Analysing Train Operating Company Data in the UK Rail Industry

Introduction

The UK railway industry works on a system where it is operated by a number of organisations. Passenger and freight services are operated by Train Operating Companies (TOC's), while the infrastructure is run by Network Rail.

In the past few years the state of the UK rail industry has been affected by COVID-19, and the Government policies implemented during this time. COVID-19 has changed the way the public view rail, as well as the way that TOCs operate.

This report aims to answer the following questions:

1. What are considered the key parameters for a TOC?
2. Has key metrics reached the same levels pre and post-COVID-19?
3. How have passengers reacted to rail usage, compared to reaction to TOCs operations?
4. Has COVID-19 changed the status of the biggest and smallest TOCs?
5. What are the immediate trends for the future of the UK rail industry?

The scope of this project looks at data before and after COVID-19, encompassing the years 2017-2023. The project will look at the trends of all of the TOCs combined while providing data visualisation to the trends for each TOC.

Data source

Data is sourced from the Office of Rail and Road (ORR) data portal https://dataportal.orr.gov.uk/. The data used is part of a suite of publications that ORR publishes as an official statistic. This means that it is produced by an organisation named by the Statistics and Registration Service Act 2007. As a result, the data used meet the highest standards of trustworthiness, quality and public value.

The TOC key statistics compendium provided by the ORR data portal is released annually, and archives the past five years worth of data. Two of these compendium data tables were sourced, giving data that span from 2017 to 2023. The latest data table published was on 3 August 2023.

Process

The project used datasets obtained as .ods files, used Colaboratory powered by Python for its data analysis, and a dashboard was created using Python and the Dash library.

The project process was grouped into 4 major tasks:

1. Cleaning the dataset: this involved checking for misspellings, inconsistent capitalizations and typos, checking for duplicate entries and blank cells and checking for consistent data format across each column.
2. Aligning the dataset with the objectives set: the dataset was aligned to be consistent with the project objectives, and involved manipulating the data (creating new variables from existing variables).
3. Data analysis: performing initial investigations on data so as to discover patterns, anomalies, and to answer the questions proposed in the project.
4. Data visualisation: this involved presenting the data analysis in a visual format.

Data Analysis

Key parameters

From the data, the key parameters recorded were:

- Number of full time equivalent (FTE) employees
- Number of station managed
- Route kilometres operated
- Number of passenger journeys
- Passenger kilometres
- Passenger train kilometres
- Delays on the rail network caused by Network Rail
- Delays on the rail network caused by other TOCs
- Delays on the rail network caused by own TOC
- Percentage of on time trains
- Cancellations score
- Number of trains planned
- Number of complaints closed
- Number of delay compensation claims closed
- Number of passenger assists

The data included the following TOCs:

- Avanti West Coast,
- c2c,
- Caledonian Sleeper,
- Chiltern Railways,
- CrossCountry,
- East Midlands Railway,
- Elizabeth line (formally TfL Rail)

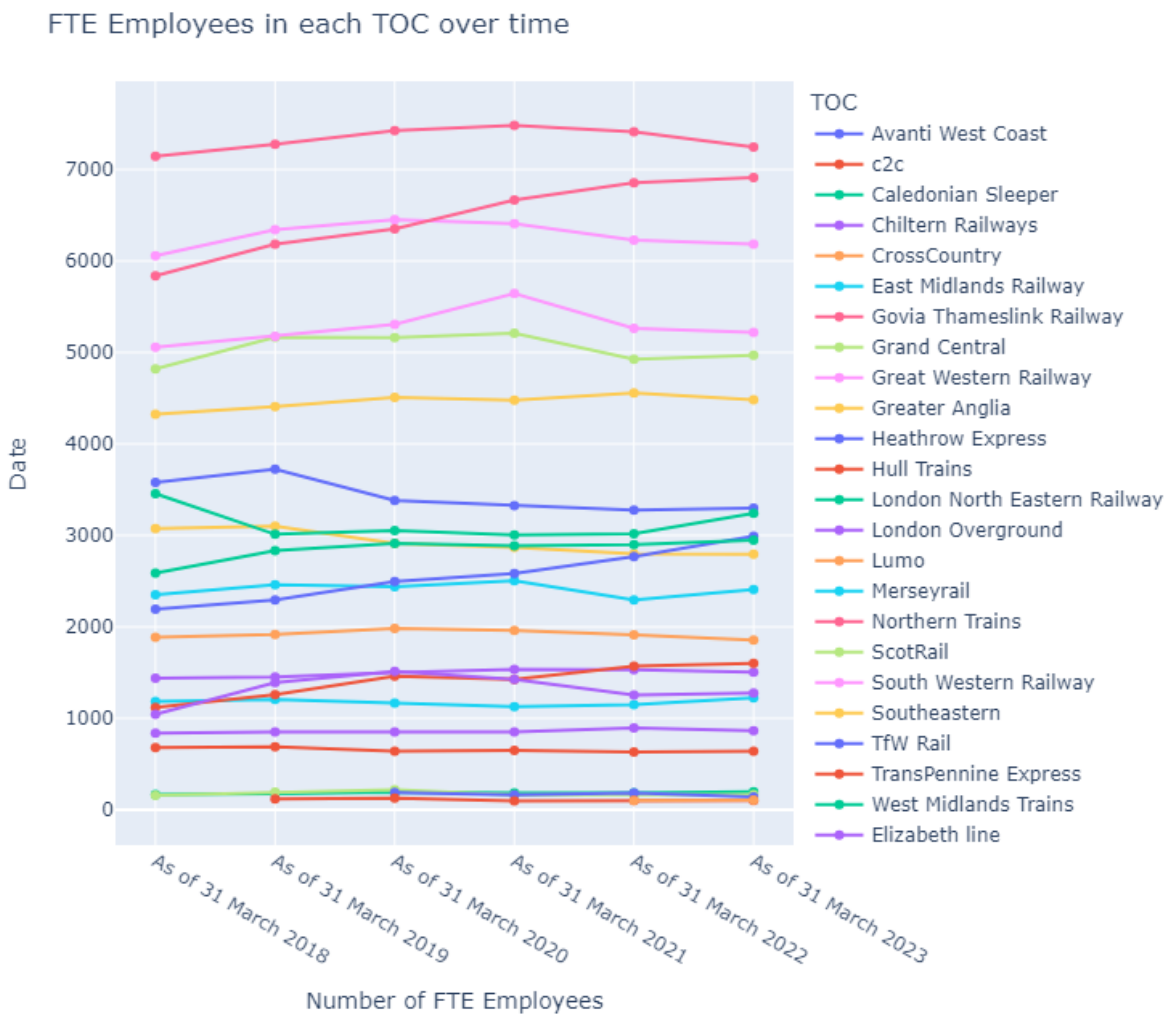
- Govia Thameslink Railway,
- Grand Central,
- Great Western Railway,
- Greater Anglia,
- Heathrow Express,
- Hull Trains,
- London North Eastern Railway,
- London Overground,
- Lumo,
- Merseyrail,
- Northern Trains,
- ScotRail,
- South Western Railway,
- Southeastern,
- TfW Rail,
- TransPennine Express,
- West Midlands Trains,

Number of full time equivalent (FTE) employees

The number of full-time equivalent (FTE) employees is calculated by comparing an employee's average number of hours worked to the average hours of a full-time worker.

Country wide, with all of the TOC's combined, there has been a general upwards trend of FTE employees. However, between March 2021 and March 2022 there was a decline. This decline has shown recovery in the year afterwards.

The first COVID-19 lockdown in the UK occurred in March 2020, and the rate of employee growth has slowed down since..

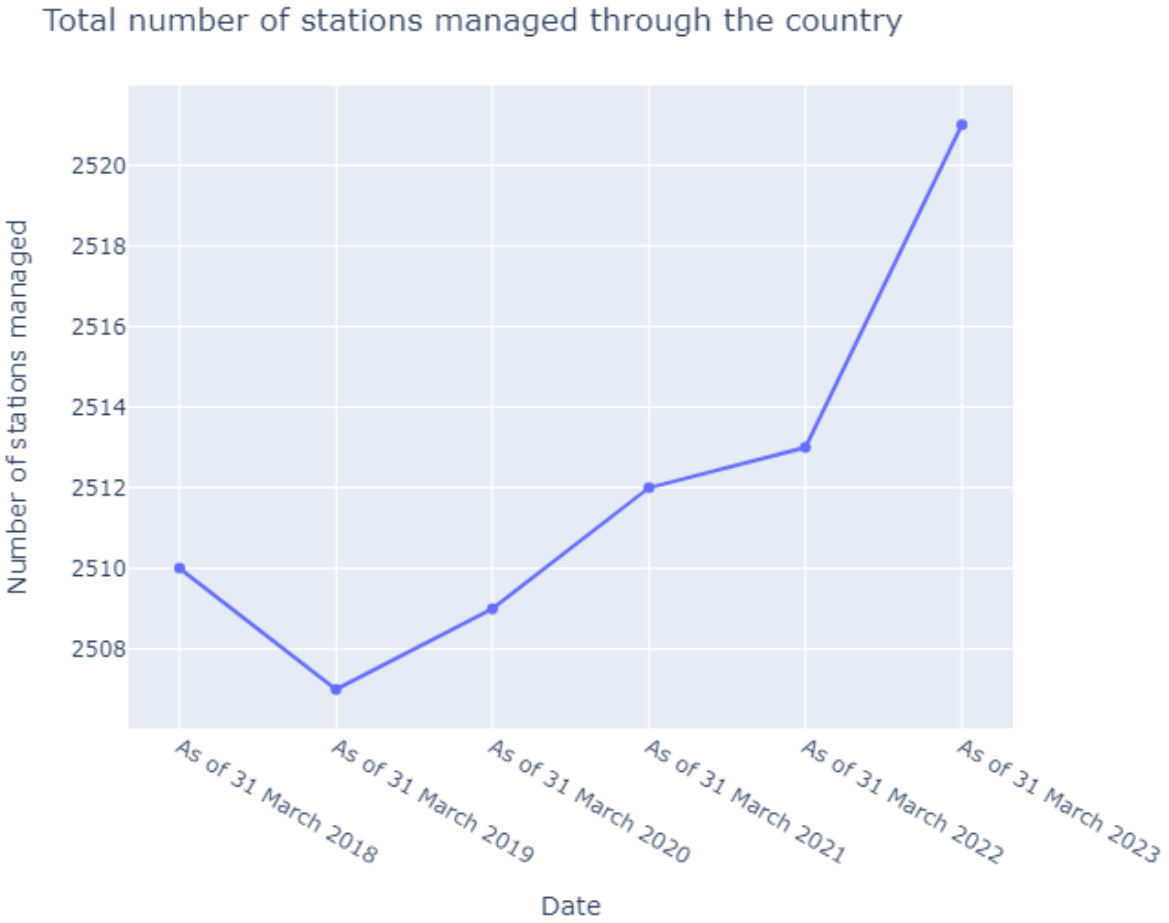


Govia Thameslink Railway consistently had the greatest number of FTE employees. However, since 31 March 2021, they have been reducing their numbers. Northern Trains have been steadily increasing their employee count, showing a possibility that they may have the greatest number of FTE employees in the future.

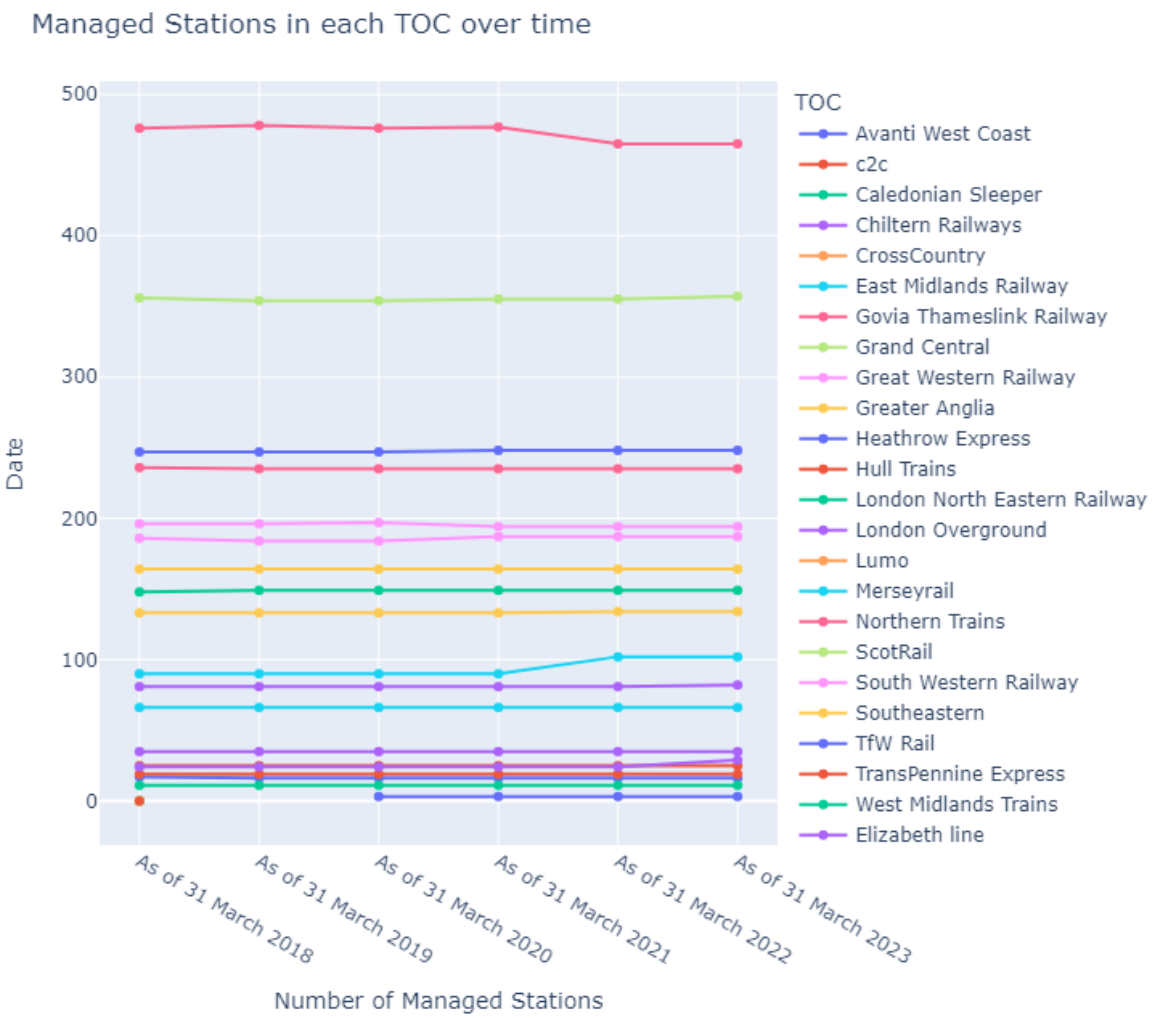
The lowest number of FTE employees was recorded by Hull Trains, who as of 31 March 2021, had 98 FTE employees. This contrasts with Govia Thameslink Railway who had 7481.27 in the same year, which was the highest value observed from the data.

Number of stations managed

Number of stations managed are stations that are called at by a mainline train service as of 31 March of the recorded annual time period. Permanently closed stations and stations where mainline services have ceased indefinitely are not included in the data.



The number of stations managed through the country has generally trended upward, with COVID-19 showing little impact to this metric.



There are multiple TOC's that do not manage any stations:

1. Caledonian Sleeper
2. CrossCountry
3. Grand Central
4. Hull Trains

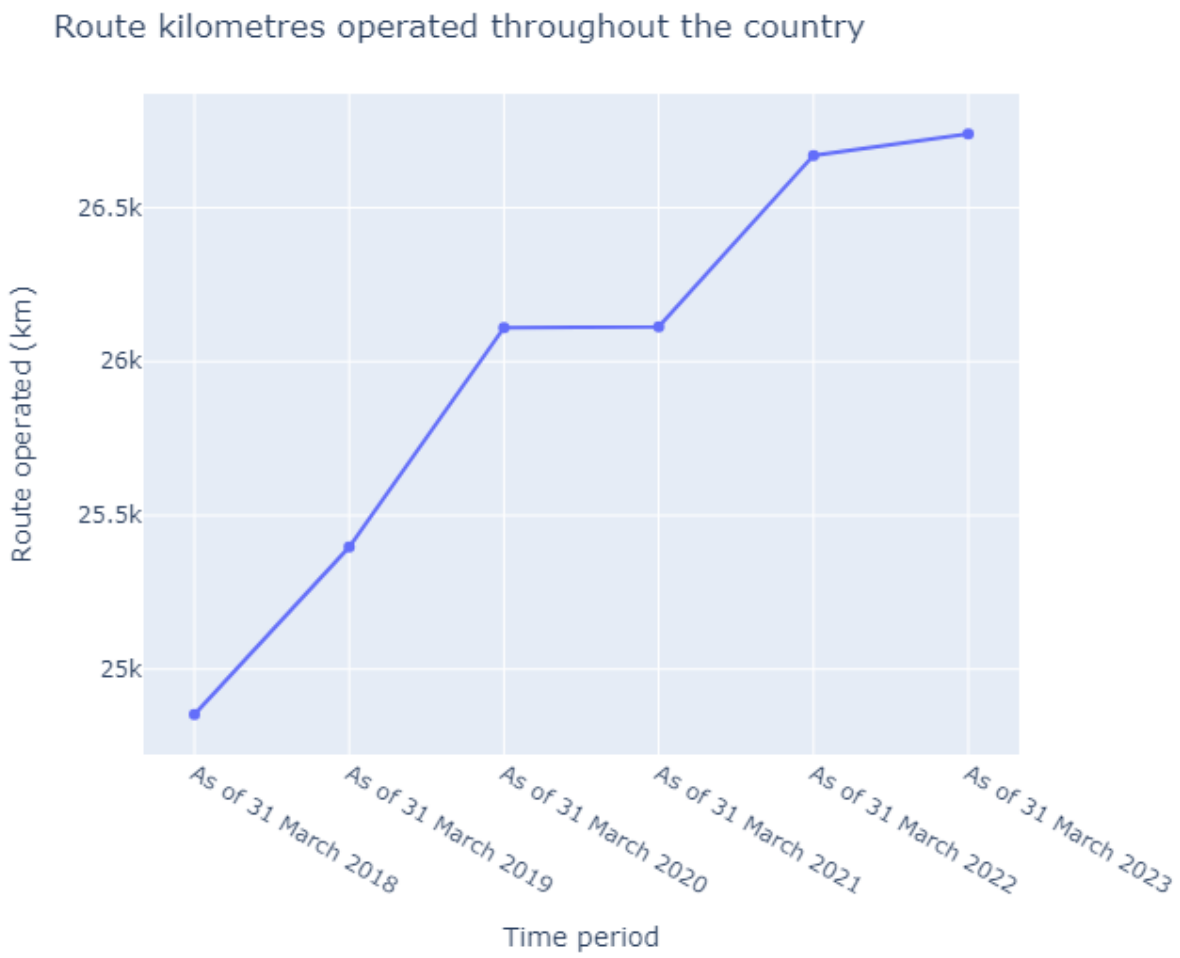
Northern Trains consistently had the greatest number of stations managed, with a peak of 478 stations as of March 2019.

Employees per Station

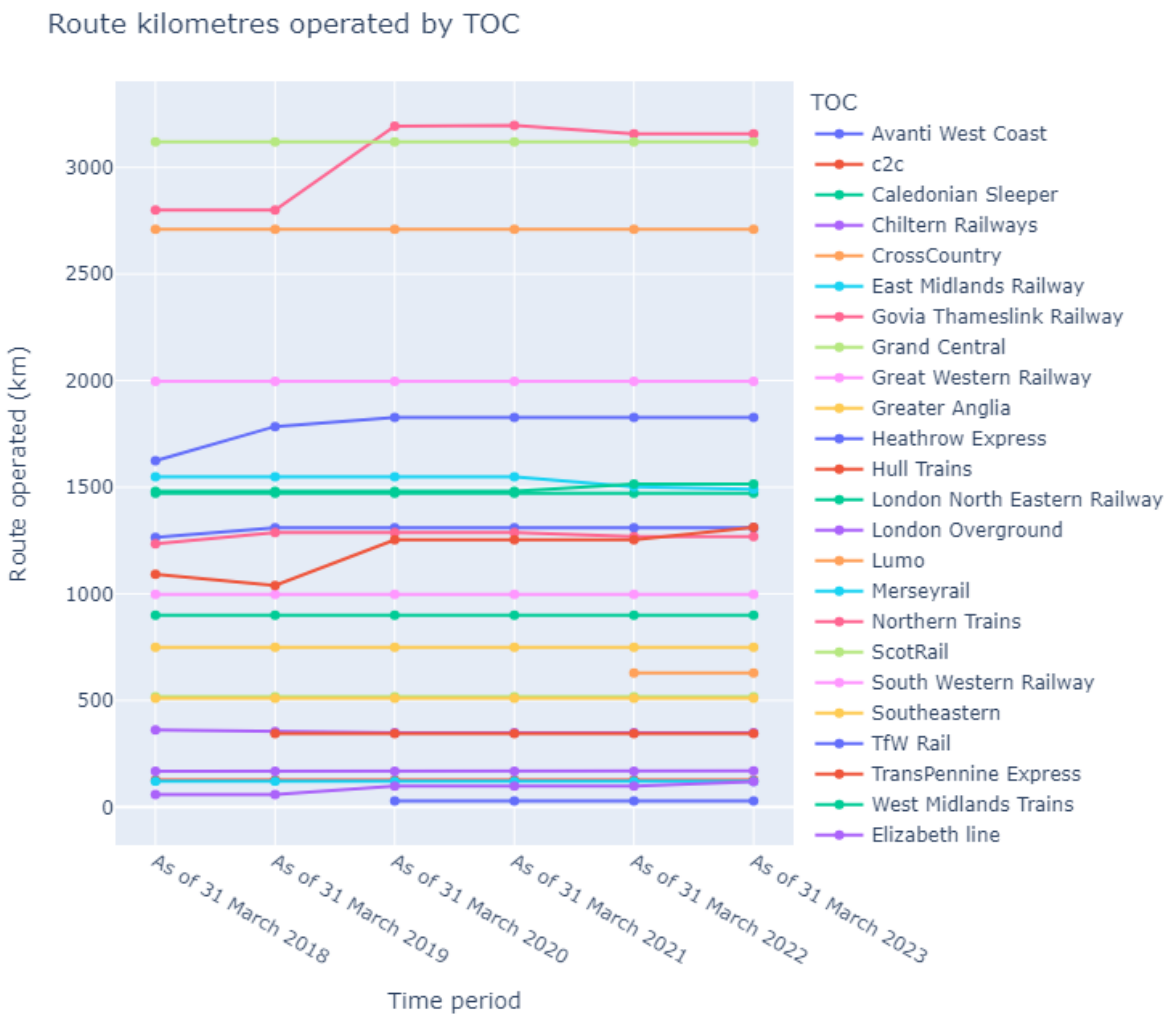
As a ratio of FTE employees to managed stations, TfW Rail had the lowest as of 31 March 2018, with 8.9 FTE employees per station. In the same time period, London North Eastern Railway had the highest, with 314.2 FTE employees per station.

Route kilometres operated

Route kilometres operated includes the total extent of route available to operate on as of 31 March. It does not take into account multiple track routes (i.e. double tracks are only counted as one route kilometre but would be two track kilometres).



Expansion of route operation was paused between March 2020 and March 2021. Route operation growth has begun again, but does not match the growth rate observed before COVID-19.

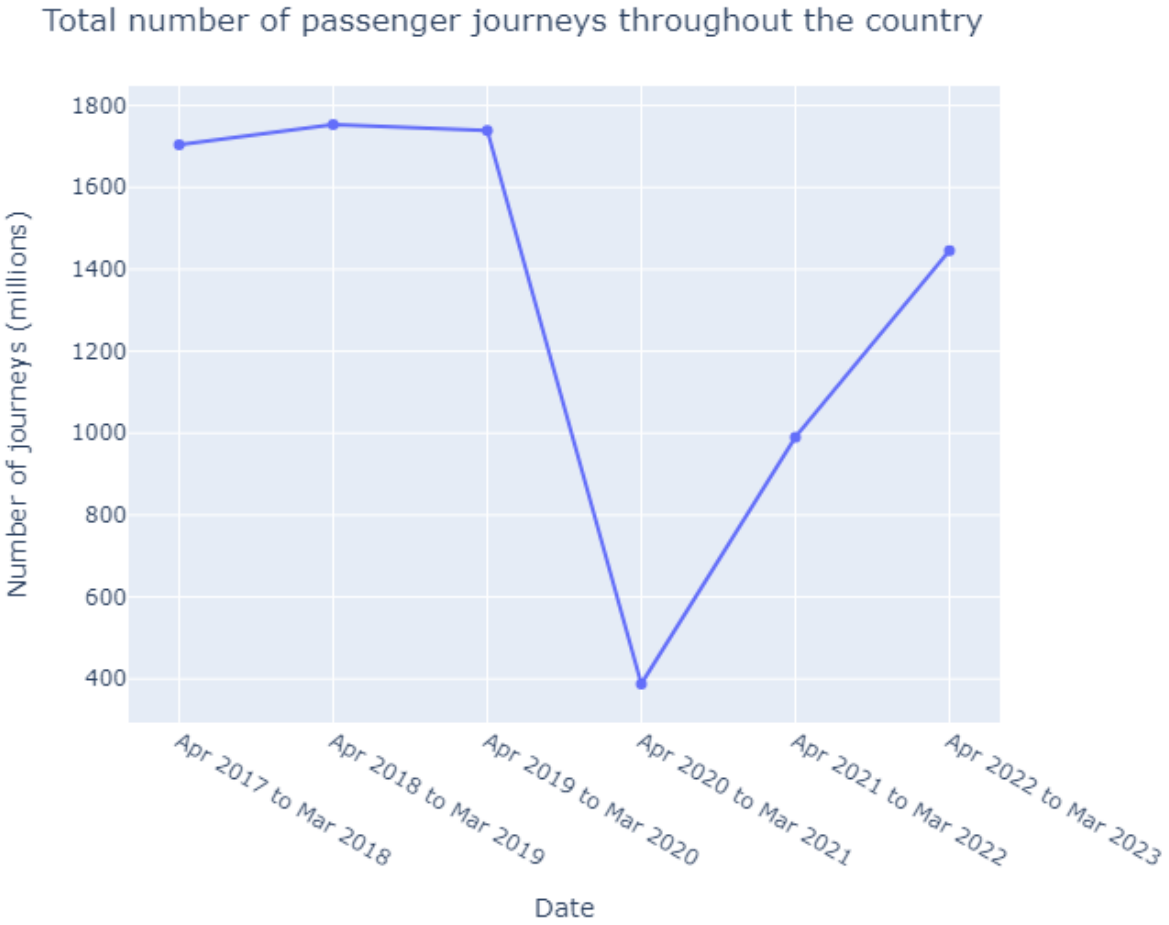


Northern Trains have become the TOC with the most train route km in operation, overtaking previous leaders ScotRail by March 2020. The peak was observed on 31 March 2021, where the TOC operated on approximately 3197 km.

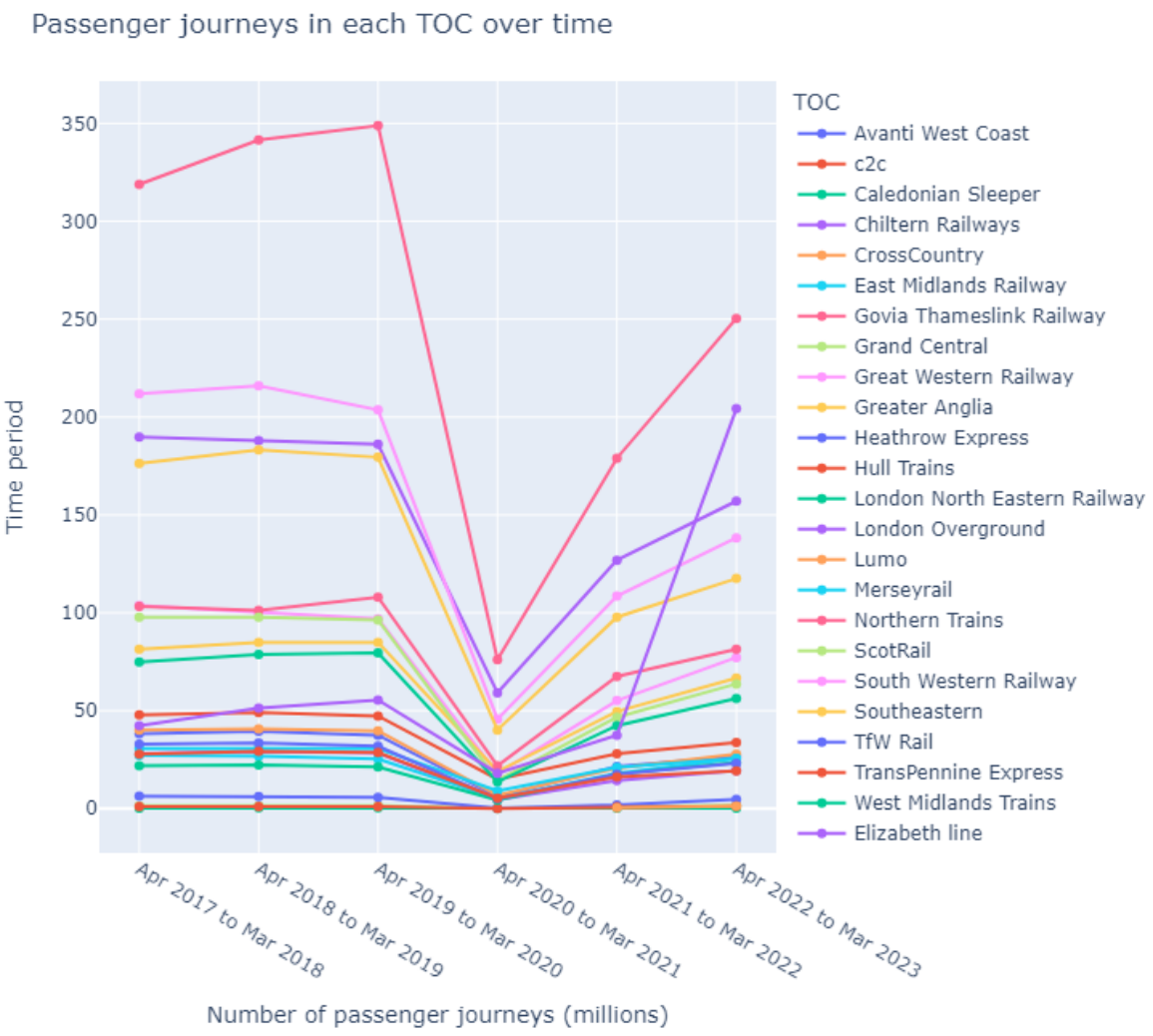
The TOC with the least amount of routes kilometres operated was Heathrow Express with a low of 29 km as of 31 March 2023.

Number of passenger journeys

Passenger journeys are estimated based on travel from an origin station to a destination station, and are recorded in millions. Where travel includes one or more changes of train, each train used is counted as one journey.



A massive decline was observed in the April 2020 to March 2021 time period, reflecting the lockdown policies implemented from March 2020 onwards. In the subsequent years, this value has recovered, but not to pre-COVID-19 levels.



All TOC's have seen a decrease in passenger journeys as a result of COVID-19. Furthermore, all have begun their recovery, but not to the levels before COVID-19. The exception to this is with the Elizabeth Line, which was formerly known as TfL Rail. The popular addition of the Elizabeth Line to the TfL network has driven growth between 2022 and 2023.

The greatest number of passenger journeys in the dataset was by Govia Thameslink between April 2019 and March 2020, with 348.8 million journeys. The lowest number was observed by Hull Trains between April 2020 to March 2021 with 0.077 million journeys.

Who are the biggest and smallest TOC's?

There are a number of possible ways to define the "Biggest" TOC's.

- Full-time equivalent (FTE) employees
- Managed Stations
- Passenger journeys
- Route kilometers operated

From the data from the most recent year, Northern Trains are in the top 5 for all four of the above metrics. Govia Thameslink Railway, ScotRail and Great Western Railway appear in the top 5 in three of the four metrics. These four TOC's can therefore be considered the biggest TOC's.

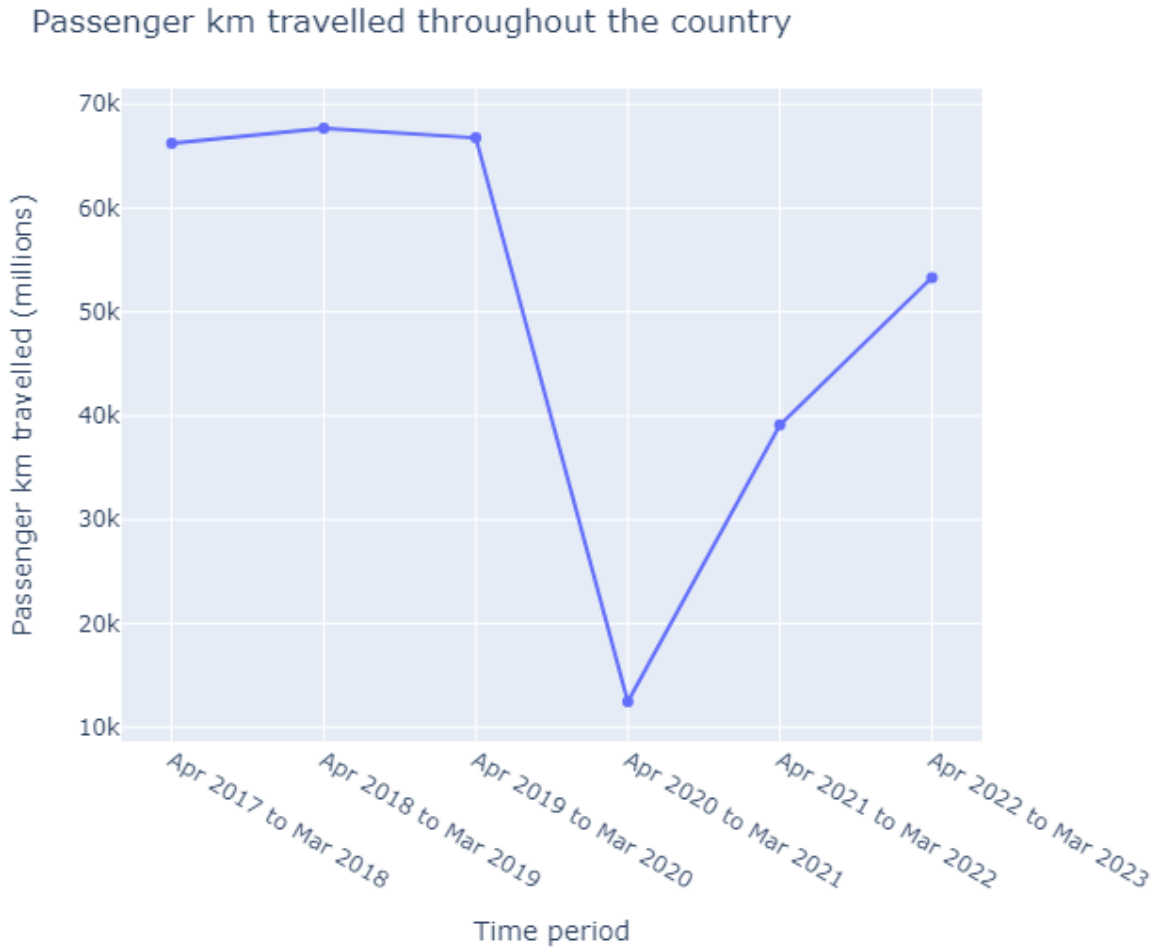
Hull Trains, Lumo, Heathrow Express, Grand Central, Caledonian Sleeper are TOC's that are present in the bottom 5 for three of the four metrics, and can therefore be considered the smallest TOC's.

From the data from 2019, before COVID-19, Northern Trains is in the top 5 for all four metrics. Govia Thameslink Railway, ScotRail and Great Western Railway appear in the top 5 for three metrics. These TOC's can be considered the biggest TOC's before and after COVID-19.

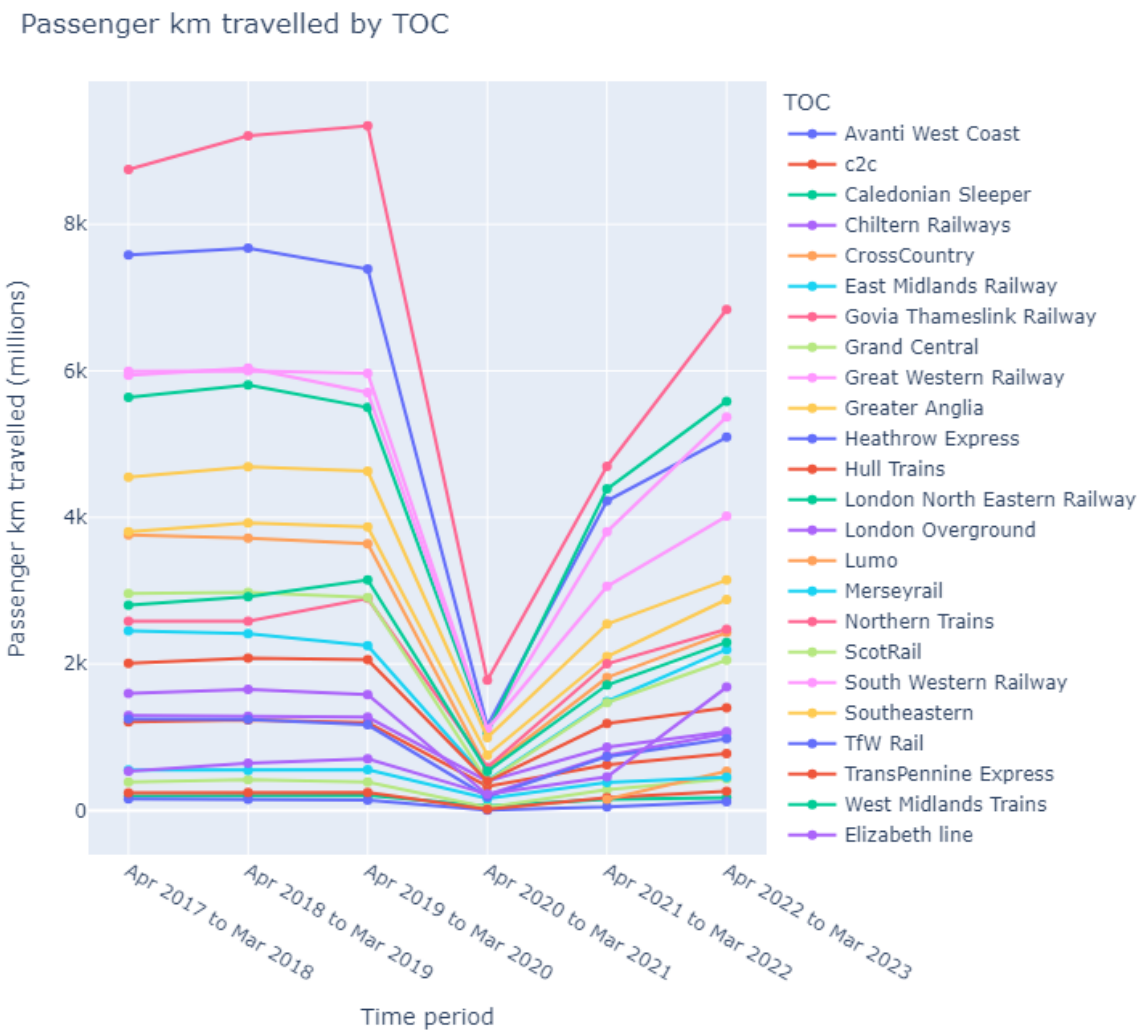
Hull Trains were in the bottom 5 in all four metrics. Caledonian Sleeper and Grand Central were in the bottom 5 in three metrics. Hull Trains, Caledonian Sleeper and Grand Central could be considered the smallest TOC's before and after COVID. Lumo, who did not exist in 2019, as well as Heathrow Express, who had complete data for 2019 were ignored when investigating the smallest TOC's. There is not enough data to confirm that Heathrow Express would be a part of this group.

Passenger kilometres

Passenger kilometres are calculated by multiplying the number of passenger journeys on a particular flow by the number of corresponding track kilometres between stations.



Similarly to passenger journeys, a massive drop in passenger kilometres travelled can be seen in the April 2020 to March 2021 time period. In the subsequent years, this value has recovered, but not to pre-COVID-19 levels.



All TOCs had experienced a drop in passenger kilometres travelled.

Govia Thameslink had the greatest value of passenger kilometres travelled in Apr Apr 2019 to Mar 2020 with 9.3 billion passenger kilometres.

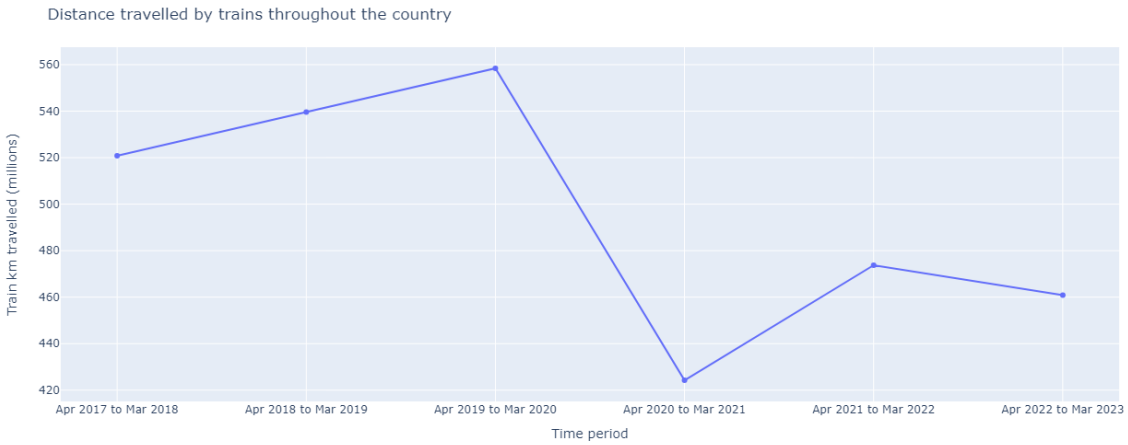
Heathrow Express had the lowest in Apr 2020 to Mar 2021 with 77 thousand passenger kilometres

Distance per journey

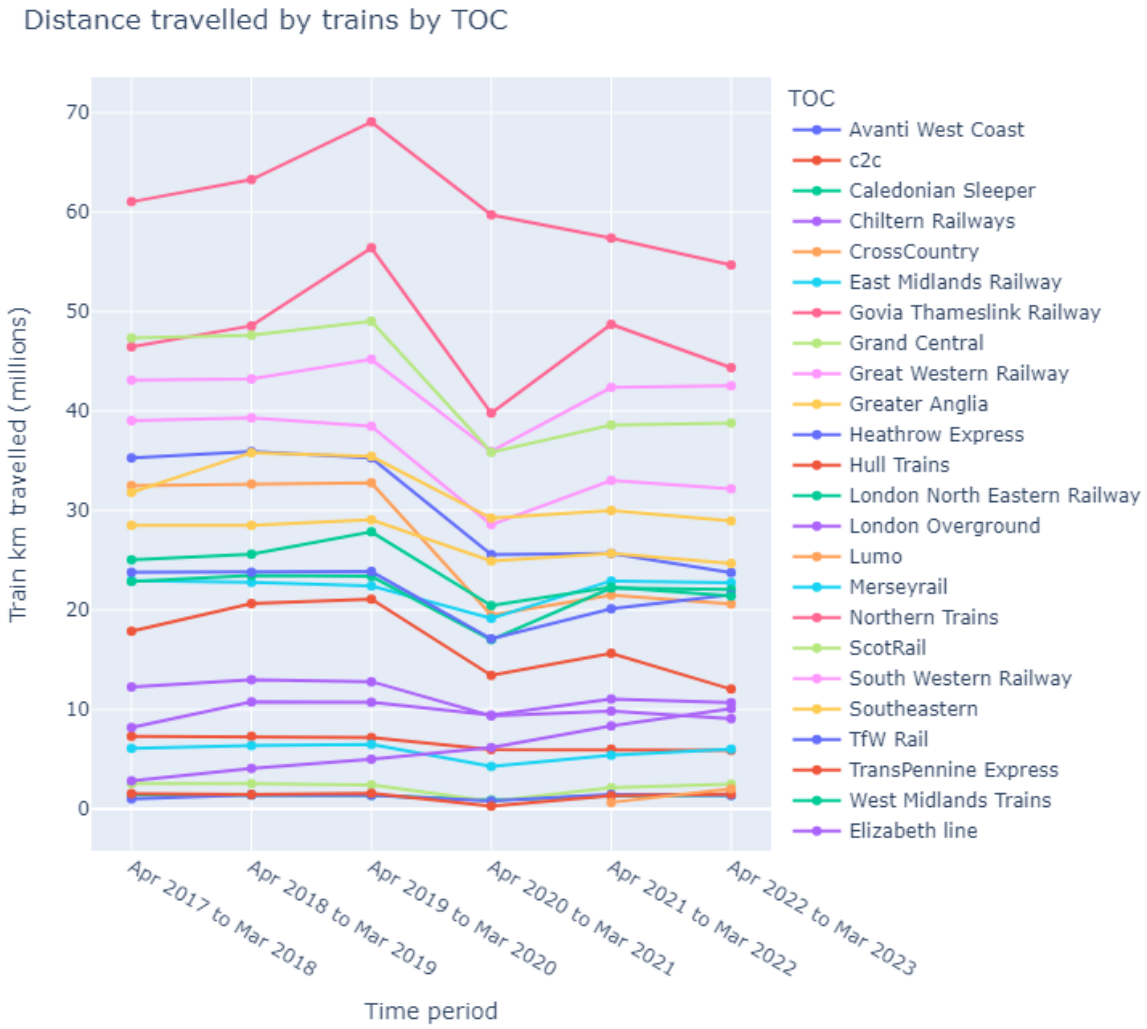
On average, London Overground passengers had the shortest journeys, averaging at 6.8 km per journey in 2020-2021. The longest journeys were experienced by Caledonian Sleeper passengers with 742 km per journey in 2020-2021

Passenger train kilometres

Passenger train kilometres refers to the number of train kilometres travelled by revenue earning passenger trains.



Similarly to passenger kilometres, a significant drop in passenger train kilometres travelled can be seen in the April 2020 to March 2021 time period. In the subsequent years, this value has struggled to recover, with a drop seen in the most recent annual data..



The majority of TOC's follow the same pattern as the country aggregate data, with the lowest value commonly seen in the April 2020 to March 2021 period, and the peak value observed in the year before.

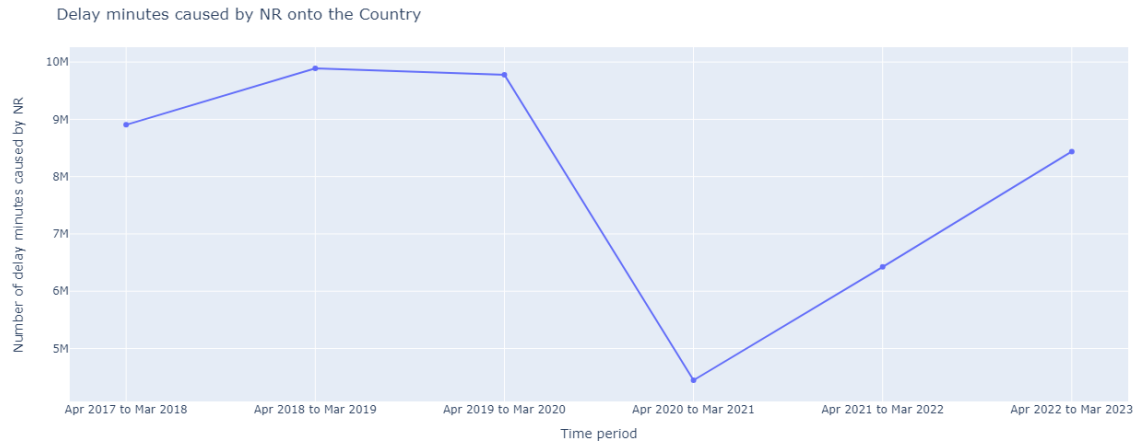
Govia Thameslink trains travelled the most, as they had the greatest number passenger train kilometres between Apr 2019 to Mar 2020 with 69 million train kilometres.

Heathrow Express had the lowest between Apr 2020 to Mar 2021 with less than 300 thousand train kilometres.

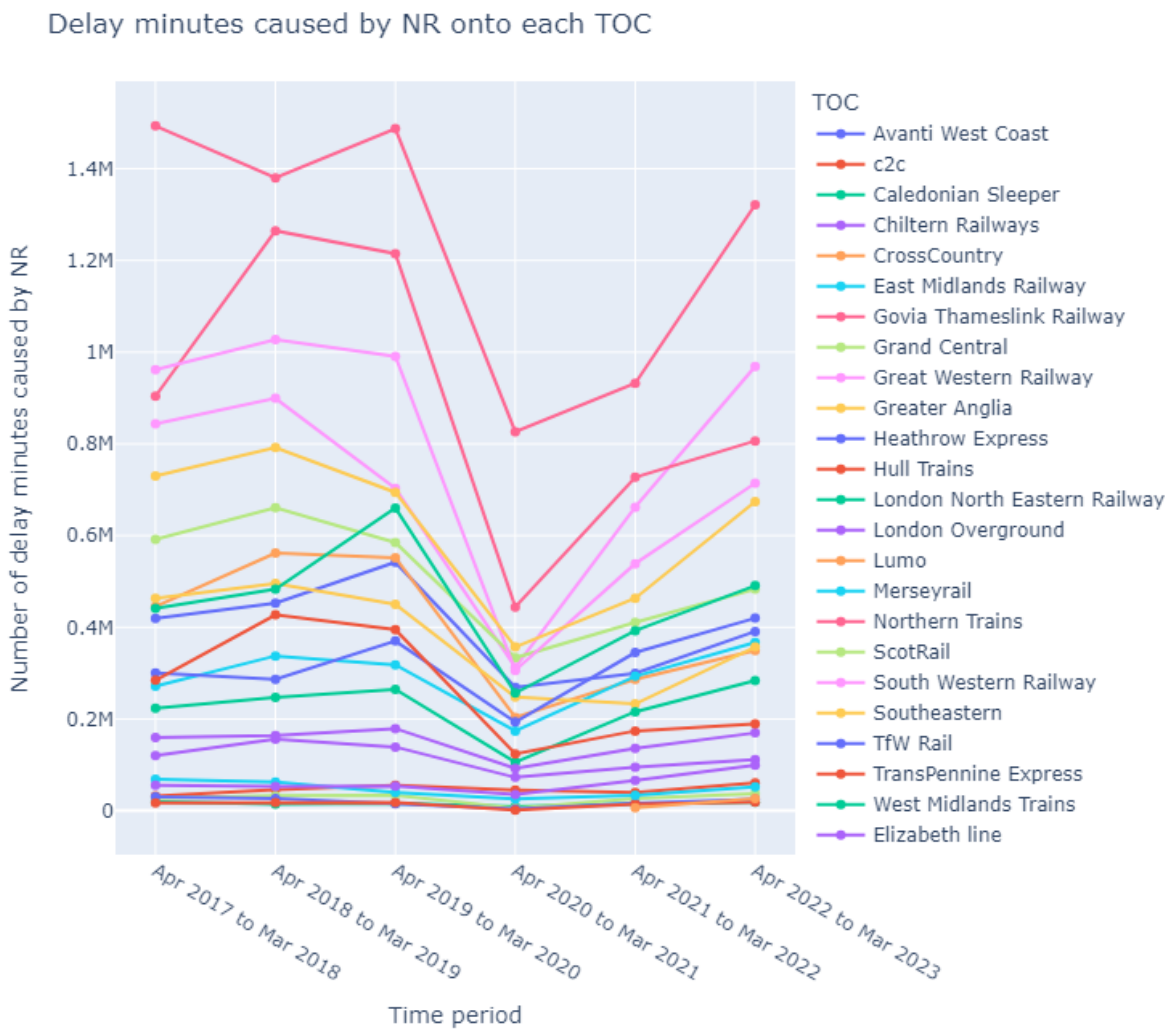
Delays on the rail network caused by Network Rail

Delay minutes are defined as the time lost between consecutive timing points on the rail network. Delay incidents producing three or more minutes of delay on Britain's railways are attributed to either Network Rail or a train operator.

Delays caused by Network Rail are delays attributed to Network Rail affecting train operating companies, such as Track, Network management, etc.



Delays caused by Network Rail fell in value during the time period between April 2020 and March 2021. However, this value has nearly reached pre-COVID-19 values.

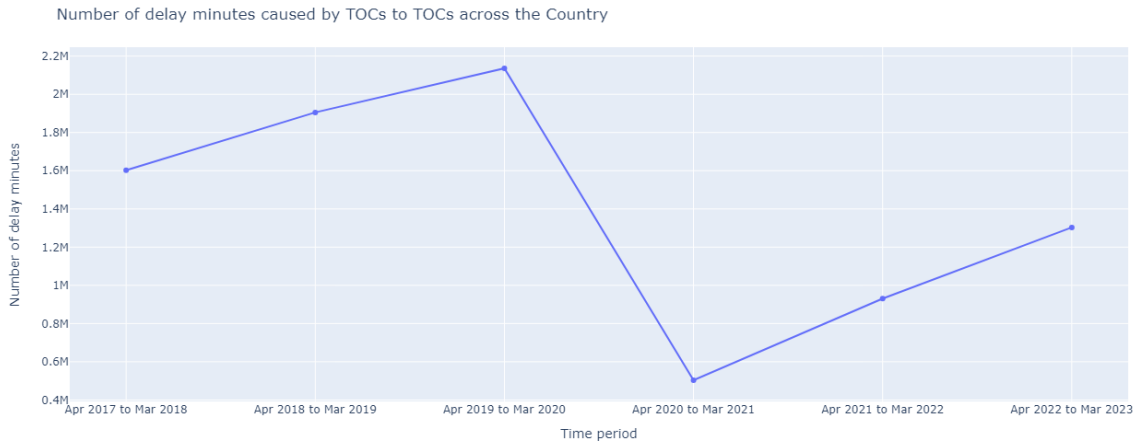


The individual TOC's have experienced a similar pattern to one another. Some, such as London North Eastern Railway and TfW Rail, are now experiencing a greater number of delay minutes compared to before COVID-19.

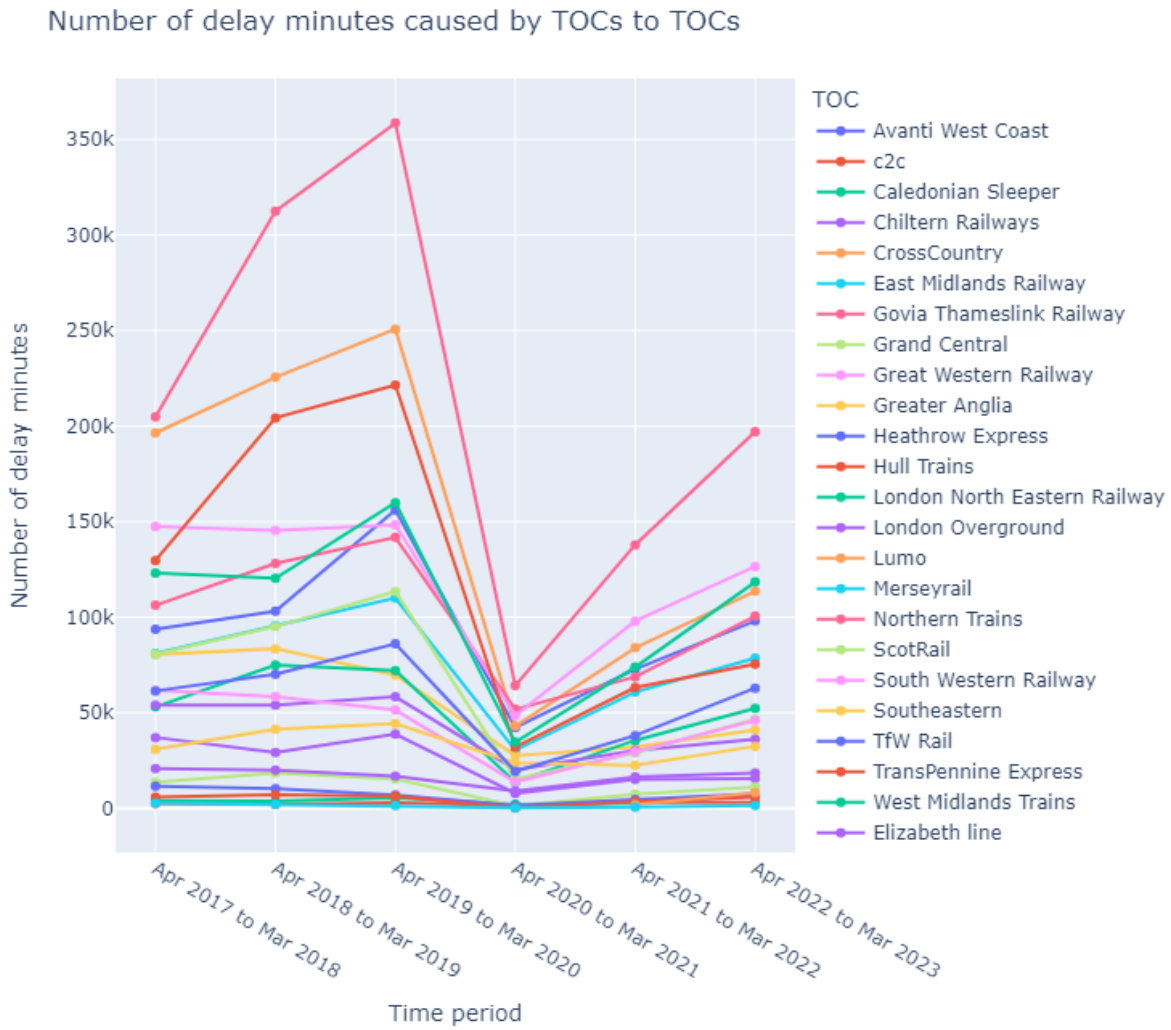
The highest instance of delay minutes in the dataset belonged to Govia Thames Railway, who had 1.5 million minutes of delays caused by Network Rail. This was observed in April 2017 to March 2018. This is in contrast with Hull Trains with 1647.5 minutes in the April 2020 to March 2021 time period.

Delays on the rail network caused by other TOCs

This type of delay are delays attributed to train operating companies affecting other train operating companies, such as via another operator's fleet, train crew



Similar to the delays caused by Network Rail, delay minutes caused by other TOC's was at their lowest between April 2020 and March 2021. It has started to increase post-COVID-19, but not at the same rate as delays caused by Network Rail.

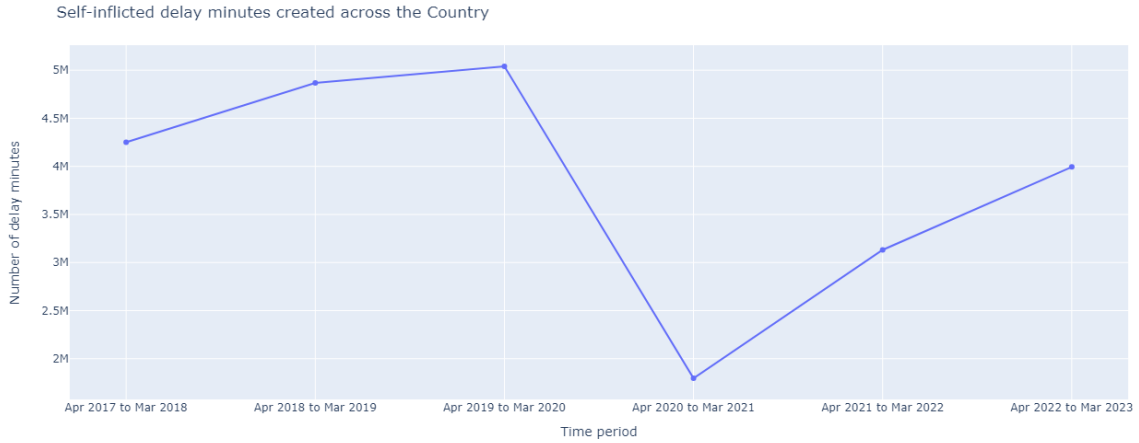


All of the TOC's experienced a fall in delay minutes during the time period between April 2020 to March 2021. The TOC who had experienced the most minutes of delays caused by other TOCs in a year was Northern Trains, with 358,614 minutes of delays,

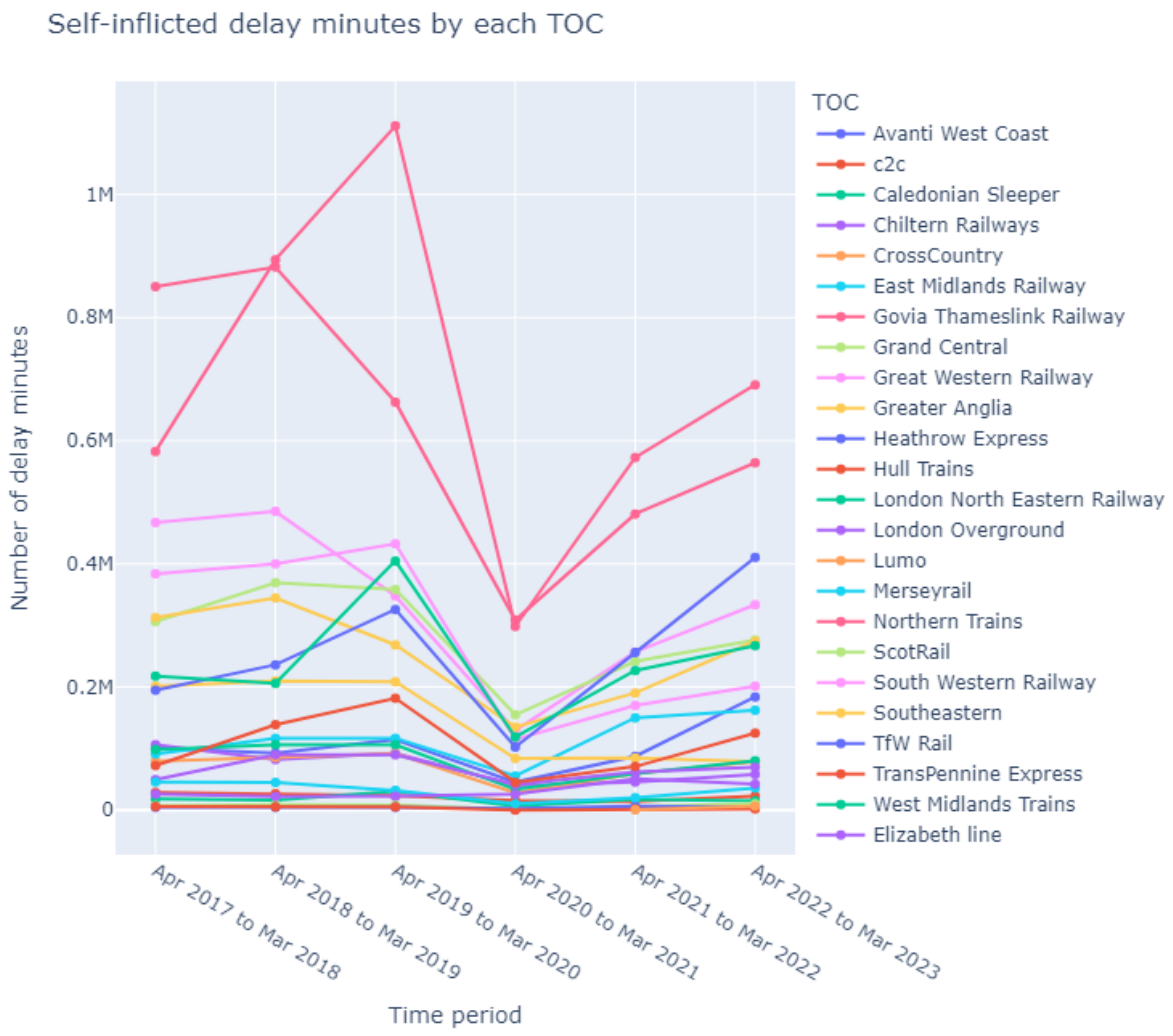
occurring between April 2019 to March 2020. The company who had the least was Hull Trains with 236.5 minutes during April 2020 to March 2021.

Delays on the rail network caused by own TOC

These delays are delays attributed to train operating companies affecting their own train operating company, for example through their own fleet, train crew, etc.



The trend is similar to the trends related to the other types of delays.

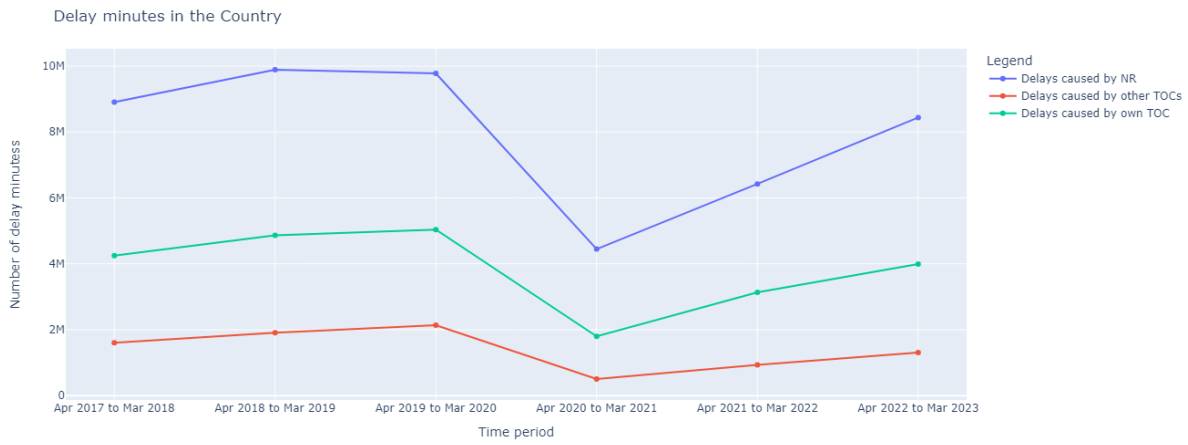


Some TOC's, such as TfW Rail and East Midlands Railway, are now experiencing a greater number of delay minutes compared to before COVID-19.

The company who had caused the most minutes of delays to themselves in a year was Northern Trains, with 1.1 million minutes of delays, occurring between April 2019 to March 2020.

The company who had the least was Hull Trains with 192 minutes during April 2020 to March 2021.

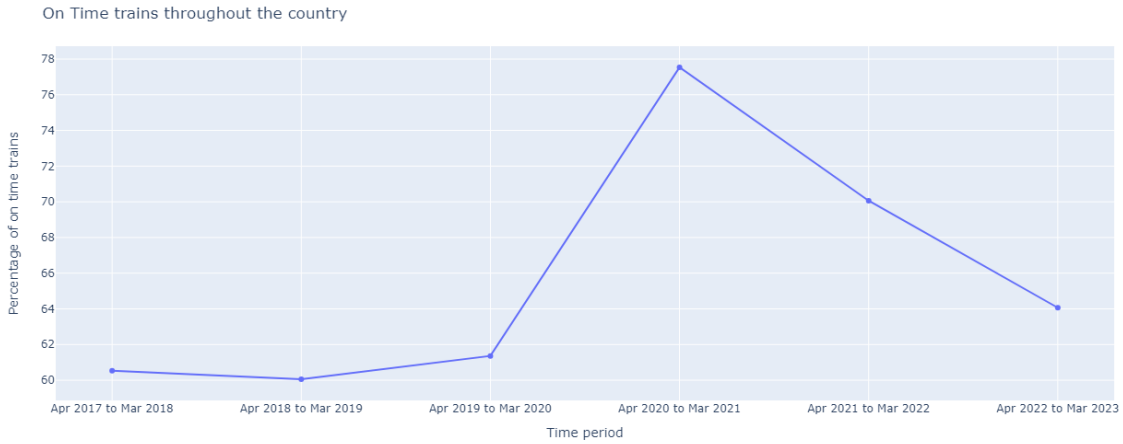
Types of Delays



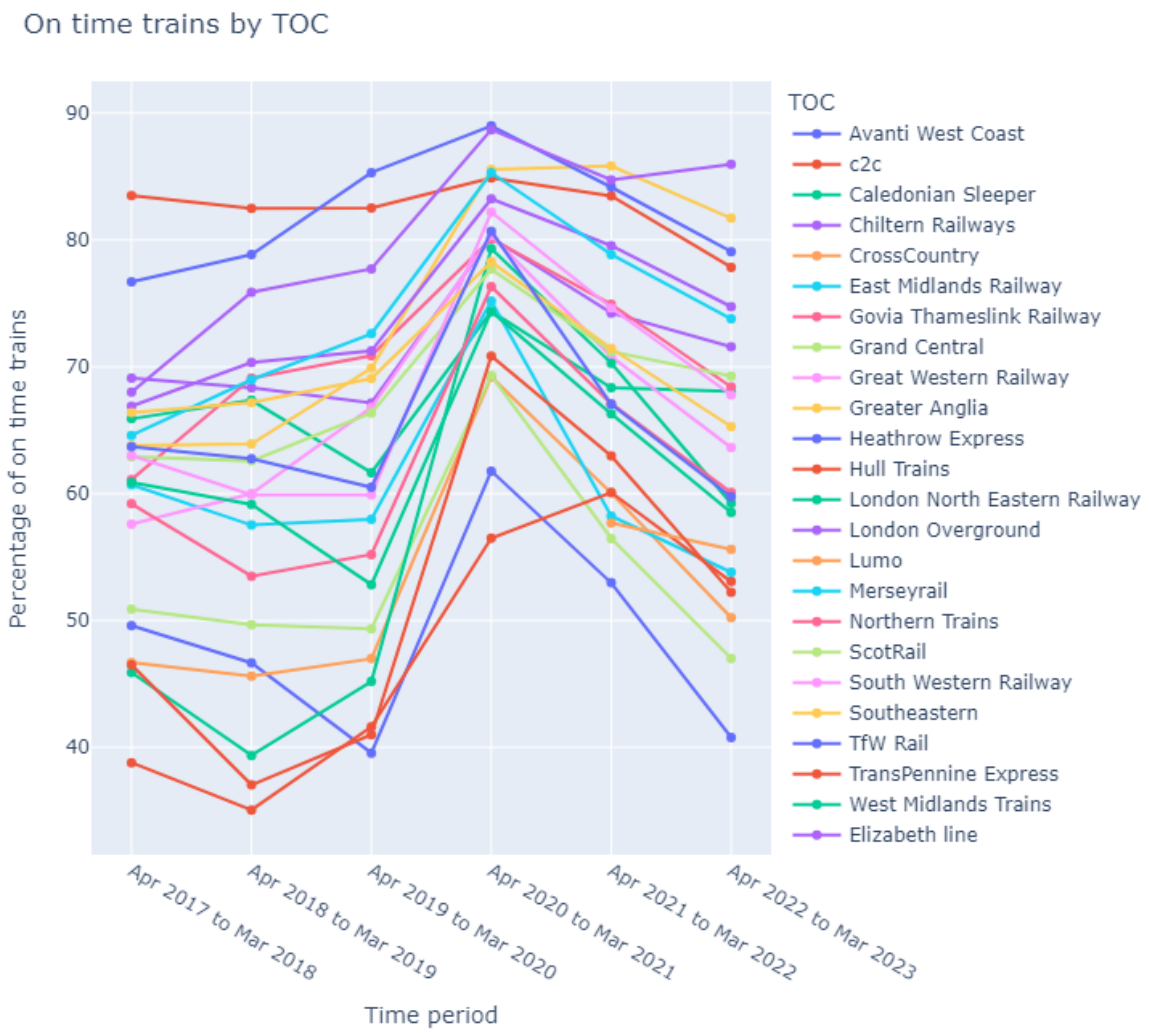
Delays caused by Network Rail consistently make up the majority of delays.

Percentage of on time trains

On Time is the percentage of recorded station stops that were early or less than one minute after the scheduled arrival time.



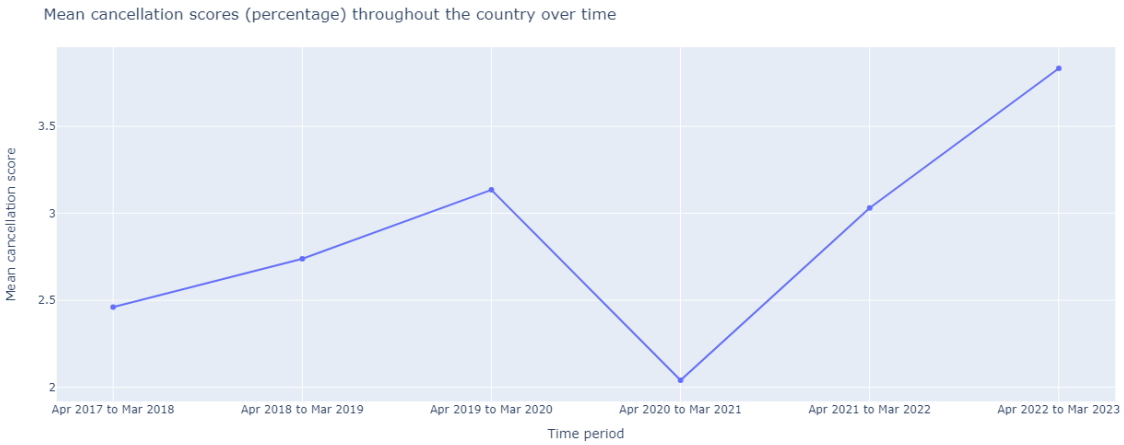
The average percentage of on time trains has achieved a peak during COVID-19, as noted by the value seen between April 2020 and March 2021. However, there has been a steady and noticeable decline in the subsequent time periods.



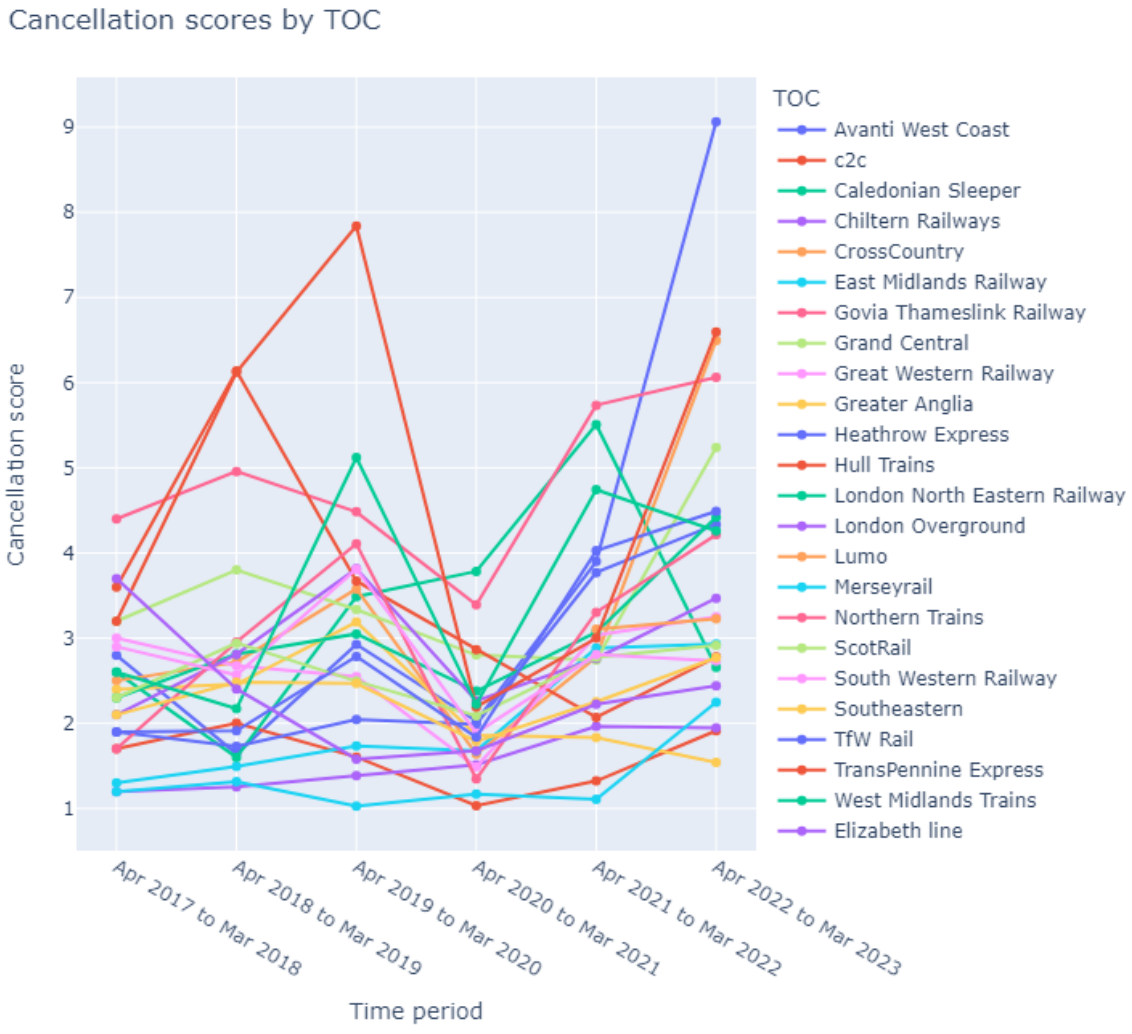
All TOC's, with the exception of c2c, also experienced a significant increase of on time trains during COVID-19. The majority of TOCs experienced a decline afterwards. Some TOC's, such as Greater Anglia and Hull Trains saw their peak value occur between April 2021 and March 2022.
Heathrow Express had the highest on time percentage score with 89 during the April 2020 to March 2021 time period. Hull Trains had the lowest with 35.1 during the April 2019 to March 2019 time period.

Cancellations score

The Cancellations score is the percentage of trains planned that were cancelled, whereby full cancellations are counted as one and part cancellations as half.



The average TOC cancellation score fell during the April 2020 to March 2021 time period. The post-COVID-19 values have grown to surpass pre-COVID-19 values.

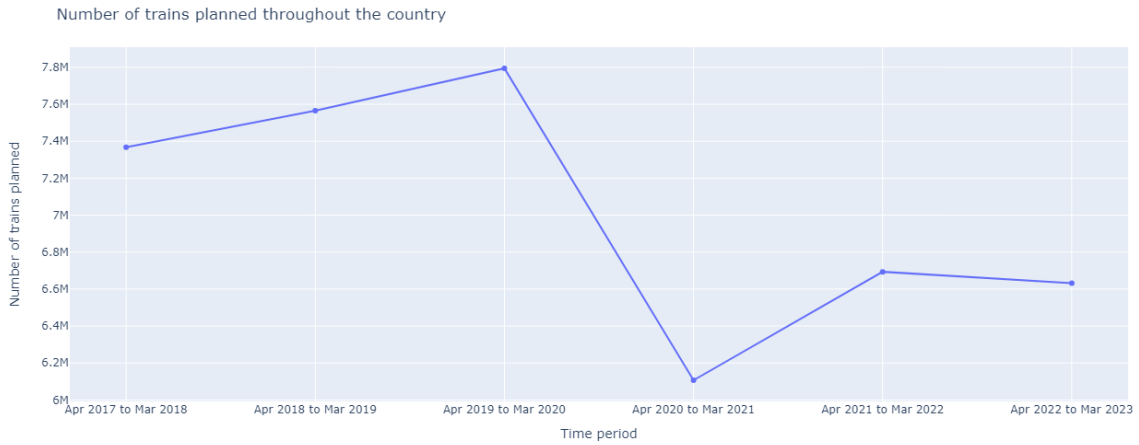


Looking at individual TOC cancellation scores, there is less of a pattern compared to other key parameters analysed. Generally, cancellation scores fell during COVID-19. Avanti West Coast was a large driver to large post-COVID-19 cancellation scores.

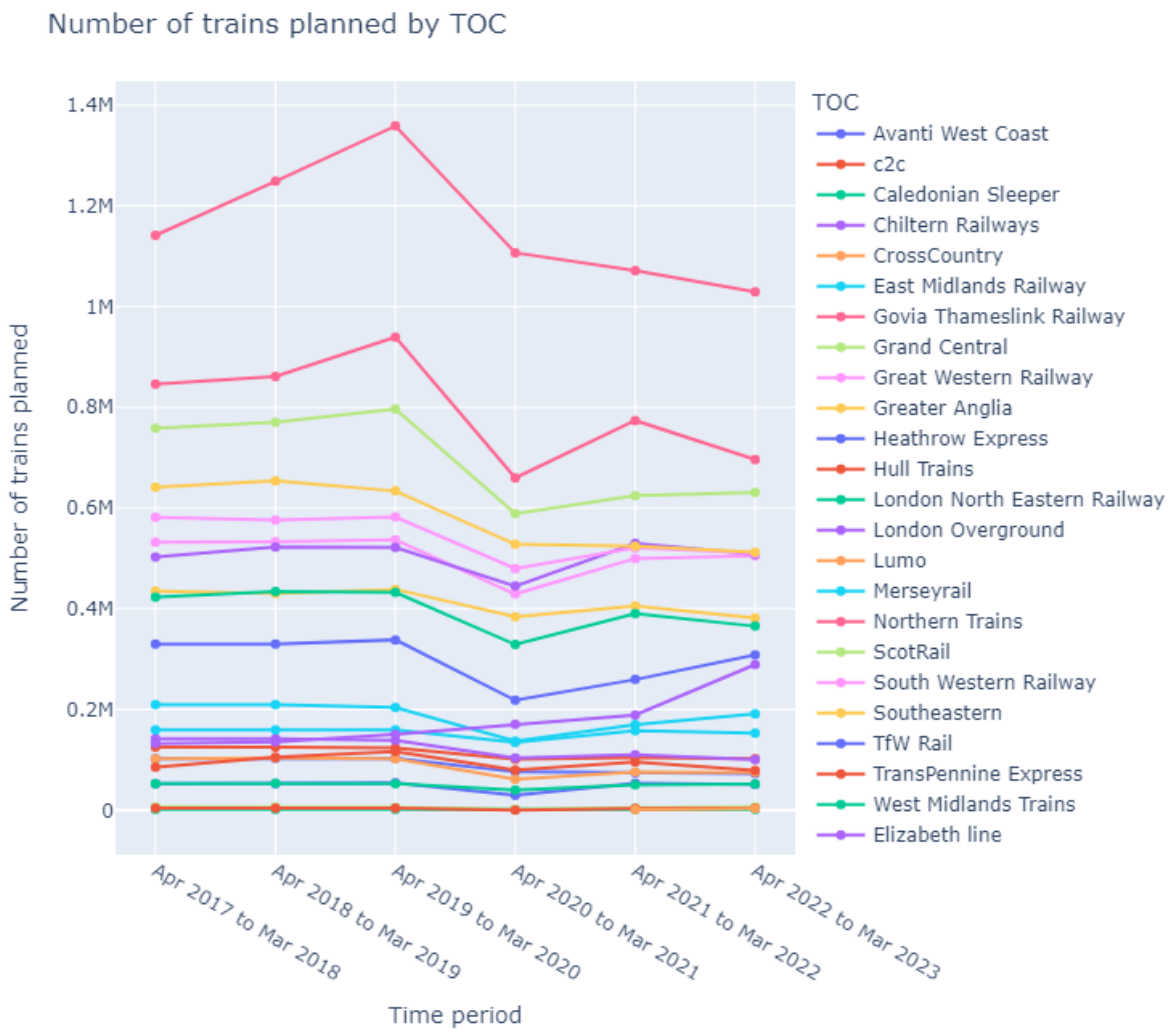
The highest cancellation score was TransPennine Express during April 2019 - March 2020. Merseyrail managed to achieve the lowest score with 1.0 during April 2019 - March 2020, with c2c achieving the same score a year after.

Number of trains planned

The number of trains planned is based on the daily schedule as agreed between the train operator and Network Rail at 22:00 on the previous evening.



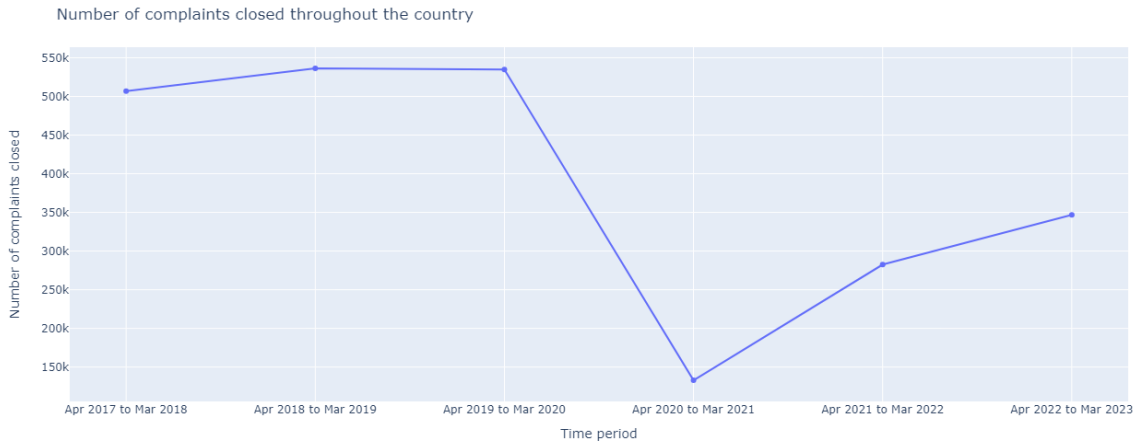
The number of trains planned throughout the country were slowly rising and then dramatically fell during COVID-19. There is little evidence that the this value will recover to pre-COVID-19 values in the near future.



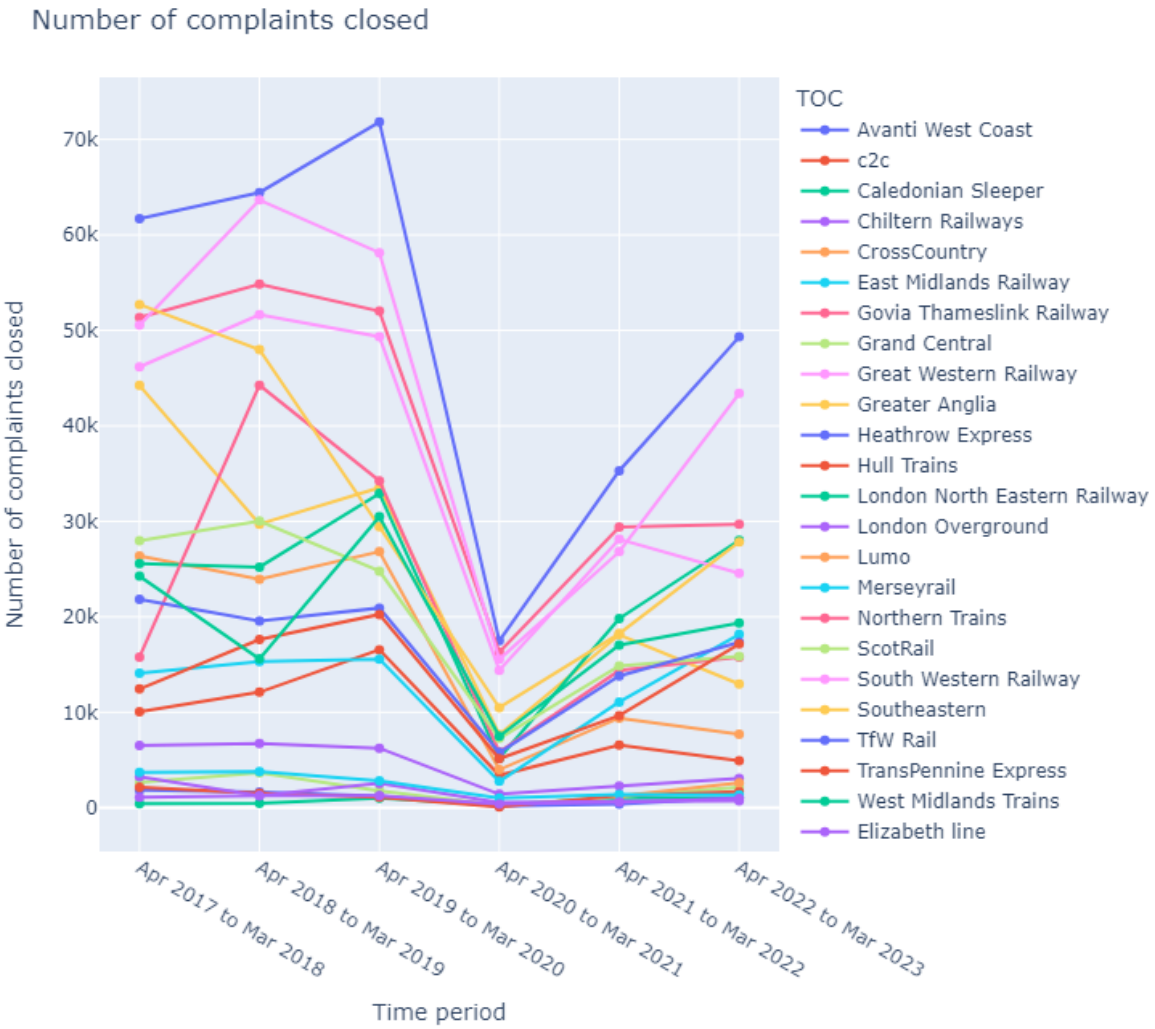
The individual TOC responses to planning trains vary. Some, such as Govia Thames Railway, are not engaged in planning more trains. Others, such as the Elizabeth Line, have planned more trains than pre-COVID-19, although that can be attributed to the opening of the new line.
Govia Thameslink Railway had the most trains planned in one year, with 1,358,603 trains planned between April 2019 to March 2020. Contrast to this is Hull Trains that planned only 506 between April 2020 to March 2021.

Number of complaints closed

Complaints are defined as ‘any expression of dissatisfaction by a customer or potential customer about service delivery or about company or industry policy’.



The number of complaints closed throughout the country fell between April 2020 and March 2021. The value has been increasing steadily in the subsequent years, but not to the levels observed pre-COVID-19..



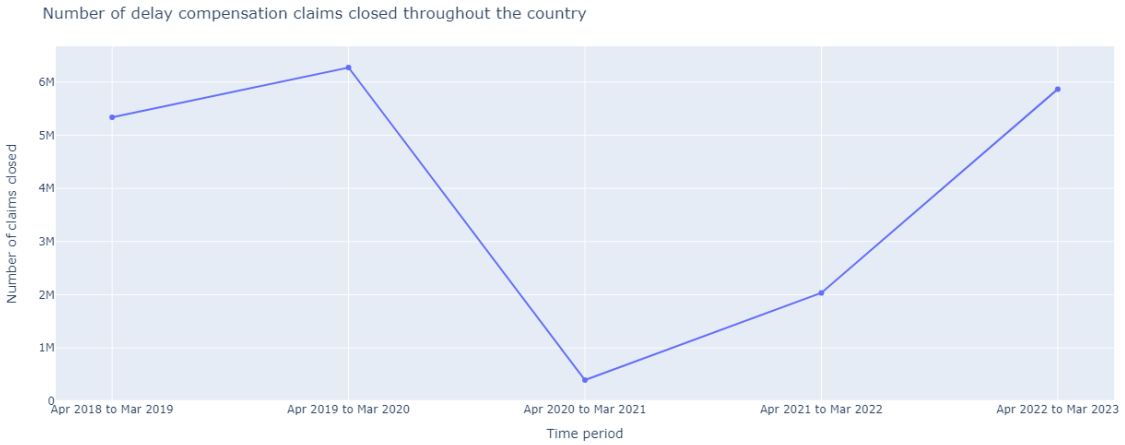
Generally, the number of complaints closed for a TOC reached a minimum during April 2020 to March 2021, with a slow increase in the years after.

Avanti West Coast had the most number of complaints closed with 71,814 during the time period between April 2019 to March 2020.

Hull trains had the least complaints closed, with 108 between April 2020 and March 2021

Number of delay compensation claims closed

Delay compensation claims closed refers to the number of claims closed when the train operator issues payment for a successful claim or when the passenger was informed that their claim was rejected.



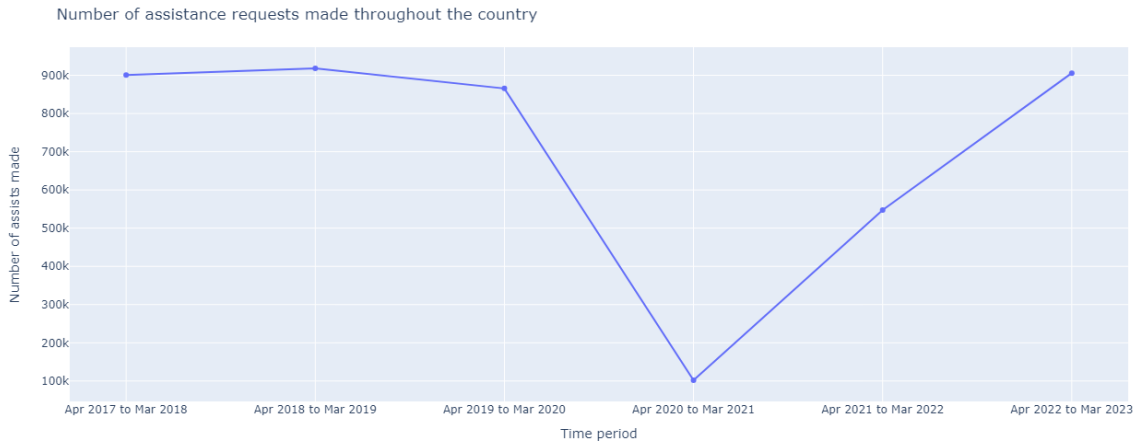
Despite a sharp drop in delay compensation claims closed during COVID-19, this value has now reached pre-COVID-19 levels.

In the dataset, Govia Thameslink Railway closed the most delay compensation claims, with 1,552,957 claims closed between April 2018 to March 2019.

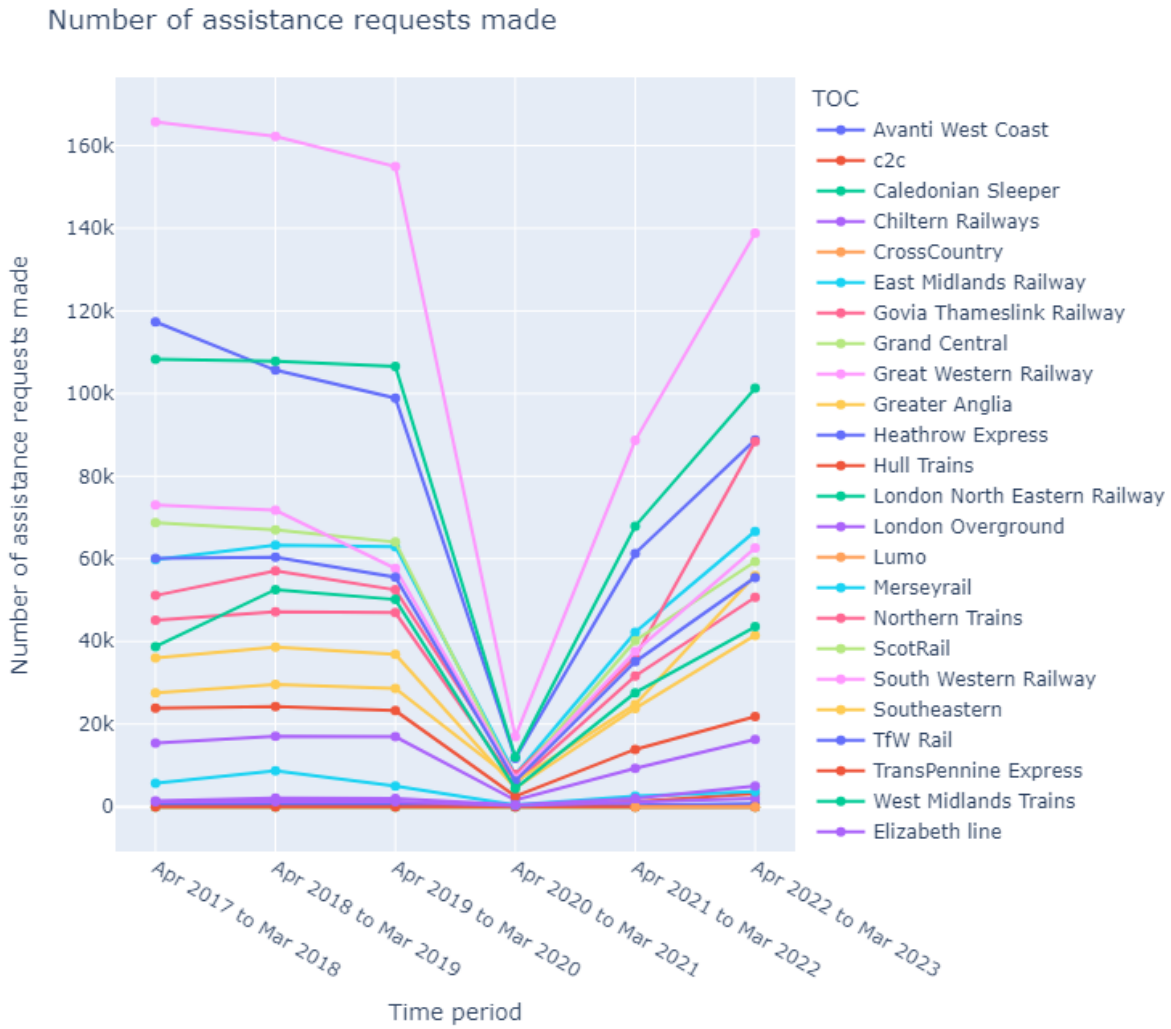
Merseyrail had the least, with 116 between April 2020 to March 2021

Number of passenger assists

Passenger assists data shows the number of assists that have been requested through the National Passenger Assistance Booking System (unbooked assistance such as ‘Turn Up and Go’ assists is not included).



The number of passenger assists requests made throughout the country fell between April 2020 and March 2021. The value has increased in the subsequent years, and is not at levels observed pre-COVID-19..



From the data, the greatest number of passenger assists requested occurred with Great Western Railway, with 165,698 assists made between April 2017 to March 2018.

Many operators did not have any passenger assists.

Conclusions

The following parameters experienced a decline during the first year of COVID-19:

1. Number of passenger journeys
2. Distance travelled by passengers
3. Distance travelled by trains
4. Complaints closed
5. Total Assist Requests
6. Delay compensation claims closed
7. Cancellations score
8. Trains planned
9. Network Rail on TOC delays
10. TOC on self delays
11. TOC on TOC delays

Out of these parameters, the following parameters have recovered afterwards to pre-COVID-19 levels:

1. Total assist requests
2. Delay compensation claims closed

The following got worse after COVID-19:

1. Mean cancellation score has increased
2. On time trains has worsened
3. Trains planned will not recover to pre-covid levels

COVID-19 did not significantly affect who were the biggest and smallest TOC's.

Less people are using trains now compared to before Covid, although this metric is currently growing. It will take a few more years until pre-Covid levels are met. However, with a trend of less trains planned, and rising cancellations, passengers may not experience the same levels of service they used to have.

Delays are primarily caused by Network Rail, regardless of Covid.

Further information

Data dashboard: https://antonyquang.pythonanywhere.com/
Data Analysis Notebook (Google Colab): Notebook
Github repository: https://github.com/AntonyQuang/Rail-Statistics-Dashboard