Diagnostic Analysis using Python

National Health Service (NHS)

MIRASHA PERERA

Table of Contents

1.0   Overview

In response to the growing demand for healthcare services, it has become imperative for the National Health Service (NHS) to assess the adequacy of its current staff, resources, and capacity. This project aims to address key questions regarding resource utilisation within the NHS network and to provide actionable insights to inform decision-making processes.

Key Questions to Address:

1. What is the actual utilisation of resources within the NHS network?

2. Are the current resources sufficient to meet the demand?

3. Is there a need to increase capacity within the NHS network?

4. If current capacity is deemed sufficient, how can utilisation be optimised?

5. What are the primary reasons for missed appointments within the NHS system?

2.0   Data Analysis Limitation

Data quality issues, appointment mode discrepancies, and variability in recorded data pose challenges to analysis accuracy. Please refer to the appendix for details.

3.0   Analytical Approach

Given the complexity of the dataset and the need for thorough preparation before data exploration, the project will follow a structured approach:

3.1   Metadata Review

The project commenced with a comprehensive review of the metadata provided, aimed at understanding the dataset's structure, variable names, types, and preprocessing steps. This included thorough scrutiny of three CSV files and one XLSX file.

3.2   Data Ingestion

Before proceeding with data processing or analysis, we imported requisite libraries, including provisions to ignore filter warnings. The dataset was imported into Python as Pandas dataframes, with variable names assigned accordingly (e.g., ad_raw, ar_raw, nc_raw,tweets).

Additional contextual data, including ICB and region names from Office for National Statistics, were sourced for enhanced visualisations and exploratory data analysis (EDA). This data underwent preprocessing within a SQL database before being imported into Python via a database connection.

The SQL table was imported as a Pandas dataframe, and preliminary checks were performed by examining the first five rows to validate data integrity.

To ensure accurate data importation, I conducted sense checks using .head() function.

3.3   Data Validation

Data Validation was done using user defined functions ('Validate_data).

Duplicates Removal and Date Formatting:

- Duplicates were identified and removed, resulting in the elimination of 21,604 duplicates from the ar dataframe.
- Furthermore, the ad['appointment_date'] column was converted to datetime format from object, promoting uniformity and facilitating temporal analysis.
- Augmentation of the dataset included the addition of a day of the week column and a day index, essential for temporal analysis.

Next steps were as follows:

4.0   Exploratory Data Analysis (EDA)

EDA started with an assessment of location diversity across sub-ICBs, ICBs, and regions, leveraging functions like len() and unique() for insightful analysis. This revealed 7 regions, 42 ICB locations, and 106 sub-ICB locations. Subsequent analysis identified the top five locations with the highest appointment records, employing value_counts() and sort_values() functions for sorting.

I analysed dataset distributions to determine the dataset with the highest record count, considering varying time periods covered. Utilising a custom user-defined function (UDF) named group_and_agg, granular analysis was performed on service settings, context types, national categories, appointment statuses, HCP types, appointment modes, and durations.

Key Findings:

- 5 service settings, 3 context types, 17 national categories, and 3 appointment statuses were identified.
- Regional data spans from January 2020 to June 2022, with national categories covering August 2021 to June 2022, and actual duration data from December 2021 to June 2022.

Additionally, I conducted Twitter analysis using the provided data. However, recognizing its indirect relevance to my primary focus, I opted not to delve deeper into this aspect.

5.0   Visualising and Uncovering Trends

Setting the stage for analysis, I fine-tuned the plot aesthetics by adopting a white plot style and a colour palette aligned with the NHS brand identity.

I utilised barplots for comparisons, backed by a tailored User-Defined Function (UDF) for precision. Lineplots depicted time-series trends, supported by another crafted UDF. Swarmplots were utilised for distributional insights, offering a clear depiction of data characteristics.

5.1   Daily Trends

Upon initial exploration, Tuesdays emerged as the busiest days, although potentially influenced by public holidays. Mondays had the highest number of holidays.(Gov.UK, n.d.)

I hypothesised Mondays as the peak appointment days. Subsequent plotting validated this hypothesis, revealing Mondays as indeed recording the highest number of appointments.
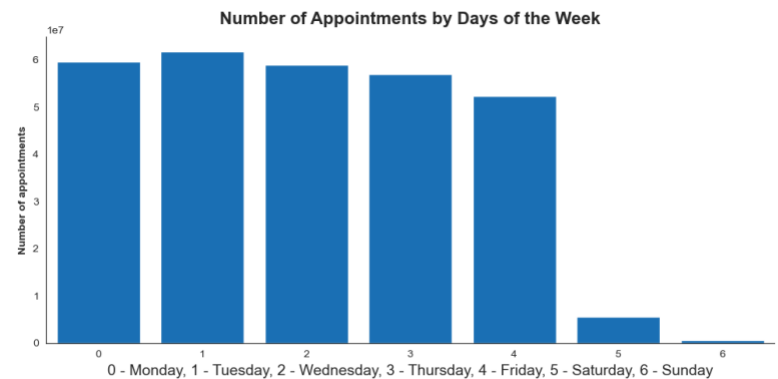

Additionally, an interesting anomaly was observed during September to November 2021, where Saturdays experienced higher-than-usual appointment volumes, accompanied by a noticeable dip in weekday appointments within the same period. This anomaly may be attributed to extended service hours during the pandemic, such as the 'Grab A Jab' weekend campaigns.
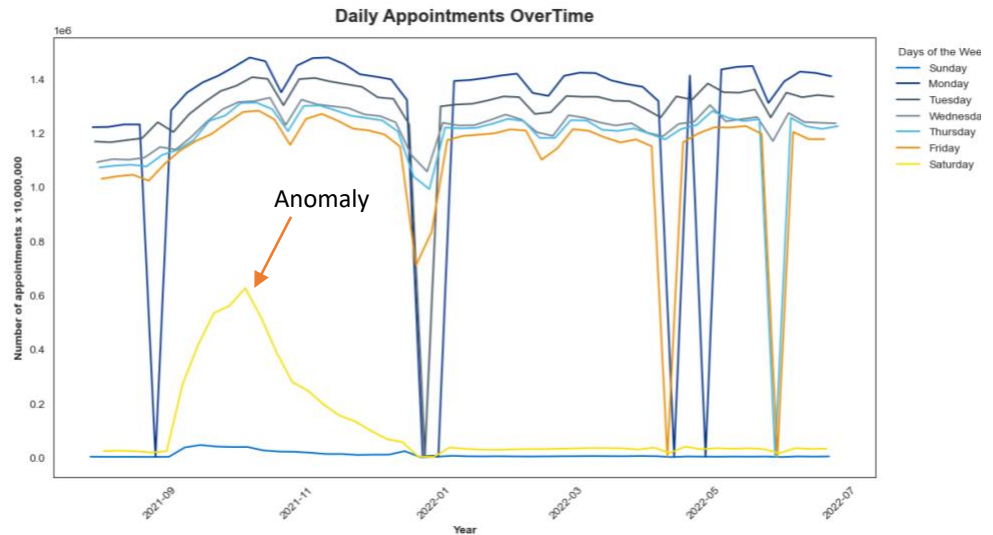
Further investigation is necessary to determine whether vaccination appointments are included in GP appointment counts, potentially impacting analysis outcomes.

5.2   Monthly Trends:

Analysis of monthly appointment trends uncovered a downward trend, possibly correlated with the COVID-19 pandemic. Notably, a 15% decline from peak appointments to June 2022 was observed, attributable to two possible factors: inclusion of vaccination appointments in total counts or the pandemic-induced increase in appointment demand. Further analysis revealed GP services as the front-runners with 87% of appointments, and routine general consultations dominating at 33%.
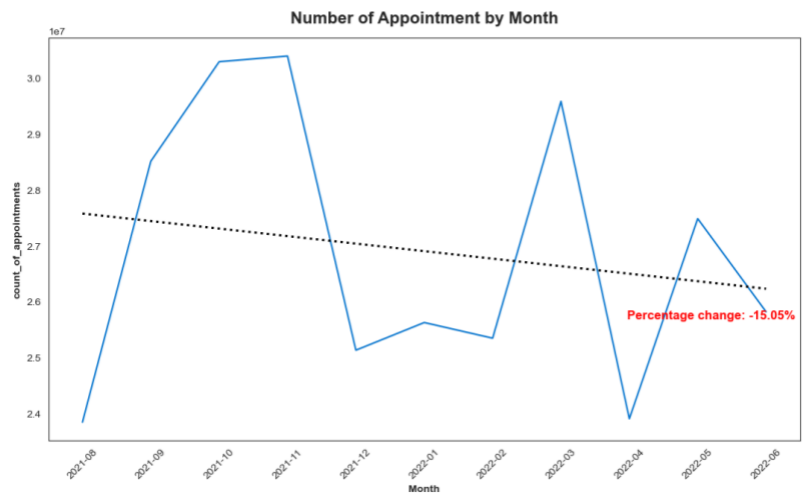

5.3   Seasonal Trends:

GP services maintained popularity across seasons, with a notable dip observed in April 2022.

5.4   Capacity Utilisation:

Referencing NHS-provided data indicating an average of 1,200,000 daily appointments for planning purposes, Monthly capacity utilisation parameter set as follows,

$$\text{Monthly capacity utilization} = \frac{\text{Total Appointments for the Month}}{\text{Monthly capacity}} \times 100$$

Monthly Capacity Utilisation
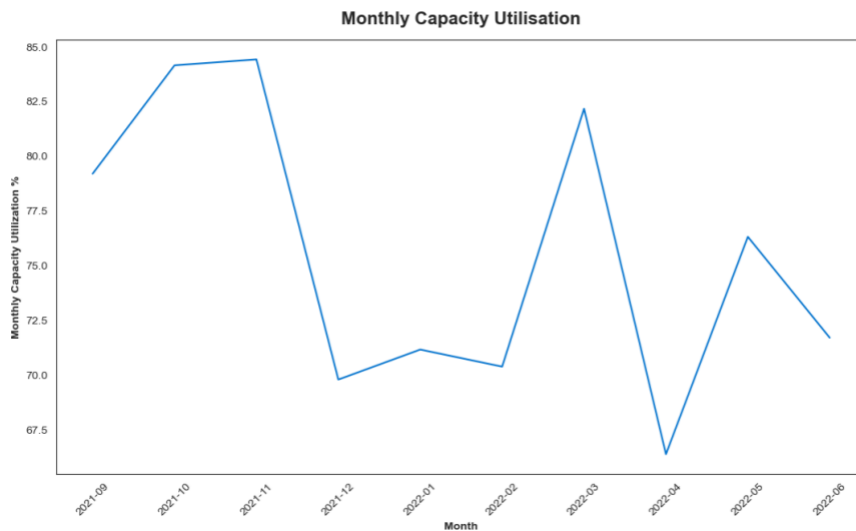
Despite a downward trend in daily utilisation, the maximum capacity reached 84% in November 2021, suggesting existing resources are under-utilised. Hence, rather than increasing staff levels, improving existing infrastructure is vital. Noteworthy is the recent monthly utilisation rate of 71%, indicating potential for efficiency improvements to utilise the capacity appropriately.

5.5   Healthcare Professional Types:
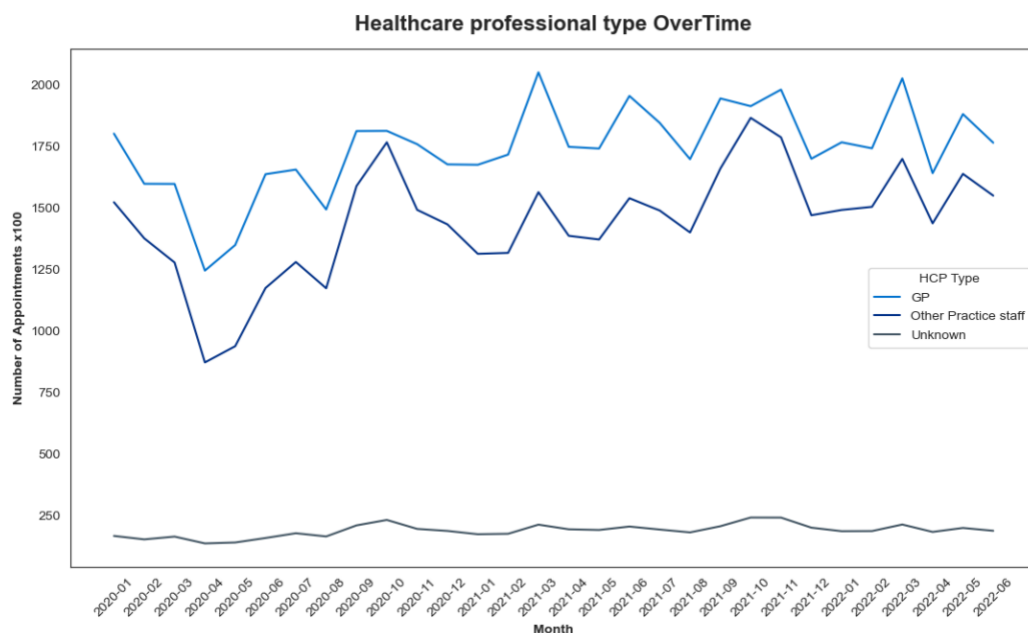
The distribution of healthcare professional types revealed GPs as the predominant category, comprising 51%, followed by Other Practice Staff at 45%, with Unknown constituting 3%. However, the broad categorization of "Other Practice Staff" limits analysis depth.

5.6   Appointment Status:

Appointment Status OverTime
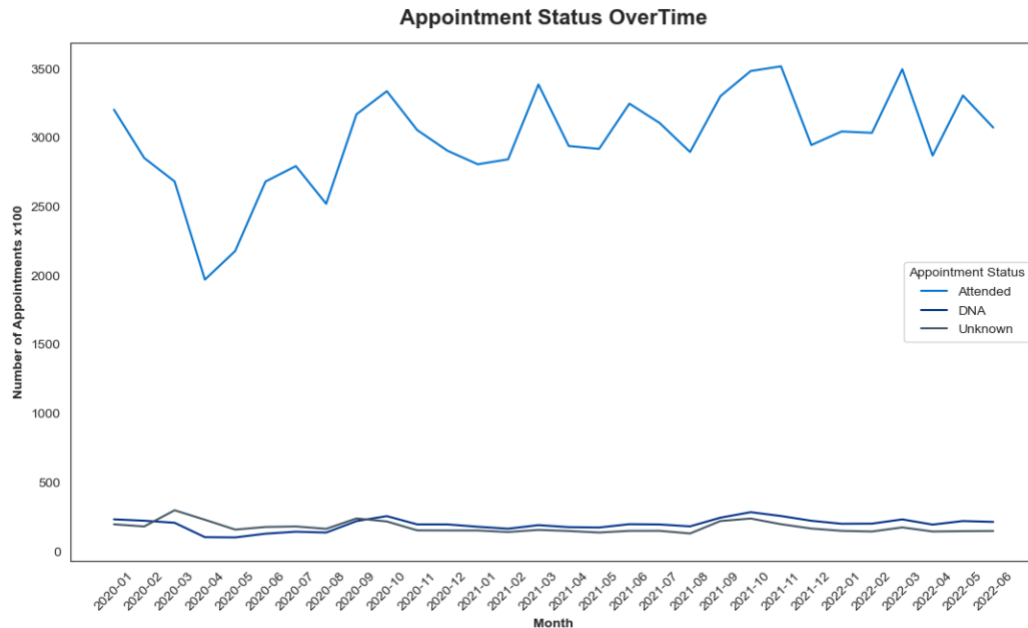
While non-attendance and unknown statuses show a declining trend, this may be attributed to the overall reduction in appointments compared to 2021. Encouragingly, most appointments are attended.

However the surveys reveal reasons for DNA (Did Not Attend) could be due to patients forgetting the appointment or never knew that the appointment was existing. (accurx, 2023)

5.7   Appointment Types:

Appointment Type OverTime
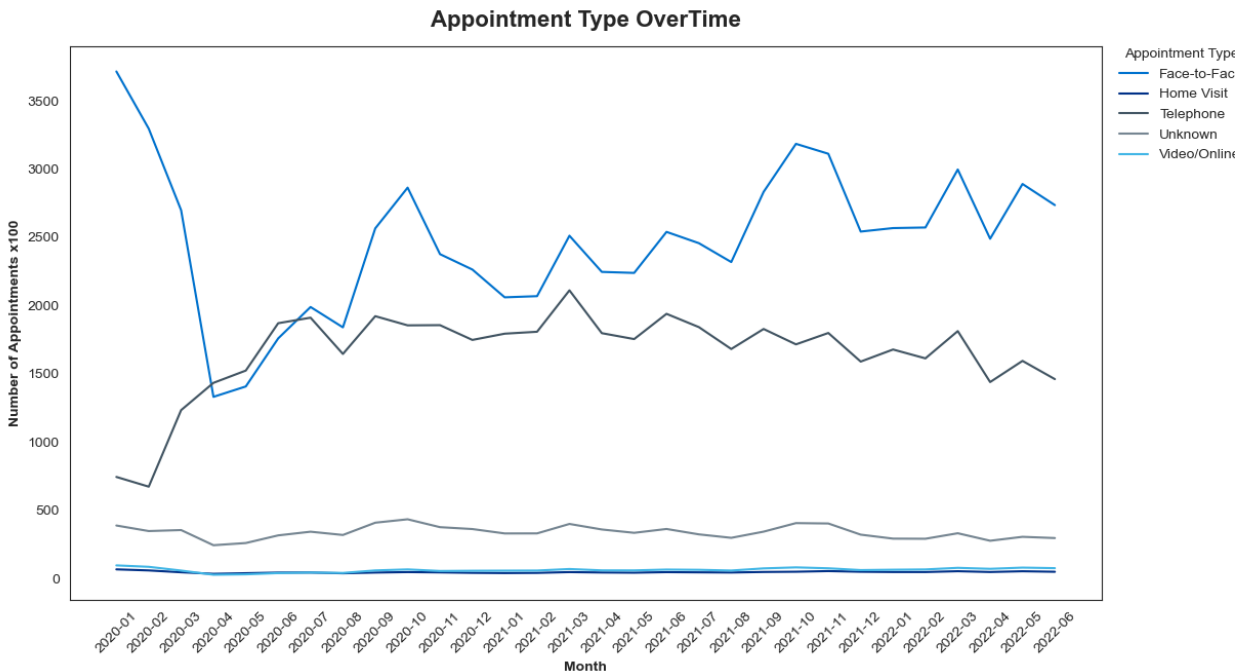
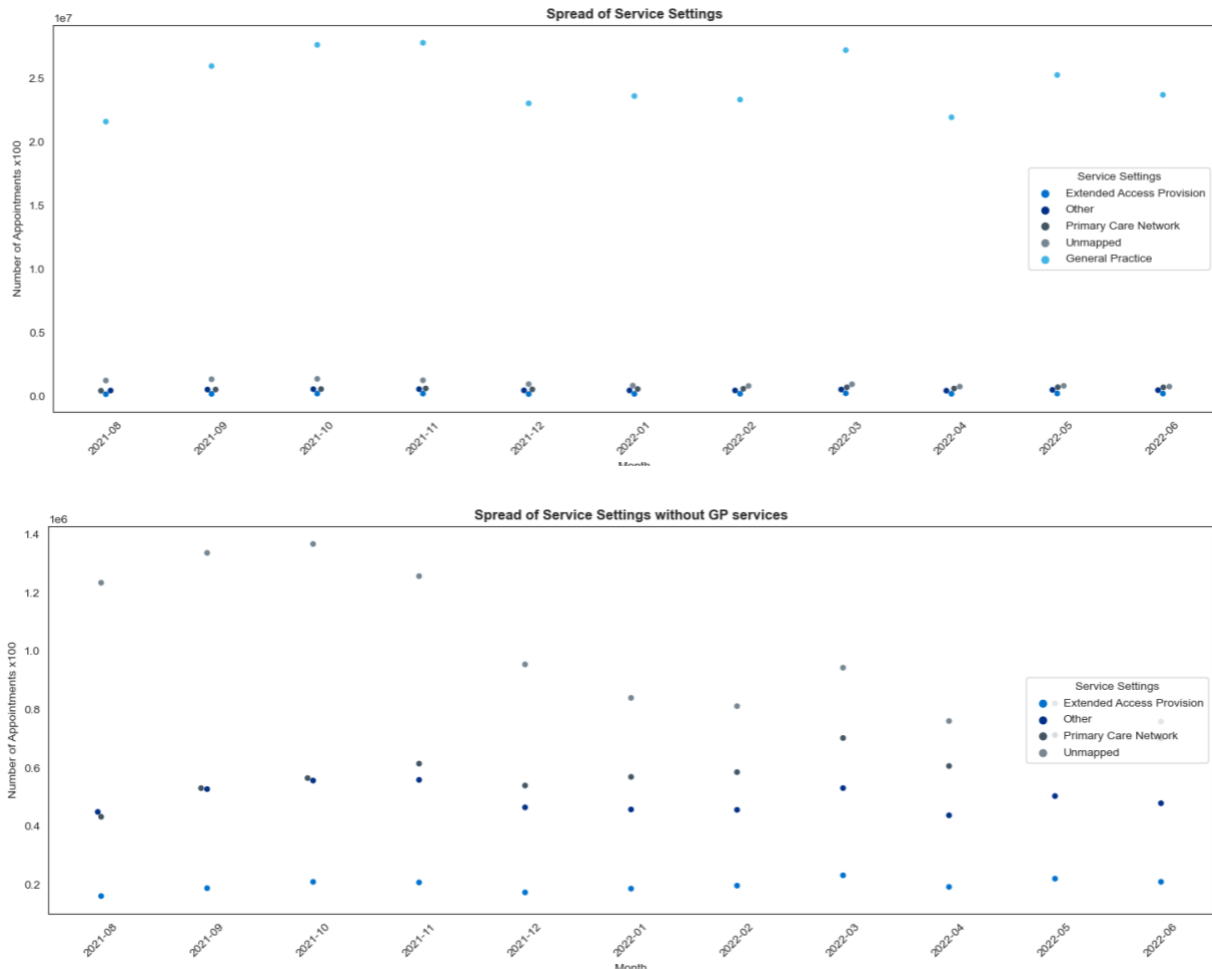Face-to-face appointments dominate, except during the beginning of the pandemic when telephone appointments surged. Post-June/July 2020, face-to-face consultations reclaimed the lead.

5.8   Booking Patterns:

An efficient booking system is evident, with most appointments made within a two-week window, and a significant proportion booked on the same day.

Time Between Book and Appointment OverTime
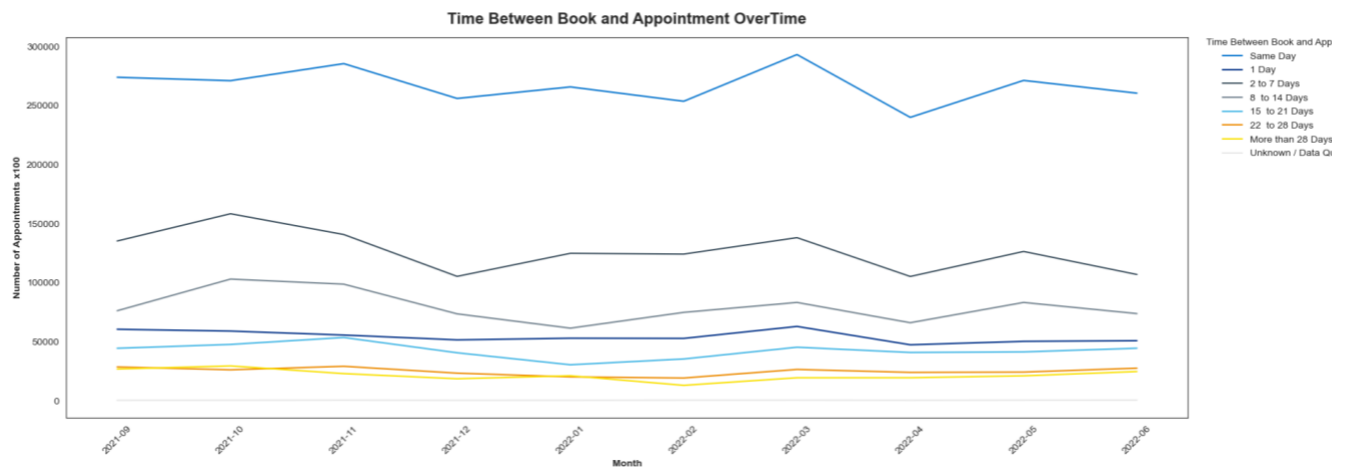
5.9   Service Setting Spread:

Spread of Service Settings

Spread of Service Settings without GP services

GP services remain consistent as the most sought-after, with fluctuations observed over time. Despite high occurrences of unmapped records, a declining trend indicates improved data quality over time.

6.0   More Insights:

- Analysis reveals that 20% of appointments took 6-10 minutes to complete, while only 5% lasted between 31-60 minutes, indicating a predominance of shorter consultations.

- Autumn and winter months experience the highest appointment volumes, with November 2021 recording 30,405,070 appointments, suggesting increased healthcare demand during colder seasons.

- Of attended appointments, 53% were face-to-face, while 3% of face-to-face appointments were not attended. Conversely, less than 1% of telephone appointments resulted in non-attendance, highlighting potential differences in patient behaviour based on appointment modality.

- DNA rates are notably higher when appointments are booked on the same day. Additionally, appointments with an unknown status is high, indicating the need for improved appointment tracking and management systems.

7.0   Recommendations:

- Allocate additional staff at the beginning of the week to manage increased appointment volumes during peak periods.
- Hire temporary staff or encourage existing staff to work overtime during autumn and winter months to address heightened demand.
- Expand categorization of "Other Practice Staff" to provide insights into which healthcare professional types are needed on specific days, facilitating more efficient resource allocation.
- Encourage the use of telephone and online appointments to reduce appointment duration and increase attendance rates.
- Implement a system where GPs can accurately record appointment start and end times, facilitating more precise analysis of appointment duration and resource utilisation.

8.0   Appendix

8.1   Data Analysis Limitations.

Data Quality:

The collected data is primarily gathered for clinical use rather than data analysis purpose therefore there's no data entry standard followed which will lead into data quality variations across practices.

Appointment mode:

The appointment mode is decided by the practices. Below data quality issues may affect the analysis.

- Video appointments recorded face-to-face.
- Many Telephone triage and home visits booked as a one slot.
- Home visits not logged in the system but in the reception.
- Practices using the Cegedim GP system are unable to supply appointment mode data.

Actual Data Variability:

The recording of the actual duration of the consultation varies depending on the practises system supplier leading to inconsistency of the data quality. Further appointment with null or extreme duration recorded as 'Unknown/Data Quality Issue', potentially affecting the analysis of the appointment lengths.

HCP Classification:

Other practice staff. Includes very broader professionals therefore unable to provide detailed analysis.

Number of appointments:

Not all practices in England are included in this release, meaning the total number of appointments is not known.

9.0   Works Cited

accurx, 2023. *Beyond forgetfulness: Why do so many people miss their hospital appointments?.* [Online]
Available at: https://www.accurx.com/post/beyond-forgetfulness-why-do-so-many-people-miss-their-hospital-appointments
Gov.UK, n.d. *UK bank holidays.* [Online]
Available at: https://www.gov.uk/bank-holidays

Office for National Statistics. "Integrated Care Boards (July 2022) Names and Codes in England."

Office for National Statistics,

https://geoportal.statistics.gov.uk/documents/25ba241a775e4a9db8e5c721ee73d85d/about.

Accessed 19 May 2024.