NHS Utilisation and Capacity Report

Overview and Business Problem

The NHS needs to expand its resourcing and infrastructure. Therefore, we must understand the current utilisation of resources in its network, and future trends. We must also assess if current resources can be made more efficient.

The NHS has provided datasets including appointment data, location data, appointment counts, service settings, and social media content. We looked to answer:

1. How many locations are there, and which locations have the most appointments?
2. How many different types of service settings, contexts, national categories, and appointment statuses exist?
3. What date range does this data cover, and what are the busiest settings and months?
4. What are the monthly and seasonal trends in appointments?
5. What is the social media sentiment regarding the NHS?
6. What are the trends in bookings and appointments?

Based on the findings, the objective is to provide recommendations to the NHS on:

1. What appointment types are most utilised
2. Where and when Missed Appointments are most prevalent(DNAs)
3. The impact of increasing staffing levels

Approach

Pre-analysis

1. Data Quality and Limitations

Quality issues and limitations, along with steps for cleaning and preparing data for use are logged in Appendix 1

2. Definitions

Definitions for Appointment Status, Appointment Mode, Service Setting, and Context Type are logged in Appendix 2, as well as the full type lists for each

Appointment Analysis

The data was analysed to determine:

- The top areas by appointment volume
- The top areas by DNAs
- Trends in Service Setting, Context Type, and National Category, by time and seasonality

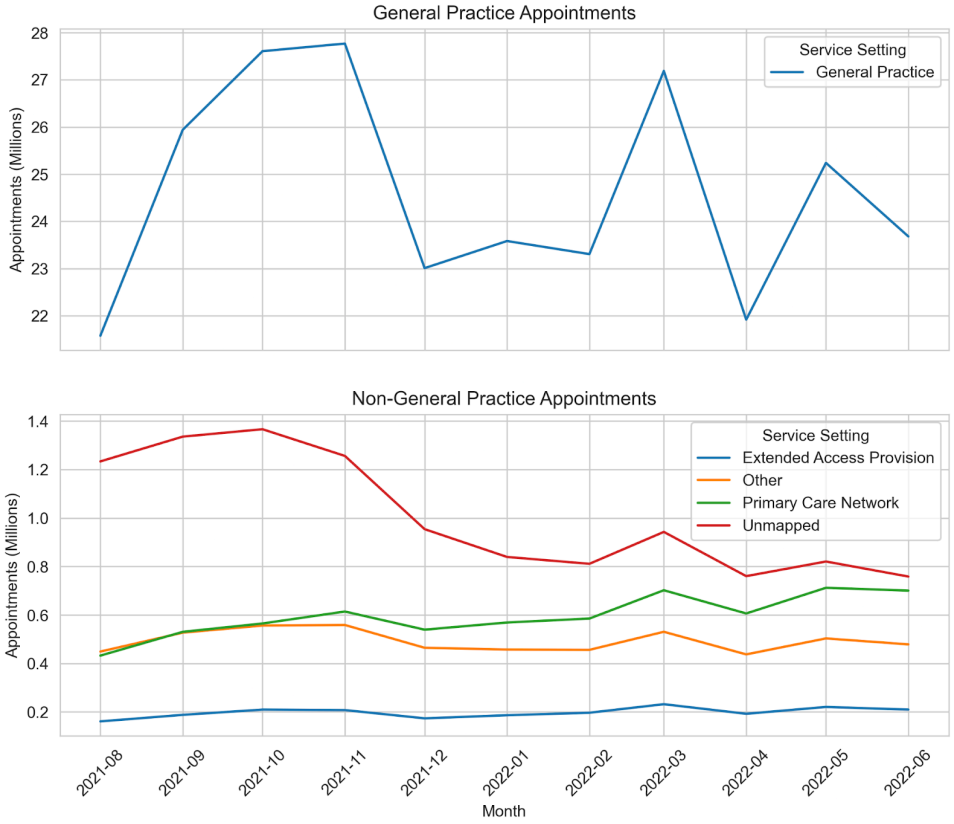
The data contains 106 Sub-ICB Locations (SICBLs) and 42 Integrated Care Boards (ICBs), validated by NHS data

- Top 5 ICBs by appointment volume (covering ~10 million people, per census data)
    - NHS North West London ICB - W2U3Z
    - NHS North East London ICB - A3A8R
    - NHS Kent and Medway ICB - 91Q
    - NHS Hampshire and Isle of Wight ICB - D9Y0V
    - NHS South East London ICB - 72Q

Trends in Service Setting and National Category

Service Setting

- **General Practice**: Appointments rose from August 2021, peaking at >25M in October 2021. This aligns with the Delta wave (and September was also the NHS's busiest September ever). Appointments dropped to ~22M for 3 months, then spiked to >25M in March 2022, likely due to eased restrictions enabling deferred care.
- Approximately 75% of appointments were in General Practise, suggesting other Service Settings may be under utilised

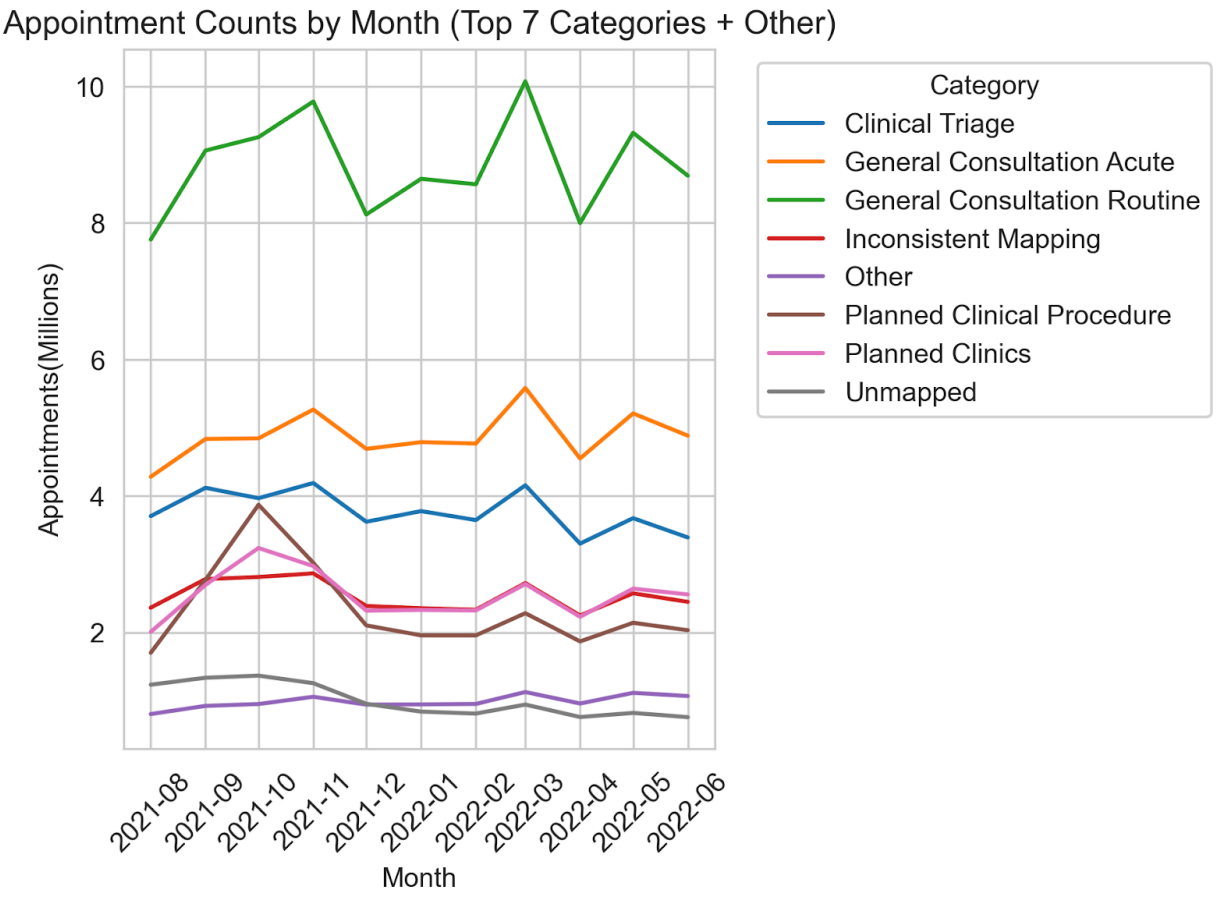


-

National Category

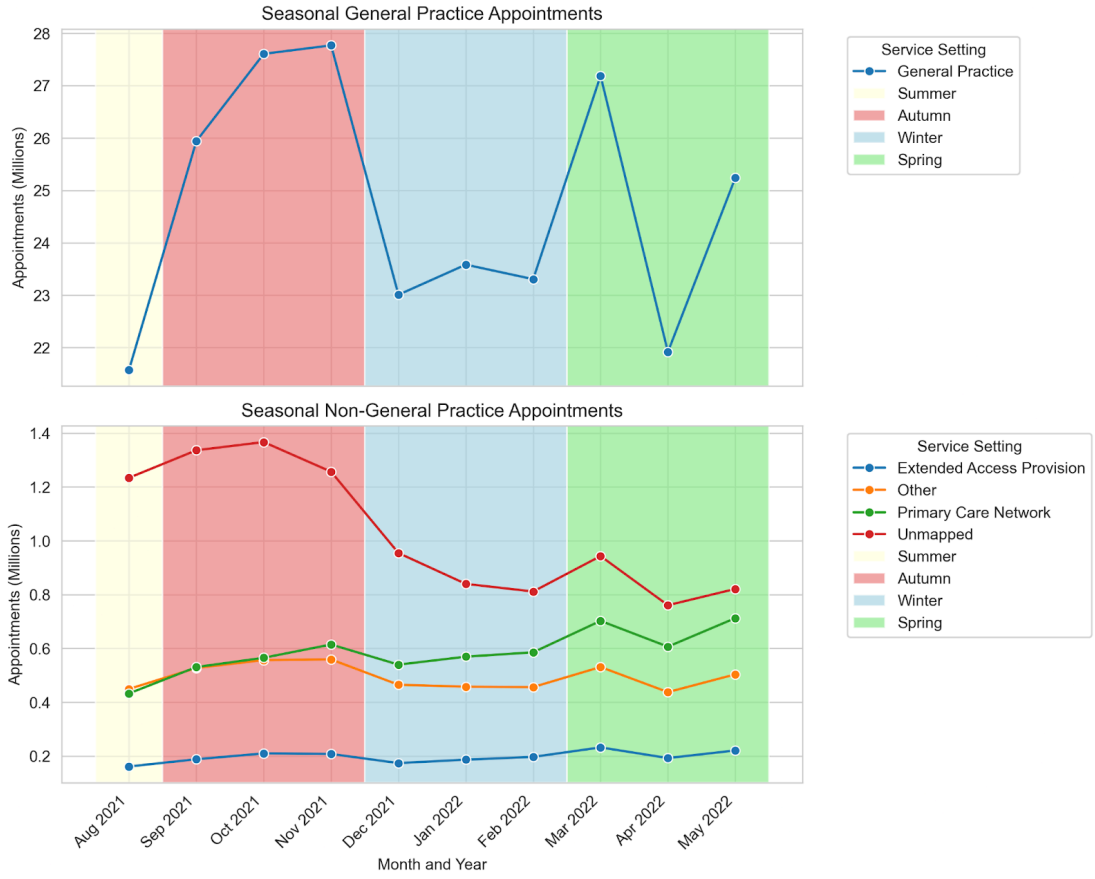
- Top 5 categories show similar trends: An August–October 2021 peak, followed by a March 2022 peak.
    - **General Consultation Routine**: Highest at ~10M appointments in October 2021, this could be due to a backlog of deferred conditions

- **Planned Clinical Procedures**: There was a sharp 50% rise from August to October 2021, which could be deferred minor procedures and non-COVID vaccinations (e.g, flu).



Seasonal Trends

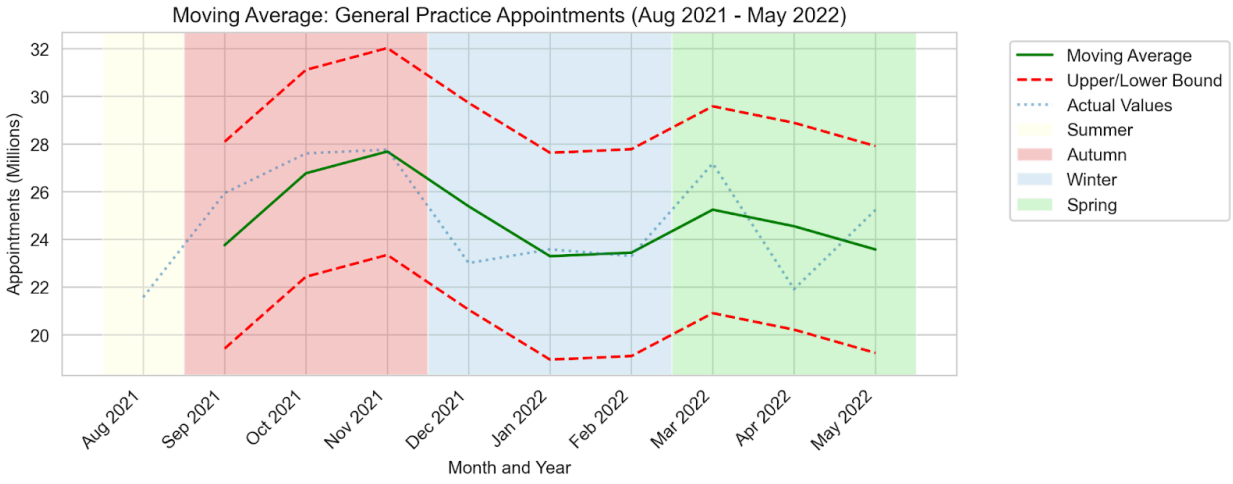
General Practice appointments started at ~22M in Summer 2021 (only August data is available), increasing through Autumn and Winter 2021, and declining by Spring



General Practice - Moving Average by Season

- The window used is 2 months, due to small dataset of 10m months, but plotted against the seasons for context
- Short term fluctuations are evened out, showing a downwards trend since November 2021

- Although actual values show sharper peaks than the trend, they still stay within the bounds, allowing for potential forecasting



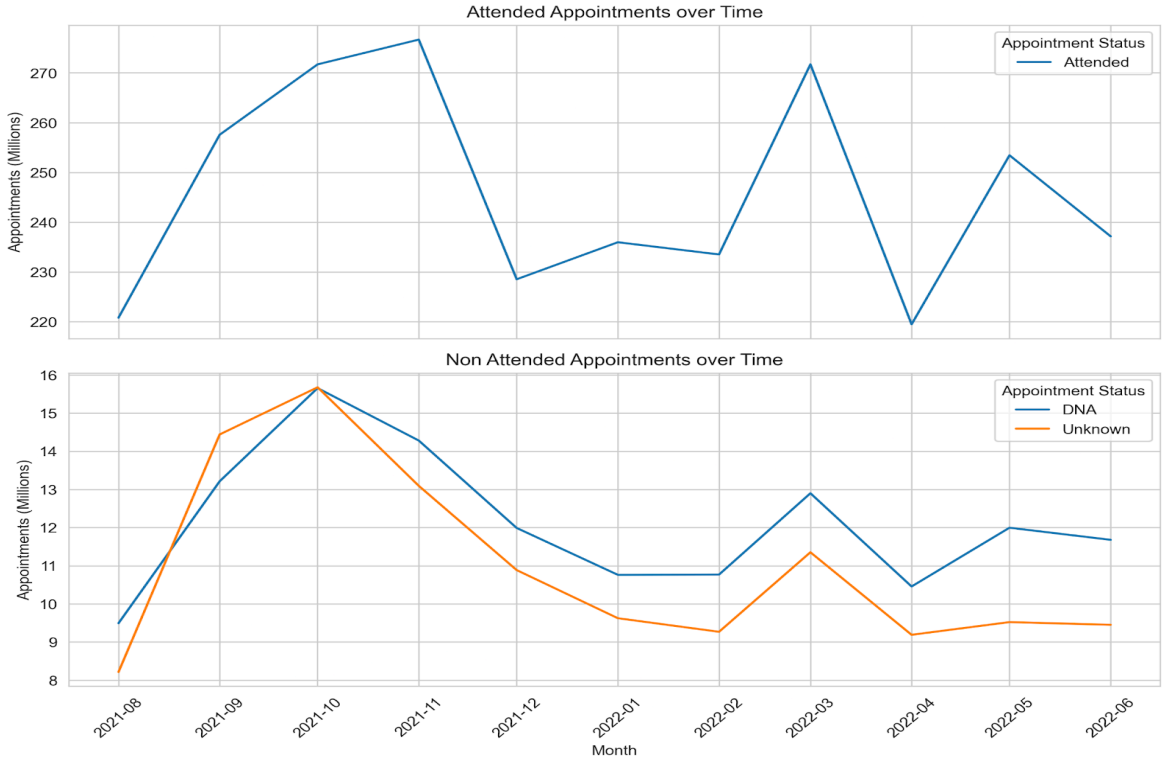
Did Not Attends(Missed Appointments)

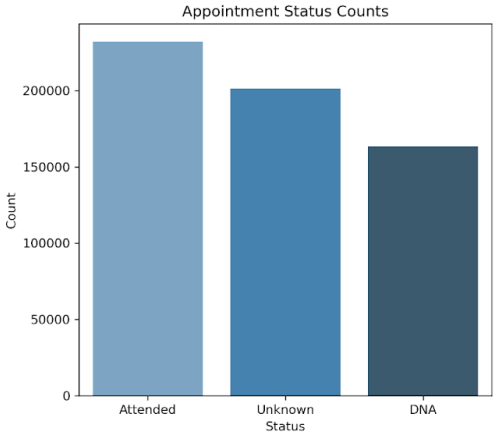
DNAs represent a large resource and time loss for the NHS. NHS England has estimated that there are 7.2 million missed general practice appointments annually, at a cost of £216 million

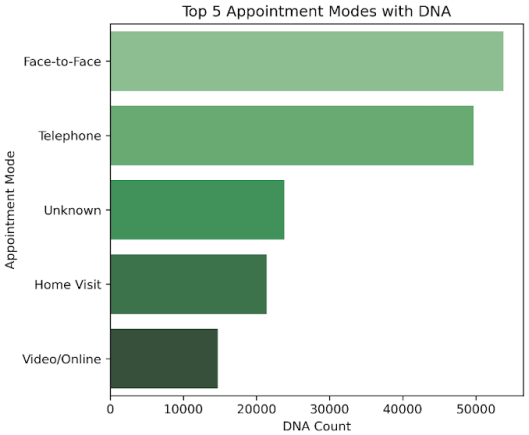
Looking at DNAs by different metrics, we see:

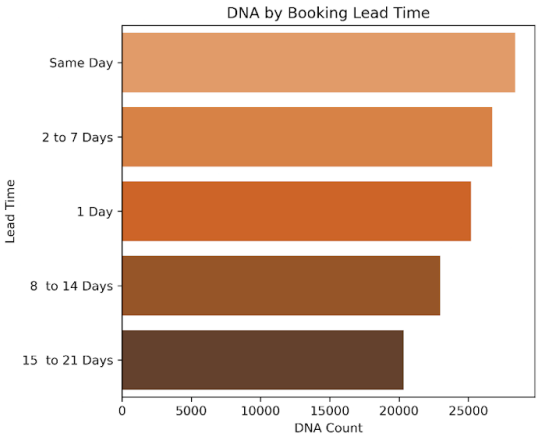
- There were around 150,000 DNAs
- Face to Face is the most commonly missed Appointment Mode, Video is the least
- Same Day bookings are most likely to be missed
- The Top 5 ICBs for DNAs are in Northern England
- There was a correlated increase in DNA's as overall attendance rose. DNAs have not yet returned to August 2021 levels

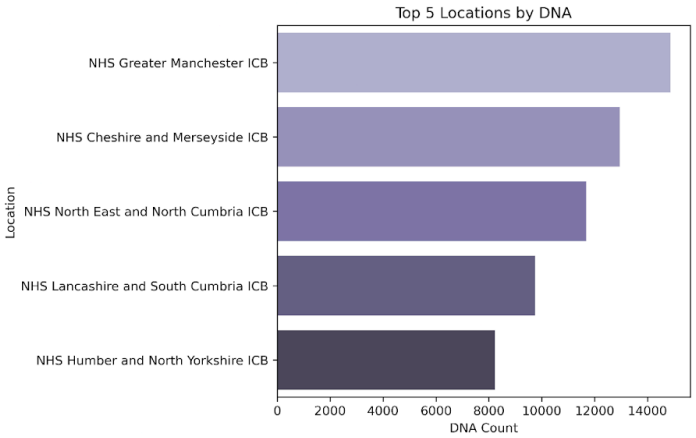
Did not Attends(DNAs):











Social Media Analysis

By analyzing tweets, we can provide insights through:

- Identifying trending topics using hashtag frequency to understand public perception
- Measuring engagement through retweets and likes to identify which messages are most relevant
- Improve recruitment by highlighting insights, such as interest in digital health services

Summary of Insights:

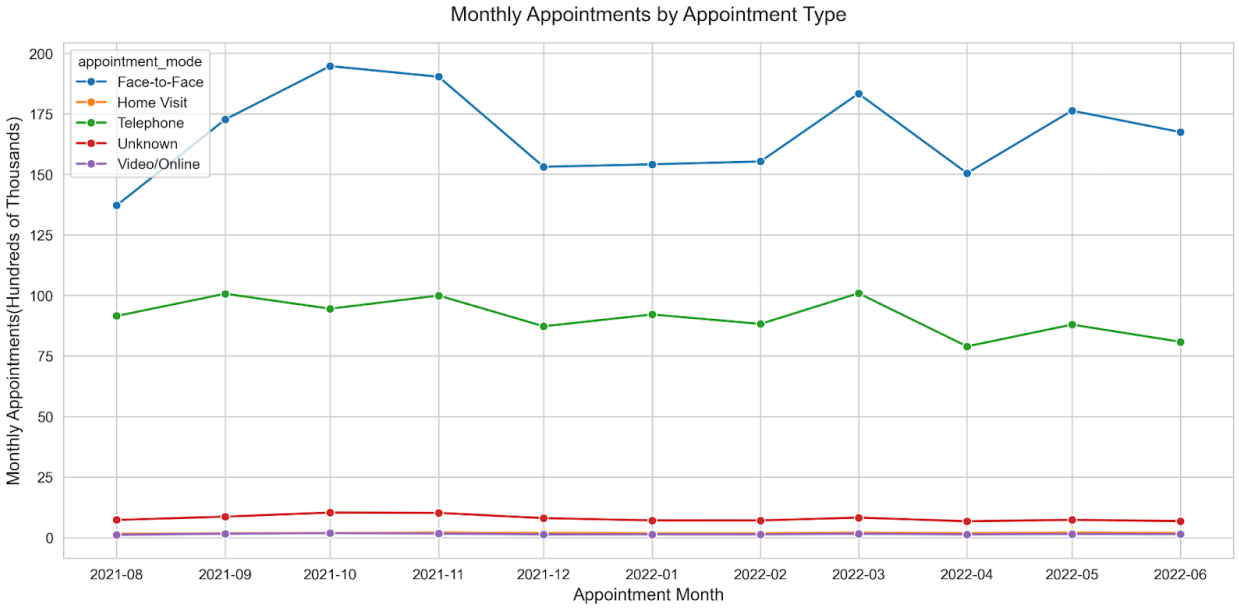
- Of 1,174 tweets analyzed (961 after removing duplicates), the median retweets (21.36) and likes (48.05) indicate moderate engagement. **Note:** Likes replaced favourites in 2015
- Top hashtags (#healthcare, #ai, #digitalhealth) suggest interest in both digital data availability and patient service through digital apps
- Retweeting involves more deliberate effort than liking, as it is seen as more an endorsement of a post, shown in a 2023 study

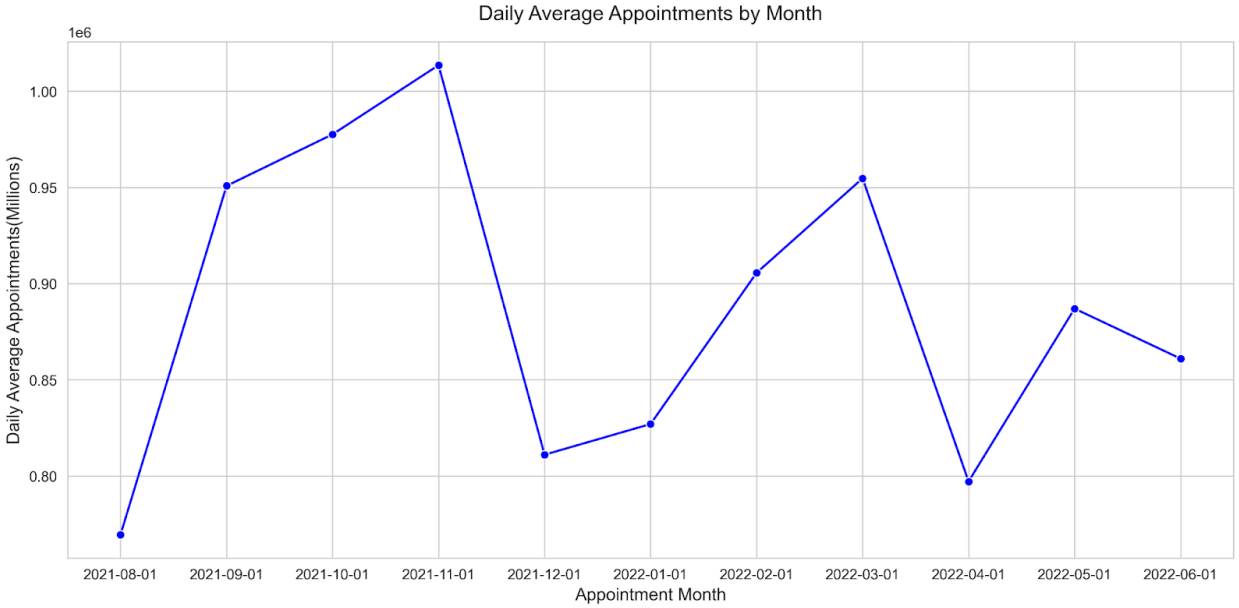
Based on popular hashtags and engagement, the NHS could leverage this by posting about NHS Digital Services , and engaging with users in the healthcare sector who would retweet and boost this message, helping to bring it into the public mind. Recruitment campaigns could also focus on the increasing use of digital services in healthcare

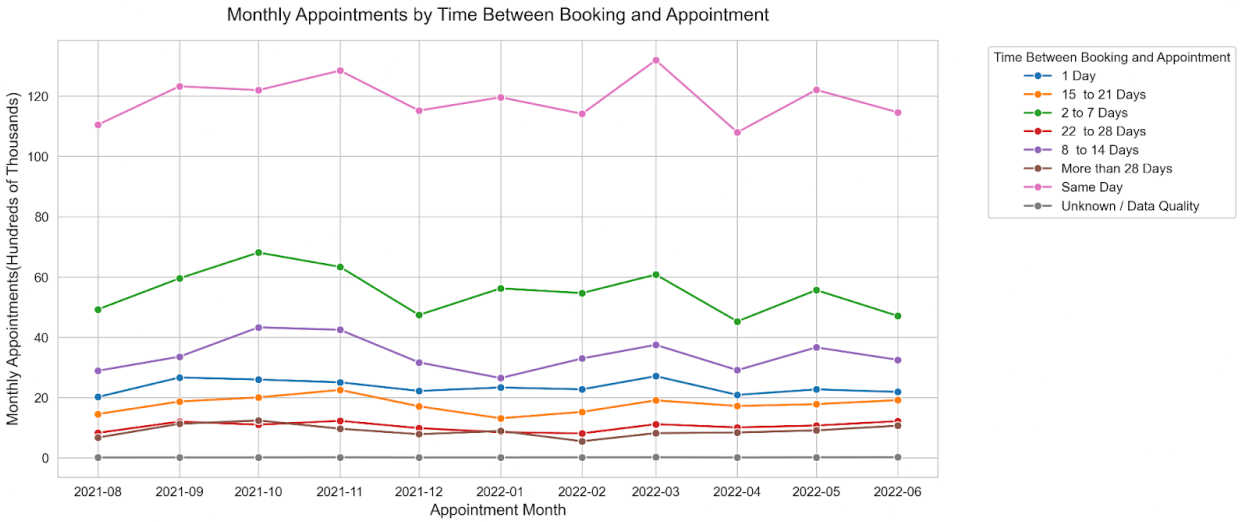
Capacity and Utilisation Analysis

Capacity was strained during peaks:

- Peak appointments: 33M per month (Oct/Nov 2021) or 1M per day
- High number of face to face appointments, which may be less efficient than other methods
- There was an increase in Same Day and '2-7 Day Before' bookings, Same Day Bookings are also likely to DNA
- Appointments dropped to 26M by June 2022 but remained above August 2021
- 1 million 10 minute appointments per day represents 166,667 staff hours. NHS data shows that of June 2024, there were 37,041 FTE GPs


Monthly Appointments by Appointment Type
appointment_mode
Face-to-Face
Home Visit
Telephone
Unknown
Video/Online
Monthly Appointments(Hundreds of Thousands)
200
175
150
125
100
75
50
25
0
2021-08
2021-09
2021-10
2021-11
2021-12
2022-01
2022-02
2022-03
2022-04
2022-05
2022-06
Appointment Month


Daily Average Appointments by Month
1e6
1.00
0.95
0.90
0.85
0.80
Daily Average Appointments(Millions)
2021-08-01
2021-09-01
2021-10-01
2021-11-01
2021-12-01
2022-01-01
2022-02-01
2022-03-01
2022-04-01
2022-05-01
2022-06-01
Appointment Month



Recommendations and findings:

Capacity, utilisation and Staffing

1. Utilisation is high but uneven, with GP appointments making up 75% of appointments. Signposting patients to other relevant services may help improve efficiency
2. Implement predictive forecasting to allocate resources for seasonal peaks
3. Increasing staffing in highest volume SICBLs would have the largest effect on handling peaks
4. The high number of face to face appointments indicate a reliance on physical locations, signposting could triage patients who have queries that could be handled through online appointment
5. Prioritize digital infrastructure, such as video appointments and online patient booking - a 10% increase in video appointments could free up 2M in person appointments during peak months

DNAs

1. Introduce or increase automated reminders, especially for same day appointments
2. Allocate extra resourcing to video appointments, which are less likely to be missed

Social Media:

- Filter out non UK data by including geolocation in the dataset
- Perform sentiment analysis to gauge people's perception of the NHS

Appendices

Appendix 1 Data Quality and Limitations:

There is no national standard for data, which may result in variations in quality and type.

- Known data issues include:
    - DNA (Did Not Attend) appointments not being correctly logged from June to November 2018 for practices using the TPP SystmOne system. This could affect DNA reports
    - Practices using the Cegedim GP system are unable to supply appointment mode data, leading to a higher instance of "Unknown" after 2019. This could affect Appointment Mode analysis
    - Healthcare Professional (HCP) type is incorrectly reported from October 2017, with only the "GP" type having consistent reporting.
    - Actual appointment duration data was introduced in April 2022, with partial data available from December 2021. This could affect duration and capacity analysis
    - Approximately 3–6% of appointments remain "Booked" with unknown attendance. This will have an effect on DNA analysis
    - Some appointment modes may be incorrectly logged, such as video conference appointments being recorded as face-to-face, or multiple video bookings logged as a single appointment. This will effect appointment mode analysis
    - National Category Data doesn't exist before August 2021, reducing ability to show seasonal trends
    - Data is largely from the Covid peak of November 2021-March 2022. This means the utilisation numbers will be highly atypical

- Time between Booking and Appointment can be influenced by other factors not present in the data (patient availability, urgency, repeat bookings)

Data Quality and cleaning:

No missing values found in the datasets.

`Appointment_date` and `appointment_month` are object types;these will need to be converted to datetime for better analysis

Preparing Data for Use

1. Check data for:
    - Missing values.
    - Outliers (e.g., appointments lasting <1 minute or >60 minutes).
    - Unmapped categories.
2. Consider grouping or excluding unreliable data, such as HCP types other than "GP."

Appendix 2 Data Definitions

Appointment Status: Includes Available, Booked, Attended, DNA, Cancelled, or Unknown.

Appointment Mode: Includes face-to-face, telephone, home visits, online/video.

Service Setting: General Practice, Primary Care Network, Extended Access Provision, Other, or Unmapped.

Context Type: Care-Related Encounter, Inconsistent Mapping, Unmapped.

Service Settings (5 types):

- Primary Care Network, General Practice, Extended Access Provision, Other, Unmapped

Context Types (3 types):

- Care Related Encounter, Unmapped, Inconsistent Mapping

National Categories (18 types):

- Patient Contact during Care Home Round, Planned Clinics, Home Visit, General Consultation (Acute/Routine), Structured Medication Review, Clinical Triage, Unmapped, others (e.g., Social Prescribing, Walk-in).

Appointment Statuses (3 types):

- Attended, DNA, Unknown

Appendix 3 Appointment analysis methodology

To determine the first and last dates in the datasets,

the min and max date was found for each:

'Actual_duration' dataset: Appointments were made between 2021-12-01 and 2022-06-30 - we know that actual duration data was first included in the April 2022 release, with data from December 2021

'Appointments_regional' dataset: Appointments were made between 2020-01-01 and 2022-06-30

'National_categories' dataset: Appointments were made between 2021-08-01 and 2022-06-30

This shows that the appointments in the data span from 2020-01-01 - 2022-06-30

To determine which service setting reported the most appointments in North West London from 1 January to 1 June 2022:

The national_categories dataset was filtered by the location and dates, as this dataset contains the service_setting value required. The showed that General Practice was the most booked appointment, at 72 bookings for this period. This can be further categorised by context type, which almost of of the GP appointments were categorised as a Care Related Encounter

To determine the month with the most appointments in the national_categories dataframe:
The month and year values were isolated from appointment_date, and the appointment_count

summed for each month. Once sorted, this showed that November 2021 had the most appointments, with just over 30.4 million, and with October 2021 at almost identical numbers.

This is supported by November 2021 being the peak of the Omicron Covid wave, and appointments to fall to 25m in December

To return the total number of appointments per month per data source:

Year and month were extracted from appointment_date were required, using dt.

New df was created, grouping count of appointments by year and month, this was the summed

The resulting data frame was sorted in Ascending date value

Appendix 4 Social Media Analysis Methodology

The data set ‘tweets.csv’ was first checked for duplicates. 91 duplicate entries were found in the ‘tweet_full_text column, these did not appear to be unique posts and were removed

Number of rows before removing duplicates: 1174
Number of rows after removing duplicates: 961

Count values was then used on both tweet_retweet_count and tweet_favorite_count, to determine the most common amount of retweets and favourites. The mean was then determined for both

Next, the tweets were ordered by amount of retweets, to determine which posts had the most engagement

They were then sorted by those that had been favourited, and retweeted. As both the tweet_favorited and tweet_retweeted columns contained only ‘False’ values, this is a possible data limitation were these values have not been correctly imported

A dataframe was created that only contained the tweet_full_text column, to prepare for the text splitting

This column was passed through a loop to extract #text from each row in tweet_full_text

As some tweets contained /n characters, some outputs contained extra text as well as the #text. The code snippet `r'#\w+` was added to treat / as literal characters and return only # text

The hashtags were then counted and put into a dataframe with the columns ‘Count’ and ‘Hashtag’

This was split into two dataframes to represent hashtags with count 10-20, and count >20. There were too many entries to comfortably fit on one bar graph while still showing all values with >10 count, so two plots were used