Background/context of the business scenario:

The NHS experiences considerable expenses that could have been prevented if patients did not miss appointments with their general practitioners (GPs). Considering the NHS's crucial role in providing affordable healthcare services to society, it is vital to examine the reasons for missed appointments. Due to the scarcity of general practitioners and staff, every available appointment should be utilised to guarantee that every patient receives treatment and consultation, regardless of their social and financial status.

According to Professor Philip Banfield, chair of the British Medical Association (BMA), the NHS should investigate the underlying cause of missed appointments through a data-driven approach before resorting to penalties. To achieve this, the NHS has posed two main questions: has there been sufficient staff and capacity within the networks, and what was the actual utilisation of resources?

Using the 5 Whys framework, we can initially break down the potential causes for missed appointments. One possible reason is that patients may have last-minute time conflicts and are not incentivized to attend their pre-scheduled appointments. Another factor could be the NHS appointment scheduling system, which may make it difficult for patients to attend their appointments. This could be due to patients' unfamiliarity with the system and a lack of awareness of the consequences of missed appointments. The scheduling system itself may lack usability features, leading to misaligned appointment scheduling. Ultimately, there is a lack of foundational data transparency, and additional data points and improved completeness around unidentified variables are necessary to gain further insights into the root cause. The NHS has focused primarily on providing healthcare consultations to the population, and data analytics has not been a primary focus.

Analytical approach: Describe the approach taken to import, clean, and analyse data in Python:

To perform the analysis, we will need to use the following libraries: pandas, seaborn, matplotlib.pyplot, and numpy. Each of these libraries serves a distinct purpose: pandas is used for creating and querying dataframes, seaborn and matplotlib.pyplot are used for visualising insights, and numpy is used for algorithmic functions.

Once we have imported the necessary libraries, we can proceed by loading the provided data from the NHS using pandas functions. The raw data, in either a CSV or Excel file format, will be stored in a variable as an object/dataframe that represents a dictionary containing series or, in simpler terms, data sorted as rows and columns.

To verify that the raw data has been loaded successfully, we can use the ".head()" function to print the first few rows of the dataframe and review the preview for any errors or formatting issues. Additionally, functions such as ".shape", ".dtypes", ".isna().sum()", ".describe()", ".info()", and ".count()" can assist in evaluating the count of rows and columns, data types, the number of null values, and descriptive statistics. These functions can be used to evaluate the preloaded data for consistency and any errors or null values.

To address the NHS questions, we utilised various Python libraries and functions:

To determine the number of unique locations, service settings, context types, national categories, and appointment statuses in the datasets, we used the ".nunique()" function on the relevant series within the dataframe. This revealed that there are 106 unique locations, 5 service settings, 3 context types, 18 national categories, and 3 appointment statuses. Additionally, we used the ".value_counts()" function to identify the locations with the highest records in each file.

To identify the date range of the provided datasets and which service settings reported the most appointments for a specific period, we examined each file individually. The "actual_duration" file covers the period from 01-12-21 to 30-06-22, the "appointments_regional" file covers the period from 02-08-21 to 30-06-2022, and the "national_categories" file covers the period from 01-12-21 to 30-06-2022. We used the "groupby()" function with arithmetic functions to derive the sum of appointments and the "print()" function for data frames or the "min()/max()" functions to identify the date ranges. Based on the national data, General Practice received the most appointments overall at 270,811,691, followed by Unmapped at 11,080,810, Primary Care Network at 6,557,386, Other at 5,420,076, and Extended Access Provision at 2,176,807.

Finally, to determine the number of appointments and records per month, we again utilised the "groupby()" function on the data frame. November 2021 had the highest recorded number of appointments at 30,405,070, while March 2022 had the highest count of records at 82,822.

Visualisation and insights:

By utilising groupby functions and manipulating the dataframes, we were able to filter the data to extract specific insights. To maintain consistency, we standardised the figure size and set the background colour to white. To visualise these insights, we employed seaborn plots to address some of the additional questions posed by NHS:

What monthly and seasonal trends are evident, based on the number of appointments for service settings, context types, and national categories?

By utilising groupby functions and dataframe manipulations, we were able to extract specific insights and visualise them using standardised figure size and white background through seaborn plots. On a national level, we grouped appointments by service setting and found that General Practice had the most appointments, averaging 25,000K per month across the period. The other service settings had below 2,500K appointments per month on average. For GP appointments, there was a downward trend in appointments towards the end of the period, as shown in Figure 1.

Looking at national context types by appointment, as seen in Figure 2, we found that Care Related Encounters had a similar demand to GP appointments. Figure 3 showed that General Consultation Routine, General Consultation Acute, Clinical Triage, Planned Clinical

Procedure, and Planned Clinics were the categories with the most appointments on average across the period.

Figures 4-7 displayed the seasonal trends for the service settings, indicating the highest demand during autumn (October 2021) for general practice appointments. Demand averaged above 1 million per day for general practice appointments during Monday to Saturday, dropping significantly on Fridays to below 40K on average.

Figure 1: National Service Setting by Appointments

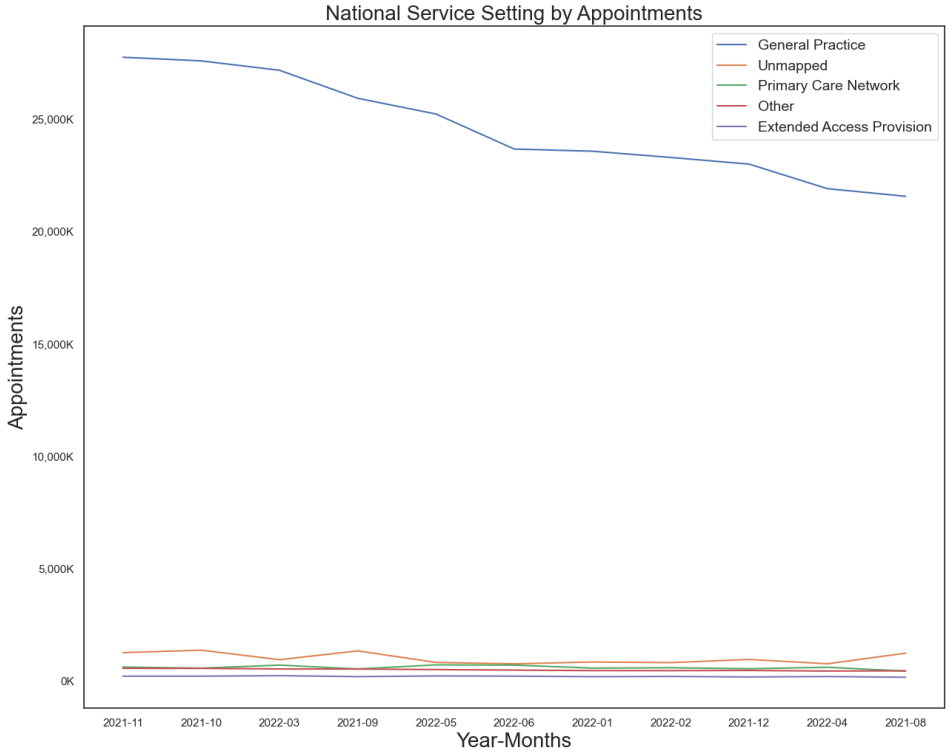


Figure 2: National Context Types by Appointments

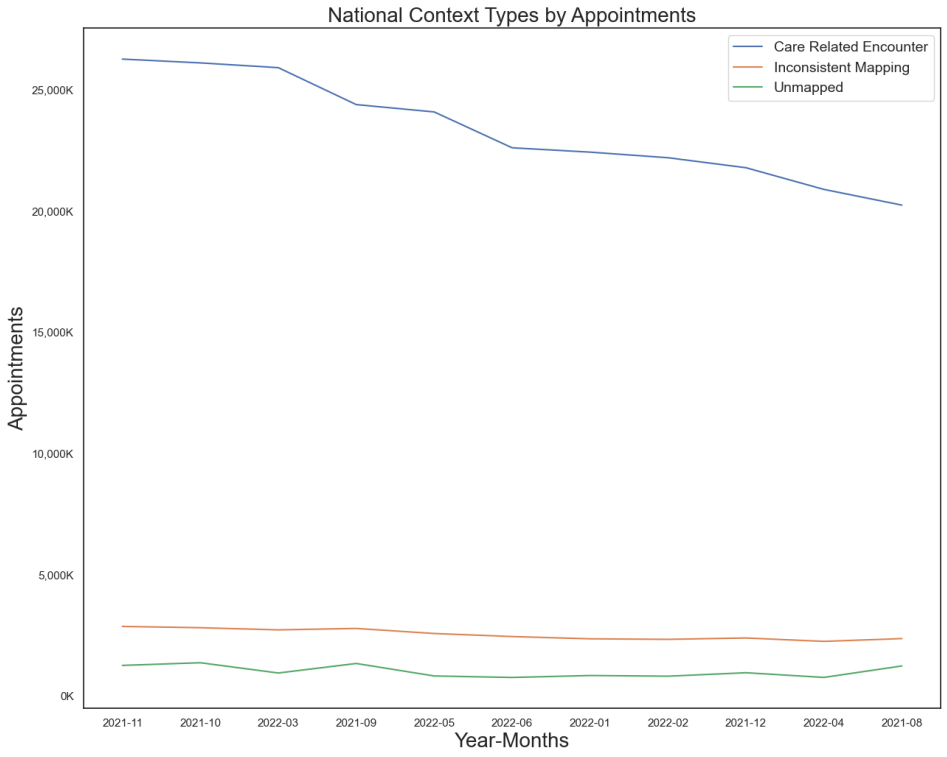


Figure 3: National Categories by Appointments

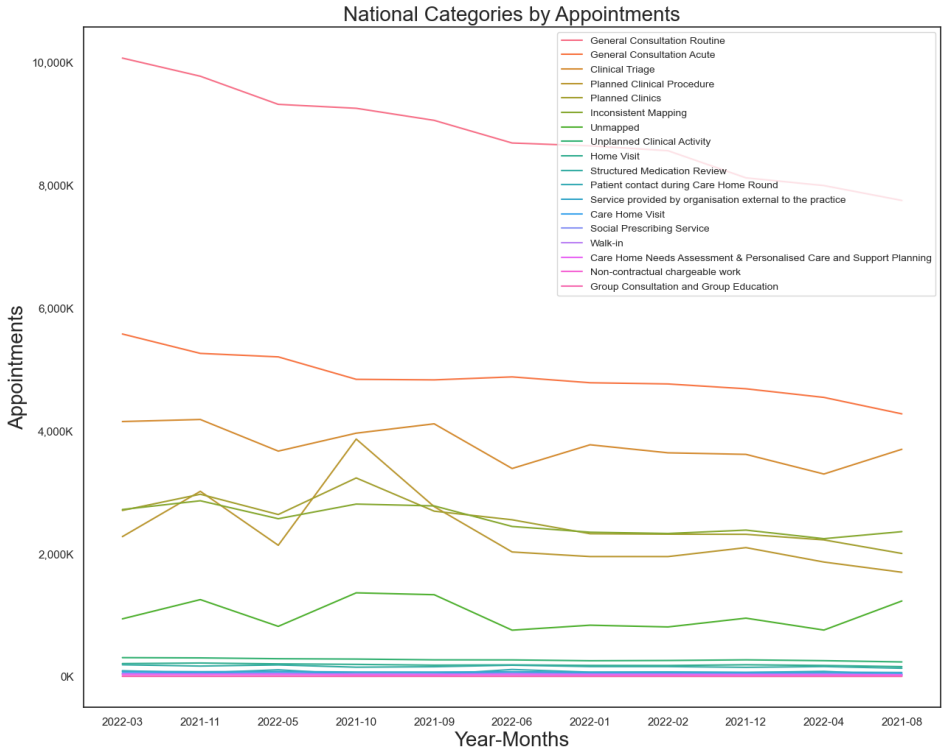


Figure 4: August 2021 National Summer Service Setting by Appointments

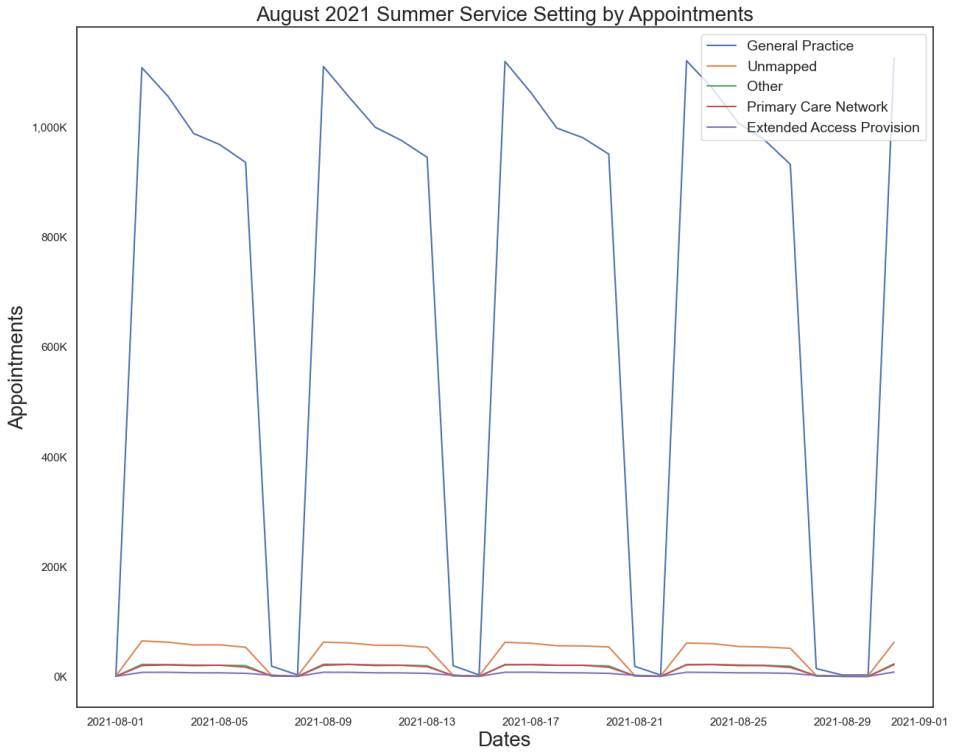


Figure 5: October 2021 Autumn Service Setting by Appointments

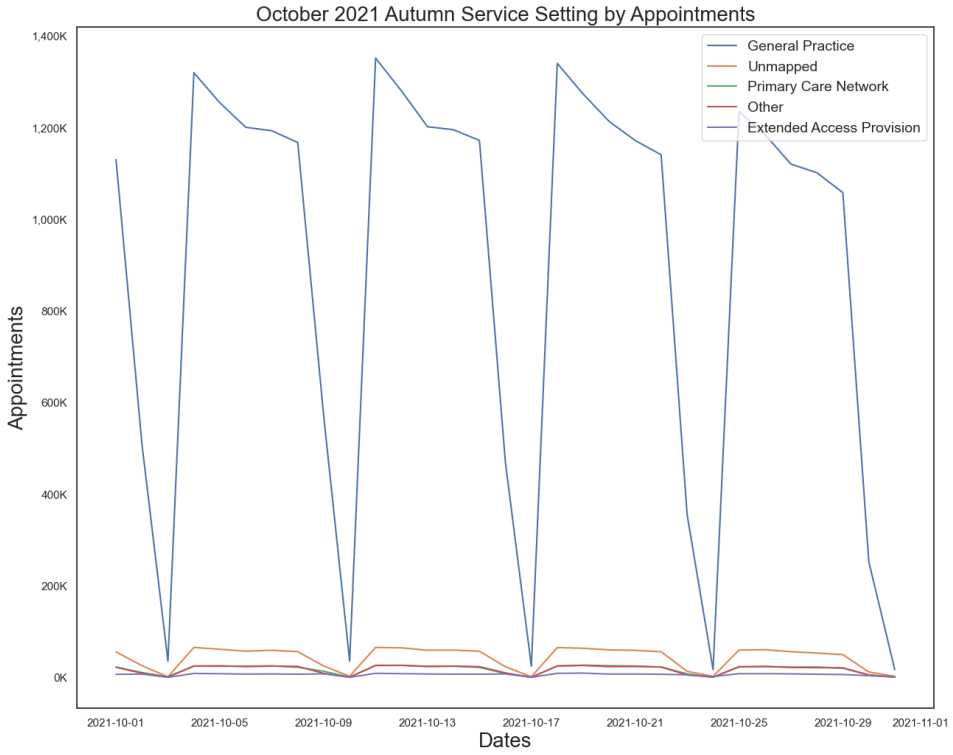


Figure 6: January 2022 Winter Service Setting by Appointments

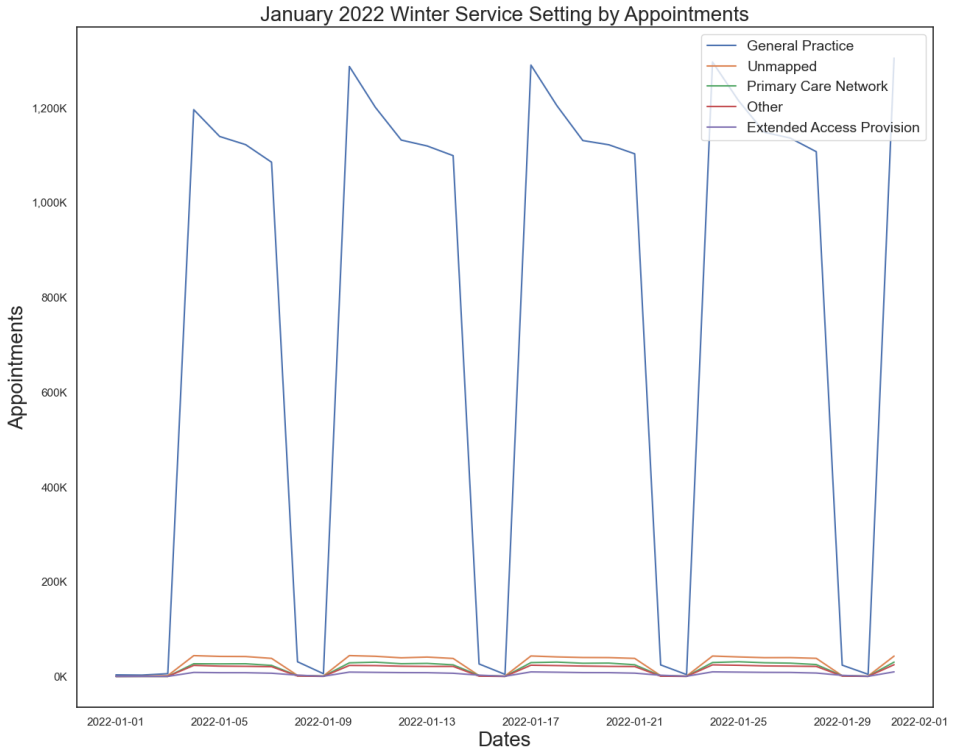
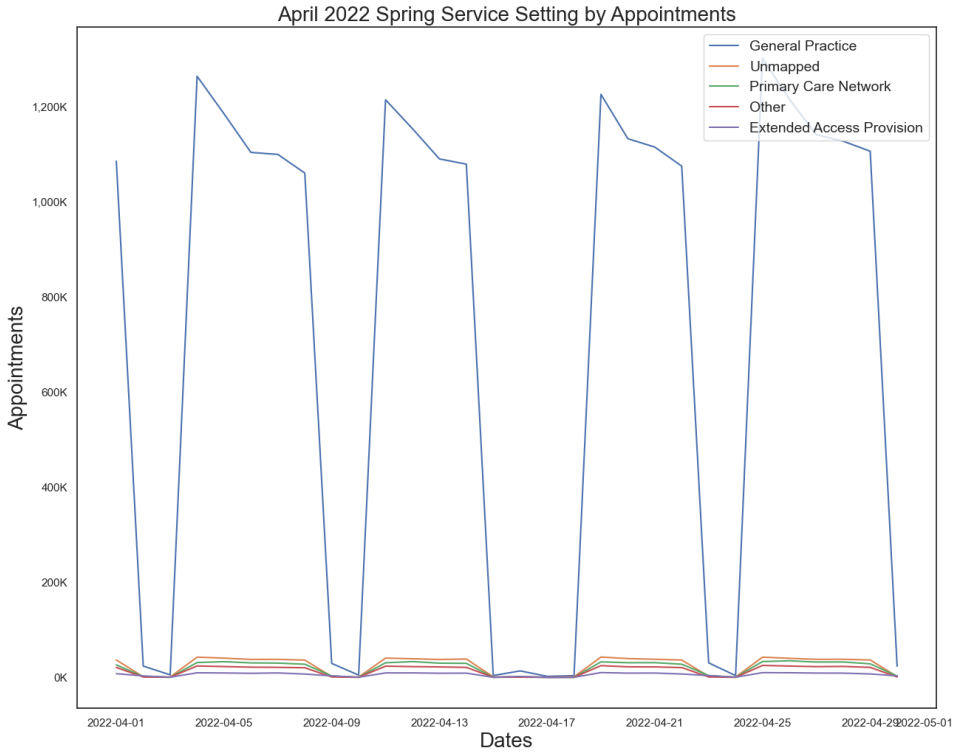


Figure 7: April 2022 Spring Service Setting by Appointments



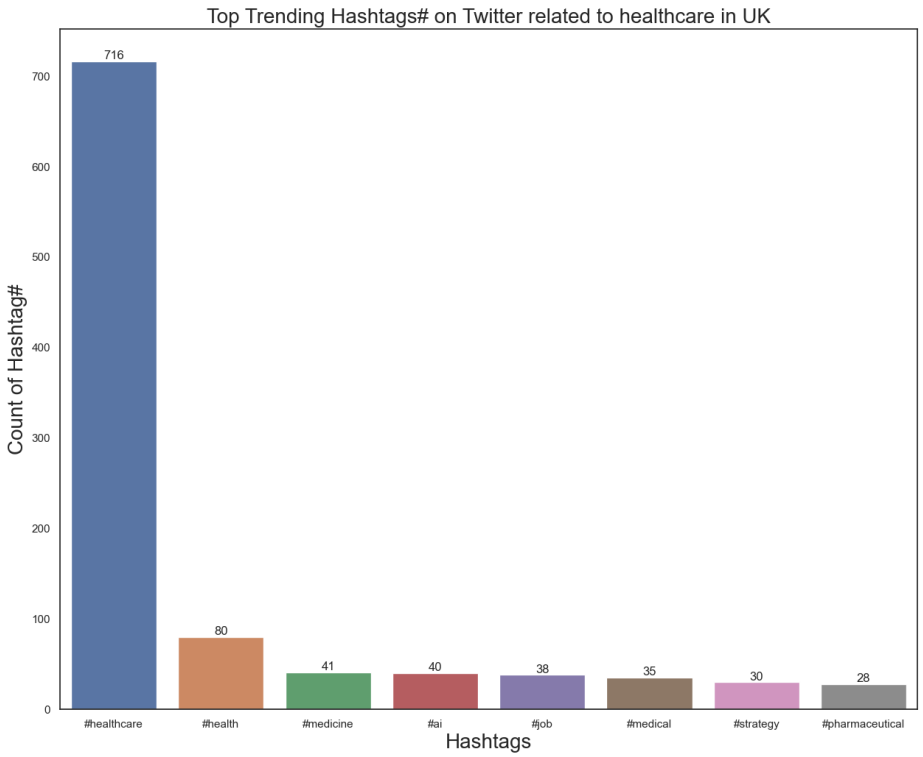
What are the top trending hashtags (#) on Twitter related to healthcare in the UK?

Businesses can effectively use social media to communicate with customers and promote their message. It is an excellent platform to increase visibility and announce business updates or product launches. Twitter, in particular, allows the use of hashtags, which are searchable keywords or phrases that group multiple tweets together. Hashtags help to extend the reach of content beyond one's own followers.

To leverage this feature, we have identified the most popular hashtags related to healthcare in the UK. We extracted hashtags from the Twitter dataframe by looping through the selected series and adding any hashtags to an empty list.

Based on our analysis, we can conclude that #healthcare and #NHS are the best hashtags to promote relevant content for the NHS on Twitter, which can help the organization connect with patients and other interested parties.

Figure 8: Top Trending Hashtags# on Twitter related to healthcare in UK



Conclusion:

To summarise our analysis and provide actionable recommendations for the future, we need to reiterate the problem statements:

The NHS incurs significant costs when patients miss GP appointments. We observed a national decrease in GP appointments over time and a significant reduction on a regional level since January 2021 (Figure 1). This is reflected in a 20% reduction in appointment utilisation from its peak to August 2021 (Figure 10). GP and practice staff appointments were in high demand (Figure 11), while online and home visits were less so. Same-day appointments were more in demand and likely to be attended (Figure 14). Unmapped service settings and context types have high utilisation per month, indicating a need for further investigation and transparency (Figure 15-16).

While the attendance rate is high, there are still around 500 daily unattended or unknown appointments. Appointment attendance has decreased over time, without an increase in the volume of unattended or unknown appointments, which may reflect seasonal demand during the winter and summer seasons.

Patients prefer face-to-face or telephone appointments scheduled closer to the consultation time (Figure 13). This presents an opportunity to develop online and video consultations for minor consultations, reducing strains on NHS staffing while serving patients' needs for last-minute consultations.

Our recommendation for the NHS is to improve its data analytics to reduce unmapped or unknown variables and improve the patient journey. Developing digital capabilities can lead to significant cost savings by better utilising staffing and resources for minor treatments.

We suggest the use of the #healthcare hashtag on Twitter to communicate with patients and increase awareness of the NHS's services, thereby reducing unattended appointments.

For future investigations, we recommend building out data analytics capabilities and segmenting NHS locations to serve patients based on their needs with the appropriate consultation service, whether digital, face-to-face, or via telephone. Evaluating the reduction in appointments during Sundays and additional staffing can increase attendance rates for patients with work obligations from Monday to Friday. This could also benefit staffing by freeing up time and resources during other days and diversifying appointments.

Figure 9: Regional Appointments since 2021-08

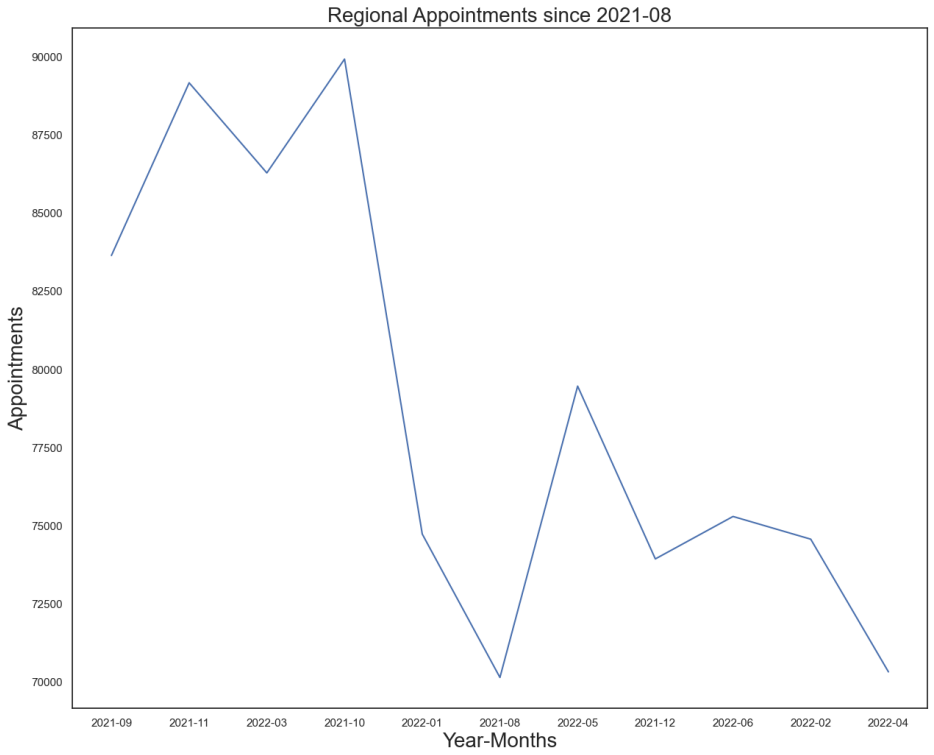


Figure 10: Regional Utilisation of Appointments since 2021-08

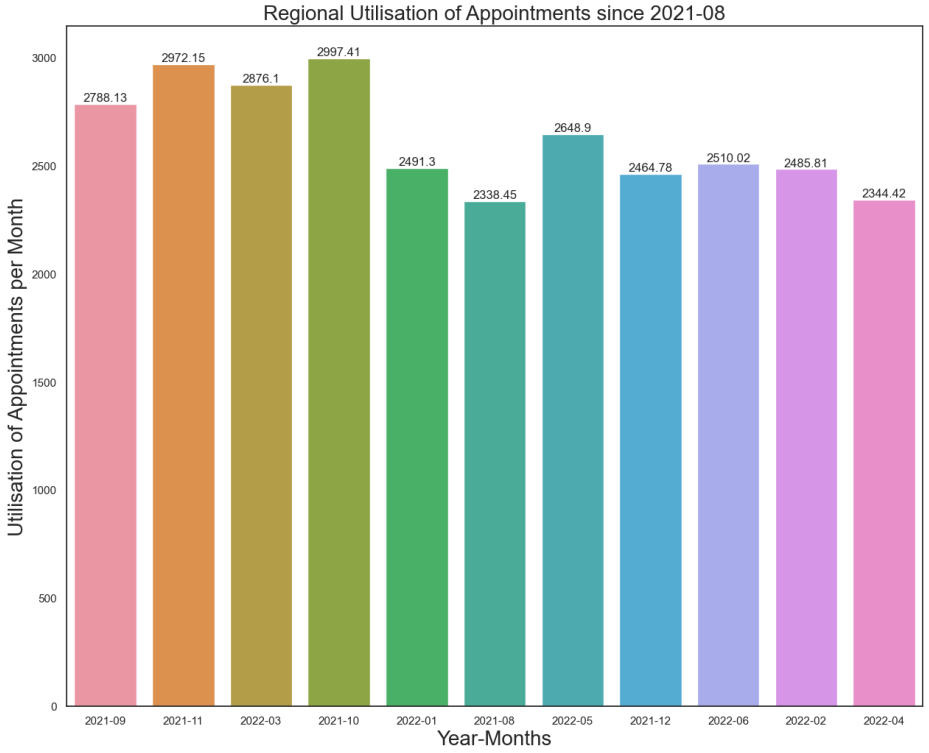


Figure 11: Regional Utilisation of Appointments by Service Setting since 2021-08

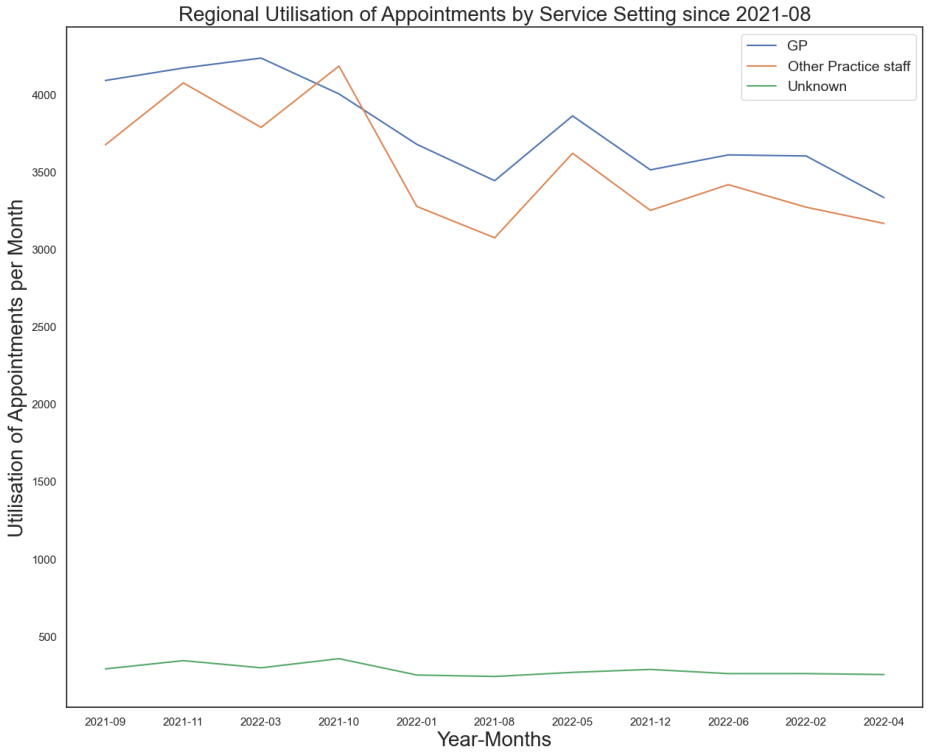


Figure 12: Regional Utilisation of Appointments by Appointment Status since 2021-08

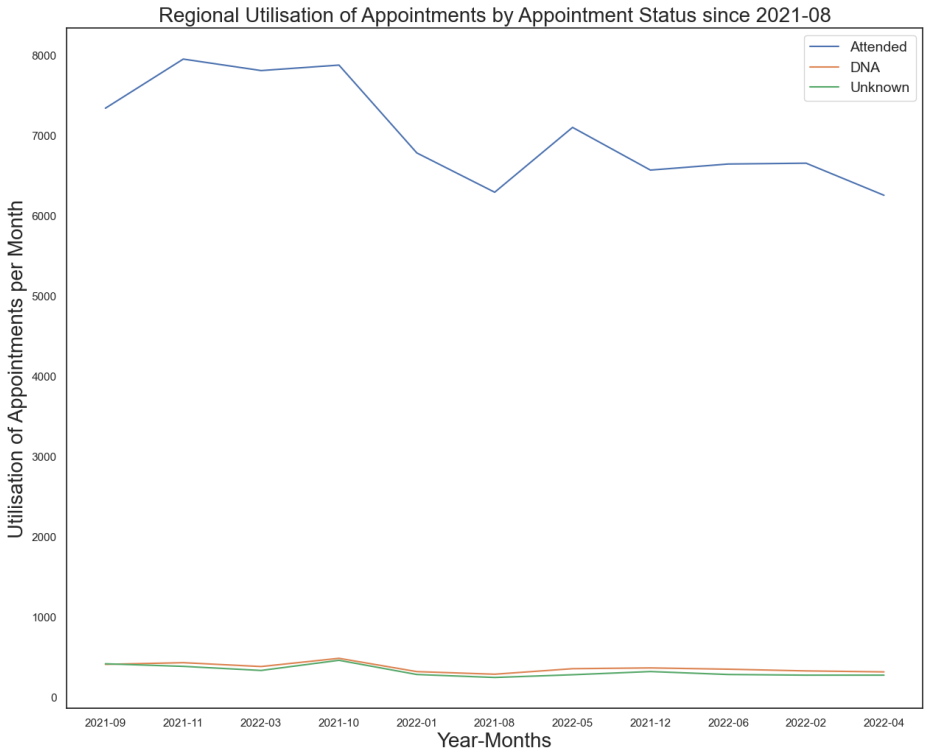


Figure 13: Regional Utilisation of Appointments by Appointment Mode since 2021-08

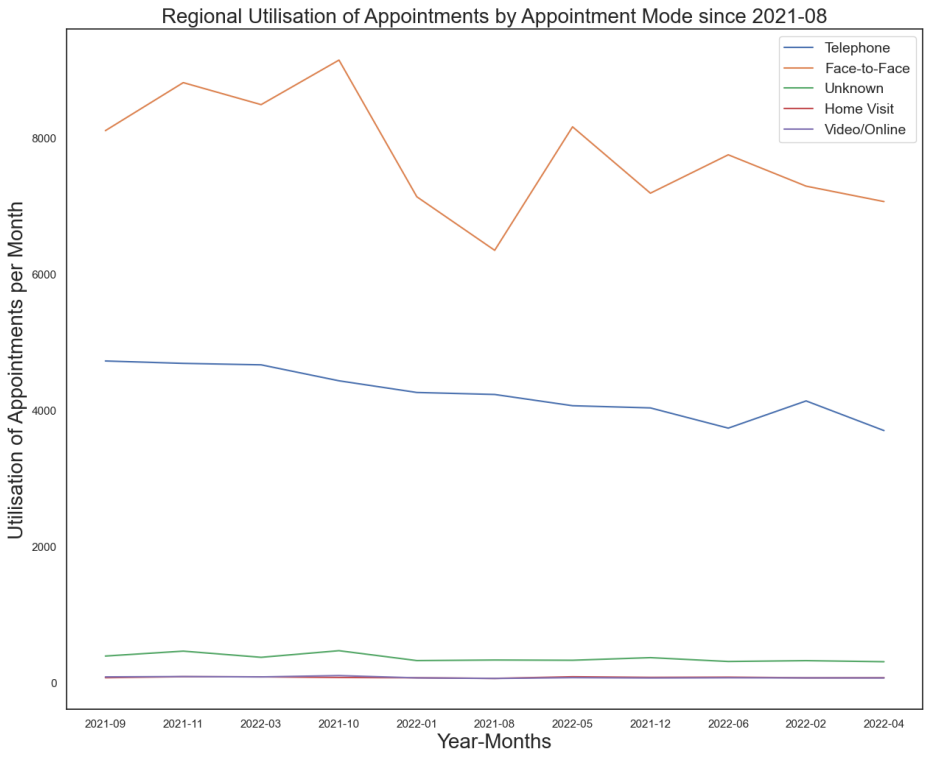


Figure 14: Regional Utilisation of Appointments by Time Between Appointments since 2021-08

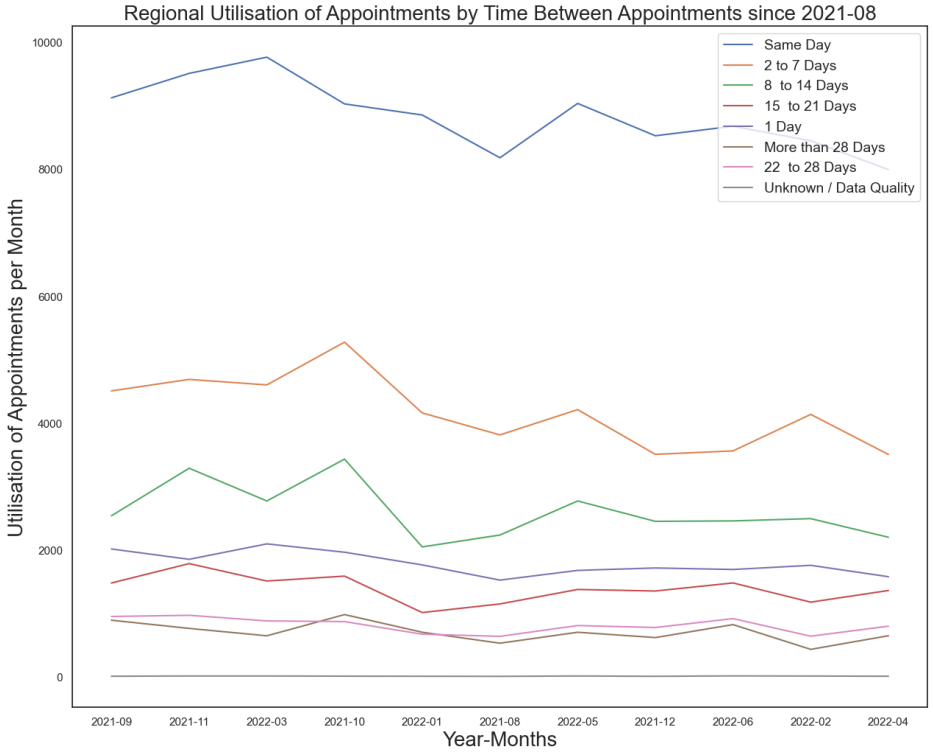


Figure 15: National Utilisation of Appointments by Service Setting

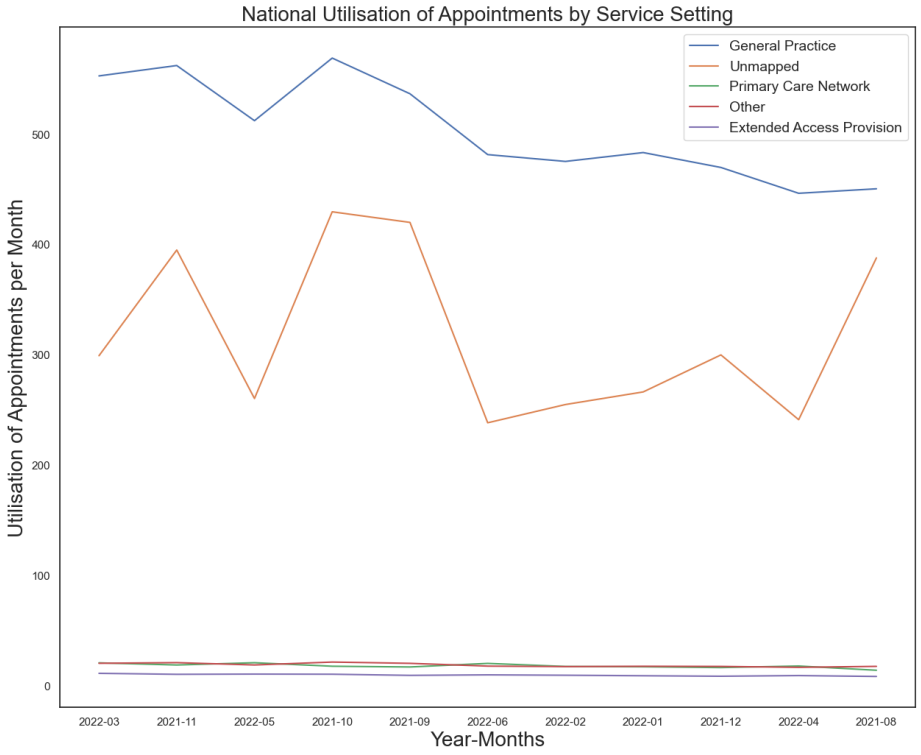


Figure 16: National Utilisation of Appointments by Context Type

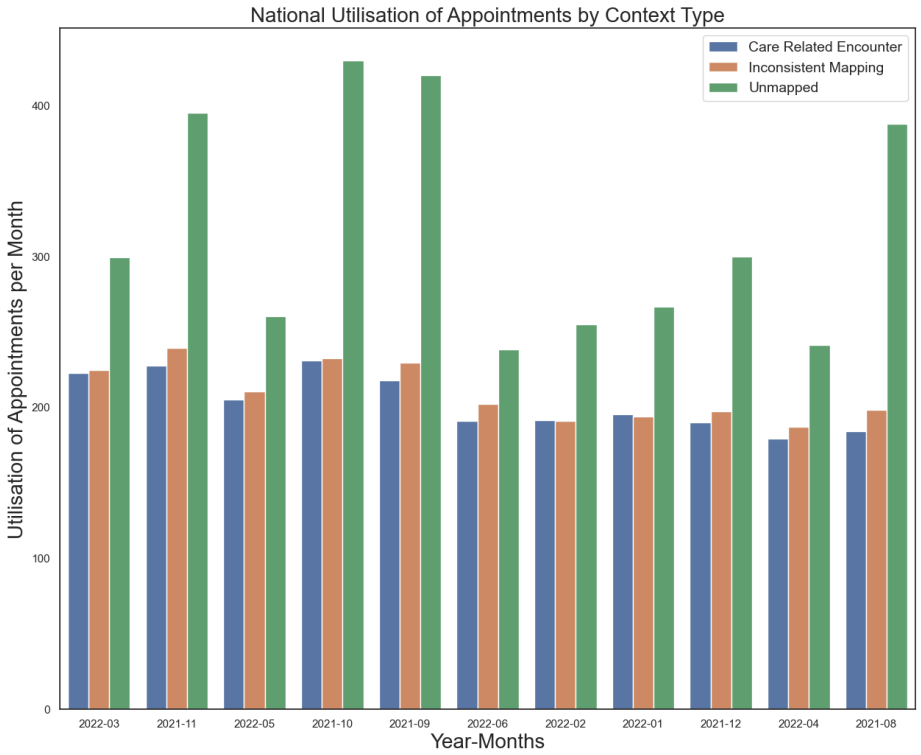


Figure 17: National Utilisation of Appointments by Category Type