. The time unit of the aggregation is taken as 1-hour. The user can first choose the railway line of his choice from the drop-down menu, and then can choose the station within that line from another drop-down menu. These drop-down menus present the user option of multiple selection, and thus user can choose more than one station for the study if desired. This module also provides the user option of comparing the entry/exit plots with last five ordinary days average in order to understand the effect of the accident. In addition, the user also has the option of choosing the target of the aggregation by either using railway lines or railway companies in place of railway stations.

In addition, the map panel in this module shows all the railway links and stations present within the 15 km. square buffer area. The location of accident is also indicated on the map by red 'X' symbol. All the selected stations by the user are shown on the map using 'station' symbol. Thus user can keep the track of selected stations and their relative locations.

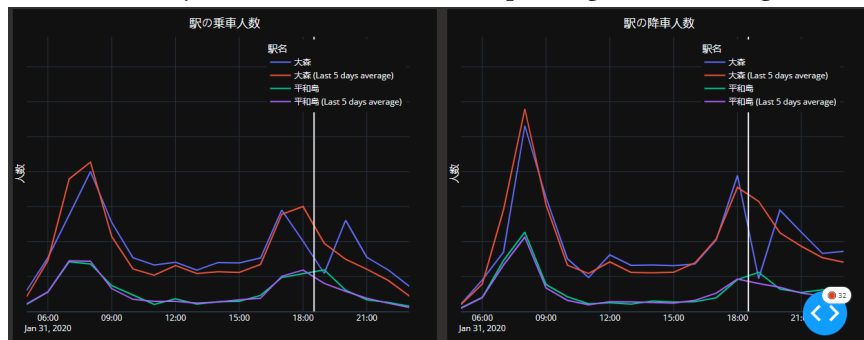
Case-Study: To further demonstrate the functionality of this module, a case study of a railway accident is presented here, which occurred on 31st January, 2020 on Keihin-Touhoku Line between Oomori and Kamata station. The trains services were suspended in Keihin-Touhoku line from 18:31 hrs to 20:01 hrs. Figure 5.2 shows the plots of aggregated get-on and get-off passenger numbers in 1-hour period for the Oomori, Kamata stations (Keihin-Touhoku Line), and Heiwajima station (Keikyu Main Line). Figure 5.3 shows the respective positions of these three stations from the Map panel. While fig. 5.2a shows the plots station-wise, fig. 5.2b on the other hand shows the plots railway line-wise. It is evident from the figures that both entry and exit numbers sharply fell during the accident time-window for Oomori and Kamata station, while numbers increased for Heiwajima station, showing that



(a) Station-wise Get-on and Get-off passenger numbers figure.



(b) Railway Line wise Get-on Get-off passenger numbers figure.



(c) Comparison of Get-on Get-off passenger numbers with last five days average figure.

Figure 5.2: Visualization of Get-on and Get-off passenger numbers for the accident of 31st January, 2020 between Oomori-Kamata stations on Keihin-Touhoku Line.

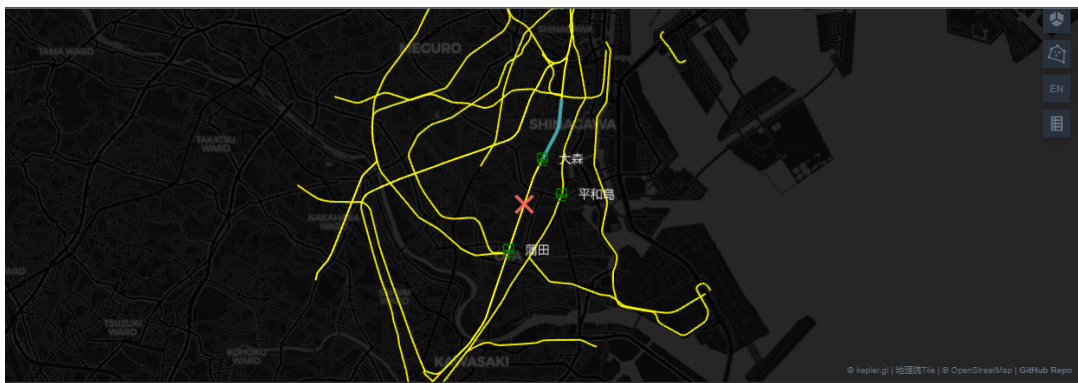


Figure 5.3: The Map panel in the Station-Analysis module.

some passengers shifted to the Keikyu main line during the accident duration. Figure 5.2c shows the plot comparison with last five ordinary days average.

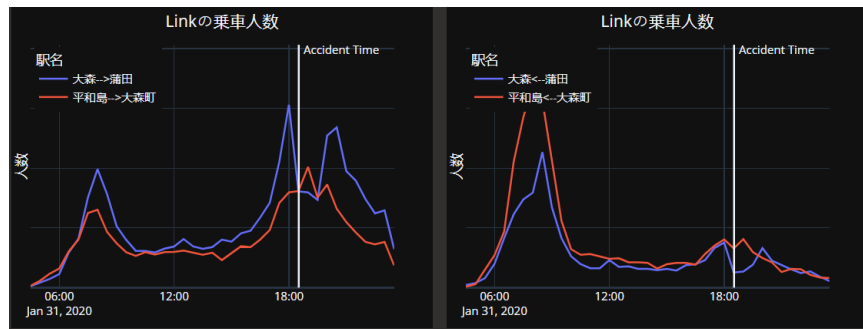
An interesting point of observation is that decline in Exit-figures have a delayed effect compared to Entry-figures. The potential reason could be that passengers take their time to exit from the stations and look for alternate means of transportation after some significant wait time only.

5.2.2 Link Analysis Module

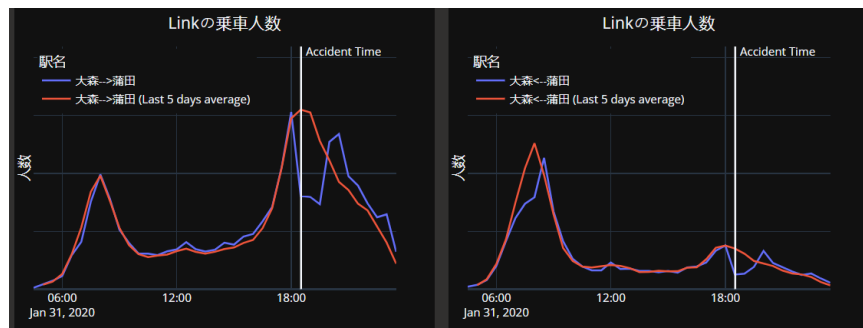
In this module, we utilize computed railway passenger numbers of railway links for analyzing and visualizing railway passenger load volume. This module provides the user the function of plotting aggregated passenger load volume in the figure panel for the chosen railway links present within the 15 km. square buffer area. The time unit of the aggregation is 30-minutes. The user can first choose the railway line of his choice from the drop-down menu, and then can choose the railway link from another drop-down menu. These drop-down menus present the user option of multiple selection, and thus user can choose more than one link for the study if desired. The module plots the passenger load volume in both direction for the given link. The user also has the option of comparing the load volume with last five days average value.

The map panel in this module shows the selected links by changed colors, and thus user can keep the track of selected links and their relative locations.

Case-Study: We once again refer to same accident case discussed in previous section 5.2.1. Figure 5.4a shows the plots of passenger load volume bidirectionally for Oomori-Kamata link (Keihin-Touhoku line) and Heiwajima-Oomorimachi link



(a) The Passenger Load Volume plots for Oomori-Kamata link (Keihin-Touhoku) and Heiwajima-Oomorimachi link (Keikyu) in both directions for date of 2020-01-31.



(b) Passenger Load Volume comparison for Oomori-Kamata link with last five days average.

Figure 5.4: Visualization of passenger load volume numbers for the accident of 31st January, 2020 between Oomori-Kamata stations on Keihin-Touhoku Line.

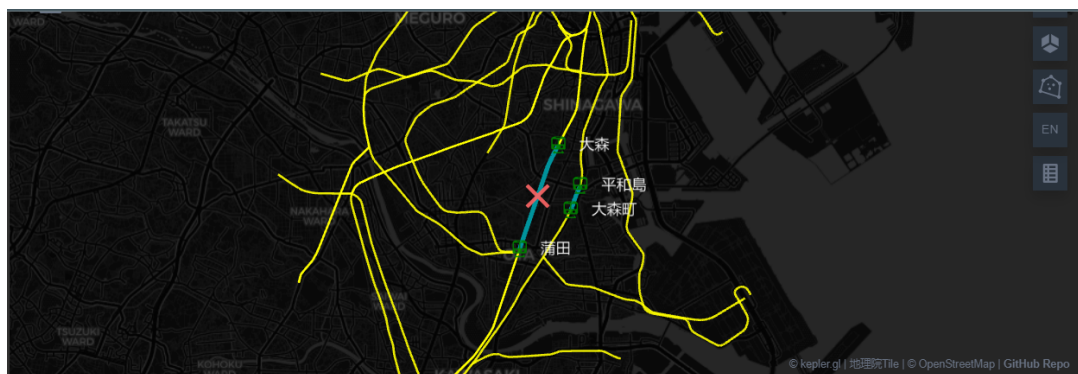


Figure 5.5: The Map panel in the Link-Analysis module.

(Keikyu line) for date of 2020-01-31 in 30-minute time interval. It is clear from the figure that there is sharp fall just after the start of the accident for Oomori-Kamata link, and the regular load volume resume only around 20:00 hrs, which suggests that is the time of resumption of the service on the affected line. While for the Heiwajima-Oomorimachi link, which is part of parallel Keikyu line, some small sudden increment can be observed during the accident duration, suggesting shifting of passengers to this line. Figure 5.4b, on the other hand, shows the comparison of passenger load volume for Oomori-Kamata link with last five day average load volume. Again, the sharp fall can be seen on the day of accident compared to the average value in this figure. Figure 5.5 shows the map panel of this module which indicates the respective position of the two links on the map.

5.2.3 Commuter Analysis Module

This module focuses on the analysis and visualization of affected railway commuters' information, and other statistics about the accident in general. For each accident, various statistical results and charts about affected railway commuters are displayed in the figure panel. The map panel is not needed in this module. The details of the available statistics and indicators are as follows:

Commute Time Increase Factor: It is the measure of increase in commute time by railway mode for the affected railway commuters on the day of the accident in comparison to last five ordinary days. If the value is more than 1, that means the mean commute time increased and was adversely affected due to the accident.

Line Load Drop Ratio: It is the measure of checking accident adversity by comparing the passenger load volume in the links of affected railway line on the day of the accident with last five ordinary days average. More the value is close to 1, higher

the drop in passenger load volume will be. This indicator is useful to analyse those accidents where commuter information is not useful.

Commuters Line-share Comparison Charts: The share of different railway lines and railway companies used by the affected railway commuters are displayed in form of pie-charts for both day of the accident and last five ordinary days. These figures give the idea about how the shares of different lines changed due to the accident.

Commuters Home/Destination Cities-share: The share of home and destination cities for the affected railway commuters are shown in form of pie-charts. These figures give the general idea about which cities' commuters were most affected by the accident. Only top ten cities are displayed.

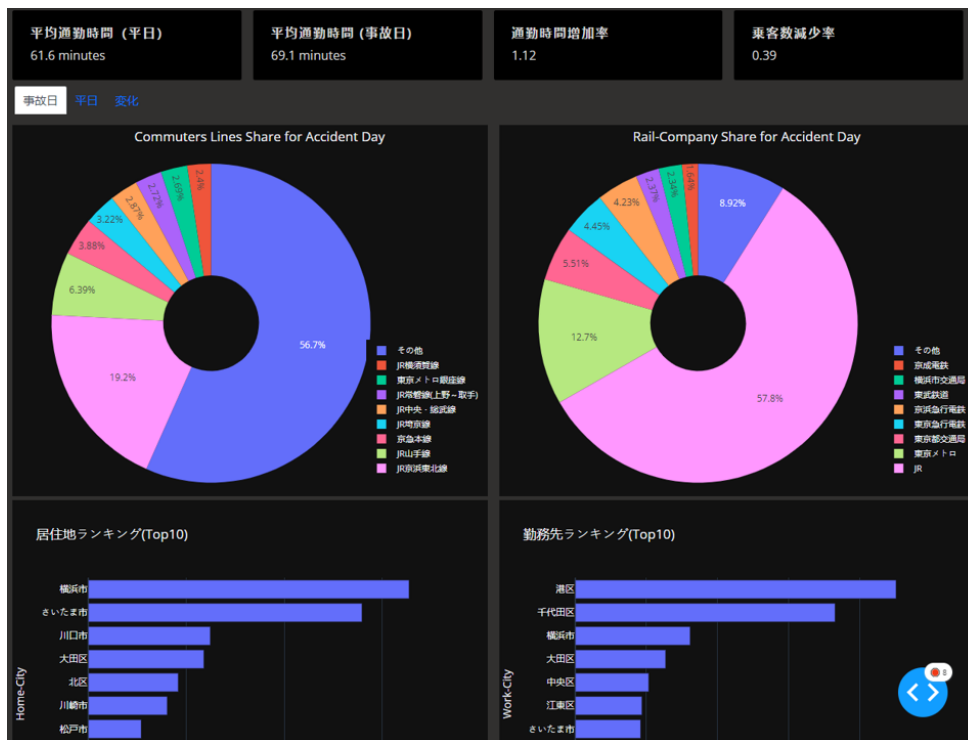
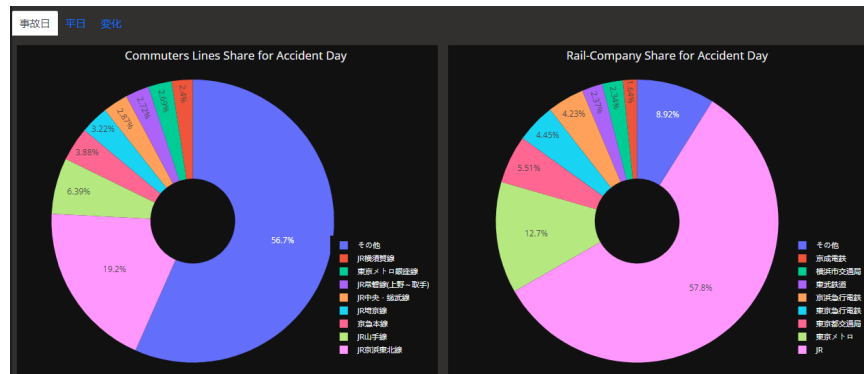
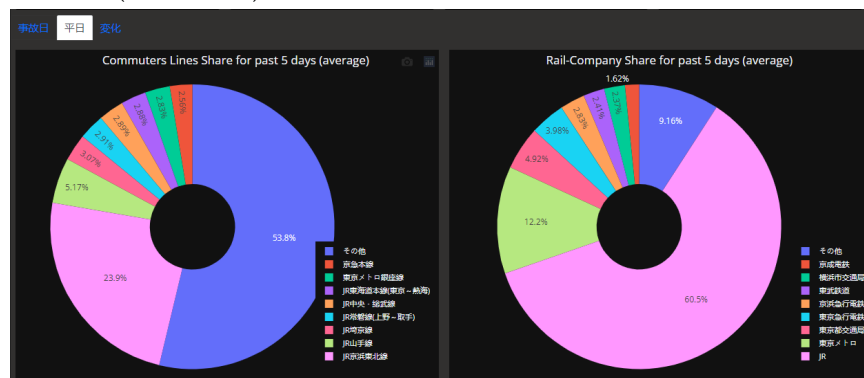


Figure 5.6: Commuter analysis and other accident statistics results.

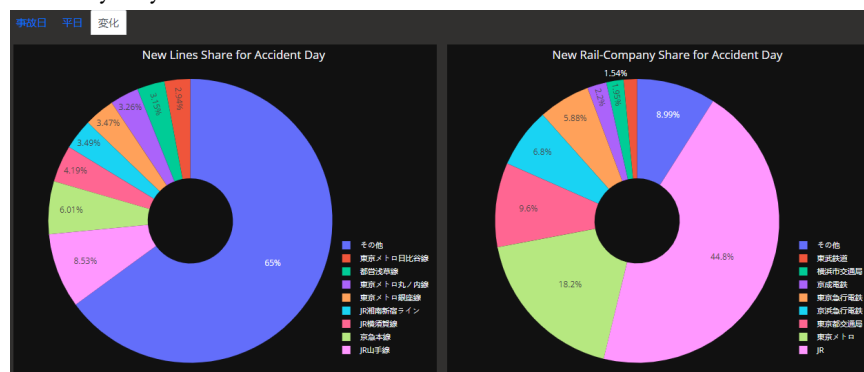
Case-Study: We once again choose the case of accident on Keihin-Touhoku line occurred on 31st January, 2020. Figure 5.6 shows the entire figure panel overview of this accident. Various indicators like average commute time on day of accident and in last five days, commute time increase factor, and line volume drop factor are



(a) Share of lines/companies for affected commuters on the day of the accident (2020-01-31).



(b) Average share of lines/companies for affected commuters in last five ordinary days.



(c) Share of new lines/companies used during accident by affected commuters.

Figure 5.7: Visualization of line/company share of affected commuters for the accident of 31st January, 2020 between Oomori-Kamata stations on Keihin-Touhoku Line.

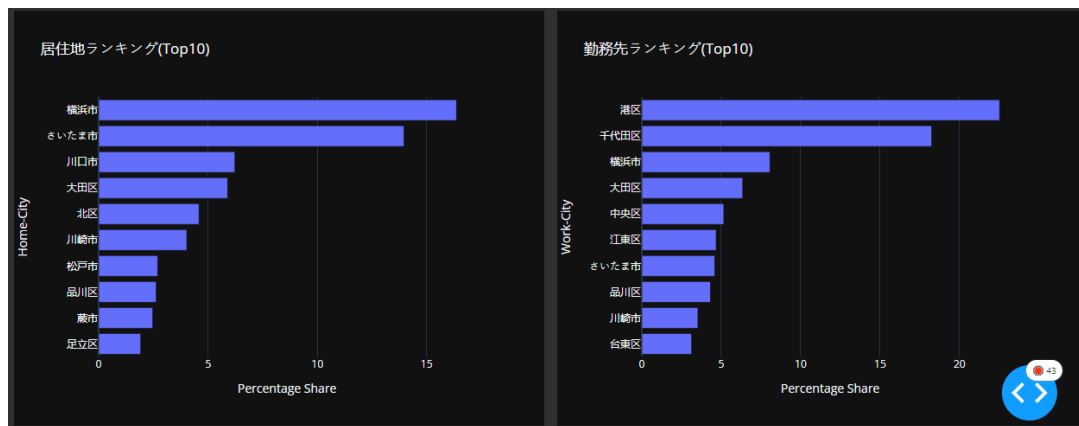


Figure 5.8: Share of home and work/destination cities for the affected railway commuters for the accident of 31st January, 2020 between Oomori-Kamata stations on Keihin Line.

displayed on the top in four cards for the ready reference. We can see that the Commute Time Increase Factor is 1.12, while Load Volume Drop Factor is 0.39. Figure 5.7 shows the line-share and company-share for the affected railway commuters on the day of the accident and for last five ordinary days. We can see from the fig. 5.7b that affected line Keihin-Touhoku line had average 23.9% share in last five days, but the same dropped to 19.2% for the day of accident (as shown in fig. 5.7a). Moreover, fig. 5.7c shows that Yamanote line saw the highest jump in share (8.53%) when it comes to new lines taken by the affected commuters, followed by Keikyuu line (6%). Figure 5.8 shows the share of top ten home and destination cities for the affected railway commuters. It is evident from the figure that commuters living in cities of Yokohama, Saitama, and Kawaguchi and commuting to Minato-ku and Chiyoda-ku were the most affected.

5.2.4 Accident Overview Module

This module mainly focuses on the visualization of link based passenger load volume using the map panel. Figure 5.9 shows the overview of the module within the dashboard. In the map panel, all the railway links present in the buffer area are shown

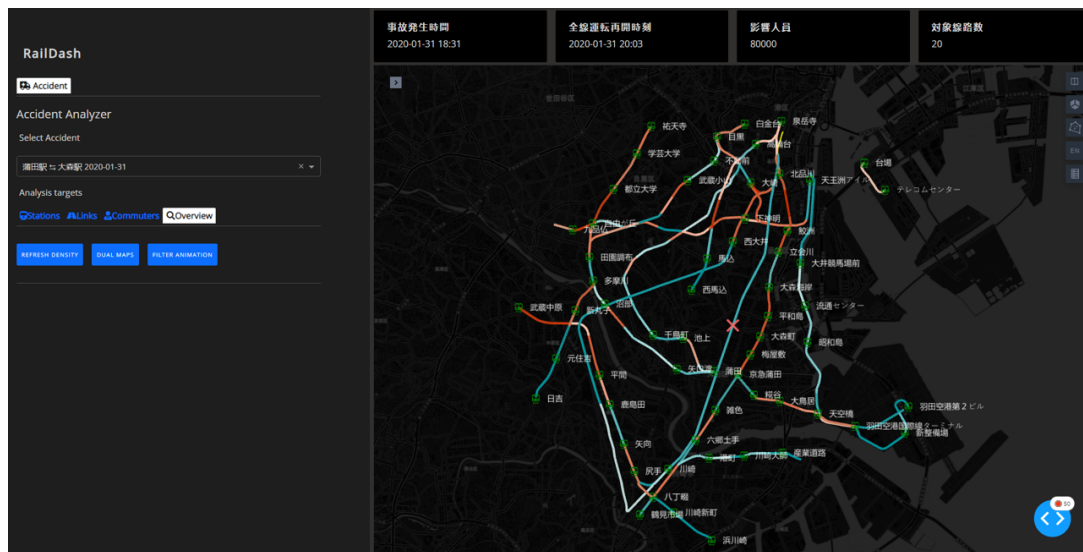


Figure 5.9: Accident-Overview Module for the accident of 31st January, 2020 between Oomori-Kamata stations on Keihin Line.

and are color-coded as per their passenger load volume numbers. The shift to color red indicates high passenger volume.

In the control panel side, we have three click options, (i) Refresh Density, (ii) Dual Maps, (iii) Filter Animation. The refresh density button refreshes the entire map figure and brings it back to its default setting. Dual maps button divides the map panel in two split parts and visualizes the passenger load volume for each direction separately in dual mode. Filter animation button creates the time-based animation of how the passenger load volume changes during the day.

Chapter 6

Conclusion

In this study, we proposed a novel and generic dashboard system for analyzing and visualizing the effect on railway traffic during big events and unforeseen incidents through big GPS trajectory data. This is the first study to use big smartphone GPS data to analyze the railway traffic during events/accidents to the best of our knowledge. By using big GPS trajectory data generated by millions of users, we are able to implement a comprehensive and multi-faceted study for the events/accidents which not only include railway passenger analysis but also mobility study of event participants and affected railway commuters. This dashboard can have a great significance for railway administrators, event organizers, and city planners to measure and estimate the impact of big events and accidents on railway traffic and also compare different big events using the Congestion Index score. It can also help them to make advance preparation, in terms of resource planning or crowd management, to manage such events/accidents in the future. Furthermore, we also demonstrate the successful integration of Plotly Dash with the Kepler.gl platform in our dashboard, enabling the user to directly interact with Kepler maps from within the dashboard console. We also showed the utility of the dashboard by providing some case studies

of real-world events and accidents.

The limitation of using GPS dataset is the potential inaccuracies in transportation mode detection and map matching. The sparsity of the GPS data also adds to this problem. Another point of discussion can be the identification criteria of the target stations, which can be subjective and may differ for different venues. In addition, the interaction between the dashboard and the Kepler map is one-way at present, i.e. the user cannot use the map panel to control graphs and figures in other panels.

In the future, we will further enhance the map visualization part for better map interaction and create a bigger database of different types of big events and accidents. We also plan to integrate the simulation of railway traffic during hypothetical accidents or suspension of train services in a particular railway line into our dashboard. Moreover, we plan to extend this research for prediction of congestion index for future events by using machine learning algorithms.

Bibliography

- [1] L. T. Authority *et al.*, “Passenger transport mode shares in world cities,” *Journeys. Singapore: Land Transport Authority*, 2011.
- [2] M. Ni, Q. He, and J. Gao, “Forecasting the subway passenger flow under event occurrences with social media,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 6, pp. 1623–1632, 2016.
- [3] M. L. Filograno, P. C. Guillén, A. Rodríguez-Barrios, S. Martín-López, M. Rodríguez-Plaza, Á. Andrés-Alguacil, and M. González-Herráez, “Real-time monitoring of railway traffic using fiber bragg grating sensors,” *IEEE Sensors Journal*, vol. 12, no. 1, pp. 85–92, 2011.
- [4] T. Kusakabe, T. Iryo, and Y. Asakura, “Estimation method for railway passengers’ train choice behavior with smart card transaction data,” *Transportation*, vol. 37, no. 5, pp. 731–749, 2010.
- [5] L. Sun, D.-H. Lee, A. Erath, and X. Huang, “Using smart card data to extract passenger’s spatio-temporal density and train’s trajectory of mrt system,” in *Proceedings of the ACM SIGKDD international workshop on urban computing*, 2012, pp. 142–148.
- [6] J. C. Herrera, D. B. Work, R. Herring, X. J. Ban, Q. Jacobson, and A. M. Bayen, “Evaluation of traffic data obtained via gps-enabled mobile phones: The mobile

- century field experiment,” *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 4, pp. 568–583, 2010.
- [7] T. Sense, “Rich monitoring of road and traffic conditions using mobile smart-phones,” Microsoft Research, Tech. Rep. MSR-TR-2008-59, Tech. Rep., 2008.
- [8] S. He, “From beautiful maps to actionable insights: Introducing kepler. gl, uber’s open source geospatial toolbox,” *Uber Engineering*, vol. 29, 2018.
- [9] S. Kawamura, Y. Tomita, M. Itoh, D. Yokoyama, M. Toyoda, and M. Kitsuregawa, “An effective use of tokyo metro passengers flow by visualization of smart card ticket ‘pasmo’ origin-destination data for public transport network to be sustainable,” *Proc. WECC*, 2015.
- [10] B. Mo, Z. Ma, H. N. Koutsopoulos, and J. Zhao, “Capacity-constrained network performance model for urban rail systems,” *Transportation Research Record*, vol. 2674, no. 5, pp. 59–69, 2020.
- [11] A. Tomoeda, M. Komatsu, I. Y. Yoo, M. Uchida, R. Takayama, R. Jiang, and K. Nishinari, “Real-time railway network simulation and alleviating congestion of crowded trains,” in *2009 ICCAS-SICE*. IEEE, 2009, pp. 4932–4936.
- [12] Y. Maekawa, A. Uchiyama, H. Yamaguchi, and T. Higashino, “Car-level congestion and position estimation for railway trips using mobile phones,” in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2014, pp. 939–950.
- [13] K. ITO and T. SAITO, “Information extraction about traffic situation of railway with web camera images.”
- [14] M. Itoh, D. Yokoyama, M. Toyoda, Y. Tomita, S. Kawamura, and M. Kitsuregawa, “Visual fusion of mega-city big data: an application to traffic and tweets data analysis of metro passengers,” in *2014 IEEE International Conference on Big Data (Big Data)*. IEEE, 2014, pp. 431–440.

- [15] C. Zhou, D. Frankowski, P. Ludford, S. Shekhar, and L. Terveen, “Discovering personally meaningful places: An interactive clustering approach,” *ACM Transactions on Information Systems (TOIS)*, vol. 25, no. 3, pp. 12–es, 2007.
- [16] L. Liao, D. J. Patterson, D. Fox, and H. Kautz, “Building personal maps from gps data,” *Annals of the New York Academy of Sciences*, vol. 1093, no. 1, pp. 249–265, 2006.
- [17] D. Ashbrook and T. Starner, “Using gps to learn significant locations and predict movement across multiple users,” *Personal and Ubiquitous computing*, vol. 7, no. 5, pp. 275–286, 2003.
- [18] W. Huimin, H. Zhongwei, G. Jifu, Z. Liyun, and S. Jianping, “Operational analysis on beijing road network during the olympic games,” *Journal of Transportation Systems Engineering and Information Technology*, vol. 8, no. 6, pp. 32–37, 2008.
- [19] T. Xia, X. Song, Z. Fan, H. Kanasugi, Q. Chen, R. Jiang, and R. Shibasaki, “Deep-railway: a deep learning system for forecasting railway traffic,” in *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*. IEEE, 2018, pp. 51–56.
- [20] E. D’Andrea and F. Marcelloni, “Detection of traffic congestion and incidents from gps trace analysis,” *Expert Systems with Applications*, vol. 73, pp. 43–56, 2017.
- [21] A. Pathak, B. K. Patra, A. Chakraborty, and A. Agarwal, “A city traffic dashboard using social network data,” in *Proceedings of the 2nd IKDD Conference on Data Sciences*, 2015, pp. 1–4.
- [22] M. Pi, H. Yeon, H. Son, and Y. Jang, “Visual cause analytics for traffic congestion,” *IEEE transactions on visualization and computer graphics*, vol. 27, no. 3, pp. 2186–2201, 2019.

- [23] K. K. Lwin, Y. Sekimoto, W. Takeuchi, and K. Zettsu, "City geospatial dashboard: Iot and big data analytics for geospatial solutions provider in disaster management," in *2019 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*. IEEE, 2019, pp. 1–4.
- [24] W. Li and W. Zhu, "A dynamic simulation model of passenger flow distribution on schedule-based rail transit networks with train delays," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 3, no. 4, pp. 364–373, 2016.
- [25] A. Witayangkurn, T. Horanont, N. Ono, Y. Sekimoto, and R. Shibasaki, "Trip reconstruction and transportation mode extraction on low data rate gps data from mobile phone," 07 2013.
- [26] S. Ikezawa, H. Kanasugi, G. Matsubara, Y. Akiyama, R. Adachi, and R. Shibasaki, "Estimation of the number of railway passengers based on individual movement trajectories," in *6th INTERNATIONAL CONFERENCE ON CARTOGRAPHY AND GIS*, 2016, p. 249.