Background

The NHS is concerned about the cost implications of missed appointments.

The government wishes to understand:

- whether there have been adequate staff and capacity in the networks? And;
- what is the actual utilization of recourses

This report is an initial approach to the data exploration, data wrangling, visualizations and identifying possible trends in the data sets.

Executive Summary

- From the initial analysis it is likely that there are insufficient staff and capacity in the networks (Figures 5, 6, & 7). The actual utilization of recourses is visualized in Figure 11 which supports this.
- The actual number of appointments on a business day (the days when General Practice are open) averages 1,259,973 whereas the NHS capacity is for 1,200,000 daily.

Analytical approach

The analytical approach was to first import and understand the data (its scale, shape type etc…).

Initial exploratory analysis was then undertaken to establish useful general information such as total number of appointments per service setting, most popular location, date ranges etc…

Further analysis was carried out relating to count of appointments per service setting, context type and national category on a monthly basis and this was then explored in relation to service setting on a seasonal basis.

Utilisation was then calculated on an average daily basis (based on business days) to measure the extent to which capacity was being used and resources ultilised.

A detailed summary of the technical process is contained in Appendix 1 – Technical Process.

The outputs from the process are detailed in Appendix 2 - Outputs.

Table	**Appendix 2**
The number of locations, service settings, context types, national categories, and appointment statuses	The responses to Question 1 and Question 3
The five locations with the highest number of records	Table 1
The date ranges	Tables 2 & 3
The service settings which reported the most appointments	Table 4
The service settings which reported the most appointments between 1 Jan – 1 Jun 2022 in the most popular location (NHS North West London)	Table 5
The number of appointments and records per month	Table 6 & Table 7 respectively.
The top 30 trending hashtags on twitter	Table 8

Visualization and insights

The following visualizations were created to increase the accessibility of the data. Additionally, the palette colorblind was used to further increase accessibility.

Visualization	Appendix 2
Monthly trends are evident, based on the number of appointments for service settings, context types and national categories.	Figures 1, 2, & 3 respectively
Seasonal trends are evident, based on the number of appointments for each service setting on a seasonal basis.	Figures 4, 5, 6 & 7
Hashtags with a count greater than 10	Figure 8
Visualizations to inform data driven decision making related to whether or not the NHS should start looking at increasing staffing levels.	Figures 9, 10 & 11
Changes to health professional types over time	Figure 12
Changes in appointment status over time	Figure 13
Changes in appointment mode over time	Figure 14
Changes in time between a booking and appointment over time	Figure 15
The spread of service settings and the spread of service settings excluding GP	Figures 16 & 17 respectively

The visualizations show both patterns and trends in the data.

The data available relating to context type only shows useful data for one defined variable (Care Related Encounter) (Figure 2) which is broadly similar in trend to the General Consultation Routine in the National Category data (Figure 3). Both follow an almost identical trend to the Service Setting trend (Figure 1) as that Context and that National Category are undertaken predominately in General Practice.

Additionally, it is clear that the vast majority of appointments are seen in General Practice and on the same day (Figure 1 & Figure 15). As they are not pre-booked appointments like Primary Care Networks or Extended Access Provision comparative analysis is not particularly useful.

Figure 17 shows the spread of service setting excluding GP which shows little value in further analyzing those settings on the scale the project requires.

A weekly trend is evident in the seasonal visualizations where the busiest day is typically Monday (with the exception of Tuesday's following bank holidays) and the number of appointments declines through the week. This is true for all seasons (Figures 4, 5, 6 & 7).

Figures 9, 10 & 11 show the importance of calculating the actual use of resources and Figure 11 shows that there is likely to be insufficient staff and/or capacity in the network.

In the absence of staffing data it has not been possible to draw conclusions or make recommendations about health professional types over time, as they are the health care professionals who attended to an appointment (Figure 12) not the number of available health care professionals.

The data used to measure changes in appointment status (Figure 13) does not indicate a significant increase in Did Not Attend.

Figure 14 and Figure 15 add little value to the project as the most significant appointment mode (Face-to-Face) and most common time between booking and appointment (Same Day) are both on the same trend as General Practice by number of appointments as General Practice appointments are typically carried out face to face and on the same day. The same day characteristic is also the reason for the large number of appointments on each Monday (having been closed for the weekend).

Patterns and predictions

The vast majority of appointments are undertaken in General Practices. The other two defined service settings are Extended Access Provision and Primary Care Networks. General Practices operate only on business days whereas Extended Access Provision and Primary Care Networks are wholly pre-booked appointments which take place seven days a week. It is clear that Mondays (or Tuesdays after a bank holiday) are by far the busiest days of the week in General Practices (Appendix 2: Figures 4, 5, 6 & 7).

The NHS has been operating above capacity since autumn 2021, winter 2021-22 and spring 2022 (Figures 5, 6, & 7 respectively).

September, October, November and March are the busiest months (Figure 9).

The method of calculating daily utilization suggested was flawed and shows average daily including weekends when General Practices (which account for almost 92% of all appointments – see Figure 16) are not open. This leads to an assumption that the NHS is operating within its daily capacity (Figure 10).

However, it is clear that when daily utilization is calculated appropriately (Figure 11) the NHS is operating above capacity which is not sustainable. (Note: this calculation has been undertaken using business days but it is clear that there was an additional break over the Christmas/New year period).

The average number of appointments on a business day (the days when General Practice are open) is 1,259,973 whereas the NHS capacity is for 1,200,000 daily.

Without data on the numbers of staff employed in each service setting it is difficult to draw a conclusion or make a recommendation regarding staffing levels.

Recommendations

The average weekly figure (1,259,973 x 5 working days) is 6,299,865. If General Practices were open seven days a week the average daily figure would be 899,981 significantly below the daily capacity figure. However, if the reduction in appointments seen in General Practice since Winter 2021-22 is due to a reduction in staffing levels in that service setting (as opposed to seasonal trends in appointments) then it is unlikely that introducing weekends into the working week would improve staff retention. In summary, the NHS does need to look at staffing levels as there does not appear to be adequate staff and/or capacity in the networks and resources are very heavily utilized.

Issues

Data relating to staffing levels and to unsuccessful attempts to book appointments (because of capacity issues) must be collected in order to draw meaningful conclusions about staffing levels.

September, October, November and March are the busiest months. While the seasonal trend is evident in Sept-Nov, why is March so busy? The data needs to be collected and analyzed over a longer timeframe as it is unclear whether the trend is cyclical (or not) over an 11 month timeframe.

Wordcount: 1,099

Appendix 1: Summary of technical process

To begin **importing and exploring the data**, Pandas, numpy and warnings were imported.

Data had been cleaned prior to importing.

The three primary datasets (actual duration, appointments_regional and national categories) were then imported using **pd.read.csv** or **pd.read.excel** as appropriate.

These datasets were sense checked to ensure they imported correctly using **.shape** and **.dtypes**, prior to checking for any missing values usings **df[df.isna().any(axis=1)]**.

The metadata of each of the datasets was determined using **info()** and descriptive statistics were determined using **.describe()**.

The **data was then explored to answer initial questions**. For example, to determine the number of locations, **.nunique()** was used on a new variable which was then **print()**.

The top five locations: was determined using **value_counts().head()**

Note: NHS North West London was the busiest location.

The date format in the datasets was then converted to datatime where necessary using **pd.to.datetime()**. The result was checked using **.info()** and the the first five rows of the datframe were viewed using **.head()**

The date range was established as follows using **agg(['min', 'max'])**.

A new dataframe was then created to begin analysing the data.

In order to establish which service setting reported the most appointments from 1 January to 1 June 2022 a **.copy()** of the dataframe was created and greater than and less than or equal to commands were used to set the date range. **Groupby()** was the used prior to **print()** the results.

This process was then repeated to determine which was the most popular service setting in NHS North West London from 1 January to 1 June 2022.

To establish which month had the highest number of appointments, the format of appointment month was first changed to datetime **pd.to_datetime()** and **groupby** was used to order the appointments months by the sum of appointments:

A **value_count()** was used to determine the total records per month, and a count was used to determine the total number of records.

In order to **visualize the data and identify initial trends** seaborn and matplotlib were imported and the figure size and plot style of the outputs were set.

A new dataframe was created from the existing dataframe with only the necessary columns, and a new column was inserted to convert the appointment_month to a string using **.astype(str)**

In order to improve visualization **.replace()** was used on the new appointment_month_str column to convert the date format into words e.g. **.replace("2022-03-01": "Mar")**

Groupby was used to aggregate on a monthly level based on the **.sum()** of appointments.

The output was converted back into a dataframe using **pd.DataFrame()** and reseting the index **.reset_index()**

The dataframe was then sorted using **sort_values()**

Three visualisations were necessary to demonstrate changes to each of service settings, context type and national categories over time and were created using seaborn.

The **sns.relplot()** was used with **kind=line**.

An order was set to allow for comparison with **hue_order =[]** and to increase accessibility the **palette="colorblind"**

A further four visualization were created to indicate the number of appointments for service setting per season.

A new dataframe was created from the previous one - pd.DataFrame() and reset_index()

Values were sorted using **.sort_values()** by appointment_date (to maintain numerical sorting) prior to **.loc[month_ss_df['appointment_month_str'].isin(['Aug','Oct', 'Jan', 'Apr'])]** being used to limited the dataset to only the required months (one month demonstrating each season).

A new variable was created to show the appointment month as the abbreviated weekday and date in month **seasons['app_date_day'] = seasons['appointment_date'].dt.strftime('%a/%d')**

Saturdays and Sundays were then dropped from the dataframe:

seasons = seasons[seasons["app_date_day"].str.contains("Sun") == False]

seasons = seasons[seasons["app_date_day"].str.contains("Sat") == False]

The decision to do this was based on the fact that the only service setting which did not operate on weekends was General Practice, which accounted for the vast majority of appointments.

It also allowed for analysis by weekday.

For each seasonal visualization the **hue_order** was set to match that in the first visualization of changes to service settings across the full date range.

Additionally, bank holidays were dropped as General Practice does not operate on bank holidays.

For example: **seasons_aug =**
seasons_aug[seasons_aug["app_date_day"].str.contains("Mon/30") == False]

The ticks on the x-axis were set to match Mondays (or Tuesdays following bank holidays) using **.set(xticks=[0, 5, 10, 15, 20])**

Gridlines were inserted using plt.grid() and a reference line was inserted to show the maximum number of appointments the NHS can accommodate per day as 1,200,000 **.refline(y = 1200000, color = 'red', lw = 3)**.

Finally, the figures were then saved using **plt.savefig('image_name.png')**

To begin **analysing the Twitter data** pandas and seaborn were imported, the figure size was set **sns.set(rc={'figure.figsize':(15, 12)})** , and its style **sns.set_style('white')** and maximum column width **pd.options.display.max_colwidth = 200** were also set.

The data was imported using **pd.read_csv()** and **.shape()**, **.dtypes()**, **.info()**, **.describe() and .head()** were run as before.

A new dataframe was created containing only the text in each tweet.

An empty list was created with the name ‘tags’. **tags = []**

The text in each tweet was looped through to create a list of values which contained the # symbol (a hashtag on a tweet.

```
for y in [x.split(' ') for x in tw_text['tweet_full_text'].values]:
  for z in y:
        if '#' in z:
        # Change to lowercase.
        tags.append(z.lower())
```

A Series was then created with the **value_count()** for each unique hashtag **tags_series = pd.Series(tags).value_counts()**

.head(30) was used to display the first 30 records.

This Series was converted into a dataframe to prepare for visualization using **pd.DataFrame()** and the index was reset using **.reset_index()**

The columns were renamed to ‘words’ and ‘count’ using **tags_data.rename(columns={'index':'words', 0:'count'}, inplace=True)**

A Seaborn barplot was created indicating records with a count >10 records.

```
tags_grp = tags_data.groupby([tags_data.index,'words','count'])
tags_data_fil = tags_grp.filter(lambda x: x['count'] > 10.)
```

This was only after removing the top 2 entries #healthcare and #health as they do not provide any information (assuming all tweets related to the National Health Service are related to healthcare/health) and the values distorted the chart. **tags_data_fil_excl = tags_data_fil.drop([0, 1])**

In order to being making recommendations, the appointments_regional.csv data was imported **pd.read_csv()**, and sense checked **.shape** and **.dtypes**

Appointment_month was converted to datatime using **pd.to_datetime()** and the date range was established **.agg(['min', 'max'])**

The data was filtered to only look at data from August 2021 onwards **ar[(ar['appointment_month'] >= '2021-08')]**

In order to establish whether the NHS should start to look at increasing staffing levels a new aggregated data set was created and grouped **groupby** and converted into a dataframe **pd.DataFrame()**

To establish number of appointments per month a further dataframe was created and **groupby** was used to group the appointment months by the **sum()** of appointments.

This was then sorted by appointment month for visualization of a trend over time **sort_values(ascending=True)**

A new variable was then created to show average appointments per day. The proposed method was **count_of_appointments/30**.

This was then rounded to remove decimal places **np.round(ar_df_fin['utilisation'], decimals = 0)**

.replace was used to convert numerical dates into strings (eg. **“2021-08”: “Aug”**), again for ease of visualization.

A seaborn lineplot was then created to show sum of count of monthly appointments, and a second lineplot was created to show average capacity utilization by month with a reference line showing

This method was flawed as the number of business days (the days on which the vast majority of appointments were booked) were significantly less than 30 in any month once weekends and bank holidays were excluded. This meant that the average produced was significantly lower than reality, and it also meant that it appeared that the NHS could accommodate all appointments.

A new dataframe was created from the national_category data set in order to visualize the actual number of appointments per day over the period. Weekends and bank holidays were excluded and there were 239 buisness days remaining. This showed the mean/average number of appointments per day **.describe()** to be 1,259,973 which is almost 60,000 appointments per day above capacity.

Another lineplot was created, with a reference line which demonstrates this.

Further lineplots using the same approach but with the appointments_regional data set to answer questions over the same period regarding changes in health professionals, changes in appointment status, changes in appointment status, changes in appointment mode, and changes in the length of time between booking and appointment. **Plt.layout=”constrained”** was used to correct formatting issues with the titles of the visualizations.

Finally, a boxplot was created from the national_category data to investigate spread of service settings over the period The General Practice data distorted this so a second boxplot was created excluding the GP data.

Appendix 2 - Outputs

Importing and exploring the data

Question 1: How many locations are there in the data set?

Number of locations : 106

Question 2: What are the five locations with the highest number of records?

Table 1

NHS North West London ICB - W2U3Z	**13007**
NHS Kent and Medway ICB - 91Q	12637
NHS Devon ICB - 15N	12526
NHS Hampshire and Isle Of Wight ICB - D9Y0V	12171
NHS North East London ICB - A3A8R	11837

Question 3: How many service settings, context types, national categories, and appointment statuses are there?

Number of service settings : 5
Number of context types : 3
Number of national categories : 18
Number of appointment status : 3

Analysing the data

Question 1: Between what dates were appointments scheduled?

Table 2

Date range in actual_duration:	
min	01/12/2021
max	30/06/2022

Table 3

Date range in national_categories:	
min	01/08/2021
max	30/06/2022

Question 2: Which service setting was the most popular from 1 January to 1 June 2022?

Table 4

Service Setting	Appointments
General Practice	248205699
Unmapped	10359163
Primary Care Network	5887086
Other	4962338
Extended Access Provision	1975589

The most popular service setting for NHS North West London from 1 January Service settings which reported most appointments from 1 January to 1 June 2022:to 1 June 2022:

Table 5

Service Setting	Appointments
General Practice	4804239
Unmapped	391106
Other	152897
Primary Care Network	109840
Extended Access Provision	98159

Question 3: Which month had the highest number of appointments?

Table 6

Appointment Month	Appointments
01/11/2021	30405070
01/10/2021	30303834
01/03/2022	29595038
01/09/2021	28522501
01/05/2022	27495508
01/06/2022	25828078
01/01/2022	25635474
01/02/2022	25355260
01/12/2021	25140776
01/04/2022	23913060
01/08/2021	23852171

Question 4: What was the total number of records per month?

Table 7

Appointment Month	Records
01/03/2022	82822
01/11/2021	77652
01/05/2022	77425
01/09/2021	74922
01/06/2022	74168
01/10/2021	74078
01/12/2021	72651
01/01/2022	71896
01/02/2022	71769
01/04/2022	70012
01/08/2021	69999
Total:	817394

Visualising and identifying initial trends

Objective 1

Figure 1 Service settings over time

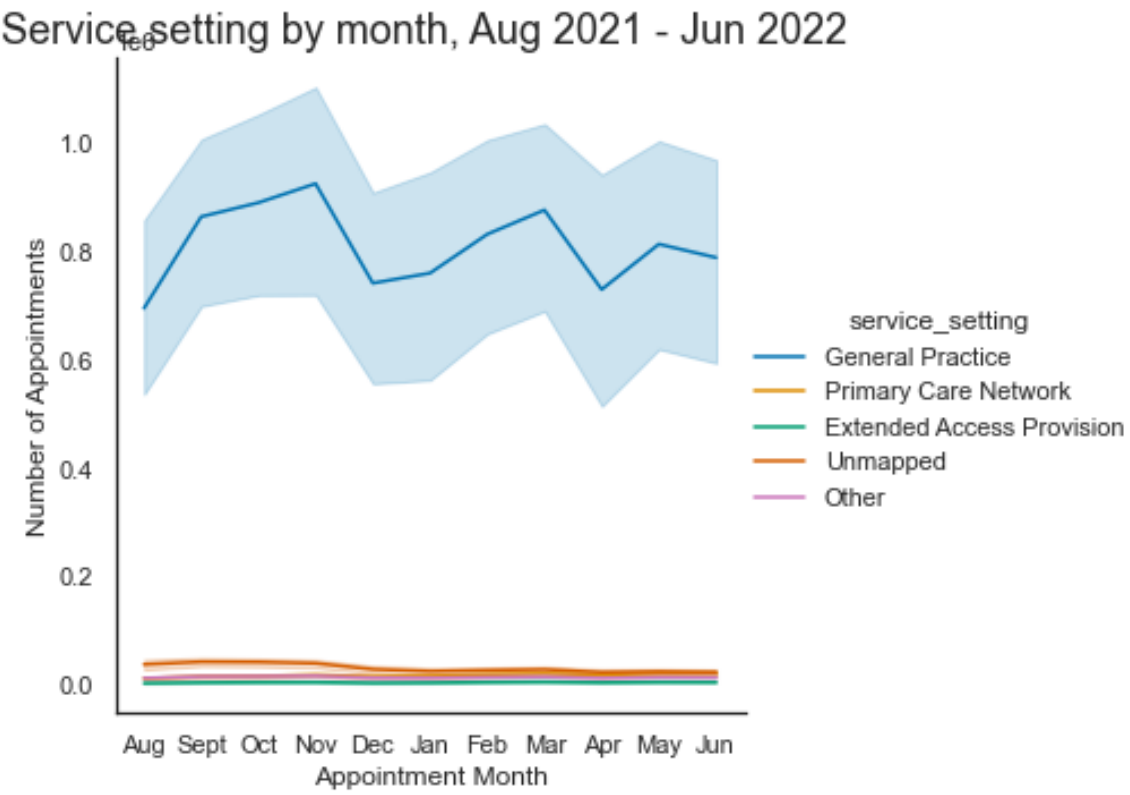


Figure 2 Context types over time

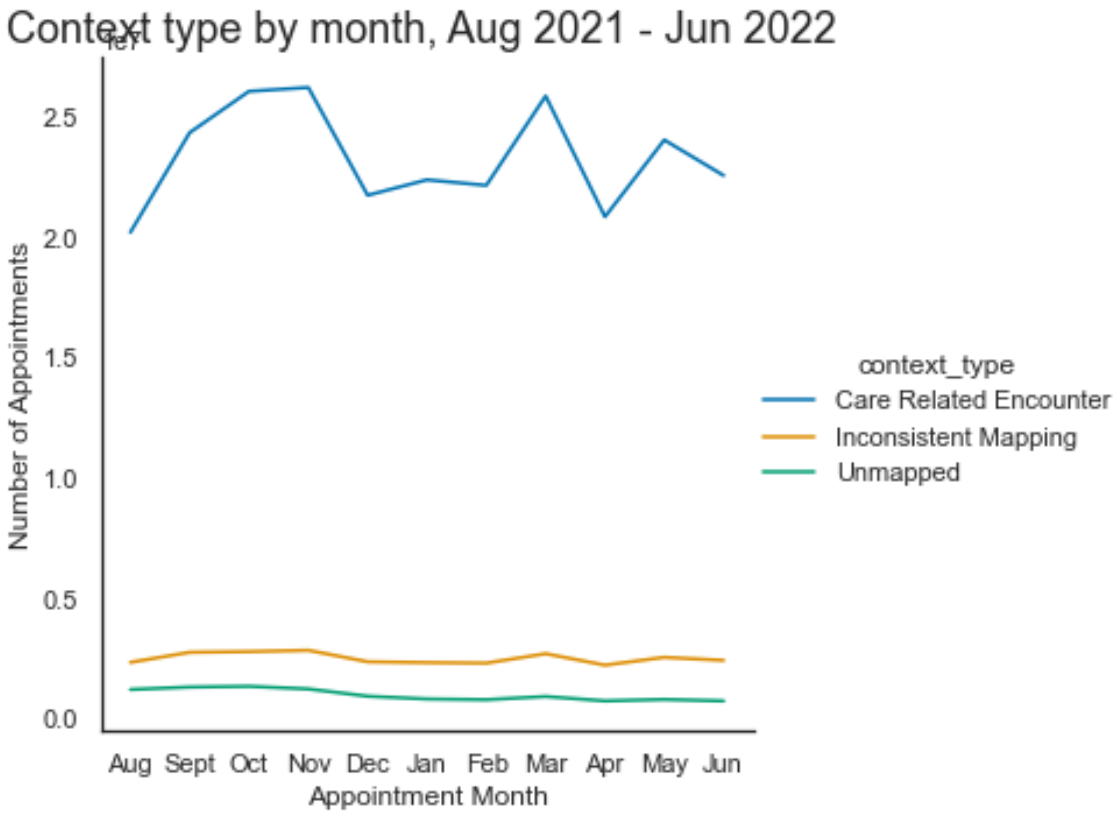
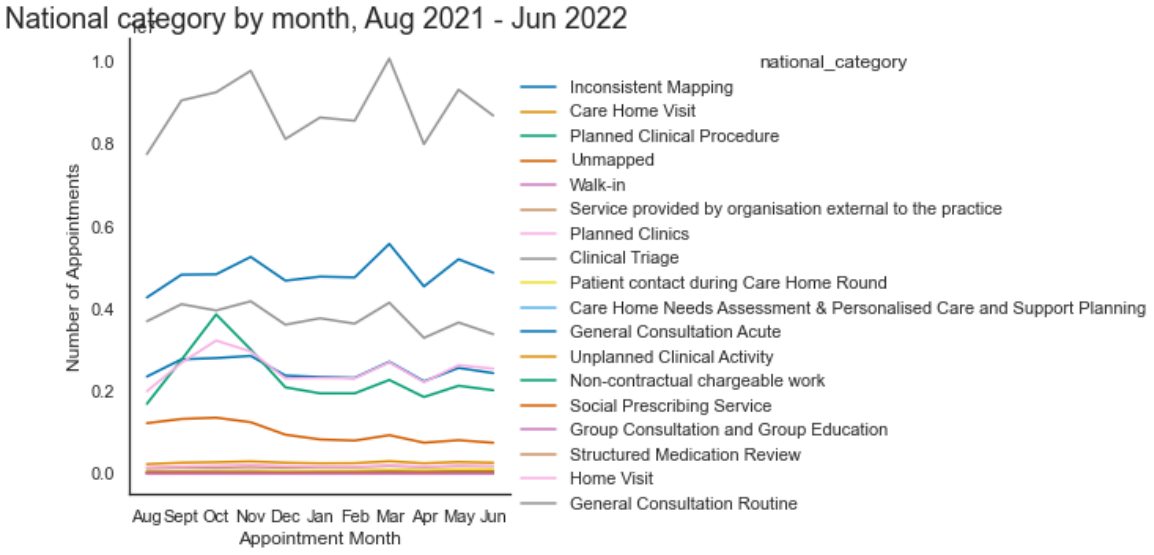


Figure 3 National categories over time



Objective 2

Create four visualisations indicating the number of appointments for service setting per season. The seasons are summer (August 2021), autumn (October 2021), winter (January 2022), and spring (April 2022).

Figure 4 Summer (August 2021):

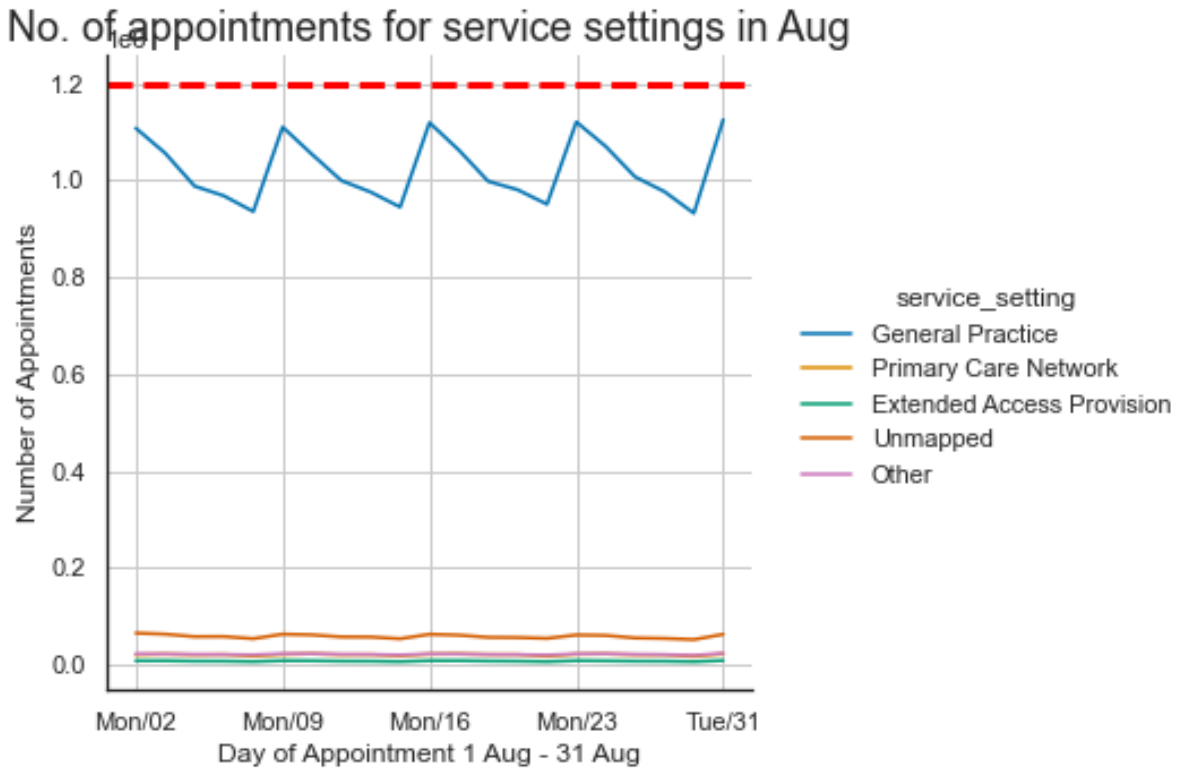


Figure 5 Autumn (October 2021):

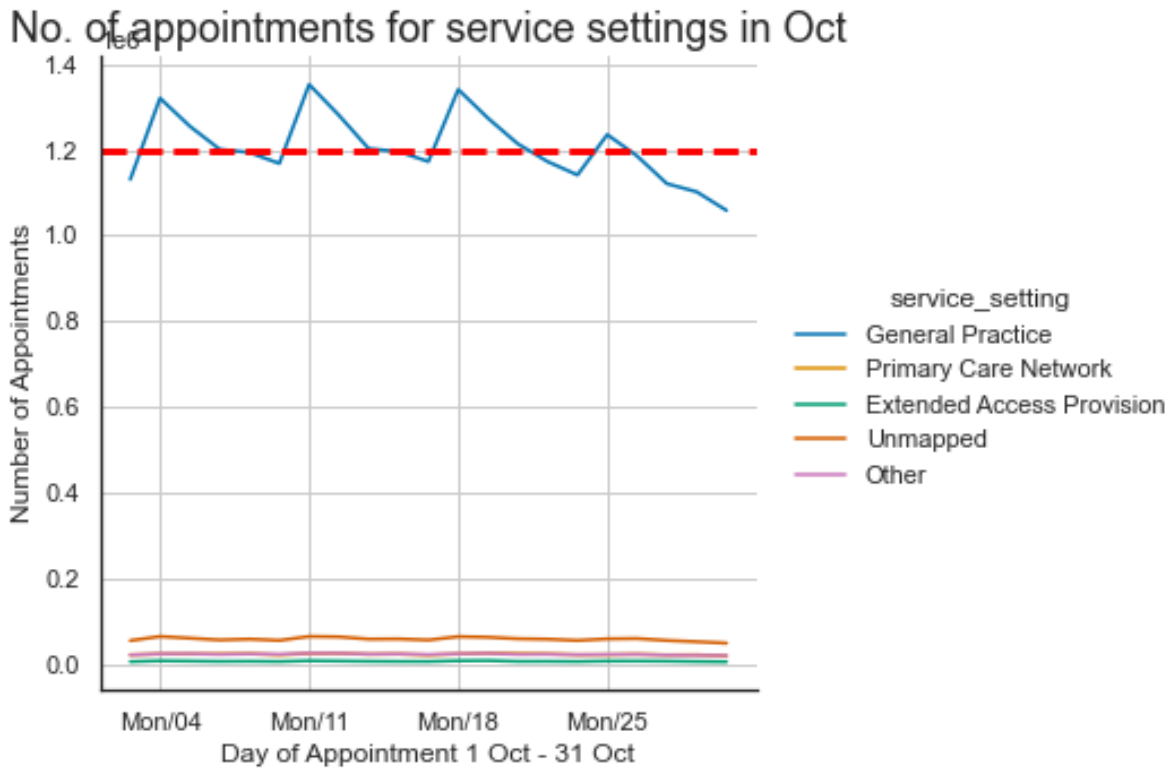


Figure 6 Winter (January 2022):

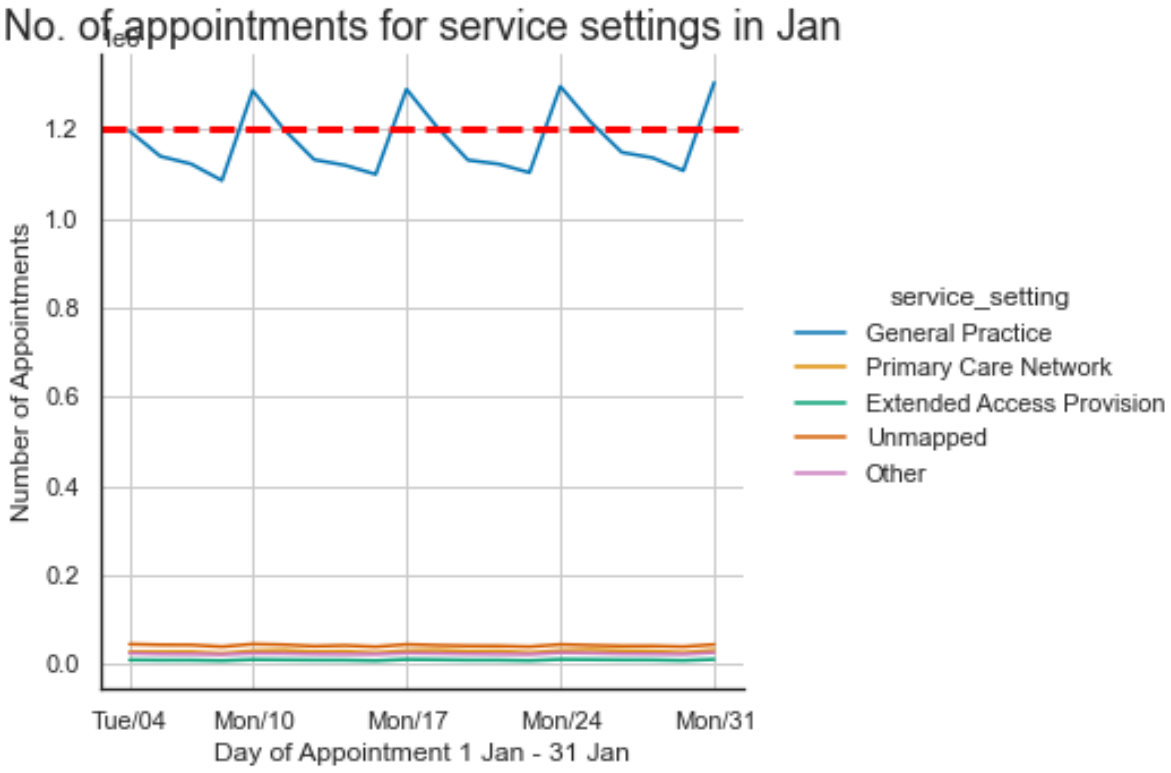
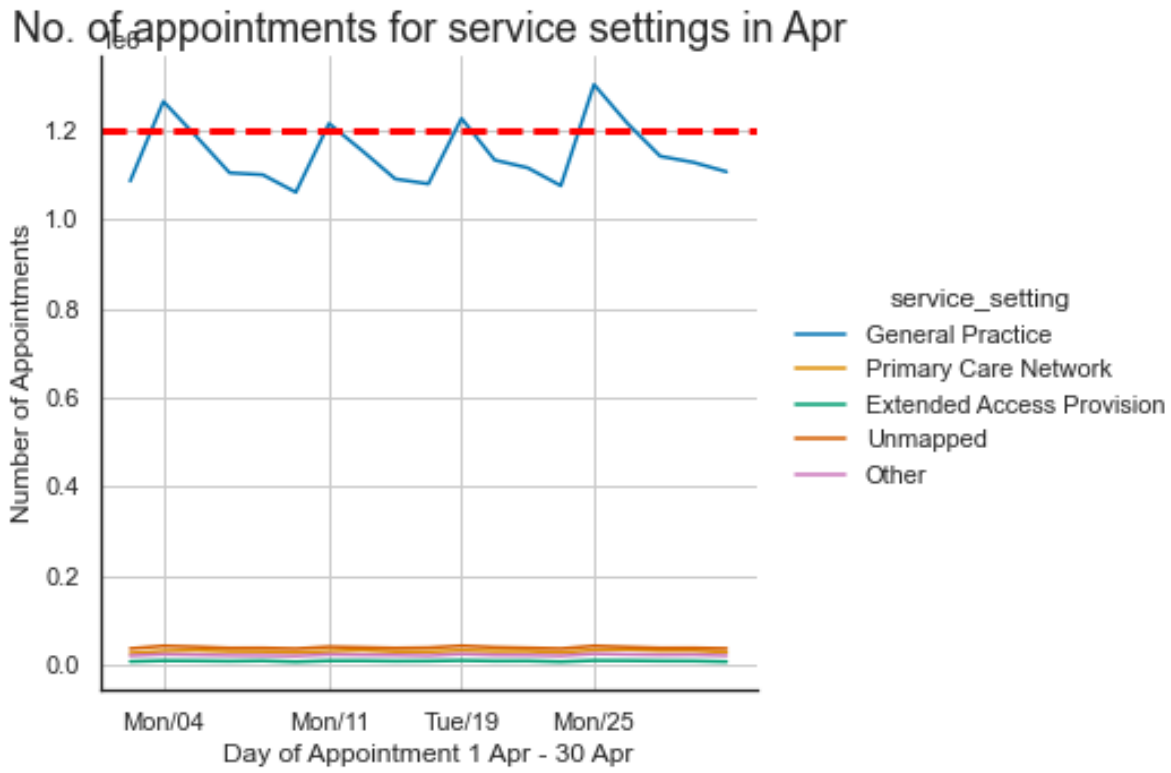


Figure 7 Spring (April 2022):



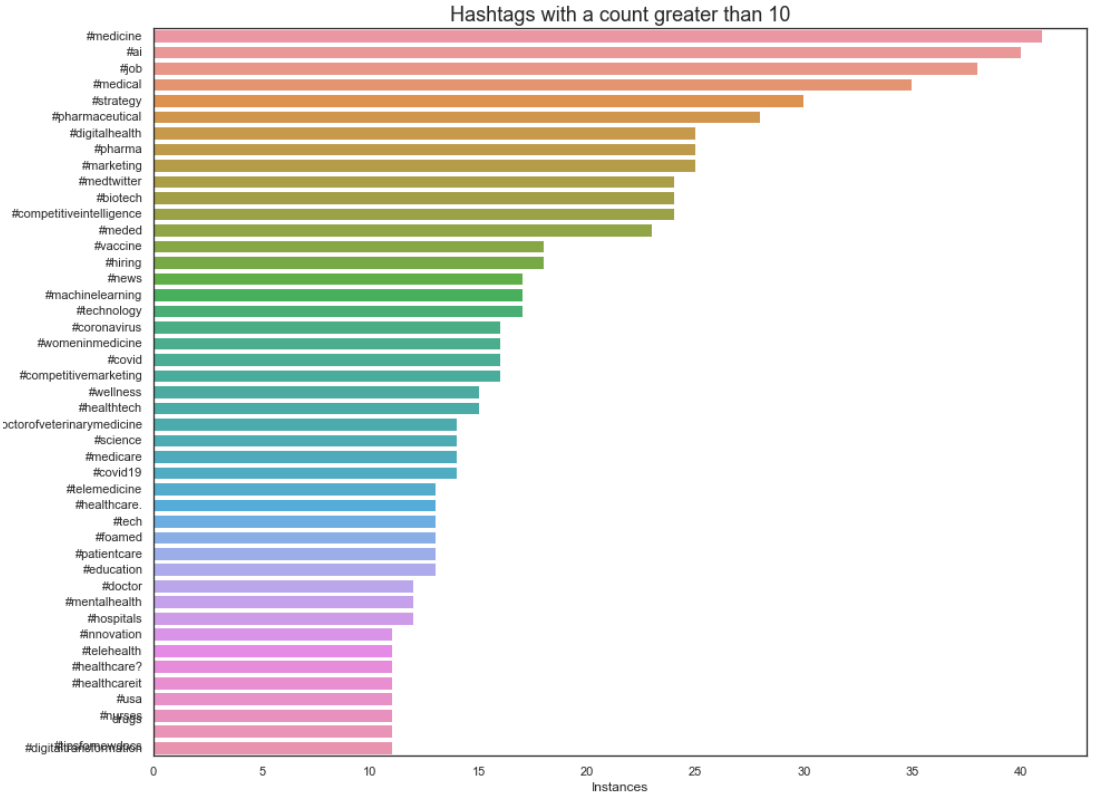
Analysing the Twitter data

Table 8 - 30 most frequently used hashtags

Hashtag	Count
#healthcare	716
#health	80
#medicine	41
#ai	40
#job	38
#medical	35
#strategy	30
#pharmaceutical	28
#digitalhealth	25
#pharma	25
#marketing	25
#medtwitter	24
#biotech	24

#competitiveintelligence	24
#meded	23
#vaccine	18
#hiring	18
#news	17
#machinelearning	17
#technology	17
#coronavirus	16
#womeninmedicine	16
#covid	16
#competitivemarketing	16
#wellness	15
#healthtech	15
#doctorofveterinarymedicine	14
#science	14
#medicare	14
#covid19	14
#telemedicine	13

Figure 8 Hashtags with a count greater than 10



Explore the tweet_retweet_count and tweet_favourite_count columns with the value_counts() function. Do yo think it is useful to look at these columns in more detail?

How will your team utilise tweets to provide feedback to the stakeholders?

Whether the tweets add value to the overall project?

How the NHS can utilise tweets?

Making recommendations

Question 1: Should the NHS start looking at increasing staff levels?

Figure 9 – sum count of monthly appointments

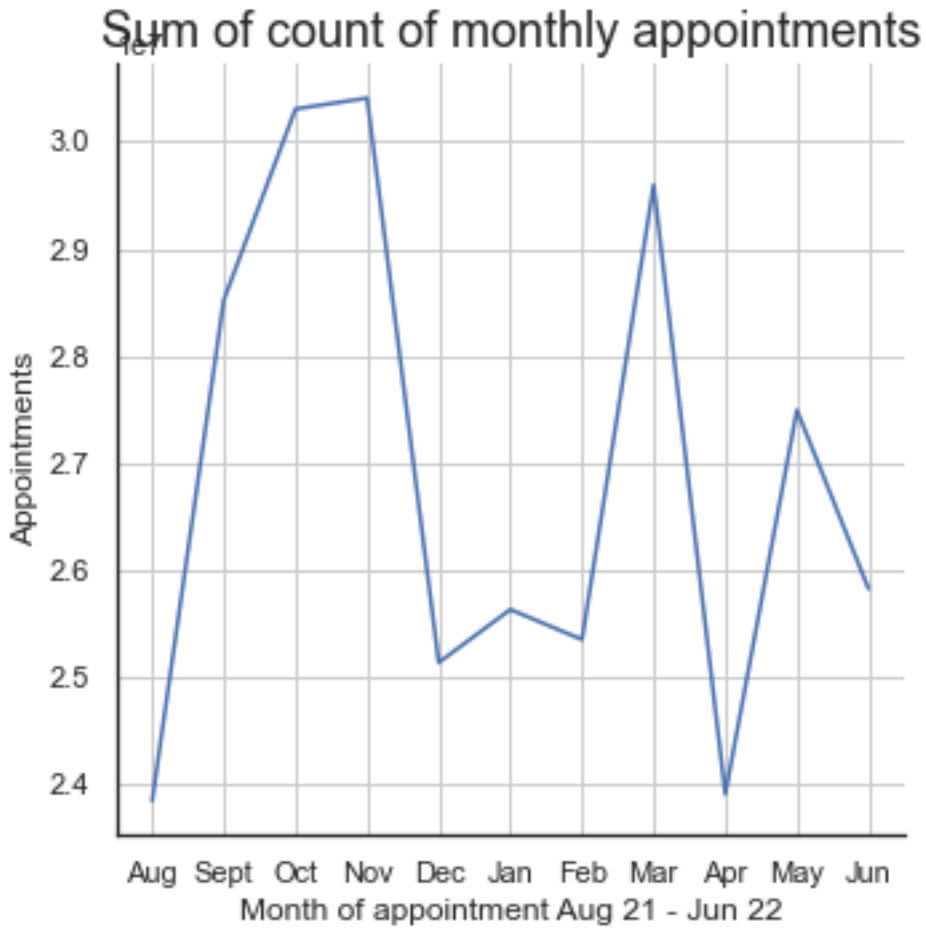


Figure 10 Daily capacity utilisation by month

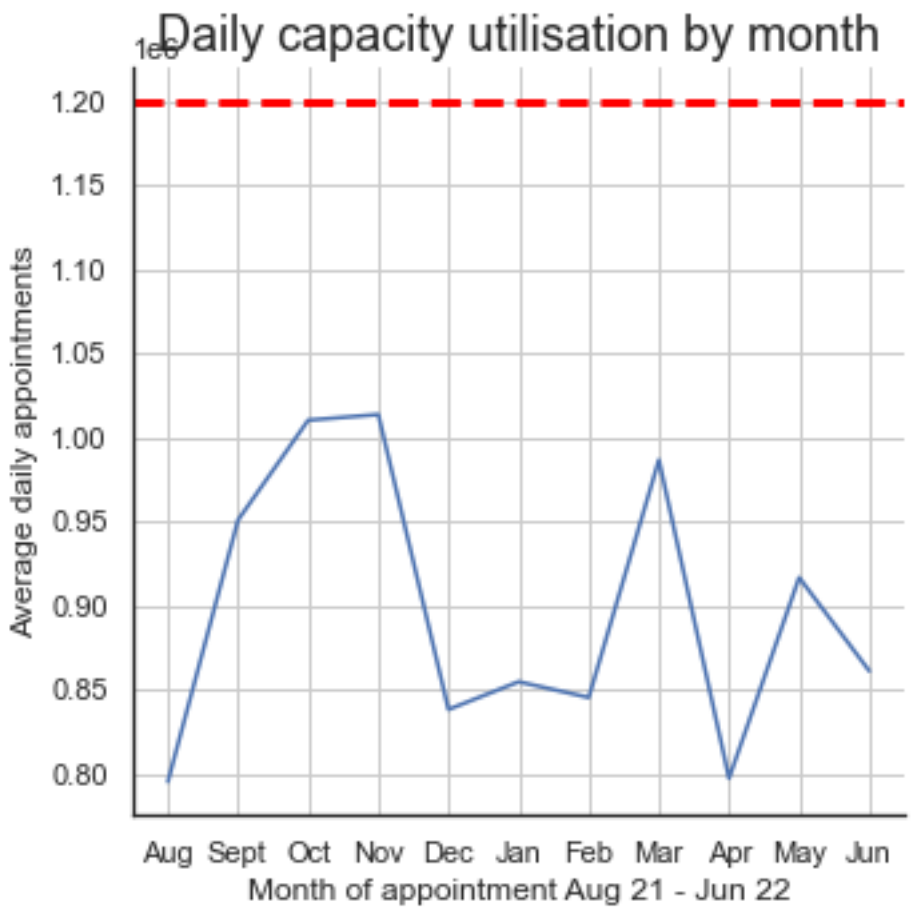
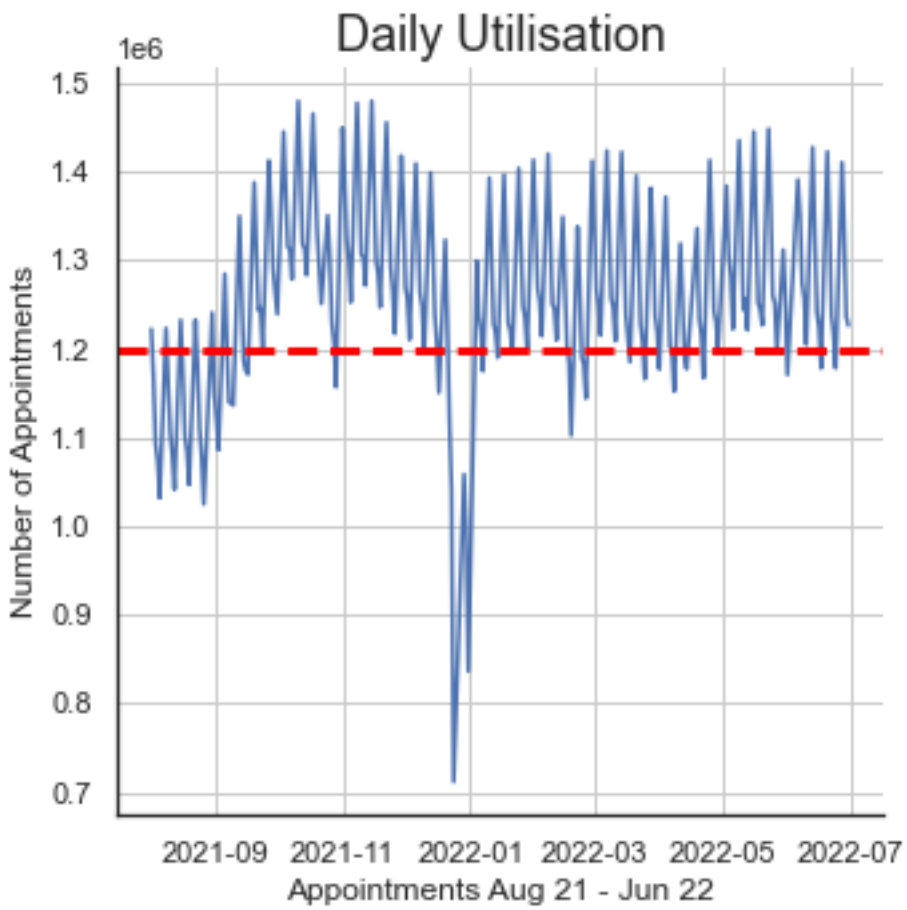


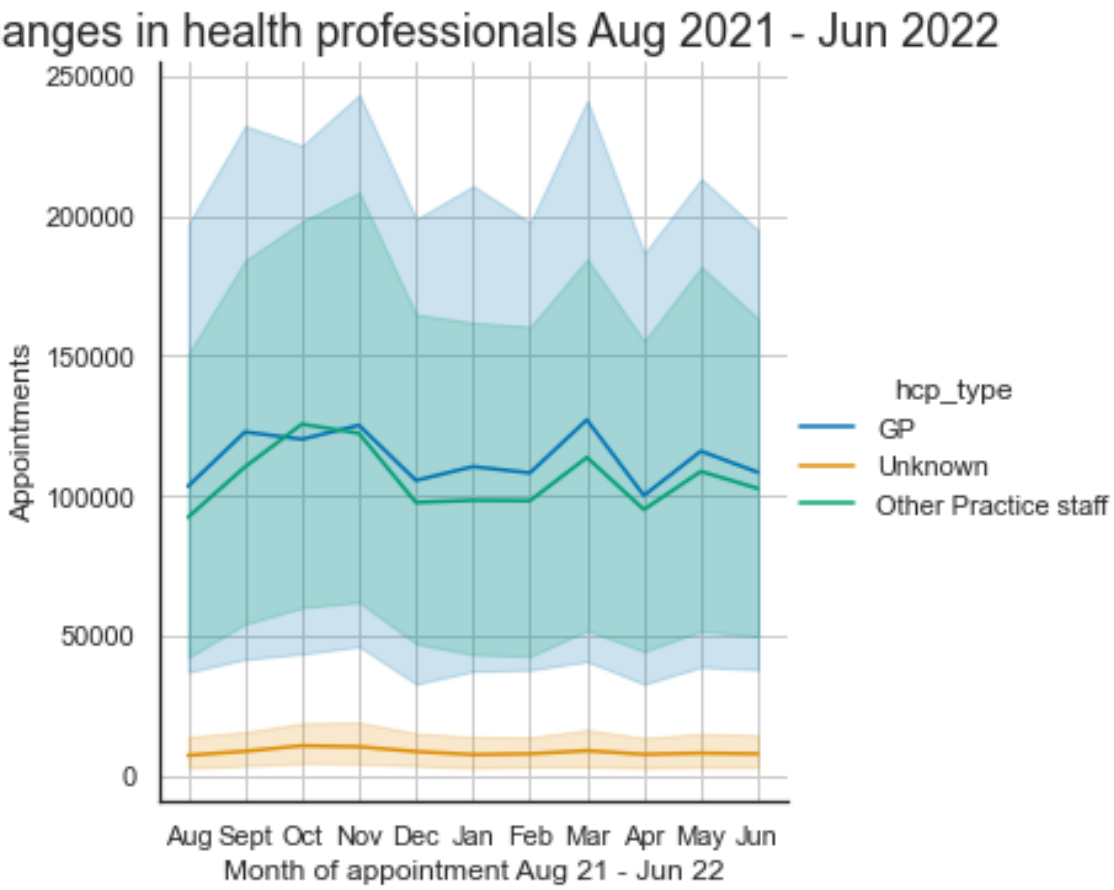
Figure 11 – Actual daily utilisation



Average number of appointments per working day: 1,259,973

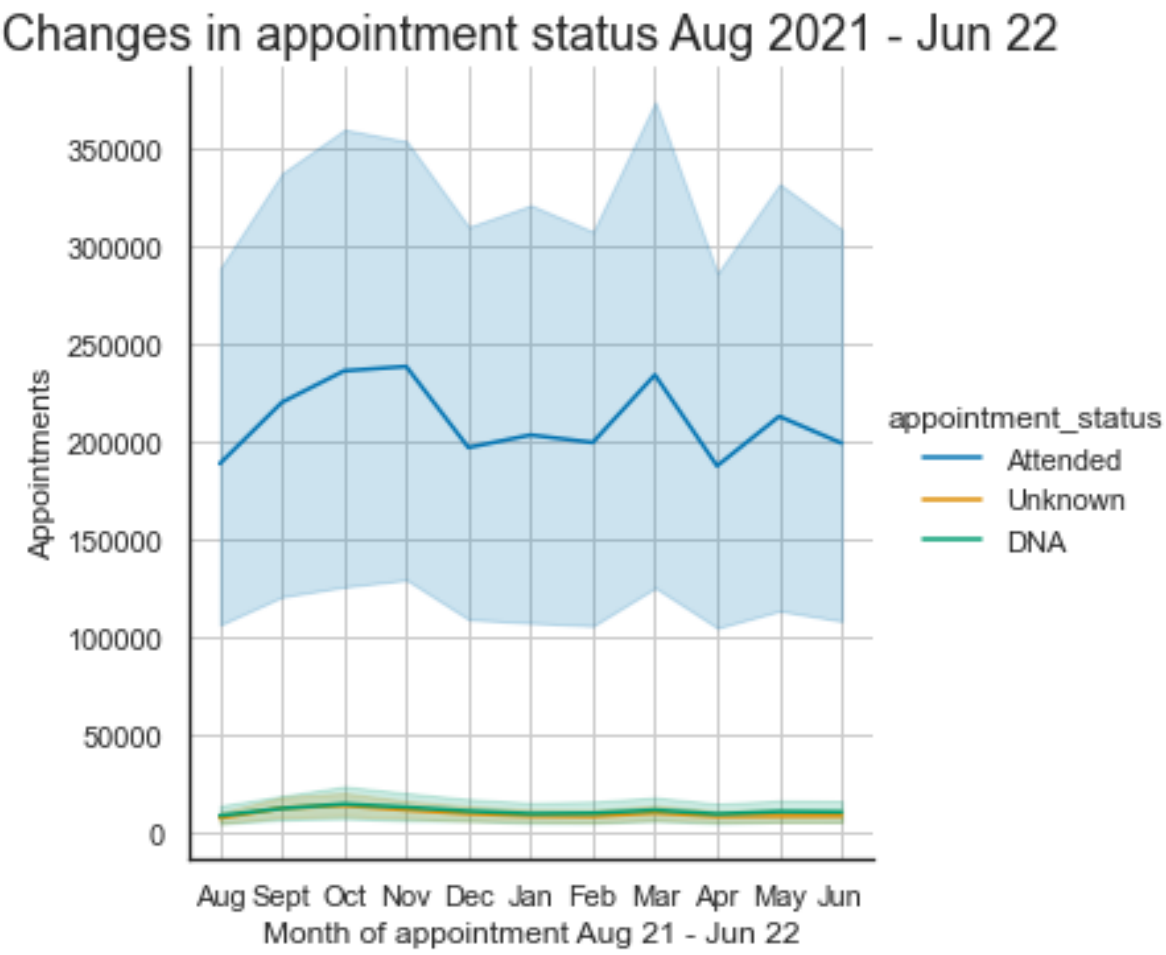
Question 2: How do the healthcare professional types differ over time?

Figure 12 Changes in health professional over time



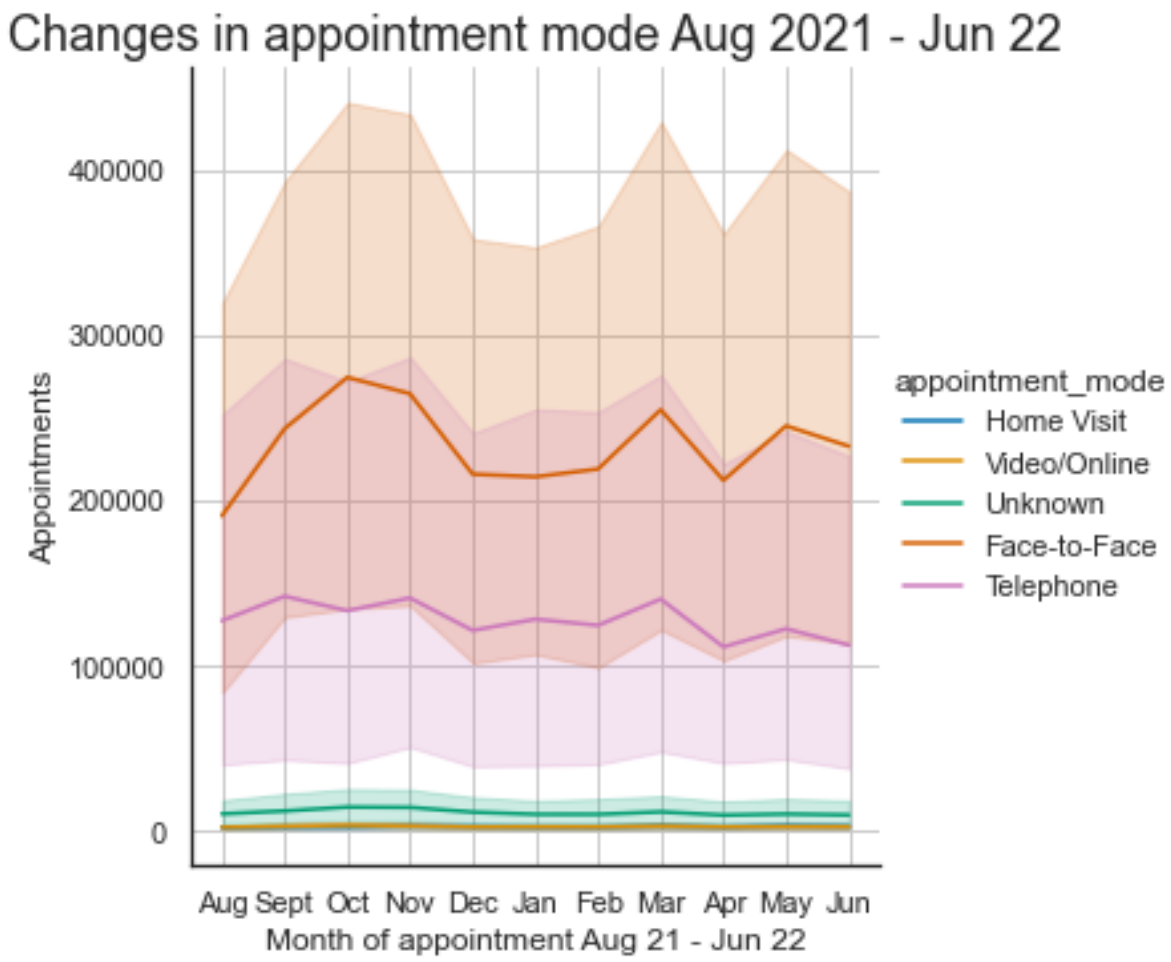
Question 3: Are there significant changes in whether or not visits are attended?

Figure 13 Changes in appointment status



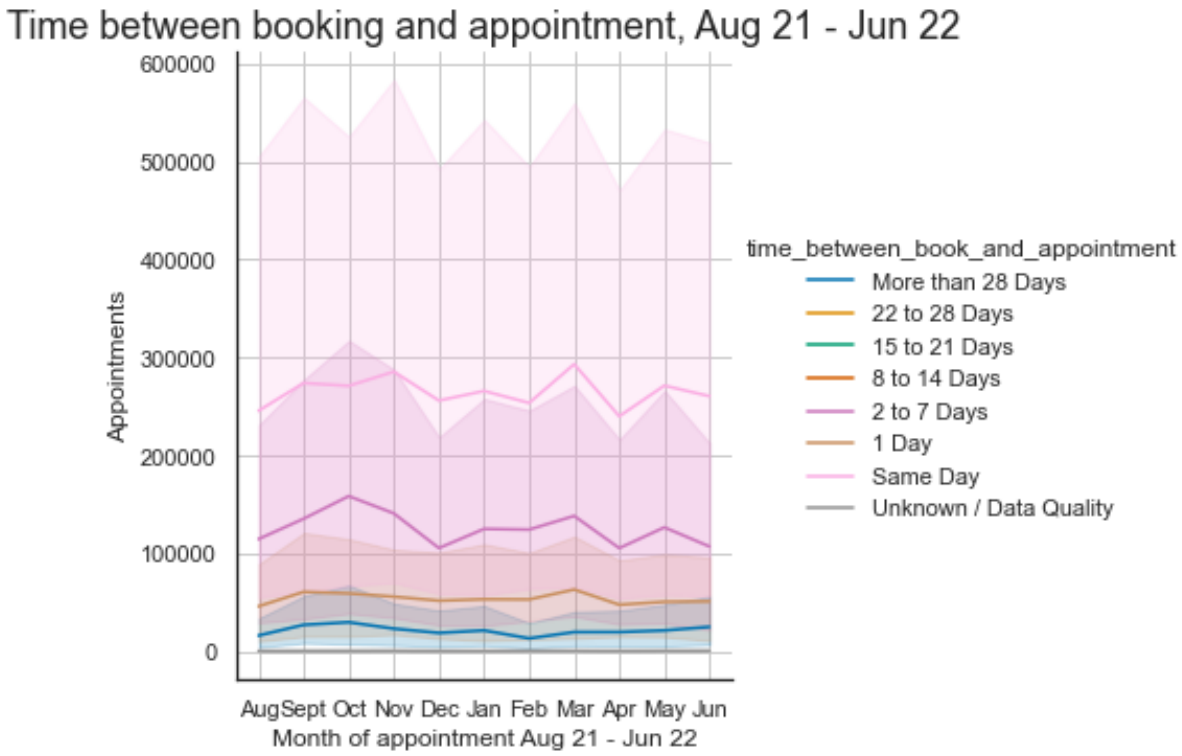
Question 4: Are there changes in terms of appointment type and the busiest months?

Figure 14 Changes in appointment type



Question 5: Are there any trends in time between booking an appointment?

Figure 15 Changes in time between booking and appointment



Question 6: How do the spread of service settings compare?

Figure 16 Spread of service settings

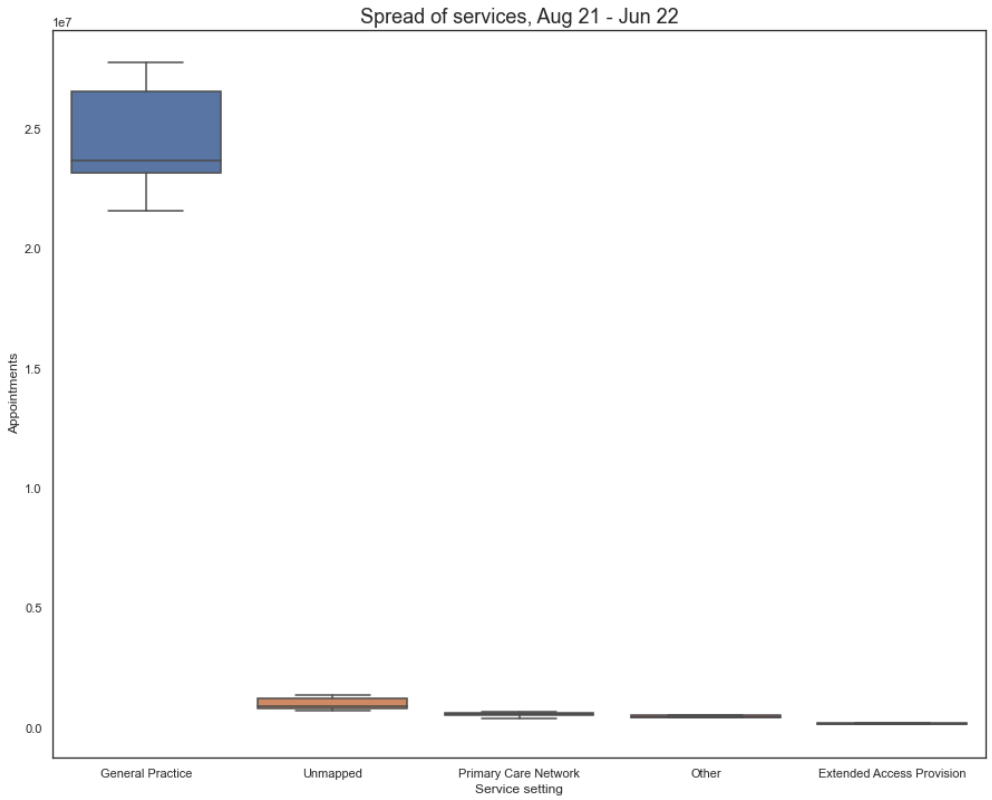


Figure 17 Spread of service setting excluding General Practice