# **NHS Data Analysis**

# 1. Introduction

The NHS faces challenges in optimising healthcare capacity, reducing missed appointments, and understanding resource utilisation to enhance service efficiency. This report presents a comprehensive data-driven analysis of NHS service usage, missed appointments, and public sentiment to aid in efficient healthcare delivery. The datasets analysed include regional appointment data, actual service durations, service categories, and social media insights from Twitter. This report is aimed at a technical audience and discusses the methodologies, findings, and validation of each aspect of the analysis.

## 1.1 Problem Statement

The NHS must expand its infrastructure and resources to meet the growing demand driven by an increasing population. Accurate budget allotment and resource planning are essential, requiring a thorough understanding of service utilisation trends. Stakeholders are divided on whether current NHS capacity is sufficient or needs expansion, and whether existing infrastructure could be optimised to improve efficiency. Additionally, the NHS incurs significant costs due to missed appointments, which need to be understood and minimised.

The key questions to be addressed include:

* Has there been adequate staffing and capacity in the networks?
* What was the actual utilisation of resources?
* What is the general perception of people based on tweets?

The goal of this analysis is to provide data-informed insights that can assist the NHS in making informed decisions about resource allocation, capacity planning, and minimising missed appointments.

## 2. Data Preparation and Cleaning

The initial phase of the analysis focused on data preparation and cleaning to ensure accuracy, consistency, and completeness of data. Four main datasets were used:

- Actual Duration: Captures time durations for healthcare activities to assess resource utilisation and identify inefficiencies.

- Appointments Regional: Contains appointment-related data across regions, including appointment counts, statuses, and modes. This dataset is critical for analysing temporal trends and understanding regional differences in healthcare usage.

- National Categories: Provides categorical information that helps analyse national-level healthcare data, such as resource allocation and service segmentation.

- Tweets: Includes public tweets related to NHS services, used to gauge public sentiment and understand public perception of healthcare services.

### 2.1 Data Cleaning Process

Data cleaning is an essential step to ensure the accuracy and reliability of any data analysis. In this project, we followed a systematic approach to cleaning the datasets:

1. **Handling missing values:** Missing values were filled with "Unknown" or imputed based on logical assumptions to ensure uniformity across the datasets. For instance, missing appointment statuses were assigned "Unknown" if no information was available. Where appropriate, statistical imputation methods such as mean, median, or mode were used to fill numerical fields to minimise data bias and maintain data integrity.

2. **Date conversion:** Date columns were converted to datetime format with error handling, ensuring consistency in date data. This conversion was crucial for accurate temporal trend analysis, allowing for the efficient grouping of data by time periods, such as months, quarters, or seasons. Additionally, date features such as month, day of the week, and year were extracted to provide more granular insights during the analysis.

3. **Data standardisation and normalisation:** Standardisation of categorical values was performed to ensure consistency. For example, appointment modes such as "Virtual" and "Online" were standardised to "Virtual" to maintain a consistent categorisation. Additionally, numerical features were normalised where required to facilitate model convergence and improve the interpretability of analysis involving machine learning models.

The code below demonstrates the data cleaning process:

# Fill missing values and convert date columns to datetime format  
datasets = {  
 "df\_appointments\_regional": df\_appointments\_regional,  
 "df\_actual\_duration": df\_actual\_duration,  
 "df\_national\_categories": df\_national\_categories,  
 "df\_tweets": df\_tweets  
}  
  
# Fill missing values with 'Unknown'  
for name, dataset in datasets.items():  
 dataset.fillna('Unknown', inplace=True)  
  
# Convert date columns to datetime format  
df\_appointments\_regional['appointment\_month'] = pd.to\_datetime(df\_appointments\_regional['appointment\_month'], errors='coerce')  
df\_actual\_duration['appointment\_date'] = pd.to\_datetime(df\_actual\_duration['appointment\_date'], errors='coerce')

# 3. Temporal Trends Analysis

The analysis aimed to identify trends in healthcare resource utilisation over time.

* **Monthly and seasonal trends**: By analysing appointment modes and statuses across different months and seasons, we identified periods of increased and decreased demand. This analysis showed that winter months generally had higher appointment volumes, likely due to seasonal illnesses.

**Insight**: Identifying these trends helps in adjusting resource allocation and staffing levels to meet peak demands effectively.

# 4. Resource Utilisation Analysis

The resource utilisation analysis provides insights into how efficiently NHS resources are being used.

* **STL Decomposition**: Seasonal-Trend decomposition was performed on the monthly count of appointments to understand the observed, trend, seasonal, and residual components. The decomposition revealed clear seasonal peaks, typically during the winter months, indicating a consistent need for increased resources during high-demand periods.

The code below demonstrates the STL process:

stl = STL(monthly\_appointments['count\_of\_appointments'], seasonal=13)  
stl\_result = stl.fit()

**Insight**: The analysis indicates consistent seasonal variations in the number of appointments. This insight can guide proactive resource allocation during high-demand seasons.

* **SARIMAX Forecast**: SARIMAX was used to forecast future appointment counts for the next 12 months. The results showed an anticipated increase in appointments during the coming winter months.

The code below demonstrates the SARIMAX forecast:

sarimax\_model = sm.tsa.SARIMAX(monthly\_appointments['count\_of\_appointments'], order=(1, 1, 1))  
sarimax\_result = sarimax\_model.fit()

**Insight**: The SARIMAX model forecast indicates that healthcare services will likely experience increased demand during winter, emphasising the importance of early planning to ensure adequate capacity.

* **Cross analysis of appointment mode and status**: A cross-analysis of appointment mode versus status was performed. In-person appointments showed a higher likelihood of being missed compared to virtual appointments.

cross\_analysis = df\_appointments\_regional.groupby(['appointment\_mode', 'appointment\_status'])['count\_of\_appointments'].sum().reset\_index()

**Insight**: Optimising resource allocation towards appointment modes with higher completion rates can enhance service efficiency and reduce the number of missed appointments.

* **Planned vs. Actual capacity**: Analysis of planned capacity versus actual utilisation rates highlighted discrepancies between planning and real-world demand.

df\_appointments\_regional['utilisation\_rate'] = (df\_appointments\_regional['count\_of\_appointments'] / df\_appointments\_regional['planned\_capacity']) \* 100  
avg\_utilisation = df\_appointments\_regional['utilisation\_rate'].mean()

**Insight**: Addressing gaps in planned versus actual utilisation will help in optimising resources and reducing inefficiencies.

**Conclusion**: The resource utilisation analysis highlighted the importance of understanding seasonal trends and optimising appointment modes to enhance service delivery efficiency. By planning resources based on these insights, NHS can improve capacity utilisation and reduce missed opportunities.

# 5. Missed Appointment Analysis

Missed appointments, known as Did Not Attend (DNA), have a significant impact on healthcare efficiency.

* **Total missed appointments**: The monthly trend of missed appointments was analysed using a line plot. Peaks in missed appointments often coincide with specific months, suggesting potential external factors like holidays.

dna\_df = df\_appointments\_regional[df\_appointments\_regional['appointment\_status'] == 'DNA']  
monthly\_dna\_counts = dna\_df.groupby(dna\_df['appointment\_month'].dt.to\_period('M')).size().reset\_index(name='Number of Missed Appointments')

**Insight**: Identifying peaks in missed appointments allows for targeted interventions during high-DNA periods, such as improved communication and reminders.

* **Service aetting analysis**: Missed appointments were broken down by service setting, revealing that certain appointment modes (e.g., face-to-face) had higher DNA rates than others (e.g., virtual).
* **Healthcare professional type analysis**: A pie chart was used to determine missed appointments by healthcare professional type, providing insights into which types are more prone to missed appointments.

**Insight**: Understanding which service settings and healthcare professional types have higher DNA rates helps in designing targeted strategies to reduce missed appointments.

**Conclusion**: The missed appointments analysis helps target interventions in high-DNA areas, such as optimising communication channels, using reminders, and focusing on virtual appointments where DNA rates are lower. This can lead to a significant reduction in inefficiencies and improve patient care.

# 6. Public Sentiment Analysis

Twitter data was analysed to understand public sentiment towards NHS services.

* **Sentiment analysis**: TextBlob was used to classify tweets as positive, negative, or neutral. The majority of tweets were neutral, with only a small fraction being positive or negative.

df\_tweets['sentiment'] = df\_tweets['tweet\_full\_text'].apply(lambda x: 'positive' if 'good' in x.lower() else ('negative' if 'bad' in x.lower() else 'neutral'))

**Insight**: Monitoring public sentiment on social media can help NHS understand patient concerns and address areas of dissatisfaction. Understanding the breakdown of sentiments also helps prioritise which issues need immediate attention.

**Conclusion**: Regular monitoring of public sentiment on social media will help the NHS identify areas for improvement and enhance patient satisfaction. By actively addressing negative sentiments, NHS can improve its public image and relationship with patients.

# 7. Summary and Recommendations

* **Resource utilisation**: The NHS should allocate resources based on the identified seasonal trends, with additional capacity during winter months. This proactive planning will help in meeting peak demands efficiently.
* **Missed appointments**: Reducing missed appointments should be prioritised, especially for in-person and high-DNA service settings. Strategies such as improved appointment reminders and encouraging virtual appointments where appropriate can help reduce DNA rates.
* **Capacity planning**: Aligning planned capacity with actual demand can enhance service delivery efficiency and reduce waste. Better forecasting and resource allocation based on data-driven insights will ensure effective service provision.
* **Public sentiment**: Regular monitoring of public sentiment on social media will help improve public relations and patient satisfaction. Addressing patient concerns in a timely manner can lead to better healthcare experiences.

# 8. Conclusion

This report presented a detailed analysis of NHS data focusing on resource utilisation, missed appointments, temporal trends, and public sentiment. The insights provided can guide the NHS in improving healthcare service efficiency, planning resources more effectively, and understanding patient needs.

By utilising data-driven approaches, NHS can better align healthcare capacity with patient demand, reduce inefficiencies, and enhance overall healthcare quality.