Course 2, Assignment
NHS Missed Appointments
Ali Fazel

Part 1 – Context

Missed appointments in the NHS end up costing the system, but these are avoidable. This project is part of a wider study to better understand why appointments are missed in order to decide how best to deal with this problem.

The aim of this project is to gain deeper insights into the efficiency and resiliency of the NHS. The key focusses include:

- understanding if there are adequate staffing levels,
- gaining a better picture of the capacity in the networks,
- deciphering what the utilisation of resources is.

Three datasets were provided to undertake this analysis:

1. Actual Duration – ad
2. Appointment Regional – ar
3. National Categories – nc

Part 2 – Analytical Approach

Using the value_counts() function it was able to determine that there are 106 locations in the dataset. F-string interpolation was used to then present a full sentence answer with this detail. Following this, sort_values() was used to find the most popular locations based on how many records were listed, 'NHS North West London ICB - W2U3Z' being the highest, with 13007 records. The top five locations were returned by using the head() function and specifying the [5] after setting 'ascending=False' in order to get highest records first.[1]

Taking a deeper dive into the 'North West London ICB' location, it was gathered that General Practice was the most popular service setting here. In order to gather this data, firstly a subset data frame was created with only the relevant columns using the loc[] function. Then conditions were added to specify the relevant dates (1 January – 1 June, 2022) using the greater than (>), less than (<), as well as the location in question using the '==' operator. Finally, value_counts() provided a list of each service setting and the corresponding number of records.[2]

In order to work out between which dates appointments had been scheduled, it was firstly necessary for all date related data to be in the datetime format. Therefore, the 'appointment_date' columns in both 'ad' and 'nc' were changed using to_datetime() from the Pandas library. Now it was possible to ascertain the minimum and maximum dates using the min() and max() functions. This would give us the range of dates in both data sets. In 'ad' appointments were scheduled between 1 December 2021 and 30 June 2022. In 'nc' they were scheduled between 1 Aug 2021 and 30 June 2022.

---

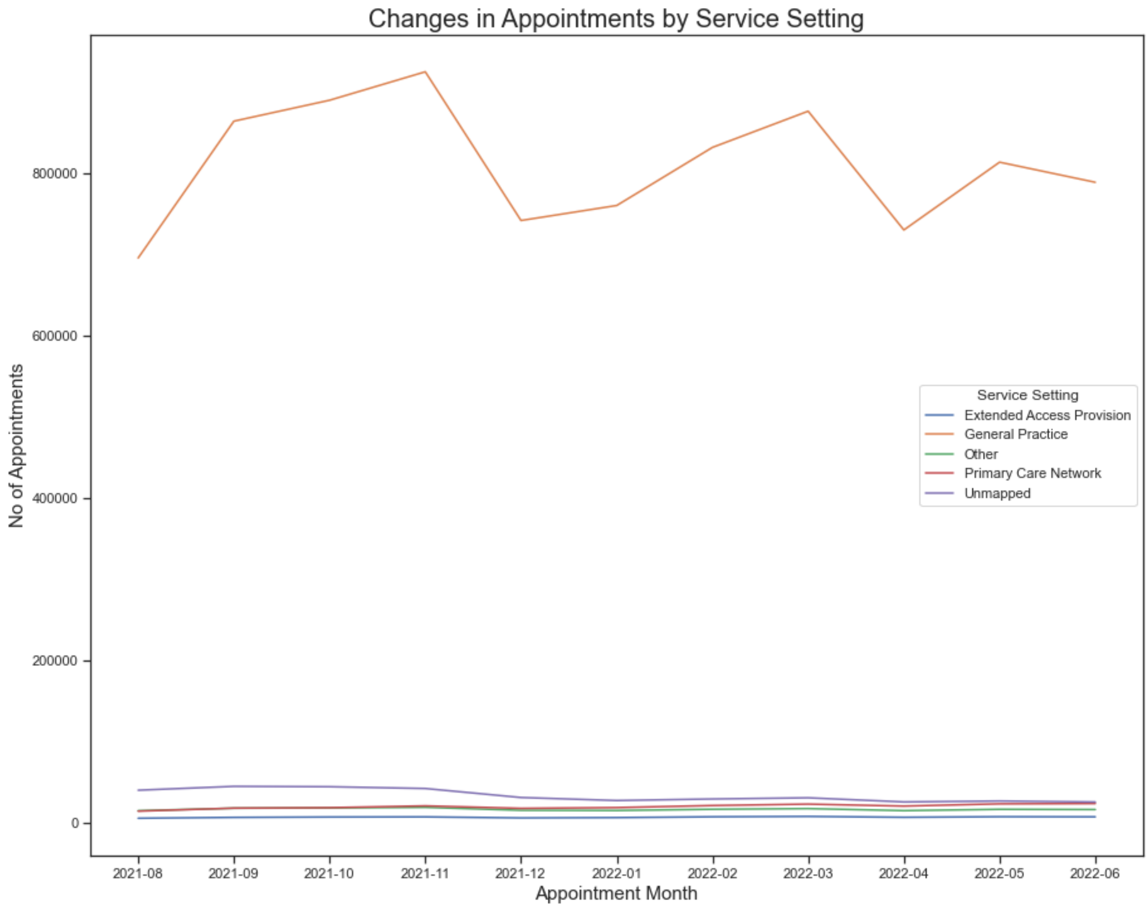
[1] See Appendix – Fig 1
[2] See Appendix – Fig 2

The month with the highest number of appointments was November 2021, with 30.4 million appointments. To get this data, firstly, 'appointment_month' in the 'ar' ('appointment regional') dataset was changed to datetime format using to_datetime(). The groupby() function was applied to the data frame to group together the month of the appointment and then the count of appointments was added together using the sum() function. The sort_values() helped to organise the data so that the highest counts appeared first. The reset_index() function was finally added to the code to create a new index for the data frame.[3]

The total number of records per month was calculated by grouping together the 'appointment_month' column in the 'ar' data frame and then applying the size() function, which counts the number of elements across a given axis. The columns were then renamed to include a 'records' column name. Overall, this showed that March 2020 had the highest number of records and May 2020, the lowest.

Part 3 – Visualisations and Insights

In order to visualise the number of appointments per month for each service setting, a data frame was firstly created which grouped 'appointment_month' and 'service_setting' followed by the sum() of the 'count_of_appointments'. This new data frame was then used to create a line plot using the Seaborn library, with month on the x-axis and count of appointments on



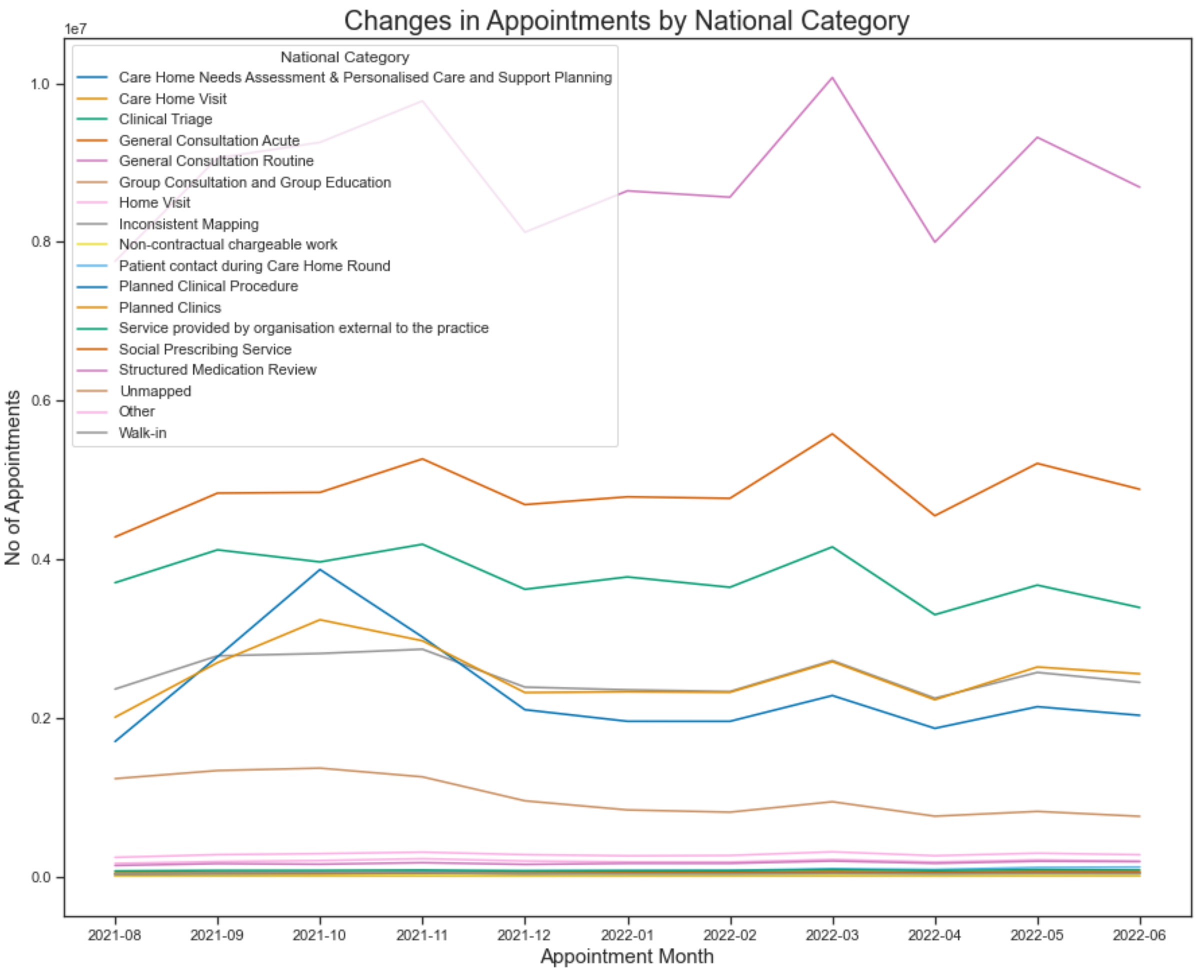
[3] See Appendix – Fig 3

the y-axis. Finally, 'service_setting' was added hue in the parameters of the lineplot() function.

Evidently, across all months, 'General Practice' appointments are by far the most popular, peaking in November 2021 and again in March 2022. There is a steady increase in GP appointments from August until November and this is symbolic for the annual winter pressures that the NHS suffers (Guardian, 2021).

To visualise the number of appointments per month for each context type, the same approach was followed as outlined above, expect the newly formed data frame was grouped by 'appointment_month' and 'context_type'. 'Care Related Encounters' were the most popular context type for appointments, once again following similar trends as we saw in service settings: a steady increase through the winter months, peaking in November 2021 and again in April 2022.

The same method was followed to visualise the number of appointments per month for each national category. As there are several lines on this line plot, the colourblind palette was applied as a parameter to distinguish between the different national categories. The set_xlabel() and set_ylabel() functions were applied to provide clear axis labels and the 'title' parameter was called in the set() function to clarify what the graph was showing. 'General Consultation Routine' was the most popular national category across all months, followed by

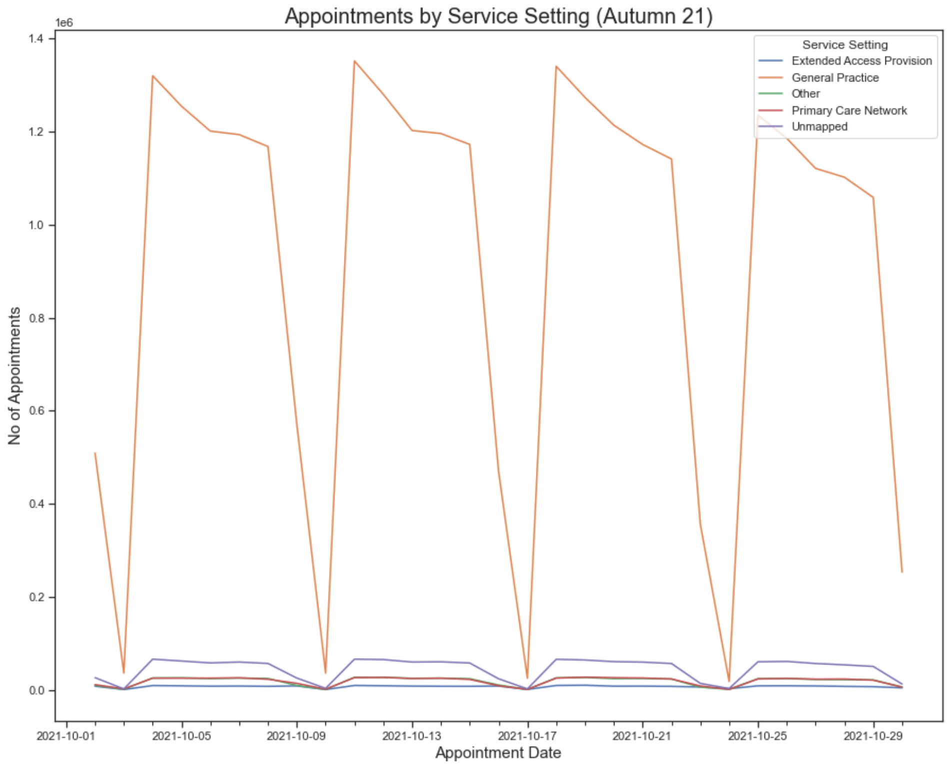


'General Consultation Acute' and 'Clinical Triage', all of which appear to have a peak in March 2022 before dropping again to relatively normal levels in April.

Notable is 'Planned Clinical Procedures' which sharply increase through the autumn overtaking both 'Planned Clinics' and 'Inconsistent Mapping' to become the fourth most popular national category by count of appointments, before falling back down to sixth in December. The visualisation could be further improved by grouping together all the lines concentrated at the bottom of the line graph as an 'Other' category. This would increase the readability and focus on the highest performing national categories only as well as reducing the size of the legend which is too large.

Turning attention to the number of appointments for each service setting by season, a new data frame was firstly created which grouped the 'nc' dataset by 'appointment_date', 'appointment_month' and 'service_setting', before applying the sum() function to add the total number of appointments in those groups. Using the loc[] function, a new data frame was created specifying each season using the < and > operators on the 'appointment_date' series. This was then used as the data to create the line plots for summer 2021, autumn 2021, winter 2022 and spring 2022. The legend location was adjusted where necessary by specifying 'loc=' in the legend() function. Additional minor ticks were added on the x-axis to make it easier for the viewer to distinguish between the dates of the month, using the set_xticks() function.

Across all four seasons, it is instantly noticeable that General Practice is by far the most popular service setting, for reasons already explained above. As expected, we see drops four times in each season, representing weekends where there are no, or close to no, appointment across all service settings. It is also worth noting that the greatest number of appointments for GP settings was seen across Autumn 2021, on 11 October where appointments peaked at almost 1.4 million.

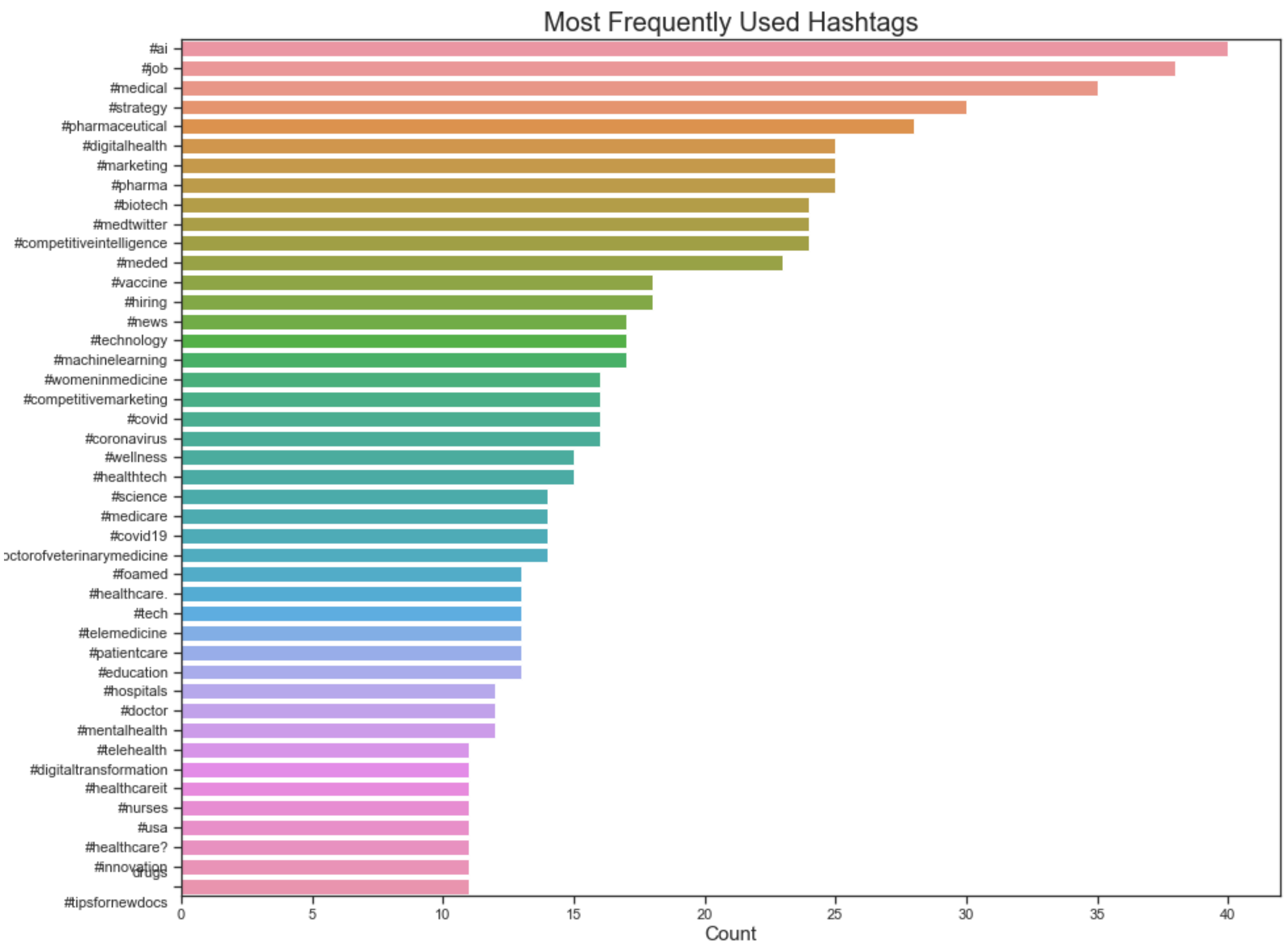


Due to the differences in appointments in service setting, the visualisation could be improved by plotting General Practice on a graph of its own – this way it will be easier to distinguish between the other types of service settings too.

Understanding Twitter data

To understand the contents of the tweets, the text was extracted into a data frame and then the hashtags were identified and counted to understand the most popular topics. This was done using a for loop which picked up '#'s and appended them to the list 'tags' that was already created.

With plotting the most frequently used hashtags, the outliers were firstly identified by working out the interquartile range and then using this to work out the upper and lower bounds, 40.5 and -3.5, respectively. This meant that '#healthcare', '#medicine' and '#health' were all removed from the data set as these had a count that was greater than 40.5. What the bar graph showed was that '#ai', '#job' and '#medical' were the most frequently used (in that order) within the accepted limits.

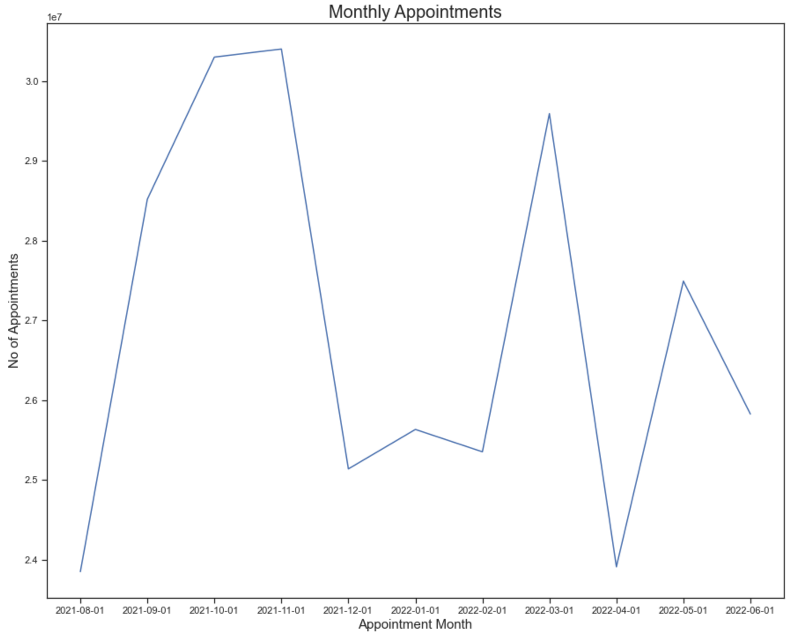


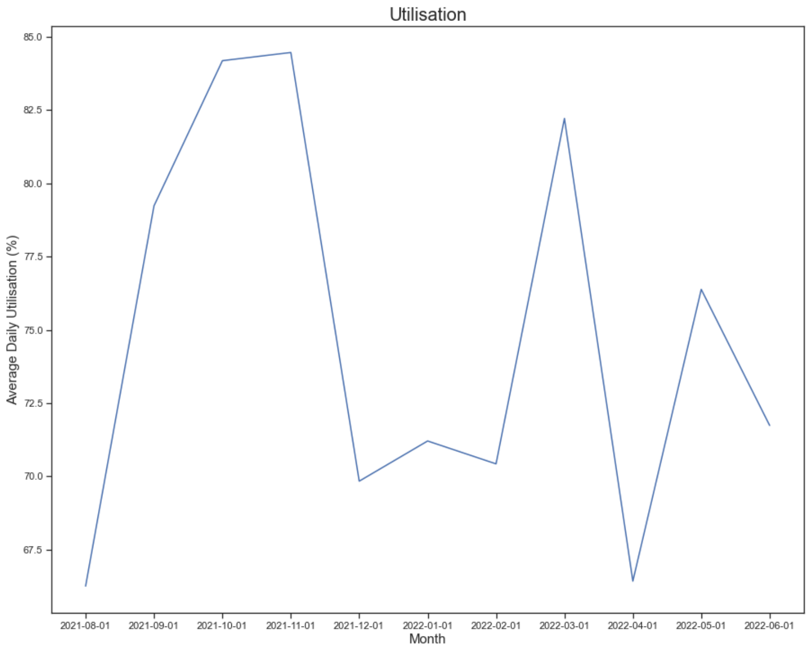
To improve the visualisations, some of the hashtags that are similar could be grouped together. For example, #covid, '#covid19' and #coronavirus are tweets related to the same issue and hence, if we are trying to gain better clarity on the popularity of topics being discussed, these can be merged. The same should apply to '#healthcare' and '#healthcare.'.

Part 4 – Patterns & Predictions

Should the NHS look at increasing staff levels?
An aggregated data frame was created to display the monthly count of appointments, the average daily utilisation of appointments as well as a percentage score for utilisation, using the ‘ar’ dataset. Using the information that the NHS can facilitate 1.2 million appointments a day, capacity utilisation could be calculated. The round() function was used here to keep the figures rounded to one decimal point ensuring the new data frame was readable. Graphs were then plotted to show the total monthly number of appointments as well as average daily utilisation per month as a figure and as a percentage.

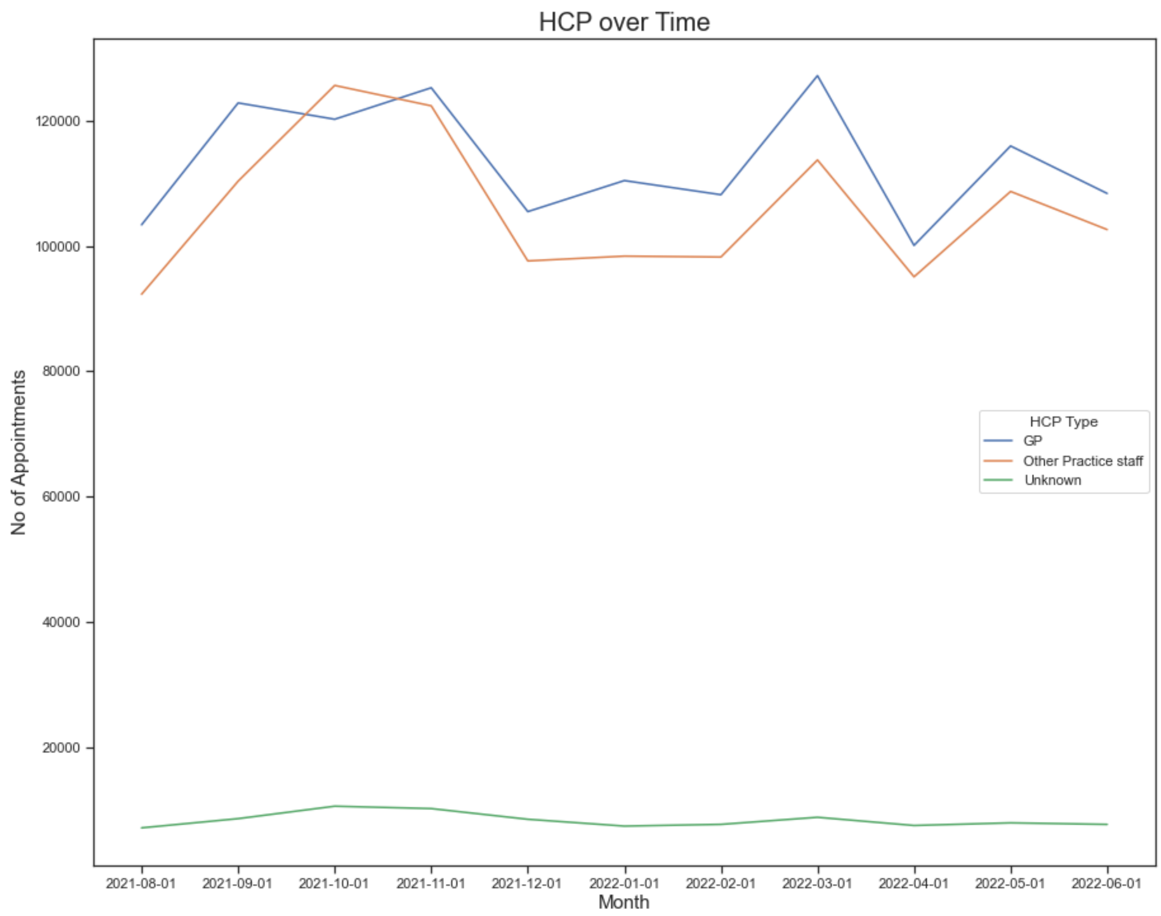
Naturally, the two graphs follow identical trends, when appointment numbers increase, utilisation percentage increases as more capacity is used.





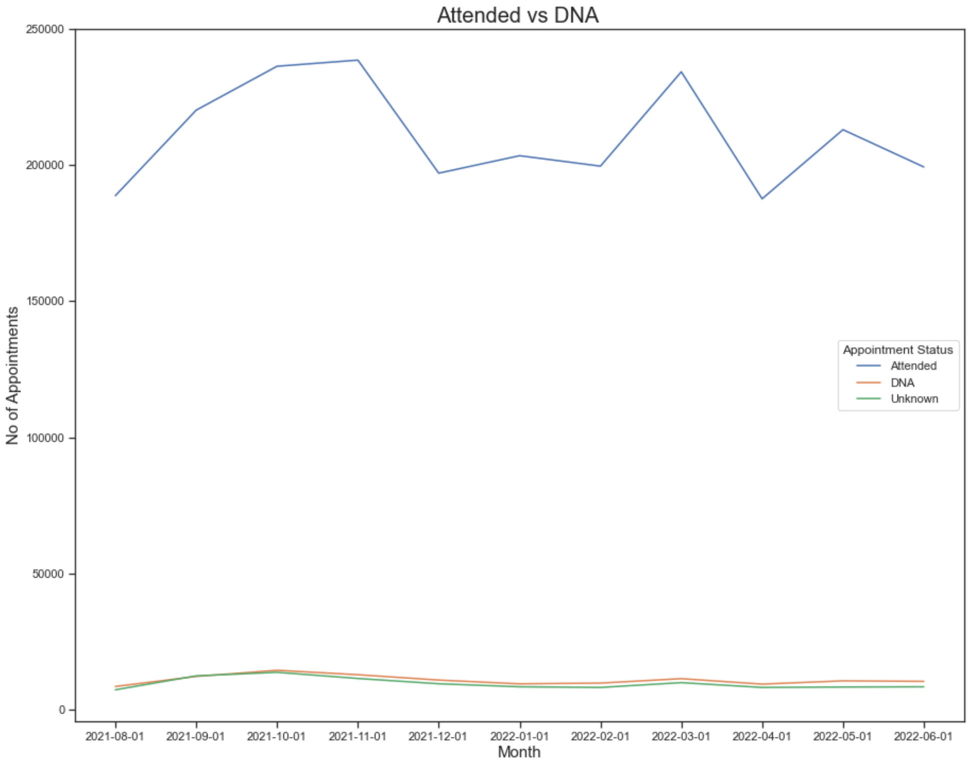
How do the healthcare professional types differ over time?

GPs are the most popular type of healthcare professional handling appointments. However, 'Other Practice Staff' also handle a significantly large number of appointments – this could include junior doctors, surgeons, clinical specialists, nurses and others. This would be standard practice outside of GP surgeries. In October 2021, 'Other Practice Staff' overtake GPs for just one month as the most popular health care professional type taking appointments. This line graph was plotted by adding 'hcp_type' to the hue parameter in the lineplot() function.



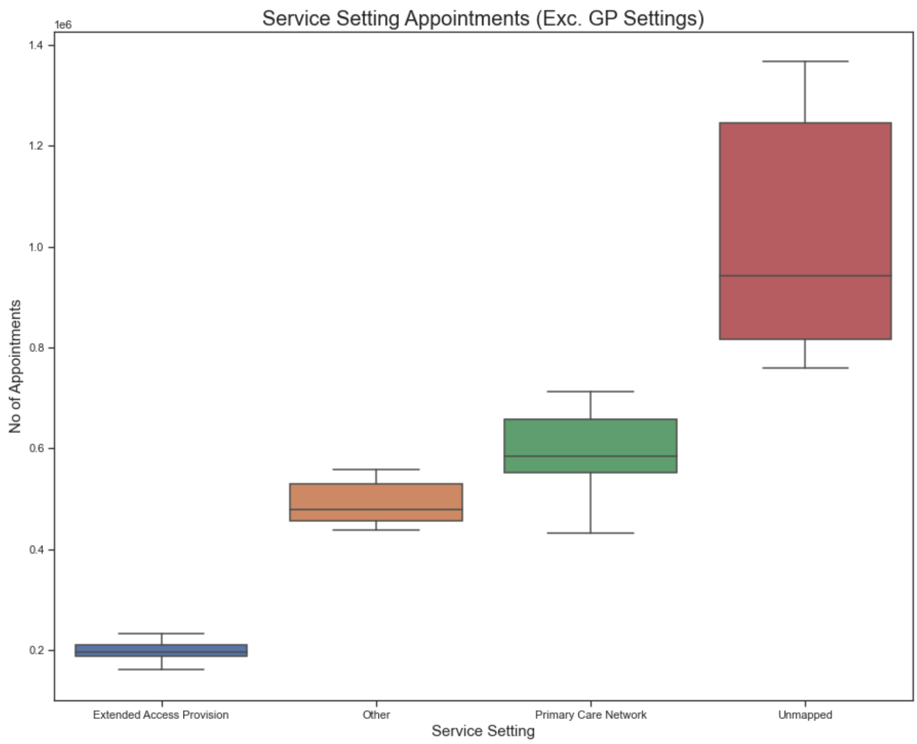
Are there significant changes in whether or not visits are attended?

There are no significant changes at all in the 'DNA's (Did Not Attend) – these stay relatively low and consistent through from August 2021 to June 2022, with a very subtle increase in October 2021. This could be attributed to the fact that there are more appointments being booked and so naturally the DNA figures will increase too. There are some notable changes in the attended appointments already delineated above – with rises through the autumn peaking in November 2021 and again in March 2022.



How do the various service settings compare?

As expected, General Practice is far greater in the number of appointments than any other service setting. Therefore, 'General Practice' was excluded from the boxplot and the remaining four service settings were plotted. This showed that 'Unmapped' was the highest, with its data most widely spread, followed by 'Primary Care Network' and 'Other' which exist in similar appointment numbers, and 'Extended Access Provision' being the least popular, with its data being highly concentrated.



Conclusions and recommendations:

'General Practice' appointments are by far the most popular. As demand for healthcare increases, people will reach out to their GPs and as such we see this steady increase through the winter; for most people, their GP is their first port of call and they will seek out an appointment there before exploring other options.

An increase in staff might help to push capacity utilisation to 100%, perhaps if there were more members of staff present, more appointments could be completed and so utilisation increases. However, further analysis of data should be undertaken to understand if the reason for capacity utilisation not exceeding 84% is due to staffing levels specifically, or whether other factors contribute to this.

To make concrete conclusions around staffing levels, more staff-related data is required, as currently these judgements are being reached based on capacity utilisation data only.

The stakeholders may want to consider if they are truly utilising all modes of appointments to maximise efficiency and capacity utilisation. The data shows low video/online appointments taking place across the NHS, relative to other appointment modes. Perhaps there is an opportunity to capitalise on this where appropriate to maximise capacity and increase appointment numbers.

The NHS should consider sentiment analysis of Twitter data to gain greater insights into patient experience, this might help to feed into decisions around staffing and utilisation.

Bibliography

The Secret Consultant, 2021, 'NHS Winter Pressure already unsustainable for patient safety', The Guardian. [Online]. Friday 29 October 2021. Available from: https://www.theguardian.com/society/2021/oct/29/nhs-winter-pressure-already-unsustainable-patient-safety

Appendix

Using the value_counts() function, it was determined that there are :
- 5 service settings, with 'General Practice' significantly the most popular,
- 3 context types, with 'Care Related Encounters' being the highest,
- 18 national categories, with 'Inconsistent Mapping', 'General Consultation Routine' and 'General Consultation Acute' being the three most popular,
- 3 appointment statuses, with 'Attended' appointments scoring highest and unattended appointments ('DNA') lowest.

Fig 1:

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

```
# Determine the top five locations based on record count.
locations.sort_values(ascending=False).head(5)
```

```
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
Name: sub_icb_location_name, dtype: int64
```

Fig 2:

```
# For each of these service settings, determine the number of records available for the period and the location.

nc_subset = nc.loc[:, ['sub_icb_location_name', 'service_setting', 'count_of_appointments', 'appointment_date']]

nc_subset_nwldn = nc_subset.loc[(nc_subset['appointment_date']>= '2022/01/01') &
                                (nc_subset['appointment_date']<= '2022/06/01') &
                                (nc_subset['sub_icb_location_name'] == 'NHS North West London ICB - W2U3Z')]

# View the output.
nc_subset_nwldn['service_setting'].value_counts()
```

```
General Practice             2104
Other                        1318
Primary Care Network         1272
Extended Access Provision    1090
Unmapped                      152
Name: service_setting, dtype: int64
```

Fig 3:

```
# Number of appointments per month == sum of count_of_appointments by month.
# Use the groupby() and sort_values() functions.
appointments_by_month = ar.groupby('appointment_month')[['count_of_appointments']]\
                         .sum().sort_values(by='count_of_appointments', ascending = False).reset_index()

appointments_by_month.head()
```

	appointment_month	count_of_appointments
0	2021-11-01	30405070
1	2021-10-01	30303834
2	2022-03-01	29595038
3	2021-09-01	28522501
4	2020-10-01	28301932