DIAGNOSTIC ANALYSIS OF THE NATIONAL HEALTH SERVICE DATA SET

22nd October 2022
Prepared by: Oksana Fedorova
e-mail address: OksanaF@email.com

1. Context of the business issue and aim of the data analytics project.

The National Health Services (NHS), a publicly funded healthcare system in England, incurs significant costs when patients miss general practitioner (GP) appointments.

This data analytics project aims to help NHS reduce or eliminate missed appointments by identifying:

a) possible appointment-related trends associated with locations, service settings, context types, national categories, and appointment statuses;
b) possible appointment-related trends connected to time series/seasons;
c) the top trending hashtags (#) on Twitter related to healthcare in the UK;
d) if there is adequate staff and capacity in the networks, and what was the actual utilisation of resources.

2. Analytical approach and discovered insights.

2.1. Analytical approach.

2.1.1. GitHub repository.

The LSE_DA_NHS_analysis repository was created on https://github.com/ website to store, update, manage project files and allow easy collaboration for the team members working on the project.

Repository URL: https://github.com/SanaFed/LSE_DA_NHS_analysis

2.1.2. Initial exploration of data files.

- The workstation was prepared by importing the necessary libraries and three data files (actual_duration.csv, appointments_regional.csv and national_categories.xlsx) in a new Python3 file.
- All three files were converted to DataFrames and sense-checked.
- No missing values were identified.
- Each DataFrame contains data for different time periods:
  - actual_duration (daily): 01/12/2021 – 30/06/2022;
  - appointments_regional (monthly): 01/2020 – 06/2022;
  - national_categories (daily): 01/08/2021 – 30/06/2022.
- Each DataFrame contains different number of appointments:
  - actual_duration: 167,980,692;
  - appointments_regional: 742,804,525;
  - national_categories: 296,046,770.
- The data set contains information about 42 Integrated Care Boards (ICBs), which are represented by 106 locations as sub-ICBs.
- Top lines by number of records by category in national_categories DataFrame:
  - 85.70% of records involved a patient ('Care Related Encounter');
  - 43.95% of records were related to GP;

- 38.90% of records were attended (= 91.24% of all appointments);
- 10.93% of records were marked as 'General Consultation Routine' (second top after 'Inconsistent Mapping').

- Sub-ICBs with the top 10 highest number of appointments per record by DataFrame:
  - actual_duration: NHS North East London ICB - A3A8R;
  - appointments_regional: 4 records - NHS North East London ICB, 6 records - NHS North West London ICB;
  - national_categories: NHS North East London ICB - A3A8R.
- Top five **sub-ICBs** in national_categories DataFrame
  with the highest number of **records** (= 7.61% of all records):

	Location (sub-ICB)	Number of records
1	NHS North West **London** ICB - W2U3Z	13007
2	NHS Kent and Medway ICB - 91Q	12637
3	NHS Devon ICB - 15N	12526
4	NHS Hampshire and Isle Of Wight ICB - D9Y0V	12171
5	NHS North East **London** ICB - A3A8R	11837

  with the highest number of **appointments** (= 15.93% of all appointments):

	Location (sub-ICB)	Number of appointments
1	NHS North West **London** ICB - W2U3Z	12142390
2	NHS North East **London** ICB - A3A8R	9588891
3	NHS Kent and Medway ICB - 91Q	9286167
4	NHS Hampshire and Isle Of Wight ICB - D9Y0V	8288102
5	NHS South East **London** ICB - 72Q	7850170

  39.96% of records and 62.73% of appointments from the above tables belong to London's ICBs.
- Top five **ICBs** in national_categories DataFrame with the highest number of appointments (= 23.81% of all appointments):

	Location (ICB)	Number of appointments
1	NHS North East and North Cumbria ICB	16882235
2	NHS West Yorkshire ICB	14358371
3	NHS Greater Manchester ICB	13857900
4	NHS Cheshire and Merseyside ICB	13250311
5	NHS North West London ICB	12142390

- Highest number of attended appointments were booked on the same day (48.45%); followed by appointments booked 2-7 days prior (20.38%) and 8-14 days prior (11.08%).
- Majority of appointments with known length were 6-10 minutes (20.12%); followed by 1-5 minutes appointments (17.03%) and 11-15 minutes (14.98%).

2.1.3. Initial exploration of data files related to GP.

- Top five **ICBs** in national_categories DataFrame with the highest number of GP appointments (= 15.90% of all GP appointments):

	Location (sub-ICB)	Service setting	Number of appointments
1	NHS North West **London** ICB - W2U3Z	General Practice	10432225
2	NHS North East **London** ICB - A3A8R	General Practice	9174258
3	NHS Kent and Medway ICB - 91Q	General Practice	8645534
4	NHS Hampshire and Isle Of Wight ICB - D9Y0V	General Practice	7407509
5	NHS South East **London** ICB - 72Q	General Practice	7395389

62.71% of appointments from the above tables belong to London ICBs.

- Top five **ICBs** in national_categories DataFrame with the highest number of GP appointments (= 23.58% of all GP appointments):

	Location (ICB)	Service setting	Number of appointments
1	NHS North East and North Cumbria ICB	General Practice	15794580
2	NHS West Yorkshire ICB	General Practice	12611739
3	NHS Cheshire and Merseyside ICB	General Practice	12602953
4	NHS Greater Manchester ICB	General Practice	12426379
5	NHS North West London ICB	General Practice	10432225

- 94.26% GP appointments were attended.
- Highest number of attended GP appointments were booked on the same day (60.15%); followed by appointments booked 2-7 days prior (16.78%) and 1 day prior (9.30%).
- Majority of all GP appointments were recorded as General Consultation Routine (34.47%).

2.1.4. Exploration of data files by time series/seasons including visualisations.

Colour-blind pallet/colours were used to create visualisations included in the report. Weekends were highlighted by darker backgrounds in visualisations representing data by day.

- Total number of records by month in national_categories DataFrame:

Year	Month	Number of records
2021	8	69999
	9	74922
	10	74078
	11	**77652**
	12	72651
2022	1	71896
	2	71769
	3	**82822**
	4	70012
	5	**77425**
	6	74168

- Top five months with the highest number of appointments in national_categories DataFrame:

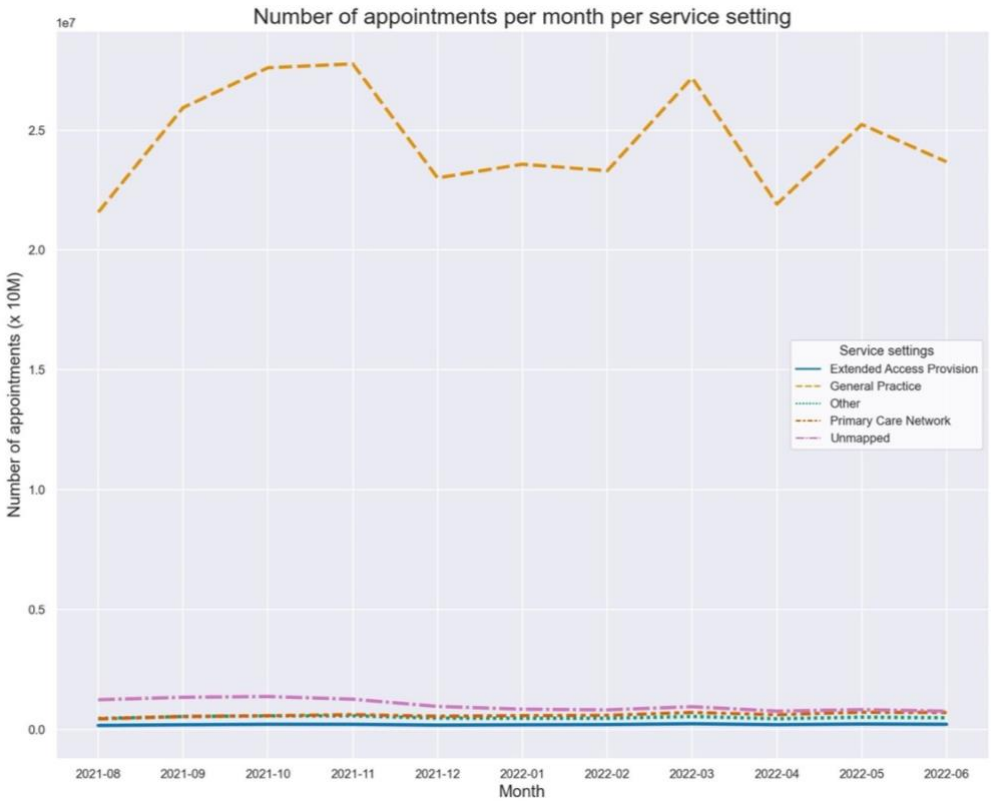
Year	Month	Number of appointments
2021	**11**	30405070
2021	**10**	30303834
2022	3	29595038
2021	**9**	28522501
2022	5	27495508

Three autumn months (30.14% of all appointments) and two spring months (19.28% of all appointments) are 5 months with the highest number of appointments. November 2021 had the highest number of appointments overall.

- Top five months with the highest number of GP appointments in national_categories DataFrame:

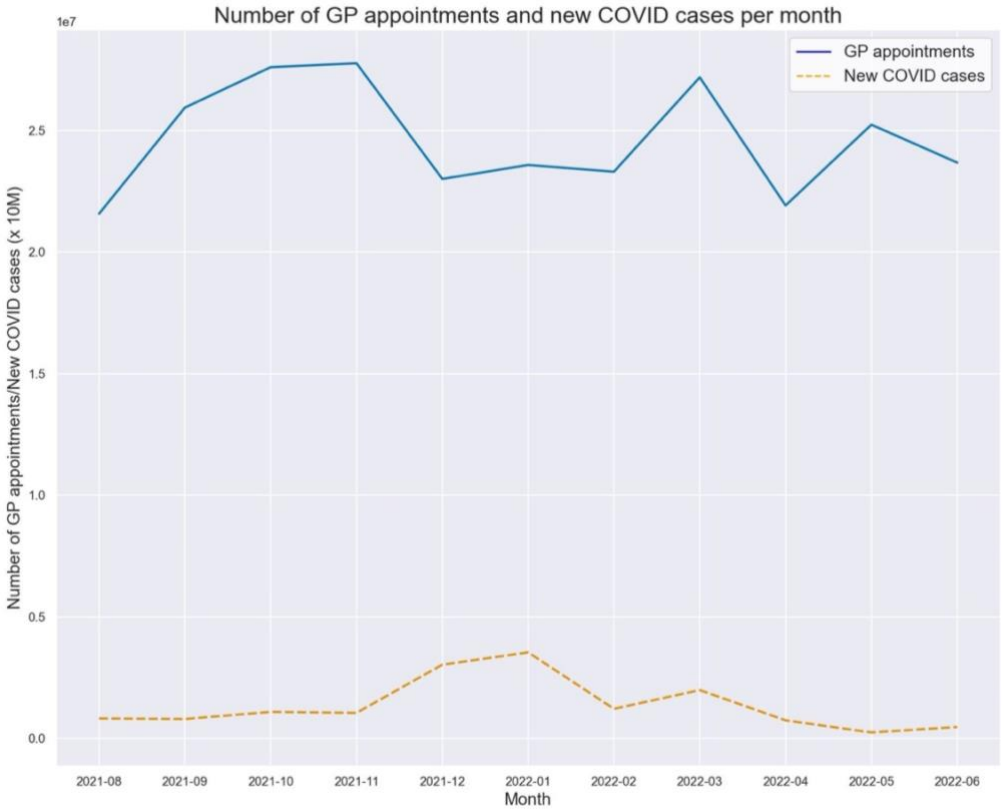
Year-Month	Service setting	Number of appointments
2021-**11**	General Practice	27767889
2021-**10**	General Practice	27606171
2022-03	General Practice	27187368
2021-**09**	General Practice	25940821
2022-05	General Practice	25238620

Three autumn months (30.03% of all GP appointments) and two spring months (19.36% of all GP appointments) are 5 months with the highest number of appointments. November 2021 had the highest number of GP appointments overall.

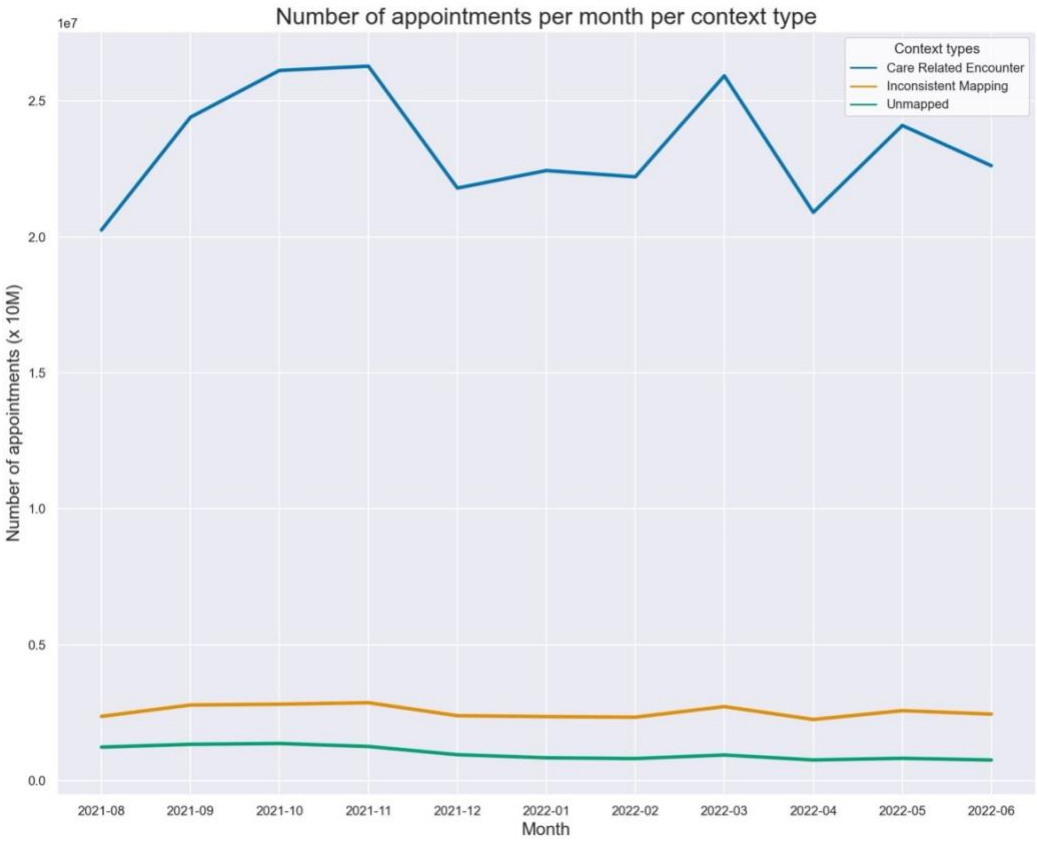


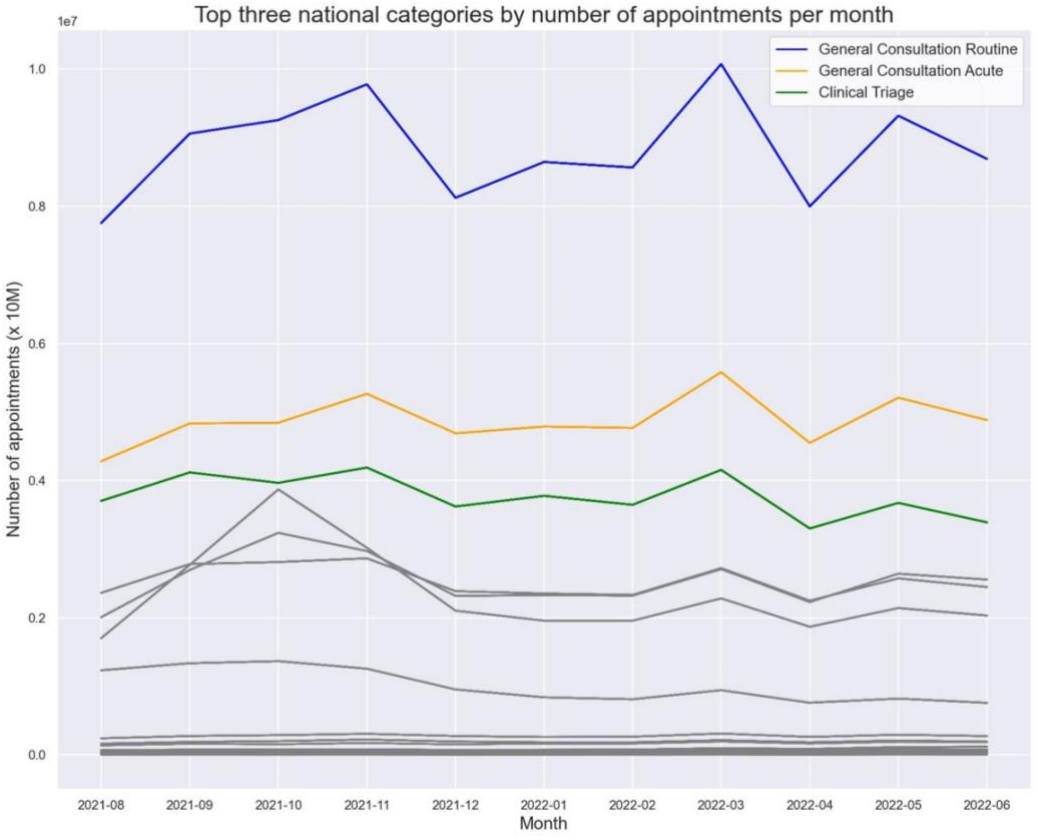
The vast majority of appointments in national_categories DataFrame are GP appointments. Three noticeable peaks related to number of GP appointments were identified by the above visualisation.

- An additional data related to new daily COVID cases were obtained from https://coronavirus.data.gov.uk/details/cases?areaType=nation&areaName=England
  Correlation between number of new COVID cases and GP appointments per month were investigated for the period: August 2021-June 2022. A pattern showing that public self-isolation caused by positive COVID testing might have reduced number of GP appointments were observed but it has to be further investigated.



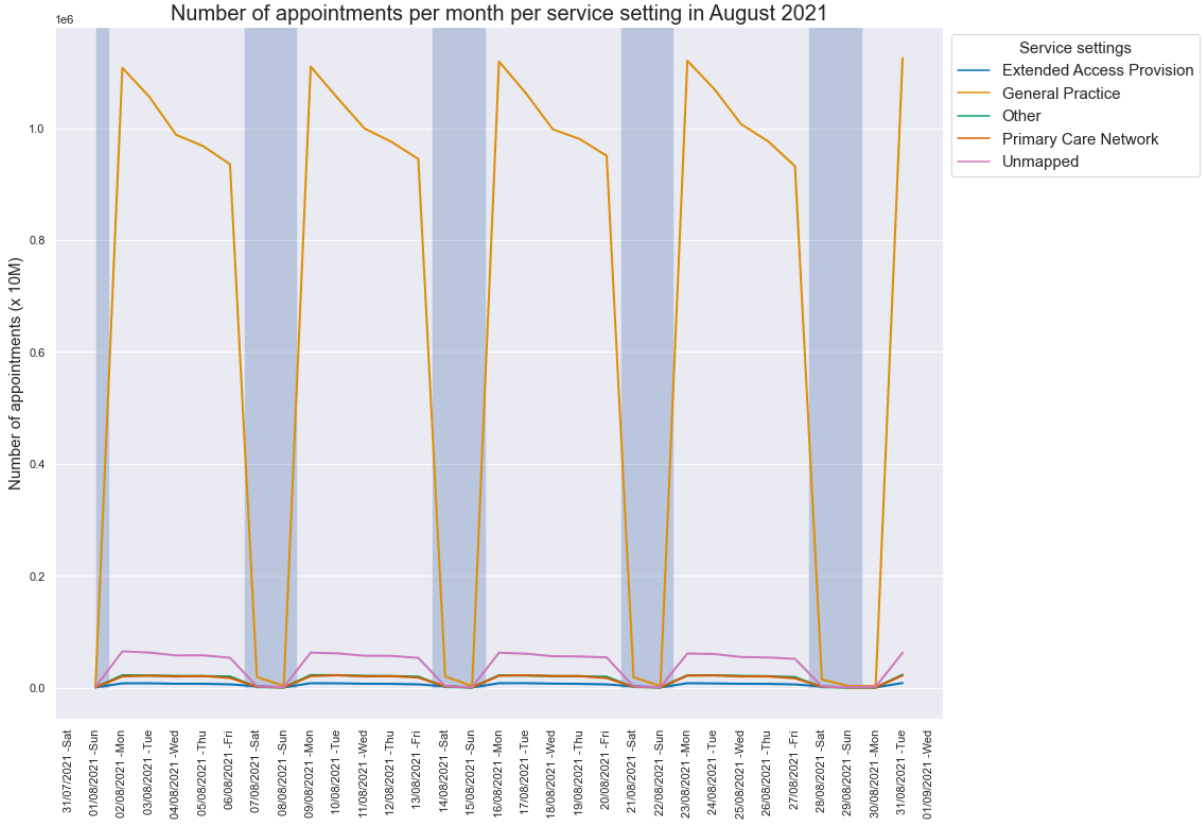
- Trends similar to the one described earlier (related to the number of GP appointments per month) were identified for the top lines of 'context type' and 'national category' groups:

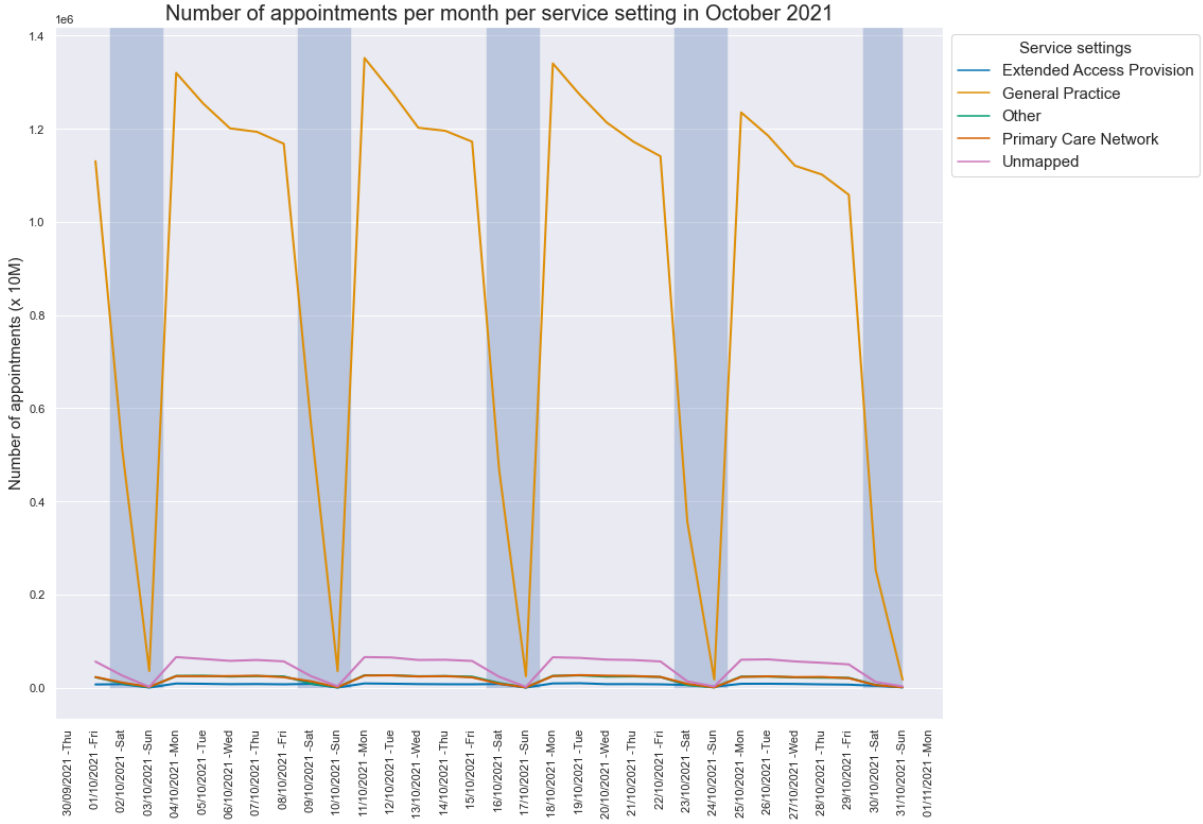


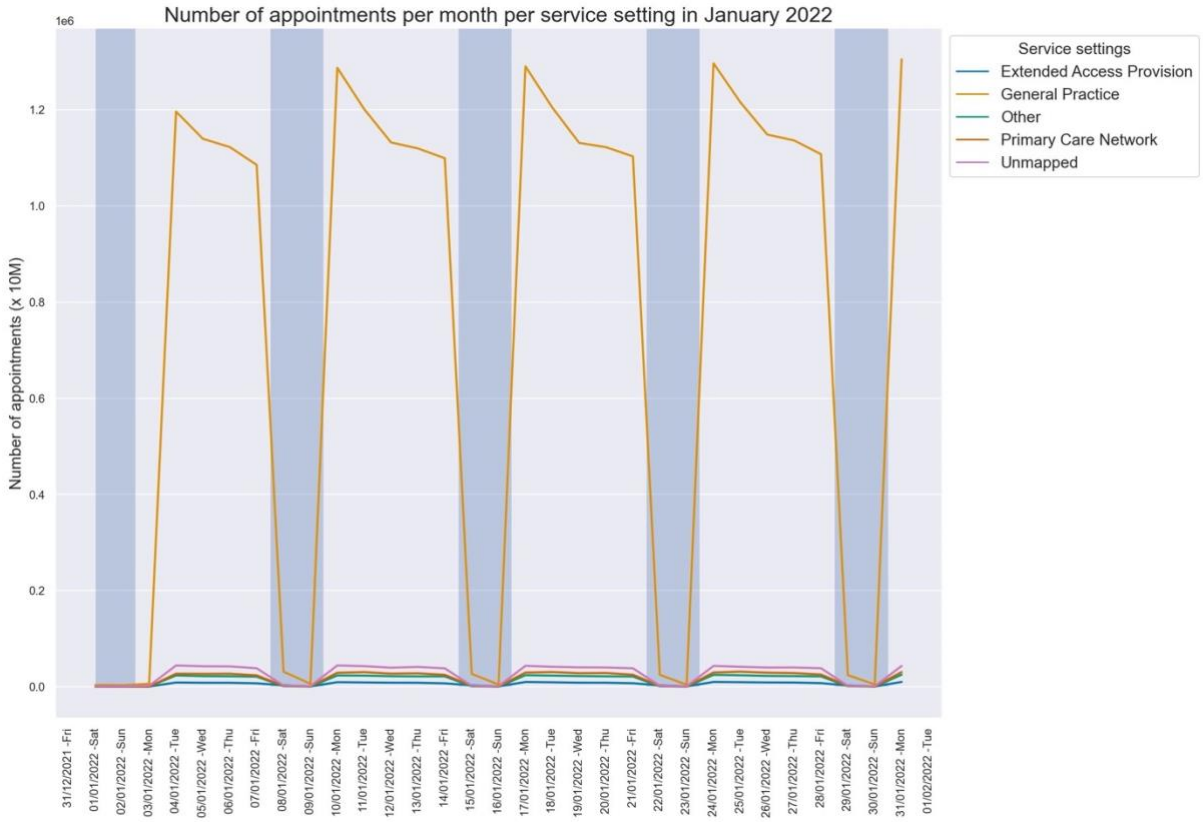


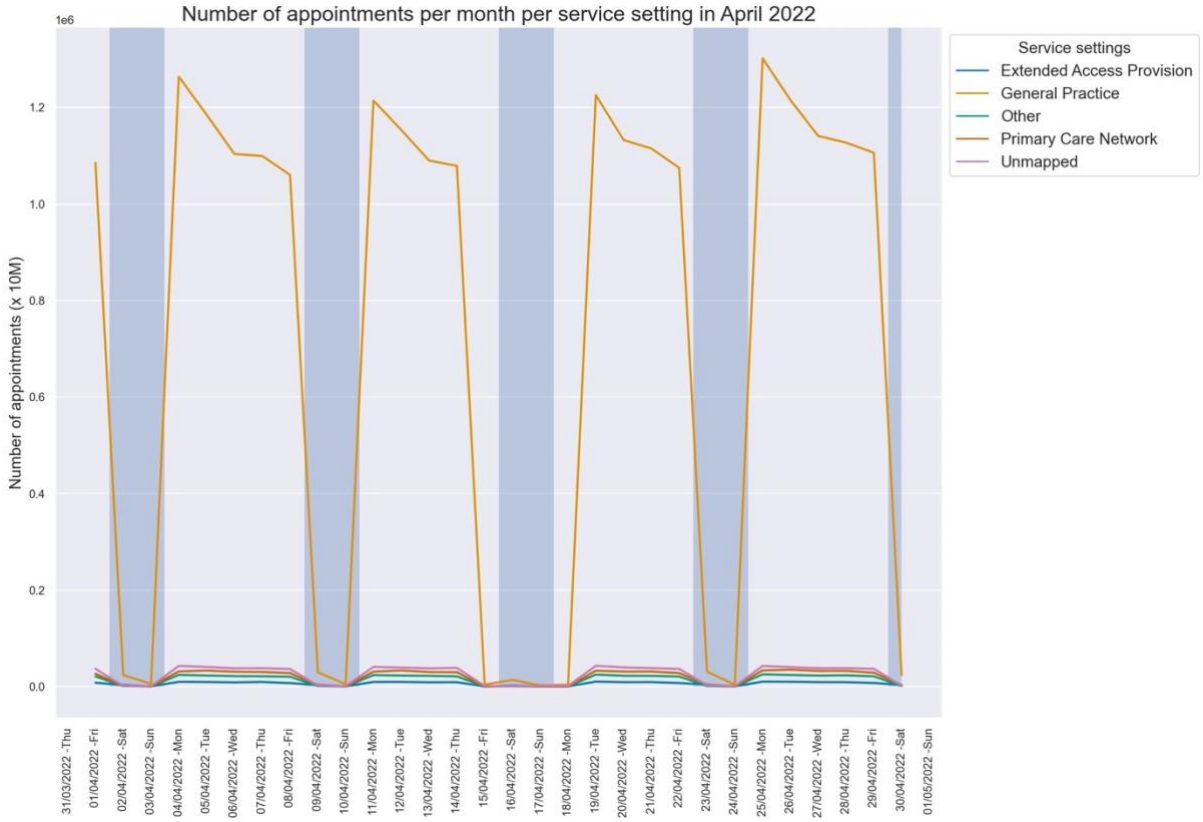
- Four months were selected to identify any seasonal patterns (August 2021, October 2021, January 2022 and April 2022). Regardless of a season, more appointments were booked for the first half of the week, noticeably fewer were booked for Saturdays, almost none were

booked for Sundays (can be caused by the majority of service settings being closed during the weekend) and none for Bank Holidays, when service settings are closed. Less appointments were booked around Bank Holidays, possibly because people often on holidays/away around this time.









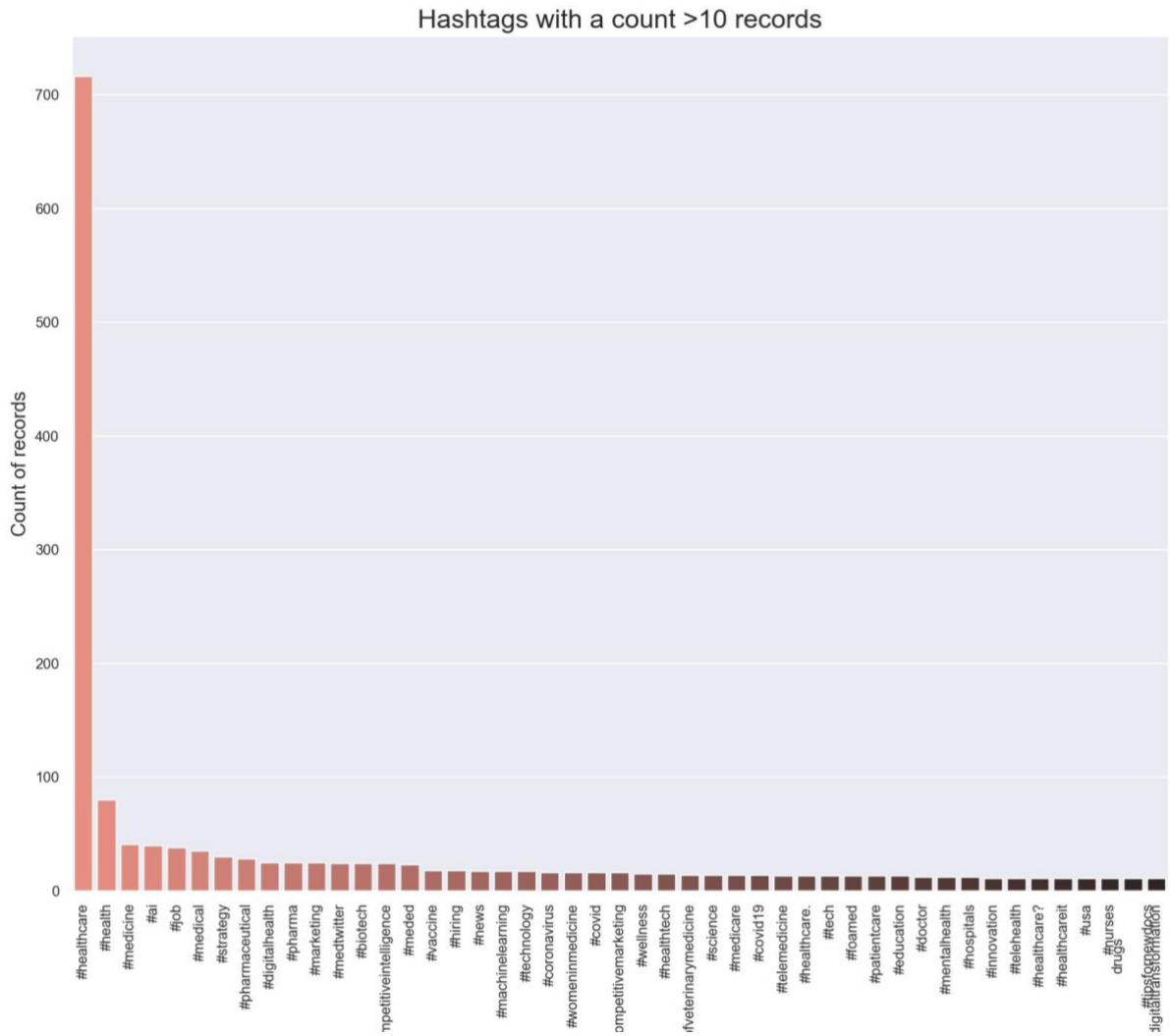
2.1.5. Exploration of Twitter data.

- The workstation was prepared by importing the necessary libraries and tweets.csv file in the existing Python3 file.
- The file was converted to a DataFrame and sense-checked.

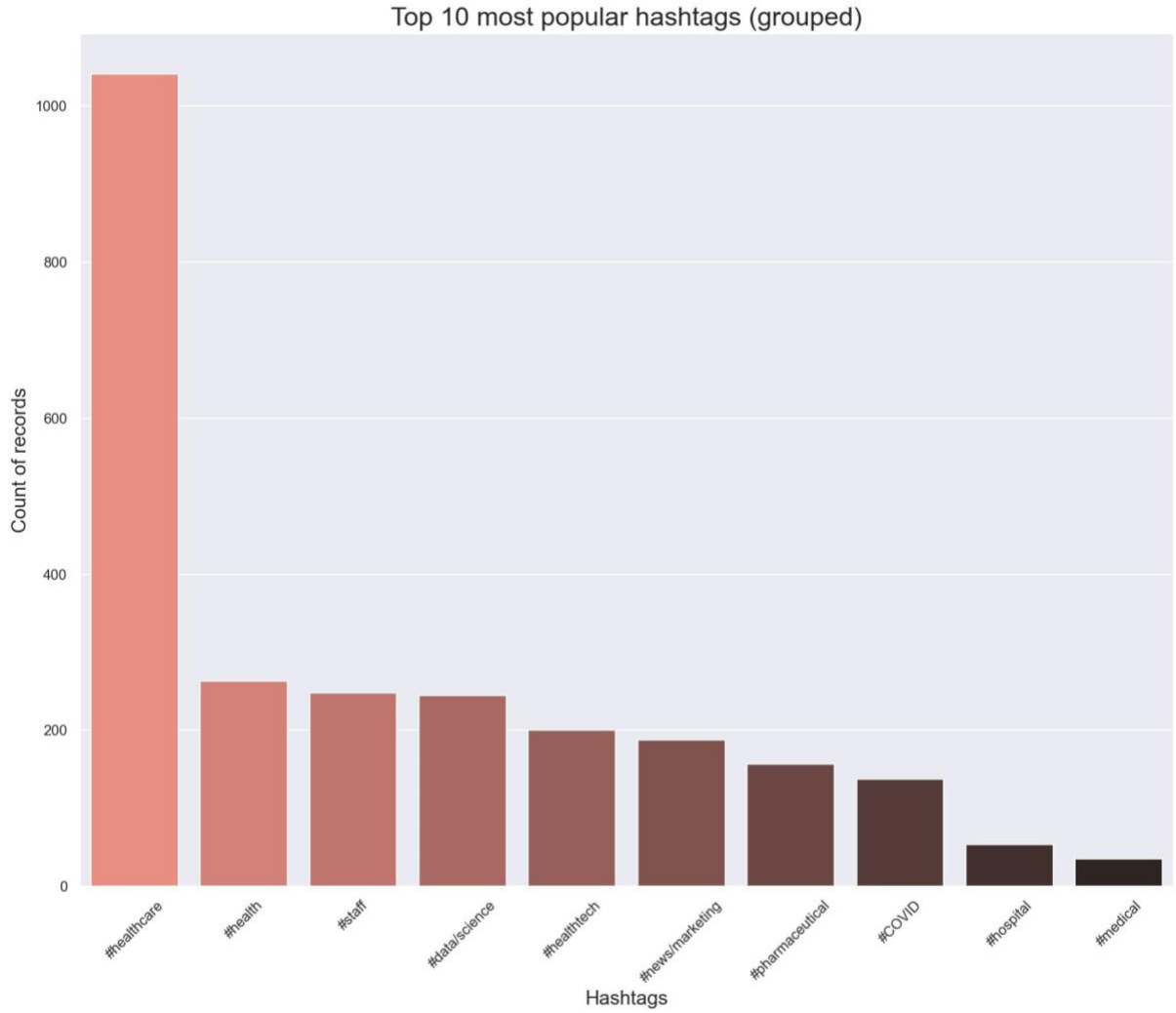
- 167 missing values were identified in the 'tweet_entities_hashtags' column.
- 1754 individual hashtags were extracted from the DataFrame and grouped in a new DataFrame by number of time used.
- #healthcare is the most frequently used hashtag in the data set (used in 16.5% of the identified cases).
- 74.86% of all hashtags were used only once in the data set.
- The data set was searched for hashtags which include words: nhs, nationalhealthservice, healthservice, uk, england. Below is the output of the search. Only one hashtag, highlighted in blue, is clearly related to healthcare in the UK. Identified hashtags represent 0.46% of all hashtags used in the data set.

	Hashtag	Number of time used
1	#gdpuk	4
2	**#uk**	1
3	**#nhs**	1
4	to:\nhttps://t.co/wukk9hk1vz\n\n#hospital	1
5	#standwithukraine	1

- Most frequently used hashtags:



- Hashtags with similar meanings were grouped in nine larger groups (first nine hashtags on the below visualisation.

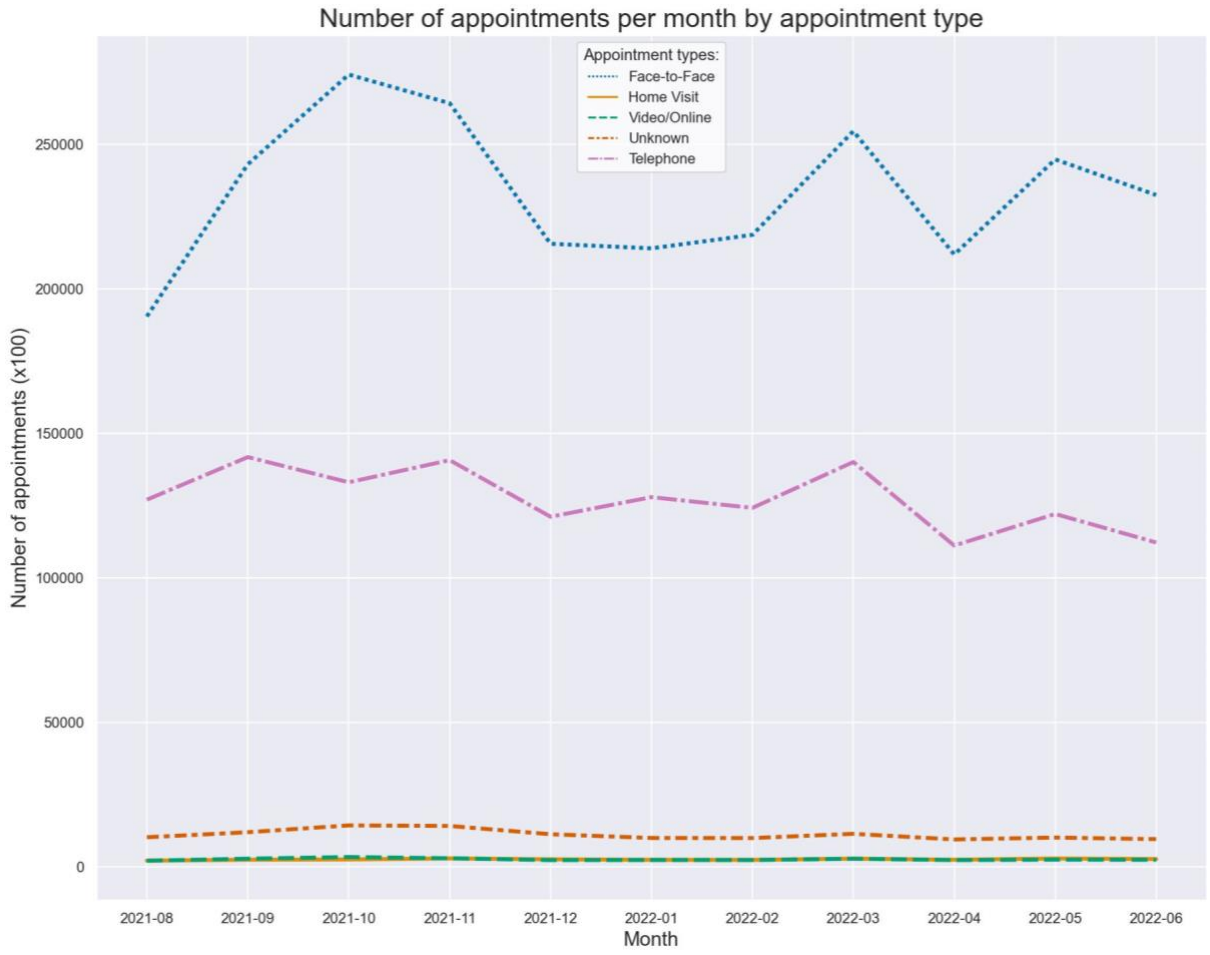


2.1.6. Exploration of data related to staff and capacity in the networks.

- Number of records by appointment mode in appointments_regional DataFrame between August 2021 and June 2022:

	Number of appointments
Face-to-Face	787
Telephone	787
Home Visit	786
Unknown	786
Video/Online	608

Below visualisation shows that despite the fact that number of records were almost split equally between different appointment modes, majority of appointments were done face-to-face followed by telephone appointments. Face-to-face appointments have more uneven spread with clear peaks comparing to other appointment types.



- According to the provided information NHS can accommodate a maximum of 1,200,000 appointments per day. The highest level of the appointment utilisation was in November 2011 and equals 84.46%:

	Month	Number of appointments	Avg appointments per day (fact)	NHS max appointments per day (plan)	Utilisation, %%
1	2021-08	23852171	795072.4	1200000	66.26
2	2021-09	28522501	950750.0	1200000	79.23
3	2021-10	30303834	1010127.8	1200000	84.18
4	**2021-11**	**30405070**	**1013502.3**	**1200000**	**84.46**
5	2021-12	25140776	838025.9	1200000	69.84
6	2022-01	25635474	854515.8	1200000	71.21
7	2022-02	25355260	845175.3	1200000	70.43
8	2022-03	29595038	986501.3	1200000	82.21
9	2022-04	23913060	797102.0	1200000	66.43
10	2022-05	27495508	916516.9	1200000	76.38
11	2022-06	25828078	860935.9	1200000	71.74

- Further investigation was done to identify level of the appointment utilisation by month by area and by location. The assumption was made that if NHS can accommodate a maximum of 1,200,000 appointments per day in total, then it should be equal 11,320.75 (1,200,000 / 106 sub-ICB locations) appointments per day per location or 28,571.43 (1,200,000 / 42 ICBs) appointments per day per area.
  Number of sub-ICB locations per month with the level of the appointment utilisation higher than 100% in national_categories DataFrame:

Year-Month	Number of sub-ICB locations
2021-10	34
2021-11	33
2022-03	33
2022-05	32
2021-09	31
2021-12	26
2022-01	26
2022-02	25
2022-06	25
2022-04	24
2021-08	23

Five months with the highest number of sub-ICB locations are the same as months identified with the highest number of appointments.

34 sub-ICB locations (= 32.08% of all sub-ICBs) and number of months when they had on average more daily appointments than maximum number calculated above.

	Sub-ICB location	Number of months
1	NHS Bath and North East Somerset Swindon and Wiltshire ICB - 92G	11
2	NHS Kent and Medway ICB - 91Q	11
3	NHS West Yorkshire ICB - 15F	11
4	NHS Sussex ICB - 70F	11
5	NHS Surrey Heartlands ICB - 92A	11
6	NHS South West London ICB - 36L	11
7	NHS South East London ICB - 72Q	11
8	NHS Bedfordshire Luton and Milton Keynes ICB - M1J4Y	11
9	NHS North West London ICB - W2U3Z	11
10	NHS North East London ICB - A3A8R	11
11	NHS North Central London ICB - 93C	11
12	NHS Norfolk and Waveney ICB - 26A	11
13	NHS Nottingham and Nottinghamshire ICB - 52R	11
14	NHS Herefordshire and Worcestershire ICB - 18C	11
15	NHS Dorset ICB - 11J	11
16	NHS Devon ICB - 15N	11
17	NHS Derby and Derbyshire ICB - 15M	11
18	NHS Coventry and Warwickshire ICB - B2M3M	11
19	NHS Cambridgeshire and Peterborough ICB - 06H	11
20	NHS Bristol North Somerset and South Gloucestershire ICB - 15C	11
21	NHS Black Country ICB - D2P2L	11
22	NHS Birmingham and Solihull ICB - 15E	11
23	NHS Hampshire and Isle Of Wight ICB - D9Y0V	11
24	NHS Lincolnshire ICB - 71E	10
25	NHS West Yorkshire ICB - 36J	9
26	NHS Northamptonshire ICB - 78H	7
27	NHS North East and North Cumbria ICB - 16C	5
28	NHS Buckinghamshire Oxfordshire and Berkshire West ICB - 10Q	5
29	NHS Frimley ICB - D4U1Y	5
30	NHS Cornwall and The Isles Of Scilly ICB - 11N	5
31	NHS Cheshire and Merseyside ICB - 27D	5
32	NHS Gloucestershire ICB - 11M	4
33	NHS North East and North Cumbria ICB - 84H	3
34	NHS Somerset ICB - 11X	1

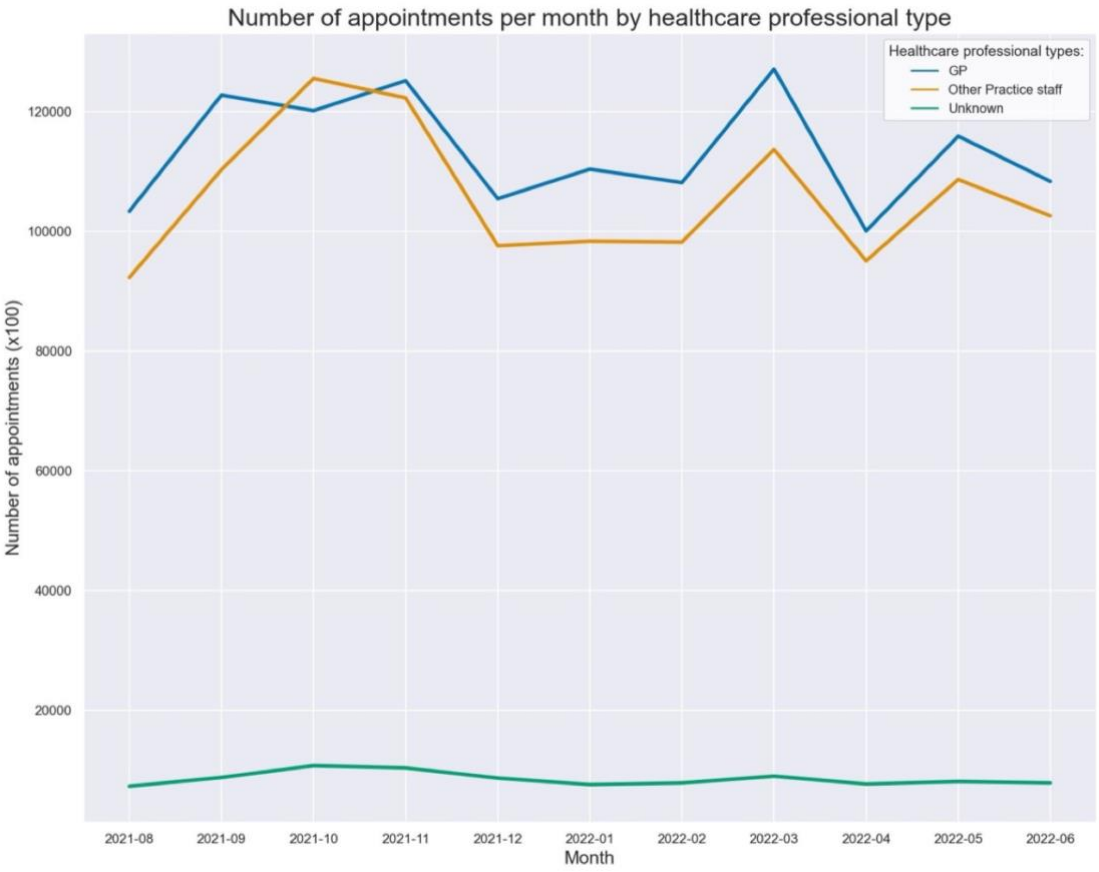
Number of ICB areas per month with the level of the appointment utilisation higher than 100% in national_categories DataFrame:

Year-Month	Number of ICB areas
2021-10	12
2021-11	12
2022-03	12
2021-09	11
2022-05	7
2022-06	6
2021-08	5
2021-12	5
2022-01	5
2022-02	5
2022-04	5

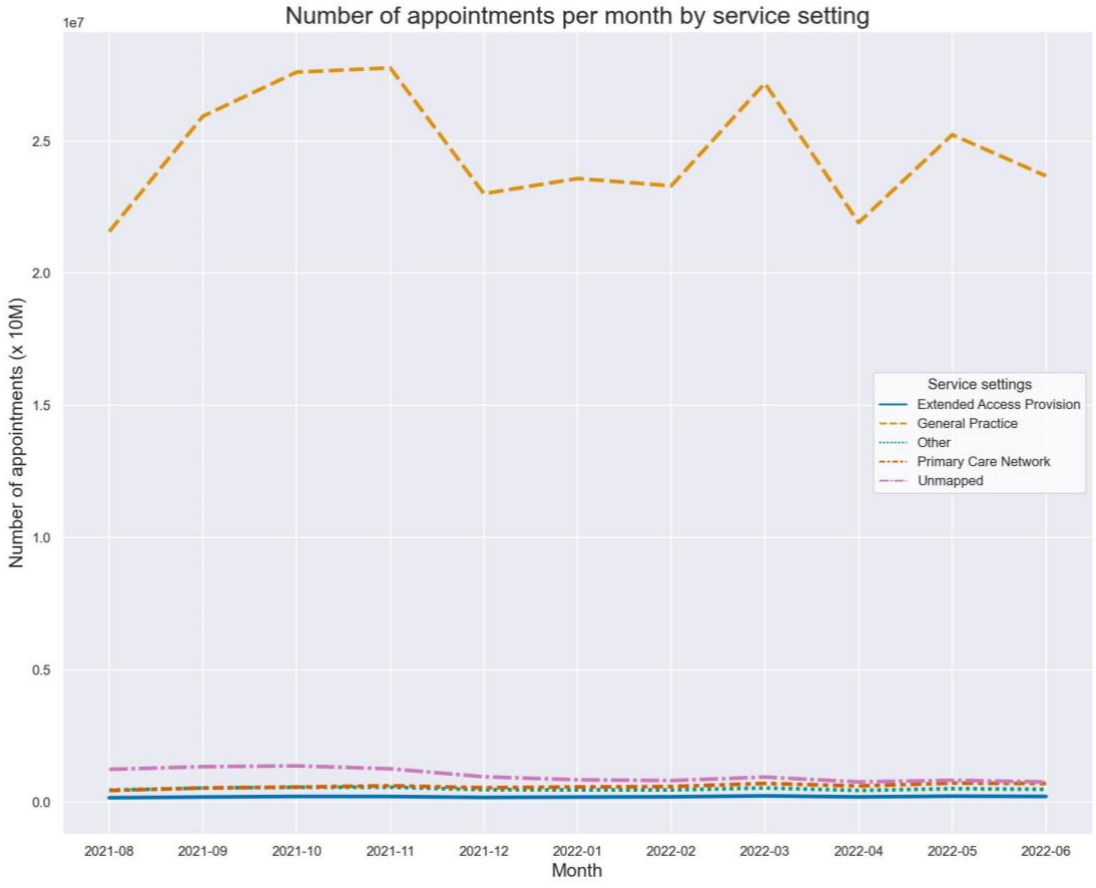
Five months with the highest number of ICB areas are the same as months identified with the highest number of appointments.
12 sub-ICB locations (= 28.57% of all ICBs) and number of months when they had on average more daily appointments than maximum number calculated above.

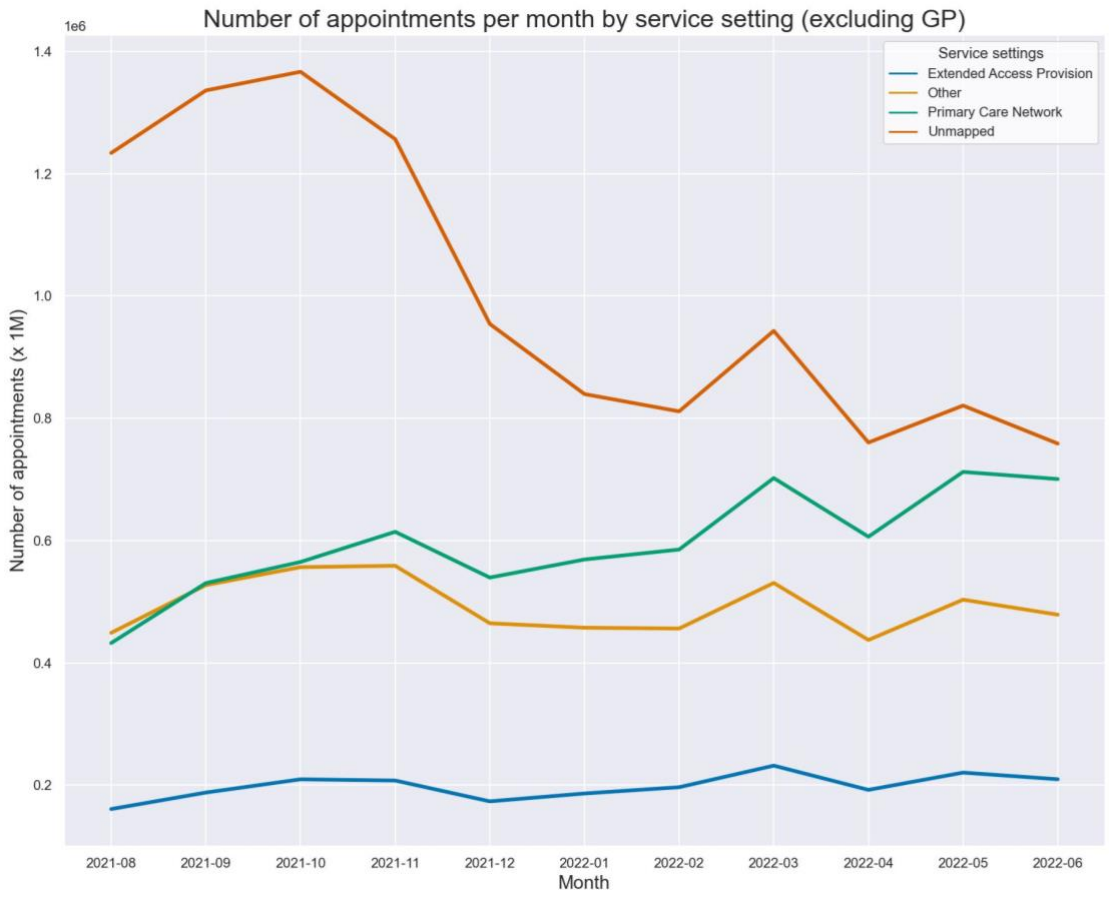
	ICB area	Number of months
1	NHS Cheshire and Merseyside ICB	11
2	NHS North West London ICB	11
3	NHS North East and North Cumbria ICB	11
4	NHS West Yorkshire ICB	11
5	NHS Greater Manchester ICB	11
6	NHS North East London ICB	6
7	NHS Hampshire and Isle Of Wight ICB	5
8	NHS Kent and Medway ICB	4
9	NHS Lancashire and South Cumbria ICB	4
10	NHS Humber and North Yorkshire ICB	4
11	NHS Sussex ICB	4
12	NHS Buckinghamshire, Oxfordshire and Berkshire West ICB	3

- According to provided data in appointments_regional DataFrame between August 2021 and June 2022 GPs (GP registrar/Locum GP/Principal GP) run the highest number of appointments per month apart from October 2021, which needs to be further investigated.



- According to provided data in national_categories DataFrame the highest number of appointments per month within identified service settings were done by GPs.





Whilst number of appointments in the above visualisation conducted by an 'unmapped' service setting has a downward trend, number of appointments conducted by other service settings have upward trends. Further investigation is needed to understand if two above trends are related.

3. Recommendations and questions for further investigation.

- In total average utilisation of service in NHS is below the set threshold suggesting that the organisation has enough capacity to handle the total amount of appointments. Nevertheless, 34 sub-ICBs and 12 ICBs had at least one month between August 2021 and June 2022 when they had on average more daily appointments than expected by NHS.
  67.65% of mentioned sub-ICBs and 41.67% of mentioned ICBs went above maximum threshold of average daily appointments in every month during the investigated period.
  Those sub-ICBs and ICBs are at higher risk of being understaffed.
  More detailed information for identified sub-ICBs and ICBs, including utilisation of service percentages, is available in Fedorova_Oksana_DA201_Assignment_Notebook.ipynb
- Autumn months (November and October 2021 in particular) had the highest number of appointments followed by two spring months (March and May 2022). Additional staff resources might be required for those seasons.
  The data did not have information about the nature/health issues of the appointments but if it can be grouped (e.g. respiratory infections related to the change of seasons), additional public education through social-media campaigns about how to avoid getting those infections might have a positive impact on reducing the number of appointments.
- More appointments were booked for the first half of the week, possibly because the majority of services are closed during the weekend, suggesting that more staff resources need to be allocated to the first half of the week. Less appointments were booked around Bank Holidays, possibly because a lot of people are away/on holidays around this time.

Majority of service settings are either closed or work reduced hours on Saturdays. Detailed information regarding opening hours were not included in the data set, therefore it was not possible to investigate the popularity of weekends among patients further.

- Majority of appointments were booked as Face-to-Face, followed by telephone appointments. Video/Online appointments were among the least popular appointment types but have more advantages than telephone appointments and closer by nature to Face-to-Face appointments. Further investigation should be conducted about the reasons why Video/Online appointments were not conducted more often and potential benefits of these appointments.
- More appointments were run by GP (GP registrar/Locum GP/Principal GP) between August 2021 and June 2022 apart from October 2021, when more appointments were run by 'Other Practice Staff'. Which needs to be further investigated.
- Highest number of attended appointments were booked on the same day (48.45%); followed by appointments booked 2-7 days prior (20.38%) and 8-14 days prior (11.08%), suggesting that the shorter the time between the date of booking the appointment and appointment date, the higher the chance that it will be attended.
- Majority of appointments with known length were 6-10 minutes (20.12%); followed by 1-5 minutes appointments (17.03%) and 11-15 minutes (14.98%). Nature of shorter appointments needs to be further investigated and any alternatives how to reduce the number of those appointments need to be explored.
- 91.24% of all appointments and 94.26% GP appointments in the data set were attended.
- Twitter search highlighted that NHS needs to encourage public to use more tailored to the organisation hashtags for getting more meaningful insights during further researches.
- Further investigation is needed to identify reasons for unmapped/inconsistent mapping/unknown data entries in the data set in order to improve outputs of further analytical projects.