Optimising Staff Levels and Capacity in the NHS Networks

&

Identifying trends in the utilisation of resources over time

National Health Services (NHS)

Data Analytics Report

Diagnostic Analysis using Python

Cristina Alonso Robles

May 2023

Methodology

1. **Background/context of the business.**

2. **Analytical Approach, using Python.**

A. **Importing and exploring the data sets.**

  o Sorting, cleaning, and organising the datasets.
  o Initial Insights.

B. **Performing further analysis.**

C. **Visualising and identifying initial trends.**

D. **Analysing Twitter data.**

E. **Final Insights.**

1. Background/context of the business

- **Organisation and Business Definition:** National Health Services (NHS), a publicly funded healthcare system in England.

- **Industry**: Health and Social Care.

- **Sector**: Public sector.

- **NHS's Business Problem:**

The NHS incurs in significant, potentially avoidable, costs when patients miss general practitioner (GP) appointments.

- **Why should the NHS try and solve this problem?**

While it is frustrating when patients do not attend, the reasons why this happens should be investigated rather than simply resorting to punishing them. Financially penalising patients inevitably impact the poorest and most vulnerable in the community (GP Practice News 2022).

Therefore, reducing or eliminating missed appointments would be beneficial financially as well as socially.

- **NHS's Business Objectives:**

1. Understand the staff level and capacity in the networks over time.

2. Identify trends in the utilisation of resources.

- **Stakeholders involved:** NHS board of directors.

- **Five Whys framework approach:**

Q1: Why the NHS incurs in significant, potentially avoidable, costs when patients miss general practitioner (GP) appointments?

A1: Because they do not know the reasons of the missed appointments.

Q2: Why the NHS do not know the reasons of the missed appointments?

A2: Because they are not correctly gathering and analysing their patient's appointments data.

Q3: Why the NHS is not correctly gathering and analysing their patient's appointments data.

A3: Because they did not have analytical resources and they thought that analysing data was a waste of time and a big investing in resources, such as data analytics tools and data analysts.

Q4: Why the NHS did not think about this earlier?

A: Because until now they had not had enough robust data highlighting the significant, costs when patients miss general practitioner (GP) appointments

Q5: Why the NHS want to change their business approach now?

A5: Because they realised that financially penalising patients inevitably impact the poorest and most vulnerable in the community, therefore, reducing or eliminating missed appointments would be beneficial financially as well as socially.

2. Analytical Approach

A. Importing and exploring the data sets

- o Importing and sense-checking the four data sets provided as well as the metadata file.
- o Determining missing and erroneous values in the data sets.

It might be worth further investigating these issues in the future and identifying the ICB and Sub ICB Locations linked to these mapping errors. For the purpose of this analysis, I will continue to work with the entire data sets that contain these mapping errors to give insights to the stakeholders on the scale of the issue.

- o **Determining the number of locations, service settings, context types, national categories, and appointment statuses in the data sets.**

1. Number of locations: 106
2. Number of service settings: 5
3. Number of context types: 3
4. Number of national categories: 18
5. Number of appointment statuses: 3

- o **Identifying the five locations with the highest number of records.**

1. NHS North West London ICB – W2U3Z – 13,007 records
2. NHS Kent and Medway ICB - 91Q   - 12,637 records
3. NHS Devon ICB – 15N – 12, 526 records
4. NHS Hampshire and Isle Of Wight ICB - D9Y0V   - 12,171 records
5. NHS North East London ICB – A3A8R – 11,837 records

B. Performing further analysis

○ **Identifying the date range of the provided data sets.**

Actual Duration Data Set

First Day Registered: 01 December 2021

Last Day Registered: 30 June 2022

National Category Data Set

First Day Registered: 01 August 2021

Last Day Registered: 30 June 2022

Appointments Regional Data Set

There is not an 'appointment date' column in the 'ar' DataFrame, hence the day is missing. Therefore, I will use only the month and the year to confirm the date range in the 'ar' DataFrame.

First Day Registered: January 2020

Last Day Registered: June 2022

Before I can give answers to the following questions, since these questions are not specific to any DataFrame, I assume that these questions refer to the three DataFrames as a whole, so I will merge the three DataFrames.

To be able to merge the three DataFrames, we should have comparable inputs in terms of 'count of appointments' aggregated to monthly counts, and no duplicate values.

Before I merge the three DataFrames, I will create a subset of each DataFrame so that I can include a new column of total_monthly_appointments.

Based on the notes from the metadata, I would use only the appointment_month column from the 'nc' DataFrame, since this column reflects the month in which the appointment is. However, when I search for duplicates, Python will identify duplicates when the appointment_date is not included in the DataFrame, and I do not feel confident removing any duplicates from the 'nc' DataFrame. So, I will include the appointment_date column.

The Kernel died when I tried to merge the DataFrames with the appointment_month columns, hence I decided to group the DataFrames without these columns.

However, after spending some time trying to merge the three DataFrames, I decided to move on with the analysis without the merge, since I was not feeling confident with the outcome that I got, since I could not see the appointment_month, which is the month in which the appointment took place.

To give answer to the next questions, I will use the National Categories DataFrame.

- o **Which service setting was the most popular for NHS North West?**
  General Practice was the most popular service setting.

- o **Which service setting reported the most appointments for a specific period?**

General Practice was the service setting that reported the most appointments from 1 January to 1 June 2022 for **NHS North West.**

- o **Number of records for each of the service settings available for NHS North West London ICB – W2U3**

  1. General Practice: 4,609 records.
  2. Other: 2,858 records.
  3. Primary Care Network: 2,791 records.
  4. Extended Access Provision: 2,415 records.
  5. Unmapped: 334 records.

- o **What is the number of appointments per month?**

Appointment Month	Total Monthly Appointments
Nov-21	30,405,070
Oct-21	30,303,834
Mar-22	29,595,038
Sep-21	28,522,501
May-22	27,495,508
Jun-22	25,828,078
Jan-22	25,635,474
Feb-22	25,355,260

Dec-21	25,140,776
Apr-22	23,913,060
Aug-21	23,852,171

o   **What is the number of records per month?**

Appointment Month	Total Records
Mar-22	82,822
Nov-21	77,652
May-22	77,425
Sep-21	74,922
Jun-22	74,168
Oct-21	74,078
Dec-21	72,651
Jan-22	71,896
Feb-22	71,769
Apr-22	70,012
Aug-21	69,999

C. Visualising and identifying initial trends

○ **What monthly and seasonal trends are evident, based on the number of appointments for service settings, context types, and national categories?**
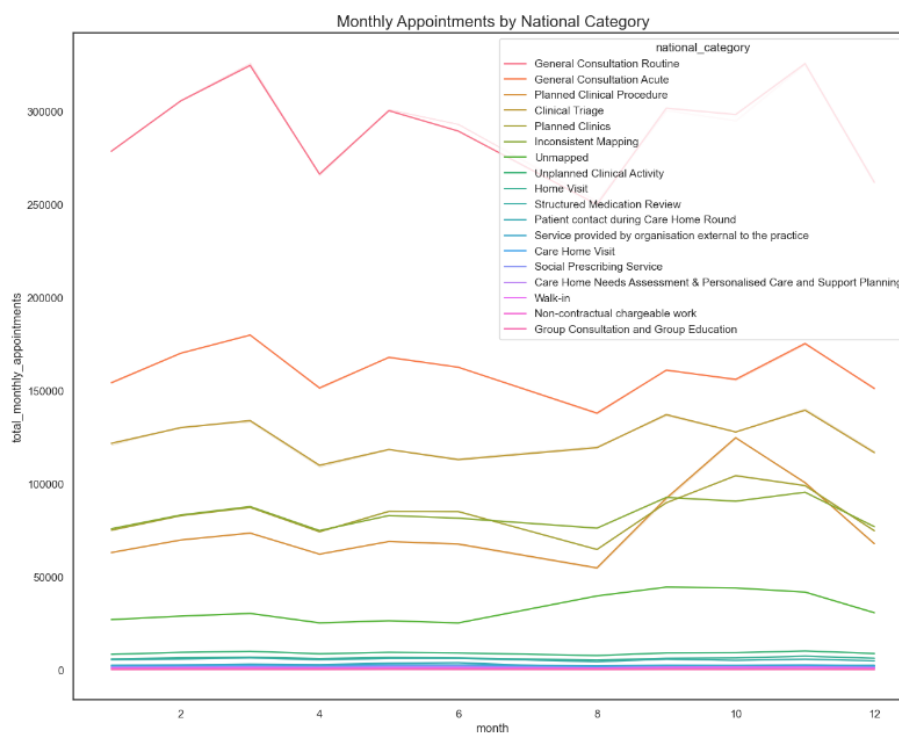
General Practice seems to be the service setting with the highest number of records in the data sets and the higher number of appointments per month, over time.
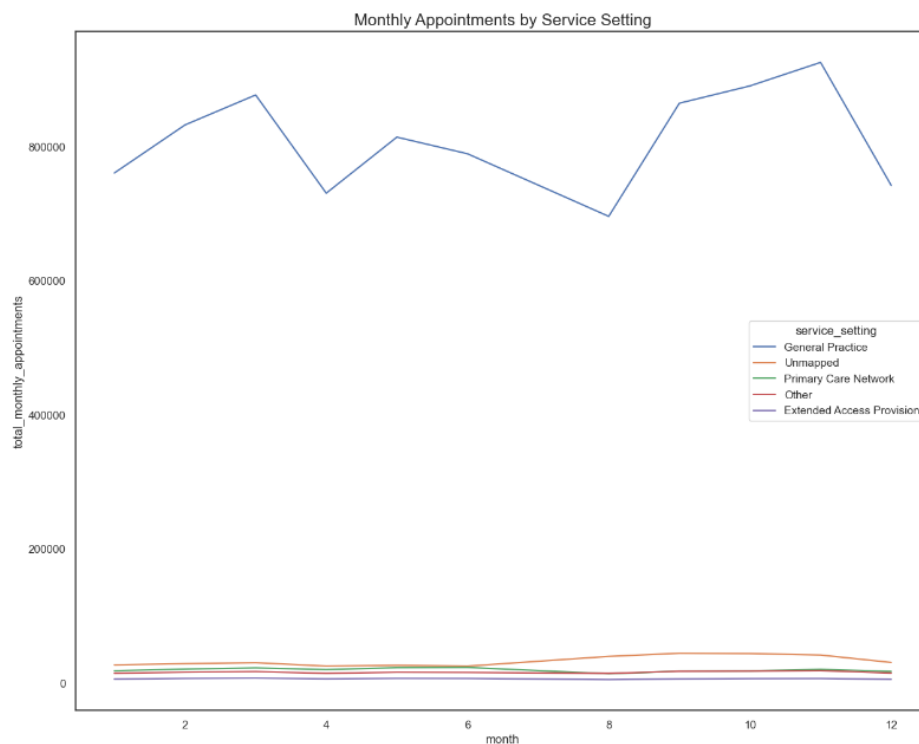
Care Related Encounter is the context type that reported the highest number of appointments per month, over time.
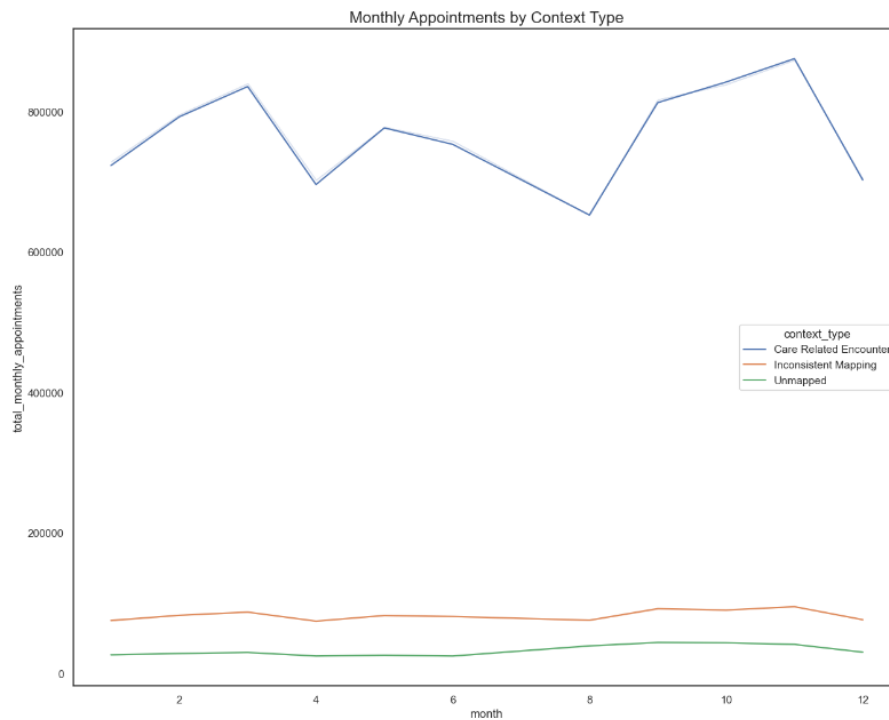

General Consultation Routine is the national category with the highest number of appointments per month over time, followed by **General Consultation Acute.**

Seasonal trend in appointments by Service Setting:

General Practice is the service setting that reported the highest number of appointments across the three seasons studied in this analysis, Summer 2021, Autumn 2021, Winter 2022, Spring 2022.

Winter 2022 - Appointments

Spring 2022 - Appointments
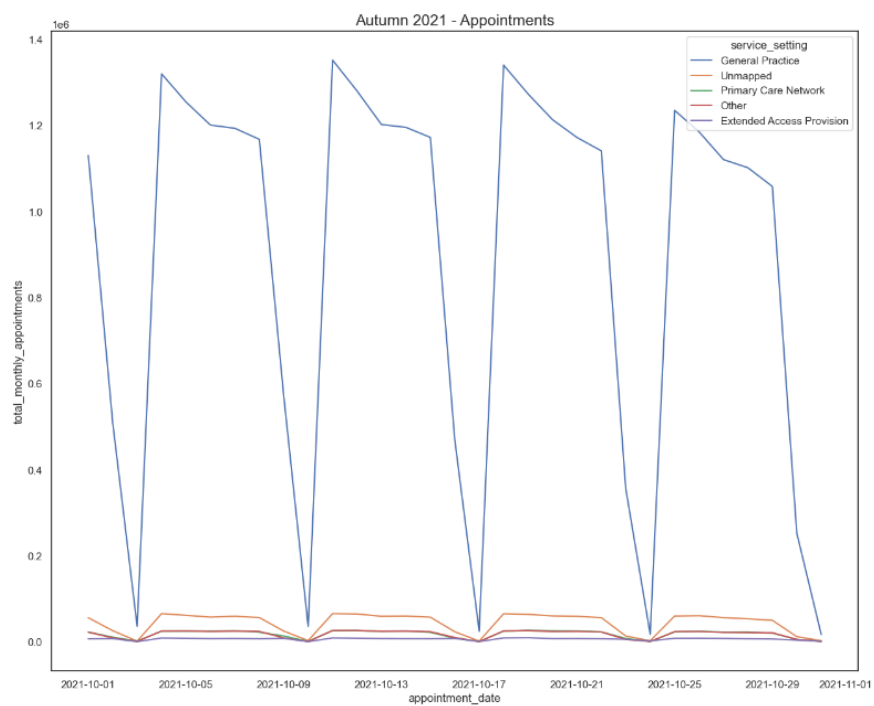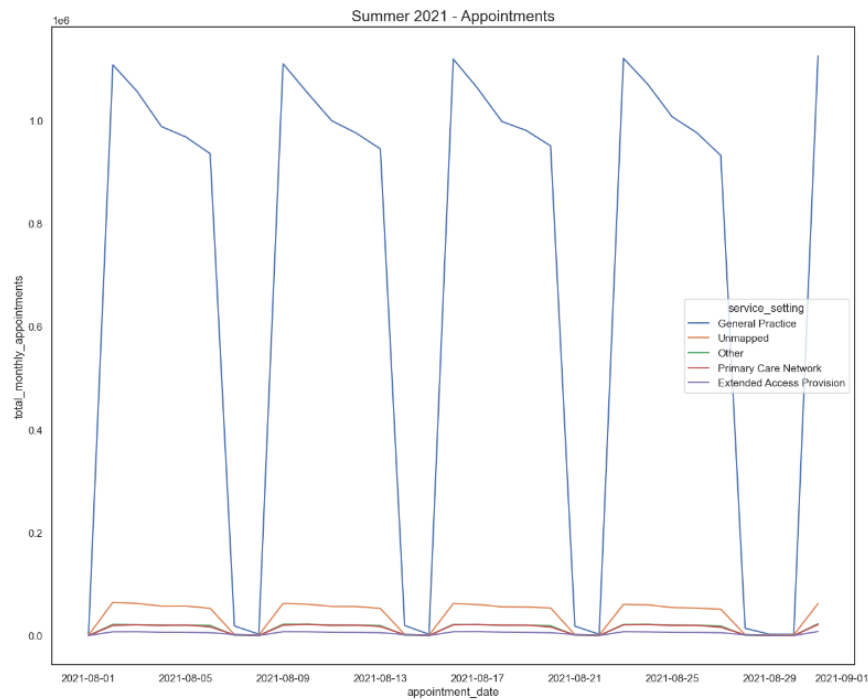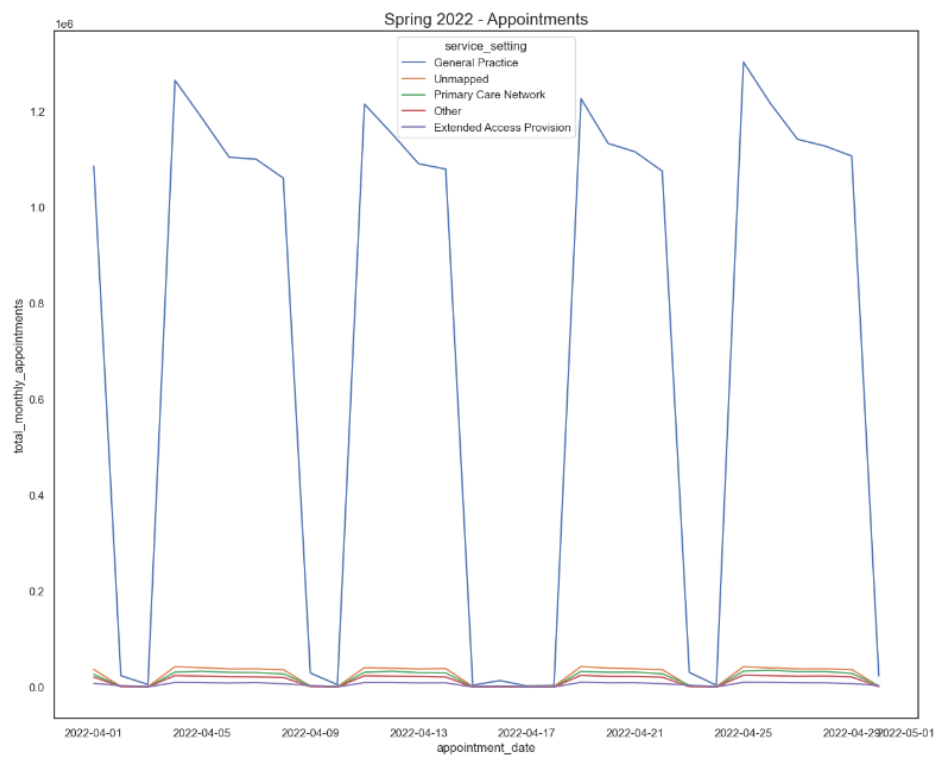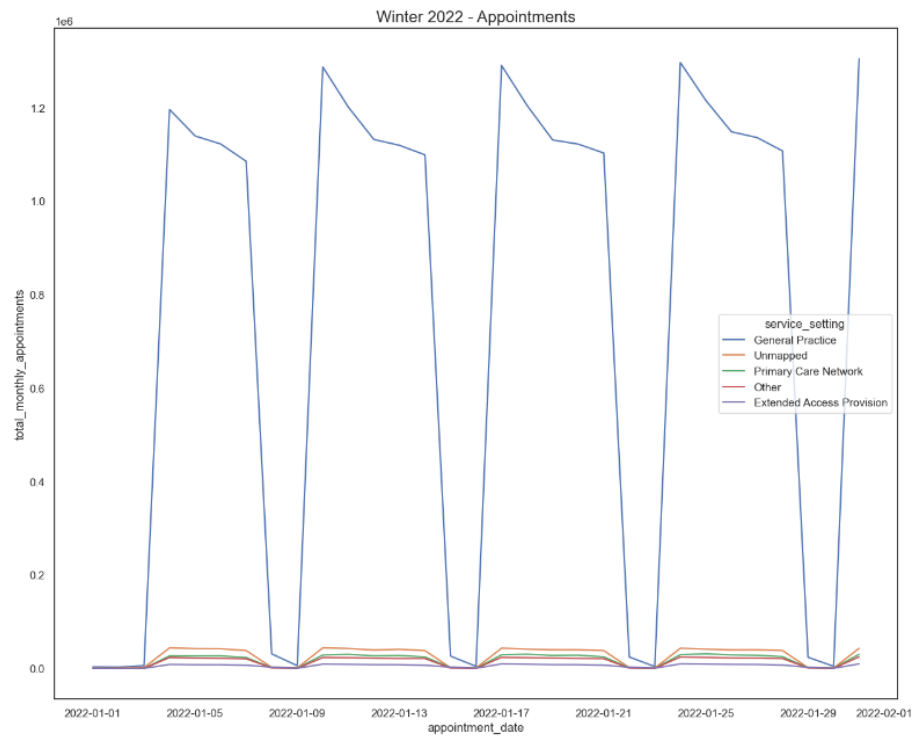
D. Analysing Twitter data

By looking at the highest number of retweets and the tweets with the highest number of 'likes', we will be able to identify the most popular tweets.

Afterwards, we will be able to look at the full text and hashtags used across the filtered tweets.

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

#healthcare, # digitalcare, #patientecare, #telemedicine, #healthtech, #telemedicine, #medtwitter, #healthcareit
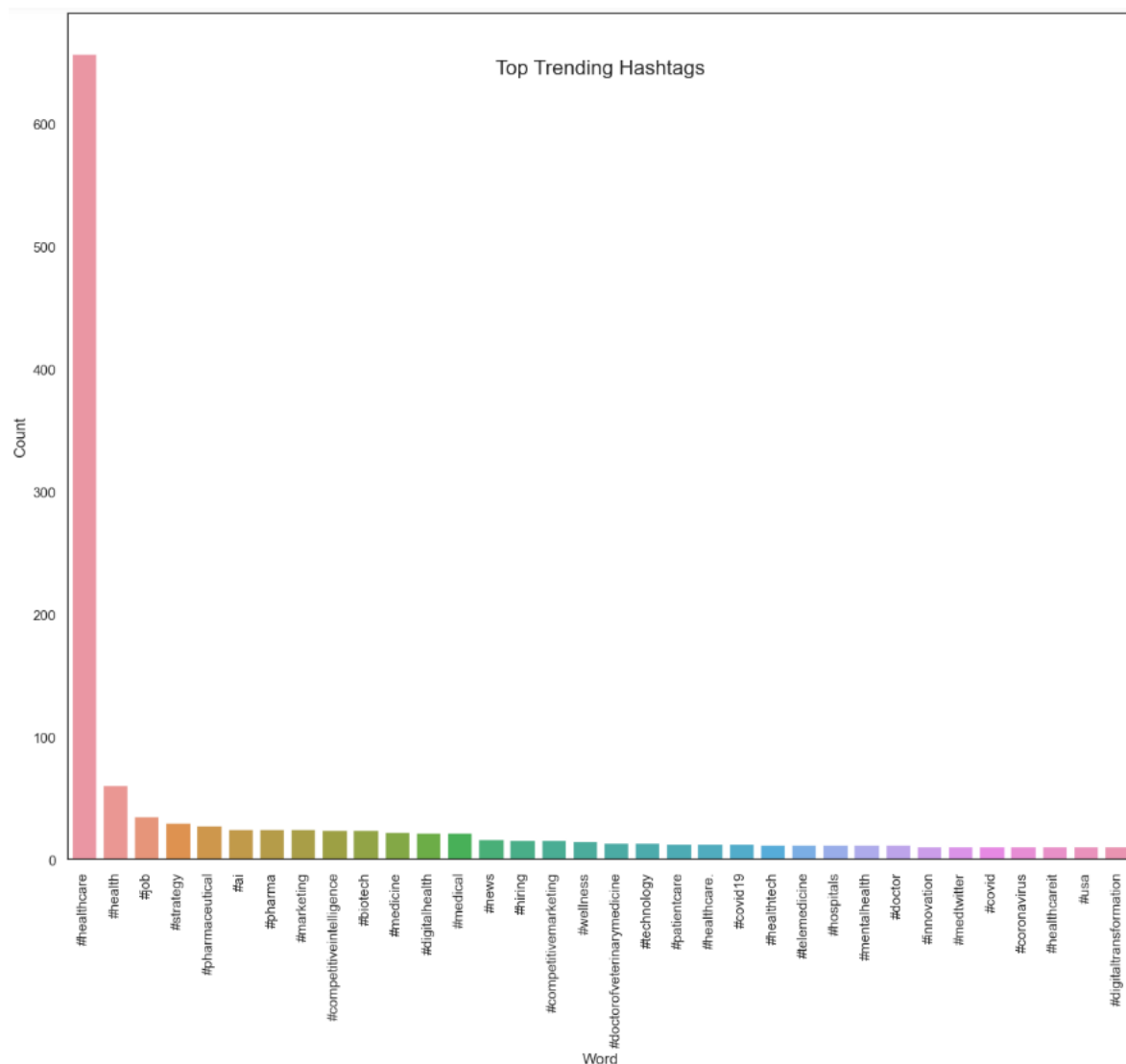
Top **#healthcare** hashtags with the highest retweets.

	tweet_id	tweet_full_text	tweet_entities	tweet_entities_hashtags	tweet_metadata	tweet_retweet_count
263	1.570000e+18	RT @UltimaLionsDen: Temitope is looking to boost efficiency in the Health industry with his app- \nWe-Care Africa. \n\nWas his pitch in good s...	{'hashtags': [], 'symbols': [], 'user_mentions': [{'screen_name': 'UltimaLionsDen', 'name': "Ultima Lions Den", 'id': 973581329119039488, 'id_str': '973581329119039488', 'indices': [3, 18]}], 'ur...	NaN	{'iso_language_code': 'en', 'result_type': 'recent'}	303
330	1.570000e+18	RT @khaleejtimes: .@BurjeelHoldings, a private healthcare services provider, has partnered with superstar @iamsrk as its new brand ambassad...	{'hashtags': [], 'symbols': [], 'user_mentions': [{'screen_name': 'khaleejtimes', 'name': 'Khaleej Times', 'id': 18860860, 'id_str': '18860860', 'indices': [3, 16]}, {'screen_name': 'BurjeelHoldin...	NaN	{'iso_language_code': 'en', 'result_type': 'recent'}	208

	tweet_id	tweet_full_text	tweet_entities	tweet_entities_hashtags	tweet_metadata	tweet_retweet_count
27	1.570000e+18	RT @khaleejtimes: .@BurjeelHoldings, a private healthcare services provider, has partnered with superstar @iamsrk as its new brand ambassad...	{'hashtags': [], 'symbols': [], 'user_mentions': [{'screen_name': 'khaleejtimes', 'name': 'Khaleej Times', 'id': 18860860, 'id_str': '18860860', 'indices': [3, 16]}, {'screen_name': 'BurjeelHoldin...	NaN	{'iso_language_code': 'en', 'result_type': 'recent'}	207
1130	1.570000e+18	RT @imedverse: I.V Drug Calculations Cheat Sheet\n\n@Pharma_Connect @imedverse #TipsForNewDocs #MedEd #MedTwitter #medicine #medical #medicar...	{'hashtags': [{'text': 'TipsForNewDocs', 'indices': [77, 92]}, {'text': 'MedEd', 'indices': [93, 99]}, {'text': 'MedTwitter', 'indices': [100, 111]}, {'text': 'medicine', 'indices': [112, 121]}, {...	#TipsForNewDocs, #MedEd, #MedTwitter, #medicine, #medical	{'iso_language_code': 'en', 'result_type': 'recent'}	169
779	1.570000e+18	RT @Khulood_Almani: 🔟#Applications of #AI in #healthcare \n\nv/@ValueCoders\n\n#digitalhealth #Healthtech #Health #DataScience #Bigdata #Analy...	{'hashtags': [{'text': 'Applications', 'indices': [21, 34]}, {'text': 'AI', 'indices': [38, 41]}, {'text': 'healthcare', 'indices': [45, 56]}, {'text': 'digitalhealth', 'indices': [76, 90]}, {'tex...	#Applications, #AI, #healthcare, #digitalhealth, #Healthtech, #Health, #DataScience, #Bigdata	{'iso_language_code': 'en', 'result_type': 'recent'}	150
790	1.570000e+18	RT @Khulood_Almani: #Healthcare #DigitalTransformation Starts with This 6️⃣Step Process\n\nv/@uptophealth\n#digitalhealth #AI #Python #DataScie...	{'hashtags': [{'text': 'Healthcare', 'indices': [20, 31]}, {'text': 'DigitalTransformation', 'indices': [32, 54]}, {'text': 'digitalhealth', 'indices': [103, 117]}, {'text': 'AI', 'indices': [118,...	#Healthcare, #DigitalTransformation, #digitalhealth, #AI, #Python	{'iso_language_code': 'en', 'result_type': 'recent'}	107

Top **#healthcare** hashtags with the highest number of 'likes'.

tweet_full_text	tweet_entities	tweet_entities_hashtags	tweet_metadata	tweet_favorite_count
Lipid-Lowering Drugs\n\n#TipsForNewDocs #MedEd #MedTwitter #medicine #medical #medicare #health #healthcare #FOAMed #ClinicalPearl #clinicaltips #MedStudents #medstudenttwitter #lipid \n\nCredit: ...	{'hashtags': [{'text': 'TipsForNewDocs', 'indices': [22, 37]}, {'text': 'MedEd', 'indices': [38, 44]}, {'text': 'MedTwitter', 'indices': [45, 56]}, {'text': 'medicine', 'indices': [57, 66]}, {'tex...	#TipsForNewDocs, #MedEd, #MedTwitter, #medicine, #medical, #medicare, #health, #healthcare, #FOAMed, #ClinicalPearl, #clinicaltips, #MedStudents, #medstudenttwitter, #lipid	{'iso_language_code': 'en', 'result_type': 'recent'}	42
You ready for JCO @_JennyCo ❤️\n\n#Healthcare data powered by @Conste11ation DAG 🔥	{'hashtags': [{'text': 'Healthcare', 'indices': [33, 44]}], 'symbols': [{'text': 'JCO', 'indices': [14, 18]}, {'text': 'DAG', 'indices': [76, 80]}], 'user_mentions': [{'screen_name': '_JennyCo', '...	#Healthcare	{'iso_language_code': 'en', 'result_type': 'recent'}	28
How health insurance works 😂 \n\n#comedy #adulting #healthcare https://t.co/ciksdeoAkb	{'hashtags': [{'text': 'comedy', 'indices': [31, 38]}, {'text': 'adulting', 'indices': [39, 48]}, {'text': 'healthcare', 'indices': [49, 60]}], 'symbols': [], 'user_mentions': [], 'urls': [], 'med...	#comedy, #adulting, #healthcare	{'iso_language_code': 'en', 'result_type': 'recent'}	20

tweet_full_text	tweet_entities	tweet_entities_hashtags	tweet_metadata	tweet_favorite_count
Our nat'l choices re: #healthcare systems aren't the continuum of public or private, but how much we want of:\n\n- fragmented or seamless\n- does simplicity or complexity well\n- prioritizes savin...	{'hashtags': [{'text': 'healthcare', 'indices': [22, 33]}, {'text': 'cdnpoli', 'indices': [270, 278]}], 'symbols': [], 'user_mentions': [], 'urls': []}	#healthcare, #cdnpoli		18
Heart Failure, Myocardial Infarction & immediate Treatment\n\n#TipsForNewDocs #MedEd #MedTwitter #medicine #medical #medicare #health #healthcare #FOAMed #ClinicalPearl #clinicaltips #MedStude...	{'hashtags': [{'text': 'TipsForNewDocs', 'indices': [64, 79]}, {'text': 'MedEd', 'indices': [80, 86]}, {'text': 'MedTwitter', 'indices': [87, 98]}, {'text': 'medicine', 'indices': [99, 108]}, {'te...	#TipsForNewDocs, #MedEd, #MedTwitter, #medicine, #medical, #medicare, #health, #healthcare, #FOAMed, #ClinicalPearl, #clinicaltips, #MedStudents, #medstudenttwitter, #lipid		17
More data that our 13+ 🇨🇦 #healthcare systems fall short of providing adequate access to care, even for those w/ a family physician.\n\nInstead of siloed solutions and further system fragmentation...	{'hashtags': [{'text': 'healthcare', 'indices': [26, 37]}], 'symbols': [], 'user_mentions': [], 'urls': [{'url': 'https://t.co/eZkz5brcYT', 'expanded_url': 'https://www.ctvnews.ca/health/canadians...	#healthcare		14
Looking forward to speaking at #ConV2X on Sep 15! Ping me for a speaker discount if interested! Register at https://t.co/v20ebbXmdO \n\n.@BHTYjournal @hedera @acoerco\n#blockchain #DLT #healthcare...	{'hashtags': [{'text': 'ConV2X', 'indices': [31, 38]}, {'text': 'blockchain', 'indices': [165, 176]}, {'text': 'DLT', 'indices': [177, 181]}, {'text': 'healthcare', 'indices': [182, 193]}, {'text'...	#ConV2X, #blockchain, #DLT, #healthcare, #innovation		13

E. Final Insights

- Using Twitter to leverage NHS customer service.

1. **How our team can utilise tweets to provide feedback to the stakeholders.**

- **Direct messages (DM).** You'll often see people refer to direct messages as "DMs." This option allows you to chat privately with an individual user. DMs have no character limit, so your response can be as detailed as necessary. To send links to interesting tweets that have relevant information for the project at hand and links to external websites.
- **Lists.** You can make public and private lists of other users you find interesting. You don't have to follow a user to include them in a list. For example, you may make a list of local news outlets or writers, your competitors, industry influencers, or important customers. You also have the option to make a list private so only you can see it.

2. **Do tweets add value to the overall project?**

Tweets provide valuable insights from the community. Also, they are up to date with the latest news and trends around the world.

3. **How can the NHS utilise tweets?**

Using Twitter to improve customer service.

- Twitter is an excellent tool for sharing updates, engaging with patients, and sharing links to relevant NHS posts.
- Using hashtags to improve health updates visibility on Twitter.
- Use Twitter to handle customer service issues, solve patients queries and keep them informed about the latest NHS updates.
- Using graphics and polls for more entertaining and interactive content.
- Use third-party Twitter tools to keep your account organized.
- Using Google Analytics to see if your Twitter ads are leading to increased website traffic and sales.

○ **Were there adequate staff and capacity in the networks?**

Important to note! The NHS can accommodate a maximum of 1,200,000 appointments per day.

Question 1: What was the actual utilisation of resources? Should the NHS start looking at increasing staff levels?

If we consider the maximum capacity of the NHS to accommodate a maximum of 1,200,000 appointments per day, we do not think it is needed at this point to increase the staff levels. However, this should regularly be monitored, over time.

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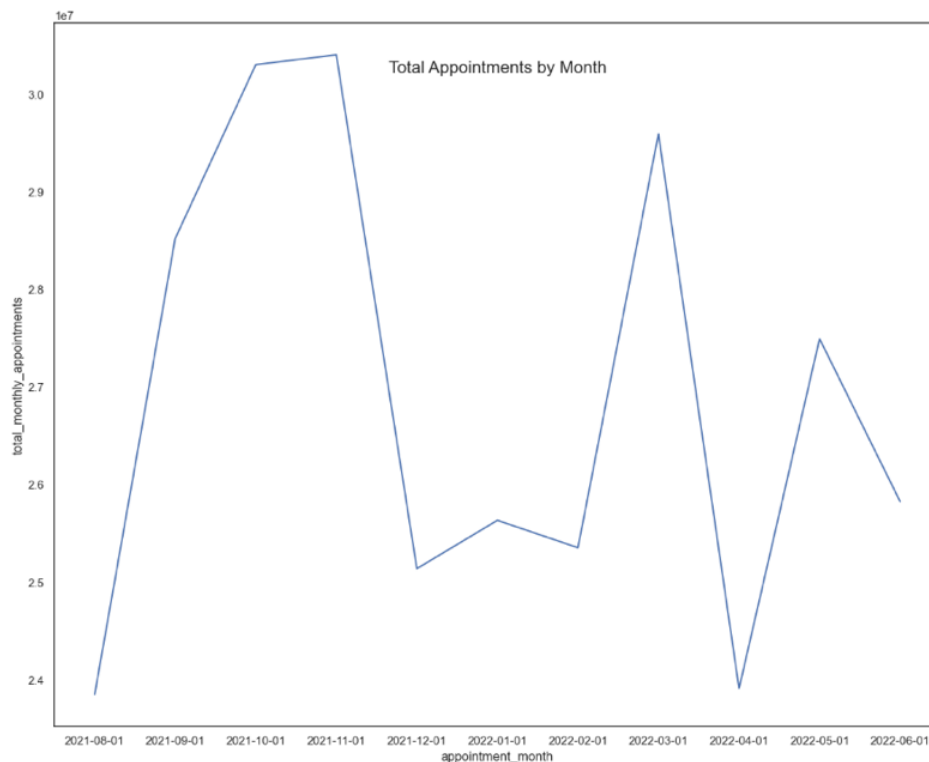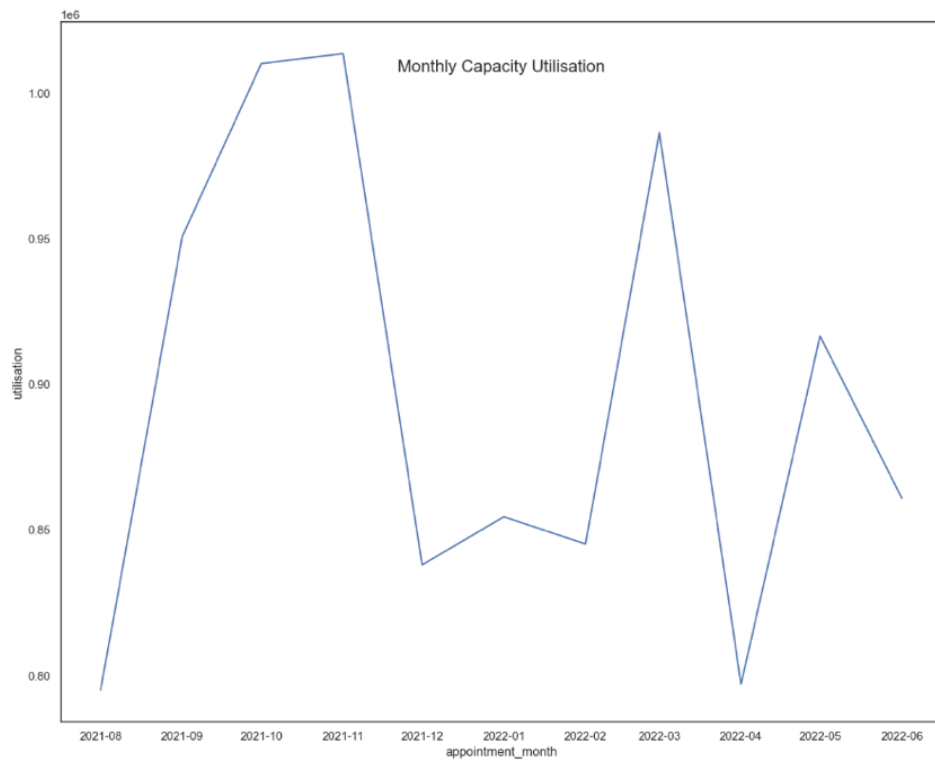
Question 2: How do the healthcare professional types differ over time?

GP is the healthcare professional with the most appointments over time, followed by Other Practices.

Healthcare Professional Types Performance Over Time
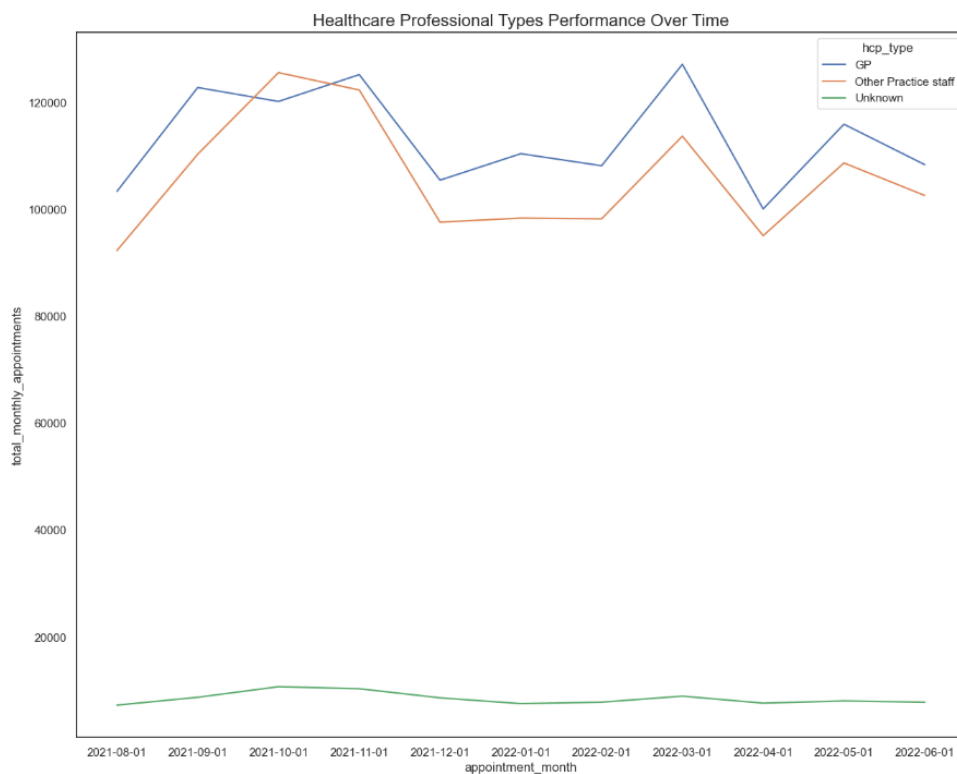
Question 3: Are there significant changes in whether visits are attended?

There are not significant changes in the appointment type over time by looking at the busiest months, and most of the appointments are attended by patients.


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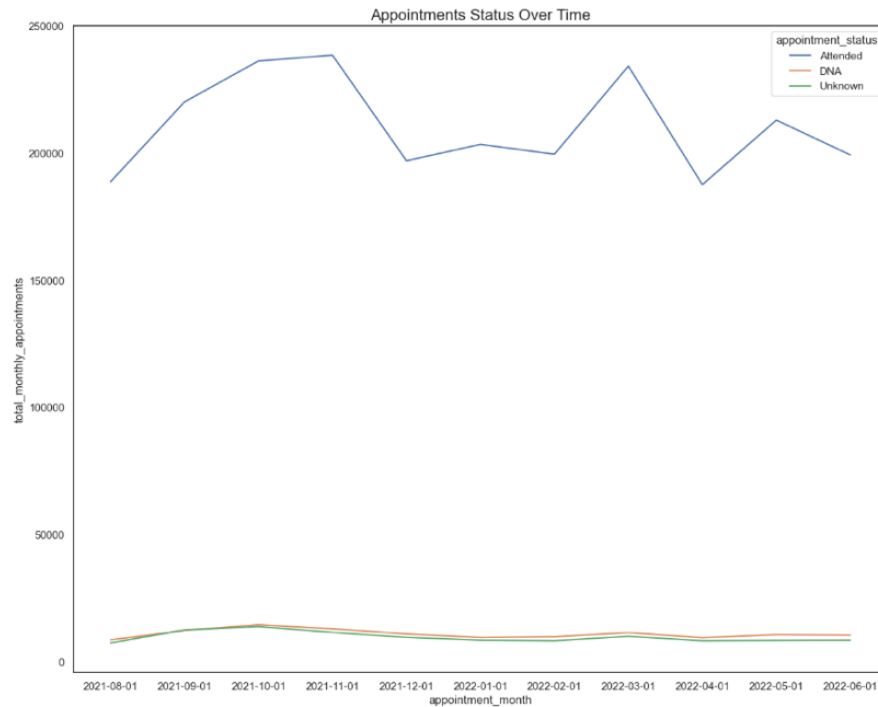
Question 4: Are there changes in terms of appointment type and the busiest months?

Most of the appointments are attended on the same day of the booking, followed by appointments that took between 2 and 7 days in being attended.
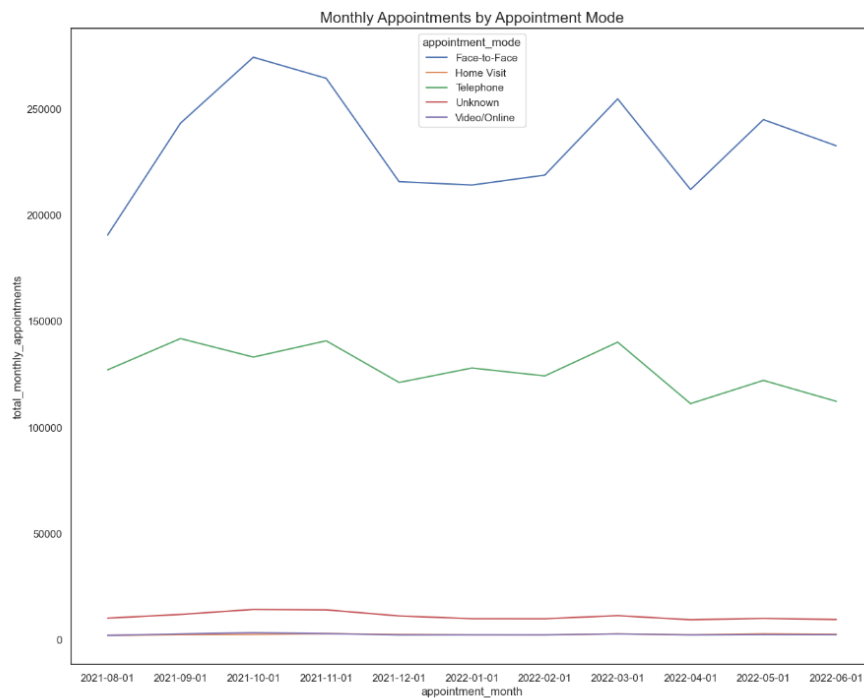

Question 5: Are there any trends in time between booking an appointment?

Most of the appointments are attended on the same day of the booking, followed by appointments that took between 2 and 7 days in being attended.
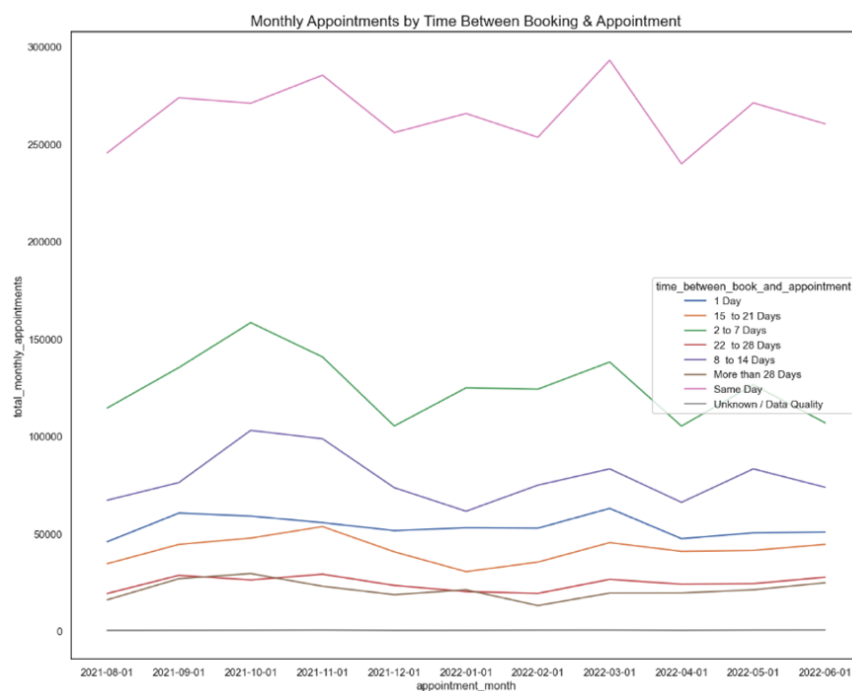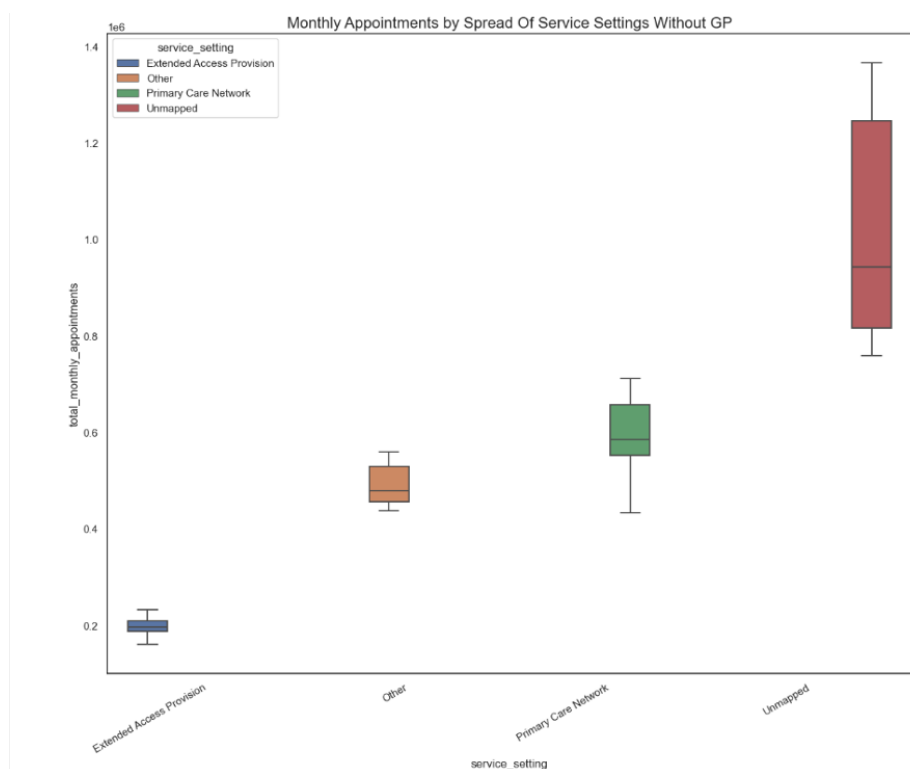
Question 6: How does the spread of service settings compare?

General Practice is by far the service setting that recorded the highest number of appointments by month.
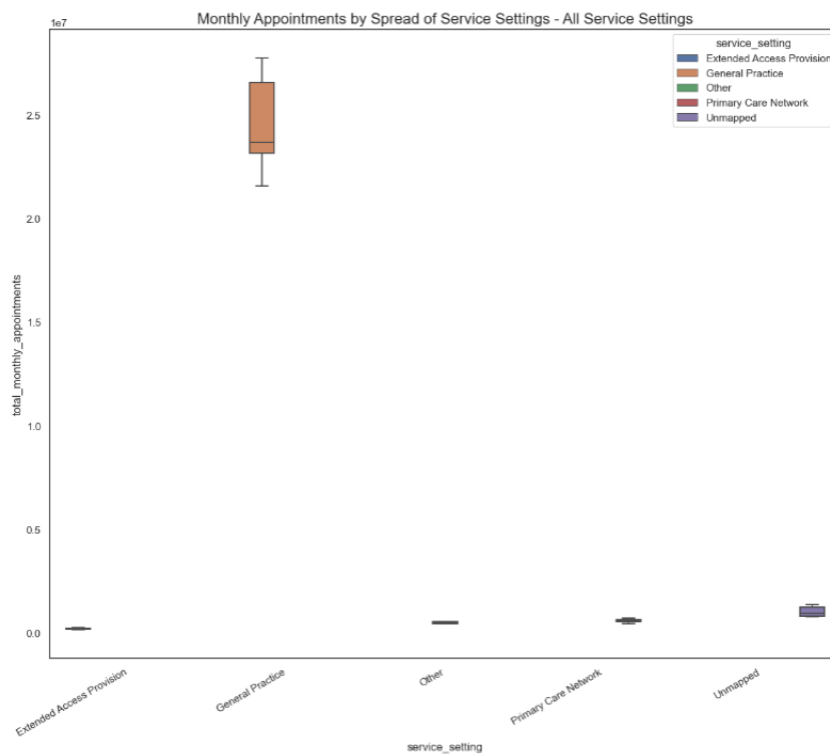
However, when we exclude 'General Practice' from the analysis, we see that the appointments categorised as 'Unmapped' show the highest number of records.

Appendix

- **Determining missing and erroneous values in the data sets.**

  o 20,161 (14.6%) rows from the 'ad' DataFrame were identified as 'Unknown / Data Quality' as the actual duration of the appointment.

  o 201,324 (33.7%) rows from the 'ar' DataFrame were identified as having an 'Unknown' appointment status.

  o 129,228 (21.6%) rows from the 'ar' DataFrame were identified as having an 'Unknown' Healthcare Professional Type.

  o 29,687(4.9%) rows from the 'ar' DataFrame were identified as having an 'Unknown / Data Quality' as the Time Between Book and Appointment.

  o 27,419 (3.4%) rows from the 'nc' DataFrame were identified as having an 'Unmapped' as the Service Setting of the Appointment.

  o 27,419 (3.4%) rows from the 'nc' DataFrame were identified as having an 'Unmapped' as the Context Type of the Appointment. These coincide with the number of rows that have an unmapped Service Setting too.

  o 89,494 (10.9%) rows from the 'nc' DataFrame were identified as having an 'Inconsistent Mapping' as the National Category of the Appointment. These coincide with the number of rows that have an unmapped Context Type too.

  o 27,419 (3.4%) rows from the 'nc' DataFrame were identified as having an 'Unmapped' National Category of the Appointment. These coincide with the number of rows that have an unmapped Service Setting and unmapped Context Type too.

Optimising Staff Levels and Capacity in the NHS Networks

&

Identifying trends in the utilisation of resources over time

National Health Services (NHS)

Data Analyst & Project lead

Cristina Alonso Robles

aroblescristina@gmail.com

https://www.linkedin.com/in/cristina-alonso-robles/

https://github.com/crialorob

May 2023