MAPPING THE MODERN NHS

Operational Intelligence & Emerging Trends

AUTHOR: DHRUVAL PATEL

25 MAY 2025

Table of Contents

The Business Scenario 2
Analytical Approach 2
2.1 Importing and Exploring Data 2
2.2 Cleaning and Processing 2
2.3 Aggregating and Preparing Data for Visualisation 3
2.4 Metric Creation 3
2.5 Combining Aggregated Data 3
Visualisation and Insights 4
3.1 Service Utilisation 4
3.2 Trends in Appointments Across Seasons 5
3.3 Attendance of Appointments 6
3.4 Monthly Appointments by HCP Types 7
3.5 Trending Hashtags (Excluding Over-Represented Hashtags) 8
Patterns And Predictions 9
4.1 Operational Capacity 9
4.2 Attendance Rates 9
4.3 Digital Transformation 9
Appendix 10

The Business Scenario

The NHS, a healthcare system in the United Kingdom, is facing increasing pressure to meet rising patient demand. With a growing population, strategic planning is required to ensure infrastructure and staffing are sufficient to deliver timely care. The stakeholders are conflicted whether to expand capacity, or improve utilisation of existing resources. Additionally, missed appointments continue to cause avoidable financial and operational burden. This analysis explores internal appointment and staffing data, alongside external sentiment platforms such as Twitter, to assess current service utilisation and inform practical, data-driven recommendations for future resource planning.

Analytical Approach

To address the NHS' key concerns about service utilisation, staffing adequacy, and missed appointments, I conducted a structured analytic approach using internal service data, and external social media insights. The process involved importing, wrangling, combining and visualising datasets to produce meaningful insights.

2.1 Importing and Exploring Data

Three internal datasets were imported using the Python library *pandas*, covering appointment volumes, attendance status, healthcare professional (HCP) roles, and service settings. External data from the social media platform Twitter (X) was incorporated to establish public sentiment and identify prevalent health-related hashtags. Figure 1 of the appendix shows the import process. An initial exploratory assessment determined the structural attributes of each dataset, including variable names, missing values, and data consistency.

2.2 Cleaning and Processing

Data validation ensured reliable analysis of the three NHS datasets. This involved addressing missing values, reviewing metadata, and standardising format – such as converting appointment dates to datetime with *pandas*. Figure 2 shows no missing values were found. Correct date formatting enabled accurate monthly grouping. Duplicate checks prevented skewed results. These steps ensured data accuracy and supported meaningful insights into appointment attendance, resource utilisation and service patterns.

2.3 Aggregating and Preparing Data for Visualisation

Aggregated data allows for grouping of key data necessary for analysis.
This meant creating variables to focus on areas of interest such as:

- Most popular service setting
- Month with the highest number of appointments
- Total appointments for each service setting

Figure 3 shows the code and result for grouping monthly appointment counts.

2.4 Metric Creation

New metrics were created to deepen understanding of service utilisation, and public sentiment. For example, a utilisation percentage enabled evaluation of capacity across the network, while attendance rates helped identify trends in missed appointments. Figure 4 shows the code and tables for these metrics (utilisation %).

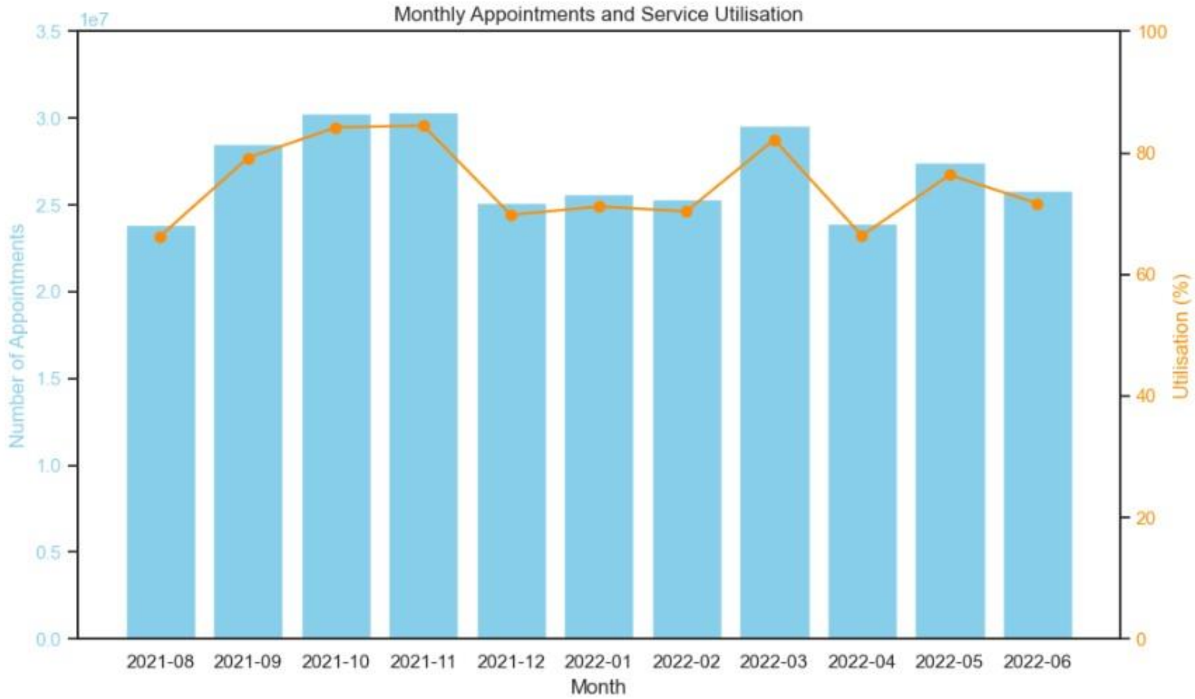
2.5 Combining Aggregated Data

Aggregating datasets using concat in *pandas* allowed combination of seasonal appointment data into a single visualisation (see Figure 5). This approach facilitated direct comparison across seasons, helping stakeholders identify periods of high demand more clearly, rather than viewing separate graphs for each season.

Visualisation and Insights

To support NHS stakeholders, determine the utilisation of resources and capacity in the network, visualisations were selected based on clarity, relevance and their ability to display trends and relationships aligned with the business objective. Each chart type maximises interpretability for non-technical audiences and drives data-driven decision making.

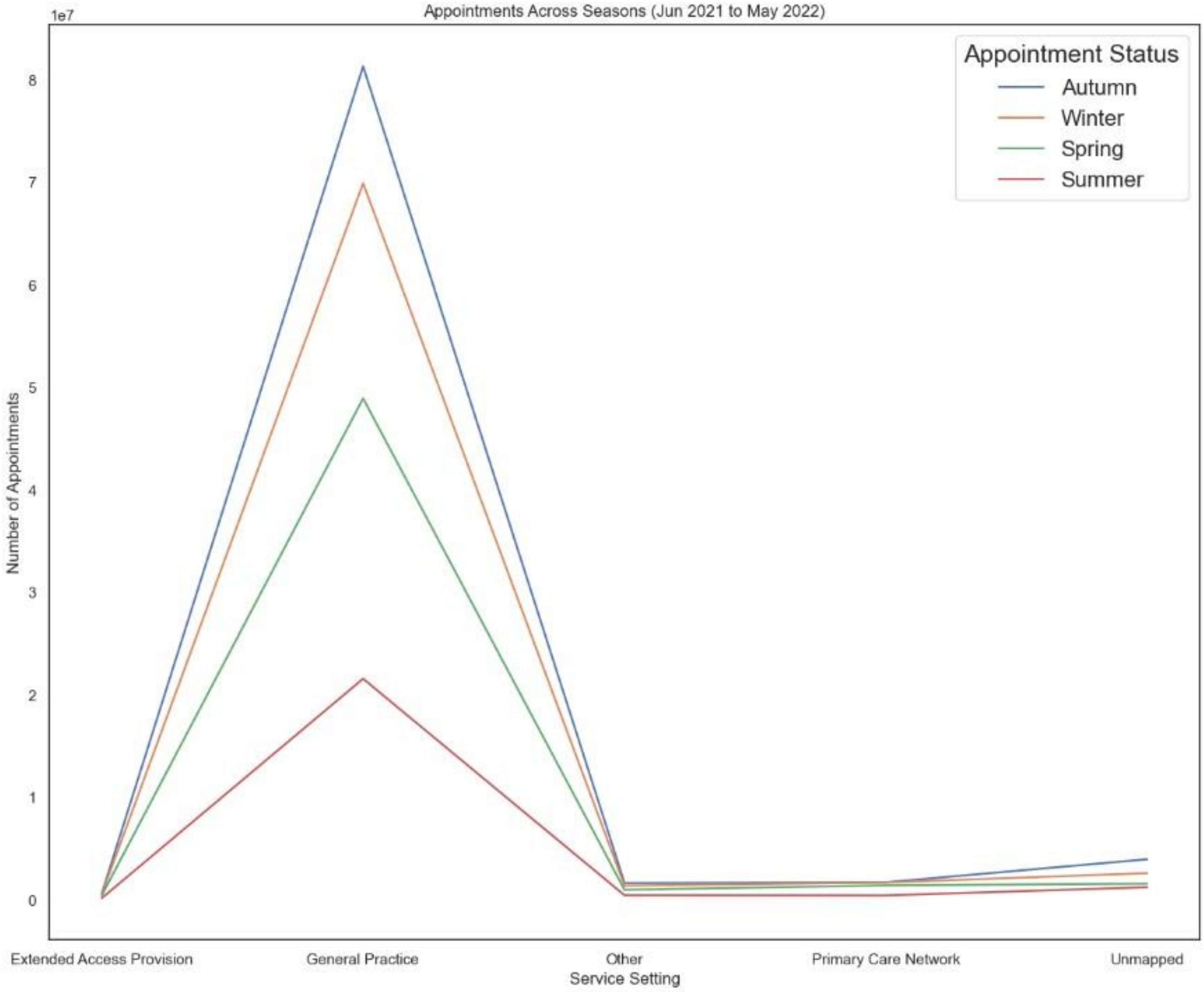
3.1 Service Utilisation



- Dual-axis chart comparing appointment volume and service utilisation monthly.
- Multiple months exceeding 80% operational capacity indicating potential strain.
- Opportunity to redistribute capacity or prepare for demand fluctuations.

A dual-axis chart enables clear comparison between service demand and operational capacity, helping stakeholders assess resource utilisation over time. By visualisation both metrics together, the chart highlights periods of potential strain and supports evaluation of whether staffing and resources are adequate. Focussing on post-COVID data ensures that planning decisions are informed by recent operational trends.

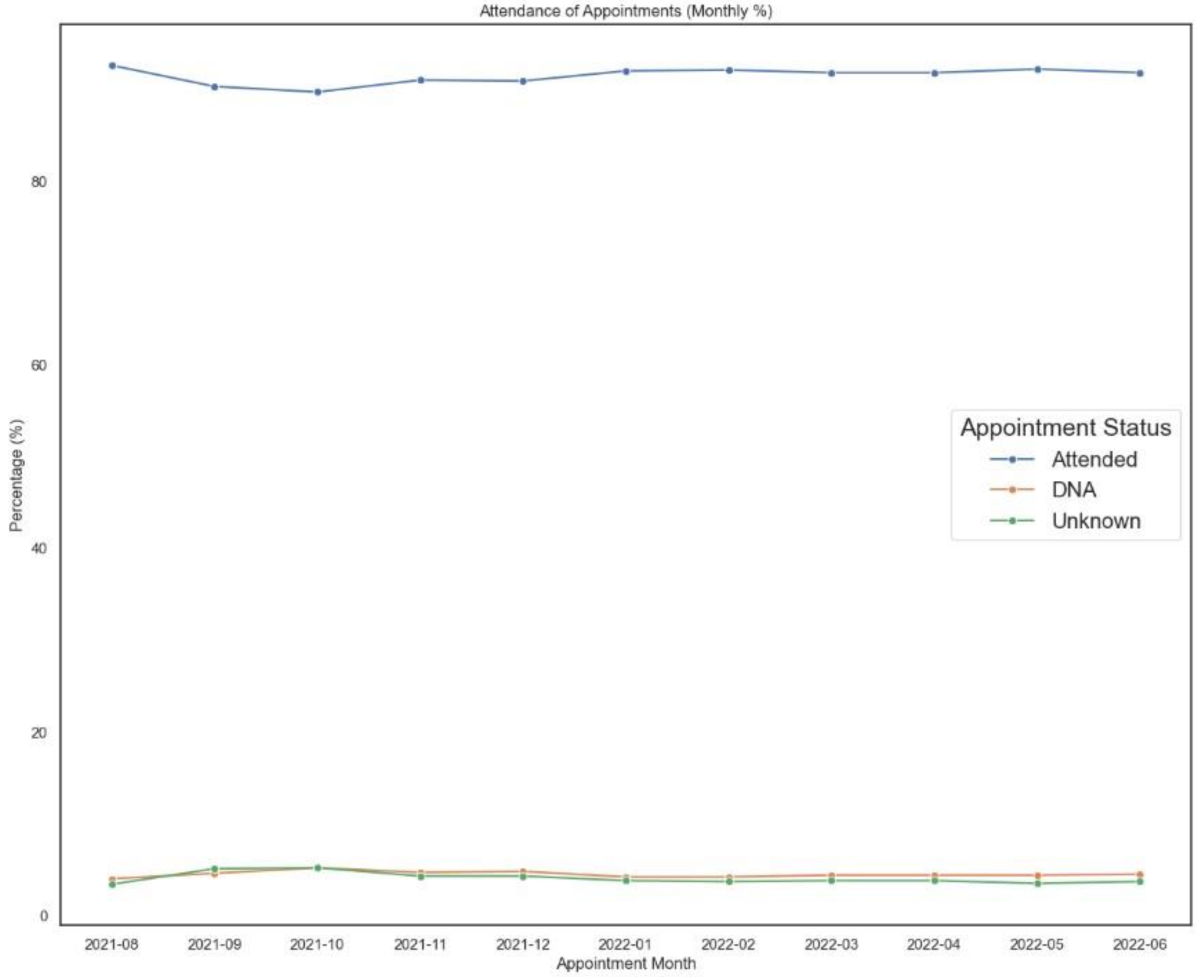
3.2 Trends in Appointments Across Seasons



- Chart created by combining data from each season using *concat*
- Autumn followed by winter show the highest volume of appointments.
- GPs conducted majority of the appointments between Jun 2021 and May 2022.

Combining the seasons into one visualisation allow stakeholders to see direct comparisons in appointment volumes for each service setting, for each season, allowing the evaluation of high demand. Figure 5 shows the code used to achieve this.

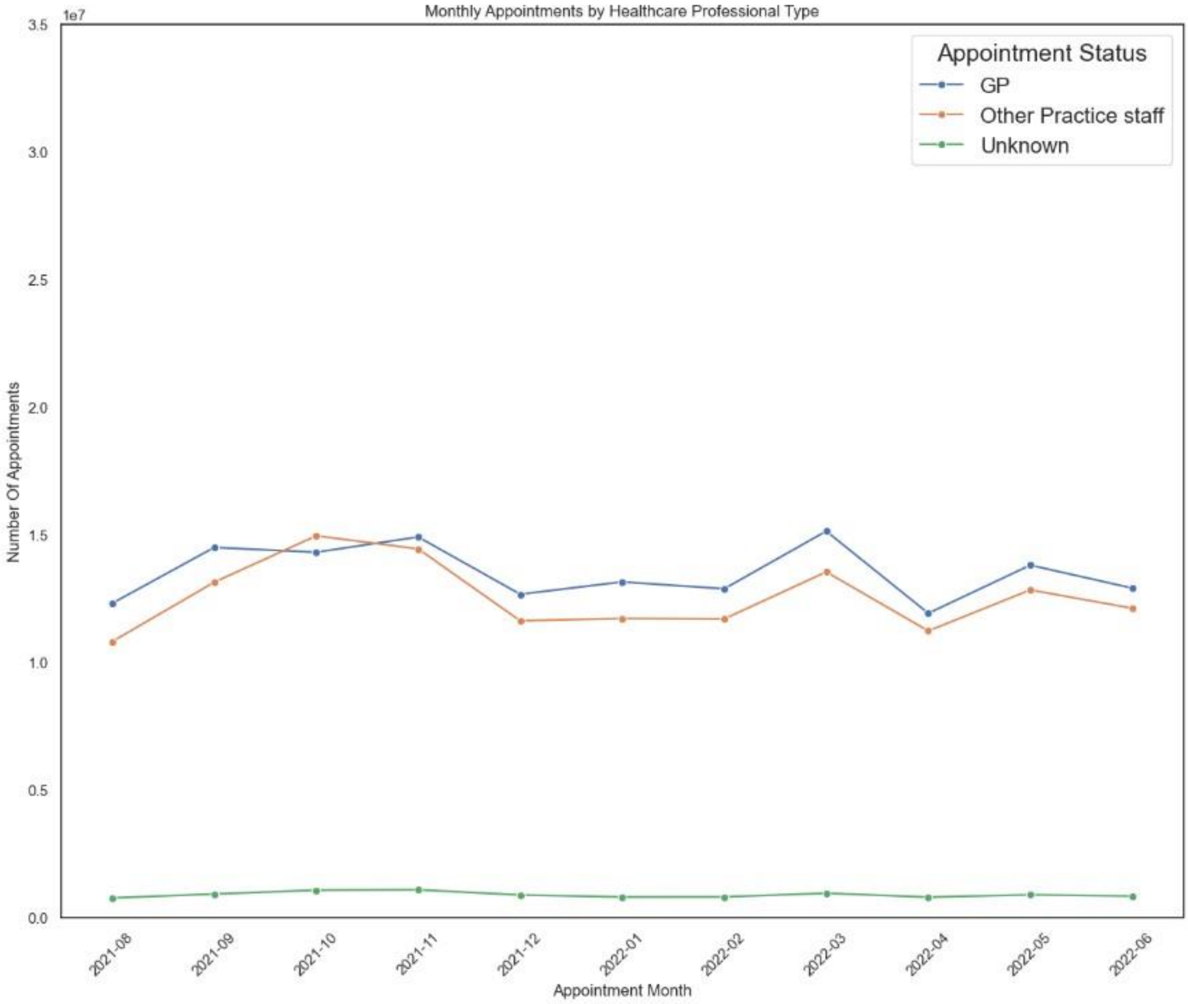
3.3 Attendance of Appointments



- Appointment status converted into a percentage rather than total
- No fluctuations in appointment statuses.
- Roughly 90% of booked appointments are attended.

By converting monthly attendance figures into percentages, this chart provides a view of attendance rate, regardless of appointment volumes. This means months with fewer appointments, are still accurately represented. Stakeholders are able to identify fluctuations in missed or unknown appointments over time, supporting efforts to reduce missed appointments. Figure 6 shows how this was achieved.

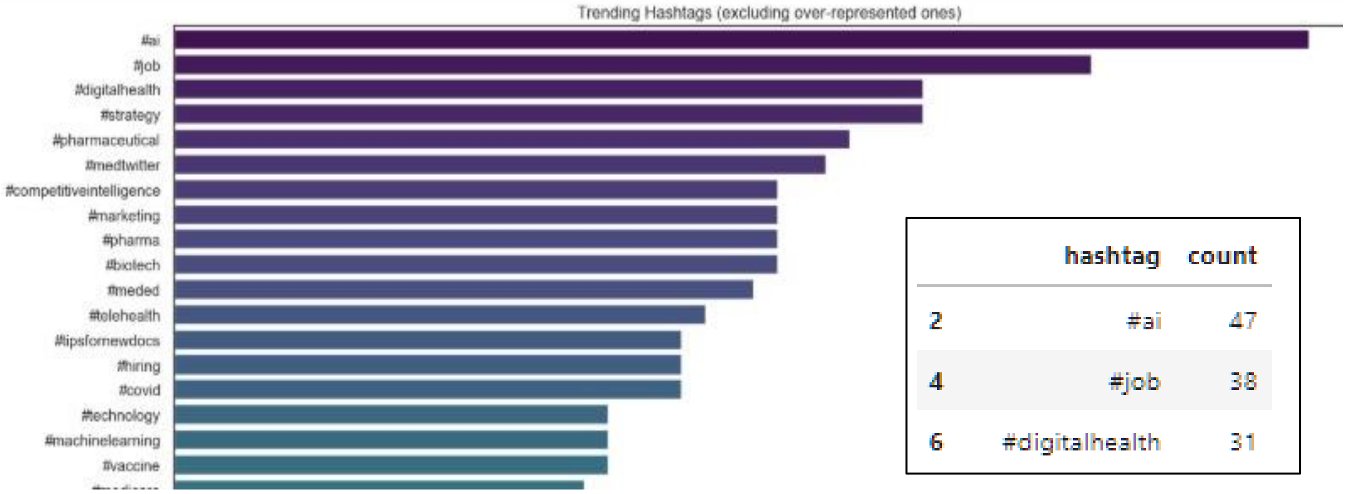
3.4 Monthly Appointments by HCP Types



- GPs consistently handle slightly more appointments than other practice staff.
- Seasonal fluctuations in GP and Practice Staff appointments.
- Unknown category remains near zero indicating analysis reliability.

By distinguishing between GP and other practice staff, stakeholders can determine whether additional staffing or resources may be needed. In addition, further derivatives of practice staff could directly pinpoint which practice staff are providing service across the network.

3.5 Trending Hashtags (Excluding Over-Represented Hashtags)



	hashtag	count
2	#ai	47
4	#job	38
6	#digitalhealth	31

- #ai, #job, #digitalhealth were the most prevalent hashtags
- Result has excluded over-represented hashtags such as #healthcare, #medical

Excluding over-represented hashtags allows users to focus on hashtags with more relevance. Based on our twitter dataset, we can see interest in ai, jobs and digital health. These pose areas of exploration which could provide benefit through employment, reduced pressure on service, and advancements in digital health.

Patterns And Predictions

The analysis aimed to support NHS stakeholders by providing insights into service demand, operational capacity, and workforce utilisation using post-COVID data. A series of visualisations highlighted key trends and performance indicators, enabling confident, data-driven decision making. External insights from Twitter were incorporated to identify areas of public interest.

4.1 Operational Capacity

Monthly service utilisation was found to frequently exceed 80%, indicating recurrent strain on capacity. With GPs handling the majority of appointments, there is an opportunity to shift workload to other qualified healthcare professionals. Seasonal analysis shows peaks in appointment volumes during autumn and winter, highlighting the need for additional staffing during periods of higher demand. The trends suggest the need to redistribute workload more evenly, or implement contingency planning.

4.2 Attendance Rates

Attendance rates remained stable, with around 90% of booked appointments attended. However, small but regular missed appointment rates permit targeted patient communication strategies. To address this, the NHS could use automated SMS and email reminders for appointments, provide cancellation/rescheduling options and opt to increase awareness of the impact of missed appointments on service delivery.

4.3 Digital Transformation

Twitter analysis revealed prevalent hashtags such as #ai, #job and #digitalheatlh, highlight key areas for further exploration. AI offers potential improvements to administrative efficiency, while job-related interest may reflect a workforce supply to support NHS demand during peak periods such as winter. Growing interest on digital health presents an opportunity to ease pressure on NHS services. Tools like remote consultations, symptom-checking apps, and digital triage improve access to care without requiring in-person appointments, enhancing patient care while reducing the burden on NHS facilities operating at high capacity.

Appendix

🧰 Prepare your workstation

```python
# Import the necessary libraries.
import pandas as pd
import numpy as np

# Ignore warnings.
import warnings
warnings.filterwarnings('ignore')
```

📥 Import The Datasets

```python
# Import the three datasets
ad = pd.read_csv('actual_duration.csv')
ar = pd.read_csv('appointments_regional.csv')
nc = pd.read_excel('national_categories.xlsx')
```

```python
# Import the 'tweets.csv' dataset
td = pd.read_csv('tweets.csv')
```

***Figure 1** – Importing internal and external datasets using pandas*

```python
# Check for missing values in ad
ad.isnull().sum()
```

```
sub_icb_location_code        0
sub_icb_location_ons_code    0
sub_icb_location_name        0
icb_ons_code                 0
region_ons_code              0
appointment_date             0
actual_duration              0
count_of_appointments        0
dtype: int64
```

```python
# Check for missing values in ar
ar.isnull().sum()
```

```
icb_ons_code                          0
appointment_month                     0
appointment_status                    0
hcp_type                              0
appointment_mode                      0
time_between_book_and_appointment     0
count_of_appointments                 0
dtype: int64
```

```python
# Check for missing values in nc
nc.isnull().sum()
```

```
appointment_date         0
icb_ons_code             0
sub_icb_location_name    0
service_setting          0
context_type             0
national_category        0
count_of_appointments    0
appointment_month        0
dtype: int64
```

***Figure 2** – Checking for missing values in the internal datasets*

```
monthly_counts = (
    nc.groupby([nc['appointment_date'].dt.year.rename('year'),
                nc['appointment_date'].dt.month.rename('month')])
               ['count_of_appointments']
    .sum()
    .sort_values(ascending = False)
)

# Step 2 - Show the results
print("Total appointments per month:")
print(monthly_counts.head())
```

```
Total appointments per month:
year  month
2021  11       30405070
      10       30303834
2022  3        29595038
2021  9        28522501
2022  5        27495508
```

***Figure 3** – Creating a variable to aggregate data to show total number of appointments per month in a DataFrame used for capacity visualisations.*

```
# Add a utilisation column to the ar_df DataFrame
ar_df['daily_appts'] = ar_df['count_of_appointments'] / 30
ar_df['utilisation'] = (ar_df['daily_appts'] / 1200000) * 100
ar_df['utilisation'] = ar_df['utilisation'].round(1)

# View the DataFrame.
ar_df.head()
```

	appointment_month	count_of_appointments	daily_appts	utilisation
0	2021-08	23852171	7.950724e+05	66.3
1	2021-09	28522501	9.507500e+05	79.2
2	2021-10	30303834	1.010128e+06	84.2
3	2021-11	30405070	1.013502e+06	84.5
4	2021-12	25140776	8.380259e+05	69.8

***Figure 4** – Adding of a utilisation column to an existing DataFrame to create meaningful analysis relevant to the considered maximum number of appointments per day (1,200,000)*

```python
# Add a 'Season' column to each DataFrame
nc_summer_appts['Season'] = 'Summer'
nc_autumn_appts['Season'] = 'Autumn'
nc_winter_appts['Season'] = 'Winter'
nc_spring_appts['Season'] = 'Spring'

# Combine all the DataFrames into one
all_seasons_df = pd.concat([
    nc_summer_appts,
    nc_autumn_appts,
    nc_winter_appts,
    nc_spring_appts
])
```

***Figure 5** – By adding a 'Season' column to each of our aggregated DataFrames, we are able to visualise all seasons on a single graph by stating the hue as Season. Using the concat function, we can combine the four DataFrames into one for visualisations as shown in '3.2 – Trends in Appointments Across Seasons'.*

```python
# Calculate the monthly total appointments
status_monthly_totals = (
    status_monthly.groupby('appointment_month')['count_of_appointments'].transform('sum')
)

# Calculate percentages
status_monthly['percentage'] = (status_monthly['count_of_appointments'] / status_monthly_totals) * 100
status_monthly['percentage'] = status_monthly['percentage'].round(1)
```

***Figure 6** – A percentage value for the number of appointments attended, did not attend and unknown provides a stronger comparison if appointments were attended. Fluctuations in appointment volumes will not cause an effect on the percentage attended.*