Elizabeth Liang – DA201 Assignment Report

Analysis of Data from the NHS

---

1. Background / Context

The National Health Service (NHS) faces increasing pressure to balance the needs of a growing population with the capacity of its services. To optimize resource use, this analysis examines appointment data and supplementary information from Twitter to identify trends in service utilization, capacity, and missed appointments—an issue that significantly drives avoidable costs. The findings aim to provide evidence-based recommendations for improving resource allocation, reducing missed appointments, and informing policy decisions.

---

2. Analytical Approach

Data Ingestion and Tools

Using Python in a Jupyter Notebook, four datasets were ingested:

- **appointments_regional.csv** – regional/monthly appointment data.
- **actual_duration.csv** – appointment-level details (dates, regions, durations).
- **national_categories.xlsx** – service setting, context type, and category data.
- **tweets.csv** – UK healthcare-related Twitter data.

Additionally, Metadata in *metadata_nhs.txt* was used for reference.

Libraries used:

- **Pandas** for cleaning, wrangling, and aggregation.
- **Seaborn & Matplotlib** for visualisations.

```python
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

Figure 1: *Importing Python Libraries for Data Analysis*

Data Wrangling

Structural Review:

- `df.info()` to look at the names of the columns and the types of data they hold.
- `df.describe()` to generate descriptive statistics.
- `df.shape` to check the number of rows and columns.
- `df.head()` to preview the first few records.
- `df.duplicated().sum()` to find duplicate records and `drop_duplicates()` to eliminate them.
- `isnull().sum()` to find missing values.

Date Formatting: Appointment dates were standardised using `.dt.strftime('%Y-%m')` to support consistent monthly aggregation.

Grouping and Aggregation: The `groupby()` function was applied to calculate the total number of appointments per month and per category. Seasonal subsets (Summer, Autumn, Winter, and Spring) were created using date filters.

Capacity Benchmarking: A benchmark column (`budgeted_utilisation`) was added to compare actual utilisation against the NHS planning figure of 1.2 million daily appointments. Monthly totals were divided by 30 to produce daily utilisation estimates.

Twitter Hashtag Analysis: Tweets were analysed by extracting hashtags with `value_counts()`. Overrepresented terms such as **#healthcare** were excluded to prevent skewing, and analysis focused on hashtags appearing more than 10 times to identify meaningful trends.

---

3. Visualisation and Insights

Seasonal and Weekly Trends

Granular seasonal analysis revealed clear weekly rhythms and seasonal peaks:

- **Summer (August):**
  Over 90% of appointments remained below NHS daily capacity during August. Mondays were busiest, with volumes falling ~10% by Wednesday and ~5% toward Friday. Very low activity occurred on Saturdays; none on Sundays or bank holidays.
- **Autumn:**
  Higher volumes than summer. Early-week appointments often surpassed NHS daily capacity. Mondays peaked, with a marked increase in Saturday appointments. Aggregated totals exceeded NHS thresholds.
- **Winter (January):**
  Consistent high demand across seasons, with Mondays consistently breaching NHS thresholds. January exemplified the strain of seasonal illness.

- **Spring:**
  Slightly lower overall volumes, but the same weekly rhythm. Totals exceeded NHS capacity by ~10% on average.

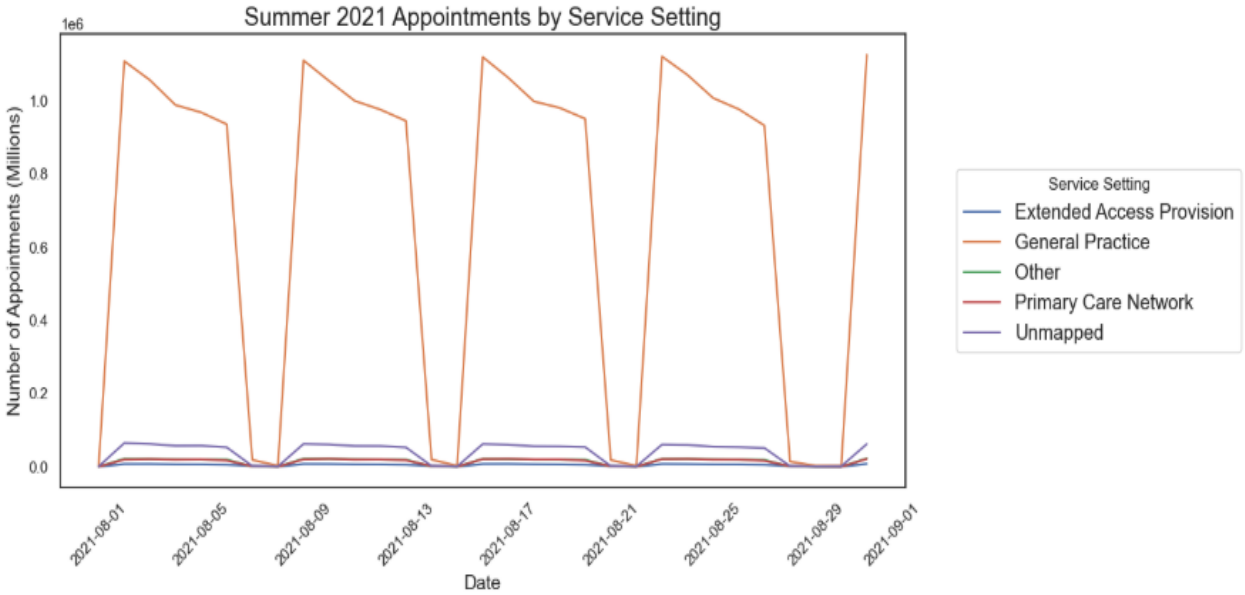

Figure 2: *Weekly Appointment Trends in Summer.*

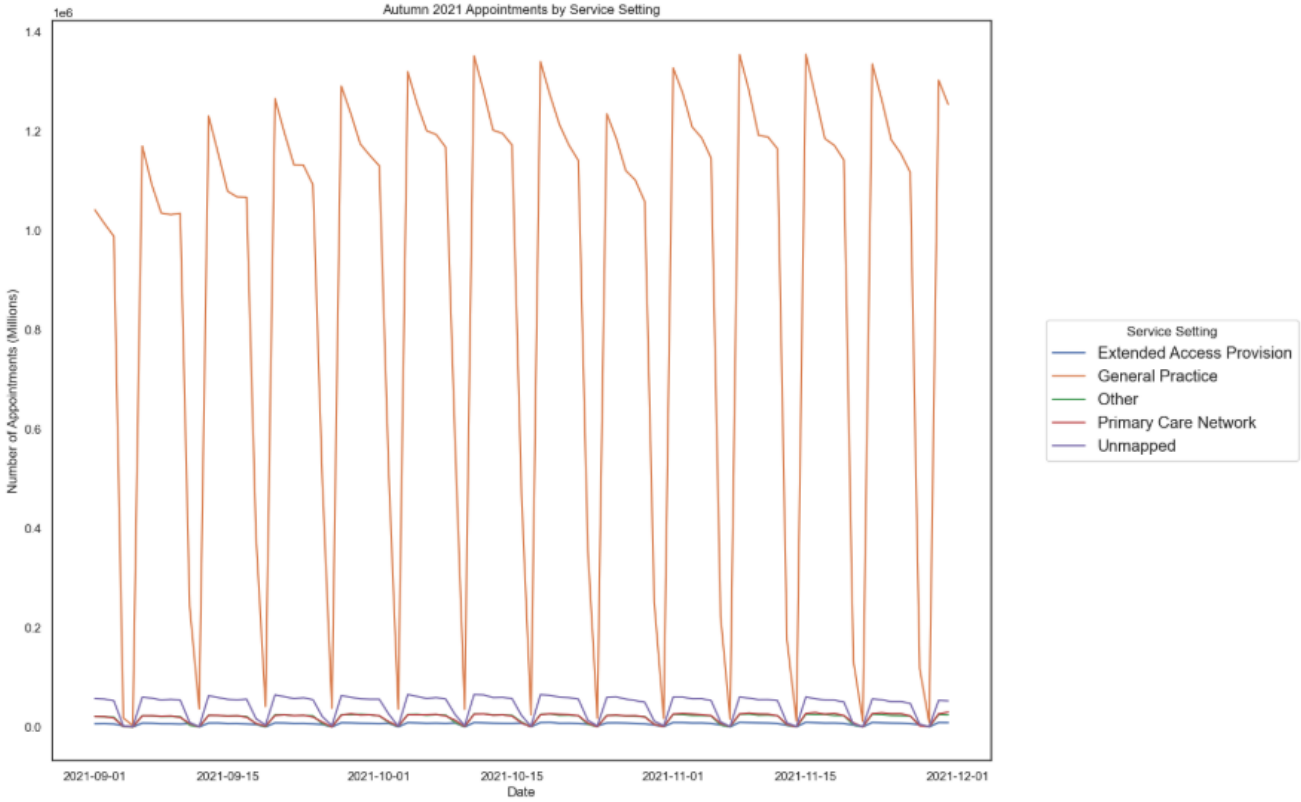

Figure 3: *Weekly Appointment Trends in Autumn.*

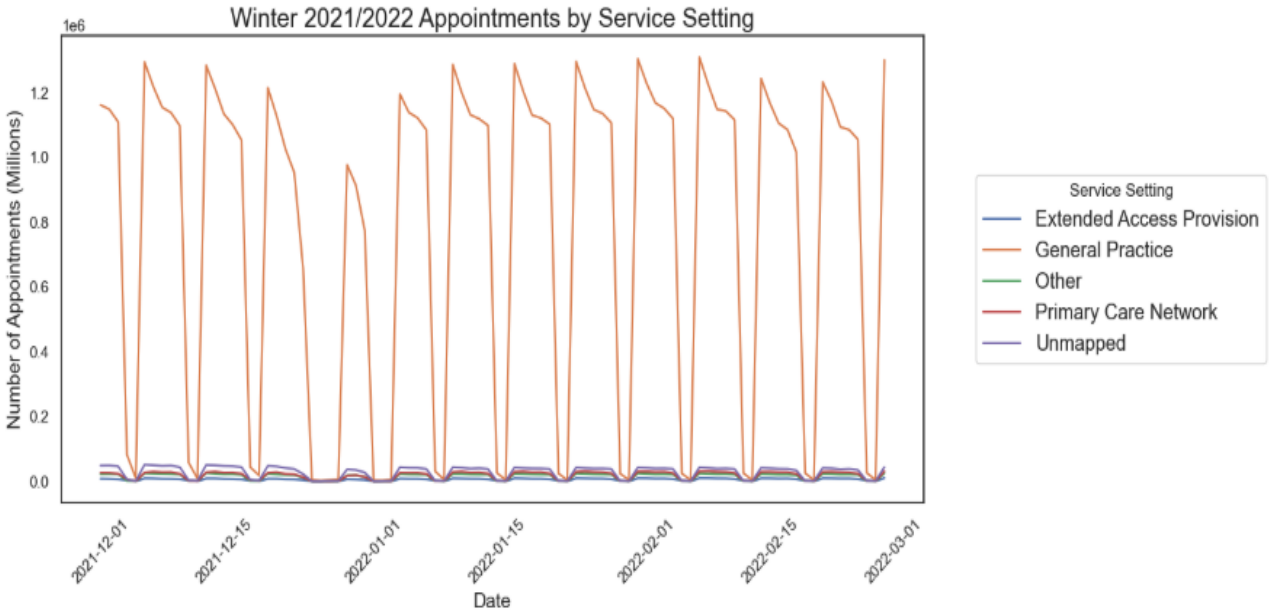

Figure 4: *Weekly Appointment Trends in Winter.*

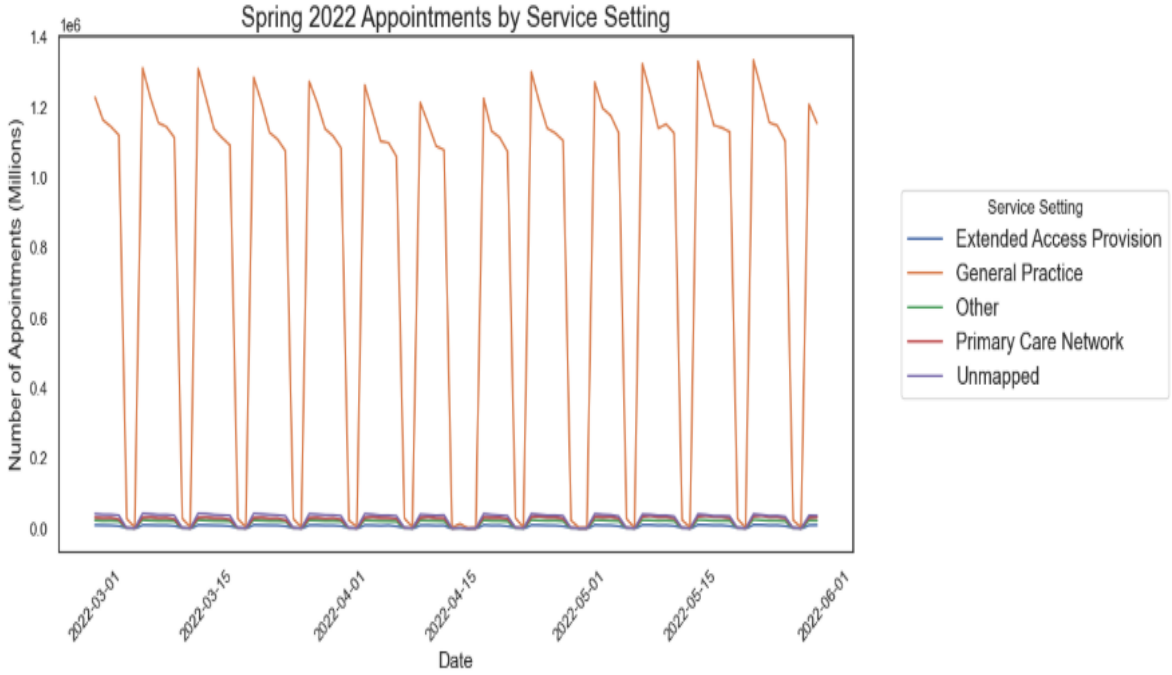

Figure 5: *Weekly Appointment Trends in Spring.*

Key Finding: Mondays are consistently the busiest, with autumn and winter driving demand above NHS capacity. This underscores the need for flexible staffing and resource planning.

Service Settings

- **Telehealth adoption** shows growth, suggesting post-COVID care shifts.

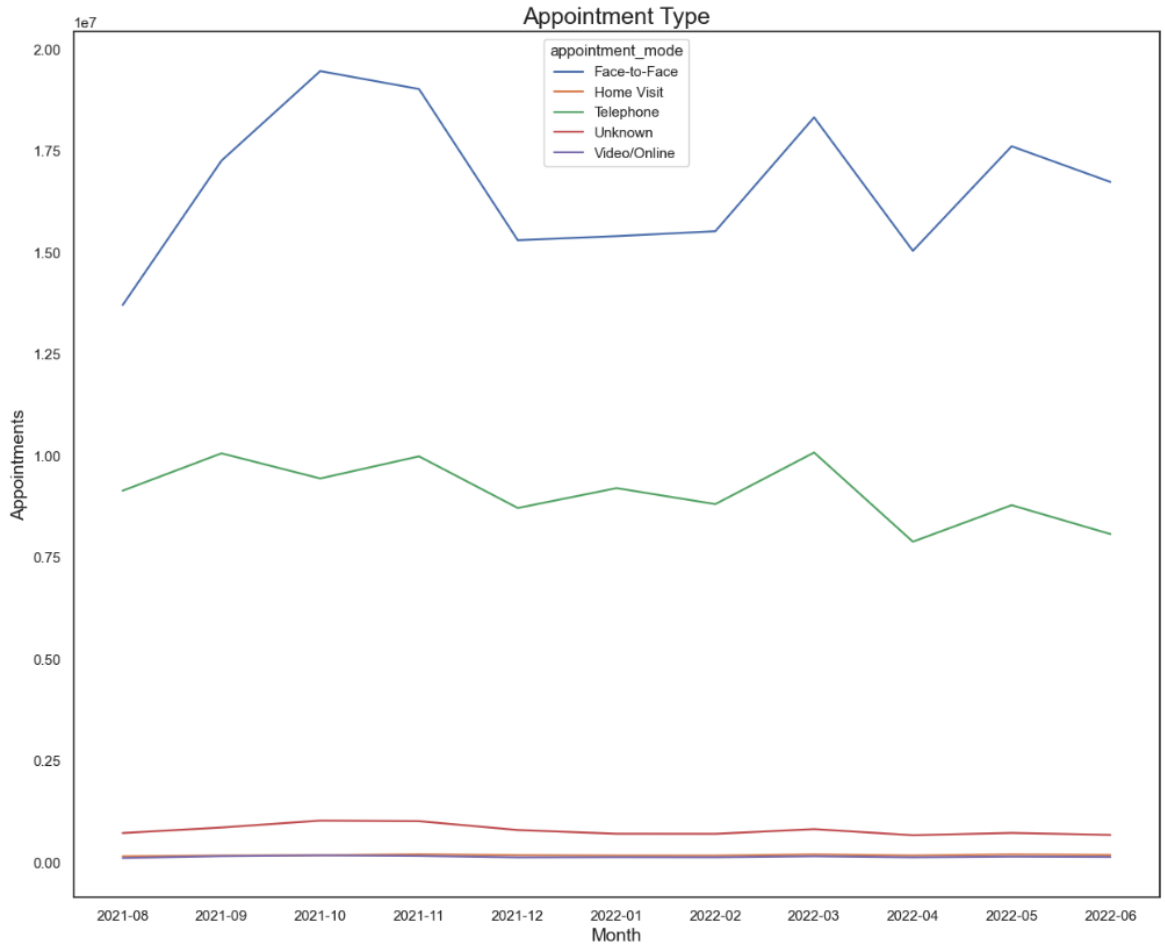

Figure 6: *Trends in Appointment Types Over Time.*

- Excluding **General Practice**, the most frequently occurring category is **'Unmapped'**, which indicates inconsistencies in how the data has been classified. This suggests that the NHS would benefit from improving its data management processes, ensuring that appointments are accurately mapped to the correct service settings. Enhancing data quality in this way would provide more reliable insights and support better-informed decision-making.

- Removing GP data via boxplots highlighted variability in specialist services, suggesting uneven demand distribution.

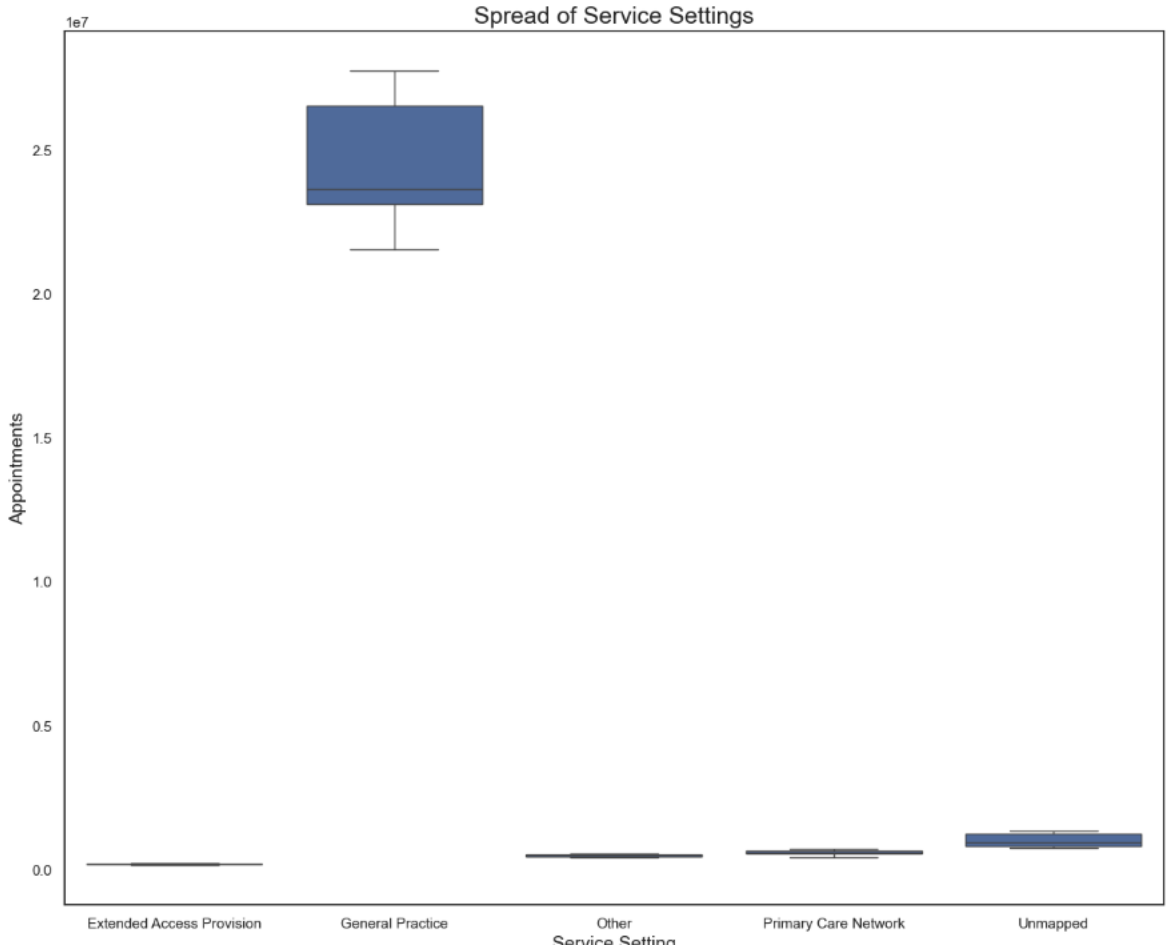

Figure 7: *Distribution of Appointments Across Service Settings (General Practice Leads Significantly)*

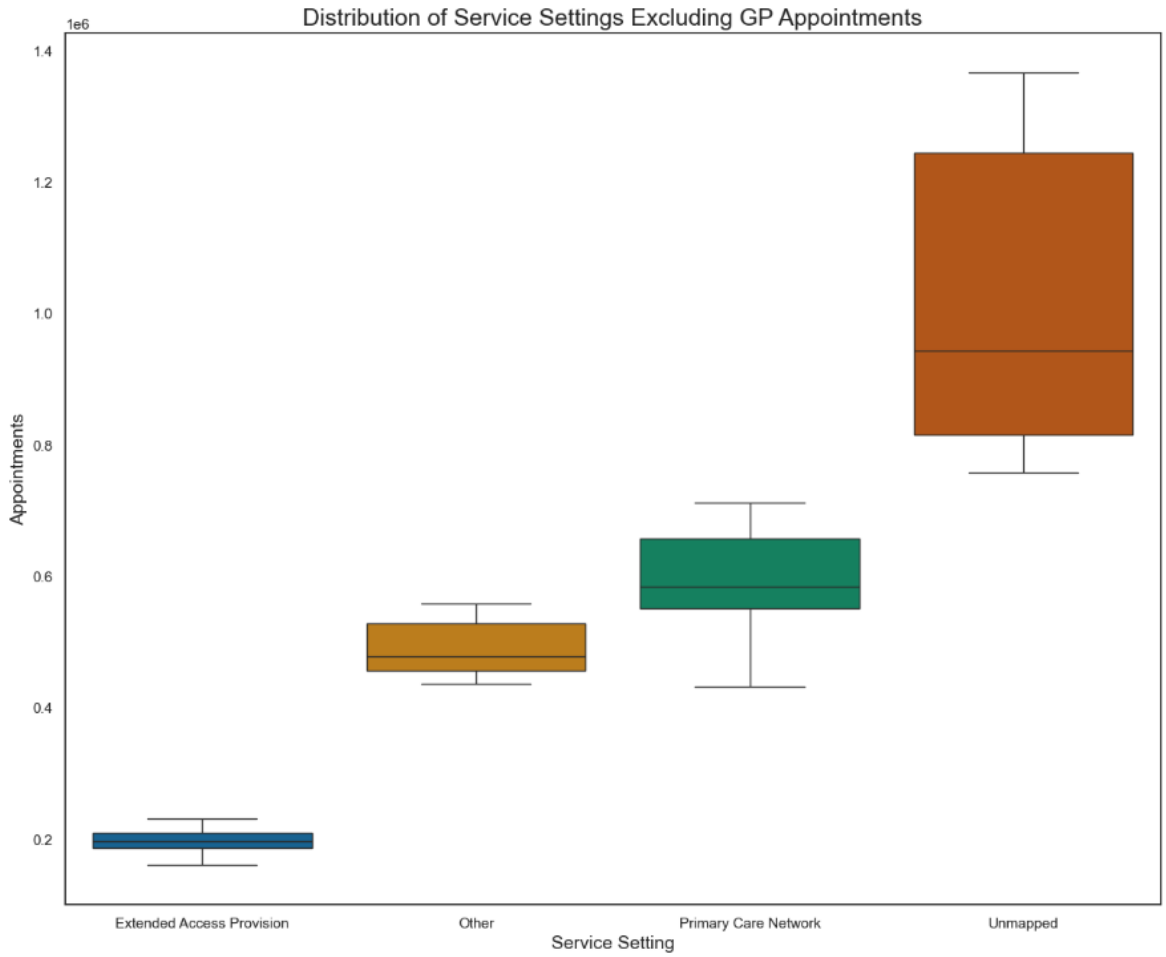

Figure 8: *Boxplot of service settings without GP appointments*

Capacity Utilisation vs. Benchmark

An analysis of capacity utilisation across the observed period highlights significant fluctuations, ranging from **65% to 85%**. Despite these variations, average utilisation consistently remained below the NHS's benchmark daily capacity of **1.2 million appointments**.

These findings suggest that, on a **monthly average level**, the NHS's existing capacity was sufficient to meet demand for appointments. However, several months exhibited clear **underutilisation**, pointing to potential opportunities for **resource reallocation** to enhance efficiency and balance workloads across the system.

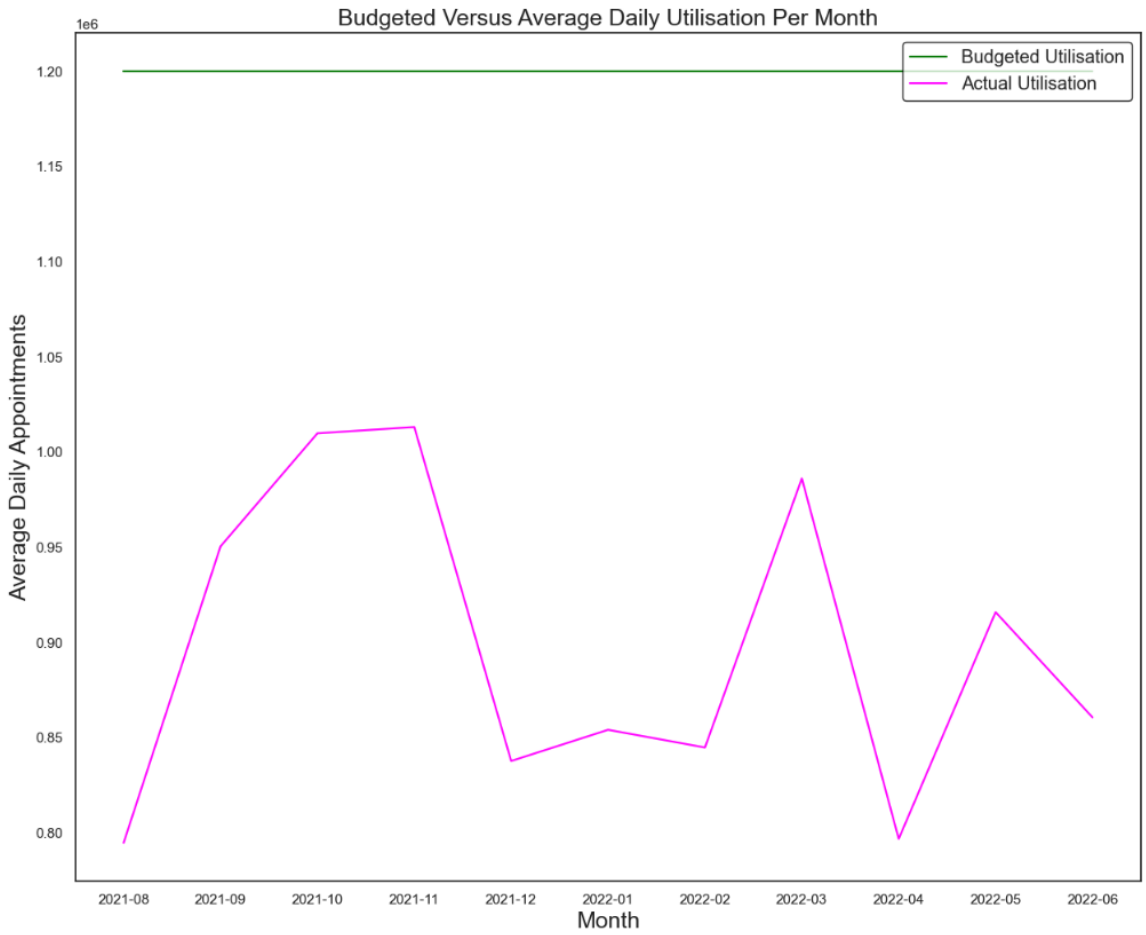

Figure 9: *Line plot comparing budgeted vs. actual utilisation.*

Twitter Insights

After filtering out #healthcare (which dominated results), themes emerged:

- **Digital Health & Innovation:** Strong interest in #ai, #digitalhealth, and #machinelearning.
- **Workforce Issues:** Hashtags like #job, #hiring, #womeninmedicine highlight recruitment and diversity concerns.
- **Clinical Focus:** Terms like #medicine and #pharma reflect professional and public discourse.
- **Public Health Response:** #covid and #vaccine remain visible, showing continued interest in health resilience.

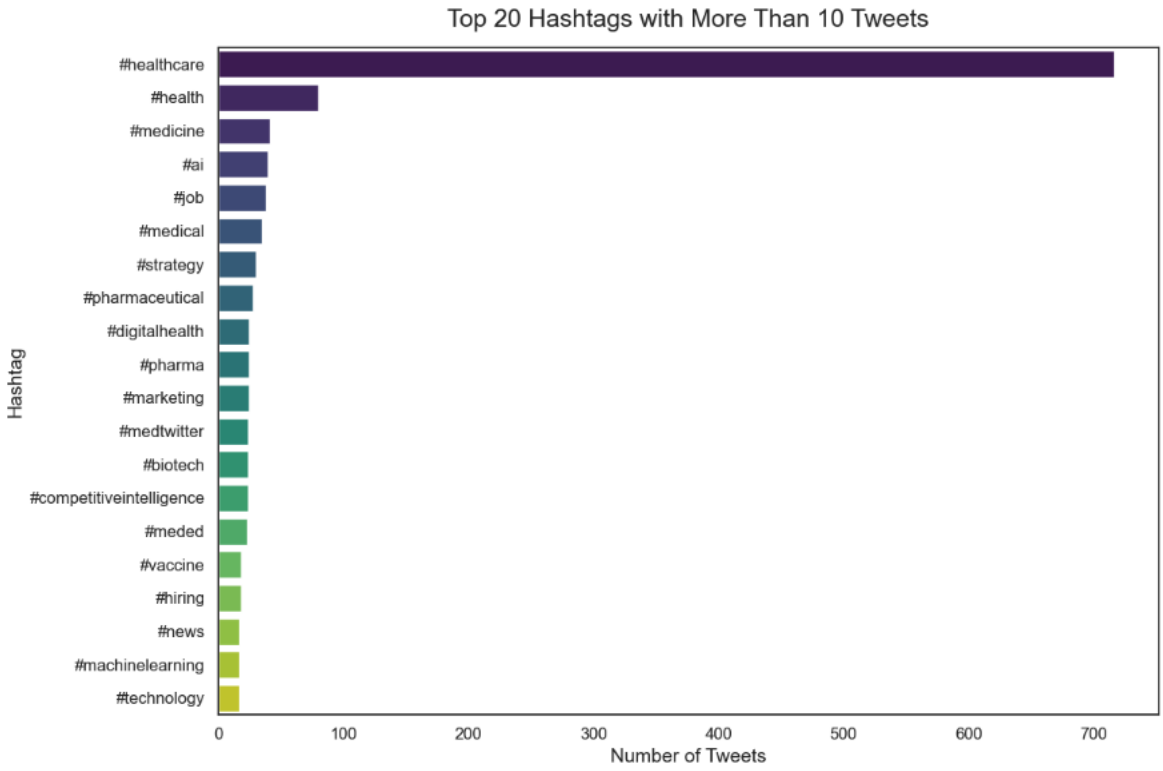

Figure 10: *Top 20 Healthcare-Related Hashtags with More Than 10 Tweets.*

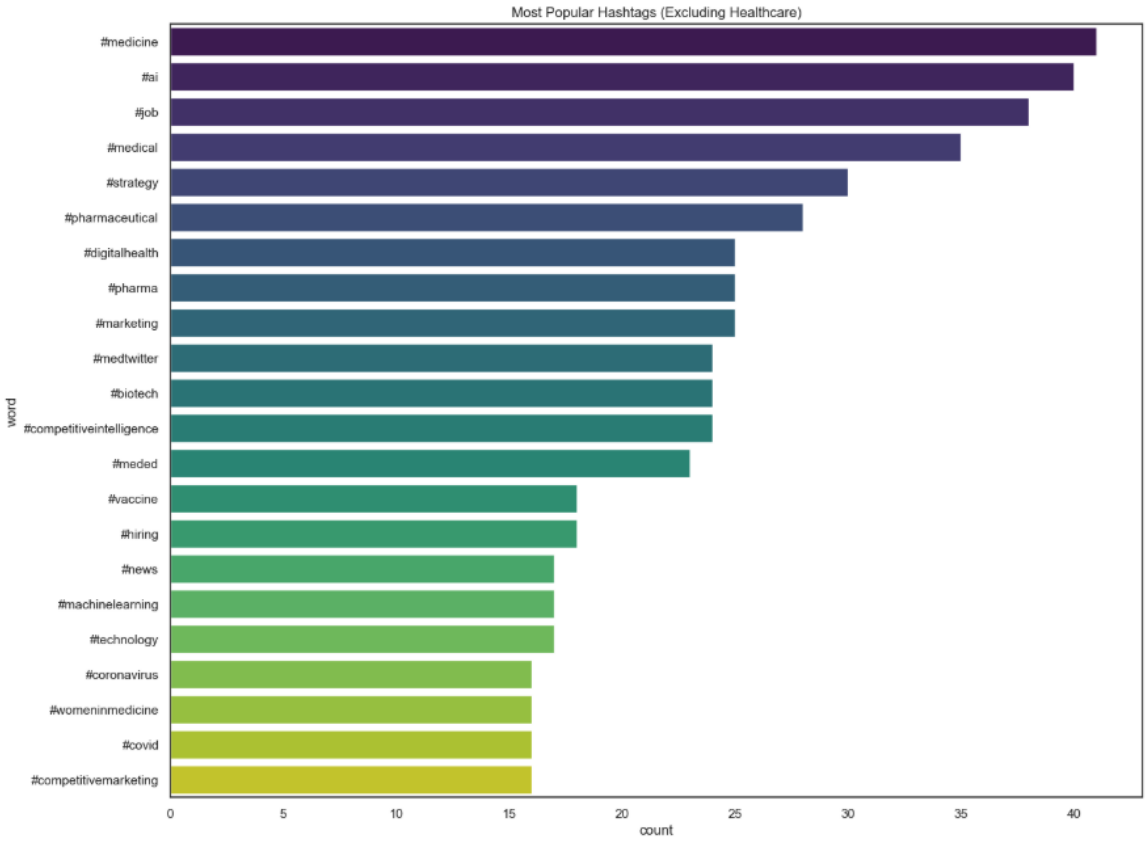

Figure 11: Popular Healthcare-Related Hashtags (Excluding #healthcare).

Implication: Social media provides valuable context on public concerns, complementing internal NHS data.

---

4. Patterns and Predictions

Identified Patterns

1. **Weekly Rhythm:** Mondays peak across all seasons; demand drops steadily during the week.
2. **Seasonal Surges:** Autumn and winter consistently push utilisation beyond capacity.
3. **GP Appointment Dominance:** Skews overall distribution; specialist services show more variability.
4. **Telehealth Growth:** A continuing trend with long-term planning implications.
5. **Missed Appointments:** Represent a significant, avoidable cost.

Reasons Behind Trends

- Seasonal illnesses drive autumn and winter peaks.
- Summer dips linked to holidays and reduced staffing.
- Telehealth expansion reflects post-pandemic adaptation.
- DNAs often due to accessibility barriers, not negligence (as emphasised by BMA).

Predictions

- **Telehealth** demand will expand further, requiring robust infrastructure.
- **Autumn peaks** will persist, needing proactive resourcing.
- **Interventions** like SMS reminders, improved booking systems, and education could reduce DNAs.

---

5. Recommendations

1. **Expand Flexible Capacity in Autumn/Winter:**
   Scale up staffing and appointment slots to match predictable seasonal peaks.
2. **Promote Telehealth Services:**
   Invest in digital infrastructure to absorb excess demand and increase patient access.
3. **Leverage Social Media Monitoring:**
   Use hashtag/sentiment analysis (#mentalhealth, #GPcrisis) to anticipate patient concerns and adapt communications.
4. **Reallocate Underutilised Resources:**
   During off-peak months, redirect resources toward preventive care and outreach to balance utilisation.
5. **Improve Data Quality:**
   Address ‘Unmapped’ service settings for more reliable insights and effective planning.

6. **Reduce Missed Appointments:**
   Implement targeted communication strategies (SMS reminders, user-friendly booking, awareness campaigns).