The National Health Services (NHS) incurs significant costs when patients miss GP appointments. Instead of introducing a penalty for the missed appointments, the government is looking for a data-informed approach how best to handle this problem. As a part of the project, our team of analysts was contracted by the NHS to answer two main questions:

- Has there been adequate staff and capacity in the network?
- What was the actual utilisation of resources?

To work on these questions, GitHub repository created. In the provided teamwork scenario, GitHub enables better collaboration between team members. Several developers can clone projects to their workstations and work on the code without affecting the original. Also, it makes possible to trace mistakes, which are unavoidable in any big project, see the history of changes, who made it and what exactly was done and roll back if needed.

Before starting analysis, necessary libraries imported to work with the data: pandas, NumPy, matplotlib and seaborn. I use pandas read_csv() and read_excel() functions to create 3 dataframes. To avoid bias in parameters estimation and reduce the risk of invalid conclusions, check the data first.

- quick check the data makes sense (df.head() method).
- look at data frame size (df.shape)
- looking at missing values (df.isna()) - no missing values found.
- checking the data type (df.dtypes).
- df.describe() method to look at descriptive statistics for the numerical columns for negative values or obvious outliners.

	ad.describe()	**ar.describe()**	**nc.describe()**
	count_of_appointments	count_of_appointments	count_of_appointments
count	137,793.00	596,821.00	817,394.00
mean	1,219.08	1,244.60	362.18
std	1,546.90	5,856.89	1,084.58
min	1.00	1.00	1.00
25%	194.00	7.00	7.00
50%	696.00	47.00	25.00
75%	1,621.00	308.00	128.00
max	15,400.00	211,265.00	16,590.00

Exploring the data: working with “nc” dataframe and applying df.nunique() method, there are 106 locations found. Five locations with the highest number of records are presented below:

NHS North West London ICB - W2U3Z	13,007
NHS Kent and Medway ICB - 91Q	12,637
NHS Devon ICB - 15N	12,526
NHS Hampshire and Isle Of Wight ICB - D9Y0V	12,171
NHS North East London ICB - A3A8R	11,837

Applying df.value_counts() method respectively we investigate service settings, context types, national categories, and appointment statuses.

There are 5 unique service settings defined:

General Practice	359,274
Primary Care Network	183,790
Other	138,789
Extended Access Provision	108,122
Unmapped	27,419

There are 3 unique context types:

Care Related Encounter	700,481
Inconsistent Mapping	89,494
Unmapped	27,419

18 unique national categories:

Inconsistent Mapping	89,494
General Consultation Routine	89,329
General Consultation Acute	84,874
Planned Clinics	76,429
Clinical Triage	74,539
Planned Clinical Procedure	59,631
Structured Medication Review	44,467
Service provided by organisation external to the practice	43,095
Home Visit	41,850
Unplanned Clinical Activity	40,415
Patient contact during Care Home Round	28,795
Unmapped	27,419
Care Home Visit	26,644
Social Prescribing Service	26,492
Care Home Needs Assessment & Personalised Care and Support Planning	23,505
Non-contractual chargeable work	20,896
Walk-in	14,179
Group Consultation and Group Education	5,341

And 3 types of appointment status:

Attended	232,137
Unknown	201,324
DNA	163,360

After checking the data format across all dataframes and making the changes where needed, we can see that in “ad” dataframe the appointment dates range from the 01 December 2021 till 30 June 2022 and in “nc” dataframe from the 01 August 2021 till the 30 June 2022.

Subset for the largest NHS North West London ICB - W2U3Z location and dates between 01 January till 01 June 2022 created where General Practice reported the largest number of records.

NHS North West London ICB - W2U3Z 01/01-01/06/2022	
Service Setting:	
General Practice	2,080
Other	1,307
Primary Care Network	1,261
Extended Access Provision	1,076
Unmapped	150

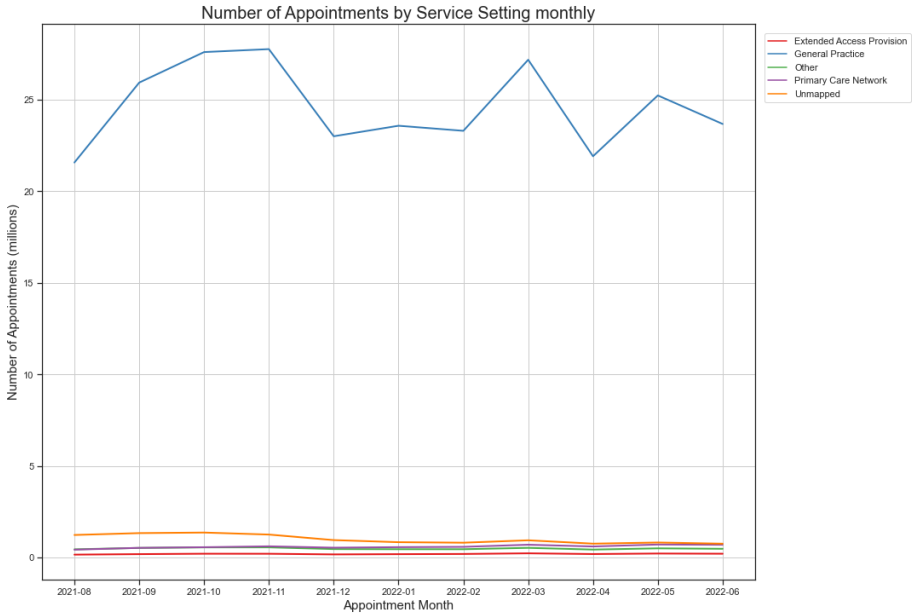
By applying df.groupby() method and aggregating count of appointments by month, we find the monthly number of appointments, and with df.sort_values() method find that November 2021 was the busiest month.

year	month	count_of_appointments
2021	11	30,405,070
2021	10	30,303,834
2022	3	29,595,038
2021	9	28,522,501
2022	5	27,495,508
2022	6	25,828,078
2022	1	25,635,474
2022	2	25,355,260
2021	12	25,140,776
2022	4	23,913,060
2021	8	23,852,171

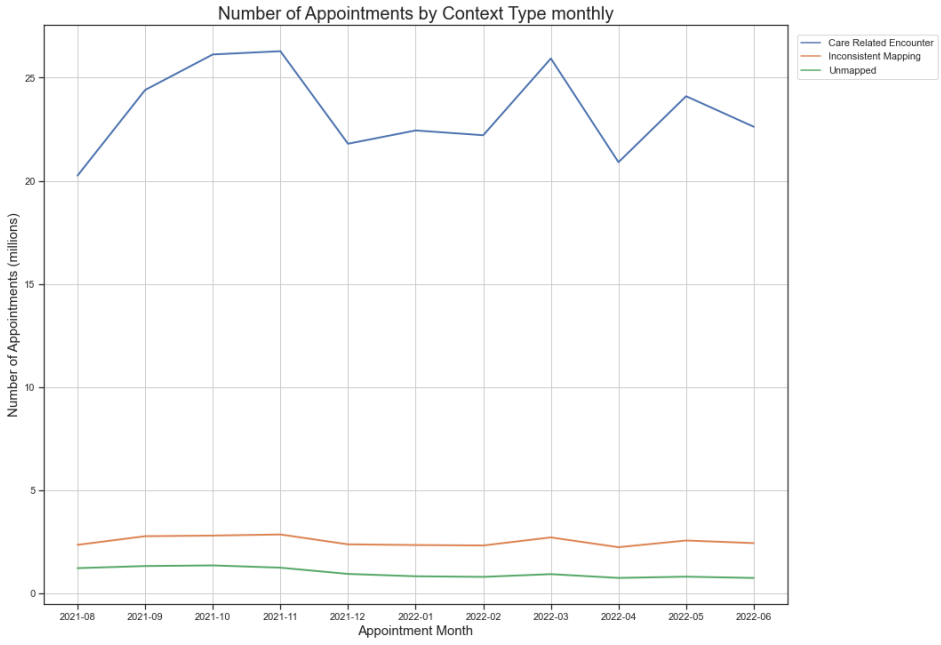
By number of records per month for the same dataframe, it is March 2022.

year	month	count_of_appointments
2022	3	82,822
2021	11	77,652
2022	5	77,425
2021	9	74,922
2022	6	74,168
2021	10	74,078
2021	12	72,651
2022	1	71,896
2022	2	71,769
2022	4	70,012
2021	8	69,999

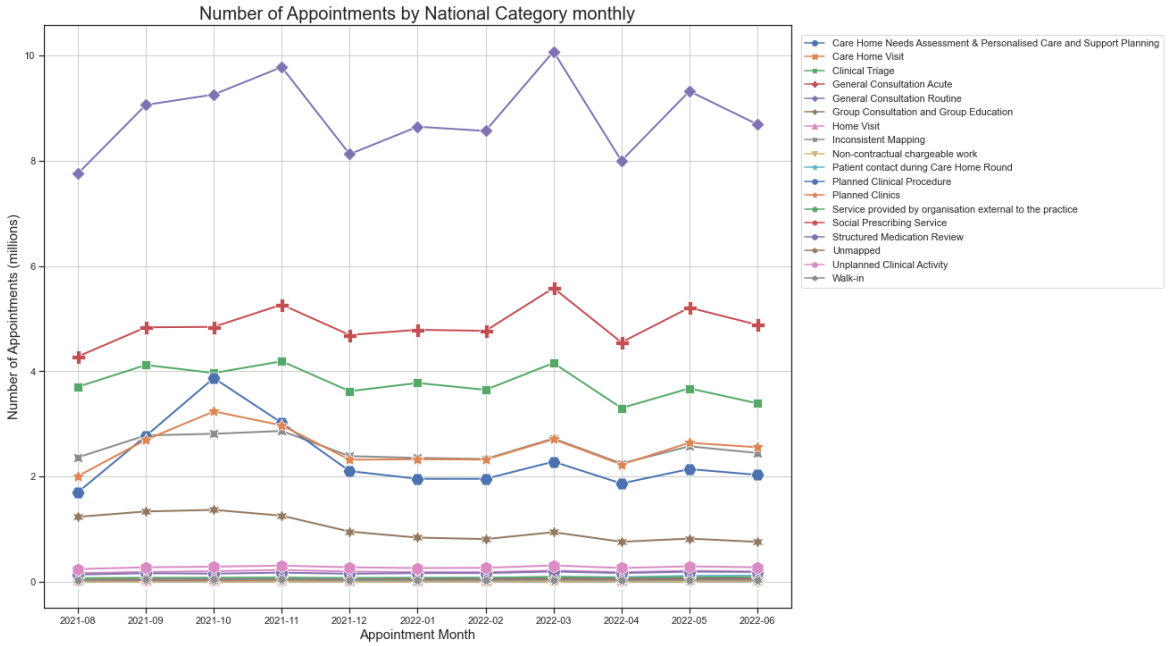
One of the best ways to determine monthly and seasonal trends and patterns is visualisation. Libraries (Matplotlib and Seaborn) imported in the beginning are used. Grouping and aggregating data (same methods used as before) we calculate monthly number of appointments by different categorical values: service settings, national categories, and context types. Lineplote was chosen to see how monthly number of appointments changes over time with ‘hue’ parameter set to respective value: service setting, national category, and context time.



Most appointments were delivered by General Practice, peaking in November 2021 and March 2022. Number of appointments delivered by other services (Primary Care Network, Extended Access Provision, practices by other providers) stays almost flat.

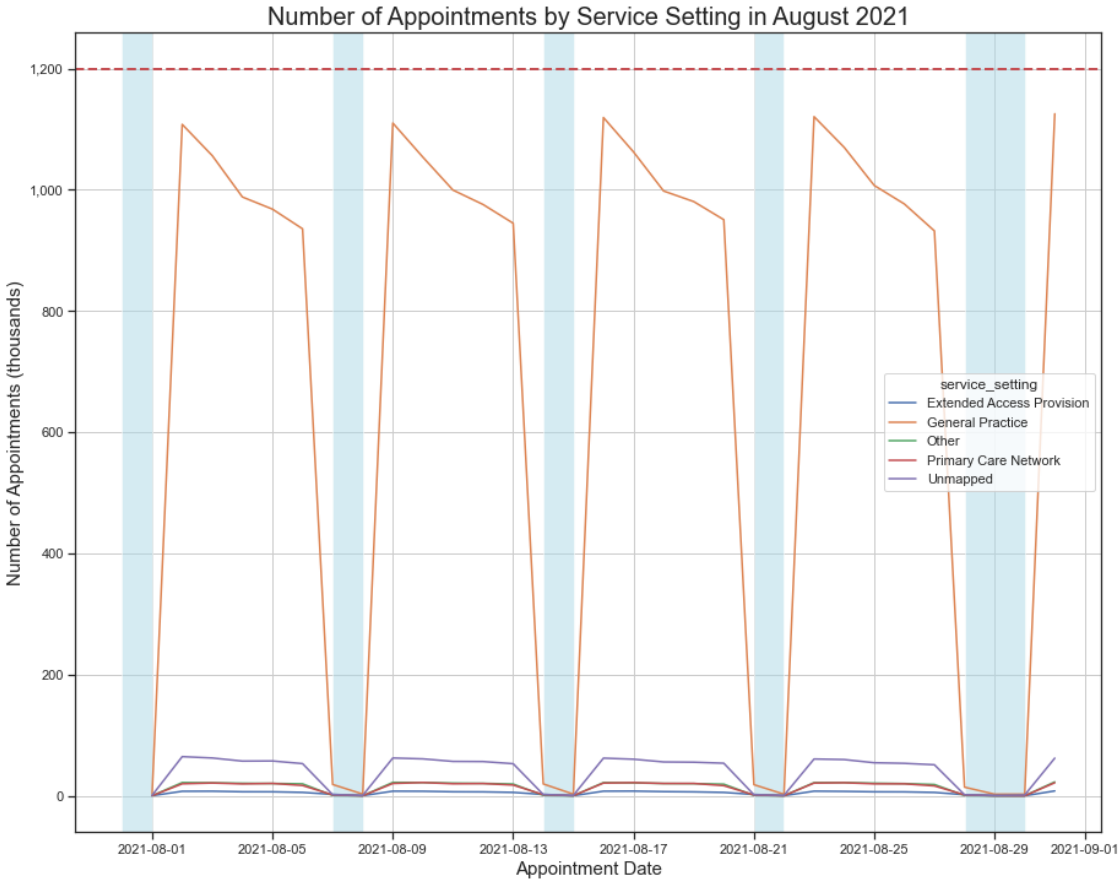


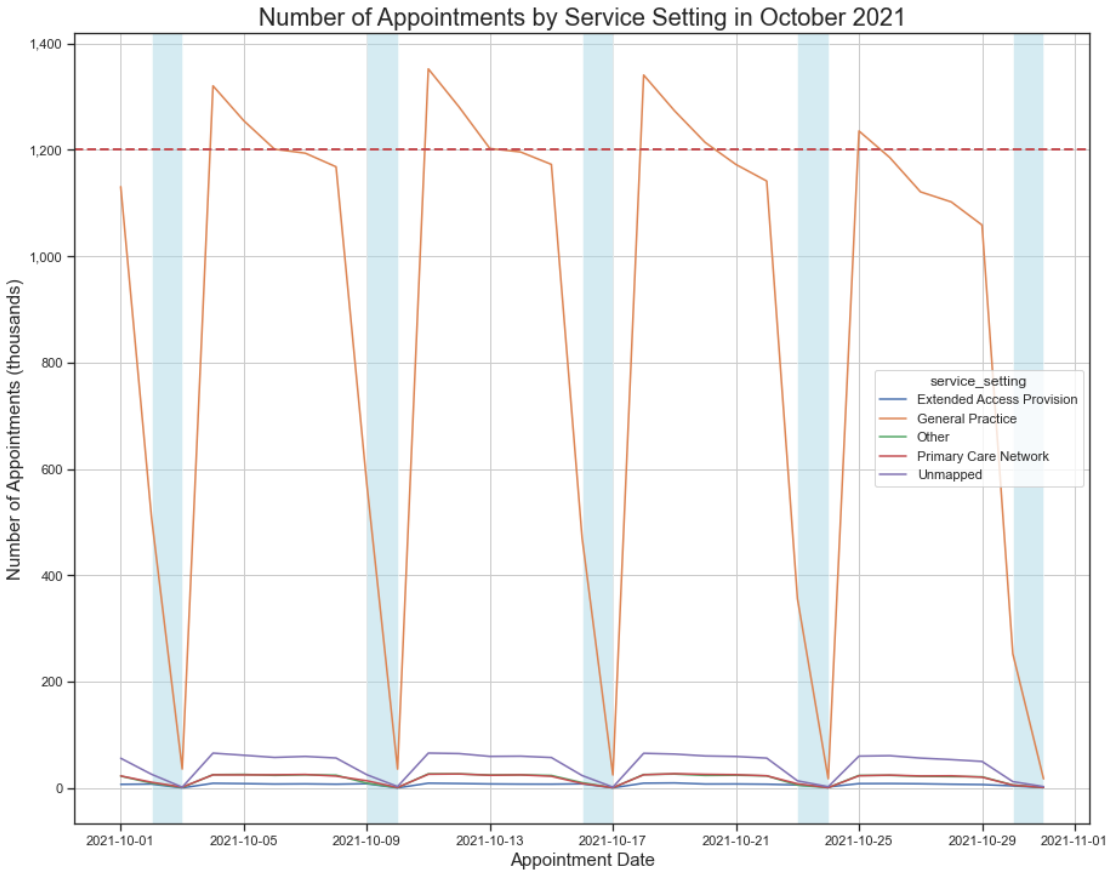
Lineplot by context type shows the similar growing trend in autumn and beginning of spring: main peaks happen by November 2021 and March 2022. The largest load of appointments related to the direct patient care - Care Related Encounter.

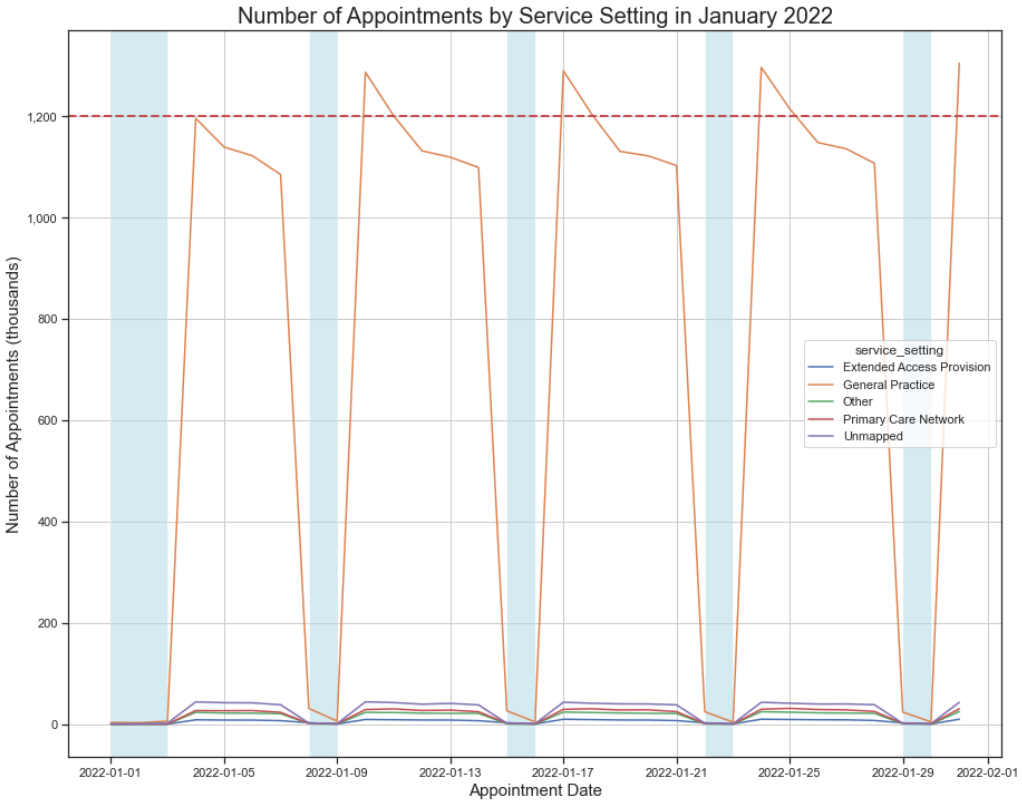


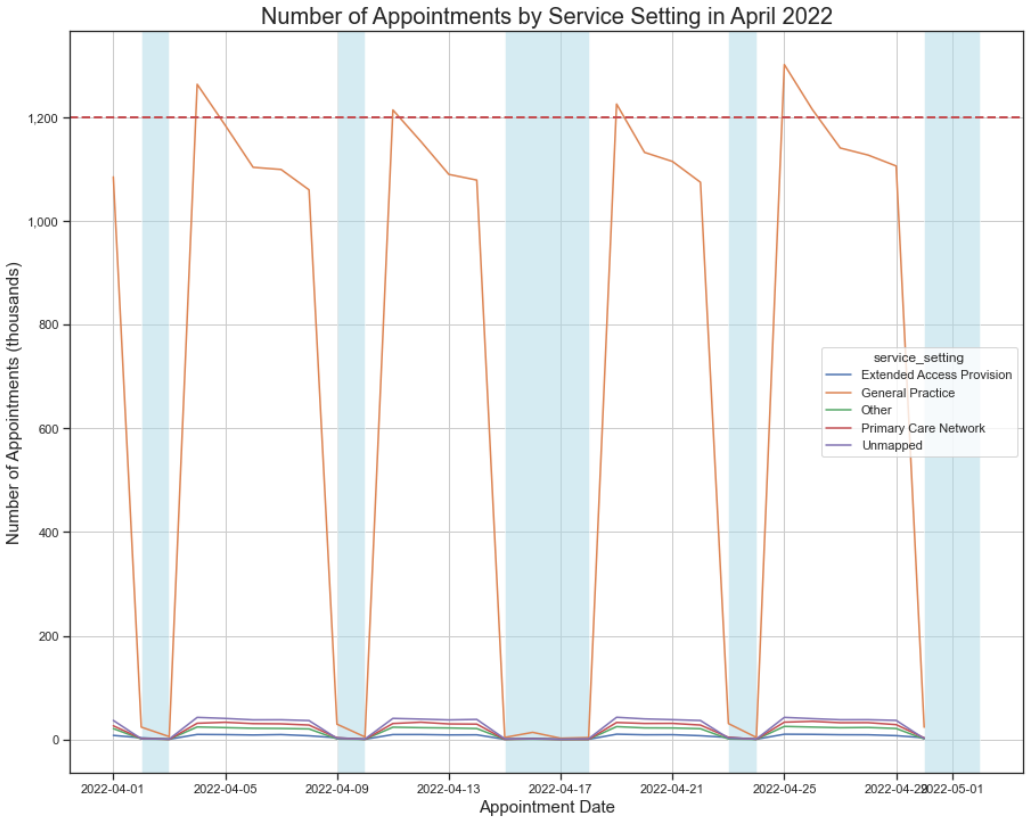
Most of the appointments were placed in General Consultation Routine or General Consultation Acute category, followed by Clinical Triage and Planned Clinical Procedure or Planned Clinic. There are similar seasonal trends as before can be seen on the graph. However, the maximum peak for the mentioned categories happens by March 2022. Interestingly, Planned Clinical Procedure has steeper growth in September-October following opposite trend (as Planned Clinics) in November, when others keep increasing. It can be explained by coping with appointments' backlog created during COVID crisis.

Four lineplots created using the same approach for August 2021, October 2021, January 2022, and April 2022 to look at daily trends. We clearly see the same weekend effect, with General Practice delivered the largest part of all appointments: the highest pressure registered on Monday (or the next day after bank holiday). As the week progress, the number decreases falling to the minimum over weekend. Depending on the season, every weekday has variation in number of appointments as well. The graphs are presented below, and the light blue areas highlight the weekends or bank holidays to underline the effect.


Number of Appointments by Service Setting in August 2021
Number of Appointments (thousands)
1,200
1,000
800
600
400
200
0
2021-08-01
2021-08-05
2021-08-09
2021-08-13
2021-08-17
2021-08-21
2021-08-25
2021-08-29
2021-09-01
Appointment Date
service_setting
Extended Access Provision
General Practice
Other
Primary Care Network
Unmapped


Number of Appointments by Service Setting in October 2021
Number of Appointments (thousands)
1,400
1,200
1,000
800
600
400
200
0
2021-10-01
2021-10-05
2021-10-09
2021-10-13
2021-10-17
2021-10-21
2021-10-25
2021-10-29
2021-11-01
Appointment Date
service_setting
Extended Access Provision
General Practice
Other
Primary Care Network
Unmapped


Number of Appointments by Service Setting in January 2022
Number of Appointments (thousands)
1,200
1,000
800
600
400
200
0
2022-01-01
2022-01-05
2022-01-09
2022-01-13
2022-01-17
2022-01-21
2022-01-25
2022-01-29
2022-02-01
Appointment Date
service_setting
Extended Access Provision
General Practice
Other
Primary Care Network
Unmapped

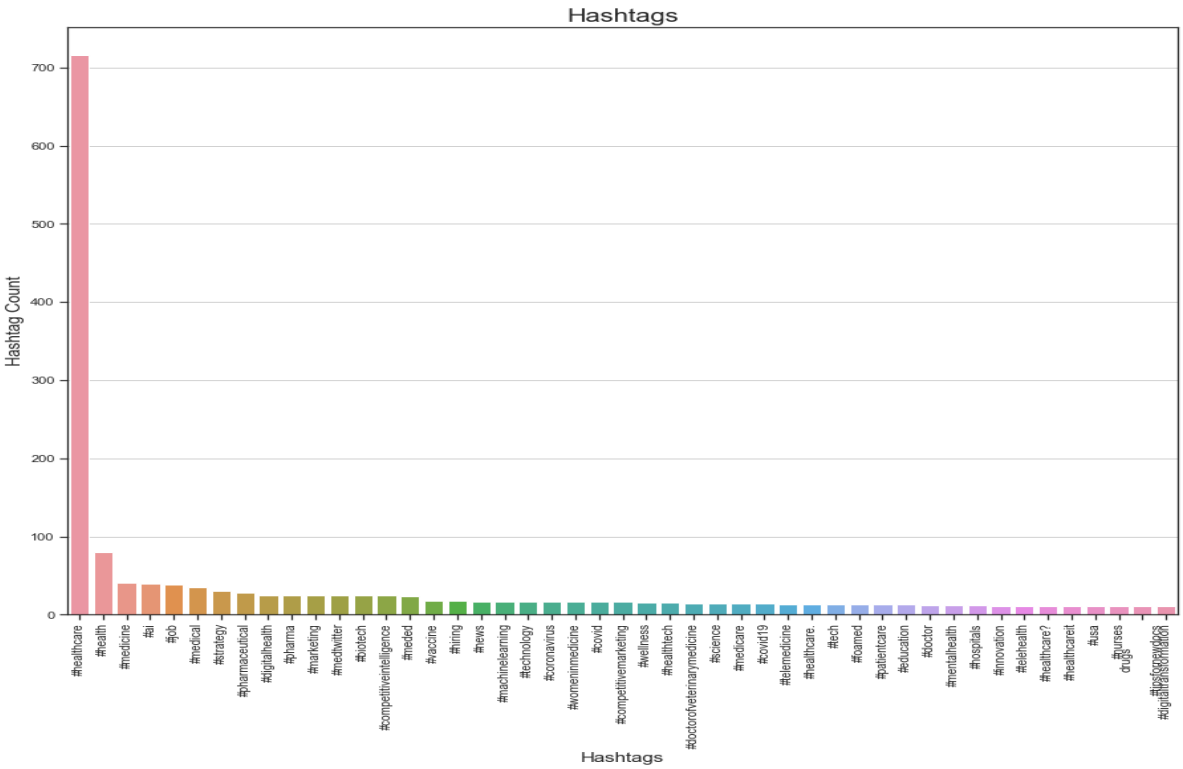

Number of Appointments by Service Setting in April 2022
Number of Appointments (thousands)
1,200
1,000
800
600
400
200
0
2022-04-01
2022-04-05
2022-04-09
2022-04-13
2022-04-17
2022-04-21
2022-04-25
2022-04-29
2022-05-01
Appointment Date
service_setting
Extended Access Provision
General Practice
Other
Primary Care Network
Unmapped

To analyse trending hashtags (#) on Twitter related to healthcare in the UK new dataframe tweets created.

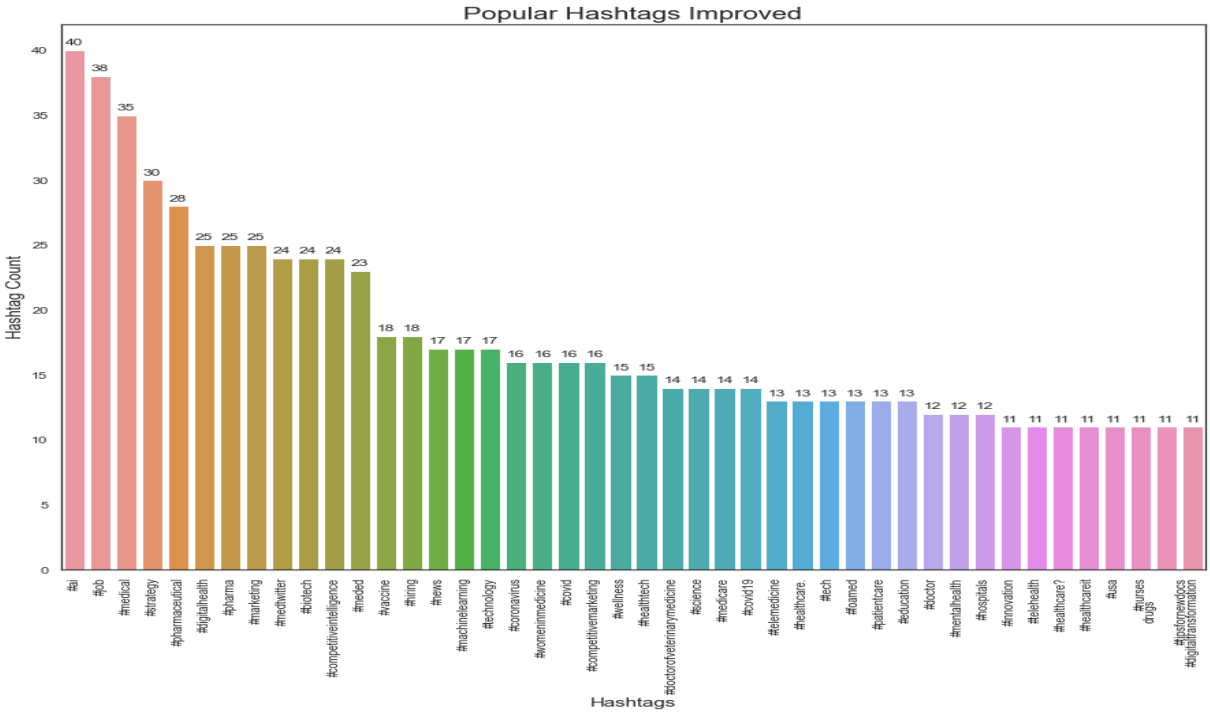
Looking at the count of retweeted and favourite tweet messages wouldn't make much sense as the most of them were not retweeted at all (526 out of 1174 were never retweeted) or favoured (1027 out of 1174 favourite zero times). Instead, we look at the hashtags used in the tweets. By knowing the popular hashtags, it is easy to see the topics people most interested and involved.

List of the hashtags with count more than 10 and the graph are presented below. For the visualisation barplot was used to see the number of times each hashtag used.

word	count	word	count
#healthcare	716	#wellness	15
#health	80	#healthtech	15
#medicine	41	#doctorofveterinarymedicine	14
#ai	40	#science	14
#job	38	#medicare	14
#medical	35	#covid19	14
#strategy	30	#telemedicine	13
#pharmaceutical	28	#healthcare.	13
#digitalhealth	25	#tech	13
#pharma	25	#foamed	13
#marketing	25	#patientcare	13
#medtwitter	24	#education	13
#biotech	24	#doctor	12
#competitiveintelligence	24	#mentalhealth	12
#meded	23	#hospitals	12
#vaccine	18	#innovation	11
#hiring	18	#telehealth	11
#news	17	#healthcare?	11
#machinelearning	17	#healthcareit	11
#technology	17	#usa	11
#coronavirus	16	#nurses	11
#womeninmedicine	16	drugs\n\n#tipsfornewdocs	11
#covid	16	#digitaltransformation	11
#competitivemarketing	16		



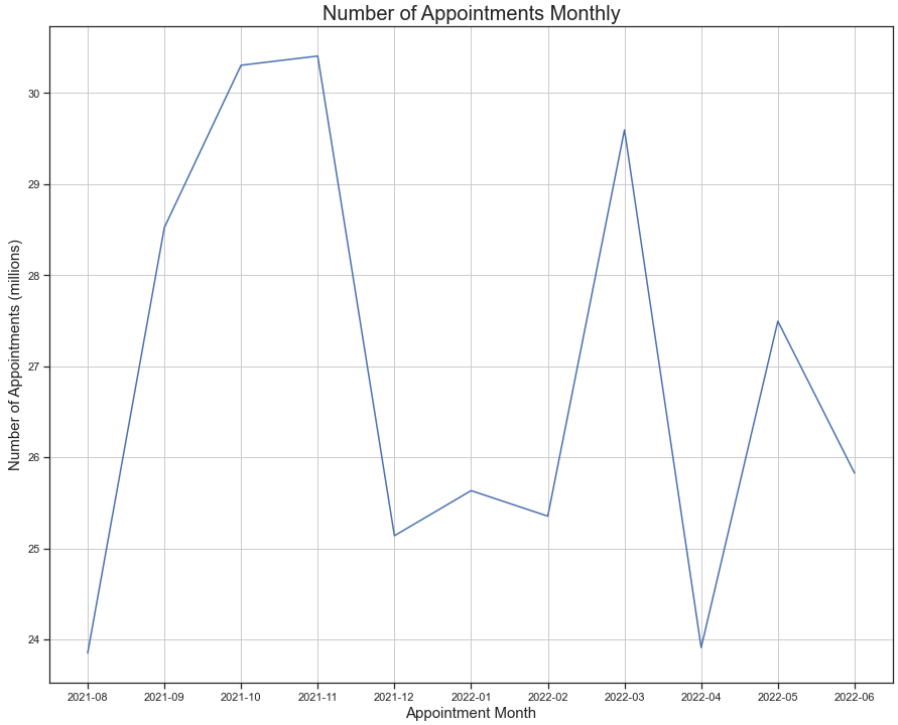
To improve the chart, Inter Quartile Range (IQR) method is used to detect the outliers. The new subset is created with the hashtag counts within the brackets of calculated upper and lower bounds. The 5 most popular hashtags are ai, job, medical, strategy and pharmaceuticals. Additionally, hashtags like #healthcare. and # healthcare? can be join, to be 6th in the list.



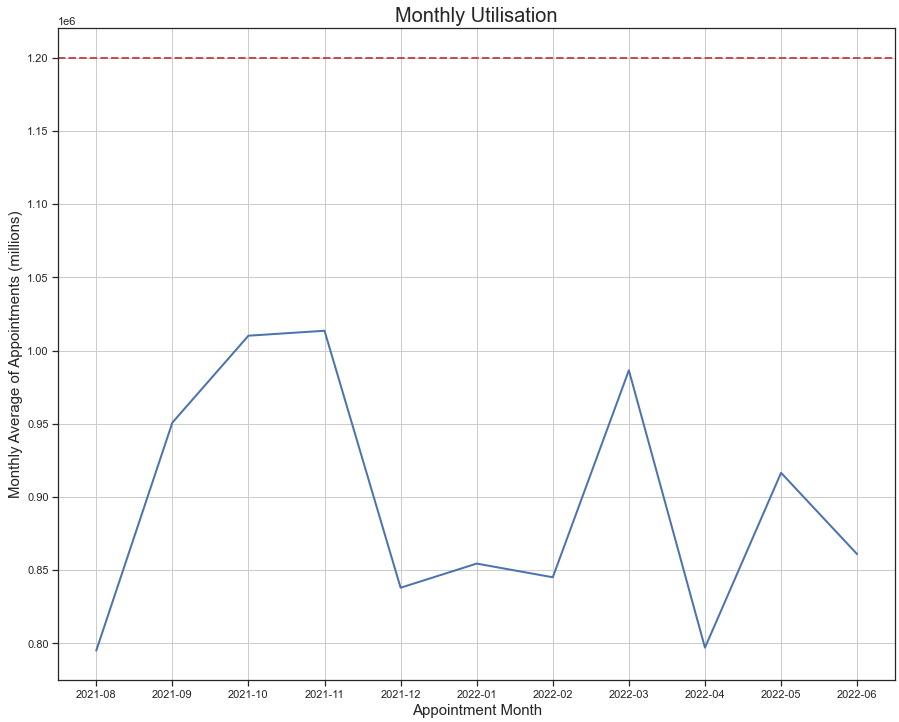
Dataframe “ar” is used for the next part. After sense check of the data, time frame was checked: January 2020 – June 2022. For the research, we filter it from August 2021 onwards.

To answer the question if the NHS needs to start looking at increasing staff level, we look at utilisation rate of services. After grouping and aggregating, monthly number of appointments and the average utilisation of services (aggregate of appointments/30) calculated. The utilisation rate calculated based on the NHS capacity of 1,200,000 appointments a day limit.

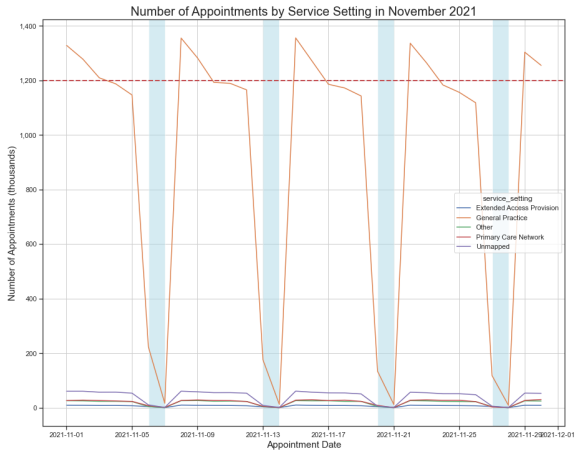
appointment_month	count_of_appointments	utilisation	utilisation rate %
2021-08	23,852,171	795,072.4	66.3%
2021-09	28,522,501	950,750.0	79.2%
2021-10	30,303,834	1,010,127.8	84.2%
2021-11	30,405,070	1,013,502.3	84.5%
2021-12	25,140,776	838,025.9	69.8%
2022-01	25,635,474	854,515.8	71.2%
2022-02	25,355,260	845,175.3	70.4%
2022-03	29,595,038	986,501.3	82.2%
2022-04	23,913,060	797,102.0	66.4%
2022-05	27,495,508	916,516.9	76.4%
2022-06	25,828,078	860,935.9	71.7%

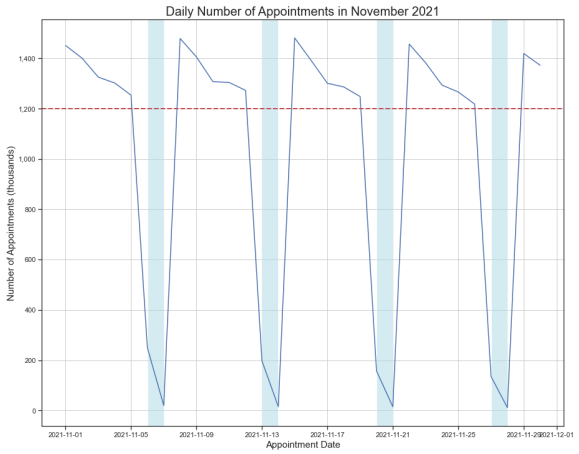


November 2021 has the highest average utilisation rate of 84.5%. The lineplot presenting the daily average is shown below.

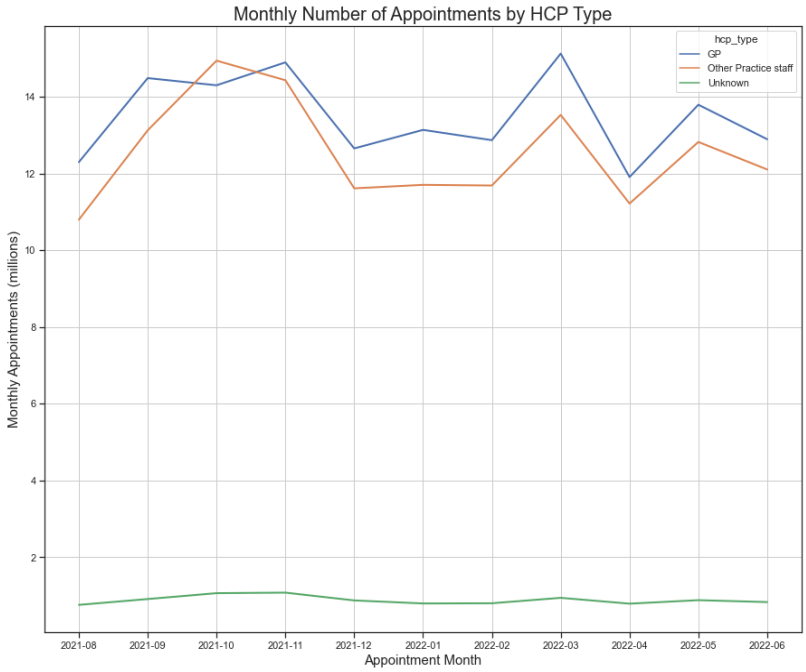


However, if we compare the results with the information from the seasonal trends, we can see that there are lots of cases number of appointments were over 1,200,000. Additional lineplots for the busiest month of November 2021 presented. Cycle effect when the number of appointments delivered by GP peaks up every Monday exceeds the capacity, happens repeatedly every week and continues till Wednesdays. For all services it is over the limit all weeks in November 2021. It gives an indication of lack of resources during peak seasons.

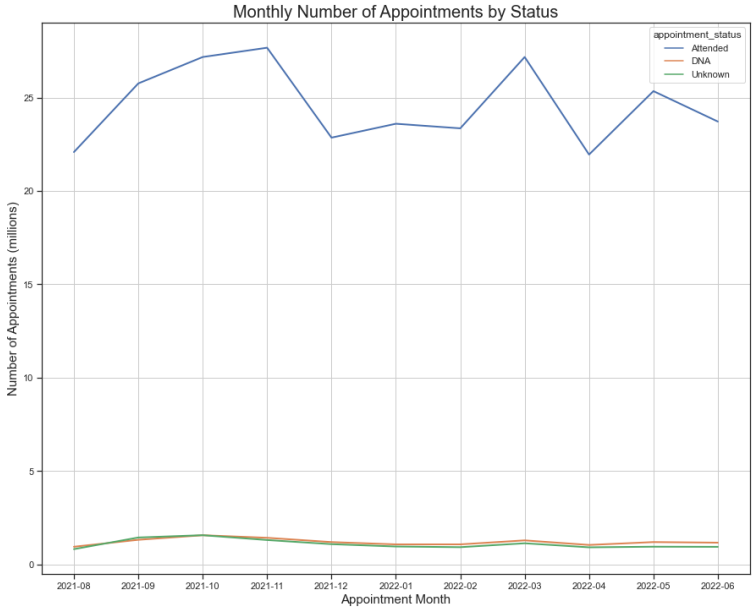




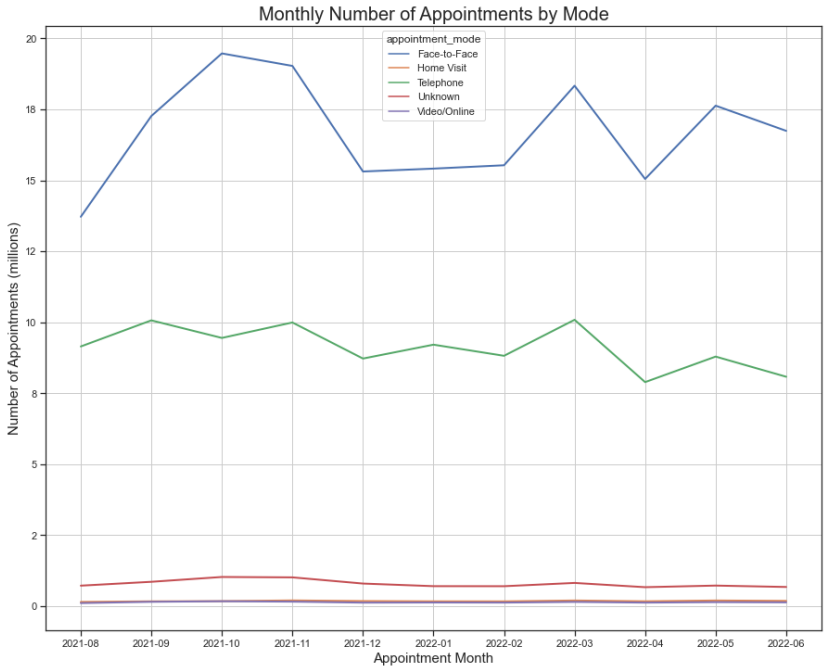
By whom the appointments were carried out. Subset created by grouping all monthly appointments by HCP type. In lineplot, HCP used as “hue”. GPs and Other Practice Staff (nurses, counsellors, osteopaths etc.) carried out similar number of appointments. Only once Other Practice Staff overran GP group - October 21.



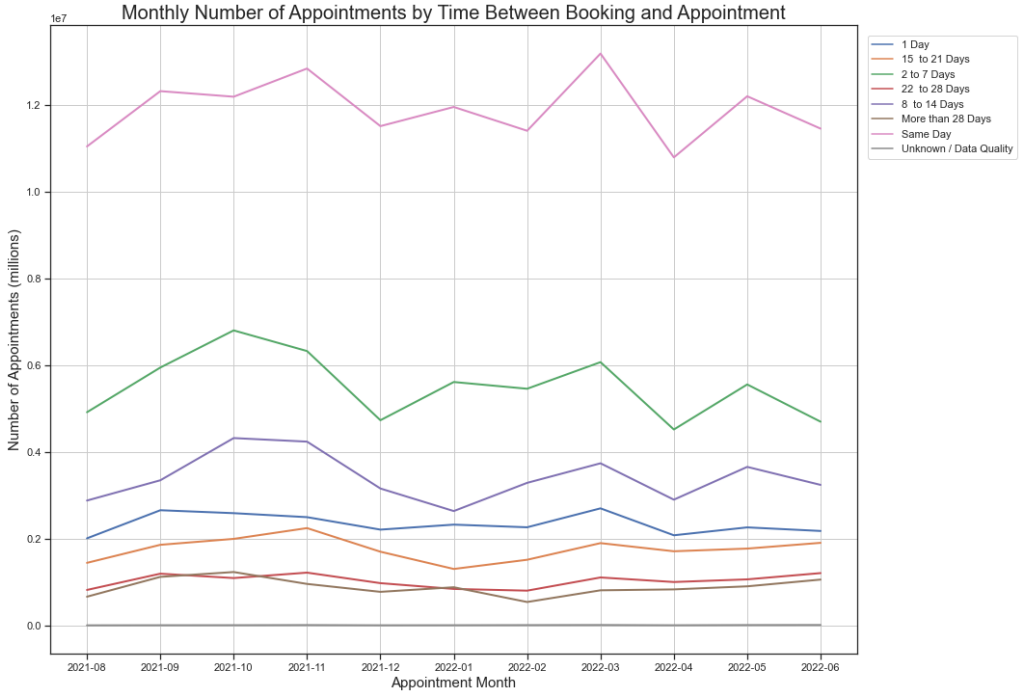
There were no significant changes in DNA (“Did Not Attend”) through the period. The same method was used grouping dataframe by month and appointment status and aggregating. The lineplot uses status as “hue” parameter.

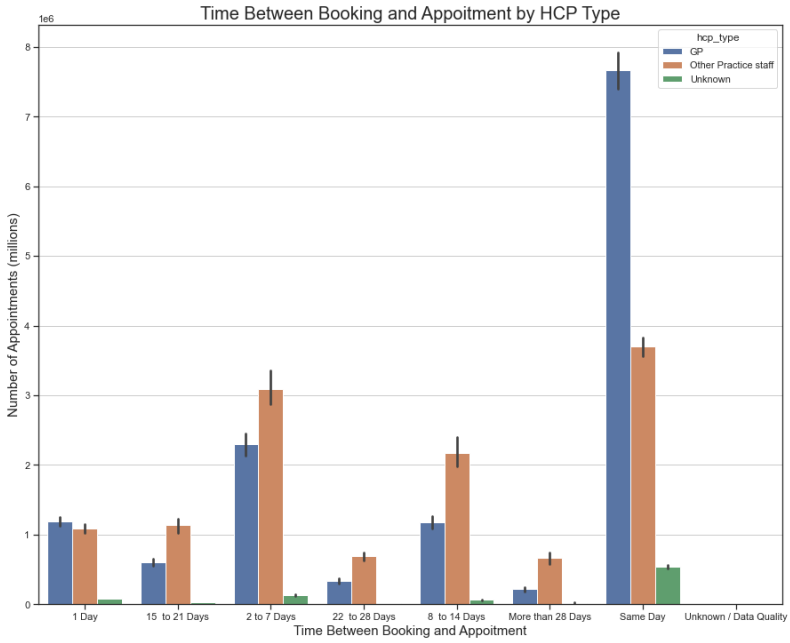


Home visits and online appointments stay almost flat, the largest share of appointments was made face-to-face or over phone. Face to face is far more sensitive to seasonal changes. Home visits and online appointments are flat.

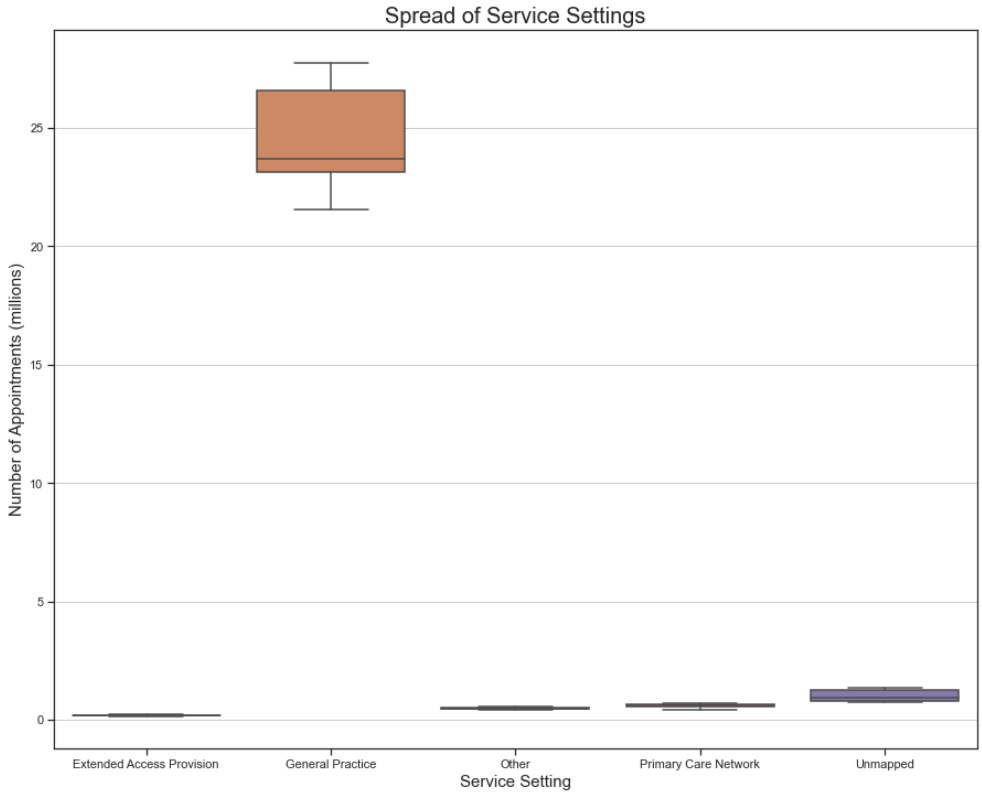


About 40% of all appointments happened the same day and delivered by GPs mostly, followed by 2-7 days mostly done by other practice staff. Over the busiest months (autumn) the share of the same day fell slightly and at the same time number of 2-7 and 8-14 days increased instead. That's the period when planned appointments increased and number of appointments delivered by Other Practice Staff overran GP.

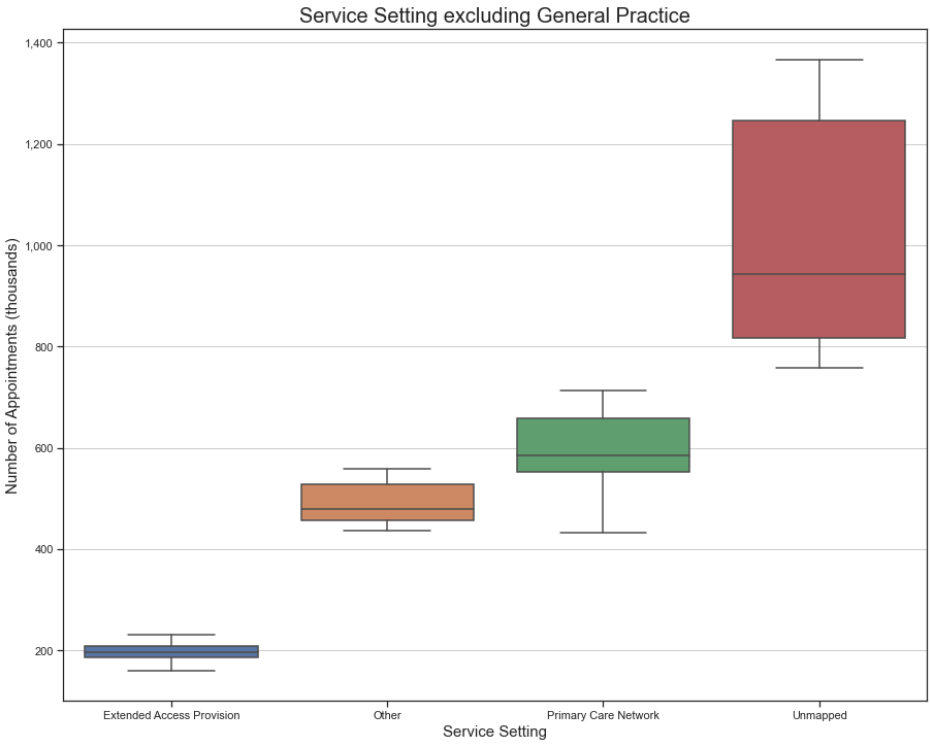




Boxplot graph is used to show the distributions of numerical values (monthly number of appointments) in different service settings. Each box presents minimum and maximum as “whiskers”, and median, lower (25%) and upper (75%) quartiles within box. All these values are much higher for General Practice group confirming they are mostly under pressure. There are no outliners outside the box, the length of the whiskers is about the same and the median is located closer to the lower quartile (positively skewed, mean greater than median or higher frequency of high monthly appointments).



The distribution of the data in other service settings is not clear from this boxplot. To improve, we exclude GP and create a new subset again grouping and aggregating number of appointments monthly.



Highest values are in “Unmapped” due to an error receiving the data or no record, followed by “Primary Care Network” (staff employed by ARRS contracts - Additional Roles Reimbursement Scheme), followed by group “Other” (not NHS) and “Extended Access Provision” when the appointments commissioned as part of extended access arrangements. The last group has smallest interquartile range meaning low dispersion in monthly appointments and not sensitivity to seasonal changes.

Conclusions and recommendations:

The number of daily appointments systematically exceeds the NHS capacity of 1,200,000 a day. It is becoming a problem during high season, especially at the busiest locations. As most of practices stay closed over weekends it doesn’t show how the things really are.

Major part of the appointments was delivered by General Practice in General Consultation Routine or Acute category, who is mostly under pressure. To solve the problem, we can hire more people to be available and increase the capacity or restructure existing system to be more flexible and cope with pressure when it is needed the most.

- Make other services like “Extend Enhanced Access” be more sensitive to seasonal changes: opening practices in the busiest locations over weekends and extend hours to reflect local needs.
- As Mondays are the busiest book less follow up appointments and leave more spaces for urgent and same day.
- Use more video/online appointments.

- Increase capacity of Planned Clinic and Procedures to prevent potential growth of General Consultation Acute appointments.
- Review and improve the data collection within the NHS to decrease the number of “Unmapped” data in the future.