The NHS wishes to reduce appointments missed by patients, as missed appointments incur significant costs. To solve this problem, the reasons for missed appointments must be understood better. To improve understanding of this issue, I will answer these questions: has there been adequate staff and capacity in the networks?; what was the actual utilisation of resources?; and several sub-questions.

I created a GitHub repository to store my project files (see Appendix Fig. 1). I used read_csv() and read_excel() to import each NHS dataset into a DataFrame. For each DataFrame, I utilised columns, shape and dtypes methods to obtain column names, number of rows and columns, and data types. I created DataFrames using df.isna() to determine the number of missing values in the datasets, and discovered no missing values. For the original DataFrames, I used describe() and info() to obtain descriptive statistics and metadata. I utilised nc[1] to count the number of locations, service settings, context types and national categories. For each column, I employed unique() to generate a list of unique values, and len() to count the unique values. I used print() to return each count alongside explanatory docstrings. I used the same approach for Appointment Statuses in ar[2]. There are 106 locations, 5 service settings, 3 context types, 18 national categories and 3 appointment statuses. To determine the five locations with the most records, I employed value_counts() to create a table of records per location, sorted in descending order. I assigned a variable to this table and used head() to display the five locations with the highest records.

After converting appointment_date's data type in ad[3] to datetime64[ns] using astype(), I utilised min() and max() on appointment_date to determine first and last appointment dates in ad and nc (I used the same approach for appointment_month in ar). Across all datasets, the reporting period covered January 2020 until 30th June 2022. Only ar encompassed the entire period. To determine which service setting reported the most appointments from 1-January-2022 – 1-June-2022 in North-West London, I created an nc subset using loc() to filter on sub_icb_location_name and appointment_date. I used groupby(), sum() and sort_values() on this subset to calculate the most popular service setting, which was clearly General Practice (GP) (c.4.8m appointments). Since only ar covered the entire reporting period, I employed groupby(), sum() and sort_values() on ar to calculate appointments per month, revealing November 2021 as the busiest month (c.30.4m appointments). I used value_counts to calculate records per month in ar and nc (I grouped ad by month and counted records per month). This analysis revealed which months and service settings NHS demand was highest for, helping answer if there was adequate capacity to meet demand.

I used groupby() and sum() to create DataFrames that calculated monthly appointments for each service setting, context type, and national category respectively. I used lineplot() to visualise monthly appointments for each categorical variable (see Fig. 2-4), as lineplots are effective at showing value changes over time. GP, Care Related Encounters and General Consultation Routine were the most popular values per categorical variable. Amongst these values, appointments were lowest in August-2021 & April-2022. GP appointments and Care Related Encounters were highest in October-November-2021 & March-2022, while General Consultation Routine appointments were highest in November-2021 & March-2022. I used groupby() and sum() to calculate daily appointments per service setting. To effectively visualise time-series, I generated 4 lineplots visualising daily appointments per service setting

[1] National Categories DataFrame
[2] Appointments Regional DataFrame
[3] Actual Duration DataFrame

in August/Summer, October/Autumn, January-2022/Winter and April/Spring (see Fig. 5-8). In each season, GP was the most popular service setting, featuring weekday appointments being significantly higher than weekday and bank holiday appointments. The weekday with the highest GP appointments was in October (which also featured the highest Saturday GP appointments), while the weekday with the lowest GP appointments was in August. Overall, NHS demand appears highest in early Autumn, and lowest in early Summer.

I used read_csv() to import 'tweets.csv' into a DataFrame, followed by a for loop to create a list of hashtags in the tweet_full_text column, and stored this list in a variable (tags) using append(). I employed Series() to convert tags into a Series (hashtags), and created this dictionary: {'word': hashtags, 'count': 1}. I utilised DataFrame() to convert this dictionary into a DataFrame (data), where each cell in the 'word' and 'count' columns contained a hashtag and 1 respectively. I created a data subset by using groupby() and sum() to count the utilisations of each hashtag, and then created an additional subset by filtering out hashtags used 10 times or less. I used barplot() to visualise hashtags used over 10 times (see Fig. 9), and created another barplot that visualised the top trending hashtags without overrepresented hashtags (i.e. "healthcare" and "health") (see Fig. 10). "job" and "hiring" were amongst the remaining top 15 trending hashtags, suggesting the NHS is understaffed. Generally, tweets can be used by the NHS to identify and act on positive and negative feedback. Trending hashtags can also be used in tweets to communicate important information to wide audiences.

To understand why missed appointments occurred from August-2021 to June-2022, I created visualisations (see Jupyter Notebook for process description). Fig. 14 shows most months feature more than 1m missed appointments. Fig. 16 indicates more patients waited 2-14 days between booking and appointments than patients who waited 1 day. Many patients thus wait over a week, and long waiting periods between booking and appointment are likely a significant reason for missed appointments: patients' issues could resolve themselves within long waiting periods, so patients no longer need to attend their appointments. Additionally, Fig. 12 indicates each month had an average daily NHS utilisation below 1.2m appointments[4]. This suggests there's inadequate staff within the NHS to meet appointment demand, and that there's extra capacity within the networks to increase appointments. Therefore, I recommend increasing staff. This should reduce waiting times between appointments and the likelihood that patients miss appointments due to their issues resolving themselves. GP appointments are the most utilised healthcare professional type and service setting (see Figs. 13 and 17), so more GPs should be hired to reduce waiting times for GP appointments and missed GP appointments. Fig. 15 shows telephone appointments are the 2nd most utilised appointment mode, so GPs should commit to more telephone appointments to reduce patients waiting for face-to-face appointments[5], and thus waiting times for face-to-face appointments and missed face-to-face appointments. For future analysis, I recommend surveying patients who missed appointments on their reasons for doing so to obtain further actionable insights.

[4] The maximum daily number of appointments the NHS can accommodate.
[5] The most utilised appointment mode.

Appendix

Fig. 1: GitHub repository

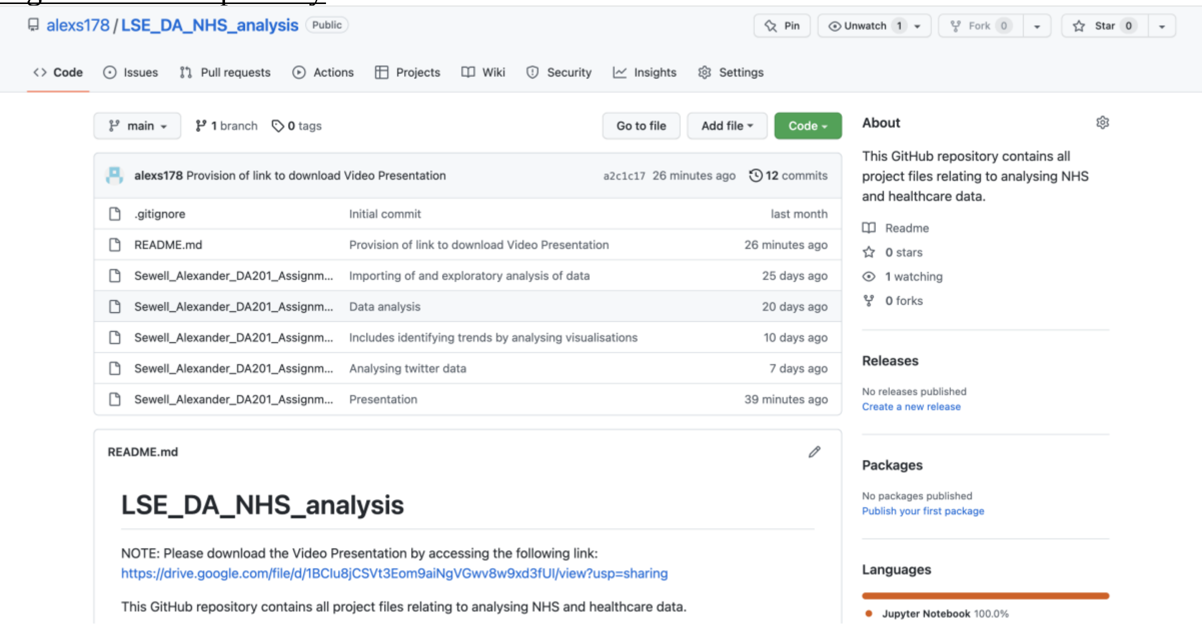

Fig. 2: Monthly Appointments per Service Setting

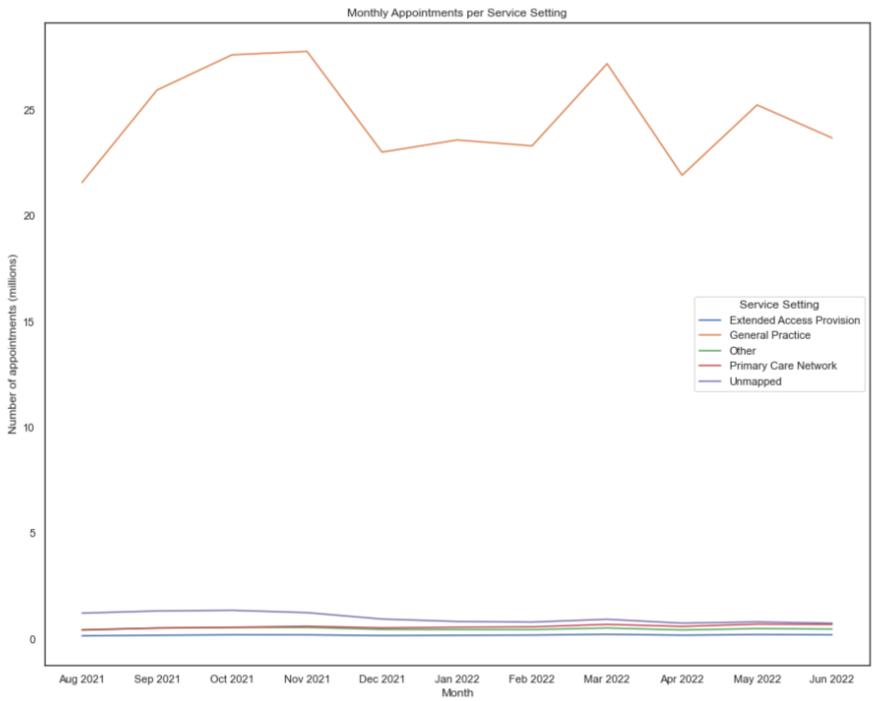

Fig. 3: Monthly Appointments per Context Type

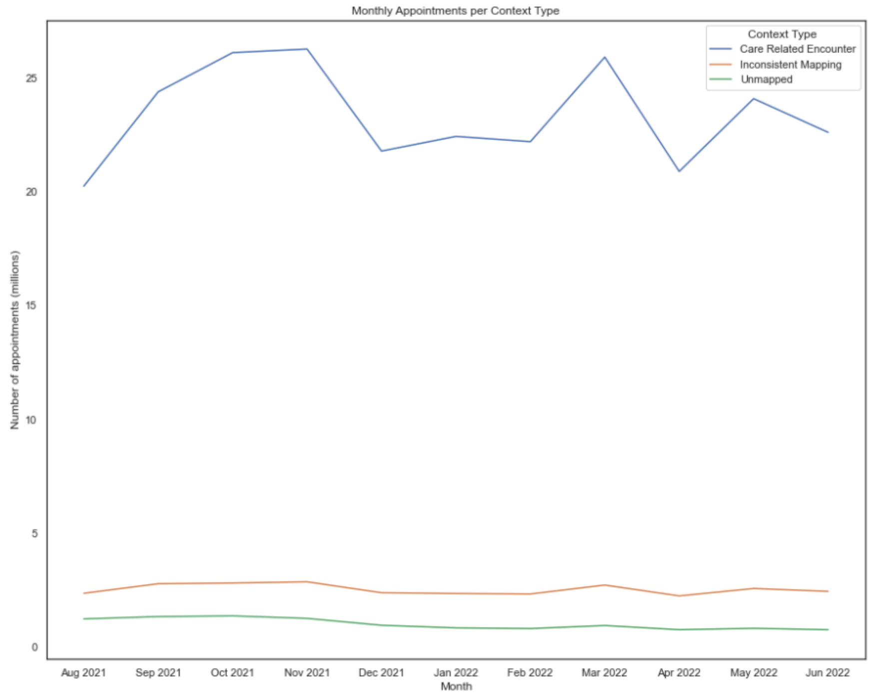

Fig. 4: Monthly Appointments per National Category

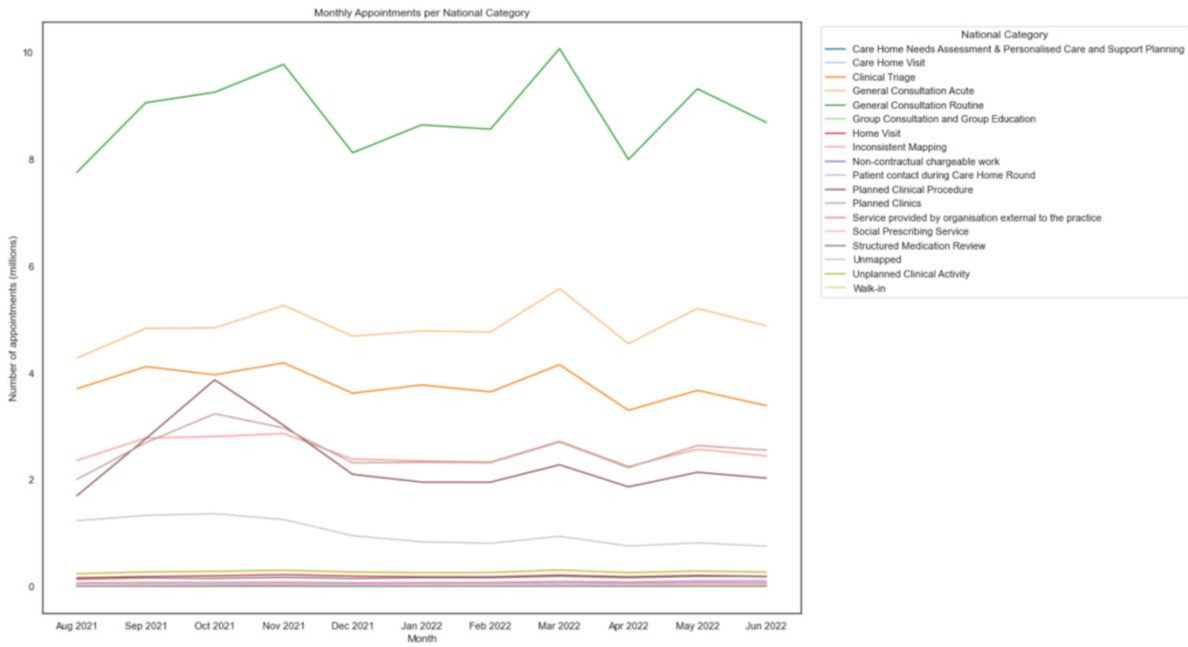

Fig. 5: Daily Appointments per Service Setting – Summer (August 2021)

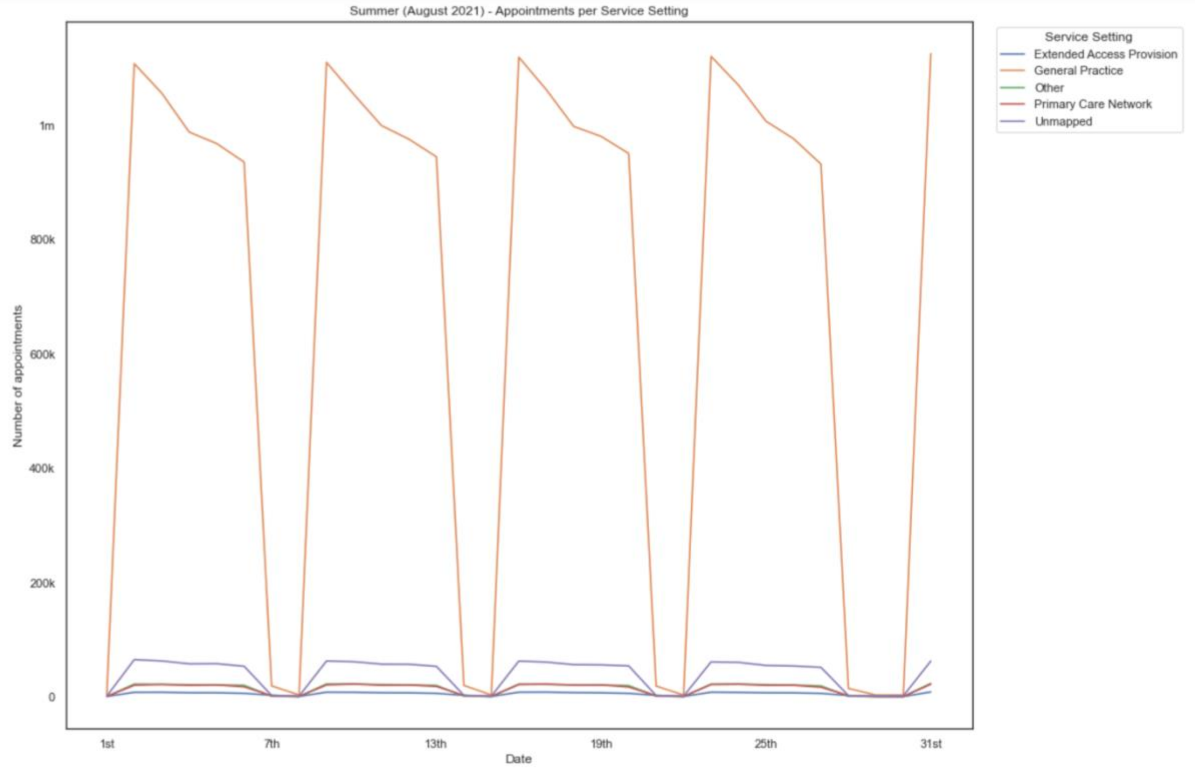

Fig. 6: Daily Appointments per Service Setting – Autumn (October 2021)

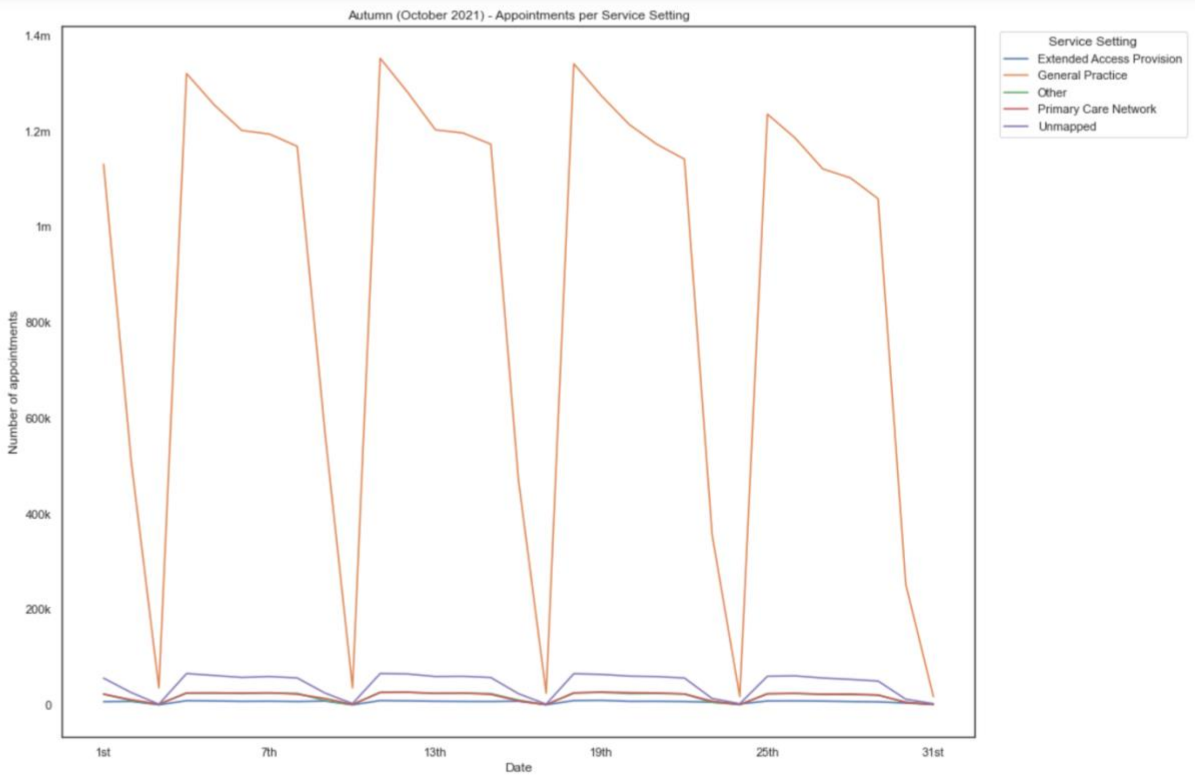

Fig. 7: Daily Appointments per Service Setting – Winter (January 2022)

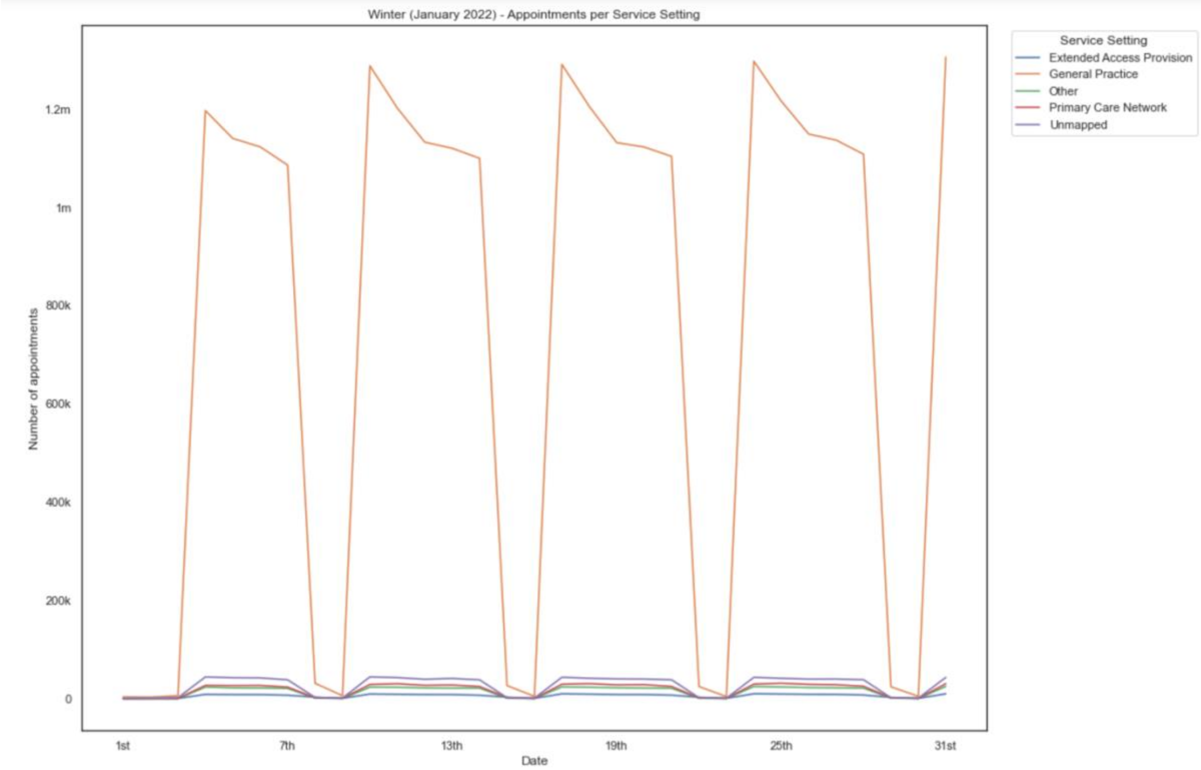

Fig. 8: Daily Appointments per Service Setting – Spring (April 2022)

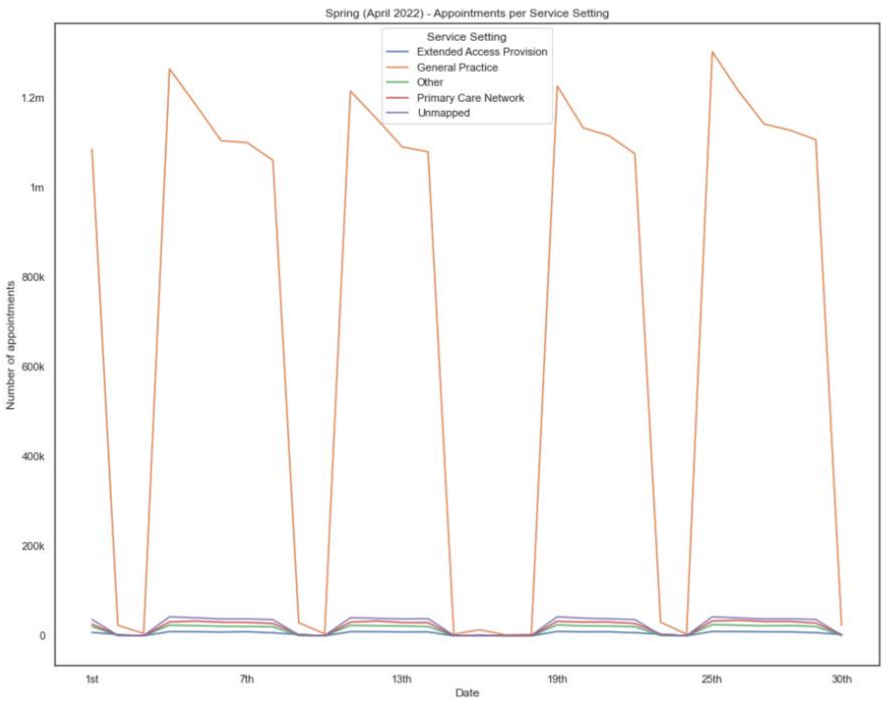

Fig. 9: Top trending hashtags on Twitter related to healthcare in UK

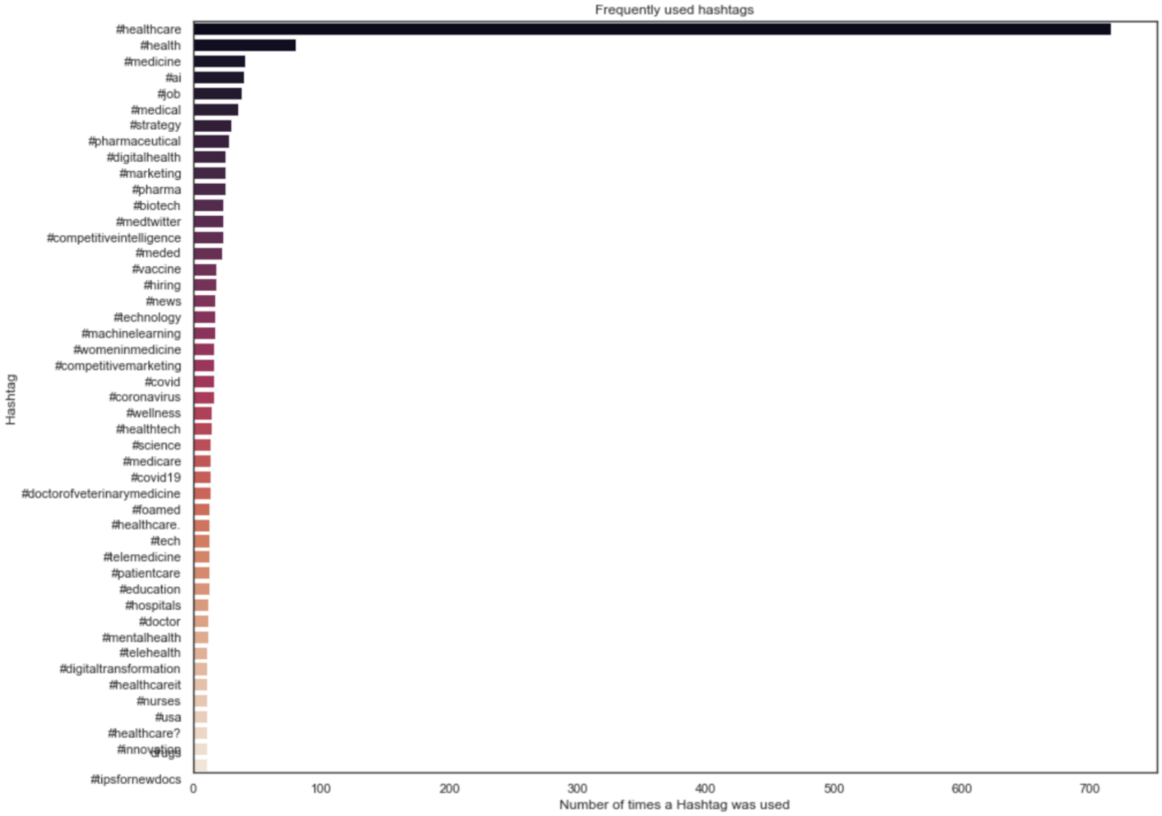

Fig. 10: Top trending hashtags without overrepresented hashtags

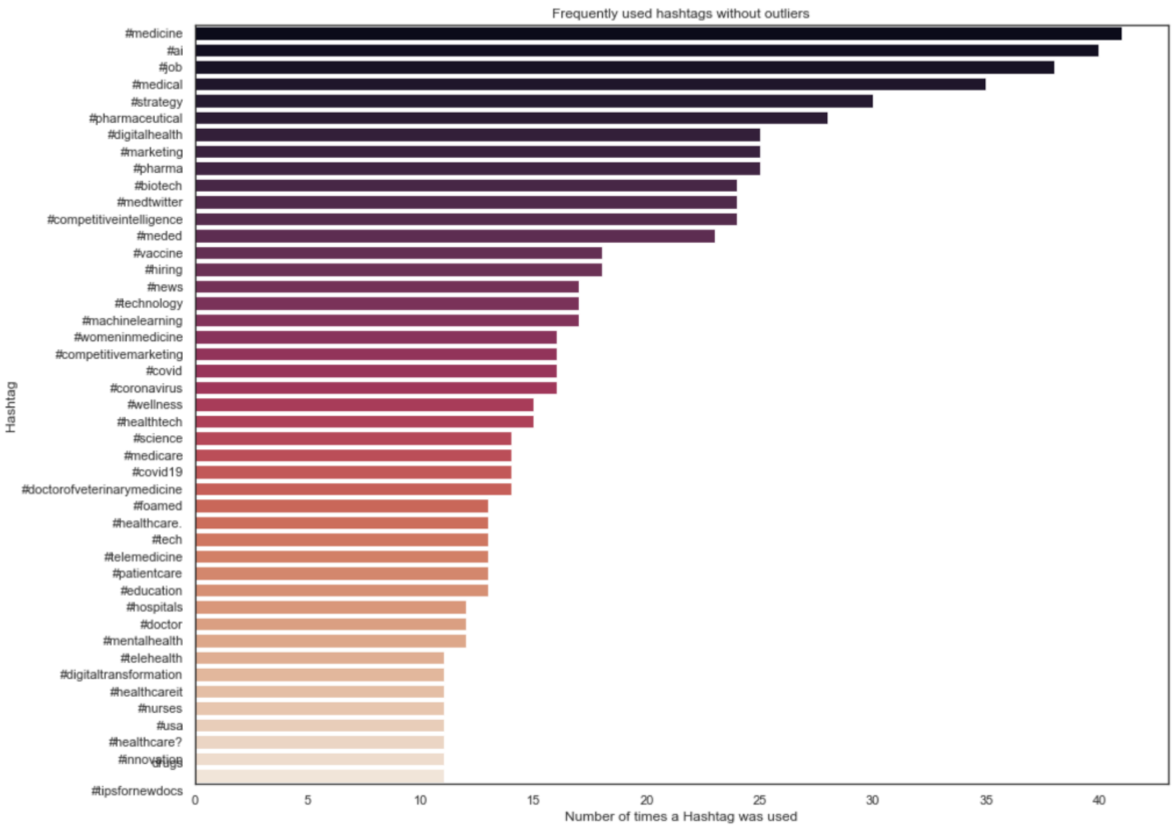

Fig. 11: Number of monthly visits

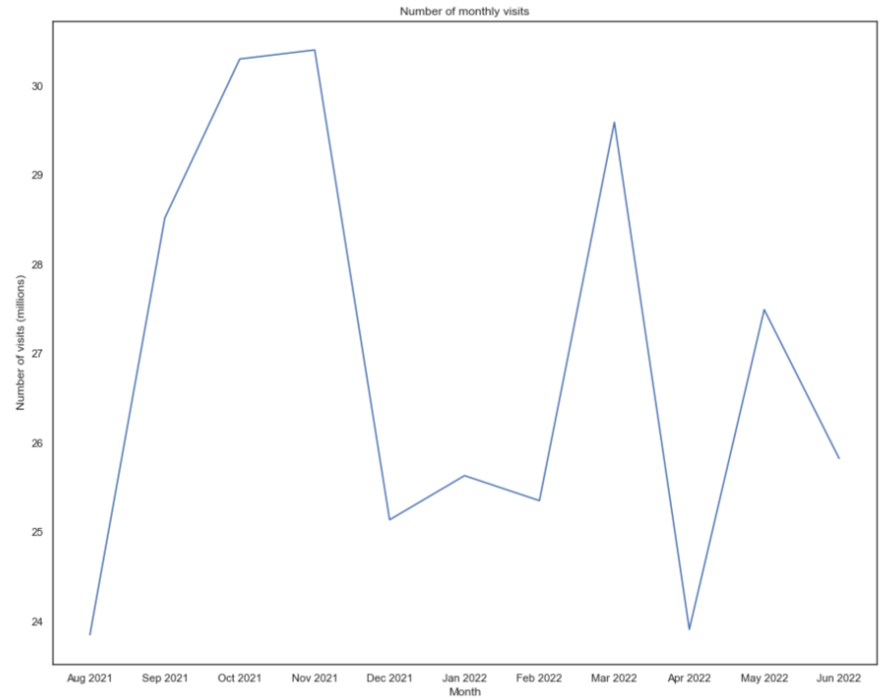

Fig. 12: Monthly capacity utilisation

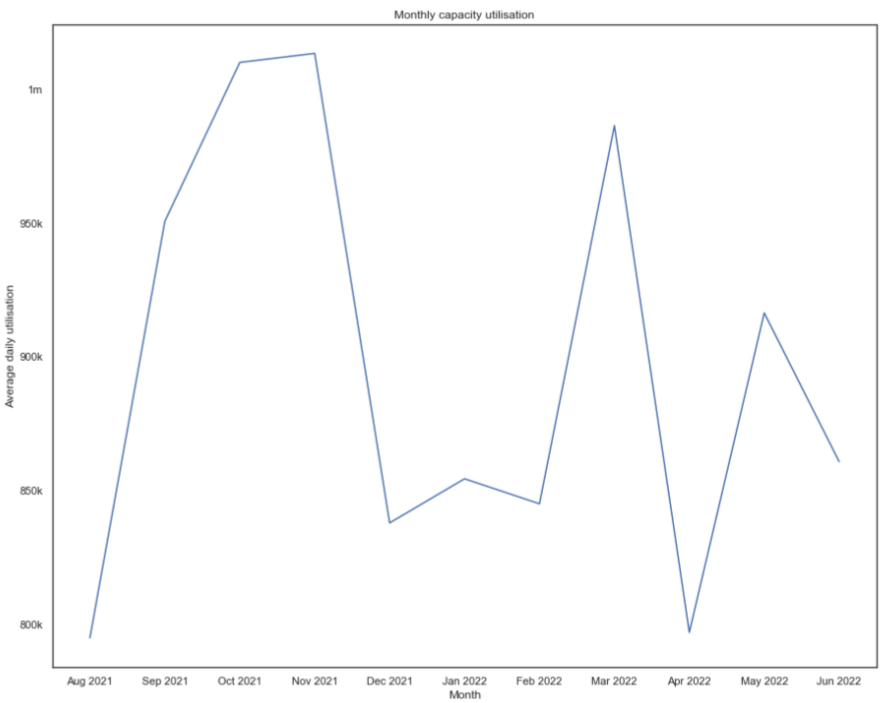

Fig. 13: Healthcare Professional Types over time

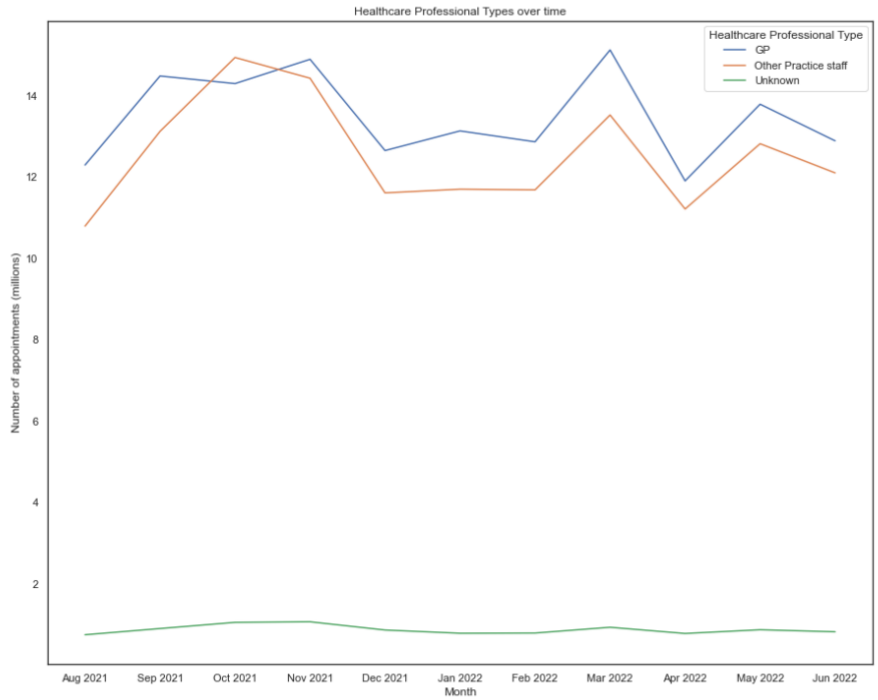

Fig. 14: Appointment Statuses over time

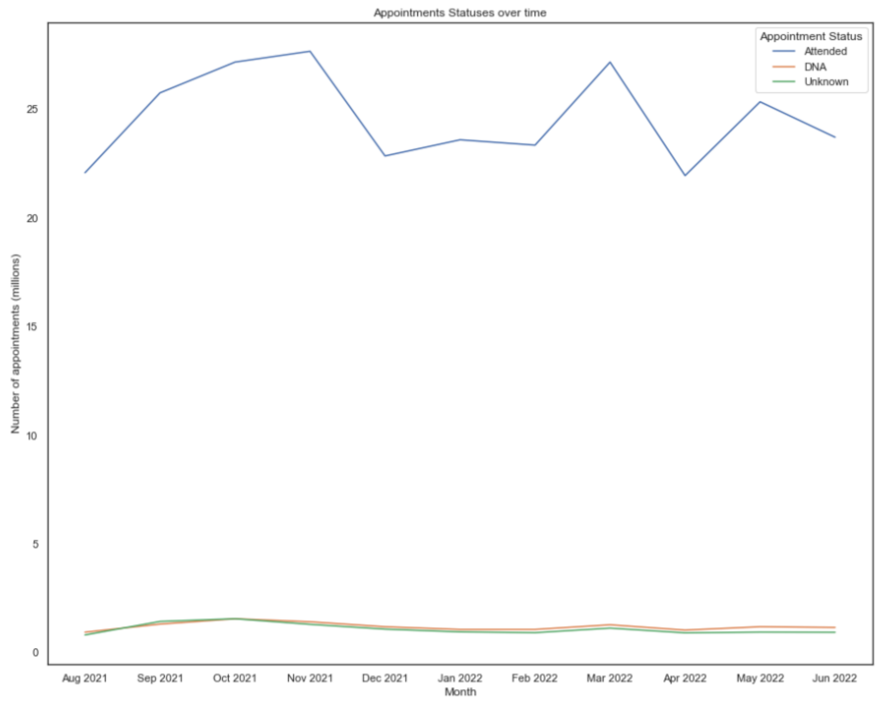

Fig. 15: Appointment Modes over time

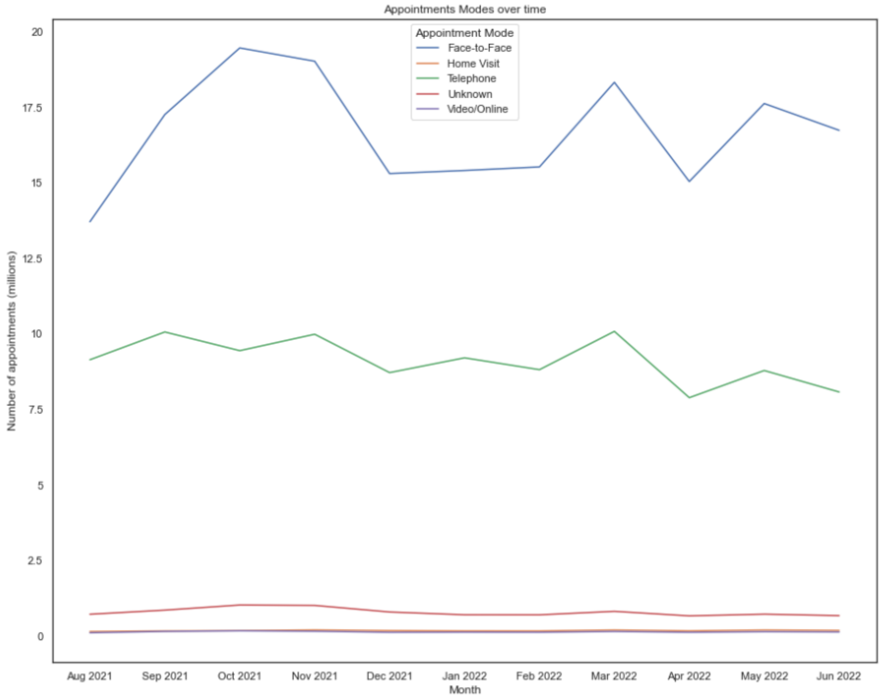

Fig. 16: Time between Booking and Appointment

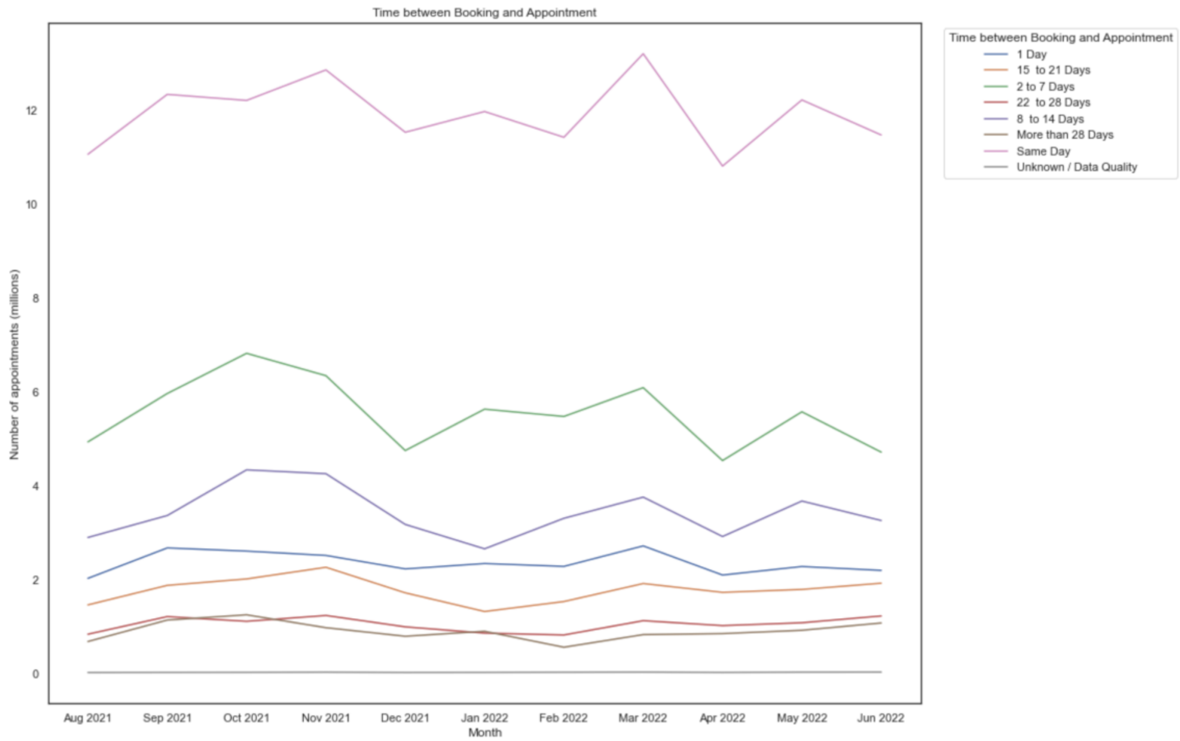

Fig. 17: Spread of Service Settings

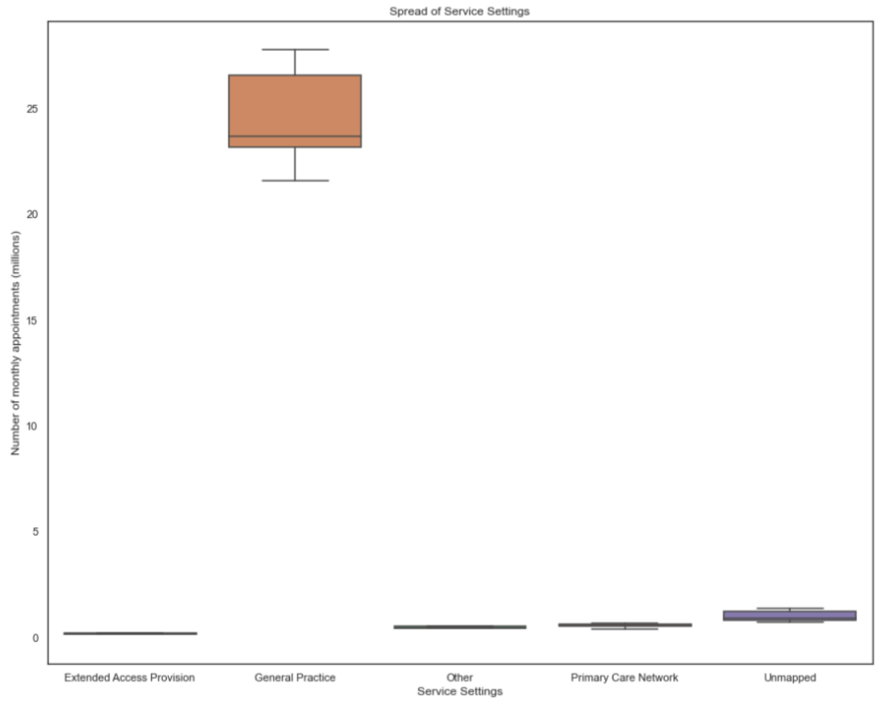

Fig. 18: Spread of Service Settings (excluding GP visits)