Diagnostic Analysis using Python Report

The NHS retained a team of analysts on a mandate to explore patients missing appointments and the influence NHS efficiency and resiliency had on this. Specifically, to understand whether there was adequate staffing within their network and what the actual utilisation of resources was. The purpose of this report is twofold: primarily to review the technical process and methodology of this study, and secondly as diagnostic exploration of the data, which provides insight on trends. Recommendations are detailed in the stakeholder presentation accompanying this report.

Initial descriptive analysis was conducted to better understand the core datasets provided by the NHS, which were as follows.

Dataset Name (from file)	Related Acronym	File Contents
National Categories	NC	Location, context type, service setting, appointment date & month.
Appointments Regional	AR	Appointment date, mode, HCP, time between booking and appointment.
Actual Duration	AD	ICB Location Codes, Appointment Date, Duration of Appointment.
Tweets	Tweets	Trending Tweets.

Analysis was structured this way initially to develop a high-level understanding of the data. There are 106 locations within the datasets, with Northwest London being the largest. All of the top 5 regions are in the South of England, by contrast, the Northwest of England contain the lowest number of records, with the smallest being Greater Manchester. Additional analysis determined the values for dates, records, and appointments within each data frame, detailed below.

Data Frame	Earliest Appointment	Latest Appointment	Count of Appointments	Count of Records
NC	August 2021	June 2022	296,046,770	817,394
AR	January 2020	June 2022	742,804,525	596,821
AD	December 2021	June 2022	167,980,692	137,793

To standardise results, analysis that involved multiple data frames was conducted during August 2021 to June 2022.

The datasets also contained numerous categories and variables. To aid readers in processing analysis, the following table serves as an illustration of the variables, highlighting their function, the maximum number of categories within each variable, and the most popular of each.

Variable Name	Number of Categories	Most Popular
Service Setting	5	General Practice
Context Type	3	Face-to-face
National Categories	18	General Consultation
Appointment Status	3	Attended

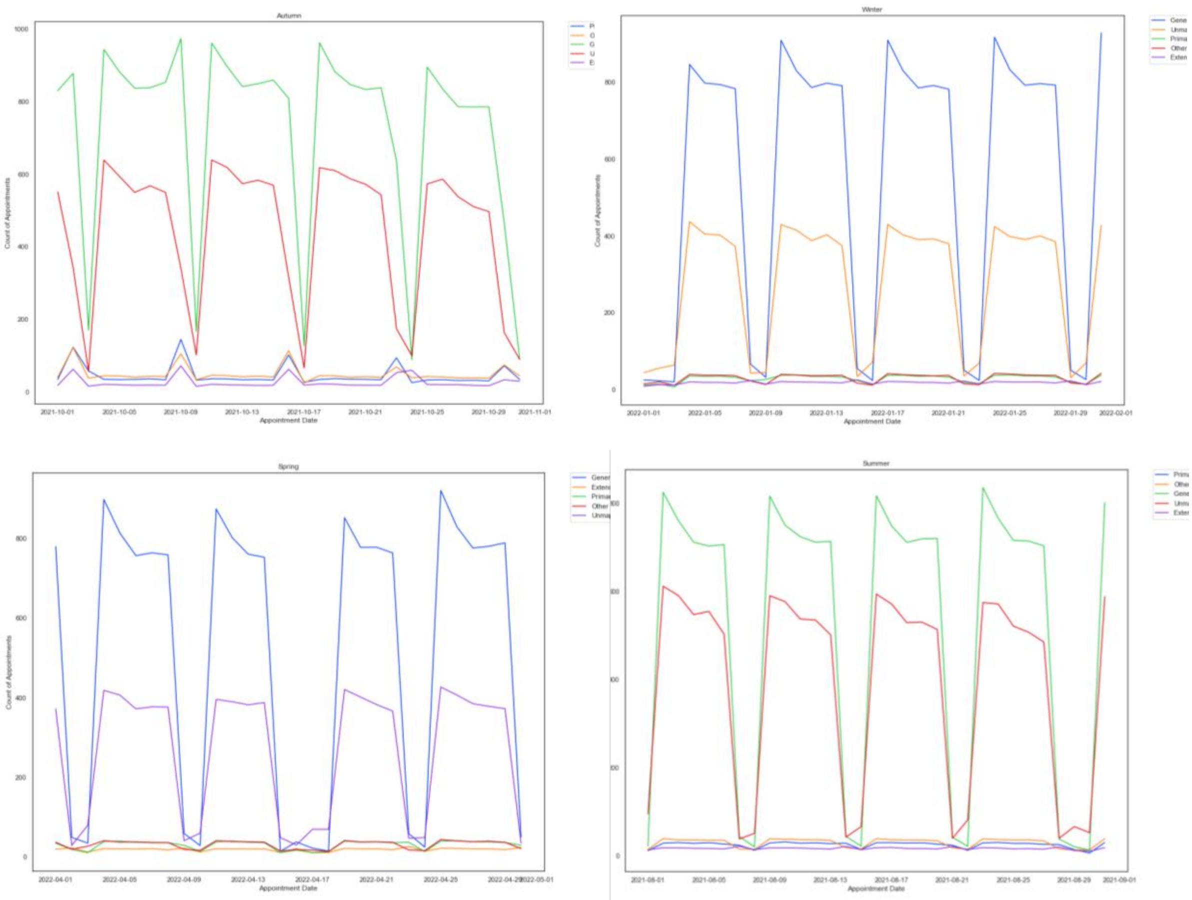
HCP Type	3	General Practitioner

To avoid errors caused by overwritten data frames, the original CSV file was reloaded at the beginning of each analytical process. The data was cleaned in an iterative, ad hoc capacity, creating new data frames, variables, and columns as necessary. Initial observations on discrepancies within the data frames were factored into subsequent analysis, as indicated in the Jupyter markdown.

The descriptive analysis was used to devise a framework for secondary analysis, for which the NHS' concerns were divided into supporting questions. Firstly, capacity utilisation was explored to determine the extent to which different services were in demand. To understand how these trends evolve over time monthly and seasonal activity was measured against the aforementioned variables. Thereafter, analysis was conducted to determine if there are trends across staffing levels, whether HCP practitioners differ over time, and changes in whether or not visits are attended.

Firstly, as the number of appointments differs by month, to better understand trends over time appointments were assessed for 1 month 'seasonal' intervals. Line plots were generated by aggregating the count of appointments, grouping by service settings, and plotted using MPL and SNS functions, filtered by each season.

Figures 1 - 4 Seasonal Activity

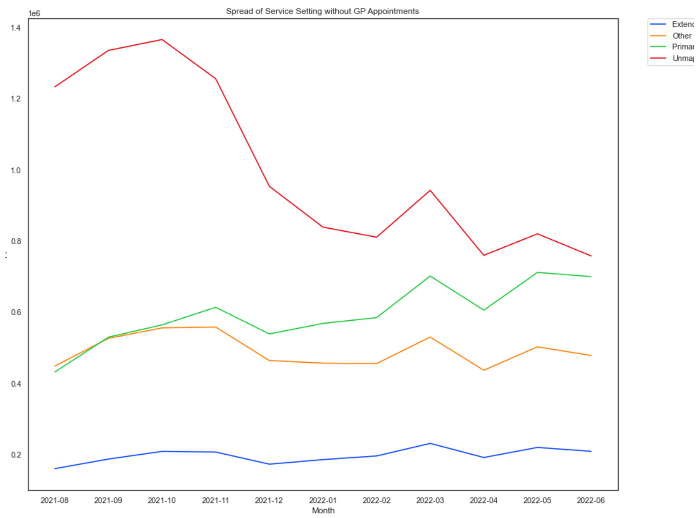


From top left to bottom right: Autumn, Winter, Spring, Summer.

Figures 1 - 4 represent daily fluctuations in the demand for appointments across service settings. The slight dip in appointments at the start of each week suggests more demand for appointments between Monday to Wednesday. Autumn was the busiest season, with the highest demand for all services.

With time showing a significant influence on the subscription of services, the next step was to analyse the shift in appointments per month for service settings (Figures 5 & 6 and Appendix 4), context types (Appendix 2), and national categories. To do so, three new data frames from the National Categories CSV were generated, each aggregating the monthly number of appointments and grouping by the respective variables, then performing additional analysis, as necessary.

Figures 5 & 6: Count of Appointments per Service Setting (with and without GP)



In Service Settings, General Practice was the most subscribed across all months by a significant margin, showing spikes in the number of appointments in keeping with the overall monthly trend highlighted in Appendix 1. Figure 6 was generated by removing General Practice appointments, which allows for a closer examination of the other Service Settings. The most noticeable trend is the significant decrease in the number of ‘unmapped’ values, which is negatively correlated with the increase in primary care network appointments.

The number of categories contained in the National Categories column made analysis challenging, so to decipher insights at the macro level the categories were remapped using a dictionary function, which reduced 18 categories to 8 (Appendix 3). These results were then replotted (see Figure 7), with the most subscribed category being Consultation. Structured Medication Review, Home Visits, and Care Home appointments all rose significantly during the 2021 - 2022 period.

To better understand whether staffing levels are appropriate the data is next broken into the type of appointment based on the healthcare practitioner (Figure 8). There is an increased demand for GP appointments in August until November, a dip between December and April, rising again in May onward. This coincides with a trend in overall appointment counts.

Figure 7: National Categories

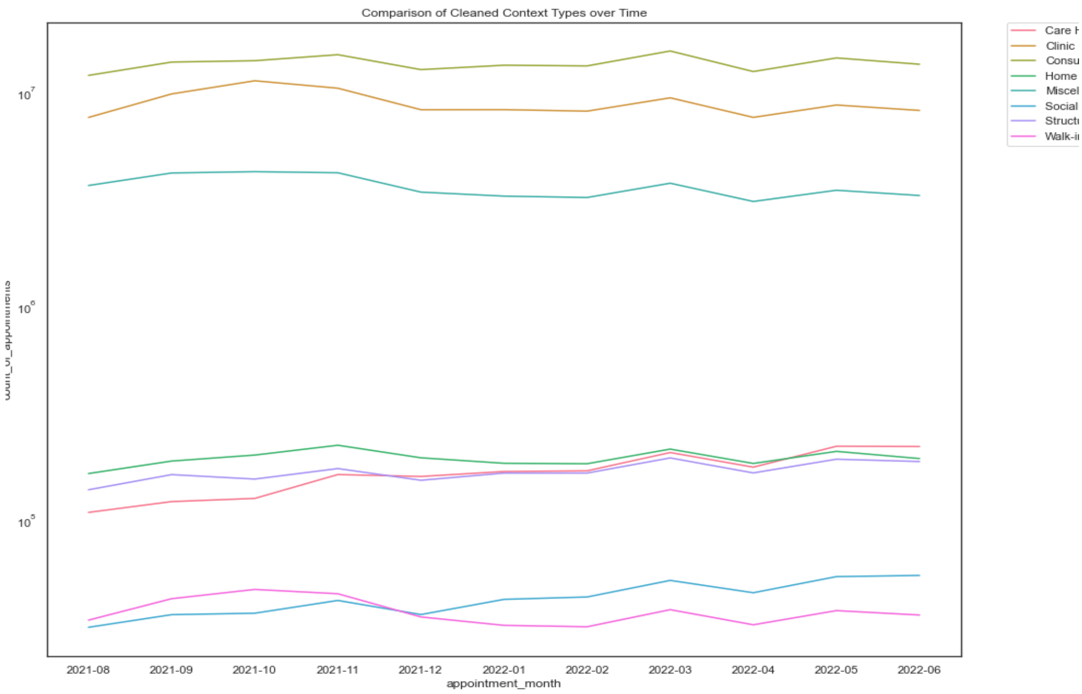
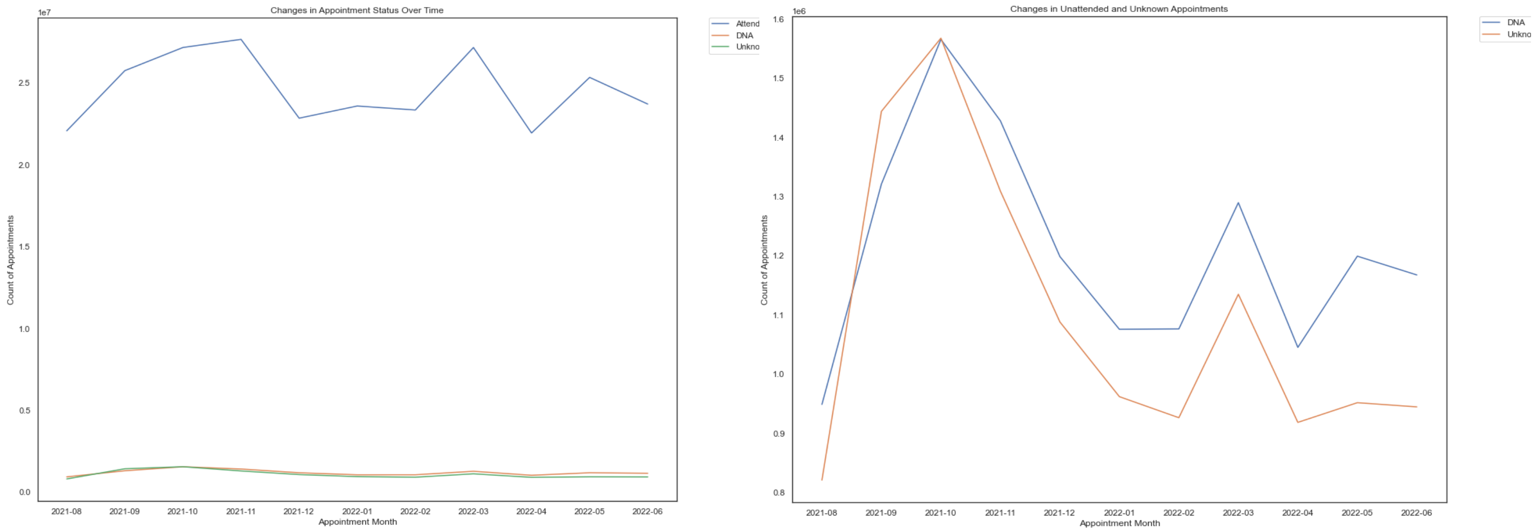


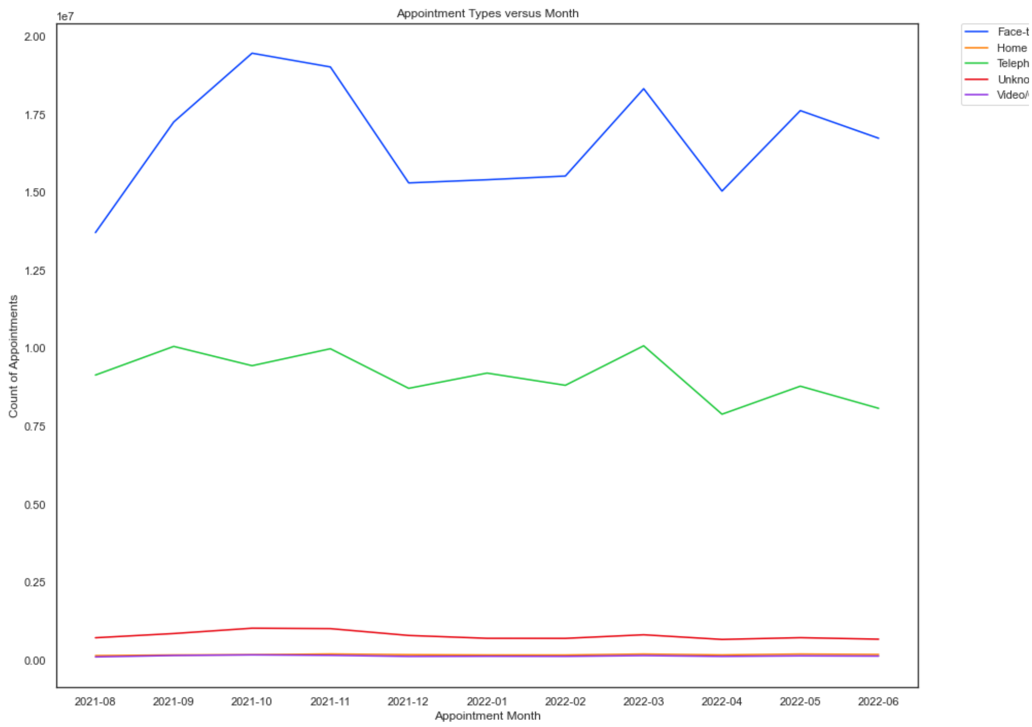
Figure 8: Changes in Healthcare Practitioners Over Time

Figures 9 & 10: Changes in Attendance Over Time



Next, changes in attendance of appointments were measured over time. The majority of appointments are attended, however there are two spikes in appointments not being attended in October 2021 and March 2022. These dates are also in line with an increase in the number of appointments, corresponding to the busiest months. During the busiest months, the demand for face-to-face appointments also increased, while the demand for telephone consultations decreased. The number of video appointments and home visits remained consistently low throughout the year (Figure 11).

Figure 11: Appointment Mode Versus Month



Lastly, another factor influencing the utilisation of capacity is the time between booking and appointment. As shown in Figure 12, the majority of appointments are booked for the same day, with more than half of all appointments in the NHS being booked 7 or fewer days in advance. Throughout the year, the number of same-day bookings increased. Although they account for a smaller proportion of appointments, the number of appointments booked 8 - 28 days in advance all increased, especially in line with surges in the count of appointments (Summer 2021 and Spring 2022), which could suggest a lack of available appointments forcing patients to book further in advance than needed at the start of the year.

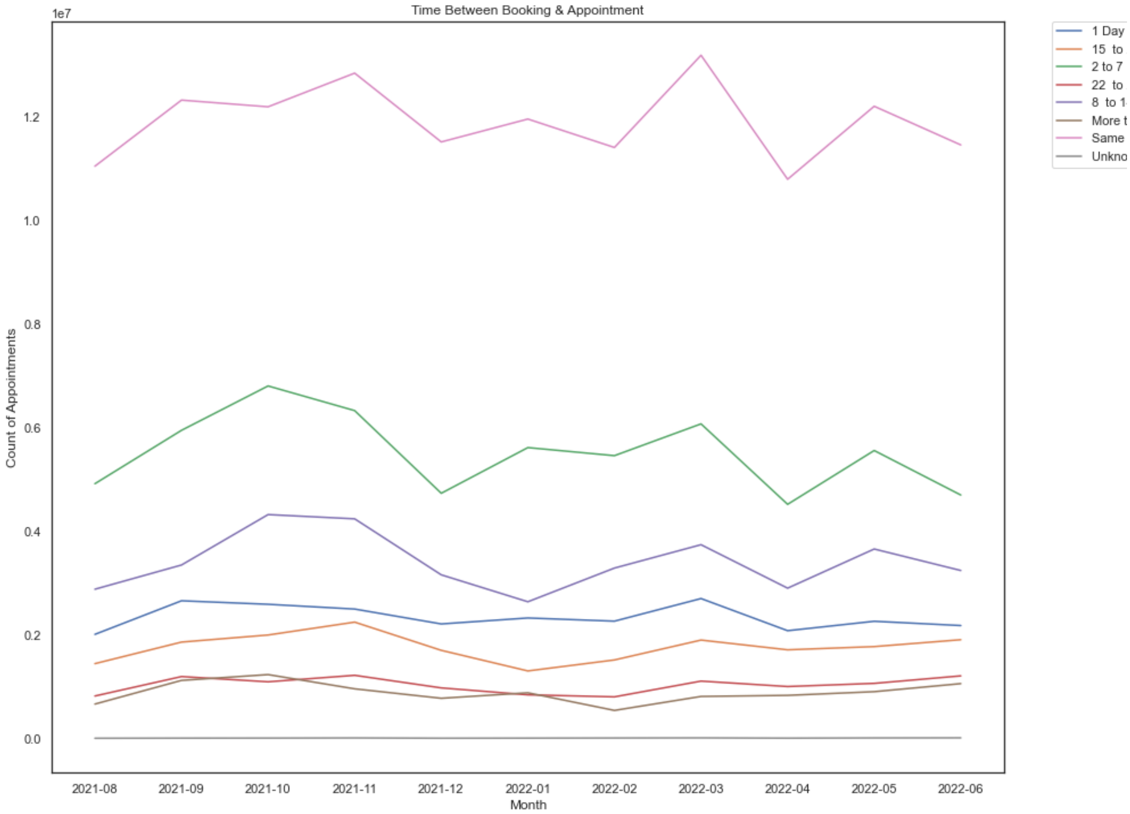


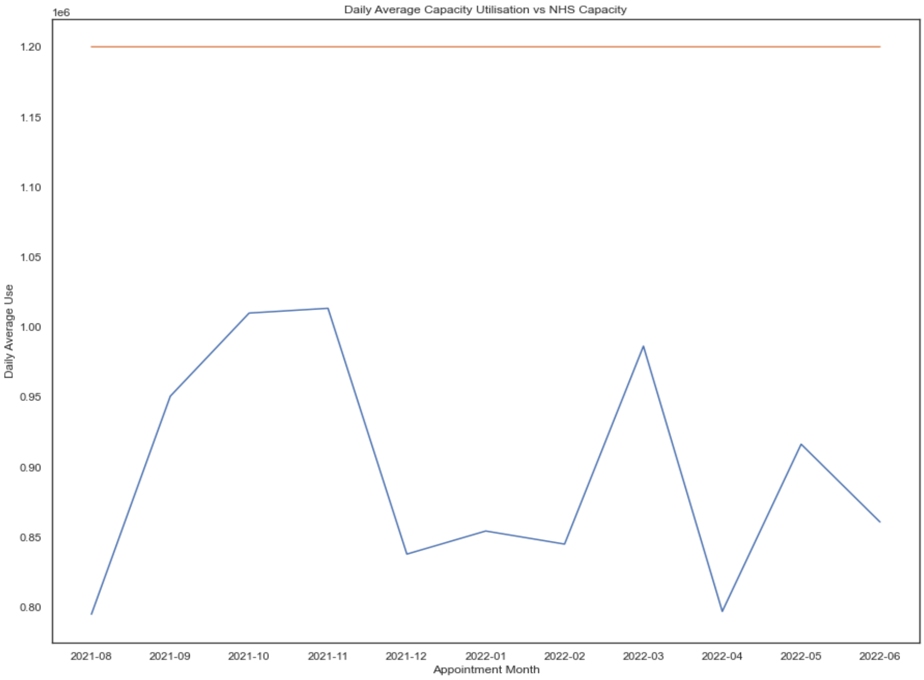
Figure 12: Time Between Booking and Appointment

Collectively, these trends demonstrate patient behaviour and service utilisation across numerous variables which can be used to determine whether the NHS should increase staff. While increasing staff levels would increase capabilities, Figure 13 demonstrates that based on the current daily averages the NHS is not necessarily oversubscribed. This study posits that alternative solutions can be derived from the trends presented in this report by increasing efficiencies in the NHS' existing capabilities.

In addition to analysing capacity utilisation, the team were also mandated to provide analysis on Twitter to determine whether there has been any activity related to healthcare. This was done by filtering the data provided in the Tweet CSV file to group by the frequency and name of a hashtag, omitting all values under 10 re-tweets or likes. The decision to omit these values was done to focus on the most impactful tweets and whether healthcare was featured among this. Once the data was filtered, results were plotted on a series of bar charts.

As demonstrated in Appendix 4, #Healthcare was the most frequently occurring hashtag, followed by #Health. The frequency of healthcare-related hashtags in this dataset suggests the public are interested in the NHS 'activities. Twitter can be used as a mechanism to connect with patients directly and convey important messages such as awareness about the cost of missed appointments. Given the insights of patients who miss the most appointments, outreaches across Twitter can be timed to predictively raise awareness.

Figure 13: Daily Average Appointments Versus NHS Capacity



In addition to using Twitter as a means of communicating with patients, several other recommendations were devised from capacity utilisation trends, which optimise NHS existing capabilities. As the NHS would prefer to avoid penalising patients for missed appointments, they predict when missed appointments are likely to occur and prevent them. Demonstrated in Figure 9, there is a clear trend between the busiest period of appointments and the most missed. Timing notifications, via Twitter or other channels ahead of this could mitigate the risk of patients missing appointments.

Additionally, there is a comparative underutilisation of telephone, video call, and home visit appointments. While additional research would be required to determine the effectiveness of these appointment settings, they could provide an additional means of meeting appointment commitments. Similarly, although the majority of appointments were booked on the same day, throughout the duration of this study the number of appointments booked greater than 2 weeks in advance increased. Although this does not necessarily imply a causal relationship, additional investigation could be conducted to ascertain whether patients are booking further in advance due to a lack of available appointments.

Ultimately, the findings of this report conclude that non-punitive measures can be taken to ensure that appointments are not missed, and these are not necessarily caused by understaffing in the NHS network. The insights presented in the report demonstrate clear behavioural patterns that can be used to optimise existing capabilities rather than increasing staff. By understanding patterns, such as when the surges in missed appointments take place, the NHS can prepare and take proactive steps to mitigate risk. By using interfaces such as Twitter to convey important messages, the NHS can increase public awareness of missed appointments without resorting to fines.

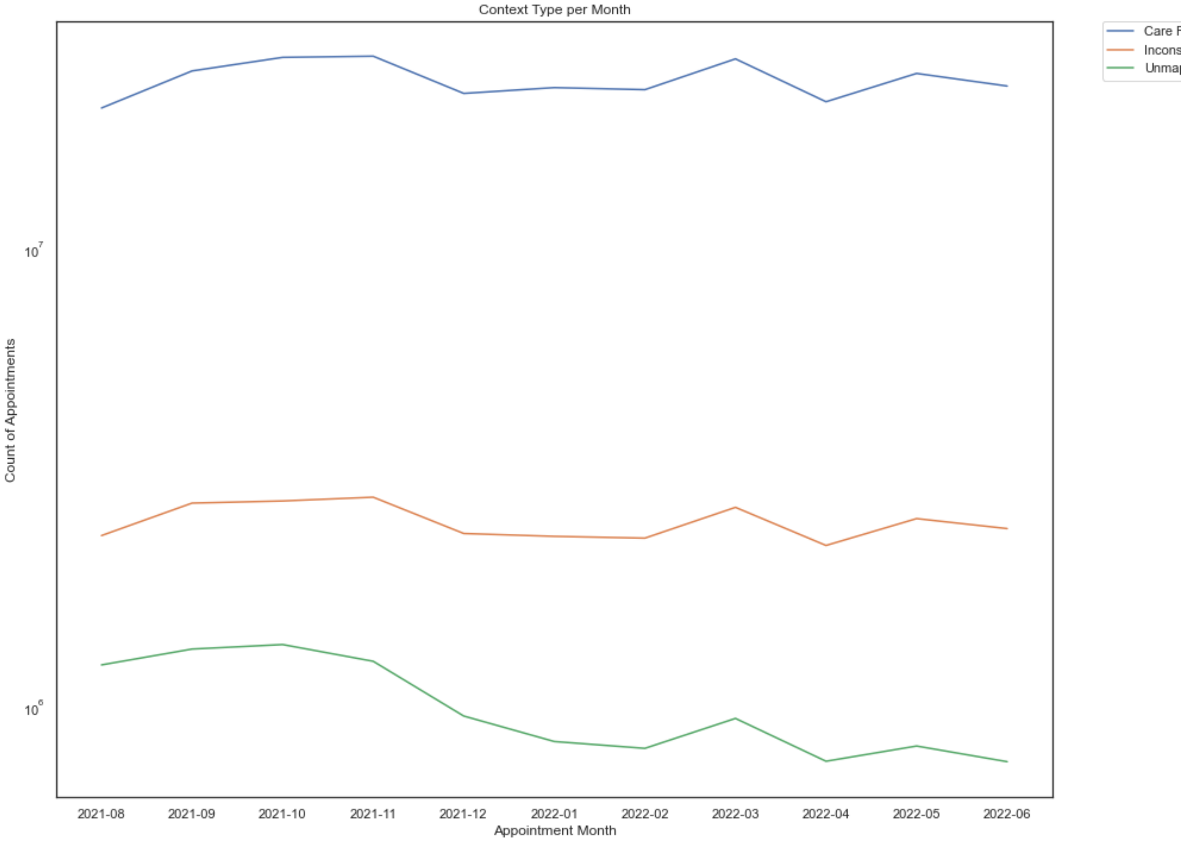
Appendix 1: Monthly Appointments versus NHS Capacity.

As a means of assessing the volume of appointments versus the monthly capacity of the NHS, a line plot was generated. The NHS cited that they have capacity within the network for 1,200,000 appointments per day, which multiplies by 30 to 36,000,000 monthly appointments.

Appendix 2: Context Types

The context type throughout the dates in question were consistent, showing minimal changes in demand. For this reason, the results are tabulated here.

To make this chart digestible, a logarithmic Y scale was chosen, which shows each context type trend in greater detail. The most frequently occurring context type was a Care Related Encounter. There was also a significant dip in unmapped context types within the same timeframe as the service settings fluctuation, again suggesting that the quality of recorded data had improved during this time period.



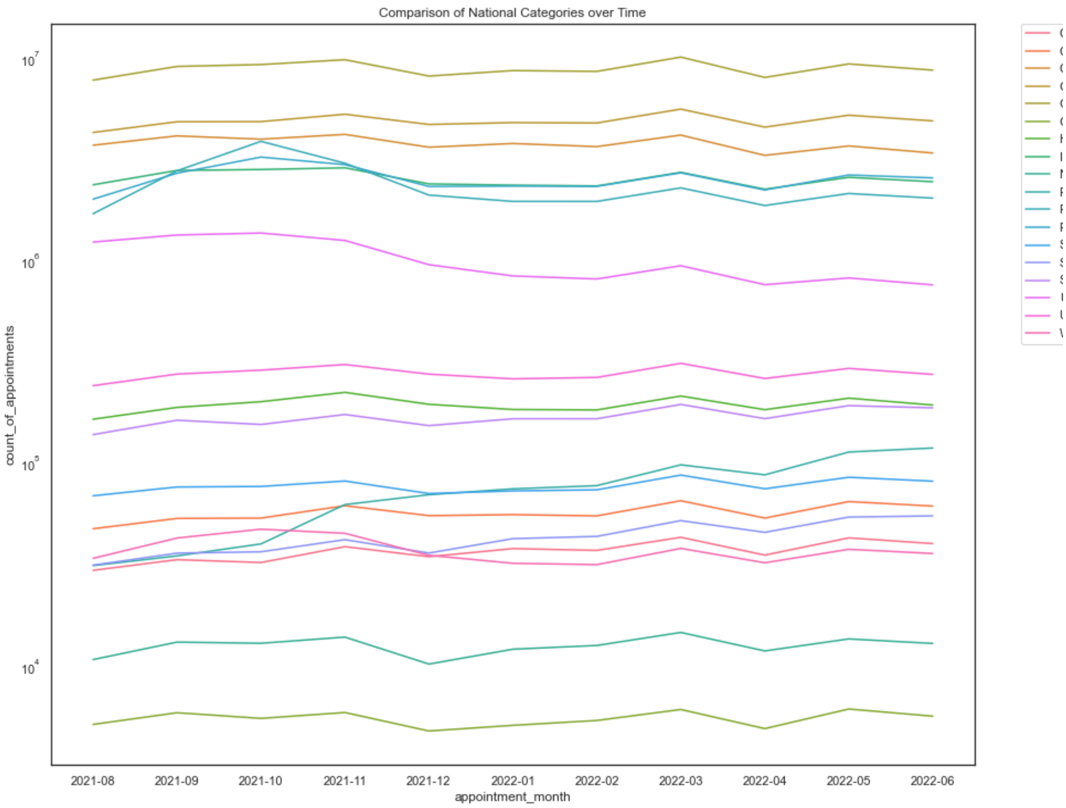
Appendix 3: Code used to remap categories in National Categories

The following code was used to remap the categories to make for easier visualisation. The original chart is plotted below.

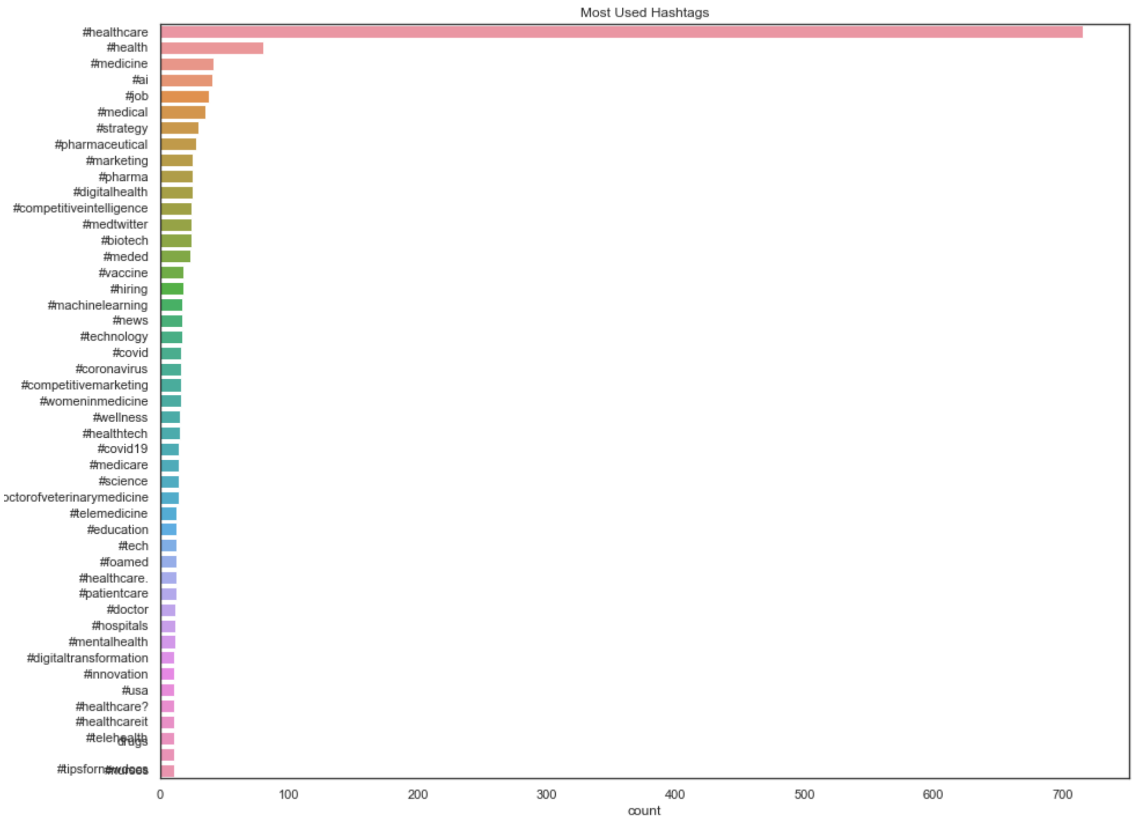
In the below chart, Patient Contact During Care Home Round appointments contributed to the spike seen in Chart XYZ, representing a steep growth in demand during the year. The number of walk-in appointments dipped significantly during the period of October 2021 - February 2022, and although increasing again in the subsequent months, the number of appointments dipped overall.

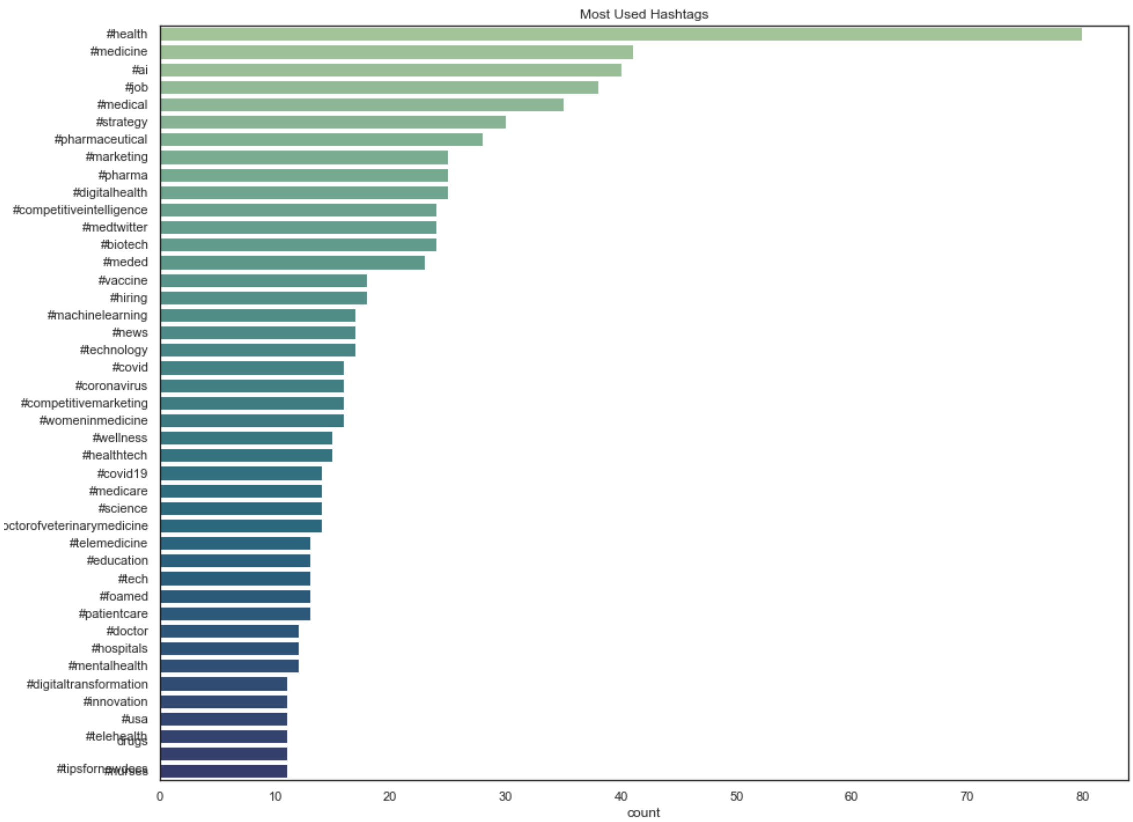
In [61]:

```
# Dictionary to remap categories.

remap_cat_dict = {
     'Care Home Needs Assessment & Personalised Care and Support Planning': 'Care Home',
     'Care Home Visit': 'Care Home',
     'Clinical Triage': 'Clinic',
     'General Consultation Acute': 'Consultation',
     'General Consultation Routine': 'Consultation',
     'Group Consultation and Group Education': 'Consultation',
     'Home Visit': 'Home Visit',
     'Inconsistent Mapping': 'Miscellaneous',
     'Non-contractual chargeable work': 'Miscellaneous',
     'Patient contact during Care Home Round': 'Care Home',
     'Planned Clinical Procedure': 'Clinic',
     'Planned Clinics': 'Clinic',
     'Service provided by organisation external to the practice': 'Miscellaneous',
     'Social Prescribing Service': 'Social Prescribing Service',
     'Structured Medication Review' : 'Structured Medication Review',
     'Unmapped': 'Miscellaneous',
     'Unplanned Clinical Activity': 'Clinic',
     'Walk-in': 'Walk-in'
}


new_cat = nc_ss['national_category']
unique_cats = sorted(nc_ss['national_category'].unique())

for cat in unique_cats:
    if cat in remap_cat_dict:
        new_cat = new_cat.replace(cat, remap_cat_dict[cat])

nc_ss['new_category'] = new_cat
```



Appendix 4: Most Frequently Used Twitter Hashtags





Appendix 5: Service Setting Box Plots

These charts are less visually appealing by comparison to the line plots in Figures 5 & 6 but are useful in determining the spread of values across service settings.