NATIONAL HEALTH SERVICES

AN ANALYSIS INTO GENERAL PRACTITIONER APPOINTMENTS

Gavin van de Bunt

ON BEHALF OF LONDON SCHOOL OF ECONOMICS

Table of Contents

Introduction and business context

This report describes the approach and findings regarding a data analysis study on General Practitioner (GP) appointments with the National Health Services (NHS). NHS is a publicly funded healthcare system in England and incurs significant, potentially avoidable costs, when patients miss GP appointments.

Therefore, the primary objective is to get a better understanding on the reasons of why patients miss GP appointments. This study is a preliminary investigation into this issue and the main outcomes will be to identify areas for further research to fully achieve the key objective.

This report supports NHS's objective by answering the following key questions:
- What are the basic characteristics of the data set provided by the NHS?
- What monthly and seasonal trends in appointment numbers and categories are evident from the data?
- What sentiments can be found on Twitter related to healthcare in the UK?
- Have the NHS been sufficiently staffed to handle demand for appointments?
- Is there any visible trend in the amount of missed NHS appointments?

The following chapters will describe the analysis approach used, the visualisations and insights retrieved from the data, and concludes with recommendations for future research.

Analytical Approach

The full analysis can be found in the accompanied Jupiter Notebook, uploaded in a public GitHub repository (https://github.com/G-Bunt/LSE_DA_NHS_Analysis). Below a brief description of the various key elements.

Data input

There are four input files that serve as the basis for the analysis in this report:
1. actual_duration.csv – Details of appointments made by patients.
2. appointments_regional.csv – Details on the type of appointments made by patients.
3. national_categories.xlsx – Details of the national categories of appointments made by patients.
4. tweets.csv – Data related to healthcare in the UK scraped from Twitter.

The client also provided metadata including available fields and a description of each field.

Import and explore the data

The Pandas library was used to import the various files and create relevant DataFrames for exploration.

Various basic functions were used to get a better understanding of the relevant DataFrames:
- .dtypes to understand the datatypes of each DataFrame. Most of the values are object/string type, including the appointment_month, which is important for later grouping and visualisation work.
- .shape: view the row and column counts.
- .isna().sum() functions to check for missing data. The metadata provided already indicated the data was cleaned, and this was confirmed as there were no missing values.
- .head() and .tail() to view first and last rows and check the values.
- .describe() to get the descriptive statistic for the numerical values in the data sets. For the three main DataFrames, this was only relevant for the count of appointments.

Further exploring the different features of the DataFrames confirmed basic key characteristics of the data, as documented in the included metadata. This was done using a combination of the ".unique" function and the "len" function, used to count the number of unique categorical values.
Location
- Location is gathered at Integrated Care Board (ICB) level across all available data. There are 42 ICBs in total. There are 106 sub-ICBs, but this granularity is not available in the "Appointments Regional" dataset.
Date ranges
- Using the ".min" and ".max" functions, the date range of the various DateFrames was checked. The date ranges differ significantly across the DataFrames. This is surprising, as it is not mentioned clearly in the metadata that the National Categories file only starts in August 2021.
Other key attributes
- There are 5 different service settings.
- There are 3 different context types.

- There are 18 different national categories.
- There are 3 different appointment statuses.

Analyse the data features

After exploring the basic attributes, a first analysis of the available data was done. Using the ".groupby" function, it was discovered that the number of appointments seemed to have two peaks:
- Late Autumn (Oct-Nov.) of 2021
- Early Spring (March) 2022

This information was used for further deep dives later in the study.

Using Seaborn plots (see figure 1 below), it was also quickly discovered that the vast majority of appointments involve the General Practice. The number of appointments follows the seasonal peaks in demand as found earlier.
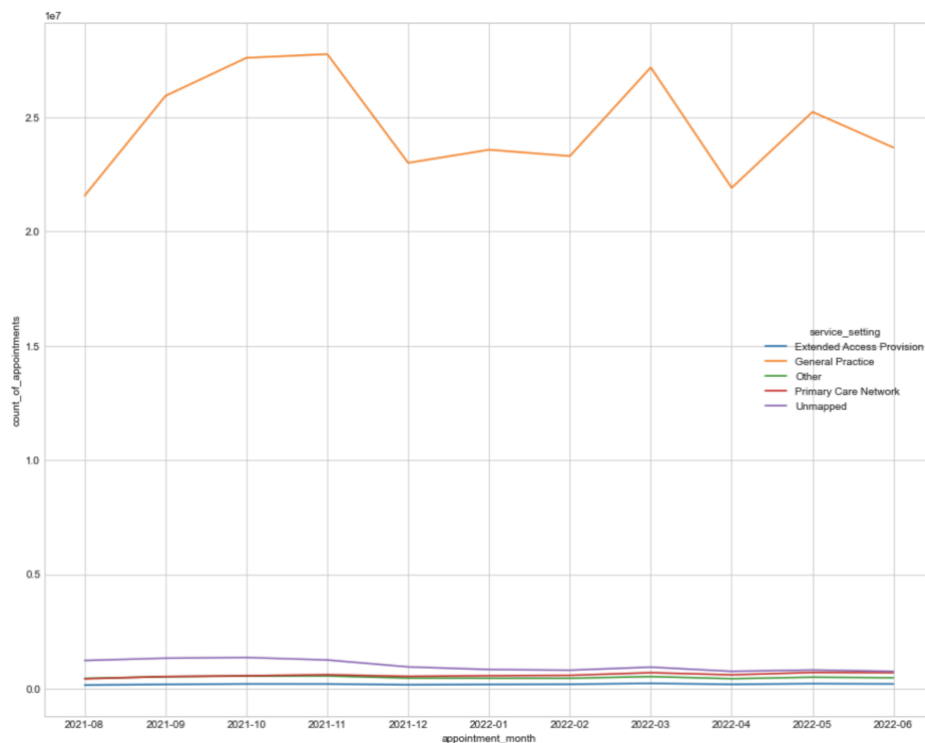

Figure 1: monthly appointments per service setting.

Plotting the different context types shows that the only consistently mapped instance is the Care Related Encounter (see Figure 2).
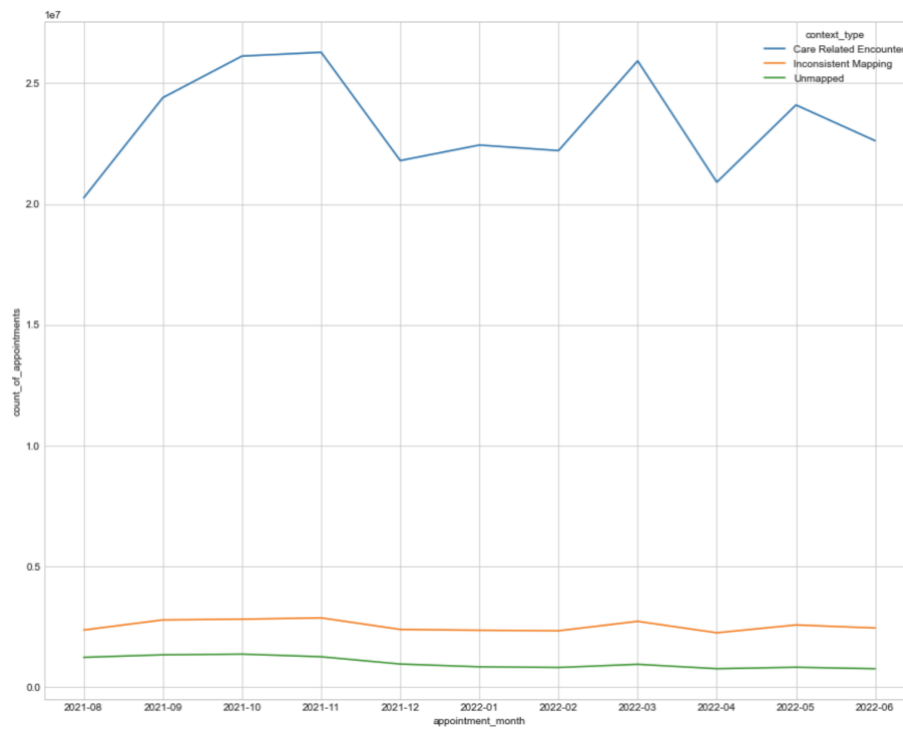
Figure 2: monthly appointments per context type.

Figure 3 shows that the majority of appointments fall within either of two General Consultation categories (Acute and Routine).
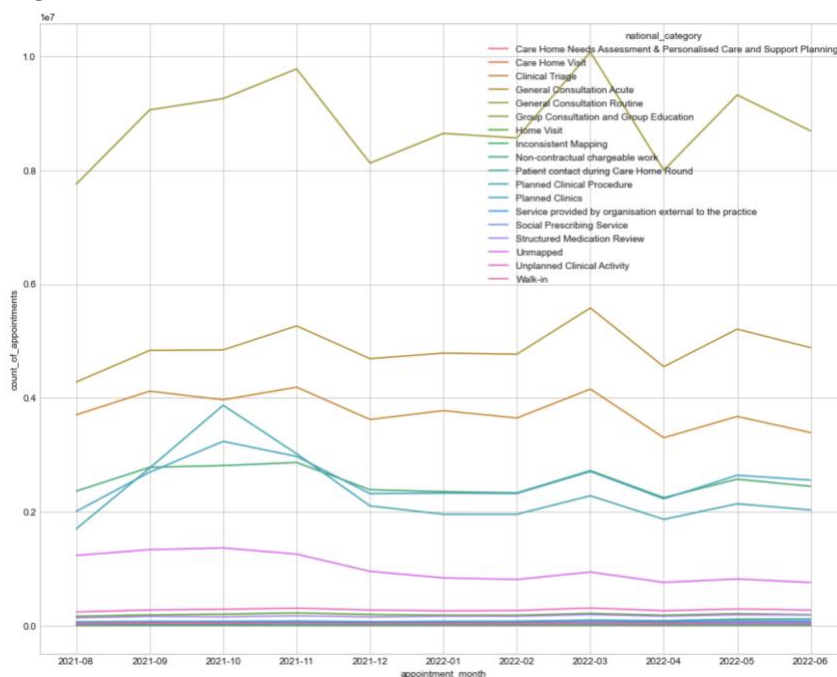

Figure 3: monthly appointments per National Category.

Visualisations, insights and trends

After exploring, describing, and building an understanding the available data set, the next step is to dive into the business objectives.

Explore sentiment on Twitter

As we are ultimately looking at an explanation of human behaviour, it was decided to explore sentiment related to Healthcare in the UK on Twitter, one of the largest social media platforms. This can help provide context to the study, in the wider social setting of the timeframe explored.

A cleaned dataset was provided and loaded with Pandas. From the created DataFrame, the top trending hashtags were explored for any potential clues for the issues identified by the NHS. While figure 4 shows that the top trending hashtags (#) are all quite generic, a further examination of the raw data shows a few clues that are worth further exploring:

1. There are several items related to employment, potentially indicating staffing issues: "job", "hiring", "nurseproblems".
2. As in any study, it is important to keep the context of the environment in mind. In this case, the time period investigated is obviously heavily impacted by an unprecedented global pandemic. This is clear from some of the hashtags as well: "vaccine", "covid", "coronavirus", "covid19".
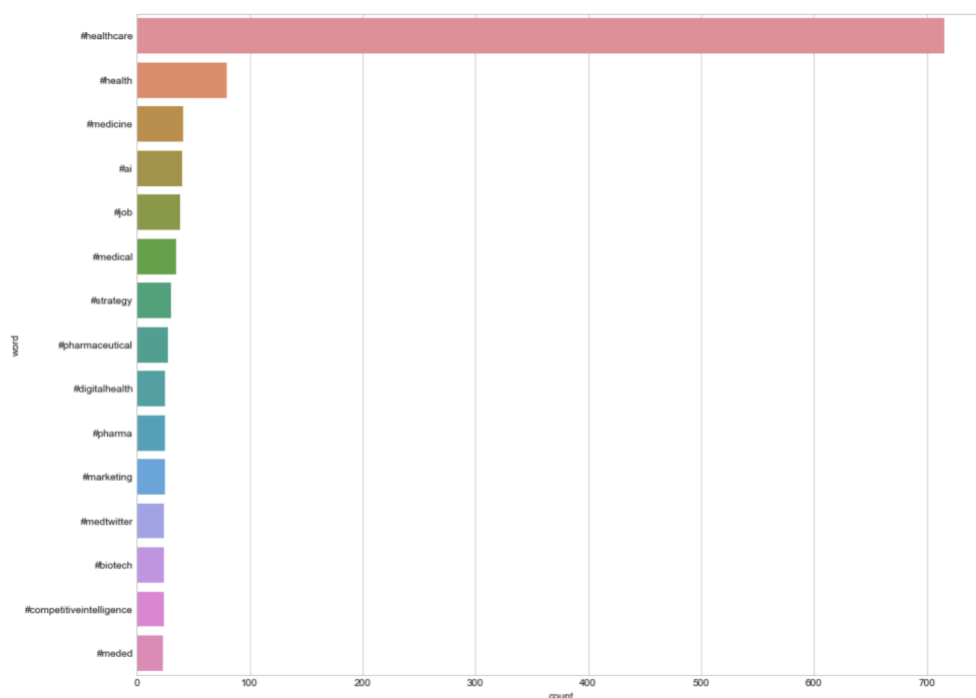

Figure 4: top trending healthcare Twitter topics.

Explore staffing and resourcing in the NHS

To aid in the analysis, a valuable piece of information was added: the NHS can, on average, accommodate a maximum of 1,200,000 patients in a single day. This opened up the opportunity to

explore whether utilization rates within the NHS show any significant trend changes over time. New columns were created:
- "utilisation": divide the monthly appointment numbers by 30 to get the approximated average daily utilization.
- "ut_rate": divide this daily utilization by the capacity of 1,2m patients.

Figure 5 below shows that, while on average the utilization never went close to 100%, the months identified earlier as peaks, do show that there is potentially significant stress on the system. It is important to keep in mind that these are all averages, which points to a few areas of further research:
- Was the NHS at maximum capacity throughout all these months? As could be seen in the Twitter analysis, the COVID19 pandemic not only negatively influenced the wellness of the public, but it is also reasonable to assume that there were significant peaks in NHS staff health-related absence. These numbers are unavailable for this study.
- Does a high utilization rate actually negatively impact the relative attendance rates of appointments? This is further explored in the section below.
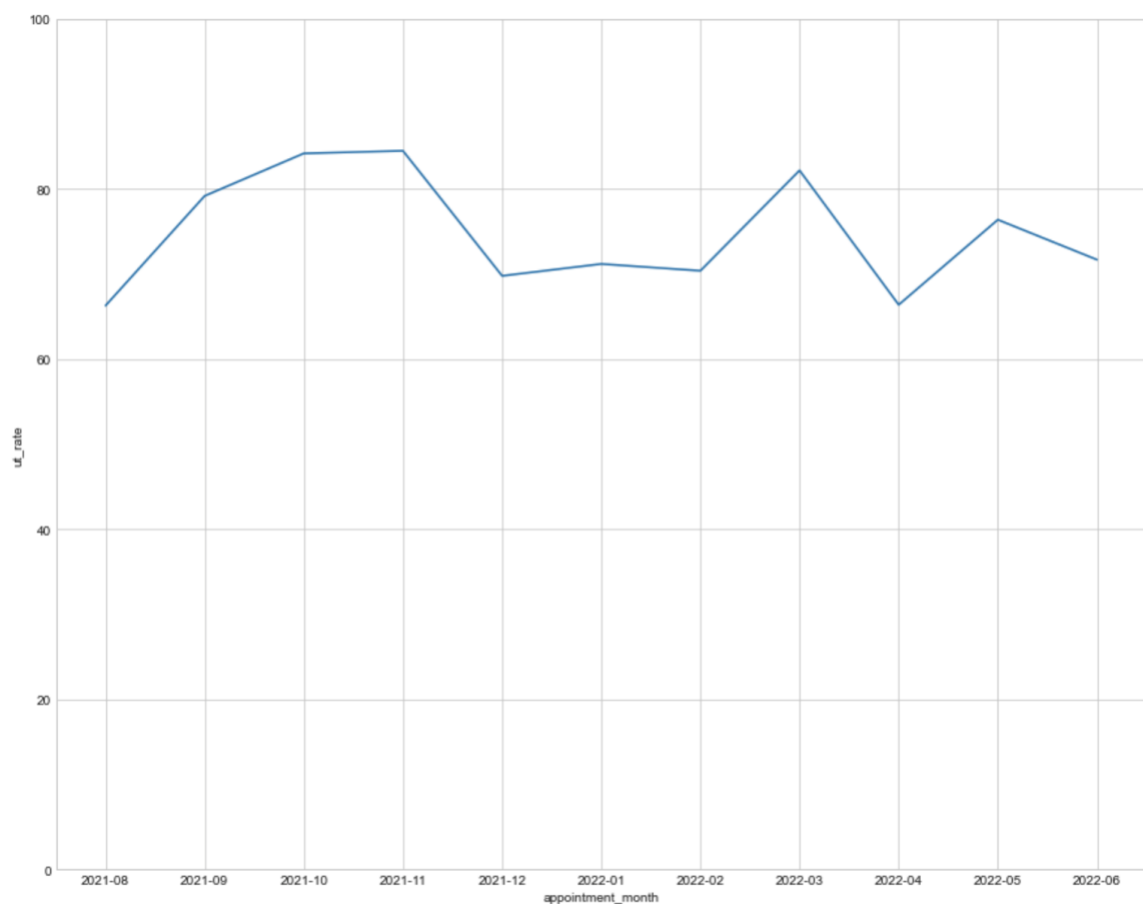

Figure 5: NHS utilization rate (%) per month.

Trends in (non)-attendance of appointments

To further dig into the main questions posed by the client, the next step is to look at the appointment status numbers. Since earlier analysis showed heavy fluctuation in overall appointment numbers, it was deemed more insightful to look at relative figures for this analysis: can we identify significant changes in the relative proportion of not-attended visits compared to the total number of visits?

Several adjustments to the provided data had to be undertaken for this:
- A new DataFrame was created to group the data by appointment month and appointment status.
- The new DataFrame was pivoted to create three columns for each of the different appointment statuses (these were identified in an earlier section).
- A new column was created to calculate the percentage of not attended visits (DNA) as part of the total number of visits ('dna_rate").
- The results of this analysis can be seen in the lineplot below (figure 6). Please note that the y-axis was reset to start at 0, to give a more realistic view of the magnitude of the fluctuation of the rate of non-attendance.
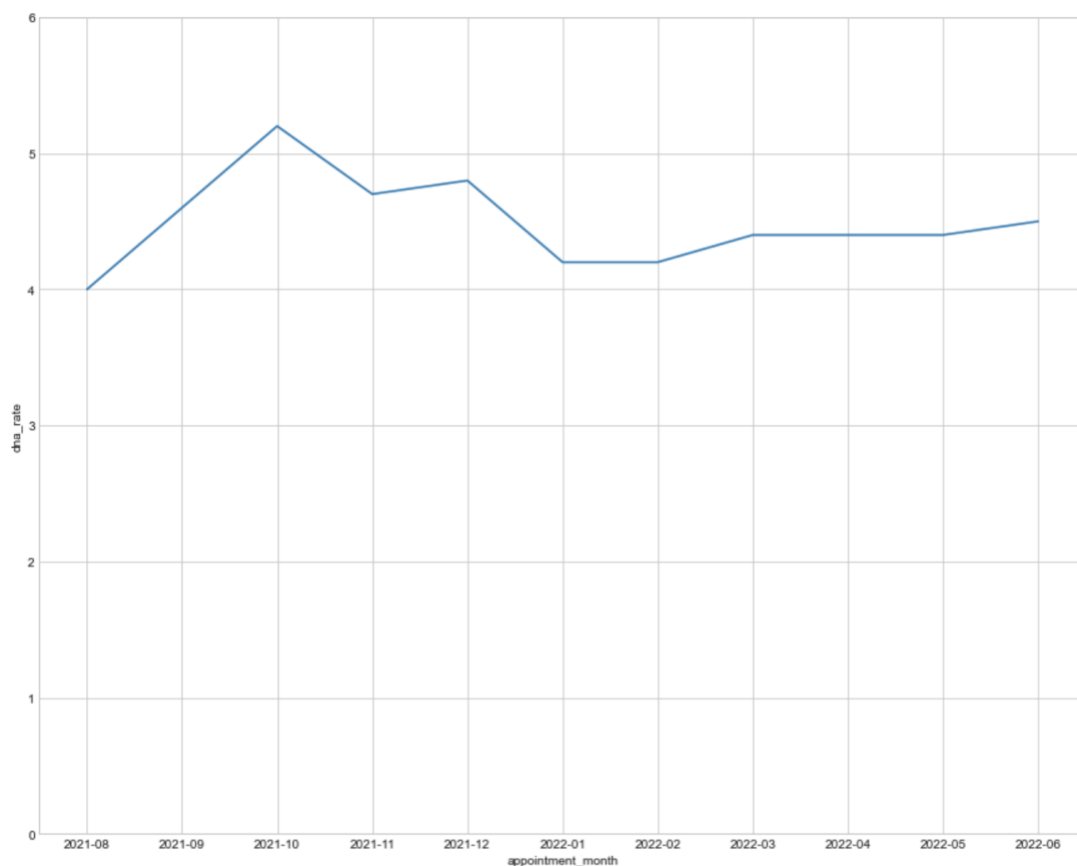

Figure 6: not attended as % of total appointments per month.

Figure 6 shows that there is a clear peak in the relative number of not attended visits, which coincides with the higher utilization numbers shown earlier. However, it is important to keep in mind that these fluctuations fall largely within a 1% boundary. Therefore, it is too early to draw definitive conclusions around the true causes behind non-attendance and further research is needed.

Conclusions and recommendations

Below the findings of the analysis will be presented, including opportunities for further research.

Findings from analysis

The aim of this study was to set the foundation to achieve the NHS's primary objective: understanding the reasons of why patients miss GP appointments.

Through this initial analysis, several key findings have been observed in the data set provided:
- There are heavy seasonal peaks in the total number of appointments scheduled with the NHS.
- Peaks in the number of non-attended visits show a similar trend as the total scheduled appointments, but the non-attended visits relative to the total appointments is mostly stable.
- Sentiment analysis through Twitter emphasizes the importance of taking the context of the global COVID19 pandemic into account. Peaks could coincide with vaccination rounds, COVID outbreaks, (end of) lockdowns, and other non-regular events.
- Utilization rates of NHS capacity fluctuate with the peaks in demand, but never reach 100% or higher. However, it is unclear whether the provided capacity figures provide an accurate picture of the situation on the ground. Taking into account the context of the global pandemic (and evidenced by several Twitter themes), it is likely that the true capacity of the NHS actually decreased during peak demand, as health workers were at increased risk of catching the COVID19 virus. This requires more research.

In any case, it is clear from the analysis that there are large swings in the demand for various NHS services across the year. While a longer time horizon is needed to fully understand the long-term trends (especially in a post-pandemic society), it is advised that the NHS approaches planning with a flexible lens: there is no one-size-fits-all approach to setting up services. Scalability to adjust to seasonal peaks is key in responding to patient demand, and if seasonal trends are found to be recurring over the years, can be planned for in advance.

Areas for further exploration

This study uncovered some of the basic characteristics of the NHS appointments, their characteristics and several trend figures over the explored time period. Further research is recommended to fully answer the NHS key objective of understanding the reasons behind not attended GP appointments:

- A more detailed breakdown into the actual capacity of the NHS (including staff availability) would help understand the true capacity figures and the strain on resources.
- Contextual data that plot the NHS data against societal events, especially related to the COVID19 pandemic, would help explore further to what extent external factors play a role in attendance of GP visits. For example, it's not hard to imagine that patients testing positive for COVID19 would cancel their face-to-face appointment last minute.
- In the long term, a patient/GP survey would greatly simplify the analysis work as well. With the large scope of the NHS, the expectation is that it would be relatively easy to gather large amounts of data in a short time.