

Technical Report

Background / Context

The National Health Service must ensure that its primary care resources are sufficient to meet public demand while avoiding unnecessary costs. Recent data indicate pressures on capacity and a persistent problem of missed general-practice appointments. These missed appointments not only affect efficiency but also represent a financial and social loss. The NHS therefore commissioned an exploratory analysis to understand utilisation trends and attendance behaviour. Using internal appointment data and public commentary from Twitter / X, the aim was to identify patterns that could inform future planning, staffing decisions, and more effective patient communication.

Analytical Approach

The analysis was conducted using Python within Jupyter Notebook, drawing primarily on three datasets supplied by the NHS: `appointments_regional.csv`, `actual_duration.csv`, and `national_categories.xlsx`. These were imported using the pandas library for data manipulation and matplotlib and seaborn for visualisation. Basic exploratory checks were carried out to confirm that the data had loaded correctly, verify the structure of each dataset, and ensure consistent column names and data types.

Some of the appointment metrics, namely context type, service setting, actual duration, and national categories, were only available from August 2021 onwards. Therefore to maintain consistency across all analyses, the datasets were filtered from this point forward. This ensured that the visualisations and comparisons were based on a uniform set of variables and avoided drawing conclusions from incomplete data.

Although merging the datasets was considered, this approach was ultimately rejected. The three files are structured at different levels of aggregation, with `appointments_regional` recorded monthly and `actual_duration` recorded daily. Combining them risked distorting patterns and introducing inconsistencies in the

results. Instead, each dataset was analysed separately, and findings were interpreted collectively where appropriate.

As the original data set was relatively consistent and free of errors, data cleaning was kept to a minimal level. This included dropping redundant columns, handling missing values, and checking for outliers. Variables such as appointment dates were converted into appropriate data types to allow grouping and time-series analysis. Markdown comments were used throughout to describe each analytical step and summarise key observations. This created a clear record of the process that another analyst could follow.

Finally, the presentation of results was organised under four main analytical lenses: Volume and Utilisation, Service Characteristics, Behavioural Insights, and External Data Exploration, in order to provide a logical flow from overall activity patterns to specific behavioural trends. This structure supported a more coherent narrative linking technical analysis to the broader business questions posed by the NHS.

Visualisation and Insights

Visualisation played an important role in helping to organise and make sense of the NHS appointment data. The data included multiple dimensions such as time, region, service type, and appointment status, so it was important to select chart types that made relationships clear without overcomplicating the message.

For the volume and utilisation analysis, simple line charts were chosen to show changes in total appointments over time. This was the most direct way to highlight broad patterns and to link them to the business question of whether overall capacity is keeping pace with demand. The addition of a capacity-utilisation chart, based on the NHS's estimated daily capacity of 1.2 million appointments, allowed the data to be viewed in relative terms and provided context for assessing whether current staffing levels appears adequate.

When examining service characteristics, a combination of line and bar charts was used to compare trends across professional types, appointment modes, and

service settings. These visualisations were selected because they make it easy to compare categories and to communicate shifts in activity to a non-technical audience. For instance, the dual-axis chart for appointment mode allowed two both total appointments and delivery method to be viewed together.

In the behavioural insights section, proportion and stacked-bar charts were used to illustrate attendance outcomes and their relationship to booking intervals. These chart types were effective for comparing relative proportions and showing how patient behaviour contributes to missed appointments. In some cases, such as the analysis of regional variation in Did Not Attend (DNA) rates, the decision was made not to include a separate visual. The data was meaningful, but presenting it verbally allowed the key message to be conveyed without overwhelming stakeholders with too many charts or details.

Finally, the Twitter dataset was explored using simple frequency plots and word counts. These were chosen to assess whether unstructured data could reveal useful external signals related to public discussion about NHS access and capacity.

Overall, each visualisation was selected for clarity, ease of understanding, and alignment with the business questions about utilisation, behaviour, and service delivery.

Patterns and Predictions

Several clear patterns emerged from the analysis. Appointment volumes varied over time, showing seasonal peaks in late autumn and early spring. These fluctuations suggest that demand for primary care changes with seasonal factors, and that capacity should continue to be monitored to help manage pressure points more effectively.

Missed appointments rose broadly in line with total appointment numbers, suggesting they are mostly volume-related rather than due to changes in behaviour. The likelihood of not attending increased when appointments were booked several weeks in advance. This indicates that longer waiting intervals contribute to patients

missing appointments, and highlights the importance of improving patient engagement and reminder systems.

There were also notable differences between regions, with some areas recording about twice the DNA rate of others. Further investigation into local characteristics may help explain these differences. Finally, improving the completeness and consistency of the available data would make future analyses more reliable and actionable.