## Background/context of the business scenario

The NHS lose vast sums of money each year to patients missing appointments. It isn't fair to fine patients who miss appointments as these fines disproportionately impact poor and vulnerable members of the community. Our team of data analysts are aiming to help the government understand what the best way to handle this problem is through the use of data. I have been tasked with conducting exploratory analysis on the data to try and answer two main questions. These questions are: has there been adequate staff and capacity in the networks and what was the actual utilisation of resources?

### Analytical approach

- The data was imported using the 'pd.read' function, which was chosen as it allows you to import data directly into a pandas dataframe making it simple to analyse. The Pandas library was selected because it's highly efficient and it allows for aggregation and grouping, making it useful for the types of analysis being conducted.
- While cleaning the data, the 'isna()' function was used to identify missing values within the dataset.
- The .nunique() function was used to explore the dataset and determine the number of unique locations, service settings, context types, national categories, and appointment statuses. Another function that was used when exploring the data was a for loop, which was used to identify the 5 locations with the highest number of records within the dataset.
- A further step used to clean the data for use was converting the dates in each dataframe to datetime. The datetime library was chosen as it allows the use of time-related calculations, like working out when the appointments were scheduled between or what month had the highest number of appointments.
- The groupby function was used throughout the project as it's a simple and easy way to group data by attributes. This was used to group the data by months and location, allowing me to analyse locations against each other and create time-series data.
- The Seaborn and Matplotlib libraries were picked to visualise the data as they have a wide range of plot types, are highly customisable and work well with Pandas dataframes.
- Given that the pandemic happened during the timeframe we're analysing, further analysis using data from the NHS Coronavirus API was conducted to assess whether Covid had skewed the data. The Covid case data was compared against the utilisation/missed appointments data to assess their correlation. If a high correlation between Covid cases and the utilisation/missed appointments data was observed, the data might not be suitable for extrapolating findings beyond the pandemic, as the current circumstances no longer reflect the conditions during the pandemic.

## Visualisation and insights

- What is the number of locations, service settings, context types, national categories, and appointment statuses in the data sets?

There are 106 locations, 5 service settings, 3 context types, 18 national categories and 3 appointment statuses in the data sets.

- What is the date range of the provided data sets, and which service settings reported the most appointments for a specific period?

The date ranges of provided datasets:

```
The appointments in the ad dataframe were scheduled between 2021-12-01 and 2022-06-30
The appointments in the nc dataframe were scheduled between 2021-08-01 and 2022-06-30
The appointments in the ar dataframe were scheduled between 2020-01 and 2022-06
```

Which service setting was the most popular for NHS North West London from 1 January to 1 June 2022?

The service setting with the highest appointment count between 2022-01 and 2022-06 was General Practice.

| service_setting | count_of_appointments |
|---|---|
| Extended Access Provision | 115052 |
| General Practice | 5719516 |
| Other | 181576 |
| Primary Care Network | 130526 |
| Unmapped | 462758 |

- What is the number of appointments and records per month?

Records per month:

| appointment_month | appointment_month | |
|---|---|---|
| 2022 | 3 | 124590 |
| | 5 | 117829 |
| | 6 | 113626 |
| 2021 | 12 | 112551 |
| 2022 | 1 | 111764 |
| | 2 | 110876 |
| | 4 | 109163 |
| 2021 | 11 | 98418 |
| | 9 | 95363 |
| | 10 | 94640 |
| | 8 | 89785 |
| 2020 | 3 | 21350 |
| | 1 | 20889 |
| | 2 | 20689 |
| | 10 | 20122 |
| | 9 | 20043 |
| 2021 | 7 | 19899 |
| | 6 | 19814 |
| 2020 | 11 | 19675 |
| | 7 | 19502 |
| 2021 | 4 | 19452 |
| 2020 | 12 | 19394 |
| 2021 | 5 | 19384 |
| | 3 | 19369 |
| | 1 | 19319 |
| 2020 | 8 | 19247 |
| | 4 | 19124 |
| 2021 | 2 | 18949 |
| 2020 | 6 | 18844 |
| | 5 | 18338 |

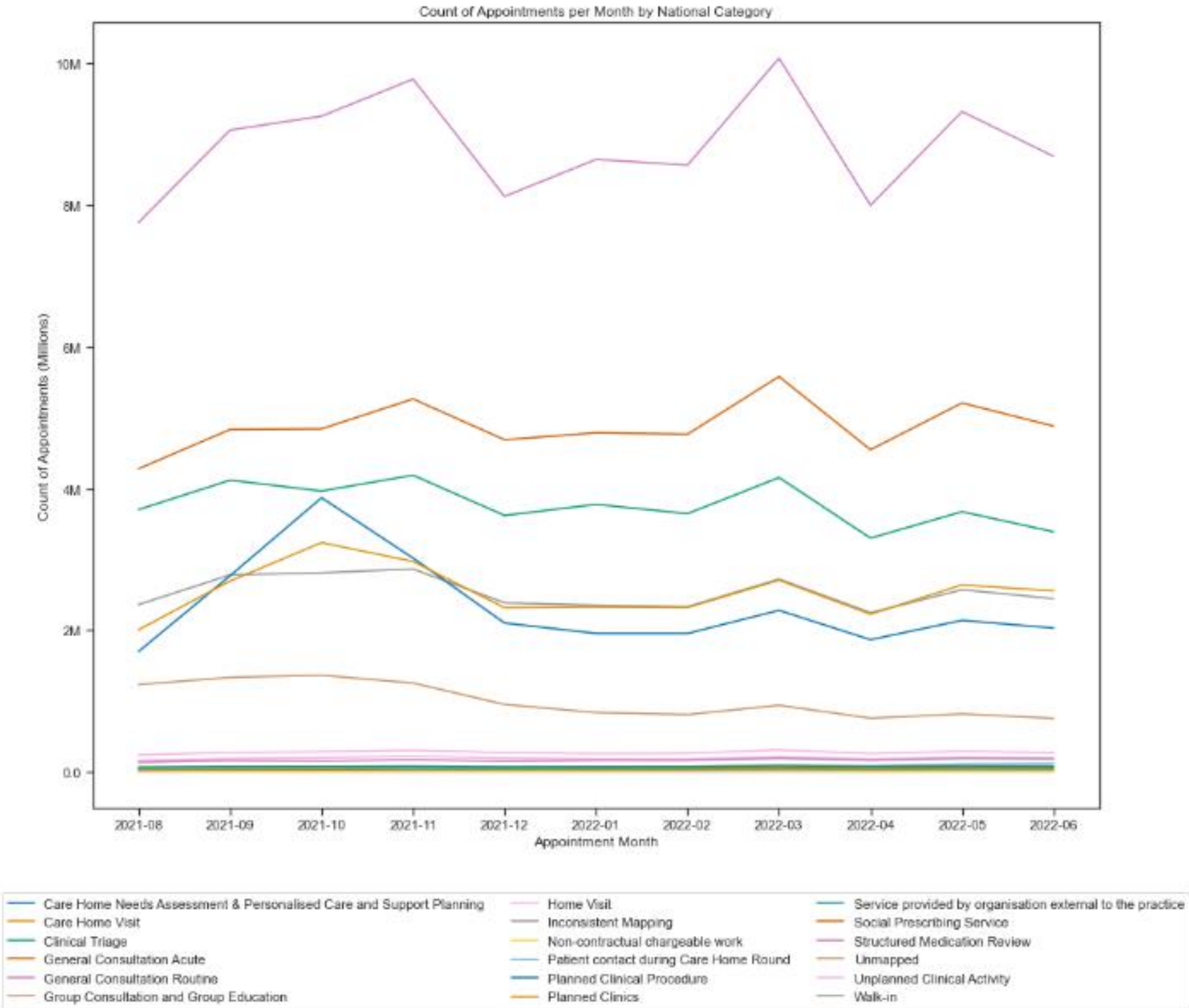
Appointments per month:

| appointment_month | appointment_month | |
|---|---|---|
| 2021 | 11 | 30405070 |
| | 10 | 30303834 |
| 2022 | 3 | 29595038 |
| 2021 | 9 | 28522501 |
| 2022 | 5 | 27495508 |
| | 6 | 25828078 |
| | 1 | 25635474 |
| | 2 | 25355260 |
| 2021 | 12 | 25140776 |
| 2022 | 4 | 23913060 |
| 2021 | 8 | 23852171 |

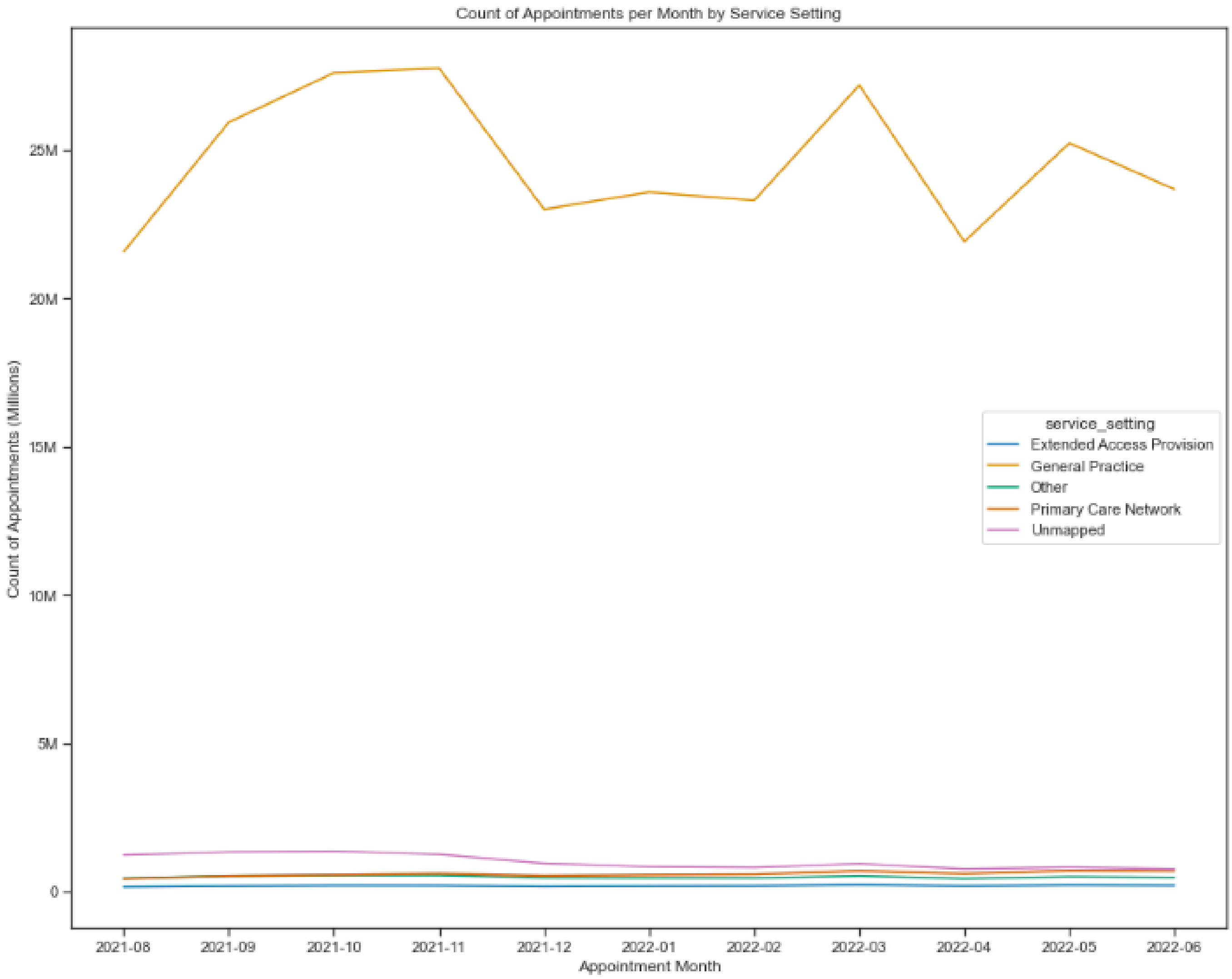
- What monthly and seasonal trends are evident, based on the number of appointments for service settings, context types, and national categories?

Seasonal trends:

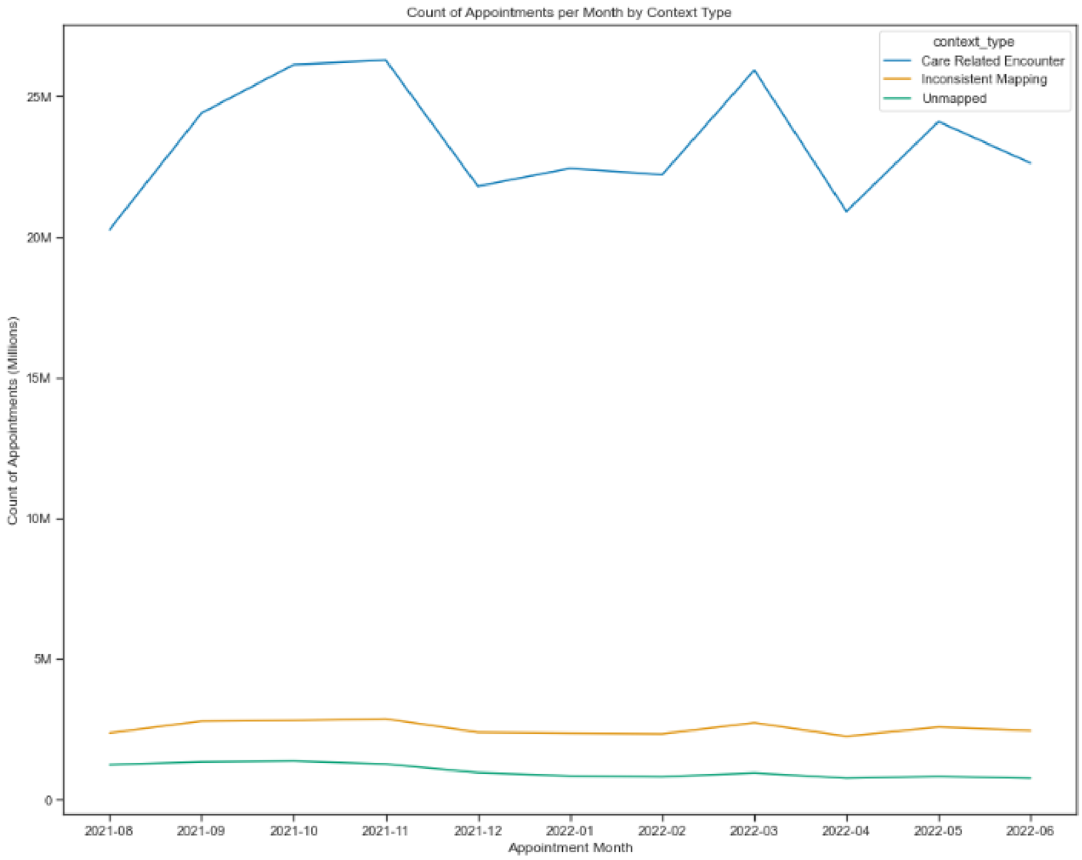
General consultation routine was the most common national category, followed by general consultation acute, and clinical triage. There is a spike in planned clinical procedures during November 2021.



The general practice (GP) service setting being by far the most common, and the number of unmapped appointments has gone down in this time.



The vast majority of appointments fall in the care related encounter context type, there are relativley few appointments which have inconsistent mapping or are unmapped(around 10-15%)



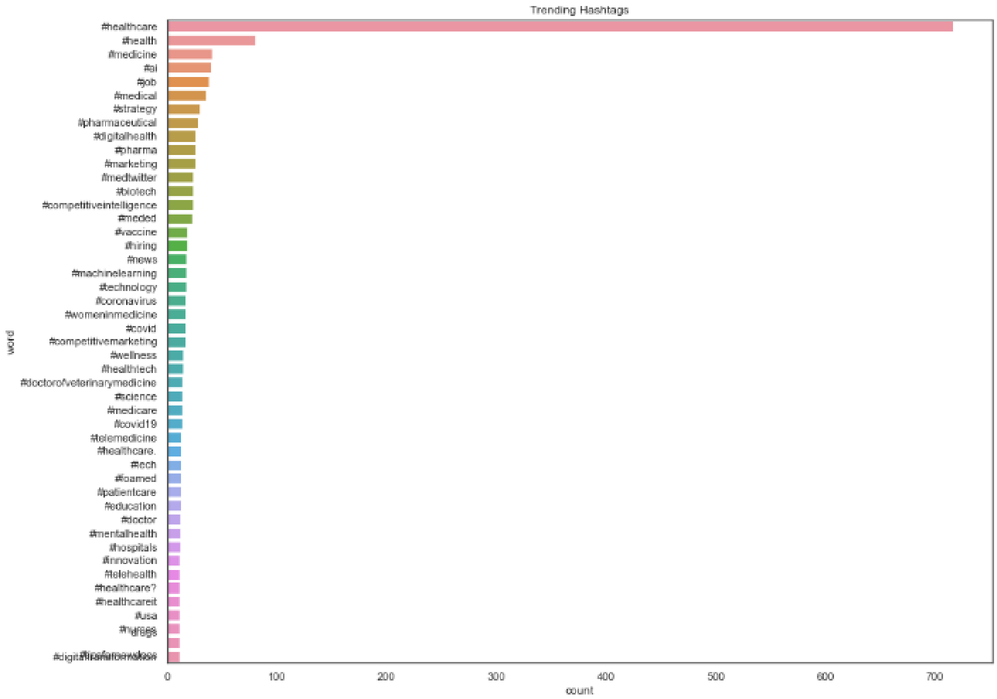

Seaborn lineplots were used to display the data as lineplots are an effective tool for displaying and analysing trends and patterns in time-series data.

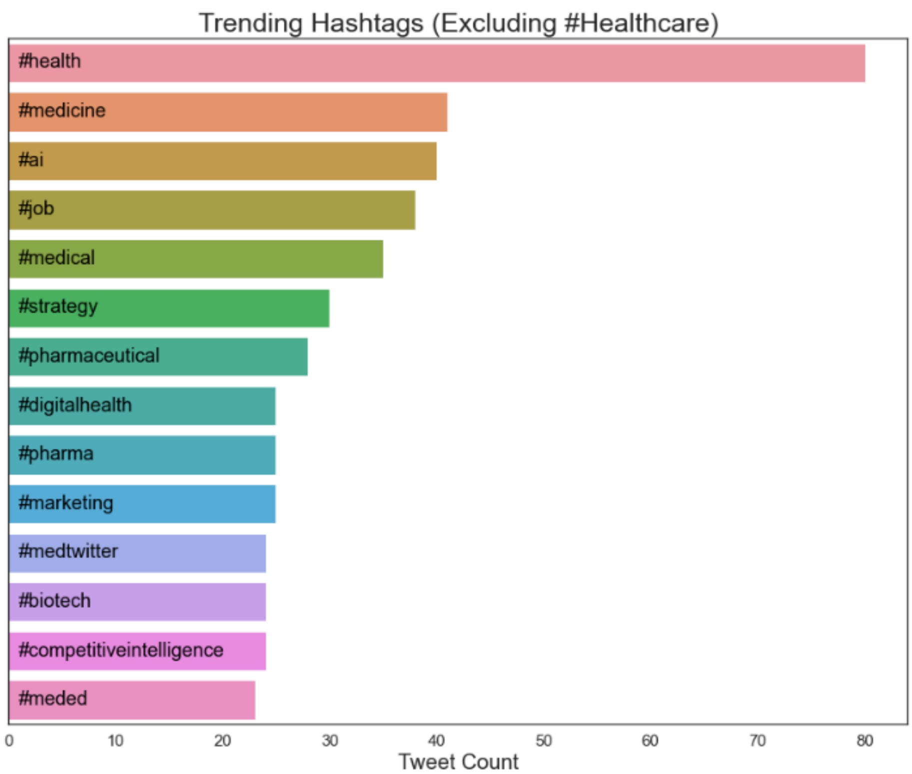
Seasonal trends:

- Autumn has the highest amount of appointments followed by Winter, Spring and Summer.

- What are the top trending hashtags (#) on Twitter related to healthcare in the UK?

By far the most used hashtag was #healthcare, then #health and #ai. #Healthcare saw more than 9x usage than the next most popular hashtag.

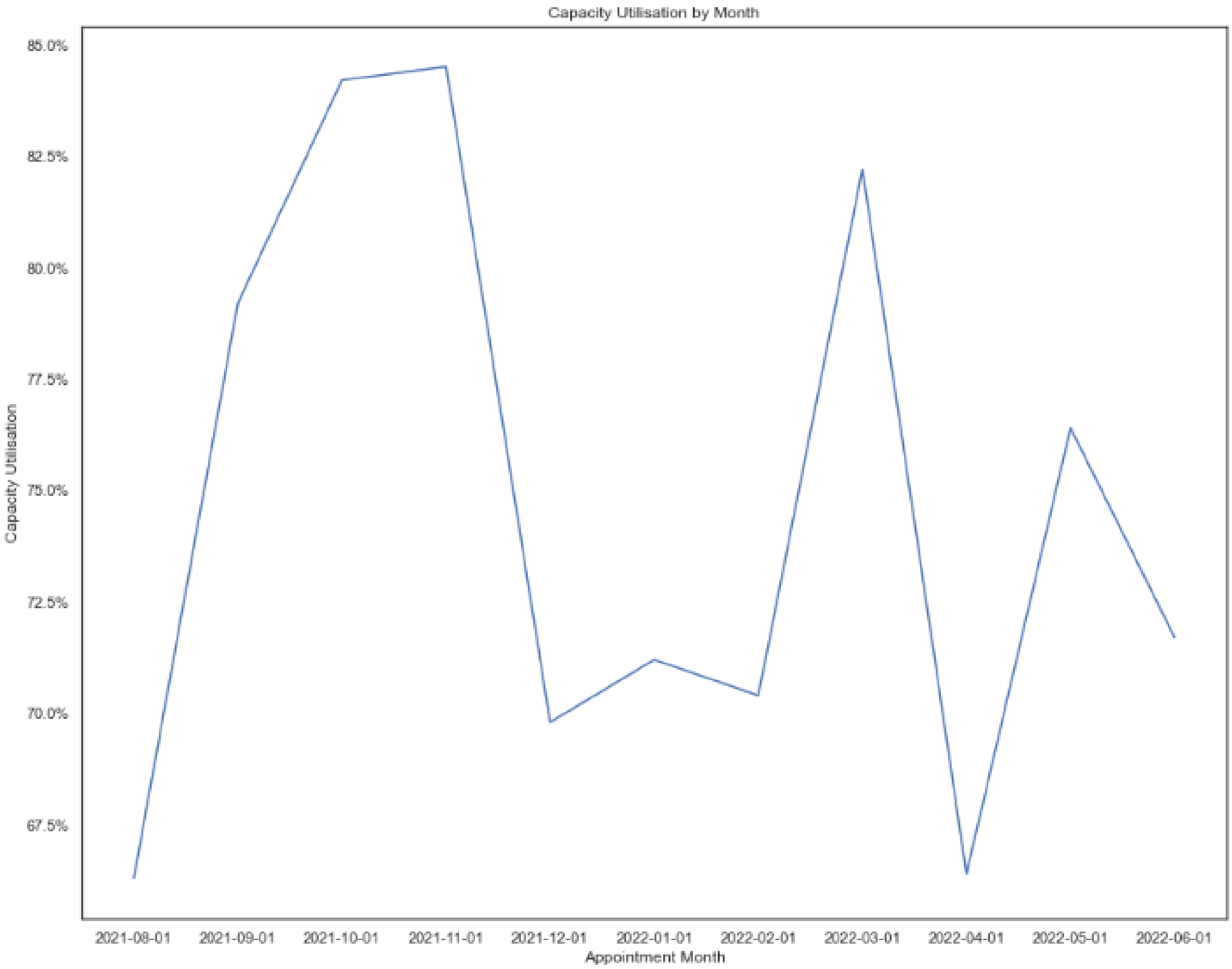


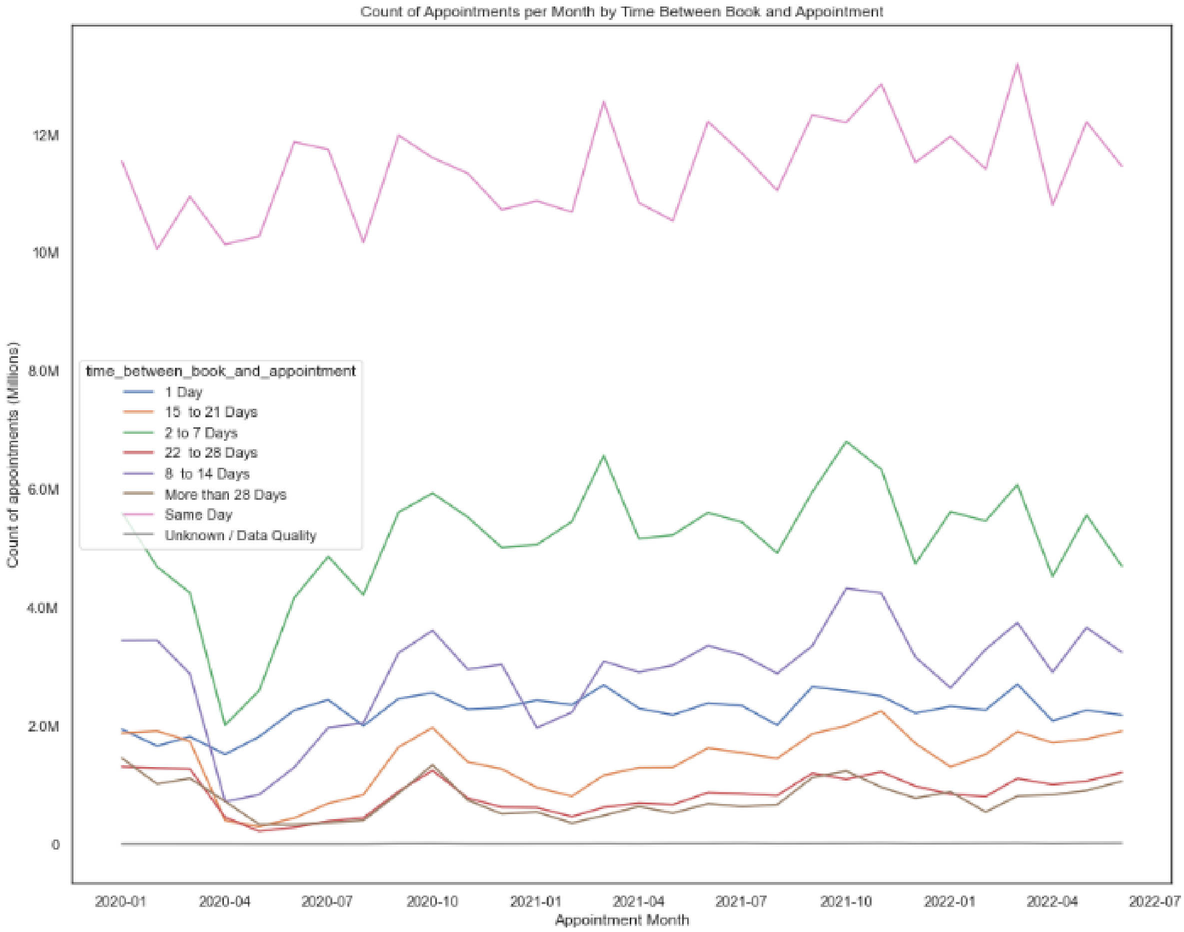

Seaborn barplot was used to display the data as barplots are an effective tool for displaying catigorical data.

- Were there adequate staff and capacity in the networks?

Capacity utilisation remains below 85% during the time period accessed, and booking time (time waiting for an appointment) remains fairly consistent, indicating there is an adequate amount of staff and capacity in the wider network.

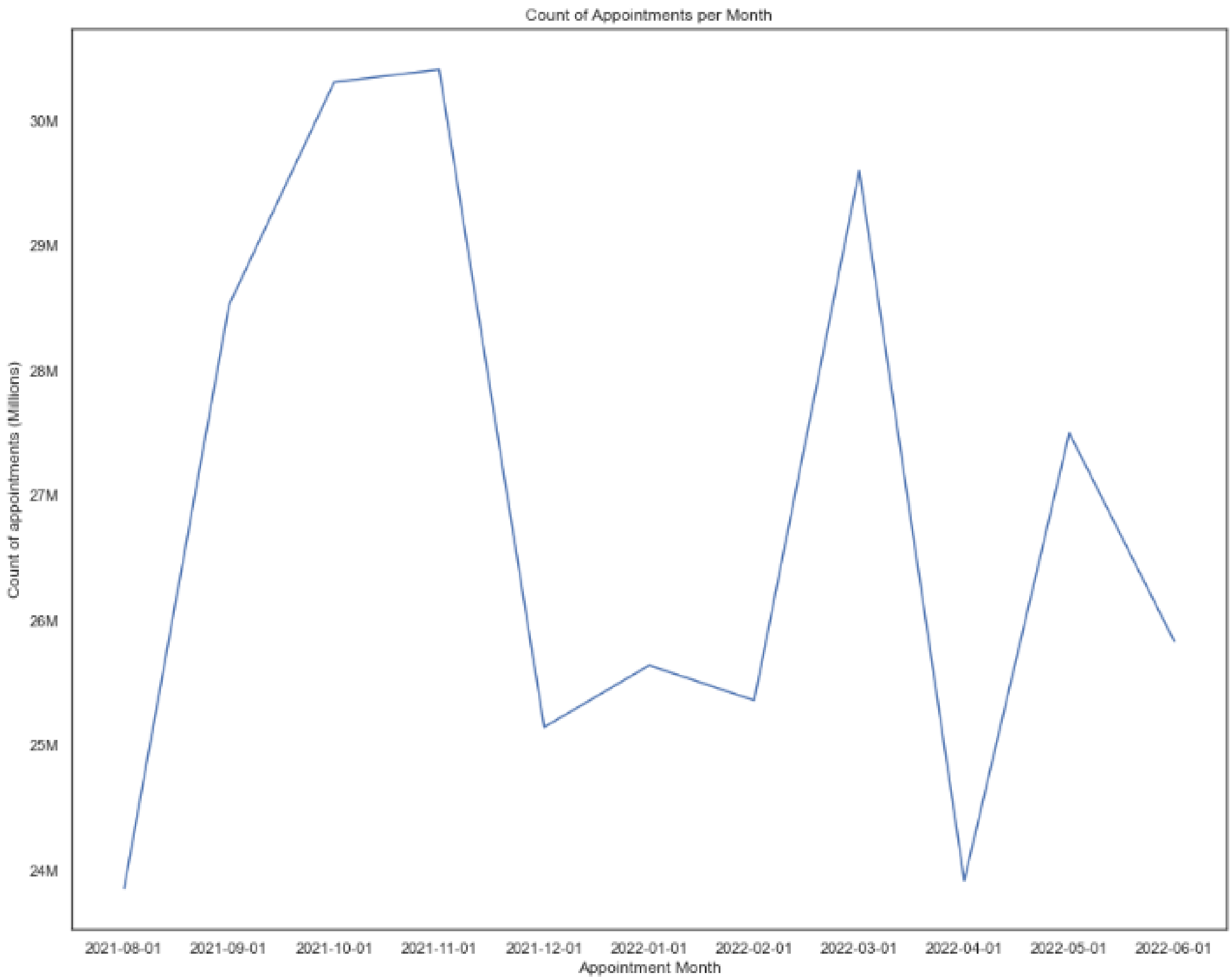


The time between booking and appointment over the whole network remained relatively stable over time.



- What was the actual utilisation of resources?

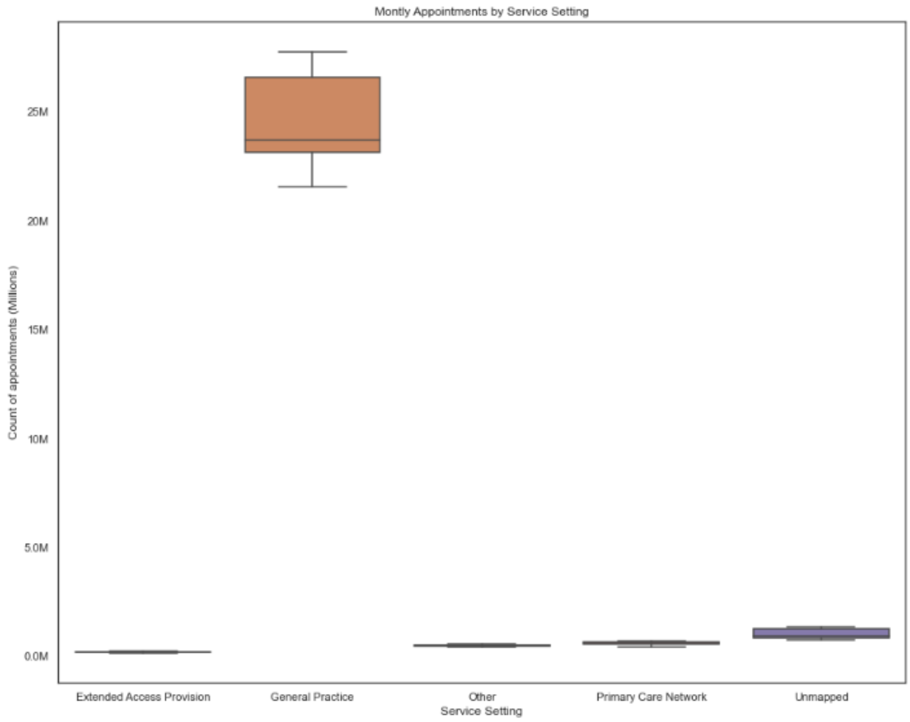
The utilisation of appointments ranges from 23.85 million to 30.4 million a month (795k – 101k daily).

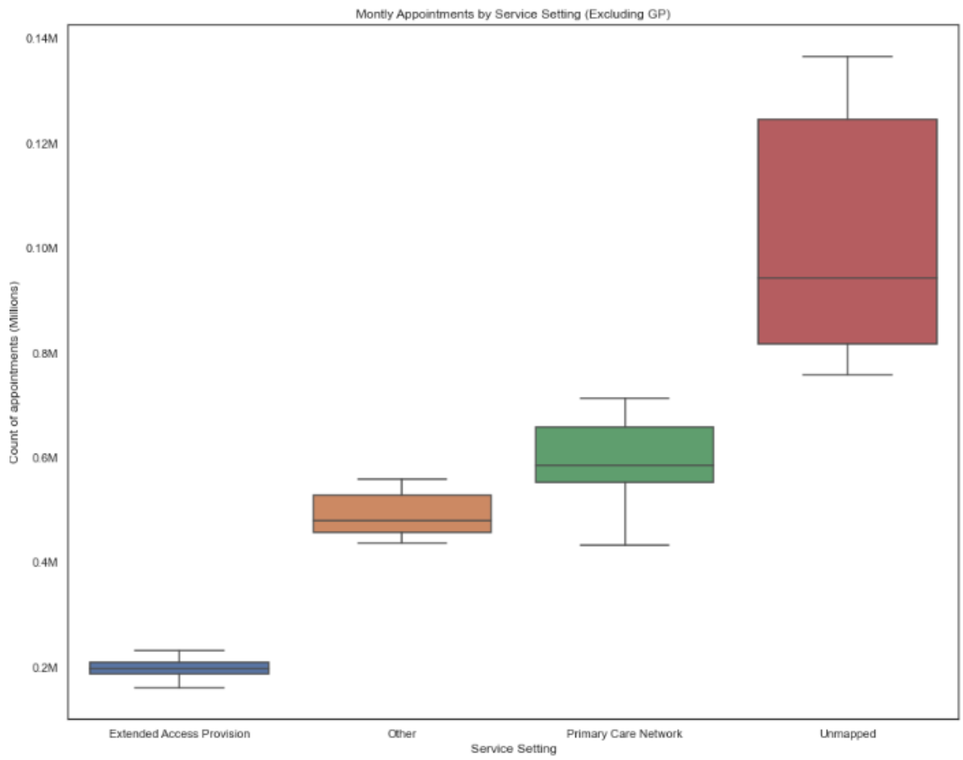


**Other visualisations and insights:**

Telephone appointments have doubled in the last 2 years

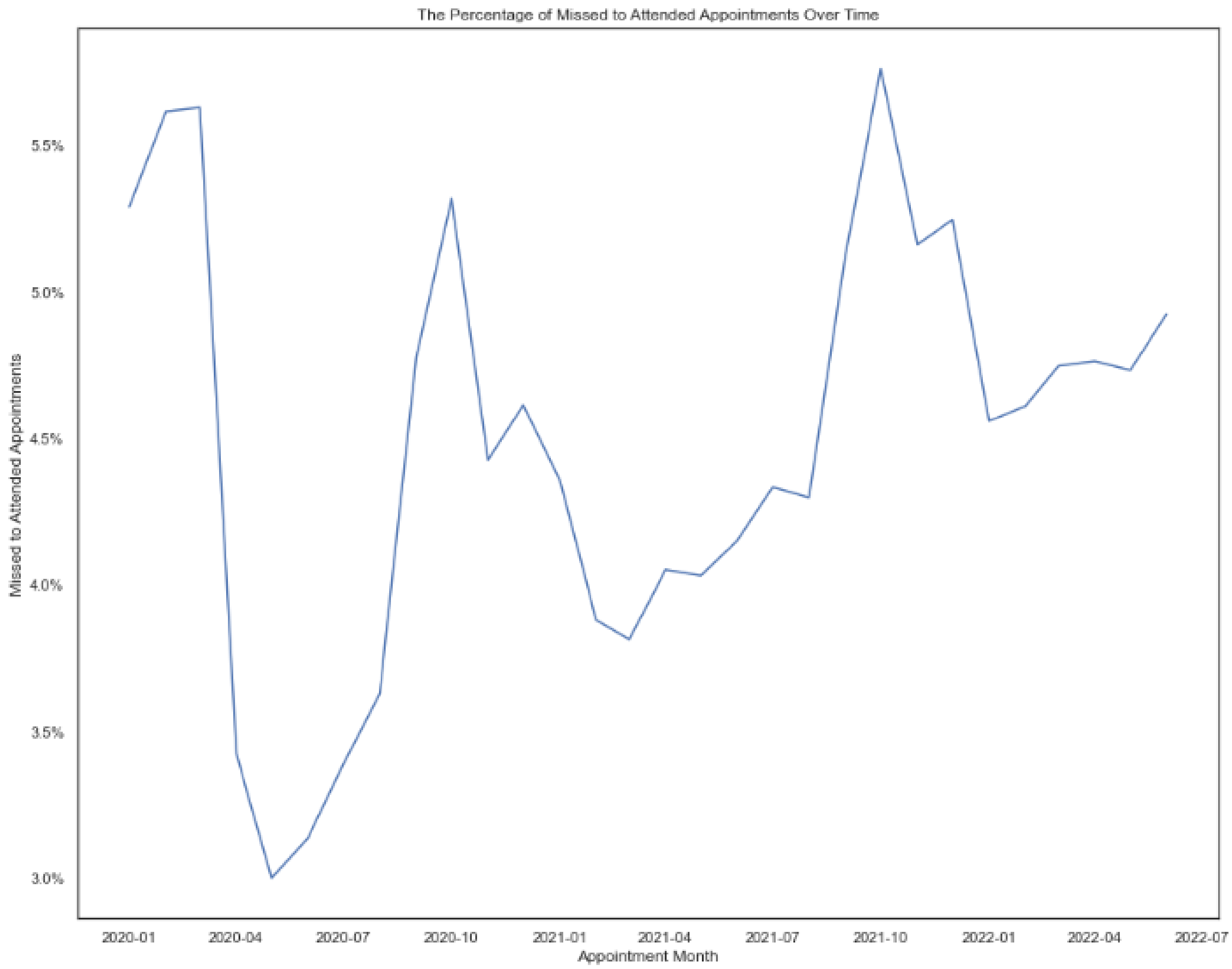
All service settings have no outliers when analysing monthly appointments showing there is a relatively consistent distribution of appointments per month and adding to the validity of the data.



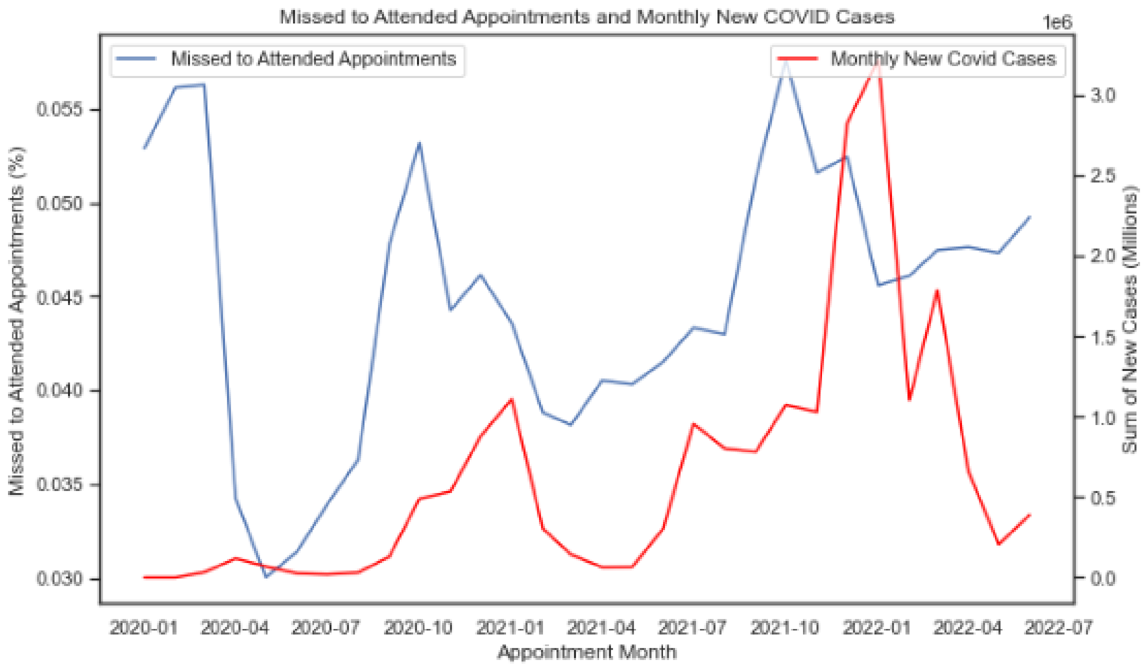



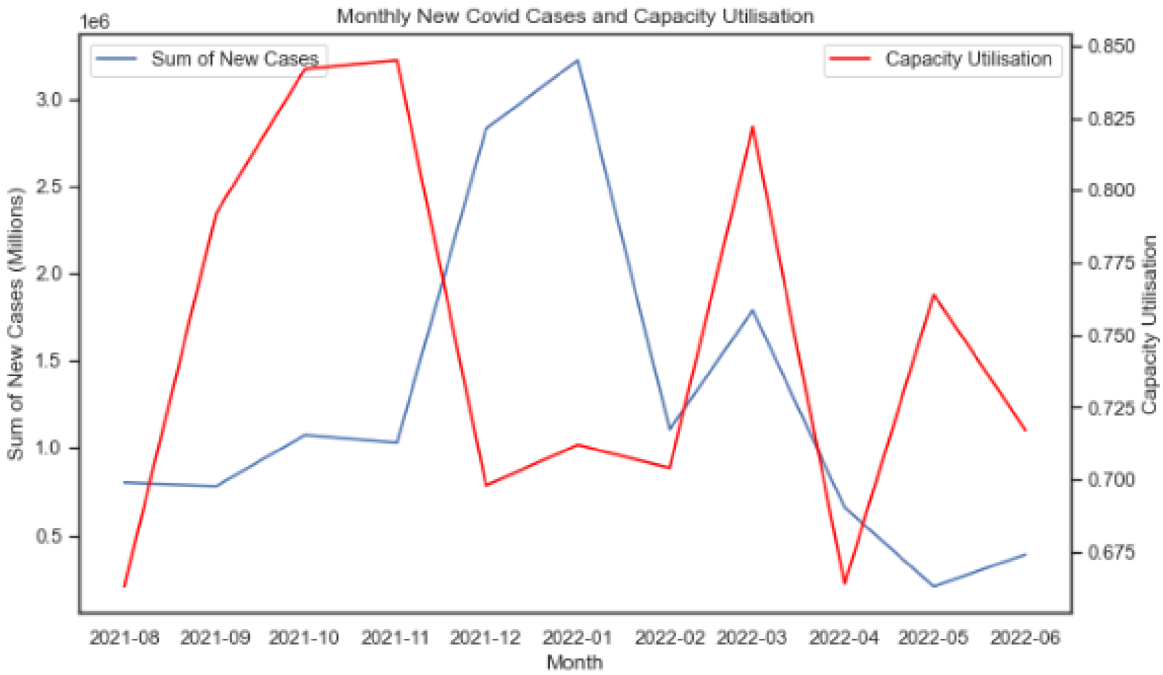

A boxplot was used to display the data as boxplots are an effective tool for analysing the distribution.

The percentage of appointments that are missed has been trending up significantly since mid 2020.

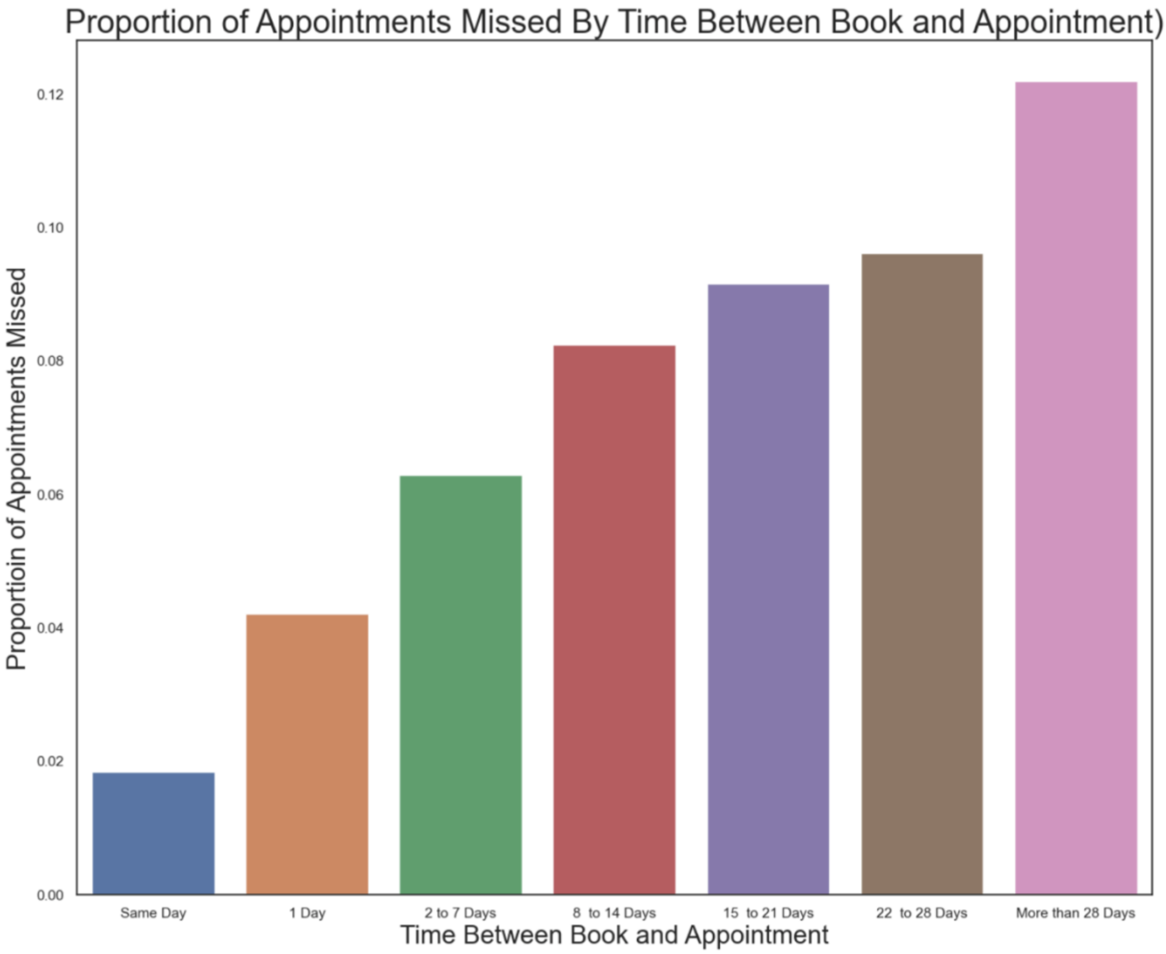


The utilisation and missed appointment data does not appear to have been impacted by covid.



The time between booking and appointment correlates with the portion of appointments that are missed.



There is a large difference in the proportion of appointments that are carried out on the same day they were booked depending on location.

```
icb_ons_code
E54000041    0.396765
E54000043    0.412949
E54000022    0.420651
E54000040    0.423616
E54000051    0.429173
E54000061    0.432939
E54000058    0.435601
E54000037    0.437750
E54000060    0.443365
E54000050    0.443977
E54000052    0.444741
E54000024    0.446418
E54000027    0.446883
E54000032    0.446970
E54000039    0.447176
E54000038    0.447360
E54000026    0.447916
E54000053    0.450534
E54000042    0.451251
E54000057    0.456481
E54000015    0.460421
E54000013    0.463329
E54000048    0.463586
E54000054    0.464397
E54000036    0.470035
E54000029    0.470812
E54000031    0.472438
E54000044    0.473337
E54000030    0.474017
E54000023    0.475151
E54000010    0.476206
E54000025    0.484305
E54000008    0.486946
E54000019    0.489883
E54000062    0.492834
E54000011    0.494010
E54000034    0.495890
E54000055    0.498175
E54000059    0.504858
E54000028    0.515756
E54000056    0.518400
E54000018    0.533602
```

## Patterns and predictions

- Covid doesn't appear to have skewed the relevant GP data.

- The overall capacity utilisation of the NHS network is at an acceptable level, as a result it appears the NHS has enough staff.

- People are more likely to miss appointments when they have to wait longer to see a GP/Doctor.

- According to an Oxford newspaper there's a high variance in the waiting times for doctors by practice. This is supported by research I did which shows the amount of appointments which happened the same day as booking varies significantly by location.
    - https://www.oxfordmail.co.uk/news/23178373.gp-practices-oxfordshire-longest-waiting-times/
- Different waiting times suggests different capacity utilisations in practices and locations. While the NHS have enough staff, they may not be in the right locations to meet the demands in appointments from patients.

- Since the amount of phone appointments doubled over the past few years, it may be possible to reduce wait times by making under-utilised GP practices take some of the phone appointments from practices that have long wait times/ high capacity utilisation.

- If the NHS can use phone appointments across practices/locations to reduce waiting times they may see a reduction in the amount of missed appointments.

- Recommendation: More research/trials must be carried out to determine whether there's causation in the relationship between waiting times and missed appointments or whether it's only correlational.

- Limitation: because the data collection methods between each GP varies, the data doesn't give a complete view of GP activity so may not give an accurate picture of GP workload. Because of this, no changes to the NHS should be actioned using the findings without further analysis.