Uneven Recovery of UK Rail Passenger Usage After COVID-19:
A Regional Data Science Analysis

Module: IJC437 Introduction to Data Science
Registration number: 250124790
Name: Kyujin Kim

Word count: 2972

1. Introduction and Aims

The COVID-19 pandemic had a profound impact on transport systems worldwide, with rail passenger demand experiencing particularly sharp declines due to lockdown measures, travel restrictions, and changes in working patterns. In the United Kingdom, rail usage fell dramatically during 2020 as commuting and long-distance travel were significantly reduced. Although passenger numbers have increased since restrictions were lifted, official statistics indicate that recovery has been uneven across different parts of the country.

National-level statistics published by the Office of Rail and Road (ORR) show a gradual recovery in overall passenger journeys following the pandemic. However, national averages may conceal substantial regional variation, as regions with different economic structures and travel patterns may follow distinct recovery trajectories. This has important implications for transport planning, as reliance on aggregate trends may overlook persistent regional disparities in rail demand.

This study examines the uneven recovery of UK rail passenger usage after COVID-19 using official regional rail statistics published by the Office of Rail and Road. The dataset provides a consistent time series of passenger journeys across UK regions, making it suitable for analysing both long-term trends and pandemic-related disruptions. By comparing post-pandemic passenger levels with a 2019 pre-pandemic baseline, the analysis moves beyond national averages to provide a clearer assessment of regional recovery patterns.

The topic was selected due to its relevance to current transport policy debates and its suitability for demonstrating a structured data science approach. The use of high-quality official data enables exploratory analysis and quantitative comparison to address a real-world problem with clear social and economic implications, while highlighting how large-scale disruptions can reshape mobility patterns unevenly across regions.

1.1 Research aim:
The aim of this study is to analyse and quantify the uneven recovery of rail passenger usage in the United Kingdom following the COVID-19 pandemic. Using official regional rail statistics, the research compares post-pandemic recovery patterns against a 2019 pre-pandemic baseline to identify whether recovery has been uniform across regions or characterised by persistent regional disparities. By examining both national and regional trends, the study seeks to clarify how aggregate recovery figures may mask underlying spatial inequalities in rail demand.

1.2 Research Questions:
To address the research aim, this study is guided by the following research questions:
- RQ1: How did national rail passenger usage in the UK change before and after the COVID-19 pandemic?
- RQ2: To what extent has the recovery of rail passenger usage been uneven across UK regions?
- RQ3: How do regional recovery patterns compare with the overall national recovery trend?

These research questions are designed to be independent of specific analytical techniques and focus on identifying broader patterns of change and inequality in post-pandemic rail usage.

2. Methodology

This study follows a structured data science process to investigate uneven recovery in UK rail passenger usage after the COVID-19 pandemic. The methodology consists of five main stages: data collection, data pre-processing and cleaning, exploratory data analysis (EDA), construction of a recovery index, and quantitative analysis of regional recovery gaps. Each stage was designed to support the research questions and ensure that the analysis is transparent, reproducible, and appropriate for the characteristics of the dataset.

2.1 Illustration of the methodology

Figure 1 illustrates the overall workflow of the data science process adopted in this study. The analysis begins with the collection of official rail passenger statistics from the Office of Rail and Road (ORR), followed by data cleaning and aggregation at the regional level. Exploratory data analysis is then conducted to understand temporal trends and regional variation. Based on insights from the EDA, passenger journeys are indexed relative to a 2019 pre-pandemic baseline. Finally, a recovery gap metric is calculated to quantify uneven recovery across regions and to support interpretation of the results.
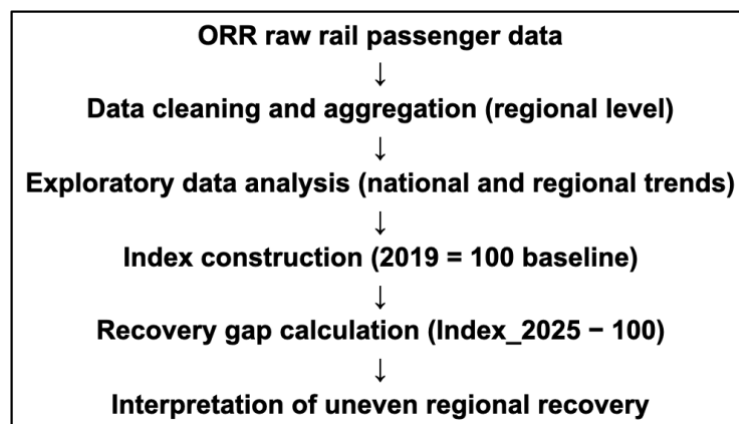

| ORR raw rail passenger data |
| ↓ |
| Data cleaning and aggregation (regional level) |
| ↓ |
| Exploratory data analysis (national and regional trends) |
| ↓ |
| Index construction (2019 = 100 baseline) |
| ↓ |
| Recovery gap calculation (Index_2025 − 100) |
| ↓ |
| Interpretation of uneven regional recovery |

Figure 1

2.2 Analysis

Before conducting the main analysis, exploratory data analysis (EDA) was used to gain an initial understanding of the dataset. This involved examining how rail passenger journeys changed over time and comparing trends across different UK regions. The EDA clearly showed a sharp decline in passenger usage during 2020, followed by a gradual recovery in subsequent years. It also revealed substantial differences in the overall scale of passenger journeys between regions.

These observations highlighted an important challenge for the analysis. Regions such as London have much higher passenger volumes than others, meaning that direct comparison of raw passenger numbers would be misleading. A region with fewer passengers could

appear to recover less simply because its baseline level is smaller. Therefore, a method was required to make recovery patterns comparable across regions.

To address this issue, passenger journeys were normalised using a pre-pandemic baseline. The year 2019 was selected as the reference year because it represents the last full year of rail usage before COVID-19 disruptions. For each region, passenger journeys in all years were expressed relative to their 2019 level, creating a recovery index where a value of 100 indicates a return to pre-pandemic usage.

This indexing approach allows recovery patterns to be compared on a common scale, regardless of regional differences in absolute passenger numbers. It provides a clearer representation of whether regions have fully recovered, partially recovered, or exceeded their pre-pandemic levels. The indexed data were then used as the basis for further analysis of uneven regional recovery, with the recovery gap defined as Index_2025 minus the 2019 baseline.

This study does not involve predictive modelling, and therefore no training or test data split was required. Instead, the analysis focuses on describing and interpreting observed patterns in historical data. All analytical steps were implemented using the R programming language and are fully documented to ensure that the results can be reproduced.

2.3 Answering the research question
The methods used in this study were chosen to directly address the research questions. National-level exploratory analysis was used to answer Research Question 1 by showing how overall rail passenger usage in the UK changed before and after the COVID-19 pandemic. This provides a general overview of the scale of the disruption and the pace of recovery at the national level.

To answer Research Question 2, regional passenger journeys were normalised using a 2019 baseline index. This allowed recovery patterns to be compared across regions on a common scale, making it possible to identify whether recovery has been uniform or uneven. By expressing post-pandemic passenger usage relative to pre-pandemic levels, the analysis highlights regions that have fully recovered, exceeded their previous levels, or continue to lag behind.

Research Question 3 was addressed by comparing regional recovery patterns with the overall national trend. This comparison helps to identify whether national-level recovery masks important regional differences. Together, these methodological steps ensure that each research question is answered in a clear and consistent way, using quantitative evidence derived from official rail passenger data.

3. Results and Discussions

3.1 Results

This section presents the results of the exploratory and quantitative analyses of post-pandemic recovery in UK rail passenger usage, focusing on national trends and regional variation.

National-level analysis shows a sharp decline in rail passenger journeys during 2020, coinciding with the onset of the COVID-19 pandemic. This represents a clear disruption to pre-pandemic trends, which had previously been stable or gradually increasing. In subsequent years, passenger numbers increased gradually, indicating a slow recovery rather than an immediate return to pre-pandemic levels.

At the regional level, recovery trajectories varied across the UK. Although all regions experienced declines during the pandemic, the pace of recovery differed. Some regions showed steady increases towards pre-pandemic passenger levels, while others continued to recover more slowly.

To enable comparison across regions, passenger journeys were indexed relative to a 2019 baseline and a recovery gap was calculated for the most recent year in the dataset. The results show that some regions have recovered to or above the 2019 baseline, while others remain substantially below it, indicating uneven recovery outcomes.

Figure 2 illustrates these recovery gaps by region, highlighting differences between areas that have recovered and those that have not.
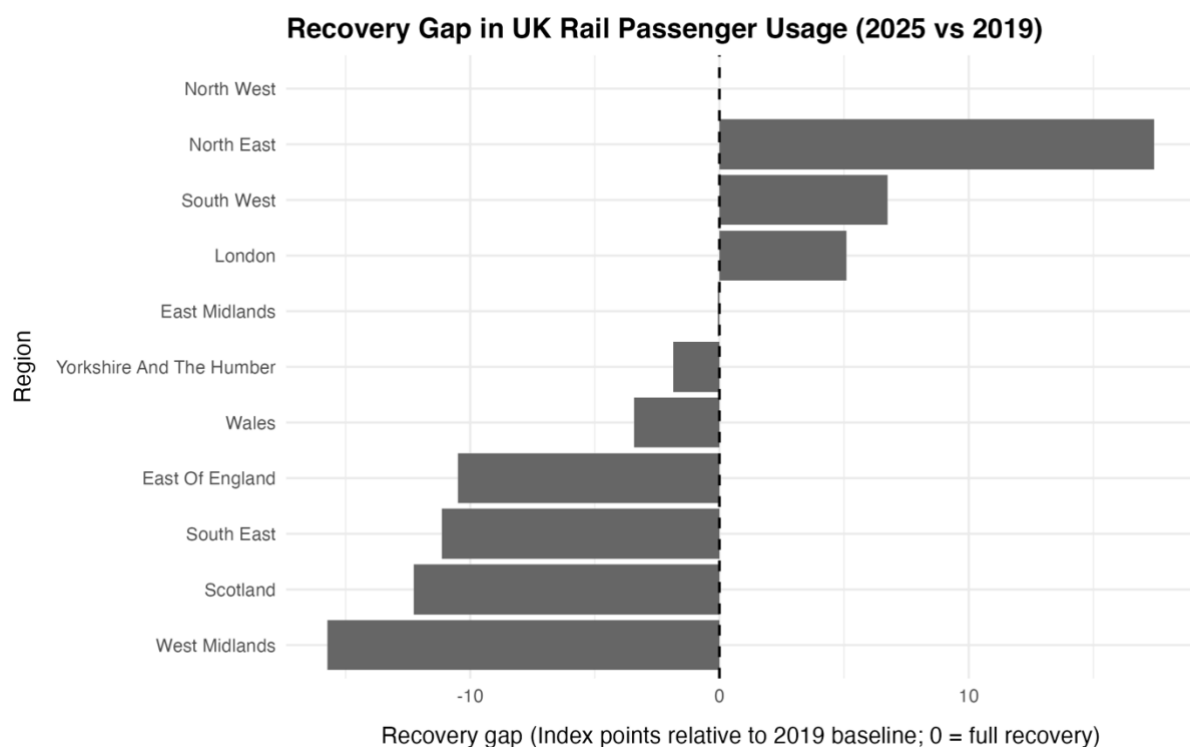

Figure 2

3.2 Discussions

The results show that post-pandemic recovery in UK rail passenger usage has not followed a uniform trajectory. Although national-level trends indicate a gradual return towards pre-pandemic levels, regional analysis reveals substantial variation in recovery outcomes, highlighting the importance of examining recovery at a regional scale rather than relying solely on aggregate national indicators.

A key implication of these findings is that national averages can obscure persistent regional disparities. Regions that have recovered more quickly or exceeded pre-pandemic passenger levels contribute disproportionately to national recovery figures, while slower recovery in other areas is less visible in aggregate statistics. As a result, national-level assessments may overestimate the extent of recovery across the country.

These uneven recovery patterns are likely influenced by structural differences between regions, including variations in economic activity, employment composition, and travel behaviour. In particular, regions with greater reliance on commuting may have been more affected by long-term changes in working practices, such as remote and hybrid working, while regions with higher levels of leisure travel may have experienced faster rebounds.

These findings have important implications for transport planning and policy. Continued reliance on national demand forecasts may fail to address local conditions, suggesting the need for more regionally sensitive approaches to post-pandemic rail planning and investment.

3.2.1 Answering the Research Questions

RQ1: How did national rail passenger usage in the UK change before and after the COVID-19 pandemic?

Answer: The analysis shows that national rail passenger usage in the UK experienced a sharp and unprecedented decline during 2020 following the onset of the COVID-19 pandemic. This decline interrupted previously stable or gradually increasing trends in passenger journeys. In the years following the easing of restrictions, national passenger numbers began to recover, but the recovery was gradual rather than immediate. Even in the most recent years of the dataset, national usage had not fully returned to a simple continuation of pre-pandemic trends, indicating that the pandemic had a lasting impact on overall rail demand.

RQ2: To what extent has the recovery of rail passenger usage been uneven across UK regions?

Answer: The results provide clear evidence that recovery has been uneven across UK regions. While some regions have recovered to, or exceeded, their 2019 pre-pandemic passenger levels, others remain below this baseline. The recovery gap analysis highlights substantial variation in recovery outcomes, demonstrating that regions have followed distinct recovery trajectories rather than a common national pattern. This confirms that regional disparities persist in post-pandemic rail usage and that recovery cannot be characterised as uniform across the country.

RQ3: How do regional recovery patterns compare with the overall national recovery trend?

Answer: Comparing regional recovery patterns with national trends reveals that national-level recovery masks important regional differences. Although aggregate national figures suggest a steady recovery, this average is influenced by stronger recovery in certain regions. At the same time, slower recovery in other regions is not fully reflected in national indicators. This comparison demonstrates that national recovery trends do not adequately capture the diversity of regional experiences and reinforces the need for regionally disaggregated analysis.

3.2.2 Relation to existing research

The findings of this study are broadly consistent with existing research on post-pandemic transport recovery, which emphasises that changes in travel behaviour are uneven across space and strongly influenced by regional economic and social factors. Previous studies have highlighted the role of remote and hybrid working in reducing commuting demand, particularly in regions with a high concentration of office-based employment. The uneven recovery observed in this analysis supports these findings, suggesting that structural differences between regions continue to shape rail demand in the post-pandemic period.

At the same time, this study extends existing research by providing a quantitative, region-level comparison using official UK rail statistics and a consistent pre-pandemic baseline. By indexing passenger journeys relative to 2019, the analysis offers a clear and interpretable measure of recovery that complements prior descriptive and qualitative studies. The results suggest that while national recovery narratives may imply a return to normality, underlying regional inequalities in rail usage persist.

In some cases, the findings also highlight potential limitations of existing national-level assessments. Studies that focus primarily on aggregate trends may underestimate the extent of uneven recovery across regions. By demonstrating how national averages can conceal slower recovery in specific areas, this study contributes to a more nuanced understanding of post-COVID transport recovery and underscores the importance of spatially disaggregated analysis in transport research and policy evaluation.
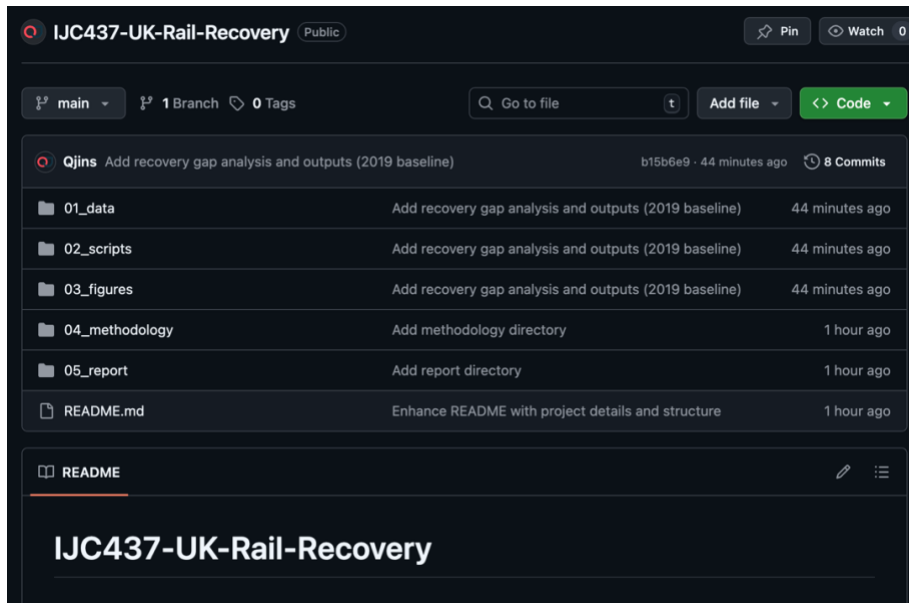
4. R Code, GitHub Pages

All code and project materials for this study are hosted on GitHub to support transparency, reproducibility, and professional presentation. The repository documents the full analytical workflow and is structured to reflect good data science practice.

The repository for this project is available at:
https://github.com/Qjins/IJC437-UK-Rail-Recovery

A related project developed for the IJC445 module, which focuses on data visualisation, is available at:
https://github.com/Qjins/IJC445-UK-Rail-Recovery-Visualisation

The repository follows a clear folder structure separating raw data, processed datasets, and analysis scripts. Raw data obtained from the Office of Rail and Road are preserved, while processed datasets are stored in standardised formats to support reproducibility. R scripts are organised by analytical task, including data cleaning, exploratory analysis, and recovery gap calculation.

The analysis was implemented in R using the tidyverse ecosystem. Clear variable naming and inline comments support readability and reuse. Instructions for reproducing the analysis, including folder structure and execution order, are provided in the repository README file.

5. Conclusions

This project examined the uneven recovery of UK rail passenger usage following the COVID-19 pandemic using official regional rail statistics. By combining exploratory data analysis with a baseline indexing approach, the study moved beyond national averages to identify regional disparities in post-pandemic recovery patterns. The analysis demonstrated that while national rail usage shows signs of recovery, this recovery has not been uniform across regions.

By focusing on regional-level data and comparing post-pandemic passenger journeys to a 2019 pre-pandemic baseline, the project provided a structured and reproducible assessment of recovery outcomes. The findings highlight the importance of spatially disaggregated analysis when evaluating large-scale disruptions and their long-term impacts on transport systems.

5.1 Reflections

5.1.1 Key findings

- Uneven regional recovery was observed across the UK, with some regions returning to or exceeding pre-pandemic passenger levels while others remained below the 2019 baseline (Section 3.1).
- National recovery trends masked important regional differences, demonstrating that aggregate indicators alone are insufficient for understanding post-pandemic rail demand (Section 3.2).
- Baseline indexing provided a clear and interpretable measure of recovery, enabling meaningful comparison across regions with different absolute passenger volumes (Sections 2.2 and 3.1).

5.1.2 Limitations, Assumptions and Weaknesses
Limitations
1. The analysis relies on regional-level aggregation, which may obscure variation within regions and prevent more granular insights at the station or route level.
2. The study uses annual data, limiting the ability to capture short-term fluctuations or seasonal recovery patterns.
3. The dataset does not include explanatory variables such as employment type or service frequency, restricting causal interpretation of recovery differences.

Assumptions
1. The year 2019 is assumed to represent a stable pre-pandemic baseline, which is reasonable given its widespread use in transport recovery studies.
2. Changes in passenger journeys are assumed to reflect underlying travel demand, rather than solely operational or reporting changes.
3. Regional boundaries are assumed to provide a meaningful spatial unit for comparison, despite internal heterogeneity.

Weaknesses
1. The analysis is descriptive rather than predictive, limiting its ability to forecast future recovery trends.
2. Potential confounding factors influencing recovery are discussed but not quantitatively modelled.
3. The study focuses exclusively on rail, without comparison to other transport modes.

5.1.3 Future Work
1. Extending the analysis to station-level or route-level data to capture finer spatial variation in recovery.
2. Integrating socio-economic and labour market data to better explain regional differences in recovery trajectories.
3. Applying longitudinal or predictive models to assess the long-term implications of post-pandemic travel behaviour changes.

Overall, this project strengthened understanding of how data science methods can be applied to real-world transport datasets to reveal uneven recovery patterns. It also highlighted the importance of methodological transparency and spatial awareness when interpreting post-pandemic trends.

5.2 Engagement

Engagement with employability sessions, alumni talks, and industry-focused events provided valuable context for interpreting the findings of this study. Discussions with industry professionals highlighted the importance of data-driven decision-making in transport planning following major disruptions such as COVID-19.

A key insight from these sessions was the emphasis placed by industry on regional and local-level evidence, rather than relying solely on national indicators. This reinforced the relevance of the project's focus on uneven recovery and the need for spatially disaggregated analysis. Discussions around flexible working aligned with the observed regional differences in rail recovery.

These engagement activities helped situate the technical analysis within a broader professional and societal context, reinforcing the role of data science not only as an analytical tool, but also as a means of informing equitable and effective policy decisions.

References

Office of Rail and Road. (2025). *Passenger rail usage statistics: Annual and quarterly data for rail passenger journeys in Great Britain.*
https://dataportal.orr.gov.uk/statistics/usage/passenger-rail-usage/

Office of Rail and Road. (2025). *Regional rail usage: passenger journeys by region in Great Britain.*
https://dataportal.orr.gov.uk/statistics/usage/regional-rail-usage/

Department for Transport. (2025). *Domestic transport usage by mode: rail, road, bus and tube (including COVID-19 recovery context).*
https://www.gov.uk/government/statistics/transport-use-during-the-coronavirus-covid-19-pandemic

Beck, M. J., & Hensher, D. A. (2020). Insights into the impact of COVID-19 on household travel and activities in Australia: The early days of easing restrictions. *Transport Policy, 99*, 95–119. https://doi.org/10.1016/j.tranpol.2020.08.004

Gkiotsalitis, K., & Cats, O. (2021). Public transport planning adaptation under the COVID-19 pandemic crisis: Literature review of research needs and directions. *Transport Reviews, 41*(3), 374–392. https://doi.org/10.1080/01441647.2020.1857886

Zhang, J., Hayashi, Y., & Frank, L. D. (2021). COVID-19 and transport: findings from a world-wide expert survey. *Transport Policy, 103*, 68–85. https://doi.org/10.1016/j.tranpol.2021.01.011